VOL. 17 · NUMBER 1
TOM НОМЕР

ACADEMY OF SCIENCES OF THE USSR
HUNGARIAN ACADEMY OF SCIENCES
CZECHOSLOVAK ACADEMY OF SCIENCES

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

АКАДЕМИЯ НАУК СССР
ВЕНГЕРСКАЯ АКАДЕМИЯ НАУК
ЧЕХОСЛОВАЦКАЯ АКАДЕМИЯ НАУК

1988

AKADÉMIAI KIADÓ, BUDAPEST
DISTRIBUTED OUTSIDE THE COMECON-COUNTRIES
BY PERGAMON PRESS, OXFORD

PROBLEMS OF CONTROL AND INFORMATION THEORY

An international bi-monthly sponsored jointly by the Presidium of the Academy of Sciences of the USSR, of the Hungarian Academy of Sciences and of the Czechoslovak Academy of Sciences. The six issues published per year make up a volume of some 480 pp. It offers publicity for original papers and short communication of the following topics:

— theory of control processes
— theory of adaptive systems
— theory of estimation and identification
— theory of controlling robot-technologic and flexible manufacturing systems
— information theory
— information-theoretic aspects of multiple access networks.

While this bi-monthly is mainly a publication forum of the research results achieved in the socialist countries, also papers of international interest from other countries are welcome.

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

Международный журнал Академии наук СССР, Венгерской Академии наук и Чехословацкой Академии наук выходит 6 раз в год общим объемом 480 печатных страниц.

В журнале публикуются оригинальные научные статьи и статьи обзорного характера по следующим проблемам управления и теории информации:

— теория процессов управления;
— теория адаптивных систем;
— теория оценивания и идентификации;
— теория управления робототехническими и гибкими производственными системами;
— теория информации;
— теория информации в области сетей с множественным доступом.

Целью журнала является ознакомление научной общественности различных стран с важнейшими проблемами, имеющими актуальный и перспективный характер, научными достижениями ученых социалистических и других стран.

Distributors

For the Soviet Union:

SOYUZPECHATY, Moscow 123 308 USSR

For Albania, Bulgaria, China, Cuba, Czechoslovakia, German Democratic Republic, Korean People's Republic, Mongolia, Poland, Rumania, Vietnam and Yugoslavia:

KULTURA Hungarian Foreign Trading Co.
P. O. Box 149 H-1389 Budapest, Hungary

For all other countries:

PERGAMON JOURNALS LTD, Headington Hill Hall, Oxford OX3 OBW, England

or

PERGAMON JOURNALS INC, Maxwell House, Fairview Park, Elmsford, NY 10523, USA

1988 Subscription Rate DM 535,— per annum including postage and insurance.

PROBLEMS OF CONTROL AND INFORMATION THEORY, VOL. 17 (1988)

SUBJECT INDEX

PROBLEMS OF CONTROL AND INFORMATION THEORY, VOL. 17 (1988)

AUTHOR INDEX

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

AKADÉMIAI KIADÓ

PUBLISHING HOUSE OF THE HUNGARIAN ACADEMY OF SCIENCES

BUDAPEST

Problems of Control and Information Theory, Vol. 17 (1), pp. 3—13 (1988)

TANDEM QUEUE WITH FINITE INTERMEDIATE WAITING ROOM AND BLOCKING IN HEAVY TRAFFIC

YA. A. KOGAN, A. A. PUKHALSKY

(*Moscow*)

(Received July 17, 1986)

A two-stage exponential network with an unlimited first-stage queue and a finite waiting room in the second stage are investigated. When the waiting room is full, the first-stage server is blocked and ceases to offer service. A request served by the second-stage server may return to the first-stage queue with a positive probability or leave the network. A limiting distribution of the steady-state queue lengths in heavy traffic is obtained explicitly. The queues are shown to be asymptotically independent. The first-stage queue behaves as an $M/G/1$ queue in heavy traffic, and the second-stage queue as an $M/M/1/m$ queue.

Introduction

We consider a two-stage tandem network shown in Fig. 1. The arrival of requests into the system is modeled by a Poisson process (of rate λ) and the service times are exponentially distributed (with rates α and β, respectively). Requests are served in each stage according to the order of their arrival. No more than m requests are allowed in the second stage (including that one in service). So, the first stage server offers service only in the case when the number of requests queued or in service in the second stage is less than m, otherwise the server is blocked and ceases to offer service. An infinite queue is allowed in the first stage. A request upon completion of its service at the second-stage server returns to the first-stage server for further service or it departs from the network with probabilities $(1-\theta)$ and θ, respectively.

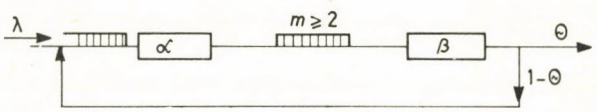


Fig. 1. A two-stage tandem network with a finite intermediate waiting room

Various computer related examples, e.g. the concentrator-processor model, are described by this network (see [1]). In [1, 2] the ergodicity conditions were given and algorithms to find the stationary state probabilities numerically were developed. Our

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

aim is to derive the heavy traffic limit distribution of the stationary queue lengths in the first and the second stage. This limiting distribution is obtained in an explicit form and is more suitable for applications. We show that the queue lengths are asymptotically independent. The first stage queue behaves as an $M/GI/1$ queue and the second stage queue behaves as an $M/M/1/m$ queue. This fact can be considered as an extension of Jackson's theorem on the product form representation of the stationary distribution [3] on networks with finite population constraints.

1. Heavy-traffic limit theorem

Following [1], we introduce polynomials $D_j(z), j \geqq 0$, which play a major role in the studying of the steady-state characteristics of the system in Fig. 1. We define

$$\begin{aligned} a(z) &= \lambda(1-z)+\alpha+\beta, \\ b(z) &= \beta((1-\theta)z+\theta), \\ D_0(z) &= \lambda, \\ D_1(z) &= \frac{1}{\alpha}(\lambda z a(z) - \alpha\beta\theta), \\ D_j(z) &= \frac{z}{\alpha}(a(z)D_{j-1}(z) - b(z)D_{j-2}(z)), \quad j \geqq 2. \end{aligned} \tag{1.1}$$

Also we denote

$$\gamma = \frac{\beta}{\alpha} \tag{1.2}$$

As shown in [1, 2] the necessary and sufficient condition for the stationary queue lengths distribution to exist is $D_m(1) < 0$, where $D_m(1)$ has the form

$$D_m(1) = \begin{cases} (m+1)\lambda - m\beta\theta, & \gamma = 1 \\ \dfrac{1-\gamma^{m+1}}{1-\gamma}\lambda - \dfrac{1-\gamma^m}{1-\gamma}\beta\theta, & \gamma \neq 1. \end{cases} \tag{1.3}$$

Therefore, the heavy traffic condition is $D_m(1) \uparrow 0$.

To state the limit theorem, we assume that the parameters of the network depend on a variable $\tau > 0$: we denote the input rate by $\lambda(\tau)$, the service rates by $\alpha(\tau)$ and $\beta(\tau)$, respectively, the departure probability by $\theta(\tau)$. The functions $\lambda(\tau)$, $\alpha(\tau)$, $\beta(\tau)$ and $\theta(\tau)$ are

assumed to be continuous for $\tau > 0$ and have limits on the right at $\tau = 0$ which are λ, α, β and θ, respectively:

$$\lim_{\tau \downarrow 0} \lambda(\tau) = \lambda, \qquad \lim_{\tau \downarrow 0} \alpha(\tau) = \alpha,$$
$$\lim_{\tau \downarrow 0} \beta(\tau) = \beta, \qquad \lim_{\tau \downarrow 0} \theta(\tau) = \theta. \tag{1.4}$$

For every $\tau > 0$ similarly to (1.1) and (1.2), we introduce polynomials $a(z, \tau)$, $b(z, \tau)$, $D_j(z, \tau)$, $j \geqq 0$, and parameter $\gamma(\tau)$. Since it is obvious in view of (1.4) that

$$D_m(1) = \lim_{\tau \downarrow 0} D_m(1, \tau), \tag{1.5}$$

then the heavy traffic condition has the form

$$D_m(1, \tau) \uparrow 0, \quad \tau \downarrow 0. \tag{1.6}$$

By $(Q_1(\tau), Q_2(\tau))$, $\tau > 0$, we denote a stationary queue lengths vector which corresponds to a specific value of τ. We denote also

$$D'_m(1) = \left. \frac{dD_m(z)}{dz} \right|_{z=1}, \tag{1.7}$$

$$p_j = \gamma^{m-j} \left(\sum_{i=0}^{m} \gamma^i \right)^{-1}, \quad 0 \leqq j \leqq m. \tag{1.8}$$

It is obvious that (p_j), $0 \leqq j \leqq m$, is the stationary queue length distribution for the $M/M/1/m$ queue that has the input rate α and the service rate β.

Theorem. Under the condition (1.6) we have

(a) $D'_m(1) > 0$,

(b) $\lim_{\tau \downarrow 0} P(-D_m(1, \tau) Q_1(\tau) \leqq x, \quad Q_2(\tau) = j) =$
$= (1 - \exp(-x/D'_m(1))) p_j, \quad x \geqq 0, \quad j = 0, 1, \ldots, m,$

(c) $\lim_{\tau \downarrow 0} (-D_m(1, \tau)) EQ_1(\tau) = D'_m(1)$.

Remark. Assertion (b) of the Theorem contains the result announced in the Introduction: the random variables $Q_1(\tau)$ and $Q_2(\tau)$ are asymptotically independent as $\tau \downarrow 0$, the limiting distribution for $Q_2(\tau)$ coinciding with that of for the stationary $M/M/1/m$ queue which has the input rate α and the service rate β, and the normalized queue $Q_1(\tau)$ being distributed asymptotically exponentially. Later on we shall indicate an $M/G/1$ queue with the input rate $\lambda(\tau)$ and with a τ-dependent service time distribution which is equivalent asymptotically to $Q_1(\tau)$.

Proof of (a). To save space and time, we set $\lambda = 1$, so that we could exploit the results of [1] (note that θ in our notation is $(1 - \theta)$ in [1]).

In view of (1.5) and (1.6), we have

$$D_m(1)=0. \tag{1.9}$$

Then using the results of [1], we see that $z=1$ is the m-th zero of $D_m(z)$ in the order of increasing (we take into account the order of each zero). Indeed, it follows from (1.3) that equation $D_m(1)=0$ implies that the inequality $\min(\alpha, \beta)\cdot\theta>1$ holds. Hence, by Lemma 2 (2, 3) from [1], the polynomial $D_m(z)$ has no more than m zeros in the interval $[0, 1]$. If the number of the zeros is less than m, then as one can see from the proof of Theorem 1 in [1], there exists a stationary queue lengths distribution for the network with parameters λ, α, β and θ. This contradicts (1.9).

Thus, $z=1$ is the zero of $D_m(z)$ nearest on the left to the point $\min(\alpha, \beta)\theta$ (see Lemma 2 (3) in [1]). Since $D_m(\min(\alpha, \beta)\cdot\theta)>0$ (see Lemma 1 (3, 4) in [1]) and the zero $z=1$ is simple (Lemma 2 (3) in [1]), then $D'_m(1)>0$ which proves (a).

Proof of (b). Since $D_m(1, \tau)<0$ for $\tau>0$ then for such τ there exists a stationary queue lengths distribution for the network with the parameters $\lambda(\tau)$, $d(\tau)$, $\beta(\tau)$ and $\theta(\tau)$. We denote for $\tau>0$

$$p_{kj}(\tau)=P(Q_1(\tau)=k, Q_2(\tau)=j), \quad k=0, 1, \ldots, \quad j=0, 1, \ldots, m, \tag{1.10}$$

We introduce the generating functions

$$P_j(z, \tau)=\sum_{k=0}^{\infty} p_{kj}(\tau)z^k, \quad j=0, 1, \ldots, m. \tag{1.11}$$

It is shown in [1] that for $\tau>0$

$$P_m(z, \tau)=-\sum_{k=0}^{m-1} p_{0k}(\tau)D_k(z, \tau)/D_m(z, \tau), \tag{1.12}$$

$$P_j(z, \tau)=A_{m-j}(z, \tau)P_m(z, \tau)+$$

$$+\sum_{k=0}^{m-j-1} A_k(z, \tau)p_{0, j+k}(\tau), \quad j=0, 1, \ldots, m-1, \tag{1.13}$$

where $A_j(z, \tau)$ are defined recursively

$$A_0(z, \tau)=1, \quad A_1(z, \tau)=\frac{z}{\alpha(\tau)}(a(z, \tau)-\alpha(\tau)),$$

$$A_j(z, \tau)=\frac{z}{\alpha(\tau)}(a(z, \tau)A_{j-1}(z, \tau)-b(z, \tau)A_{j-2}(z, \tau)), \quad j\geqq 2. \tag{1.14}$$

Lemma. The functions $p_{0j}(\tau)$, $j=0, 1, \ldots, m-1$ are continuous in τ for $\tau>0$ and right-continuous at $\tau=0$, if we set $p_{0j}(0)=0$.

The proof is given in the Appendix.

Denote for $\tau>0$

$$S_m(z,\tau)=-\sum_{k=0}^{m-1} p_{0k}(\tau)D_k(z,\tau), \tag{1.15}$$

$$S_j(z,\tau)=A_{m-j}(z,\tau)S_m(z,\tau)+$$

$$+D_m(z,\tau)\sum_{k=0}^{m-j-1} A_k(z,\tau)p_{0,k+j}(\tau), \quad j=0,1,\ldots,m-1, \tag{1.16}$$

$$S(z,\tau)=\sum_{k=0}^{m} S_k(z,\tau), \tag{1.17}$$

$$P(z,\tau)=\sum_{k=0}^{m} P_k(z,\tau). \tag{1.18}$$

We define $S(z,\tau)$ and $D_m(z,\tau)$ at $\tau=0$ by setting

$$S(z,0)=0, \quad D_m(z,0)=D_m(z). \tag{1.19}$$

Then the functions $S(z,\tau)$ and $D_m(z,\tau)$ with all their partial derivatives with respect to z are continuous in (z,τ) for $z\in C$, $\tau\geqq 0$ (at $\tau=0$ the limit on the right is implied). That follows from the definitions of $S(z,\tau)$ and $D_m(z,\tau)$, properties of $\lambda(\tau)$, $\alpha(\tau)$, $\beta(\tau)$, and $\theta(\tau)$ and the assertion of the Lemma. Therefore, we have the following expansions for $\tau>0$:

$$S(z,\tau)=S(1,\tau)+(z-1)\left.\frac{\partial S(z,\tau)}{\partial z}\right|_{z=1}+C_1(z,\tau)(z-1)^2, \tag{1.20}$$

$$D_m(z,\tau)=D_m(1,\tau)+(z-1)\left.\frac{\partial D_m(z,\tau)}{\partial z}\right|_{z=1}+C_2(z,\tau)(z-1)^2, \tag{1.21}$$

with the functions $C_1(z,\tau)$ and $C_2(z,\tau)$ bounded for $|z-1|<\varepsilon$, $0<\tau<\delta$.

We note (see (1.12), (1.13) and (1.15)–(1.18)) that for $\tau>0$ holds

$$P(z,\tau)=S(z,\tau)/D_m(z,\tau) \tag{1.22}$$

so that $S(1,\tau)=D_m(1,\tau)$ (since $P(1,\tau)=1$, $\tau>0$, in view of (1.11) and (1.18)) and (1.19)–(1.21) result in

$$\lim_{\substack{z\to 1\\ \tau\downarrow 0}}\frac{S(z,\tau)-D_m(1,\tau)}{z-1}=0,$$

$$\lim_{\substack{z\to 1\\ \tau\downarrow 0}}\frac{D_m(z,\tau)-D_m(1,\tau)-(z-1)D_m'(1)}{z-1}=0.$$

Substituting into the latter relations $z=\exp(-itD_m(1,\tau))$ we have in view of the fact that $D_m(1,\tau)\uparrow 0$, $\tau\downarrow 0$, that

$$\lim_{\tau\downarrow 0}\frac{S(\exp(-itD_m(1,\tau)),\tau)}{D_m(1,\tau)}=1,$$

$$\lim_{\tau\downarrow 0}\frac{D_m(\exp(-itD_m(1,\tau)),\tau)}{D_m(1,\tau)}=1-itD'_m(1),$$

and then making use of (1.22), we derive that

$$\lim_{\tau\downarrow 0}P(\exp(-itD_m(1,\tau)),\tau)=\frac{1}{1-itD'_m(1)}. \tag{1.23}$$

Since $P(1,\tau)=1$, $\tau>0$, then (1.13), (1.18) and the Lemma imply that

$$\liminf_{\substack{z\to 1\\ \tau\downarrow 0}}|P_m(z,\tau)|>0$$

and then, from (1.13) we get

$$\lim_{\substack{z\to 1\\ \tau\downarrow 0}}\frac{P_j(z,\tau)}{P_m(z,\tau)}=\lim_{\tau\downarrow 0}A_{m-j}(1,\tau)=\gamma^{m-j},\quad j=0,1,\ldots,m,$$

(the latter equality can easily be obtained from (1.14)). So, using (1.18) and (1.23) we finally have

$$\lim_{\tau\downarrow 0}E(\exp(-itD_m(1,\tau)Q_1(\tau))I(Q_2(\tau)=j))=$$

$$=\lim_{\tau\downarrow 0}P_j(\exp(-itD_m(1,\tau)),\ \tau)=$$

$$=p_j/(1-itD'_m(1)),\quad j=0,1,\ldots,m.$$

This proves (b) because $(1-itv)^{-1}$ is the characteristic function of the exponential distribution with the mean $1/v$. To prove (b) we write, making use of (1.22)

$$\left.\frac{\partial S(z,\tau)}{\partial z}\right|_{(1,\tau)}=D_m(1,\tau)\left.\frac{\partial P(z,\tau)}{\partial z}\right|_{(1,\tau)}+P(1,\tau)\left.\frac{\partial D_m(z,\tau)}{\partial z}\right|_{(1,\tau)},\quad \tau>0$$

and observe that

$$\left.\frac{\partial P(z,\tau)}{\partial z}\right|_{(1,\tau)} = EQ_1(\tau), \quad P(1,\tau)=1, \quad \tau>0,$$

$$\lim_{\tau\downarrow 0}\left.\frac{\partial S(z,\tau)}{\partial z}\right|_{(1,\tau)}=0, \quad \lim_{\tau\downarrow 0}\left.\frac{\partial D_m(z,\tau)}{\partial z}\right|_{(1,\tau)}=D'_m(1).$$

This completes the proof.

2. External queue length distribution

In this section we give an explicit formula for $D'_m(1)$ and we clarify the essence of assertions (b) and (c) of the Theorem.

From definitions (1.1) and (1.2) we easily obtain for $j\geqq 2$

$$D'_j(1)-D'_{j-1}(1)=\gamma(D'_{j-1}(1)-D'_{j-2}(1))+$$

$$+D_j(1)-\frac{1}{\alpha}(\lambda D_{j-1}(1)+\beta(1-\theta)D_{j-2}(1)).$$

This results in the equality ($m\geqq 2$)

$$D'_m(1)=D'_1(1)\sum_{j=0}^{m-1}\gamma^j+\sum_{k=2}^{m}\left(\sum_{j=0}^{m-k}\gamma^j\right)D_k(1)-$$

$$-\frac{\lambda}{\alpha}\sum_{k=1}^{m-1}\left(\sum_{j=0}^{m-k-1}\gamma^j\right)D_k(1)-\frac{\beta(1-\theta)}{\alpha}\sum_{k=0}^{m-2}\left(\sum_{j=0}^{m-k-2}\gamma^j\right)D_k(1).$$

Using (1.3) and observing that

$$D'_1(1)=\lambda(1+\gamma)-\frac{\lambda^2}{\alpha}$$

we reduce the expression for $D'_m(1)$ to the form

$$D'_m(1)=\begin{cases} m\left(1-\dfrac{m+2}{6(m+1)}\theta\right)\beta\theta, & \gamma=1, \\ \left[1-\gamma^m+\dfrac{m(\gamma^{m+2}-1)+(m+2)\gamma(\gamma^m-1)}{(1-\gamma^{m+1})^2}\theta\gamma^m\right]\dfrac{\beta\theta}{1-\gamma}, & \gamma\neq 1. \end{cases} \tag{2.1}$$

We denote

$$\mu = \begin{cases} \dfrac{m}{m+1}\beta\theta, & \gamma = 1, \\ \dfrac{1-\gamma^m}{1-\gamma^{m+1}}\beta\theta, & \gamma \neq 1, \end{cases} \tag{2.2}$$

$$\delta = \begin{cases} \dfrac{m}{m+1}\left(1 - \dfrac{m+2}{3(m+1)}\theta\right)\beta\theta, & \gamma = 1, \\ 1-\gamma^m - 2\dfrac{m(1-\gamma^{m+2}) - (m+2)\gamma(1-\gamma^m)}{(1-\gamma^{m+1})^2}\theta\gamma^m \Big] \dfrac{\beta\theta}{1-\gamma^{m+1}}, & \gamma \neq 1. \end{cases} \tag{2.3}$$

Parameters $\mu(\tau)$ and $\delta(\tau)$ are defined similarly (β, γ and θ are substituted by $\beta(\tau)$, $\gamma(\tau)$ and $\theta(\tau)$, respectively, in the relations (2.2) and (2.3)).

The parameters μ and δ can be interpreted as follows. Let the queue at the input of the network in Fig. 1 be infinite at time $t=0$. We denote by D_t the number of departures from the network up to time t. Then for any initial distribution of the second stage queue the following relations hold [4]

$$\mu = \lim_{t\to\infty} ED_t/t, \tag{2.4}$$

$$\delta = \lim_{t\to\infty} \operatorname{Var} D_t/t. \tag{2.5}$$

Comparison of (1.3) and (2.2) shows that the ergodicity condition $D_m(1)<0$ is equivalent to the condition $\rho<1$ where $\rho=\lambda/\mu$ is the load and the heavy traffic condition (1.6) has the form $\rho(\tau)\uparrow 1$, $\tau\downarrow 0$, $\rho(\tau)=\lambda(\tau)/\mu(\tau)$. We observe also that, in view of (2.1) and (2.3)

$$D'_m(1) = \frac{\mu+\delta}{2}\sum_{i=0}^{m}\gamma^i.$$

This leads to the conclusion that assertions (b) and (c) of the Theorem imply that, if $\rho(\tau)\uparrow 1$ as $\tau\downarrow 0$ then

$$\lim_{\tau\downarrow 0} P((1-\rho(\tau))Q_1(\tau) \leqq x) = 1 - \exp\left(-\frac{2x}{1+\delta/\mu}\right), \quad x \geqq 0, \tag{2.6}$$

$$\lim_{\tau\downarrow 0} E(1-\rho(\tau))Q_1(\tau) = \frac{1+\delta/\mu}{2}.$$

Consider the $M/G/1$ queue which is dependent on $\tau > 0$: the input rate is $\lambda(\tau)$, the mean and the variance of the service time are equal to $1/\mu(\tau)$ and $\delta(\tau)/(\mu(\tau))^3$, respectively. Then, if $\rho(\tau) = \lambda(\tau)/\mu(\tau) \uparrow 1$ as $\tau \downarrow 0$, then this system is in heavy traffic as $\tau \downarrow 0$ and, in view of (2.6), its asymptotical behaviour is equivalent to that of $Q_1(\tau)$ (see [3, ch. 4]).

3. Concluding remarks

1. The Theorem also holds for a network with service repetition in the first stage. A request is served in the first stage even in the case when m requests are queued or in service in the second stage. But if, by the time the request completes its service, it cannot proceed to the second stage, it restarts being served at the first-stage server. This cycle is repeated until by the time of service completion the second-stage queue contains less than m requests (including that one in service). In view of the memoryless property of an exponentially distributed random variable, the network with service repetition is described by the same Chapman–Kolmogorov equations (this was observed in [2]). Therefore, the Theorem holds for this network with no modifications.

2. From (2.2) and (2.3) we readily obtain

$$\lim_{m\to\infty} \frac{\delta}{\mu} = \begin{cases} 1-\theta/3, & \gamma = 1, \\ 1, & \gamma \neq 1 \end{cases}$$

and then, from assertion (a) of the Theorem, we have for large m and $\rho \approx 1$

$$EQ \approx \begin{cases} \dfrac{1-\theta/6}{1-\rho}, & \gamma = 1, \\ \dfrac{1}{1-\rho}, & \gamma \neq 1. \end{cases}$$

This means that the disbalance of the network ($\gamma \neq 1$) results in an abrupt increasing of the expected input queue length.

Appendix

Proof of the Lemma. Since $D_m(1, \tau) < 0$ for $\tau > 0$, any of the polynomials $D_m(z, \tau)$ for every $\tau > 0$ has $(m-1)$ zeros in the closed unit disc (taking into account their order), all of which being real (see [1]). Furthermore (see Lemma 2 in [1]), the zero $z = 0$ is of order $\lfloor m/2 \rfloor$ while the rest $\lceil m/2 \rceil - 1$ zeros are positive, simple and all belong to $(0, 1)$ ($\lfloor x \rfloor$ is the largest integer not greater than x; $\lceil x \rceil$ is the smallest integer not less than x). We denote these zeros as they increase: $0 < \eta_1(\tau) < \eta_2(\tau) < \ldots < \eta_{\lceil m/2 \rceil - 1}(\tau) < 1$. The polynomial $D_m(z)$ also has the zero $z = 0$ of order $\lfloor m/2 \rfloor$, $(\lceil m/2 \rceil - 1)$ real simple zeros $0 < \eta_1 < \eta_2 < \ldots < \eta_{\lceil m/2 \rceil - 1} < 1$, and the simple zero $z = 1$ (this was demonstrated while proving assertion (a) of the Theorem).

Since all the zeros $\eta_k(\tau)$, $1 \leqq k \leqq \lceil m/2 \rceil - 1$, $\tau > 0$, are simple and the coefficients of the polynomials $D_m(z, \tau)$ are continuous in τ for $\tau > 0$, and they converge to the coefficients of $D_m(z)$ as $\tau \downarrow 0$, the implicit function theorem states (see, e.g. [5]) that the functions $\eta_k(\tau)$, $1 \leqq k \leqq \lceil m/2 \rceil - 1$, are continuous for $\tau > 0$ and that

$$\lim_{\tau \downarrow 0} \eta_k(\tau) = \eta_k, \quad 1 \leqq k \leqq \lceil m/2 \rceil - 1.$$

It was shown in Theorem 1 of [1] that probabilities $p_{0j}(\tau)$, $j = 0, 1, \ldots, m-1$, $\tau > 0$, are uniquely determined by normalization and by the requirement that every zero of $D_m(z, \tau)$ in the closed unit disc be a zero of $\sum_{j=0}^{m-1} p_{0j}(\tau) D_j(z, \tau)$.

The normalization condition is obtained from (1.12), (1.13) and (1.18). It has the form (recall that $A_j(1, \tau) = (\gamma(\tau))^j$):

$$\sum_{j=0}^{m-1} \left(\sum_{i=0}^{j} (\gamma(\tau))^i D_m(1, \tau) - \sum_{i=0}^{m} (\gamma(\tau))^i D_i(1, \tau) \right) p_{0j}(\tau) = D_m(1, \tau).$$

Thus, the probabilities $p_{0j}(\tau)$, $j = 0, 1, \ldots, m-1$, represent the unique solution to the system of linear equations:

$$\sum_{j=0}^{m-1} D_j(\eta_k(\tau), \tau) x_j = 0, \qquad 1 \leqq k \leqq \lceil m/2 \rceil - 1, \tag{A.1}$$

$$\sum_{j=0}^{m-1} \frac{\partial^k D_j(z, \tau)}{\partial z^k}\bigg|_{(0,\tau)} x_j = 0, \quad 0 \leqq k \leqq \lfloor m/2 \rfloor - 1, \tag{A.2}$$

$$\sum_{j=0}^{m-1} \left(\sum_{i=0}^{j} (\gamma(\tau))^i D_m(1, \tau) - \sum_{i=0}^{m} (\gamma(\tau))^i D_i(1, \tau) \right) x_j - D_m(1, \tau) = 0, \tag{A.3}$$

where by definition $\partial^0 D_j(z, \tau)/\partial z^0 = D_j(z, \tau)$.

In other words, the vector $p_0(\tau) = (p_{01}(\tau), \ldots, p_{0,m-1}(\tau))$ for $\tau > 0$ is the implicit function $x(\tau) = (x_0(\tau), \ldots, x_{m-1}(\tau))$ which is given by an equation of the form

$$C_1(\tau) x + C_2(\tau) = 0, \tag{A.4}$$

where $C_1(\tau)$ is an $m \times m$-matrix, $C_2(\tau)$ is a column vector of length m. Since, for every $\tau > 0$, the solution to the system (A.1)–(A.3) is unique, then the matrix $C_1(\tau)$ is nondegenerate. In addition the functions $C_1(\tau)$, $C_2(\tau)$ are continuous at any $\tau > 0$. So, by the implicit function theorem, the function $x(\tau)$ is continuous for $\tau > 0$.

Now we show that

$$\lim_{\tau \downarrow 0} p_{0j}(\tau) = 0, \quad j = 0, 1, \ldots, m-1. \tag{A.5}$$

To this end we set

$$C_1(0) = \lim_{\tau \downarrow 0} C_1(\tau), \quad C_2(0) = \lim_{\tau \downarrow 0} C_2(\tau)$$

(the limits exist in view of (1.4)). Since $D_m(1, \tau) \uparrow 0$ ($\tau \downarrow 0$), so the vector $\mathbf{0} = (0, \ldots, 0)$ solves equation (A.4) for $\tau = 0$. If we prove that this is the unique solution, then the matrix $C_1(0)$ is nondegenerate and, in view of the implicit function theorem, equation (A.4) considered in the right-hand vicinity of $\tau = 0$ defines a function $x(\tau)$, which is equal to 0 at $\tau = 0$, continuous for $\tau > 0$, and right-continuous at $\tau = 0$. That will prove (A.5) since $p_0(\tau) = x(\tau)$, $\tau > 0$.

So, we consider the system of equations

$$\sum_{j=0}^{m-1} D_j(\eta_k) x_j = 0, \qquad 1 \leqq k \leqq \lceil m/2 \rceil - 1, \tag{A.6}$$

$$\sum_{j=0}^{m-1} \frac{\partial^k D_j(2)}{\partial z^k}\bigg|_0 x_j = 0, \quad 0 \leqq k \leqq \lfloor m/2 \rfloor - 1, \tag{A.7}$$

$$\sum_{j=0}^{m-1} D_j(1) x_j = 0, \tag{A.8}$$

which is obtained from (A.1)–(A.3) by taking limit as $\tau \downarrow 0$. Relations (A.6) and (A.7) mean that the polynomial $\sum_{j=0}^{m-1} D_j(z)x_j$ has $\eta_1, \eta_2, \ldots, \eta_{\lceil m/2 \rceil - 1}$ as its zeros and the zero $z=0$ of order $\lfloor m/2 \rfloor$, and relation (A.8) means that the polynomial has an additional zero $z=1$. Therefore, if x_j, $j=0, 1, \ldots, m-1$, is a nontrivial solution to (A.6)–(A.8), then any zero of the polynomial $D_m(z)$ in the closed unit disc is a zero of $\sum_{j=0}^{m-1} D_j(z)x_j$, i.e. a nondegenerate stationary queue lengths distribution exists for the network with the parameters λ, α, β and θ, that contradicts Theorem 2 in [1].

The proof is completed.

References

1. *Konheim, A. G., Reiser, M.*, A queueing model with finite waiting room and blocking. J. ACM, vol. **23,** pp. 328–341, April 1976.
2. *Bocharov, P. P., Al'bores, F. K.*, On a two-stage exponential queueing system with internal losses or blocking. Problems of Control and Information Theory, vol. **9,** pp. 363–379, 1980.
3. *Aven, O. I., Gurin, N. N., Kogan, Ya. A.*, Performance Evaluation and Optimization of Computer Systems. Nauka, Moscow 1982 (in Russian).
4. *Bocharov, P. P., Naumov, V. A.*, Analysis of hyper-exponential two-phase system with limited waiting room. In: Informatsionnye seti i ikh struktura. Nauka, Moscow 1976, pp. 168–181 (in Russian).
5. *Schwartz, L.*, Analyse mathematique, I. Hermann, Paris 1967.

Двухфазная система с ограниченным промежуточным накопителем и блокировками при большой нагрузке

Я. А. КОГАН, А. А. ПУХАЛЬСКИЙ

(Москва)

Рассматривается двухфазная экспоненциальная система с неограниченным числом мест для ожидания в первой фазе и конечным накопителем во второй. При заполненном накопителе происходит блокировка прибора первой фазы с остановкой его работы. Заявка, обслуженная прибором второй фазы, с положительной вероятностью возвращается в очередь первой фазы или покидает систему. Для режима большой нагрузки в явном виде получено предельное распределение стационарных длин очередей в фазах системы. Показано, что очереди в первой и второй фазах асимптотически независимы, причем первая из них ведет себя как очередь в системе $M(G)1$, а вторая — как очередь в системе $M/M/1/m$.

Я. А. Коган
Институт проблем управления.
СССР, 11342 Москва ГСП-7,
Профсоюзная, 65

Problems of Control and Information Theory, Vol. 17 (1), pp. 15—22 (1988)

PREDICTION IN ZERO-ONE RANDOM SEQUENCES

M. JANŽURA

(Prague)

(Received January 4, 1987)

A method for prediction in sequences of random variables which assume only two values (say zero and one) is derived. The method is based on the application of the higher order Markovian model with pair interactions and it suits well in case of rather short sequence of observed data and long range of dependence. The obtained result is formulated in a way which is easy to be implemented and a numerical example with simulated data is included.

1. Introduction

Random sequences with binary state space occur in many problems of decision making and control. The state space may be given e.g. by {yes, no}, or $\{+, -\}$, or {black, white}, and so forth. For the sake of simplicity we shall use the $\{0, 1\}$ representation of the state space, which may be interpreted as the absence and the presence, respectively, of some phenomenon.

The problem we shall deal with in this paper is to predict the following value in a given zero-one sequence. The sequence is supposed to be generated by a discrete time stochastic process which assumes only values zero and one.

While choosing the prediction procedure, we shall follow the principle saying that the best one is the prediction which minimizes the sum of the error probabilities (which corresponds to the assumption of uniform prior distribution in the Bayesian approach). Thus, if we knew the distribution of the generating random process, there would be practically no problem. Therefore, the main problem consists in the identification of the model, i.e. the estimation of the unknown distribution.

In order to obtain reasonable results, we restrict our further considerations to the class of stationary R-Markovian random sequences. Thus, due to the R-Markov property, it is sufficient to estimate the transition probabilities (of the range R). But, anyhow, it means to estimate probabilities of all the 2^{R+1} possible configurations. And if the observed sequence is not long enough to compare with the number 2^{R+1}, we can hardly expect that the usual "frequency" approach may give a really good estimate.

In such situation we may, of course, decrease the range R of the Markov property. But in this paper we shall introduce another approach which consists in simplifying the dependence structure. An experience of physicists says that almost all

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

systems in nature may be described with the aid of the pair (two-body) interactions between its elements. Following this idea, we hope that without a substantial loss of generality we may assume that the unknown distribution is a Gibbs distribution with interactions of range R, as defined in frame of statistical physics.

With this assumption it is sufficient to estimate only the "covariances" (the analogy with the Gaussian random processes is straightforward). Since for every covariance there are only four possible configurations, they may be estimated quite well even from a short observed sequence.

The method obtained in this way seems a bit complicated and time-consuming when being implemented. But the final formula is quite simple, and, moreover, we must realize that we shall probably have time enough in the situations when the method should be applied (long time intervals between observations).

2. Model identification

Let us denote by $X=\{0, 1\}$ the state space, by $\mathscr{F}=\exp X$ the σ-algebra of all its subsets, and by $\mathbf{Z}$ the set of all integers. By the R-Markovian model with pair interactions we mean any probability measure μ on the product space $(X, \mathscr{F})^{\mathbf{Z}}$ satisfying

$$\mu(x_j|x_{\mathbf{Z}\setminus\{j\}})=\frac{\exp\left\{x_j\left(U_0+\sum_{i=1}^{R}(x_{j+i}+x_{j-i})U_i\right)\right\}}{1+\exp\left\{U_0+\sum_{i=1}^{R}(x_{j+i}+x_{j-i})U_i\right\}}$$

for every $j\in\mathbf{Z}$ and a.e. $x\in X^{\mathbf{Z}}$ $[\mu]$, where $U=(U_0, \ldots, U_R)\in\mathscr{R}^{R+1}$ is $(R+1)$-tuple of parameters called (pair) interactions. (Here the short notation $\mu(x_j|x_{\mathbf{Z}\setminus\{j\}})$ is used for the conditional distribution.)

For every $U\in\mathscr{R}^{R+1}$ there is exactly one probability measure μ_U on $(X, \mathscr{F})^{\mathbf{Z}}$ satisfying the above condition (see e.g. Section I.2.1 in Mayer (1980)). Such probability measure μ_U is called Gibbs distribution with respect to the interactions $U\in\mathscr{R}^{R+1}$ in frame of statistical physics.

The uniquely defined Gibbs distribution is stationary (Proposition 5.4 in Preston (1976)), i.e.

$$\mu_U T^{-1}=\mu_U,$$

where T: $X^{\mathbf{Z}}\to X^{\mathbf{Z}}$ is the shift on $X^{\mathbf{Z}}$ defined through $T(x)_j=x_{j+1}$ for every $j\in\mathbf{Z}$, $x\in X^{\mathbf{Z}}$.

Moreover, according to Proposition 4.1 in Preston (1976), the Gibbs distribution μ_U is ergodic, i.e. its restriction to the σ-algebra of invariant sets assumes only values

zero or one. Precisely, if $\mu_U(F)>0$ then $\mu_U(F)=1$ for every $F\in\mathscr{S}=\{E\in\mathscr{F}^{\mathbf{Z}};\ T^{-1}E=E\}$.

Let us denote $C_i=\{x\in X^{\mathbf{Z}};\ x_0\cdot x_i=1\}\in\mathscr{F}^{\mathbf{Z}}$ for every $i=0,\ldots,R$.

Further, let $\lambda_{\max}(M_U)$ be the uniquely defined (due to the well-known Perron–Frobenius theorem) strictly positive eigenvalue, greater in absolute value than all other eigenvalues, of the so-called transfer matrix M_U which is the strictly positive-valued $2^R\times 2^R$-matrix with elements given by the formula

$$M_U(x_{[1,R]},z_{[1,R)})=\exp\left\{\sum_{i=0}^{R}U_i\left(\sum_{j=1}^{R-i}x_jx_{j+i}+\sum_{i=R-i+1}^{R}x_jz_{j+i-R}\right)\right\}$$

for every $x_{[1,R]},\ z_{[1,R]}\in X^{[1,R]}$.

Now, following Mayer (1980), Section I.2.1, we obtain that $p(U)=$ $=R^{-1}\log\lambda_{\max}(M_U)$ is the free energy function, and, according to the variational principle for Gibbs distributions (cf. e.g. Proposition 8.1 in Preston (1976)), we conclude that

$$\min_{U\in\mathscr{R}^{R+1}}\left(p(U)-\sum_{i=0}^{R}U_i\mu_{U^0}(C_i)\right)=p(U^0)-\sum_{i=0}^{R}U_i^0\mu_{U^0}(C_i)$$

for every $U^0\in\mathscr{R}^{R+1}$.

The latter result is used for estimating the interactions on base of an observed zero-one sequence $\bar{x}_1,\ldots,\bar{x}_n$ which is assumed to be generated by a Gibbs distribution μ_{U^0}. For the terms $\mu_{U^0}(C_i)$ we substitute their simple "frequency" estimates

$$\hat{\mu}_{U^0}(C_i)=(n-i)^{-1}\sum_{j=1}^{n-i}\bar{x}_j\bar{x}_{j+i},\qquad i=0,\ldots,R,$$

and, minimizing the expression $p(U)-\sum_{i=0}^{R}U_i\hat{\mu}_{U^0}(C_i)$, we obtain a consistent estimate $\hat{U}$ of the unknown interactions U^0 (for details cf. Janžura (1986)).

3. Prediction

It is not difficult to see that the "bilateral" Markov property of our model yields the usual "unilateral" Markov property. For the case $R=1$ cf. e.g. Preston (1976), Section 5, while the case $R>1$ will be treated as follows. Aggregating the segments of length R, we obtain a 1-Markovian model with state space given by the set $X^{[1,R]}$. Here the statement holds, and the claimed property of the original model we obtain by summing over the appropriate set of configurations.

But, unfortunately, the connection between the interactions and the "unilateral", conditional distributions, which we need for constructing the optimal prediction, is not so straightforward.

Let $\hat{U}$ be the estimate of interactions obtained from the observed sequence, let $\hat{\mu}=\mu_{\hat{U}}$ be the corresponding Gibbs distribution, and $\hat{\lambda}$ the greatest eigenvalue of the transfer matrix $\hat{M}=M_{\hat{U}}$. Further, we denote by $\hat{r}$ the corresponding left eigenvector (all its elements being also strictly positive).

Keeping the "aggregation" idea, we obtain, due to the Theorem 5.37 in Preston (1976), that

$$\hat{\mu}(x_{[1,R]}|y_{[1,R]})=\frac{\hat{M}(x_{[1,R]}|y_{[1,R]})\cdot\hat{r}(x_{[1,R]})}{\hat{\lambda}\cdot\hat{r}(y_{[1,R]})}$$

is the matrix of the transition probabilities corresponding to the aggregated model.

Since $\hat{\mu}$ obeys the unilateral Markov property, we may write

$$\hat{\mu}(x_{[1,R]}|y_{[1,R]})=\prod_{i=1}^{R}\hat{\mu}(x_i|x_{i+1},\ldots,x_R,y_1,\ldots,y_i)$$

for every $x_{[1,R]}, y_{[1,R]}\in X^{[1,R]}$.

Our problem is how to calculate $\hat{\mu}(x_0|x_1,\ldots,x_R)$ for every $x_0\in X$, $x_{[1,R]}\in X^R$ (note that $\hat{\mu}(x_0|x_1,\ldots,x_R)>0$) with the aid of the two relations above.

Now, since we can identify each $x_{[1,R]}\in X^{[1,R]}$ with the set $A\subset[1,R]$ for which it holds: $x_i=1$ iff $i\in A$, we shall write rather $\hat{\mu}(x_0|A)$ and $\hat{\mu}(B|A)$ instead of $\hat{\mu}(x_0|x_1,\ldots,x_R)$ and $\hat{\mu}(y_{[1,R]}|x_{[1,R]})$, respectively. Further, we denote $\vec{A}=(A+1)\cap$ $\cap[1,R]$ for every $A\subset[1,R]$.

By standard considerations we obtain a system of constants $\{V_B\}_{B\subset[1,R]}$ such that

$$\frac{\hat{\mu}(1|A)}{\hat{\mu}(0|A)}=\exp\left\{\sum_{B\subset A}V_B\right\}$$

holds for every $A\subset[1,R]$.

Note that

$$V_B=\sum_{C\subset B}(-1)^{|B\setminus C|}\log\frac{\mu(1|C)}{\mu(0|C)}\qquad\text{for every}\quad B\subset[1,R].$$

In the following theorem we shall express the system $\{V_B\}_{B\subset[1,R]}$ with the aid of the known characteristics of our model.

Theorem 1. For every $B\subset[1,R]$ it holds

$$V_B=\sum_{C\subset B}(-1)^{|B\setminus C|}\log\left((\hat{\lambda})^{\frac{1}{R}}\frac{\hat{r}(C)}{\hat{r}(\vec{C})}-1\right).$$

Proof. For every $B \subset [1, R]$ it holds

$$\hat{\mu}(\emptyset \mid B) = \frac{\hat{r}(\emptyset)}{\hat{\lambda}\hat{r}(B)};$$

and, from the other side,

$$\hat{\mu}(\emptyset \mid B) = \left[\prod_{i=1}^{R} \left(1 + \exp \left\{ \sum_{C \subset \{B + R - i\} \cap [1, R]} V_C \right\} \right) \right]^{-1}.$$

Similarly

$$\hat{\mu}(\emptyset \mid \vec{B}) = \left[\prod_{i=0}^{R-1} \left(1 + \exp \left\{ \sum_{C \subset \{B + R - i\} \cap [1, R]} V_C \right\} \right) \right]^{-1}$$

and therefore

$$\frac{\hat{\mu}(\emptyset \mid B)}{\hat{\mu}(\emptyset \mid \vec{B})} = \frac{r(\vec{B})}{r(B)} = \frac{(1 + \exp(V_\emptyset))}{\left(1 + \exp\left(\sum_{C \subset B} V_C \right) \right)}.$$

Wherefrom we obtain

$$\sum_{C \subset B} V_C = \log \left((1 + \exp(V_\emptyset)) \frac{r(B)}{r(\vec{B})} - 1 \right).$$

Since especially $\hat{\mu}(\emptyset \mid \emptyset) = \frac{1}{\hat{\lambda}} = \frac{1}{(1 + \exp V_\emptyset)^R}$, the proof is finished by standard arguments.

Corollary. For every $A \subset [1, R]$ it holds that

$$\frac{\hat{\mu}(1 \mid A)}{\hat{\mu}(0 \mid A)} = (\hat{\lambda})^{\frac{1}{R}} \frac{\hat{r}(A)}{r(\vec{A})} - 1.$$

Proof. The statement follows immediately from the proof of the preceding theorem.

Let us recall that the optimal prediction (in the sense of minimal sum of the error probabilities) is based on checking if the expressed above ratio exceeds one.

4. Implementation

At the beginning we are given an observed zero-one sequence $\bar{x}_1, \ldots, \bar{x}_n$ which is assumed to be generated by an R-Markovian (R fixed) model with pair interactions.

We calculate the estimates $\hat{\mu}(C_i) = (n-1)^{-1} \sum_{j=1}^{n-i} \bar{x}_j \bar{x}_{j+i}$ for $i = 0, \ldots, R$.

The model is identified by minimizing the convex function

$$R^{-1} \log \lambda_{\max}(M_U) - \sum_{i=0}^{R} U_i \hat{\mu}(C_i).$$

Let $\hat{U}$ be the vector of interactions for which the minimum is reached. From the preceding section it follows that we actually need the greatest eigenvalue $\hat{\lambda} = \lambda_{\max}(M_{\hat{U}})$ and the corresponding left eigenvector $\hat{r}$. But, numerical calculation of $\hat{\lambda}$ involves simultaneous calculation of the corresponding eigenvector.

Thus, when the minimization procedure is stopped, all the needed values are immediately available. To be quite fair, let us note that with some small probability the minimum of the convex function above may not exist. Anyhow, a stopping rule of the minimization algorithm will give some result which may be considered as an estimate of the interactions in such a singular case.

Then, according to Theorem 1 in the preceding section, we may calculate the system $\{V_B\}_{B \subset [1,R]}$ and formulate the prediction rule:

$$\text{if} \quad \sum_{B \subset [1,R]} V_B \prod_{i \in B} \bar{x}_{n-i+1} > 0 \qquad \text{we predict} \quad \bar{x}_{n+1} = 1,$$

otherwise we predict $\bar{x}_{n+1} = 0$ $\left(\text{setting } \prod_{i \in \emptyset} \bar{x}_{n-i+1} = 1\right)$.

The system $\{V_B\}_{B \subset [1,R]}$ offers deep insight into the dependence structure and indicates which values in the past are important for the prediction.

On the other hand, for fast prediction the result contained in the corollary to Theorem 1 seems to be more convenient.

Let $A_n \subset [1, R]$ be the set of indices satisfying: $i \in A_n$ iff $\bar{x}_{n-i+1} = 1$ for every $i \in [1, R]$.

Then, according to the corollary, the prediction procedure assumes the form

$$\text{if} \quad (\hat{\lambda})^{\frac{1}{R}} \frac{\hat{r}(A_n)}{\hat{r}(\bar{A}_n)} > 2 \qquad \text{we predict} \quad \bar{x}_{n+1} = 1,$$

otherwise we predict $\bar{x}_{n+1} = 0$.

5. Example

Now, let us introduce a simple example to illustrate how the proposed method works.

We generated 100 successive values of a 3-Markov chain with the transition probabilities given by the following table

x_1	x_2	x_3	$\log P(1\|x_1, x_2, x_3)/P(0\|x_1, x_2, x_3)$
0	0	0	0.5
0	0	1	−0.5
0	1	0	1.5
0	1	1	0.5
1	0	0	1.0
1	0	1	0
1	1	0	2.0
1	1	1	1.0

The first 30 values formed the training set and the remaining 70 values were used to check the quality of the prediction. We repeated the test three times with various initial values and compared the frequency of errors (FE1) in prediction with the result (FE2) obtained by the usual method based on calculating the empirical distribution. We applied the two models with $R=2$ and $R=3$, respectively, and arranged the results into a table:

	FE1 (FE2)		
$R=2$	0.34 (0.39)	0.27 (0.27)	0.37 (0.37)
$R=3$	0.29 (0.44)	0.27 (0.34)	0.34 (0.56)

One can see that in this case the proposed method gives really slightly better results. Moreover, FE1 is getting smaller while FE2 is getting greater when the range R of the model is increased.

6. Concluding remarks

I. There is one more reason for the R-Markovian model with pair interactions since the corresponding Gibbs distribution assumes the maximal entropy rate among the distributions with the given covariances. Thus, the approach introduced here may be viewed as an application of the maximum entropy principle for the approximation of an unknown distribution.

II. The approach employed in this paper is based on the theory of Gibbs random fields, initiated by the pioneering work of Dobrushin and Ruelle. Many useful basic results are contained in Preston's book, while Mayer deals mostly with the one-dimensional models which are in particular relevant here for the considered problem.

References

1. *Dobrushin, R. L.*, The description of a random field by conditional probabilities and conditions on its regularity. Teorija Verojatn. i Ee Prim. **12,** 1968, 197–224.
2. *Janžura, M.*, Estimating Interactions in Binary Data Sequences. Kybernetika, Vol. **22,** 1986, 227–284.
3. *Mayer, D. H.*, The Ruelle–Araki Transfer Operator in Classical Mechanics (Lecture Notes in Physics, No. *123*). Springer Verlag, Berlin, 1980.
4. *Preston, C.*, Random Fields (Lecture Notes in Math., No. *534*). Springer Verlag, Berlin, 1976.
5. *Ruelle, D.*, Statistical Mechanics. Rigorous Results. Benjamin, New York, 1969.

Прогноз случайных последовательностей нулей и единиц

М. ЯНЖУРА

(Прага)

В статье обсуждается метод прогнозирования последовательностей случайных величин с двумя значениями. Метод основан на применении марковской модели высшего порядка с парными взаимодействиями. Подобный же метод можно использовать как в случае меньшего количества наблюдений, так и при наличии зависимости на длинных временных интервалах. Получаемый результат сформулирован таким образом, чтобы облегчить его применение. Приведен пример, использующий результаты численного моделирования.

M. Janžura
Institute of Information Theory and Automation,
Czechoslovak Academy of Sciences,
Pod vodárenskou věží 4,
182 08 Prague 8,
Chechoslovakia

Problems of Control and Information Theory, Vol. 17 (1), pp. 23—32 (1988)

REDUCED MULTI-STEP ALGORITHMS FOR IDENTIFICATION OF LINEAR PLANTS

V. I. SALYGA, O. G. RUDENKO, V. L. OBRUCHEV

(Moscow)

(Received March 24, 1986)

At present multi-step adaptive identification algorithms, permitting to get mathematical description of an object under conditions of minimal prior information, find broad application in the synthesis of complex object adaptive control. The main advantages attracting researchers' attention to procedures of this kind, are the simplicity of calculations (inverse matrix is no more necessary), a sufficiently small volume of operation memory, the possibility to identify in the process of object operation that is in real time. In contrast to the method of least squares in recurrent form, reduced multi-step algorithms may be used when the number of observations is smaller than that of the evaluated parameters.

A multi-step identification algorithm is discussed where recognition of the data on some preceding steps results in a significant reduction of the model development time, provided the measurements are not noise-contaminated. However, because they are, as the algorithm memory (number of steps stored) expands, the algorithm properties deteriorate in that the r.m.s. identification error increases. This error is reduced if the estimates are allowed to be biased and if reduced or shortened estimates are obtained. The feasible reduction parameter range is determined with which the r.m.s. error is smaller than that of the multi-step algorithm. The optimal reduction parameter value is discussed with which the rate of convergence is maximal. Much attention is given to the design of computationally more convenient recurrent forms of a reduced multi-step algorithm. One form is obtained by using the common part of the observation matrix which results from estimation at two subsequent steps. In another form, direct inversal of the observation matrix is replaced by its recurrent computation.

A major efficiency indicator for identification algorithms is the rate of convergence. Many papers [1–7] discuss a plant described by the equation

$$y_n = c^{*T} x_n + \xi_n, \tag{1}$$

where y_n is the output signal; $x_n = (x_{1n}, x_{2n}, \ldots, x_{Nn})^T$ is the vector of input signals; $c^* = (c_1^*, c_2^*, \ldots, c_N^*)^T$ is the vector of desired parameters; and ξ_n is a random error. This plant is identified by multi-step adaptive algorithms which in many cases significantly

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

accelerate identification; it is true that the structure complexity insignificantly increases because a larger amount of data has to be processed.

Several papers [2, 4–7] report a multi-step procedure

$$c_n = c_{n-1} + X_n^{(S)}[X_n^{(S)T} X_n^{(S)}]^{-1} E_n^{(S)}, \tag{2}$$

where $X_n^{(S)} = (x_n, x_{n-1}, \ldots, x_{n-S+1})$ is an $N \times S$ matrix:

$$E_n^{(S)} = (y_n - c_{n-1}^T x_n, y_{n-1} - c_{n-1}^T x_{n-1}, \ldots, y_{n-S+1} - c_{n-1}^T x_{n-S+1})^T$$

is an $S \times 1$ dimensional vector; and S is the number of algorithm steps (memory depth).

In this algorithm it is assumed that 1) $\{x_{ik}\}$ are independent Gaussian quantities with $M\{x_{ik}\} = 0$ and identical variance $M\{x_{ik} x_{jm}\} = \sigma_x^2 \delta_{ij} \delta_{km}$, and 2) ξ_n is a Gaussian quantity with $M\{\xi_n\} = 0$ and $M\{\xi_n^2\} = \sigma_\xi^2$.

Some properties of multi-step algorithms

Introducing an identification error $\Theta_i = c_i - c^*$ and with the algorithm (2) becomes

$$\Theta_n = (I_N - P_n^{(S)})\Theta_{n-1} + B_n^{(S)} \xi_n^{(S)}, \tag{3}$$

where I_N is an $N \times N$ identity matrix; $P_n^{(S)} = X_n^{(S)}[X_n^{(S)T} X_n^{(S)}]^{-1} X_n^{(S)T}$ is a matrix of orthogonal mapping on the linear hull S of the vectors $x_n, x_{n-1}, \ldots, x_{n-s+1}$; $B_n^{(S)} = X_n^{(S)}[X_n^{(S)T} X_n^{(S)}]^{-1}$ is an $N \times S$ matrix; and $\xi_n^{(S)} = (\xi_n, \xi_{n-1}, \ldots, \xi_{n-S+1})$ is an $S \times 1$ dimensional vector.

The assumptions on the properties of x_n and ξ_n, the properties of mapping operators [8], and the following properties of the matrices $P_n^{(S)}$ and $B_n^{(S)}$ [7]:

$$M_{x_n/x_{n-1}, \ldots, \xi}\{P_n^{(S)}\} = \left(1 - \frac{1}{m}\right) P_{n-1}^{(S-1)} + \frac{1}{m} I_N;$$

$$M_{\{x\}}\{B_n^{(S)T} B_n^{(S)}\} = M_{\{x\}}\{[X_n^{(S)T} X_n^{(S)}]^{-1}\} = \frac{1}{\sigma_x^2 m} I_S;$$

$$M_{\{x\}}\{B_n^{(S)T} P_n^{(S)} B_n^{(S)}\} = \frac{1}{N-S} I_S^{(S-1)},$$

where $m = N - S + 1$; and $I_S^{(S-1)} = \left(\begin{array}{c|c} I_{S-1} & 0 \\ \hline 0 & 0 \end{array}\right)$ makes it possible to study the properties of algorithm (2). The following equations are true [7]

$$\lim_{n \to \infty} M\{\Theta_n\} = 0, \tag{4}$$

$$M\{\|\Theta_n\|^2\}=k_S\|\Theta\|^2\left(1-\frac{1}{m}\right)^{n-S}+\frac{N+S-1}{N-S-1}\frac{\sigma_\xi^2}{\sigma_x^2}\left(1-\left(1-\frac{1}{m}\right)^{n-S}\right) \tag{5}$$

where k_S is a coefficient which depends on the iteration technique at the first S steps. Thus if at these steps the estimates are updated by the Kaczmarz algorithm, then $k_S=(1-N^{-1})^S$; if at the first step a single-step (Kaczmarz) algorithm is used, at the second step, a two-step algorithm, is used, etc., then $k_S=\left(1-\frac{1}{N}\right)\ldots$ $\ldots\left(1-\frac{1}{N-S+2}\right)$. If at the first S steps the data is accumulated and the estimate is not derived until after $x_S, x_{S-1}, \ldots, x_1$ are available, then $k_S=\left(1-\frac{S}{N}\right)$ [5].

From expression (4) the estimate, provided by algorithm (2), is seen to be asymptotically unbiased. From formula (5) it is easy to determine the domain where the algorithm converges

$$\lim_{n\to\infty} M\{\|\Theta_n\|^2\}=\frac{N+S-1}{N-S-1}\frac{\sigma_\xi^2}{\sigma_x^2}. \tag{6}$$

If follows from equations (5) and (6) that the rate of convergence (the first added in (5)) increases with S but the convergence domain (6) expands.

One way to reduce this domain is to replace unbiased estimates by biased estimates (in the case of normal sample an estimate from the class of all estimates can be obtained whose means square error is the smallest but which is biased). This approach also works in improving the stability of estimates in the case of multi-collinearity (when they should not necessarily be unbiased, the estimates are stable even with singular observation matrices). Because multi-collinearity increases the estimate length, shortened or reduced estimates should be considered.

Reduced multi-step algorithms and their properties

By analogy with reduced estimates obtained in the method of least squares and thoroughly explored in the literature (e.g. [9]), let us consider estimates obtained by multi-step algorithms (2).

A reduced S-step algorithm has the form

$$c_n^\lambda=c_{n-1}^\lambda+X_n^{(S)}[X_n^{(S)T}X_n^{(S)}]^{-1}\hat{E}_n^{(S)}, \tag{7}$$

where $\hat{E}_n^{(S)}=(\lambda y_n-c_{n-1}^{\lambda T}x_n, \ldots, \lambda y_{n-S+1}-c_{n-1}^{\lambda T}x_{n-S+1})^T$ is an $S\times 1$ dimensional algorithm and $0<\lambda\leqq 1$ is the reduction factor.

The properties of this algorithm are studied with the above assumptions on the statistical properties of useful signals and noise.

Substracting c^* from both sides of equation (7), write an algorithm for identification errors Θ

$$\Theta_n^\lambda = (I_N - P_n^{(S)})\Theta_{n-1}^\lambda - (1-\lambda)P_n^{(S)}c^* + \lambda B_n^{(S)}\xi_n^{(S)}. \tag{8}$$

With an allowance for statistical properties of mapping operators and statistical properties of useful signals and noise we have, following the averaging of (8) and iteration

$$M\{\Theta_n^\lambda\} = k_S\left(1-\frac{1}{m}\right)^{n-S}\Theta_0 - (1-\lambda)\left(1-\left(1-\frac{1}{m}\right)^{n-S}\right)c^*,$$

whence it follows that as the identification time grows

$$\lim_{n\to\infty} M\{\Theta_n^\lambda\} = -(1-\lambda)c^*,$$

or

$$\lim_{n\to\infty} M\{c_n^\lambda\} = \lambda c^*. \tag{9}$$

Consequently, the estimate c_n^λ is biased.

Multiplying both sides of equation (8) by $\Theta_n^{\lambda T}$ rightwards we have, following the averaging of the resultant equation

$$M\{\|\Theta_n^\lambda\|^2\} = M\{\Theta_{n-1}^{\lambda T}(I_N - P_n^{(S)})\Theta_{n-1}^\lambda\} +$$
$$+(1-\lambda)^2 M\{c^{*T}P_n^{(S)}c^*\} + \lambda^2 M\{\xi_n^{(S)T}B_n^{(S)T}B_n^{(S)}\xi_n^{(S)}\}.$$

By analogy with [7] we have

$$M\{\|\Theta_n^\lambda\|^2\} = k_S\left(1-\frac{1}{m}\right)^{n-S}\|\Theta_0\|^2 + (1-\lambda)^2\left(1-\left(1-\frac{1}{m}\right)^{n-S}\right)\|c^*\|^2 +$$
$$+\lambda^2\frac{N+S-1}{N-S-1}\frac{\sigma_\xi^2}{\sigma_x^2}\left(1-\left(1-\frac{1}{m}\right)^{n-S}\right),$$

$$\lim_{n\to\infty} M\{\|\Theta_n^\lambda\|^2\} = (1-\lambda)^2\|c^*\|^2 + \lambda^2\frac{N+S-1}{N-S-1}\frac{\sigma_\xi^2}{\sigma_x^2}. \tag{10}$$

It is obvious from equation (10) that by choosing in (7) the reduction factor $0<\lambda<1$ an estimate c_n^λ can be obtained with $M\{\|\Theta_n^\lambda\|^2\} < M\{\|\Theta_n\|^2\}$. The range of λ

can be obtained more accurately in the following way. Since it is required that $M\{\|\Theta_n^\lambda\|^2\} < M\{\|\Theta_n\|^2\}$, equations (10) and (6) lead to the inequality

$$(1-\lambda)^2\|c^*\|^2 + \lambda^2 \frac{N+S-1}{N-S-1}\frac{\sigma_\xi^2}{\sigma_x^2} < \frac{N+S-1}{N-S-1}\frac{\sigma_\xi^2}{\sigma_x^2},$$

which is easily seen to hold with

a) $0<\lambda<1$ if $\|c^*\|^2 < \dfrac{N+S-1}{N-S-1}\dfrac{\sigma_\xi^2}{\sigma_x^2}$;

b) $\dfrac{(N-S-1)\sigma_x^2\|c^*\|^2-(N+S-1)\sigma_\xi^2}{(N-S-1)\sigma_x^2\|c^*\|^2+(N+S-1)\sigma_\xi^2} < \lambda < 1$, otherwise.

The optimal reduction factor λ^{opt} is obtained by minimizing equation (10). Following differentiation with respect to λ and equating the derivative to zero we have an equation, linear for the desired parameter, whose solution yields

$$\lambda^{\text{opt}} = \frac{(N-S-1)\sigma_x^2\|c^*\|^2}{(N-S-1)\sigma_x^2\|c^*\|^2+(N+S-1)\sigma_\xi^2}. \tag{11}$$

Since $\delta^2 M\{\|\Theta_n^\lambda\|^2\}/\delta\lambda^2 < 0$, the resultant equation minimizes (10). This value of λ^{opt} is associated with

$$M\{\|\Theta_\infty^\lambda\|\} \underset{\min}{=} \frac{(N-S-1)\|c^*\|^2\sigma_\xi^2(3\|c^*\|^2(N+S-1)+(N-S-1)\sigma_\xi^2}{(N+S-1)\sigma_x^2((N-S-1)\sigma_\xi^2+(N+S-1)\|c^*\|^2\sigma_x^2}.$$

Equation (11) is, however, impracticable because λ^{opt} is a function of unknown parameters c^*, σ_x^2, and σ_ξ^2 which should be replaced by their estimates c_n^λ, $\hat{\sigma}_x^2$, and $\hat{\sigma}_\xi^2$ when equation (7) is used; in other words, λ^{opt} of equation (11) should be replaced by a stochastic reduction factor

$$\hat{\lambda}_n^{\text{opt}} = \frac{(N-S-1)\hat{\sigma}_x^2\|c_n^\lambda\|^2}{(N-S-1)\hat{\sigma}_x^2\|c_n^\lambda\|^2+(N+S-1)\hat{\sigma}_\xi^2}$$

Because $\hat{\lambda}_n^{\text{opt}}$ is a function of time t, the resultant value of $\hat{\lambda}_n^{\text{opt}}$ should be used to determine the estimate c_{n+1}^λ which leads to $\hat{\lambda}_{n+1}^{\text{opt}}$, etc.

The above formulae suggest that in the absence of noise ($\sigma_\xi^2 = 0$) $\lambda^{\text{opt}} = 1$ while $M\{\|\Theta_\infty\|^2\} = 0$, or the parameters c^* are determined accurately. The resultant expressions for λ show that the optimal value λ^{opt} falls against S and the ratio σ_ξ^2/σ_x^2.

The recurrent form of reduced S-step algorithms

When algorithms (2) and (7) are used, the inverse matrix $[X_n^{(S)T} X_n^{(S)}]^{-1}$ has to be calculated at every step of the iterative identification process. The computer load is reduced by recurrent recomputation of this matrix. To obtain a recurrent form of the algorithm (7) (the algorithm (2) is a particular case of (7) with $\lambda = 1$), let us decompose the matrix $X_n^{(S)}$ into blocks $[x_n | X_{n-1}^{(S-1)}]$ where $X_{n-1}^{(S-1)} = (x_{n-1}, \ldots, \ldots, x_{n-S+1})$ is an $N \times (S-1)$ matrix. By virtue of the rules for multiplication and inversal of block matrices, we have, following simple transformations

$$c_n^{\lambda} = c_{n-1}^{\lambda} + \frac{R_{n-1}^{(S-1)} x_n}{\alpha_n} \left(\lambda y_n - c_{n-1}^{\lambda T} x_n \right), \tag{12}$$

where

$$R_{n-1}^{(S-1)} = I_N - X_n^{(S)} [X_n^{(S)T} X_n^{(S)}]^{-1} X_n^{(S)T}.$$

Using block-wise representation of the matrix $X_{n-1}^{(S-1)} = [X_{n-1} | X_{n-2}^{(S-2)}]$ leads to a recurrent form of computing $R_{n-1}^{(S-1)}$. Following simple transformations we have finally

$$R_{n-S+i}^{(i)} = R_{n-S+i-1}^{(i-1)} - \frac{R_{n-S+i-1}^{(i-1)} x_{n-S+i} x_{n-S+i}^T R_{n-S+i-1}^{(i-1)}}{\alpha_{n-S+i}}, \tag{13}$$

where

$$\alpha_{n-S+i} = x_{n-S+i}^T R_{n-S+i-1}^{(i-1)} x_{n-S+i}, \quad i = \overline{1, S-1}, \quad R_{n-S}^{(0)} = I_N.$$

Consequently, direct inversal of the matrix $X_n^{(S)T} X_n^{(S)}$ is replaced by its recurrent computation whereby at every step of the identification process the $(S-1)$-st iteration has to be performed in compliance with (13). The recurrent form of the algorithm with $\lambda = 1$ is given in [2].

The factors in equations (12) and (13) are ratios of Gramm determinants which indicate the linear dependence of the input vectors. If at some i-th step α_i vanishes, then this value of x_i is not used in the algorithm which proceeds to the following value x_{i+1}.

The computing stability of the algorithm (7) is increased by introducing a regularizing supplement

$$c_n^{\lambda} = c_{n-1}^{\lambda} + X_n^{(S)} [X_n^{(S)T} X_n^{(S)} + \delta^2 I_S]^{-1} E_n^{(S)}. \tag{14}$$

This does not bias the estimate further and the factors α do not vanish. Consequently, the estimate c_n^{λ} can be improved at every iteration by using all the observations.

To obtain a recurrent form of equation (14), the original $N \times 1$-dimensional input vectors x are replaced by expanded $(N+S) \times 1$, dimensional vectors $\tilde{x}$. Then $\tilde{x}_n, \tilde{x}_{n-1}, \ldots, \tilde{x}_{n-S+1}$ have the form

$$\tilde{x}_n^T = (x_n^T, 0, 0, \ldots, \delta), \quad \tilde{x}_{n-1}^T = (x_{n-1}^T, 0, 0, \ldots, \delta, 0), \ldots,$$

$$\ldots, \tilde{x}_{n-S+1}^T = (x_{n-S+1}^T, \delta, 0, \ldots, 0).$$

or every measured vector is supplemented with an $(S-1)$-st zero component and one non-zero component δ. The associated matrices $\tilde{X}_n^{S)}$ of the vectors $\tilde{x}_n, \tilde{x}_{n-1}, \ldots, x_{n-S+1}$, have the form

$$\tilde{X}_n^{(S)} = \left[\frac{X_n^{(S)}}{\delta I_S}\right].$$

As before, decomposing $X_n^{(S)}$ and $\tilde{X}_n^{(S)}$ into blocks, etc., we have, following simple transformations

$$c_n^{\lambda} = c_{n-1}^{\lambda} + \frac{1}{\tilde{\alpha}_n}(x_n - \tilde{P}_{n-1}^{(S-1)}\tilde{x}_n)(e_n - \tilde{x}_n^T \tilde{c}_{n-1}^{\lambda}), \tag{15}$$

$$\tilde{c}_n^{\lambda} = \tilde{c}_{n-1}^{\lambda} + \frac{1}{\tilde{\alpha}_n}\tilde{R}_{n-1}^{(S-1)}\tilde{x}_n(e_n - \tilde{x}_n^T c_{n-1}^{\lambda}), \tag{16}$$

where

$$\tilde{\alpha}_n = \tilde{x}_n^T R_{n-1}^{(S-1)} x_n; \quad e_n = \lambda y_n - x_n^T c_{n-1}^{\lambda};$$

$$\tilde{P}_{n-1}^{(S-1)} = X_{n-1}^{(S-1)}[\tilde{X}_{n-1}^{(S-1)}\tilde{X}_{n-1}^{(S-1)}]^{-1}\tilde{X}_{n-1}^{(S-1)T} \text{ is an } N \times (N+S) \text{ matrix;}$$

$$\tilde{R}_{n-1}^{(S-1)} = I_{N+S} - \tilde{X}_{n-1}^{(S-1)}[\tilde{X}_{n-1}^{(S-1)T}\tilde{X}_{n-1}^{(S-1)}]^{-1}\tilde{X}_{n-1}^{(S-1)T} \text{ is an } (N+S) \times N+S) \text{ matrix}$$

which can be recurrently computed

$$\tilde{P}_{n-S+i}^{(i)} = \tilde{P}_{n-S+i-1}^{(i-1)} + \frac{1}{\tilde{\alpha}_{n-S+i}}(x_{n-S+i} - \tilde{P}_{n-S+i-1}^{(i-1)}\tilde{x}_{n-S+i}) \times$$

$$\times \tilde{x}_{n-S+i}^T \tilde{R}_{n-S+i-1}^{(i-1)}, \tag{17}$$

$$\tilde{R}_{n-S+i}^{(i)} = \tilde{R}_{n-S+i-1}^{(-1)} - \frac{1}{\tilde{\alpha}_{n+S+i}}\tilde{R}_{n-S+i-1}^{(i-1)}\tilde{x}_{n-Si}\tilde{x}_{n-S+i}^T\tilde{R}_{n-S+i-1}^{(i-1)}, \tag{18}$$

where

$$\tilde{\alpha}_{n-S+i} = \tilde{x}_{n-S+i}^T \tilde{R}_{n-S+i-1}^{(i-1)}\tilde{x}_{n-S+i}, \quad i = \overline{1, S-1},$$

$$\tilde{R}_{n-S}^{(0)} = I_{N+S}, \quad \tilde{P}_{n-S}^{(0)} = 0.$$

Estimates cannot be computed by formulae (12), (13) and (15)–(18) before time $n \geqq S$. Until that time they are obtained by these formulae where, however, S is variable.

A recurrent form of the algorithm (7) other than the above one, can be obtained by recognizing the specific way in which the observation matrix $X_n^{(S)}$ is obtained. Because at every step the latest observation (the last column of the matrix $X_{n-1}^{(S)}$, is rejected and a new observation x_n is included in the matrix

$$X_{n-1}^{(S)} = (x_{n-1}, \ldots, x_{n-S}) = [X_{n-1}^{(S-1)} | x_{n-S}],$$

$$X_n^{(S)} = (x_n, \ldots, x_{n-S+1}) = [x_n | X_{n-1}^{(S-1)}].$$

Using the notation

$$A_{n-1}^{(S)} = X_{n-1}^{(S)T} X_{n-1}^{(S)}, \qquad A_n^{(S)} = X_n^{(S)T} X_n^{(S)},$$

$$D_{n-1}^{(S)} = [A_{n-1}^{(S-1)}]^{-1}, \qquad D_n^{(S)} = [A_n^{(S)}]^{-1}$$

and block-wise representation of these matrices

$$A_{n-1}^{(S)} = \begin{bmatrix} [D_{n-1}^{(S-1)}]^{-1} & X_{n-1}^{(S-1)} x_{n-S} \\ x_{n-S}^T X_{n-1}^{(S-1)} & \|x_{n-S}\|^2 \end{bmatrix},$$

$$A_n^{(S)} = \begin{bmatrix} \|x_n\|^2 & x_n^T X_{n-1}^{(S-1)} \\ X_{n-1}^{(S-1)T} x_n & [D_{n-1}^{(S-1)}]^{-1} \end{bmatrix},$$

$$D_{n-1}^{(S)} = \begin{bmatrix} F_{n-1} & b_{n-1} \\ b_{n-1}^T & d_{n-1} \end{bmatrix}, \qquad D_n^{(S)} = \begin{bmatrix} \hat{d}_n & \hat{b}_n^T \\ \hat{b}_n & \hat{F}_n \end{bmatrix}. \tag{19}$$

Simple transformations result in a recurrent form of the multi-step algorithm

$$c_n = c_{n-1} + X_n^{(S)} D_n^{(S)} E_n^{(S)},$$

where the matrix $D_n^{(S)}$ of equation (19) is computed by the formulae

$$D_{n-1}^{(S-1)} = F_{n-1} - \frac{F_{n-1} X_{n-1}^{(S-1)T} x_{n-S} x_{n-S}^T X_{n-1}^{(S-1)} F_{n-1}}{x_{n-S}^T (I_N + X_{n-1}^{(S-1)} F_{n-1} X_{n-1}^{(S-1)T}) x_{n-S}},$$

$$\hat{F}_n = D_{n-1}^{(S-1)} + \frac{D_{n-1}^{(S-1)} X_{n-1}^{(S-1)} x_n x_n^T X_{n-1}^{(S-1)} D_{n-1}^{(S-1)}}{x_n^T (I_N - X_{n-1}^{(S-1)} D_{n-1}^{(S-1)} X_{n-1}^{(S-1)T}) x_n},$$

$$\hat{b}_n = D_{n-1}^{(S-1)} X_{n-1}^{(S-1)T} x_n \hat{d}_n,$$

$$\hat{d}_n = [x_n^T (I_N - X_{n-1}^{(S-1)} D_{n-1}^{(S-1)} X_{n-1}^{(S-1)T}) x_n]^{-1}.$$

Conclusions

When the estimates are allowed to be biased, reduced multi-step algorithms lead to estimates whose r.m.s. errors are smaller than in the case of unbiased estimates.

An expression for the optimal value of the reduction coefficient results in a minimal r.m.s. error.

Recurrent forms of reduced multi-step algorithms simplify identification in real time and are easily implemented in software.

References

1. *Tsypkin, Ya. Z.*, Adaptation and Learning in Automatic Systems. Moscow: Nauka, 1968 (in Russian).
2. *Avedyan, E. D.*, Modified Kaczmarz Algorithms for Estimating the Parameters of Linear Plants. Avtomatika i telemekhanika, 1978, No. *5*.
3. *Bodnya, V. G., Rurua, A. A., Chadeev, V. M.*, A Two-Step Algorithm for Identification of Linear Plants. Avtomatika i telemekhanika, 1982, No. *8*.
4. *Rudenko, O. G., Salyga, V. I.*, Multi-Step Identification Algorithms. Abstracts of Papers Presented to the All-Union Symposium on Identification of Non-Standard Plants in Instrumentation. Moscow, 1975 (in Russian).
5. *Rudenko, O. G., Salyga, V. I., Shtonda, A. N.*, A Multi-Step Adaptive Identification Algorithm. In: Control and Modeling of Metallurgic Processes. Moscow: Metallurgiya, 1980 (in Russian).
6. *Ishchenko, A. A., Rudenko, O. G.*, Multi-Step Algorithms for Identification of Linear Plants. Avtometriya, 1983, No. *6* (in Russian).
7. *Ishchenko, A. A., Liberol', V. D., Rudenko, O. G.*, Mapping Algorithms for Identification of Linear Plants. Doklady AN USSR, Ser. A 1985, No. *7* (in Russian).
8. *Seber, G. A.*, Linear regression analysis. (New York etc.) 1977.
9. *Demidenko, Ye. Z.*, Linear and Nonlinear Regression. Moscow: Finansy i statistika. 1981.
10. *Albert, A.*, Regression and the Moor-Penrose pseudoinverse. Academic Press, New York, 1972.

Редуцированные многошаговые алгоритмы идентификации линейных объектов

В. И. САЛЫГА, О. Г. РУДЕНКО, В. Л. ОБРУЧЕВ

(Москва)

Рассматривается многошаговый алгоритм идентификации, позволяющий путем учета информации о ряде предыдущих шагов добиться существенного сокращения времени построения модели при отсутствии помех измерений. Однако наличие помех приводит к тому, что с увеличением памяти алгоритма (числа учитываемых шагов) свойства алгоритма ухудшаются — увеличивается среднеквадратичная ошибка идентификации. В работе показано, что уменьшить среднеквадратичную ошибку можно, отказавшись от требования несмещенности оценок и используя идею построения редуцированных или укороченных оценок. Определены допустимые границы изменения параметра редукции, обеспечивающего меньшую среднеквадратичную ошибку

по сравнению с ошибкой, даваемой многошаговым алгоритмом. Рассмотрен вопрос определения оптимального значения параметра редукции, обеспечивающего максимальную скорость сходимости алгоритма. Значительное внимание уделено построению более удобных в вычислительном отношении рекуррентных форм редуцированного многошагового алгоритма.

При построении одной формы используется общая часть матрицы наблюдения, получающаяся при формировании оценки на двух соседних шагах. Во второй форме вместо непосредственного обращения матрицы наблюдения используется ее рекуррентное вычисление.

В. И. Салыга
Министерство высшего и среднего специального
образования СССР
СССР 113833 Москва М-230, ГСП,
Люсиновская ул., 51

Problems of Control and Information Theory, Vol. 17 (1), pp. 33—36 (1988)

A T-FAMILY EXTREMAL PROBLEM IN HAMMING SPACE

H. K. AYDINIAN

(*Yerevan*)

(Received December 30, 1986)

In this paper we consider (T, m, Δ)-code families and give the least upper bound of cardinality of such codes. Besides a class of optimal equidistant code families is given.

1. Introduction

Let E^m be the set of binary sequences of length m and d denote the Hamming metric in E^m. Then let we have a set of positive integers $\Delta = \{\delta_{ij}\}$; $\delta_{ij} = \delta_{ji} (i \neq j)$; $i, j = \overline{1, T}$ $(T \geqq 2)$.

The family $(A_1, \ldots, A_T)$; $A_1, \ldots, A_T \subset E^m$ is said to be a (T, m, Δ)-code family or (T, m, Δ)-system if for any $i, j \in \{1, \ldots, T\}$ $(i \neq j)$ and

$$\forall a^i \in A_i \forall a^j \in A_j d(a^i, a^j) = \delta_{ij}.$$

If $\delta_{ij} = \delta$ for each $i, j \in \{1, \ldots, T\}$ we shall call that system an equidistant code family and denote it by (T, m, δ).

Let us consider the function

$$M(T, m, \Delta) \triangleq \max_{\delta} \left\{ \prod_{u=1}^{T} |A_u| : (A_1, \ldots, A_T) \text{ is a } (T, m, \Delta)\text{-system} \right\}.$$

The study of (T, m, Δ)-systems and the function $M(T, m, \Delta)$ is motivated by the problem of lower boundary of the two-way complexity (in the sense of Yao [1]) of the Hamming distance function.

Ahlswede et al. proved (by induction on m) the following:

Theorem [2]. For each natural m

$$\max_{\delta} M(2, m, \delta) = 2^{2n}, \qquad \text{if} \quad m = 2n, \quad \text{or} \quad m = 2n+1.$$

The maximum assumed for $\delta = n$.

In [3] the authors propose to consider this problem for $T > 2$, however, the proof method used in [2] does not give us a solution of the problem. In this paper a solution of this problem for arbitrary T is given.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

2. The main result

Theorem. i) For any (T, m, Δ)-system one has the relation

$$M(T, m, \Delta) \leqq 2^m$$

ii) for $k = 1, 2, \ldots$; $T \leqq n \leqq 2T$; n is an arbitrary number, for which there exists an Hadamard matrix H_n; holds the equation

$$M(T, m = nk, m/2) = 2^m. \tag{1}$$

To prove this theorem, we use the next

Lemma.

$$M(T, 2n, n) \leqq 2^{2n}. \tag{2}$$

Proof. First we prove the lemma for $T = 2$.

Let (A_1, A_2) be a $(2, 2n, n)$-system. Replace 0 by -1 in the vectors of E^{2n}. Next, let us consider the sequences obtained from E^{2n} to be vectors of Euclidean space R^n. Denote by (B_1, B_2) the pair obtained from (A_1, A_2), respectively. Now note that the condition

$$d(a^1, a^2) = n \quad (a^1 \in A_1,\ a^2 \in A_2)$$

implies

$$\langle b^1, b^2 \rangle = 2n - 2d(a^1, a^2) = 0 \quad (b^1 \in B_1, b^2 \in B_2),$$

where $\langle b^1, b^2 \rangle$ is the scalar product of the vectors' b^1, b^2 corresponding to a^1 and a^2. Therefore, the linear envelopes L_1, L_2 of the B_1 and B_2 are mutually orthogonal planes of R^{2n}.

It is easy to see that $|B_1| \leqq 2^{\dim L_1}$, $|B_2| \leqq 2^{\dim L_2}$ and $\dim L_1 + \dim L_2 \leqq 2n$. Hence we obtain

$$|B_1|\,|B_2| \leqq 2^{2n}.$$

If now $T > 2$, the planes L_1 and L^* (L^* is the linear envelope of the $B_2 \cup \ldots \cup B_T$) are mutually orthogonal and we get the relation (2) inductively. Thus the lemma is proved.

Proof of the Theorem. Let $(A_1, \ldots, A_T)$ be a (T, m, Δ)-system and H_n is an Hadamard matrix of order $n \geqq T$. Denote by $\bar{a}$ $(a \in E^m)$ the complement of a, i.e. $\bar{a} = a + 1^m$ where 1^m is all-one vector, "$+$" means addition in the Hamming space.

If $A \subseteq E^m$ then $\bar{A} = \{\bar{a} : a \in A\}$.

Next choose a $T \times n$ submatrix (h_{uv}) of the H_n and define the $T \times n$ matrix (α_{uv}) as follows

$$\alpha_{uv} = \begin{cases} A_u, & \text{if } h_{uv} = 1 \\ \bar{A}_u, & \text{if } h_{uv} = -1, \end{cases}$$

where A_u ($u=\overline{1, T}$) is represented as a $|A_u| \times m$ matrix. Now define $\mathscr{A}_i$ ($i=\overline{1, T}$) as the Cartesian product $\alpha_{i1} \times \ldots \times \alpha_{in}$. As far as an Hadamard matrix H_n determines an equidistant $(n, n, d=n/2)$-code [4, chapter 2] it is clear that the family $(\mathscr{A}_1, \ldots, \mathscr{A}_T)$ is a $(T, mn, mn/2)$-equidistant code family. Using the lemma, we obtain

$$\prod_{i=1}^{T} |\mathscr{A}_i| = \prod_{i=1}^{T} |A_i|^n \leqq 2^{mn}.$$

Hence, we proved the first part of our theorem.

Now we shall give a construction of optimal equidistant code families (it means one has relation (1) for arbitrary T. Let we have an $(n, 2n, d=n/2)$-Hadamard code $\mathscr{C}_n$ [4]. Consider a partition of the $\mathscr{C}_n$ on t classes $\mathscr{C}_n = \bigcup_{i=1}^{t} A_i$, such that each class A_i ($i \in \{1, \ldots, t\}$) with an element a^i also contains its complement $\bar{a}^i$. It is clear that the family $(A_1, \ldots, A_t)$ is a $(t, n, n/2)$-system. In order that the $(A_1, \ldots, A_t)$ be optimal, we must have the following conditions:

$$\sum_{i=1}^{t} l_i = 2n$$

$$\prod_{i=1}^{t} l_i = 2^n = 2^{(l_1 + \ldots l_t)/2}$$

where $l_i = |A_i|$, $i=\overline{1, t}$.

Since for natural l one has $2^{l/2} = l$ only if $l=2$ or 4 we conclude that our construction is optimal iff $|A_i| = 2$ or 4. Next let us show that for arbitrary T there exists an Hadamard code $\mathscr{C}_n$, from which one can construct an optimal $(T, n, n/2)$-system. Obviously, for given natural T, there exists an Hadamard matrix H_n, where $T \leqq n \leqq 2T$. On the other hand, a partition of the Hadamard code $\mathscr{C}_n$ gives us (as mentioned above) an optimal $(T, n, n/2)$-system, if the following relations hold:

$$\begin{cases} 4t_1 + 2t_2 = 2n \\ t_1 + t_2 = T \end{cases}$$

where t_1, t_2 are non-negative integers. But these equations have non-negative solutions iff $T \leqq n \leqq 2T$. Therefore, we can get an optimal $(T, n, n/2)$-system for each $T \geqq 2$. Then it is clear that for any natural k, the family $(A_1^k, \ldots, A_T^k)$ is a $(T, m=nk, m/2)$-equidistant code family and

$$\prod_{i=1}^{T} |A_i^k| = \prod_{i=1}^{T} |A_i|^k = 2^{kn} = 2^m.$$

Thus the proof of the Theorem is complete.

Let us give now an example of an optimal $(3, 4k, 2k)$ system.

$$\mathscr{C}_4 = \{0000,\ 1111,\ 1100,\ 0011,\ 0110,\ 1001,\ 0101,\ 1010\}$$

$$A_1 = \{0000,\ 1111\}^k, \quad A_2 = \{1010,\ 0101\}^k, \quad A_3 = \{1100,\ 0011,\ 0110,\ 1001\}^k,$$

$$|A_1||A_2||A_3| = 2^{4k}.$$

References

1. *Yao, A. C.*, Some complexity questions related to distributive computing. 11th ACM Symp. on Theory of Computing (1979), pp. 209–213.
2. *Ahlswede, R., Gamal, A. El, Pang, K. F.*, A two family extremal problem in Hamming space. Discrete Math., **49** (1984), pp. 1–5.
3. *Ahlswede, R., Mörs, M.*, Inequalities for code pair. Submitted to European J. Combinatorics.
4. *MacWilliams, F. J., Sloane, N. J. A.*, The Theory of Error Correcting Codes. North Holland, Amsterdam, 1977.

Экстремальная задача T-семейств кодов в пространстве Хэмминга

А. К. АЙДИНЯН

(Ереван)

Пусть E^m — множество бинарных последовательностей длины m, а d — расстояние Хэмминга в E^m. Пусть далее имеем множество натуральных чисел $\Delta = \{\delta_{ij}\}$; $\delta_{ij} = \delta_{ji}$, $i \neq j$; $i, j = \overline{1, T}$ $(T \geqq 2)$.

Семейство непустых множеств $(A_1, \ldots, A_T)$; $A_1, \ldots, A_T \subset E^m$ назовем (T, m, Δ)-семейством кодов, или (T, m, Δ)-системой, если при любых $i, j \in \{1, \ldots, T\}$ $(i \neq j)$ и

$$\forall a^i \in A_i\ \forall\, a^i \in A_j d(a^i, a^j) = \delta_{ij}.$$

Рассмотрим функцию

$$M(T, m, \Delta) \triangleq \max \left\{ \prod_{u=1}^{T} |A_u| : (A_1, \ldots, A_T) \text{ — есть } (T, m, \Delta)\text{-система} \right\}.$$

Изучение (T, m, Δ)-систем и функции $M(T, m, \Delta)$ связано с задачами нахождения сложностных оценок при распределенном вычислении [1].

В работе [2] показано (индукцией по m), что при любом $m \in N$

$$\max_{\delta} M(2, m, \delta) = 2^{2n}, \text{ при } m = 2n \text{ или } m = 2n + 1.$$

В другой работе [3], касающейся также $(2, m, \delta)$-систем, авторами предложено решить аналогичную задачу для $T > 2$. В данной статье, используя элементарные свойства пространства E^m, эта задача решена для произвольного T.

А. К. Айдинян
ВЦ Арм. ССР
СССР, 375044 Ереван 44,
ул. П. Севака, 1

Problems of Control and Information Theory, Vol. 17 (1), pp. 37—47 (1988)

MINIMAX FUNCTION APPROXIMATION BY NOISY OBSERVATIONS UNDER NONPARAMETRIC UNCERTAINTY

P. P. GUSAK, L. R. SORKIN

(*Moscow*)

(Received July 18, 1986)

This paper suggests two approaches to the problem of minimax function approximation by noisy observations under nonparametric uncertainty.

Both approaches imply lack of apriori knowledge on the structure of the function to be approximated using only some of its qualitative properties such as its Lipschitzian nature, constrained nonlinearity, etc. In general, the suggested statement is a certain generalization of problems of parametric uncertainty which are solved with the use of classical regression techniques. The proposed method is intended to regularize the solution of an ill-stated approximation problem.

The approaches considered by the authors differ in the nature of the qualitative assumptions in the system, the metric used, and the data processing techniques applied. The first approach is oriented toward on-line observation processing, while the second one is based on data accumulation.

Application of mathematical models to control system design is often hindered by the following major reasons: on one hand, the behaviour of a on-line control plant is always affected by a number of inaccountable factors leading to an uncertainty in the mathematical description that is essentially unremovable; on the other hand, computational considerations sometimes do not allow one to deal with nonlinearities and high dimension of the models.

In the first case, we are forced to use inadequate description because of incompleteness of our knowledge on the functional connections and numerical values of parameters, while in the second case inadequacy results from our deliberate simplification of the model.

In both cases, however, the researcher should deal with a different problem whose solution is supposed to substitute that of the original problem.

In this situation the following seems to be of paramount importance:

— the guaranteed boundaries of the control system performance when using an approximate model (the problem of analysis), and

— the way to choose a subset from the set of approximate models for which the guaranteed control performance meets the pre-specified requirements (the problem of synthesis).

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

This paper suggests two approaches to the above problems.

Both approaches imply lack of apriori knowledge on the structure of the function to be approximated using only some of its qualitative features such as its Lipschitzian nature, constrained nonlinearity, etc. In general, the suggested statement is a certain generalization of problems of parametric uncertainty which are solved with the use of classical regression techniques. The proposed method is intended to regularize the solution of an ill-stated approximation problem.

The approaches considered by the authors differ in the nature of qualitative assumptions in the system, the metric used, and the data processing techniques applied. The first approach is oriented toward on-line observation processing, while the second one is based on data accumulation.

Under the assumed uncertainty, the minimax solution to the problem seems most natural.

Introduce the following notation:

$x_t \in \mathbf{R}^l$ is an approximated function which belongs to some known nonparametric class $\mathcal{N}$, $x_t \in \mathcal{N}$;

$\hat{x}_t \in \mathbf{R}^l$ is the approximating function, $\hat{x}_t \in \mathcal{P}$, where $\mathcal{P}$ is the given parameter class;

$\rho(x_t, \hat{x}_t)$ is the distance, i.e. the error of approximation.

Let us call $\hat{x}_t^*$ the minimax approximation of x_t if

$$\hat{x}_t^* = \arg\min_{\hat{x}_t \in \mathcal{P}} \max_{x_t \in \mathcal{N}} \rho(x_t, \hat{x}_t), \quad t \in T. \tag{1}$$

The steps of solving the stated problem are

1°. Estimate $\hat{\rho}(\hat{x}_t) \triangleq \max_{x_t \in \mathcal{N}} \rho(x_t, \hat{x}_t)$ an internal problem.

2°. Estimate $\hat{x}_t^* = \arg\min_{\hat{x}_t \in \mathcal{P}} \hat{\rho}(\hat{x}_t)$ an external problem.

Approach I

The function may be specified in different ways, in particular, the differential equation apparatus may be used for the purpose. In this case, let $\mathcal{N}$ be a class of approximated functions x_t specified by the differential equation

$$\begin{aligned} &dx_t + f_t(x_t)dt = w_t dt, \\ &Ex_0 = \bar{x}_0; \qquad \operatorname{cov} x_0 = \bar{P}_0 \end{aligned} \tag{2}$$

where $f(\cdot)$ is an unknown continuous Lipschitz operator specified by its belonging to the class

$$f_t \in \hbar_{A_t, a_t} \triangleq \{f: \|f(z) - A_t z\| \leqq a_t, \quad \forall z \in Z; \quad t \in T\}, \tag{3}$$

where $\|l\| \triangleq l^T l$, Z is the definition domain of $f(\cdot)$, $w_t \in \mathbf{R}^l$ is a known function, and A_t and a_t are a given matrix and scalar, respectively.

The model of observations is given by the equation

$$dy_t = H_t x_t dt + D_t d\zeta_t, \tag{4}$$

where ζ_t is a Wiener process.

The elements of matrices A_t, H_t and D_t are limited and continuous, D_t being a full rank matrix. The class of functions (3) was introduced in [1].

Let the approximating functions class $\mathscr{P}$ be given by a set of differential equations of the form

$$d\hat{x}_t + A_t \hat{x}_t dt = G_t du_t + w_t dt, \quad \hat{x}_0 = \bar{x}_0 \tag{5}$$

parametrized with respect to matrix G_t. The notation du_t is understood as

$$du_t \triangleq dy_t - H_t \hat{x}_t dt. \tag{6}$$

Thus, the unknown function x_t is approximated in the control space by choosing an appropriate control $G_t du_t$.

The error of such an approximation $\rho(x_t, \hat{x}_t)$ is defined as

$$\rho(x_t, \hat{x}_t) \triangleq E(x_t - \hat{x}_t)(x_t - \hat{x}_t)^T.$$

The design of the optimal approximation could rest upon the knowledge of matrix ρ_t, but since the assumed uncertainty makes this knowledge inaccessible, it seems expedient to seek for a majorant matrix* $\hat{\rho}_t$ such that $\hat{\rho}_t \geqq \rho_t$, $t \in T$ uniform for the entire class $\mathscr{N}$ and depending on the control action or, to be more specific, on the matrix-parameter G_t. Evidently, the value $\hat{\rho}_t$ being a guaranteed estimate of the approximation quality gives a solution to the internal problem.

Introduce the following differential equation

$$\frac{d\hat{\rho}_t}{dt} = -\left(A_t - \frac{a_t}{2} I + G_t H_t\right)\hat{\rho}_t - \hat{\rho}_t\left(A_t - \frac{a_t}{2} I + G_t H_t\right)^T + \\ + a_t I + G_t D_t D_t^T G_t^T, \quad \hat{\rho}_0 = \bar{P}. \tag{7}$$

* The inequality between symmetric matrices Ω_1 and Ω_2 means that $\Omega_1 \leqq \Omega_2$ is equivalent to some nonnegative defined matrix $(\Omega_2 - \Omega_1)$.

Theorem 1. For an arbitrary matrix G_t, $t \in T$, the solution of eq. (7) gives an upper boundary $\hat{\rho}_t$ for the second moment errors matrix ρ_t, i.e. $\rho_t \leqq \hat{\rho}_t$.

A proof of this theorem and Theorems 2 and 3 is given in the Appendix. Note that the right-hand part of (7) is convexly (quadratically) dependent upon matrix G_t and, consequently, a matrix G_t^* exists for any fixed time $t \in T$ such that it realizes, in a certain sense, the minimal property of matrix $\frac{d}{dt}\hat{\rho}_t$ which meaningfully corresponds to the "closest" clinging of the upper boundary $\hat{\rho}_t$ to matrix ρ_t at any fixed time t. This, evidently, solves the external problem.

Theorem 2. For any fixed $t \in T$ and an arbitrary $\tilde{G}_t$ the following inequality is true:

$$\frac{d}{dt}\hat{\rho}_t(G_t^*) \leqq \frac{d}{dt}\hat{\rho}_t(\tilde{G}_t), \tag{8}$$

where $G_t^* \triangleq \hat{\rho}_t H_t (D_t D_t^T)^{-1}$.

An appropriate choice of $G_t = G_t^*$ allows one to find the minimal increment $d\hat{\rho}_t^*$ guaranteed for any $\tilde{G}_t$.

Using the obtained results, the unknown element $\hat{x}_t$ from the approximating class $\mathscr{P}$ may be presented in the form of a solution to the following differential equation

$$d\hat{x}_t + [A_t + \hat{\rho}_t H_t (D_t D_t^T) H_t]\hat{x}_t dt = \hat{\rho}_t H_t (D_t D_t^T)^{-1} dy_t + w_t dt. \tag{9}$$

The robustness of the obtained estimate is obviously correlated with the diameter of class $\mathscr{N}$ specified by the parameter $\hat{a} \triangleq \max_{\tau \leqq t} a_\tau$. Ideally, when $\hat{a}_t \equiv 0$, class $\mathscr{N}$ consists of only one element and the upper boundary $\hat{\rho}_t$ of the second moments matrix coincides with the covariance matrix ρ_t while G_t together with eq. (2) sets an r.m.s. optimal approximation of function x_t at each instant of time t [2].

Let us try to estimate the estimation error changing rate with the decrease of $\hat{a}_t$. The following theorem is true:

Theorem 3. For any $t \in T$,

$$\|\hat{\rho}_t - \rho_t^0\| \leqq L_t \hat{a}_t, \tag{10}$$

where $\rho_t^0 \triangleq \rho_t|_{\hat{a}_t \equiv 0}$, and L_t is a function independent of $\hat{a}_t$.

The last result allows one to make a conclusion on a continuous correspondence between the second moments matrix and the diameter of class $\mathscr{N}$ in point $\hat{a}_t = 0$, i.e. on the fact that small magnitudes of parameter $\hat{a}_t$ correspond to small deviations of the suggested approximation of $\hat{x}_t$ from the optimal one.

Although this result gives certain idea on the roughness of the suggested approximation its value is somewhat reduced due to the time dependence of the

coefficient in the right-hand part of inequality (10). However, a specification of conditions for t-uniform satisfaction of inequality (10) may strengthen this result which also establishes the fact of t-uniform boundness (dissipation) of the second moments matrix upper bound.

Theorem 4. Let the following conditions be satisfied

1°. $\operatorname{Re} \lambda_i\left(A_t - \frac{a_t}{2} I + G_t H_t\right) > 0, \quad t \in T, \quad i = 1, \ldots, l.$

2°. A pair of matrices $\left[A_t - \frac{a_t}{2} I + G_t H_t, S_t\right]$ is uniformly observable where S_t is a matrix such that

$$S^T S = a_t I + G_t D_t D_t^T G_t^T.$$

Then eq. (7) has a t-uniform bounded solution.

The proof of this assertion is given in [3].

Concluding this section the following is noted. The classical theorem on the continuous dependence of the solution on the parameter is known to establish a qualitative characteristic of the differential equation behaviour under parametric uncertainty of the right-hand part of the equation.

The results suggested are attempted to obtain a relationship between the solution of the differential equation and the uncertainty measure in the nonparametric case. An established fact here is that a small value of diameter of nonparametric class $\mathcal{N}$ corresponds to a small change of L_2-quality of linear approximation of the reference equation solution.

Approach II

Let now $\mathcal{N}$, $\mathcal{P}$ and ρ be given in a different manner.

Class $\mathcal{N}$ consists of scalar functions satisfying the Lipschitz condition such that $\forall t, t+\tau \in T$

$$\|x_t - x_{t+\tau}\| \leqq M\tau.$$

Class $\mathcal{P}$ is a set of monotonic and concave (convex) functions $\{\hat{x}_t(\alpha), t \in T, \alpha \in \mathfrak{A}\}$, α is the unknown vector valued parameter, and $\mathfrak{A}$ is the given set.

Such a statement of classes $\mathcal{N}$ and $\mathcal{P}$ features both the mathematical tradition [4] and a meaningful background. Thus, the constant M implies a limit rate of changing x_t; the qualitative properties of functions such as monotonicity and concavity are often physically induced and provide certain mathematical convenience [5].

Let us make use of the following observation model:

$$y_t = \chi_{[t \in \mathring{T}]} x_t + \xi_t,$$

ξ_t being the noise with a finite distribution density, i.e. $|\zeta_t| \leqq \delta$ a.s., and the value δ being known. Find the quality of approximation ρ_t in the following way:

$$\rho(x_t, \hat{x}_t) \triangleq \max_{t \in T} |x_t - \hat{x}_t|.$$

Knowing $\mathring{T}$, M and δ, one may construct a polyhedron Ω (Fig. 1) incorporating the domain of feasible values of x_t. Consider those of its vertices $\omega^i \in \Omega$ which belong to the minimal convex hull ω and denote as x_ω those of $x_t \in \mathcal{N}$ whose value domain includes at least one ω^i. The following theorem gives a solution to the internal problem.

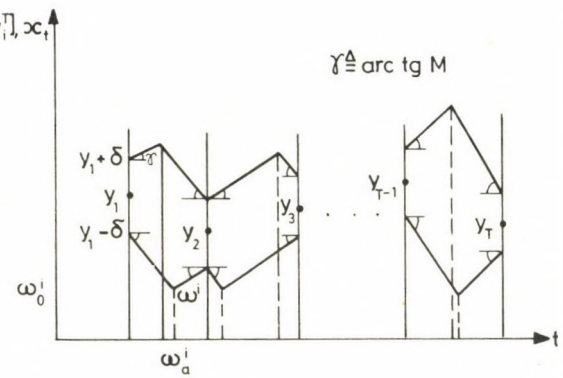


Fig. 1

Theorem 5

$$\forall \hat{x}_t \in \mathscr{P}, \quad x_t \in \mathcal{N}, \quad \exists x_\omega \in \mathcal{N}$$

so that

$$\rho(x_\omega, \hat{x}_t) \geqq \rho(x_t, \hat{x}_t).$$

The minimax approximation is given by the following theorem on the external problem.

Theorem 6. The solution of the external problem is a Chebyshev approximation of vertices $\omega^i \in \omega$ in class $\mathscr{P}$.

If we denote the ordinates and the abscissas of these vertices as ω_Y^i and ω_X^i, respectively, this approximation may be found by solving the following problem:

$$\alpha^* = \arg\min_{\alpha \in \mathfrak{A}} \max_{i = \overline{-1, I}} |\omega_Y^i - \hat{x}(\alpha, \omega_X^i)|, \tag{11}$$

I being the number of vertices ω^i. Figure 1 directly verifies the validity of theorems 5 and 6.

Problem (11) is reduced to the equivalent problem of linear programming [6] if function $\hat{x}_t(\alpha)$ is linear with respect to α and $\mathfrak{A}$ is a convex polyhedron:

$$\alpha^* = \arg\min_{\alpha \in \mathfrak{A}} Z,$$

$$\omega_Y^i - \hat{x}(\alpha, \omega_X^i) \leqq Z, \qquad i = \overline{1, I}, \tag{12}$$

$$\hat{x}(\alpha, \omega_X^i) - \omega_Y^i \leqq Z, \qquad i = 1, I,$$

where Z is an artificial variable.

Problem (12) may be solved both by standard techniques and by ad hoc discrete Chebyshev approximation algorithms [6].

Note that a nonparametric way of statement may be extended toward function $\hat{x}_t$ by analogy with [7]. Stating $\hat{x}_t$ by a class of monofinic concave (convex) linear upper and/or lower bounded functions, $\hat{x}_t^*$ may be found by solving a problem of linear programming similar to (9). In this case, x_t^* is a piecewise-linear function with nodes at $t = \omega_X^i$ [8].

Thus it turns out that in order to obtain a minimax (i.e. guaranteed) Chebyshev approximation one must extend the initial set of observations. The coordinates of points added with this extention depend (Fig. 1) on the initial set, the Lipschitz constant M and the noise level δ. The Chebyshev approximation of the extended set is minimax.

Note in conclusion that the suggested results may prove useful for those application cases when the physical object does not allow any assumptions to be made on the structure of the relationship under study or when this structure is too complicated to employ the standard approximation techniques.

The obtained results are compact in their algorithmic realization and do not require much laboursome computation.

Appendix (Proofs)

Proof of Theorem 1. Introduce the notation $\tilde{x}_t \triangleq x_t - \hat{x}_t$; then, subtracting (5) from (2) we obtain

$$d\tilde{x}_t + [f_t(x_t) - A_t\hat{x}_t]dt = -G_t(H_t x_t dt - dw_t + D_t d\xi_t),$$

$$E\tilde{x}_0 = 0, \tag{A.1}$$

$$\operatorname{cov} \tilde{x}_0 = \bar{P}_0.$$

Elementary manipulation yields

$$d\tilde{x}_t + [\tilde{f}_t + (A_t + G_t H_t)\tilde{x}_t]dt = G_t D_t d\zeta_t, \tag{A.2}$$

where $\tilde{f}_t \triangleq f_t(x_t) - A_t x_t$ and, as follows from condition (3), $\|\tilde{f}_t\| \leqq a_t$.

The proximity measure $\rho_t \triangleq E\tilde{x}_t\tilde{x}_t^T$ may be specified with the help of Ito's formula in the form of the following differential equation:

$$d\rho_t = E(\tilde{x}_t dx_t + d\tilde{x}_t\tilde{x}_t^T + d\tilde{x}_t d\tilde{x}_t^T). \tag{A.3}$$

In terms of the derivatives, and with regard for (A.2), this equation takes up the form

$$\frac{d}{dt}\rho_t = G_t D_t D_t^T G_t^T - E[\tilde{x}_t\tilde{f}_t^T + \tilde{f}_t\tilde{x}_t^T] - \rho_t(A_t + G_t H_t)^T - (A_t + G_t H_t)\rho_t. \tag{A.4}$$

To estimate the term in brackets in the right-hand part of (A.4), we may use the following relationship

$$E(\tilde{x}_t\tilde{f}_t^T + \tilde{f}_t\tilde{x}_t^T) \leqq a_t(\rho_t + I) \tag{A.5}$$

whose validity results from elementary manipulation over the obvious inequality

$$E(a_t\tilde{x}_t - \tilde{f}_t)(a_t\tilde{x}_t - \tilde{f}_t)^T \geqq 0$$

recognizing the fact that $E\tilde{f}_t\tilde{f}_t^T \leqq a_t^2 I$ in accordance with condition (3).

(A.4) and (A.5) allow one to easily obtain the following differential inequality with respect to ρ_t:

$$\frac{d\rho_t}{dt} \leqq -\left[\left(A_t - \frac{a_t}{2}I\right) + G_t H_t\right]\rho_t - \rho_t\left[\left(A_t - \frac{a_t}{2}I\right) + G_t H_t\right] + a_t I + G_t D_t D_t^T G_t^T,$$

$$\rho_0 = \bar{P}_0$$

or the differential equation

$$\frac{d\rho_t}{dt} = -\left[\left(A_t - \frac{a_t}{2}I\right) + G_t H_t\right]\rho_t - \rho_t\left[\left(A_t - \frac{a_t}{2}I\right) + G_t H_t\right]^T + a_t I + G_t D_t D_t^T G_t^T - \beta_t,$$

$$\rho_0 = \bar{P}_0, \tag{A.6}$$

where β_t is a nonnegative defined matrix.

Subtraction of equation (A.6) from equation (7) yields

$$\frac{d\rho_t}{dt} = -\left[\left(A_t - \frac{a_t}{2}\right)I + G_t H_t\right]\tilde{\rho}_t - \tilde{\rho}_t\left[\left(A_t - \frac{a_t}{2}I\right] + G_t H_t\right]^T + \beta_t,$$

$$\tilde{\rho}_0 = 0, \tag{A.7}$$

where $\tilde{\rho}_t \triangleq \hat{\rho}_t - \rho_t$.

The assertion of Theorem 1 is obviously equivalent to the nonnegative certainty of the solution to equation (A.7) while the latter results from the following lemma.

Lemma. For equation

$$\frac{dP_t}{dt} = \alpha_t P_t + P_t\alpha_t^T + \beta_t, \qquad P_0 = \bar{P},$$

where $\bar{P}_t$ is a symmetric nonnegative defined matrix and β_t, a nonnegative defined matrix, the following holds: $\bar{P}_t$ is nonnegative defined, $t \in [0, T]$.

Proof of the Lemma. a) Consider the uniform equation and make sure that its solution is obtainable in the form

$$\bar{P}_t = \Phi_{0,t}\bar{P}\Phi_{0,t}^T,$$

where $\Phi_{0,t}$ is a fundamental matrix such that

$$\frac{d}{dt}\Phi_{0,t} = \alpha_t\Phi_{0,t}; \qquad \Phi_{t,t} = I.$$

Thus indeed is so, since

$$\frac{dP_t}{dt} = \alpha_t \Phi_{0,t} \bar{P} \Phi_{0,t}^T + \Phi_{0,t} \bar{P} \Phi_{0,t}^T \alpha_t^T = \alpha_t \bar{P}_t + \bar{P}_t \alpha_t^T,$$

b) a particular solution to the nonuniform equation $\frac{d}{dt} P_t = \alpha P_t + P_t \alpha_t^T + \beta_t$ will be found with the help of the variation technique for $\bar{P}$, namely

$$\frac{d}{dt} P_t = (\Phi_{0,t} \bar{P} \Phi_{0,t})^T = \alpha_t \Phi_{0,t} \bar{P} \Phi_{0,t}^T + \Phi_{0,t} \bar{P} \Phi_{0,t}^T \alpha_t^T + \Phi_{0,t} \frac{d\bar{P}}{dt} \Phi_{0,t}^T.$$

Substituting this expression into the lefthand part of the initial equation and making the necessary manipulation, we obtain

$$\frac{d\bar{P}}{dt} = \Phi_{0,t}^{-1} \beta_t (\Phi_{0,t}^T)^{-1},$$

or

$$\bar{P} = \int_0^t \Phi_{0,s}^{-1} \beta_s (\Phi_{0,s}^{-1})^T ds$$

and hence it follows that a particular solution $\tilde{P}_t$ of the initial equation may be found as

$$\tilde{P}_t = \int_0^t \Phi_{0,t} \Phi_{0,s}^{-1} \beta_s (\Phi_{0,s}^T)^{-1} \Phi_{0,t}^T ds = \int_0^t \Phi_{s,t} \beta_s \Phi_{s,t}^T ds.$$

The general solution will accordingly take up the form

$$P_t = \bar{P}_t + \tilde{P}_t = \Phi_{0,t} \bar{P} \Phi_{0,t}^T + \int_0^t \Phi_{s,t} \beta_s \Phi_{s,t}^T ds. \tag{A.8}$$

It is obvious that $P_t \geqq 0$ by virtue of the assumption $\bar{P} \geqq 0$ and $\beta_s \geqq 0$, $0 \leqq t \leqq \infty$.
The lemma is proved.
Theorem 1 is proved.
Proof of Theorem 2. Consider the matrix polynomial

$$Q_t(G_t) \triangleq -G_t H_t \hat{\rho}_t - \hat{\rho}_t H_t^T G_t^T + G_t D_t D_t^T G_t^T, \tag{A.9}$$

which coincides with the right-hand part of equation (6) with an accuracy to G_t-independent terms.
Let us make sure that the following is true of an arbitrary vector q of the corresponding dimension:

$$q^T Q_t(G_t^*) q = \min_{G_t} q^T Q_t(G_t) q, \qquad 0 \leqq t < \infty.$$

To see this, we may modify (A.8) to obtain

$$q^T Q_t(G_t) q = -q^T G_t H_t \hat{\rho}_t q - q^T \hat{\rho}_t H_t G_t^T q + q^T G_t D_t D_t^T G_t^T q =$$
$$= -s_t^T H_t \rho_t q - q^T \hat{\rho}_t H_t^T S_t + S_t^T D_t D_t^T S_t,$$

where $s_t \triangleq G_t^T q$.
Zeroing the first variation in S_t we have

$$-H_t \hat{\rho}_t q - q^T \hat{\rho}_t H_t^T + 2 D_t D_t^T G_t^T q = 0,$$

wherefrom it follows that

$$G_t^* = \hat{\rho}_t H_t (D_t D_t^T)^{-1}.$$

Theorem 2 is proved.

Proof of Theorem 3. Assuming $a_t \equiv 0$ in equation (7) and subtracting this from (7) we have

$$\frac{d}{dt}\tilde{\rho}_t = -(A_t + G_t H_t)\tilde{\rho}_t - \tilde{\rho}_t(A_t + G_t H_t)^T + a_t(I + \hat{\rho}_t), \qquad \tilde{\rho}_0 = 0$$

where $\tilde{\rho}_t \triangleq \hat{\rho}_t - \mathring{\rho}_t$.

Using the result of the lemma, in particular, formula (A.8), we obtain

$$\tilde{\rho}_t = \int_0^t \Phi_{s,t} a_s (I + \hat{\rho}_s) \Phi_{s,t}^T ds \leqq \hat{a}_t \int_0^t \Phi_{s,t}(I + \hat{\rho}) \Phi_{s,t}^T ds,$$

which yields that $\|\tilde{\rho}_t\| \leqq L_t \hat{a}_t$, where

$$L_t \triangleq \left\| \int_0^t \Phi_{s,t}(I + \hat{\rho}_s) \Phi_{s,t}^T ds \right\|$$

Theorem 3 is proved.

References

1. *Gusak, P. P.*, Autom. and Remote Control, vol. **42,** (1981), No. *4*, pp. 1–466.
2. *Kalman, R. E.*, J. Basic Eng. Trans. **82**D (1960), 33.
3. *Anderson, B. D. O., Moore, J. B.*, SIAM J. Contr., vol. **7** (1961), *3*, p. 398.
4. *Nikol'skii, S. M.*, On the best approximation of the Lipschitz condition-satisfying functions. Izv. AN SSSR, Mathematical Series, vol. **10** (1961), pp. 295–332.
5. *Itzkovich, E. L., Sorkin, L. R.*, The choice of equipment type models in solving the linear problem of static optimization for a continuous production. Avtomatika i Telemekhanika, (1984), No. *8*, pp. 142–149.
6. *Arthanari, T. S., Dodge, Y.*, Mathematical programming in statistics. J. Wiley & Sons, New York, 1981.
7. *Nemirovskii, A. S., Polyak, B. T., Tzybakov, A. B.*, Estimates of the maximum likelihood type for nonparametric regression. Doklady AN SSSR, vol. **273** (1983), No. *6*, pp. 1310–1314.
8. *Sorkin, L. R.*, A guaranteed approach to problems of nonparametric identification of control plants statics. Izvestiya AN SSSR, Technical Cybernetics (1985), No. *3*, pp. 204–207.

Минимаксное приближение функции по зашумленным наблюдениям в условиях непараметрической неопределенности

П. П. ГУСАК, Л. Р. СОРКИН
(Москва)

В данной работе предлагаются два подхода к решению задачи минимаксного приближения функций в условиях непараметрической неопределенности.

Оба подхода предполагают отсутствие априорной информации о структуре аппроксимируемой функции, используя лишь некоторые качественные характеристики функции типа липшицевость, ограниченная нелинейность и др. В целом предлагаемая постановка является некоторым обобщением по отношению к задачам с параметрической неопределенностью, где используются классические регрессионные методы, и служит для регуляризации решения задачи аппроксимации по некорректному условию.

Рассматриваемые здесь подходы при этом различаются характером качественных предположений о системе, принятой метрикой, а также способом обработки данных. Первый подход ориентирован на обработку наблюдений в режиме реального времени, в то время как второй базируется на накопленных данных.

В условиях предположенной неопределенности представляется естественным минимаксный принцип решения задачи.

П. П. Гусак, Л. Р. Соркин
Институт проблем управления
СССР 117342 Москва, ГСП-7,
Профсоюзная ул., 65

Problems of Control and Information Theory, Vol. 17 (1), pp. 49—61 (1988)

GENERALIZED HERMITE POLYNOMIALS AND ESTIMATION OF KERNELS FOR DISCRETE *I/O* WIENER MODELS

G. TERDIK

(*Debrecen*)

(Received January 10, 1987)

In this paper we generalize the classical Hermite polynomials to give a complete polynomial system in the Hilbert space of stationary processes measurable with respect to a Gaussian stationary input. By the help of the multiple Wiener–Ito integral the spectral representation for the GHAW polynomials of a Gaussian stationary series is given.

The statistical examination and the identification of linear stationary models has been studied by several authors, Bokor and Keviczky (1982), Arató (1982). Recently, there has been an increasing interest in nonlinear models, studied both in the theory, Schetzen (1980), and in the applications, Várlaki and Terdik (1985).

The paper discusses the square integrable regular functionals of a stationary Gaussian series. The discrete Wiener models with stationary Gaussian input are given as a subordinated series. The estimation of the coefficients of the Wiener model with a Gaussian AR input is constructed. Using the central limit theorem for the functionals of strong mixing sequences we prove the asymptotic normality of the estimators.

1. Generalized Hermite–Appel–Wick polynomials

The classical Hermite polynomials form a complete orthonormal system in the Hilbert space of all square integrable random variables measurable with respect to a standard Gaussian one. The Hermite polynomials with parameter (μ, σ^2) are defined by differentiating the nonstandard Gaussian density function, Szegő (1948), Hida (1970). This method can be generalized, Grad (1949), to get the N-dimensional Hermite polynomial system being a complete orthogonal one in the Hilbert space generated by N independent standard Gaussian random variables. In case of arbitrary Gaussian vector this method does not work. The Appel polynomials which are not else than the Hermite polynomials with parameters (μ, σ^2), were defined in three rules by Shutterly (1963), Campbell (1964), Bonnet (1964). We generalize this method to give a complete polynomial system in the Hilbert space of N jointly Gaussian random variables. Polynomials of this type were mentioned by Ibragimov, Rozanov (1970, p. 28) and will be referred to as the Generalized Hermite–Appel–Wick (GHAW) polynomials. The Wick polynomials, Major (1982), as well as the Q-polynomials, Schetzen (1980), are special cases of the GHAW polynomials.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

Let $X_1, X_2, \ldots, X_n$ be jointly Gaussian random variables with $EX_i = 0$ and covariance matrix $(C_{X_iX_j})$, $i, j = 1, \ldots, n$. Then the GHAW polynomials are defined by the following way

(a) $A_0 = 1$

(b) $\dfrac{\partial}{\partial X_i} A_k(X_1, \ldots, X_k) = A_{k-1}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_k)$ $\quad i = 1, 2, \ldots, k;$ $\quad k = 1, \ldots, n.$

(c) $EA_k(X_1, \ldots, X_k) = 0$ $\quad k = 1, 2, \ldots, n.$

Some remarks (see Terdik (1985))

1. The first three GHAW polynomials are given by the formulae

$$A_0 = 1; \quad A_1(X) = X; \quad A_2(X_1, X_2) = X_1X_2 - C_{X_1X_2};$$

$$A_3(X_1, X_2, X_3) = X_1X_2X_3 - C_{X_1X_2}X_3 - C_{X_2X_3}X_1 - C_{X_3X_2}X.$$

2. $X_1 = X_2 = X_n = X$ then $A_n(X, \ldots, X) = A_n(X)$ where $A_n(X)$ is the nth Hermite polynomial with parameters $(0, \sigma^2)$ given by the recursive formula

$$A_{n+1}(X) = XA_{n-1}(X) - n\sigma^2 A_{n-2}(X); \quad A_0 = 1; \quad A_{-1} = 0.$$

3. The polynomial $A_n(\mathbf{X}) = A_n(X_1, \ldots, X_n)$ depends on the covariances $C_{X_iX_j}$; $i, j = 1, 2, \ldots, n$.

Let $A_{\mathbf{k}}(\mathbf{X})$, $\mathbf{X} = (X_1, \ldots, X_n)$, $\mathbf{k} = (k_1, \ldots, k_n) \in \mathbf{Z}^n$, $\mathbf{k} \geqq 0$ denote the GHAW polynomial $A_{|\mathbf{k}|}\left(\underbrace{X_1 \ldots X_1}_{k_1}, \ldots, \underbrace{X_n, \ldots, X_n}_{k_n}\right)$, $|\mathbf{k}| = \Sigma k_i$. If $X_1, \ldots, X_n$ are independent then

$$A_{\mathbf{k}}(\mathbf{X}) = \prod_i A_{k_i}(X_i) = Q_{\mathbf{k}}(\mathbf{X}).$$

The product of independent Hermite polynomials are called by Schetzen (1980) Q-polynomials. One can get the GHAW polynomials by the recursive formula

$$A_k(X_1, \ldots, X_k) = X_m A_{k-1}(X_1, \ldots, X_{m-1}, X_{m+1}, \ldots, X_k) -$$

$$- \sum_{i \neq m} C_{X_iX_m} A_{k-2}(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_{m-1}, X_{m+1}, \ldots, X_k).$$

The GHAW polynomials of order n have the generating function

$$G(\mathbf{X}, \mathbf{t}) = \sum_{\substack{\mathbf{k} \geqq 0 \\ \mathbf{k} \in \mathbf{Z}^n}} (i)^{|\mathbf{k}|} \frac{\mathbf{t}^{\mathbf{k}}}{\mathbf{k}!} A_{\mathbf{k}}(\mathbf{X}) = \frac{e^{i(\mathbf{X}, \mathbf{t})}}{\varphi_{\mathbf{X}}(\mathbf{t})}$$

where $\varphi_{\mathbf{X}}(\mathbf{t})$ is the characteristic function of $\mathbf{X}$. From this it follows that

$$A_{\mathbf{k}}(\mathbf{X})=(-i)^{|\mathbf{k}|}\left[\left(\frac{\partial}{\partial \mathbf{t}}\right)^{\mathbf{k}} G(\mathbf{X},\mathbf{t})\right]_{\mathbf{t}=0}.$$

It is easy to show that the following properties of GHAW polynomials hold

$$A_{\mathbf{k}}(X_1,\ldots,\alpha X_i,\ldots,X_n)=\alpha^{k_i}A_{\mathbf{k}}(\mathbf{X}), \tag{1.1}$$

$$A_{\mathbf{k}}(X_1,\ldots,X_i^{(1)}+X_i^{(2)},\ldots,X_n)= \tag{1.2}$$

$$\sum_{l=0}^{k_i}\binom{k_i}{l}A_{k_1,\ldots,l,k_i-l,\ldots,k_n}(X_1,\ldots,X_i^{(1)},X_i^{(2)},\ldots,X_n).$$

Let $z_1,\ldots,z_n,X_1,\ldots,X_m$ be jointly Gaussian variables with mean zero and covariances $C_{z_ix_j}$. The covariance of GHAW polynomials $A_n(\mathbf{z})$ and $A_m(\mathbf{X})$ is

$$EA_n(\mathbf{z})A_m(\mathbf{X})=\delta_m^n\sum_{n!}\prod_i C_{z_ix_{i_l}}, \tag{1.3}$$

where the summation is extended for all possible permutation $i_1,\ldots,i_n$ of numbers $1, 2, \ldots, n$.

Moreover

$$EA_{\mathbf{k}}(\mathbf{z})A_{\mathbf{l}}(\mathbf{X})=\delta_{|\mathbf{k}|}^{|\mathbf{l}|}\mathbf{l}!\sum{}^{*}\prod_p\binom{k_p}{j_{1,p}\cdots j_{m,p}}\prod_p\prod_q C_{x_pz_q}^{j_{p,q}},$$

where the summation $\sum^{*}$ is extended for all integers

$$j_{p,q}\geqq 0,\quad p=1,2,\ldots,n,\quad q=1,2,\ldots,m,\quad \sum_p j_{p,q}=k_q,$$

and $\sum_q j_{p,q}=l_p$, see Terdik (1985). From (1.3) follows that the GHAW polynomials of different degree are orthogonal.

2. The spectral representation of the GHAW polynomials

Let us now turn to the Hilbert space H defined by the stationary Gaussian input z_t, $t\in T$, i.e. the space of the square integrable random variables measurable with respect to the σ-algebra $\mathcal{B}(T)=\mathcal{B}(z_t, t\in T)$. Let us regard the following decomposition of H

$$H=\sum_{m=0}^{\infty}\oplus H_m$$

where $\oplus$ denotes the direct sum and H_m denotes the Hilbert space generated by the mth degree polynomials of z_t, $t\in T$. The random variables $Q_{\mathbf{k}}(\boldsymbol{\varepsilon})=A_{k_1}(\varepsilon_{t_1})\ldots A_{k_n}(\varepsilon_{t_n})$ form a

complete orthogonal basis in H_m if $|\mathbf{k}|=m$ where $\varepsilon_t, t\in T$ is a complete orthogonal basis in H_1. We shall refer to $Q_{\mathbf{k}}(\boldsymbol{\varepsilon})$ as the Q-polynomial of degree m and order n.

Lemma 2.1. The subspace H_m of H is generated by the set of random variables $\{A_{\mathbf{k}}(z_{t_1}, \ldots, z_{t_n}), |\mathbf{k}|=m, \mathbf{k}=(k_1, \ldots, k_n)$ and $t_1, \ldots, t_n \in T\}$.

Proof. As any polynomial P_l of degree $l \leqq m-1$ can be written in the form

$$P_k(\mathbf{X}) = \sum_{|\mathbf{k}| \leqq m-1} C_{\mathbf{k}} A_{\mathbf{k}}(\mathbf{X}), \qquad \text{where} \quad \mathbf{X}=(X_1, \ldots, X_n)$$

so for any $y_1, \ldots, y_p \in H_1$, $t_1, \ldots, t_n \in T$ and $\mathbf{k}$, $|\mathbf{k}|=m$

$$EP_l(\mathbf{y}) A_{\mathbf{k}}(z_{t_1}, \ldots, z_{t_n}) = 0,$$

i.e., $A_{\mathbf{k}}(z_{t_1}, \ldots, z_{t_n}) \in H_m$.

From the other hand, an orthogonal system $\varepsilon_1, \varepsilon_2 \ldots$ in H_1 can be chosen such that each of the elements is finite linear combination of the input. Using properties (1.1) and (1.2), it is clear that any Q-polynomial $Q_{\mathbf{k}}(\boldsymbol{\varepsilon})$ of degree $\mathbf{m}$ is linear combination of GHAW polynomials of degree m so $\{A_{\mathbf{k}}(z_{t_1}, \ldots, z_{t_n}), |\mathbf{k}|=m, t_1, \ldots, t_n \in T\}$ is a dense set in H_m. Q.e.d.

Let the spectral representation of z_t be

$$z_t = \int_D e^{i\lambda t} z(d\lambda).$$

Here the range of integration is $D=[-\pi, \pi]$ or $(-\infty, \infty)$, respectively, to the discrete and continuous cases. We assume that the spectral measure $G_z(d\lambda) = E|z(d\lambda)|^2$ is nonatomic.

Theorem 2.1. The spectral representation of the GHAW polynomial is

$$A_{\mathbf{k}}(z_{t_1}, \ldots, z_{t_n}) = \int_{D^{|\mathbf{k}|}} \exp\left(i \sum_{j=1}^{n} t_j \sum_{s=K_{j-1}+1}^{K_j} \lambda_s \right) z(d\boldsymbol{\lambda})$$

$$K_j = \sum_{l=1}^{j} k_l, \quad K_0 = 0, \tag{2.1}$$

where the integral is the multiple Wiener–Ito integral, Major (1981).

Proof. For $|\mathbf{k}|=1$ this representation is the spectral representation of z_{t_i}. So it is enough to show that the recursive formula for the GHAW polynomials is valid for the right-hand side of (2.1) too. From Ito's formula it follows that

$$\int_{D^N} \exp\left(i(s, \boldsymbol{\lambda}) z(d\boldsymbol{\lambda})\right) = z_{s_1} \int_{D^{N-1}} \exp\left(i \sum_{j=2}^{N} s_j \lambda_j \right) z(d\boldsymbol{\lambda}) -$$

$$- \sum_{j=2}^{N} C_{z_{s_1} z_{s_j}} \int_{D^{N-2}} \exp\left(i \sum_{l \neq 1, j} \lambda_l s_l \right) z(d\boldsymbol{\lambda}).$$

Q.e.d.

A special case

Let the input be an autoregressive (AR) time series of order p, i.e.

$$\sum_{t=0}^{p} b_k z_{t-k} = P(\mathscr{L}^{-1}) z_t = \varepsilon_t \qquad (b_0 = 1)$$

where $\varepsilon_t \in N(0, \sigma_\varepsilon^2)$ is the white noise, P is the polynomial and $\mathscr{L}^{-1}$ denotes the backshift operator for z_t, i.e. $\mathscr{L}^{-1} z_t = z_{t-1}$. As ε_t is a complete orthogonal system in H_1 and

$$\varepsilon_t = \int_D e^{i\lambda t} P(e^{-i\lambda}) z(d\lambda),$$

so $e^{i\lambda t} P(e^{-i\lambda})$, $t \in \mathbf{Z}$ is a complete orthonormal system in L^2_G, moreover, from the Ito's formula it follows that

$$Q_{\mathbf{k}}(\varepsilon_{t_1}, \ldots, \varepsilon_{t_n}) = \int_{D^{|\mathbf{k}|}} P_{|\mathbf{k}|}(e^{-i\boldsymbol{\lambda}}) \exp\left(i \sum_l t_l \sum_{s=k_{l-1}+1}^{N_l} \lambda_s \right) z(d\boldsymbol{\lambda})$$

where

$$P_{|\mathbf{k}|}(e^{-i\boldsymbol{\lambda}}) = \prod_{j=1}^{|\mathbf{k}|} P(e^{-i\lambda_j}).$$

3. Definition of the Wiener model

Let the input z_t, $t \in T$, $Ez_t = 0$ be a stationary Gaussian process with stochastical spectral measure $z(d\boldsymbol{\lambda})$ and spectral measure $G_z(d\lambda)$. Let $g_n \in L^2_{G_z^n}$ where $G_z^n(d\boldsymbol{\lambda}) = G_z(d\lambda) \ldots G_z(d\lambda_n)$ such that

$$g_n(\mathbf{s}) = 0 \quad \text{if} \quad \mathbf{s} \notin D_+^n = \{\mathbf{s} | s_i \geqq 0\}, \quad \text{(causability)}$$

$$g_n(\mathbf{s}) = \operatorname{sym} g_n(\mathbf{s}) = \sum_{n!} g_n(\pi\mathbf{s})/n! \quad \text{(symmetry)}$$

and

$$g = (g_0, g_1, \ldots, g_n \ldots) \in \prod_{n=0}^{\infty} L^2_{G_z^n},$$

i.e.

$$\|g\|^2 = \sum_{n=0}^{\infty} \frac{\|g_n\|^2_{G_z^n}}{n!} < \infty \quad \text{(finite memory)}.$$

Moreover let $\hat{g}_n$ denote the Fourier transform of g_n. Then the Wiener model of a nonlinear system y_t with the kernel $g=(g_0, \ldots, g_n, \ldots)$ is defined by the multiple Wiener–Ito integral

$$y_t = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{D^n} \hat{g}_n(\boldsymbol{\lambda}) e^{it \sum_{j=1}^{n} \lambda_j} z(d\boldsymbol{\lambda}) + \xi_t \tag{3.1}$$

where the additive noise ξ_t is independent of the input z_t and

$$E_{\xi_t} = 0, \quad \operatorname{cov}(\xi_t, \xi_s) = \delta_t^s \sigma_\xi^2.$$

Some remarks

1. It is easy to see that $\hat{g} = (\hat{g}_0, \ldots, \hat{g}_n \ldots) \in \prod_{n=0}^{\infty} L^2_{G^n}$, so the definition (3.1) is meaningful.

2. $y_t \in H$ for all $t \in T$, $Ey_t = g_0$

3. y_t is stationary, i.e.

$$Ey_t y_s = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{D^n} |\hat{g}_n(\boldsymbol{\lambda})|^2 e^{i(t-s) \sum_j \lambda_j} G_z^n(d\boldsymbol{\lambda}) + \delta_s^t \sigma_\xi^2.$$

4. If the model is discrete, i.e. $T = \mathbf{Z}$ then we refer to it as discrete and if $T = R$ then as the continuous Wiener model.

From definition (3.1) we easily obtain the so called detailed Wiener modell, Schetzen (1980), p. 398). Let $\varphi_1, \varphi_2 \ldots$ be a complete orthogonal system in $L^2_{G_z}$ and $\varepsilon_t = \int_D \varphi_t(\lambda) z(d\lambda)$. Then for whitened input

$$y_t = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{D^n} \hat{g}_n(\boldsymbol{\lambda}) e^{it \sum_j \lambda_j} z_\varepsilon(d\boldsymbol{\lambda}) + \xi_t =$$

$$= \sum_{n=0}^{\infty} \sum_{\mathbf{s} \geq 0} g_n(\mathbf{s}) A_n(\varepsilon_{t-s_1}, \ldots, \varepsilon_{t-s_n}) + \xi_t$$

for the discrete case and

$$y_t = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{D^n} A_n(\varepsilon_{t-s_1}, \ldots, \varepsilon_{t-s_n}) g_n(\mathbf{s}) d\mathbf{s} + \xi_t$$

for the continuous one.

We remark that when the input process is the Wiener process then $\varepsilon_t(t) = \int_0^\infty \hat{\varphi}_t(s) dw_{t-s}$ and

$$y_t = \sum_{n=0}^{\infty} \sum_{\mathbf{k} \in \mathbf{Z}_+^n} g_{\mathbf{k}} A_n(\varepsilon_{k_1}(t), \ldots, \varepsilon_{k_n}(t)) + \xi_t.$$

4. Discrete Wiener model with stationary AR input

Now, we concentrate on the discrete case but remark that the continuous one is similar as well. Let the input process z_t, $t \in \mathbf{Z}$ be Gaussian stationary AR process of order p, i.e.

$$P(\mathscr{L}^{-1}) z_t = \varepsilon_t$$

where P is the characteristic polynomial associated to the process z_t and $\mathscr{L}^{-1}$ is the backshift operator. In this case the spectral measure of z_t can be written in the form:

$$z(d\lambda) = \frac{1}{P(e^{-i\lambda})} z_\varepsilon(d\lambda), \quad G_z(d\lambda) = \frac{1}{|P(e^{-i\lambda})|^2} d\lambda.$$

Now, the Wiener model with kernel g is the following

$$y_t = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{D^n} \hat{g}_n(\boldsymbol{\lambda}) \prod_{j=1}^{n} \frac{e^{it\lambda_j}}{P(e^{-i\lambda_j})} z_\varepsilon(d\boldsymbol{\lambda}) + \xi_t =$$

$$= \sum_{n=0}^{\infty} \sum_{\mathbf{s} \in \mathbf{Z}_t^n} \mathring{g}_n(\mathbf{s}) \int_{D^n} e^{-i(\mathbf{s}, \boldsymbol{\lambda})} \prod_{j=1}^{n} e^{it\lambda_j} z_\varepsilon(d\boldsymbol{\lambda}) + \xi_t =$$

$$= \sum_{n=0}^{\infty} \sum_{\mathbf{s} \in \mathbf{Z}_t^n} \mathring{g}_n(\mathbf{s}) Q_n(\varepsilon_{t-s_1}, \ldots, \varepsilon_{t-s_n}).$$

We remark that here

$$\hat{g}_n(\boldsymbol{\lambda}) = P_n(e^{-i\boldsymbol{\lambda}}) \sum_{\mathbf{s} \in \mathbf{Z}_t^n} \mathring{g}_n(\mathbf{s}) e^{-i(\boldsymbol{\lambda}, \mathbf{s})},$$

where

$$P_n(e^{-i\boldsymbol{\lambda}}) = \prod_j P(e^{-i\lambda_j}).$$

The simple calculation

$$\|g_n\|_G^2 = \int_{D^n} |\hat{g}_n(\boldsymbol{\lambda})|^2 \frac{1}{|P_n(e^{-i\boldsymbol{\lambda}})|^2} d\boldsymbol{\lambda} = \sum_{\mathbf{s} \in \mathbf{Z}_+^n} \mathring{g}_n^2(\mathbf{s})$$

shows that $\mathring{g}_n(\mathbf{s}) \in l^2$. Let us introduce the Fourier transform of $f_n(\mathbf{s})$ as follows

$$\tilde{f}_n(\boldsymbol{\lambda}) = \sum_{\mathbf{s} \in \mathbf{Z}_+^n} f_n(\mathbf{s}) e^{-i(\boldsymbol{\lambda}, \mathbf{s})}$$

The projection y_t^n of y_t onto the subspace H^n by Lemma 2.1 can be expressed by the GHAW polynomials of degree n, i.e.

$$y_t^n = \sum_{\mathbf{s} \in \mathbf{Z}_t^n} f_n(\mathbf{s}) A_n(z_{t-s_1}, \ldots, z_{t-s_n}) = \sum_{\mathbf{s} \in \mathbf{Z}_+^n} f_n(\mathbf{s}) \int_{\dot{D}^n} e^{it \sum_j \lambda_j - i(\boldsymbol{\lambda}, \mathbf{s})} z(d\boldsymbol{\lambda}) =$$

$$= \int_{\dot{D}^n} \tilde{f}_n(\boldsymbol{\lambda}) e^{it \sum_j \lambda_j} z(d\boldsymbol{\lambda}).$$

We obtained another spectral representation of y_t with kernel f and $\tilde{f} = \hat{g}$, $\tilde{f} = (\tilde{f}_0, \ldots, \tilde{f}_n \ldots)$ For example when z_t is a first order AR process with parameter b then

$$\mathring{g}_n(\mathbf{s}) = \sum_{\mathbf{k} \in \mathbf{Z}_+^n} b^{|\mathbf{k}|} f_n(\mathbf{s} + \mathbf{k})$$

and

$$f_n(\mathbf{s}) = g_n(\mathbf{s}) + b^n g_n(\mathbf{s} + \mathbf{1}).$$

The use of $\hat{g}$ results in an orthogonal Q-polynomial representation, i.e.

$$y_t = \sum_{n=0}^{\infty} \sum_{\mathbf{s} \in \mathbf{Z}_+^n} \mathring{g}_n(\mathbf{s}) Q_n(\varepsilon_{t-s_1}, \ldots, \varepsilon_{t-s_n}) + \xi_t$$

but based on so-called whitened input ε_t. Using the kernel f with Fourier transform $\hat{g} = \tilde{f}$, we get the GHAW polynomial representation, i.e.

$$y_t = \sum_{n=0}^{\infty} \sum_{\mathbf{s} \in \mathbf{Z}_+^n} f_n(\mathbf{s}) A_n(z_{t-s_1}, \ldots, z_{t-s_n}) + \xi_t$$

and we are on the base of input z_t, but unfortunately the GHAW polynomials with the same degree are not orthogonal.

For the statistical investigation one have to choose one of the two possible standpoint in Fig. 1. Standing in the position of the input series z_t, we will estimate the kernel f.

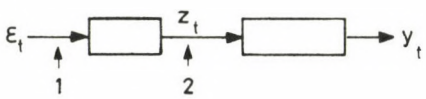


Fig. 1

5. An estimator for kernel f

Under the assumptions of the previous section, we consider the estimator

$$\bar{f}_m(\mathbf{r}) = P_m(\mathscr{L})P_m(\mathscr{L}^{-1})\overline{A_m(z_{t-r_1}, \ldots, z_{t-rn})y_t} =$$

$$= P_m(\mathscr{L})P_m(\mathscr{L}^{-1})\frac{1}{T}\sum_{t=1}^{T} A_m(z_{t-r_1}, \ldots, z_{t-1_m})y_t,$$

where $P_m(\mathscr{L}) = \prod_{j=1}^{m} P(\mathscr{L}_j)$ and $\mathscr{L}_j^{-1}$ is the backshift operator for jth variable of A_m, i.e. $\mathscr{L}_j^{-1} A_m(z_{t-1_1}, \ldots, z_{t-r_m}) = A_m(z_{t-r_1}, \ldots, z_{t-r_j-1}, \ldots, z_{t-r_m})$.

Theorem 5.1. $\bar{f}_m(\mathbf{r})$ is an unbiased estimator for $f_m(\mathbf{r})$.

Proof

$$EP_m(\mathscr{L})P_m(\mathscr{L}^{-1})A_m(z_{t-r_1}, \ldots, z_{t-r_m})y_t =$$

$$= \sum_{n=0}^{\infty} \sum_{\mathbf{s}\in\mathbf{Z}_+^n} f_n(\mathbf{s}) E \int_{D^n} e^{it\sum_j \lambda_j - i(\boldsymbol{\lambda}, \mathbf{s})} z(d\boldsymbol{\lambda}) \int_{D^m} e^{it\sum_j \lambda_j - i(\boldsymbol{\lambda}, \mathbf{r})} |P_m(e^{-i\lambda})|^2 z(d\boldsymbol{\lambda})$$

$$= \sum_{\mathbf{s}\in\mathbf{Z}_+^n} f_m(\mathbf{s}) \int_{D^n} e^{i(\boldsymbol{\lambda}, \mathbf{r}-\mathbf{s})} d\boldsymbol{\lambda} = f_m(\mathbf{r}).$$

Q.e.d.

Let us now introduce the random variable K_t when $m, \mathbf{r}$ are fixed by the following formula

$$K_t = P_m(\mathscr{L})P_m(\mathscr{L}^{-1})A_m(z_{t-\mathbf{r}})y_t$$

where $A_m(z_{t-\mathbf{r}}) = A_m(z_{t-r_1}, \ldots, z_{t-r_m})$. The estimator $\bar{f}_m(\mathbf{r})$ is the mean of K_t, $t = 1, 2, \ldots, T$.

Theorem 5.2. The time series K_t, $t \in \mathbf{Z}$ is stationary and its spectral measure is absolute continuous with respect to the Lebesgue measure.

Proof. First we show that $K_t \in H$, i.e. $EK_t^2 < \infty$, $t \in \mathbf{Z}$ as

$$P_m(\mathscr{L}^{-1})A_m(z_{t-\mathbf{r}}) = Q_m(\varepsilon_{t-\mathbf{r}}).$$

It is enough to prove that $E(Q_m(\varepsilon_{t-\mathbf{t}})y_t) < \infty$. When it is not confusing, we shall write $Q_m(t-\mathbf{s}) \ldots A_m(t-\mathbf{s})$ instead of $Q_m(\varepsilon_{t-s_1}, \ldots, \varepsilon_{t-s_m}) \ldots A_m(z_{t-s_1}, \ldots, z_{t-s_m})$.

$$EQ_m^2(t-\mathbf{r})y_t^2 = \sum_{n=0}^{\infty}\sum_{k=0}^{\infty}\sum_{\mathbf{s}\in\mathbf{Z}_+^n}\sum_{\mathbf{q}\in\mathbf{Z}_+^k} g_n(\mathbf{s})g_k(\mathbf{q})EQ_n(t-s)Q_k(t-\mathbf{q})Q_m^2(t-\mathbf{r}) +$$

$$+ \sigma_\xi^2 = g_m^2(\mathbf{r})EQ_m^4(t-\mathbf{r}) + \sum_{n=0}^{\infty}\sum_{\substack{\mathbf{s}\in\mathbf{Z}_+^n \\ \mathbf{s}\neq\mathbf{r}}} g_n^2(\mathbf{s})EQ_n^2(t-\mathbf{s})EQ_m^2(t-\mathbf{r}) +$$

$$+2\sum_{\substack{k=0\\(\mathbf{r}_1,\mathbf{r}_2)=\mathbf{r}}}^{m} g_k(\mathbf{r}_1)g_{m-k}(\mathbf{r}_2)EQ_m^2(t-\mathbf{r})Q_k(t-r_1)Q_{m-k}(t-\mathbf{r}_2)$$

$$+2\sum_{n=0}^{\infty}\sum_{\mathbf{s}\in\mathbf{Z}_+^n} g_{n+m}(\mathbf{s},\mathbf{r})g_n(\mathbf{s})EQ_m^2(t-\mathbf{r})Q_n(t-\mathbf{s})Q_{n+m}(t-(\mathbf{s},\mathbf{r}))+$$

$$+\sigma_\xi^2 \leqq C\|g\|^2+\sigma_\xi^2$$

where C is a constant.

Taking into consideration that

$$Q_m(t-\mathbf{r})Q_n(t-\mathbf{s})=\sum_{k=0}^{n+m}\sum_{0\leqq\mathbf{q}\leqq\mathbf{R}} d_k(\mathbf{q})Q_k(t-\mathbf{q}),$$

where $R_i=\max(r_i, s_i)$, we see that

$$Q_m(t-\mathbf{r})y_t=\sum_{n=0}^{\infty}\sum_{\mathbf{s}\in\mathbf{Z}_+^n} k_n(\mathbf{s})Q_n(t-\mathbf{s})+Q_m(t-\mathbf{r})\xi_t,$$

where the coefficients $k_n(\mathbf{s})$ depend on m, r but do not on t. From this it follows that K_t has the following spectral representation

$$K_t=\sum_{n=0}^{\infty}\int_{\hat{D}^n}\hat{k}_n(\boldsymbol{\lambda})z(d\boldsymbol{\lambda})+P_m(\mathscr{L})Q_m(t-\mathbf{r})\xi_t.$$

From

$$P_m(\mathscr{L})Q_m(t-\mathbf{r})=\sum_{0\leqq k_j\leqq p} \Pi b_{k_j}Q_m(t-\mathbf{r}-\mathbf{k}),$$

it follows that

$$EP_m(\mathscr{L})Q_m(t_1-\mathbf{r})P_m(\mathscr{L})Q_m(t_2-\mathbf{r})=\sum_{0\leqq k_j\leqq p-|t_1-t_2|}\Pi b_{k_j}\Pi b_{k_j},$$

i.e., stationarity of K_t when ξ_t is stationary.

Theorem 5.3. The estimator $\bar{f}_m(\mathbf{r})$ is consistent.

Proof. As K_t can be represented by the sum of orthogonal random variables therefore its spectral measure is continuous (Linnik–Ibragimov (1965) Theorem 16.7.8). So, the statement of the theorem follows from the ergodic theorem for stationary processes with continuous spectral measure at zero, Hannan (1970).

6. Asymptotic normality of estimator $\bar{f}_m(\mathbf{r})$

If the input process z_t is a Gaussian stationary AR process of order p, then it satisfies the strong mixing condition, Linnik–Ibragimov (1965). The $\bar{f}_m(\mathbf{r})$ is a functional of the input z_t and belongs to H. So, it is reasonable to use the central limit theorem for functionals of strong mixing sequences.

We restrict ourselves to the case of first order AR process with parameter b, ($|b|<1$) but remark that the argument is analogous in the cases when $p \geqq 1$ and

$$|\mathrm{cov}\,(z_0, z_t)| \leqq c|b|^t, \quad |b|<1.$$

Let the strong mixing coefficient of z_t be denoted by

$$\alpha(k) = \sup_{\substack{A \in \mathfrak{M}_{-\infty}^{0} \\ B \in \mathfrak{M}_{k}^{\infty}}} |P(AB) - P(A)P(B)|$$

where $\mathfrak{M}_k^l$ is the σ-algebra generated by $z_k, \ldots, z_l$.

Theorem 6.1. Let us assume that for some $\delta > 0$

1. $E|y_0|^{2+\delta} < \infty$.

2. The kernels f_n are decreasing, uniformly and exponentially, in every direction, i.e.

$$\sum_{s_i \leqq k} \sum_{\substack{s_j \geqq 0 \\ i \neq j}} f_n^2(\mathbf{s}) \leqq c\gamma^k, \qquad i = 1, 2, \ldots, n$$

where c is a constant and $0 < \gamma < 1$.

3. The input z_t is a stationary Gaussian AR process of order 1 with parameter b, $|b|<1$.

Then

$$\sqrt{T}\bar{f}_m(\mathbf{r}) \underset{T \to \infty}{\overset{\mathscr{D}}{}} \mathscr{N}(f_m(\mathbf{r}), \sigma^2)$$

where

$$\sigma^2 = EK_0^2 + 2\sum_{t=1}^{\infty} EK_0K_t.$$

Proof. We shall show that the assumptions of the central limit theorem for functionals of strong mixing sequences (Ibragimov–Linnik (1965) Theorem 18.6.2) are valid for K_t as a functional of z_t. For this we have to prove that for some $\delta_1 > 0$

(1) $E|K_0|^{2+\delta_1} < \infty$

(2) $\sum_{k=1}^{\infty} |E\{|K_0 - E(K_0|\mathscr{M}_{-k}^{k})|^{\frac{2+\delta_1}{1+\delta_1}}\}|^{\frac{1+\delta_1}{2+\delta_1}} < \infty$

(3) $\sum_{k=1}^{\infty} \alpha(k)^{\delta_1/(2+\delta_1)} < \infty \Xi$

To prove (1) we use Hölder's inequality. Let $\delta_1 = \delta/4$ then

$$E|y_0Q_m(-\mathbf{r})|^{2+\delta_1} \leqq (E|y_0|^{(2+\delta_1)(1+\delta_1)})^{\frac{1}{1+\delta_1}} (E|Q_m(-\mathbf{r})|^{(2+\delta_1)\frac{1+\delta_1}{\delta_1}})^{\frac{\delta_1}{1+\delta_1}} \leqq$$

$$\leqq (E|y_0|^{2+\delta})^{\frac{1}{1+\delta_1}} (E|Q_m(-\mathbf{r})|^{3\delta_2})^{1/\delta_2} < \infty,$$

where $\delta_2 = [1 + 1/\delta_1] + 1$.

We show that

$$\sum_{k=1}^{\infty} |E\{|K_0 - E(K_0|\mathfrak{M}^k_{-k})|^{\frac{2+\delta_1}{1+\delta_1}}\}|^{\frac{1+\delta_1}{2+\delta_1}} < \infty.$$

Let $\mathfrak{N}^k_k$ denote the σ-algebra generated by ε_t, $t \in [-k+1, k-1]$. Then $\mathfrak{M}^k_{-k} \supseteq \mathfrak{N}^k_{-k}$ and

$$E(K_0 - E(K_0|\mathfrak{M}^k_{-k}))^2 \leqq E|K_0 - E(K_0|\mathfrak{N}^k_{-k})|^2.$$

If $\max r_i < k-1$, then by the Hölder inequality

$$(E|y_0 P_m(\mathscr{L})Q_m(-\mathbf{r}) - E(y_0 P_m(\mathscr{L})Q_m(-\mathbf{r})|\mathfrak{M}^k_{-k})|^{\frac{2+\delta_1}{1+\delta_1}})^{\frac{1+\delta_1}{2+\delta_1}} \leqq$$

$$\leqq [\{E(|P_m(\mathscr{L})Q_m(-\mathbf{r})|^{2\left(1+\frac{2}{\delta_1}\right)})\}^{\frac{\delta_1}{2+2\delta_1}} E(|y_0 - E(y_0|\mathfrak{M}^k_{-k})|^2)^{\frac{2+\delta_1}{2+2\delta_1}}]^{\frac{1+\delta_1}{2+\delta_1}} \leqq$$

$$\leqq C_1 (E|y_0 - E(y_0|\mathfrak{M}^k_{-k})|^2)^{1/2}.$$

From assumption 2 of the theorem, it follows that

$$E|y_0 - E(y_0|\mathfrak{M}^k_{-k})|^2 = \sum_{n=1}^{\infty} \sum_{\mathbf{s} \in S^n_k} g^2_n(\mathbf{s}) E Q^2_n(\mathbf{s}) \leqq C\gamma^k \|g\|^2_2$$

where $S^n_k = \{\mathbf{s} \in \mathbf{Z}^n_+ | s_i > k \text{ for some } i\}$.

To show (3) we need the Kolmogorov–Rozanov theorem [7] by which

$$\alpha(n) \leqq \rho(n) \leqq 2\pi\alpha(n),$$

where $\rho(n) = \sup_{x,u} Exu$ and the supremum is taken over all $x \in H_1(z_t, t \leqq 0)$ and $u \in H_1(z_t, t \geqq n)$, $Eu^2 = Ex^2 = 1$.

The Kolmogorov–Rozanov Theorem and the following lemma guarantee (3):

Lemma 7.1. $\rho(n) \leqq |^b|^n$

Proof. Let $x = \sum_{j \geqq 0} a_{-j} z_{-j}$ and $u = \sum_{j \geqq n} d_j z_j$ then

$$Exu = b^n \sum_{j \geqq 0} \sum_{k \geqq n} a_{-j} d_k b^{j+k-n} = b^n \int_{-\pi}^{\pi} a(\lambda) d(\lambda) \frac{1}{|1 + be^{-i\lambda}|^2} \, d\lambda \tag{9.1}$$

where

$$a(\lambda) = \sum_{j \geqq 0} a_{-j} e^{ij\lambda}, \quad d(\lambda) = \sum_{k=n}^{\infty} d_k e^{-i(k-n)\lambda}$$

and

$$\int_{-\pi}^{\pi} |a(\lambda)|^2 / |1 + be^{-i\lambda}|^2 d\lambda = \int_{-\pi}^{\pi} |d(\lambda)|^2 / |1 + be^{-i\lambda}|^2 d\lambda = 1$$

using the Cauchy inequality for (9.1) we obtain that

$$\rho(n) \leqq |b|^n.$$

References

1. *Arató, M.*, Linear Stochastic Systems with Constant Coefficients. Lecture Notes in Control and Inf. Sci. **45,** 1982, Springer-Verlag.
2. *Bonnet, G.*, Transformation des signaux aleatoires, Ann. Télécom., **19,** *9–10*, 1964, pp. 203–220.
3. *Bokor, J., Keviczky, L.*, Structural properties and structure estimation of vector difference equations, Int. J. Control, 1982, **36,** *3*, 461–475.
4. *Campbell, L. L.*, On a Class of Polynomials Useful in Probability Calculations, I.E.E.E. Trans. Inf. Theory, **IT-10,** *3* (1964), p. 255.
5. *Grad, H.*, Note on *N*-dim. Hermite Polynomials, Comm. on Pure and Appl. Maths. **2** (Dec), 1949, pp. 325–330.
6. *Hannan, E. J.*, Multiple Time Series. Wiley, New York, 1970.
7. *Hida, T.*, Stationary Stochastic Processes, Princeton 1970.
8. *Ibragimov, I. A., Linnik, Yu. V.*, Independent and Stationary Random Variables, Nauka, Moscow, 1965 (in Russian).
9. *Ibragimov, I. A., Rozanov, Yu. A.*, Gaussian Stochastic Process, Nauka, Moscow, 1970 (in Russian).
10. *Major, P.*, Multiple Wiener–Ito Integrals, Lectures notes in maths, vol. **849,** New York, Springer-Verlag, 1981.
11. *Schetzen, M.*, The Volterra and Wiener Theories of Nonlinear Systems, New York, Wiley, 1980.
12. *Shutterly, H. B.*, General Results in Mathematical Theory of Random Signals and Noise in Nonlinear Devices, I.E.E.E. Trans. on Inf. Theory, **IT-9** (1963), pp. 74–84.
13. *Szegő, G.*, Orthogonal Polynomials, Edward, Michigan, 1948.
14. *Terdik, G.*, Expectation of Nonlinear Functions of Gaussian Random Variables, Publ. Math. Debrecen, 1986. in print
15. *Terdik, G.*, Transfer Functions and Conditions for Stationarity of Bilinear Models with Gaussian Residuals, Proc. R. Soc., London. **A400,** (1985), pp. 315–330.
16. *Várlaki, P., Terdik, G.*: Identification of Raibman Kernels in Zadeh functional series representation of nonlinear systems, Problems of Control and Information Theory (1985) No. *3*.

Обобщённые полиномы Эрмита и оценки ядер для дискретных *I/O* моделей Винера

Г. ТЕРДИК

(Дебрецен)

В статье обсуждаются квадратично-интегрируемые регулярные функционалы стационарных гауссовских рядов. Дискретные винеровские модели со стационарным гауссовским входом описываются как подчинённые ряды. В статье конструируются оценки коэффициентов винеровских моделей с гауссовским AR входом. В главной теореме статьи доказывается нормальность оценок при помощи центральной предельной теоремы для функционалов сильно перемещиванных последовательностей.

G. Terdik
Department of Mathematics
University of Debrecen
H-4010 Debrecen, Pf. 12
Hungary

Problems of Control and Information Theory, Vol. 17 (1), pp. 63—70 (1988)

ASYMPTOTIC DISTRIBUTION OF MAXIMUM LIKELIHOOD ESTIMATORS FOR NONSTATIONARY AUTOREGRESSIVE PROCESSES

S. VERES

(*Budapest*)

(Received December 12, 1986)

Asymptotic distribution of maximum likelihood (ML) estimators for continuous time autoregressive processes is computed without the stability assumptions of the autoregressive operator. It is proved that using appropriate normalizing terms the parameter estimates are the linear combinations of asymptotically normally and Cauchy distributed random variables.

1. Introduction

The present paper is devoted to one of the simplest estimation problems requiring both the ergodic and nonergodic type of statistical inference for parameters. The statistical inference for nonergodic processes has been studied by Anderson [1] for explosive autoregressive (AR) processes. The estimation of the supercritical Galton–Watson process has been dealt with by Heyde [7] and Basawa and Scott [5], where the asymptotic distribution of estimators was proved to be a mixture of normal distributions. An early computation of the exact distribution of the estimator of the first order continuous time process was given by Arató–Kolmogorov–Sinai [2], Arató [3], and Arató–Benczúr [4], and in the discrete time case the asymptotic distribution was computed by Rao [10], [11]. Concerning the asymptotic distributions and the efficiency of estimators, many results were presented in the monograph of Basawa and Scott [5].

This paper is dealing with a special type of nonergodic statistical inference, where the novelty is that the autoregressive process to be estimated has both stable and explosive roots in its characteristic polynomial. Here the roots are called explosive or stable according to that they have positive or negative real parts, respectively. Let us consider processes satisfying the stochastic differential equation

$$\begin{aligned} d\xi_t^{(p-1)} + [a_0\xi_t + a_1\xi_t^{(1)} + \ldots + a_{p-1}\xi_t^{(p-1)}]dt = b dW_t\,, \\ \xi_0^{(i)} = 0\,, \qquad i = 0, 1, \ldots, p-1\,. \end{aligned} \tag{1.1}$$

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

It can easily be proved by a Baxter [6] type theorem that b can consistently be estimated on the basis of observations on any finite time interval. Let $t_i^{(n)}=iT/n$ $(i=1, \ldots, n)$ be a sequence of subdivision points of $[0, T]$ with $\sup_i |t_{i+1}^{(n)}-t_i^{(n)}|\to 0$ as $n\to\infty$. Then, with probability 1

$$\lim_{n\to\infty}\sum_{i=1}^{n}[\xi^{(n-1)}(t_i^{(n)})-\xi^{(n-1)}(t_{i-1}^{(n)})]^2=bT.$$

Therefore, we assume $b=1$ in (1.1) throughout this paper.

Denote by μ_W^T the measure generated on $(C[0, T], \mathscr{B})$ ($\mathscr{B}$ is the Borel σ-algebra generated by the uniform metric on $C[0, T]$) by a Wiener process $\{W_t\}_{t\geq 0}$ and denote by μ_ξ^T the measure generated by ξ_t satisfying (1.1). Then, using the notation

$$x_t=[\xi_t, \xi_t^{(1)}, \ldots, \xi_t^{(p-1)}]^T \qquad \text{and} \qquad a=[a_0, a_1, \ldots, a_{p-1}]^T$$

$$\frac{d\mu_\xi}{d\mu_W}(\xi)=\exp\left(\int_0^T a^T x_t d\xi_t^{(n-1)}-\frac{1}{2}\int_0^T (a^T x_t)\right)^2 dt$$

and therefore, the conditional (under condition $x_0=0$) maximum likelihood estimate of a is

$$\hat{a}_T=\left(\int_0^T x_t x_t^T dt\right)^{-1}\left(\int_0^T x_t d\xi_t^{(n-1)}\right). \tag{1.2}$$

It can easily be seen that x_t satisfies also the equation

$$dx_t=Ax_t dt+BdW_t \tag{1.3}$$

where

$$A=\begin{bmatrix} 0 & 1 & 0 & \ldots & & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & & \ddots & & \\ 0 & 0 & \ldots & & 0 & 1 \\ -a_0 & -a_1 & \ldots & & & -a_{p-1} \end{bmatrix}; \qquad B=\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

The set of eigenvalues of A can usually be decomposed into three parts: a stable part $S=\{\lambda_i: \operatorname{Re}\lambda_i<0\}$, a limiting part $L=\{\lambda_i: \operatorname{Re}\lambda_i=0\}$ and explosive part $E=\{\lambda_i: \operatorname{Re}\lambda_i=0\}$. For the sake of simplicity assume throughout this paper that $\operatorname{Re}(\lambda_1)<\operatorname{Re}(\lambda_2)<\ldots<\operatorname{Re}(\lambda_n)$.

A can be transformed by a Wandermonde matrix V to diagonal form

$$\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix} = V^{-1}AV$$

where

$$V = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_{p-1} \\ \vdots & \vdots & & \vdots \\ \lambda_1^{p-1} & \lambda_2^{p-1} & & \lambda_p^{p-1} \end{bmatrix} \tag{1.4}$$

2. Computation of the asymptotic distributions

The following theorem gives the asymptotic distribution of maximum likelihood estimators in the mixed stationary-explosive case.

Theorem 1. Assume that the matrix A given in (1.4) has eigenvalues only on the set $S \cup E$. Define the normalizing matrix N_T by

$$M_T = \begin{bmatrix} \sqrt{T} & 0 & \dots & & & 0 \\ 0 & \sqrt{T} & 0 & & & \vdots \\ & & & 0 & \dots & 0 \\ 0 & 0 & e^{\lambda_k T} & & & \vdots \\ \vdots & & \ddots & & & \\ & & & e^{\lambda_{k+1} T} & & 0 \\ & & & \ddots & & \\ 0 & \dots & & 0 & & e^{\lambda_p T} \end{bmatrix}, \quad N_T = M_T V^*, \qquad T = 2, 3, \dots$$

where $\lambda_1, \dots, \lambda_{k-1} \in S$ and $\lambda_k, \lambda_{k+1}, \dots, \lambda_p \in E$.

Let $\hat{a}_T$ be the ML estimator defined in (1.2). Then in distribution

$$N_T \cdot (\hat{a}_T - a) \xrightarrow{d} \Sigma^{-1/2} Z \qquad \text{as} \quad T \to \infty,$$

where Σ is a positive definite matrix-valued random variable with Laplace transform

$$E \exp(\operatorname{tr} \Sigma S) = \det [I - 2\bar{S}\bar{B}]^{-1/2},$$

where $S = (S_{ij})$ is a symmetric matrix and

$$\bar{S}_{ij} = S_{ij}/(\lambda_i + \lambda_j), \qquad \bar{B}_{ij} = (V^{-1}BV)_{ij}/(\lambda_i + \lambda_j), \qquad i, j = 1, \ldots, n$$

with the Wandermonde matrix V given in (1.4), and Z is an $N(0, I)$-distributed random vector, which is independent of Σ.

Before proving Theorem 1, we need some lemmas.

Lemma 1. The Laplace transform of $\Sigma_T = \int_0^T \eta_t \eta_t^T dt$, where η_t is the solution of the differential equation

$$d\eta_t = C\eta_t dt + R dW_t, \qquad \eta_0 = 0$$

is

$$E \exp(\operatorname{tr} \Sigma_T S) = e^{T \operatorname{tr}(RD)} \det (I - 2D\Gamma(T))^{-1/2}$$

where D is the solution of the matrix Riccati equation

$$DC + C^*D - 2DRD = S$$

and

$$\Gamma(T) = \int_0^T e^{(C - 2RD)s} \operatorname{Re}^{(C - 2RD)^* s} ds.$$

The proof of this lemma is based on a vectorial version of a method given by Novikov [9] in Remark 2. The derivation of the above formulae for the vectorial case was given in Koncz [8].

Lemma 2. Assume that the processes η_t^1, η_t^2, $t \geqq 0$ are strong solutions of the stochastic differential equations

$$d\eta_t^1 = \lambda_1 \eta_t^1 dt + b_1 dW_t, \qquad \eta_0^1 = 0$$

$$d\eta_t^2 = \lambda_2 \eta_t^2 dt + b_2 dW_t, \qquad \eta_0^2 = 0$$

where λ_1, λ_2, b_1, b_2 are complex numbers with $\operatorname{Re} \lambda_1 > 0$, $\operatorname{Re} \lambda_2 > 0$. Then the sequence

$$\zeta_T^{12} = e^{-(\lambda_1 + \lambda_2)T} \int_0^T \eta_t^1 \eta_t^2 dt$$

converges in L_2 to a random variable $\zeta_\infty^{12} \in L_2$. Some straightforward but tedious computation shows that ζ_T^{12}, $T \geqq 0$, is a Cauchy sequence, and the statement of Lemma 1 follows.

Proof of Theorem 1. Multiplying equation (1.3) by V^{-1} to the left we obtain the new equation

$$d\eta_t = \Lambda' \eta_t dt + B' dW_t$$

where

$$\eta_t = V^{-1}x_t, \qquad \Lambda = V^{-1}AV, \qquad B' = V^{-1}B.$$

Then

$$\hat{a} - a = (V)^{-1}\left(\int_0^T \eta_t \eta_t^T dt\right)_p^{-1}\left(\int_0^T \eta_t dw_t\right) \tag{2.3}$$

where A_p denotes the last column of a matrix A. We compute first the Laplace transform of the asymptotic distribution of $e^{-\Lambda T}\int_0^T \eta_t\eta_t^T dt\, e^{-\Lambda T}$. By Lemma 1 we have that there is a limit random matrix Σ which is constant in its first $k-1$ rows and columns with

$$\Sigma_{ij} = \lim_{T\to\infty}\frac{1}{T}\int_0^T \eta_t^i\eta_t^j dt = \frac{(V^{-1})_{in}(V^{-1})_{jn}}{\lambda_i+\lambda_j}, \qquad 1 \leqq i,j \leqq k-1.$$

Also the variables Σ_{ij}, $i \leqq k-1, j \geqq k$ are constant, since

$$\overline{\lim_{T\to\infty}}\frac{e^{-\lambda_j T}}{\sqrt{T}}\int_0^T (\eta_t^i\eta_t^j - E\eta_t^i\eta_t^j)dt \leqq$$

$$\leqq \overline{\lim_{T\to\infty}} D^2\left(\frac{1}{\sqrt{T}}\int_0^T \eta_t^i dt\right) D^2\left(e^{-\lambda_j T}\int_0^T (\eta_t^j)^2 dt\right) = 0\cdot c_j,$$

where

$$c_j = \frac{1}{8}(V^{-1})_{jn}^2/|\operatorname{Re}\lambda_j|^2$$

and

$$\Sigma_{ij} = \lim_{T\to\infty}\frac{e^{-\lambda_j T}}{\sqrt{T}}\int_0^T E(\eta_t^i\eta_t^j)dt =$$

$$= \lim_{T\to\infty}\frac{e^{-\lambda_j T}}{\sqrt{T}}\int_0^T \frac{(V^{-1})_{in}(V^{-1})}{\lambda_i+\lambda_j} jn(e^{(\lambda_i+\lambda_j)}-1)dt =$$

$$= \lim_{T\to\infty}\frac{(V^{-1})_{in}(V^{-1})_{jn}}{(\lambda_i+\lambda_j)}\frac{e^{-\lambda_j T}}{\sqrt{T}}(e^{(\lambda_i+\lambda_j)T}-1) = 0$$

since $\operatorname{Re}\lambda_i < 0$ and $\operatorname{Re}\lambda_j > 0$.

The remainder right lower quadruple part of Σ is composed of random variables, common distribution can be computed in the following way. Introduce the notations

$$\Lambda' = \operatorname{diag}(\lambda_k, \ldots, \lambda_p) \qquad \Sigma'_T = (\Sigma_{ij})_{i,j=k,k}^{p,p}$$

$$\eta'_t = (\eta_k, \ldots, \eta_p)^T \qquad S' = (S_{ij})_{i,j=1,1}^{p,p}$$

$$B'_{ji} = 0, \quad \text{if} \quad i < p \quad \text{and} \quad B'_{jn} = (v^{-1})_{jn},$$

$$B' = V^{-1}B$$

and substitute in Lemma 2 $C = \Lambda'$, $R_T = B_\Lambda$, $\eta_t = \eta'_t$. Then the Laplace transform of $e^{-\Lambda T} \int_0^T \eta'_t (\eta'_t)^T dt\, e^{-\Lambda T}$ is

$$E \exp\left(\operatorname{tr}\left(e^{-\Lambda T} \int_0^T \eta'_t (\eta'_t)^T dt\, e^{-\Lambda T} S\right)\right) =$$

$$= E \exp\left(\operatorname{tr}\left(\int_0^T \eta'_t (\eta'_t)^T dt (e^{-\Lambda T} S e^{-\Lambda T})\right)\right)$$

and, therefore, the equation

$$D_T \Lambda' + (\Lambda')' D_T + D_T B' D_T = e^{-\Lambda' T} S' e^{-\Lambda' T} \tag{2.5}$$

has to be solved. Since the left-hand side of (2.5) tends uniformly to 0, and the solution of (2.17) is close to the solution of

$$D\Lambda' + (\Lambda')D = e^{-\Lambda' T} S' e^{-\Lambda' T}$$

which is $D_T \sim e^{-\Lambda' T} \bar{S} e^{-\Lambda' T}$ with $(\bar{S})_{ij} = S'_{ij}/(\lambda_i + \lambda_j)$, $i, j = k, \ldots, n$, therefore

$$\lim_{T \to \infty} E \exp(\operatorname{tr} \Sigma'_T e^{-\Lambda' T} S' e^{-\Lambda' T}) =$$

$$= \lim_{T \to \infty} e^{T \operatorname{tr}(BD_T)} \det(I - 2D_T(T))^{-1/2} =$$

$$= \lim_{T \to \infty} \det(I - 2e^{-\Lambda' T} \bar{S} e^{-\Lambda' T} e^{\Lambda' T} \bar{B} e^{\Lambda' T}) =$$

$$= \lim_{T \to \infty} \det(e^{-\Lambda' T})^{-1/2} \det(I - 2\overline{SB})^{-1/2} \det(e^{-\Lambda' T})^{-1/2} =$$

$$= \det(I - 2\overline{SB})^{-1/2}$$

Furthermore, it can easily be proved that

$$\Gamma(T) = \int_0^T e^{(\Lambda' - 2B'D_T)s} B' e^{(\Lambda' - 2B'D_T)'s} \sim$$

$$\sim \int_0^T e^{A's} B' e^{A's} ds \sim e^{AT} \bar{B} e^{AT}$$

and

$$\lim_{T \to \infty} e^{T \operatorname{tr}(BD_T)} = 1 .$$

The following theorem is a straightforward vectorial extension of Theorem 2, Feigin [12].

Theorem (Feigin). Assume that $M = \{M_t, H_t, 0 < t\}$ is a vector-valued square integrable martingale and let $I_T = E(M_T M_T^T)$ be positive definite. Suppose that the following conditions are satisfied as $T \to \infty$

(i) $\operatorname{tr}(I_T) \to \infty$

(ii) $E\{I_T^{-1/2} \sup_{t \leqq T} | \Delta M_t | \} \to \infty$

(iii) $I_T^{-1/2} [M]_T I_T^{-1/2} \to \Sigma$

where Σ is a positive definite matrix valued random variable. Then

$$\mathcal{L}(I_T^{-1/2} M_T) \to \mathcal{L}(x) \qquad \text{(stably)}$$

where

$$E(e^{i\lambda^T x}) = E[e^{-1/2 \lambda^T \Sigma \lambda}] .$$

Applying this theorem for $M_T = \int_0^T \eta_t dw_t$, we have that for $I_T = E \int_0^T \eta_t \eta_t^T dt$

$$\| I_T - M_T E \Sigma M_T \| \to 0 \qquad \text{as} \quad T \to \infty ,$$

which implies condition (i). It can be concluded from the above derivations that also

$$I_T^{-1/2} [M]_T I_T^{-1/2} = I_T^{-1/2} \int_0^T \eta_t \eta_t^T dt \, I_T^{-1/2} \xrightarrow{P} \Sigma \qquad \text{as} \quad T \to \infty$$

is satisfied. Condition (ii) can be derived from the standard properties of Wiener integrals.

Thus it follows from (2.3) that $N_T(\hat{a}_T - a)$ stably converges to a random variable with Laplace transform $E(e^{-1/2 \lambda^T \Sigma \lambda})$ implying the statement of our theorem.

References

1. *Anderson, T. W.*, On asymptotic distribution of estimates of parameters of stochastic difference equations. Ann. Statist., **30,** *3* (1959), pp. 676–687.
2. *Arató, M., Kolmogorov, A. N., Sinai, Ja. G.*, Estimation of the parameters of a complex stationary Gaussian–Markov process. Dokl. Akad. Nauk SSSR **146** (1962), pp. 1368–1371.

3. *Arató, M.*, On the statistical examination of continuous state Markov processes, III; Magyar Tud. Akad. Mat. Fiz. Oszt. Közl. **14** (1964), pp. 1317–1333. English trans., Selected Trans. Math. Statist. and Probability, **14,** Amer. Math. Soc., Providence (1978), pp. 253–267.
4. *Arató, M., Benczúr, A.*, Точные формулы для плотностей мер элементарных Гауссовских процессов. Studia Sci. Mathemat. Hungarica **5,** *4* (1976), pp. 445–456.
5. *Basawa, I. V., Scott, D. J.*, Asymptotic Optimal Inference for Nonergodic Models. Lecture Notes in Statistics **17** (1983), Springer, New York–Berlin.
6. *Baxter, G.*, A strong limit theorem for Gaussian processes. Proc. Amer. Math. Soc., **7** (1956), pp. 522–525.
7. *Heyde, C. C.*, Remarks on efficiency in estimation for branching processes Biometrika **62** (1975) pp. 49–55.
8. *Koncz, K.*, Distribution of sufficient statistics for continuous time ARMA processes; (in Hungarian), Alkalmazott Matematikai Lapok (1986).
9. *Novikov, A. A.*, Об оценках параметров диффузионных процессов. Studia Sci. Mathemat. Hungarica **7,** *2* (1972), pp. 201–209.
10. *Rao, M. M.*, Inference in stochastic processes; Zeitschrift für Wahrscheinlichkeitstheorie **5,** *4* (1966), pp. 317–335.
11. *Rao, M. M.*, Probability Theory and Its Applications. Academic Press, New York–London, 1984.
12. *Feigin, P. D.*, Central limit theorem for semimartingales. Stochastic Processes and Its Appl. **19,** *2* (1985), pp. 125–134.

Асимптотическое распределение оценок коэффициентов авторегрессивных процессов

Ш. ВЕРЕШ

(Будапешт)

В статье выведены асимптотические распределения оценок максимального правдоподобия в том случае, когда авторегрессивные операторы имеют смешанные — стабильные и нестабильные — собственные значения. Тогда нормирующие матрицы составлены из экспоненциальных и линейных выражений.

Асимптотическое распределение это смесь нормальных и Коши распределений.

S. Veres
Department of Probability Theory and Mathematical Statistics
Eötvös Loránd University, Budapest
H-1088, Múzeum krt. 6–8
Budapest
Hungary

РУССКИЙ ПЕРЕВОД

Проблемы Управления и Теории Информации, том **17**, *номер 1 (1988)*

ЭКСТРЕМАЛЬНАЯ ЗАДАЧА T-СЕМЕЙСТВА КОДОВ В ПРОСТРАНСТВЕ ХЭММИНГА

А. К. АЙДИНЯН

(Ереван)

В статье рассматриваются (T, m, Δ)-семейства кодов и дается точная верхняя граница мощности таких кодов. Кроме того, приводится класс оптимальных экзидистантных семейств кодов.

1. Введение

Пусть E^m — множество бинарных последовательностей длины m, а d — обозначает расстояние Хэмминга в E^m. Далее пусть множество натуральных чисел $\Delta = \{\delta_{ij}\}$; $\delta_{ij} = \delta_{ji}$; $i \neq j$; $i,\ j = \overline{1, T}$ $(T \geqq 2)$. Семейство $(A_1, \ldots, A_T)$; $A_1, \ldots, A_T \subset E^m$ называется (T, m, Δ)-семейством кодов, или (T, m, Δ)-системой, если для любых $i, j \in \{1, \ldots, T\}$ $(i \neq j)$ и

$$\forall a^i \in A_i \quad \forall a^j \in A_j \quad d(a^i, a^i) = \delta_{ij}.$$

Если при любых $i, j \in \{1, \ldots, T\}$ $\delta_{ij} = \delta$, то мы будем называть такую систему эквидистантным семейством кодов и будем обозначать через (T, m, δ).

Рассмотрим следующую функцию

$$M(T, m, \Delta) \triangleq \max \left\{ \prod_{u=1}^{T} |A_u| : (A_1, \ldots, A_T) \text{ есть } (T, m, \Delta)\text{-система} \right\}.$$

Изучение (T, m, Δ)-систем и функции $M(T, m, \Delta)$ связано с задачами нахождения сложностных оценок при распределенном вычислении (см. работу Яо [1]). Алсведе и др. была доказана (индукцией по m) следующая

Теорема [2]. Для любого натурального m

$$\max M(2, m, \delta) = 2^{2n}, \qquad \text{при} \quad m = 2n,$$

$$\text{или} \quad m = 2n + 1,$$

причем максимум достигается при $\delta = n$. В работе [3] авторы предлагают рассмотреть эту задачу для $T > 2$. Однако, метод доказательства, приведенный в [2], в этом случае не приводит к успеху. В данной статье эта задача решена для произвольного T.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

2. Основной результат

Теорема. 1) Для произвольной (T, m, Δ)-системы имеет место неравенство

$$M(T, m, \Delta) \leqq 2^m .$$

2) При $k = 1, 2, \ldots ; T \leqq n \leqq 2T$; n — произвольное число, для которого существует матрица Адамара H_n выполняется равенство

$$M(T, m = nk, m/2) = 2^m . \tag{1}$$

Для доказательства теоремы используется

Лемма.

$$M(T, 2n, n) \leqq 2^{2n} . \tag{2}$$

Доказательство. Приведем сначала доказательство леммы для $T = 2$.

Пусть имеем $(2, 2n, n)$-систему (A_1 , A_2). В векторах E^m заменим нули на -1. Полученные векторы будем рассматривать как векторы Евклидова пространства E^{2n}. Обозначим через (B_1 , B_2) пару, полученную из (A_1 , A_2) соответственно. Заметим теперь, что условие

$$d(a^1, a^2) = n(a^1 \in A_1 , a^2 \in A_2)$$

эквивалентно условию

$$\langle b^1, b^2 \rangle = 2n - 2d(a^1, a^2) = 0 \qquad (b^1 \in B_1 , b^2 \in B_2) ,$$

где $\langle b^1, b^2 \rangle$ — скалярное произведение векторов b^1, b^2 соответствующих векторам a^1 и a^2. Поэтому линейные оболочки L_1 , L_2 множеств B_1 и B_2 являются взаимно ортогональными плоскостями пространства R^{2n}. Нетрудно увидеть, что $| B_1 | \leqq 2^{\dim L_1}$, $| B_2 | \leqq 2^{\dim L_2}$ а кроме того $\dim L_1 + \dim L_2 \leqq 2n$.

Таким образом, мы получаем

$$| B_1 | \, | B_2 | \leqq 2^{2n} .$$

Если теперь $T > 2$, то плоскости L_1 и L^* (L^*-линейная оболочка $B_2 \cup \ldots \cup B_T$) взаимно ортогональны и (2) мы получаем индуктивно. Лемма доказана.

Доказательство теоремы. Пусть имеем (T, m, Δ)-систему $(A_1 , \ldots , A_T)$, а $\mathscr{C}_n$ — матрица Адамара порядка n. Обозначим через $\bar{a} (a \in E^m)$ дополнение вектора a, т.е. $\bar{a} = 1^m + a$, где 1^m вектор из одних единиц длины m; „$+$" означает сложение в пространстве Хэмминга. Если $A \subseteq E^m$, то $\bar{A} \triangleq \{\bar{a} : a \in A\}$.

Выберем далее $T \times n$-подматрицу (h_{uv}) матрицы H_n и определим матрицу (α_{uv}) размера $T \times n$ следующим образом

$$\alpha_{uv} = \begin{cases} A_u , & \text{если} \quad h_{uv} = 1 \\ \bar{A}_u , & \text{если} \quad h_{uv} = -1 , \end{cases}$$

где A_u $(u=\overline{1,T})$ представлено в виде $|A_u|\times m$-матрицы. Определим теперь $\mathscr{A}_i$ $(i=\overline{1,T})$ как декартово произведение $\alpha_{i1}\times\ldots\times\alpha_{in}$. Поскольку матрица Адамара H_n определяет $(n, n, d=n/2)$-код [4, гл. 2] ясно, что семейство $(\mathscr{A}_i, \ldots, \mathscr{A}_T)$ является $(T, mn, mn/2)$ — эквидистантным семейством кодов. Используя лемму, получим

$$\prod_{i=1}^{T}|\mathscr{A}_i|=\prod_{i=1}^{T}|A_i|^n\leqq 2^{mn}.$$

Итак, мы доказали первую часть теоремы. Теперь дадим конструкцию семейства оптимальных эквидистантных кодов (для которых выполняется равенство (1)) для произвольного T.

Пусть имеем $(n, 2n, d=n/2)$-код Адамара $\mathscr{C}_n$ [4] Рассмотрим разбиение $\mathscr{C}_n$ на t классов $\mathscr{C}_n=\bigcup_{i=1}^{t}A_i$, такое, что каждый класс $A_i(i\in\{1,\ldots,t\})$ вместе с каждым элементом a^i содержит также его дополнение $\bar{a}^i$. Ясно, что семейство $(A_1,\ldots,A_t)$ является $(t, n, n/2)$-системой. Для того, чтобы $(A_1,\ldots,A_t)$ было оптимальным должны выполняться следующие соотношения

$$\sum_{i=1}^{t}l_i=2n$$

$$\prod_{i=1}^{t}l_i=2^n=2^{(l_1+\ldots+l_t)/2},\qquad \text{где}\quad l_i=|A_i|;\ i=\overline{1,t}.$$

Поскольку при натуральном l равенство $2^{l/2}=l$ имеет место только при $l=2$ или 4, заключаем, что наша конструкция оптимальна в том и только том случае, когда $|A_i|=2$, или $|A_i|=4$. Теперь покажем, что для произвольного Tсуществует код Адамара $\mathscr{C}_n$, из которого можно построить оптимальную $(T, n, n/2)$-систему. Очевидно, что при любом натуральном Tсуществует матрица Адамара H_n, где $T\leqq n\leqq 2T$. С другой стороны, разбиение кода $\mathscr{C}_n$ дает нам (как указано выше) оптимальную $(T, n, n/2)$-систему, если выполняются равенства

$$\begin{cases}4t_1+2t_2=2n\\ t_1+t_2=T,\end{cases}$$

где t_1, t_2 — неотрицательные целые числа. Очевидно, что эта система уравнений относительно t_1 и t_2 имеет целое неотрицательное решение лишь при $T\leqq n\leqq 2T$. Поэтому для любого $T\geqq 2$ можно получить оптимальную $(T, n, n/2)$-систему. Далее ясно, что при любом $k\in\mathbf{N}$ семейство $(A_1^k,\ldots,A_T^k)$ является $(T, m=nk, m/2)$-эквидистантным семейством кодов:

$$\prod_{i=1}^{T}|A_i^k|=\prod_{i=1}^{T}|A_i|^k=2^{kn}=2^m.$$

Теорема полностью доказана.

Приведем пример оптимальной $(3, 4k, 2k)$-системы

$$\mathscr{C}_4 = \{0000, 1111, 1100, 0011, 0110, 1001, 0101, 10{,}10\}$$

$$A_1 = \{0000, 1111\}^k,\ A_2 = \{1010, 0101\}^k,\ A_3 = \{1100, 0011, 0110, 1001\}^k$$

$$|A_1 \| A_2 \| A_3| = 2^{4k}.$$

Литература

1. *Yao, A. C.*, Some complexity questions related to distributive computing. 11th ACM Symp. on Theory of Computing. (1979), pp. 209–213
2. *Ahlswede, R., El Gamal, A., Pang, K. F.*, A two family extremal problem in Hamming space. Discrete Math., **49** (1984), pp. 1–5.
3. *Ahlswede, R., Mörs, M.*, Inequalities for code pair. Submitted to Discrete Math.
4. *Mac Williams, F. J., Sloane, N. J. A.*, The Theory of Error Correcting Codes. Amsterdam: North Holland, 1977.

PRINTED IN HUNGARY

Akadémiai Kiadó és Nyomda Vállalat, Budapest

NOTE TO CONTRIBUTORS

Two copies of the *manuscript* (each complete with figures, tables and references) are to be sent to

E.D. TERYAEV coordinating editor
Department of Mechanics and Control Processes
Academy of Sciences of the USSR
Leninsky Prospect 14, Moscow V-71, USSR

or to

L. GYÖRFI
Technical University of Budapest
H-1111 Budapest, Stoczek u. 2, Hungary

Authors are requested to retain a third copy of the submitted typescript to be able to check the proofs.

The papers, preferably in English or Russian, should be typed double spaced on one side of good-quality paper with wide margins (4–5 cm). The first page of the paper should carry the title, the author(s)' names and the name of the town where they are active. The name and address of the author to whom the proofs should be sent should be given at the end of the paper. An *abstract* should head the paper. English papers should also have a Russian abstract.

The papers should not exceed 15 pages (25 × 50 characters per page) including tables and references. The proper location of the tables and figures must be indicated on the margin.

Mathematical notations should follow up-to-date usage. Equations longer than half a line should not be incorporated in the text. In-text equations must be typed on a single line except that one level of subscripting and/or superscripting is permissible. Use / instead of horizontal bars. Displayed equations should be written so as to require the fewest possible lines. Therefore use "exp" for the exponential function whenever the exponent requires more than a single line. Matrices should, if possible, not be written in full. Use subscript notations instead such as $A = ||a_{ij}||$. Write diagonal matrices as diag $(d_1, d_2, \ldots d_n)$.

The authors will be sent galley proofs to be returned by next mail. Rejected manuscripts will be returned. Authors will receive 100 reprints free of charge. Additional reprints may be ordered.

К СВЕДЕНИЮ АВТОРОВ

Рукописи статей в трех экземплярах на русском языке и в трех на английском следует направлять по адресу: 117312 Москва В-312, просп. 60 летия Октября, 9, МНИИПУ. Редакция журнала «Проблемы управления и теории информации» (зав. редакцией Н. И. Родионова, тел. 208-60-19).

Объём статьи не должен превышать 15 печатных страниц (25 строк по 50 букв). Статье должна предшествовать аннотация объемом 50–100 слов и приложено резюме–реферат объемом не менее 10–15% объема статьи на русском языке в трех экземплярах, на котором напечатан служебный адрес автора (фамилия, название учреждения, адрес).

При написании статьи авторам надо строго придерживаться следующей формы: введение (постановка задачи), основное содержание, примеры практического использования, обсуждение результатов, выводы и литература.

Статьи должны быть отпечатаны с промежутком в два интервала, последовательность таблиц и рисунков должна быть отмечена на полях. Математические обозначения рекомендуется давать в соответствии с современными требованиями и традициями. Разметку букв следует производить только во втором экземпляре и русского, и английского варианта статьи.

Авторам высылается верстка, которую необходимо незамедлительно проверить и возвратить в редакцию.

После публикации авторам высылаются бесплатно 100 оттисков их статей.

Рукописи непринятых статей возвращаются авторам.

CONTENTS · СОДЕРЖАНИЕ

HU ISSN 0370-2529 Index: 26 660

VOL. 17 · NUMBER 2
TOM HOMEP

ACADEMY OF SCIENCES OF THE USSR
HUNGARIAN ACADEMY OF SCIENCES
CZECHOSLOVAK ACADEMY OF SCIENCES

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

АКАДЕМИЯ НАУК СССР
ВЕНГЕРСКАЯ АКАДЕМИЯ НАУК
ЧЕХОСЛОВАЦКАЯ АКАДЕМИЯ НАУК

1988

AKADÉMIAI KIADÓ, BUDAPEST
DISTRIBUTED OUTSIDE THE COMECON-COUNTRIES
BY PERGAMON PRESS, OXFORD

PROBLEMS OF CONTROL AND INFORMATION THEORY

An international bi-monthly sponsored jointly by the Presidium of the Academy of Sciences of the USSR, of the Hungarian Academy of Sciences and of the Czechoslovak Academy of Sciences. The six issues published per year make up a volume of some 480 pp. It offers publicity for original papers and short communication of the following topics:

— theory of control processes
— theory of adaptive systems
— theory of estimation and identification
— theory of controlling robot-technologic and flexible manufacturing systems
— information theory
— information-theoretic aspects of multiple access networks.

While this bi-monthly is mainly a publication forum of the research results achieved in the socialist countries, also papers of international interest from other countries are welcome.

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

Международный журнал Академии наук СССР, Венгерской Академии наук и Чехословацкой Академии наук выходит 6 раз в год общим объемом 480 печатных страниц.

В журнале публикуются оригинальные научные статьи и статьи обзорного характера по следующим проблемам управления и теории информации:

— теория процессов управления;
— теория адаптивных систем;
— теория оценивания и идентификации;
— теория управления робототехническими и гибкими производственными системами;
— теория информации;
— теория информации в области сетей с множественным доступом.

Целью журнала является ознакомление научной общественности различных стран с важнейшими проблемами, имеющими актуальный и перспективный характер, научными достижениями ученых социалистических и других стран.

Distributors

For the Soviet Union:

SOYUZPECHATY, Moscow 123 308 USSR

For Albania, Bulgaria, China, Cuba, Czechoslovakia, German Democratic Republic, Korean People's Republic, Mongolia, Poland, Rumania, Vietnam and Yugoslavia:

KULTURA Hungarian Foreign Trading Co.
P. O. Box 149 H-1389 Budapest, Hungary

For all other countries:

PERGAMON JOURNALS LTD, Headington Hill Hall, Oxford OX3 OBW, England

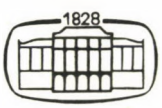
or

PERGAMON JOURNALS INC, Maxwell House, Fairview Park, Elmsford, NY 10523, USA

1988 Subscription Rate DM 535,— per annum including postage and insurance.

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ



AKADÉMIAI KIADÓ

PUBLISHING HOUSE OF THE HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST

Problems of Control and Information Theory, Vol. 17 (2), pp. 71—76 (1988)

DISCRETE VERSION OF MODEL-IN-THE SYSTEM CONTROL

V. STREJC

(Prague)

(Received March 8, 1987)

The paper is concerned with the synthesis of the discrete version of the model-following the precept using model-in-the-system and quadratic cost function. Control plant and its model are assumed to be linear. Solution yields one feedback controller and two feed-forward controllers.

1. Introduction

There are many day-to-day control applications where a model-following technique is recommendable in order to reach the desired control quality. The problem of this kind was formulated in the past, namely for linear continuously acting systems. References [1] and [6] prove that the model-following concept is not new but it is still open for up-to-date modifications. This contribution is devoted to the discrete version of the state space approach of model-following control of linear systems. The model of the system to be controlled may be applied in different ways as for example model-in-the-system decoupling control, model-in-the-performance-index or as model-in-the-system. The latter case is the most important one and will be described in more detail.

For the sake of brevity and clarity only single-input–single-output time-invariant systems are considered in this article. Extension for multi-input–multi-output time varying systems is straightforward.

The objective of the formulation is to provide a good match between the dynamics of the plant and of the conceptual model minimizing the squared error between the outputs of the plant and the model under the constraint imposed on the controlling variable and thus forcing the plant to follow the model output.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

2. Formulation

Consider a controllable and observable plant described by the following state equations

$$\begin{aligned} x_p(k+1) &= F_p x_p(k) + g_p u_p(k) \\ y_p(k) &= c_p^T x_p(k) \end{aligned} \tag{2.1}$$

the model of the plant

$$\begin{aligned} x_m(k+1) &= F_m x_m(k) + g_m u_m(k) \\ y_m &= c_m^T x_m(k) \end{aligned} \tag{2.2}$$

and the generator of the command variable

$$\begin{aligned} \omega(k+1) &= W\omega(k) \\ w(k) &= c_w^T \omega(k) \end{aligned} \tag{2.3}$$

for the discrete sampling instants $k = 0, 1, 2, \ldots$. Here $u \in \mathbf{R}$ and $u_m \in \mathbf{R}$ are the scalar correcting variables, $w \in \mathbf{R}$ is the command variable, $y_p \in \mathbf{R}$ and $y_m \in \mathbf{R}$ are the plant and model plant outputs and $x_p \in \mathbf{R}$, $x_m \in \mathbf{R}$ and $\omega \in \mathbf{R}$ are the state vectors of dimensions n_p, n_m and n_w, while $n_w \leqq n_m \leqq n_p$. Matrices in (2.1), (2.2) and (2.3) have compatible dimensions.

Defining a combined state vector

$$z = \begin{bmatrix} x_p \\ x_m \\ \omega \end{bmatrix} \tag{2.4}$$

of the dimension $n = n_p + n_m + n_w$ and taking into account that $u_m = w$ then the overall system is described by the state equations

$$\begin{aligned} z(k+1) &= Fz(k) + gu(k) \\ y(k) &= C^T z(k) \end{aligned} \tag{2.5}$$

where

$$F=\begin{bmatrix} F_p & 0 & 0 \\ 0 & F_m & g_m c_w^T \\ 0 & 0 & W \end{bmatrix}; \qquad g=\begin{bmatrix} g_p \\ 0 \\ 0 \end{bmatrix}; \qquad C^T=\begin{bmatrix} c_p & 0 & 0 \\ 0 & c_m^T & 0 \\ 0 & 0 & c_w^T \end{bmatrix};$$

$$y=\begin{bmatrix} y_p \\ y_m \\ w \end{bmatrix}; \qquad u=u_p. \tag{2.6}$$

The problem is to find the structure of the overall system and the controllers generating correcting variable u such that the cost function

$$J=\sum_{k=0}^{\infty}[z^T(k)\Phi z(k)+u^T(k)Ru(k)] \tag{2.7}$$

reaches its minimum where $\Phi=[\Phi_{ij}]$, $i, j=1, 2, \ldots, n$ and R are in general constant symmetric weighting matrices. Of course, for single-input–single-output systems, R is a scalar.

The elements of the weighting matrix Φ can be determined in conformity with the requirements imposed on the overall system representing the most important part of the cost function (2.7). It is required that

$$J_0=\sum_{k=0}^{\infty}[(y_p(k)-y_m(k))^T Q(y_p(k)-y_m(k))+u^T(k)Ru(k)]=$$

$$=\sum_{k=0}^{\infty}[(c_p^T x_p(k)-c_m^T x_m(k))^T Q(c_p^T x_p(k)-c_m^T x_m(k))+u^T(k)Ru(k)] \tag{2.8}$$

is a minimum. Expanding the first term in the sum (2.8) yields the following matrix expression

$$x_p^T(k)c_p Q c_p^T x_p(k)-x_p^T(k)c_p Q c_m^T x_m(k)-x_m^T(k)c_m Q c_p^T x_p(k)+$$
$$+x_m^T(k)c_m Q c_m^T x_m(k)$$

so that the weighting matrix Φ can be defined as

$$\Phi=\begin{bmatrix} c_p Q c_p^T & -c_p Q c_m^T & 0 \\ -c_m Q c_p^T & c_m Q c_m^T & 0 \\ 0 & 0 & 0 \end{bmatrix}=\begin{bmatrix} \Phi_{11} & \Phi_{12} & 0 \\ \Phi_{21} & \Phi_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}. \tag{2.9}$$

3. Solution

By inspection of Eqs. (2.5) and (2.7) it may be found that the state description of the proposed system preserves the form of the general discrete *LQ* problem, the solution of which is well known and published repeatedly in the corresponding professional literature. Using the respective results we may write that

$$u_0(k) = -[g^T P g + R]^{-1} g^T P F z(k) \tag{3.1}$$

where u_0 is the optimum correcting variable, while P is an $(n \times n)$ symmetric matrix obtained by the solution of the following nonlinear matrix Riccati difference equation

$$P = \Phi + F^T P g [g^T P g + R]^{-1} g^T P F + F^T P F. \tag{3.2}$$

Denoting

$$D = g^T P g + R$$

we have

$$D = [g_p^T \quad 0 \quad 0] \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix} \begin{bmatrix} g_p \\ 0 \\ 0 \end{bmatrix} + R = g^T P_{11} g + R. \tag{3.3}$$

Hence, with this result, the optimum correcting variable is

$$u_0(k) = -D^{-1} g^T P F = -D^{-1} [g^T P_{11}, g^T P_{12}, g^T P_{13}] \begin{bmatrix} F_p & 0 & 0 \\ 0 & F_m & g_m c_w^T \\ 0 & 0 & W \end{bmatrix} \begin{bmatrix} x_p(k) \\ x_m(k) \\ \omega(k) \end{bmatrix} =$$

$$\begin{aligned} &= -D^{-1} g^T P_{11} F_p x_p(k) - D^{-1} g^T P_{12} F_m x_m(k) - \\ &\quad - D^{-1} g^T [P_{12} g_m c_w^T + P_{13} W] \omega(k) = \\ &= -K_1 x_p(k) - K_2 x_m(k) - K_3 \omega(k) = \\ &= -K z(k). \end{aligned} \tag{3.4}$$

It is evident that the structure of the controller has three terms. There are the normal feedback gains on the plant state variables and in addition to these the model-in-the system control scheme incorporates feed-forward gains on both the model states, x_m, and the input states, w. The overall system is shown in Fig. 1 where E^{-1} is one sampling interval back shift operator.

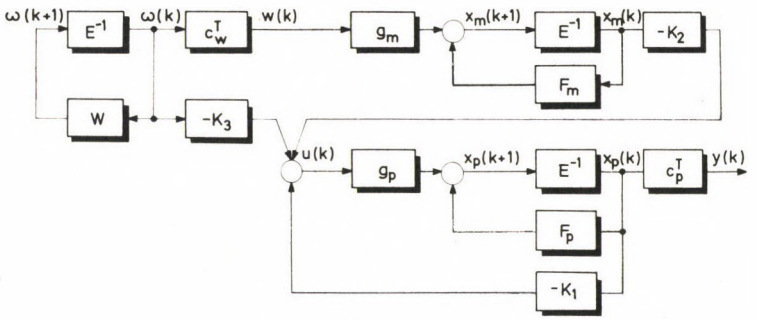


Fig. 1

Riccati equation (3.2) can be solved as a whole using appropriate procedure or it can be splitted in the following three equations

$$P_{11}=\Phi_{11}+F_p^T P_{11} g_p D^{-1} g_p^T P_{11} F+F_p^T P_{11} F_p \tag{3.5}$$

$$P_{12}=\Phi_{12}+F_p^T P_{11} g_p D^{-1} g_p^T P_{12} F_m+F_p^T P_{12} F_m \tag{3.6}$$

$$P_{13}=\Phi_{13}+F_p^T P_{11} g_p D^{-1} g_p^T [P_{12} g_m c_w^T+P_{13} W]+F_p^T P_{13} W. \tag{3.7}$$

Equation (3.5) is quadratic matrix equation as Eq. (3.2). Unique positive definite symmetric solution for P_{11} can be obtained provided that Φ_{11} is nonnegative definite and D symmetric positive definite. Substituting solution P_{11} into Eq. (3.6) yields a linear matrix equation in P_{12}. Finally, substituting P_{11} and P_{12} into Eq. (3.7) gives again a linear matrix equation in P_{13}.

Both last equations are of the general form

$$AX+XB=C. \tag{3.8}$$

They will have a unique solution for every Φ_{12} and Φ_{13} (see Ref. [2], p. 239) if the following conditions are satisfied.

$$\lambda_i+\mu_j\neq 0; \qquad i,j=1,2,\ldots \tag{3.9}$$

where λ_i are the eigenvalues of the square matrix A of the dimension $(n_p \times n_p)$ and μ_j are the eigenvalues of the square matrix B of the dimension $(n_m \times n_m)$ if P_{12} is sought for or of the dimension $(n_w \times n_w)$ for P_{13}, respectively. It is easy to prove that μ_j are eigenvalues of F_m^{-1} in the first case and of the matrix W^{-1} in the second case.

For more details concerning the solution of the Riccati equation (3.2) and (3.5) the reader is referred to [4] and [5]. A comprehensive theory of the matrix linear equation (3.8) is presented in [3].

If $W=0$, then the optimum control is solved for a step input. In practical terms the system performs quite well even for more general inputs, because the step input exites all frequencies and is, therefore, a fairly general command variable. Besides this

block, P_{13} of the matrix P is expressed by an explicit form in this case. In view of (2.9), $\Phi_{13}=0$, so that

$$P_{13}=F_p^T P_{11} g_p D^{-1} g_p^T P_{12} g_m c_w^T. \tag{3.10}$$

4. Conclusion

Design of a discrete control system extended by a model-in-the-system is the main objective of this article. Unlike the high gain controllers used in some current control systems increasing by the high gain, the total bandwidth and exciting higher order dynamic modes, the model-in-the-control-system differs from the conventional system by its incorporation of feed-forward links from all the model and input states. The links reduce the magnitude of the system state feedback gains and provide good model-following even if the plant parameters are uncertain.

References

1. *Asseo, S. J.*, Application of optimal control to "perfect" model following. J.A.C.C. Proceedings, June 1968, pp. 1056–1070.
2. *Bellman, R.*, Introduction to Matrix Analysis. McGraw-Hill Book Company, New York, 1970.
3. *Kučera, V.*, The matrix equation $AX+XB=C$. SIAM J. Appl. Math. **26** (1974), pp. 15–25.
4. *Strejc, V.*, State Space Theory of Discrete Linear Control. John Wiley and Sons, Chichester, 1981.
5. *Strejc, V.*, Riccati equation solved by matrix triangulization. Problems of Control and Information Theory **13** (1984), pp. 221–228.
6. *Tyler, J. S., Jr.*, The characteristics of model following systems as synthesized by optimal control. IEEE Trans. on Automatic Control **AC-9** (1964), pp. 485–498.

Дискретная версия модели в системе

В. СТРЕЙЦ

(Прага)

В статье рассматривается решение дискретной версии системы, следящей за моделью, которая применяет непосредственное включение модели в систему и критерий качества следования квадратичный. Объект управления и модель предполагаются линейными. Результат решения даёт один регулятор в обратной связи к объекту управления и два регулятора в прямой связи.

V. Strejc
Institute of Information Theory and Automation
Czechoslovak Academy of Sciences
Pod vodárenskou věží 4
182 08 Praha 8
Czechoslovakia

Problems of Control and Information Theory, Vol. 17 (2), pp. 77—90 (1988)

THE MATHEMATICAL MODEL OF FLEXIBLE MACHINING-CELL

YU. M. SOLOMENTSEV, E. B. FROLOV

(*Moscow*)

(Received March 12, 1987)

The mathematical model of the manufacturing process in the flexible automated machining-cell is defined and applied to the optimal control problem of the production schedule.

1. Introduction

On the basis of thorough study of the technological process with a full range of workpieces, a flexible manufacturing system (FMS) is developed, the goal of which is to ensure the desired machining quality and productivity [1], [2].

The economic effects of FMS depend critically on the functioning of the system with *internal* and *external* perturbations in its current production programme [2] such as:

1) failure of machines and auxiliary equipments, shortage of processing tool-wear, maintenance of some machines etc.,
2) delays in billet loading, orders of high priority, changes in range of processed workpieces and others.

Then the FMS control system [3] must timely compute a new sequence rule of loading the pieces in processing to achieve maximum economic benefits of flexible manufacturing. The FMS effectiveness largely depends on cyclic regularity of assembly of units and products, which contain the pieces manufactured by the system [1], [4]. And those are of various priorities due to their belonging to the same or different assembly items.

The FMS rescheduling problem is known to be NP-complete, thus with a great variety of piece-types, in varying quantities, simultaneously processed, its exact combinatorial solution is practically impossible.

Formalized models of manufacturing processes are applied to the problem of production schedule correction. The mathematical models of FMS described in literature fall into two basic groups [5], [6]:

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

(A) Evaluative Models that predict (evaluate) the performance of FMS and its subsystems in different situations. The dispatch control algorithms (piece input sequence [7], [8], processing routes) are to be known a priori.

(B) Generative Models, developed in order to generate optimal control modes of FMS schedule correction (machine setting-up moments, rerouting of pieces in-process taking into account alternate operations [7], [9], etc.) with a given set of FMS performance criteria.

Evaluative models based on various mathematical methods (queuing theory [6], [10], time-series analysis [5], [11], stochastic approximation techniques and perturbation analysis [5], [12], the Petri net methods [6], [13], and others [5], [14]–[16]), combined with modern software [17]–[20] (from programming languages, such as FORTRAN and PASCAL to GPSS, SIMCRIPT, MAP/I, ADA and etc.) help to find "weak points" and error decisions in manufacturing programme, to optimize the parameters of the flexible system and partially improve the set of control algorithms.

Useful during the stages of FMS design and long-term planning evaluative models fail to be efficient in the short-term production control problems [5] (with various perturbations in the current shift plan).

To generate such kind of control, the mathematical models of the second group are applied with their greater part being based on a large body of heuristics that exists for scheduling [21]–[23]. Some special problems dealing with control of flexible manufacturing systems can be successfully solved by traditional operations research methods (see, e.g. [6], [7], [24]–[31]).

However, improved results are achieved with adoption of control theory technique, involving formal description of the corresponding technological processes in the form of dynamic control system. An attempt to apply for this purpose, a vector space discrete-time model [32]–[38] fails as great dimension of the problem makes it computationally extremely demanding.

The authors of this work propose abandoning enormous calculation of processing trajectories (i.e. different stages of pieces in-process routes). The main aim of the paper is to determine the optimal number of various-type piece inputs routed for processing (setting-up moments) within the time cycle of the equipment performance [24]. A mathematical model of machining technology process in the form of a discrete control system is introduced for the purpose. The further application of the discrete maximum principle offers an effective solution of the production schedule correction problem in a flexible automated machining-cell.

2. Mathematical model of the FMS cell production

As an illustration we shall develop a mathematical model for a flexible automated machining-cell, consisting of four machines. Consider the machining of four kinds of workpieces that differ both in material and sizes. The manufacturing is of cyclic nature [4], [36], when the finished pieces are produced as a set of batches consisting of various numbers of units of one particular type. The set is periodically performed at given equal intervals, called cycle time.

The technological process is designed so as to provide the rough-machining of the accessible surfaces of piece-types 1, 2 and 3, 4 by M_1 (milling machine) and M_2 (lathe), respectively, during the first stage. Then, during the second processing stage pieces of types 1, 3 and 2, 4 are launched onto machining-centres M_3 and M_4, correspondingly, where being repositioned, these pieces undergo finish-machining. Thus, each machine M_j ($j=1, 2, 3, 4$) can process two types of pieces. A formal scheme of the technological process is presented in Fig. 1.

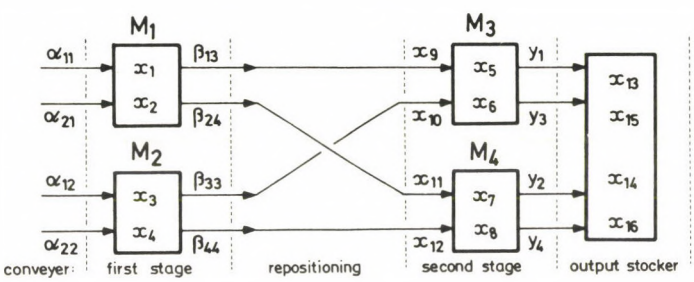


Fig. 1

Partially processed by M_1, pieces of types 1 and 3 in quantities β_{13} and β_{24} (the value of β_{il} depends on cyclic productivity of machine M_j in i-type pieces; l — machine, onto which the pieces are routed for further processing) get to load/unload area for repositioning.

The numbers $x_1(k)$ and $x_2(k)$ of pieces referring to types 1 and 2, waiting during the k-th cycle to be machined by M_1, satisfy the following balance equation:

$$\begin{aligned} x_1(k+1) &= x_1(k)+\alpha_{11}(k)-\beta_{13}(k) \\ x_2(k+1) &= x_2(k)+\alpha_{21}(k)-\beta_{24}(k). \end{aligned} \tag{2.1}$$

Similarly, for pieces of types 2, 4 and the second machine M_2 we obtain

$$\begin{aligned} x_3(k+1) &= x_3(k)+\alpha_{32}(k)-\beta_{33}(k) \\ x_4(k+1) &= x_4(k)+\alpha_{42}(k)-\beta_{44}(k). \end{aligned} \tag{2.2}$$

Here the variables $x_3, x_4, \alpha_{32}, \alpha_{42}, \beta_{33}, \beta_{44}$ have the same sense as the variables x_1, x_2, $\alpha_{11}, \alpha_{21}, \beta_{13}, \beta_{24}$ for machine M_1 in (2.1).

In the load/unload area repositioning of workpieces takes place: β_{13} items turn into x_9 pieces of type 1, β_{24} items — into x_{11} pieces of type 2 ($\beta_{13}=x_9$, $\beta_{24}=x_{11}$) etc.

In the next $(k+1)$-th time-cycle workpieces in quantities $x_{4+i}(k+1)$ $(i=1, 2, 3, 4)$ again arrive at the processing zone of the machining-centres M_3 and M_4.

Finally, finished workpieces of various ranges in quantities $y_i(k)$ (where i-type of pieces, $i=1, 2, 3, 4$) are routed to the output stocker of the cell.

Having developed similar (2.1) and (2.2) balance equations for the machining-centres M_3 and M_4, we obtain processing state equations for workpieces of types 1, 3 and 2, 4 respectively, in accordance with the mentioned technology:

$$\begin{aligned}
x_1(k+1)&=x_1(k)+\alpha_{11}(k)-\beta_{13}(k)\\
x_2(k+1)&=x_2(k)+\alpha_{21}(k)-\beta_{24}(k)\\
x_3(k+1)&=x_3(k)+\alpha_{32}(k)-\beta_{33}(k)\\
x_4(k+1)&=x_4(k)+\alpha_{42}(k)-\beta_{44}(k)\\
x_5(k+1)&=x_5(k)+x_9(k)-y_1(k)\\
x_6(k+1)&=x_6(k)+x_{10}(k)-y_3(k)\\
x_7(k+1)&=x_7(k)+x_{11}(k)-y_2(k)\\
x_8(k+1)&=x_8(k)+x_{12}(k)-y_4(k).
\end{aligned} \tag{2.3}$$

Assume that the cycle time is T and let t_j be the loss of working time (downtime period) in the cycle due to the new set-up of the j-th machine. Then

$$1-t_j/T=(\gamma_{ij}+\gamma_{lj})/T, \tag{2.4}$$

where γ_{ij} and γ_{lj} represent the duration of time in which pieces of types i and l are processed by machine M_j $(j=1, 2, 3, 4)$ within the entire cycle time T $(1\leqq i, l\leqq 4)$.

Following [36], consider the values

$$\begin{cases} u_1(k)=\gamma_{11}(k)/T, & u_2(k)=\gamma_{42}(k)/T,\\ u_3(k)=\gamma_{13}(k)/T, & u_4(k)=\gamma_{44}(k)/T \end{cases} \tag{2.5}$$

as control variables.

Their values u_j $(j=1, 2, 3, 4)$ demonstrate what part of the cycle-time M_1 and M_3 process type 1 pieces, and M_2, M_4-type 4 pieces (see Fig. 1).

Let S_{ij} be the maximum amount of i-type workpieces that can be processed by machine M_j during the cycle. In case of M_j-failure $S_{ij}=0$.

Due to (2.4) and (2.5), the output of machines M_1 and M_2 is:

$$\begin{aligned}
\beta_{13}(k) &= S_{11}u_1(k) \\
\beta_{24}(k) &= S_{21}(1-t_1/T-u_1(k)) \\
\beta_{33}(k) &= S_{32}(1-t_2/T-u_2(k)) \\
\beta_{44}(k) &= S_{42}u_2(k).
\end{aligned} \tag{2.6}$$

The number of pieces of varying types arriving at the FMS cell output stocker within the k-th time cycle is expressed by the formula:

$$\begin{aligned}
y_1(k) &= S_{13}u_3(k) \\
y_2(k) &= S_{24}(1-t_4/T-u_4(k)) \\
y_3(k) &= S_{33}(1-t_3/T-u_3(k)) \\
y_4(k) &= S_{44}u_4(k).
\end{aligned} \tag{2.7}$$

As vacant pallets are again refurnished with new billets, coming by input/output conveyer 5 for the first stage of processing, the following relations are valid:

$$\alpha_{11}+\alpha_{21}+\alpha_{32}+\alpha_{42}=\sum_{j=1}^{4} y_j \leqq d_0, \tag{2.8}$$

(here d_0 is input/output capacity of the conveyer), while the total number of pallets in the machining-cell is limited:

$$\begin{aligned}
x_1+x_2+x_5+x_7+(S_{11}-S_{21})u_1+(1-t_1/T)S_{21} &\leqq d_{12} \\
x_3+x_4+x_6+x_8+(S_{42}-S_{32})u_2+(1-t_2/T)S_{32} &\leqq d_{34}
\end{aligned}. \tag{2.9}$$

It is proposed in (2.9) that different pallets in quantities d_{12} and d_{34} are used for billets of types 1, 2 and 3, 4, correspondingly.

Additional constraints on the system variables are imposed by their physical restrictions:

$$\begin{aligned}
&x_i \geqq 0 && i=1,2,3,4,5,6,7,8, \\
&0 \leqq u_j \leqq 1 && j=1,2,3,4, \\
&\text{the values} && S_{ij}u_j \text{ and } S_{lj}(1-t_j/T-u_j) \text{ are integers,} \\
& && j=1,2,3,4, \quad 1<i,\ l<4.
\end{aligned} \tag{2.10}$$

Thus, according to (2.5), control variable $u_j(k)$, when $0<u_j(k)<1$, in fact determines setting-up moments of machine M_j during the k-th time cycle. $S_{lj}u_j(k)$ is the number of inputs of l-type pieces to M_j ($l=1,4$).

In conformity with the mentioned technology the manufacturing process in the analysed FMS cell is described, as follows from (2.3) and (2.6), in the form of a discrete-time control system

$$
\begin{aligned}
x_1(k+1) &= x_1(k) - S_{11}u_1(k) + \alpha_{11}(k) \\
x_2(k+1) &= x_2(k) - S_{21}(1 - t_1/T - u_1(k)) + \alpha_{21}(k) \\
x_3(k+1) &= x_3(k) - S_{32}(1 - t_2/T - u_2(k)) + \alpha_{32}(k) \\
x_4(k+1) &= x_4(k) - S_{42}u_2(k) + \alpha_{42}(k) \\
x_5(k+1) &= x_5(k) + x_9(k) - S_{13}u_3(k) \\
x_6(k+1) &= x_6(k) + x_{10}(k) - S_{33}(1 - t_3/T - u_3(k)) \\
x_7(k+1) &= x_7(k) + x_{11}(k) - S_{24}(1 - t_4/T - u_4(k)) \\
x_8(k+1) &= x_8(k) + x_{12}(k) - S_{44}u_4(k) \\
x_9(k+1) &= S_{11}u_1(k) \\
x_{10}(k+1) &= S_{32}(1 - t_2/T - u_2(k)) \\
x_{11}(k+1) &= S_{21}(1 - t_1/T - u_1(k)) \\
x_{12}(k+1) &= S_{42}u_2(k)
\end{aligned}
\tag{2.11}
$$

with output (2.7) and constraints (2.8), (2.9), (2.10). Here variables x_1, x_2 and x_3, x_4 denote the number of pieces of varying types waiting to be processed by machines M_1 and M_2, respectively, x_5, x_6 and x_7, x_8 the number of pieces queuing for finish processing (after repositioning) by machining centres M_3 and M_4. Auxiliary variables x_9, x_{11} and x_{10}, x_{12} characterize the production capacity of M_1 and M_2, correspondingly, in the preceding time-cycle.

System (2.11) described the production state $x = (x_1, \ldots, x_{12})$ of the machining cell as a function of input flows $\alpha_{ij}(k)$ and control vector values $u = (u_1, u_2, u_3, u_4)$ for all $k = 0, 1, 2, \ldots$ The initial state:

$$
\begin{aligned}
&x_i(0) = x_{i0}, \qquad i = 1, 2, 3, 4, 5, 6, 7, 8, \\
&x_9(0) = x_{10}(0) = x_{11}(0) = x_{12}(0) = 0.
\end{aligned}
$$

3. Shift plan perturbations in the FMS cell

Any perturbations in the assigned manufacturing plan, occurring for organizational reasons [2] (lack of needed billets, necessary tools and devices, selection of high priority workpieces, directive modification in the range of productions etc.) cause, in the long run, changes in the current shift plan of the machining cell.

The FMS cell rescheduling problem is NP-complete [1], [33], and thus it cannot be successfully solved in the real time [27].

On the other hand, analysing the dynamic nature of the given technological processes in the form (2.11) makes it possible to obtain a solution of the production schedule correction problem related to a FMS cell. The following assumptions are to be made:

C1: There are no failure-prone machines in the cell: $S_{ij} = \text{const} > 0$;

C2: There is not more than one setting-up moment of machine M_j $(j = 1, 2, 3, 4)$ within a single cycle; consequently, the loss of the working time due to machine setting-up is defined as:

$$\begin{aligned} t_j(k) &= \tau_j \delta_j(k, k-1) \\ \delta_j(k, k-1) &= \operatorname{sgn} \{\sigma_j + (1-\sigma_j) \mid \operatorname{sgn} u_j(k) - \operatorname{sgn} u_j(k-1) \mid \} \end{aligned} \tag{3.1}$$

where $\sigma_j = \operatorname{sgn} \{u_j(k)(1-u_j(k))\}$,

τ_j – setting-up time of M_j,

δ_j – the total number of M_j set-ups;

(if $u_j(k) = 1$ or $u_j(k) = 0$, then $\sigma_j = 0$, and so machine M_j is not subjected to resetting only on condition that in the previous cycle the control variable $u_j(k-1)$ is of the same value as $u_j(k)$,

C3: the constraint (2.8) is not essential—the input/output capacity d_0 of conveyer 5 is big enough;

C4: the constraints (2.10) are relaxed; it is assumed that

$$x_i(k) > 0 \qquad i = 1, 2, 3, 4, 5, 6, 7, 8, \quad k = 1, 2, \ldots, \tag{3.2}$$

(i.e. in the FMS cell there are always queuing workpieces of various types available for processing by M_j, $j = 1, 2, 3, 4$) like the relationship (2.10) $0 < u_j(k) < 1$, $j = 1, 2, 3, 4$, but he values $S_{ij}u_j(k)$ and $S_{lj}(1 - t_j/T - u_j(k))$ are real. $1 < l$, $i < 4$, $j = 1, 2, 3, 4$.

Condition C4, in particular, means that workpieces of a certain type may be lacking only at the beginning of the first cycle, i.e. the initial state of system (2.11) will have some zero components. With all the other values of $k = 1, 2, 3, \ldots, N$, the condition $x_i(k) > 0$ mentioned in (3.2) is valid, i.e. pallet-loading control service guarantees the proper input flow of billets $\alpha_{11}, \alpha_{21}, \alpha_{32}, \alpha_{42}$.

Planned machine-loading index, characterizing the expected productivity of the machining cell, specifies constraints for the total number of set-ups within the planning horizon $[0, N]$:

$$T_N = \sum_{j=1}^{4} \sum_{k=0}^{N-1} t_j(k) \leqq T_{np}, \tag{3.3}$$

here T_N and T_{np} are real and planned downtime periods of the cell equipment (caused by set-ups) during N cycles.

To equations (2.11) the state equations of the output stocker must be added

$$\begin{aligned}
x_{13}(k+1) &= x_{13}(k) + S_{13}u_3(k) \\
x_{14}(k+1) &= x_{14}(k) + S_{24}(1 - t_4/T - u_4(k)) \\
x_{15}(k+1) &= x_{15}(k) + S_{33}(1 - t_3/T - u_3(k)) \\
x_{16}(k+1) &= x_{16}(k) + S_{44}u_4(k) \\
x_{12+j}(0) &= 0, \qquad j = 1, 2, 3, 4,
\end{aligned} \tag{3.4}$$

here $x_{12+j}(k)$ is the number of finished j-type pieces, collected in the cell output stocker within k cycles.

Note that the variables t_j present in (2.9), (2.10), (2.11), (3.3), (3.4), are derived from (3.1). That is why the system (2.11), (3.4) is nonlinear.

Thus, the state of the manufacturing process is described by a discrete-time control system (2.11), (3.1), (3.4) with constraints (2.8), (2.9), (3.2), (3.3), and production characteristics during N cycles are defined by the minimized functional

$$I = \Phi(x(0), x(N)) + \sum_{k=0}^{N-1} \varphi_k(x(k), u(k), k). \tag{3.5}$$

Functional (3.5) is the optimality criterion of the manufacturing process within the given planning horizon $[0, N]$.

Obviously, the values of the state vectors $x(k)$ and $x(k+1)$ of system (2.11), (3.4) make it possible to calculate downtime periods of the cell machines in the previous cycle:

$$\begin{aligned}
t_1(k) &= T\{1 - x_9(k+1)/S_{11} - x_{11}(k+1)/S_{21}\} \\
t_2(k) &= T\{1 - x_{10}(k+1)/S_{32} - x_{12}(k+1)/S_{42}\} \\
t_3(k) &= T\{(x_{13}(k) - x_{13}(k+1))/S_{13} + (x_{15}(k) - x_{15}(k+1))/S_{33} + 1\} \\
t_4(k) &= T\{(x_{14}(k) - x_{14}(k+1))/S_{24} + (x_{16}(k) - x_{16}(k+1))/S_{44} + 1\}.
\end{aligned} \tag{3.6}$$

In view of (3.6) the left part of relationship (3.3) limiting the loss of working time during N cycles takes the form:

$$T_N(x) = T\sum_{k=0}^{N-1}[2 - x_9(k)/S_{11} - x_{10}(k)/S_{32} - x_{11}(k)/S_{21} - x_{12}(k)/S_{42}] + 2N - \\ - T(x_{13}(N)/S_{13} + x_{14}(N)/S_{24} + x_{15}(N)/S_{33} + x_{16}(N)/S_{44}). \tag{3.7}$$

It is clear that the downtime restriction (3.3) may be duly ensured by including T_N in functional (3.5).

If the altered shift plan of the FMS cell (a set of 4 different type batches of pieces) is represented as demand vector $z = (z_1, z_2, z_3, z_4)$ where z_j is the needed number of j-type pieces in the batch, then the trivial production schedule correction problem for the machining cell within N cycles involves the optimality criterion

$$I = \sum_{j=1}^{4} c_j(x_{12+j}(N) - z_j)^2 + c_0 T_N(x), \tag{3.8}$$

here c_j are penalty weights for failed shift plan in j-type pieces, c_0 is the average downtime manufacturing cost of cell machines.

Perturbations in the shift plan can also be expressed by alterations of functional (3.5). For example, the task "to implement the shift plan z for N cycles so that the processing time of type-four workpieces were minimum" presupposes minimizing the following functional:

$$I = \sum_{j=1}^{4} c_j(x_{12+j}(N) - z_j)^2 + \sum_{k=0}^{N-1} kS_{44}u_4(k) + c_0 T_N(x). \tag{3.9}$$

A similar problem is treated in [7], [35].

On the other hand, the requirement "to guarantee the fulfilment of the final task working to the planned production schedule $z_j(k)$ $j = 1, 2, \ldots, N$" means the minimization of the functional

$$I = \sum_{k=0}^{N}\sum_{j=1}^{4} c_j(x_{12+j}(k) - z_j(k))^2 + c_0 T_N(x). \tag{3.10}$$

Other forms of functionals, used in control problems that arise from automated machining processes, can be found in contemporary literature [7], [8], [24], [27], [36].

Consequently, even with a fixed value of vector z, any change in requirements to the nature of the manufacturing process (alteration in the form of functional (3.5), e.g. (3.8), (3.9) or (3.10)) is, in fact, a perturbation of the FMS cell shift plan and calls for production schedule correction.

4. Optimal correction of the FMS cell production schedule

The problem of the optimal production schedule correction for the flexible automated machining-cell analyzed in assumptions C1–C4 is equivalent to the control problem of a discrete-time system (2.11), (3.1), (3.4) with constraints (2.9), (3.2) and functional (3.5). The initial state $x(0)$ is considered to be known:

$$x_i(0)=x_{i0}\,, \qquad x_{i+8}(0)=0, \qquad i=1,2,3,4,5,6,7,8\,.$$

Following the discrete maximum principle [37], [39] the desired optimal control $u(k)$ maximizes the Hamilton function for every $k=0,1,\ldots,N-1$.

For the given system we shall introduce a Hamiltonian [39] by the formula

$$H(x(k),u(k),\psi(k+1))=\sum_{i=1}^{16}\psi_i(k+1)f_i(x(k),u(k),k)-$$

$$-\varphi_k(x(k),u(k),k)+\mu_1\xi_1(x(k),u(k))+\mu_2\xi_2(x(k),u(k))\,, \tag{4.1}$$

where f_i are the right parts of system (2.11) written as

$$x_i(k+1)-x_i(k)=f_i(x(k),u(k),k) \qquad (i=1,\ldots,16)\,,$$

the function $\varphi_k(x(k),u(k),k)$ is present in (3.5), and $\xi_1\,,\xi_2$ are defined by:

$$\begin{cases}\xi_1=x_1+x_2+x_5+x_7+(S_{11}-S_{21})u_1+(1-t_1/T)S_{21}-d_{12}\\ \xi_2=x_3+x_4+x_6+x_8+(S_{42}-S_{32})u_2+(1-t_2/T)S_{32}-d_{34}\,.\end{cases} \tag{4.2}$$

$\mu_1\,,\mu_2$ are auxiliary Lagrange multipliers, related to restrictions (2.9).

$$\begin{cases}\mu_j(k)=0 & \text{if} \quad \xi_j(x(k),u(k))<0\,,\\ \mu_j(k)<0 & \text{if} \quad \xi_j(x(k),u(k))>0\,, \qquad j=1,2\,.\end{cases}$$

$\psi_i(k+1)$ in (4.1) are the state variables of the adjoint system

$$\psi_i(k+1)-\psi_i(k)=-\partial H/\partial x_i(x(k),u(k),\psi(k+1)) \tag{4.3}$$

$$i=1,2,\ldots,16\,.$$

Adjoint system (4.3) related to (2.11), (3.4) can be written as

$$\begin{aligned}
&\psi_i(k)=\psi_i(k+1)+\mu_1(k)-\partial\varphi_k/\partial x_i && i=1,2,5,7\,,\\
&\psi_j(k)=\psi_j(k+1)+\mu_2(k)-\partial\varphi_k/\partial x_j && j=3,4,6,8\,,\\
&\psi_i(k)=\psi_{i-4}(k+1)-\partial\varphi_k/\partial x_i && i=9,10,11,12\,,\\
&\psi_j(k)=\psi_j(k+1)-\partial\varphi_k/\partial x_j && j=13,14,15,16\,.
\end{aligned} \tag{4.4}$$

The terminal constraints for (4.4) are derived from transversal conditions [39]:

$$\psi(N) = -\partial\Phi(x(0), x(N))/\partial x(N)\,.$$

Thus, the problem of the optimal production schedule correction for the analyzed cell (with assumptions C1–C4) results in determining the vector of optimal control variables $u(k)=(u_1(k), u_2(k), u_3(k), u_4(k))$ which maximizes the Hamiltonian function (4.1) for every $k=0, 1, \ldots, N-1$, $x(k)$ and $\psi(k+1)$ are the solutions of system (2.11), (3.1), (3.4) and system (4.4), correspondingly [37], [39], [40].

Finally, it must be noted that the functions φ_k in (3.5) being convex in u (or independent of the variable u—see (3.8) and (3.10)), the control variables u_1 and u_2 from the maximized Hamiltonian (4.1) have their ultimate values

$$\begin{cases} u_1 = \min\{1, [d_{12} - x_1 - x_2 - x_5 - x_7 - (1 - t_1/T)S_{21}]/(S_{11} - S_{21})\} \\ u_2 = \min\{1, [d_{34} - x_3 - x_4 - x_6 - x_8 - (1 - t_2/T)S_{32}]/(S_{42} - S_{32})\} \end{cases} \tag{4.5}$$

if and only if the following conditions are valid

$$\begin{cases} S_{11}(\psi_9(k+1) - \psi_1(k+1)) - S_{21}(\psi_{11}(k+1) - \psi_2(k+1)) \geqq -\partial\varphi_k/\partial u_1(x(k), u(k)) \\ S_{42}(\psi_{12}(k+1) - \psi_4(k+1)) - S_{32}(\psi_{10}(k+1) - \psi_3(k+1)) \geqq -\partial\varphi_k/\partial u_2(x(k), u(k))\,. \end{cases}$$

So, the optimal values of the multipliers μ_1 and μ_2 from (4.1) are calculated by formulas [37]:

$$\mu_1(k) = \min\left\{0, \left[\frac{\partial\varphi_k}{\partial u_1}(x(k), u(k)) + S_{11}(\psi_9(k+1) - \psi_1(k+1)) - \right.\right.$$
$$\left.\left. - S_{21}(\psi_{11}(k+1) - \psi_2(k+1))\right]/(S_{21} - S_{11})\right\}, \tag{4.6}$$

$$\mu_2(k) = \min\left\{0, \left[\frac{\partial\varphi_k}{\partial u_2}(x(k), u(k)) + S_{42}(\psi_{12}(k+1) - \psi_4(k+1)) - \right.\right.$$
$$\left.\left. - S_{32}(\psi_{10}(k+1) - \psi_3(k+1))\right]/(S_{32} - S_{42})\right\}. \tag{4.7}$$

Substitution of (4.6) and (4.7) into (4.1) ensures (with constraints (2.9) and (3.2)) the conditions $\partial H/\partial u_1 = 0$ and $\partial H/\partial u_2 = 0$ along the trajectory $x(k)$ and the desired control $u(k)$, $k=0, 1, \ldots, N-1$. The values $u_j(k)$ $(j=1, 2, 3, 4)$ generated with the help of the described procedure specify setting-up moments of machine M_j.

The optimal number of i- and l-type piece-inputs routed for processing by machine M_j within the k-cycle is integers nearest to $S_{ij}u_j(k)$ and $S_{lj}(1 - t_j(k)/T - u_j(k))$, $j=1, 2, 3, 4$. Their timely computation allows an *effective* solution of the production schedule correction problem in the flexible automated machining cell.

5. Conclusions

1. The production schedule correction problem for a flexible machining-cell, being NP-complete combinatorial problem, can in some cases be efficiently solved by control theory methods.

2. The mathematical machining-cell model, yielding an effective solution of the production schedule correction problem, is developed as a result of describing piece-processing technology (by means of available equipment) as a discrete-time control system.

3. Optimal correction of the manufacturing process in an FMS cell is obtained sequentially during several cycles of the equipment performance. The proper number of workpiece inputs and machine setting-up moments are defined by determining optimal control for a discrete-time system (with constraints to control and state variables) on the basis of the discrete maximum principle.

References

1. Новые методы управления производством за рубежом. Экспресс-информация "Приборы, средства автоматизации и системы управления", **ТС-3**, вып. *1*, Москва, 1986.
2. *Соломенцев Ю. М., Кутин А. А., Шептунов С. А.*, Оценка гибкости автоматизированной станочной системы. Вестник машиностроения, 1984, No. *1*, 38–40.
3. *Соломенцев Ю. М.*, Возможности систем управления в гибком автоматизированном производстве. Микропроцессорные средства и системы, 1984, No. *2*, 73–74.
4. *Лищинский Л. Ю. М.*, Циклический производственный процесс в гибком автоматизированном производстве. Станки и инструмент, 1986, No. *7*, 2–6.
5. *Gershwin, S., Hildebrant, R., Suri, R., Mitter, S.*, A control perspective on recent trends in manufacturing systems. IEEE Control Systems Magazine, 1986, Vol. **6,** No. *2*, 3–15.
6. Робототехника и гибкие автоматизированные производства. Кн. **5.** Моделирование робототехнических систем и гибких автоматизированных производств. (Пантюшин С. В., Назаретов В. М., Тягунов О. А. и др.) Высшая школа, Москва, 1986.
7. Робототехника и гибкие автоматизированные производства. Кн. **3.** Управление робототехническими системами и гибким автоматизированным производств. (Макаров И. М., Рахманкулов В. З., Назаретов В. М. и др.). Высшая школа, Москва, 1986.
8. *Wroblewski, K., Krawczynski, R.*, Priority rules in production flow control. Material Flow, 1985, No. *2*, 167–177.
9. *Wilhelm, W. E., Shin, H. M.*, Effectiveness of alternative operations in a flexible manufacturing, system. International Journal of Production Research, 1985, Vol. **23,** No. *1*, 65–79.
10. *Buzacott, J. A., Shanthikumar, J. G.*, Models for understanding flexible manufacturing systems. AIIE Transactions, 1980, Vol. **12,** No. *2*, 339–350.
11. *Law, A. M.*, Statistical analysis of simulation output data. Operations Research, 1983, Vol. **31,** No. *6*, 983–1029.
12. *Ho, Y. C., Cao, X. R.*, Perturbation analyzis and optimization of queuing networks. Journal of Optimization Theory and Applications, 1983, Vol. **40,** No. *4*, 559–582.
13. *Dubois, D., Stecke, K.*, Using Petry Nets to represent production processes. Proceedings of the IEEE Conference on Decision and Control, 1983.

14. *Yao, D., Buzacott, J. A.*, Modeling the performance of flexible manufacturing systems. International Journal of Production Reseach, 1985, Vol. **23,** No. *5*, 945–959.
15. *Ho, Y. C., Eyeler, M., Chien, T. T.*, A gradient technique for general buffer storage design in a serial production line. International Journal of Production Research, 1979, Vol. **17,** No. *6*, 557–580.
16. *Suri, R.*, An overview of evaluative models for flexible manufacturing systems. Annals of Oper. Research, 1985.
17. *Schriber, T.*, The use of GPSS/H in modeling a typical flexible manufacturing system. Annals of Oper. Research, 1985.
18. *Musselman, K.*, Computer simulation: a design tool for FMS. Manufacturing Engineering, 1984, Vol. **92,** No. *9*, 117–120.
19. *Bevans, J. P.*, First choose an FMS simulator. American Machinist, May 1982, 143–145.
20. *Micheletti, G., Cantino, V., Fenoglio, F.*, An application of simulation theory in FMS. Robotica, 1985, Vol. **3,** No. *2*, 253–257.
21. *Graves, S.* A review of production scheduling. Operations Research, 1981, Vol. **29,** No. *4*, 641–675.
22. *Russell, R., Taylor, B.*, An evaluation of scheduling policies in a dual resource constrained assembly shop. IIE Transactions, 1985, Vol. **17,** No. *3*, 219–232.
23. *Yamamoto, M., Nof, S. Y.*, Scheduling/rescheduling in the manufacturing operating system environment. International Journal of Production Research, 1985, Vol. **23,** No. *4*, 705–722.
24. *Смоляр Л. И.*, Модели оперативного планирования в дискретном производстве. Москва, Наука, 1978.
25. *Блехерман М. Х.*, Организационно-технологическое группирование деталей в ГПС. Вестник машиностроения, 1986, No. *6*, 37–42.
26. *De Luca, A.*, Optimal production planning for FMS: an "Optimum batching" algorithm. Proceedings of the 3-rd International Conference on Flexible Manufacturing Systems. Boeblingen, 1984, 323–332.
27. *Hildebrant, R.*, Scheduling flexible machining systems using mean value analysis. Proceedings, IEEE Conference on Decision and Control, 1980, 701–706.
28. *Stecke, K. E.*, Formulation and solution of nonlinear integer production planning problems for flexible manufacturing systems. Management Science, 1983, Vol. **29,** No. *2*, 273–288.
29. *Shanker, K., Tzen, Y.*, A loading and dispatching problem in a random manufacturing system. International Journal of Production Research, 1985, Vol. **23,** No. *3*, 579–595.
30. *Kusiak, A.*, Flexible manufacturing systems: a structural approach. International Journal of Production Research, 1985, Vol. **23,** No. *6*, 1057–1073.
31. *Graves, S.*, Using Lagrangian techniques to solve hierarchical production planning problems. Management Science, 1982, Vol. **28,** No. *3*, 260–275.
32. *Axsater, S.*, Control theory concepts in production and inventory control. International Journal on Systems Science, 1985, Vol. **16,** No. *2*, 161–169.
33. *Olsder, G. J., Suri, R.*, Time-optimal control of parts-routing in a manufacturing system with failure-prone machines. Proceedings, IEEE Conference on Decision and Control, 1980, 722–727.
34. *Teny, R. M., Kochar, A. K.*, A vector space model approach to production planning, multi-cell and multi-stage manufacturing system. International Journal on Systems Science, 1984, Vol. **15,** No. *11*, 1169–1187.
35. *Janiak, A.*, Optimalizacja czasowa sekwencji kompleksow operacji w sterowaniu dyskretnymi procesami produkcyjnymi (in Polish). Politechnika Slaska Zeszyty Naukowe, 1980, No. 650, 57–65.
36. *Gruver, W. A., Narasimhan, S. L.*, Optimal scheduling of multi-stage continuous flowshops. INFOR, 1981, Vol. **19,** No. *4*, 318–329.
37. *Kleindofer, P. B., Kriebels, C. H., Thompson, G. L., Kleindofer, G. B.*, Discrete Optimal Control of production plans. Management Science, 1975, Vol. **22,** No. *3*, 261–273.
38. *Hitomi, K., Nakamura, N.* Optimal production planning for a multiproduct, multistage production system. International Journal of Production Research, 1976, Vol. **14,** No. *2*, 199–213.
39. *Сейдж Э. П. Уайт Ч. С.*, Оптимальное управление системами. Москва, Радио и связь, 1982.
40. *Outrata, J. V.*, Minimization of nonsmooth nonregular functions: application to discrete-time optimal control problems. Problems of Control and Information Theory, 1984, Vol. **13,** No. *6*, 413–424.

Математическая модель участка гибкой производственной системы

Ю. М. СОЛОМЕНЦЕВ, Е. В. ФРОЛОВ

(Москва)

Экономические показатели работы ГПС существенно зависят от функционирования системы при возникновении *внутренних* и *внешних* отклонений в ее текущей производственной программе. Система управления ГПС должна оперативно рассчитать новый порядок запуска деталей на обработку, так, чтобы экономическая эффективность гибкого производства была бы максимальной.

Перерасчет производственного расписания ГПС является NP-трудной проблемой, поэтому для большого числа одновременно обрабатываемых разнотипных деталей ее точное решение с помощью комбинаторных методов практически недостижимо.

Для задач коррекции производственной программы ГПС используются формальные модели, описывающие технологические процессы обработки деталей. Имеющиеся в литературе математические модели гибких систем можно разделить на две основные группы:

A. Имитационные модели (Evaluative Models),
B. Модели для генерации управления (Generative Models).

Используемые на этапах проектирования ГПС и долгосрочного планирования, имитационные модели оказываются малоэффективными в задачах оперативного управления производством (при различных отклонениях в текущем сменно-суточном задании). Для такого управления применяются математические модели второй группы.

Авторы настоящей работы предлагают отказаться от трудоемких вычислений технологических траектории (отдельных этапов обработки деталей различной номенклатуры). Основная цель — определение оптимального *числа запусков* деталей различных типов (моментов переналадок станков) в пределах временного цикла работы оборудования. Для этого строится математическая модель технологического процесса механообработки в виде некоторой дискретной управляемой системы. Последующее применение дискретного принципа максимума позволяет эффективно решить задачу коррекции производственной программы гибкого автоматизированного участка механообработки.

Ю. М. Соломенцев
Московский станкоинструментальный институт
СССР, Москва, Вадковский пер., 3а

Problems of Control and Information Theory, Vol. 17 (2), pp. 91—98 (1988)

ELEMENTARY PROOF OF RUEPPEL'S LINEAR COMPLEXITY CONJECTURE AND A CONSTRUCTION OF SEQUENCES WITH HIGH LINEAR COMPLEXITY AND GOOD CORRELATION PROPERTIES[1]

NGUYEN QUANG A

(Hanoi)

(Received March 8, 1987)

We prove that, for all $n>0$, the subsequence consisting of the first n digits of the semi-infinite sequence $s=(s_0, s_1, \ldots)$ defined over a (finite or infinite) field F such that s_i is nonzero if and only if $i+1$ is a power of 2, has linear complexity $\lfloor (n+1)/2 \rfloor$. If F is the finite field $GF(2)$, our results prove the conjecture of Rueppel, and in this case we have exactly one semi-infinite sequence having the properties mentioned. If F has more than two elements we have infinitely many sequences having these properties. Based on these sequences, we give a construction for sequences having high linear complexity and good correlation properties.

1. Introduction

In 1963 J. L. Massey introduced the following semi-infinite binary sequence in connection with self-orthogonal convolution codes [1]

$$\mathbf{s}=(s_0, s_1, s_2, \ldots)=(1, 1, 0, 1, 0, 0, 0, 1, \ldots), \quad s_i=1 \text{ iff } i+1 \text{ is a power of 2.} \tag{1}$$

We generalize the Massey's sequence over an arbitrary (finite or infinite) field F as follows

$$\mathbf{s}=(s_0, s_1, s_2, \ldots)=(t, x, 0, y, 0, 0, 0, z, \ldots), \quad s_i \in F \text{ and } s_i \neq 0 \text{ iff } i+1 \text{ is a power of 2.} \tag{2}$$

Let $\mathbf{s}(n)=(s_0, s_i, \ldots, s_{n-1})$, $s_i \in F$, denote an arbitrary sequence of length n over F. A linear feedback shift register, LFSR (see Fig. 1) will be identified with its connection polynomial

$$C(x)=1+c_1x+c_2x^2+\ldots+c_Lx^L, \quad c_i \in F.$$

[1] A version of this paper has been accepted for presentation at EUROCRYPT'86, 20–22 May 1986, Lingköping, Sweden.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

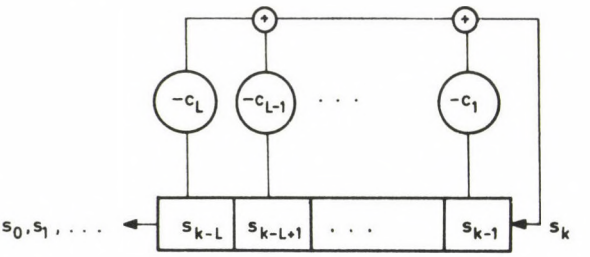


Fig. 1

The LFSR $C(x)$ is said to generate $\mathbf{s}(n)$ if the initial contents of the LFSR can be chosen in such a way that $s_0, s_1, \ldots, s_{n-1}$ are the first n output digits from the LFSR, i.e. with $n > L$

$$s_k + c_1 s_{k-1} + \ldots + c_L s_{k-L} = 0, \qquad L \leqq k < n. \tag{3}$$

The *linear complexity* of a finite or semi-infinite sequence is the length of the shortest LFSR that generates the sequence or, equivalently, is the smallest degree L for which there exists an LFSR $C(x)$ that generate the sequence.

Let $L_s(n)$ denote the linear complexity of the subsequence consisting of the first n digits, i.e. $\mathbf{s}(n)$, of a sequence $\mathbf{s}$. R. A. Rueppel [2] conjectured that if $\mathbf{s}$ is defined by (1) then

$$L_s(n) = \lfloor (n+1)/2 \rfloor \tag{4}$$

where $\lfloor a \rfloor$ denotes the integer part of a, and showed that the "linear complexity profile" given by (4) virtually coincides with the high probability linear complexity profile of a truly random binary sequence.

Z. D. Dai [3] proved the conjecture. Her proof utilizes properties of an element in an extension field of the field of rational functions over $GF(2)$. Later, the author gave a more elementary proof of the conjecture. It turns out that our proof is valid without any change for the case of sequences defined by (2), too. We summarize our result as

Theorem 1. Let $\mathbf{s}$ be semi-infinite sequence defined by (2) over arbitrary field F. Let $\mathbf{s}(n)$ be the subsequence consisting of the first n digits of $\mathbf{s}$. Then $\mathbf{s}(n)$ has linear complexity given by (4).

It seems quite remarkable that we have infinitely many sequences having unbounded linear complexity but the linear complexity of each of their subsequences $\mathbf{s}(n)$ can be precisely determined regardless of the underlying field F. Obviously, the Rueppel's conjecture follows as a corollary of Theorem 1 for $F = GF(2)$.

2. Proof of theorem 1

First, we prove a lemma.

Lemma. Let **s** be defined by (2), define for all nonnegative integer k the following $(k+1)\times(k+1)$ matrix over F

$$\mathbf{A}_k = \begin{bmatrix} s_0 & s_1 & \cdot & \cdot & \cdot & s_k \\ s_1 & s_2 & \cdot & \cdot & \cdot & s_{k+1} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ s_k & s_{k+1} & \cdot & \cdot & \cdot & s_{2k} \end{bmatrix}. \tag{5}$$

Then $\mathbf{A}_k$ are nonsingular for all k.

Proof. Since between k and $2k$ there is exactly one number, say $k+j$ with $0 \leqq j < k$, such that $k+j-1$ is a power of 2, the last row as well as the last column contains exactly one nonzero element. By deleting the last row and the last column we get $\mathbf{A}_{k-1}$ which has the same property just mentioned, thus $\mathbf{A}_k$ has the following form ($a = s_{k+1} \neq 0$):

$$\left[\begin{array}{c|c|c} \mathbf{A}_{j-1} & \mathbf{X} & \mathbf{0} \\ \hline \mathbf{Z} & \mathbf{Y} & \begin{matrix} \mathbf{0} & a \\ a & \mathbf{0} \end{matrix} \\ \hline \mathbf{0} & \begin{matrix} \mathbf{0} & a \\ a & \end{matrix} & \mathbf{0} \end{array}\right].$$

If $j=0$, i.e. $s_k \neq 0$ then the matrix is an upper triangular matrix with nonzero elements on the diagonal running from the bottom left corner to the upper right corner, hence the matrix is nonsingular.

The statement of the lemma holds trivially for $k=0$, 1 and we have shown that it is true for all $k = 2^m - 1$. Now by induction suppose that it is true for 0, 1, 2, ..., $k-1$. We have to show that it is true for k, too.

Now by Laplace expansion of the determinant we have

$$|\det \mathbf{A}_k| = |a^{k-j+1} \det \mathbf{A}_{j-1}|$$

where we have expanded the determinant alternating by the last row, then by the last column. Since a is nonzero and $\mathbf{A}_{j-1}$ is nonsingular by assumption, it follows that $\mathbf{A}_k$ is nonsingular and the lemma is proved.

Remark. In the binary case the lemma shows that if the sum $s_k + \ldots + s_{2k} = 1$ then matrices defined by (5) are nonsingular for all k. It is conjectured that the same remains true if

$$s_k + \ldots + s_{2k} = 1 \bmod 2$$

i.e., the number of 1's between k and $2k$ is an odd number. If the conjecture were true, we could construct many other binary sequences having the same linear complexity as that of the Massey sequence.

Proof of Theorem 1. Obviously, the theorem holds trivially for $n = 1$ and $n = 2$. By induction suppose that

$$L_s(m) = \lfloor (m+1)/2 \rfloor, \qquad \text{for all} \quad 1 \leqq m \leqq n$$

we have to show that

$$L_s(n+1) = \lfloor (n+1+1)/2 \rfloor = L_s(n-1) + 1.$$

First we show that $L_s(n+1) \geqq L_s(n-1) + 1$.

Since $L_s(n+1) \geqq L_s(n) \geqq L_s(n-1)$ $L_s(n+1) < L_s(n-1)$ is impossible, so by contrary suppose that

$$L_s(n+1) = L_s(n-1). \tag{6}$$

(i) Now if n is odd, $n = 2k+1$, then (6) implies

$$k = L_s(n+1) \geqq L_s(n) = k+1$$

a contradiction.

(ii) If n is even, $n = 2k$, then (6) and (3) imply that

$$c_0 s_j + c_1 s_{j-1} + \ldots + c_k s_{j-k} = 0, \qquad \text{for all} \quad j, k \leqq j \leqq 2k \tag{7}$$

where $c_0 = 1$ and $c_k \neq 0$. However, the lemma implies that these equations have unique solution for $(c_0, \ldots, c_k)$ and this solution is the zero vector, a contradiction again.

Thus,

$$L_s(n+1) \geqq L_s(n-1) + 1.$$

However, Theorem 2 of [4] implies that $L_s(n+1) = L_s(n-1) + 1$. The proof is completed.

3. Construction of family of sequences with high linear complexity and good correlation properties

The sequences defined by (2) are very sparse. If $f(n)$ denotes the number of nonzero elements in $\mathbf{s}(n)$ then

$$f(n)=\lfloor \log n \rfloor + 1$$

where log denotes the logarithm to the base two.

In engineering applications one needs families of sequences with high linear complexity and certain other good properties. The idea of our construction is to mix the sparse sequences defined by (2) with sequences known to have good additional properties. We hope that the resulted sequences will inherite good properties of both components.

The inheritability of high linear complexity is ensured by the following result which is a consequence of Theorem 2 in [5].

Theorem 2. Let

$$\mathbf{a}=(a_0, a_1, \ldots, a_{n-1})$$

and

$$\mathbf{b}=(b_0, b_1, \ldots, b_{n-1})$$

be two sequences over a field F with linear complexity L_a and L_b, respectively. Define sequence $\mathbf{d}$ as follows

$$\mathbf{d}=\mathbf{a}+\mathbf{b}=(d_0, d_1, \ldots, d_{n-1}), \qquad d_i=a_i+b_i$$

then $\mathbf{d}$ has linear complexity L_d such that

$$L_d \geqq |L_a - L_b|.$$

In the sequel suppose that F is the finite field $GF(p)$ with p being a prime number. We define the correlation of two sequences as follows [5]. Let w be a fixed primitive complex p-th root of unity, e.g. $w=\exp(2\pi j/p)$, $j=-1$. The correlation function is a function from the set of integers to the complex field defined as follows

$$\theta_{a,b}(k): \mathbf{N} \to \mathbf{C}$$

$$\theta_{a,b}(k)=\sum_{i=0}^{n-1} w^{a_i - b_{i+k}}$$

where the index is taken modulo n.

The inheritability of good correlation properties is ensured by the following result.

Theorem 3. Let **u** and **v** be two, not necessarily distinct, subsequences of length n of sequences defined by (2). Let **a** and **b** be two arbitrary sequences of length n. Define

$$\mathbf{a}' = \mathbf{a} + \mathbf{u}; \quad \mathbf{b}' = \mathbf{b} + \mathbf{v}$$

then

$$|\theta_{a',b'}(k)| \leqq |\theta_{a,b}(k)| + 4(\lfloor \log n \rfloor + 1).$$

Proof. For each k define

$$J_k = \{i : u_i - v_{i+k} \neq 0, \quad 0 \leqq i \leqq n-1\}$$

then the number of elements of J_k cannot exceed twice the number of nonzero elements of u, hence $|J_k| \leqq 2 \log n + 2$. Now write down the defining expression of correlation between **a**′ and **b**′. We have

$$\theta_{a',b'}(k) = \theta_{a,b}(k) + \sum_{i \in J_k} w^{a_i - b_{i+k}}(w^{u_i - v_{i+k}} - 1)$$

and the theorem follows.

Based on the results of Theorems 1–3 we can easily construct sequences having high linear complexity and good correlation properties.

Construction of sequences with good correlation properties

i. Autocorrelation

Let **G** be the family of M-sequences over $GF(p)$ with period $n = p^m - 1$. Let **S** be the set of $(p-1)^{f(n)}$ distinct subsequences of sequences defined by (2). Define

$$\mathbf{Q} = (\mathbf{a} + \mathbf{u} : \mathbf{a} \in \mathbf{G}, \ \mathbf{u} \in \mathbf{S}). \tag{8}$$

It is well known that for each $\mathbf{a} \in \mathbf{G}$ [5]

$$\theta_a(0) = n; \ \theta_a(k) = -1 \text{ if } k \neq 0 \bmod n$$
$$L_a = m. \tag{9}$$

For sequences **x** in **Q** we have

$$\theta_x(0) = n; \ \theta_x(k) \leqq 4(\log n + 1) + 1$$
$$L_x \geqq p^m/2 - m. \tag{20}$$

Observe that the ratio of maximum out-of-phase value to the peak of the autocorrelation remains relatively small, the linear complexity, however, increases dramatically. It is worth to mention that the number of sequences in **Q** is equal to product of numbers of sequences in **G** and **S**.

ii. Auto- and crosscorrelation

Now let **G** in (8) be a set of sequences having good auto- and crosscorrelation over $GF(2)$, and **S** consists of a single binary sequence of the same length defined by (2). Then the number of modified sequences is the same as that of the original sequences. For example, we can choose **G** to be the set of Gold sequences, Kasami sequences or bent sequences [6].

Let $\theta_{\max}$ denote the maximum value of out-of-phase autocorrelation and crosscorrelation. The properties of these families of sequences [6] and that of the modified families are summarized in the following table.

Table 1

Family: **G**	KASAMI	BENT	GOLD
Period	2^m-1	2^m-1	2^m-1
$\|\mathbf{G}\|$	$2^{m/2}$	$2^{m/2}$	2^m+1
m	$2k$	$4k$	$2k+1$
$\theta_{\max}=$	$2^{m/2}+1$	$2^{m/2}+1$	$2^{(m+1)/2}+1$
$\theta^*_{\max}\leqq$	$2^{m/2}+1+4m$	$2^{m/2}+1+4m$	$2^{(n+1)/2}+1+4m$
L	$3m/2$	$\leqq \sum_{i=1}^{m/4}\binom{m}{i}$	$2m$
$L^*\geqq$	$2^{m-1}-3m/2$	$2^{m-1}-\sum_{i=1}^{m/4}\binom{m}{i}$	$2^{m-1}-2m$

$\theta_{\max}$, $\theta^*_{\max}$ = maximum correlation values; $|\mathbf{G}|$ = number of sequences;
L, L^* = linear complexity; * refers to the modified sequences.

Again, we see that the maximum correlation value changes very little, while the linear complexity increases dramatically. For example [6], if $m=48$, then the maximum correlation values of Kasami and bent families as well as their modified families are practically the same, while the linear complexity of Kasami family is 72, and that of the modified Kasami family is of order 10^{14}; complexity of bent sequences is of order 10^{11}, and that of modified bent sequences is of order 10^{14}. It is clear from the table that the complexity of modified bent sequences is also much greater than that of the bent sequences.

These sequences should be of considerable interest in communication theory.

Acknowledgement

Before sending an earlier version of this paper for considering for presentation at EUROCRYPT 86, Prof. J. L. Massey informed me about [3] and was so kind to send me a copy of the manuscript of Dai's paper. I also thank the reviewers of EUROCRYPT 86 for their comments on the proof of the lemma.

References

1. *Massey, J. L.*, Threshold Decoding. Cambridge, MA: M.I.T. Press, 1963.
2. *Rueppel, R. A.*, "New Approach to Stream Ciphers". Ph. D. Dissertation, ETH, Zurich, Switzerland, Dec. 1984.
3. *Dai, Z. D.*, Proof of Rueppel's Linear Complexity Conjecture. IEEE Trans. Inf. Theory, May, 1985, pp. 440–443.
4. *Massey, J. L.*, Shift-Register Synthesis and BCH Decoding. IEEE Trans. Inf. Theory, **IT-15**, pp. 122–127, Jan. 1969.
5. *Zierler, N.*, Linear Recurring Sequences. J. Siam. Appl. Math., vol. 7, pp. 31–48, 1959.
6. *Olsen, J. D.*, *Scholtz, R. A.* and *Welch, L. R.*, Bent-Function Sequences. IEEE Trans. Inf. Theory, **IT-28**, pp. 858–868, Nov. 1982.

Элементарное доказательство утверждения о линейной сложности и конструкции последовательностей с высокой сложностью и хорошими корреляционными свойствами

НГУЕН КВАНГ А

(Ханой)

Изучаются свойства последовательностей над произвольным полем F. Обозначим через μ множество всех последовательностей $S=(S_0, S_1, S_2, \ldots)$, обладающих следующими свойствами: 1) $S_i \in F$, и $S_i \neq 0$ тогда и только тогда, когда для некоторого $k \geq 0$ целого числа справедливо равенство $i+1=2^k$. Для $S \in \mu$ $S(n)$ означает последовательность $(S_0, S_1, \ldots, S_{n-1})$, $L_S(n)$ линейную сложность $S(n)$.

Руппель высказал предположение о линейной сложности последовательности из μ, каждый член которой есть 0 или 1. Дей доказал это предположение.

В статье дается простое доказательство более сильного утверждения для любой последовательности.

Далее дается оценка линейной сложности суммы двух произвольных конечных последовательностей, а также функции корреляции двух сумм $a+u$ и $b+v$, где a и b произвольные последовательности длины n над полем F, $\{u, v\} \subseteq \{S(n) | S \in \mu\}$.

Предлагается способ построения последовательностей с большой линейной сложностью и хорошими корреляционными свойствами.

Nguyen Quang A
Institute for Telecommunication and Electronics
Technical University of Budapest
H-1111 Budapest, Stoczek u. 2.
Hungary
and
Vietnamese Academy of Sciences
Hanoi, Vietnam

Problems of Control and Information Theory, Vol. 17 (2), pp. 99—105 (1988)

E-CAPACITY UPPER BOUND FOR A CHANNEL WITH RANDOM PARAMETER

E. A. AROUTUNIAN, M. E. AROUTUNIAN

(Yerevan)

(Received March 26, 1987)

Average *E*-capacity upper bound for discrete memoryless channel with side information at the transmitter [1] is constructed.

The discrete memoryless channel with random parameter having input alphabet X and output alphabet Y is a channel whose transition probabilities depend on a parameter s with values in a finite set S. The values of parameter s are changing independently at different instances with the same distribution $Q(s)$ on S. The choice of parameter s does not depend on the input signal x.

Let $\mathbf{x}=(x_1, \ldots, x_n)$ be the input sequence of length n, $\mathbf{y}=(y_1, \ldots, y_n)$ the output sequence, $\mathbf{s}=(s_1, \ldots, s_n)$ the the channel states sequence, and let the set of conditional probabilities $W(y/x, s)$, $y \in Y$, $x \in X$, $s \in S$ be given. The channel is supposed to be stationary and memoryless which means that

$$W^n(\mathbf{y}/\mathbf{x}, \mathbf{s}) = \prod_{i=1}^{n} W(y_i/x_i, s_i),$$

$$Q^n(s) = \prod_{i=1}^{n} Q(s_i).$$

Consider the case when the states of the channel are known at the encoder and unknown at the decoder.

The channel having random parameter with additional information at the encoder was first considered by Shannon [1] in 1948. He mentioned that the case when the states of channel are known in each moment at the encoder and are unknown at the decoder is the most interesting. He studied the situation, when choosing the input symbol x_i one knows the states s_j of the channel for $j \leqq i$, however the states s_j for $j > i$ are unknown. Gelfand and Pinsker [2] determined the average capacity for the channel with random parameter in the case, when for the choice of the codeword $(x_1, \ldots, x_n)$ one need all states $(s_1, \ldots, s_n)$. In this paper the sphere packing bound for E-capacity of the channel with random parameter for the situation considered in [2] is constructed.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

This bound is also true for the situation considered by Shannon [1]. For the proof of this bound the method suggested in [3] for discrete memoryless channel is used. The results of this paper were reported at the conference held in Yerevan, 1986 [4]. The sphere packing bound just as random coding and expurgated bounds for E-capacity of the compound channel were constructed in [5].

A block (n, M)-code for the channel

$$K=\{W^n(\cdot/\cdot, \mathbf{s}): X^n \to Y^n, s \in S^n\}$$

is given by encoding $f: M \times S^n \to X^n$ and decoding $g: Y^n \to M$. The average probability of error e is

$$e(K, f, g)=\frac{1}{|M|}\sum_{i=1}^{|M|}\sum_{s\in S^n} Q^n(\mathbf{s})W^n(Y^n - g^{-1}(m)/f(m, \mathbf{s}), \mathbf{s})$$

(by $|A|$ we denote the cardinality of the finite set A).

The average E-capacity of the channel K is defined as $C_a(E)=\overline{\lim_{n\to\infty}}\frac{1}{n}\log M(K, E)$, where

$$M(K, E)=\sup_{f,g}\{|M|: \hat{e}(K, f, g)\leqq \exp(-nE)\}. \tag{1}$$

For random variables X, Y, S with values on X, Y, S, for distribution Q on S, conditional distribution P on X and channels $V: X \times S \to Y$, $W: X \times S \to Y$ we shall use the following notations. For entropy

$$H(Q)=H(S)=-\sum_s Q(s)\log Q(s),$$

conditional entropies

$$H(P/Q)=H(X/S)=-\sum_{s,x} Q(s)P(x/s)\log P(x/s),$$

$$H(V/Q, P)=H(Y/XS)=-\sum_{s,x,y} Q(s)P(x/s)V(y/x, s)\log V(y/x, s),$$

mutual informations

$$I(Q, P)=I(X \wedge S)=\sum_{s,x} Q(s)P(x/s)\log\frac{P(x/s)}{QP(x)},$$

$$I(Q, P, V)=I(Y \wedge XS)=\sum_{s,x,y} Q(s)P(x/s)V(y/x, s)\log\frac{V(y/x, s)}{QPV(y)}$$

where

$$QP(x)=\sum_{s} Q(s)P(x/s), \quad QPV(y)=\sum_{s,x} Q(s)P(x/s)V(y/x,s),$$

conditional divergence

$$D(V\|W/Q,P)=\sum_{s,x,y} Q(s)P(x/s)V(y/x,s)\log\frac{V(y/x,s)}{W(y/x,s)}.$$

Theorem. For the channel K when $E>0$

$$C_a(E)\leqq R_{sp}(E), \tag{2}$$

where

$$R_{sp}(E)=\max_{P}\min_{V:D(V\|W/Q,P)\leqq E}\{I(Q,P,V)-I(Q,P)\}.$$

Here we need some combinatorial notions ([6, 7]).

Denote by $N(s/\mathbf{s})$ the number of occurrences of $s\in S$ in $\mathbf{s}$. The type of a sequence $\mathbf{s}\in S^n$ is the distribution Q on S defined by

$$Q(s)=\frac{1}{n}N(s/\mathbf{s}), \quad \text{for every} \quad s\in S.$$

The set of sequences of type Q in S^n is denoted by T_Q. We say that $\mathbf{x}\in X^n$ has conditional type P given $\mathbf{s}\in S^n$ if $N(s,x/\mathbf{s},\mathbf{x})=N(s/\mathbf{s})P(x/s)$ for every $s\in S, x\in X$. For any given $\mathbf{s}\in S^n$ and stochastic matrix $P:S\to X$, the set of sequences $\mathbf{x}\in X^n$ having conditional type P given s is denoted by $T_P(\mathbf{s})$. Similarly,

$$T_V(\mathbf{x},\mathbf{s})=\{\mathbf{y}: N(s,x,y/\mathbf{s},\mathbf{x},\mathbf{y})=N(s,x/\mathbf{s},\mathbf{x})V(y/x,s), \quad s\in S, x\in X, y\in Y\}.$$

One can easily prove that the number of different types of sequences in S^n is upper bounded by $(n+1)^{|S|}$, the number of different types of sequences in $S^n\times X^n\times Y^n$ is bounded by $(n+1)^{|S||X||Y|}$.

We shall use some well-known combinatorial inequalities [6, 7].

$$(n+1)^{-|S|}\exp(nH(Q))\leqq|T_Q|\leqq\exp(nH(Q)) \tag{3}$$

$$(n+1)^{-|S||X|}\exp(nH(P/Q)\leqq|T_P(\mathbf{s})|\leqq\exp(nH(P/Q)), \qquad \mathbf{s}\in T_Q \tag{4}$$

$$(n+1)^{-|S||X||Y|}\exp(nH(V/Q,P))\leqq|T_V(\mathbf{s},\mathbf{x})|\leqq\exp(nH(V/Q,P)), \quad \mathbf{s}\in T_Q, \mathbf{x}\in T_P(\mathbf{s}). \tag{5}$$

If $\mathbf{s}\in T_Q$, $\mathbf{x}\in T_P(\mathbf{s})$, $\mathbf{y}\in T_V(\mathbf{s},\mathbf{x})$, then

$$W^n(\mathbf{y}/\mathbf{x},\mathbf{s})=\exp\{-n(D(V\|W/Q,P)+H(V/Q,P))\}. \tag{6}$$

Proof of the theorem. Let (f, g) be a code with rate $R = \frac{1}{n}\log|M|$. The condition $\bar{e}(K, f, g) \leqq \exp(-nE)$ means that

$$\sum_{i=1}^{|M|} \sum_{\mathbf{s}\in S^n} Q^n(\mathbf{s}) W^n(Y^n - g^{-1}(m)/f(m, \mathbf{s}), \mathbf{s}) \leqq \sum_{i=1}^{|M|} \sum_{\mathbf{s}\in S^n} Q^n(\mathbf{s}) \exp(-nE).$$

Since $Q^n(T_Q) \geqq Q^n(T_{Q'})$ for every Q', we come to

$$\sum_{i=1}^{|M|} \sum_{\mathbf{s}\in T_Q} Q^n(\mathbf{s}) W^n(Y^n - g^{-1}(m)/f(m, \mathbf{s}), \mathbf{s}) \leqq$$

$$\leqq \sum_{i=1}^{|M|} \sum_{s\in T_Q} Q^n(s) \exp(-n(E - \delta_n^1))$$

here $\delta_n^1 \to 0$, when $n \to \infty$.

Simplifying, we have for some stochastic matrix $V: X \times S \to Y$

$$\sum_{\mathbf{s}\in T_Q} \sum_{x(m,\mathbf{s})\in f(M,\mathbf{s})} \{|T_V(\mathbf{x}(m, \mathbf{s}), \mathbf{s})| - |g^{-1}(m) \cap T_V(\mathbf{x}(m, \mathbf{s}), \mathbf{s})|\} \times$$

$$\times W^n(\mathbf{y}/\mathbf{x}, \mathbf{s}) \leqq \sum_{\mathbf{s}\in T_Q} \sum_{\mathbf{x}(m,\mathbf{s})\in f(M,s)} \exp(-n(E - \delta_n^1)). \tag{7}$$

If the conditional type P is such that

$$\sum_{\mathbf{s}\in T_Q} |f(M, \mathbf{s}) \cap T_P(\mathbf{s})| \geqq |T_Q| \exp(nR)(n+1)^{-|X|}$$

then (7) will turn into the following inequality

$$\sum_{s\in T_Q} \sum_{\mathbf{x}(m,s)\in f(M,\mathbf{s})\cap T_P(\mathbf{s})} \{|T_V(\mathbf{x}(m, \mathbf{s}), \mathbf{s})| - |g^{-1}(m) \cap T_V(\mathbf{x}(m, \mathbf{s}), \mathbf{s})|\},$$

$$W^n(y/x, s) \leqq \sum_{s\in T_Q} \sum_{x(m,s)\in f(M,s)\cap T_P(s)} \exp(-n(E - \delta_n^1 - \delta_n^2)),$$

where $\delta_n^2 \to 0$, when $n \to \infty$. We shall denote $\delta_n = \delta_n^1 + \delta_n^2$. So

$$\sum_{\mathbf{s}\in T_Q} \sum_{\mathbf{x}(m,\mathbf{s})\in f(M,\mathbf{s})\cap T_P(\mathbf{s})} |g^{-1}(m) \cap T_V(\mathbf{x}(m, \mathbf{s}), \mathbf{s})| \geqq$$

$$\geqq \sum_{\mathbf{s}\in T_Q} \sum_{\mathbf{x}(m,\mathbf{s})\in f(M,\mathbf{s})\cap T_P(\mathbf{s})} \left\{|T_V(\mathbf{x}(m, \mathbf{s}), \mathbf{s})| - \frac{\exp(-n(E - \delta_n))}{W^n(\mathbf{y}/\mathbf{x}, \mathbf{s})}\right\}. \tag{8}$$

It is obvious that

$$|T_{QPV}| \geqq \sum_{\mathbf{s} \in T_Q} \sum_{\mathbf{x}(m,\mathbf{s}) \in f(M,\mathbf{s}) \cap T_P(\mathbf{s})} |g^{-1}(m) \cap T_V(\mathbf{x}(m,\mathbf{s}),\mathbf{s})|. \tag{9}$$

Now from (5), (6), (8), (9) we obtain

$$|T_{QPV}| \geqq \sum_{\mathbf{s} \in T_Q} \sum_{\mathbf{x}(m,\mathbf{s}) \in f(M,\mathbf{s}) \cap T_P(\mathbf{s})} [(n+1)^{-|S||X||Y|} \exp(nH(V/Q,P)) -$$

$$-\exp(n(D(V\|W/Q,P)+H(V/Q,P)=E+\delta_n))].$$

It remains to consider that for different s codewords of one message may be the same. The set of sequences $\mathbf{s}$ for fixed $\mathbf{x}$ is $T_{\bar{P}}(\mathbf{x})$, where $\bar{P}: X \to S$ is such that $Q(s)P(x/s) = QP(x)P(s/x)$. Thus

$$|T_{QPV}| \geqq \frac{\sum_{s \in T_Q} |f(M,\mathbf{s}) \cap T_P(\mathbf{s})|}{\max_{\mathbf{x} \in T_P(\mathbf{s}),\, \mathbf{s} \in T_Q} T_{\bar{P}}(\mathbf{x})} \{(n+1)^{-|S||X||Y|} \exp(nH(V/Q,P)) -$$

$$-\exp(n(D(V\|W/Q,P)+H(V/Q,P)-E+\delta_n))\}.$$

Now from (3) and (4)

$$\frac{1}{|T_Q|} \sum_{\mathbf{s} \in T_Q} |f(M,\mathbf{s}) \cap T_P(\mathbf{s})| \leqq$$

$$\leqq \frac{\exp(n(I(Q,P,V)-I(Q,P)))}{(n+1)^{-|S|}((n+1)^{-|S||X||Y|}-\exp(n(D\|V)(W/Q,P)-E+\delta_n))}. \tag{10}$$

The right side of the inequality can be decreased by the choice of V, leaving the denominator positive.

It is obvious that

$$|M| \leqq (n+1)^{|X|} \max_P \frac{1}{|T_Q|} \sum_{s \in T_Q} |f(M,s) \cap T_P(s)|. \tag{11}$$

The bound (2) follows from (1), (10) and (11).

Now let C_0 be the zero-error capacity of the channel with random parameter. We obtain the upper bound for C_0 from the theorem with $E \to \infty$.

Corollary.

$$C_0 \leqq \lim_{E \to \infty} C(E) \leqq \lim_{E \to \infty} R_{sp}(E) = \max_P \min_V (I(Q,P,V)-I(Q,P)),$$

where the minimum is taken by all V for which $V(y/x,s)=0$ if $W(y/x,s)=0$.

References

1. *Shannon, C.*, Channels with side information at the transmitter. IBM Res. Developm., 1948, **2,** *4*, pp. 289–293.
2. *Gelfand, S. I., Pinsker, M. S.*, Coding for channel with random parameters. Problems of Control and Information Theory, 1980, **9,** *1*, pp. 19–31.
3. *Арутюнян Е. А.*, Комбинаторный метод построения верхней границы E-пропускной способности. Межвуз. сб. науч. трудов, Матем., Ереван 1982, вып. *1*, стр. 213–220.
4. *Арутюнян Е. А., Арутюнян М. Е.*, Верхняя граница E-пропускной способности для дискретного канала со случайными параметрами. Материалы республиканской научно-практической конференции по методике преподавания математики и механики в ВУЗе, Ереван 1986, стр. 53–54.
5. *Арутюнян М. Е.*, Границы E-пропускной способности составных каналов. В печати.
6. *Csiszár, I., Körner, J.*, Information Theory. Coding Theorems for Didscrete Memoryless Systems. Akadémiai Kiadó, Budapest, 1981.
7. *Dueck, G., Körner, J.*, Reliability function of a discrete memoryless channel at rates above capacity. IEEE Trans. Inform. Theory, 1979, **25,** pp. 82–85.

Верхняя граница E-пропускной способности канала со случайным параметром

Е. А. АРУТЮНЯН, М. Е. АРУТЮНЯН

(Ереван)

Пусть X, Y и S конечные множества и K семейство дискретных каналов без памяти $W: X \to Y$, где $s \in S$ состояние канала. Значения параметра s изменяются независимо в разные моменты времени с одним и тем же распределением $Q(s)$ на S. Выбор параметра s не зависит также от входного сигнала x. Рассматривается ситуация, когда при выборе кодового слова $(x_1, \ldots, x_n)$ известны состояния канала $(s_1, \ldots, s_n)$, а при декодировании состояния канала неизвестны.

Блоковый (n, M)-код задается кодированием $f: M \times S^n \to X^n$ и декодированием $g: Y^n \to M$. Средняя вероятность ошибки равна

$$e(K, f, g) = \frac{1}{M} \sum_{i=1}^{M} \sum_{\mathbf{s} \in S^n} Q^n(\mathbf{s}) W^n(Y^n - g^{-1}(m)/f(m, \mathbf{s}), \mathbf{s}).$$

E-пропускная способность канала K для средней вероятности ошибки определяется как

$$C_a(E) = \overline{\lim_{n \to \infty}} \frac{1}{n} \log M(K, E),$$

где

$$M(K, E) = \sup_{f, g} \{M : e(K, f, g) \leqq \exp(-nE)\}.$$

Используются обозначения для: средней взаимной информации

$$I(Q, P) = \sum_{s, x} Q(s) P(x/s) \log \frac{P(x/s)}{QP(x)}$$

$$I(Q, P, V) = \sum_{s, x, y} Q(s) P(x/s) V(y/x, s) \log \frac{V(y/x, s)}{QPV(y)}$$

и условной дивергенции

$$D(V\|W/Q,P)=\sum_{s,x,y}Q(s)P(x/s)V(y/x,s)\log\frac{V(y/x,s)}{W(y/x,s)}.$$

Теорема. Для канала K, когда $E>0$,

$$C_a(E)\leqq R_{sp}(E),$$

где

$$R_{sp}(E)=\max_{P}\min_{V:D(V\|W/Q,P)\leqq E}\{I(Q,P,V)-I(Q,P)\}.$$

В работе доказана эта теорема. Как следствие из теоремы получена верхняя граница нулевой пропускной способности канала со случайным параметром.

Е. А. Арутюнян
М. Е. Арутюнян
Ереванский Государственный университет
СССР, 375049, Ереван
ул. Мравяна, 1.

Problems of Control and Information Theory, Vol. 17 (2), pp. 107—115 (1988)

TWO-DIMENSIONAL FISHER'S TEST

J. JEŽEK, A. OTÁHAL

(*Prague*)

(Received March 12, 1987)

A generalization of Fisher's test for two-dimensional data is derived and an asymptotic formula for critical values is given. As an example, an application of the test in image processing is presented.

Introduction

In harmonic analysis of time series, Fisher's test is used for revealing periodicities: the maximum of the normalized periodogram is compared with its critical value; if it exceeds the critical value, the hypothesis that the time series in question represents random noise is rejected, the corresponding frequency is accepted as a real one and some interpretation is attributed to it. The test derivation can be found in the original paper of Fisher (1929) or in many books on time series, e.g. Hannan (1970), Anderson (1971). As exact formulas for critical values are rather complicated, either asymptotics or tables are usually used. The latter are published e.g. by Nowroozi (1967) and Shimshoni (1971) who also discuss practical use of the test and its effectivity in context of estimation periods of eigenvibration of the Earth from seismic records.

Recently, fast progress in hardware makes possible effective processing of two-dimensional data as well. We mean especially (mini- and micro-) computer systems that are oriented onto image processing (and are, of course, used also for processing of geophysical, geochemical, geological data etc.). A two-dimensional generalization of Fisher's test is a problem naturally arising in this situation.

In our paper we derive such a generalization; its application range we see in studying wave processes and revealing periodicities, similarly as in the one-dimensional time-series case. We have in mind e.g. the study of sea-waves in oceanography or elimination of periodic-character disturbance in image data. The latter case will be discussed in more detail below.

Sometimes the spectrum of two-dimensional data is modified in a heuristic way so as to obtain some information or to improve readability of the data. Objectivity of such manipulations can be increased and better insight into the nature of the processed data can be achieved by means of the presented test.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

1. Periodogram

Let a matrix $X=(X(m,n): m=1,\ldots,M, n=1,\ldots,N)$ of reals be given. We define a (two-dimensional) periodogram

$$I(j,k)=\frac{1}{4\pi^2 MN}\left|\sum_{m=1}^{M}\sum_{n=1}^{N}(X(m,n)-\bar{X})\times\exp\{-2\pi i(jm/M+kn/N)\}\right|^2 \qquad (1)$$

where $j=1,\ldots,M, k=1,\ldots,N$ and $\bar{X}$ is the arithmetic mean of X,

$$\bar{X}=M^{-1}N^{-1}\sum_{m=1}^{M}\sum_{n=1}^{N}X(m,n).$$

All our later considerations could be, and in fact in a more simple way, carried out for M, N odd. On the other hand, powers of 2 are typical values for M, N in image processing. So we everywhere below suppose that M, N are even positive integers.

We define the following sets of indices

$$\mathbf{K}=\{(j,k): j=1,\ldots,M/2-1, k=1,\ldots,N/2-1\}\cup$$
$$\cup\{(j,k): j=M/2, k=1,\ldots,N/2-1\}\cup$$
$$\cup\{(j,k): j=M, k=1,\ldots,N/2-1\},$$
$$\mathbf{L}=\{(M/2,N),(M,N/2),(M/2,N/2),(M,N)\},$$
$$\mathbf{M}=\mathbf{K}\cup\mathbf{L},$$
$$\mathbf{N}=\{(j,k): j=1,\ldots,M, k=1,\ldots,N\}.$$

Further we define, for every $(j,k)\in\mathbf{N}$,

$$t_1(j)=\begin{cases}M-j & \text{for } j=1,\ldots,M-1\\ M & \text{for } j=M\end{cases}$$

$$t_2(k)=\begin{cases}N-k & \text{for } k=1,\ldots,N-1\\ N & \text{for } k=N\end{cases}$$

$$T(j,k)=(t_1(j),t_2(k)).$$

Remark 1.1. It is easy to see that the following statements are true:

(i) $I(j,k)=I(T(j,k))$ holds for every $(j,k)\in\mathbf{N}$,

(ii) $T(N)=\mathbf{N}$,

(iii) for every $(j,k)\in\mathbf{N}$, it holds $(j,k)=T(j,k)$ if and only if $(j,k)\in\mathbf{L}$,

(iv) $\mathbf{N}=\mathbf{K}\cup T(\mathbf{K})\cup\mathbf{L}$ where the union is disjoint.

2. Test Derivation

Our approach to the test derivation generalizes that of Hannan (1970), ch. VII. 2. For $(j, k) \in \mathbf{N}$ we denote

$$A(j,k) = \frac{1}{2\pi\sqrt{MN}} \sum_{m=1}^{M} \sum_{n=1}^{N} (X(m,n) - \bar{X}) \cos 2\pi(jm/M + kn/N), \tag{2}$$

$$B(j,k) = \frac{1}{2\pi\sqrt{MN}} \sum_{m=1}^{M} \sum_{n=1}^{N} (X(m,n) - \bar{X}) \sin 2\pi(jm/M + kn/N). \tag{3}$$

Comparing (1), (2) and (3) we see that

$$I(j,k) = A^2(j,k) + B^2(j,k)$$

holds for every $(j, k) \in \mathbf{N}$.

We will consider the hypothesis

(H) $X(m, n), (m, n) \in \mathbf{N}$ are independent identically distributed random variables, their common distribution being Gaussian with parameters μ, σ^2.

For the test derivation we need the joint distribution of $(A(m, n), B(m, n) : (m, n) \in \mathbf{N})$. As both A's and B's are linear combinations of Gaussians, they are again Gaussians and it is sufficient to calculate their expectations and (co)variances. From (2) and (3) it is obvious that the expectations of all A's and B's are equal to zero. The following Theorem together with its Corollary take care of the rest.

Theorem 2.1. Let $(j, k), (p, q) \in \mathbf{N}$. Then the following relations hold:

$$E\{A(j,k) \cdot A(p,q)\} =$$

$$= \frac{\sigma^2}{8\pi^2 MN} (\delta(j,p)\delta(k,q) + \delta(j,t_1(p))\delta(k,t_2(q)) -$$

$$- 2\delta(j,M)\delta(k,N)\delta(p,M)\delta(q,N)), \tag{4}$$

$$E\{B(j,k) \cdot B(p,q)\} =$$

$$= \frac{\sigma^2}{8\pi^2 MN} (\delta(j,p)\delta(k,q) - \delta(j,t_1(p))\delta(k,t_2(q))), \tag{5}$$

$$E\{A(j,k) \cdot B(p,q)\} = 0 \tag{6}$$

where $\delta(\cdot, \cdot)$ is the Kronecker symbol; for the sake of clarity the way of writing its arguments is somewhat unusual.

Proof. On the left-hand sides of (4), (5), (6) we substitute for A's and B's according to their definitions (2), (3). Having in mind that X's are independent and identically distributed Gaussians, we, after some tedious but immediate calculations, come to the right-hand sides of (4), (5) and (6).

Corollary 2.2. Let the sets of indices $\mathbf{K}, \mathbf{L}, \mathbf{M}$ be given as above. Then

(i) for $(j, k), (p, q) \in \mathbf{M}, (j, k) \neq (p, q)$ it holds

$$E\{A(j, k) \cdot A(p, q)\} = 0\,,$$

$$E\{B(j, k) \cdot B(p, q)\} = 0\,,$$

(ii) for $(j, k) \in \mathbf{K}$ it is

$$E\{A^2(j, k)\} = E\{B^2(j, k)\} = \frac{\sigma^2}{8\pi^2 MN}\,,$$

(iii) for $(j, k) \in \mathbf{L}$ it holds

$$E\{A^2(j, k)\} = \begin{cases} 0 & \text{for} \quad (j, k) = (M, N) \\ \dfrac{\sigma^2}{4\pi^2 MN} & \text{otherwise} \end{cases}$$

$$E\{B^2(j, k)\} = 0\,.$$

Proof. All the statements follow from the comparison of the above Theorem with Remark 1.1.

Now we are able to define test statistics and to determine their distributions. For the sake of brevity, let us write a instead of (j, k). We denote

$$y(a) = I(a) / \sum_{b \in \mathbf{K}} I(b)\,,$$

$$T_1 = \max_{a \in \mathbf{M}} y(a)\,,$$

$$T_r = \max \{y(a) : a \in \mathbf{M} \setminus \{a_1, \ldots, a_{r-1}\}\}$$

$$\text{for} \quad r = 2, \ldots, MN/2 + 2$$

where a_s is given by $T_s = y(a_s)$ for $s = 1, \ldots, MN/2 + 1$. That is, T_r is the r-th largest value of the normalized periodogram.

Theorem 2.3. Under holding of (H), statistics T_r has, for $M, N \to \infty$, an asymptotic distribution which is independent of r, and is given by

$$P\{T_r > (x + \ln D)/D\} \to 1 - \exp\{-e^{-x}\} \tag{7}$$

where $D = MN/2 - 2$.

Proof. With regard to the above considerations, the proof is, except for some minor and obvious alterations, the same as that of Theorem VII.6.17 of Hannan (1970) where a similar result for the one-dimensional case is proved.

The critical value for the confidence level α that corresponds to (7) is

$$c(\alpha)=\frac{1}{D}\ln\frac{-D}{\ln(1-\alpha)}. \tag{8}$$

These asymptotic critical values for typical confidence levels and for dimensions usual in image processing are given in Table 1.

Table 1. Asymptotic critical values $\times 10^4$

Image dimensions	D	α	0.1	0.05	0.01
64 × 64	2046		48.26	51.78	59.74
128 × 128	8190		13.75	14.63	16.62
256 × 256	32766		3.86	4.08	4.58
512 × 512	131070		1.07	1.13	1.25

It is not easy to give an exact assessment of how close an approximation we get using asymptotics according to Theorem 2.3, because direct comparison with the exact values meets some difficulties. In fact, straightforward calculation of the exact critical values is subjected to rounding errors; and the larger the value of D, the stronger this rounding-errors effect is. On the other hand, we may expect a fairly good approximation for large values of D: e.g. for $D=3000$, the exact critical value on 0.05 significance level is equal to 3.650×10^{-3} (cf. Shimshoni (1971), Table 1(d)) while the asymptotic one by (8) is 3.659×10^{-3}, the relative error being 0.27%.

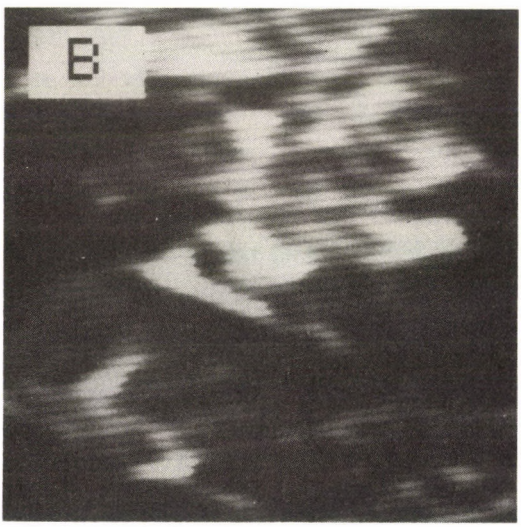
3. Application Example

So as to illustrate the application of the two-dimensional Fisher's test, we present an example of model-data analysis in remote sensing. Calculations and graphics are due to FOURFIVE 85 package implemented in PERICOLOR 2000E system.

Processed data are shown in Figs 1A and 2A (both images are the same). There are periodic disturbances in the data that are not difficult to recognize. We note that distrubances of this type (so-called "stripes") are often met with in satellite data where they appear as a result of poorly matched sensors.

We present two types of processing the data. The first one consists in filtering away all the frequencies that the above test points out as significant. In the second one


A


B

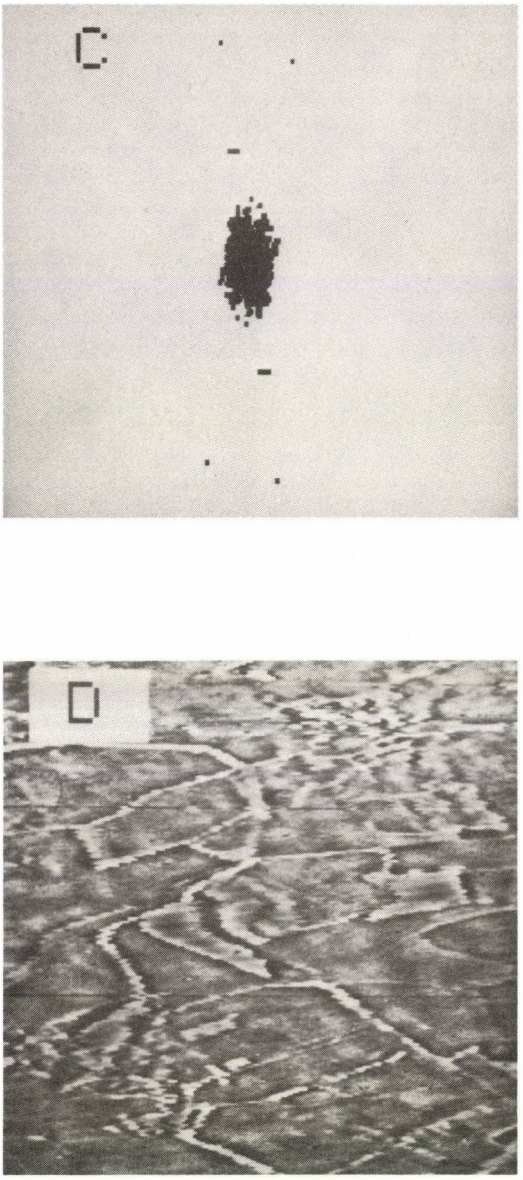


Fig. 1. A: original data, B: Fourier inverse of significant frequencies, C: significant frequencies, D: complement of B

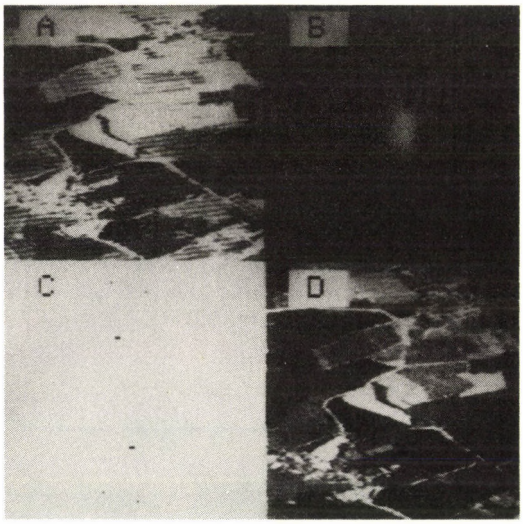


Fig. 2. A: original data, B: logarithmic amplitude spectrum, C: singular "disturbing" frequencies, D: enhanced image

we again find the significant frequencies. Then we filter away only singular significant ones that we feel to correspond to the disturbance.

Figure 1A shows the significant frequencies on the 0.05 level; their inverse Fourier transform is visualized in Fig. 1B. The difference between the last image and the original data is in Fig. 1D; we see that the structure effects are emphasized there.

Figure 2B presents the periodogram in the logarithmic scale or, equivalently, the logarithmic amplitude spectrum. Figure 2C shows those significant frequencies we have chosen as responsible for the disturbance and the enhanced image after filtering them away is shown in Fig. 2D. Improvement in the image quality is easily discernible.

As mentioned above, periodic disturbance of the type dealt with here are often met with in satellite data. That is, the two-dimensional Fisher test could be taken for a quantitative criterion both in quality evaluation and in enhancement of satellite data.

Acknowledgement

The authors wish to express their gratitude to V. Bezvoda for fruitful discussions and to S. Saic for his help in the example data processing.

References

Anderson, T. W., The Statistical Analysis of Time Series. Wiley, New York 1971.

Bezvoda, V., Ježek J., Saic, S., Segeth, K., FOURFIVE 85, Package of the programs for Fourier transform, filtering and visualizing of two-dimensional data (in Czech). Unpublished manuscript. Faculty of Sciences of Charles University, Prague 1985.

Fisher, R. A., Test of significance in harmonic analysis. Proc. Roy. Soc., A, **125** (1929), 54–59.

Hannan, E. J., Multiple Time Series. Wiley, New York 1970.

Nowroozi, A. A., Table for Fisher's test of significance in harmonic analysis. Geophys. J. R. Astr. Soc., **12** (1967), 517–520.

Otáhal, A., Generalization of the Fisher's test for two-dimensional periodogram. Res. Report No. 1423, Inst. of Inf. Theory and Automation of the Czechoslovak Academy of Sciences, Prague 1986.

Shimshoni, M., On Fisher's test of Significance in harmonic analysis, Geophys. J. R. Astr. Soc. **23** (1971), 373–377.

Выделение двумерных периодичностей

Й. ЕЖЕК, А. ОТАХАЛ

(Прага)

В статье обобщается метод выделения периодичностей, основой которого является теорема Фишера, для случая двумерных данных. Приведено асимптотическое распределение статистик, соответствующих проверке значимости наибольших значений спектра. В качестве примера дается применение метода к обработке образов.

J. Ježek
Department of Mathematics and Computer Science
Faculty of Sciences, Charles University
Albertov 6
128 43 Praha 2
Czechoslovakia

A. Otáhal
Institute of Information Theory and Automation
Czechoslovak Academy of Sciences
Pod vodárenskou věží 4
182 08 Praha 8
Czechoslovakia

РУССКИЙ ПЕРЕВОД

Проблемы Управления и Теории Информации, том **17**, *номер 2 (1988)*

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ УЧАСТКА ГИБКОЙ ПРОИЗВОДСТВЕННОЙ СИСТЕМЫ

Ю. М. СОЛОМЕНЦЕВ, Е. Б. ФРОЛОВ

(*Москва*)

Для гибкого автоматизированного участка механообработки построена математическая модель соответствующего производственного процесса, рассмотрена задача оптимального управления производственной программой участка.

1. Введение

На основании тщательной проработки технологического процесса для всей номенклатуры изготавливаемых деталей создается гибкая производственная система (ГПС), задача которой — обеспечить требуемое качество обработки и заданную производительность [1], [2].

Экономические показатели работы ГПС существенно зависят от функционирования системы при возникновении *внутренних* и *внешних* отклонений в ее текущей производственной программе [2], таких как:

1) выход из строя основного или вспомогательного оборудования, нехватка режущего инструмента, профилактический ремонт некоторых станков и т. п.,
2) несвоевременная поставка заготовок, появление приоритетных заказов, изменение номенклатуры обрабатываемых деталей и пр.

При этом система управления ГПС [3] должна оперативно рассчитать новый порядок запуска деталей на обработку, так, чтобы экономическая эффективность гибкого производства была бы максимальной. Экономические показатели ГПС во многом зависят от ритмичности сборки узлов и изделий, в состав которых входят изготавливаемые системой детали [1], [4]. Последние, как правило, имеют различный приоритет в зависимости от их принадлежности к одному или разным изделиям.

Перерасчет производственного расписания ГПС является NP-трудной проблемой, поэтому для большого числа одновременно обрабатываемых разнотипных деталей ее точное решение с помощью комбинаторных методов практически недостижимо.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

Для задач коррекции производственной программы ГПС используются формальные модели, описывающие технологические процессы обработки деталей. Имеющиеся в литературе математические модели гибких систем можно разделить на две основные группы [5], [6]:

A. Имитационные модели (Evaluative Models), цель которых — прогнозирование поведения ГПС и ее отдельных подсистем в различных ситуациях. Алгоритмы оперативно-диспетчерского управления (приоритеты запуска деталей [7], [8], их технологические маршруты) считаются известными.

B. Модели для генерации управления (Generative Models), предназначенные для формирования оптимальных режимов коррекции программы ГПС (переналадок оборудования, запуска деталей на обработку с учетом альтернативных технологических маршрутов [7], [9] и пр.) по заданным критериям ее функционирования.

Имитационные модели, построенные на основе различных математических методов (теории массового обслуживания [6], [10], анализа временных рядов [5], [11], стохастической аппроксимации и вариационного анализа [5], [12], методах сетей Петри [6], [13] и др. [5], [14]–[16]), в сочетании с современными программными средствами [17]–[20] (от языков типа FORTRAN или PASCAL до GPSS, SIMCRIPT, MAP/I, ADA и т.д.) позволяют выявить "узкие места" и ошибочные решения в реализации производственной программы, оптимизировать параметры гибкой системы, частично улучшить набор управляюших алгоритмов.

Используемые на этапах проектирования ГПС и долгосрочного планирования, имитационные модели оказываются малоэффективными в задачах оперативного управления производством [5] (при различных отклонениях в текушем сменно-суточном задании).

Для генерации такого управления применяются математические модели второй группы, причем большая их часть основана на различных эвристических процедурах теории расписаний [21]–[23]. Ряд частных задач, связанных с управлением гибкой автоматизированной системой, удается решить традиционными методами исследования операций (см., например, [6], [7], [24]–[31]).

Однако, наилучшие результаты можно получить на основе методов теории управления, предполагающих наличие формального описания соответствующих технологических процессов в виде динамической управляемой системы. Попытка использовать для этой цели модель в виде дискретной конечно-разностной системы, заданной в пространстве состояний [32]–[38], наталкивается на серьезные вычислительные трудности в связи с большой размерностью задачи.

Авторы настоящей работы предлагают отказаться от трудоемких вычислений технологических траекторий (отдельных этапов обработки деталей различной номенклатуры). Основная цель — определение оптимального *числа запусков* деталей различных типов (моментов переналадок станков) в пределах временного цикла работы оборудования [24]. Для этого строится математическая модель технологического процесса механообработки в виде некоторой дискретнои управляемой системы. Последующее применение дискретного принципа максимума позволяет эффективно решить задачу коррекции производственной программы гибкого автоматизированного участка механообработки.

2. Математическая модель производства на участке ГПС

В качестве иллюстрации построим математическую модель гибкого автоматизированного участка механообработки, состоящего из четырех станков. Рассмотрим механическую обработку деталей четырех номенклатур, различающихся как видом материала, так и типоразмерами. Производство имеет циклический характер [4], [36], когда продукция выпускается в виде набора партий, состоящих из различного числа деталей определенного типа. Этот набор периодически повторяется через определенные равные промежутки времени, именуемые временем цикла.

Технологический процесс построен таким образом, что на первой стадии механообработки на станках M_1 (фрезерном) и M_2 (токарном) производится черновая обработка доступных поверхностей деталей 1 и 2-го, а также 3 и 4-го типов соответственно. Затем, на второй стадии механообработки детали 1, 3 и 2, 4 типов поступают соответственно на многоцелевые станки M_3 и M_4, где от новых технологических баз производится их окончательная обработка. Таким образом, каждый станок M_j $(j=1, 2, 3, 4)$ может обрабатывать два типа деталей. Условная схема техпроцесса ызображена на рисунке. Оптимальные размеры α_{ij} партий заготовок (i — номер типа, j — номер станка) считаются известными функциями от k [4], [25], [26].

Частично обработанные на станке M_1 детали 1 и 3-го типов в количествах β_{13} и β_{24} (величина β_{il} определяется цикловой производительностью станка M_j по деталям i-го типа, l — номер станка, на котором будет продолжена механообработка этих деталей) попадают на позицию смены технологических баз.

Число $x_1(k)$ и $x_2(k)$ деталей 1 и 2-го типов, ожидающих в течение k-го цикла обработки на станке M_1, удовлетворяет уравнению баланса:

$$\begin{aligned} x_1(k+1) &= x_1(k)+\alpha_{11}(k)-\beta_{13}(k) \\ x_2(k+1) &= x_2(k)+\alpha_{21}(k)-\beta_{24}(k). \end{aligned} \tag{2.1}$$

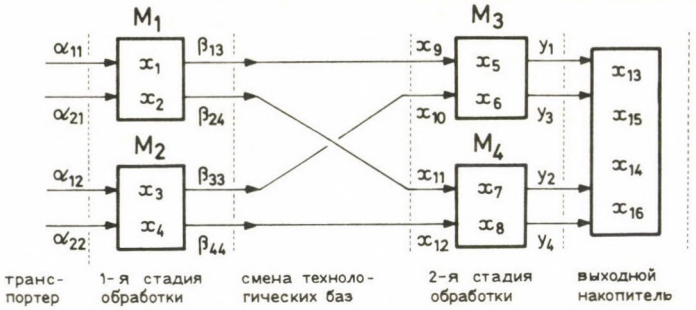


Рис. 1

Аналогично для деталей 2, 4 типов и второго станка M_2 получаем

$$\begin{aligned} x_3(k+1) &= x_3(k)+\alpha_{32}(k)-\beta_{33}(k) \\ x_4(k+1) &= x_4(k)+\alpha_{42}(k)-\beta_{44}(k), \end{aligned} \tag{2.2}$$

здесь переменные x_3, x_4, α_{32}, α_{42}, β_{33}, β_{44} имеют тот же смысл, что и переменные x_1, x_2, α_{11}, α_{21}, β_{13}, β_{24} для станка M_1 в (2.1).

При смене технологических баз происходит переустановка деталей: β_{13} штук деталей превращается в x_9 штук деталей 1-го типа, β_{24} штук деталей — в x_{11} штук деталей 2-го типа ($\beta_{13}=x_9$, $\beta_{24}=x_{11}$) и т.д.

В следующем $k+1$-м временном цикле детали в количествах $x_{4+i}(k+1)$ ($i=1,2,3,4$) вновь подаются в рабочую зону многоцелевых станков M_3 и M_4. Наконец, в выходной накопитель участка поступают готовые детали различной номенклатуры в количествах $y_i(k)$, где i — номер типа, $i=1,2,3,4$.

Построив аналогичные (2.1) и (2.2) уравнения баланса для станков M_3 и M_4, получим уравнения состояния процесса маханообработки деталей соответственно 1, 3 и 2, 4 типов:

$$\begin{aligned} x_1(k+1) &= x_1(k)+\alpha_{11}(k)-\beta_{13}(k) \\ x_2(k+1) &= x_2(k)+\alpha_{21}(k)-\beta_{24}(k) \\ x_3(k+1) &= x_3(k)+\alpha_{32}(k)-\beta_{33}(k) \\ x_4(k+1) &= x_4(k)+\alpha_{42}(k)-\beta_{44}(k) \\ x_5(k+1) &= x_5(k)+x_9(k)-y_1(k) \\ x_6(k+1) &= x_6(k)+x_{10}(k)-y_3(k) \\ x_7(k+1) &= x_7(k)+x_{11}(k)-y_2(k) \\ x_8(k+1) &= x_8(k)+x_{12}(k)-y_4(k). \end{aligned} \tag{2.3}$$

Предположим, что время цикла равно T, и пусть t_j — величина потерь рабочего времени в цикле за счет переналадки j-го станка. Тогда

$$1-t_j/T=(\gamma_{ij}+\gamma_{lj})/T, \tag{2.4}$$

где γ_{ij} и γ_{lj} — длительности времен, в течение которых детали i-го и l-го типов обрабатывались на станке M_j $(j=1, 2, 3, 4)$ в пределах цикла длительности T $(1\leqq i, l\leqq 4)$.

Следуя [36], рассмотрим в качестве управляющих переменных величины

$$\begin{cases} u_1(k)=\gamma_{11}(k)/T, & u_2(k)=\gamma_{42}(k)/T, \\ u_3(k)=\gamma_{13}(k)/T, & u_4(k)=\gamma_{44}(k)/T. \end{cases} \tag{2.5}$$

Значения управляющих переменных u_j $(j=1, 2, 3, 4)$ показывают, какую часть времени цикла станки M_1 и M_3 обрабатывают детали 1-го типа, M_2 и M_4 — детали 4-го типа (см. рисунок 2).

Обозначим через S_{ij} максимальную цикловую производительность станка M_j по деталям i-го типа. Если станок M_j вышел из строя, то $S_{ij}=0$.

В сиду (2.4) и (2.5) на выходе станков M_1 и M_2 имеем:

$$\begin{aligned} \beta_{13}(k)&=S_{11}u_1(k) \\ \beta_{24}(k)&=S_{21}(1-t_1/T-u_1(k)) \\ \beta_{33}(k)&=S_{32}(1-t_2/T-u_2(k)) \\ \beta_{44}(k)&=S_{42}u_2(k). \end{aligned} \tag{2.6}$$

Число деталей разных типов, поступивших в выходной накопитель участка ГПС в течение k-го временного цикла выражается формулой

$$\begin{aligned} y_1(k)&=S_{13}u_3(k) \\ y_2(k)&=S_{24}(1-t_4/T-u_4(k)) \\ y_3(k)&=S_{33}(1-t_3/T-u_3(k)) \\ y_4(k)&=S_{44}u_4(k). \end{aligned} \tag{2.7}$$

Поскольку на освободившиеся палеты вновь устанавливаются заготовки, поступающие по вход-выходному транспортеру на первую стадию обработки, то справедливо неравенство

$$\alpha_{11}+\alpha_{21}+\alpha_{32}+\alpha_{42}=\sum_{j=1}^{4} y_j\leqq d_0, \tag{2.8}$$

(здесь d_0 — пропускная способность транспортера), причем общее число палет, находящихся на участке механообработки, ограничено:

$$\begin{aligned} &x_1+x_2+x_5+x_7+(S_{11}-S_{21})u_1+(1-t_1/T)S_{21}\leqq d_{12}\\ &x_3+x_4+x_6+x_8+(S_{42}-S_{32})u_2+(1-t_2/T)S_{32}\leqq d_{34}\,. \end{aligned} \tag{2.9}$$

В (2.9) предполагается, что для заготовок 1, 2-го и 3, 4-го типов используются различные палеты в количествах d_{12} и d_{34} соответственно.

Дополнительные ограничения на переменные системы определяются их физическим смыслом:

$$\begin{aligned} &x_i\geqq 0 && i=1,2,3,4,5,6,7,8\,,\\ &0\leqq u_j\leqq 1 && j=1,2,3,4\,,\\ &\text{значения} && S_{ij}u_j \text{ и } S_{lj}\,(1-t_j/T-u_j) \text{ — целые числа},\\ & && j=1,2,3,4\,,\quad 1<i\,,\quad l<4\,. \end{aligned} \tag{2.10}$$

Таким образом, величина $u_j(k)$ в случае $0<u_j(k)<1$ фактически определяет в силу (2.5) момент переналадки станка M_j в течение k-го временного цикла его работы. $S_{lj}u_j(k)$ есть число запусков деталей l-го типа на j-й станок ($l=1,4$).

Производственный процесс на рассматриваемом участке ГПС описывается в соответствии с заданной технологией согласно (2.3) и (2.6) с помощью дискретной системы управления

$$\begin{aligned} x_1(k+1)&=x_1(k)-S_{11}u_1(k)+\alpha_{11}(k)\\ x_2(k+1)&=x_2(k)-S_{21}(1-t_1/T-u_1(k))+\alpha_{21}(k)\\ x_3(k+1)&=x_3(k)-S_{32}(1-t_2/T-u_2(k))+\alpha_{32}(k)\\ x_4(k+1)&=x_4(k)-S_{42}u_2(k)+\alpha_{42}(k)\\ x_5(k+1)&=x_5(k)+x_9(k)-S_{13}u_3(k)\\ x_6(k+1)&=x_6(k)+x_{10}(k)-S_{33}(1-t_3/T-u_3(k))\\ x_7(k+1)&=x_7(k)+x_{11}(k)-S_{24}(1-t_4/T-u_4(k))\\ x_8(k+1)&=x_8(k)+x_{12}(k)-S_{44}u_4(k)\\ x_9(k+1)&=S_{11}u_1(k)\\ x_{10}(k+1)&=S_{32}(1-t_2/T-u_2(k))\\ x_{11}(k+1)&=S_{21}(1-t_2/T-u_1(k))\\ x_{12}(k+1)&=S_{42}u_2(k) \end{aligned} \tag{2.11}$$

с выходом (2.7) и ограничениями (2.8), (2.9), (2.10). Здесь переменные x_1, x_2 и x_3, x_4 означают число деталей разных типов, ожидающих обработки на станках M_1 и M_2 соответственно, x_5, x_6 и x_7, x_8 — число деталей, ожидающих окончательной механообработки (от новых технологических баз) на многоцелевых станках M_3 и M_4. Вспомогательные переменные x_9, x_{11} и x_{10}, x_{12} характеризуют производительность станков M_1 и M_2 соответственно в течение предыдущего временного цикла.

Система (2.11) выражает состояние $x=(x_1, \ldots, x_{12})$ производства на участке как функцию входных потоков $\alpha_{ij}(k)$ и вектора управления $u=(u_1, u_2, u_3, u_4)$ для всех $k=0, 1, 2, \ldots$ Начальное состояние:

$$x_i(0)=x_{i0} \qquad i=1, 2, 3, 4, 5, 6, 7, 8,$$

$$x_9(0)=x_{10}(0)=x_{11}(0)=x_{12}(0)=0.$$

3. Отклонения в сменно-суточном задании участка ГПС

Любые отклонения от намеченного производственного плана, возникающие по организационным причинам [2] (отсутствие требуемых заготовок, необходимого инструмента и приспособлений, выделение группы приоритетных деталей, директивное изменение номенклатуры изделий и т.д.), приводят, в конечном счете, к изменению текущего сменно-суточного задания участка механообработки.

Перерасчет производственного расписания участка является NP-трудной задачей [1], [33] и поэтому не может быть эффективно осуществлен в реальном масштабе времени [27].

С другой стороны, учет динамики реализуемых технологических процессов в виде системы (2.11) позволяет получить решение задачи коррекции производственной программы участка ГПС при следующих предположениях:

C1: Оборудование участка не выходит из строя: $S_{ij}=\text{const}>0$;

C2: В течение одного временного цикла допускается не более одной переналадки станка M_j $(j=1, 2, 3, 4)$; следовательно, потеря рабочего времени за счет переналадки станка определяется величиной

$$\begin{cases} t_j(k)=\tau_j\delta_j(k, k-1) \\ \delta_j(k, k-1)=\operatorname{sgn}\{\sigma_j+(1-\sigma_j)\,|\operatorname{sgn} u_j(k)-\operatorname{sgn} u_j(k-1)|\}, \end{cases} \tag{3.1}$$

где $\sigma_j=\operatorname{sgn}\{u_j(k)(1-u_j(k))\}$,

τ_j — длительность переналадки M_j,

δ_j — число переналадок станка M_j:

(если $u_j(k)=1$ или $u_j(k)=0$, то $\sigma_j=0$, и станок M_j не переналаживается лишь при условии, что в предыдущем цикле управление $u_j(k-1)$ имело то же значение, что и $u_j(k)$;

С3: ограничение (2.8) не является существенным — пропускная способность транспортера d_0 достаточно велика;

С4: ограничения (2.10) ослаблены: предполагается, что

$$x_i(k)>0 \qquad i=1,2,3,4,5,6,7,8\,, \quad k=1,2,\ldots, \tag{3.2}$$

(т.е. на участке всегда имеются детали разных типов, доступные для обработки на станках M_j) как и в (2.10) $0\leqq u_j(k)\leqq 1, j=1,2,3,4$, однако значения $S_{ij}u_j(k)$ и $S_{ij}(1-t_j/T-u_j(k))$ — действительные числа, $1\leqq l, i\leqq 4$, $i=1,2,3,4$.

Условие С4, в частности, означает, что детали какого-либо типа могут отсутствовать лишь в начале первого временного цикла: начальное состояние системы (2.11) может иметь нулевые координаты. При всех остальных значениях $k=1,2,3,\ldots N$ выполняется указанное в (3.2) требование $x_i(k)>0$, т.е. служба контроля загрузки палет обеспечивает надлежащий входной поток заготовок $\alpha_{11}, \alpha_{21}, \alpha_{32}, \alpha_{42}$.

Плановый коэффициент загрузки оборудования, характеризующий ожидаемую производительность участка механообработки, задает ограничение на общее число переналадок станков в течение интервала планирования $[q, N]$:

$$T_N=\sum_{j=1}^{4}\sum_{k=0}^{N-1} t_j(k)\leqq T_{np} \tag{3.3}$$

здесь T_N и T_{np} фактическое и плановое время простоя оборудования участка (связанное с переналадками) за N циклов его работы.

Добавим к уравнениям (2.11) уравнения состояния выходного накопителя

$$\begin{aligned}
x_{13}(k+1)&=x_{13}(k)+S_{13}u_3(k)\\
x_{14}(k+1)&=x_{14}(k)+S_{24}(1-t_4/T-u_4(k))\\
x_{15}(k+1)&=x_{15}(k)+S_{33}(1-t_3/T-u_3(k))\\
x_{16}(k+1)&=x_{16}(k)+S_{44}u_4(k)\\
x_{12+j}(0)&=0 \qquad j=1,2,3,4,
\end{aligned} \tag{3.4}$$

здесь $x_{12+j}(k)$ — число готовых деталей j-го типа, поступивших в выходной накопитель участка за k временных циклов.

Отметим, что величины t_j, фигурирующие в (2.9), (2.10), (2.11), (3.3), (3.4), вычисляются в силу (3.1). Поэтому система (2.11), (3.4) — нелинейна.

Таким образом, состояние производственного процесса описывается дискретном управляемой системой (2.11), (3.1), (3.4) с ограничениями (2.8), (2.9), (3.2), (3.3), а характер производства в течение N временных циклов выражается минимизируемым функционалом

$$I = \Phi(x(0),\ x(N)) + \sum_{k=0}^{N-1} \varphi_k(x(k),\ u(k),\ k). \tag{3.5}$$

Функционал (3.5) является критерием оптимальности процесса производства на заданном интервале планирования $[0, N]$.

Нетрудно видеть, что по значениям вектора состояния $x(k+1)$ и $x(k)$ системы (2.11), (3.4) можно судить о величинах простоев станков участка в предыдущем временном цикле:

$$\begin{aligned}
t_1(k) &= T\{1 - x_9(k+1)/S_{11} - x_{11}(k+1)/S_{21}\} \\
t_2(k) &= T\{1 - x_{10}(k+1)/S_{32} - x_{12}(k+1)/S_{42}\} \\
t_3(k) &= T\{(x_{13}(k) - x_{13}(k+1))/S_{13} + (x_{15}(k) - x_{15}(k+1))/S_{33} + 1\} \\
t_4(k) &= T\{(x_{14}(k) - x_{14}(k+1))/S_{24} + (x_{16}(k) - x_{16}(k+1))/S_{44} + 1\}.
\end{aligned} \tag{3.6}$$

С учетом (3.6) левая часть ограничения (3.3) на время простоя станков в течение N временных циклов имеет вид:

$$\begin{aligned}
T_N(x) = T \sum_{k=0}^{N-1} [2 - x_9(k)/S_{11} - x_{10}(k)/S_{32} - x_{11}(k)/S_{21} - x_{12}(k)/S_{42}] + 2N - \\
- T(x_{13}(N)/S_{13} + x_{14}(N)/S_{24} + x_{15}(N)/S_{33} + x_{16}(N)/S_{44}).
\end{aligned} \tag{3.7}$$

Нетрудно заметить, что ограничение на время переналадки оборудования (3.3) может быть заведомо учтено, если включить T_N в минимизируемый функционал (3.5).

Если измененное сменно-суточное задание участка ГПС (набор из 4-х партий деталей) выражается вектором потребности $z = (z_1, z_2, z_3, z_4)$, где z_j — требуемое количество деталей j-го типа в партии, то простейшая задача коррекции производственной программы участка в течение N временных циклов имеет критерий оптимальности вида

$$I = \sum_{j=1}^{4} c_j (x_{12+j}(N) - z_j)^2 + c_0 T_N(x), \tag{3.8}$$

здесь c_j — штрафные коэффициенты за невыполненное сменносуточное задание по j-му типу деталей, c_0 — средняя себестоимость одного станко-часа простоя оборудования участка.

Изменение сменно-суточного задания может также выражаться в изменении функционала (3.5). Например, требование «обеспечить выполнение задания z за N циклов так, чтобы детали 4-го типа были обработаны за минимально возможное время» подразумевает минимизацию функционала вида:

$$I = \sum_{j=1}^{4} c_j (x_{12+j}(N) - z_j)^2 + \sum_{k=0}^{N-1} kS_{44}u_4(k) + c_0 T_N(x). \tag{3.9}$$

Сходная постановка задачи рассматривалась в [7], [35].

С другой стороны, требование: «обеспечить выполнение конечного задания, придерживаясь известного производственного графика $z_j(k)$ $j=1, 2, \ldots, N$» означает минимизацию функционала

$$I = \sum_{k=0}^{N} \sum_{j=1}^{4} c_j (x_{12+j}(k) - z_j(k))^2 + c_0 T_N(x). \tag{3.10}$$

Другие виды функционалов, применяемые в задачах управления автоматизированной механообработкой, можно найти в [7], [8], [24], [27], [36].

Таким образом, даже при фиксированном значении вектора z изменение требований к характеру производственного процесса (изменение вида функционала (3.5), например, в виде (3.8), (3.9) или (3.10)) является, по-существу, изменением сменно-суточного задания участка ГПС и требует коррекции его производственной программы.

4. Оптимальная коррекция производственной программы участка ГПС

Проблема оптимальной коррекции производственной программы рассмотренного гибкого автоматизированного участка механообработки в предположениях С1–С4 равносильна задаче управления дискретной системой (2.11), (3.1), (3.4) с ограничениями (2.9), (3.2) и функционалом (3.5). Начальное состояние $x(0)$ системы считается заданным:

$$x_i(0) = x_{i0}, \qquad x_{i+8}(0) = 0, \qquad i = 1, 2, 3, 4, 5, 6, 7, 8.$$

В соответствии с дискретным принципом максимума [37], [39] искомое оптимальное управление $u(k)$ доставляет максимальное значение функции Гамильтона для всех значений $k=0, 1, \ldots, N-1$.

Для рассматриваемой системы построим, следуя [39], гамильтониан по формуле

$$H(x(k),\ u(k),\ \psi(k+1)) = \sum_{i=1}^{16} \psi_i(k+1) f_i(x(k),\ u(k),\ k) -$$

$$- \varphi_k(x(k),\ u(k),\ k) + \mu_1 \xi_1(x(k),\ u(k)) + \mu_2 \xi_2(x(k),\ u(k)), \tag{4.1}$$

где f_i — правые части системы (2.11), записанной в виде

$$x_i(k+1) - x_i(k) = f_i(x(k),\ u(k),\ k), \qquad (i = 1, \ldots, 16),$$

функция $\varphi_k(x(k),\ u(k),\ k)$ содержится в функционале (3.5), а ξ_1 и ξ_2 определяются следующим образом:

$$\begin{aligned} \xi_1 &= x_1 + x_2 + x_5 + x_7 + (S_{11} - S_{21})u_1 + (1 - t_1/T)S_{21} - d_{12} \\ \xi_2 &= x_3 + x_4 + x_6 + x_8 + (S_{42} - S_{32})u_2 + (1 - t_2/T)S_{32} - d_{34} \end{aligned} \tag{4.2}$$

μ_1, μ_2 — вспомогательные множители Лагранжа, связанные с ограничениями (8.9)

$$\begin{aligned} &\mu_j(k) = 0 \quad \text{при условии} \quad \xi_j(x(k),\ u(k)) < 0, \\ &\mu_j(k) < 0 \quad \text{при условии} \quad \xi_j(x(k),\ u(k)) > 0, \qquad j = 1, 2. \end{aligned}$$

Переменные $\psi_i(k+1)$ из (4.1) являются решением сопряженной системы уравнений [24]

$$\psi_i(k+1) - \psi_i(k) = -\partial H / \partial x_i(x(k),\ u(k),\ \psi(k+1)) \tag{4.3}$$

$$i = 1, 2, \ldots, 16.$$

Для системы (2.11), (3.4) выражение (4.3) принимает вид

$$\begin{aligned} &\psi_i(k) = \psi_i(k+1) + \mu_1(k) - \partial\varphi_k/\partial x_i \qquad i = 1, 2, 5, 7, \\ &\psi_j(k) = \psi_j(k+1) + \mu_2(k) - \partial\varphi_k/\partial x_j \qquad j = 3, 4, 6, 8, \\ &\psi_i(k) = \psi_{i-4}(k+1) - \partial\varphi_k/\partial x_i \qquad i = 9, 10, 11, 12, \\ &\psi_j(k) = \psi_j(k+1) - \partial\varphi_k/\partial x_j \qquad j = 13, 14, 15, 16. \end{aligned} \tag{4.4}$$

Граничные условия для сопряженной системы (4.4) определяются из условий трансверсальности [39]:

$$\psi(N) = -\partial\Phi(x(0),\ x(N))/\partial x(N).$$

Таким образом, задача оптимальной коррекции производственной программы участка (в условиях С1–С4) сводится к нахождению вектора оптимального управления $u(k) = (u_1(k),\ u_2(k),\ u_3(k),\ u_4(k))$, доставляющего максимальное

значение функции Гамильтона (4.1) при всех $k=0, 1, \ldots, N-1$, $x(k)$ и $\psi(k+1)$ — есть решения системы (2.11), (3.1), (3.4) и системы (4.5) соответственно [37], [39], [40].

В заключение отметим, что если функции $\varphi_k(x(k), u(k), k)$ из (3.5) выпуклы по $u(k)$ (или не зависят от переменной $u(k)$ — случаи (3.8) и (3.10)), то при максимизации гамильтониана (4.1) управляющие параметры u_1 и u_2 принимают максимальное значение

$$\begin{aligned} u_1 &= \min\{1,\ [d_{12}-x_1-x_2-x_5-x_7-(1-t_1/T)S_{21}]/(S_{11}-S_{21})\} \\ u_2 &= \min\{1,\ [d_{34}-x_3-x_4-x_6-x_8-(1-t_2/T)S_{32}]/(S_{42}-S_{32})\} \end{aligned} \tag{4.5}$$

тогда и только тогда, когда выполнены условия

$$\begin{aligned} S_{11}(\psi_9(k+1)-\psi_1(k+1))-S_{21}(\psi_{11}(k+1)-\psi_2(k+1)) &\geqq -\partial\varphi_k/\partial u_1(x(k),\ u(k)) \\ S_{42}(\psi_{12}(k+1)-\psi_4(k+1))-S_{32}(\psi_{10}(k+1)-\psi_3(k+1)) &\geqq -\partial\varphi_k/\partial u_2(x(k),\ u(k)). \end{aligned}$$

Поэтому оптимальные значения множителей μ_1 и μ_2 из (4.1) вычисляются по формулам [37]:

$$\mu_1(k)=\min\left\{0,\ \left[\frac{\partial\varphi_k}{\partial u_1}(x(k),\ u(k))+S_{11}(\psi_9(k+1)-\psi_1(k+1))-\right.\right.$$
$$\left.\left.-S_{21}(\psi_{11}(k+1)-\psi_2(k+1))\right]/(S_{21}-S_{11})\right\}, \tag{4.6}$$

$$\mu_2(k)=\min\left\{0,\ \left[\frac{\partial\varphi_k}{\partial u_2}(x(k),\ u(k))+S_{42}(\psi_{12}(k+1)-\psi_4(k+1))-\right.\right.$$
$$\left.\left.-S_{32}(\psi_{10}(k+1)-\psi_3(k+1))\right]/(S_{32}-S_{42})\right\}. \tag{4.7}$$

Использование (4.6) и (4.7) в (4.1) гарантирует при ограничениях (2.9) и (3.2) выполнение условий $\partial H/\partial u_1=0$ и $\partial H/\partial u_2=0$ вдоль траектории $x(k)$ и искомого управления $u(k)$, $k=0, 1, \ldots, N-1$. Полученные таким образом величины $u_j(k)$ ($j=1, 2, 3, 4$) определяют моменты переналадок станков M_j.

Оптимальное число запусков деталей i-го и l-го типов на станок M_j в течение k-го цикла его работы есть целые части величин $S_{ij}u_j(k)$ и $S_{lj}(1-t_j(k)/T-u_j(k))$. Их своевременное вычисление позволяет *эффективно* решать задачу коррекции производственной программы гибкого автоматизированного участка механообработки.

5. Выводы

1. Задача коррекции производственной программы гибкого участка механообработки, являясь NP-трудной комбинаторной задачей, может быть в некоторых случаях эффективно решена методами теории оптимального управления.

2. Математическая модель участка механообработки, позволяющая получить эффективное решение задачи коррекции его производственной программы, основана на описании технологии изготовления деталей (с помощью имеющегося оборудования) в виде дискретной управляемой системы.

3. Оптимальная коррекция производственного процесса на участке ГПС осуществляется последовательно в течение нескольких временных циклов работы его оборудования. Соответствующее число запусков деталей на станки и моменты их переналадок определяются в результате решения задачи оптимального управления дискретной системой (при ограничениях на управление и состояние) на основе дискретного принципа максимума.

Литература

1. Новые методы управления производством за рубежом. Экспресс-информация «Приборы, средства автоматизации и системы управления», **ТС-3**, вып. *1*, М. 1986.
2. *Соломенцев Ю. М., Кутин А. А., Шептунов С. А.* Оценка гибкости автоматизированной станочной системы. Вестник машиностроения, 1984, *1*, 38–40.
3. *Соломенцев Ю. М.* Возможности систем управления в гибком автоматизированном производстве. Микропроцессорные средства и системы, 1984, *2*, 73–74.
4. *Лищинский Л. Ю.* Циклический производственный процесс в гибком автоматизированном производстве. Станки и инструмент, 1986, *7*, 2–6.
5. *Gershwin, S., Hildebrant, R., Suri, R., Mitter, S.*, A control perspective on recent trends in manufacturing systems. IEEE Control Systems Magazine, 1986, Vol. **6,** No. *2*, 3–15.
6. Робототехника и гибкие автоматизированные производства. Кн. 5. Моделирование робототехнических систем и гибких автоматизированных производством. (Пантюшин С. В., Назаретов В. М., Тягунов О. А. и др.) Высшая школа. М. 1986.
7. Робототехника и гибкие автоматизированные производства. Кн. 3. Управление робототехническими системами и гибким автоматизированным производством. (Макаров И. М., Рахманкулов В. З., Назаретов В. М. и др.). Высшая школа. М. 1986.
8. *Wroblewski, K., Krawczynski, R.*, Priority rules in production flow control. Material Flow, 1985, No. *2*, 167–177.
9. *Wilhelm, W. E., Shin, H. M.*, Effectiveness of alternative operations in a flexible manufacturing system. International Journal of Production Research, 1985, Vol. **23,** No. *1*, 65–79.
10. *Buzacott, J. A., Shanthikumar, J. G.*, Models for understanding flexible manufacturing systems. AIIE Transactions, 1980, Vol. **12,** No. *2*, 339–350.
11. *Law, A. M.*, Statistical analysis of simulation output data. Operations Research, 1983, Vol. **31,** No. *6*, 983–1029.
12. *Ho, Y. C., Cao, X. R.*, Perturbation analysis and optimization of queuing networks. Journal of Optimization Theory and Applications, 1983, Vol. **40,** No. *4*, 559–582.

13. *Dubois, D., Stecke, K.*, Using Petry Nets to represent production processes. Proceedings of the IEEE Conference on Decision and Control, 1983.
14. *Yao, D., Buzacott, J. A.*, Modeling the performance of flexible manufacturing systems. International Journal of Production Research, 1985, Vol. **23,** No. *5*, 945–959.
15. *Ho, Y. C., Eyeler, M., Chien, T. T.*, A gradient technique for general buffer storage design in a serial production line. International Journal of Production Research, 1979, Vol. **17,** No. *6*, 557–580.
16. *Suri, R.*, An overview of evaluative models for flexible manufacturing systems. Annals of Oper. Research, 1985.
17. *Schriber, T.*, The use of GPSS/H in modelling a typical flexible manufacturing system. Annals of Oper. Research, 1985.
18. *Musselman, K.*, Computer simulation: a design tool for FMS. Manufacturing Engineering, 1984, Vol. **92,** No. *9*, 117–120.
19. *Bevans, J. P.*, First choose an FMS simulator. American Machinist, May 1982, 143–145.
20. *Micheletti, G., Cantino, V., Fenoglio, F.*, An application of simulation theory in FMS. Robotica, 1985, Vol. **3,** No. *2*, 253–257.
21. *Graves, S.*, A review of production scheduling. Operations Research, 1981, Vol. **29,** No. *4*, 641–675.
22. *Russell, R., Taylor, B.*, An evaluation of scheduling policies in a dual resource constrained assembly shop. IIE Transactions, 1985, Vol. **17,** No. *3*, 219–232.
23. *Yamamoto, M., Nof, S. Y.*, Scheduling/rescheduling in the manufacturing operating system environment. International Journal of Production Research, 1985, Vol. **23,** No. *4*, 705–722.
24. *Смоляр Л. И.*, Модели оперативного планирования в дискретном производстве. М.: Наука, 1978.
25. *Блехерман М. Х.*, Организационно-технологическое группирование деталей в ГПС. Вестник машиностроения, 1986, *6*, 37–42.
26. *De Luca, A.*, Optimal production planning for FMS: an "Optimum batching" algorithm. Proceedings of the 3-rd International Conference on Flexible Manufacturing Systems. Boeblingen, 1984, 323–332.
27. *Hildebrant, R.*, Scheduling flexible machining systems using mean value analysis. Proceedings, IEEE Conference on Decision and Control, 1980, 701–706.
28. *Stecke, K. E.*, Formulation and solution of nonlinear integer production planning problems for flexible manufacturing systems. Management Science, 1983, Vol. **29,** No. *2*, 273–288.
29. *Shanker, K., Tzen, Y.*, A loading and dispatching problem in a random manufacturing system. International Journal of Production Research, 1985, Vol. **23,** No. *3*, 579–595.
30. *Kusiak, A.*, Flexible manufacturing systems: a structural approach. International Journal of Production Research, 1985, Vol. **23,** No. *6*, 1057–1073.
31. *Graves, S.*, Using Lagrangian techniques to solve hierarchical production planning problems. Management Science, 1982, Vol. **28,** No. *3*, 260–275.
32. *Axsater, S.*, Control theory concepts in production and inventory control. International Journal on Systems Science, 1985, Vol. **16,** No. *2*, 161–169.
33. *Olsder, G. J., Suri, R.*, Time-optimal control of parts-routing in a manufacturing system with failure-prone machines. Proceedings, IEEE Conference on Decision and Control, 1980, 722–727.
34. *Teny, R. M., Kochar, A. K.*, A vector space model approach to production planning, multi-cell and multi-stage manufacturing system. International Journal on Systems Science, 1984, Vol. **15,** No. *11*, 1169–1187.
35. *Janiak, A.*, Optimalizacja czasowa sekwencji kompleksow operacji w sterowaniu dyskretnymi procesami produkcyjnymi. Politechnika Slaska Zeszyty Naukowe, 1980, No. *650*, 57–65.
36. *Gruver, W. A., Narasimhan, S. L.*, Optimal scheduling of multi-stage continuous flowshops. INFOR, 1981, Vol. **19,** No. *4*, 318–329.
37. *Kleindofer, P. B., Kriebels, C. H., Thompson, G. L., Kleindofer, G. B.*, Discrete Optimal Control of production plans. Management Science, 1975, Vol. **22,** No. *3*, 261–273.
38. *Hitomi, K., Nakamura, N.*, Optimal production planning for a multiproduct, multistage production system. International Journal of Production Research, 1976, Vol. **14,** No. *2*, 199–213.
39. *Сейдж Э. П., Уайт Ч. С.*, Оптимальное управление системами. М. Радио и связь, 1982.
40. *Outrata, J. V.*, Minimization of nonsmooth nonregular functions: application to discrete-time optimal control problems. Problems of Control and Information Theory, 1984, Vol. **13,** No. *6*, 413–424.

PRINTED IN HUNGARY

Akadémiai Kiadó és Nyomda Vállalat, Budapest

NOTE TO CONTRIBUTORS

Two copies of the *manuscript* (each complete with figures, tables and references) are to be sent to

E.D. TERYAEV coordinating editor
Department of Mechanics and Control Processes
Academy of Sciences of the USSR
Leninsky Prospect 14, Moscow V-71, USSR

or to

L. GYÖRFI
Technical University of Budapest
H-1111 Budapest, Stoczek u. 2, Hungary

Authors are requested to retain a third copy of the submitted typescript to be able to check the proofs.

The papers, preferably in English or Russian, should be typed double spaced on one side of good-quality paper with wide margins (4–5 cm). The first page of the paper should carry the title, the author(s)' names and the name of the town where they are active. The name and address of the author to whom the proofs should be sent should be given at the end of the paper. An *abstract* should head the paper. English papers should also have a Russian abstract.

The papers should not exceed 15 pages (25 × 50 characters per page) including tables and references. The proper location of the tables and figures must be indicated on the margin.

Mathematical notations should follow up-to-date usage. Equations longer than half a line should not be incorporated in the text. In-text equations must be typed on a single line except that one level of subscripting and/or superscripting is permissible. Use / instead of horizontal bars. Displayed equations should be written so as to require the fewest possible lines. Therefore use "exp" for the exponential function whenever the exponent requires more than a single line. Matrices should, if possible, not be written in full. Use subscript notations instead such as $A = ||a_{ij}||$. Write diagonal matrices as diag $(d_1, d_2, \ldots d_n)$.

The authors will be sent galley proofs to be returned by next mail. Rejected manuscripts will be returned. Authors will receive 100 reprints free of charge. Additional reprints may be ordered.

К СВЕДЕНИЮ АВТОРОВ

Рукописи статей в трех экземплярах на русском языке и в трех на английском следует направлять по адресу: 117312 Москва В-312, просп. 60 летия Октября, 9, МНИИПУ. Редакция журнала «Проблемы управления и теории информации» (зав. редакцией Н. И. Родионова, тел. 208-60-19).

Объём статьи не должен превышать 15 печатных страниц (25 строк по 50 букв). Статье должна предшествовать аннотация объемом 50–100 слов и приложено резюме–реферат объемом не менее 10–15% объема статьи на русском языке в трех экземплярах, на котором напечатан служебный адрес автора (фамилия, название учреждения, адрес).

При написании статьи авторам надо строго придерживаться следующей формы: введение (постановка задачи), основное содержание, примеры практического использования, обсуждение результатов, выводы и литература.

Статьи должны быть отпечатаны с промежутком в два интервала, последовательность таблиц и рисунков должна быть отмечена на полях. Математические обозначения рекомендуется давать в соответствии с современными требованиями и традициями. Разметку букв следует производить только во втором экземпляре и русского, и английского варианта статьи.

Авторам высылается верстка, которую необходимо незамедлительно проверить и возвратить в редакцию.

После публикации авторам высылаются бесплатно 100 оттисков их статей.

Рукописи непринятых статей возвращаются авторам.

CONTENTS · СОДЕРЖАНИЕ

HU ISSN 0370-2529 Index 26 660

VOL. 17 · NUMBER 3
TOM ГОМЕР

ACADEMY OF SCIENCES OF THE USSR
HUNGARIAN ACADEMY OF SCIENCES
CZECHOSLOVAK ACADEMY OF SCIENCES

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

АКАДЕМИЯ НАУК СССР
ВЕНГЕРСКАЯ АКАДЕМИЯ НАУК
ЧЕХОСЛОВАЦКАЯ АКАДЕМИЯ НАУК

1988

AKADÉMIAI KIADÓ, BUDAPEST
DISTRIBUTED OUTSIDE THE COMECON-COUNTRIES
BY PERGAMON PRESS, OXFORD

PROBLEMS OF CONTROL AND INFORMATION THEORY

An international bi-monthly sponsored jointly by the Presidium of the Academy of Sciences of the USSR, of the Hungarian Academy of Sciences and of the Czechoslovak Academy of Sciences. The six issues published per year make up a volume of some 480 pp. It offers publicity for original papers and short communication of the following topics:

- theory of control processes
- theory of adaptive systems
- theory of estimation and identification
- theory of controlling robot-technologic and flexible manufacturing systems
- information theory
- information-theoretic aspects of multiple access networks.

While this bi-monthly is mainly a publication forum of the research results achieved in the socialist countries, also papers of international interest from other countries are welcome.

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

Международный журнал Академии наук СССР, Венгерской Академии наук и Чехословацкой Академии наук выходит 6 раз в год общим объемом 480 печатных страниц.

В журнале публикуются оригинальные научные статьи и статьи обзорного характера по следующим проблемам управления и теории информации:

- теория процессов управления;
- теория адаптивных систем;
- теория оценивания и идентификации;
- теория управления робототехническими и гибкими производственными системами;
- теория информации;
- теория информации в области сетей с множественным доступом.

Целью журнала является ознакомление научной общественности различных стран с важнейшими проблемами, имеющими актуальный и перспективный характер, научными достижениями ученых социалистических и других стран.

Distributors

For the Soviet Union:

SOYUZPECHATY, Moscow 123 308 USSR

For Albania, Bulgaria, China, Cuba, Czechoslovakia, German Democratic Republic, Korean People's Republic, Mongolia, Poland, Rumania, Vietnam and Yugoslavia:

KULTURA Hungarian Foreign Trading Co.
P. O. Box 149 H-1389 Budapest, Hungary

For all other countries:

PERGAMON JOURNALS LTD, Headington Hill Hall, Oxford OX3 OBW, England

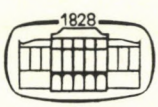
or

PERGAMON JOURNALS INC, Maxwell House, Fairview Park, Elmsford, NY 10523, USA

1988 Subscription Rate DM 535,— per annum including postage and insurance.

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ



AKADÉMIAI KIADÓ

PUBLISHING HOUSE OF THE HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST

Problems of Control and Information Theory, Vol. 17 (3), pp. 117—123 (1988)

FAMILIES OF SEQUENCES WITH OPTIMAL GENERALIZED HAMMING CORRELATION PROPERTIES[1]

NGUYEN QUANG A

(Budapest, Hanoi)

(Received April 2, 1987)

Generalized Hamming correlation function is introduced. Bounds on out-of-phase autocorrelation and crosscorrelation are given. Family of sequences with optimal generalized Hamming correlation properties is constructed. As a corollary it is shown that the majority of the codewords of a Reed–Solomon code have no inner period. Several interesting applications of sequences with good generalized Hamming correlation properties are mentioned.

1. Introduction

In many communication situations we need a set of sequences having not only good Hamming correlation properties but also some additional properties. In seeking for sequences which can be used in frequency-hopping systems and in the construction of protocoll sequences for multiple-access collision channels we found that it is useful to introduce a generalized Hamming correlation function to describe such sequences.

The concept of generalized Hamming correlation is introduced in the next section where some bounds on out-of-phase autocorrelation and crosscorrelation are given. In the last section a construction of families of sequences with optimal generalized Hamming correlation properties is given. As a by-product this gives also a direct construction of a subcode of the Reed–Solomon code over $GF(q)$. The subcode has a very interesting property; its codewords have no inner period and the subcode has nearly as many codewords as in the original Reed–Solomon code if q is large. It turns out that such interesting subcodes can also be constructed from a large class of BCH codes [2]. Such a subcode is, thus, a self-synchronizing code with degree of comma freedom [6] at least $q-2k+1$ if the original Reed–Solomon code has k information digits and rate $<1/2$.

[1] This paper has been presented partly at the IEEE International Symposium on Information Theory, 5–9 October, 1986, Michigan, USA.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

2. The generalized Hamming correlation function

Let G be a finite Abelian group, the group operation is denoted by $+$, its neutral element is denoted by 0. Let x and y be sequences of length n over G, and $\mathbf{N}$ be the set set of integers.

The function of two variables defined below is called the *generalized Hamming correlation function* of two sequences x and y over group G:

$$H_{xy}(.,.): G \times \mathbf{N} \to \mathbf{N}$$

$$H_{xy}(a, i) = m(x - S^i y, a) \tag{1}$$

$$= \sum_{j=0}^{n-1} h[x(j) - y(j+i), a)$$

where $j+i$ *is taken modulo* n, $m(z, a)$ is the multiplicity of a in z, i.e. the number of times a occurs in z, the difference is understood componentwise, S denotes the cyclic shift operator, and

$$h[u, v] = \begin{cases} 1, & \text{if } u = v \\ 0, & \text{otherwise.} \end{cases}$$

Note that $H_{xy}(0, i)$ is the conventional Hamming correlation function [4].

Following Lempel and Greenberger [4], we use the following quantities for measuring the performance of sequences

$$H(x) = \max_{\substack{0 < i < n \\ a \in G}} H_{xx}(a, i) \tag{2}$$

$$H(x, y) = \max_{\substack{0 \leq i < n \\ a \in G}} H_{xy}(a, i) \tag{3}$$

$$M(x, y) = \max \{H(x), H(y), H(x, y)\}. \tag{4}$$

We have the following bounds on $H(x)$ and $H(x, y)$.

Lemma. If x and y are two distinct sequences of length n over G then

$$H(x) \geqq \left\lceil \frac{n}{q} \right\rceil$$

$$H(x, y) \geqq \left\lceil \frac{n}{q} \right\rceil$$

where q is the number of elements of G and $\lceil x \rceil$ is the smallest integer greater than or equal to x.

Proof. Let $\bar{H}(x)$ and $\bar{H}(x, y)$ be average out-of-phase values of $H_x(a, i)$ and average value of $H_{x,y}(a, i)$, respectively. We have

$$(n-1)q\bar{H}(x)=\sum_{a\in G}\sum_{i=1}^{n-1}H_x(a,i)=\sum_{a\in G}\sum_{i=1}^{n-1}\sum_{j=0}^{n-1}h[x(j)-x(j+i),a]=$$

$$=\sum_{i=1}^{n-1}\sum_{j=0}^{n-1}\sum_{a\in G}h[x(j)-x(j+i),a]=n(n-1).$$

The last equality follows from the fact that the inner sum is always 1 for any j and i. By the same way it follows that

$$qn\bar{H}(x,y)=n^2.$$

Since the maximum value is at least as large as the average value, the Lemma is proved.

Remark, It is clear that $H(x)\geqq 1$ and $H(x, y)\geqq 1$, so the bound of the Lemma is nontrivial only for the case $n>q$.

Let x be a sequence, the vector $S^i x$ is a cyclic shift of x, the vector $x+u\mathbf{1}$ is called a *translation* of x, the vector $S^i x+u\mathbf{1}$ is called a *shifted translation* of x. The possible number of cyclic shifts is n and the number of translation is q. A vector x of length n over G is called of *maximal period* if it has exactly nq distinct shifted translations. In the sequel we are interested only in the collection of maximal period sequences. For such a family we have the following bound.

Theorem 1. Let H be a family of sequences with length n over group G such that each sequence is of maximum period and if a sequence $x\in H$ then its proper shifted translations, $S^i x+u\mathbf{1}$, $i\neq 0$ or $u\neq 0$, do not belong to H. Let

$$M(H)=\max_{\substack{x\in H\\ y\in H\\ x\neq y}}M(x,y)$$

then we have

$$M(H)\geqq\max\left\{\left\lceil\frac{n}{q}\right\rceil;\lceil\log_q(n|H|)\rceil\right\}$$

where $|H|$ denotes the number of sequences in H and $q=|G|$.

Proof. Let C be the set of all shifted translations of sequences in H. Since each sequence in H has maximum period, the number of elements in C is

$$|C|=qn|H|.$$

Now consider C as a nonlinear code over G and let $d(C)$ denote the minimum Hamming distance of this code. The Singleton bound [5, p. 544] gives

$$d(C)\leqq n-\log_q|C|+1.$$

Now for any $x \in H$, $y \in H$ and $x \neq y$ we have

$$H_{x,y}(a, i) = \sum_{j=0}^{n-1} h[x(j) - y(j+i), a]$$

$$= n - d(x, z)$$

where

$$z = S^i y + a\mathbf{1},$$

$$z \in C, \qquad z \neq x,$$

$d(x, z)$ denotes the Hamming distance between x and z.

Thus,

$$M(H) = n - d(C)$$

$$\geqq n - n + \log_q(qn|H|) - 1$$

$$= \log_q(n|H|).$$

By the Lemma the proof is completed.

In the next section we give a construction that achieves the above bound.

A construction of families of sequences with optimal generalized Hamming correlation properties

Let G be the additive group of a Galois field $GF(q)$ where q is a prime power. Let β be a primitive element of $GF(q)$, $2 \leqq k \leqq n = q - 1$ be a fixed integer, and $\mathbf{A}$ be a $k \times n$ matrix defined as follows

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \beta & \beta^2 & \dots & \beta^{n-1} \\ 1 & \beta^2 & \beta^4 & \dots & \beta^{2(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \beta^{k-1} & \beta^{2(k-1)} & \dots & \beta^{(n-1)(k-1)} \end{bmatrix} \tag{5}$$

Theorem 3. Let E_r be the set of q^{k-2} sequences of length n over G defined as follows

$$E_r = \{x : x = a\mathbf{A}, a = (0, 1, a_2, \dots, a_{k-1})\} \tag{6}$$

where a_i, $i = 2, \dots, k-1$, is taken from G.

Then for all pairs of distinct x, y in E_r we have

$$M(x, y) \leqq k - 1. \tag{7}$$

Proof. First note that E_r is just a subcode of the Reed–Solomon code of parameter $[n, k]$ over $GF(q)$ with distance $d \geqq n-k+1$. Now let x, y be distinct elements of E_r, i.e.

$$x = a\mathbf{A} \qquad \text{and} \qquad y = b\mathbf{A}; \quad a \neq b.$$

(i) By definition we have

$$H(x) = \max_{\substack{u \in G \\ i \neq 0}} m(x - S^i x, u) = \max_{x'} m(x - x', 0) =$$

$$= \max_{x'} \{n - d(x, x')\}$$

where $x' = u\mathbf{1} + S^i x, \qquad \mathbf{1} = (1, 1, \ldots, 1)$

$$x' = a'\mathbf{A}$$

$$a' = (u, \beta^{-1}, a_2\beta^{-2i}, \ldots, a_{k-1}\beta^{-(k-1)i}).$$

This implies that x' is a codeword not belonging to E_r, hence $x \neq x'$, i.e. $d(x, x') \geqq \geqq n-k+1$ or

$$H(x) \leqq k-1; \; H(y) \leqq k-1. \tag{8}$$

(ii) In the same way as above we have

$$H(x, y) = \max_{u \in G} m(x - S^i y, u) = \max_{y'} (n - d(x, y')\}$$

where $y' = b'\mathbf{A}'$ and

$$b' = (u, \beta^{-1}, b_2\beta^{-2i}, \ldots, b_{k-1}\beta^{-(k-1)i}).$$

By assumption at least one $a_j \neq b_j$, $j \geqq 2$. This implies that for all i and u, $b' \neq a$, so we have

$$H(x, y) \leqq k-1. \tag{9}$$

This and (8) complete the proof.

Let E_m be defined as follows:

$$E_m = \{y = u\mathbf{1} + S^i x : x \in E_r, \; 0 \leqq i < n, \; u \in G\}. \tag{10}$$

Two vectors in E_m are said to be in the same *equivalence class* iff they are shifted translations of each other. With this partition, the set E_m is decomposed into equivalence classes. A subset of E_m that contains one single element from each equivalence class is called the *representative set* of E_m.

From the proof of Theorem 2 we have the following result.

Corollary 1. E_m defined by (10) is a subcode of the Reed–Solomon code that is closed under cyclic shift and translation operations. Each element of E_m is a codeword of maximum period. E_r is a representative set of E_m. Consequently, E_m contains

$$|E_m| = nq^{k-1} = (q-1)q^{k-1}$$

distinct codewords.

The families defined in Theorem 2 are optimal according to Theorem 1.

Corollary 2. The families of sequences defined in Theorem 2 are optimal for all $q > 2$ and $2 \leqq k \leqq q-1$.

$$M(E_r) = k-1$$

Proof. By Theorem 1 and Theorem 2

$$\begin{aligned} k-1 \geqq M(E_r) &\geqq \lceil \log_q [(q-1)q^{k-2}] \rceil \\ &= \left\lceil k-1+\frac{\ln(1-q^{-1})}{\ln q} \right\rceil \\ &\geqq \left\lceil k-1-\frac{1}{(q-1)\ln q} \right\rceil = k-1. \end{aligned}$$

Our results show a very interesting property of the Reed–Solomon code which can be stated as follows. In a Reed–Solomon code almost all codewords have no inner period. More precisely, a codeword x of a cyclic code with length n is said to have *inner period* if there exists some i, $0 < i < n$, such that $S^i x = x$.

Corollary 3. Let T be the number of codewords of the Reed–Solomon code over $GF(q)$ and T_0 be maximum number of codewords which have no inner period, then

$$1 > \frac{T_0}{T} \geqq 1 - \frac{1}{q}.$$

In other words, for large q almost all codewords have no inner period.

Proof. It is obvious, by Theorem 2, see also Corollary 1, that

$$T_0 \geqq (q-1)q \cdot q^{k-2}.$$

The statement of the Corollary follows immediately.

Families of sequences with good generalized Hamming correlation properties can be used in many interesting areas. In [1] we have shown how to use families of sequences with optimal generalized Hamming correlation properties to construct signature codes for frequency-hopped systems. These codes have good acquisition (synchronization) properties, can be decoded easily, and result in higher spectral efficiency than the Einarsson code [3]. In [2] optimal families of sequences have been used to construct nonlinear binary cyclic codes with constant weight which can be used as protocol sequences for collision channels without feedback.

Corollary 3 shows a very interesting feature of the Reed–Solomon code, i.e. almost all of its codewords have maximal period. It is shown in [2] that the same is true for a large class of linear codes.

It is worthwhile to note that optimal families of sequences as that given in Theorem 2, and families of optimal sequences do not mean the same. The problem of finding families of optimal sequences according to the Lemma, see also [4], is an open problem. Another open question is to determine the exact number of codewords having no inner period in the Reed–Solomon code.

References

1. *Nguyen, Q. A*, On High Spectral Efficiency Coding of Frequency-Hopped Channels. Int. Symp. on EMC, EMC-86, Wroclaw, June 1986.
2. *Nguyen Q. A, Györfi, L., Massey, J. L.*, Some constructions of cyclic sets, cyclic constant weight binary codes. Submitted to IEEE Trans. Inf. Theory.
3. *Einarsson, G.*, Address Assignment for a Time-Frequency-Coded, Spread-Spectrum System. The Bell Syst. Tech. J. Vol. **57**, No. 7, pp. 1241–1255, Sept. 1980.
4. *Lempel, A., Greenberger, H.*, Families of Sequences with Optimal Hamming Correlation Properties, IEEE Trans. Inf. Theory, **IT-20**, pp. 90–94, Jan. 1974.
5. *MacWilliams, F. J., Sloane, N. J. A., The Theory of Error-Correcting Codes.* North Holland Publishing Co. 1977.
6. *Solomon, G.*, Self-Synchronizing Reed–Solomon Codes. IEEE Trans. on Information Theory, July 1968, pp. 608–609.

Семейства последовательностей с оптимальными обобщёнными корреляционными свойствами Хэмминга

НГУЕН КВАНГ А

(Ханой)

Вводятся обобщённые корреляционные функции Хэмминга. Даются границы для автокорреляции и взаимокорреляции. Даются конструкции семейства последовательностей с оптимальными обобщенными корреляционными функциями Хэмминга.

Nguyen Quang A
Technical University of Budapest
H-1111 Budapest, XI., Stoczek u. 2.
Hungary
and
Institute of Computer Science and Cybernetics
Vietnamese Academy of Sciences
Hanoi
Vietnam

Problems of Control and Information Theory, Vol. 17 (3), pp. 125—136 (1988)

SET POINT CONTROL AND OFFSET COMPENSATION IN DISCRETE *LQ* ADAPTIVE CONTROL

J. BÖHM

(Prague)

(Received February 1, 1987)

The set point control occurs frequently in practice resulting in d.c. value in the system output. The use of feedback controllers not designed with respect to d.c. value can lead to steady state error, the offset. The extension of *LQ* design oriented to adaptive regulators for systems with d.c. value is dealt with in the paper. Some atention is paid to the offset problem and its compensation. The problem is solved in detail for the square root method of *LQ* design.

1. Introduction

Linearization of the mathematical description of a real system leads to a linear difference equation relating the input u_t and the output y_t of the system in the form

$$y_t = \sum_{i=1}^{MR} a_i y_{t-i} + \sum_{i=0}^{MR} b_i u_{t-i} + k + v_t \tag{1}$$

where b_i, a_i are coefficients of the system,

v_t represents stochastic disturbance,

k represents all bias in the system, either inherent in the system itself or caused by a disturbance.

Presence of the d.c. value in the system output is the result of $k \neq 0$. It is known that the feedback compensation of a d.c. value without an integration term causes offset. The same situation occurs when a nonzero set point is assumed.

Standard textbooks frequently suppose $k = 0$. A suitable choice of the scaling of variables can actually make any d.c. value equal to zero for a known constant system.

In the case of control of real systems, d.c. value should be admitted and special measures must be taken to cope with the offset problem. The usual measure is to incorporate the integrator action in the open loop. An alternative approach uses the system model in an incremental form

$$\Delta y_t = \sum_{i=1}^{MR} a_i \Delta y_{t-i} + \sum_{i=0}^{MR} b_i \Delta u_{t-i} \tag{2}$$

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

and the controller based on such a model automatically compensate any possible d.c. value.

In both cases integration action is added to the open loop explicitly. This changes closed loop properties including stability conditions and robustness.

An alternative approach to solving the problem with d.c. value for adaptive regulators based on model (1) up to the second order, is treated in [1] and the offset compensation is dealt with in [2].

In this article the solution is extended to the case of the general model (1). First, the minimization of the general quadratic criterion will be performed, including d.c. value and variable set point. An analysis of the result is made and a detailed solution of the set point control with offset compensation will be given for the square root optimization procedure of the quadratic criterion used in *LQ* general purpose adaptive regulators.

2. Minimization of the quadratic criterion in the presence of d.c. value and nonzero set point

The system described by the linear multivariate regression model (1) will be considered in the form

$$y_t = Pz_t + k + e_t \tag{1'}$$

where $y_t' = (y_t^1, y_t^2, \ldots, y_t^{MY})$ $\quad k' = [k_1, k_2, \ldots, k_{MY}]$

$u_t = (u_t', u_t^2, \ldots, u_t^{MU})$

$e_t =$ white noise, $E\{e_t\} = 0$, $\operatorname{cov}(e_t) = R$

$P = [B_0, A_1, B_1, \ldots, A_{MR}, B_{MR}]$

$x_{t-1}' = [y_{t-1}, u_{t-1}, \ldots, y_{t-MR}, u_{t-MR}]$

$z_t' = [u_t, x_{t-1}]$

A_i, B_i are matrix coefficients of corresponding dimensions. For the time being let us take them to be known. When applied to adaptive regulators, A_i, B_i will be substituted by their estimates obtained by some identification method. (See e.g. [5].)

For the successive minimization of a quadratic criterion the state space form of the regression model (1) will be more convenient

$$x_t = P_x x_{t-1} + P_u u_t + \hat{k} + \Gamma e_t \tag{1''}$$

with

$$P_x = \begin{bmatrix} & & P & & & & P_u = & b_0 & \Gamma = & I_y & \tilde{k} = & k \\ 0 & 0 & . & . & 0 & 0 & & I_u & & 0 & & 0 \\ 1 & 0 & . & . & 0 & 0 & & 0 & & 0 & & 0 \\ 0 & 1 & 0 & . & . & 0 & 0 & 0 & & 0 & & 0 \\ . & . & & . & . & . & . & . & & . & & . \\ 0 & 0 & . & . & . & 0 & 0 & 0 & & 0 & & 0 \end{bmatrix}$$

I_y, I_u are MY^*MY or MU^*MU unit matrices
P_x has dimension $IR \times IR$ where $IR = MR^*(MU + MY)$
P_u has dimension $MU \times IR$.

Standard solution of the LQ problem for $k = 0$

Let us briefly recapitulate the standard solution of the *LQ* problem. The quadratic criterion is

$$J_T \min_{u_1 \ldots u_T} \frac{1}{T} E\left\{ \sum_{t=1}^{T} (y_t' Q_y y_t + u_t' Q_u u_t)/u_0, y_0 \right\} \tag{3}$$

or in the state form

$$J_T = \min_{u_1 \ldots u_T} \frac{1}{T} E\left\{ \sum_{t=1}^{T} x_t' Q_x x_t / x_0 \right\} \tag{3'}$$

$$Q_x = \begin{bmatrix} Q_y & 0 & 0 & . & . & 0 \\ 0 & Q_u & 0 & . & . & 0 \\ 0 & 0 & 0 & . & . & 0 \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & 0 \end{bmatrix}$$

$E\{.../x\}$ stands for conditional expectation.

The part $y_t' Q_y y_t$ measures the distance of the output from the desired value 0 and $u_t' Q_u u_t$ penalizes the "control effort" to achieve the desired value over the interval of planning T. The result of optimization for $T \to \infty$ is mostly needed for real processes.

The solution is well known. It is obtained e.g. via the solution of the corresponding matrix Riccati equation and can be written in the following form

$$u_t = -(P_u' \tilde{G}_{t+1} P_u + Q_u)^{-1} P_u' \tilde{G}_{t+1} P_x \; x_{t-1} = -L_t x_{t-1} \tag{4}$$

with

$$G_t = \tilde{G}_{t+1} - L_t' P_u' \tilde{G}_{t+1} P_u L_t - L_t' Q_u L_t \tag{5}$$

$$\tilde{G}_{t+1} = G_{t+1} + Q_x. \tag{6}$$

Minimization of the quadratic criterion for variable set point w and $k \neq 0$

The following criterion is then to be minimized

$$J_T = {}^1/T \min_{u_1 \ldots u_T} E\left\{ \sum_{t=1}^{T} (x_t - \Gamma w_t)' Q_x (x_t - \Gamma w_t) \right\} \tag{7}$$

Criterion (7) can be minimized successively starting from the end of the interval $t = T$ to $t = 1$ by a method similar to dynamic programming described in detail in [3, 4].

The process of successive minimizations can be expressed as an evolution of the cost-to-go for $t = T, T-1, \ldots, 1$. It can be written as

$$K^*(t, T) = \min_{u_t} E\{(x_t - \Gamma w_t)' Q_x (x_t - \Gamma w_t) + K^*(t+1, T)/y_{t-1}, u_t\} \tag{8}$$

with

$$J_* = 1/T K^*(1, T) \qquad K^*(T+1, T) = 0. \tag{9}$$

Let us follow one step of optimization in detail. The optimal cost-to-go can be expected in the general quadratic form. Then

$$K^*(t+1, T) = x_t' G_{t+1} x_t + 2M_{t+1}' x_t + N_{t+1}. \tag{10}$$

After performing the expectation in (8) considering (10) and using model (1'') we obtain

$$\begin{aligned} K(t, T) = {} & x_{t-1}' H_{xx,t} x_{t-1} + 2u_t' H_{ux,t} x_{t-1} + u_t' H_{uu,t} u_t + 2\tilde{k}' G_{t+1} P_x x_{t-1} + \\ & + 2k_{t+1}' P_u u_t + k_{t+1}' \tilde{k} + \operatorname{tr}(\Gamma' G_{t+1} \Gamma R) + 2M_{t+1}' P_u u_t + \\ & + 2M_{t+1}' P_x x_{t-1} + 2M_{t+1}' \tilde{k} + N_{t+1} + (\tilde{k} - \Gamma w_t)' Q_x (\tilde{k} - \Gamma w_t) + \\ & + 2(\tilde{k} - \Gamma w_t)' Q_x P_x x_{t-1} + 2u_t' P_u' Q_x (\tilde{k} - \Gamma w_t). \end{aligned} \tag{11}$$

The optimal u_t can be obtained by deriving $K(t, T)$. So from

$$\frac{\partial K(t, T)}{\partial u_t} = 0$$

it follows

$$u_t^* = -H_{uu,t}^{-1}(H_{ux,t}x_{t-1} + P_u'G_{t+1}\tilde{k} + P_u'M_{t+1} + P_u'Q_x(\tilde{k} - \Gamma w_t) \tag{12}$$

where

$$H_{xx,t} = P_x'\tilde{G}_{t+1}P_x$$

$$H_{ux,t} = P_u'\tilde{G}_{t+1}P_x$$

$$H_{uu,t} = P_u'\tilde{G}_{t+1}P_u$$

$$\tilde{G}_{t+1} = G_{t+1} + Q_x.$$

Substituting (12) into (11) the optimal $K^*(t, T)$ will be obtained in the form of (10) with

$$G_t = H_{xx,t} - H_{ux,t}'H_{uu,t}^{-1}H_{ux,t} \tag{13a}$$

$$M_t' = (M_{t+1}' + \tilde{k}'G_{t+1} + (\tilde{k} - \Gamma w_t)'Q_x) \tag{13b}$$

$$(P_x - P_uH_{uu,t}^{-1}H_{ux,t})$$

$$N_t = N_{t+1} + 2M_{t+1}'\tilde{k} + \tilde{k}'G_{t+1}\tilde{k} + (\tilde{k} - \Gamma w_t)'Q_x(\tilde{k} - \Gamma w_t) +$$

$$+ \operatorname{tr}\Gamma'G_{t+1}\Gamma R - (\tilde{k}'G_{t+1} + M_{t+1}' + (\tilde{k}' - \Gamma w_t)'Q_x)$$

$$P_uH_{uu,t}^{-1}P_u'(G_{t+1}\tilde{k} + M_{t+1} + Q_x(\tilde{k} - \Gamma w_t)). \tag{13c}$$

Comparing (13a) with (5) we can see that the evolution of G is the same and does not depend on k or $w = 0$, $k = 0$. From (12) we can deduce that u can be expressed in the form

$$u_t^* = u_{t,f} + u_{t,k} + u_{t,w} \tag{14}$$

where $u_{t,f}$ is given by (4) and represents the feedback for $k = 0$ and $w = 0$, while $u_{t,k}$ and $u_{t,w}$ are the part of the input corresponding with k and w_t.

Let us follow the behaviour of criterion (10) for $T \to \infty$. It is known that in a stabilizable system G_t converges to steady state matrix G_∞. The steady state values of M_t and N_t can be searched only for $w_t \to w$ for $T \to \infty$.

The evolution of M_t is the following

$$M_t' = M_{t+1}' + \tilde{k}'\tilde{G}_{t+1}\Omega - (\Gamma w)'Q_x\Omega \tag{15}$$

where Ω is the closed loop matrix

$$\Omega = (P_x - P_uH_{uu}^{-1}H_{ux})$$

Equation (15) is the first order multivariate linear equation for M. For stable Ω, (15) is stable and the steady state solution for $k = \text{const.}$ and $w_t = w = \text{const.}$ is given by

$$M_\infty = (I - \Omega)^{-1}\Omega(\tilde{G}_\infty\tilde{k} - Q_x\Gamma w). \tag{16}$$

The absolute term N_t grows but its increment $\Delta N_t = N_t - N_{t+1}$ is limited as all components of N are finite constants or variables converging to finite values. We can write

$$N_1 = 2N_T + \sum_{t=1}^{T-1} \Delta N_t. \tag{17}$$

Finally, the value of the criterion (7) is

$$J_T = 1/T\left\{x_0' G_1 x_0 + 2M_1' x_0 + 2N_T + \sum_{r=1}^{T-1} \Delta N_t\right\} \tag{18}$$

and the infinite horizon criterion will be

$$J_\infty = \lim_{T\to\infty} J_T = \Delta N_\infty. \tag{18'}$$

Using (13) the ΔN_∞ is given by the following formula

$$\Delta N_\infty = \tilde{k}'(1-\Omega)^{-1}(1+\Omega)\tilde{G}\tilde{k} - 2\tilde{k}'(1-\Omega)^{-1}Q_x\Gamma w + (\Gamma w)'Q_x\Gamma w -$$
$$-(\tilde{G}\tilde{k} - Q_x\Gamma w)'(1-\Omega)^{-1}P_u H_{uu}^{-1}P_u'(1-\Omega)^{-1}(\tilde{G}\tilde{k} - Q_x\Gamma w). \tag{19}$$

From the condition for control without offset it follows that $J_\infty = 0$. Consequently, ΔN_∞ is equal to zero. To avoid offset we will aim at deriving such control that $\Delta N_\infty = 0$.

It is known that for standard LQ synthesis $\Delta N_\infty = 0$ only for $Q_u = 0$. Using $Q_u \neq 0$ the minimum of the criterion is reached even if $E\{y_\infty - w\} \neq 0$. The reason is that in the criterion the demand for $E\{y_\infty - w\} = 0$ is not explicitly expressed.

Minimization of the criterion with an additional condition

The condition for zero offset is usually defined as

$$\lim_{T\to\infty} E\{y_t - w_t\} = 0. \tag{20}$$

As our optimization problem is formulated for a finite horizon criterion (3) it would be better to reformulate the condition in terms of finite horizon. The criterion

$$E\left\{1/T\sum_{t=1}^{T}(y_t - w_t)\right\} = 0 \tag{20'}$$

is well suited to be incorporated to criterion (3) using the Lagrange multipliers. It even reflects the demand for finite time "offset" and for T tending to infinity passes to criterion (20).

The criterion to be minimized is then

$$J_T = 1/T \min_{u_1 \ldots, u_T} E\left\{ \sum_{t=1}^{T} (x_t - \Gamma w_t)' Q_x (x_t - \Gamma w_t) + \lambda (y_t - w_t)) \right. \tag{21}$$

where λ is a Lagrange multiplier.

The minimization of this criterion can be performed in the same way as in the previous case. The general step of minimization of (21) gives

$$\begin{aligned} u_t^* &= -H_{uu}^{-1}(H_{ux} x_{t-1} + P_u'(G_{t+1}\tilde{k} + M_{t+1} + Q_x(\tilde{k} - \Gamma w_t) + 1/2\Gamma\lambda)) \\ &= u_x + u_k + u_w + u_\lambda . \end{aligned} \tag{22}$$

Optimal u_t is similar to (12), there is just one more term, namely u_λ. The quadratic term will not be influenced by λ but the other terms become a function of λ. The optimal value of λ is obtained from condition (20). First, $x_\infty = x_t$ for $t \to \infty$ is calculated using (21), (1″) and $x_\infty - w = 0$ gives the equation for λ. The result is

$$\begin{aligned} \lambda = (\Gamma'(I-\Omega)^{-1} P_u H_{uu}^{-1} P_u' \Gamma)^{-1} (w - \Gamma'(I-\Omega)^{-1} [\tilde{k} - \Gamma w - \\ - P_u H_{uu}^{-1} P_u' (I-\Omega)^{-1} (\tilde{G}\tilde{k} - Q_x \Gamma w)]) . \end{aligned} \tag{23}$$

In this way the problem was theoretically solved for the case of constant k and w.

Remark. There is one possibility to calculate λ even for finite T, namely for $T=1$. In that case λ becomes a function of x_t. For different t we obtain different λ_t according to x_t but it will ensure $E\{y_t - w_t\} = 0$ in each step and arbitrary k and w. However, the stability of the closed loop is not guaranteed for such criterion.

3. An alternative modification of the criterion

Relation (23) is not well suited for practical calculations. From the previous analysis we have seen that for the compensation of the offset it is necessary to add the term u_λ that is a function of k and w. Let us consider that u_t for $t \to \infty$ will be composed of

$$\tilde{u}_t = u_t + u_0 \tag{24}$$

where u_0 is sought so that the condition for nonzero offset may be fulfilled. Due to (24) the penalization of u will now have the form $(u_t - u_0)' Q_u (u_t - u_0)$ and the criterion to be minimized is now

$$J_T = 1/T \min_{u_1 \ldots u_t} E\left\{ \sum_{t=1}^{T} (y_t - w_t)' Q_y (y_t - w_t) + (u_t - u_0)' Q_u (u_t - u_0)/x_0 \right\} \tag{25}$$

and in the state space form

$$J_T = 1/T \min_{u_1 \ldots u_T} E\left\{ \sum_{t=1}^{T} (x - \Gamma_{w,u_0})' Q (x - \Gamma_{w,u_0})/x_0 \right\} \tag{25'}$$

where

$$\Gamma'_{w,u_1} = [w, u_0, 0, 0, \ldots, 0]. \tag{26}$$

The minimization differs from (7) only in term Γ_{w,u_0} and the results previously obtained can be directly used. The optimal control is

$$u_t^* = H_{uu}^{-1}(P'_u(\tilde{G}_{t+1}\tilde{k} - Q_x\Gamma_{w,u_0} + M_{t+1}) + H_{uu}^{-1}H_{ux}x_{t-1}. \tag{27}$$

The part N becomes now a function of u_0. From (19) u_0 can be chosen in such a way that $\Delta N_\infty = 0$. However, the resulting formula is also complicated. In the following part of the paper the problem will be solved in the framework of square root technique, and it will be shown that the solution of the offset problem is much more accessible in this way.

4. Set point control in the square root minimization technique

In the previous part we have solved the problem of offset for $k = \text{const}$, $w = \text{const}$ and obtained some teoretical results that are rather complicated for practical use.

The square root method is very efficient for real time adaptive algorithms where higher numerical stability and computational efficiency is required mainly to ensure the nonnegative definiteness of the minimized quadratic form even when computing with reduced word length.

The square root LQ synthesis is described in detail in [4] where the basic optimization procedure is derived and the variable set point control is dealt with, however, the problem of the offset is not solved.

To present the results of the offset compensation we will need to recapitulate the basic features of the square root approach and the results obtained. After that a very simple idea of how the offset can be compensated will be presented and solved for the single input single output case.

The square root minimization is similarly based on dynamic programming but the matrices G, N and M in (10) are propagated throughout the whole recursion in the factorized form. The following ideas are implemented in the square root method.

1) The regression model (1) is used to describe the system and the criterion (25) is used to express our demands.

2) The general quadratic form of the cost-to-go (10) is expressed by one square symmetric matrix S and the augmented vector $\tilde{x}' = [x', 1]$ in the form

$$K^*(t, T) = \tilde{x}_{t-1}\tilde{S}_t\tilde{x}_{t-1} \tag{30}$$

and (8) have the form

$$K(t, T) = E\{\tilde{x}_t(\tilde{S}_{t+1} + \tilde{Q}_x)\tilde{x}_t/\tilde{x}_{t-1}u_t\}. \tag{31}$$

In this case the regression model will be written in the form

$$\tilde{x}_t = \tilde{P}\tilde{z}_t \qquad \text{where} \qquad \tilde{P} = \begin{bmatrix} & & P & & & k \\ 1 & 0 & . & . & 0\ 0 & 0 \\ 0 & 1 & 0 & . & 0\ 0 & 0 \\ . & . & . & . & .\ .\ & . \\ 0 & 0 & . & . & 0\ 0 & 1 \end{bmatrix}$$

3) The matrices $\tilde{S}_t$ and $\tilde{Q}_x$ are factorized into the Cholesky square root factors S, Q

$$S'S = \tilde{S} \qquad S = \tilde{S}^{1/2}$$
$$Q'Q = \tilde{Q}_x \qquad Q = \tilde{Q}_x^{1/2}. \tag{32}$$

For our purpose the upper triangular form will be suitable. Factors of Q have special form

$$Q = \tilde{Q}_x^{1/2} = \begin{bmatrix} Q_y^{1/2} & 0 & 0 & . & . & 0 & -Q_y^{1/2} & w \\ 0 & Q_u^{1/2} & 0 & . & . & 0 & -Q_u^{1/2} & u_0 \\ 0 & 0 & 0 & . & . & 0 & 0 & \\ . & . & . & . & . & . & . & \\ 0 & 0 & 0 & . & . & 0 & 0 & \end{bmatrix}$$

Under these conditions $K(t, T)$ can be written as

$$K^*(t, T) = \min_{u_t} E\left\{\tilde{x}_t \begin{bmatrix} S_{t+1} \\ Q \end{bmatrix}' \begin{bmatrix} S_{t+1} \\ Q \end{bmatrix} \tilde{x}_t / \tilde{x}_{t-1}, u_t\right\} \tag{33}$$

where $\begin{bmatrix} S \\ Q \end{bmatrix}$ means the extension of S by rows of Q.

4) The expectation in (33) is calculated by

$$\begin{bmatrix} S \\ Q \end{bmatrix} \tilde{P}$$

as

$$K^*(t, T) = \min_{u_t} \tilde{z}_t' \tilde{P}' \begin{bmatrix} S \\ Q \end{bmatrix}' \begin{bmatrix} S \\ Q \end{bmatrix} \tilde{P} \tilde{z}_t$$

$$= \min_{u_t} \tilde{z}_t' H' H \tilde{z}_t. \tag{34}$$

5) The minimization step is performed in such a way that $\bar{H}$ is converted to the triangular matrix $\tilde{H}$ by the orthogonal transformation, using the identity

$$\tilde{z}_t'\bar{H}'\bar{H}\tilde{z}_t = \tilde{z}_t'\bar{H}'T'T\bar{H}\tilde{z}_t = \tilde{z}_t'\tilde{H}'\tilde{H}\tilde{z}_t$$

with

$$T'T = I \qquad \text{and} \qquad T\bar{H} = \tilde{H}.$$

Due to the structure of $\tilde{z}$ and the triangular form of $\tilde{H}$, u_t influences only the MU first elements of $\tilde{H}\tilde{z}_t$. To minimize $K(t, T)$, it is sufficient to choose such u_t^* that it makes these elements equal to zero. I.e.

$$\tilde{H}_{uu}u_t^* = -\tilde{H}_{ux}\tilde{x}_{t-1} \tag{35}$$

where the decomposition of H to the submatrices corresponds with the composition of the vector $\tilde{z}_t$.

The rest of $\tilde{H}$, i.e. $\tilde{H}_{xx} = S_t$ represent the optimal cost-to-go

$$K^*(t, T) = \tilde{x}_{t-1}'S_t'S_t\tilde{x}_{t-1}.$$

The matrix $S(\tilde{H}_{xx})$ can be further decomposed to the parts S_{x1}, S_{xx} and S_{11}.

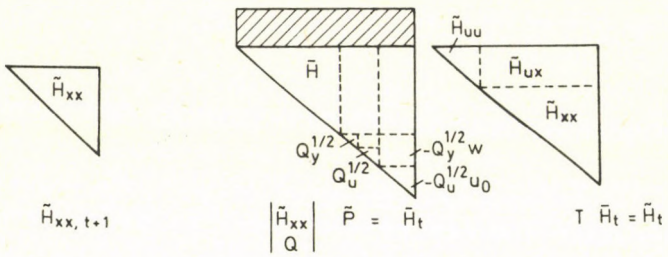


Fig. 1

The evolution of S_{x1} and S_{11} is also apparent from the diagram. We can follow that in the expectation step S_{x1} is shifted upwards and the vacant elements are filled with

$$-Q_y^{1/2}w_t, \quad -Q_u^{1/2}u_0.$$

During the minimization step the column will be changed by reductions of elements indicated by hatching.

Now we are getting to the main point. The absolute part of the criterion causing offset is formed by those parts of the reduced rows that remain after $IR-1$ reductions. The IRth reduction only signifies the accumulation of residuals. Now the question arises if it is possible to choose the values of u_0 in S_{x1} so that the residuals vanish.

The answer is positive. The method of its implementation is based on the algorithm of evolution of S_{x1}. The idea will be explained in a simple case of one input, one output ($MU=1$, $MY=1$) and for the regression model of order MR.Then $\rho=2*MR+2$ and there are ρ coefficients c_i, s_i determining the successive reduction of one row. The last column of H_{xx} will be influenced by the last element of the reduced row. Let us indicate it by r. After the first $\rho-3$ reductions the first $\rho-3$ elements of S_{x1} are changed and r is also changed to $r_{\rho-3}$.

Further reductions of r have the following form

$$S_{x1-\rho-2}=-c_{\rho-2}Q_y w_t+s_{\rho-2}r_{\rho-3}$$

$$r_{\rho-2}=c_{\rho-2}r_{\rho-3}+s_{\rho-2}Q_y w_t$$

$$S_{x1-1}=-c_{\rho-1}Q_u u_{0,t}+s_{\rho-1}r_{\rho-2}$$

$$r_{\rho-1}=c_{\rho-1}r_{\rho-2}+s_{\rho-1}Q_u u_{0,t}$$

the condition for zero increment of the absolute term is

$$r_{\rho-1}=0$$

from which

$$Q_u u_{0,t}=c_{\rho-1}/s_{\rho-1}r_{\rho-2}=c_{\rho-1}/s_{\rho-1}(c_{\rho-2}r_{\rho-3}+s_{\rho-2}Q_y w_t).$$

For the multivariate case the complexity of the solution grows with dimensionality of input but it always leads to the systems of linear equations.

The properties of the method are as follows:

1) It makes the absolute term of the quadratic form equal to zero to fulfil $\lim_{t\to\infty}(y_t-w)=0$.

2) Given solution influences only the part of u_t that represents k, w, u_0. The roots of the closed loop are not affected.

3) It can be used for arbitrary w and varying k. No assumptions about their form have been made. Of course u_0 will vary as well.

4) It is a simple extension of the existing square root algorithm.

5) Modification for $w\neq$const can be deduced, the complexity of which does not grow much for the multivariate case.

5. Conclusion

The problems of the set point control, absolute term k in regression model and the offset have been discussed. It has been shown that the presence of $k\neq 0$ and the set point control is treated in a similar way and the solution is a rather simple extension of known results. The offset originates from the incorrect application of

quadratic criterion that usually does not reflect the demand for zero offset. Two ways of modification of the quadratic criterion have been proposed and the offset compensation has been solved for $k = \text{const}$, $w = \text{const}$.

Then the offset problem has been solved in the square root optimization method that is used for the design of adaptive regulators. The offset compensation is not limited to $k = \text{const}$, $w = \text{const}$ in this case. The methods of the offset compensations are so simple that regulators based on normal regression model can compete with the incremental ones or even prove superior.

References

1. *Peterka, V., Böhm, J., Halousková, A., Karný, M., Maršík, J.*, Algorithms for the Adaptive Microprocessor Control of Technological Processes (in Czech). UTIA CSAV. Prague 1982.
2. *Rosol, I.*, Adaptive control with restrictive conditions. Report no 1314 UTIA 1985 (in Czech)
3. *Peterka, V.*, Digital Control of Processes with Stochastic Disturbance and Uncertain Characteristics (in Czech). Doctor of Science Dissertation. UTIA CSAV, Prague 1975.
4. *Karný, M., Halousková, A., Böhm, J., Kulhavý, R., Nedoma, P.*, Design of Linear Quadratic Adaptive Control: Theory and Algorithm for Practice. Supplement to Kybernetika, 3–6, Prague 1985.
5. *Peterka, V.*, Bayesian approach to system identification. In: Trends and Progress in System Identification (P. Eykhoff, ed.) Pergamon Press, Oxford 1981, 239–304.

Управление и компенсация отклонения с дискретными адаптивными регуляторами, основанными на линейно-квадратическом синтезе

Й. БОМ

(Прага)

Работа посвящена построению адаптивных регуляторов для линейных управляющих систем с постоянной составляющей. Регулятор обеспечивает управление и компенсацию возмущений с ненулевым средним значением без помощи интеграционного звена.

Указан способ выбора квадратического критерия оптимальности, приводящего к соответствующему управлению, обеспечивающему нулевое отклонение в установленном состоянии. Предложенные процедуры вычисления основаны на методе разложения матриц в квадратные корни.

J. Böhm
Institute of Information Theory and Automation
Pod vodárenskou věží 4
18208 Prague 8
Czechoslovakia

Problems of Control and Information Theory, Vol. 17 (3), pp. 137—144 (1988)

DYNAMICS OF THE SET OF VIABLE TRAJECTORIES TO A DIFFERENTIAL INCLUSION: THE EVOLUTION EQUATION

A. B. KURZHANSKII, T. F. FILIPPOVA

(*Sverdlovsk*)

(Received June 22, 1987)

The paper deals with the specification of all solutions to a differential inclusion that satisfy a preassigned restriction on the state space variables (the assembly of viable trajectories). The equation describing the evolution of the attainability domain for a given inclusion under phase constraints is proposed.

1. Introduction

The present paper is devoted to the problems of constructing mathematical models describing the dynamics of complex nonlinear systems with uncertainty. The latter means that the disturbances in system inputs and in the current measurement are uncertain. They are taken to be unknown in advance with respective information being restricted to only a set-membership description of their values. It is assumed also that there is no statistical information on the unknowns.

Under such assumptions the system state at each instant is not single-valued. Therefore, one cannot describe the total dynamics of the studied system by a solution of a differential equation as it was in classical control theory. The natural formalization of the problem considered below is fulfilled on the basis of ideas and methods of differential inclusions theory [1–3]. According to this formalization in place of an isolated process trajectory we have an assembly (or a bundle) of solutions of a differential inclusion that arise from a given set of possible initial state vectors. Proposing a differential inclusion to be an adequate model we must take into account the available measurements of system parameters. The latter may be expressed mathematically in the introduction of the so called informational domains [4–5]. These domains are generated by all the solutions to a differential inclusion that are consistent with the results of on-line observations. In other words, considering the measurement procedure as a phase constraint on the state space variables, we must reduce the set of inclusion trajectories to those that are viable with respect to the phase restriction [1–3].

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

As it is known [6–7], the tube of viable trajectories may be described by a new differential inclusion whose right-hand side is formed with the aid of a tangent cone to the phase restriction. We develop here another approach to the problem under consideration in order to avoid the cumbersome procedure of constructing the above-mentioned cone-valued mapping.

Note that time-cross-sections of the set of viable trajectories may be considered as analogies of the phase vectors for the standard control system since they absorb all available information on the system behaviour. Then the problem of discovering the evolution law for these "phase states" of the uncertain process is relevant. This paper is devoted to the solution of this problem and continues the investigations [1–3]. It is demonstrated that cross-sections of the viable solution assembly satisfy a special evolution equation. This equation generalizes the integral funnel equation [8–9] and coincides with the latter in the absence of phase constraints.

2. Statement of the problem

Let R^n be the n-dimensional Euclidean space. For $x, y \in R^n$ let $x'y$ (or (x, y)) denote the usual inner product of x and y with the prime as the transpose, $\|x\| = (x'x)^{1/2}$, $S = \{x \in R^n : \|x\| \leqq 1\}$. Denote conv R^n to be the set of convex compact subsets of R^n, $h(A, B)$ to be the Hausdorff metric for $A, B \in \text{conv } R^n$.

Consider the following differential inclusion

$$\dot{x} \in F(t, x) \qquad (t_0 \leqq t \leqq T) \tag{2.1}$$

where $x \in R^n$, F is a continuous mapping from $[t_0, T] \times R^n$ into conv R^n. We will assume the Lipschitz condition for F to be satisfied ($L > 0$):

$$h(F(t, x), \quad F(t, y)) \leqq L\|x - y\|, \quad \forall x, y \in R^n.$$

Assuming set $X_0 \in \text{conv } R^n$ to be given, denote $x[t] = x(t; t_0, x_0)$ $(t_0 \leqq t \leqq T)$ to be the Carathéodory-type solution to (2.1) that starts at $x[t_0] = x_0 \in X_0$. We further require all solutions $\{x(t; t_0, x_0) | x_0 \in X_0\}$ to be extendable until instant T [10].

Let $Y(t)$ be a continuous mapping from $[t_0, T]$ into conv R^n, $X_0 \subseteq Y(t_0)$.

Definition 2.1 [1–3]. A trajectory $x[t] = x(t; t_0, x_0)$ $(x_0 \in X_0, \ t_0 \leqq t \leqq T)$ of differential inclusion (2.1) will be said to be viable on $[t_0, \tau]$ $(\tau \leqq T)$ if

$$x[t] \in Y(t) \qquad \text{for all} \quad t \in [t_0, \tau]. \tag{2.2}$$

For every $x_0 \in X_0$ the set of all viable on $[t_0, \tau]$ trajectories $x(\cdot\,; t_0, x_0)$ will be denoted as $X(\cdot\,; \tau, t_0, x_0)$

$$X(\cdot\,; \tau, t_0, X_0) = U\{X(\cdot\,; \tau, t_0, x_0) | x_0 \in X_0\},$$

and its cross-section at instant τ as $X(\tau; t_0, x_0)$ and $X(\tau; t_0, X_0)$, respectively.

Let $X^*(\cdot; t_0, X_0)$ be the solution assembly of a differential inclusion (2.1). Under our assumptions the set

$$M = U\{X^*(t; t_0, X_0) | t_0 \leqq t \leqq T\}$$

is compact in R^n [9, 10]. Let us denote the graph of the map $F(t, \cdot)$ as $gr_t F$ (t is fixed):

$$gr_t F = \{(x, y) \in R^n \times R^n: \quad y \in F(t, x)\}$$

and the interior of $A \subseteq R^n$ as int A.

Assumption A. 1. For some convex compact $D \subset R^n$ such that $M \subset \text{int } D$, the set $D \cap gr_t F$ is convex for every $t \in [t_0, T]$.

2. There exists a solution $x_*[\cdot]$ of inclusion (2.1) such that $x_*[t_0] \in X_0$ and $x_*[t] \in \text{int } Y(t)$ for arbitrary $t \in [t_0, T]$.

Note that under Assumption A, the bundle $X(\cdot; \tau, t_0, X_0)$ of viable trajectories is a convex compact subset in the space $C[t_0, T]$ of all continuous n-vector functions, its τ-cross-section $X(\tau; t_0, X_0)$ is a convex compact subset of R^n.

It is known that sets $X(t; t_0, X_0)$ satisfy a semigroup property:

$$X(\tau; t_0, X_0) = X(\tau; s, X(s; t_0, X_0)).$$

Therefore, they define a generalized dynamic system. The construction of an adequate evolution model describing this system is the subject of the present paper.

3. Evolution equation

We demand further one of the following hypotheses to be satisfied.

Assumption B. The graph $gr Y$ is a convex compact subset of R^{1+n}.

Assumption C. For every $l \in R^n$ the support function

$$\rho(l | Y(t)) = \max \{l'y | y \in Y(t)\}$$

is differentiable in t and its derivative $d/dt\rho(l | Y(t))$ is continuous in (t, l).

The following statement is true:

Theorem 3.1 Suppose Assumption A is fulfilled and the restriction map $Y(\cdot)$ satisfies either Assumption B or Assumption C. Then τ-cross-section $X[\tau] = X(\tau; t_0, X_0)$ of the set $X(\cdot; \tau, t_0, X_0)$ of all viable trajectories to a differential inclusion (2.1) is the solution of the following evolution equation:

$$\lim_{\sigma \to 0+} \sigma^{-1} h(X[\tau + \sigma], \bigcup_{x \in X[\tau]} (x + \sigma F(\tau, x)) \cap Y(\tau + \sigma)) = 0, \tag{3.1}$$

$$X[t_0] = X_0, \qquad t_0 \leqq \tau \leqq T.$$

The proof of this main theorem follows from a number of Lemmas given in the next section. Concluding this paragraph we make some remarks.

Remarks. 1. Assume the hypotheses of Theorem 3.1 to be valid. Then the set-valued mapping $X[\tau]=X(\tau;\, t_0,\, X_0)$ is continuous.

2. Replace Assumption A (2) in Theorem 3.1 by the following one: there exists a solution $x_*[\,\cdot\,]$ of (2.1) such that for almost every $t\in[t_0,\, T]$ $x_*[t]\in \operatorname{int} Y(t)$. Then the evolution equation (3.1) for $X[\tau]$ will be fulfilled almost everywhere on the interval $[t_0,\, T]$. In this case $X[\tau]$ may be discontinuous on a set $\{\tau\}$ of measure zero. (As it is known [4], function $X[\tau]$ is continuous from the left and upper semicontinuous from the right in every point $\tau\in[t_0,\, T]$).

4. Auxiliary constructions

Here we give some preliminary results each of which substitutes a successive step in the proof of the principal Theorem 3.1.

Let $\tau\in[t_0,\, T]$ be fixed, $X[\tau]=X(\tau;\, t_0,\, X_0)$.

First, we have the following estimate:

Lemma 4.1. Suppose Assumption A is true. Then for every $\varepsilon>0$ there exists a $\sigma_*>0$ such that

$$X[\tau+\sigma]\subseteq \bigcup_{x\in X[\tau]} (x+\sigma F(\tau,\, x))\cap Y(\tau+\sigma)+\varepsilon\sigma S \tag{4.1}$$

for all $\sigma\in[0,\, \sigma_*]$.

The next estimate is more precise. Denote for every $\sigma>0$

$$\begin{aligned} Z_\sigma(\tau)=\bigcup_{x\in X[\tau]} \{z\in R^n:\ \ & \exists\, v_x(\,\cdot\,)\in L_1[\tau,\, \tau+\sigma], \\ z=x=\int_\tau^{\tau+\sigma} & v_x(\xi)\,d\xi, \qquad v_x(s)\in F(\tau,\, x), \\ +\int_\tau^s & v_x(\xi)\,d\xi\in Y(s), \qquad \tau\leqq s\leqq\tau+\sigma\} \end{aligned} \tag{4.2}$$

(here $L_1[\tau,\, \tau+\sigma]$ is the space of all integrable on $[\tau,\, \tau+\sigma]$ n-vector functions).

From the definition we have

$$Z_\sigma(\tau)=\bigcup\{Z_\sigma(\tau,\, x_0)\,|\,x_0\in X[\tau]\}$$

where $Z_\sigma(x_0,\tau)$ is the $(\tau+\sigma)$-cross-section of the tube of viable solutions to the differential inclusion

$$\frac{dx}{dt}\in F(\tau,\, x_0), \qquad x(\tau)=x_0,$$

$$x(t)\in Y(t) \qquad (\tau\leqq t\leqq\tau+\sigma).$$

Lemma 4.2. Under Assumption A for every $\varepsilon>0$ there exists $\sigma_*>0$ such that for all $\sigma\in(0,\sigma_*]$ the following inclusions hold:

$$X[\tau+\sigma]\subseteq Z_\sigma(\tau)+\varepsilon\sigma S \tag{4.3}$$

$$Z_\sigma(\tau)\subseteq \bigcup_{x\in X[\tau]}(x+\sigma F(\tau,x))\cap Y(\tau+\sigma)+\varepsilon\sigma S. \tag{4.4}$$

Let $Z'_\sigma(\tau)$ be the subset of $Z_\sigma(\tau)$ related to the constant functions $v_x(\cdot)$ in (4.2):

$$Z'_\sigma(\tau)=\bigcup_{x\in X[\tau]}\{x+\sigma v\,|\,v\in F(\tau,x)\cap(\bigcap_{0<s\leq\sigma}s^{-1}(Y(\tau+s)-x))\}.$$

Lemma 4.3. 1. Suppose Assumption A is true. Then for every $\varepsilon>0$ there exists $\sigma_*>0$ such that for all $\sigma\in(0,\sigma_*]$

$$Z_\sigma(\tau)\subseteq X[\tau+\sigma]+\varepsilon\sigma S. \tag{4.5}$$

2. If Assumption B holds then for every $\sigma>0$

$$\bigcup_{x\in X[\tau]}(x+\sigma F(\tau,x))\cap Y(\tau+\sigma)\subseteq Z'_\sigma(\tau)\subseteq Z_\sigma(\tau). \tag{4.6}$$

Combining Lemmas 4.2 and 4.3 we obtain

Corollary 4.1. Under Assumption A

$$\lim_{\sigma\to0+}\sigma^{-1}h(X[\tau+\sigma],\quad Z_\sigma(\tau))=0.$$

Now introduce some auxiliary constructions. Define for an arbitrary closed set $C\subseteq R^n$ a contingent cone $T_C(x)$ $(x\in C)$:

$$T_C(x)=\left\{v\in R^n\colon \liminf_{\sigma\to0+}\sigma^{-1}d(x+\sigma v,C)=0\right\}$$

and for the multivalued mapping $Y(\cdot)$ a contingent derivative $DY(t,\ y)(\alpha)$ $(\alpha\in R^1,\ (t,\ y)\in grY)$:

$$DY(t,y)(\alpha)=\{v\in R^n\colon\ (\alpha,v)\in T_{grY}(t,y)\}\ \ [6,7]$$

(here $d(x,C)=\min\{\|x-c\|\,|\,c\in C\}$).

Determine

$$V(t,y)=DY(t,y)(1) \tag{4.7}$$

for $(t,\ y)\in grY$. Under Assumption C for all $(t,\ y)\in grY$ the set $V(t,y)$ is closed and convex in R^n [6].

Following [11] consider a local approximation $Y_\sigma(\tau)$ for the set-valued map $Y(\cdot)$ near a fixed point τ:

$$Y_\sigma(\tau)=\bigcap_{y\in Y(\tau)}(y+\sigma V(\tau,y)),\qquad Y_0(\tau)=Y(\tau). \tag{4.8}$$

Lemma 4.4 [11]. 1. Let Assumption C be fulfilled. Then for all $\sigma > 0$ the following equality is true

$$Y_\sigma(\tau) = \{z \in R^n : l'z \leqq \rho(l \mid Y(\tau)) + \sigma d/d\tau \rho(l \mid Y(\tau)),\ \forall\, l \in R^n\}.$$

2. Suppose assumptions A (2) and C hold.

Then for every $\varepsilon > 0$ there exists a $\sigma_* > 0$ such that for all $\sigma \in (0, \sigma_*]$

$$\begin{aligned} Y(\tau+\sigma) &\subseteq Y_\sigma(\tau) + \varepsilon\sigma S \\ Y_\sigma(\tau) &\subseteq Y(\tau+\sigma) + \varepsilon\sigma S. \end{aligned} \tag{4.9}$$

Note that the graph of the map $Y_\sigma(\tau)$ (as a function of σ) is convex. Therefore, in view of Lemmas 4.1–4.4 we have established

Lemma 4.5. Let Assumptions A and C be satisfied. Then

$$\lim_{\sigma \to 0+} \sigma^{-1} h(X[\tau+\sigma],\ \bigcup_{x \in X[\tau]} (x + \sigma F(\tau, x)) \cap Y_\sigma(\tau)) = 0.$$

Summarizing the results of Lemmas 4.1–4.5 we arrive to Theorem 3.1.

5. The linear system

Consider the following system

$$\dot{x} \in A(t)x + P(t) \qquad (t_0 \leqq t \leqq T) \tag{5.1}$$

where $x \in R^n$, $A(t)$ is a continuous $n \times n$-matrix function, $P(t)$ is a continuous map from $[t_0, T]$ into conv R^n.

Let us keep the same notation $X(\cdot\,; \tau, t_0, X_0)$ for the assembly of all solutions of (5.1) viable on $[t_0, \tau]$ with respect to restriction $Y(\cdot)$;

$$X[\tau] = X(\tau; t_0, X_0) = X(\tau; \tau, t_0, X_0).$$

Assumption A (1) is fulfilled automatically for the linear inclusion (5.1). Hence we propose only

Assumption A'. There exists a solution $x_*[\cdot]$ of (5.1) such that

$$x_*[t] \in \operatorname{int} Y(t), \qquad \forall\, t \in [t_0, T].$$

The following result is a direct consequence of Theorem 3.1, it generalizes also Theorem 4.1 [3].

Theorem 5.1. Suppose Assumption A' is fulfilled. If the map $Y(\cdot)$ satisfies either Assumption B or Assumption C then the set-valued function $X[\tau] = X(\tau; t_0, X_0)$ is

the solution of the evolution equation

$$\lim_{\sigma \to 0+} \sigma^{-1} h(X[\tau+\sigma],\ ((E+\sigma A(\tau))X[\tau]+\sigma P(\tau)) \cap Y(\tau+\sigma)) = 0 \tag{5.2}$$

$$X[t_0] = X_0, \qquad t_0 \leqq \tau \leqq T$$

(here E is the identity $n \times n$-matrix).

One may prove an analogy of Theorem 4.2 [3] (the convexity of grY is not now supposed).

Theorem 5.2. Under Assumptions A′ and C the support function $\rho(l|X[t])$ is differentiable in t from the right and the exact formula for the derivative is

$$d^+/dt(\rho(l|X[t])) = \min\{\rho(q'A(t)|\partial_l \rho(l|X[t]) + \\ + \rho(q|P(t)) + d/dt(\rho(l-q|Y(t))|q \in Q(t,l)\}$$

where

$$Q(t,l) = \{q \in R^n:\ \rho(l-q|Y(t) = \rho(l|X[t]) - \rho(q|X[t])\},$$

$$\partial_l \rho(l|X[t]) = \{x \in X[t]:\ l'x = \rho(l|X[t])\}.$$

6. Concluding remark

The proposed evolution equation (see Theorems 3.1, 5.1) is useful for the approximation procedures related to the solution of control and observation problems for uncertain systems [4, 5].

References

1. *Kurzhanskii, A. B.*, On analytical description of a set of viable trajectories of a differential system, Dokl. Akad. Nauk SSSR, 1986, **287,** *5*, pp. 1047–1050 (in Russian).
2. *Kurzhanskii, A. B., Filippova, T. F.*, On the description of the set of viable trajectories of a differential inclusion, Dokl. Akad. Nauk SSSR, 1986, **289,** *1*, pp. 38–41 (in Russian).
3. *Kurzhanskii, A. B., Filippova, T. F.*, On the description of the set of viable trajectories of a control system, Different. Uravn., 1987, **23,** *8*, pp. 1303–1315 (in Russian).
4. *Kurzhanskii, A. B.*, Control and observation under uncertainty. Moscow, Nauka, 1977 (in Russian).
5. *Kurzhanskii, A. B.*, Dynamic control system estimation under uncertainty conditions, I. Problems of Control and Information Theory, 1980, **9,** *6*, pp. 395–406; II. Problems of Control and Information Theory, 1981, **10,** *1*, pp. 33–42.
6. *Aubin, J.-P., Cellina, A.*, Differential inclusions, Heidelberg, Springer-Verlag, 1984.
7. *Aubin, J.-P., Ekeland, I.*, Applied nonlinear analysis, New York, Academic Press, 1984.
8. *Panasyuk, A. I., Panasyuk, V. I.*, Asymptotic magistral optimization of control systems. Minsk, Nauka i Tekhnika, 1986 (in Russian).
9. *Tolstonogov, A. A.*, Differential inclusions in Banach space. Novosibirsk, Nauka, 1986 (in Russian).

10. *Blagodatskikh, V. I., Filippov, A. F.*, Differential inclusions and optimal control, Trudy Matem. Inst. Akad. Nauk SSSR, **169,** Moscow, Nauka, 1985 (in Russian)
11. *Demyanov, V. F., Lemaréchal, C., Zowe, J.*, Approximation to a setvalued mapping, I: a proposal, Appl. Math. Optim., 1986, **14,** *3*, pp. 203–214.

Динамика пучка выживающих траекторий дифференциального включения: эволюционное уравнение

А. Б. КУРЖАНСКИЙ, Т. Ф. ФИЛИППОВА

(Свердловск)

Работа посвящена задачам описания ансамбля траекторий дифференциального включения, сохраняющихся (выживающих) в пределах заданного множества в фазовом пространстве. Известно, что трубку выживающих траекторий можно описать как решение нового дифференциального включения, правая часть которого определяется при помощи контингентного конуса к фазовому ограничению. В данной работе развивается иной подход к рассматриваемой задаче, позволяющий избежать операции построения касательных конусов. Установлено, что сечения пучков выживающих траекторий (множества достижимости включения при ограничениях на пространственные переменные) удовлетворяют эволюционному уравнению специального типа.

А. Б. Куржанский, Т.Ф. Филиппова
Институт математики и механики
Уральского отделения АН СССР
620219, Свердловск, ГСП-384, ул. С. Ковалевской, 16

Problems of Control and Information Theory, Vol. 17 (3), pp. 145—158 (1988)

ESTIMATION OF STATES AND PARAMETERS OF STOCHASTIC NONLINEAR SYSTEMS WITH MEASUREMENTS CORRUPTED BY CORRELATED NOISE

J. R. RAOL, N. K. SINHA

(*Bangalore*) (*Hamilton*)

(Received January 2, 1987)

We consider the problem of estimation of states and parameters of stochastic nonlinear systems described by difference equations when the measurements are corrupted by correlated noise. The solutions to the filtering problem are provided for certain general as well as a few special cases.

1. Introduction

In many practical systems the measurements are corrupted by coloured noise. Occasionally some of the measurements would be very accurate and hence they can be considered as perfect observations. These two cases represent singular problems in as much as the correlation matrix of the measurements noise can be shown to be singular [1].

We address the problem of state and parameter estimation for discrete-time systems when the noise in measurements is described by a stochastic difference equation. The noise can enter the measurements linearly or nonlinearly. We obtain quite general results based on Pugachev's nonlinear filtering theory [2–7, 9] and its ramifications [10, 11].

Pugachev's approach is based on the criterion of the minimization of the mean-square error [MSE] in a class of measurements that satisfy stochastic differential or difference equations. The theory is based on an arbitrary assignment of the so-called structural functions of the filter. The estimator has variable coefficients as gains which are determined by processing all the a priori information on filtered as well as measured processes.

In [10, 11] the nonlinear filtering problem for stochastic differential systems with measurements containing correlated noise was solved. We obtain similar results for systems described by difference equations.

It must be noted here that linear filtering problems for singular noise in continuous and discrete systems have been solved in [12–16, 26].

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

2. Statement of the problem

Let us consider the stochastic system represented by the difference equation

$$Y_{j+1} = g(Y_j, j) + p(Y_j, j)V_j \tag{1}$$

and the discrete measurements by the equation

$$W_j = l(Y_j, U_j, j). \tag{2}$$

The measurements noise U is considered to be correlated and is given by the difference equation

$$U_{j+1} = b(U_j, j) + d(U_j, j)V_j. \tag{3}$$

In these equations we have

Y_j — $n \times 1$ stochastic state process
W_j — $m \times 1$ measurement process
U_j — $m \times 1$ measurement noise process
g, l, b — known nonlinear vector-valued functions
p, d — known nonlinear matrix valued functions
V_j — vector-valued sequence of independent random numbers with zero-mean and covariance matrix Q_j.

Given Eqs (1) to (3) and the observations at instants $j = 1, 2, \ldots, N$, it is required to obtain the optimal estimate $\hat{Y}_j$ of the process Y_j in the class of functionals of the measurement process W_j.

We see from Eqs (1) and (3) that the two processes Y and U can be jointly represented as:

$$X_{j+1} = f(X_j, j) + q(X_j, j)V_j \tag{4}$$

$$W_j = h(X_j, j) \tag{5}$$

wherein

$$\left.\begin{aligned} X_j &= [Y_j^T, U_j^T]^T \\ f(X_j, j) &= [g(Y_j, j)^T, b(U_j, j)^T]^T \\ q(X_j, j) &= [p(Y_j, j)^T, d(U_j, j)^T]^T \end{aligned}\right\} \tag{6}$$

and

$$h(X_j, j) = l(Y_j, U_j, j).$$

Here f, q and h are known nonlinear functions of appropriate dimensions. The nonlinear functions are such that the expectations, to be defined later on, exist.

Thus the problem of Eqs (1) to (3) reduces to that of Eqs (4) to (6). From Eq.(5) it can be seen that the measurement process depends on a process that is not available. Also it does not contain an independent noise process V_j explicitly.

The filtering problem posed by Eqs (4) to (6) can be considered as a special case of a more general problem [10, 11] described by the following Eqs:

$$X_{j+1} = f(X_j, Z_j, j) + q(X_j, Z_j, j)V_j \tag{7}$$

$$Z_{j+1} = h(X_j, Z_j, j) \tag{8}$$

The problem is then to obtain an estimator $\hat{X}$ for the state X utilizing the known samples of process Z. We provide the solution to this problem in the next section.

3. Solution of the problem

A. *General Solution*

Let the estimator for the system of Eqs (7) and (8) be given by the difference equation

$$\hat{X}_{j+1} = K\zeta(Z_j, Z_{j+1}, \hat{X}_j, j) + \gamma \tag{9}$$

where

K, γ — optimal (j-dependent) filter gains

ζ — preassignable vector-valued nonlinear function of appropriate dimension.

The function ζ determines the structure of the nonlinear estimator (NE). Alternatively it is possible to introduce the concept of permissible estimates in order to make the comparison of accuracy and performance of the filters having similar structures feasible [7, 17].

The time-dependent optimal gains K and γ are determined from the condition of minimum of MSE

$$E\{X_j - \hat{X}_j)^T (X_j - \hat{X}_j)\}$$

and are given by (see the Appendix):

$$KM = L \tag{10}$$

and

$$\gamma = E_0 - KE_1$$

where

$$\left.\begin{aligned} L &= E\{(X_{j+1} - EX_{j+1})\zeta(Z_j, Z_{j+1}, \hat{X}_j, j)^T \\ M &= E\{(\zeta(Z_j, Z_{j+1}, \hat{X}_j, j) - E_1)\zeta(Z_j, Z_{j+1}, \hat{X}_j, j)^T \\ E_0 &= E\{X_{j+1}\};\ E_1 = E\{\zeta(Z_j, Z_{j+1}, \hat{X}_j j)\} \end{aligned}\right\}. \tag{11}$$

Here 'E' stands for the mathematical expectation and 'T' for matrix/vector transposition.

In order to evaluate the indicated expectations it is sufficient to know the joint characteristics function of the processes X_j, Z_j and $\hat{X}_j$, which is given as:

$$\Phi_j(\lambda_1, v, \mu) = E \exp\{i\lambda_1^T[f(X_{j-1}, Z_{j-1}, j-1) + q(X_{j-1}, Z_{j-1}, j-1)V_{j-1}] + \\ + iv^T h(X_{j-1}, Z_{j-1}, j-1) + i\mu^T[K_{j-1}\zeta(Z_{j-1}, Z_j, \hat{X}_{j-1}, j-1) + Y_{j-1}]\}. \quad (12)$$

The evaluation of expectations in Eq. (11) by utilizing the characteristic function Eq. (12) and substitution of the optimal gains in (9) completely solves the general filtering problem as posed in Section 2.

The above results can be used to estimate all the components of the vector X or to obtain estimates of some of the components of X by making the preassignable structural function ζ dependent on the corresponding subset of the vector $\hat{X}$.

B. Correlated measurement noise

As mentioned earlier, the general solution can be utilized to obtain the solution of the filtering problem with singular measurement noise for $Z_{j+1} = W_j$.

Let the system be described by Eqs. (1) to (3). The estimator for the process Y is then given as

$$\hat{Y}_{j+1} = K\zeta(Z_j, W_j, Y_j, j) + \gamma \quad (13)$$

where ζ is independent of $\hat{U}$.

The optimal values of the gains K and γ can be obtained from Eqs (10) to (13) by establishing the following relations between functions (see Eqs (4) to (8)):

$$\left.\begin{aligned} \lambda = [\lambda_1^T \lambda_2^T]^T;\ q(X_j, Z_j, j) &= [p(Y_j, j)^T d(U_j, j)^T]^T \\ f(X_j, Z_j, j) &= [g(Y_j, j)^T, b(U_j, j)^T]^T \\ h(X_j, Z_j, j) &= l(Y_j, U_j, j) \\ \zeta(Z_j, Z_{j+1}, \hat{X}_j, j) &= \zeta(Z_j, W_j, \hat{Y}_j, j) \end{aligned}\right\}. \quad (14)$$

Consequently the intermediate gains are given by

$$\left.\begin{aligned} L &= E\{(Y_{j+1} - EY_{j+1})\zeta(Z_j, W_j, \hat{Y}_j, j)^T\} \\ M &= E\{(\zeta(Z_j, W_j, \hat{Y}_j, j) - E_1)\zeta(Z_j, W_j, \hat{Y}_j, j)^T\} \end{aligned}\right\}$$

and

$$E_0 = E\{Y_{j+1}\};\ E_1 = E\{\zeta(Z_j, W_j, \hat{Y}_j, j)\} \quad (15)$$

Using the relationships shown in Eq. (14) we can obtain from Eq. (12), the joint characteristic function of the processes Y_j, U_j, Z_j and $\hat{Y}_j$:

$$\Phi_j(\lambda_1, \lambda_2, v, \mu) = E \exp\{i\lambda_1^T[g(Y_{j-1}, j-1) + p(Y_{j-1}, j-1)V_{j-1}] \\ + i\lambda_2^T[b(U_{j-1}, j-1) + d(U_{j-1}, j-1)V_{j-1}] + iv^T l(Y_{j-1}, U_{j-1}, j-1) \\ + i\mu^T[K_{j-1}\zeta(Z_{j-1}, l(Y_{j-1}, U_{j-1}, j-1), \hat{Y}_{j-1}, j-1) + Y_{j-1}]\}. \quad (16)$$

Thus Eq. (10), with the characteristic function given by Eq. (16) to evaluate the expectations in Eq. (15), completely solves the problem of discrete-time nonlinear filtering with measurements contaminated by coloured noise.

The issues related to selection of appropriate structural function ζ and computation of optimal gains are addressed in Section 7 and [7, 17, 18].

4. Special cases

In order to elucidate the results of the previous section, we consider certain special structural functions and derive filters for linear systems.

A. Linear Filter from General Solution

Let the linear system be described by

$$X_{j+1} = FX_j + BV_j \tag{17}$$

$$W_j = HX_j$$

where

F, B, H — known system matrices of appropriate dimensions
V_j — zero-mean WGN sequences with correlation matrix Q
X_j, W_j — state and measurement processes.

Let the structural function be selected as:

$$\zeta = [\hat{X}_j^T, W_j^T]^T.$$

Consequently we have from Eq. (9)

$$\hat{X}_{j+1} = K'[\hat{X}_j^T, W_j^T]^T + Y$$

$$= G\hat{X}_j + KW_j + Y \tag{18}$$

where $K' = [G|K]$.

The optimal values of the gains are obtained by using Eq. (17) and ζ in Eqs (10) and (11):

$$GP_{\hat{X}_j} + KHP_{\hat{X}_j} = FP_{\hat{X}_j}$$

$$GP_{\hat{X}_j}H^T + KHP_{\hat{X}_j}H^T = FP_{X_j}H^T$$

and solving the above equations we obtain

$$G = F - KH$$

$$K = FPH^T(HPH^T)^{-1} \tag{19}$$

where P_X, $P_{\hat{X}}$ are the variance-covariance matrices of vector X and $\hat{X}$ and $P = P_X - P_{\hat{X}}$.

The vector γ is given by

$$\gamma = FE\{X_j\} - GE\{\hat{X}_j\} - KHE\{X_j\}.$$

Substituting the value of G from Eq. (19) and noting that $E\{X_j\} = E\{\hat{X}_j\}$ we see that $\gamma = 0$.

Substituting Eq. (19) in (18) we obtain

$$\hat{X}_{j+1} = F\hat{X}_j + K[W_j - H\hat{X}_j]. \tag{20}$$

The covariance matrix equation for the state error is easily obtained and is given by

$$\begin{aligned} P_{j+1} &= FP_jF^T - KHP_jF^T - FP_jH^TK^T + KHP_jH^TK^T + BQB^T \\ &= (F - KH)P_j(F - KH)^T + BQB^T. \end{aligned}$$

After substituting Eq. (19) in the above equation and simplifying we obtain

$$P_{j+1} = (F - KH)P_jF^T + BQB^T. \tag{21}$$

Equations (19) to (21) describe the linear filtering algorithm.

B. Correlated Noise Case

Let the linear system be described by

$$Y_{j+1} = FY_j + BV_j \tag{22}$$

$$W_j = HY_j + DU_j \tag{23}$$

$$U_{j+1} = AU_j + SV_j \tag{24}$$

where F, B, H, D, A and S are known matrices of appropriate dimensions. The correlated measurement noise U_j is described by difference Eq. (24).

We choose the structural function ζ as

$$\zeta(Z_j, W_j, \hat{Y}_j, j) = [\hat{Y}_j^T, W_j^T]^T \tag{25}$$

and consequently the estimator, Eq. (13), takes the form

$$\hat{Y}_{j+1} = G\hat{Y}_j + KW_j + \gamma. \tag{26}$$

The intermediate gains L and M can be obtained by using Eqs (22) to (24) in Eqs (10) and (11):

$$L = [FP_{Y_j\hat{Y}_j} | FP_{Y_j}H^T + FP_{Y_jU_j}D^T] \tag{27}$$

and

$$M = \begin{bmatrix} P_{\hat{Y}_j} & P_{\hat{Y}_j}H^T + P_{\hat{Y}_jU_j}D^T \\ HP_{\hat{Y}_j} + DP_{\hat{Y}_jU_j} & HP_{\hat{Y}_j}H^T + DP_{U_j}D^T + HP_{Y_jU_j}D^T + DP_{U_jY_j}H^T \end{bmatrix}.$$

Then optimal gains G and K can be obtained from the following equations:

$$(F-G-KH)P_{\hat{Y}_j}-KDP_{\hat{Y}_jU_j}=0$$

$$(F-KH)(P_{Y_j}H^T+P_{Y_jU_j}D^T)-G(P_{\hat{Y}_j}H^T+P_{\hat{Y}_jU_j}D^T)-KD(P_{U_jY_j}H^T+P_{U_j}D^T)=0$$

$$\gamma=(F-G-KH)E\{Y_j\}-KDE\{U_j\}. \tag{28}$$

In order to obtain the variance-covariance matrices in Eq. (28), we utilize Eqs (22), (24) and (26) to get

$$\begin{bmatrix} Y_{j+1} \\ U_{j+1} \\ \hat{Y}_{j+1} \end{bmatrix} = \begin{bmatrix} F & 0 & 0 \\ 0 & A & 0 \\ KH & KD & G \end{bmatrix} \begin{bmatrix} Y_j \\ U_j \\ \hat{Y}_j \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \gamma \end{bmatrix} + \begin{bmatrix} B \\ S \\ 0 \end{bmatrix} V_j. \tag{29}$$

From the above Eq. (29) we obtain the following equation for the covariance matrix $\mathbf{P}_j$ for the random vector $[Y_j^T U_j^T \hat{Y}_j^T]^T$:

$$\mathbf{P}_{j+1}=\mathbf{F}\mathbf{P}_j\mathbf{F}^T+\mathbf{B}Q_j\mathbf{B}^T \tag{30}$$

where $\mathbf{F}$ is the block matrix in Eq. (29), $\mathbf{B}=[B^TS^T0^T]^T$ and 0^T is a null matrix with appropriate dimensions.

Equations (26),(28) and (30) provide the solution to the filtering problem with correlated measurement noise. It may be noted that this solution does not need differencing of the measurement data as required by certain existing linear filtering results [1, 10, 11].

5. Joint state and parameter estimation

For simplicity let the scalar system be described by

$$X_{j+1}=-\vartheta X_j+b_1V_1 \tag{31}$$

$$W_j=X_j+U_j \tag{32}$$

$$U_{j+1}=-aU_j+b_2V_2 \tag{33}$$

where

ϑ — an unknown parameter

b_1, b_2, a — known constants.

We represent the unknown parameter ϑ as a stochastic process.

$$\Theta_{j+1}=\Theta_j. \tag{34}$$

From Eqs (31) and (34) we obtain

$$\begin{bmatrix} X_{j+1} \\ \Theta_{j+1} \end{bmatrix} = \begin{bmatrix} -\Theta_j X_j \\ \Theta_j \end{bmatrix} + \begin{bmatrix} b_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}. \tag{35}$$

Also from Eq. (33) we obtain

$$U_{j+1} = -aU_j + [0 \quad b_2] \begin{bmatrix} V_1 \\ V_2 \end{bmatrix}. \tag{36}$$

We specify the optimal estimator as:

$$[\hat{X}_{j+1} \hat{\Theta}_{j+1}]^T = K\zeta(\hat{X}_j, \hat{\Theta}_j, W_j, j) + \gamma. \tag{37}$$

We then select K, ζ and γ as

$$K = \begin{bmatrix} k_{11} & k_{12} & \beta_1 \\ k_{21} & k_{22} & \beta_2 \end{bmatrix}; \quad \zeta = [\hat{X}_j, \hat{\Theta}_j, W_j]^T$$

and

$$\gamma = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}. \tag{38}$$

The intermediate gains are obtained as (dropping j for simplicity):

$$L_j = \begin{bmatrix} -P_{\Theta X \hat{X}} & -P_{\Theta X \hat{\Theta}} & -P_{\Theta X W} \\ P_{\Theta \hat{X}} & P_{\Theta \hat{\Theta}} & P_{\Theta W} \end{bmatrix}_j$$

$$M_j = \begin{bmatrix} P_{\hat{X}} & P_{\hat{X}\hat{\Theta}} & P_{\hat{X}W} \\ P_{\hat{\Theta}\hat{X}} & P_{\Theta} & P_{\hat{\Theta}W} \\ P_{W\hat{X}} & P_{W\hat{\Theta}} & P_W \end{bmatrix}_j$$

and

$$K = L_j M_j^{-1} \tag{41}$$

$$E_0 = \begin{bmatrix} -E\{\Theta_j X_j\} \\ E\{\Theta_j\} \end{bmatrix}$$

$$E_1 = [E\{\hat{X}_j\} E\{\hat{\Theta}_j\} E\{X_j + U_j\}]^T. \tag{42}$$

The entries in Eqs (39) and (40) are the variance and covariance matrices of the random variables X_j, $\hat{X}_j$, Θ_j, $\hat{\Theta}_j$, W_j and $\Theta_j X_j$. In order to simplify these equations we can use the following relations

$$P_{X_j W_j} = P_{\hat{X}_j X_j} + P_{\hat{X}_j U_j}$$

$$P_{\Theta_j W_j} = P_{\hat{\Theta}_j X_j} + P_{\hat{\Theta}_j U_j} \tag{43}$$

$$P_{W_j} = P_{X_j} + P_{U_j}$$

$$P_{X_j \hat{X}_j} = P_{\hat{X}_j}; \; P_{X_j W_j} = P_{\hat{X}_j W_j} \tag{44}$$

and

$$P_{X_j\hat{\Theta}_j}=P_{\hat{X}_j\hat{\Theta}_j}.$$

The relationships in Eq. (43) follow from Eq. (32) and those in Eq. (44) follow from $P_{X\zeta}=P_{\hat{X}\zeta}$ (see (A5)).

The joint one-dimensional characteristic function of the processes X_j, Θ_j, U_j, $\hat{X}_j$ and $\hat{\Theta}_j$ is expressed as:

$$\begin{aligned}\Phi_j(\lambda_1,\lambda_2,v,\mu_1,\mu_2)=\psi_{j-1}(\lambda)E\exp\{&i\lambda_1(-\Theta_{j-1}X_{j-1})+i(\mu_1\beta_1+\mu_2\beta_2)\hat{X}_{j-1}\\&+i(\mu_1\beta_1+\mu_2\beta_2-va)+i(\mu_1k_{11}+\mu_2k_{21})\hat{X}_{j-1}\\&+i(\mu_1k_{12}+\mu_2k_{22})\hat{\Theta}_{j-1}+i\mu_1Y_1+i\mu_2Y_2\}\end{aligned}\tag{45}$$

where $\psi_{j-1}(\lambda)$ is the characteristic function of the random process

$$V_{j-1}=[b_1V_1|b_2V_2]^T \quad \text{and} \quad \lambda^T=[\lambda_1 v].$$

We see from Eq. (45) that the characteristic function $\Phi_j(\lambda_1,\lambda_2,v,\mu_1,\mu_2)$ is given in terms of the random variables X_{j-1}, Θ_{j-1}, U_{j-1}, $\hat{X}_{j-1}$, $\hat{\Theta}_{j-1}$ and the probability distribution of the variable V_{j-1}. This latter distribution is assumed to be known. Hence Eqs (39) to (42) and (45) provide recursive procedure for determining Φ_j, K_j and γ_j at each step with the knowledge of Φ_{j-1}, K_{j-1} and γ_{j-1}. The choice of initial conditions is thoroughly discussed in [9].

6. Some generalizations

In this section we obtain the expressions for the optimal gains for the estimators described by [6, 10, 11]:

$$\hat{X}_{j+s}=\delta_j\zeta_j(Z_j,Z_{j+1},\hat{X}_j,\ldots,\hat{X}_{j+s-1})+\gamma_j\tag{46}$$

for the system described by Eqs (7) and (8).

Here s — the order which is specified a priori

ζ_j — preassignable nonlinear structural function

δ_j, γ_j — optimal gains.

These gains are given by

$$\begin{aligned}\delta_jM_j&=L_j\\ \gamma_j&=E_{0j}-\delta_jE_{1j}\end{aligned}\tag{47}$$

where

$$\begin{aligned}L_j&=E\{(X_{j+s}-EX_{j+s})\zeta_j(Z_j,Z_{j+1},\hat{X}_j,\ldots,\hat{X}_{j+s-1})^T\}\\ M_j&=E\{(\zeta_j(Z_j,Z_{j+1},\hat{X}_j,\ldots,\hat{X}_{j+s-1})-E_{1j})\zeta_j(Z_j,Z_{j+1},\hat{X}_j,\ldots,\hat{X}_{j+s-1})^T\}\\ E_{0i}&=E\{X_{j+s}\};\ E_{1j}=E\{\zeta_j(Z_j,Z_{j+1},\hat{X}_j,\ldots,\hat{X}_{j+s-1})\}.\end{aligned}\tag{48}$$

The one-dimensional joint characteristic function of the processes X_j, Z_j, $\hat{X}_j$, ..., $\hat{X}_{j+s-1}$ is given by

$$\Phi_j(\lambda, v, \mu) = E \exp \{i\lambda^T [f(X_{j-1}, Z_{j-1}, j-1) + q(X_{j-1}, Z_{j-1}, j-1) V_{j-1}]$$
$$+ iv^T h(X_{j-1}, Z_{j-1}, j-1) + i \sum_{r=1}^{s-1} {}^{jT}_{r} \hat{X}_{j+r-1}$$
$$+ i\mu_s^T [\delta_{j-1} \zeta_{j-1}(Z_{j-1}, Z_j, \hat{X}_{j-1}, \ldots, \hat{X}_{j-s2}) + \gamma_{j-1}]\}. \tag{49}$$

In the above equation $\mu_1, \mu_2, \ldots \mu_s$ are the columns of the matrix μ. We note here, that since V_{j-1} is independent of other random processes, we have recursive method for solving Eq. (49) and obtaining the optimal gains from Eqs (47) and (48) based on the knowledge of the a priori distribution of the variable V_{j-1} and the characteristic function $\Phi_{j-1}(\lambda, v, \mu)$ of the random processes X_{j-1}, Z_{j-1}, $\hat{X}_{j-1}$, ..., $\hat{X}_{j-s-2}$. The estimates $\hat{X}_1, \ldots, \hat{X}_s$ of the variables $X_1, \ldots, X_s$ should be selected arbitrarily [6]. Also the joint characteristic function $\Phi_1(\lambda, v, \mu)$ of the variables X_1, Z_1, $\hat{X}_1$, ..., $\hat{X}_s$ be selected such that proper solution of the above filtering problem is obtained without any ambiguity [6].

We have seen that when the measurements contain correlated noise, the system is described by the set of equations (1) to (3). We can now obtain the solution to this filtering problem for any value of 's' from the above results. In this case the estimator is given by

$$\hat{Y}_{j+s} = \delta_j \zeta_j(Z_j, W_j, \hat{Y}_j, \ldots, \hat{Y}_{j+s-1}) + \gamma_j. \tag{50}$$

The optimal values of the variable gains are obtained from Eqs (47) to (49) by the following replacements:

$$X_j \rightarrow Y_j; \; h(X_j, Z_j, j) \rightarrow l(Y_j, U_j, j)$$

and

$$\zeta(Z_j, Z_{j+1}, \hat{X}_j, \ldots, \hat{X}_{j+s-1}) \rightarrow \zeta(Z_j, W_j, \hat{Y}_j, \ldots, \hat{Y}_{j+s-1}).$$

The required expectations can be evaluated by knowing the joint one-dimensional characteristic function of the random processes $Y_j, U_j, Z_j, \hat{Y}_j, \ldots, \hat{Y}_{j+s-1}$, which is given by

$$\Phi_j(\lambda_1, \lambda_2, v, \mu) = E \exp \{i\lambda_1^T [g(Y_{j-1}, j-1) + p(Y_{j-1}, j-1) V_{j-1}]$$
$$+ i\lambda_2^T [b(U_{j-1}, j-1) + d(U_{j-1}, j-1) V_{j-1}]$$
$$+ iv^T l(Y_{j-1}, U_{j-1}, j-1) + i \sum_{r=1}^{s-1} \mu_r^T \hat{Y}_{j+r-1} \tag{51}$$
$$+ i\mu_s^T [\delta_{j-1} \zeta_{j-1}(Z_{j-1}, l(Y_{j-1}, U_{j-1}, j-1), \hat{Y}_{j-1}, \ldots, \hat{Y}_{j+s+2}) + \gamma_{j-1}]\}.$$

As per the remarks given after Eq. (49) the recursive procedure is established to solve Eqs (47), (48) and (51), thereby yielding the optimal filter gains δ_j and y_j.

7. Computational aspects

From the results of the previous sections it is apparent that two major issues are involved in the design of the nonlinear estimators: 1) the proper selection of the structural function ζ, and 2) computations of the optimal gains.

Due to the nature of the estimator equations we obtain conditionally optimal filters and hence there is an additional degree of freedom for selection of the structural functions. As yet no systematic procedure seems to have been established. However based on the studies in [19] some guidelines are available. In most cases the structural function is selected such that it is very similar to the system under consideration.

The problem of selection of structural functions has not been fully resolved. For further details, one may refer to [27].

In [6, 7, 9, 17, 18] certain special functions were selected in order to obtain some conventional and simple filtering structures for related problems. Such studies may also be extended for the results of this paper. Based on such choices it may be possible to establish interconnections with other familiar results [14, 16, 26]. It must be noted here that, the results of the present paper are general in nature, though linear filtering solutions have been obtained as special cases. A recent study [20, 21] may be useful in selecting the structural functions for the estimators proposed in this paper and, in general, for Pugachev's nonlinear estimators [6, 7, 10, 11]. Further research is required in this direction.

The second aspect is the evaluation of the expectations for computing the optimal gains of the filters. These expectations are functionals of the characteristic function Φ_j. These equations, in general, represent complex functional difference equations. Some approximation methods can be used for this purpose. Such methods are discussed in [9, 17, 18, 22–25].

Once a proper choice of the structural function ζ is made, the optimal gains K, G, γ (or δ) can be obtained by any of the methods described in [9, 19, 23] utilizing only a priori information on the various processes involved. Thus it is possible to design an optimal filter completely before collecting actual measurement data from the system. The subsequent computation of the estimates involves only matrix-vector operations making the actual estimation processor very efficient. The precomputed gains can be stored and fectched from microcomputer's memory when the observation data become available for processing. The estimator retains sequential nature. The results of this paper can be used for wide variety of problems: identification, joint state and parameter estimation, stochastic model building and control [5, 9].

7. Conclusions

The conditionally optimal solutions for nonlinear filtering problem described by stochastic difference equations with measurements containing correlated noise have been obtained. We have obtained filters for some general and special linear cases. The filter for joint state and parameter estimation for correlated noise case has been derived for a scalar case. The advantages of the present method are: the measurement model can be linear or nonlinear in the correlated noise and the filter's conditionally optimal gains can be precomputed before actually processing the measurement data. This makes estimation procedure very efficient and simple while incorporating the observations sequentially. Some computational aspects have been discussed.

References

1. *Maybeck, P. S.* (1982), Stochastic Models, Estimation and Control. Academic Press, New York.
2. *Pugachev, V. S.* (1978a), Estimation of variables and parameters in stochastic systems described by differential equations. Soviet Math. Dokl., **19,** 967.
3. *Pugachev, V. S.* (1978b), Recursive estimation of variables and parameters in stochastic systems described by autoregression equations. Soviet Math. Dokl., **19,** 991.
4. *Pugachev, V. S.* (1978c), Recursive estimation of variables and parameters in stochastic systems described by difference equations. Soviet Math. Dokl., **19,** 1495.
5. *Pugachev, V. S.* (1979a), Estimation of state and parameters of continuous nonlinear systems. Autom. and Remote Control, **40,** 831.
6. *Pugachev, V. S.* (1979b), Estimation of variables and parameters in discrete-time nonlinear systems. Autom. and Remote Control, **40,** 512.
7. *Pugachev, V. S.* (1982), Conditionally optimal estimation in stochastic differential systems. Automatica, **18,** 685.
8. *Pugachev, V. S.* (1965), Theory of Random Functions and its Application to Control Problems. Pergamon Press, New York.
9. *Pugachev, V. S.* (1984), Probability Theory and Mathematical Statistics for Engineers. Pergamon Press, New York.
10. *Silujanova, I. D.* (1980), Estimating the variables and parameters in stochastic systems described by differential equations in a particular case. Soviet Math. Dokl., **21,** 152.
11. *Silujanova, I. D.* (1980), State and parameter estimation of nonlinear systems using observations containing colored noise. Autom. and Remote Control, **41,** 1380.
12. *Bryson, A. E., Johansen, D. E.* (1965), Linear filtering for time-varying systems using measurements containing colored noise. IEEE Trans. Automat. Control, **AC-10,** No. *1*, 4.
13. *Bucy, R. S.* (1967), Optimal filtering for correlated noise. J. of Math. Anal. and App. **20,** 1.
14. *Bryson, A. E., Henrikson, L. J.* (1968), Estimation using sampled data containing sequentially correlated noise. J. Spacecraft and Rockets, **6***(6)*, 662.
15. *Stear, E. B., Stubberud, A. R.* (1968), Optimal filtering for Gauss–Markov noise. Int. J. Control. **8***(2)*, 123.
16. *Deyst, J. J.* (1969), A derivation of the optimum continuous linear estimator for systems with correlated measurement noise. AIAA Journal, **7***(11)*, 2116.
17. *Raol, J. R., Sinha, N. K.* (1986), On Pugachev's filtering theory for stochastic nonlinear systems, presented at the Second IFAC Symposium on Stochastic Control (Vilnius, U.S.S.R.).
18. *Raol, J. R., Sinha, N. K.* (1986), A nonlinear filter for estimation of states of a continuous-time system with discrete measurements, presented at the Second IFAC Symposium on Stochastic Control (Vilnius, U.S.S.R.).

19. *Kazakov, I. E., Malchikov, S. V.* (1981), Approximate design of Pugachev filters of specified complexity. Autom. and Remote Control, **42,** 1309.
20. *Ahlbehrendt, N., Madiger, B.* (1981), A new method for synthesis of nonlinear parameter and state estimation for noisely disturbed processes. Proc. of the VIIIth IFAC Congress, Kyoto, Japan, V, 99.
21. *Ahlbehrendt, N., Draeger, U., Madiger, B.* (1985), Synthesis of estimation algorithms for stochastic disturbed systems by the help of Lyapunov functions. Problems of Control and Information Theory, **14**(*1*), 47.
22. *Dashevskii, M. L., Shin, V. I.* (1981), Design of conditionally optimal filters in continuous dynamic systems. Autom. and Remote Control, **42,** 1309.
23. *Demukh, V. I.* (1965), An approximate method for analysing accuracy of nonlinear systems. Autom. and Remote Control, **26,** 1012.
24. *Dashevskii, M. L., Liptser, R. Sh.* (1967), Use of conditional semiinvariants in nonlinear filtering of Markov processes. Autom. and Remote Control, **28,** 63.
25. *Malchikov, S. V.* (1970), Approximate method of determining the phase coordinate distributions of nonlinear automatic systems. Autom. and Remote Control, **31,** 717.
26. *Sinha, N. K., Tom, A.* (1978), Adaptive state estimation for systems with colored observation noise. Int. J. Syst. Sc., **9,** 31.
27. *Raol, J. R., Sinha, N. K.* (1987), Conditionally optimal state estimation for systems governed by difference equations. Can. Elec. Eng. J., **12,** 71.

Appendix

For stochastic difference system described by Eqs (7) and (8), we have the following difference equation for the estimator:

$$\hat{X}_{j+1} = K_j \zeta_j(Z_j, Z_{j+1}, \hat{X}_j, j) + \gamma_j, \tag{A1}$$

Then using the mean-square regression theory [6, 8, 9], we obtain

$$E\{[X_{j+1} - E\{X_{j+1}\}]\zeta(Z_j, Z_{j+1}, \hat{X}_j, j)^T\} = K_j E\{[\zeta(Z_j, Z_{j+1}, \hat{X}_j, j) - \\ - E\{\zeta(Z_j, Z_{j+1}, \hat{X}_j, j)\}]\zeta(Z_j, Z_{j+1}, \hat{X}_j, j)^T\} \tag{A2}$$

and

$$\gamma_j = E\{X_{j+1}\} - KE\{\zeta(Z_j, Z_{j+1}, \hat{X}_j, j)\}. \tag{A3}$$

Defining L and M as in (11) we obtain for optimal gain the following relation

$$K_j = L_j M^{j-1}. \tag{A4}$$

Also we have the condition

$$E\{(\hat{X}_{j+1} - X_{j+1})\zeta_j(Z_j, Z_{j+1}, \hat{X}_j, j)^T\} = 0. \tag{A5}$$

The characteristic function Φ_j for the random processes X_j, Z_j and $\hat{X}_j$ can be obtained by substituting Eqs (7) to (9) in the following equation

$$\Phi_j(\lambda_1, v, \mu) = E \exp\{i\lambda_1^T X_j + iv^T Z_j + i\mu^T \hat{X}_j\}. \tag{A6}$$

When the estimator is specified by the following structure

$$\hat{X}_{j+1} = G\xi(Z_j, \hat{X}_j, j) + K\eta(Z_j, \hat{X}_j, j)Z_{j+1} + \gamma \tag{A7}$$

the optimal gains G, K and y can be obtained by representing the structural function as

$$\zeta(Z_j, Z_{j+1}, \hat{X}_j, j) = [\xi(Z_j, \hat{X}_j, j)^T | Z_{j+1}{}^T \eta(Z_j, \hat{X}_j, j)^T]^T$$

in (A2) and (A3).

Similarly the characteristic functional equation (12) can be modified. This representation is useful in obtaining various special cases as discussed in the text.

Оценка состояний и параметров стохастических нелинейных систем при измерениях с коррелированным мешающим шумом

Й. Р. РАОЛ, Н. К. СИНХА

(Бангалоре) (Хамилтон)

Рассматривается проблема оценки состояний и параметров стохастических нелинейных систем, описываемых разностными уравнениями, когда при измерениях мешающий шум является коррелированным. Даются рещения проблемы фильтрации для некоторых общих и частных случаев.

J. R. Raol
Systems Engineering Division
National Aeronautical Laboratory
Bangalore — 560017
India

N. K. Sinha
Department of Electrical and Computer Engineering
McMaster University
Hamilton
Canada
L8S 4L7

Problems of Control and Information Theory, Vol. 17 (3), pp. 159—170 (1988)

THE NETS OF CONJUGATED SHIFT REGISTERS

Z. GRODZKI, M. ŁATKO

(Lublin)

(Received April 10, 1987)

The nets of right-hand side and left-hand side k-registers will be considered. The notion of similarity of such nets will be introduced. Two nets are said to be similar if there is a one-to-one mapping between their transition graphs such that corresponding to each other cycles consist of the same states. An algebraic method for the construction of the whole similarity class of nets will be adopted from [5].

0. Introduction

k-registers are technical arrangements which generate pseudoperiodic binary sequences with cycle length less than or equal to 2^k. With respect to easy technical realization the linear k-registers have been used most frequently. However, the theory of these registers is worked out in a satisfactory way for the practice [3] but from the mathematical point of view there are still open problems. It is possible to distinguish some subclass of linear k-registers generating of only pseudorandom sequences with "pure" period of length 2^k-1. These ones have been used in certain areas such as automatic control, criptology, radar, radiotechnics and many others [1, 3, 6].

In modern electronics k-registers are used rather as nets (sequential or parallel) than as singular ones. In coding theory, especially in the theory of error-correcting codes, the nets of k-registers are taken often as coders and decoders. The reader is referred to Blahut's monograph [2]. Additionally, the nets of k-registers are needed in the construction of integrated circuits.

A new class of nets of parallel k-registers will be considered here. Such nets consists of two k-registers, righthand side and left-hand side, with the same feedback function. The output sequences of a right-hand (left-hand) side k-register are the right-hand (left-hand) side binary infinite sequences. The properties of both kind of k-registers in terms of graph theory will be given here.

A necessary and sufficient condition for all connected components of the transition graphs of both k-registers to be a cycle will be given and proved. Only the nets with the above property will be considered in this paper.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

The notion of similarity between two k-registers will be introduced. Two arbitrary k-registers (right-hand side, left-hand side or mixed) are said to be similar if and only if there is a one-to-one correspondence between transition graphs of these k-registers such that the connected components corresponding to each other consist of the same states (with different ordering). Then the notion of similarity is extended to the nets of k-registers. An algebraic method is adopted from [5] for a construction of the whole class of similar nets of k-registers to the given one.

1. Preliminaries

Nonempty sets will be denoted always by upper case letters and their elements by lower case ones (with or without subscripts).

The positive integers will be denoted by lower case Latin letters i, j, k, m, p.

Let $B=\{0, 1\}$ and let B^k $(k>1)$ denote the k-th Cartesian product of B. The elements of B will be denoted by lower case Latin letters t, u, v, x, y (with subscripts) and of B^k — by lower case boldface Latin letters $\mathbf{t}, \mathbf{u}, \mathbf{v}, \mathbf{x}, \mathbf{y}$ (with or without subscripts).

The functions of mappings of B^k into B will be denoted by lower case script letters $\mathscr{f}, \mathscr{g}, \mathscr{h}$ and functions of mappings of B^k into B^k — by upper case italic letters $\mathscr{F}, \mathscr{H}$. In most cases symbols $\mathscr{f}, \mathscr{F}$ denoting the functions will be written with the subscripts r, l and s as follows: $\mathscr{f}^r, \mathscr{f}^l, \mathscr{f}^s, \mathscr{F}^r, \mathscr{F}^l$ and $\mathscr{F}^s$.

The symbol + denotes the addition modulo 2.

2. The right-hand side and left-hand side k-registers

A formal definition of right-hand side and left-hand side k-registers and their transition graph will be recalled from [3].

Given Boolean functions $\mathscr{f}^r$ and $\mathscr{f}^l$ of B^k into B, let us define new functions $\mathscr{F}^r$ and $\mathscr{F}^l$ of B^k into B^k as follows:

$$\mathscr{F}^r(t_1 \ldots t_k)=t_2 \ldots t_k \mathscr{f}^r(t_1 \ldots t_k), \tag{2.1}$$

$$\mathscr{F}^l(t_1 \ldots t_k)=\mathscr{f}^l(t_1 \ldots t_k)t_1 \ldots t_{k-1} \tag{2.2}$$

for all $t_1 \ldots t_k \in B^k$.

Then a pair $(B^k, \mathscr{F}^r)$ is said to be a right-hand side k-register and $(B^k, \mathscr{F}^l)$ — a left-hand side one.

Let $(B^k, \mathscr{F}^s)$ be a k-register.[1] B^k is said to be the set of states, $\mathscr{F}^s$ — the transition function and $\mathscr{f}^s$ — the feedback function of this k-register.

[1] Further $(B^k, \mathscr{F}^s)$ will denote a right-hand side k-register $(B^k, \mathscr{F}^r)$ or a left-hand side one $(B^k, \mathscr{F}^l)$.

A state $\mathbf{u} \in B^k$ is called a successor of a state $\mathbf{t} \in B^k$ in $(B^k, \mathscr{F}^s)$ if and only if there is $i \geqq 1$ such that $\mathbf{u} = (\mathscr{F}^s)^i(\mathbf{t})$, where $(\mathscr{F}^s)^i$ denotes the i-th iteration of $\mathscr{F}^s$.

Remark 2.1. With respect to the technical realization of k-registers, their transition function $\mathscr{F}^r$ and $\mathscr{F}^l$ by means of the feedback functions f^r and f^l have been defined.

However, it is possible to define immediately these functions as follows:

$$\text{if } \mathscr{F}^r(t_1 \ldots t_k) = u_i \ldots u_k \quad \text{then } u_i = t_{i+i} \quad \text{for all } i = 1, \ldots, k-1; \tag{2.3}$$

$$\text{if } \mathscr{F}^l(t_1 \ldots t_k) = u_i \ldots u_k \quad \text{then } u_{i+1} = t_i \quad \text{for all } i = 1, \ldots, k-1. \tag{2.4}$$

Conversely, each transition function $\mathscr{F}^s$ allows to define unambiguously a feedback function f^s.

Each k-register $(B^k, \mathscr{F}^s)$ determines the existence of a directed graph $\mathbf{F}_k^s$ (the transition graph of this k-register). The nodes of this graph are all the elements of B^k, and the edge goes from a state $\mathbf{t}$ to a state $\mathbf{u}$ if and only if $\mathbf{u} = \mathscr{F}^s(\mathbf{t})$. Conversely, each transition graph of a k-register determines its feedback function as well.

Example 2.1. Define the right-hand side and left-hand side 3-registers by means of their transition graphs (Fig. 2.1) as follows:

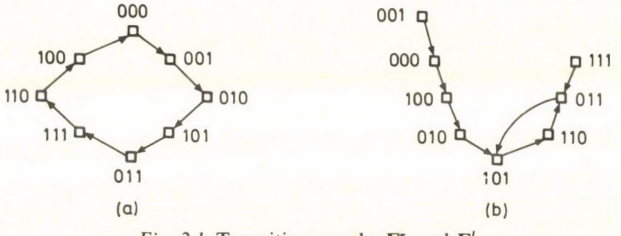


Fig. 2.1. Transition graphs $\mathbf{F}_3^r$ and $\mathbf{F}_3^l$

These graphs determine uniquely their feedback function

$$f^r(x_1x_2x_3) = f^l(x_1x_2x_3) = x_1 + x_2 + \bar{x}_2\bar{x}_3,$$

where $\bar{x}_i = x_i + 1$. ∎

Lemma 2.1. For an arbitrary right-hand side k-register $(B^k, \mathscr{F}^r)$ with feedback function f^r, the following conditions are equivalent:

$$\text{every connected component of transition graph } \mathbf{F}_k^r \text{ is a cycle;} \tag{2.5}$$

$$\mathscr{F}^r \text{ is a mapping of } B^k \text{ onto } B^k; \tag{2.6}$$

$$f^r(0t_2 \ldots t_k) \neq f^r(1t_2 \ldots t_k); \tag{2.7}$$

$$f^r(t_1 \ldots t_k) = t_1 + g(t_2 \ldots t_k), \tag{2.8}$$

where g is a function of B^{k-1} into B.

Proof has been given in [3] ▮

Lemma 2.2. For an arbitrary left-hand side k-register $(B^k, \mathscr{F}^l)$ with feedback function f^l, the following conditions are equivalent:

every connected component of transition graph $\mathbf{F}_k^l$ is a cycle; (2.9)

$$\mathscr{F}^l \text{ is a mapping of } B^k \text{ onto } B^k; \tag{2.10}$$

$$f^l(t_1 \dots t_{k-1}0) \neq f^l(t_1 \dots t_{k-1}1); \tag{2.11}$$

$$f^l(t_1 \dots t_k) = h(t_1 \dots t_{k-1}) + t_k, \tag{2.12}$$

where h is a function of B^{k-1} into B.

Proof is analogical to the proof of Lemma 2.1. and will be omitted here. ▮

Let us denote by $\mathscr{R}_k^r$ and $\mathscr{R}_k^l$ the sets of all right-hand side and left-hand side k-registers satisfying conditions (2.7) and (2.11), respectively. Only elements of $\mathscr{R}_k^r$ and $\mathscr{R}_k^l$ will be considered here.

It is easy to show that for each k-register $(B^k, \mathscr{F}^s)$ the transition function $\mathscr{F}^s$ determines some equivalence relation in B^k, called a shift relation. Every equivalence class of such a relation which is determined by any element $\mathbf{t} \in B^k$ covers of all successors of $\mathbf{t}$ in this k-register. A quotient set of this relation will be denoted by $B^k/\mathscr{F}^s$.

Now consider a k-register $(B^k, \mathscr{F}^s)$ and a set $A \subseteq B^{k-1}$.

If $s = r$ then we define a new function $f_A^r : B^k \to B$ as follows:

$$f_A^r(x_1 \dots x_k) = \begin{cases} f^r(x_1 x_2 \dots x_k) + 1 & \text{when} \quad x_2 \dots x_k \in A \\ f(x_1 x_2 \dots x_k) & \text{otherwise.} \end{cases} \tag{2.13}$$

If $s = 1$ then we define a new function $f_A^l : B^k \to B$ as follows:

$$f_A^l(x_1 \dots x_k) = \begin{cases} f^l(x_1 \dots x_{k-1} x_k) + 1 & \text{when} \quad x_1 \dots x_{k-1} \in A \\ f^l(x_1 \dots x_{k-1} x_k) & \text{otherwise.} \end{cases} \tag{2.14}$$

As for the functions f^r and f^l conditions (2.7) and (2.11) are satisfied, it follows immediately from (2.13) and (2.14) that the same conclusion holds for functions f_A^r and f_A^l. Then f_A^s is the feedback function of some k-register of class $\mathscr{R}_k^s$. Let $\mathscr{F}_A^s$ denote the transition function of this k-register.

On the other hand, for arbitrary k-registers $(B^k, \mathscr{F}^s)$ and $(B^k, \mathscr{H}^s)$ there is a unique set $A \subseteq B^{k-1}$ such that $\mathscr{F}^s = \mathscr{H}_A^s$ and $\mathscr{H}^s = \mathscr{F}_A^s$. Such a set will be denoted by $\mathrm{Dif}(\mathscr{F}^s, \mathscr{H}^s)$.

3. Similarity of the k-registers

A subclass of $\mathscr{R}_k^r \cup \mathscr{R}_k^l$ of all k-registers such that their quotient sets of shift relation are identical will be considered. Such registers will be called similar.

The main problem is to construct the whole similarity class which is determined by a given k-register $(B^k, \mathscr{F}^s)$. All similar k-registers are able to transform the same information. In this section several properties of the similarity relation will be given. The construction of whole similarity class to the given k-register in the next section will be given.

Two k-registers $(B^k, \mathscr{F}^{s^1})$ and $(B^k, \mathscr{H}^{s^2})$ $(s^1, s^2 \in \{r, 1\})$ are said to be similar $((B^k, \mathscr{F}^{s^1}) \approx (B^k, \mathscr{H}^{s^2}))$ if and only if $B^k/\mathscr{F}^{s^1} = B^k/\mathscr{H}^{s^2}$.

At the beginning we shall show that each similarity class consists of the same number of right-hand side and left-hand side k-registers.

For this purpose let us consider a k-register $(B^k, \mathscr{F}^s)$. Define a new function $\hat{f}^s: B^k \to B$ as follows

$$\hat{f}^s(x_1 \dots x_k) = \begin{cases} f^s(x_k x_1 \dots x_{k-1}) & \text{when} \quad s = r \\ f^s(x_2 \dots x_k x_1) & \text{when} \quad s = 1. \end{cases} \tag{3.1}$$

It follows from (2.8) that

$$\hat{f}^r(x_1 \dots x_k) = f^r(x_k x_1 \dots x_{k-1}) = x_k + g(x_1 \dots x_{k-1}),$$

then from (2.12) $\hat{f}^r$ is a feedback function of a left-hand side k-register. Analogously, if $s = 1$ then from (2.12) it follows that

$$\hat{f}^l(x_1 \dots x_k) = f^l(x_2 \dots x_k x_1) = x_1 + h(x_2 \dots x_k),$$

then from (2.8) $\hat{f}^l$ is a feedback function of a right-hand side k-register.

Two k-registers $(B^k, \mathscr{F}^s)$ and $(B^k, \hat{\mathscr{F}}^s)$ for which the feedback functions f^s and $\hat{f}^s$ are connected with each other by (3.1) are said to be dual.

Lemma 3.1. For arbitrary pairs of dual k-register $(B^k, \mathscr{F}^s)$, $(B^k, \hat{\mathscr{F}}^s)$ and $(B^k, \mathscr{H}^s)$ $(B^k, \hat{\mathscr{H}}^s)$ the following equality holds:

$$\mathrm{Dif}(\mathscr{F}^s, \mathscr{H}^s) = \mathrm{Dif}(\hat{\mathscr{F}}^s, \hat{\mathscr{H}}^s).$$

Proof immediately follows from the definition of dual k-registers. ▌

Theorem 3.1. Each two dual k-registers $(B^k, \mathscr{F}^s)$ and $(B^k, \hat{\mathscr{F}}^s)$ are similar.

Proof. Let $\mathbf{F}_k^s$ and $\hat{\mathbf{F}}_k^s$ be the transition graphs of $(B^k, \mathscr{F}^s)$ and $(B^k, \hat{\mathscr{F}}^s)$, respectively We show that if $(\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_p)$ is a cycle in $\mathbf{F}_k^s$ then $(\mathbf{t}_p, \dots, \mathbf{t}_2, \mathbf{t}_1)$ is a cycle in $\hat{\mathbf{F}}_k^s$, i.e. if $\mathscr{F}^s(\mathbf{t}_i) = \mathbf{t}_{i+1}$, then $\hat{\mathscr{F}}^s(\mathbf{t}_{i+1}) = \mathbf{t}_i$, for $1 \leqq i \leqq p$. Let $\mathbf{t}_i = t_1 \dots t_k$.

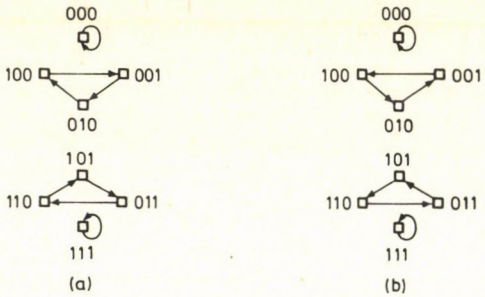


Fig. 3.1. Transition graphs $\mathbf{F}^r_3$ and $\hat{\mathbf{F}}^r_3$

For the right-hand side case we have:

$$\mathscr{F}^r(\mathbf{t}_i)=\mathscr{F}^r(t_1\ldots t_k)=t_2\ldots t_k f^r(t_1\ldots t_k)=\mathbf{t}_{i+1} \quad \text{and} \quad \hat{\mathscr{F}}^r(\mathbf{t}_{i+1})=$$

$$=\hat{\mathscr{F}}^r(t_2\ldots t_k f^r(t_1\ldots t_k))=\hat{f}^r(t_2\ldots t_k f^r(t_1\ldots t_k))t_2\ldots t_k=$$

$$f^r(f^r(t_1\ldots t_k)t_2\ldots t_k)t_2\ldots t_k=(f^r(t_1\ldots t_k)+g(t_2\ldots t_k))$$

$$t_2\ldots t_k=(t_1+g(t_2\ldots t_k)+g(t_2\ldots t_k))t_2\ldots t_k=t_1t_2\ldots t_k=\mathbf{t}_i.$$

For the left-hand side case we have:

$$\mathscr{F}^l(\mathbf{t}_i)=\mathscr{F}^l(t_1\ldots t_k)=f^l(t_1\ldots t_k)t_1\ldots t_{k-1}=\mathbf{t}_{i+1}$$

and

$$\hat{\mathscr{F}}^l(\mathbf{t}_{i+1})=\hat{\mathscr{F}}^l(f^l(t_1\ldots t_k)t_1\ldots t_{k-1})=t_1\ldots t_{k-1}\hat{f}^l(f^l(t_1\ldots$$

$$\ldots t_k)t_1\ldots t_{k-1})=t_1\ldots t_{k-1}f^l(t_1\ldots t_{k-1}f^l(t_1\ldots t_k))=t_1\ldots$$

$$\ldots t_{k-1}(f^l(t_1\ldots t_k)+h(t_1\ldots t_{k-1}))=t_1\ldots t_{k-1}(h(t_1\ldots t_{k-1})+$$

$$+h(t_1\ldots t_{k-1})+t_k)=t_1\ldots t_{k-1}t_k=\mathbf{t}_i. \blacksquare$$

Example 3.1. Let $(B^3, \mathscr{F}^r)$ be a right-hand side 3-register with feedback function $f^r(x_1x_2x_3)=x_1$. Then transition graph $\mathbf{F}^r_3$ has the cycles: $T_1=(000)$, $T_2=(001, 010, 100)$, $T_3=(011, 110, 101)$ and $T_4=(111)$ (Fig. 3.1a). The feedback function $\hat{f}^r$ of a dual 3-register $(B^3, \hat{\mathscr{F}}^r)$ has a form $\hat{f}^r(x_1x_2x_3)=f^r(x_3x_1x_2)=x_3$. The graph $\hat{\mathbf{F}}^r_3$ has the cycles: $T'_1=(000)$, $T'_2=(100, 010, 001)$, $T'_3=(101, 110, 011)$ and $T'_4=(111)$ (Fig. 3.1b). As cycles T_i and T'_i, for $i=1, 2, 3, 4$, contain the same nodes, then $(B^3, \mathscr{F}^r)$ and $(B^3, \hat{\mathscr{F}}^r)$ are similar. ▮

4. Construction of the similarity class

The similarity class of an arbitrary k-register consists of pairs of dual k-registers. From the considerations of the previous section it follows that it is sufficient to determine a subclass of all similar right-hand side or left-hand side k-registers.

Now we shall give a construction of the similarity subclass of right-hand side k-registers.

Given k-register $(B^k, \mathscr{F}^r)$ let us define the set

$$P_{\mathscr{F}^r} = \{\mathbf{t} \in B^{k-1}: \text{there is a cycle in } \mathbf{F}_k^r \text{ consisting of the nodes } 0\mathbf{t} \text{ and } 1\mathbf{t}\}.$$

Let us define the relation $I_{\mathscr{F}^r} \subseteq P_{\mathscr{F}^r} \times P_{\mathscr{F}^r}$ (interlacing relation) as follows:

$$(\mathbf{x}, \mathbf{y}) \in I_{\mathscr{F}^r} \text{ if and only if there is a cycle } T \text{ in } \mathbf{F}_k^r$$

$$\text{such that } T = (u\mathbf{x}, \ldots, v\mathbf{y}, \ldots, \bar{u}\mathbf{x}, \ldots, \bar{v}\mathbf{y}, \ldots),$$

where $\bar{u} = u + 1$, $\bar{v} = v + 1$ and $u, v \in \{0, 1\}$.

Let $\chi_{\mathscr{F}^r}$ denote the characteristic function of the relation $I_{\mathscr{F}^r}$, i.e.

$$\chi_{\mathscr{F}^r}(\mathbf{x}, \mathbf{y}) = \begin{cases} 1 & \text{when} \quad (\mathbf{x}, \mathbf{y}) \in I_{\mathscr{F}^r} \\ 0 & \text{otherwise.} \end{cases}$$

The range of the function $\chi_{\mathscr{F}^r}$ will be presented in the form of a square matrix $M_{\mathscr{F}^r}$. For any $A \subset P_{\mathscr{F}^r}$ the submatrix of $M_{\mathscr{F}^r}$ which is determined by the pairs $(\mathbf{u}, \mathbf{v}) \in A \times A$ will be denoted by $M_{\mathscr{F}^r|A}$.

Theorem 4.1. [5] For an arbitrary right-hand side k-register $(B^k, \mathscr{F}^r)$ and a set $A \subset B^{k-1}$

$$(B^k, \mathscr{F}^r) \approx (B^k, \mathscr{F}_A^r) \text{if and only if matrix } M_{\mathscr{F}^r|A} \text{ is nonsingular over } GF(2). \quad (4.1)$$

Example 4.1. For feedback function $f^r(x_1x_2x_3x_4) = x_1 + x_4$ of $(B^4, \mathscr{F}^r)$, the graph $\mathbf{F}_4^r$ consists the cycles (0000), (0001, 0011, 0111, 1111, 1110, 1101, 1010, 0101, 1011, 0110, 1100, 1001, 0010, 0100, 1000) and the set $P_{\mathscr{F}^r} = \{001, 010, 011, 100, 101, 110, 111\}$. The matrix $M_{\mathscr{F}^r}$ which is determined by the relation $I_{\mathscr{F}^r}$ takes the following form (rows and columns are lexicographically ordered in the set $P_{\mathscr{F}^r}$):

$$M_{\mathscr{F}^r} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

By determining all nonsingular over $GF(2)$ submatrices of $M_{\mathscr{F}^r}$ we obtain 15 sets $A_i \subset P_{\mathscr{F}^r}$, for example $A_1 = \{001, 010\}$, $A_2 = \{001, 010, 011, 100\}$, $A_3 = \{001, 010, 011, 100, 101, 110\}$, that for any A_i, $i = 1, \ldots 15$, 4-register $(B^4, \mathscr{F}_{A_i}^r)$is similar to $(B^4, \mathscr{F}^r)$. ▮

5. Properties of the conjugated k-registers

The connection between transition graphs of two conjugated k-registers will be investigated.

Two k-registers $(B^k, \mathscr{F}^r)$ and $(B^k, \mathscr{F}^l)$, for which the feedback functions f^r and f^l are identical, are called conjugated. A pair of conjugated k-registers will be denoted by $\langle B^k, (\mathscr{F}^r, \mathscr{F}^l)\rangle$.

Theorem 5.1. For arbitrary conjugated k-registers $(B^k, \mathscr{F}^r)$ and $(B^k, \mathscr{F}^l)$ $(k>2)$ with feedback function f the following conditions are equivalent:

all connected components of the transition graphs $\mathbf{F}_k^r$ and $\mathbf{F}_k^l$ are the cycles; (5.1)

$\mathscr{F}^r$ and $\mathscr{F}^l$ are the mapping of B^k onto B^k; (5.2)

the feedback function f has the form: (5.3)

$$f(x_1x_2 \ldots x_{k-1}x_k)=x_1+f^*(x_2 \ldots x_{k-1})+x_k$$

where f^* is a function of B^{k-2} into B.

Proof immediately follows from Lemmas 2.1 and 2.2. ∎

Now a connection between the sets $B^k/\mathscr{F}^r$ and $B^k/\mathscr{F}^l$ for arbitrary pair of conjugated k-registers $(B^k, \mathscr{F}^r)$ and $(B^k, \mathscr{F}^l)$ will be investigated. The sum $\sum$ will not be taken modulo 2.

Lemma 5.1. [3] If $f^r(x_1 \ldots x_k)=x_1+g(x_2 \ldots x_k)$ is a feedback function of a right-hand side k-register $(B^k, \mathscr{F}^r)$ $(k>2)$, then the parity of the number of cycles of $\mathbf{F}_k^r$ is the same as the parity of the number $\sum_{\mathbf{t}\in B^{k-1}} g(\mathbf{t})$.

Theorem 5.2. For any pair $\langle B^k, (\mathscr{F}^r, \mathscr{F}^l)\rangle$ of conjugated k-registers, for which $\mathscr{F}^r$ and $\mathscr{F}^l$ are the mappings of B^k onto B^k, the quotient sets $B^k/\mathscr{F}^r$ and $B^k/\mathscr{F}^l$ consist of the even number of elements.

Proof. Let $f(x_1 \ldots x_k)=x_1+f^*(x_2 \ldots x_{k-1})+x_k$ be the feedback function of k-registers mentioned above. For the right-hand side case we have $f(x_1 \ldots x_k)=x_1+$ $+(x_k+f^*(x_2 \ldots x_{k-1}))$ and then the number

$$\sum_{x_2 \ldots x_k \in B^{k-1}} (x_k+f^*(x_2 \ldots x_{k-1}))=\sum_{x_2 \ldots x_{k-1}\in B^{k-2}} (0+f^*(x_2 \ldots x_{k-1}))+$$

$$+\sum_{x_2 \ldots x_{k-1}\in B^{k-2}} (1+f^*(x_2 \ldots x_{k-1}))=2^{k-2}$$

is even.

From Lemma 5.1 it follows that card $(B^k/\mathscr{F}^r)$ is even. For $(B^k, \mathscr{F}^l)$ there exists a dual k-register $(B^k, \hat{\mathscr{F}}^l)$ such that card $(B^k/\mathscr{F}^l) = \text{card}\,(B^k, \hat{\mathscr{F}}^l)$. Since

$$\hat{f}(x_1 \ldots x_k) = f(x_2 \ldots x_k x_1) = x_2 + f^*(x_3 \ldots x_k) + x_1 = x_1 + (x_2 + f^*(x_3 \ldots x_k))$$

then

$$\sum_{x_2 \ldots x_k \in B^{k-1}} (x_2 + f^*(x_3 \ldots x_k)) = \sum_{x_3 \ldots x_k \in B^{k-2}} (0 + f^*(x_3 \ldots x_k)) +$$

$$+ \sum_{x_3 \ldots x_k \in B^{k-2}} (1 + f^*(x_3 \ldots x_k)) = 2^{k-2}$$

and from Lemma 5.1 it follows that the number card $(B^k/\mathscr{F}^l)$ is even. ∎

Corollary 5.1. For $k > 2$ there are do not exist two conjugated k-registers such that their transition graphs have cycles of length 2^k.

6. Similarity of the pairs of conjugated k-registers

The purpose of this section is to adopt the algebraic methods which have been demonstrated in section 4 for the construction of the whole class of similar pairs of conjugated k-registers to the given one. Only elements of $\mathscr{R}^r \times \mathscr{R}^l$ $(k > 2)$ will be considered here.

A pair $\langle B^k, (\mathscr{F}^r, \mathscr{F}^l)\rangle$ is said to be a net of k-registers $(B^k, \mathscr{F}^r)$ and $(B^k, \mathscr{F}^l)$ and will be denoted by $\mathbf{N}^{k,f}$, where f is the feedback function of these k-registers.

For arbitrary nets $\mathbf{N}^{k,f} = \langle B^k, (\mathscr{F}^r, \mathscr{F}^l)\rangle$ and $\mathbf{N}^{k,h} = \langle B^k, (\mathscr{H}^r, \mathscr{H}^l)\rangle$ let us define a set

$$\text{Dif}\,(\mathbf{N}^{k,f}, \mathbf{N}^{k,h}) = \{x \in B^{k-2}: f^*(x) \neq h^*(x)\}$$

where f^* and h^* are the functions which are connected with f and h by equality (5.3).

Let us see that the following equalities hold:

$$\text{Dif}\,(\mathscr{F}^r, \mathscr{H}^r) = \text{Dif}\,(\mathbf{N}^{k,f}, \mathbf{N}^{k,h}) \times \{0, 1\} \tag{6.1}$$

$$\text{Dif}\,(\mathscr{F}^l, \mathscr{H}^l) = \{0, 1\} \times \text{Dif}\,(\mathbf{N}^{k,f}, \mathbf{N}^{k,h}). \tag{6.2}$$

For a net $\mathbf{N}^{k,f}$ and an arbitrary set $A \subset B^{k-2}$ symbol $\mathbf{N}_A^{k,f}$ denotes a net such that $\text{Dif}\,(\mathbf{N}^{k,f}, \mathbf{N}_A^{k,f}) = A$.

Now, we shall extend the notion of similarity of k-registers to the nets of k-registers. For this purpose let us consider two nets $\mathbf{N}^{k,f} = \langle B^k, (\mathscr{F}^r, \mathscr{F}^l)\rangle$ and $\mathbf{N}^{k,h} = \langle B^k, (\mathscr{H}^r, \mathscr{H}^l)\rangle$.

$\mathbf{N}^{k,f}$ is said to be right-hand side similar to $\mathbf{N}^{k,h}$ (notation: $\mathbf{N}^{k,f} \underset{r}{\approx} \mathbf{N}^{k,h}$) if and only if $(B^k, \mathscr{F}^r) \approx (B^k, \mathscr{H}^r)$.

$\mathbf{N}^{k,f}$ is said to be left-hand side similar to $\mathbf{N}^{k,h}$ (notation: $\mathbf{N}^{k,f} \underset{l}{\approx} \mathbf{N}^{k,h}$) if and only if $(B^k, \mathscr{F}^l) \approx (B^k, \mathscr{H}^l)$.

$\mathbf{N}^{k,f}$ is said to be both-hand side similar to $\mathbf{N}^{k,\mathscr{A}}$ (notation: $\mathbf{N}^{k,f} \underset{}{\approx} \mathbf{N}^{k,\mathscr{A}}$) if and only if $\mathbf{N}^{k,f} \underset{r}{\approx} \mathbf{N}^{k,\mathscr{A}}$ and $\mathbf{N}^{k,f} \underset{l}{\approx} \mathbf{N}^{k,\mathscr{A}}$.

Theorem 6.1. For an arbitrary net $\mathbf{N}^{k,f} = \langle B^k, (\mathscr{F}^r, \mathscr{F}^l) \rangle$ and a set $A \subset B^{k-2}$ the following conditions are satisfied:

$$\mathbf{N}^{k,f} \underset{r}{\approx} \mathbf{N}_A^{k,f} \text{ if and only if } M_{\mathscr{F}^r|A\times\{0,1\}} \text{ is nonsingular over } GF(2); \tag{6.3}$$

$$\mathbf{N}^{k,f} \underset{l}{\approx} \mathbf{N}_A^{k,f} \text{ if and only if } M_{\hat{\mathscr{F}}^l|\{0,1\}\times A} \text{ is nonsingular over } GF(2); \tag{6.4}$$

$\mathbf{N}^{k,f} \approx \mathbf{N}_A^{k,f}$ if and only if $M_{\mathscr{F}^r|A\times\{0,1\}}$ and $M_{\hat{\mathscr{F}}^l|\{0,1\}\times A}$ are nonsingular over $GF(2)$. (6.5)

Proof. It follows immediately from the properties of dual k-registers and Theorem 4.1. ∎

Example 6.1. Consider a net of 5-registers $\mathbf{N}^{5,f} = \langle B^5, (\mathscr{F}^r, \mathscr{F}^l) \rangle$ which is defined by means of feedback function $f(x_1x_2x_3x_4x_5) = x_1 + x_2 + x_4 + x_5$. Since $P_{\mathscr{F}^r} = P_{\hat{\mathscr{F}}^l} = B^4 - \{0000\}$ then matrices $M_{\mathscr{F}^r}$ and $M_{\hat{\mathscr{F}}^l}$ have a form (rows and columns are lexicographically ordered in $B^4 - \{0000\}$):

$$M_{\mathscr{F}^r} = \begin{bmatrix}
0&1&1&0&1&1&0&1&1&0&0&0&1&0&0\\
1&0&0&1&0&0&0&1&0&0&0&0&0&0&0\\
1&0&0&0&0&0&1&1&0&0&0&1&0&1&0\\
0&1&0&0&0&0&0&0&1&0&0&0&0&0&0\\
1&0&0&0&0&0&0&1&1&1&0&0&0&0&0\\
1&0&0&0&0&0&1&1&1&0&0&0&0&1&0\\
0&0&1&0&0&1&0&0&1&0&1&1&1&1&0\\
1&1&1&0&1&1&0&0&1&0&0&0&1&0&0\\
1&0&0&1&1&1&1&1&0&0&0&1&1&1&0\\
0&0&0&0&1&0&0&0&0&0&0&0&0&0&0\\
0&0&0&0&0&0&1&0&0&0&0&0&1&1&0\\
0&0&1&0&0&0&1&0&1&0&0&0&0&1&0\\
1&0&0&0&0&0&1&1&1&0&1&0&0&1&0\\
0&0&1&0&0&1&1&0&1&0&1&1&1&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0
\end{bmatrix}$$

$$M_{\hat{\mathcal{F}}^1} = \begin{bmatrix}
0&0&0&1&0&1&0&1&1&1&1&1&0&0&0\\
0&0&0&1&0&0&0&0&1&0&0&0&0&0&0\\
0&0&0&0&0&0&1&0&1&0&0&1&0&1&0\\
1&1&0&0&0&0&0&1&0&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&1&0&0&0&0&0\\
1&0&0&0&0&0&1&1&1&0&0&0&0&1&0\\
0&0&1&0&0&1&0&0&1&0&1&1&1&1&0\\
1&0&0&1&0&1&0&0&1&1&1&1&0&0&0\\
1&1&1&0&0&1&1&1&0&1&1&0&0&1&0\\
1&0&0&0&1&0&0&1&1&0&0&0&0&0&0\\
1&0&0&0&0&0&1&1&1&0&0&0&1&1&0\\
1&0&1&0&0&0&1&1&0&0&0&0&0&1&0\\
0&0&0&0&0&0&1&0&0&0&1&0&0&1&0\\
0&0&1&0&0&1&1&0&1&0&1&1&1&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0
\end{bmatrix}$$

There are 15 sets $A_i \subset B^3$ for which $\mathbf{N}^{5,f} \underset{r}{\approx} \mathbf{N}^{5,f}_{A_i}$, 15 sets $D_i \subset B^3$ for which $\mathbf{N}^{5,f} \underset{l}{\approx} \mathbf{N}^{5,f}_{D_i}$ and 9 sets $E_i \subset B^3$ for which $\mathbf{N}^{5,f} \approx \mathbf{N}^{5,f}_{E_i}$. The list of these sets is given in Table I (elements of B^3 are represented in decimal form). ▮

Table I

i	A_i	D_i	E_i
1	{3}	{1}	{1, 2, 4}
2	{4}	{6}	{1, 3, 4}
3	{2, 4}	{1, 2}	{1, 4, 6}
4	{4, 6}	{1, 3}	{1, 2, 5, 6}
5	{1, 2, 4}	{1, 2, 4}	{2, 3, 4, 5}
6	{1, 3, 4}	{1, 2, 6}	{1, 2, 3, 4, 5}
7	{1, 4, 6}	{1, 3, 4}	{1, 2, 3, 5, 6}
8	{2, 3, 4}	{1, 4, 6}	{1, 2, 4, 5, 6}
9	{1, 2, 3, 4}	{1, 2, 4, 6}	{2, 3, 4, 5, 6}
10	{1, 2, 5, 6}	{1, 2, 5, 6}	
11	{2, 3, 4, 5}	{2, 3, 4, 5}	
12	{1, 2, 3, 4, 5}	{1, 2, 3, 4, 5}	
13	{1, 2, 3, 5, 6}	{1, 2, 3, 5, 6}	
14	{1, 2, 4, 5, 6}	{1, 2, 4, 5, 6}	
15	{2, 3, 4, 5, 6}	{2, 3, 4, 5, 6}	

7. Final remarks

The aim of this paper is to give a construction of the whole class of the pairs of conjugated k-registers which are similar to the given pair. The similarity notion corresponds to the property that each pair of k-registers is able to send the same information (each state is a code of any information). This is of great importance for the practice, in particular for criptology where the conjugated k-registers can be used.

On the other hand, the similarity notion of the nets corresponds to a well-known notion of symmetry which has been fruitfully explored in certain areas.

In opinion of the authors the notion of the net of conjugated k-registers allows to extend the theory of shift registers. The investigations on this problem can be continued.

We put forward only one problem. Construct a whole class $\mathbf{M}_k$ of the nets of conjugated k-registers (or even a singular net) with the following property: all the connected components of the transition graphs of both conjugated k-registers form cycles of length 2^k-1 and 1. What is the cardinality of the class $\mathbf{M}_k$?

References

1. *Aleksiejev, A. I., Szeremietiev, A. G., Tuzov, G. I., Glazov, I.*, Theory and application of pseudorandom signals (In Russian); Izd. Nauka, Moscow, 1969.
2. *Blahut, R. E.*, Theory and practice of error control codes. Addison–Wesley Publishing Company 1984.
3. *Golomb, S.W.*, Shift register sequences. Holden Day, San Francisco 1967.
4. *Grodzki, Z.*, Synthesis problem for deterministic controlled (k, m)-shift registers (to appear in Demonstratio Mathematica).
5. *Łatko, M.*, The similitude of shift registers (to appear in Demonstratio Mathematica).
6. *Ronse, Ch.*, Feedback shift registers, Lecture Notes in Computer Science. (Eds G. Gross, J. Hartmanis), Springer Verlag, Berlin 1984.

Сети сопряженных k_*-регистров сдвига

З. ГРОДЗКИ, М. ЛАТКО

(Люблин)

Рассматриваются сети левосторонних и правосторонних k-регистров сдвига.

Две сети называются похожими, если существует их взаимно однозначное отображение такое, что отвечающие друг другу циклы в графах перехода таких сетей имеют одни и те же состояния.

Адаптируется алгебраический метод для конструкции семейства похожих сетей.

Z. Grodzki
M. Łatko
Nadbystrzycka 38a
Lublin

Problems of Control and Information Theory, Vol. 17 (3), pp. 171—172 (1988)

THE THIRD INTERNATIONAL WORKSHOP ON INFORMATION THEORY "CONVOLUTIONAL CODES; MULTI-USER COMMUNICATION"

The workshop was held from May 24 to 30, 1987 on the basis of the cooperation between the Institute for Information Transmission Problems of the USSR Academy of Sciences, Lund University and Linköping University. This cooperation is realized within the frame of the agreement between the USSR Academy of Sciences and the Royal Swedish Academy.

Unlike the two previous workshops, the number of talks concerning the multi-user information theory have substantially decreased. The multi-user communication topic was mainly represented by papers on random-access communication and protocol sequence constructions.

The upper bound of the capacity of a slotted random access system with ternary feedback was improved by Tsybakov and Likhanov (USSR). This new upper bound, the sharpest known so far, is 0.568. Vvedenskaya and Pinsker (USSR) investigated the capacity of a random channel access algorithm obeying the first come first served (FCFS) rule. It was shown that the capacity of the part and try type algorithms, which are an important subclass of the FCFS procedures and are well known to the researchers of this area, is very close to that of the FCFS algorithms. The difference is of order 10^{-6}. Mikhailov (USSR) presented the stability analysis of a channel access algorithm suitable for random access type communication in multi-hop networks. Khasminskii (USSR) demonstrated, how dynamic programming can be used to find the optimal parameters of some contention resolution algorithms for random access channels.

The problem of finding non random protocols and code constructions for different multi-user communication situations proved to be of increasing importance.

Mathys (USA) constructed a code for a multiple-access adder channel with N binary inputs. It was assumed that only $T \ll N$ user are active at any given time. Assuming that the elements of the active subset are known, the problem was to construct a set of N uniquely decodable code combination. Codes with sum rates strictly greater than 1 was given. The easy way of the decoding was also demonstrated.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

Győrfi, Győri (Hungary) and A (Vietnam) focused their attention on a nonadaptive protocol construction for conflict resolution. Assuming ALOHA type binary feedback (success, no success), a set of protocol sequences was constructed for T users having only $M = 3$ as active among them. The nonadaptivity implies that the feedback is of limited importance. Its use is confined to the stopping of further transmissions as a packet has been transmitted successfully.

In the field of the coding theory, the growing interest in the complex modulation-coding approach was experienced.

The paper by Zinoviev, Zyablov, Litsin and Portnoi (USSR) on "Cascade methods of Conctructing and Decoding of Codes In Euclidean Spaces", may well attract the attention of the communication engineers. Instead of treating the modulation and coding separetly, which is still common in the engineering practice, they handle them jointly to find optimal modulation-coding procedures.

The importance of the connection between modulation and coding was also demonstrated in the paper by Burnashev (USSR) and Biglieri (Italy). They gave an upper bound on the minimum distance of trellis code with PSK modulation.

The talk given by T. Kailath (USA) was of outstanding interest. He covered the problems of decoding convolutional codes by Viterbi decoder in VLSI. As the main goal is to implement decoders for high-speed digital communication, the highest possible level of parallel processing should be reached. On the other hand, the interconnection of the simple processors the network of which executes the decoding is limited by the technology. To circumvent this problem, graph theoretic methods have been invoked. The application of the Suffle-Exchange graphs proved to be very fruitful.

I. Kerekes

Problems of Control and Information Theory, Vol. 17 (3), pp. 173—174 (1988)

MELECON '87 — THE MEDITERRANEAN ELECTROTECHNICAL CONFERENCE, MARCH 24–27, 1987, ROME, ITALY

The theme of this biennial scientific event "Developments in Telecommunication and Energy Systems" has attracted the attention of a large number of scientists not only from the Mediterranean countries but also from beyond this area. The above general theme has been sub-divided into four parallel sessions, among which the results presented in the area "Communication and Computer Networks" are probably interesting for the readers of this journal.

The first session was devoted to the performance evaluation of local area networks and chaired by such famous representatives of this field as E. N. Protonotarios (Greece) and K. Kuemmerle (Switzerland). The latter was also an invited lecturer and his interesting talk called "Local area networks: state of the art and trends". The topic of the second session was: "Implementation and application of computer networks", the chairman was L. Fratta (Italy). The third session was chaired by K. E. Drangeid (Switzerland) on the topic "Design and planning of integrated services networks". Several interesting contributions were included into the fourth session under the name "Multiservice networks".

The majority of papers dealt with theoretical and/or simulational investigations of multiple access algorithms, with the emphasis on protocols suitable for integrating various kinds of information (e.g. voice, interactive data, pictures etc.). Some of them will be briefly mentioned below.

M. Ajmone Marsan and R. D. De Lavoco (Italy) analysed the well known CSMA/CD (carrier sense multiple access with collision detection) algorithm, extended for the "multi-channel" case (i.e., when a number of parallel accessible transmission media, viz. cables etc. is to be shared among a population of users).

The CSMA/CD random access protocol is well suitable for transmitting interactive data traffic but modifications are needed when one wishes to handle periodic (e.g. voice) traffic, as well. An interesting extension of this protocol and an experimental system based on it was analysed by J. Dunlop (United Kingdom).

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

A class of new, hybrid multiple access protocols is also of interest where the traffic of periodic sources is given higher priority and it is transmitted in a scheduled fashion while data packets are transmitted during the time freed up by voice according to the CSMA/CD algorithm. Two protocol versions belonging to this class were analysed by L. Fratta (Italy) and Cs. Szabó (Hungary).

A new type of integrated communication systems called metropolitan area networks (MAN's) is emerging nowadays which is characterized by a tree-like topology and broadband signaling instead of baseband. New protocols are required for this new class of communication systems. A number of contributions were devoted to this and related questions, e.g., the one given by G. K. Kehagias et al. of Greece ("Protocols for broadband local area networks"), the paper of V. J. Sgardoni et al., "Accessing protocols for hierarchical metropolitan cable networks" (Greece).

Specific sessions were organized on communication protocols responsible for routing the packets and the packet flow control in packet-switched networks.

In addition to the above mentioned papers and their companions of analytical character, there was an other interesting category of those papers reporting about new integrated-services network prototypes. All this shows that, in this area, the new theoretical results on multiple access protocols are being rapidly incorporated in practical communication systems.

Cs. Szabó

Problems of Control and Information Theory, Vol. 17 (3), pp. 175—176 (1988)

BOOK REVIEWS

Hyde, G. (ed.): Review of Radio Science, 1984–1986. International Union of Radio Science (URSI), Brussels, Belgium, 1987.

"Radio Science is the underpinning of a communication/information revolution that is fueled by telephone, television and radio navigation, and explosively developed by satellite communication, space exploration, fiber optics, and computers to deal with information as well as arithmetic." This statement, reproduced from the Preface by A. P. Mitra, President, URSI particularly well emphasizes why this very Union, originally mostly devoted to radio propagation studies and measurements, became so actively involved also in the field of Communications and particularly in Information Theory and related disciplines.

Only one of the nine chapters in this Review, viz. Chapter 3, Signals and Systems is likely to be really of interest to the readers of this Journal, however a fair part of this very chapter certainly represents one of the most ambitious accounts of the progress in Information Theory, Digital and Optical Signal Processing, the basic methodologies of Optical Fiber and Radio Communications and Radio Networks, made in the period November 1, 1984 to October 30, 1986. Chapter 3 includes eight, concise but very thoughtfully written, sections and, at the end of the Chapter, almost 200 references (confined almost entirely to internationally acknowledged journals and alike).

The chapter titles of the Review and the section titles specifically of Chapter 3 read as follows:

Chapter 1: Electromagnetic Metrology, Chapter 2: Fields and Waves, Chapter 3: Signals and Systems, Chapter 4: Electronic and Optical Devices, Chapter 5: Electromagnetic Noise and Interference, Chapter 6: Wave Propagation and Remote Sensing, Chapter 7: Ionospheric Radio Propagation: Waves in Plasmas. Chapter 8: Radio Astronomy, Chapter 9: Bioeffects of Electromagnetic Waves.

Section 3.1: Introduction, Section 3.2: Digital Signal Processing, Section 3.4: Circuits and Systems, Section 3.5: Information Theory, Section 3.6: Radio Communications, Section 3.7: Optical Fiber Communications, Section 8: Communication Networks.

Comparing the latter list with the present main fields of the Problems of Control and Information Theory (as stated on the cover of any 1987 issue), Section 5 and 8 within Chapter 3 may be of most immediate interest to those readers of this journal who are active in Information Theory, Cooperative Queueing and Distributed Control.

Chapter 3 was compiled from nearly 240 standard size MS pages and more than 1000 references, submitted by 23 National Committees. While many of the world widely well known contributions and contributors of the reported period appear in Section 3.5 and 3.8, the compression of 9:1 page ratio, a reference ratio of 6:1 and the differences in the time, tastes and interests of the National Committees, almost unavoidable caused some unfortunate overlooks and non-optimal selections. Because of this, the present book reveiewer particularly welcomes that the reviews of at least four of the National Committees, particularly active in the field, were also simultaneously published and distributed by the URSI at the time of the XXIIth General Assembly. Thus at least the final step of the compilation can be personally traced in detail by anybody interested in any particular topic.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

Considering Section 3.5 with 3.8 more distinctly, three internationally relevant key events are pointed out in the field during the considered period: the two IEEE International Symposia on Information Theory in Brighton (England), June 1985, and in Ann Arbor, MI (USA), in October 1986, and the Sixth International Symposium on Information Theory, arranged mainly by the Institute for Information Transmission Problems of the USSR Academy of Sciences in Tashkent (USSR) in September 1984. It is also appropriately emphasized that the majority of the contributions to Information Theory and related methodologies were reported in the Problemy Peredachi Informatsii and in the IEEE Transactions of Information Theory. The latter also published two excellent special issues during the reported period: "Linear Adaptive Filtering" (March 1984) and "Random Access Communications" (March 1985).

Section 3.5 includes four brief subsections on Multiuser Communication, Coding and Cryptography, Source Coding and Information and Statistics, and is commenting on 21 contributions. Section 3.8 has got three sub-sections on Communication Protocols, Local Area Networks and Switching Systems, and includes specific comments altogether on thirty two research papers.

While reviewing on such a broad field, with so many relevant contributions just during couple of years, can hardly be appropriate in every respect, the present reviewer assumes that in the field of Signals and Systems, specifically within the fields of Section 3.5 and 3.8, closest to the majority of the readers of this journal, the International Union of Radio Science provided again a key reference to most of the most recent advances: the Review of Radio Science, as well as its simultaneously published national background reviews, provide much useful information to workers in Information Theory, Cooperative Queueing and Distributed Control in these days, in several respects, even to those having nothing to do with Radio Science itself.

S. Csibi

L. Devroye "A Course in Density Estimation" Birkhauser, Boston–Basel–Stuttgart, 1987.

This book was written as course notes during the summer quarter of 1986, when the author taught a course on density estimation in the Department of Statistics at Stanford University; therefore it is a kind and didactic textbook for a graduate course on density estimation and the ideas and techniques are illustrated by figures and each chapters is followed by exercises.

The subject of the book is the L_1 theory of density estimation the same as that of the Devroye, Györfi (1985) book. However, it is not a research monograph. The book contains some selected basic problems of the fields as follows:

I Distances between densities
II Density estimation and derivation of measures
III Consistency of kernel estimate
IV Robustness
V Minimax bounds
VI Minimum distance estimators
VII Rate of convergence of kernel estimates
VIII Case study: monotone densities on [0, 1]
IX Relative stability

Reference

L. Devroye, L. Györfi "Nonparametric Density Estimation: the L_1 Wiew", Wiley, 1985.

L. Györfi

Проблемы управления и теории информации, том **17**, *номер 3*

РУССКИЙ ПЕРЕВОД

ДИНАМИКА ПУЧКА ВЫЖИВАЮЩИХ ТРАЕКТОРИЙ ДИФФЕРЕНЦИАЛЬНОГО ВКЛЮЧЕНИЯ: ЭВОЛЮЦИОННОЕ УРАВНЕНИЕ

А. Б. Куржанский, Т. Ф. Филиппова

(*Свердловск*)

Рассматривается ансамбль траекторий дифференциального включения, сохраняющихся (выживающих) до некоторого предписанного момента времени в пределах заданного множества в фазовом пространстве. Приводится уравнение, описывающее эволюцию областей достижимости данного дифференциального включения при ограничениях на фазовые переменные.

1. Введение

Настоящая работа посвящена вопросам построения математических моделей, описывающих динамику сложных нелинейных систем, функционирующих в условиях неопределенности. Характер неопределенности в рассматриваемых процессах состоит в том, что параметры и входные воздействия в системах, а также измеряемые по ходу развития процесса характеристики исследуемого объекта известны лишь с точностью до некоторых непредсказуемых ошибок. Предполагаются заданными ограничивающие множества, содержащие указанные возмущения (ошибки), при этом какая-либо статистическая информация о неопределенных факторах в системах отсутствует.

При данных информационных предположениях состояние системы в текущий момент времени определяется неоднозначно. Поэтому динамика изучаемого объекта не может быть описана отдельной траекторией — решением соответствующего дифференциального уравнения, как в классической теории управления движением. Рассматриваемая задача допускает естественную формализацию, основанную на использовании аппарата теории дифференциальных включений [1–3]. При данной формализации место точно определенной классической траектории движения занимает ансамбль (пучок) решений дифференциального включения, исходящих из заданного множества возможных начальных состояний системы. Важным моментом является наличие в системе текущей информации, получаемой в результате наблюдений

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

(измерений). Математически последнее обстоятельство выражается в том, что из всего пучка возможных движений необходимо выделить лишь те, что совместимы с полученным на данный момент времени объемом информации, т. е. построить так называемое информационное множество [4, 5]. Иначе говоря, рассматривая систему наблюдения как фазовое ограничение на траектории включения, требуется найти множество всех выживающих к некоторому моменту (сохраняющихся на заданном промежутке времени в пределах фазового ограничения) решений дифференциального включения [1–3].

Известно [6, 7], что трубку выживающих траекторий можно описать, как множество решений нового дифференциального включения, правая часть которого определяется при помощи касательного конуса к фазовому ограничению. В данной работе развивается иной подход к рассматриваемой задаче, позволяющий избежать достаточно сложной операции построения указанного отображения с коническими значениями. Отметим, что сечения пучков выживающих траекторий (информационных множеств) по существу являются аналогами фазовых состояний системы, поскольку вбирают в себя по определению все имеющиеся информационные и динамические характеристики системы. В связи с этим представляет интерес задача поиска уравнения, описывающего эволюцию во времени данных «фазовых состояний» системы, решение этой задачи и составляет цель настоящей работы. Установлено, что сечения пучков выживающих траекторий удовлетворяют специальному эволюционному уравнению, которое обобщает уравнение интегральной воронки дифференциального включения [8, 9] и совпадает с последним в том случае, когда фазовые ограничения отсутствуют.

2. Постановка задачи

Пусть R^n означает n-мерное евклидово пространство, (x, y) (или $x'y$) — скалярное произведение векторов $x, y \in R^n$ (штрих-символ транспонирования); $\|x\| = (x, x)^{1/2}$, $S = \{x \in R^n : \|x\| \leqq 1\}$. Обозначим $\operatorname{conv} R^n$ совокупность всех выпуклых компактных подмножеств R^n, $h(A, B)$ — метрика Хаусдорфа ($A, B \in \operatorname{conv} R^n$).

Рассмотрим дифференциальное включение

$$\dot{x} \in F \subset t, x) \qquad (t_0 \leqq t \leqq T), \tag{2.1}$$

где $x \in R^n$, F — непрерывное отображение из $[t_0, T] \times R^n$ в $\operatorname{conv} R^n$. Предположим, что выполнено следующее условие Липшица ($L > 0$):

$$h(F(t, x), \quad F(t, y)) \leqq L\|x - y\|, \quad \forall x, y \in R^n.$$

Пусть дано множество $X_0 \in \operatorname{conv} R^n$. Для каждого $x_0 \in X_0$ обозначим $x[t] = x(t; t_0, x_0)$ $(t_0 \leqq t \leqq T)$ решение в смысле К. Каратеодори включения (2.1)

— абсолютно непрерывную функцию времени, удовлетворяющую соотношениям: $x[t_0]=x_0$, $\dot{x}[t]\in F(t,x[t])$ при почти всех $t\in[t_0,T]$. Будем считать, что выполнено какое-либо из условий, гарантирующих продолжимость решений $\{x(t;t_0,x_0)|x_0\in X_0\}$ на весь промежуток $[t_0,T]$ [10].

Пусть $Y(t)$ — непрерывная многозначная функция, определенная на $[t_0,T]$, со значениями в conv R^n, $X\subseteq Y(t_0)$.

Определение [1, 2]. Траекторию $x[\cdot]=x(\cdot;t_0,x_0)$ $(x_0\in X_0)$ дифференциального включения (2.1) назовем выживающей к моменту $\tau(t_0<\tau\leqq T)$, если

$$x[t]\in Y(t) \tag{2.2}$$

при всех $t\in[t_0,\tau]$.

Для каждого $x_0\in X_0$ множество всех выживающих к моменту τ траекторий $x(\cdot;t_0,x_0)$ обозначим символом $X(\cdot;\tau,t_0,x_0)$. Положим

$$X(\cdot;\tau,t_0,X_0)=U\{X(\cdot;\tau,t_0,x_0)|x_0\in X_0\},$$

$$X(\tau;t_0,X_0)=X(\tau;\tau,t_0,X_0).$$

Пусть $X^*(\cdot;t_0,X_0)$ означает ансамбль всех траекторий дифференциального включения (2.1), исходящих в начальный момент t_0 из множества X_0. В силу предположений относительно правой части F включения (2.1) множество

$$M=U\{X^*(t;t_0,X_0)|t_0\leqq t\leqq T\}$$

компактно в R^n [9, 10]. Обозначим gr_tF график отображения $F(t,\cdot)$ (t фиксировано):

$$gr_tF=\{\{x,y\}\in R^n\times R^n:\quad y\in F(t,x)\}.$$

int A — внутренность множества $A\subseteq R^n$.

Примем следующее

Предположение A. 1. Для некоторого выпуклого компакта $D\subset R^n$ такого, что $M\subset\text{int}\,D$, при каждом $t\in[t_0,T]$ множество $D\cap gr_tF$ выпукло.

2. Существует решение $x_*[\cdot]$ включения (2.1), удовлетворяющее условиям: $x_*[t_0]\in X_0$,

$$x_*[t]\in\text{int}\,Y(t),\quad\forall t\in[t_0,T].$$

Отметим, что при предположении A пучок $X(\cdot;\tau,t_0,X_0)$ выживающих к моменту τ траекторий включения (2.1) является непустым, выпуклым и компактным подмножеством пространства $C[t_0,T]$ непрерывных n-векторных функций, а его τ-сечение $X(\tau;t_0,X_0)$ непусто, выпукло и компактно в R^n.

Известно, что множества $X(t;t_0,X_0)$ удовлетворяют полугрупповому свойству

$$X(\tau;t_0,X_0)=X(\tau;s,X(s;t_0,X_0)).$$

Следовательно, они порождают обобщенную динамическую систему. Целью настоящей работы является построение адекватной эволюционной модели, описывающей указанную систему.

3. Эволюционное уравнение

Будем предполагать, что многозначное отображение Y удовлетворяет одному из следующих условий.

Предположение B. График $gr\,Y$ отображения Y-выпуклый компакт в R^{1+n}.

Предположение C. При каждом $l \in R^n$ опорная функция $\rho(l|\,Y(t))$ отображения $Y(t)$,

$$\rho(l|\,Y(t)) = \max\{l'y \mid y \in Y(t)\},$$

дифференцируема по t, и производная $d/dt\ \rho(l|\,Y(t))$ непрерывна по совокупности переменных (t, l).

Справедлива

Теорема 3.1. Пусть выполнено предположение A. Если отображение Y удовлетворяет предположению B или C, то τ-сечение $X[\tau] = X(\tau;\ t_0,\ X_0)$ ансамбля $X(\cdot\,;\tau,\ t_0,\ X_0)$ выживающих к моменту τ траекторий включения (2.1) является решением следующего эволюционного уравнения

$$\lim_{\sigma \to 0+} \sigma^{-1} h\Big(X[\tau+\sigma],\ \bigcup_{x \in X[\tau]} (x + \sigma F(\tau, x)) \bigcap Y(\tau+\sigma)\Big) = 0, \tag{3.1}$$

$$t_0 \leqq \tau \leqq T, \qquad X[t_0] = X_0.$$

Доказательство теоремы вытекает из серии лемм, приведенных в следующем разделе. В заключение данного параграфа сделаем несколько замечаний.

Замечания. 1. В предположениях теоремы 3.1 многозначная функция $X[\tau] = X(\tau;\ t_0,\ X_0)$ непрерывна по $\tau \in [t_0,\ T]$.

2. Если в теореме 3.1 предположение A (2) заменить следующим: существует решение $x_*[\,\cdot\,]$ включения (2.1) такое, что

$$x_*[t] \in \operatorname{int} Y(t)$$

при почти всех $t \in [t_0,\ T]$, то функция $X[\tau] = X(\tau;\ t_0,\ X_0)$ будет удовлетворять эволюционному уравнению (3.1) почти в каждой точке $\tau \in [t_0,\ T]$. В этом случае $X[\tau]$ может бытъ разрывной на множестве меры нулъ. (Известно, что $X[\tau]$ непрерывна слева и полунепрерывна сверху справа в каждой точке $\tau \in [t_0,\ T]$ [4]).

4. Вспомогательные результаты

Пусть момент $\tau \in [t_0, T]$ фиксирован, $X[\tau] = X(\tau; t_0, X_0)$. Основываясь на предположении о существовании внутренней (по отношению к фазовому ограничению) траектории дифференциального включения (2.1), нетрудно получить следующий результат.

Лемма 4.1. Пусть выполнено предположение A. Тогда для любого $\varepsilon > 0$ найдется $\sigma_* > 0$ такое, что при всех $\sigma \in [0, \sigma_*]$ справедливо включение

$$X[\tau+\sigma] \subseteq \left(\bigcup_{x \in X[\tau]} (x + \sigma F(\tau, x)) \cap Y(\tau+\sigma) \right) + \varepsilon\sigma S. \tag{4.1}$$

Уточним полученную оценку (4.1). Обозначим для каждого $\sigma > 0$

$$Z_\sigma(\tau) = \bigcup_{x \in X[\tau]} \{ z \in R^n : \quad \exists\, v_x(\cdot) \in L_1[\tau, \tau+\sigma],$$

$$Z = x + \int_\tau^{\tau+\sigma} v_x(\xi)\, d\xi, \qquad v_x(s) \in F(\tau, x), \tag{4.2}$$

$$x + \int_\tau^s v_x(\xi)\, d\xi \in Y(s), \qquad \tau \leqq s \leqq \tau+\sigma \}.$$

(Здесь $L_1[\tau, \tau+\sigma]$ — пространство суммируемых на отрезке $[\tau, \tau+\sigma]$ n-векторных функций). По определению

$$Z_\sigma(\tau) = U\{ Z_\sigma(x_0, \tau) \mid x_0 \in X[\tau] \},$$

где $Z_\sigma(x_0, \tau) - (\tau+\sigma)$ — сечение пучка выживающих к моменту $\tau+\sigma$ траекторий $x(\cdot)$ включения

$$\frac{dx}{dt} \in F(\tau, x_0), \qquad x(\tau) = x_0,$$

$$x(t) \in Y(t) \qquad (\tau \leqq t \leqq \tau+\sigma).$$

Лемма 4.2. Пусть выполнено предположение A. Тогда для любого $\varepsilon > 0$ найдется $\sigma_* > 0$ такое, что при всех $\sigma \in [0, \sigma_*]$ верных включения

$$X[\tau+\sigma] \subseteq Z_\sigma(\tau) + \varepsilon\sigma S \tag{4.3}$$

$$Z_\sigma(\tau) \subseteq \left(\bigcup_{x \in X[\tau]} (x + \sigma F(\tau, x)) \cap Y(\tau+\sigma) \right) + \varepsilon\sigma S. \tag{4.4}$$

Пусть $Z'_\sigma(\tau)$ означает подмножество $Z_\sigma(\tau)$, отвечающее постоянным функциям $v_x(t) \equiv \text{const}$ в определении (4.2):

$$Z'_\sigma(\tau) = \bigcup_{x \in X[\tau]} \{ x + \sigma v \mid v \in F(\tau, x) \cap (\bigcap_{0 < s \leqq \sigma} s^{-1}(Y(\tau+s) - x)) \}.$$

Лемма 4.3. 1. Пусть выполнено предположение А. Тогда для произвольного $\varepsilon>0$ можно указать $\sigma_*>0$ такое, что при любом $\sigma\in(0,\sigma_*]$

$$Z_\sigma(\tau)\subseteq X[\tau+\sigma]+\varepsilon\sigma S. \tag{4.5}$$

2. Пусть выполнено предположение В. Тогда при любом $\sigma>0$

$$\bigcup_{x\in X[\tau]}(x+\sigma F(\tau,x))\cap Y(\tau+\sigma)\subseteq Z'_\sigma(\tau)\subseteq Z_\sigma(\tau). \tag{4.6}$$

Сравнивая леммы 4.2 и 4.3, получаем

Следствие 4.1. Пусть выполнено предположение А. Тогда

$$\lim_{\sigma\to 0+}\sigma^{-1}h(X[\tau+\sigma],\ Z_\sigma(\tau))=0.$$

Введем ряд вспомогательных конструкций. Определим для произвольного замкнутого множества $C\subseteq R^n$ контингентный конус $T_c(x)$ $(x\in C)$:

$$T_c(x)=\{v\in R^n:\ \liminf_{\sigma\to 0+}\sigma^{-1}\min\{\|x+\sigma v-c\|:c\in C\}=0\}$$

и контингентную производную $DY(t,y)(\alpha)$ $(\alpha\in R^1)$ отображения $Y(\cdot)$ в точке $(t,y)\in grY$

$$DY(t,y)(\alpha)=\{v\in R^n:\ (\alpha,v)\in T_{grY}(t,y)\} \qquad [6,7].$$

Положим при $(t,y)\in grY$

$$V(t,y)=DY(t,y)(1). \tag{4.7}$$

При выполнении предположения С при любых $(t,y)\in grY$ множество $V(t,y)$ непусто, выпукло и замкнуто в R^n [6].

Зафиксируем момент $\tau\in[t_0,T]$ и построим локальную аппроксимацию $Y_\sigma(\tau)$ отображения $Y(\tau+\sigma)$ при малых $\sigma>0$ [11]:

$$Y_\sigma(\tau)=\bigcap_{y\in Y(\tau)}(y+\sigma V(\tau,y)) \tag{4.8}$$

$$Y_0(\tau)=Y(\tau).$$

Лемма 4.4. [11] 1. Пусть выполнено предположение С. Тогда при всех $\sigma\geqq 0$ справедливо равенство

$$Y_\sigma(\tau)=\{z\in R^n:\ l'z\leqq\rho(l|Y(\tau))+\sigma\frac{d}{d\tau}\rho(l|Y(\tau)),\ \forall\, l\in R^n\}.$$

2. Если выполнены предположения А (2) и С, то для любого $\varepsilon>0$ существует $\sigma_*>0$ такое, что при всех $\sigma\in(0,\sigma_*]$ имеют место включения

$$\begin{aligned} Y(\tau+\sigma)&\subseteq Y_\sigma(\tau)+\varepsilon\sigma S\\ Y_\sigma(\tau)&\subseteq Y(\tau+\sigma)+\varepsilon\sigma S. \end{aligned} \tag{4.9}$$

Заметим, что график отображения $Y_\sigma(\tau)$ (как многозначной функции от σ при фиксированном τ) является выпуклым. Поэтому из лемм 4.1–4.4 следует

Лемма 4.5. Пусть выполнены предположения A и C. Тогда

$$\lim_{\sigma\to 0+} \sigma^{-1}h \subset X[\tau+\sigma], \quad \bigcup_{x\in X[\tau]} (x+\sigma F(\tau, x))\cap Y_\sigma(\tau)) = 0.$$

Суммируя результаты лемм 4.1–4.5, приходим к заключению о справедливости теоремы 3.1.

5. Линейная система

Рассмотрим систему

$$\dot{x} \in A(t)x + P(t) \qquad (t_0 \leqq t \leqq T), \tag{5.1}$$

где $x \in R^n$, $A(t)$ — непрерывная $n\times n$-матричная функция, $P(t)$ — непрерывное отображение из $[t_0, T]$ в conv R^n.

Сохраним обозначение $X(\cdot\,; \tau, t_0, X_0)$ предыдущих разделов работы для ансамбля выживающих (по отношению к ограничению $Y(\cdot)$) к моменту τ решений $x[\cdot]$ системы (5.1);

$$X[\tau] = X(\tau;\ t_0, X_0) = X(\tau;\ \tau, t_0, X_0).$$

Условие A(1) в данном случае выполнено.

Предположение A′. Существует траектория $x_*[\cdot]$ включения (5.1) такая, что

$$x_*[t] \in \operatorname{int} Y(t), \qquad \forall\, t \in [t_0, T].$$

Следующий результат вытекает из теоремы 3.1 и обобщает теорему 4.1 [3] (с учетом замечания 2 § 3 настоящей работы).

Теорема 5.1. Пусть выполнено предположение A′. Если многозначное отображение $Y(\cdot)$ удовлетворяет одному из предположений B или C, то функция $X[\tau] = X(\tau;\ t_0, X_0)$ является решением уравнения:

$$\lim_{\sigma\to 0+} \sigma^{-1}h(X[\tau+\sigma],\ ((E+\sigma A(\tau))X[\tau] + \\ +\sigma P(\tau))\cap Y(\tau+\sigma)) = 0, \tag{5.2}$$

$$X[t_0] = X_0, \qquad t_0 \leqq \tau \leqq T$$

(здесь E — единичная $n\times n$-матрица).

По аналогии с теоремой 4.2 [3] доказывается более сильное утверждение (выпуклость $gr\,Y$ не требуется).

Теорема 5.2. Пусть выполнены предположения A′ и C. Тогда опорная функция $\rho(l|X[t])$ при любом $l \in R^n$ дифференцируема справа по $t \in [t_0, T]$ и

$$\frac{d^+}{dt}\rho(l|X[t]) = \min\{\rho(q'A(t)|\partial_l\rho(l|X[t])) +$$

$$+\rho(q|P(t)) + \frac{d}{dt}\rho(l-q|Y(t))|q \in Q(t,l)\},$$

$$Q(t,l) = \{q \in R^n:\ \rho(l-q|Y(t)) = \rho(l|X[t]) - \rho(q|X[t])\},$$

$$\partial_l\rho(l|X[t]) = \{x \in X[t]:\ l'x = \rho(l|X[t])\}.$$

6. Заключение

Предложенное в теоремах 3.1 и 5.1 зволюционное уравнение может бытъ использовано при построении аппроксимирующих процедур в задачах гарантированного оценивания динамики управляемых систем по результатам наблюдений в условиях помех [4, 5].

Литература

1. *Куржанский А. Б.* Об аналитическом описании пучка выживающих траекторий дифференциальной системы. Докл. Акад. Наук СССР, 1986, **287,** *5,* 1047–1050.
2. *Куржанский А. Б., Филиппова Т. Ф.* Об описании множества выживающих траекторий дифференциального включения. Докл. Акад. Наук СССР, 1986, **289,** *1,* 38–41.
3. *Куржанский А. Б., Филиппова Т. Ф.* Об описании пучка выживающих траекторий управляемой системы. Дифференц. уравнения, 1987, 8.
4. *Куржанский А. Б.* Управление и наблюдение в условиях неопределенности. М.: Наука, 1977.
5. *Kurzhanskii, A. B.*, Dynamic control system estimation under uncertainty conditions. I. Problems of Control and Information Theory, 1980, **9,** *6* pp. 395–406; II. 1981, **10,** *1,* pp. 33–42.
6. *Aubin, J.-P., Cellina, A.*, Differential inclusions. Heidelberg: Springer-Verlag, 1984.
7. *Aubin, J.-P., Ekeland, I.*, Applied nonlinear analysis. New York: Academic Press, 1984.
8. *Панасюк А. И., Панасюк В. И.* Асимптотическая магистральная оптимизация управляемых систем. Минск, «Наука и техника», 1986.
9. *Толстоногов А. А.* Дифференциальные включения в банаховом пространстве. Новосибирск, «Наука», 1986.
10. *Благодатских В. И., Филиппов А. Ф.* Дифференциальные включения и оптимальное управление. Труды Матем. института АН СССР, 169, Москва, «Наука», 1985.
11. *Demyanov, V. F., Lemaréchal, C., Zowe, J.*, Approximation to a set-valued mapping. I: a proposal. Appl. Math. Optim. 1986, **14,** *3*, pp. 203–214.

PRINTED IN HUNGARY

Akadémiai Kiadó és Nyomda Vállalat, Budapest

NOTE TO CONTRIBUTORS

Two copies of the *manuscript* (each complete with figures, tables and references) are to be sent to

E.D. TERYAEV coordinating editor
Department of Mechanics and Control Processes
Academy of Sciences of the USSR
Leninsky Prospect 14, Moscow V-71, USSR

or to V. STREJC
UTIA ČSAV
182 08 Prague 8
Pod Vodárenskou věží 4, Czechoslovakia

or to L. GYÖRFI
Technical University of Budapest
H-1111 Budapest, Stoczek u. 2, Hungary

Authors are requested to retain a third copy of the submitted typescript to be able to check the proofs.

The papers, preferably in English or Russian, should be typed double spaced on one side of good-quality paper with wide margins (4–5 cm). The first page of the paper should carry the title, the author(s)' names and the name of the town where they are active. The name and address of the author to whom the proofs should be sent should be given at the end of the paper. An *abstract* should head the paper. English papers should also have a Russian abstract.

The papers should not exceed 15 pages (25 × 50 characters per page) including tables and references. The proper location of the tables and figures must be indicated on the margin.

Mathematical notations should follow up-to-date usage. Equations longer than half a line should not be incorporated in the text. In-text equations must be typed on a single line except that one level of subscripting and/or superscripting is permissible. Use / instead of horizontal bars. Displayed equations should be written so as to require the fewest possible lines. Therefore use "exp" for the exponential function whenever the exponent requires more than a single line. Matrices should, if possible, not be written in full. Use subscript notations instead such as $A = ||a_{ij}||$. Write diagonal matrices as diag $(d_1, d_2, \ldots d_n)$.

The authors will be sent galley proofs to be returned by next mail. Rejected manuscripts will be returned. Authors will receive 100 reprints free of charge. Additional reprints may be ordered.

К СВЕДЕНИЮ АВТОРОВ

Рукописи статей в трех экземплярах на русском языке и в трех на английском следует направлять по адресу: 117312 Москва В-312, просп. 60 летия Октября, 9, МНИИПУ. Редакция журнала «Проблемы управления и теории информации» (зав. редакцией Н. И. Родионова, тел. 208-60-19).

Объём статьи не должен превышать 15 печатных страниц (25 строк по 50 букв). Статье должна предшествовать аннотация объемом 50–100 слов и приложено резюме–реферат объемом не менее 10–15% объема статьи на русском языке в трех экземплярах, на котором напечатан служебный адрес автора (фамилия, название учреждения, адрес).

При написании статьи авторам надо строго придерживаться следующей формы: введение (постановка задачи), основное содержание, примеры практического использования, обсуждение результатов, выводы и литература.

Статьи должны быть отпечатаны с промежутком в два интервала, последовательность таблиц и рисунков должна быть отмечена на полях. Математические обозначения рекомендуется давать в соответствии с современными требованиями и традициями. Разметку букв следует производить только во втором экземпляре и русского, и английского варианта статьи.

Авторам высылается верстка, которую необходимо незамедлительно проверить и возвратить в редакцию.

После публикации авторам высылаются бесплатно 100 оттисков их статей.

Рукописи непринятых статей возвращаются авторам.

CONTENTS · СОДЕРЖАНИЕ

HU ISSN 0370-2529 Index: 26 660

VOL. TOM 17 · NUMBER HOMEP 4

ACADEMY OF SCIENCES OF THE USSR
HUNGARIAN ACADEMY OF SCIENCES
CZECHOSLOVAK ACADEMY OF SCIENCES

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

АКАДЕМИЯ НАУК С С С Р
ВЕНГЕРСКАЯ АКАДЕМИЯ НАУК
ЧЕХОСЛОВАЦКАЯ АКАДЕМИЯ НАУК

1988

AKADÉMIAI KIADÓ, BUDAPEST
DISTRIBUTED OUTSIDE THE COMECON-COUNTRIES
BY PERGAMON PRESS, OXFORD

PROBLEMS OF CONTROL AND INFORMATION THEORY

An international bi-monthly sponsored jointly by the Presidium of the Academy of Sciences of the USSR, of the Hungarian Academy of Sciences and of the Czechoslovak Academy of Sciences. The six issues published per year make up a volume of some 480 pp. It offers publicity for original papers and short communication of the following topics:

— theory of control processes
— theory of adaptive systems
— theory of estimation and identification
— theory of controlling robot-technologic and flexible manufacturing systems
— information theory
— information-theoretic aspects of multiple access networks.

While this bi-monthly is mainly a publication forum of the research results achieved in the socialist countries, also papers of international interest from other countries are welcome.

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

Международный журнал Академии наук СССР, Венгерской Академии наук и Чехословацкой Академии наук выходит 6 раз в год общим объемом 480 печатных страниц.

В журнале публикуются оригинальные научные статьи и статьи обзорного характера по следующим проблемам управления и теории информации:

— теория процессов управления;
— теория адаптивных систем;
— теория оценивания и идентификации;
— теория управления робототехническими и гибкими производственными системами;
— теория информации;
— теория информации в области сетей с множественным доступом.

Целью журнала является ознакомление научной общественности различных стран с важнейшими проблемами, имеющими актуальный и перспективный характер, научными достижениями ученых социалистических и других стран.

Distributors

For the Soviet Union:

SOYUZPECHATY, Moscow 123 308 USSR

For Albania, Bulgaria, China, Cuba, Czechoslovakia, German Democratic Republic, Korean People's Republic, Mongolia, Poland, Rumania, Vietnam and Yugoslavia:

KULTURA Hungarian Foreign Trading Co.
P. O. Box 149 H-1389 Budapest, Hungary

For all other countries:

PERGAMON JOURNALS LTD, Headington Hill Hall, Oxford OX3 OBW, England

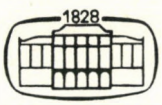
or

PERGAMON JOURNALS INC, Maxwell House, Fairview Park, Elmsford, NY 10523, USA

1988 Subscription Rate DM 535,— per annum including postage and insurance.

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ



AKADÉMIAI KIADÓ

PUBLISHING HOUSE OF THE HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST

Problems of Control and Information Theory, Vol. 17 (4), pp. 177—205 (1988)

A SOLUTION METHOD FOR SLP RECOURSE PROBLEMS WITH ARBITRARY MULTIVARIATE DISTRIBUTIONS — THE INDEPENDENT CASE

K. FRAUENDORFER, P. KALL*

(Zürich)

(Received October 2, 1986)

One way of solving stochastic linear programs with recourse is to approximate the objective function by a piecewise linear function — substituting the arbitrarily distributed random variables by discretely distributed ones. We determine sequences of discrete distributions such that the corresponding optimal objective values converge to the original optimal value and the discrete supports increase in such a way, that the total computational work can still be afforded. This requires to evaluate probabilities and conditional expectations which can be a difficult task for arbitrary distributions. Hence, in this case we replace the given distribution by a discrete one resulting from sampling and then apply the above-mentioned procedure. How large the sampling should be taken depends mainly on the variance of the recourse function. We determine an upper bound for that variance on the basis of the information yielded by the algorithm. Numerical examples illustrate the results.

1. Introduction

Problems of the form

$$\min \left[f(x) + \int_{\Omega} Q(x, \xi(\omega)) P(d\omega) \right] \tag{1.1}$$

$$\text{s.t. } x \in X$$

represent a broad class within the field of stochastic optimization. ξ is a random vector mapping of the probability space (Ω, A, P) into $(\mathbf{R}^K, B^K, F)$ where B^K is the Borel field on $\mathbf{R}^K$ and F is the corresponding distribution function.

$f: X \to \mathbf{R}$ is a real valued function defined on $X \subset \mathbf{R}^{n_1}$ and $Q: \mathbf{R}^{n_1} \times \Xi \to \mathbf{R} \cup \{+\infty\}$ is an extended real valued function, where $\Xi \subset \mathbf{R}^K$ is the support of the probability

* Part of the work for this paper could be carried out by the second author at the Mathematics Research Center, University of Wisconsin–Madison, with financial support by the National Science Foundation under Grant No. DCR-850 22 02. The hospitality and support by these institutions are greatly appreciated.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

measure induced by ξ (i.e. Ξ is the smallest closed set in $\mathbf{R}^K$ such that $P(\xi(\omega) \in \Xi) = 1$). We assume throughout that Ξ is a convex set.

The term

$$\int_{\Xi} Q(x, \xi) dF =: E(Q(x, \xi))$$

can be interpreted as the expected cost subject to the random vector ξ and a given decision $x \in X$.

If f is a linear function, X a convex polyhedral set and Q given as the optimal value of a linear program

$$\begin{aligned} Q(x, \xi) &= \min \langle q(\xi), y \rangle \\ \text{s.t.} \quad & Wy = h(\xi) - T(\xi) \cdot x \\ & y \geqq 0, \end{aligned} \tag{1.2}$$

where W is a fixed $(m_2 \times n_2)$-matrix and

$$\begin{aligned} q(\xi) &= q_0 + q_1\xi_1 + \ldots + q_K\xi_K, \\ h(\xi) &= h_0 + h_1\xi_1 + \ldots + h_K\xi_K, \\ T(\xi) &= T_0 + T_1\xi_1 + \ldots + T_K\xi_K, \end{aligned} \tag{1.3}$$

$$\left.\begin{array}{ll} q_i & n_2\text{-vectors} \\ h_i & m_2\text{-vectors} \\ T_i & (m_2 \times n_1)\text{-matrices} \end{array}\right\} \quad i = 0, \ldots, K,$$

$\xi_1, \ldots, \xi_K$ independent random variables,

we obtain from (1.1) a special class of problems, the so called stochastic linear programming problems (SLP) with fixed recourse.

SLP recourse problems can be expressed as follows:

$$\begin{aligned} \min \Psi(x) &:= \langle c, x \rangle + \int_{\Xi} Q(x, \xi) dF \\ \text{s.t.} \quad & Ax = b \\ & x \geqq 0 \end{aligned} \tag{1.4}$$

with

$$\begin{aligned} Q(x, \xi) &= \min \langle q(\xi), y \rangle \\ \text{s.t.} \quad & Wy = h(\xi) - T(\xi)x \\ & y \geqq 0, \end{aligned} \tag{1.4.1}$$

where

$$c \in \mathbf{R}^{n_1},\ b \in \mathbf{R}^{m_1} \text{ and } A \text{ of order } m_1 \times n_1.$$

The function $Q(x, \xi)$ is called the recourse function as it refers to the recourse problem (1.4.1), which has decision x as a parameter and involves the random coefficients $q(\xi)$, $h(\xi)$ and $T(\xi)$.

The practical meaning is as follows: At the first stage, when deciding upon x, we have no information about the values of ξ. Hence, looking for an optimal first stage decision x, we have to take into account not only the direct first stage cost $\langle c, x\rangle$, but also the expected value of the future recourse cost. If there is no feasible solution to (1.4.1) we assume $Q(x, \xi) = \infty$, thereby inducing the constraint $P(\{\xi | Q(x, \xi) = \infty\}) = 0$ on the decision x of the first stage.

The expected cost $E(Q(x, \xi))$ is caused by the fact, that in general it is impossible to require that the equality $T(\xi) \cdot x = h(\xi)$ is satisfied for each realization of the stochastic parameter ξ. Hence the "best" recourse decision y — also called second stage decision — may be interpreted as a correction, which generates minimal additional costs $Q(x, \xi)$ at the moment, when the random vector ξ is realized.

If W has the additional property, that $\{Wy : y \geqq 0\} = \mathbf{R}^{m_2}$ (i.e. the correction Wy in (1.4.1) can compensate for any error, whatever x, $h(\xi)$ and $T(\xi)$ may be) we have an SLP problem with complete fixed recourse. In this paper we confine ourselves to the case of complete fixed recourse, which is no essential restriction for most practical problems.

Let us now briefly discuss the basic properties, which are important for solving this class of problems.

By duality theory of linear programming we know that $Q(x, \xi) > -\infty$ (i.e. the recourse problem is bounded from below) if and only if there is a $u \in \mathbf{R}^{m_2}$ such that $W^T u \leqq q(\xi)$. Since the case $Q(x, \xi) = -\infty$ is of no interest for us, we shall assume from now on, that this condition is satisfied for each realization of the random vector $q(\xi)$.

Under this assumption the recourse function possesses the following properties [3]:

i) For any fixed ξ the function

$$x \to Q(x, \xi)$$

is a piecewise linear and convex function on

$$X := \{x | Ax = b, x \geqq 0\};$$

ii) For any fixed $x \in X$ and for any fixed $q(\xi) \equiv q^0$ the function

$$\xi \to Q(x, \xi)$$

is piecewise linear and convex.

Similarly, if we consider only $q(\xi)$ as random and the right-hand side of (1.4.1) as fixed, we obtain:

iii) For any fixed $x \in X$ and for any constant $h(\xi) \equiv h_0$, $T(\xi) \equiv T_0$, the function

$$\xi \to Q(x, \xi)$$

is piecewise linear and concave.

Under the additional assumption, that the random parameters $q(\xi)$, $h(\xi)$ and $T(\xi)$ have finite second moments, we get:

iv) The function

$$\tilde{Q}(x) = \int_{\Xi} Q(x, \xi) dF$$

is finite and convex on X.

In our case, considering only affine linear transformations in (1.3), the compactness of Ξ, therefore, ensures already the finiteness of $\tilde{Q}(x)$.

v) If ξ is a discretely distributed random variable with a finite number of realizations $\xi^1, \ldots, \xi^L$, the linear two-stage problem (1.4) takes on the form of a large-scale LP problem with dual block-angular structure:

$$\begin{aligned} \min \Psi(x) := \langle c, x \rangle + \sum_{l=1}^{L} p^l \langle q^l, y^l \rangle \qquad (1.5) \\ \text{s.t. } Ax = b \\ T^l x + W y^l = h^l, \qquad l = 1, \ldots, L \\ x, y^l \geqq 0, \end{aligned}$$

where

$$q^l = q(\xi^l),\ h^l = h(\xi^l),\ T^l = T(\xi^l),$$

$$p^l = P(\xi(\omega) = \xi^l) \quad \text{and} \quad p^l > 0 \,\forall\, l, \sum_{l=1}^{L} p^l = 1.$$

Minimizing $\Psi(x)$ in (1.5) implies (by $p^l > 0 \ \forall\, l$) automatically that the recourse costs $\langle q^l, y^l \rangle$ — subject to realization ξ^l — are minimized.

Observe that the number of realizations defines the number of blocks and hence the size of the LP (1.5).

A detailed discussion of the properties of stochastic linear programming problems can be found in [1] and [3].

Looking at properties ii) and iii) we realize, that for ensuring convexity, concavity, respectively, of the recourse function $Q(x, \xi)$ with respect to ξ, we have to

keep either q or h and T fixed, respectively. As there is no essential difference in treating convex or concave functions, from now on we will concentrate on the case, where $h(\xi)$ and $T(\xi)$ are random and q is held fixed!

One way of solving SLP problems with recourse is to approximate the objective function by a piecewise linear function — substituting the continuously distributed random vector by a discretely distributed one.

In the second section we shortly review how to generate appropriate discrete distributions $\hat{F}$, $\tilde{F}$, which enable us to evaluate lower and upper bounds for the optimal objective value, see e.g. [1] and [5]. From property v) we know, that the discrete support defines the size of the LP (1.5); in section 3 we determine sequences of discrete distributions $\hat{F}^j$, $\tilde{F}^j$, whose corresponding optimal objective values converge to the original optimal value and whose discrete supports $\hat{\Xi}^j$, $\tilde{\Xi}^j$ increase in such a way, that the total computational work can still be afforded. We shall see, that the algorithm presented in section 3 requires the evaluation of probabilities and conditional expectations, which can be a difficult task for arbitrary distributions. Replacing the given distribution by a discrete one resulting from sampling avoids these difficulties. In the fourth section we discuss how large the sampling size should be taken to obtain a sufficiently accurate confidence interval for the expected recourse cost. In the fifth section numerical examples will illustrate the convergence and the computational effort needed for solving SLP problems with complete fixed recourse.

2. Inner and outer linearization of SLP objectives

Several proposals have been made for approximating linear two-stage problems, see e.g. [1], [3], [4], [5], [8]. In this section we shortly review a particular approximation method on which our algorithm is based. We intend to approximate the expected recourse cost $\tilde{Q}(x) = \int_{\Xi} Q(x, \xi) dF$ by approximating the original distribution F by appropriate discrete distributions $\hat{F}$, $\tilde{F}$ yielding lower and upper bounds for $\tilde{Q}(x)$ — for any feasible x.

We assume that the support Ξ of ξ is compact to ensure the finiteness of $\tilde{Q}(x)$; further we assume that $Q(x, \xi)$ is convex in ξ, which can easily be achieved by holding q fixed.

As we consider not only the original distribution F, but also the approximations $\hat{F}$, $\tilde{F}$, it is meaningful to define:

$$\tilde{Q}(x, F) := \int_{\Xi} Q(x, \xi) dF,$$

$$\Psi(x, F) := \langle c, x \rangle + \int_{\Xi} Q(x, \xi) dF.$$

For obtaining lower and upper bounds for $Q(x, F)$, $\Psi(x, F)$ resp., we look for discrete distributions $\hat{F}$, $\tilde{F}$ for which

$$\tilde{Q}(x, \hat{F}) \leqq \tilde{Q}(x, F) \leqq \tilde{Q}(x, \tilde{F}) \quad \forall x,$$

and hence

$$\Psi(x, \hat{F}) \leqq \Psi(x, F) \leqq \Psi(x, \tilde{F}) \quad \forall x.$$

The supports of $\hat{F}$, $\tilde{F}$ are denoted as $\hat{\Xi}$, $\tilde{\Xi}$, respectively.

Since Ξ is a convex set and $Q(x, \xi)$ is convex in ξ, we get from Jensen's inequality:

$$Q(x, \hat{\xi}) \leqq \int_{\Xi} Q(x, \xi) dF \quad \forall x, \tag{2.1}$$

where $\hat{\xi} = E(\xi / \xi \in \Xi)$.

Considering a given partition $S = (\Xi_1, \ldots, \Xi_L)$ of Ξ into convex cells Ξ_l applying (2.1) on each cell Ξ_l, $l = 1, \ldots, L$, we obtain:

$$\sum_{l=1}^{L} Q(x, \hat{\xi}^l) p^l \leqq \sum_{l=1}^{L} \int_{\Xi_l} Q(x, \xi) dF = \tilde{Q}(x, F) \tag{2.2}$$

where

$$\hat{\xi}^l = E(\xi \mid \xi \in \Xi_l) \qquad \text{and} \quad p^l = P(\xi(\omega) \in \Xi_l).$$

(To avoid unnecessary discussions, we assume $p^l > 0 \; \forall l$, implying $\hat{\xi}^l \neq \hat{\xi}^k$ for $l \neq k$.)

Hence, for the discrete random variable $\hat{\xi}$ defined by

$$\hat{\xi} := \hat{\xi}^l \quad \text{with probability} \quad p^l, \quad l = 1, \ldots, L, \tag{2.3}$$

with distribution $\hat{F}$ and support $\hat{\Xi} = \bigcup_{l=1}^{L} \hat{\Xi}_l = \bigcup_{l=1}^{L} \{\hat{\xi}^l\}$, we get

$$\int_{\hat{\Xi}} Q(x, \hat{\xi}) d\hat{F} = \tilde{Q}(x, \hat{F}) \leqq \tilde{Q}(x, F) \quad \forall x. \tag{2.4}$$

$\tilde{Q}(x, \hat{F})$ approximates the expected recourse costs $\tilde{Q}(x, F)$ piecewise linearly from below — as $\hat{\xi}$ is discrete — and so $\Psi(x, \hat{F}) = \langle c, x \rangle + \tilde{Q}(x, \hat{F})$ can be looked at as an outer linearization for the original objective $\Psi(x, F)$.

To get upper bounds we may use the Edmundson–Madansky inequality (*E–M*). Given a one-dimensional random variable ξ with support $\Xi = [\alpha, \beta]$ and expectation $\hat{\xi} = E(\xi \mid \xi \in \Xi)$, and assuming that $Q(x, \xi)$ is convex in ξ, for the discrete random variable $\tilde{\xi}$ with distribution $\tilde{F}$ and support $\tilde{\Xi} = \{\alpha, \beta\}$ defined by

$$\begin{aligned} \tilde{\xi} &= \alpha \quad \text{with probability} \quad p_1 = \frac{\beta - \hat{\xi}}{\beta - \alpha}, \\ \tilde{\xi} &= \beta \quad \text{with probability} \quad p_2 = \frac{\hat{\xi} - \alpha}{\beta - \alpha}, \end{aligned} \tag{2.5}$$

it follows due to $(E\text{–}M)$ that

$$\tilde{Q}(x, F) = \int_{\Xi} Q(x, \xi) dF \leqq \int_{\tilde{\Xi}} Q(x, \tilde{\xi}) d\tilde{F} = \tilde{Q}(x, \tilde{F}). \tag{2.6}$$

For a K-dimensional random vector ξ with stochastically independent components ξ_j and a K-dimensional interval $\Xi = \bigtimes_{j=1}^{K} [\alpha_j, \beta_j]$ as support (or containing the support), this implies for the discrete random vector $\tilde{\xi}$ with *independent* components, distribution F, and support

$$\tilde{\Xi} = \bigtimes_{j=1}^{K} \{\alpha_j, \beta_j\} = \{\text{all vertices of } \Xi\},$$

defined according to (2.5) by replacing $\tilde{\xi}$, α, β, $\hat{\xi}$ by $\tilde{\xi}_j$, α_j, β_j, $\hat{\xi}_j$, respectively — $\hat{\xi}_j = E(\xi_j | \xi_j \in [\alpha_j, \beta_j])$ — and for any $Q(x, \xi)$ convex in ξ, that

$$\tilde{Q}(x, F) \leqq \tilde{Q}(x, \tilde{F}). \tag{2.7}$$

This follows from (2.6) and the fact that independence allows for iterated integration.

Finally, if the K-dimensional half-open interval Ξ, containing the support of F, is partitioned into half-open subintervals

$$\Xi_l = \bigtimes_{j=1}^{K} [\alpha_j^l, \beta_j^l) \qquad l = 1, \ldots, L,$$

we may construct in an analogous way discrete random vectors $\tilde{\xi}^l$ with *independent* components, distributions $\tilde{F}_l$, and supports

$$\tilde{\Xi}_l = \bigtimes_{j=1}^{K} \{\alpha_j^l, \beta_j^l\} = \{\text{all vertices of } \Xi_l\}$$

by replacing in (2.5) $\tilde{\xi}$, α, β, $\hat{\xi}$ by $\tilde{\xi}_j^l$, α_j^l, β_j^l, $\hat{\xi}_j^l$, respectively — $\hat{\xi}_j^l = E(\xi_j | \xi_j \in [\alpha_j^l, \beta_j^l))$.

Applying $(E\text{–}M)$ to the conditional distribution $F(\xi | \xi \in \Xi_l) =: \tilde{F}_l$ now yields immediately

$$\frac{1}{p^l} \int_{\Xi_l} Q(x, \xi) dF \leqq \int_{\tilde{\Xi}_l} Q(x, \tilde{\xi}^l) d\tilde{F}_l \tag{2.8}$$

where $p^l = P(\xi \in \Xi_l) > 0$ according to (2.2), and therefore

$$\tilde{Q}(x, F) = \sum_{l=1}^{L} \int_{\Xi_l} Q(x, \xi) dF \leqq \sum_{l=1}^{L} p^l \int_{\tilde{\Xi}_l} Q(x, \tilde{\xi}^l) d\tilde{F}_l. \tag{2.9}$$

Hence, we have implicitly found a discrete random vector $\tilde{\xi}$ with support $\tilde{\Xi}$ and distribution $\tilde{F}$,

$$\tilde{\Xi} = \bigcup_{l=1}^{L} \tilde{\Xi}_l, \qquad \tilde{F} = \sum_{l=1}^{L} p^l \tilde{F}_l, \tag{2.10}$$

which due to (2.9) yields an upper bound:

$$\tilde{Q}(x, F) \leqq \tilde{Q}(x, \tilde{F}). \tag{2.11}$$

$\tilde{Q}(x, \tilde{F})$ approximates the expected recourse cost $\tilde{Q}(x, F)$ piecewise linearly from above (since $\tilde{\xi}$ is discrete) and, therefore, $\Psi(x, \tilde{F}) = \langle c, x \rangle + \tilde{Q}(x, \tilde{F})$ is an inner linearization of $\Psi(x, F)$.

For detailed discussions of (*E–M*) bounds as well as further upper bounds we refer to [1], [2], [5].

Remark: If $Q(x, \xi)$ is linear in ξ on $\mathrm{cl}[\Xi_l]$, then obviously

$$\int_{\hat{\Xi}_l} Q(x, \hat{\xi}) d\hat{F} = \int_{\Xi_l} Q(x, \xi) dF = p^l \cdot \int_{\tilde{\Xi}_l} Q(x, \tilde{\xi}^l) dF_l. \tag{2.12}$$

3. The solution method

For solving a linear two-stage problem we will use an important feature of Jensen's and Edmundson–Madansky's inequalities, namely their monotonicity: Considering partitions S^j into half-open intervals Ξ_l^j, $l = 1, \ldots, L_j$, if S^2 is a refinement of the partition S^1 (i.e. S^2 results from S^1 by partitioning some of its cells Ξ_l^1), then the lower and the upper bound obtained for S^2 are at least as sharp as, the previous ones.

The simplest and most obvious technique of refining is to subdivide each cell Ξ_l^1, $l = 1, \ldots, L_1$, by hyperplanes orthogonal to all coordinate axes in $\mathbf{R}^K$. If the cells Ξ_l^1 are (half-open) hypercubes in $\mathbf{R}^K$, this strategy divides each Ξ_l^1 into 2^K smaller cubes, hence after j steps we shall get $2^{K \cdot j}$ cells. The discrete supports $\hat{\Xi}^j$, $\tilde{\Xi}^j$ of our discrete random variables $\hat{\xi}^j$, $\tilde{\xi}^j$ — subject to partition S^j — strongly increase with respect to their number of realizations; hence after a small number of refining steps we shall no longer be able to solve the approximating LP's (1.5) corresponding to the distributions $\hat{F}^j$, $\tilde{F}^j$ introduced in Section 2 before.

However, a more careful analysis of our problem shows that the computational effort can be considerably reduced by dividing only some properly chosen subcells along appropriate directions.

We start from a certain partition $S^j := \{\Xi_l^j, l = 1, \ldots, L_j\}$ of Ξ into half-open intervals. The discrete random variables $\hat{\xi}^j$, $\tilde{\xi}^j$ — with their supports $\hat{\Xi}^j$, $\tilde{\Xi}^j$ and discrete distributions $\hat{F}^j$, $\tilde{F}^j$ — approximate the original random variable ξ, and we already know that

$$\Psi(x, \hat{F}^j) \leqq \Psi(x, F) \leqq \Psi(x, \tilde{F}^j).$$

Hence a solution $\hat{x}^j$ of

$$\min_{x \in X} \Psi(x, \hat{F}^j) \tag{3.1}$$

(i.e. a LP of type (1.5)), yields a lower bound $\Psi(\hat{x}^j, \hat{F}^j)$ for

$$\min_{x \in X} \Psi(x, F) =: \Psi(x^*, F).$$

Similarly

$$\min_{x \in X} \Psi(x, \tilde{F}^j) \tag{3.2}$$

(i.e. a LP of type (1.5)), yields a solution $\tilde{x}^j$ and hence — in some sense — a minimal upper bound $\Psi(\tilde{x}^j, \tilde{F}^j)$ for $\Psi(x^*, F)$.

We have to observe that determining $\hat{x}^j$ requires much less effort than finding $\tilde{x}^j$, since the number of blocks in the LP (1.5) corresponding to $\tilde{F}^j$ — i.e. the number of vertices in all Ξ_l^j — is a multiple of the number of blocks in the LP (1.5) corresponding to $\hat{F}^j$ — i.e. the number of conditional expectations $\hat{\xi}^{j,l}$ of all Ξ_l^j.

There is another way, that yields an upper bound for $\Psi(x^*, F)$, assuming that $\hat{x}^j$ has already been determined. Since $\Psi(x, \tilde{F}^j)$ is an inner linearization of $\Psi(x, F)$, we obtain an upper bound by $\Psi(\hat{x}^j, \tilde{F}^j)$ which is in general less sharp than $\Psi(\tilde{x}^j, \tilde{F}^j)$ but available with much less effort.

To determine $\Psi(\hat{x}^j, \tilde{F}^j)$ we have to solve the recourse problem

$$\begin{aligned} Q(\hat{x}^j, \tilde{\xi}^{j,\nu}) := \min \ & \langle q, y \rangle \\ \text{s.t.} \ & Wy = h(\tilde{\xi}^{j,\nu}) - T(\tilde{\xi}^{j,\nu})\hat{x}^j \\ & y \geqq 0 \end{aligned} \tag{3.3}$$

for all $\tilde{\xi}^{j,\nu} \in \tilde{\Xi}^j$.

So, instead of solving one large LP (1.5) for x^j and y_ν^j we have to solve all much smaller subproblems associated with the angular blocks.

Hence we get:

$$\Psi(\hat{x}^j, \hat{F}^j) \leqq \Psi(x^*, F) \leqq \Psi(\tilde{x}^j, \tilde{F}^j) \leqq \Psi(\hat{x}^j, \tilde{F}^j)$$

where

$\Psi(\hat{x}^j, \hat{F}^j)$ results from (3.1) and

$\Psi(\hat{x}^j, \tilde{F}^j) = \langle c, \hat{x}^j \rangle + \int_{\tilde{\Xi}^j} Q(\hat{x}^j, \tilde{\xi}) d\tilde{F}^j$ from (3.3).

There is another fact, which turns out to be an essential advantage, if we consider $\Psi(\cdot, \tilde{F}^j)$ at $\hat{x}^j$. From (2.2), (2.3) and (2.9) we have

$$\Psi(\hat{x}^j, \hat{F}^j) = \langle c, \hat{x}^j \rangle + \sum_{l=1}^{L_j} \int_{\hat{\Xi}_l^j} Q(\hat{x}^j, \hat{\xi}) d\hat{F}^j, \tag{3.4}$$

$$\Psi(\hat{x}^j, \tilde{F}^j) = \langle c, \hat{x}^j \rangle + \sum_{l=1}^{L_j} p^{j,l} \int_{\tilde{\Xi}_l^j} Q(\hat{x}^j, \xi) d\tilde{F}_l^j, \tag{3.5}$$

with

$$p^{j,l} = p(\xi \in \Xi_l^j).$$

So, we actually have lower and upper bounds for the expected costs for the same approximate solution $\hat{x}^j$, subject to the cells Ξ_l^j, $l = 1, \ldots, L_j$:

$$\Psi_l(\hat{x}^j, \hat{F}^j) := \frac{1}{p^{j,l}} \int_{\hat{\Xi}_l^j} Q(\hat{x}^j, \hat{\xi}) d\hat{F}^j =$$

$$= Q(\hat{x}^j, \hat{\xi}^{j,l}) \leqq \frac{1}{p^{j,l}} \int_{\hat{\Xi}_l^j} Q(\hat{x}^j, \xi) dF \leqq \tag{3.6}$$

$$\leqq \int_{\tilde{\Xi}_l} Q(\hat{x}^j, \tilde{\xi}) d\tilde{F}_l^j =: \Psi_l(\hat{x}^j, \tilde{F}^j).$$

If it turns out that for some $\Xi_{l_0}^j$

$$\Psi_{l_0}(\hat{x}^j, \tilde{F}^j) - \Psi_{l_0}(\hat{x}^j, \hat{F}^j) > \varepsilon \cdot \Psi_{l_0}(\hat{x}^j, \hat{F}^j), \text{ (see } ^1) \tag{3.7}$$

then $\Xi_{l_0}^j$ is not approximated sufficiently accurate with respect to the tolerance ε (relative error); so we have to improve the approximation by partitioning this cell $\Xi_{l_0}^j$.

There exists another criteria for further subdividing, proposed by H. Gassmann:

$$\Psi_{l_0}(\hat{x}^j, \tilde{F}^j) - \Psi_{l_0}(\hat{x}^j, \hat{F}^j) > \frac{\varepsilon}{p^{j,l_0} \cdot 2^{\mu}} \cdot M \tag{3.8}$$

where μ denotes the number of divisions that were necessary to obtain $\Xi_{l_0}^j$, p^{j,l_0} the probability of $\Xi_{l_0}^j$, and

M is some over all lower bound for $\Psi(x^*, F)$, e.g. $M = \Psi(\hat{x}^j, \hat{F}^j)$.

The practical meaning of (3.8) is the following: cells Ξ_l^j, for which $p^{j,l_0} \cdot 2^{\mu} > 1$, are expected to be more important than cells for which $p^{j,l_0} \cdot 2^{\mu} < 1$.

The set T^j of cells Ξ_l^j, which have to be partitioned, can now be defined by:

$$T^j := \{\Xi_l^j / (3.7), (3.8) \text{ resp., is satisfied}\}. \tag{3.9}$$

Remark: Performing (3.1) and (3.2) would yield similarly as lower and upper bounds

$$\Psi_l(\hat{x}^j, \hat{F}^j) \quad \text{and} \quad \Psi_l(\tilde{x}^j, \tilde{F}^j), \quad \text{respectively,}$$

but as such bounds correspond to different decisions $\hat{x}^j$ and $\tilde{x}^j$, it would be not significant to compare them according to (3.7) or (3.8). Assume that the functions

[1] Here we assume $\Psi_{l_0}(\hat{x}^j, \hat{F}^j) > 0$; otherwise this relation must either be completed by taking into account sign $[\Psi_{l_0}(\hat{x}^j, \tilde{F}^j)]$ or replaced by considering just the absolut error ε.

$Q(\hat{x}^j, \cdot)$ and $Q(\tilde{x}^j, \cdot)$ are linear on a cell Ξ_l^j, then we know that we do not achieve any improvement by partitioning this cell, independent of how much the lower and upper bound differ. Indeed, computational results justified to consider $\Psi(\cdot, \tilde{F}^j)$ at $\hat{x}^j$, as we will see in section 4.

Refining strategies

So far, (3.7) or (3.8) enables us to determine the set of cells T^j, which have to be subdivided to improve the approximation; but we still do not know how to choose the dividing hyperplane or cut. The efficiency of cuts in different directions depends on the nonlinearity of the function $Q(\hat{x}^j, \xi)$ with respect to the coordinates $\xi_1, \ldots, \xi_K$ of ξ. As we see from the example in Fig. 3.1, no improvement can be achieved by splitting Ξ_l with a cut orthogonal to ξ_2 since $Q(\hat{x}^j, \cdot)$ is linear in ξ_2. On the other hand, if we cut Ξ_l by a hyperplane orthogonal to ξ_1, we may obtain two subcells, in which $Q(\hat{x}^j, \xi)$ will be linear in ξ, and our next upper and lower bound, subject to these subcells, might become exact: Hence it seems plausible to choose for each $\Xi_l^j \in T^j$ that coordinate along which $Q(\hat{x}^j, \cdot)$ is "mostly nonlinear".

We consider an arbitrary cell $\Xi_l^j \in T^j$ and define $\mathscr{C}^l$ as the set of pairs of adjacent vertices $(\xi_{v_1}^l, \xi_{v_2}^l)$ — i.e. $\xi_{v_1}^l$ and $\xi_{v_2}^l$ differ in exactly one coordinate:

$$\mathscr{C}^l := \{(\xi_{v_1}^l, \xi_{v_2}^l) \mid \xi_{v_1}^l, \xi_{v_2}^l \in \tilde{\Xi}_l^j \text{ and adjacent}\}.$$

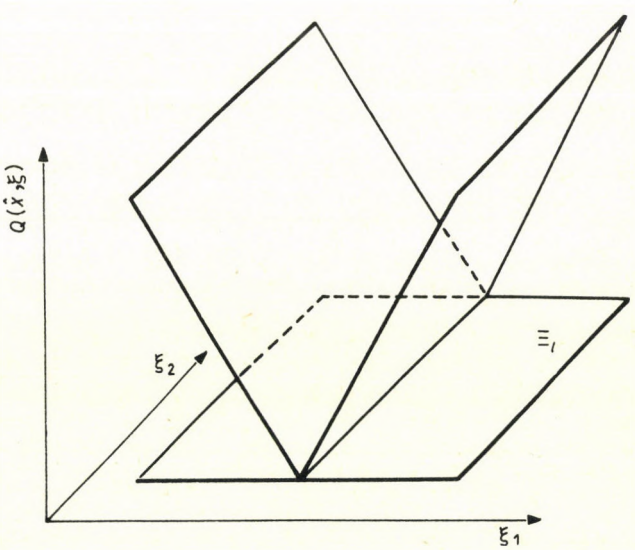


Fig. 3.1

Similarly, we define:

$$\mathscr{C}_k^l := \{(\xi_{v_1}^l, \xi_{v_2}^l) \mid \xi_{v_1}^l \in \tilde{\Xi}_l^j \text{ and } \xi_{v_2}^l \in \tilde{\Xi}_l^j \text{ differ exactly in the } k\text{th coordinate}\}.$$

So, we have:

$$\mathscr{C}^l = \bigcup_{k=1}^{K} \mathscr{C}_k^l.$$

For evaluating $\Psi(\hat{x}^j, \tilde{F}^j)$ we already solved (see (3.3))

$$Q(\hat{x}^j, \xi_v^l) = \min \langle q, y \rangle \quad \text{s.t. } Wy = h(\xi_v^l) - T(\xi_v^l)\hat{x}^j, \quad y \geqq 0 \tag{3.10}$$

$$\text{for all } \xi_v^l \in \tilde{\Xi}_l^j,\ v = 1, \ldots, 2^K;\ l = 1, \ldots, L_j.$$

From duality theory we know that the dual solution $\pi'(\xi_v^l)$ of (3.10) is a measure of sensitivity of $Q(\hat{x}^j, \cdot)$ with respect to the right-hand side of (3.10).

If the multipliers $\pi(\xi_v^l)$ coincide at each vertex of Ξ_l^j, $Q(\hat{x}^j, \cdot)$ is linear on Ξ_l^j. But in this case $\Xi_l^j \notin T^j$ as the already obtained upper and lower bound are equal and represent the exact expected recourse cost on Ξ_l^j for the decision $\hat{x}^j$.

Let us consider the pair of multipliers $(\pi(\xi_{v_1}^l), \pi(\xi_{v_2}^l))$ corresponding to $(\xi_{v_1}^l, \xi_{v_2}^l) \in \mathscr{C}_k^l$:

If

$$\pi(\xi_{v_1}^l) = \pi(\xi_{v_2}^l) \qquad \forall (\xi_{v_1}^l, \xi_{v_2}^l) \in \mathscr{C}_k^l,$$

then $Q(\hat{x}^j, \cdot)$ is linear on all edges $[\xi_{v_1}^l, \xi_{v_2}^l]$ such that $(\xi_{v_1}^l, \xi_{v_2}^l) \in \mathscr{C}_k^l$.

This does not necessarily imply the linearity of $Q(\hat{x}^j, \cdot)$ in Ξ_l^j; but we may expect that a cut orthogonal to coordinate k is not efficient, since linearity of $Q(\hat{x}^j, \cdot)$ occurs on all edges of Ξ_l^j orthogonal to coordinate k (see Fig. 3.1 for $k = 2$).

Now, we would be able to define a partition-coordinate k by the following rule:

> We determine out of all adjacent vertices that pair of multipliers $(\pi(\xi_{v_1}^l), \pi(\xi_{v_2}^l))$ which differ most with respect to $\|\cdot\|_2$ and, if $(\xi_{v_1}^l, \xi_{v_2}^l) \in \mathscr{C}_k^l$, define that coordinate k as partition-coordinate. (3.11)

Still some improvements of (3.11) can be expected, if we look at Fig. 3.2a and Fig. 3.2b:

In Fig. 3.2a, obviously no essential improvement can be achieved by division of coordinate k_1 in $[\xi_1^l, \xi_2^l]$. It seems to be more efficient to divide coordinate k_2 in $[\xi_3^l, \xi_4^l]$ — illustrated in Fig. 3.2b — although the slopes, determined by the multipliers, coincide in ξ_1^l and ξ_3^l as well as in ξ_2^l and ξ_4^l.

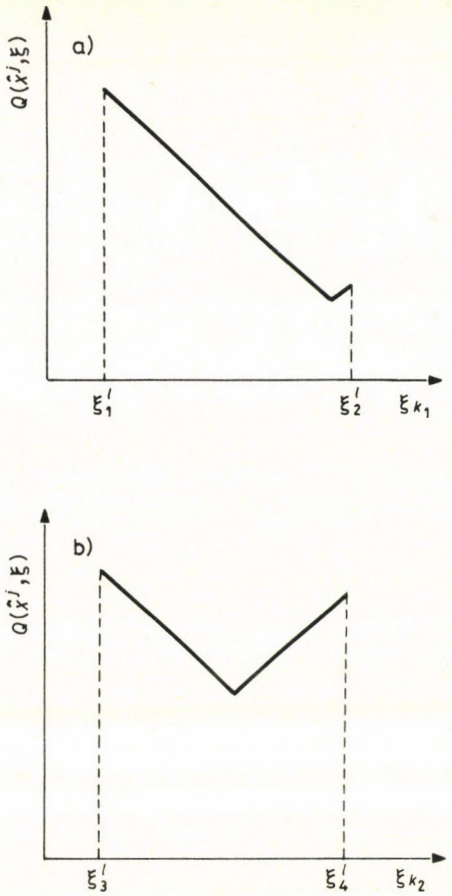


Fig. 3.2

To take into consideration this possible outcome, we observe that all multipliers $\pi(\xi_{\nu_i}^l)$ are dual feasible for (3.10) with arbitrary $\xi_\nu^l \in \tilde{\Xi}_l^j$.

Hence we get:

$$\langle \pi(\xi_{\nu_i}^l), h(\xi_\nu^l) - T(\xi_\nu^l) \cdot \hat{x}^j \rangle \leqq Q(\hat{x}^j, \xi_\nu^l) \quad \forall\, \xi_{\nu_i}^l, \xi_\nu^l \in \tilde{\Xi}_l^j. \tag{3.12}$$

For any pair of adjacent vertices $(\xi_{\nu_1}^l, \xi_{\nu_2}^l) \in \mathscr{C}^l$ it follows:

$$\langle \pi(\xi_{\nu_1}^l), h(\xi_{\nu_2}^l) - T(\xi_{\nu_2}^l) \cdot \hat{x}^j \rangle \leqq Q(\hat{x}^j, \xi_{\nu_2}^l), \tag{3.13.1}$$

$$\langle \pi(\xi_{\nu_2}^l), h(\xi_{\nu_1}^l) - T(\xi_{\nu_1}^l) \cdot \hat{x}^j \rangle \leqq Q(\hat{x}^j, \xi_{\nu_1}^l). \tag{3.13.2}$$

If equality holds either in (3.13.1) or (3.13.2) we have linearity of $Q(\hat{x}^j, \cdot)$ in $[\xi_{\nu_1}^l, \xi_{\nu_2}^l]$; nevertheless $\pi(\xi_{\nu_1}^l)$ and $\pi(\xi_{\nu_2}^l)$ may differ, e.g. in the case of degeneracy.

Defining now

$$\Delta(\xi^l_{\nu_1}, \xi^l_{\nu_2}) := \min\,[Q(\hat{x}^j, \xi^l_{\nu_2}) - \langle \pi(\xi^l_{\nu_1}), h(\xi^l_{\nu_2}) - T(\xi^l_{\nu_2}) \cdot \hat{x}^j \rangle, \quad (3.14)$$
$$Q(\hat{x}^j, \xi^l_{\nu_1}) - \langle \pi(\xi^l_{\nu_2}), h(\xi^l_{\nu_1}) - T(\xi^l_{\nu_1}) \cdot \hat{x}^j \rangle],$$

we may understand $\Delta(\xi^l_{\nu_1}, \xi^l_{\nu_2})$ as a measure of nonlinearity of $Q(\hat{x}^j, \cdot)$ in $[\xi^l_{\nu_1}, \xi^l_{\nu_2}]$.

Evaluating

$$\max_{(\xi^l_{\nu_1}, \xi^l_{\nu_2}) \in \mathscr{C}^l} \Delta(\xi^l_{\nu_1}, \xi^l_{\nu_2}) =: \Delta(\xi^l_{\mu_1}, \xi^l_{\mu_2}), \quad \forall\, \Xi^j_l \in T^j, \quad (3.15)$$

we obtain those edges $[\xi^l_{\mu_1}, \xi^l_{\mu_2}]$ of $\Xi^j_l \in T^j$ on which $Q(\hat{x}^j, \cdot)$ is "mostly nonlinear". We choose that coordinate k for subdividing Ξ^j_l, for which $(\xi^l_{\mu_1}, \xi^l_{\mu_2}) \in \mathscr{C}^l_k$. It proves to be meaningful to cut at $\hat{\xi}^l_k = E(\xi^l_k \mid \xi_k \in [\xi^l_{\mu_1}, \xi^l_{\mu_2}])$ and so the cutting plane cp^l — subject to Ξ^j_l — is given by

$$cp^l := \{\xi \mid \xi_k = \hat{\xi}^l_k,\ \xi \in \Xi^j_l\}. \quad (3.16)$$

What actually is done by (3.14) and (3.15) is illustrated in Fig 3.3.:

In Fig. 3.3a we have:

$$g_1 := Q(\hat{x}^j, \xi^l_1) = \langle \pi(\xi^l_1), h(\xi^l_1) - T(\xi^l_1) \cdot \hat{x}^j \rangle,$$
$$\bar{g}_1 := \langle \pi(\xi^l_2), h(\xi^l_1) - T(\xi^l_1) \cdot \hat{x}^j \rangle;$$
$$g_2 := Q(\hat{x}^j, \xi^l_2) = \langle \pi(\xi^l_2), h(\xi^l_2) - T(\xi^l_2) \cdot \hat{x}^j \rangle,$$
$$\bar{g}_2 := \langle \pi(\xi^l_1), h(\xi^l_2) - T(\xi^l_2) \cdot \hat{x}^j \rangle$$

and hence $\Delta(\xi^l_1, \xi^l_2) = g_2 - \bar{g}_2$.

Analogously, from Fig. 3.3b we get

$$\Delta(\xi^l_3, \xi^l_4) = g_4 - \bar{g}_4.$$

Applying (3.15) we get $\Delta(\xi^l_3, \xi^l_4)$ and, therefore, the coordinate k_2, whose partition lets expect greatest improvement in cell Ξ^j_l.

Performing (3.7), (3.8), resp., and (3.14)–(3.16) we obtain a new partition S^{j+1} which is — by construction — a refinement of S^j.

Now we can summarize the procedure as follows:

Algorithm

STEP 1: $j := 1$, $L_1 := 1$, $S^1 = \{\Xi^1_l\} = \{\Xi\}$; (3.17.1)

STEP 2: determine $\hat{\Xi}^j$, $\hat{\xi}^j$, $\hat{F}^j$ according to (2.3) (3.17.2)
and $\tilde{\Xi}^j$, $\tilde{\xi}^j$, $\tilde{F}^j$ according to (2.10);

STEP 3: solve (3.1) to get $\hat{x}^j$ and the lower bound $\Psi(\hat{x}^j, \hat{F}^j)$; (3.17.3)
solve (3.3) to get the upper bound $\Psi(\hat{x}^j, \tilde{F}^j)$;

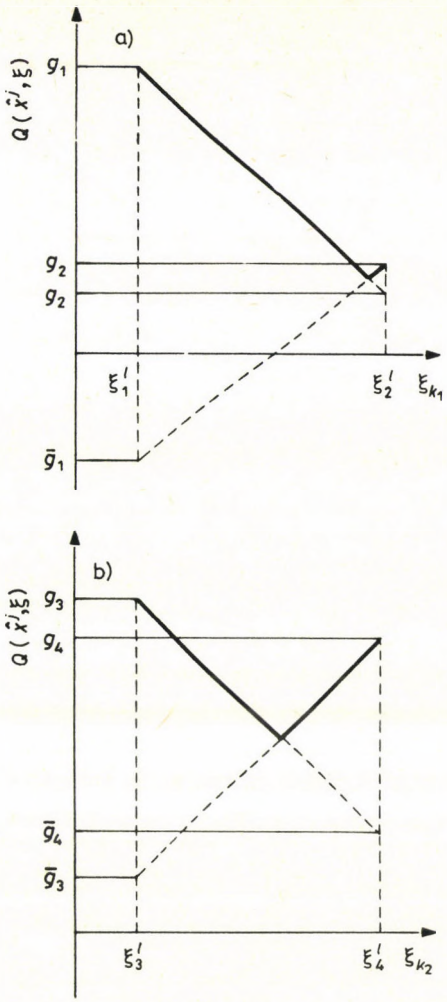


Fig. 3.3

if $\Psi(\hat{x}^j, \tilde{F}^j) - \Psi(\hat{x}^j, \hat{F}^j) \leqq \varepsilon \cdot \Psi(\hat{x}^j, \hat{F}^j) \Rightarrow \hat{x}^j$ is sufficiently accurate, STOP;
if not, evaluate $\Psi_l(\hat{x}^j, \tilde{F}^j)$, $\Psi_l(\hat{x}^j, \hat{F}^j)$ according to (3.6);

STEP 4: determine T^j according to (3.9); (3.17.4)
evaluate the partition coordinates according to (3.14)–(3.15)
and determine the cutting planes cp^l according to (3.16),
for all $\Xi_l^j \in T^j$.

STEP 5: determine S^{j+1}, L_{j+1};
$j := j+1$; go to STEP 2. (3.17.5)

This procedure is characterized by the following feature: In cells $\Xi_l^j \notin T^j$, subject to which $\Psi(\hat{x}^j, F)$ is approximated sufficiently well, we do not strive for a better approximation for ξ; in such cells it is sufficient to consider the realization $\hat{\xi}^{j,l}$ as a representative of Ξ_l^j.

The error, obtained by substituting ξ by $\hat{\xi}^j$, $\tilde{\xi}^j$ is — at least for the decision $\hat{x}^j$ — localized in step 4, by determining T^j, the set of cells, in which ξ may be approximated quite well by $\hat{\xi}^j$, $\tilde{\xi}^j$, but not sufficiently well with respect to the behaviour of the function $Q(\hat{x}^j, \cdot)$. Roughly speaking, we enforce the convergence of $\hat{\xi}^j$, $\tilde{\xi}^j$ in probability to ξ only in the cells $\Xi_l^j \in T^j$.

4. Considering arbitrary distributions F

Step 2 in the above-mentioned procedure requires to evaluate probabilities of half-open intervals Ξ_l and conditional expectations $E(\xi | \xi \in \Xi_l)$ which can be a difficult task for arbitrary distributions. Replacing the given distribution by a discrete one, resulting from sampling in Ξ will avoid these difficulties. In doing so an additional error occurs. The question we are discussing in this section is, how large the sample should be chosen, to keep the error negligible as compared to the tolerance ε we prescribe for the optimal value of the SLP problem.

If $(\xi_i, i = 1, \ldots, N)$ denotes the sample of size N, we get

$$\Psi(x, F) \approx \langle c, x \rangle + \frac{1}{N} \cdot \sum_{i=1}^{N} Q(x, \xi_i),$$

$$I^N(x) := \frac{1}{N} \sum_{i=1}^{N} Q(x, \xi_i) \approx \int_{\Xi} Q(x, \xi) dF =: I(x). \tag{4.1}$$

Applying the central limit theorem, we obtain

$$P\left(|I^N(x) - I(x)| \leqq \frac{\lambda \sigma(x)}{\sqrt{N}} \right) = \Phi(\lambda) - \Phi(-\lambda) + Q\left(\frac{1}{\sqrt{N}} \right) \tag{4.2}$$

where $\sigma(x)$ is the standard deviation of $Q(x, \xi)$, i.e.

$$[\sigma(x)]^2 = \int_{\Xi} Q^2(x, \xi) dF - [I(x)]^2, \tag{4.3}$$

and $\Phi(\lambda) - \Phi(-\lambda)$ is the area below the density of the Gaussian Standard-Normal distribution in the interval $[-\lambda, \lambda]$.

So, λ gives us the probability, with which $I^N(x)$ is in the confidence interval $CI_\lambda(x)$,

$$CI_\lambda(x) := \left[I(x) - \frac{\lambda \sigma(x)}{\sqrt{N}}, I(x) + \frac{\lambda \sigma(x)}{\sqrt{N}} \right].$$

For a given probability $[\Phi(\lambda)-\Phi(-\lambda)]$ the length of $CI_\lambda(x)$ depends on $\sigma(x)$ and on N. As $\sigma(x)$ is determined by $Q(x,\cdot)$, the only possibility for keeping $CI_\lambda(x)$ small enough, is to choose the sampling size N properly large; to do so, it requires some information about $\sigma(x)$.

Of course, we do not know $\sigma^2(x)$ exactly but we could estimate it by the sampling variance $\hat{\sigma}^2(x)$,

$$\sigma^2(x) \approx \hat{\sigma}^2(x) := \frac{1}{N}\sum_i Q^2(x,\xi_i) - [I^N(x)]^2, \quad \text{(see [2])} \tag{4.4}$$

if we knew the values of $Q(x,\cdot)$ at all realizations ξ_i, $i=1,\ldots,N$, of the sampling.

To avoid these (many) evaluations we strive for an upper bound $\tilde{\sigma}(x)$ of $\hat{\sigma}(x)$ based on information we have from our algorithm, i.e. the values of $Q(x,\cdot)$ at the vertices and the conditional expectations of the cells Ξ_l^j.

Replacing ξ by the "sample variable" η with the distribution $P(\eta=\xi_i)=\frac{1}{N}$, $i=1,\ldots,N$, and applying the algorithm (3.17.1)–(3.17.5) for η instead of ξ we obtain in the j-th iteration step:

$S^j=\{\Xi_1^j,\ldots,\Xi_{L_j}^j\}$;
$\hat{\eta}^l = E(\eta/\eta\in\Xi_l^j)$, $\quad \hat{\Xi}^j = \bigcup_{l=1}^{L_j}\{\hat{\eta}^l\}$,
$p^l = P(\eta\in\Xi_l^j)$;
$\tilde{\eta}$ with support $\tilde{\Xi}^j$ and
$\tilde{p}^\mu$, the probability $P(\tilde{\eta}=\tilde{\eta}^\mu)$,
$\tilde{\eta}^\mu\in\tilde{\Xi}^j$ (according to (2.10));
$\tilde{\eta}_\nu^l\in\tilde{\Xi}_l^j$, $\nu=1,\ldots,2^K$, the vertices of Ξ_l^j,
$\tilde{p}_\nu^l$, the probability for $\tilde{\eta}=\tilde{\eta}_\nu^l$ according to $\tilde{F}_l$.

Further we get (see (2.2) and (2.9)):

$$L^N(\hat{x}^j) := \sum_{l=1}^{L_j} Q(\hat{x}^j,\hat{\eta}^l)\cdot p^l =: \sum_{l=1}^{L_j} L_l^N(\hat{x}^j), \tag{4.5}$$

$$U^N(\hat{x}^j) := \sum_{l=1}^{L_j}\left(\sum_{\nu=1}^{2^K} Q(\hat{x}^j,\hat{\eta}_\nu^l)\right)\cdot \tilde{p}_\nu^l\cdot p^l =: \sum_{l=1}^{L_j} U_l^N(\hat{x}^j). \tag{4.6}$$

Obviously we have:

$$L^N(\hat{x}^j) \leqq I^N(\hat{x}^j) \leqq U^N(\hat{x}^j).$$

[2] We omit the factor $\frac{N}{N-1}$, since N is supposed to be large.

As we know the dual solution $\pi(\hat{\eta}^l)$ of the recourse problem subject to $\hat{\eta}^l$ and $\hat{x}^j$, we are able to approximate from below $Q(\hat{x}^j, \eta)$ on Ξ_l^j by a linear function $\hat{Q}_l(\hat{x}^j, \eta)$, namely

$$\hat{Q}_l(\hat{x}^j, \eta) := \langle \pi(\hat{\eta}^l), h(\eta) - T(\eta) \cdot \hat{x}^j \rangle \quad \forall \eta \in \Xi_l^j,$$

for which follows (see Fig. 4.1):

$$\hat{Q}_l(\hat{x}^j, \eta) \leqq Q(\hat{x}^j, \eta) \qquad \forall \eta \in \Xi_l^j. \tag{4.7}$$

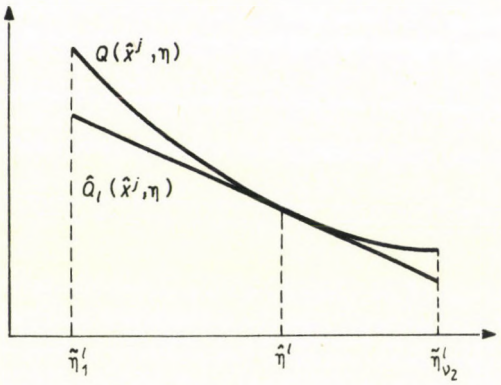


Fig. 4.1

It is easy to verify that

$$\hat{L}^N(\hat{x}^j) := \frac{1}{N} \sum_{i=1}^{N} \hat{Q}(\hat{x}^j, \xi_i) = L^N(\hat{x}^j) \tag{4.8}$$

where

$$\hat{Q}(\hat{x}^j, \xi_i) = \hat{Q}_l(\hat{x}^j, \xi_i) \text{ for } \xi_i \in \Xi_l^j.$$

We consider now

$$\operatorname{Var}(\hat{Q}(\hat{x}^j, \eta)) := \frac{1}{N} \sum_i \hat{Q}^2(\hat{x}^j, \xi_i) - \left[\frac{1}{N} \sum_i \hat{Q}(\hat{x}^j, \xi_i) \right]^2. \tag{4.9}$$

Bounding the terms $|\hat{\sigma}^2(\hat{x}^j) - \operatorname{Var}(\hat{Q}(\hat{x}^j, \eta))|$ and $\operatorname{Var}(\hat{Q}(\hat{x}^j, \eta)$ from above yields an upper bound for $\hat{\sigma}^2(\hat{x}^j)$ and hence for $\hat{\sigma}(\hat{x}^j)$ as

$$\hat{\sigma}^2(\hat{x}^j) \leqq |\hat{\sigma}^2(\hat{x}^j) - \operatorname{Var}(\hat{Q}(\hat{x}^j, \eta))| + \operatorname{Var}(\hat{Q}(\hat{x}^j, \eta)). \tag{4.10}$$

From (4.4), (4.8) and (4.9) and observing that adding a constant M to a random

variable — in our case $Q(\hat{x}^j, \eta)$ and $\hat{Q}(\hat{x}^j, \eta)$ — does not affect the variance, we get

$$\hat{\sigma}^2(\hat{x}^j) - \mathrm{Var}\,(\hat{Q}(\hat{x}^j, \eta)) = \frac{1}{N} \sum_{i=1}^{N} \{[Q(\hat{x}^j, \xi_i) + M]^2 - [\hat{Q}(\hat{x}^j, \xi_i) + M]^2\} -$$

$$- [I^N(\hat{x}^j) + M]^2 + [L^N(\hat{x}^j) + M]^2 =$$

$$= \frac{1}{N} \sum_{l=1}^{L_j} \sum_{\xi_i \in \Xi_l^j} \{Q(\hat{x}^j, \xi_i) - \hat{Q}_l(\hat{x}^j, \xi_i)\} \{Q(\hat{x}^j, \xi_i) + \hat{Q}_l(\hat{x}^j, \xi_i) + 2M\} -$$

$$- [I^N(\hat{x}^j)]^2 + [L^N(\hat{x}^j)]^2 + 2M[L^N(\hat{x}^j) - I^N(\hat{x}^j)]. \tag{4.11}$$

For

$$\rho = \min_l \min_{\tilde{\eta}_\nu^l \in \tilde{\Xi}_l^j} \hat{Q}_l(\hat{x}^j, \tilde{\eta}_\nu^l)$$

and

$$M = \begin{cases} |\rho| & \text{if } \rho < 0 \\ 0 & \text{else} \end{cases}, \tag{4.12}$$

it is obvious from (4.7) that

$$\{Q(\hat{x}^j, \xi_i) - \hat{Q}_l(\hat{x}^j, \xi_i)\} \geqq 0$$

and

$$\{Q(\hat{x}^j, \xi_i) + \hat{Q}_l(\hat{x}^j, \xi_i) + 2M\} \geqq 0 \qquad \forall\, \xi_i \in \Xi_l^j \quad \forall\, l.$$

Therefore, we conclude from (4.11)

$$|\hat{\sigma}(\hat{x}^j) - \mathrm{Var}\,(\hat{Q}(\hat{x}^j, \eta))| \leqq$$

$$\leqq \frac{1}{N} \sum_{l=1}^{L_j} \sum_{\xi_i \in \Xi_l^j} \{Q(\hat{x}^j, \xi_i) - \hat{Q}_l(\hat{x}^j, \xi_i)\} \{Q(\hat{x}^j, \xi_i) + \hat{Q}_l(\hat{x}^j, \xi_i) + 2M\} +$$

$$+ |[I^N(\hat{x}^j)]^2 - [L^N(\hat{x}^j)]^2| + 2M[I^N(\hat{x}^j) - L^N(\hat{x}^j)]. \tag{4.13}$$

The first term herein can be bounded from above by bounding from above the factors of all products, since they are positive.

$\{Q(\hat{x}^j, \eta) - \hat{Q}_l(\hat{x}^j, \eta)\}$ being a convex function on Ξ_l^j, it attains its maximum in a vertex of Ξ_l^j. Hence, we define

$$\gamma_{li} := \max_{\tilde{\eta}_\nu^l \in \tilde{\Xi}_l^j} \{Q(\hat{x}^j, \tilde{\eta}_\nu^l) - \hat{Q}_l(\hat{x}^j, \tilde{\eta}_\nu^l)\}$$

as upper bound for

$$\{Q(\hat{x}^j, \xi_i) - \hat{Q}_l(\hat{x}^j, \xi_i)\},\ \xi_i \in \Xi_l^j.$$

Then the first term in (4.13) is bounded from above by

$$\frac{1}{N}\sum_{l=1}^{L_j}\gamma_l \sum_{\xi_i\in\Xi_l^j}\{Q(\hat{x}^j,\xi_i)+\hat{Q}_l(\hat{x}^j,\xi_i)+2M\}\leqq$$

$$\leqq \sum_{l=1}^{L_j}\gamma_l\{U_l^N(\hat{x}^j)+L_l^N(\hat{x}^j)\}+2M\sum_{l=1}^{L_j}\gamma_l\cdot p^l.$$

Further, we have for the second term of (4.13)

$$\{I^N(\hat{x}^j)-L^N(\hat{x}^j)\}\cdot|I^N(\hat{x}^j)+L^N(\hat{x}^j)|\leqq$$

$$\leqq\{U^N(\hat{x}^j)-L^N(\hat{x}^j)\}\cdot|I^N(\hat{x}^j)+L^N(\hat{x}^j)|\leqq$$

$$\leqq\{U^N(\hat{x}^j)-L^N(\hat{x}^j)\}\{|I^N(\hat{x}^j)|+|L^N(\hat{x}^j)|\}\leqq$$

$$\leqq\{U^N(\hat{x}^j)-L^N(\hat{x}^j)\}\cdot\{\max[|L^N(\hat{x}^j)|,\ |U^N(\hat{x}^j)|]+|L^N(\hat{x}^j)|\},$$

and for the third term

$$2M\{I^N(\hat{x}^j)-L^N(\hat{x}^j)\}\leqq 2M\{U^N(\hat{x}^j)-L^N(\hat{x}^j)\}.$$

Putting all this together we get

$$|\sigma^2(\hat{x}^j)-\mathrm{Var}\,(\hat{Q}(\hat{x}^j,\eta))|\leqq\sum_{l=1}^{L_j}\gamma_l\{U_l^N(\hat{x}^j)+L_l^N(\hat{x}^j)\}+2M\sum_{l=1}^{L_j}\gamma_l p^l+$$

$$+\{U^N(\hat{x}^j)-L^N(\hat{x}^j)\}\cdot\{\max[|L^N(\hat{x}^j)|,\ |U^N(\hat{x}^j)]+L^N(\hat{x}^j)|\}+$$

$$+2M\cdot\{U^N(\hat{x}^j)-L^N(\hat{x}^j)\}. \tag{4.14}$$

So, it remains to estimate $\mathrm{Var}\,(\hat{Q}(\hat{x}^j,\eta))$; from (4.8) and (4.9), we get:

$$\mathrm{Var}\,(\hat{Q}(\hat{x}^j,\eta))=\frac{1}{N}\sum_i\hat{Q}^2(\hat{x}^j,\xi_i)-[L^N(\hat{x}^j)]^2. \tag{4.15}$$

As $\hat{Q}(\hat{x}^j,\eta)=\hat{Q}_l(\hat{x}^j,\eta)\ \forall\,\eta\in\Xi_l^j$ and so is linear in Ξ_l^j, $\hat{Q}^2(\hat{x}^j,\cdot)$ is convex in Ξ_l^j. Now, applying the Edmundson–Madansky inequality yields:

$$\frac{1}{N}\sum_{\xi_i\in\Xi_l^j}\hat{Q}_l^2(\hat{x}^j,\xi_i)\leqq\left[\sum_{v=1}^{2^K}\hat{Q}_l^2(\hat{x}^j,\eta_v^l)\cdot p_v^l\right]\cdot p^l \tag{4.16}$$

and hence

$$\mathrm{Var}\,(\hat{Q}(\hat{x}^j,\eta))\leqq\sum_{l=1}^{L_j}\left[\sum_{v=1}^{2^K}\hat{Q}_l^2(\hat{x}^j,\tilde{\eta}_v^l)\tilde{p}_v^l\right]\cdot p^l-[L^N(\hat{x}^j)]^2:=\Lambda(\hat{x}^j). \tag{4.17}$$

Summarizing we get an upper bound $\tilde{\sigma}^2(\hat{x}^j)$ for $\hat{\sigma}^2(\hat{x}^j)$:

$$\hat{\sigma}^2(\hat{x}^j) \leqq \sum_{l=1}^{L_j} [\gamma_l \{U_l^N(\hat{x}^j) + L_l^N(\hat{x}^j)\} + 2M\gamma_{ll}p^l] +$$
$$+ \{U^N(\hat{x}^j) - L^N(\hat{x}^j)\} \{\max [|L^N(\hat{x}^j)|, |U^N(\hat{x}^j)|] + |L^N(\hat{x}^j)|\} +$$
$$+ 2M\{U^N(\hat{x}^j) - L^N(\hat{x}^j)\} + \Lambda(\hat{x}^j) =: \tilde{\sigma}^2(\hat{x}^j). \tag{4.18}$$

This bound may be tightened in certain cases as shown in Appendix I. We have to stress that (4.18) enables us, to decide a posteriori, i.e. after a certain sampling size N has been chosen and the algorithm stopped at iteration step j, whether

$$CI_\lambda(\hat{x}^j) = \left[I^N(\hat{x}^j) - \frac{\lambda\tilde{\sigma}(\hat{x}^j)}{\sqrt{N}},\ I^N(\hat{x}^j) + \frac{\lambda\tilde{\sigma}(\hat{x}^j)}{\sqrt{N}} \right]$$

is small enough. If it turns out that N has to be increased, this should be done by taking $\lambda \cdot \tilde{\sigma}(\hat{x}^j)$ into consideration; after determining a proper sampling size, we have to modify the conditional expectations and the probabilities of the cells Ξ_l^j — which refer to the final partition S^j — and continue the algorithm (3.17.1)–(3.17.5), if

$$\Psi(\hat{x}^j, \tilde{\tilde{F}}^j) - \Psi(\hat{x}^j, \hat{\tilde{F}}^j) > \varepsilon$$

where $\tilde{\tilde{F}}^j$, $\hat{\tilde{F}}^j$ denote the modified distributions, obtained from the new sampling.

5. Numerical experience

The method described in section 3 has been implemented on the IBM 3033 computer of the University of Zürich for SLP with $q(\xi) = q^0$. Parts of MPSX/370 are used as subroutines to solve problems of type (3.1)–(3.3). We have to point out that the LPs with dual blockangular structure have been treated as usual LPs, i.e. the special structure has not been taken into consideration. Hence the computer times, listed below, can be regarded as an upper bound for solving the corresponding SLP's.

The algorithm has been tested in different versions and on different problem-sizes.

First, we considered two problem-sizes P_I, P_II:

P_I: $A(6/16)$,
$T(\xi)\,(8/16)$, $W(8/20)$;

P_II: $A(12/28)$,
$T(\xi)\,(15/28)$, $W(15/30)$

where the matrices A, $T(\xi)$, W and the vectors c, q, b, $h(\xi)$ were generated randomly.

In the first instance, the random vector ξ was assumed to be 3-dimensional, i.e. $\xi=(\xi_1, \xi_2, \xi_3)$ and uniformly distributed on $\bigtimes_{i=1}^{3}[0,1]$; hence the conditional expectations and probabilities of the cells Ξ_l could easily be evaluated exactly.

In section 3 two ways have been proposed for determining an upper bound for the optimal objective value:

(3.2) yielding $\Psi(\tilde{x}^j, \tilde{F}^j)$, and

(3.3) yielding $\Psi(\hat{x}^j, \tilde{F}^j)$, respectively.

According to these two possibilities the algorithm has been implemented in two different versions V1, V2:

Version V1: STEP 1, STEP 2 analogue to (3.17.1), (3.17.2);
STEP 3: solve (3.1) obtaining $\hat{x}^j$, $\Psi(\hat{x}^j, \hat{F}^j)$;
solve (3.2) obtaining $\tilde{x}^j$, $\Psi(\tilde{x}^j, \tilde{F}^j)$;
if $\Psi(\tilde{x}^j, \tilde{F}^j)-\Psi(\hat{x}^j, \hat{F}^j)\leqq\varepsilon\cdot\Psi(\hat{x}^j, \hat{F}^j)\Rightarrow$
$\Rightarrow\tilde{x}^j$ is sufficiently accurate,
STOP;
if not:
STEP 4: determine T^j, according to
$T^j:=\{\Xi_l^j|$ basic recourse vector y_v^l 'close to' degeneration$\}$;
choose the partition-coordinate cyclically;
STEP 5: analogous to (3.17.5).
Version V2: STEP 1, STEP 2, STEP 3 analogous to (3.17.1)–(3.17.3);
STEP 4: determine T^j according to (3.9);
choose the partition-coordinate according to (3.11);
STEP 5: analoguous to (3.17.5).

Remark: In STEP 3 of V1, V2 we used for our test series

$$\varepsilon:=0.05. \tag{5.1}$$

15 test examples have been generated randomly for each problem-size P_{I} and P_{II}, which have been solved by version V1 and V2. The results are summarized in Table 5.1.

The values corresponding to the headings $\bar{m}$ and s, respectively, represent the mean error bound in % and the deviation of the relative error respectively, obtained in the several iterations for the fore-mentioned 15 test examples:

In the first iterations the mean error and its deviation, by applying V1 is less than by applying V2. This is caused by the fact that performing STEP 3 in V1

Table 5.1

	P_I				P_{II}			
	V1		V2		V1		V2	
Iteration	$\bar{m}$	s	$\bar{m}$	s	$\bar{m}$	s	$\bar{m}$	s
1	585.4	1947.2	2286.5	7111.4	116.7	69.2	352.6	444.1
2	93.5	231.8	242.7	472.6	91.7	46.6	284.3	486.6
3	39.7	71.1	103.2	187.5	43.6	27.3	61.3	94.3
4	9.3	9.4	9.7	8.8	9.9	4.8	12.1	7.9
5	6.2	8.0	5.0	3.9	8.9	4.7	7.1	5.0
6	5.0	8.3	2.3	2.0	CPU-Time Limit		3.7	3.4
7	3.6	6.9	1.5	1.6	(11 min)		2.7	3.2
number of partitions (mean/deviation)	(26.0/20.1)		(25.4/12.3)		(?/?)		(32.4/15.6)	
CPU-Time (min) (mean/deviation)	(3.5/3.3)		(1.5/0.85)		$\gg$ 11 min		(4.0)/2.3)	

determines the minimal upper bound, while performing STEP 3 in V2 determines an upper bound at the initial solution $\hat{x}^j$. In the subsequent iterations the sign changes, caused by a more efficient partitioning strategy, as lower and upper bounds for the expected costs — subject to each cell — refer to the same decision $\hat{x}^j$ in V2. Hence the importance of evaluating the *minimal* upper bound in V1 strongly decreases.

Comparing the number of partitionings, we observe that its mean value in V1 and V2 is almost the same for the smaller problem size P_I. So, from the point of view of partitioning, determining the sharpest upper bound and performing a less significant partitioning strategy requires the same effort as performing the more efficient partitioning strategy by determining a weaker upper bound. For the larger problems P_{II}, we realize that evaluating the sharpest upper bound causes too much computational work, as it already exceeds the CPU-time limit — which was set to 11 min — within the first iterations. We expected — and indicated earlier — that V2 is far less expensive than V1, as solving the larger LP (3.2) in V1 is replaced by solving a number of small recourse problems (3.3) in V2.

Next we applied V2 on problems P_I, P_{II} with considering "arbitrary" distributions for the independent random variables ξ_1, ξ_2, ξ_3. We have chosen the following distributions:

$$\xi_1 \sim F(t) = 1 - e^{-\lambda t}, \quad \lambda := 3, \tag{5.2.1}$$

(i.e. exponential distribution)

$$\xi_2 \sim N(0, 1), \tag{5.2.2}$$

(i.e. Gaussian Normal Distribution)

$$\xi_3 \sim F(t) = 1 - e^{-\lambda t}, \quad \lambda := 6, \tag{5.2.3}$$

(i.e. exponential distribution)

and replaced $\xi = (\xi_1, \xi_2, \xi_3)$ by a discrete random variable $\xi_i = (\xi_{1,i}, \xi_{2,i}, \xi_{3,i})$, resulting from sampling with size $N := 100^3$.

In this phase we compared the two possibilities of choosing the partition-coordinate — i.e. (3.11) and (3.15), discussed in section 3.

Version V3: STEP 1–STEP 5 analogue to (3.17.1)–(3.17.5).

Remark: the tolerance ε is again chosen according to (5.1).

The results are summarized in Table 5.2.

We see that little improvements are achieved concerning mean error and its deviation and concerning the required CPU-time. Considering the number of partitions an essential improvement is obtained, which proves that partitioning according to (3.15) is more efficient than according to (3.11).

We also compared (3.7) and (3.8) — for determining the set T^j — subject to P_{I} and P_{II}, but as no essential difference has been observed, the results are not listed separately.

Table 5.2

	P_{I}				P_{II}			
	V2		V3		V2		V3	
Iteration	$\bar{m}$	s	$\bar{m}$	s	$\bar{m}$	s	$\bar{m}$	s
1	–	–	–	–	–	–	–	–
2	128.7	187.5	126.5	218.4	235.6	364.9	223.4	352.5
3	49.7	79.5	50.5	74.0	67.3	64.0	70.1	68.8
4	18.8	13.6	18.5	13.7	36.8	28.0	27.1	17.8
5	11.3	8.4	9.8	7.3	14.6	7.4	12.1	5.1
6	6.1	4.1	5.7	3.5	7.9	5.0	6.6	2.3
7	3.8	2.6	3.7	1.8	4.1	1.6	3.7	1.1
number of partitions (mean/deviation)	(52.4/27.4)		(25.9/19.4)		(67.5/35.4)		(31.8/11.5)	
CPU-Time (min) (mean/deviation)	(2.6/1.2)		(1.8/1.5)		(7.5/1.4)		(5.6/3.1)	

The next question we are interested in, is whether the chosen sampling size $N = 100^3$ is sufficiently large, subject to our tolerance level of 5%.

We have set $\lambda = 3$ which corresponds to the probability

$$[\Phi(\lambda) - \Phi(-\lambda)] = 0.9975,$$

and considered three different sampling sizes:

$$N_1 = 10^3, \quad N_2 = 100^3, \quad N_3 = 1000^3.$$

Evaluating the upper bound for $\hat{\sigma}(\hat{x}^j)$ according to (4.18) we obtained the following:

For each of the 15 test examples — subject to P_{I} and P_{II} —

$$\text{error by sampling} \leqq 5\% \quad \text{for } N_1,$$

$$\text{error by sampling} \leqq 0.1\% \quad \text{for } N_2,$$

$$\text{error by sampling} \leqq 0.005\% \quad \text{for } N_3,$$

with probability 0.9975 which confirms that in this special case the error resulting from sampling is negligible subject to the tolerance level of 5% for $N = 100^3$.

Finally, we applied V3 on additional problem-sizes P_{III}, P_{IV}.

P_{III}: $A(15/60)$,
$T(\xi)$ (30/60), $W(30/70)$;
$\xi = (\xi_1, \xi_2, \xi_3)$ distributed according to (5.2),
sampling size $N = 100^3$.

P_{IV}: $A(6/16)$,
$T(\xi)$ (8/16), $W(8/20)$;
$\xi = (\xi_1, \xi_2, \xi_3, \xi_4)$
with ξ_1, ξ_2, ξ_3 distributed according to (5.2),
$\xi_4 \sim N(0, 1)$ i.e. Gaussian Normal Distribution;
sampling size $N := 100^4$.

The results are listed in Table 5.3.

Finally, we have to mention that the implementation of our algorithm can be improved essentially with respect to CPU-time and storage by taking advantage of the properties of LP^j (3.1):

In extensive tests with Minos 5.0 it turned out that solving LP^J ab ovo, required comparably many pivot steps as solving $\mathrm{LP}^1, \ldots, \mathrm{LP}^J$ by transmitting all the basic columns of the optimal basis solution of LP^{j-1} to $\mathrm{LP}^j, j = 2, \ldots, J$. (Minos 5.0 verifies the regularity of the so obtained matrix and adds unit columns and exchanges columns, respectively if needed.) This is mainly caused by the fact that the corrections of the first stage solutions $\hat{x}^j$ and the matrices T^l, respectively, decrease with j, i.e. with a

Table 5.3

	V3			
	P_{III}		P_{IV}	
Iteration	$\bar{m}$	s	$\bar{m}$	s
1	–	–	–	–
2	492.4	677.3	89.5	110.1
3	320.2	796.1	33.2	18.8
4	15.3	5.7	22.1	13.8
5	7.9	3.1	11.8	7.4
6	4.8	1.3	7.6	4.3
7	3.1	0.8	5.6	2.8
8			4.1	1.7
9			3.3	1.2
number of partitions (mean/deviation)	(30.0/11.1)		(50.2/41.1)	
CPU-Time (min) (mean/deviation)	(40.9/22.5)		(6.2/5.2)	

finer partitioning. We have to stress that this reduces the numerical effort considerably as pivoting becomes the more expensive, the more cells (i.e. blocks in (3.1)) are created. Further, as we used dense matrices T^l, W concerning the LU-factorization of Minos 5.0, no fill-in was observed.

Another possibility of taking advantage of the special structure of (3.1) is to apply the basis reduction method described in [7], which was shortly reviewed in [4] where it was pointed out that the computer-time per pivot step will be reduced from $0(L_j^2)$ to $0(L_j)$, which is considerable as L_j tends to become large.

Finally, according to first tests by the author of [6], the regularized decomposition may be expected to be efficient in solving (3.1), at least for sparse matrices W and $T(\xi)$.

References

1. *Birge, J., Wets, R. J. B.*, Designing Approximation Schemes for Stochastic Optimization Problems, in particular for Stochastic Programs with Recourse. Working Paper WP-83-111, IIASA, Laxenburg 1983.
2. *Hausch, P. B., Ziemba, W. T.*, Bounds on the Value of Information in Uncertain Decision Problems. Stochastics **10** (1983), pp. 181–217.
3. *Kall, P.*, Stochastic Linear Programming. Springer-Verlag, Berlin 1976.

4. *Kall, P.*, Computational Methods for Solving Two-stage Stochastic Linear Programming Problems. Z. Angew. Math. Phys. **30** (1979), pp. 261–271.
5. *Kall, P., Stoyan, D.*, Stochastic Linear Programming Problems with Recourse, Including Error Bounds. Math. Operationsforsch. Statist., Ser. Optimization **13** (1982), pp. 431–447.
6. *Ruszczynski, A.*, A Regularized Decomposition Method for Minimizing a Sum of Polyhedral Functions. Manuscript, Inst. f. OR, Universität Zürich (1985).
7. *Strazicky, B.*, On an Algorithm for Solution of the Two-stage Stochastic Programming Problem. Methods of Operations Research **XIX** (1974), pp. 142–156.
8. *Wets, R.*, Stochastic Programming: Solution Techniques and Approximation Schemes. In Mathematical Programming: The State of the Art, eds. A. Bachem, M. Grötschel and B. Korte, Springer-Verlag, Berlin 1983, pp. 566–603.

Appendix I

Tightening the bound $\tilde{\sigma}(\hat{x}^j)$ for $\hat{\sigma}(\hat{x}^j)$.)
Applying the triangle inequality, we get from (4.4), (4.8) and (4.9)

$$|\hat{\sigma}^2(\hat{x}^j) - \mathrm{Var}\,(\hat{Q}(\hat{x}^j, \eta))| \leqq \tag{I.4.11}$$

$$\leqq \frac{1}{N} \sum_{l=1}^{L_j} \left| \sum_{\xi_i \in \Xi_l^j} (Q(\hat{x}^j, \xi_i) - \hat{Q}_l(\hat{x}^j, \xi_i))\,(Q(\hat{x}^j, \xi_i) + \hat{Q}_l(\hat{x}^j, \xi_i)) \right| +$$

$$+ |(I^N(\hat{x}^j))^2 - (L^N(\hat{x}^j))^2|.$$

The second term on the right-hand side of (I.4.11) may be bounded from above due to

$$|(I^N(\hat{x}^j))^2 - (L^N(\hat{x}^j))^2| = \tag{I.4.12}$$

$$= [I^N(\hat{x}^j) - L^N(\hat{x}^j)] \cdot |I^N(\hat{x}^j) + L^N(\hat{x}^j)| \leqq$$

$$\leqq [U^N(\hat{x}^j) - L^N(\hat{x}^j)] \cdot \{|I^N(\hat{x}^j)| + |L^N(\hat{x}^j)|\} \leqq$$

$$\leqq [U^N(\hat{x}^j) - L^N(\hat{x}^j)] \cdot \{\max [|L^N(\hat{x}^j)|, |U^N(\hat{x}^j)|] + |L^N(\hat{x}^j)|\} =: \mu(\hat{x}^j).$$

To bound the first term on the right-hand side of (I.4.11), we have to distinguish between different cases:

case α) $\min\limits_{\eta \in \Xi_l^j} \hat{Q}_l(\hat{x}^j, \eta) \geqq 0$:

$$\frac{1}{N} \sum_{\xi_i \in \Xi_l^j} (Q(\hat{x}^j, \xi_i) - \hat{Q}_l(\hat{x}^j, \xi_i))\,(Q(\hat{x}^j, \xi_i) + \hat{Q}_l(\hat{x}^j, \xi_i)) \leqq$$

$$\leqq \left[\max_{\eta \in \Xi_l^j} (Q(\hat{x}^j, \eta) - \hat{Q}_l(\hat{x}^j, \eta)) \right] \left[\frac{1}{N} \sum_{\xi_i \in \Xi_l^j} Q(\hat{x}^j, \xi_i) + \hat{Q}_l(\hat{x}^j, \xi_i) \right] \leqq$$

$$\leqq \gamma_l (U_l^N(\hat{x}^j) + L_l^N(\hat{x}^j))$$

where $\gamma_l := \max\limits_{\eta \in \Xi_l^j} [Q(\hat{x}^j, \eta) - \hat{Q}_l(\hat{x}^j, \eta)]$ (see Fig. 4.1).

case β) $\max\limits_{\eta \in \Xi_l^j} Q(\hat{x}^j, \eta) \leqq 0$:

As this case implies

$$|L_l^N(\hat{x}^j)| \geqq \left| \sum_{\xi_i \in \Xi_l^j} \frac{1}{N} Q(\hat{x}^j, \xi_i) \right|,$$

we get similar to case α)

$$\frac{1}{N}\left|\sum_{\xi_i\in\Xi_l^j}(Q(\hat{x}^j,\xi_i)-\hat{Q}_l(\hat{x}^j,\xi_i))(Q(\hat{x}^j,\xi_i)+\hat{Q}_l(\hat{x}^j,\xi_i))\right|\leqq\gamma_l\cdot 2|L_l^N(\hat{x}^j)|.$$

case γ) $\max\limits_{\eta\in\tilde{\Xi}_l^j} Q(\hat{x}^j,\eta)>0$ $\min\limits_{\eta\in\tilde{\Xi}_l^j}\hat{Q}_l(\hat{x}^j,\eta)<0$:

Applying the triangle inequality yields

$$\frac{1}{N}\left|\sum_{\xi_i\in\Xi_l^j}(Q(\hat{x}^j,\xi_i)-\hat{Q}_l(\hat{x}^j,\xi_i))(Q(\hat{x}^j,\xi_i)+\hat{Q}_l(\hat{x}^j,\xi_i))\right|\leqq$$

$$\leqq\sum_{\xi_i\in\Xi_l^j}\frac{1}{N}|Q(\hat{x}^j,\xi_i)-\hat{Q}_l(\hat{x}^j,\xi_i)|\cdot[|Q(\hat{x}^j,\xi_i)|+|\hat{Q}_l(\hat{x}^j,\xi_i)|]\leqq$$

$$\leqq\gamma_l\sum_{\xi_i\in\Xi_l^j}\frac{1}{N}[|Q(\hat{x}^j,\xi_i)|+|\hat{Q}_l(\hat{x}^j,\xi_i)|].$$

As for all $\xi_i\in\Xi_l^j$

$$|Q(\hat{x}^j,\xi_i)|\leqq\max\left[\max_{\eta\in\tilde{\Xi}_l^j}Q(\hat{x}^j,\eta);\quad\min_{\eta\in\tilde{\Xi}_l^j}\hat{Q}_l(\hat{x}^j,\eta)\right]=:\tau_l(\hat{x}^j)$$

and

$$|\hat{Q}_l(\hat{x}^j,\xi_i)|\leqq\max\left[\max_{\eta\in\tilde{\Xi}_l^j}\hat{Q}_l(\hat{x}^j,\eta);\quad\min_{\eta\in\tilde{\Xi}_l^j}\hat{Q}_l(\hat{x}^j,\eta)\right]=:\hat{\tau}_l(\hat{x}^j),$$

we may conclude

$$\frac{1}{N}\left|\sum_{\xi_i\in\Xi_l^j}(Q(\hat{x}^j,\xi_i)-\hat{Q}_l(\hat{x}^j,\xi_i))\cdot(Q(\hat{x}^j,\xi_i)+\hat{Q}_l(\hat{x}^j,\xi_i))\right|\leqq$$

$$\leqq\gamma_l\cdot p^l\cdot(\tau_l(\hat{x}^j)+\hat{\tau}_l(\hat{x}^j)).$$

Defining now

$$\delta_l(\hat{x}^j)=\begin{cases}\gamma_l(U_l^N(\hat{x}^j)+L_l^N(\hat{x}^j)) & \text{if case } \alpha)\\ \gamma_l\cdot 2|L_l^N(\hat{x}^j)| & \text{if case } \beta)\\ \gamma_l p^l(\tau_l(\hat{x}^j)+\hat{\tau}_l(\hat{x}^j)) & \text{if case } \gamma),\end{cases}\tag{I.4.13}$$

we get from (I.4.11), (I.4.12) and (I.4.13)

$$|\hat{\sigma}^2(\hat{x}^j)-\operatorname{Var}(\hat{Q}(\hat{x}^j,\eta))|\leqq\sum_l\delta_l(\hat{x}^j)+\mu(\hat{x}^j).\tag{I.4.14}$$

Bounding $\operatorname{Var}(\hat{Q}(\hat{x}^j,\eta))$ due to (4.17), and applying (I.4.14) we obtain an upper bound $\tilde{\sigma}^2(\hat{x}^j)$ for $\hat{\sigma}^2(\hat{x}^j)$

$$\hat{\sigma}^2(\hat{x}^j)\leqq\sum_l\delta_l(\hat{x}^j)+\mu(\hat{x}^j)+\Lambda(\hat{x}^j)=:\tilde{\tilde{\sigma}}^2(\hat{x}^j).\tag{I.4.15}$$

Метод решения стохастической линейной программы с произвольными мультивариационными распределениями — независимый случай

К. ФРАУЭНДОРФЕР, П. КАЛЛ

(Цюрих)

Рассматривается модифицированная стохастическая линейная программа, в которой целевая функция аппроксимируется кусочно-линейной функцией, а случайные переменные моделируются дискретными.

K. Frauendorfer
P. Kall
Institut für Operations Research und
Mathematische Methoden der Wirtschaftswissenschaften
der Universität Zürich
Weinbergstrasse 59
CH-8006 Zürich

Problems of Control and Information Theory, Vol. 17 (4) pp. 207—221 (1988)

SYNTHESIS OF OPTIMAL GUARANTEEING CONTROL

V. E. TRET'JAKOV

(Sverdlovsk)

(Received February 9, 1987)

In the present paper an algorithm for calculation of optimal controlling actions in the problem of positional control of systems perturbed by uncontrollable disturbance is presented. It is shown that combination of three components, namely, 1) extremal displacement onto concomitant point, 2) constructions of stochastic synthesis, 3) introduction of an additional phase coordinate, allows us to construct a function which approximates the game value and is differentiable with respect to the phase variable.

Here extremal displacement onto concomitant point is reduced to the extremal displacement in the direction opposite to the gradient of the approximating function. It is established that if a quality index is the norm of the phase vector at the instant of process termination, the present approach leads to a realized calculating procedure. The obtained results make clear the connection of stochastic program synthesis [1–3] with known generalization of the method of dynamic programming [4, 5].

1. Statement of the problem

Let us consider a problem of positional control under the condition of minimum of guaranteed result over the index

$$\gamma = |x[\vartheta]| \tag{1.1}$$

for the system

$$\frac{dx}{dt} = A(t)x + B(t)u + C(t)v, \quad t_0 \leqq t \leqq \vartheta. \tag{1.2}$$

Here x is an n-dimensional phase vector of the object; $|x|$ is its Euclidean norm, t_0 and ϑ are fixed instants of time; $A(t)$, $B(t)$, $C(t)$ are continuous matrix-valued functions; u is an r-dimensional vector of control, v is an s-dimensional vector of disturbance which are constrained by the restrictions $u \in \mathscr{P}$, $v \in \mathscr{Q}$, respectively, where $\mathscr{P}$ and $\mathscr{Q}$ are given compact sets.

Notions of strategy $u(\cdot) = \{u(t, x, \varepsilon) \in \mathscr{P},\ t_0 \leqq t \leqq \vartheta, x \in E^n, \varepsilon > 0\}$ and control law

$$U = \{u(\cdot), \varepsilon, \Delta_\delta\{t_i\}\} \tag{1.3}$$

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

are the main tools for solving the considered problem. Motion $x[t_*[\cdot]\vartheta = \{x[t], t_* \leqq t \leqq \vartheta\}$ of the system (1.2) is generated from arbitrary position $\{t_*, x_*\}$, $x[t_*] = x_*$, $t_* \in [t_0, \vartheta)$, $x_* \in E^n$ in steps $[t_i, t_{i+1}]$, $i = 1, \ldots, k$ of partitioning

$$\Delta_\delta\{t_i\} = \{t_* = t_1;\ t_i,\ t_{i+1} - t_i \leqq \delta;\ t_{k+1} = \vartheta\} \tag{1.4}$$

by the control law U (1.3) and some realization of uncontrolled disturbance

$$v(t_*[\cdot]\vartheta] = \{v[t] \in \mathcal{Q}, t_* < t \leqq \vartheta\} \tag{1.5}$$

which is any measurable function of time.

These notions constitute the underlying mathematical formalization of the minimum problem for ensured result proposed in monograph [1]. According to this formalization the problem of calculating of the optimal ensured result

$$\rho^0(t_*, x_*) = \rho[u^0(\cdot);\ t_*, x_*] = \min_{u(\cdot)} \overline{\lim_{\varepsilon \to 0}} \lim_{\delta \to 0} \sup_{\Delta_\delta\{t_i\}} \sup_{v(t_*[\cdot]\vartheta]} |x[\vartheta]| \tag{1.6}$$

and of searching of the universal optimal strategy $u^0(\cdot)$ is set. The problem has a solution.

At any position $\{t, x\}$ the optimal ensured result for the index γ (1.1) can be calculated by the method of stochastic program synthesis as a value of stochastic program extremum ([1], p. 306)

$$\rho^0(t, x) = \sup_{\Delta\{\tau_j\}} \sup_{\|l(\cdot)\| \leqq 1} \kappa(t, x; \Delta\{\tau_j\}, l(\cdot)). \tag{1.7}$$

Here

$$\Delta\{\tau_j\} = \{t = \tau_1, \ldots, \tau_j < \tau_{j+1}, \ldots, \tau_{q+1} = \vartheta\} \tag{1.8}$$

is an arbitrary partitioning of the segment $[t, \vartheta]$ which is not connected with the partitioning (1.4) in any way. Value $l(\cdot)$ is Borel-measurable on cube $\Omega = \{\omega = \{\xi_1, \ldots, \xi_q\}, 0 \leqq \xi_j < 1, j = 1, \ldots, q\}$ n-dimensional vector-valued function which is naturally treated as a vectorial random variable and then the symbol $\|l(\cdot)\|$ should be understood as Eucledian norm in Hilbert space of random variables ([6], p. 279). The scalar function κ in expression (1.7) is determined by relation (41.8) in monograph [1] (see p. 315) and there is no necessity to write out this relation explicitly. We only note that by the replacement

$$l(\cdot) = m + b(\cdot), \quad m = M\{l(\omega)\}, \quad M\{b(\omega)\} = 0, \tag{1.9}$$

where M denotes the mathematical expectation, problem (1.7) is reduced to the following one ([7], p. 187)

$$\rho^0(t, x) = \max_{|m| \leqq 1} [\langle X'(\vartheta, t)m \cdot x\rangle + g(t, m)] \tag{1.10}$$

Here $X(\vartheta, t)$ is the fundamental matrix of the uniform system (1.2), the upper prime denotes transposition, and $g(t, m)$ is determined by the equality

$$g(t,m)=\sup_{\Delta\{\tau_j\}}\left[\sup_{\|b(\cdot)\|\leqq(1-|m|^2)^{1/2}} M\left\{\int_t^{\vartheta}\min_{u\in\mathscr{P}}\max_{v\in\mathscr{Q}}<[m+\right.\right.$$

$$\left.\left.+n(\tau,\omega)]\cdot X(\vartheta,\tau)\,[B(\tau)u+C(\tau)v]>d\tau\right\}\right], \tag{1.11}$$

where the values

$$n(\tau,\omega)=n[\xi_1,\ldots,\xi_j]=\int_{0\leqq\xi_{j+1}<1}\cdots\int_{0\leqq\xi_q<1} b[\xi_1,\ldots,\xi_q]\,d\xi_{j+1}\ldots d\xi_q;$$

$$\tau_j\leqq\tau<\tau_{j+1},\quad j=1,\ldots,q \tag{1.12}$$

are conditional mathematical expectations of n-dimensional vectorial random variable $b(\omega)=b[\xi_1,\ldots,\xi_q]$. We can show (and it is essential for the further consideration) that function $g(t, m)$ (1.11) is concave with respect to argument m. Then all function maximized in (1.10) will be concave with respect to m at every fixed position $\{t, x\}$.

The optimal strategy $u^0(\cdot)$ on the basis of function $\rho^0=\rho^0(t, x)$ (1.10) is found by the rule of extremal displacement ([1], p. 210)

$$\langle s_u(t,x,\varepsilon)\cdot B(t)u^0(t,x,\varepsilon)\rangle=\min_{u\in\mathscr{P}}\langle s_u(t,x,\varepsilon)\cdot B(t)u\rangle \tag{1.13}$$

in the direction opposite to the vector $s_u(t, x, \varepsilon)$ which is determined from the condition

$$\rho^0(t,x-s_u(t,x,\varepsilon))=\min_{|s|\leqq\beta(\varepsilon,t)}\rho^0(t,x-s), \tag{1.14}$$

where

$$\beta(\varepsilon,t)=[\varepsilon(1+[t-t_0])]^{1/2}\exp\lambda[t-t_0], \tag{1.15}$$

$$\lambda=\max_{t_0\leqq t\leqq\vartheta}\max_{|x|=1}|A(t)x| \tag{1.16}$$

Thus, to calculate the optimal controlling action

$$u^0[t]=u^0(t_i,x[t_i],\varepsilon),\quad t_i<t\leqq t_{i+1} \tag{1.17}$$

on each step $(t_i, t_{i+1}]$, $i=1,\ldots,k$, it is required to solve two minimum problems (1.14) and (1.13) at $t=t_i$, $x=x[t_i]$ that in general case can lead to the necessity of repeated counts of $\rho^0(t_i, x[t_i]-s)$ by formulae (1.10), (1.11) at points of n-dimensional sphere $|s|\leqq\beta(\varepsilon, t)$. From the point of view of the computation this problem is the most time-consuming one.

In the present paper an approach is suggested which allows us to pass by solving of problem (1.14). It turns out that due to the addition of one more fictitious coordinate x_{n+1}, we succeed with the help of expressions (1.10), (1.11) in constructing of some function $\rho_\varepsilon^0(t, x)$ which approximizes the function of optimal ensured result $\rho^0(t, x)$ (1.6) and is differentiable with respect to x. In turn, it allows us to simplify an algorithm of calculation of optimal controlling actions $u^0[t]$ (1.17) and to reduce it on each step $[t_i, t_{i+1}]$, $i=1, \ldots, k$ to the extremal displacement

$$\left\langle \left[\frac{\partial \rho_\varepsilon^0(t, x)}{\partial x}\right]_{t=t_i,\, x=x[t_i]} \cdot B(t_i)u^0(t_i, x[t_i], \varepsilon)\right\rangle =$$

$$= \min_{u\in\mathcal{P}} \left\langle \left[\frac{\partial \rho_\varepsilon^0(t, x)}{\partial x}\right]_{t=t_i,\, x=x[t_i]} \cdot B(t_i)u\right\rangle \tag{1.18}$$

in the direction opposite to the gradient

$$\partial\rho_\varepsilon^0(t, x)/\partial x = \{\partial\rho_\varepsilon^0/\partial x_1, \ldots, \partial\rho_\varepsilon^0/\partial x_n\},\ t=t_i,\ x=x[t_i]$$

of the function $\rho_\varepsilon^0(t, x)$ constructed above.

2. Approximating function and its properties

Consider the function

$$\rho_\varepsilon^0(t, x) = \max_{|m|\leq 1} H(t, x, \varepsilon, m) \tag{2.1}$$

$$H(t, x, \varepsilon, m) = -\beta(\varepsilon, t)(1+|X'(\vartheta, t)m|^2)^{1/2} +$$

$$+\langle X'(\vartheta, t)m \cdot x\rangle + g(t, m) \tag{2.2}$$

defined at all $t\in[t_0, \vartheta]$ and all $x\in E^n$.

We stress that the calculation of this function can be effectively realized in a number of concrete cases including the case when $\gamma=|x[\vartheta]|$ which is analyzed here in detail. We shall not dwell on sources of obtaining of the function. As it is already noted, these sources are based on the ideas of stochastic synthesis and introduction of additional regulating phase coordinate x_{n+1}. We shall only mark the principal properties of the function $\rho_\varepsilon^0(t, x)$ (2.1), (2.2).

1^0. Thanks to the strict concavity of the function $H(t, x, \varepsilon, m)$ (2.2) with respect to m, the maximum in (2.1) is achieved on the unique vector

$$m^0 = m^0(t, x, \varepsilon). \tag{2.3}$$

2^0. In consequence of the uniqueness of the vector m^0 (2.3) vector-valued function $m^0 = m^0(t, x, \varepsilon)$ is continuous with respect to the arguments $t\in[t_0, \vartheta]$ $x\in E^n$.

3°. Due to properties 1°, 2°, the function $\rho_\varepsilon^0(t, x)$ (2.1) is differentiable with respect to x, i.e. the gradient $\partial\rho_\varepsilon^0(t, x)/\partial x$ exists. Moreover,

$$\frac{\partial\rho_\varepsilon^0(t, x)}{\partial x} = X'(\vartheta, t)m^0(t, x, \varepsilon) \tag{2.4}$$

and on account of the condition $|m^0(t, x, \varepsilon)| \leqq 1$, the gradient $\partial\rho_\varepsilon^0(t, x)/\partial x$ (2.4) continuous with respect to $\{t, x\}$ is uniformly bounded at $t \in [t_0, \vartheta]$, $x \in E^n$.

4°. In consequence of the equality $g(\vartheta, m) = 0$ for $t = \vartheta$, the function $\rho_\varepsilon^0(t, x)$ (2.1) satisfies the boundary condition

$$\rho_\varepsilon^0(\vartheta, x) = \begin{cases} -[\beta^2(\varepsilon, \vartheta) - |x|^2]^{1/2}, & |x| \leqq \dfrac{\beta(\varepsilon, \vartheta)}{\sqrt{2}} \\ |x| - \sqrt{2}\beta(\varepsilon, \vartheta), & |x| > \dfrac{\beta(\varepsilon, \vartheta)}{\sqrt{2}} . \end{cases} \tag{2.5}$$

5°. The function $\rho_\varepsilon^0(t, x)$ (2.1) approximates the function of the optimal guaranteed result $\rho^0(t, x)$ (1.10) in the following sense: at all $\{t, x\}$ the estimate

$$\beta(\varepsilon, t_0) \leqq \rho^0(t, x) - \rho_\varepsilon^0(t, x) \leqq K\beta(\varepsilon, \vartheta) \tag{2.6}$$

is valid, where K is a positive constant.

6°. Let us choose the control law

$$U^0 = \{u^0(\cdot), \varepsilon, \Delta_\delta\{t_i\}\} \tag{2.7}$$

which forms actions $u^0[t] = u^0(t_i, x[t_i], \varepsilon)$, $t_i < t \leqq t_{i+1}$, $i = 1, \ldots, k$ from the condition of extremal displacement (1.18) on the basis of constructed function $\rho_\varepsilon^0(t, x)$ (2.1). And let some unknown disturbance $v(t_*[\cdot]\vartheta]$ (1.5) be realized. Then the following estimate expressing the property of u-stability of the function $\rho_\varepsilon^0(t, x)$ occurs along the motion $x[t_*[\cdot]\vartheta]$ generated by the law U^0 (2.7) on each step $[t_i, t_{i+1}]$, $i = 1, \ldots, k$

$$\Delta\rho_\varepsilon^0 = \rho_\varepsilon^0(t_{i+1}, x[t_{i+1}]) - \rho_\varepsilon^0(t_i, x[t_i]) \leqq$$

$$\leqq \alpha(\varepsilon, \delta)[t_{i+1} - t_i]; \quad \lim_{\delta \to 0, \varepsilon \to 0} \alpha(\varepsilon, \delta) = 0 \tag{2.8}$$

for arbitrary realization of disturbance $v(t_*[\cdot]\vartheta]$ (1.5).

The given estimate can also be expressed in the following differential form

$$D_t^+(\rho_\varepsilon^0(t, x)) + \left\langle \frac{\partial\rho_\varepsilon^0}{\partial x} \cdot [A(t)x + B(t)u^0(t, x, \varepsilon) + C(t)v] \right\rangle \leqq \vartheta(\varepsilon), \tag{2.9}$$

where the symbol $D_t^+(\rho_\varepsilon^0(t, x))$ designates the upper right-hand with respect to t the derivative number of the function $\rho_\varepsilon^0(t, x)$ (2.1). Here relation (2.9) replaces the corresponding differential inequality from the method of dynamic programming and makes the connection of this method with the constructions of the method of stochastic program synthesis clear.

These estimates are directly derived from the definition of the function $\rho_\varepsilon^0(t, x)$ (2.1) which gives the inequality

$$\Delta\rho_\varepsilon^0 \leqq H(t_{i+1}, x[t_{i+1}], \varepsilon, m^0(t_{i+1}, x[t_{i+1}], \varepsilon)) - \\ - H(t_i, x[t_i], \varepsilon, m^0(t_{i+1}, x[t_{i+1}], \varepsilon)) \tag{2.10}$$

Here properties 1^0–3^0 of the function $\rho_\varepsilon^0(t, x)$ (2.1) are used. Making the further treatment of the right-hand side of inequality (2.10), the following estimate of the variation of $g(t, m^0)$ (1.11) is the most essential

$$\Delta g = g(t_{i+1}, m^0(t_{i+1}, x[t_{i+1}], \varepsilon)) - g(t_i, m^0(t_{i+1}, x[t_{i+1}], \varepsilon)) \leqq$$

$$\leqq - \int_{t_i}^{t_{i+1}} \min_{u \in \mathcal{P}} \max_{v \in \mathcal{Q}} < m^0(t_i, x[t_i], \varepsilon) \cdot X(\vartheta, t)\,[B(t)u + C(t)v] > dt \leqq$$

$$\leqq - \left[\min_{u \in \mathcal{P}} \max_{v \in \mathcal{Q}} \left\langle \left[\frac{\partial \rho_\varepsilon^0(t, x)}{\partial x} \right]_{t = t_i,\, x = x[t_i]} \cdot [B(t_i)u + C(t_i)v] \right\rangle \right] \cdot \\ \cdot (t_{i+1} - t_i) + \alpha_1(\delta)\,(t_{i+1} - t_i), \tag{2.11}$$

where $\alpha_1(\delta) \to 0$ for $\delta \to 0$.

Then using the condition of extremal displacement (1.18) and taking into account estimates (2.11), we obtain the required estimate (2.8) from inequality (2.10).

Successively comparing inequalities (2.8) at $i = 1, \ldots, k$, we obtain that on the motion $x[t_*[\cdot]\vartheta]$ generated by the control law U^0 (2.7) and arbitrary disturbance $v(t_*[\cdot]\vartheta]$ (1.5), the inequality

$$\rho_\varepsilon^0(\vartheta, x[\vartheta] \leqq \rho_\varepsilon^0(t_*, x_*) + \alpha(\varepsilon, \delta)\,[\vartheta - t_0] \tag{2.12}$$

is fulfilled where $\alpha(\varepsilon, \delta) \to 0$ for $\varepsilon \to 0$, $\delta \to 0$.

Then taking into consideration properties 4^0, 5^0 of the function $\rho_\varepsilon^0(t, x)$ (2.1) and the boundary condition $\rho^0(\vartheta, x) = |x|$ for the function $\rho^0(t, x)$ (1.10), we come to a conclusion that for any arbitrary small $\chi > 0$ the control law U^0 (2.7) ensures the inequality

$$|x[\vartheta]| = \gamma[t_*, x_*;\ U^0,\ v(t_*[\cdot]\vartheta]] \leqq \rho^0(t_*, x_*) + \chi \tag{2.13}$$

if only in the law U^0 (2.7) ε and δ are chosen sufficiently small depending on χ.

On the other hand it is known [1] that one can find a law of forming of disturbance

$$V^0 = \{v^0(\cdot),\ \varepsilon \leqq \varepsilon(\chi),\ \Delta_\delta\{t_i\} : \delta \leqq \delta(\varepsilon)\} \tag{2.14}$$

based on some optimal strategy

$$v^0(\cdot) = \{v^0(t, x, \varepsilon) \in Q,\ t_0 \leqq t \leqq \vartheta,\ x \in E^n,\ \varepsilon > 0\}, \tag{2.15}$$

which gives the result

$$\gamma[t_*, x_*; u(t_*[\cdot]\vartheta], V^0] \geqq \rho^0(t_*, x_*) - \chi \tag{2.16}$$

for arbitrary realization of control $u(t_*[\cdot]\vartheta]$ including, maybe, the realization which is formed in steps by the control law U^0 (2.7). Comparing (2.13) with (2.16), we convince ourselves that the variable $\rho^0(t_*, x_*)$ in (2.13) cannot be decreased by any arbitrarily small constant and, therefore, it delivers the optimal guaranteed result $\rho[u^0(\cdot);\ t_*,\ x_*]$ in the sense of definition (1.6).

According to condition (1.18) ensuring estimate (2.8) the optimal universal strategy $u^0(\cdot)$ is found from the condition of extremal displacement

$$\left\langle \frac{\partial \rho_\varepsilon^0}{\partial x} \cdot B(t)u^0(t, x, \varepsilon) \right\rangle = \min_{u \in \mathcal{P}} \left\langle \frac{\partial \rho_\varepsilon^0}{\partial x} \cdot B(t)u \right\rangle \tag{2.17}$$

in the direction opposite to the gradient $\partial \rho_\varepsilon^0 / \partial x$ of the function $\rho_\varepsilon^0(t, x)$ (2.1).

In conclusion of this section we note that according to the method of stochastic program synthesis the optimal strategy $v^0(\cdot)$ (2.15) can also be found on the basis of the function $\rho^0(t, x)$ (1.10) by the rule of extremal displacement

$$\langle s_v(t, x, \varepsilon) \cdot C(t)v^0(t, x, \varepsilon) \rangle = \min_{v \in \mathcal{Q}} \langle s_v(t, x, \varepsilon) \cdot C(t)v \rangle \tag{2.18}$$

but now in the direction opposite to the vector $s_v(t, x, \varepsilon)$ determined from the condition

$$\rho^0(t, x - s_v(t, x, \varepsilon)) = \max_{|s| \leqq \beta(\varepsilon, t)} \rho^0(t, x - s). \tag{2.19}$$

However, here one fails to avoid solution of the problem (2.19), as it was done in the case of problem (1.14). At all events one must keep in mind that the extremal displacement (2.18) cannot be replaced by the one in the direction opposite to the gradient $\partial \rho_\varepsilon^0 / \partial x$ of the function $\rho_\varepsilon^0(t, x)$ (2.1). Thus, the problem of construction of the strategy $v^0(\cdot)$ (2.15) is more difficult. In this connection the remark on the possibility of forming of the optimal counteraction $v^0[t]$, $t_i < t \leqq t_{i+1}$, $i = 1, \ldots, k$, relying on maximizing the stochastic nonanticipatory program $v^0(t, \omega)$, $t_* < t \leqq \vartheta$, $\omega \in \Omega$ which is extracted from the solution of the problem (1.10) is of essential interest.

It turns out that on the motion generated by the control law U^0 (2.7) and maximizing program $v^0(t, \omega)$ one can obtain the result γ (1.1) arbitrarily close to $\rho^0(t_*, x_*)$ with probability arbitrarily close to 1.

3. Model problem

We shall illustrate the suggested approach with the help of the convergence-evasion problem ([5], p. 51) of two objects.

Let two mass points $m^{(1)}$ and $m^{(2)}$ with masses m_1 and m_2, correspondingly, move in three-dimensional space. Positions of these points in the space are determined by its radius vectors

$$r^{(k)}[t] = \{r_1^{(k)}[t], r_2^{(k)}[t], r_3^{(k)}[t]\},$$
$$k = 1, 2; \quad t_0 \leqq t \leqq \vartheta. \tag{3.1}$$

The points move under the action of forces of environment resistance proportional to the points' velocities $dr^{(k)}/dt$, $k = 1, 2$ and attractive (or repulsive) forces proportional to distances of the points to the origin coordinates. The first, $m^{(1)}$ and the second, $m^{(2)}$ points are controlled by forces $u^{(1)}$ and $u^{(2)}$, correspondingly. At every instant of time t the forces are bounded

$$|u^{(k)}[t]| \leqq \mu_k, \qquad \mu_k = \text{const}, \qquad k = 1, 2. \tag{3.2}$$

The equations of points' motion have the following form

$$m_k \frac{d^2 r^{(k)}}{dt^2} + \alpha_k \frac{dr^{(k)}}{dt} + \beta_k r^{(k)} = u^{(k)}, \qquad k = 1, 2 \tag{3.3}$$

where m_k and α_k are some non-negative numbers and β_k, $k = 1, 2$, are arbitrary ones.

We assume that the aim of the first player who commands the choice of control $u^{(1)}$ is the convergence of the point $m^{(1)}$ with the point $m^{(2)}$ at given instant of time, i.e. by the choice of his controlling actions $u^0[t]$ (1.17) the first player intends to make the distance from the point $m^{(2)}$

$$\gamma = |r^{(1)}[\vartheta] - r^{(2)}[\vartheta]| \tag{3.4}$$

as small as possible at instant of time ϑ.

At instants $t_i = 1, \ldots, k$, needed for him the first player has at his disposal information on realizing states of system (3.3) but he knows nothing about the intentions of the second player who commands the choice of control $u^{(2)}$. Perhaps, harmful to the first player, the second one will intend to increase the distance γ (3.4) as much as possible.

The considered differential game can be reduced by the known method ([5], p. 53) to the minimum problem of ensured result result with respect to index $\gamma = |x[\vartheta]|$ (1.1) for the system

$$\frac{dx}{dt} = \varphi(t)u - \psi(t)v, \tag{3.5}$$

$$u = \frac{1}{m_1}u^{(1)}, \quad v = \frac{1}{m_2}u^{(2)}, \quad \frac{1}{m_1}\mu_1 = \mu, \quad \frac{1}{m_2}\mu_2 = \nu, \tag{3.6}$$

where $x = \{x_1, x_2, x_3\}$ is a three-dimensional vector, connected with phase coordinates of the points $m^{(1)}$ and $m^{(2)}$ by the relations

$$\begin{aligned}
x_1 &= x_{11}^{(1)}[\vartheta, t]r_1^{(1)} + x_{12}^{(1)}[\vartheta, t]\dot{r}_1^{(1)} - \\
&\quad - x_{11}^{(2)}[\vartheta, t]r_1^{(2)} - x_{12}^{(2)}[\vartheta, t]\dot{r}_1^{(2)}; \\
x_2 &= x_{11}^{(1)}[\vartheta, t]r_2^{(1)} + x_{12}^{(1)}[\vartheta, t]\dot{r}_2^{(1)} - \\
&\quad - x_{11}^{(2)}[\vartheta, t]r_2^{(2)} - x_{12}^{(2)}[\vartheta, t]\dot{r}_2^{(2)}; \\
x_3 &= x_{11}^{(1)}[\vartheta, t]r_3^{(1)} + x_{12}^{(1)}[\vartheta, t]\dot{r}_3^{(1)} - \\
&\quad - x_{11}^{(2)}[\vartheta, t]r_3^{(2)} - x_{12}^{(2)}[\vartheta, t]\dot{r}_3^{(2)}.
\end{aligned} \tag{3.7}$$

Here functions $x_{11}^{(k)}[\vartheta, t]$, $x_{12}^{(k)}[\vartheta, t]$, $k = 1, 2$ are elements of some fundamental matrix calculated with the initial parameters of the system

$$a_1^{(1)} = \frac{\alpha_1}{m_1}, \quad a_2^{(1)} = \frac{\beta_1}{m_1}, \quad a_1^{(2)} = \frac{\alpha_2}{m_2}, \quad a_2^{(2)} = \frac{\beta_2}{m_2}. \tag{3.8}$$

For example, depending on parameters (3.8), the following three expressions are possible for the function $x_{12}^{(k)}[\vartheta, t]$, $k = 1, 2$

(a) $[a_1^{(k)}]^2 - 4a_2^{(k)} < 0$,

$$x_{12}^{(k)}[\vartheta, t] = \frac{2}{\sqrt{4a_2^{(k)} - [a_1^{(k)}]^2}} \exp\left[-\frac{a_1^{(k)}}{2}(\vartheta - t)\right] \times$$

$$\times \sin\left[\frac{1}{2}\sqrt{4a_2^{(k)} - [a_1^{(k)}]^2}(\vartheta - t)\right]; \tag{3.9}$$

(b) $[a_1^{(k)}]^2 - 4a_2^{(k)} > 0$,

$$x_{12}^{(k)}[\vartheta, t] = \frac{2}{\sqrt{[a_1^{(k)}]^2 - 4a_2^{(k)}}} \exp\left[-\frac{a_1^{(k)}}{2}(\vartheta - t)\right] \times$$

$$\times \operatorname{sh}\left[\frac{1}{2}\sqrt{[a_1^{(k)}]^2 - 4a_2^{(k)}}(\vartheta - t)\right]; \tag{3.10}$$

(c) $[a_1^{(k)}]^2 - 4a_2^{(k)} = 0$,

$$x_2^{(k)} = (\vartheta - t)\exp\left[-\frac{a_1^{(k)}}{2}(\vartheta - t)\right]. \tag{3.11}$$

If $a_2^{(k)} = 0$, i.e. the attractive force acting on the kth point is absent, then

$$x_{12}^{(k)}[\vartheta, t] = \frac{1}{a_1^{(k)}}[1 - \exp[-a_1^{(k)}(\vartheta - t)]]. \tag{3.12}$$

And if the resistance force is also absent, i.e. $a_1^{(k)} = 0$, then

$$x_{12}^{(k)}[\vartheta, t] = \vartheta - t. \tag{3.13}$$

Functions $\varphi(t)$ and $\psi(t)$, $t_0 \leqq t \leqq \vartheta$ in equation (3.5) are determined by the equalities

$$\varphi(t) = x_{12}^{(1)}[\vartheta, t], \qquad \psi(t) = x_{12}^{(2)}[\vartheta, t]. \tag{3.14}$$

Construct the equation

$$h(\tau) \equiv \nu|\psi(\tau)| - \mu|\varphi(\tau)| = 0. \tag{3.15}$$

It has the root $\tau_0^* = \vartheta$. Let this equation have some more real roots τ_s^*, $s = 1, \ldots, p$ on the half-open interval $[t_0, \vartheta)$, so that $\tau_{s+1}^* < \tau_s^*$.

Then, after solution of problems (1.11), (1.10) the following explicit expression for the calculation of the optimal guaranteed result

$$\rho^0(t_*, x_*) = \max\left\{|x_*| + \int_{t_*}^{\vartheta} h(\tau)d\tau; \int_{\tau_j^*}^{\vartheta} h(\tau)d\tau, \quad j : \tau_j^* \in [t_*, \vartheta]\right\}. \tag{3.16}$$

If, for instance, $p = 1$ and the condition

$$h(\tau) < 0, \quad t_0 < \tau < \tau_1^*; \qquad h(\tau) > 0, \quad \tau_1^* < \tau < \vartheta \tag{3.17}$$

is fulfilled, then from equality (3.16) the following simple formula is obtained

$$\rho^0(t_*, x_*) = \begin{cases} |x_*| + \int_{t_*}^{\vartheta} h(\tau)d\tau, & \text{if } t_* \geqq \tau_1^* \text{ or if } t_* < \tau_1^* \text{ and } r(t_*, x_*) > 0; \\ \int_{\tau_1^*}^{\vartheta} h(\tau)d\tau, & \text{if } t_* < \tau_1^* \text{ and } r(t_*, x_*) \leqq 0, \end{cases} \tag{3.18}$$

where

$$r(t_*, x_*) = |x_*| + \int_{t_*}^{\tau_1^*} h(\tau)d\tau, \quad t_* < \tau_1^* \tag{3.19}$$

Note that function $\rho^0(t, x)$ (3.18) is not differentiable with respect to the vector-valued argument x. In the considered case after necessary calculation according to the suggested formulas (2.1), (2.2), (1.11), we can write out the expression for $\rho_\varepsilon^0(t_*, x_*)$ which defines here the approximating function $\rho_\varepsilon^0(t, x)$ (2.1) in explicit form.

With regard to (3.18), (3.19), this expression takes the following form

$$\rho^0(t_*, x_*) = \begin{cases} \rho^0(t_*, x_*) - \sqrt{2}\beta(\varepsilon, t_*), \text{ if } t_* \geqq \tau_1^* \text{ and} \\ |x_*| \geqq \dfrac{1}{\sqrt{2}}\beta(\varepsilon, t_*) \text{ or if } t_* < \tau_1^* \text{ and} \\ r(t_*, x_*) \geqq \dfrac{1}{\sqrt{2}}\beta(\varepsilon, t_*); \\ \rho^0(t_*, x_*) - |x_*| - [\beta^2(\varepsilon, t_*) - |x_*|^2]^{1/2}, \\ \text{if } t_* \geqq \tau_1^* \text{ and } 0 \leqq |x_*| < \dfrac{1}{\sqrt{2}}\beta(\varepsilon, t_*); \\ \rho^0(t_*, x_*) - \beta(\varepsilon, t_*), \text{ if } t_* < \tau_1^* \text{ and} \\ \int_{t_*}^{\tau_1^*} h(\tau)d\tau \leqq r(t_*, x_*) \leqq 0; \\ \rho^0(t_*, x_*) - r(t_*, x_*) - [\beta^2(\varepsilon, t_*) - r^2(t_*, x_*)]^{1/2}, \\ \text{if } t_* < \tau_1^* \text{ and } 0 < r(t_*, x_*) < \dfrac{1}{\sqrt{2}}\beta(\varepsilon, t_*). \end{cases} \tag{3.20}$$

In accordance with the stated theory, boundary condition (2.5) is realized for the function $\rho_\varepsilon^0(t, x)$ (3.20) which has the gradient $\partial\rho_\varepsilon^0(t, x)/\partial x$. Estimate (2.6) is valid and differential inequality (2.9) is fulfilled for the function.

Calculating the gradient of the function $\rho_\varepsilon^0(t, x)$ (3.20) and carrying out the extremal displacement (1.18) in the direction opposite to the gradient, we find that in the considered case the optimal strategy $u^0(\cdot)$ has the following explicit form

$$u^0(t, x, \varepsilon) = \begin{cases} -\mu\dfrac{x}{|x|}\operatorname{sgn}\varphi(t), & |x| \neq 0 \\ 0, & |x| = 0 \end{cases} \tag{3.21}$$

Further analysis shows that in the considered example the optimal strategy $u^0(\cdot)$ preserves its form (3.21) not only if condition (3.17) is realized but in the general case of arbitrary functions $\varphi(t)$ and $\psi(t)$, $t_0 \leqq t \leqq \vartheta$, too.

Here we succeed in constructing of the optimal strategy $v^0(\cdot)$ (2.15) on the basis of the extremal displacement (2.18). Like the strategy $u^0(\cdot)$ (3.21), $v^0(\cdot)$ (2.15) is obtained independent of the parameter ε and has the following form

$$v^0(t, x, \varepsilon) = \begin{cases} -v \dfrac{x}{|x|} \operatorname{sgn} \psi(t), & |x| \neq 0 \\ v^0 : |v^0| = v \ , & |x| = 0 \end{cases} \tag{3.22}$$

The mentioned pair of strategies $\{u^0(\cdot)\,(3.21), v^0(\cdot)\,(3.22)\}$ is said [1–3] to make up the saddle point of differential game which has the value $\rho^0(t_*, x_*)$ calculated by formula (3.16). Optimal control laws U^0 (2.7) and V^0 (2.14) are based on optimal strategies $u^0(\cdot)$. (3.21) and $v^0(\cdot)$ (3.22) were tested on electronic computer at the following initial data:

$$t_0 = 0, \quad \vartheta = 5, \quad t_* = 0, \quad m_1 = 1.0, \quad m_2 = 0.5, \quad \alpha_1 = 0.1,$$

$$\alpha_2 = 0.4, \quad \beta_1 = \beta_2 = 4, \quad \mu_1 = \mu_2 = 1, \quad r_{1*}^{(1)} = r_{2*}^{(1)} = 0,$$

$$r_{3*}^{(1)} = 0.7, \quad \dot{r}_{1*}^{(1)} = \dot{r}_{2*}^{(1)} = -0.2, \quad \dot{r}_{3*}^{(1)} = 0, \quad r_{1*}^{(2)} = r_{2*}^{(2)} = r_{3*}^{(2)} = 0,$$

$$\dot{r}_{1*}^{(2)} = \dot{r}_{2*}^{(2)} = 0.2, \quad \dot{r}_{3*}^{(2)} = 0.0.$$

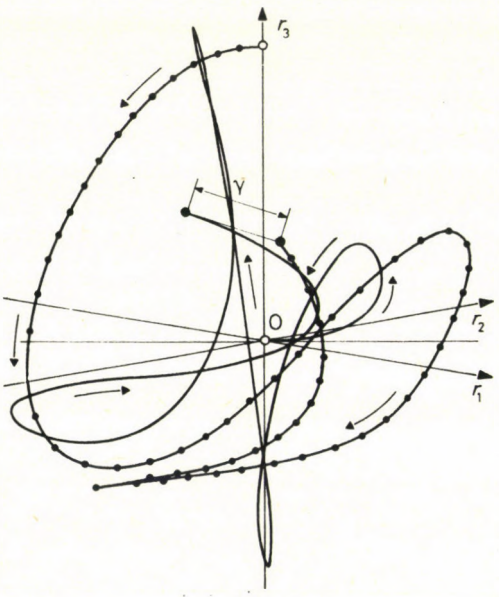


Fig. 1

Case (a) (3.9) corresponds to these data. And equation (3.15) has seven real roots on the segment [0, 5] not coinciding with root $\tau_0^* = \vartheta = 5$, and the game-value calculated by formula (3.16) is equal to $\rho^0 = 0.094$. The trajectories of the object $m^{(1)}$ (denoted by a solid line with circles) and object $m^{(2)}$ (solid line) generated by the optimal control laws are shown on Fig. 1. Initial data and projection type are selected so that shown trajectories entirely lie in the picture plane. In both control laws step δ was taken constant and equal to 0.001. According to the theory the distance between the points at instant of time ϑ turned out to be equal to $\gamma = 0.095 \approx \rho^0 = 0.094$.

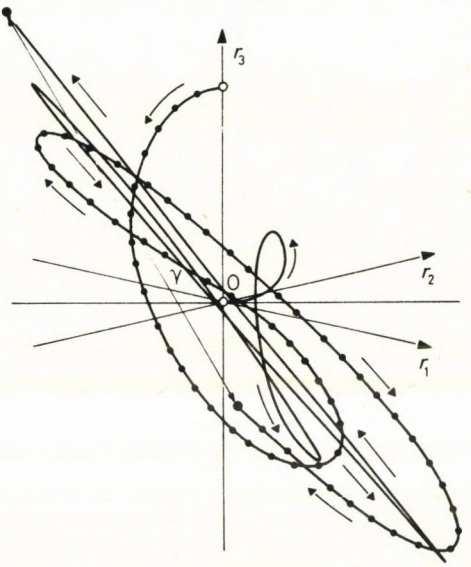


Fig. 2

On Fig. 2 trajectories of the objects are represented in the case when the second player evades convergence in optimal way, and the first one at every instant of time $t \in [0, 5]$ acts upon the object $m^{(1)}$ by the force of maximal value μ_1 and directed to the object $m^{(2)}$. Since actions of the first player are not optimal in this case, we have that $\gamma = 1.311 > \rho^0 = 0.094$.

Finally, on Fig. 3 there are objects' trajectories starting from the position

$$t_* = 0, \quad r_*^{(1)} = 0.0, \quad r_{2*}^{(1)} = 0.1, \quad r_{3*}^{(1)} = 0.2, \quad \dot{r}_{1*}^{(1)} = 0.5, \quad \dot{r}_{2*}^{(1)} = \dot{r}_{3*}^{(1)} = 0;$$

$$r_{1*}^{(2)} = 0.0, \quad r_{2*}^{(2)} = 0.1, \quad r_{3*}^{(2)} = 0.0, \quad \dot{r}_{1*}^{(2)} = 0.5, \quad \dot{r}_{2*}^{(2)} = \dot{r}_{3*}^{(2)} = 0$$

in the case when the motion of the first object $m^{(1)}$ are generated by the optimal control law, and the second player sets $v(t, x) \equiv 0$, i.e. object $m^{(2)}$ moves in a logarithmic

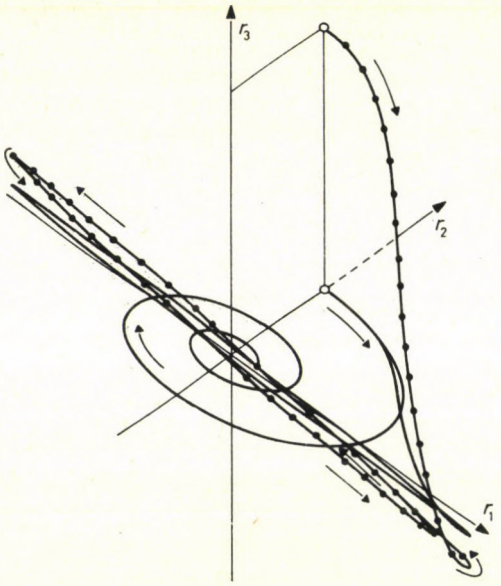


Fig. 3

spiral in the plane $\{r_1, r_2\}$ under the action of the given forces. As it could be expected, in this case we obtained that $\gamma = 0.000 < \rho^0 = 0.094$.

The results of numerical experiments show that the optimal control law constructed on the basis of the approach suggested in the paper, can be reliably realized on electronic computers.

References

1. *Krasovskii, N. N.*, Control of dynamic system. Moscow: Nauka, 1985, 520 pp. (in Russian)
2. *Krasovskii, N. N., Tret'jakov, V. E.*, On program synthesis of a positional control. Dokl. Akad. Nauk SSSR, 1982, **264**, 6, pp. 1309–1312 (in Russian); English transl. in Soviet Math. Dokl., 1982, **25**, 3.
3. *Krasovskii, A. N., Tret'jakov, V. E.*, Stochastic program synthesis in differential game with integral payoff. Prikl. mat. meh., 1984, **48**, 6, pp. 883–891 (in Russian).
4. *Subbotin, A. I., Subbotina, N. N.*, On sensitivity of the value function of the differential game with the integral-terminal payoff. Problems of Control and Information Theory, 1983, **12**, 3, pp. 153–166.
5. *Krasovskii, N. N., Tret'jakov, V. E.*, Control problems with guaranteed result. Sverdlovsk: Sredne-Uralskoje Knizhnoje Izdatelstvo, 1986. 64 pp. (in Russian)
6. *Shirjaev, A. N.*, Probability. Moscow: Nauka, 1980. 576 pp. (in Russian)
7. *Krasovskii, N. N.*, Determinate strategy and stochastic programs. Prikl. mat. meh., 1985, **49**, 2, pp. 179–198 (in Russian).

Синтез оптимального гарантирующего управления

В. Е. ТРЕТЬЯКОВ

(Свердловск)

В работе предлагается алгоритм вычисления оптимальных управляющих воздействий в задаче позиционного управления при условии минимума гарантированного результата относительно показателя

$$\gamma = |x[\vartheta]|$$

для системы

$$\frac{dx}{dt} = A(t)x + B(t)u + C(t)v, \quad t_0 \leqq t \leqq \vartheta,$$

где x — фазовый вектор объекта; u и v — соответственно векторы управления и помехи, стесненные геометрическими ограничениями.

На основе правила экстремального сдвига объекта на сопутствующую точку, конструкций стохастического программного синтеза и введения дополнительной регуляризирующей фазовой координаты строится функция, которая аппроксимирует оптимальный гарантированный результат и является дифференцируемой по фазовым переменным. При этом универсальная оптимальная стратегия получается путем экстремального сдвига против градиента аппроксимирующей функции. В ряде случаев данный подход приводит к реализуемой вычислительной процедуре. Приводится иллюстрирующий пример.

Полученные результаты проясняют связь метода стохастического программного синтеза с известными обобщениями метода динамического программирования.

В. Е. Третьяков
Уральский государственный университет им. А. М. Горького
СССР, 620083, Свердловск, К-83, пр. Ленина, 51.

Problems of Control and Information Theory, Vol. 17 (4), pp. 223—230 (1988)

BOUNDS ON CONSTANT WEIGHT BINARY SUPERIMPOSED CODES

NGUYEN QUANG A, T. ZEISEL

(Hanoi), *(Budapest)*

(Received October 7, 1987)

The asymptotic bounds on the minimal length of binary superimposed codes are investigated. We give an improvement for the Dyachkov–Rykov upper bound. The bounds of the classical coding theory for this kind of codes are applied, too. The Reed–Solomon code with componentwise binary mapping can asymptotically reach the lower bound under certain condition.

1. Basic definitions and known bounds

Let $A=(x_i)_1^M$ be a set of N-dimensional binary vectors, the Boolean sum of $x \in A$ is an N-dimensional binary vector $y=\bigvee_{i=1}^{M} x_i$ defined by

$$y_j=\begin{cases}1 & \text{if } x_{i_j}=1 \text{ for some } i=1,2,\ldots,M\\ 0 & \text{if } x_{i_j}=0 \text{ for all } i=1,2,\ldots,M\end{cases}$$

where $x_i=(x_{i_1},\ldots,x_{i_N})$, $i=1,2,\ldots,M$ and $y=(y_1,\ldots,y_N)$.

We say, x vector covers y if $x \vee y=x$.

A code C is *Zero false dropping* (*ZFD*) of order M if the Boolean sum of any M-subset of C can cover an $x \in C$ if and only if x belongs to the subset [1]. The cardinality of C is denoted by T.

There are lower and upper bounds of the minimal length $N(M, T)$ for given T and M [2], [3], [4].

Our aim is to investigate the asymptotic behaviour of these bounds looking for the minimal length of the code as T tends to infinity and M is fixed value.

First we consider the known results.

Theorem 1 (Bassalygo bound [2], [3]). For all $T \geqq M$

$$N(M, T) \geqq \min\left\{T, \frac{(M+1)(M+2)}{2}\right\}.$$

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

The main consequence of the theorem is that for any $M > \sqrt{2T}$, $N(T, M) = T$, so in this case the trivial code is the optimal, whose codewords create a $T \times T$ diagonal matrix.

Theorem 2 (Dyachkov–Rykov [2], [3]). For fixed $M \gg 1$ and $T \to \infty$

$$N(M, T) \geqq \frac{M^2}{2 \log M} \log T$$

(for log the base is 2 and for ln the base is e).

Theorem 3 (Dyachkov–Rykov [2], [3]). For fixed $M \gg 1$ and $T \to \infty$

$$N(M, T) \leqq \frac{e}{\log e} M^2 \log T \approx 1.8842\, M^2 \log T.$$

2. Main results

We get a better upper bound if we use a random code with alphabet $1, 2, \ldots, L$, and a componentwise transformation of the codewords into binary vectors by means of the following simple mapping:

$$\begin{aligned} 1 &\text{ ----- } 0, 0, \ldots 0, 1 \\ 2 &\text{ ----- } 0, 0, \ldots 1, 0 \\ &\cdot \qquad\quad \cdot \qquad \cdot \\ L &\text{ ----- } 1, 0, \ldots 0, 0 \end{aligned} \tag{1}$$

The interesting feature of this random construction is that it is an upper bound on the minimal length of the constant weight ZFD code.

Theorem 4. For any size T and order M

$$N(M, T) \leqq K(M, T)\,(M+1)^2 \log T \tag{2}$$

where

$$K(M, T) = \min_{M \leqq L} \ln 2 \quad \frac{1}{-\frac{M+1}{L} \ln\left(1 - \left(1 - \frac{1}{L}\right)^M\right)} + \frac{L}{(M+1)^2 \log T} \quad .$$

Proof. Let us examine a random code with alphabet $1, 2, \ldots L$ and length N/L having independent, uniformly distributed components, mapped by (1) to binary sequences with length N where $L \geqq M$. Then

$$P(C \text{ is not } ZFD) \leqq \binom{T}{M} (T - M) \left[1 - \left(1 - \frac{1}{L}\right)^M\right]^{N/L} \leqq$$

$$\leqq \exp\left[(M+1) \ln T + \left(\frac{N}{L} - 1\right) \ln\left[1 - \left(1 - \frac{1}{L}\right)^M\right]\right]. \tag{3}$$

If this probability is less than one then a *ZFD* code of order M exists, which is guaranteed if N satisfies (2).

Corollary 1. Asymptotically, as M is fixed and $T\to\infty$

$$N(M, T)\leqq K(M)(M+1)^2 \log T(1+o(1)) \tag{4}$$

where

$$K(M)=\min_{M\leqq L}\frac{\ln 2}{-\frac{M+1}{L}\ln\left(1-\left(1-\frac{1}{L}\right)^M\right)}\leqq\frac{\ln 2}{-\alpha\ln(1-e^{-\alpha})}\leqq\frac{\ln 2}{\ln\frac{e}{e-1}}\approx 1.5112$$

where $\alpha=\dfrac{M+1}{\left\lfloor\frac{M+1}{\ln 2}\right\rfloor}$, and

$$\limsup_{M\to\infty}K(M)=\frac{1}{\ln 2}\approx 1.4427.$$

Proof. Let us choose $L=\lfloor(M+1)\ln 2\rfloor$ in (2) and use the inequality

$$\left(1-\frac{1}{L}\right)^M\geqq\exp\left(-\frac{M+1}{L}\right)\quad\text{if}\quad M\leqq L.\qquad\text{(see [5])}$$

If $M\to\infty$ then α tends to $\ln 2$ and the proof is complete.

For comparison the results of Dyachkov–Rykov, we calculate $K(M)$ for small M in Table 1.

Table 1

M	$K(M)(M-1)^2$	Dyachkov–Rykov constant
3	19.2448	24.8762
4	31.2615	40.5487
5	46.2371	59.9883
6	64.0031	83.1955

This upper bound is slightly better than the Dyachkov–Rykov bound and we can see that code concatenation is a good way to construct superimposed codes.

Kautz and Singleton proposed constant weight binary codes to construct superimposed codes.

Let us denote the parameters of the code in the following way:

w — the weight of the code

c — the maximal Hamming correlation: $c=\max\limits_{i\neq j}(x_i x_j)$ where $x_i, x_j\in C$ (where $(x_i x_j)$ is the usual scalar product).

Theorem 5 (Kautz–Singleton). A binary constant weight code with parameters T, w, c, has the $ZFD(N, M, T)$ property if the order

$$M \leqq \left\lfloor \frac{w-1}{c} \right\rfloor.$$

We shall call the code satisfying this property Kau–Sin code. Let us take the weight $w = \alpha N/M$ where $\alpha = \alpha(M)$ and in accordance with Theorem 5, let $c = \left\lfloor \frac{w-1}{M} \right\rfloor$.

In this way the task is to construct a large constant weight binary code with weight $\alpha N/M$ and maximal Hamming correlation $\lfloor (w-1)/M \rfloor$.

There are the following bounds for these codes with parameters T, N, w, c:

Johnson bound 1 [1], [10]: for all code with parameters T, N, w, c

$$T \leqq \frac{\binom{N}{c+1}}{\binom{w}{c+1}}. \tag{5}$$

Johnson bound 2 [11]: for all code with parameters T, N, w, c and $w^2 > Nc$

$$T \leqq N \frac{w-c}{w^2 - Nc}. \tag{6}$$

Corollary of the Gilbert–Varshamov bound (see in Theorem 7 [12]) there exists a code with parameters T, N, w, c if

$$T \geqq \frac{\binom{N}{w}}{\sum_{i=0}^{w-c-1} \binom{w}{i}\binom{N-w}{i}}. \tag{7}$$

In order to simplify the formulas, in the sequel we ignore the symbol of integer part if it does not result in a confusion.

Theorem 6. For $w \geqq 2\frac{N}{M}$ Kau–Sin code does not exist.

Proof. Using the expression $w = \alpha \frac{N}{M}$ with $\alpha \geqq 2$, in accordance with the condition.

Satisfying the Kau–Sin code condition $c \leqq \left\lfloor \frac{w-1}{M} \right\rfloor < \frac{w}{M}$.

Considering $D = w^2 - Nc$

$$D = \left(\frac{\alpha N}{M}\right)^2 - N\left\lfloor \frac{w-1}{M} \right\rfloor \geqq \left(\frac{\alpha N}{M}\right)^2 - N\frac{\alpha N}{M^2} = \alpha \frac{N^2}{M^2}(\alpha - 1) > 0 \text{ if } \alpha > 1.$$

In this case the Johnson bound 2 (6) must be valid.

Because expression (6) is increasing with c for $D>0$, we get

$$T \leqq N \frac{\frac{\alpha N}{M} - \frac{\alpha N}{M^2}}{\left(\frac{\alpha N}{M}\right)^2 - \frac{\alpha N^2}{M^2}} = \frac{M-1}{\alpha - 1} \leqq M-1 \quad \text{if} \quad \alpha \geqq 2.$$

This is a contradiction which completes the proof. The proof of Theorem 6 implies the following:

Corollary 2. If $M>1$ is fixed, then for all $\alpha>1$ there is no Kau–Sin code with $T> \frac{M-1}{\alpha-1}$, thus for $T\to\infty$ we must have $\alpha \leqq 1$.

Theorem 7. If a Kau–Sin code exists as $T\to\infty$, $M \gg 1$ fixed, then its length:

$$N(M, T) \geqq \frac{M^2}{\alpha \log M} \log T \tag{8}$$

where $\alpha = \frac{wM}{N} \leqq 1$.

Proof. For the asymptotic behaviour the following lemma is required which is a straightforward consequence of the Stirling formula:

Lemma 1. For any sequence k_n where $\lim_{n\to\infty} k_n/n = q$

$$\lim_{n\to\infty} \frac{1}{n} \log \binom{n}{k_n} = h(q)$$

where $h(x) = -x \log x - (1-x)\log(1-x)$ is the binary entropy function.

Applying Lemma 1 and Johnson bound 1 (5)

$$N(M, T) \geqq \frac{M^2}{\alpha \log \frac{M}{\alpha} - (M^2-\alpha)\log\left(1-\frac{\alpha}{M}\right) + \alpha(M-1)\log\left(1-\frac{1}{M}\right)} \log T.$$

For the case $M \gg 1$

$$N(M, T) \geqq \frac{M^2}{\alpha \log M} \log T.$$

This completes the proof.

This result is slightly weaker then Dyachkov–Rykov lower bound (see Corollary 5 in [2]), since because of Corollary 2, α must be less than 1.

An upper bound can be given by means of Gilbert–Varshamov bound (7).

Theorem 8. If $T \to \infty$, $M \gg 1$ fixed, there exists a Kau–Sin code with length

$$N(M, T) \leqq 4.28 M^2 \log T. \tag{9}$$

This result is the same one that was given by Dyachkov and Rykov via random coding (Theorem 10 in [2]). We give another proof using the Gilbert–Varshamov bound. The proof of the theorem requires the following lemma:

Lemma 2. If $w = \alpha N/M$, $c = \alpha N/M^2 - 1$, $\alpha < 1$ then:

$$\frac{\binom{N}{w}}{\sum_{i=0}^{w-c-i} \binom{w}{i}\binom{N-w}{i}} \leqq \frac{\binom{N}{w}}{(w-c)\binom{w}{c+1}\binom{N-w}{w-c-1}}.$$

Proof. The following equality can be easily checked:

$$\sum_{i=0}^{w-c-1} \binom{w}{i}\binom{N-w}{i} = \sum_{i=c+1}^{w} \binom{w}{i}\binom{N-w}{w-i}.$$

Let us examine the sequence

$$a_i = \binom{w}{i}\binom{N-w}{w-i}.$$

We prove that this sequence is decreasing in $c+1 \leqq i \leqq w$

$$\frac{a_{i+1}}{a_i} = \frac{(w-i)^2}{(i+1)(N-2w+i+1)} \leqq 1.$$

As $w = \alpha N/M$, $c = \alpha N/M^2 - 1$, $\alpha < 1$ we get:

$$\alpha \frac{\alpha - 1}{M^2} N^2 + \left(\frac{2\alpha}{M} - \frac{2\alpha}{M^2} - 1\right) N - 1 \leqq 0.$$

If $\alpha < 1$, the roots are negative and the condition is satisfied. Substituting all components for the maximal a_{c+1} in the summation of the expression, the statement of the Lemma is given.

Proof of Theorem 8. From the Gilbert–Varshamov bound, Lemma 1 and Lemma 2:

$$N(M, T) \leqq \frac{M^2}{M^2 \log \frac{1 - 2\frac{\alpha}{M} + \frac{\alpha}{M^2}}{1 - \frac{\alpha}{M}} + 2\alpha M \log \frac{\left(1 - \frac{\alpha}{M}\right)\left(1 - \frac{1}{M}\right)}{1 - 2\frac{\alpha}{M} + \frac{\alpha}{M^2}} + \alpha \log \frac{1 - 2\frac{\alpha}{M} + \frac{\alpha}{M^2}}{\alpha\left(1 - \frac{1}{M^2}\right)}} \log T.$$

If $M \gg 1$

$$N(M, T) \leqq \frac{\ln 2}{-\alpha \ln \alpha - \alpha(1-\alpha)} M^2 \log T.$$

Taking the minimum as $\alpha \approx 0.2031$, the statement of the theorem is given.

In [4] a slightly weaker result is derived from the Gilbert–Varshamov bound (7).

In combinatorics some of the graph- and set-covering problems are similar to superimposed codes. We can find some various bounds in this field, too ([6], [7]), but unfortunately it can be proven that these bounds are not better than the results mentioned in this paper.

If we allow that $M \to \infty$ much more slowly than $T \to \infty$ then the lower bound can be approached asymptotically up to a constant factor.

Theorem 9. If $M < aT^{1/k}$ for some integer $k > 2$ and constant a then as $M \to \infty$ there exists a Kau–Sin code for which asymptotically

$$N(M, T) = (k-1)\frac{M^2}{\log M} \log T. \tag{10}$$

Proof. Kautz and Singleton [1] as well as Dyachkov and Rykov [2] constructed *ZFD* code via *MDS* code with length n and k information characters by the use of trivial binary mapping (1) with the following parameters:

$$T = q^k$$

$$N = qn$$

$$M = \frac{n-1}{k-2}.$$

Let k be fixed as $T \to \infty$, the proof is a straightforward consequence of Zinoviev's result [9] and Theorem 7.

Finally, we summarize the results and conclusions. We saw as $T \to \infty$ and M is fixed, the bounds for superimposed codes differ only by a constant from the bounds given for constant weight binary codes characterized by maximal Hamming correlation. We saw the good choice of the weight for a Kau–Sin code is $\alpha N/M$ where $\alpha < 1$. In Theorem 4 there is an improvement of the Dyachkov–Rykov upper bound by means of code concatenation.

References

1. *Kautz, W. H., Singleton, R. C.*, Nonrandom Binary Superimposed Codes. IEEE Trans. Inf. Theory, **IT-10**, *4*, pp. 363–377, 1964.
2. *Dyachkov, A. G., Rykov, V.V.*, A Survey of Superimposed Code Theory. Problems of Control and Information Theory, **12**, *4*, pp. 229–244, 1983.
3. *Dyachkov, A. G., Rykov, V. V.*, Bounds on the Length of Disjunctive Codes. Probl. Pered. Inf., **18**, *3*, pp. 166–171, 1982.
4. *Ericson, T.*, Performance Limitation for Multiple Access Communication Without Feedback. Report in University Linköping, Sweden, 1986.

5. *Győrfi, L., Kerekes, I.*, A Block code for Noiseless Asynchronous Multiple Access OR Channel. IEEE Trans. on Inf. Theory, **IT-27**, *6*, pp. 363–377, 1981.
6. *Erdős, P., Frankl, P., Füredi, Z.*, Families of Finite Sets in which No Set Is Covered by the Union of *r* Others. Israel J. of Mat., **51**, *1–2*, pp. 79–89, 1985.
7. *Rödl, V.*, On a Packing and covering Problem. Eur. J. Combinatorics, **5**, 1984.
8. *Erdős, P., Hanani, H.*, On a Limit Theorem in Combinatorial Analysis. Publ. Math. Debrecen, **10**, pp. 10–13, 1963.
9. *Zinoviev, V.*, Cascade Equal-Weight Codes and Maximal Packing. Problems of Control and Information Theory, **12**, *1*, pp. 3–10, 1983.
10. *Johnson, S. M.*, A New Upper Bound for Error Correcting Codes. IEEE Trans on Inf. Theory **IT-8**, pp. 203–207, 1962.
11. *MacWilliams, Sloane, N. J.*, The Theory of Error Correctiong Codes. North Holland Publ. Comp., 1977.
12. *Graham, R. L., Sloane, N. J.*, Lower Bounds for Constant Weight codes. IEEE. Trans. on Inf. Theory, **IT-26**, pp. 37–43, 1980.

Границы для двоичных наложенных кодов с постоянным весом

НГУЕН КВАНГ А, Т. ЗЕЙЗЕЛ
(Ханой, Будапешт)

Рассматриваются асимптотические границы для минимальной длины двоичных наложенных кодов. Дается улучшение верхней границы Дьячкова–Рыкова. Границы классической теории кодирования применяются также для этого случая.

Nguyen Quang A
T. Zeisel
Technical University of Budapest
H-1111 Budapest, Stoczek u. 2.
Hungary

Problems of Control and Information Theory, Vol. 17 (4), pp. 231—238 (1988)

ON THE LIPSCHITZEAN DEPENDENCE OF TRAJECTORIES OF MULTI-INPUT TIME DEPENDENT BILINEAR SYSTEMS ON CONTROLS

S. ČELIKOVSKÝ

(Prague)

(Received February 9, 1987)

A multi-input time dependent bilinear system is considered. An estimate for the Lipschitzean dependence of trajectories of the bilinear systems on controls in certain norms is developed. This estimate is then used to study some interesting properties of attainable sets of bilinear systems. Possible numerical applications of the Lipschitzean dependence are briefly described.

1. Introduction

Let us consider the following control system

$$\dot{x} = (A(t) + \sum_{k=1}^{m} B_k(t)u^k(t))x + C(t)u(t)$$

$$x(t_0) = x_0, \qquad t \in [t_0, t_1] \subset R, \tag{1}$$

which will be further denoted as a multi-input time dependent bilinear system. Here $A(t)$, $B_k(t)$, $k = \overline{1, m}$, are $(n \times n)$-dimensional matrix-valued functions, $C(t)$ is $(n \times m)$-dimensional matrix-valued function. The control (or input) $u = (u^1, \ldots, u^m)^T \in \in R^m$ is assumed to be a measurable function on every finite time interval $[t_0, t_1]$ and such that, for each $k = 1, \ldots, m$ almost everywhere (further a.e.) on $[t_0, t_1]$, holds $a^k \leqq u^k(t) \leqq b^k$, where a^k, b^k, $k = 1, \ldots, m$ are given real numbers. Such a control is further denoted as an admissible one. Finally, $x \in R^n$ is the vector of the state variables, and $x_0 \in R^n$ is the given initial state of the system.

Bilinear systems have been studied widely and have been shown to be an important extension to linear systems with various possible applications (see e.q. [1], [2], [3]). The aim of this contribution is to derive an estimate which establishes Lipschitzean dependence of trajectories of bilinear systems on controls in certain norms, namely

$$\max_{t \in [t_0, t_1]} \|x_1(t) - x_2(t)\|_{R^n} \leqq \sum_{k=1}^{m} L_k \max_{t \in [t_0, t_1]} \left| \int_{t_0}^{t} u_1^k(s)ds - \int_{t_0}^{t} u_2^k(s)ds \right|. \tag{2}$$

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

Here $x_1(t)$ and $x_2(t)$, $t \in [t_0, t_1]$ are trajectories corresponding to admissible controls $u_1(s)=(u_1^1(s), \ldots, u_1^m(s))^T$ and $u_2(s)=(u_2^1(s), \ldots, u_2^m(s))^T$, $s \in [t_0, t_1]$, respectively; L_k, $k=1, \ldots, m$, are certain constants, depending only on the parameters of system (1).

Estimate (2) is then used to study some important properties of the so-called attainable set of bilinear systems and some ideas of further possible applications of this estimate are given.

2. Estimate for Lipschitzean dependence

We derive estimate (2) by generalization of the corresponding result for single input time dependent bilinear systems. This case is considered in [4, Theorem 3]:

Theorem 1. Let us consider the time-dependent bilinear system with single input:

$$\dot{x}=(A(t)+B(t)u(t))x+C(t)u(t)+f(t),$$

$$x(t_0)=x_0, \quad t \in [t_0, t_1], \quad x_0 \in R^n, \quad u(t) \in [a, b] \quad \text{a.e. on} \quad [t_0, t_1], \tag{4}$$

where $A(t)$, $B(t)$ are $(n \times n)$-dimensional matrix-valued functions, $c(t)$, $f(t)$ are functions with values in R^n. We impose the following assumptions:

(i) $B(t)$, $c(t)$ are absolutely continuous and almost everywhere on $[t_0, t_1]$

$$\|B(t)\|_s \leqq B^M, \quad \|B'(t)\|_s \leqq B^{DM}, \quad \|c(t)\|_{R^n} \leqq c^M,$$

$$\|c'(t)\|_{R^n} \leqq c^{DM}.$$

(ii) $A(t)$, $f(t)$ are essentially bounded measurable functions and a.e. on $[t_0, t_1]$

$$\|A(t)\|_s \leqq A^M, \quad \|f(t)\|_{R^n} \leqq f^M.$$

Here $\|\cdot\|_s$ stands for the spectral matrix norm and $\|\cdot\|_{R^n}$ for the Euclidean vector norm in R^n.

(iii) For every $t', t'' \in [t_0, t_1]$ matrices $B(t')$ and $B(t'')$ commute, i.e., $B(t')B(t'')= = B(t'')B(t')$.

Let us $x_1(t)$ and $x_2(t)$ be trajectories of system (4) for admissible controls $u_1(t)$ and $u_2(t)$, respectively. Then

$$\max_{t \in [t_0, t_1]} \|x_1(t)-x_2(t)\|_{R^n} \leqq K \max_{t \in [t_0, t_1]} \left| \int_{t_0}^{t} u_1(s)ds - \int_{t_0}^{t} u_2(s)ds \right|, \tag{5}$$

where

$$K=K(\|x_0\|_{R^n}, u_p, A^M, B^M, c^M, B^{DM}, c^{DM}, f^M, t_0, t_1)=$$

$$=K_1K_2\|x_0\|_{R^n}+2K_1K_2^2(K_3+K_4K_5u_p(t_1-t_0))K_4K_6(t_1-t_0)+$$

$$+K_2^2K_4(1+K_4K_5(1+B^Mu_p(t_1-t_0)))K_6(t_1-t_0)+$$

$$+K_2^2K_4(K_3+K_4K_5u_p(t_1-t_0))B^{DM}c^M(t_1-t_0)+$$

$$+K_4(1+K_4K_5(1+B^Mu_p(t_1-t_0)))c^M+2K_1K_2^2(t_1-t_0)f^M. \tag{6}$$

Here the following notations are used:

$$K_1=(B^M+(t_1-t_0)B^{DM})(1+A^M(t_1-t_0))+B^MA^M(t_1-t_0) \tag{7}$$

$$K_2=\exp((B^Mu_p+A^M)(t_1-t_0)) \tag{8}$$

$$K_3=\{\exp(B^Mu_p(t_1-t_0))-1\}/B^M \tag{9}$$

$$K_4=\exp(B^Mu_p(t_1-t_0)) \tag{10}$$

$$K_5=u_p(t_1-t_0)^2B^{DM} \tag{11}$$

$$K_6=A^Mc^M+B^{DM}c^Mu_p(t_1-t_0)+c^{DM} \tag{12}$$

$$u_p=\max\{|a|,|b|\}. \tag{13}$$

Proof of this theorem can be found in [4] in detail.

Remark 1. Condition (iii) is the only commutativity condition which is neccessary. Note that there are no commutativity conditions between $A(t)$ and $B(t)$. Condition (iii) is fulfilled e.g. when $B(t)=B\cdot g(t)$, where B is an arbitrary constant matrix and $g(t)$ is an arbitrary scalar function. So, we can see that assumptions of Theorem 1 may be considered as quite general.

Now we generalize Theorem 1 to the multi-input case.

Theorem 2. Let us consider system (1) with initial condition $x(t_0)=x_0$. We impose the following assumptions:

(i) $B_k(t)$, $k=1,\ \ldots,\ m$, $C(t)$ are absolutely continuous and almost everywhere on $[t_0, t_1]$

$$\|B_k(t)\|_s\leqq B_k^M,\quad \|B_k'(t)\|_s\leqq B_k^{DM},\quad \|c_k(t)\|_{R^n}\leqq c_k^M,$$

$\|c_k'(t)\|_{R^n}\leqq c_k^{DM}$, where $c_k(t)$ is the kth column of $C(t)$.

(ii) $A(t)$ is an essentially bounded measurable function and almost everywhere on $[t_0, t_1]$: $\|A(t)\|_s\leqq A^M$.

(iii) For every $k=1,\ \ldots,\ m$, $t', t''\in[t_0, t_1]$ it holds that $B_k(t'')B_k(t')=B_k(t')B_k(t')$.

Let $x_1(t)$ and $x_2(t)$ be trajectories of system (4) for admissible controls $u_1(t)=(u_1^1(t),\ \ldots, u_1^m(t))^T$ and $u_2(t)=(u_2^1(t),\ \ldots,\ u_2^m(t))^T$, respectively. Then estimate (2) is valid, where

$$L_k=K\left(\|x_0\|_{R^n}, u_p^k, A^M+\sum_{i\neq k}B_i^Mu_p^i, B_k^M, c_k^M, B_k^{DM}, c_k^{DM}, \sum_{i\neq k}c_i^Mu_p^i, t_0, t_1\right). \tag{14}$$

Here K is given by (6)–(13), and $u_p^k=\max\{|a^k|,|b^k|\}$, $k=1,\ \ldots,\ m$.

Proof. We introduce a chain of admissible controls $\tilde{u}_1(t)$, $\tilde{u}_2(t)$, ..., $\tilde{u}_{m+1}(t)$, where $\tilde{u}_1(t)=u_1(t)$, $\tilde{u}_{m+1}(t)=u_2(t)$ and

$$\tilde{u}_i(t)=(u_2^1(t), \ldots, u_2^{i-1}(t), u_1^i(t), \ldots, u_1^m(t))^T, \quad i=2, \ldots, m.$$

Let us denote the respective trajectories as $\tilde{x}_1(t)$, ..., $\tilde{x}_{m+1}(t)$. Then we can write

$$x_1(t)-x_2(t)=\sum_{i=1}^{m}(\tilde{x}_i(t)-\tilde{x}_{i+1}(t)). \tag{15}$$

From (15) it follows that

$$\|x_1(t)-x_2(t)\|_{R^n} \leqq \sum_{i=1}^{m}\|\tilde{x}_i(t)-\tilde{x}_{i+1}(t)\|_{R^n}. \tag{16}$$

Now we obtain an estimate for each $\|\tilde{x}_i(t)-\tilde{x}_{i+1}(t)\|_{R^n}$, $i=1, \ldots, m$. By construction $\tilde{x}_i(t)$ and $\tilde{x}_{i+1}(t)$ are solutions of system (1) with controls $\tilde{u}_i(t)$ and $\tilde{u}_{i+1}(t)$, respectively. We can rewrite system (1) as follows

$$\dot{x}=(\tilde{A}_i(t)+\tilde{B}_i(t)u^i(t))x+\tilde{c}_i(t)u^i(t)+\tilde{f}_i(t), \qquad x(t_0)=x_0, \tag{17}$$

where $\tilde{A}_i(t)=A(t)+\sum_{k\neq i}u^k(t)B_k(t)$, $\tilde{B}_i(t)=B_i(t)$,

$\tilde{c}_i(t)$ is the ith column of $C(t)$ and

$\tilde{f}_i(t)=\sum_{k\neq i}\tilde{c}_k(t)u^k(t)$.

So, we consider system (1) as a system with single-input — ith component $u^i(t)$ of multi-input $(u^1(t), \ldots, u^i(t), \ldots, u^m(t))^T$. Observe that $\tilde{u}_i(t)$ and $\tilde{u}_{i+1}(t)$ differs from each other only in the ith component, i.e., we can consider $\tilde{x}_i(t)$ and $\tilde{x}_{i+1}(t)$ as solutions of system (17) with single inputs $u_1^i(t)$ and $u_2^i(t)$, respectively. Then, according to Theorem 1, we obtain

$$\max_{t\in[t_0,t_1]}\|\tilde{x}_i(t)-\tilde{x}_{i+1}(t)\|_{R^n} \leqq \tilde{K}_i \max_{t\in[t_0,t_1]}\left|\int_{t_0}^{t}u_1^i(s)ds-\int_{t_0}^{t}u_2^i(s)ds\right|. \tag{18}$$

Taking into the account that

$$\|\tilde{A}_i(t)\|_s \leqq A^M+\sum_{k\neq i}B_k^M u_p^k,$$

$$\|\tilde{B}_i(t)\|_s \leqq B_i^M, \quad \|\tilde{c}_i(t)\|_{R^n} \leqq c_i^M, \quad \|\tilde{c}_i'(t)\|_{R^n} \leqq c_i^{DM}, \quad \|\tilde{f}_i\|_{R^n} \leqq \sum_{k\neq i}c_k u_p^k, \|\tilde{B}_i'(t)\|_s \leqq B_i^{DM},$$

we may conclude that $\tilde{K}_i=L_i$ given by (14).

Now the proof of the theorem evidently follows from (16) and (18).

3. Properties of attainable sets for bilinear systems

In this section we use the result of the preceding section in order to study some important properties of the so-called attainable sets for bilinear systems. First, let us recall the definition of an attainable set.

Definition 1. Let us consider system (1) with initial state $x(t_0)=x_0$. We call an attainable set for this system at time $t\in[t_0, t_1]$ the set of all points y from R^n for which there exists an admissible control $u(s)$ on $[t_0, t_1]$, such that for the respective trajectory $x(s)$, $s\in[t_0, t]$, holds $x(t)=y$. We denote attainable set at time t by $X(t_0, t, x_0)$.

Definition 2. Let us consider system (1) with initial state $x(t_0)=x_0$. By $X^*(t_0,t, x_0)$ we denote the set of all points $y\in R^n$ for which there exists a piecewise constant control $u(s)$ on $[t_0, t]$, $u^i(s)\in\{a^i, b^i\}$, $i=1, \ldots, m$, for all $s\in[t_0, t]$, such that for the respective trajectory $x(s)$, $s\in[t_0, t]$, it holds that $x(t)=y$.

It is obvious that $X^*(t_0, t, x_0)\subset X(t_0, t, x_0)$.

The following theorem establishes the fundamental property of an attainable set for bilinear systems (1).

Theorem 3. Let system (1) with initial state $x(t_0)=x_0$ be given. Then for any $t\in[t_0, t_1]=X^*(t_0, t, x_0)$ is dense in $X(t_0, t, x_0)$.

In order to prove the assertion of Theorem 3, we need to establish that for any admissible control $u(s)$, $s\in[t_0, t_1]$, and for any $\varepsilon>0$ there exists $u_\varepsilon(s)$ piecewise constant on $[t_0, t]$ and $u_\varepsilon^i(s)\in\{a^i, b^i\}$, $i=1, \ldots, m$, for all $s\in[t_0, t_1]$, such that for the corresponding trajectories $x(t)$, $x_\varepsilon(t)$ of system (1) holds:

$$\|x(t)-x_\varepsilon(t)\|_{R^n}\leqq\varepsilon.$$

This fact evidently follows from estimate (2) and the following lemma.

Lemma 1. Let us consider a scalar function $u(s)$ measurable on a closed interval $[t_0, t_1]$ such that $u(s)\in[a, b]$ a.e. on $[t_0, t_1]$. Let us divide closed interval $[t_0, t_1]$ into l closed equal subintervals $[t_0+(i-1)h, t_0+ih]$, $i=1, 2, \ldots, l$, $h=(t_1-t_0)/l$. Then there exists a function $u^*(s)$ with the following properties:

1) $u^*(s)$ is constant on each subinterval of the form $[t_0+(i-1)h, t_0-ih]$,
2) $u^*(s)\in\{a, b\}$ for all $s\in[t_0, t_1]$
3) for all $t\in[t_0, t_1]$

$$\left|\int_{t_0}^{t}u(s)ds-\int_{t_0}^{t}u^*(s)ds\right|\leqq\frac{b-a}{2}h. \tag{19}$$

The proof of this lemma is performed in [5] in detail. From Lemma 1 and Theorem 2 we can, in fact, derive stronger results as we did in Theorem 3.

Definition 3. Let us consider system (1) with initial state $x(t_0)=x_0$. By $X_l^*(t_0, t, x_0)$ we denote the set of all points $y \in R^n$ for which there exists a control $u(s)$, $u^i(s) \in \{a^i, b^i\}$, $i=1, \ldots, m$, for all $s \in [t_0, t_1]$ and constant on each subinterval of the form $[t_0+(j-1)h, t_0+jh]$, $j=1, \ldots, l$, $h=(t_1-t_0)/l$ and such that for the respective trajectory $x(s)$, $s \in [t_0, t_1]$, it holds that $x(t)=y$.

It is obvious that $X_l^*(t_0, t, x_0) \subset X^*(t_0, t, x_0)$ for each positive integer l. Moreover, evidently

$$X^*(t_0, t, x_0) = \bigcup_{l=1}^{\infty} X_l^*(t_0, t, x_0).$$

The following theorem is again direct consequence of Lemma 1 and Theorem 2.

Theorem 4. Let system (1) with initial state $x(t_0)=x_0$ be given. Then for any $t \in [t_0, t_1]$ and any positive integer l

$$h\{X(t_0, t, x_0),\ X_l^*(t_0, t, x_0)\} \leqq \frac{t_1-t_0}{l} \sum_{k=1}^{m} L_k \frac{(b^k-a^k)}{2}. \tag{20}$$

Here $h\{X, Y\}$ is Hausdorff distance between sets X, Y, i.e.

$$h\{X, Y\} = \max\left\{ \sup_{x \in X} \inf_{y \in Y} \|x-y\|_{R^n}, \sup_{y \in Y} \inf_{x \in X} \|x-y\|_{R^n} \right\}.$$

Theorem 4 gives the possibility to approximate an attainable set by its certain finite subset.

4. Numerical applications

In this short section the author aims to describe briefly some experiences with numerical applications of estimate (2).

First and the most simple idea is to use Theorem 4 and to compute some approximation of the attainable set of the bilinear system, but there are practical difficulties — the set $X_l^*(t_0, t, x_0)$, which approximates the attainable set, has 2^{lm} points.

Second idea is to construct a simple algorithm for solution of system (1) with arbitrary control. This algorithm is based on the following theorem which is a consequence of Theorem 2 and Lemma 1.

Theorem 5. Let us consider an arbitrary admissible control $u(s)=(u^1(s), \ldots, u^m(s))^T$ for system (1) on the time interval $[t_0, t_1]$ and let us denote the corresponding trajectory of system (1) with initial condition $x(t_0)=x_0$ by $x(t)$. Further, let $[t_0, t_1]$ be divided into l subintervals as in Lemma 1, let $u^*(s)$ be a control constructed to the control $u(s)$ by applying Lemma 1 to each component $u^k(s)$, $k=1, \ldots, m$, and let

$x^*(t)$ be the corresponding trajectory of system (1) with initial condition $x(t_0)=x_0$. Then

$$\max_{t\in[t_0,t_1]} \|x(t)-x^*(t)\|_{R^n} \leqq \frac{t_1-t_0}{l} \sum_{k=1}^{m} L_k \frac{b_k-a_k}{2}. \tag{21}$$

Here L_k, $k=1, \ldots, m$, are given by (14) and (6)–(13).

The proof of Lemma 1 has a constructive character, i.e. it gives a simple algorithm, how to construct for any function $u(s)$ the appropriate function $u^*(s)$. For controls of type $u^*(s)$ we can relatively easily compute the appropriate solution $x^*(t)$ which can be then considered as an approximate solution of system (1). This algorithm is described in [5] in detail (for time independent systems with single input), where concrete examples of its application are also given,

The most elaborated area of application of estimate (2) is shown in [4], where the preceding algorithm is incorporated into the gradient method for the solution of the optimal control problem. This method is also described in detail for time independent bilinear systems with single input, although there are no theoretical obstacles to extend this method to more general bilinear systems. As a result of the applications of the described method, suboptimal control is obtained which has only two values: a and b. Concrete examples are also included in [4] together with additional numerical experiences.

5. Conclusions

The aim of this contribution was to evaluate the estimate for Lipschitzean dependence of trajectories of multiinput time dependent bilinear systems on controls. Then this estimate was used to study some properties of attainable sets of bilinear systems. In the last section numerical experience with estimate (2) is reported.

References

1. *Mohler, R. R.*, Bilinear Control Processes. Academic Press, New York–London, 1973.
2. *Mohler, R. R., Kolodziej, W. J.*, An overview of bilinear systems theory and applications. IEEE Transaction on Systems, Man and Cybernetics, **10**, 1980, pp. 683–688.
3. *Mohler, R. R., Bruni, C., Gandolfi, A.*, A systems approach to immunology. Proceedings of the IEEE, **8**, 1980, pp. 964–988.
4. *Čelikovský, S.*, On the continuous dependence of trajectories of bilinear systems on controls and its applications. Kybernetika, submitted for publication.
5. *Čelikovský, S.*, On the representation of trajectories of bilinear systems and its applications. Kybernetika, **3**, 1987, pp. 198–213.

О липшицевской зависимости траекторий неавтономной билинейной многомерной системы на управлении

С. ЧЕЛИКОВСКИ

(Прага)

Рассматриваются билинейные неавтономные системы с многомерным управлением, т. е.

$$\dot{x} = (A(t) + \sum_{k=1}^{m} B_k(t) u^k(t)) x + C(t) u(t),$$

$$x(t_0) = x_0, \quad t \in [t_0, t_1], \quad x \in R^n, \quad u \in R^m, \quad a^k \leqq u^k(t) \leqq b^k.$$

Выводится следующая оценка липшицевской зависимости траекторий билинейной системы на управлении относительно некоторых норм:

$$\max_{t \in [t_0, t_1]} \|x_1(t) - x_2(t)\|_{R^n} \leqq \sum_{k=1}^{m} L_k \max_{t \in [t_0, t_1]} \left| \int_{t_0}^{t} u_1^k(s) ds - \int_{t_0}^{t} u_2^k(s) ds \right|.$$

Здесь $x_1(t)$, $x_2(t)$ суть траектории билинейной системы, исходящие из точки $x(t_0)$ с управлениями соответственно $u_1(s) = (u_1^1(s), \ldots, u_1^m(s))$ и $u_2(s) = (u_2^1(s), \ldots, u_2^m(s))$, $s \in [t_0, t_1]$, L_k, $k = 1, \ldots, m$, некоторые постоянные.

Эта оценка затем используется для изучения некоторых свойств множества достижимости билинейных систем. В заключение коротко описаны возможные численные применения липшицевской зависимости.

S. Čelikovský
Institute of Information Theory and Automation
Czechoslovak Academy of Sciences
Pod vodárenskou věží 4
18208 Praha 8
Czechoslovakia

Problems of Control and Information Theory, Vol. 17 (4), pp. 239—251 (1988)

A COMPARATIVE ANALYSIS OF SOME ALGORITHMS FOR THE IDENTIFICATION OF PARAMETERS OF A LINEAR DYNAMIC OBJECT

V. N. BUKOV, YA. KHORVAT

(*Moscow*)

(Received March 11, 1987)

The paper deals with the problem of real-time identification of a model of a controlled entity for adaptive control. A brief description of six different algorithms is given and the results of their numerical behaviour are presented for one object. The influence of various factors on the accuracy and rate of convergence of the estimates is analysed.

1. Formulation of the problem

Modern theory and procedures of control are focussed on dynamic objects operating under conditions of internal and external uncertainties. Control of such objects involves estimation of their structure, parameters, full state vector and uncontrolled disturbances. Approximate values (estimates) of these constants and functions must be computed promptly and in the system's normal operating conditions. Thus, the need arises for effective algorithms which establish the contents and sequence of mathematical operations for processing a posteriori data (current measurements of the object's inputs and outputs). These algorithms having been synthesized taking into account a priori data on the structure of the object, point of application of the uncontrolled disturbances, on their statistic characteristics, etc.

The effectiveness of the algorithms depends on their ability to function under various specific conditions and is usually related to certain qualities of the algorithms. These include the rate of convergence, the accuracy of the estimates obtained, the suitability for operation in a loop, numerical stability, real-time realization, and so on.

This paper carries out a comparative analysis of some algorithms for identification of parameters of a mathematical model by the accuracy of estimates, rate of convergence and labour input for the realization using a standard large digital computer. Estimation of state is considered if by chance it is an integral part of the algorithms.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

The linear model [1] was taken as the object in the increments

$$\begin{aligned}
\Delta\dot{V} &= a_x^V \Delta V + a_x^\theta \Delta\Theta + a_x^\alpha \Delta\alpha,\\
\Delta\dot{\Theta} &= a_y^V \Delta V + a_y^\theta \Delta\Theta + a_y^\alpha \Delta\alpha + a_y^{\delta_B} \Delta\delta_B,\\
\Delta\dot{\omega}_z &= a_{mz}^V \Delta V + a_{mz}^{\omega_z} \Delta\omega_z + a_{mz}^\alpha \Delta\alpha + a_{mz}^{\delta_B} \Delta\delta_B,\\
\Delta\dot{\alpha} &= \Delta\omega_z - \Delta\dot{\Theta}
\end{aligned} \tag{1.1}$$

where $V, \Theta, \omega_z, \alpha$ are components of a state vector, δ_B is the control force. Identification is required for the coefficients a_x^V, a_x^θ, a_x^α, a_y^V, a_y^θ, a_y^α, a_{mz}^V, $a_{mz}^{\omega_z}$, and a_{mz}^α.

2. Brief information about the algorithms under study

A great number of various algorithms have been used for identification of parameters and estimation of a state vector as a controlled entity. This paper examines the following algorithms:

1) Krasovsky's suboptimal algorithm [2].

2) A discrete and continuous algorithm [3].

3) The nonrecursive algorithm of the method of least squares [4].

4) The recursive algorithm of the method of least squares [4].

5) The algorithm for nongradient random search [5].

6) The algorithm for adaptive filtering using the ideas of the projection net method [6].

In the authors' opinion, the algorithms listed above reflect the great variety of algorithms that can be found in the literature.

Algorithm 1. Krasovsky's suboptimal algorithm. A particular case is under consideration: identification of the parameters of object (1.1) without estimating the state vector. It is assumed that the state vector $x = [V\Theta\omega_z\alpha]^T$ is exactly measurable and that the parameter vector only slightly varies during the transition process.

In this case the equation for the object under study and observation can be written as

$$x(t) = f(x, a, u, t), \quad \dot{a}(t) = 0, \quad z(t) = h_z(x, a, u, t) + \xi_z \tag{2.1}$$

where x is an n-dimensional state vector, a is an r-dimensional vector of the parameters to be identified, u is an m-dimensional control vector, z is an l-dimensional observation vector, $f(x, a, u, t)$ is an n-dimensional vector-valued function, differentiable with respect to x and a, for the given arguments, $h_z(x, a, u, t)$ is an l-dimensional vector-valued function, differentiable with respect to x and a, for the given arguments, and ξ_z is an l-dimensional vector of the measurement interference which is assumed to be white noise.

The algorithm for identification of the parameter vector a is written as

$$\hat{a} = R_a \left(\frac{\partial h_z}{\partial \hat{a}} \right)^T S_z^{-1} [z - h_z(x, \hat{a}, u, t)],$$

$$R_a = \left[R_a^{-1}(0) + \int_0^t \left(\frac{\partial h_z}{\partial \hat{a}} \right)^T S_z^{-1} \frac{\partial h_z}{\partial \hat{a}} \, d\tau \right]^{-1} \tag{2.2}$$

where $S_z = \operatorname{diag}(S_{z1}, \ldots, S_{zl})$ is the specified diagonal matrix of noise intensity ξ_{zi} assumed to be non-singular, $\partial h_z / \partial \hat{a}$ is the matrix of partial derivatives $\partial h_{zi} / \partial \hat{a}_j$ (an element of the i-th row and the j-th column) taken at the point $a = \hat{a}$, R_a is a covariational matrix (some of its approximation) of the parameter a estimation errors, $R_a(0)$ is the specified covariational matrix of diagonal form, and $h_z(x, \hat{a}, u, t)$ is an estimate of the observation vector or, to be more exact, the observation vector calculated on the basis of parameter estimates.

Algorithm 2. The discrete and continuous algorithm for identification of a continuous system. This algorithm is a variant of that for discrete non-linear filtering, consistent with continuous equations of the system being identified.

The object under study is produced by combining the equation of the continuous system being identified and the equation for the parameters being estimated

$$\dot{x} = f(x, a, t), \qquad \dot{a} = 0. \tag{2.3}$$

Let the p-dimensional vector y generalize the state of system (2.3):

$$y^T = [x^T a^T], \qquad p = n + r. \tag{2.4}$$

Let the observations also be made at discrete moments of time t_k ($k = 1, 2, \ldots$) in constant steps $T = t_{k+1} - t_k$. Then the discrete equations of observations in the linear case assume the form

$$z_k = H_k y_k + \xi_k \tag{2.5}$$

where H_k is the observation matrix of dimension $l \times p$ and ξ_k is an l-dimensional vector of measurement noises with the correlation matrix R of dimension $l \times l$.

When completed, the operations of linearization and transfer from the continuous system to an equivalent discrete scheme result in a sequence of computing operations of the discrete and continuous method for identification of the vector of the parameters a of the continuous object (2.3) under discrete observations (2.5).

1. Calculation of the prediction for the state vector $\hat{y}(t_{k+1}|t_k)$ of the block of the transition matrix $F_1(t_{k+1}, t_k)$ by simultaneous solution of the set of equations

$$\dot{\hat{x}} = f(\hat{x}, \hat{a}, t),$$

$$\dot{F}_1(t, t_k) = \frac{\partial f(\hat{x}, \hat{a}, t)}{\partial \hat{x}} F_1(t, t_k) + \left[0 \frac{\partial f(\hat{x}, \hat{a}, t)}{\partial \hat{a}} \right]. \tag{2.6}$$

System (2.6) is solvable over the range $[t_k, t_{k+1}]$ with the initial values being $\hat{x}(t_k) = \hat{x}_k$, $F_1(t_k, t_k) = [E_{n\times n} 0_{n\times r}]$ (E is a unit matrix). It is assumed in the solution that $\hat{a}(t) = \hat{a}(t_k) = \hat{a}_k$ for $t \in [t_k, t_{k+1}]$.

The solution of the first equation in system (2.6) determines the prediction of the combined state and parameter vector

$$\hat{y}_{k+1|k} = [\hat{x}^T(t_{k+1}|t_k)\hat{a}^T(t_k)]^T.$$

2. Calculation of the correlation matrix of the prediction error

$$V_{k+1|k} = F(t_{k+1}, t_k) V_k F^T(t_{k+1}, t_k) \tag{2.7}$$

where

$$F(t_{k+1}, t_k) = \begin{bmatrix} F_1(t_{k+1}, t_k)_{n\times p} \\ 0_{r\times n} E_{r\times r} \end{bmatrix}$$

and V_k is the correlation matrix of the filtering error at the k-th step of the algorithm.

3. Calculation of the matrix-valued amplification factor

$$K_{k+1} = V_{k+1|k} H_{k+1}^T (H_{k+1} V_{k+1|k} H_{k+1}^T + R_{k+1})^{-1}. \tag{2.8}$$

4. Calculation of the correlation matrix of the filtering error

$$V_{k+1} = (K_{k+1} H_{k+1} - E) V_{k+1|k} (K_{k+1} H_{k+1} - E)^T + K_{k+1} R_{k+1} K_{k+1}^T. \tag{2.9}$$

5. Calculation of the current estimate of the state and parameter vector

$$\hat{y}_{k+1} = \hat{y}_{k+1|k} + K_{k+1}(z_{k+1} - H_{k+1}\hat{y}_{k+1|k}). \tag{2.10}$$

The aggregative index $k+1|k$ shows that measurements up to and including the k-th step are used for calculating the prediction at the $(k+1)$-th step. The index $k+1$ corresponds to the estimates obtained by all the measurements up to and including the $(k+1)$-th step.

Algorithm 3. The nonrecursive algorithm of the method of least squares. Object (2.4), when linear over the parameters, can be written as

$$\dot{x} = \varphi(x, t) a \tag{2.11}$$

where $\varphi(x, t)$ is the basis function matrix of the state vector with dimension $n \times r$. In

a particular case, the nonzero elements of this matrix can be elements of the state vector.

If each of the parameters a_j occurs only in one of the equations for the components $\dot{x}_i$, i.e. the equations for the components $\dot{x}_i$ are non-interacting with respect to parameters, we can confine ourselves to consideration of

$$\dot{x}_i = \varphi_i(x, t)a_i \tag{2.12}$$

where φ_i is a row composed of r_i nonzero elements of a row of the matrix φ and a_i is an r_i-dimensional vector of the parameters of the i-th scalar equation in (2.11).

If a series of measurements of $\dot{x}_i$ is made at s moments of time, the following relation can be written:

$$z_i = \Phi a_i + \eta_i \tag{2.13}$$

where z_i is an s-dimensional vector whose elements are successive observations of $\dot{x}_i$ at the moments t_k ($k = \overline{1, s}$), η_i is an s-dimensional vector of accidental mistakes of observations at the moments t_k, Φ is a matrix of dimension $s \times r_i$ whose rows are formed by the row φ_i at different measurement moments t_k ($k = \overline{1, s}$), i.e.

$$\Phi^T = [\varphi_i^T(x, t_1) \ldots \varphi_i^T(x, t_s)]. \tag{2.14}$$

The best (in the sense of least squares when observations are of the same accuracy) estimate $\hat{a}_i$ of the vector a_i for $s > r_i$ satisfies the relation

$$\hat{a}_i = (\Phi^T \Phi)^{-1} \Phi^T z_i. \tag{2.15}$$

We can have in a particular case $\Phi^T = [x^T(t_1) \ldots x^T(t_s)]$.

Algorithm 4. The recursive algorithm of the method of least squares. For object (2.12), the estimate $\hat{a}_i$ of the vector a_i at the k-th step can be obtained according to the following recurrent relation:

$$\hat{a}_{i \cdot k} = \hat{a}_{i \cdot k-1} + P_k \varphi_{i \cdot k}^T (z_{i \cdot k} - \varphi_{i \cdot k} \hat{a}_{i \cdot k-1}) \tag{2.16}$$

where $P_k =$ is a matrix of dimension $r_i \times r_i$ also obtained successively on the basis of the relation

$$P_k = P_{k-1} - P_{k-1} \varphi_{i \cdot k}^T (1 + \varphi_{i \cdot k} P_{k-1} \varphi_{i \cdot k})^{-1} \varphi_{i \cdot k} P_{k-1}. \tag{2.17}$$

The original values of $\hat{a}_{i \cdot 0}$ should be selected as zeroes while P_0 should be diagonal with large elements if possible.

Algorithm 5. The algorithm for nongradient random search. Object (1.1) is represented as (2.11). The procedure of random search for the vector a_i is performed such that $\hat{a}_i$ belongs to a specified bounded space Ω.

The algorithm in question assumes that the sought for values of the parameters are random quantities (with an a priori mathematical expectation vector $m_{0 \cdot i}$ and

correlation matrix $K_{0\cdot i}$) distributed by a certain (normal) law. The search run is carried out on the basis of these a priori data according to the following scheme. Realization of the vector $\hat{a}_i$ with the given distribution law is worked out. This realization and s measurements at the t_k moments of inputs (elements of the matrix $\varphi_i(x, t_k)$ in this case) and outputs (the scalars z_i (2.13) in this case) are used to compute the value (observations are supposed to be of equal accuracy)

$$J = \sum_{k=1}^{s} [z_i(t_k) - \varphi_i(x, t_k)\hat{a}_i]^2. \tag{2.18}$$

Then the inequality

$$J \leqq C \tag{2.19}$$

is verified. Here C is a given constant. If this condition is met, the run is considered to be successful and the mathematical expectation and correlation matrix worked out at the last step become a priori ones for the next step. If condition (2.19) is not met, the run is considered to be unsuccessful and a new realization is worked out again on the basis of the previous values of $m_{0\cdot i}$, $K_{0\cdot i}$. The process of the search continues until the specified quality of successful runs is achieved. The procedure of search results in such mathematical expectation vector and correlation matrix of random values $\hat{a}_i$ where with the probability of conformity with (2.19) is the largest.

Algorithm 6. The algorithm for adaptive filtering using the ideas of the projection net method. By "adaptive observer" we will mean henceforth devices designed for evaluating unknown parameters and for the recovery of unmeasurable phase coordinates of an object in the absence of measurements noise. Adaptive filters are devices for combined (simultaneous) estimation of the state vector components and uncontrolled disturbances.

1. *Adaptive observers.* Object (1.1) and the observation equation for the linear case can be represented as

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad z(t) = Hx(t) \tag{2.20}$$

where A, B are matrices of unknown constant coefficients of dimension $n \times n$ and $n \times m$, respectively, and H is the observation matrix of dimension $l \times n$ of full rank.

The algorithm which enables values of the unknown elements of the matrices A and B, of the components of the original state vector $x(t_0)$ and of nonmeasurable components of the state vector $x(t)$ to be found on the basis of observations of $u(t)$ and $z(t)$ over the interval $[t_0, t_0 + T]$ of a specified duration, is synthesized in the following manner.

The matrix A is split by the formula

$$A = C + DH \tag{2.21}$$

where C is a familiar non-singular constant Hurwitz matrix of dimension $n \times n$ in the form

$$C = \begin{bmatrix} -C_{n-1} & 1 & 0 & \dots & 0 \\ -C_{n-2} & 0 & 1 & \dots & 0 \\ \vdots & & & & \vdots \\ -C_1 & 0 & 0 & \dots & 1 \\ -C_0 & 0 & 0 & \dots & 0 \end{bmatrix} \tag{2.22}$$

which has different latent roots, D is an unknown constant matrix of dimension $n \times l$. Representation (2.21) is true only if the space of the rows of matrix H coincides with that of the matrix $A - C$.

Using (2.21) and from (2.20) we get

$$\dot{x}(t) = Cx(t) + Dz(t) + Bu(t), \qquad x(t_0) = x_0 \tag{2.23}$$

which we rewrite as

$$\dot{x}(t) = Cx(t) + Z(t)a_D + V(t)a_B, \qquad x(t_0) = x_0. \tag{2.24}$$

The following notation is adopted in (2.24): $Z(t) = \operatorname{diag}(z^T(t), \dots, z^T(t))$ is a matrix of dimension $n \times nl$ of the output coordinates being measured, $V(t) = \operatorname{diag}(u^T(t), \dots, u^T(t))$ is a matrix of dimension $n \times nm$ of the inputs being measured, $a_D = [D_{|1|}, \dots, D_{|n|}]^T$ is an nl-dimensional vector of the unknown parameters of the matrix D, where $D_{|i|}$ is the i-th row of D, and $a_B = [B_{|1|}, \dots, B_{|m|}]^T$ is an nm-dimensional vector of the unknown parameters composed of the rows of B.

A solution of (2.24) takes the form

$$x(t) = K(t)x(t_0) + R(t)a_D + S(t)a_B \tag{2.25}$$

where the variable matrices $K(t)$, $R(t)$ and $S(t)$ satisfy the differential equations

$$\begin{aligned} &\dot{K}(t) = CK(t), && K(t_0) = E, && \dim K(t) = n \times n, \\ &\dot{R}(t) = CR(t) + Z(t), && R(t_0) = 0, && \dim R(t) = n \times nl, \\ &\dot{S}(t) = CS(t) + V(t), && S(t_0) = 0, && \dim S(t) = n \times nm. \end{aligned} \tag{2.26}$$

By multiplying the two sides of (2.25) from the left by the observation matrix H, we obtain

$$z(t) = Q(t)y \tag{2.27}$$

where $Q(t) = [HK(t)\ HR(t)\ HS(t)]$ is a block matrix of dimension $l \times n(1 + l + m)$ and $y = [x^T(t_0)a_D^T a_B^T]^T$ is an $n(1 + l + m)$-dimensional vector of unknown parameters.

Having solved the set of algebraic equations (2.27), we get the estimate $\hat{y}$. The estimate of the state vector $\hat{x}(t)$ can be obtained on the basis of (2.25) and the estimate of the elements of A can be found from (2.21).

2. *Adaptive filters.* The equation of the object is represented as

$$\dot{x}(t)=Ax(t)+Bu(t), \qquad z(t)=Hx(t)+\xi(t). \tag{2.28}$$

The above-described manipulation yields

$$\dot{x}(t)=Cx(t)+Dz(t)+Bu(t)-D\xi(t), \qquad x(t_0)=x_0, \tag{2.29}$$

We introduce a net with the step $h=t_{i+1}-t_i$ and the total number of steps q over the observation interval and we approximate $\xi=[\xi_1 \ \dots \ \xi_l]^T$ in the linear form

$$\xi(t)=N\psi(t) \tag{2.30}$$

where $\psi(t)$ is a q-dimensional vector of the familiar finite functions $\psi_j=h^{1/2}\psi\left(\frac{t}{h}-j\right)$ and N is a matrix of unknown constant coefficients. And each of the functions $\psi_j(t)$ is zero out of the range $[t_{j-1}, t_{j+1}]$ and is a polynomial to the power d in each of the ranges $[t_{j-1}, t_j]$ and $[t_j, t_{j+1}]$.

Substitution of (2.30) into (2.29) and a little manipulation yield

$$\dot{x}(t)=Cx(t)+Z(t)a_D+V(t)a_B+F(t)a_N, \qquad x(t_0)=x_0 \tag{2.31}$$

where (2.24) is supplemented with $a_N=[(DN)_{11}(DN)_{12} \ \dots \ (DN)_{nq}]$, a column whose elements are those of the product of the matrices DN:

$$F(t)=\begin{bmatrix} \psi_1 \ \dots \ \psi_q & & 0 \\ & \ddots & \\ 0 & & \psi_1 \ \dots \ \psi_q \end{bmatrix}.$$

Here F is a block-diagonal matrix of dimension $n \times nq$ with diagonal blocks in the form of the transposed vector $\psi(t)$.

A solution of (2.31) has the form:

$$x(t)=K(t)x(t_0)+R(t)a_D+S(t)a_B+M(t)a_N \tag{2.32}$$

where the matrices $K(t)$, $R(t)$, $S(t)$, and $M(t)$ satify the differential equations

$$\begin{aligned} \dot{K}(t)&=CK(t), & K(t_0)&=E, \\ \dot{R}(t)&=CR(t)+Z(t), & R(t_0)&=0, \\ \dot{S}(t)&=CS(t)+V(t), & S(t_0)&=0, \\ \dot{M}(t)&=CM(t)+F(t), & M(t_0)&=0. \end{aligned} \tag{2.33}$$

By multiplying the two sides of (2.32) from the left by the observation matrix, we get a set of algebraic equations

$$z(t)=Q'(t)y'. \tag{2.34}$$

Here $Q'(t)=H[K(t)R(t)S(t)M(t)]$ is a block matrix consisting of the well-known time functions and $y'=[x_0^T a_D^T a_B^T a_N^T]^T$ is the vector of unknown coefficients.

Having solved (2.34), we get the estimate $\hat{y}'$ using that one which can determine $\hat{x}(t)$ from (2.32) as well as estimates of uncontrolled disturbances and unknown parameters using (2.31) and (2.30).

3. The research arrangements

Numerical modelling of the algorithms listed above for identification was carried out using an ES-1060 computer. The Runge–Kutta numerical method with constant integration steps was used to solve the differential equations of the object and algorithms. The method of singular decomposition of matrices was adopted for solution of the sets of algebraic equations.

The process was excited with step ($\delta_B=5$ and 10 grades) and harmonic ($\delta_B=\delta_{B\cdot m}\sin\dfrac{2\pi}{T}t-\delta_{B\cdot 0}$ for $\delta_{B\cdot m}=5$ and 10 grades), $T=T_1$, $0.5T_1$ and $2T_1$ where T_1 is the time of identification) inputs. Noises of the information pickups were assumed to be white with zero mathematical expectations and with dispersions

$$\sigma_{\dot{V}}^2=1\ \mathrm{m}^2/c^4,\quad \sigma_{\dot{\vartheta}}^2=64\cdot 10^{-6}\ 1/c^2,\quad \sigma_{\dot{\omega}_z}^2=9\cdot 10^{-6}\ 1/c^4,\quad \sigma_{\dot{\alpha}}^2=25\cdot 10^{-6}\ 1/c^2,$$

$$\sigma_{\dot{V}}^2=1\ \mathrm{m}^2/c^2,\quad \sigma_{\vartheta}^2=64\cdot 10^{-6},\quad \sigma_{\omega_z}^2=9\cdot 10^{-6}\ 1/c^2,\quad \text{and}\quad \sigma_{\alpha}^2=25\cdot 10^{-6}$$

and they were simulated with the Gauss subroutine.

4. Results of numerical modelling

Numerical modelling of identification algorithms has established that the shape, amplitude, and period of input and the measurement noises affect the accuracy of estimates obtained and the rate of convergence of the algorithm. These effects vary for different parameters as well as for different algorithms.

In algorithm 1, with the initial conditions $\hat{a}(0)=0$, the coefficient estimation error ($\delta_B=5$ grades without noise) does not exceed ± 4 percent. And the process converged in 0.2 ... 2.2 s. $a_{mz}^{\omega_z}$ was the fastest to estimate while a_x^V was the slowest. The appearance of measurement noise disturbed the accuracy of the estimation of coefficients in the first equation of (1.1) to an unacceptable level and the errors of

other estimates increased to ± 30 percent. The process converged in 0.8 ... 3.6 s. An increase in input amplitude somewhat reduced the noise effects. For harmonic input, the best estimates were obtained when the input period T was equal to the identification time T_1.

Algorithm 2 estimated the state vector and the parameter vector simultaneously. It practically gave no estimation for the coefficients in the first equation of (1.1). The accuracy and convergence of the process were greatly affected by the selection of the initial values of the estimates. If initial errors were less than ± 30 percent, the process converged in 1.5 ... 3.0 s and the estimation errors did not go beyond the range of $\pm(5 \ldots 20)$ percent. The shape, amplitude and period of input affected only the estimates of a_y^V, a_y^0 and a_{mz}^V. The latter were the most accurate with a harmonic input having the period $T = T_1$. And the state vector was estimated quite accurately.

The operation of algorithm 3 was not particularly affected by the shape, amplitude and period of input if at least $r+1$ measurements had been made, where r is the number of parameters to be identified. The algorithm operated practically without errors if there was no noise and the rate of convergence was governed by the time of information collection. The noise affected considerably the accuracy of estimation of the coefficients from the first equation but this effect diminished as the number of measurements increased. 20 measurements for estimation of three coefficients made it possible to ignore the influence of noise.

Algorithm 4 estimated the coefficients of the object's second and third equation practically without errors, the process converging in 0.2 ... 1.2 s. The accuracy of estimates of the first equation was $\pm(8 \ldots 20)$ percent. The shape, amplitude and period had only a slight influence and the convergence improved with larger amplitude. The appearance of noise increased the coefficient estimation errors to an unacceptable level with the other estimate being worsened a little and the time of convergence increased to 1.4 ... 3.2 s. Harmonic input with an amplitude of 10 grades and period $T = T_1$ proved to be preferable.

The operation of algorithm 5 depended greatly on the selection of a priori mathematical expectations and covariational matrices. This dependence exceeded that of the type of input. In the absence of noise the number of runs required was 20 ... 80 while with noise it increased to 80 ... 1200. The accuracy of estimates was $\pm(10 \ldots 40)$ percent in both cases.

Algorithm 6 estimated the state vector and the coefficients simultaneously. In the absence of noise and with linear independence of outputs for the step input, it ensured the accuracy of estimates at the level of ± 2 percent and for the harmonic input ($T = T_1$) at the level of ± 1 percent. The state vector was estimated with the same accuracy. When affected by noise, the estimates of the coefficients became unacceptable although the accuracy of the state vector estimation remained high enough, i.e. at the level of $\pm(10 \ldots 22)$ percent.

Figure 1 shows the results of identification of the coefficients from the third equation of object (1.1) with various algorithms for the step input $\delta_B = 5$ grades. The horizontal line marks the levels of the coefficient's true levels. Figures 2–4 show the graphs of the processes of identification of the parameters by algorithms 1, 2, and 4.

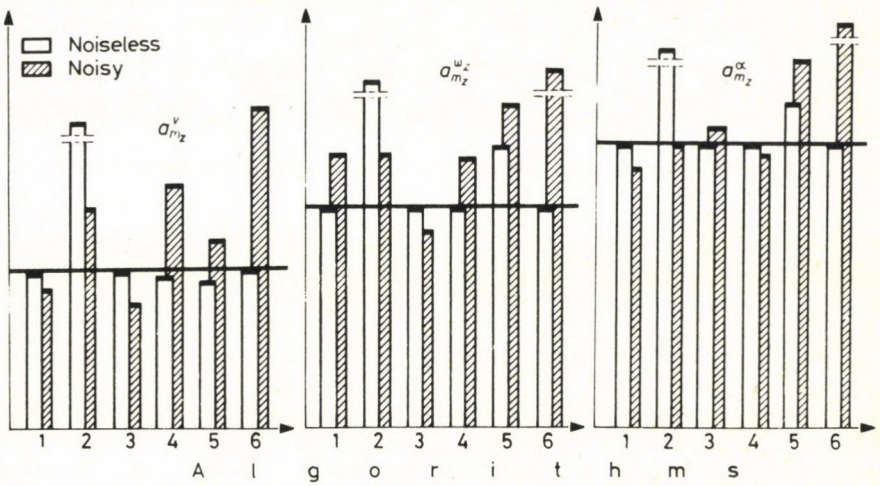


Fig. 1

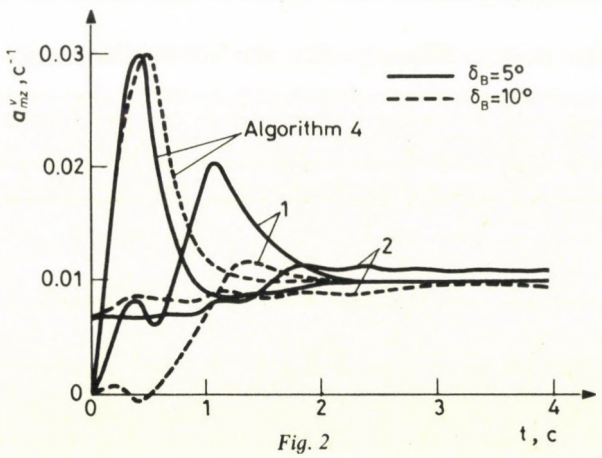


Fig. 2

The results of an analysis of computer resources required for realization of the algorithms are listed in Table 1. It indicates mathematical expectations of the main storage and permanent storage capacity and the number of operations per one step of an algorithm. The data were obtained for a standard 16-digit large computer. Nonrecurrency of algorithms 3 and 6 justifies to some extent their high cost.

Table 1

Algorithm	1	2	3	4	5	6
Permanent storage capacity, words	32 000	38 700	25 100	4900	3200	11 900
Main storage capacity, words	120	141	97	25	39	107
Number of operations, thousands	63.8	73.3	48.1	9.4	10.3	124.0

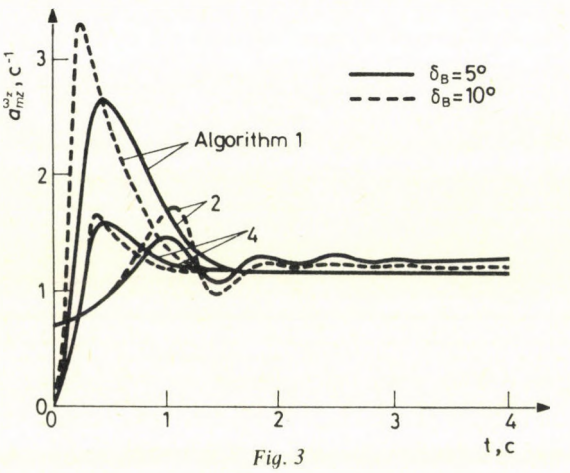


Fig. 3

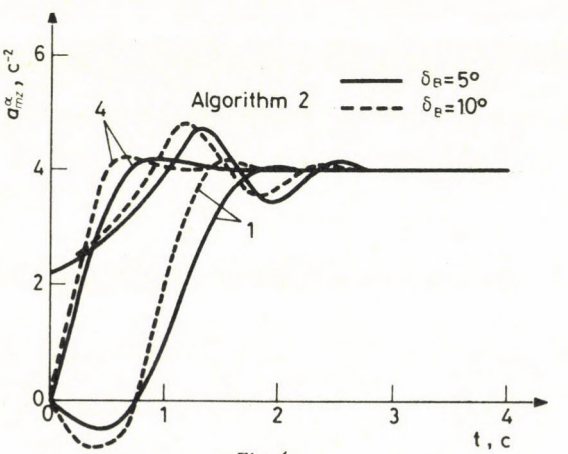


Fig. 4

Thus, the shape of input affected considerably the accuracy and convergence of algorithms 1, 2 and 6. The input amplitude affected algorithms 1, 2, and 4. Harmonic input with the period $T = T_1$ proved to be the best here. Measurement noise affected most of all the accuracy and convergence of algorithms 2 and 6. Algorithm 3 and 4 proved to be the most effective in terms of accuracy, rate of convergence and machine costs under the above-described conditions for identification of coefficients of the mathematical models of a plane's longitudinal motion.

References

1. *Krasovsky, A. A.*, Systems of automatic flight control and their analytical construction. Moscow, Nauka, 1973.
2. *Krasovsky, A. A., Bukov, V. N., Shendrik, V. S.*, Universal algorithms for optimal control of continuous processes. Moscow, Nauka, 1977.
3. *Kulifeev, Yu. B.*, Discrete and continuous method of identification of continuous systems. Izv. AN SSSR, Mekhanika tverdogo tela, 1981, No. 5.
4. *Grop, D.*, Methods of Systems Identification. Moscow, Mir, 1979.
5. *Gladkov, D. I.*, Optimization of systems with nongradient random search. Moscow, Energoatomizdat, 1984.
6. *Prokopov, B. I.*, On construction of adaptive observers. Avtomatika i telemekhanika, 1981, No. 5.

Сравнительный анализ некоторых алгоритмов идентификации параметров линейного динамического объекта

В. Н. БУКОВ, Я. ХОРВАТ

(Москва)

Рассматривается задача выполняемой в реальном времени идентификации модели объекта управления для целей адаптивного управления. Приводится краткое описание шести различных алгоритмов и результаты их численного исследования для одного объекта.

Анализируется влияние различных факторов на точность и скорость сходимости.

В. Н. Буков
Научный совет по комплексной программе «Кибернетика»
СССР, 117333 Москва, ул. Вавилова, 40

РУССКИЙ ПЕРЕВОД

Проблемы управления и теории информации, том **17**, *номер 4 (1988)*

СИНТЕЗ ОПТИМАЛЬНОГО ГАРАНТИРУЮЩЕГО УПРАВЛЕНИЯ

В. Е. Третьяков

(*Свердловск*)

В данной работе указывается алгоритм вычисления оптимальных управляющих воздействий в задаче позиционного управления системой, возмущаемой неконтролируемой помехой. Показывается, что объединение трех подходов: 1) экстремального сдвига на сопутствующую точку, 2) конструкций стохастического синтеза, 3) введения дополнительной фазовой координаты позволяет эффективно построить функцию, которая аппроксимирует цену игры и является дифференцируемой по фазовым переменным.

Экстремальный сдвиг на сопутствующую точку сводится при этом к экстремальному сдвигу против градиента аппроксимирующей функции. Устанавливается, что в случае показателя качества, являющегося нормой фазового вектора в момент окончания процесса, данный подход приводит к реализуемой вычислительной процедуре. Полученные результаты проясняют также связь стохастического программного синтеза [1–3] с известными обобщениями метода динамического программирования [4, 5].

1. Постановка задачи

Рассмотрим задачу позиционного управления при условии минимума гарантированного результата по показателю

$$\gamma = |x[\vartheta]| \tag{1.1}$$

для системы

$$\frac{dx}{dt} = A(t)x + B(t)u + C(t)v, \quad t_0 \leqq t \leqq \vartheta. \tag{1.2}$$

Здесь x — n-мерный фазовый вектор объекта; $|x|$ — его евклидова норма; t_0 и ϑ — фиксированные моменты времени; $A(t)$, $B(t)$, $C(t)$ — непрерывные матрицы-функции; u — r-мерный вектор управления, v — s-мерный вектор помехи, стесненные соответственно ограничениями $u \in \mathcal{P}$, $v \in \mathcal{Q}$, где $\mathcal{P}$ и $\mathcal{Q}$ — заданные компакты.

Основными инструментами для решения рассматриваемой задачи будут понятия стратегии $u(\cdot)=\{u(t, x, \varepsilon)\in\mathscr{P}, t_0\leqq t\leqq\vartheta, x\in E^n, \varepsilon>0\}$, закона управления

$$U=\{u(\cdot), \varepsilon, \Delta_\delta\{t_i\}\} \tag{1.3}$$

и движения $x[t_*[\cdot]\vartheta]=\{x[t], t_*\leqq t\leqq\vartheta\}$, порождаемого из произвольной позиции $\{t_*, x_*\}$, $x[t_*]=x_*$, $t_*\in[t_0, \vartheta)$, $x_*\in E^n$ по шагам $[t_i, t_{i+1}]$, $i=1, \ldots, k$ разбиения

$$\Delta_\delta\{t_i\}=\{t_*=t_1, t_i, t_{i+1}-t_i\leqq\delta, t_{k+1}=\vartheta\} \tag{1.4}$$

законом управления U (1.3) и какой-то реализацией неконтролируемой нами помехи

$$v(t_*[\cdot]\vartheta]=\{v[t]\in\mathscr{Q}, t_*<t\leqq\vartheta\} \tag{1.5}$$

— произвольной измеримой функцией времени.

Эти понятия лежат в основе математической формализации задачи на минимум гарантированного результата, предложенной в монографии [1]. В соответствии с этой формализацией ставится задача о вычислении оптимального гарантированного результата ([1], с. 71)

$$\rho^0(t_*, x_*)=\rho[u^0(\cdot); t_*, x_*]=\min_{u(\cdot)}\overline{\lim_{\varepsilon\to 0}}\lim_{\delta\to 0}\sup_{\Delta_\delta\{t_i\}}\sup_{v(t_*[\cdot]\vartheta]}|x[\vartheta]| \tag{1.6}$$

и об отыскании универсальной оптимальной стратегии $u^0(\cdot)$. Эта задача имеет решение.

Оптимальный гарантированный результат для показателя γ (1.1) в произвольной позиции $\{t, x\}$ может быть вычислен методом стохастического программного синтеза как величина стохастического программного экстремума ([1], с. 306)

$$\rho^0(t, x)=\sup_{\Delta\{\tau_j\}}\sup_{\|l(\cdot)\|\leqq 1}\kappa(t, x, \Delta\{\tau_j\}, l(\cdot)). \tag{1.7}$$

Здесь

$$\Delta\{\tau_j\}=\{t=\tau_1, \ldots, \tau_j<\tau_{j+1}, \ldots, \tau_{q+1}=\vartheta\} \tag{1.8}$$

— произвольное разбиение отрезка $[t, \vartheta]$, никак не связанное с разбиением (1.4). Величина $l(\cdot)$ есть измеримая по Борелю на кубе $\Omega=\{\omega=\{\xi_1, \ldots, \xi_q\}, 0\leqq\xi_j<1, j=1, \ldots, q\}$ n-мерная вектор-функция, которая естественно трактуется как векторная случайная величина, и тогда символ $\|l(\cdot)\|$ следует понимать как евклидову норму в гильбертовом пространстве случайных величин ([6], с. 279). Скалярная функция κ в выражении (1.7) определена соотношением (41.8) в монографии [1] (см. с. 315) и здесь нет необходимости выписывать это соотноше-

ние явно. Заметим лишь, что путем замены

$$l(\cdot)=m+b(\cdot), \quad m=M\{l(\omega)\}, \quad M\{b(\omega)\}=0, \tag{1.9}$$

где M означает математическое ожидание, задача (1.7) сводится к задаче ([7], с. 187)

$$\rho^0(t,x)=\max_{|m|\leq 1}\left[\langle X'(\vartheta,t)m\cdot x\rangle+g(t,m)\right]. \tag{1.10}$$

Здесь $X(\vartheta,t)$ — фундаментальная матрица однородной системы (1.2), верхний знак $'$ означает транспонирование и величина $g(t,m)$ определяется равенством

$$g(t,m)=\sup_{\Delta\{\tau_j\}}\left[\sup_{||b(\cdot)||\leq(1-|m|^2)^{1/2}} M\left\{\int_t^{\vartheta}\min_{u\in\mathcal{P}}\max_{v\in\mathcal{Q}}\langle[m+\right.\right.$$

$$\left.\left.+n(\tau,\omega)]\cdot X(\vartheta,\tau)\,[B(\tau)u+C(\tau)v]\rangle\,d\tau\right\}\right], \tag{1.11}$$

где величины

$$n(\tau,\omega)=n[\xi_1,\ldots,\xi_j]=\int_{0\leq\xi_{j+1}<1}\cdots\int_{0\leq\xi_q<1} b[\xi_1,\ldots,\xi_q]\,d\xi_{j+1}\ldots d\xi_q;$$

$$\tau_j\leqq\tau<\tau_{j+1}, \qquad j=1,\ldots,q \tag{1.12}$$

имеют смысл условных математических ожиданий n-мерной векторной случайной величины $b(\omega)=b[\xi_1,\ldots,\xi_q]$. Можно показать, и это для дальнейшего существенно, что функция $g(t,m)$ (1.11) по аргументу m вогнута. Тогда вогнутой по аргументу m при каждой фиксированной позиции $\{t,x\}$ будет и вся максимизируемая в (1.10) функция.

Оптимальная стратегия $u^0(\cdot)$ на основе функции $\rho^0=\rho^0(t,x)$ (1.10) находится по правилу экстремального сдвига ([1], с. 210)

$$\langle s_u(t,x,\varepsilon)\cdot B(t)u^0(t,x,\varepsilon)\rangle=\min_{u\in\mathcal{P}}\langle s_u(t,x,\varepsilon)\cdot B(t)u\rangle \tag{1.13}$$

в направлении против вектора $s_u(t,x,\varepsilon)$, определяемого из условия

$$\rho^0(t,x-s_u(t,x,\varepsilon))=\min_{|s|\leq\beta(\varepsilon,t)}\rho^0(t,x-s), \tag{1.14}$$

где

$$\beta(\varepsilon,t)=[\varepsilon(1+[t-t_0])]^{1/2}\exp\lambda[t-t_0], \tag{1.15}$$

$$\lambda=\max_{t_0\leq t\leq\vartheta}\max_{|x|=1}|A(t)x|. \tag{1.16}$$

Таким образом, для вычисления на каждом шаге $[t_i, t_{i+1}]$, $i=1, \ldots, k$ оптимального управляющего воздействия

$$u^0[t]=u^0(t_i, x[t_i], \varepsilon), \quad t_i<t\leqq t_{i+1} \tag{1.17}$$

требуется последовательно решить две задачи на минимум (1.14) и (1.13) при $t=t_i$, $x=x[t_i]$, что в общем случае может привести к необходимости многократного подсчета величины $\rho^0(t_i, x[t_i]-s)$ по формулам (1.10), (1.11) в точках n-мерного шара $|s|\leqq\beta(\varepsilon, t_i)$. С вычислительной точки зрения эта задача является наиболее трудоемкой.

В данной статье предлагается подход, позволяющий обойти решение задачи (1.14). Оказывается, что благодаря добавлению к фазовому вектору x еще одной фиктивной координаты x_{n+1}, удается с помощью выражений (1.10), (1.11) построить некоторую функцию $\rho_\varepsilon^0(t, x)$, аппроксимирующую функцию оптимального гарантированного результата $\rho^0(t, x)$ и являющуюся дифференцируемой по аргументу x. Это позволяет в свою очередь существенно упростить алгоритм вычисления оптимальных управляющих воздействий $u^0[t]$ (1.17) и свести его на каждом шаге $[t_i, t_{i+1}]$, $i=1, \ldots, k$ к экстремальному сдвигу

$$\left\langle\left[\frac{\partial\rho_\varepsilon^0(t, x)}{\partial x}\right]_{t=t_i, x=x[t_i]} \cdot B(t_i)u^0(t_i, x[t_i], \varepsilon)\right\rangle =$$

$$= \min_{u\in\mathscr{P}} \left\langle\left[\frac{\partial\rho_\varepsilon^0(t, x)}{\partial x}\right]_{t=t_i, x=x[t_i]} \cdot B(t_i)u\right\rangle \tag{1.18}$$

против градиента $\partial\rho_\varepsilon^0(t, x)/\partial x=\{\partial\rho_\varepsilon^0/\partial x_1, \ldots, \partial\rho_\varepsilon^0/\partial x_n\}$, $t=t_i$, $x=x[t_i]$ построенной в статье функции $\rho_\varepsilon^0(t, x)$.

2. Аппроксимирующая функция и ее свойства

Рассмотрим функцию

$$\rho_\varepsilon^0(t, x)=\max_{|m|\leqq 1} H(t, x, \varepsilon, m), \tag{2.1}$$

$$H(t, x, \varepsilon, m)=-\beta(\varepsilon, t)(1+|X'(\vartheta, t)m|^2)^{1/2}+$$

$$+\langle X'(\vartheta, t)m\cdot x\rangle+g(t, m), \tag{2.2}$$

определенную при всех $t\in[t_0, \vartheta]$ и всех $x\in E^n$.

Подчеркнем, что вычисление этой функции в ряде конкретных случаев, в том числе и в разбираемом здесь подробно случае $\gamma=|x[\vartheta]|$,может быть осуществлено эффективно. Не будем останавливаться на истоках получения

этой функции. Как уже отмечалось, эти истоки базируются на идеях стохастического синтеза и введении дополнительной регуляризующей фазовой координаты x_{n+1}. Отметим лишь основные свойства функции $\rho_\varepsilon^0(t, x)$ (2.1), (2.2)

1^0. Благодаря строгой вогнутости функции $H(t, x, \varepsilon, m)$ (2.2) по аргументу m, максимум в (2.1) достигается на единственном векторе

$$m^0 = m^0(t, x, \varepsilon) \tag{2.3}$$

2^0. Вследствие единственности вектора m^0 (2.3) вектор-функция $m^0 = m^0(t, x, \varepsilon)$ непрерывна по аргументам $t \in [t_0, \vartheta]$, $x \in E^n$.

3^0. Благодаря свойствам 1^0, 2^0 функция $\rho_\varepsilon^0(t, x)$ (2.1) является дифференцируемой по аргументу x, т. е. существует градиент $\partial\rho_\varepsilon^0(t, x)/\partial x$. При этом

$$\frac{\partial\rho_\varepsilon^0(t,x)}{\partial x} = X'(\vartheta, t)m^0(t, x, \varepsilon) \tag{2.4}$$

и вследствие условия $|m^0(t, x, \varepsilon)| \leqq 1$ градиент $\partial\rho_\varepsilon^0(t, x)/\partial x$ (2.4), непрерывный по $\{t, x\}$, равномерно ограничен при $t \in [t_0, \vartheta]$, $x \in E^n$.

4^0. Вследствие равенства $g(\vartheta, m) = 0$ функция $\rho_\varepsilon^0(t, x)$ (2.1) при $t = \vartheta$ удовлетворяет краевому условию

$$\rho_\varepsilon^0(\vartheta, x) = \begin{cases} -[\beta^2(\varepsilon, \vartheta) - |x|^2]^{1/2}, & |x| \leqq \dfrac{\beta(\varepsilon, \vartheta)}{\sqrt{2}} \\ |x| - \sqrt{2}\beta(\varepsilon, \vartheta), & |x| > \dfrac{\beta(\varepsilon, \vartheta)}{\sqrt{2}} \end{cases} \tag{2.5}$$

5^0. Функция $\rho_\varepsilon^0(t, x)$ (2.1) аппроксимирует функцию оптимального гарантированного результата $\rho^0(t, x)$ (1.10) в следующем смысле: при всех $\{t, x\}$ имеет место оценка

$$\beta(\varepsilon, t_0) \leqq \rho^0(t, x) - \rho_\varepsilon^0(t, x) \leqq K\beta(\varepsilon, \vartheta), \tag{2.6}$$

где K — некоторая положительная константа.

6^0. Пусть выбран закон управления

$$U^0 = \{u^0(\cdot), \varepsilon, \Delta_\delta\{t_i\}\}, \tag{2.7}$$

формирующий воздействия $u^0[t] = u^0(t_i, x[t_i], \varepsilon)$, $t_i < t \leqq t_{i+1}$, $i = 1, \ldots, k$ из условия экстремального сдвига (1.18) на основе построенной функции $\rho_\varepsilon^0(t, x)$ (2.1). Пусть при этом реализуется какая-то неизвестная помеха $v(t_*[\cdot]\vartheta]$ (1.5).

Тогда вдоль движения $x[t_*[\cdot]\vartheta]$, порожденного законом U^0 (2.7), на каждом шаге $[t_i, t_{i+1}]$, $i=1, \ldots, k$ будет иметь место следующая оценка, выражающая свойство u-стабильности функции $\rho_\varepsilon^0(t, x)$,

$$\Delta\rho_\varepsilon^0 = \rho_\varepsilon^0(t_{i+1}, x[t_{i+1}]) - \rho_\varepsilon^0(t_i, x[t_i]) \leqq$$

$$\leqq \alpha(\varepsilon, \delta)[t_{i+1} - t_i]; \quad \lim_{\varepsilon\to 0, \delta\to 0} \alpha(\varepsilon, \delta) = 0, \tag{2.8}$$

какова бы ни была на этом шаге реализация помехи $v(t_*[\cdot]\vartheta]$ (1.5).

Данную оценку можно выразить также в следующей дифференциальной форме

$$D_t^+(\rho_\varepsilon^0(t, x)) + \left\langle \frac{\partial\rho_\varepsilon^0}{\partial x} \cdot [A(t)x + B(t)u^0(t, x, \varepsilon) + C(t)v] \right\rangle \leqq v(\varepsilon), \tag{2.9}$$

где символом $D_t^+(\rho_\varepsilon^0(t, x))$ обозначено верхнее по t производное число функции $\rho_\varepsilon^0(t, x)$ (2.1). Соотношение (2.9) заменяет здесь соответствующее дифференциальное неравенство из метода динамического программирования и проясняет связь этого метода с построениями из метода стохастического программного синтеза.

Эти оценки выводятся непосредственно из определения функции $\rho_\varepsilon^0(t, x)$ (2.1), дающего неравенство

$$\Delta\rho_\varepsilon^0 \leqq H(t_{i+1}, x[t_{i+1}], \varepsilon, m^0(t_{i+1}, x[t_{i+1}], \varepsilon)) -$$

$$- H(t_i, x[t_i], \varepsilon, m^0(t_{i+1}, x[t_{i+1}], \varepsilon)). \tag{2.10}$$

При этом используются свойства 1^0–3^0 функции $\rho_\varepsilon^0(t, x)$ (2.1). При дальнейшей обработке правой части неравенства (2.10) наиболее существенной является следующая оценка изменения величины $g(t, m)$ (1.11)

$$\Delta g = g(t_{i+1}, m^0(t_{i+1}, x[t_{i+1}], \varepsilon)) - g(t_i, m^0(t_{i+1}, x[t_{i+1}], \varepsilon)) \leqq$$

$$\leqq -\int_{t_i}^{t_{i+1}} \min_{u\in\mathscr{P}} \max_{v\in\mathscr{Q}} \langle m^0(t_i, x[t_i], \varepsilon) \cdot X(\vartheta, t)[B(t)u + C(t)v] \rangle dt \leqq$$

$$\leqq -\left[\min_{u\in\mathscr{P}} \max_{v\in\mathscr{Q}} \left\langle \left[\frac{\partial\rho_\varepsilon^0(t, x)}{\partial x} \right]_{t=t_i, x=x[t_i]} \cdot [B(t_i)u + C(t_i)v] \right\rangle \right] \cdot$$

$$\cdot (t_{i+1} - t_i) + \alpha_1(\delta)(t_{i+1} - t_i), \tag{2.11}$$

где $\alpha_1(\delta)\to 0$ при $\delta\to 0$.

Используя далее условие экстремального сдвига (1.18), с учетом оценки (2.11) из неравенства (2.10) получаем требуемую оценку (2.8).

Последовательно сравнивая неравенства (2.8) при $i=1, \ldots, k$, получаем, что на движении $x[t_*[\cdot]\vartheta]$, порожденном законом управления U^0 (2.7) и

произвольной помехой $v(t_*[\cdot]\vartheta]$ (1.5), выполняется неравенство

$$\rho_\varepsilon^0(\vartheta, x[\vartheta]) \leqq \rho_\varepsilon^0(t_*, x_*) + \alpha(\varepsilon, \delta)[\vartheta - t_0], \tag{2.12}$$

где $\alpha(\varepsilon, \delta) \to 0$ при $\varepsilon \to 0$, $\delta \to 0$.

Учитывая затем свойства 4^0, 5^0 функции $\rho_\varepsilon^0(t, x)$ (2.1) и краевое условие $\rho^0(\vartheta, x) = |x|$ для функции $\rho^0(t, x)$ (1.10), приходим к выводу, что закон управления U^0 (2.7) для любого, сколь угодно малого $\chi > 0$ обеспечивает неравенство

$$|x[\vartheta]| = \gamma[t_*, x_*; \quad U^0, v(t_*[\cdot]\vartheta]] \leqq \rho^0(t_*, x_*) + \chi, \tag{2.13}$$

если только ε и δ в законе U^0 (2.7) выбраны в зависимости от χ достаточно малыми.

С другой стороны известно [1], что можно указать закон формирования помехи

$$V^0 = \{v^0(\cdot), \ \varepsilon \leqq \varepsilon(\chi), \ \Delta_\delta : \delta \leqq \delta(\varepsilon)\}, \tag{2.14}$$

основанный на некоторой оптимальной стратегии

$$v^0(\cdot) = \{v^0(t, x, \varepsilon) \in \mathcal{P}, t_0 \leqq t \leqq \vartheta, x \in E^n, \varepsilon > 0\}, \tag{2.15}$$

который дает результат

$$\gamma[t_*, x_*; \ u(t_*[\cdot]\vartheta], V^0] \geqq \rho^0(t_*, x_*) - \chi, \tag{2.16}$$

какова бы ни была реализация управления $u(t_*[\cdot]\vartheta]$, в том числе, быть может и та, которая по шагам формируется законом управления U^0 (2.7).

Сопоставляя (2.13) и (2.16), убеждаемся, что величина $\rho^0(t_*, x_*)$ в (2.13) не может быть уменьшена ни на какую угодно малую постоянную $\xi > 0$ и потому она доставляет оптимальный гарантированный результат $\rho[u^0(\cdot); \ t_*, \ x_*]$ в смысле определения (1.6).

При этом в соответствии с условием (1.18), обеспечивающем оценку (2.8), универсальная стратегия $u^0(\cdot)$ находится из условия экстремального сдвига

$$\left\langle \frac{\partial \rho_\varepsilon^0}{\partial x} \cdot B(t) u^0(t, x, \varepsilon) \right\rangle = \min_{u \in \mathcal{P}} \left\langle \frac{\partial \rho_\varepsilon^0}{\partial x} \cdot B(t) u \right\rangle \tag{2.17}$$

в направлении против градиента $\partial \rho_\varepsilon^0 / \partial x$ функции $\rho_\varepsilon^0(t, x)$ (2.1).

В заключение параграфа заметим, что оптимальная стратегия $v^0(\cdot)$ (2.15) в соответствии с методом стохастического программного синтеза может быть найдена на основе функции $\rho^0(t, x)$ (1.10) также по правилу экстремального сдвига

$$\langle s_v(t, x, \varepsilon) \cdot C(t) v^0(t, x, \varepsilon) \rangle = \min_{v \in \mathcal{P}} \langle s_v(t, x, \varepsilon) \cdot C(t) v \rangle, \tag{2.18}$$

но уже в направлении против вектора $s_v(t, x, \varepsilon)$, определяемого из условия

$$\rho^0(t, x - s_v(t, x, \varepsilon)) = \max_{|s| \leq \beta(\varepsilon, t)} \rho^0(t, x - s). \tag{2.19}$$

Однако обойти решение задачи (2.19) так же, как это было сделано в случае задачи (1.14), здесь не удается. Во всяком случае нужно иметь в виду, что экстремальный сдвиг (2.18) не может быть заменен здесь на экстремальный сдвиг вдоль вектора градиента $\partial \rho_\varepsilon^0 / \partial x$ функции $\rho_\varepsilon^0(t, x)$ (2.1). Таким образом, задача построения стратегии $v^0(\cdot)$ (2.15) является более трудной. В этой связи представляет существенный интерес возможность формировать оптимальные контрвоздействия $v^0[t]$, $t_i < t \leqq t_{i+1}$, опираясь на максимизирующую стохастическую неупреждающую программу $v^0(t, \omega)$, $t_* < t \leqq \vartheta$, $\omega \in \Omega$, извлекаемую из решения задачи (1.10). Оказывается, что на движении, порожденном законом управления U^0 (2.7) и максимизирующей программой $v^0(t, \omega)$, можно с вероятностью, сколь угодно близкой к единице, получить результат γ (1.1), сколь угодно близкий к величине $\rho^0(t_*, x_*)$.

3. Модельная задача

Проиллюстрируем предлагаемый подход на задаче сближения уклонения ([5], с. 51) двух объектов.

Пусть в трехмерном пространстве движутся две материальные точки $m^{(1)}$ и $m^{(2)}$ с массами m_1 и m_2 соответственно. Положение этих точек в пространстве определяется их радиусами — векторами

$$r^{(k)}[t] = \{r_1^{(k)}[t], r_2^{(k)}[t], r_3^{(k)}[t]\},$$
$$k = 1, 2; \quad t_0 \leqq t \leqq \vartheta. \tag{3.1}$$

Движение точек происходит под действием сил сопротивления окружающей среды, пропорциональных скоростям точек $dr^{(k)}/dt$, $k = 1, 2$, и сил притяжения (или отталкивания), пропорциональных расстояниям точек до начала координат. Первая точка $m^{(1)}$ управляется силой $u^{(1)}$, вторая точка — силой $u^{(2)}$. В каждый момент времени t силы ограничены по величине

$$|u^{(k)}[t]| \leqq \mu_k, \qquad \mu_k = \text{const}, \qquad k = 1, 2. \tag{3.2}$$

Уравнения движения точек имеют вид

$$m_k \frac{d^2 r^{(k)}}{dt^2} + \alpha_k \frac{dr^{(k)}}{dt} + \beta_k r^{(k)} = u^{(k)}, \qquad k = 1, 2, \tag{3.3}$$

где m_k и α_k — некоторые неотрицательные, а β_k — произвольные числа.

Примем, что целью первого игрока, распоряжающегося выбором управления $u^{(1)}$, является сближение в заданный момент времени ϑ точки $m^{(1)}$ с точкой $m^{(2)}$, т. е. первый игрок стремится выбором своих управляющих воздействий $u^0[t]$ (1.17) сделать по возможности наименьшим в момент времени ϑ расстояние до точки $m^{(2)}$

$$\gamma = |r^{(1)}[\vartheta] - r^{(2)}[\vartheta]|. \tag{3.4}$$

При этом первый игрок располагает в нужные для него моменты времени t_i, $i = 1, \ldots, k$ сведениями о реализующихся состояниях системы (3.3), но ему ничего неизвестно о намерениях второго игрока, распоряжающегося выбором управления $u^{(2)}$. Возможно, что второй игрок будет стремиться во вред первому максимально увеличить это расстояние γ (3.4).

Известным способом ([5], с. 53) рассматриваемая дифференциальная игра может быть редуцирована к задаче на минимум гарантированного результата по показателю $\gamma = |x[\vartheta]|$ (1.1) для системы

$$\frac{dx}{dt} = \varphi(t)u - \psi(t)v, \tag{3.5}$$

$$u = \frac{1}{m_1}u^{(1)}, \quad v = \frac{1}{m_2}u^{(2)}, \quad \frac{\mu_1}{m_1} = \mu, \quad \frac{\mu_2}{m_2} = \nu, \tag{3.6}$$

где $x = \{x_1, x_2, x_3\}$ — трехмерный вектор, связанный с фазовыми координатами точек $m^{(1)}$ и $m^{(2)}$ соотношениями

$$\begin{aligned}
x_1 &= x_{11}^{(1)}[\vartheta, t]r_1^{(1)} + x_{12}^{(1)}[\vartheta, t]r_1^{*(1)} - \\
&\quad - x_{11}^{(2)}[\vartheta, t]r_1^{(2)} - x_{12}^{(2)}[\vartheta, t]r_1^{*(2)}; \\
x_2 &= x_{11}^{(1)}[\vartheta, t]r_2^{(1)} + x_{12}^{(1)}[\vartheta, t]r_2^{*(1)} - \\
&\quad - x_{11}^{(2)}[\vartheta, t]r_2^{(2)} - x_{12}^{(2)}[\vartheta, t]r_2^{*(2)}; \\
x_3 &= x_{11}^{(1)}[\vartheta, t]r_3^{(1)} + x_{12}^{(1)}[\vartheta, t]r_3^{*(1)} - \\
&\quad - x_{11}^{(2)}[\vartheta, t]r_3^{(2)} - x_{12}^{(2)}[\vartheta, t]r_3^{*(2)}.
\end{aligned} \tag{3.7}$$

Здесь функции $x_{11}^{(k)}[\vartheta, t]$, $x_{12}^{(k)}[\vartheta, t]$, $k = 1, 2$ — элементы некоторой фундаментальной матрицы, вычисляемой по исходным параметрам системы

$$a_1^{(1)} = \frac{\alpha_1}{m_1}, \quad a_2^{(1)} = \frac{\beta_1}{m_1}, \quad a_1^{(2)} = \frac{\alpha_2}{m_2}, \quad a_2^{(2)} = \frac{\beta_2}{m_2}. \tag{3.8}$$

Например, для функций $x_{12}^{(k)}[\vartheta, t]$, $k = 1, 2$ в зависимости от значений

параметров (3.8) возможны следующие три выражения

а)
$$[a_1^{(k)}]^2 - 4a_2^{(k)} < 0, \tag{3.9}$$

$$x_{12}^{(k)}[\vartheta, t] = \frac{2}{\sqrt{4a_2^{(k)} - [a_1^{(k)}]^2}} \exp\left[-\frac{a_1^{(k)}}{2}(\vartheta - t)\right] \cdot$$

$$\cdot \sin\left[\frac{1}{2}\sqrt{4a_2^{(k)} - [a_1^{(k)}]^2}\,(\vartheta - t)\right];$$

б)
$$[a_1^{(k)}]^2 - 4a_2^{(k)} > 0, \tag{3.10}$$

$$x_{12}^{(k)}[\vartheta, t] = \frac{2}{\sqrt{[a_1^{(k)}]^2 - 4a_2^{(k)}}} \exp\left[-\frac{a_1^{(k)}}{2}(\vartheta - t)\right] \cdot$$

$$\cdot \operatorname{sh}\left[\frac{1}{2}\sqrt{[a_1^{(k)}]^2 - 4a_2^{(k)}}(\vartheta - t)\right];$$

в)
$$[a_1^{(k)}]^2 - 4a_2^{(k)} = 0, \tag{3.11}$$

$$x_{12}^{(k)}[\vartheta, t] = (\vartheta - t)\exp\left[-\frac{a_1^{(k)}}{2}(\vartheta - t)\right].$$

Если же $a_2^{(k)} = 0$, т. е. сила притяжения, действующая на k-тую точку, отсутствует, то

$$x_{12}^{(k)}[\vartheta, t] = \frac{1}{a_1^{(k)}}[1 - \exp[-a_1^{(k)}(\vartheta - t)]]. \tag{3.12}$$

Если к тому же отсутствует и сила сопротивления, т. е. $a_1^{(k)} = 0$, то

$$x_{12}^{(k)}[\vartheta, t] = \vartheta - t. \tag{3.13}$$

Функции $\varphi(t)$ и $\psi(t)$, $t_0 \leqq t \leqq \vartheta$ в уравнении (3.5) определяются равенствами

$$\varphi(t) = x_{12}^{(1)}[\vartheta, t], \qquad \psi(t) = x_{12}^{(2)}[\vartheta, t]. \tag{3.14}$$

Построим уравнение

$$h(\tau) \equiv \nu|\psi(\tau)| - \mu|\varphi(\tau)| = 0. \tag{3.15}$$

Оно имеет корень $\tau_0^* = \vartheta$. Пусть это уравнение на полуинтервале $[t_0, \vartheta)$ имеет еще p действительных корней τ_s^*, $s = 1, \ldots, p$, так что $\tau_{s+1}^* < \tau_s^*$. Тогда после решения задач (1.11), (1.10) для вычисления оптимального гарантированного результата получается следующее явное выражение

$$\rho^0(t_*, x_*) = \max\left\{ |x_*| + \int_{t_*}^{\vartheta} h(\tau)d\tau;\right.$$

$$\left. \int_{\tau_j^*}^{\vartheta} h(\tau)d\tau,\ j : \tau_j^* \in [t_*, \vartheta]\right\}. \tag{3.16}$$

Так что, если, например, $p=1$ и при этом выполняется условие

$$h(\tau)<0, \quad t_0<\tau<\tau_1^*; \qquad h(\tau)>0, \quad \tau_1^*<\tau<\vartheta \tag{3.17}$$

то из равенства (3.16) получается следующая простая формула

$$\rho^0(t_*, x_*)=\begin{cases} |x_*|+\int\limits_{t_*}^{\vartheta} h(\tau)d\tau, \text{ если } t_*\geqq\tau_1^* \text{ или} \\ \text{если } t_*<\tau_1^* \text{ и } r(t_*, x_*)>0; \\ \int\limits_{\tau_1^*}^{\vartheta} h(\tau)d\tau, \text{ если } t_*<\tau_1^* \text{ и } r(t_*, x_*)\leqq 0, \end{cases} \tag{3.18}$$

где обозначено

$$r(t_*, x_*)=|x_*|+\int\limits_{t_*}^{\tau_1^*} h(\tau)d\tau, \quad t_*<\tau_1^*. \tag{3.19}$$

Заметим, что функция $\rho^0(t, x)$ (3.18) не дифференцируема по векторному аргументу x. После необходимых вычислений по предложенным формулам (2.1), (2.2), (1.11) в рассматриваемом случае можно в явном виде выписать и выражение $\rho_\varepsilon^0(t_*, x_*)$, определяющее здесь аппроксимирующую функцию $\rho_\varepsilon^0(t, x)$ (2.1).

Это выражение с учетом (3.18), (3.19) принимает вид

$$\rho_\varepsilon^0(t_*, x_*)=\begin{cases} \rho^0(t_*, x_*)-\sqrt{2}\beta(\varepsilon, t_*), \text{ если } t_*\geqq\tau_1^* \text{ и} \\ |x_*|\geqq\dfrac{1}{\sqrt{2}}\beta(\varepsilon, t_*) \text{ или если } t_*<\tau_1^* \\ \text{и } r(t_*, x_*)\geqq\dfrac{1}{\sqrt{2}}\beta(\varepsilon, t_*); \\ \rho^0(t_*, x_*)-|x_*|-[\beta(\varepsilon, t_*)-|x_*|^2]^{1/2}, \\ \text{если } t_*\geqq\tau_1^* \text{ и } 0\leqq|x_*|<\dfrac{1}{\sqrt{2}}\beta(\varepsilon, t_*); \\ \rho^0(t_*, x_*)-\beta(\varepsilon, t_*), \text{ если } t_*<\tau_1^* \text{ и} \\ \int\limits_{t_*}^{\tau_1^*} h(\tau)d\tau\leqq r(t_*, x_*)\leqq 0; \\ \rho^0(t_*, x_*)-r(t_*, x_*)-[\beta^2(\varepsilon, t)- \\ -r^2(t_*, x_*)]^{1/2} \text{ если } t_*<\tau_1^* \text{ и} \\ 0<r(t_*, x_*)<\dfrac{1}{\sqrt{2}}\beta(\varepsilon, t_*). \end{cases} \tag{3.20}$$

В согласии с изложенной теорией для функции $\rho_\varepsilon^0(t, x)$ (3.20) выполняется краевое условие (2.5), она имеет градиент $\partial\rho_\varepsilon^0(t, x)/\partial x$, для нее справедлива оценка (2.6) и выполняется дифференциальное неравенство (2.9).

Вычисляя градиент функции $\rho_\varepsilon^0(t, x)$ (3.20) и производя против него экстремальный сдвиг (1.18), находим, что оптимальная стратегия $u^0(\cdot)$ в рассматриваемом случае имеет следующий явный вид

$$u^0(t, x, \varepsilon) = \begin{cases} -\mu \dfrac{x}{|x|} \operatorname{sgn} \varphi(t), & |x| \neq 0 \\ 0, & |x| = 0. \end{cases} \tag{3.21}$$

Дополнительный анализ показывает, что оптимальная стратегия $u^0(\cdot)$ в рассматриваемом примере сохраняет вид (3.21) не только при условии (3.17), но и в общем случае произвольных функций $\varphi(t)$ и $\psi(t)$, $t_0 \leqq t \leqq \vartheta$.

Здесь удается также на основе экстремального сдвига (2.18) построить и оптимальную стратегию $v^0(\cdot)$ (2.15). Как и стратегия $u^0(\cdot)$ (3.21), она получается не зависящей от параметра ε и имеет следующий вид

$$v^0(t, x, \varepsilon) = \begin{cases} -v \dfrac{x}{|x|} \operatorname{sgn} \psi(t), & |x| \neq 0 \\ v^0 : |v^0| = v, & |x| = 0. \end{cases} \tag{3.22}$$

Говорят [1–3], что указанная пара стратегий $\{u^0(\cdot)$ (3.21), $v^0(\cdot)$ (3.22)$\}$ составляет седловую точку позиционной дифференциальной игры, имеющей цену $\rho^0(t_*, x_*)$, вычисляемую по формуле (3.16).

Оптимальные законы управления U^0 (2.7) и V^0 (2.14), основанные на оптимальных стратегиях $u^0(\cdot)$ (3.21) и $v^0(\cdot)$ (3.22), были испытаны на ЭВМ при следующих исходных данных

$$t_0 = 0, \quad \vartheta = 5, \quad t_* = 0, \quad m_1 = 1.0, \quad m_2 = 0.5, \quad \alpha_1 = 0.1,$$

$$\alpha_2 = 0.4, \quad \beta_1 = \beta_2 = 4, \quad \mu_1 = \mu_2 = 1, \quad r_{1*}^{(1)} = r_{2*}^{(1)} = 0, \quad r_{3*}^{(1)} = 0.7,$$

$$\dot{r}_{1*}^{(1)} = \dot{r}_{2*}^{(1)} = -0.2, \quad \dot{r}_{3*}^{(1)} = 0, \quad r_{1*}^{(2)} = r_{2*}^{(2)} = r_{3*}^{(2)} = 0,$$

$$\dot{r}_{1*}^{(2)} = \dot{r}_{2*}^{(2)} = 0.2, \quad \dot{r}_{3*}^{(2)} = 0.$$

Этим данным отвечает случай а) (3.9). При этом уравнение (3.15) имеет на отрезке [0, 5] семь действительных корней, отличных от корня $\tau_0^* = \vartheta = 5$, и цена игры, подсчитанная по формуле (3.16), составляет $\rho^0 = 0{,}094$. На рис. 1 приведены траектории объекта $m^{(1)}$ (линия с точками) и объекта $m^{(2)}$ (сплошная

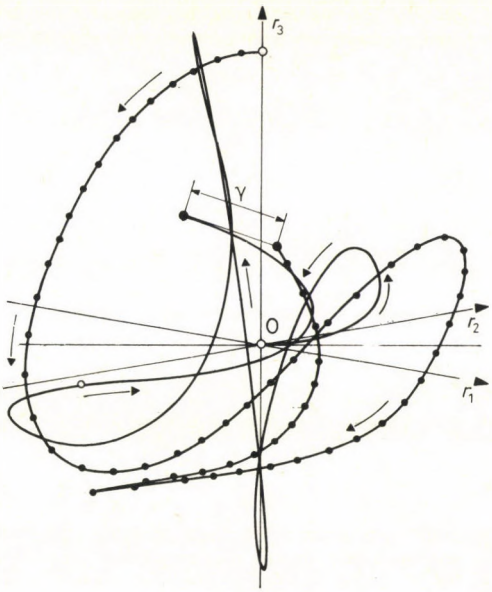


Рис. 1

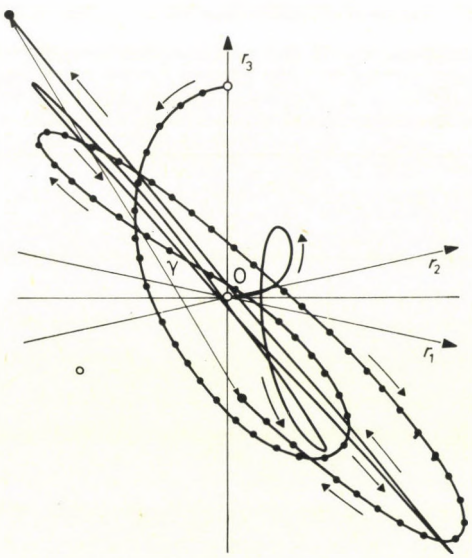


Рис. 2

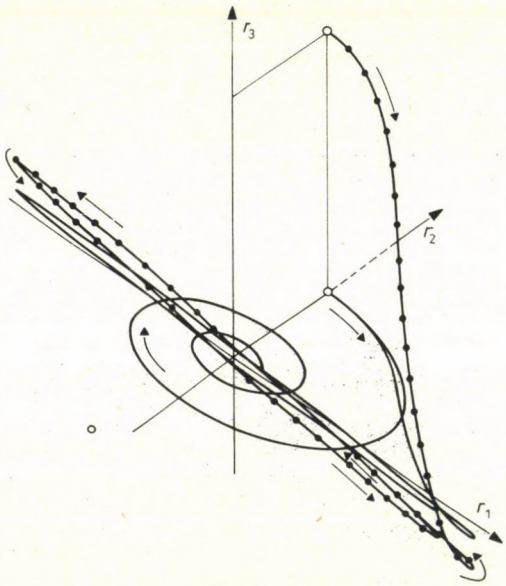


Рис. 3

линия), порожденные оптимальными законами управления. Исходные данные и вид проекции подобраны так, что изображенные траектории целиком расположены в плоскости рисунка. В обоих законах управления шаг δ был выбран постоянным и равным 0,001. В согласии с теорией расстояние между точками в момент времени ϑ получилось равным $\gamma = 0{,}095 \approx \rho^0 = 0{,}094$.

На рис. 2 представлены траектории объектов в случае, когда второй игрок оптимально уклоняется от сближения, а первый игрок в каждый момент времени $t \in [0,5]$ прикладывает к объекту $m^{(1)}$ силу, максимальной величины μ_1 и направленную прямо на объект $m^{(2)}$. В этом случае, поскольку действия первого игрока неоптимальны, получилось, что $\gamma = 1{,}311 > \rho^0 = 0{,}094$.

И наконец, на рис. 3 изображены траектории объектов, начинающиеся из позиций

$$t_* = 0, \quad r_{1*}^{(1)} = 0, \quad r_{2*}^{(1)} = 0{,}1, \quad r_{3*}^{(1)} = 0{,}2, \quad \dot{r}_{1*}^{(1)} = 0{,}5, \quad \dot{r}_{2*}^{(1)} = \dot{r}_{3*}^{(1)} = 0;$$

$$r_{1*}^{(2)} = 0, \quad r_{2*}^{(2)} = 0{,}1, \quad r_{3*}^{(2)} = 0, \quad \dot{r}_{1*}^{(2)} = 0{,}5, \quad \dot{r}_{2*}^{(2)} = \dot{r}_{3*}^{(2)} = 0$$

в случае, когда движение первого объекта $m^{(1)}$ порождается оптимальным законом управления, а второй игрок полагает $v(t, x) \equiv 0$, т. е. объект $m^{(2)}$ движется в плоскости $\{r_1, r_2\}$ под действием заданных сил по логарифмической спирали. В этом случае, как и следовало ожидать, получилось, что $\gamma = 0.000 < \rho^0 = 0.094$.

Результаты вычислительного эксперимента показывают, что оптимальный закон управления, построенный на основе предложенного в статье подхода, может быть устойчиво реализован на ЭВМ.

Литература

1. *Красовский Н. Н.* Управление динамической системой. М.: Наука, 1985, 520 с.
2. *Красовский Н. Н., Третьяков В. Е.* О программном синтезе позиционного управления. Докл. АН СССР, 1982, т. **264,** № *6;* с. 1309–1312.
3. *Красовский А. Н., Третьяков В. Е.* Стохастический программный синтез в дифференциальной игре с интегральной платой. Прикл. матем. и мех., 1984, т. **48,** вып. *6,* с. 883–891.
4. *Субботин А. Н., Субботина Н. Н.* Свойства дифференцируемости функции цены дифференциальной игры с интегрально-терминальной платой. Проблемы управления и теории информации, 1983, т. **12** *(3),* с. 153–166.
5. *Красовский Н. Н., Третьяков В. Е.* Задачи управления с гарантированным результатом. Свердловск, Средне-Уральское книжное издательство, 1986, 64 с.
6. *Ширяев А. Н.* Вероятность. М.: Наука, 1980, 576 с.
7. *Красовский Н. Н.* Детерминированная стратегия и стохастические программы. Прикл. матем. и мех., 1985, т. **49,** вып. *2,* с. 179–198.

PRINTED IN HUNGARY

Akadémiai Kiadó és Nyomda Vállalat, Budapest

NOTE TO CONTRIBUTORS

Two copies of the *manuscript* (each complete with figures, tables and references) are to be sent to

E.D. TERYAEV coordinating editor
Department of Mechanics and Control Processes
Academy of Sciences of the USSR
Leninsky Prospect 14, Moscow V-71, USSR

or to V. STREJC
UTIA ČSAV
182 08 Prague 8
Pod vodárenskou věží 4, Czechoslovakia

or to L. GYÖRFI
Technical University of Budapest
H-1111 Budapest, Stoczek u. 2, Hungary

Authors are requested to retain a third copy of the submitted typescript to be able to check the proofs.

The papers, preferably in English or Russian, should be typed double spaced on one side of good-quality paper with wide margins (4–5 cm). The first page of the paper should carry the title, the author(s)' names and the name of the town where they are active. The name and address of the author to whom the proofs should be sent should be given at the end of the paper. An *abstract* should head the paper. English papers should also have a Russian abstract.

The papers should not exceed 15 pages (25 × 50 characters per page) including tables and references. The proper location of the tables and figures must be indicated on the margin.

Mathematical notations should follow up-to-date usage. Equations longer than half a line should not be incorporated in the text. In-text equations must be typed on a single line except that one level of subscripting and/or superscripting is permissible. Use / instead of horizontal bars. Displayed equations should be written so as to require the fewest possible lines. Therefore use "exp" for the exponential function whenever the exponent requires more than a single line. Matrices should, if possible, not be written in full. Use subscript notations instead such as $A = ||a_{ij}||$. Write diagonal matrices as diag $(d_1, d_2, \ldots d_n)$.

The authors will be sent galley proofs to be returned by next mail. Rejected manuscripts will be returned. Authors will receive 100 reprints free of charge. Additional reprints may be ordered.

К СВЕДЕНИЮ АВТОРОВ

Рукописи статей в трех экземплярах на русском языке и в трех на английском следует направлять по адресу: 117312 Москва В-312, просп. 60 летия Октября, 9, МНИИПУ. Редакция журнала «Проблемы управления и теории информации» (зав. редакцией Н. И. Родионова, тел. 208-60-19).

Объём статьи не должен превышать 15 печатных страниц (25 строк по 50 букв). Статье должна предшествовать аннотация объемом 50–100 слов и приложено резюме–реферат объемом не менее 10–15% объема статьи на русском языке в трех экземплярах, на котором напечатан служебный адрес автора (фамилия, название учреждения, адрес).

При написании статьи авторам надо строго придерживаться следующей формы: введение (постановка задачи), основное содержание, примеры практического использования, обсуждение результатов, выводы и литература.

Статьи должны быть отпечатаны с промежутком в два интервала, последовательность таблиц и рисунков должна быть отмечена на полях. Математические обозначения рекомендуется давать в соответствии с современными требованиями и традициями. Разметку букв следует производить только во втором экземпляре и русского, и английского варианта статьи.

Авторам высылается верстка, которую необходимо незамедлительно проверить и возвратить в редакцию.

После публикации авторам высылаются бесплатно 100 оттисков их статей.

Рукописи непринятых статей возвращаются авторам.

CONTENTS · СОДЕРЖАНИЕ

HU ISSN 0370-2529 Index 26 660

VOL. 17 · NUMBER 5
TOM HOMEP

ACADEMY OF SCIENCES OF THE USSR
HUNGARIAN ACADEMY OF SCIENCES
CZECHOSLOVAK ACADEMY OF SCIENCES

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

АКАДЕМИЯ НАУК СССР
ВЕНГЕРСКАЯ АКАДЕМИЯ НАУК
ЧЕХОСЛОВАЦКАЯ АКАДЕМИЯ НАУК

1988

AKADÉMIAI KIADÓ, BUDAPEST
DISTRIBUTED OUTSIDE THE COMECON-COUNTRIES
BY PERGAMON PRESS, OXFORD

PROBLEMS OF CONTROL AND INFORMATION THEORY

An international bi-monthly sponsored jointly by the Presidium of the Academy of Sciences of the USSR, of the Hungarian Academy of Sciences and of the Czechoslovak Academy of Sciences. The six issues published per year make up a volume of some 480 pp. It offers publicity for original papers and short communication of the following topics:

— theory of control processes
— theory of adaptive systems
— theory of estimation and identification
— theory of controlling robot-technologic and flexible manufacturing systems
— information theory
— information-theoretic aspects of multiple access networks.

While this bi-monthly is mainly a publication forum of the research results achieved in the socialist countries, also papers of international interest from other countries are welcome.

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

Международный журнал Академии наук СССР, Венгерской Академии наук и Чехословацкой Академии наук выходит 6 раз в год общим объемом 480 печатных страниц.

В журнале публикуются оригинальные научные статьи и статьи обзорного характера по следующим проблемам управления и теории информации:

— теория процессов управления;
— теория адаптивных систем;
— теория оценивания и идентификации;
— теория управления робототехническими и гибкими производственными системами;
— теория информации;
— теория информации в области сетей с множественным доступом.

Целью журнала является ознакомление научной общественности различных стран с важнейшими проблемами, имеющими актуальный и перспективный характер, научными достижениями ученых социалистических и других стран.

Distributors

For the Soviet Union:

SOYUZPECHATY, Moscow 123 308 USSR

For Albania, Bulgaria, China, Cuba, Czechoslovakia, German Democratic Republic, Korean People's Republic, Mongolia, Poland, Rumania, Vietnam and Yugoslavia:

KULTURA Hungarian Foreign Trading Co.
P. O. Box 149 H-1389 Budapest, Hungary

For all other countries:

PERGAMON JOURNALS LTD, Headington Hill Hall, Oxford OX3 OBW, England

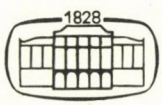
or

PERGAMON JOURNALS INC, Maxwell House, Fairview Park, Elmsford, NY 10523, USA

1988 Subscription Rate DM 535,— per annum including postage and insurance.

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ



AKADÉMIAI KIADÓ

PUBLISHING HOUSE OF THE HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST

Problems of Control and Information Theory, Vol. 17 (5), pp. 253—266 (1988)

MINIMUM-DISTANCE AND GNOSTICAL ESTIMATORS

I. VAJDA

(*Prague*)

(Received October 1, 1987)

This paper is based on the result of Fabian [3] who showed that three estimators of Kovanic [6], derived from non-statistical (gnostical) considerations, are particular cases of α-estimators which are minimum distance-type estimators proposed by us in [10]. Using known statistical properties of the general α-estimators, we establish here the corresponding statistical properties of the three gnostical estimators.

1. Notation

$\mathbf{R}$ is the real line and $\mathbf{R}^* = \mathbf{R} \cup \{-\infty, +\infty\}$ the extended real line. We consider on $\mathbf{R}^*$ the same ordering topology and arithmetic as Halmos [13] in his Prerequisites. In particular, we assume

$$0.\pm\infty = 0, \quad -1.\pm\infty = \mp\infty, \quad \text{and} \quad \pm\infty + t = \pm\infty, \quad t \in \mathbf{R}.$$

Each Borel subset $B \subset \mathbf{R}^*$ is considered to be a measurable space, with the σ-algebra of Borel subsets of B. Conversely, in a measurable space $(B, \mathscr{B})$ with a Borel subset $B \subset \mathbf{R}^*$, $\mathscr{B}$ is assumed to be the σ-algebra of Borel subsets of B.

The topological space $\mathbf{R}^*$ may be viewed as a two-point compactification of $\mathbf{R}$. By $\mathbf{R}^0$ we denote the one-point compactification $\mathbf{R} \cup \{\infty\}$ (cf. Kelley [14]). Thus we distinguish between ∞ and $+\infty$.

Let $(\mathscr{X}, \mathscr{B}, \mu)$ be a measure space and $\varphi : \mathscr{X} \mapsto \mathbf{R}^*$ a measurable function. If φ is nonnegative and not μ-integrable (i.e. the integral $\int_{\mathscr{X}} \varphi \, d\mu$ is not finite) then we put $\int_{\mathscr{X}} \varphi \, d\mu = +\infty$. We shall say that the integral $\int_{\mathscr{X}} \varphi \, d\mu$ exists if either the positive part φ^+ or the negative part φ^- of φ, is μ-integrable, in which case we put

$$\int_{\mathscr{X}} \varphi \, d\mu = \int_{\mathscr{X}} \varphi^+ d\mu - \int_{\mathscr{X}} \varphi^- d\mu.$$

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

$\mathbf{P}$ denotes the set of all probability measures on $\mathbf{R}$, and $\mathbf{P}_{emp}$ denotes the subset of $\mathbf{P}$ containing all *empirical probability measures*

$$P_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}, \qquad \mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{R}^n, \quad n = 1, 2, \ldots \tag{1}$$

where, here and in the sequel, δ_x for $x \in \mathbf{R}$ denotes the probability measure on $\mathbf{R}$ with all mass concentrated at x. The index x in P_n is omitted, in accordance with the common convention of statistical literature.

Q denotes an absolutely continuous probability measure on $\mathbf{R}$. It is supposed that the density q of Q is even, bounded, continuous, and unimodal. The mode of q is obviously at 0. $\mathcal{Q}$ is the family of probability measures $(Q_\theta \mid \theta \in \mathbf{R})$, where Q_θ has the density $q(x-\theta)$.

2. α-estimators of location

Let us consider continuous functions f_α: $[0, \infty) \mapsto \mathbf{R}^*$, $0 \leqq \alpha \leqq 1$, defined by the condition that for every $t > 0$

$$f_\alpha(t) = \begin{cases} \dfrac{1-t^\alpha}{\alpha}, & \text{if} \quad 0 < \alpha \leqq 1 \\ -\ln t, & \text{if} \quad \alpha = 0. \end{cases}$$

Since q is assumed to be bounded, the integrals

$$D_\alpha(P, \theta) = \int_{\mathbf{R}} f_\alpha(q(x-\theta))\, dP(x)$$

$$= \begin{cases} \dfrac{1}{\alpha}\left(1 - \displaystyle\int_{\mathbf{R}} q(x-\theta)^\alpha dP(x)\right), & \text{if} \quad 0 < \alpha \leqq 1 \\ -\displaystyle\int_{\mathbf{R}} \ln q\,(x-\theta)\, dP(x), & \text{if} \quad \alpha = 0 \end{cases} \tag{2}$$

exist for all $P \in \mathbf{P}$ and all $\theta \in \mathbf{R}$, and take on values from $\mathbf{R}$ if $0 < \alpha \leqq 1$ and from $\mathbf{R} \cup \{+\infty\}$ if $\alpha = 0$.

Let us denote by $f_\alpha(P_1 \| P_2)$ the divergences of probability measures $P_1, P_2 \in \mathbf{P}$ defined by the convex functions f_α in accordance with Csiszár [2]. Let us say that a sequence of σ-algebras $(\mathcal{B}^{(\nu)} \mid \nu = 1, 2, \ldots)$ is exhaustive if $\mathcal{B}^{(1)} \subset \mathcal{B}^{(2)} \subset \ldots$ and if the union of these σ-algebras generates the σ-algebra of Borel subsets of $\mathbf{R}$. Denote by the superscript $^{(\nu)}$ restrictions of probability measures from $\mathbf{P}$ on the σ-algebra $\mathcal{B}^{(\nu)}$.

Lemma 1. Let us consider $0 \leqq \alpha < 1$, a natural number n, and an absolutely continuous measure $P \in \mathbf{P}$, and let us suppose that q is differentiable, with a derivative

q', and that

$$\sup_{x \in \mathbf{R}} \frac{|q'(x)|}{q(x)} < \infty . \tag{3}$$

Then there exists an exhaustive sequence of σ-algebras $(\mathscr{B}^{(v)} | v = 1, 2, \ldots)$ such that for P^n-almost every $\mathbf{x} \in \mathbf{R}^n$ there exists a sequence $(l_v | v = 1, 2, \ldots)$ of increasing affine functions such that P_n defined by (1) satisfies the relation

$$\lim_{v \to \infty} \sup_{\theta \in \mathbf{R}} |l_v(f_\alpha(Q_\theta^{(v)} || P_n^{(v)})) - D_\alpha(P_n, \theta)| = 0.$$

Proof. It follows from the assumptions that it holds

$$\sup_{x \in \mathbf{R}} \left| \frac{d}{d\theta} f_\alpha(q(x - \theta)) \right| < \infty .$$

Hence the assumptions of the Lemma in our paper [10] hold. It follows from that Lemma that for P^n-almost every $\mathbf{x} \in \mathbf{R}^n$ there exists $(\mathscr{B}^{(v)} | v = 1, 2, \ldots)$ such that the assertion of Lemma 1 holds. It is easy to see that the construction of $(\mathscr{B}^{(v)} | v = 1, 2, \ldots)$ can be made independent of $\mathbf{x} \in \mathbf{R}^n$ — for details we refer to (12.15) in Vajda [12]. □

Remark. Some densities q considered in the present paper do satisfy assumption (3). The standard normal density $q(x) = (2\pi)^{-1/2}$ exp. $\{-x^2/2\}$ does not do so. If (3) is not satisfied then it follows from (12.15) in Vajda [12] that the assertion of Lemma 1 remains true with the last relation replaced by the following weaker relation

$$\lim_{v \to \infty} |l_v(f_\alpha(Q_\theta^{(v)} || P_n^{(v)})) - D_\alpha(P_n, \theta)| = 0, \quad \theta \in \mathbf{R}. \qquad \square$$

It follows from Lemma 1 that, for large v, the location $\theta \in \mathbf{R}$ which minimizes the divergence $f_\alpha(Q_\theta^{(v)} || P_n^{(v)})$ practically coincides with the location minimizing the function $D_\alpha(P_n, \theta)$. The affine functions $l_v(f_\alpha(Q_\theta^{(v)} || P_n^{(v)})$ just increase the contrast between the infimum and supremum of the argument function $f_\alpha(Q_\theta^{(v)} || P_n^{(v)})$, which becomes too flat as it tends to the constant $f_\alpha(0)$ for $v \to \infty$ (cf. Theorem 1 in Vajda [9]). Thus we are obtaining from Lemma 1 a sanction for the interpretation of the estimators we are now going to introduce as minimum distance estimators (more precisely, as minimum f_α-divergence estimators).

For $0 \leqq \alpha \leqq 1$, by an *α-estimator of location with projection parent Q* we mean a mapping $T^\alpha : \mathbf{P} \mapsto \mathbf{R}$ such that

$$T^\alpha(P) \in \arg\min D_\alpha(P, \theta) \tag{4}$$

and

$$\text{for every } n=1, 2, \ldots, \text{ the restriction } T_n^\alpha(\mathbf{x})=T^\alpha(P_n) \text{ on } \mathbf{R}^n, \text{ defined via (1), is measurable.} \tag{5}$$

Up to some technical details, the last estimator is a particular case of the α-estimator of arbitrary parameter defined by relations (2), (3) in our paper [10]. Let us also point out that, by (2), if $P=P_n$ is defined by a sample $\mathbf{x}=(x_1, \ldots, x_n)$ in accordance with (1), then

$$D_\alpha(P_n, \theta)=\begin{cases} \dfrac{1}{\alpha}\left(1-\dfrac{1}{n}\sum_{i=1}^{n} q(x_i-\theta)^\alpha\right), & \text{if } 0<\alpha\leqq 1 \\ -\sum_{i=1}^{n} \ln q(x_i-\theta), & \text{if } \alpha=0. \end{cases} \tag{6}$$

One can see from the lower formula that T^0 is nothing, but the well-known maximum likelihood estimator of location with the projection parent Q.

One can also see from (6) that all estimators T^α are closely related to the M-estimators of location of Huber [4, 5], provided his function $\rho: \mathbf{R} \mapsto \mathbf{R}$ is given by

$$\rho(x)=\begin{cases} \dfrac{1}{\alpha}(1-q(x)^\alpha), & \text{if } 0<\alpha\leqq 1 \\ -\ln q(x), & \text{if } \alpha=0. \end{cases}$$

The difference is of technical nature only: in the two cited publications, the M-estimators are formally defined by weaker conditions than (4), (5). These conditions are making the theory easier (cf. [4]), but the applicability of such a theory to what is computed in practical applications is less clear. This is the reason why we had to develop the theory of α-estimators in our own terms (cf. [11] and more recently [12]). This theory is close to the theory of Pfanzagl [8] who considered M-estimators under similar conditions as (4), (5) above. In the sequel we refer to the results of [11, 12].

3. Gnostical estimators of location

Kovanic [6] considered for $n=1, 2, \ldots$ vectors of data $\mathbf{z}=(z_1, \ldots, z_n)\in(0, \infty)^n$ and proposed to estimate an unknown value $z_0>0$ by means of mappings $\tau_n^\alpha: (0, \infty)^n \mapsto (0, \infty)$ for $\alpha \in \{0, 1/2, 1\}$, where

$$\tau_n^\alpha(\mathbf{z}) \in \arg\min \Delta_\alpha(\mathbf{z}, z_0), \qquad 0\leqq\alpha\leqq 1, \tag{7}$$

for

$$\Delta_\alpha(\mathbf{z}, z_0)=\begin{cases} -\sum_{i=1}^{n}\left(\dfrac{2}{\left[\left(\dfrac{z_i}{z_0}\right)^2+\left(\dfrac{z_i}{z_0}\right)^{-2}\right]^2}\right)^\alpha, & \text{if } 0<\alpha\leqq 1 \\ -\sum_{i=1}^{n}\ln\dfrac{2}{\left[\left(\dfrac{z_i}{z_0}\right)^2+\left(\dfrac{z_i}{z_0}\right)^{-2}\right]^2}, & \text{if } \alpha=0. \end{cases}$$

To verify this statement take at first into account the fact that if (7) holds, then $z_0=\tau_n^\alpha(\mathbf{z})$ is a solution of the equation

$$\frac{d}{dz_0}\Delta_\alpha(\mathbf{z}, z_0)=0.$$

This equation can be rewritten as follows:

$$\sum_{i=1}^{n}\frac{\left(\dfrac{z_i}{z_0}\right)^2-\left(\dfrac{z_i}{z_0}\right)^{-2}}{\left(\dfrac{z_i}{z_0}\right)^2+\left(\dfrac{z_i}{z_0}\right)^{-2}}=0, \qquad \text{if } \alpha=0$$

$$\frac{d}{dz_0}\sum_{i=1}^{n}\frac{1}{\left(\dfrac{z_i}{z_0}\right)^2+\left(\dfrac{z_i}{z_0}\right)^{-2}}=0, \qquad \text{if } \alpha=\frac{1}{2}$$

$$\frac{d}{dz_0}\sum_{i=1}^{n}\frac{1}{\left[\left(\dfrac{z_i}{z_0}\right)^2+\left(\dfrac{z_i}{z_0}\right)^{-2}\right]^2}=0, \qquad \text{if } \alpha=1,$$

which are the three equations in the lines 5, 7 and 8 of Table 2 on p. 311 of [6].

Using the hyperbolic cosine function $\operatorname{ch} x$, $x\in\mathbf{R}$, we see from the definition of $\Delta_\alpha(\mathbf{z}, z_0)$ that (7) is equivalent to

$$\tau_n^\alpha(\mathbf{z})\in\arg\min\frac{1}{n}\sum_{i=1}^{n}f_\alpha\left(\frac{1}{2\operatorname{ch}^2(2\ln(z_i/z_0))}\right),$$

where f_α, $0\leqq\alpha\leqq 1$, was defined at the beginning of Section 2. Transforming the data and parameter by

$$2\ln z_i=x_i, \qquad i=1,\ldots,n,\ 2\ln z_0=\vartheta, \tag{8}$$

and denoting by P_n the measure given by (1) for the transformed data vector

$\mathbf{x}=(x_1, \ldots, x_n)$ one obtains that (7) is equivalent with

$$2 \ln \tau_n^\alpha(\mathbf{z}) \in \arg\min D_\alpha(P_n, \theta),$$

where $D_\alpha(P_n, \theta)$ is defined by (6) for

$$q(x)=\frac{1}{2 \operatorname{ch}^2 x}, \qquad x \in \mathbf{R}. \tag{9}$$

It follows from here that if, for some $0 \leqq \alpha \leqq 1$, there exists the α-estimator T^α defined by (4), (5), then

$$\tau_n^\alpha(\mathbf{z})=\exp\{T_n^\alpha(\mathbf{x})/2\} \tag{10}$$

for all $\mathbf{z} \in (0, \infty)^n$ and $\mathbf{x} \in \mathbf{R}^n$ related by the one-one relation (8) and for all $n=1, 2, \ldots$

It follows from the last statement that Kovanic's *gnostical estimators* τ_n^α, $\alpha \in \{0, 1/2, 1\}$, are equivalent (via the one-one relations (8), (10)) to the restrictions T_n^α of α-estimators of location with the special projection parent density (9), for the special values $\alpha \in \{0, 1/2, 1\}$.

It follows from this characterization of the gnostical estimators that these may be identified with the α-estimators of location T^α with the projection parent given by (9), for $\alpha \in \{0, 1/2, 1\}$. Similar conclusion has been first established by Fabian [3].

Let us note that density (9) satisfies the relation

$$\frac{q'(x)}{q(x)}=-2 \operatorname{th} x, \qquad x \in \mathbf{R}, \tag{11}$$

so that condition (3) holds.

4. Properties of gnostical estimators

In this section we are interested in the estimators T^α, $0 \leqq \alpha \leqq 1$, with the projection parent density (9). Particular emphasis will be placed on values $\alpha \in \{0, 1/2, 1\}$ (see Section 3).

For every probability measure P on $(\mathbf{R}, \mathscr{B})$ we denote by P^∞ the probability measure $P \times P \times \ldots$ on $(\mathbf{R} \times \mathbf{R} \times \ldots, \mathscr{B} \times \mathscr{B} \times \ldots)$.

We shall consider a family $\mathscr{P}=(P_\theta | \theta \in \mathbf{R})$ of probability measures on $\mathbf{R}$ with densities $p(x-\theta)$, where p is even, bounded, unimodal, almost everywhere differentiable function. Let $p'(x)$ be the derivative of $p(x)$ if p is differentiable at x, and 0 otherwise. We shall assume that $p'(x) \neq 0$ almost everywhere.

Theorem 1. For every $0 \leqq \alpha \leqq 1$, the estimator T^α exists in the sense specified in (4) and (5).

Proof. It holds

$$\int_{\mathbf{R}} q(x-\theta)dP(x)>0, \qquad \theta\in\mathbf{R}, \quad P\in\mathbf{P},$$

and

$$\lim_{\theta\to\infty} q(x-\theta)=0, \qquad x\in\mathbf{R}.$$

Further, each t from the topological space $\mathbf{R}^0$ has neighborhood N_t ($N_\infty=\mathbf{R}-B$ for a compact set $B\subset\mathbf{R}$!) such that

$$\int_{\mathbf{R}} \inf_{\theta\in N_t} f_\alpha(q(x-\theta))dP(x)>-\infty, \qquad t\in\mathbf{R}^0, \ P\in\mathbf{P}.$$

and

$$\int_{\mathbf{R}} \sup_{\theta\in N_t} f_\alpha(q(x-\theta))dP(x)<+\infty, \qquad t\in\mathbf{R}, \quad P\in\mathbf{P}.$$

Hence all assumptions of Theorem 1.1 in our paper [11] hold (for the function f_α considered there replaced by the present function $-f_\alpha$). Therefore, the desired assertion follows from that Theorem. □

Theorem 2. For every $0\leqq\alpha\leqq 1$, the estimator T^α is strongly consistent in the sense that it holds for every $\theta\in\mathbf{R}$

$$\lim_{n\to\infty} T_n^\alpha(\mathbf{x})=\theta \quad P_\theta^\infty\text{-a.s.} \tag{12}$$

Proof. (1) We shall prove that it holds $\arg\min D_\alpha(P_0, \theta)=\{0\}$.

Let us define (cf. (11))

$$\psi(x-\theta)=\frac{d}{d\theta} f_\alpha(q(x-\theta))=-2q(x-\theta)^\alpha \operatorname{th}(x-\theta), \quad x,\theta\in\mathbf{R}. \tag{13}$$

It holds

$$D'_\alpha(P_\theta,\theta)=\frac{d}{d\theta} D_\alpha(P_0,\theta)=\int_{\mathbf{R}} \psi(x-\theta)dP_0(x)=$$

$$=\int_{\mathbf{R}} \psi(x)p(x+\theta)dx \tag{14}$$

and

$$D''_\alpha(P_0,\theta)=\frac{d}{d\theta} D'_\alpha(P_0,\theta)=\int \psi(x)p'(x+\theta)dx. \tag{15}$$

The function $\psi(x)p(x)$ is odd so that it holds $D'_\alpha(P_0, \theta)=0$. The function $\psi(x)p'(x)$ is even and almost everywhere positive. Therefore, it holds $D''_\alpha(P_0, 0)>0$. If we prove that $D'_\alpha(P_0, \theta)\neq 0$ for every $\theta\neq 0$, then the desired assertion follows from Corollary 2.5 of Mäkaläinen, Schmidt and Styan [7].

Let us consider $\theta<0$. The distribution function F corresponding to p is convex on $(-\infty, 0]$, concave on $[0, \infty)$, and skew-symmetric on $\mathbf{R}$ in the sense

$$F(-x)=1-F(x), \qquad x\geqq 0.$$

It holds

$$\frac{F(-x+\theta)+F(x+\theta)}{2}\geqq F(\theta), \qquad x\geqq 0. \tag{16}$$

Indeed, if $x+\theta\leqq 0$, then this inequality follows from the convexity of F on $(-\infty, 0]$. If $x+\theta>0$, then the concavity of F on $[0, \infty)$ implies

$$F(x+\theta)\geqq \frac{-2\theta}{x-\theta}F(0)+\frac{x+\theta}{x-\theta}F(x-\theta)$$

and the convexity of F on $(-\infty, 0]$ implies

$$F(\theta)\leqq \frac{-\theta}{x-\theta}F(-x+\theta)+\frac{x}{x-\theta}F(0).$$

Substituting in these inequalities

$$F(0)=\frac{F(-x+\theta)+F(x-\theta)}{2}$$

one obtains (16) again. On the other hand, it follows from (14)

$$D'_\alpha(P_0, \theta)=\int_0^\infty \psi(x)d[F(x+\theta)+F(-x+\theta)]$$

$$=2(1-F(\theta))\int_0^\infty \psi(x)dG(x),$$

where

$$G(x)=\frac{F(-x+\theta)-F(x+\theta)-2F(\theta)}{2(1-F(\theta))}.$$

It follows from (16) that G is a probability distribution function on $[0, \infty)$. Since ψ is negative on $[0, \infty)$ and $0<F(\theta)<1$, it holds

$$D'_\alpha(P_0, \theta)>0.$$

Analogically, one proves that if $\theta>0$, then

$$D'_\alpha(P_0, \theta)<0.$$

Therefore, the desired assertion is proved.

(2) Let us consider $t \in \mathbf{R}$. Since it holds $D_\alpha(P_t, \theta) = D_\alpha(P_0, \theta - t)$, it follows from the result of part (1) that it holds

$$\arg\min D_\alpha(P_t, \theta) = \{t\}.$$

It follows from here in particular that it holds

$$T^\alpha(P_\theta) = \theta, \qquad \theta \in \mathbf{R}.$$

(3) It follows from Theorem 2.1 in our paper [11] and from the results of part (2), that it holds for every $\theta \in \mathbf{R}$

$$\lim_{n \to \infty} T_n^\alpha(\mathbf{x}) = T^\alpha(P_\theta) \qquad P_\theta\text{-a.s.}$$

Applying the second result of part (2) again, we obtain the desired assertion. □

Let us define for $0 \leqq \alpha \leqq 1$ the influence curve of the estimator T^α at the probability measure $P \in \mathbf{P}$ in accordance with Huber [5] by

$$\mathrm{IC}(x, T^\alpha, P) = \lim_{\varepsilon \downarrow 0} \frac{T^\alpha((1-\varepsilon)P + \varepsilon\delta_x) - T^\alpha(P)}{\varepsilon}, \quad x \in \mathbf{R},$$

provided the limits exist.

In the next theorem we make use of the following relation

$$\int_{\mathbf{R}} \psi'(x) dP_0(x) = \int_{\mathbf{R}} \psi(x) p'(x) dx > 0, \tag{17}$$

where ψ is defined by (13) and

$$\psi'(x-\theta) = \frac{d}{d\theta}\psi(x-\theta) = 2q(x-\theta)^\alpha \frac{1 - 2\alpha \operatorname{sh}^2(x-\theta)}{\operatorname{ch}^2(x-\theta)}, \quad x, \theta \in \mathbf{R},$$

(so that $\psi'(x) = -d\psi(x)/dx$!). Relation (17) follows from the per partes integration rule and from the fact that the function $\psi(x)p'(x)$ is even and almost everywhere positive. The finiteness of the integrals in (17) follows from the fact that $\psi'(x)$ is bounded.

Theorem 3. For every $0 \leqq \alpha \leqq 1$ and $t \in \mathbf{R}$ the influence curve of T^α at P_t exists and satisfies the relation

$$\mathrm{IC}(x, T^\alpha, P_t) = \frac{\dfrac{\operatorname{sh}(x-t)}{\operatorname{ch}^{2\alpha+1}(x-t)}}{-\displaystyle\int_{\mathbf{R}} \frac{\operatorname{sh} x}{\operatorname{ch}^{2\alpha+1} x} p'(x) dx}, \qquad x \in \mathbf{R}.$$

Proof. Let us consider $x \in \mathbf{R}$ and let us denote

$$\left.\begin{aligned} P_\varepsilon^* &= (1-\varepsilon)P_t + \varepsilon\delta_x \\ D'_\alpha(P,\theta) &= \frac{d}{d\theta} D_\alpha(P,\theta) \\ D''_\alpha(P,\theta) &= \frac{d}{d\theta} D'_\alpha(P,\theta) \end{aligned}\right\} \quad P \in \mathbf{P},\ \theta \in \mathbf{R}.$$

It holds

$$\lim_{\varepsilon \downarrow 0} \frac{D'(P_\varepsilon^*, T^\alpha(P_t))}{\varepsilon} = \psi(x-t).$$

Since $D'_\alpha(P_\varepsilon^*, T^\alpha(P_\varepsilon^*)) = 0$, it follows that, for $\varepsilon \downarrow 0$,

$$D'_\alpha(P_\varepsilon^* T^\alpha(P_t)) - D'_\alpha(P_\varepsilon^*, T^\alpha(P_\varepsilon^*)) = \varepsilon\psi(x-t) + o(\varepsilon).$$

Further, the relation

$$\lim_{\varepsilon \downarrow 0} \sup_{\theta \in \mathbf{R}} |D_\alpha(P_\varepsilon^*, \vartheta) - D_\alpha(P_t, \theta)| = 0$$

implies that $T^\alpha(P_\varepsilon^* \to T^\alpha(P_t)$ as $\varepsilon \downarrow 0$.

By the mean-value theorem, the left-hand side of the relation preceding the last one equals

$$D''_\alpha(P_\varepsilon^*, \theta_\varepsilon)(T^\alpha(P_t) - T^\alpha(P_\varepsilon^*)),$$

where $\theta_\varepsilon \to T^\alpha(P_t)$ as $\varepsilon \downarrow 0$. Further, it holds

$$\lim_{\varepsilon \downarrow 0} D''_\alpha(P_\varepsilon^*, \theta_\varepsilon) = D''_\alpha(P_t, T^\alpha(P_t)) = D''_\alpha(P_t, t) \tag{18}$$

(cf. part (2) of the proof of Theorem 2), where

$$D''_\alpha(P_t, t) = \int_{\mathbf{R}} \psi'(x-t)dP_t(x) = \int_{\mathbf{R}} \psi'(x)dP_0(x) > 0 \qquad \text{(cf. (17))}. \tag{19}$$

We deduce from here that, for all sufficiently small $\varepsilon > 0$, $D''_\alpha(P_\varepsilon^*, \theta_\varepsilon) > 0$ and

$$T^\alpha(P_t) - T^\alpha(P_\varepsilon^*) = \frac{\varepsilon\psi(x-t) + \sigma(\varepsilon)}{D''_\alpha(P_\varepsilon^*, \theta_\varepsilon)}.$$

One obtains from this relation, from (18) and (19), and from the definition of $\mathrm{IC}(x, T^\alpha, P_t)$, that it holds

$$\mathrm{IC}(x, T^\alpha, P_t) = -\frac{\psi}{\int_{\mathbf{R}} \psi'(x)dP_0(x)}.$$

The desired assertion now follows from (17). □

Example 1. Let us consider the probability density $p(x)$ from the family of *hyperbolic secant* densities, i.e. let

$$p(x)=c(a)\,\mathrm{sech}^a x=\frac{c(a)}{\mathrm{ch}^a x}, \qquad a>0. \tag{20}$$

The norming constant is given by

$$c(a)=\frac{2^{1-a}\Gamma(a)}{\Gamma\left(\frac{a}{2}\right)^2},$$

where Γ is the gamma function, so that in particular

$$c(1)=\frac{1}{\pi}, \qquad c(2)=\frac{1}{2}.$$

The density (9) is from this family for $a=2$.

The denominator in Theorem 3 is equal to

$$-\int_{\mathbf{R}}\frac{\mathrm{sh}\,x}{\mathrm{ch}^{2\alpha+1}x}\cdot(-a)c(a)\frac{\mathrm{sh}\,x}{\mathrm{ch}^{a+1}x}\,dx$$

$$=ac(a)\cdot\int_{\mathbf{R}}\frac{\mathrm{sh}^2 x}{\mathrm{sh}^{2\alpha+a+2}x}\,dx.$$

Using the relation

$$\int_0^\infty\frac{\mathrm{sh}^2 x}{\mathrm{ch}^\beta x}\,dx=\frac{1}{\beta-1}\int_0^\infty\frac{1}{\mathrm{ch}^{\beta-2}x}\,dx, \qquad \beta>2,$$

which follows from the per partes integration rule, one obtains that the denominator is equal to

$$\frac{ac(a)}{(2\alpha+a+1)c(2\alpha+a)}.$$

Therefore we obtain the following explicit formulas

$$\mathrm{IC}(x, T^\alpha, P_t)=\frac{(2\alpha+a+1)c(2\alpha+a)}{ac(a)}\cdot\frac{\mathrm{sh}\,(x-t)}{\mathrm{ch}^{2\alpha+1}(x-t)},$$

$$\mathrm{IC}(x, T^0, P_t) = \frac{a+1}{a} \cdot \mathrm{th}\,(x-t),$$

$$\mathrm{IC}(x, T^{1/2}, P_t) = \frac{a+2}{2} \cdot \frac{\Gamma\left(\frac{a}{2}\right)^2}{\Gamma\left(\frac{a+1}{2}\right)^2} \cdot \frac{\mathrm{sh}\,(x-t)}{\mathrm{ch}^2\,(x-t)},$$

$$\mathrm{IC}(x, T^1, P_t) = \frac{(a+1)\,(a+3)}{a^2} \cdot \frac{\mathrm{sh}\,(x-t)}{\mathrm{ch}^3\,(x-t)}. \qquad \square$$

Theorem 4. For every $0 \leqq \alpha \leqq 1$ and $\theta \in \mathbf{R}$, the random variable $\sqrt{n}(T_n^\alpha(\mathbf{x}) - \theta)$ tends under the distribution P_θ^∞ in probability law to the normal random variable $N(0, \sigma^2)$, the variance of which is given by

$$\sigma^2 = \int_{\mathbf{R}} \mathrm{IC}(x, T^\alpha, P_0)^2 dP_0(x) \tag{21}$$

where $\mathrm{IC}(x, T^\alpha, P_0)$ is given by Theorem 3.

Proof. Let us write briefly T_n instead of $T(P_n) = T_n(\mathbf{x})$ (cf. (5)). It holds $T^\alpha(P_\theta) = \theta$ and $D'(P_n, T_n) = 0$, so that

$$\sqrt{n}[D'_\alpha(P_n, \theta) - D'_\alpha(P_n, T_n)] = n^{-1/2} \sum_{i=1}^{n} \psi(x_i - \theta) + o_p(1)$$

where $\sigma_p(1)$ tends in probability to zero. By the mean-value theorem, there exists P_θ^∞-a.s. a sequence $(\theta_n | n = 1, 2, \ldots)$ with the limit θ for which (since T_n tends P_θ^∞-a.s. to θ by Theorem 2)

$$D'_\alpha(P_n, \theta) - D'_\alpha(P_n, T_n) = D''_\alpha(P_n, \theta_n)\,(\theta - T_n).$$

Therefore, it holds

$$M_n \cdot \sqrt{n}(T_n - \theta) = n^{-1/2} \sum_{i=1}^{n} \mathrm{IC}(x_i, T^\alpha, P_\theta) + \sigma_p(1),$$

where

$$M_n = \frac{D''_\alpha(P_n, \theta_n)}{D''_\alpha(P_\theta, \theta)} \qquad \text{(cf. (19))}$$

tends in probability to 1. Hence it follows from the central limit theorem and from the Cramér–Slutskij theorem (see e.g. p. 186 in Anděl [1]) that the desired assertion holds. $\square$

Theorem 5. The variance in Theorem 4 satisfies the inequality

$$\sigma^2 \geqq \frac{1}{I(p)}, \tag{22}$$

where

$$I(p)=\int_{\mathbf{R}}\left(\frac{p'(x)}{p(x)}\right)^2 p(x)dx>0 \tag{23}$$

is the Fisher information for the family $\mathscr{P}=(P_\theta|\theta\in\mathbf{R})$ under consideration. The sign of equality takes place iff almost everywhere

$$\frac{\operatorname{sh} x}{\operatorname{ch}^{2\alpha+1}x}=\text{const}\cdot\frac{p'(x)}{p(x)}. \tag{24}$$

Proof. Let us denote the left-hand side of (24) by $\varphi(x)$. By Theorem 4,

$$\sigma^2=\frac{\int_{\mathbf{R}}\varphi(x)^2p(x)dx}{\left(\int_{\mathbf{R}}\varphi(x)p'(x)dx\right)^2} \qquad \text{(cf. (17))},$$

so that the desired inequality, as well as the condition for equality, follow from the Schwartz inequality. □

Example 2. For the family $\mathscr{P}$ considered in Example 1 we obtain, using the formula for influence curve evaluated there,

$$\sigma^2=\frac{(2\alpha+a+1)^2c(2\alpha+a)^2}{a^2c(a)^2}\int_{\mathbf{R}}\frac{\operatorname{sh}^2 x}{\operatorname{ch}^{4\alpha+2}x}\cdot\frac{c(a)}{\operatorname{ch}^a x}dx$$

$$=\frac{(2\alpha+a+1)^2c(2\alpha+a)^2}{a^2c(a)(4\alpha+a+1)c(4\alpha+a)}.$$

For $\alpha=0$, i.e. for the estimator T^0, we obtain the asymptotic variance

$$\sigma^2=\frac{a+1}{a^2}.$$

It follows from (23) that in this case

$$I(p)=\frac{a^2}{a+1}.$$

Therefore, equality (22) takes place, i.e. T^0 is asymptotically efficient, for the family $\mathscr{P}$ under consideration with arbitrary $a>0$. This conclusion can be obtained more easily from Theorem 5, since, for $\alpha=0$, the left-hand side of (24) is equal to $\operatorname{th} x$ and

the density (20) satisfies the relation

$$\frac{p'(x)}{p(x)} = a \operatorname{th} x$$

for every $a > 0$. □

References

1. *Anděl, J.* (1978), Mathematical Statistics (in Czech). SNTL, Praha.
2. *Csiszár, I.* (1963), Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. Publ. Math. Inst. Hungar. Acad. Sci. **8,** Ser. A, 85–108.
3. *Fabian, Z.* (1987), Point estimation in case of small data sets. Trans. 10th Prague Conf. on Information Theory, Academia, Praha.
4. *Huber, P. J.* (1967), The behaviour of maximum likelihood estimates under nonstandard conditions. Proc. 5th Berkeley Symp. on Math. Statist. and Probab., Vol. **1,** Calif. Univ. Press, Berkeley.
5. *Huber, P. J.* (1981), Robust Statistics. J. Wiley, New York.
6. *Kovanic, P.* (1984), Gnostical theory of small samples of real data. PCIT **13,** 309–319.
7. *Mäkaläinen, T., Schmidt, K., Styan, G. P. H.* (1981), On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples. Ann. Statist. **8,** 758–767.
8. *Pfanzagl, J.* (1969), On the measurability and consistency of minimum contrast estimators. Metrika **14,** 249–272.
9. *Vajda, I.* (1972), On the f-divergence and singularity of probability measures. Period. Math. Hungar. **2,** 223–234.
10. *Vajda, I.* (1983), A new general approach to minimum distance estimation. Trans. 9th Prague Conf. on Information Theory, 103–112. Academia, Praha.
11. *Vajda, I.* (1986), Efficiency and robustness control via distorted maximum likelihood estimation. Kybernetika **22,** 47–67.
12. *Vajda, I.* (1988), Theory of Statistical Inference and Information. Reidel, Dordrecht.
13. *Halmos, P. R.* (1964), Measure Theory. Acad. Press, New York.
14. *Kelley, J. L.* (1957), General Topology. Van Nostrand, Princeton.

Оценки с минимальным расстоянием и гностические оценки

И. ВАЙДА

(Прага)

Эта статья основана на результате Фабиана [3], который показал, что три оценки Кованица [6], которые выведены из так называемых гностических соображений, являются частным случаем оценок с минимальным расстоянием, введенных в нашей статье [10]. Используя известные свойства наших оценок, в статье доказываются статистические свойства указанных гностических оценок.

I. Vajda
Institute of Information Theory and Automation
Czechoslovak Academy of Sciences
Pod vodárenskou věží 4
182 08 Praha
Czechoslovakia

Problems of Control and Information Theory, Vol. 17 (5), pp. 267—284 (1988)

THE DISCOUNTED CONTROLLED MARKOV CHAIN WITH DELAY

A. B. PIUNOVSKI

(*Moscow*)

(Received February 24, 1987)

A controlled Markov chain with a delay in control is investigated. Necessary and sufficient conditions for the control policy to be optimal in the case of semi-continuous models are given. The τ-models' class definition is given for which methods are being developed, analogous to the dynamic programming method. Characteristic properties of the τ-models are shown; one particular example is discussed.

1. Introduction

Papers [1–5] are devoted to controlled Markov processes without delay, wherein the dynamic programming principle is proved for a broad model class, the Bellman equation is derived and necessary and sufficient conditions for the optimality of the control policy are obtained. However, very often the control on the output of the regulator has a delay. For example, the problem of remote control of moving objects; the problem of service of technical systems in which latent refusals can appear and so on. As distinct from the works mentioned, the present article investigates the points of existence and construction of optimal policy in homogeneous models with delayed control on an infinite horison. In [6, 7], analogous models with finite horizon are studied. Investigation of such models, as a rule, results in substantial problems, particularly involved in the violation of the Markov property of a process studied, even for the simplest control policies.

In Section 2 the basic definitions are provided that in the case of zero delay are reduced to the agreed-upon notions [1–5]. In particular, a definition of the Markov stationary selectors (nonrandomized policies) class is introduced, which subsequently appears sufficient to decide problems of the optimal synthesis. The present article describes the class of semi-continuous models and τ-models. For semi-continuous models, the existence of an uniformly optimal B-selector is found (Theorem 2), as well as necessary and sufficient conditions for an arbitrary control policy to be optimal (Theorem 3) are determined. General semi-continuous models do not allow the dynamic programming method to be directly applied, and an optimal B-selector may

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

be non-Markovian. Therefore, in Section 4 a new class of τ-models is introduced, wherein the synthesis problem can be decided by means of methods analogous to the dynamic programming technique. Hypotheses H_1 and H_2 play a crucial role in the description of the τ-models. The hypotheses mentioned above are based on the assumption that a "predictable" probability distribution of the current state can serve as sufficient statistic for choosing the control, and on the assumption that the optimal reward functional has an integral form (Theorem 6).

Note that the τ-model is not necessarily semi-continuous. However, if it is the case, there exists an optimal stationary Markov B-selector (Lemma 1, Theorems 4, 5) which can be built into solving the "Bellman equation" from the hypothesis H_2 formulation. As an example, the task is discussed to optimize a linear system with square criterion. It is shown that if we neglect the non-zero delay then the value of the expected profit decreases.

It should be noted that in the case of zero delay, the results obtained convert into the known facts from the theory of dynamic programming [1–5].

2. Definitions and denotations

Let a whole non-negative constant τ and arbitrary Borel spaces $(X, \mathscr{B}(X))$ — the state space, and $(A, \mathscr{B}(A))$ — the control space be given. The direct product

$$H_t = \begin{cases} A^t, & \text{at} \quad 1 \leqq t < \tau, \\ A^\tau \times X \times (A \times X)^{t-\tau}, & \text{at} \quad t \geqq \tau \end{cases}$$

is called the history space. It is evident that, with all $t = 1, 2, \ldots$, H_t, $\Omega \triangleq H_\infty$, and $H \triangleq \bigcup_{t=1}^{\infty} H_t$ are Borel spaces with σ-algebras of Borel sets $\mathscr{B}(H_t)$, $\mathscr{F} = \mathscr{B}(H_\infty)$, $\mathscr{B}(H)$, respectively.

Further, if $(E, \mathscr{B}(E))$ is a Borel space, symbols $\mathscr{A}(E)$, $\mathscr{U}(E)$ denote the analytical and universally measurable σ-algebras in E. As it is known from [1], a space of probability measures on the Borel space E is Borel; this is denoted by $P(E)$. Next, we suppose that a weak topology in $P(E)$ is fixed.

The properties of Borel spaces and of probability measures there on, also of A-functions and A-measurable and universally measurable mappings are stated in [1, 2] in detail.

Elements of the space Ω are trajectories, i.e. realizations of control and state sequences: $\omega = \{a_1, \ldots, a_\tau, \xi_0, a_{\tau+1}, \xi_1, \ldots\}$ which are written in the order of their appearance in front of the observer. Note that the state ξ_t is formed under the influence of the control a_t and is observed only after the selection of the control $a_{t+\tau}$. In the case of $\tau = 0$ we have the classical scheme [1, 2]. Further, the elements ξ_t, a_t of a

sequence $\omega \in \Omega$ are designed by $\xi_t(\omega)$, $a_t(\omega)$, i.e. they are considered as functions (projections) from Ω into X and A. Similarly $h_t(\omega) = \{a_1, \ldots, a_t, \xi_0, a_{\tau+1}, \ldots, \xi_{t-\tau}\}$ is the part of the trajectory which is observable after the selection of the control a_t. Further the argument ω, as a rule, is omitted.

Symbol $\mathscr{F}_t$ denotes a σ-algebra in Ω, the preimage of $\mathscr{B}(H_{t+\tau})$ with respect to the projection $h_{t+\tau}: \Omega \to H_{t+\tau}$. So, $\{\xi_t\}_{t \geq 0}$ are $\mathscr{F}_t$-matched random sequences. The values of parameter t are natural as a rule; however sometimes it will be convenient to denote the σ-algebra $\sigma\{a_s, S \leqq t-\theta\}$ in Ω by the symbol $\mathscr{F}_{-\theta}$ which is in agreement with the notations given earlier ($\theta = 1, 2, \ldots, \tau$); $\mathscr{F}_{-\tau} = \{\varnothing, \Omega\}$ is the trivial σ-algebra.

Suppose that the followings are given: 1) the initial distribution $P^0(\cdot)$ — a probability measure on X from some class $\tilde{P}(X) \subseteq P(X)$; 2) the one-step transition probability $P_a(x, \Gamma)$ — a probability measure on X (under arbitrary, fixed $a \in A$, $x \in X$)) $\mathscr{B}(A \times X)$ — measurable at any fixed $\Gamma \in \mathscr{B}(X)$; 3) the discount factor $\beta \in (0, 1)$, and A-function of reward $r(x, a)$.

The object $Z = \{X, A, P, r\}$ is called the model.

Definition. A sequence of universally measurable stochastic kernels $\mu_t(\Gamma | P^0, h_{t-1})$ on A is called the control policy $\pi = \{\mu_t\}_{t=1}^{\infty}$. At $t = 1$, μ_1 is a probability measure which can depend on the initial distribution $P^0(\cdot)$ only.

Theorem 1. Each policy $\pi = \{\mu_t\}_{t=1}^{\infty}$ at an arbitrary fixed $P^0 \in P(X)$ corresponds to a unique probability measure P^π on Ω, such that

$$1)\ P^\pi\{a_1 \in \Gamma_1^A, \ldots, a_{t-1} \in \Gamma_{t-1}^A, a_t \in \Gamma_t^A\} =$$

$$= \int_{\Gamma_1^A} \cdots \int_{\Gamma_{t-1}^A} \mu_t(\Gamma_t^A | P^0, a_1 \ldots a_{t-1}) \mu_{t-1}(da_{t-1} | P^0, a_1 \ldots a_{t-2}) \times$$

$$\times \ldots \times \mu_1(da_1 | P^0), \qquad 1 \leqq t \leqq \tau;$$

$$2)\ P^\pi\{a_1 \in \Gamma_1^A, \ldots, a_\tau \in \Gamma_\tau^A, \xi_0 \in \Gamma_0^X, a_{\tau+1} \in \Gamma_{\tau+1}^A, \ldots, \xi_t \in \Gamma_t^X\} =$$

$$= \int_{\Gamma_1^A} \cdots \int_{\Gamma_\tau^A} \int_{\Gamma_0^X} \int_{\Gamma_{\tau+1}^A} \cdots \int_{\Gamma_{\tau+t}^A} P_{a_t}(\xi_{t-1}, \Gamma_t^X) \mu_{\tau+t}(da_{\tau+t} | P^0, a_1 \ldots$$

$$\ldots a_\tau \xi_0 a_{\tau+1} \ldots \xi_{t-1}) \times \ldots \times \mu_{\tau+1}(da_{\tau+1} | P^0, a_1 \ldots a_\tau \xi_0) P^0(d\xi_0) \times$$

$$\times \mu_\tau(da_\tau | P^0, a_1 \ldots a_{\tau-1}) \times \ldots \times \mu_1(da_1 | P^0), \qquad t = 0, 1, 2, \ldots$$

$$3)\ P^\pi\{a_1 \in \Gamma_1^A, \ldots, a_\tau \in \Gamma_\tau^A, \xi_0 \in \Gamma_0^X, a_{\tau+1} \in \Gamma_{\tau+1}^A, \ldots, \xi_t \in \Gamma_t^X,$$

$$a_{\tau+t+1} \in \Gamma_{\tau+t+1}^A\} = \int_{\Gamma_1^A} \cdots \int_{\Gamma_\tau^A} \int_{\Gamma_0^X} \int_{\Gamma_{\tau+1}^A} \cdots \int_{\Gamma_{\tau+t}^A} \int_{\Gamma_t^X} \mu_{\tau+t+1}(\Gamma_{\tau+t+1}^A |$$

$$| P^0, a_1 \ldots a_\tau \xi_0 a_{\tau+1} \ldots \xi_t) P_{a_t}(\xi_{t-1}, d\xi_t) \mu_{\tau+t}(da_{\tau+t} | P^0, a_1 \ldots$$

$$\ldots a_\tau \xi_0 a_{\tau+1} \ldots \xi_{t-1}) \times \ldots \times \mu_{\tau+1}(da_{\tau+1} | P^0, a_1 \ldots a_\tau \xi_0) P^0(d\xi_0) \times$$

$$\times \mu_\tau(da_\tau | P^0, a_1 \ldots a_{\tau-1}) \times \ldots \times \mu_1(da_1 | P^0), \qquad t = 0, 1, 2, \ldots$$

Here $\Gamma_i^A \in \mathscr{B}(A)$, $\Gamma_i^X \in \mathscr{B}(X)$ are arbitrary sets.

The proof follows simply from the Ionesku–Tulcea theorem [1, 2].

The randomized control a_t is specified by the kernel $\mu_t(\cdot\,|P^0,\ h_{t-1})$ that is, when choosing a control a_t, only information about P^0 and values $a_1\ a_2,\ \ldots,\ a_\tau,\ \xi_0,\ a_{\tau+1},\ \xi_1,\ \ldots,\ a_{t-1},\ \xi_{t-\tau-1} (t \geqq \tau+1)$ may be employed. In other words, the control a_t is an $\mathscr{U}_{t-\tau-1}$-measurable random element, where $\mathscr{U}_t$ is an universal σ-algebra $\mathscr{F}_t$-extension. Consequently, the value τ is the control delay. With $\tau=0$, we obtain the model without delay that was studied in [1–4].

According to Theorem 1, with P^0 fixed each policy π specifies a matched random process $(a_t,\ \xi_t)$ on the stochastic basis $(\Omega,\ \mathscr{F},\ P^\pi,\ (\mathscr{F}_t)_{t\geq 0})$. The σ-algebras $\mathscr{F}_t$ are filled up with sets of zero P^π-measure. Symbol M^π denotes the P^π-measure integral.

Definition. Symbols $\pi^t = \{\mu_\theta\}_{\theta=1}^t,\ \pi_t = \{\mu_\theta\}_{\theta=t+1}^\infty$ denote the "parts" of a policy π.

It is obvious that (similarly to Theorem 1) any fixed π^t by a given $P^0 \in P(X)$ determines the unique probability measure $P_H^{\pi^t}$ on H_t, the preimage of which on Ω is denoted by P^{π^t} $(t=1,\ 2,\ \ldots)$. Let $P^0 \in P(X)$ be fixed and $\{\psi_T(\omega)\}_{T=1}^\infty$ be a sequence of random variables such that $P^\pi-$ almost surely (further: $P^\pi-$a.s.) the finite $\lim\limits_{T\to\infty} M^\pi[\psi_T(\omega)|\mathscr{F}_{t-\tau}]$ is defined for all policies π $(t=0,\ 1,\ 2,\ \ldots)$. Let π^t be fixed and π_t be variable. Then $\mathscr{F}_{t-\tau}$-measurable random variables $\lim\limits_{T\to\infty} M^\pi[\psi_T(\omega)|\mathscr{F}_{t-\tau}]$ form the set for which π_t is a parameter. So, we can introduce the operation P^{π^t} — ess sup by the set $\{\pi_t\}$. The described construction is denoted further by

$$P^{\pi^t} - \operatorname*{ess\,sup}_{\pi_t} \lim_{T\to\infty} M^{\pi_t}[\psi_T(\omega)|\mathscr{F}_{t-\tau}]$$

(we assume that the essential supremem is correctly defined). The meaning of the record $P^{\pi^t} - \operatorname*{ess\,sup}_{\pi_t} M^{\pi_t}[\psi(\omega)|\mathscr{F}_{t-\tau}]$ is similar.

Definition. With P^0 fixed, a policy π is called admissible if $\forall\, t \geqq 0\ \forall\, \theta = t+1$, $t+2,\ \ldots$ finite conditional expectations $M^\pi[r(\xi_{\theta-1},\ a_\theta)|\mathscr{F}_{t-\tau}]$ are given $P^\pi-$a.s. and there exists a finite $(P^\pi-\text{a.s.})$ $\lim\limits_{T\to\infty} M^\pi\left[\sum\limits_{\theta=t+1}^{T} \beta^\theta \cdot r(\xi_{\theta-1},\ a_\theta)|\mathscr{F}_{t-\tau}\right]$. The class of admissible policies is denoted by Π.

The value $W(\pi) \triangleq \lim\limits_{T\to\infty} M^\pi\left[\sum\limits_{t=1}^{T} \beta^t r(\xi_{t-1},\ a_t)\right]$ at a given P^0 is called the policy $\pi \in \Pi$ value. A policy $\pi^* \in \Pi$ is called optimal (for initial distribution P^0) if $W(\pi^*) = \sup\limits_{\pi\in\Pi} W(\pi)$. A policy π^* is called uniformly optimal (ε-optimal) if it is admissible for all $P^0 \in \tilde{P}(X)$ and $W(\pi^*) = \sup\limits_{\pi\in\Pi} W(\pi)$ $(W(\pi^*) > \sup\limits_{\pi\in\Pi} W(\pi) - \varepsilon)$ simultaneously at various initial distributions $P^0 \in \tilde{P}(X)$.

A policy $\pi=\{\mu_t\}_{t=1}^{\infty}$ is called a selector (nonrandomized policy), if $\forall\, t>0$ the measure μ_t is centered at the point $a_t=\varphi(P^0, h_{t-1})$. The selector is called Markov if $\varphi(P^0, h_{t-1})=\varphi(t, P_{t-1}^{\varphi})$. Here and below

$$P_{t-1}^{\pi}(\Gamma, P^0, \omega)\triangleq P^{\pi}\{\xi_{t-1}\in\Gamma|\mathscr{F}_{t-\tau-1}\},\quad \Gamma\in\mathscr{B}(X) \tag{1}$$

is the conditional probability. A Markov selector is called stationary if $\varphi(t, P_{t-1}^{\varphi})=\varphi(t\wedge(\tau+1), P_{t-1}^{\varphi})$. Here and below $a\wedge b\triangleq\min\{a, b\}$. According to the introduced denotations the conditional probability P_{t-1}^{π} is defined by the "part" π^{t-1} of a policy. So we shall use the notation $P^{\pi^{t-1}}$ sometimes.

By B_{t-1} we shall denote the family $P_{t-1}^{\pi}(\cdot, P^0, \omega)$ under diverse P^0, π, ω. Further, all the families B_t are assumed to coincide: $B_t\equiv B$ and $\tilde{P}(X)\subseteq B$; $B\in\mathscr{B}(P(X))$ and $\forall\, P\in B\ \forall\, a\in A$ there exists a finite integral $\int_X r(y, a)P(dy)$.

Definition. If a policy (selector) is given by Borel kernels (functions) then it is called B-policy (B-selector). The Markov selector φ is called A-selector if the mapping $\varphi(t,\cdot): B\to A$ is $\mathscr{A}(B)$-measurable at all $t=1, 2, \ldots$. Further denote by Π^m a set of all the admissible stationary Markov B-selectors.

3. Semicontinuous models

A model Z is called semi-continuous if the next conditions are satisfied:

C1. Function r is bounded from above- and upper-semicontinuous.

C2. The transition probability $P_a(x, \Gamma)$ defines a continuous stochastic kernel $P: A\times X\to P(X)$.

C3. Set A is a compact.

Remark. Value "$-\infty$" is permissible for the upper-semi-continuous functions.

It is worthwile for the semi-continuous model not to limit the class of admissible policies because all of the conditional expectations and limits can be correctly defined on the extended straight line by means of the standard operation of the bottom "cutting-off" the reward function.

As it is shown in Section 6, $\forall\, t\geqq 0\ \forall\, P^0$ the random value

$$P^{\pi^{t+\tau}}\text{-}\operatorname*{ess\,sup}_{\pi_{t+\tau}}\lim_{T\to\infty} M^{\pi_{t+\tau}}\left[\sum_{\theta=t+1}^{T}\beta^{\theta}\cdot r(\xi_{\theta-1}, a_{\theta})|\mathscr{F}_t\right]P^{\pi^{t+\tau}}\text{-a.s.}=V_t(\xi_t, a_{t+1}, \ldots, a_{t+\tau})$$

depends on enumerated arguments only. Complete investigation of semi-continuous models at $\tau=0$ is presented in [1, 2, 5].

As noted above, in the case of $\tau>0$ it can happen that an ε-optimal Markov selector does not exist. However, the following statements are true.

Theorem 2. In the semi-continuous model, there exists a uniformly optimal B-selector π^*, and $\forall P^0 \in \tilde{P}(X)$ $W(\pi^*) = \sup_{\pi} W(\pi) = \sup_{(a_1, \ldots, a_\tau) \in A^\tau} \int_X V_0(\xi_0, a_1, \ldots, a_\tau) P^0(d\xi_0)$.

Theorem 3. With any fixed initial distribution $P^0 \in \tilde{P}(X)$, in order that the policy $\pi = \{\mu_t\}_{t=1}^{\infty}$ be optimal in the semi-continuous model, it is necessary and sufficient the following conditions be satisfied:

$$1)\ \lim_{T \to \infty} M^\pi \left[\sum_{\theta=t+1}^{I} \beta^\theta \cdot r(\xi_{\theta-1}, a_\theta) | \mathscr{F}_t \right] P^\pi - \text{a.s.} = V_t(\xi_t, a_{t+1}, a_{t+\tau}) \quad \forall t \geqq 0.$$

$$2)\ M^\pi[V_0(\xi_0, a_1, \ldots, a_\tau)] = \sup_{(a_1, \ldots, a_\tau) \in A^\tau} \int_X V_0(\xi_0, a_1, \ldots, a_\tau) P^0(d\xi_0).$$

The proofs are given in Section 6.

4. τ-models

Let D denote the linear space of the real universally measurable functions $F(\cdot)$ on X such that for $\forall P \in B$ there exists the finite integral $\int_X F(y) P(dy)$.

Assume that the initial distribution $P^0 \in \tilde{P}(X)$ is fixed.

Let $\{F_i\}_{i=0}^{\tau}$ be a set of functions from the class D. Here we need the following conditions:

C4. a) $\forall a \in A$ $\forall x \in X$ there exists the finite integral $\int_X F_i(y) P_a(x, dy)$; $i = 0, 1, \ldots, \tau$;

b) $\forall \pi \in \Pi$ $\forall \theta = 1, 2, \ldots$, there exist the finite conditional expectations $M^\pi[F_{\theta \wedge \tau}(\xi_\theta) | \mathscr{F}_{t-\tau}]$, $t = 0, 1, \ldots, \theta$;

$$M^\pi \left[\int_X F_{\theta \wedge \tau}(y) P_{a_\theta}(\xi_{\theta-1}, dy) | \mathscr{F}_{t-\tau} \right], \quad t = 0, 1, \ldots, \theta - 1.$$

c) $\forall \pi \in \Pi$ $\forall t = 0, 1, 2, \ldots$, there exists $\lim_{T \to \infty} M^\pi[\beta^T F_\tau(\xi_T) | \mathscr{F}_{t-\tau}]$.

For example, C4 are met for bounded functions F_i.

Hypothesis H_1. There exist a policy $\pi^* \in \Pi$ and a set of functions $\{F_i(y)\}_{i=0}^{\tau} \in D$ for which conditions C4 are satisfied such that

$$1)\qquad \forall t = 1, 2, \ldots \quad M^{\pi^*} \left[r(\xi_{t-1}, a_t^*) - \frac{1}{\beta} F_{(t-1) \wedge \tau}(\xi_{t-1}) + \right.$$

$$\left. + \int_X F_{t \wedge \tau}(y) P_{a_t^*}(\xi_{t-1}, dy) | \mathscr{F}_{t-\tau-1} \right] = 0 \quad P^{\pi^*} - \text{a.s.} \qquad (2)$$

2) $$\forall \pi \in \Pi \quad M^{\pi}\Big|r(\xi_{t-1}, a_t) - \frac{1}{\beta}\cdot F_{t-1)\wedge\tau}(\xi_{t-1}) + \\ + \int_X F_{t\wedge\tau}(y) P_{a_t}(\xi_{t-1}, dy)\,\Big|\,\mathscr{F}_{t-\tau-1}\Big] \leqq 0 \quad P^{\pi}-\text{a.s.}$$

3) $$\forall \pi \in \Pi \quad \lim_{T\to\infty} M^{\pi}[\beta^T F_{\tau}(\xi_T)] \geqq \lim_{T\to\infty} M^{\pi^*}[\beta^T \cdot F_{\tau}(\xi_T)].$$

Hypothesis H_2. There exists a set of A-functions $\{F_i(y)\}_{i=0}^{\tau}$ for which conditions C4 are satisfied such that

1) $$\forall \pi \in \Pi \quad \forall t = 1, 2, \ldots, P^{\pi} - \operatorname*{ess\,sup}_{a\in A}\left\{M^{\pi}\left[r(\xi_{t-1}, a) - \frac{1}{\beta}\cdot F_{(t-1)\wedge\tau}(\xi_{t-1}) + \right.\right. \\ \left.\left. + \int_X F_{t\wedge\tau}(y) P_a(\xi_{t-1}, dy)\,|\mathscr{F}_{t-\tau-1}\right]\right\} = 0 \quad P^{\pi}-\text{a.s.} \tag{3}$$

2) $$\forall \pi \in \Pi \quad \lim_{T\to\infty} M^{\pi}[\beta^T \cdot F_{\tau}(\xi_T)] = 0.$$

Definition. Models wherein hypothesis H_2 holds, we shall call τ-models. (Analogous definition is given in [6] for models with finite horizon.)

Remark. The following two assumptions are behind hypotheses H_1, H_2:

1) when choosing control on each step t, the predictable probability distribution P_{t-1}^{π} (1) is a sufficient statistic;

2) the optimal reward functional $V_t(\cdot)$ introduced in Section 3, is expressed in a certain fashion through the measure integral $P_{t+\tau}^{\pi^{t+\tau}}$ (see Theorem 6), the integrand functions are the same as the functions F_i from the formulations of the hypotheses.

Note that equality (3) in H_2 plays the role of the Bellman equation, with which the functions F_i can be determined. In the case of zero delay, hypotheses H_1, H_2 hold, for example, for semi-continuous models with a bounded reward function; with function F_0 being in agreement with the decision of the classical Bellman equation [5]. It is natural to call F_i Bellman functions.

Lemma 1. Let the model Z be semi-continuous and $\{F_i(y)\}_{i=0}^{\tau}$ — arbitrary, bounded from above upper-semi-continuous functions on X. Then

1) $$\forall \pi \;\; \forall t > 0 \quad P^{\pi} - \operatorname*{ess\,sup}_{a\in A}\left\{M^{\pi}\left[r(\xi_{t-1}, a) - \right.\right. \\ \left.\left. - \frac{1}{\beta} F_{(t-1)\wedge\tau}(\xi_{t-1}) + \int_X F_{t\wedge\tau}(y) P_a(\xi_{t-1}, dy)|\mathscr{F}_{t-\tau-1}\right]\right\}$$

is attained under a certain control $\varphi(P_{t-1}^{\pi})$ which is (Borel) measurable, dependent on the conditional probability measure $P_{t-1}^{\pi} \in B$ only;

2) the supremum value is a bounded from above upper-semi-continuous function on B;

3) for the stationary Markov B-selector $\varphi(t, P^{\varphi}_{t-1}) \triangleq \varphi_{t \wedge (\tau+1)}(P^{\varphi}_{t-1})$ under all $t>0$, the following equalities are satisfied:

$$M^{\varphi}\left[r(\xi_{t-1}, a_t) - \frac{1}{\beta} F_{(t-1)\wedge\tau}(\xi_{t-1}) + \int_X F_{t\wedge\tau}(y) P_{a_t}(\xi_{t-1}, dy) \,|\, \mathscr{F}_{t-\tau-1} \right] P^{\varphi}\text{-a.s.} =$$

$$= P^{\varphi}\text{-ess sup}_{a\in A} \left\{ M^{\varphi}\left[r(\xi_{t-1}, a) - \frac{1}{\beta} F_{(t-1)\wedge\tau}(\xi_{t-1}) + \right.\right.$$

$$\left.\left. + \int_X F_{t\wedge\tau}(y) P_a(\xi_{t-1}, dy) \,|\, \mathscr{F}_{t-\tau-1} \right] \right\}.$$

(there exists "a measurable choice").

The proof is given in Section 7.

Note, if Lemma 1 is applicable in the τ-model, with $\varphi \in \Pi$, then hypothesis H_1 holds.

Theorem 4. If in the model Z hypothesis H_1 holds, then the policy π^* is optimal, and

$$W(\pi^*) = \int_X F_0(y) P^0(dy) - \lim_{T\to\infty} M^{\pi^*}[\beta^T \cdot F_{\tau}(\xi_T)].$$

The proof is given in Section 8.

Theorem 5. Let in the τ-model, for policy $\pi^* \in \Pi$, equality (2) be satisfied (P^{π^*}-a.s.) for $t=1, 2, \ldots$. Then the policy π^* is optimal, and $W(\pi^*) = \int_X F_0(y) P^0(dy)$.

The proof follows from Theorem 4 by noting that for set $\{F_i\}_{i=0}^{\tau}$ and policy π^* hypothesis H_1 is satisfied.

Let for the τ-model, the following conditions be satisfied:

C5. a) $\forall \pi \lim_{T\to\infty} M^{\pi}[\beta^T \cdot F_{\tau}(\xi_T)] = 0$;

b) conditions C4 are satisfied for all policies (not only for $\pi \in \Pi$);

c) for all π, $t=1, 2, \ldots$ equality (3) is satisfied (P^{π}-a.s.).

In this case Theorem 5 can be inverted: if policy π^* is optimal, then for it equality (2) is satisfied (P^{π^*}-a.s.). The proof is a from-the-converse one using Lemma 3 and the Note to it. In this case the following statements are true:

1) for each $\varepsilon>0$, there exists an ε-optimal stationary Markov A-selector;

2) if in Section 3 of Lemma 1 there exists an universally measurable (measurable) choice then there exists an optimal stationary Markov selector (B-selector).

Suppose that all the policies are admissible. For example, model Z is semi-continuous or the reward function r is uniformly bounded.

Definition. At a fixed initial distribution P^0 assume

$$\Phi_t(h_t) \triangleq P^{\pi_t} - \operatorname{ess\,sup}_{\pi_t} \lim_{T\to\infty} M^{\pi_t}\left[\sum_{\theta=t+1}^{T} \beta^\theta \cdot r(\xi_{\theta-1}, a_\theta) | \mathscr{F}_{t-\tau}\right], \quad t=0, 1, 2, \ldots$$

(for $h_t \in H_t$ the supremum is taken by the kernels $\pi_t = \{\mu_\theta\}_{\theta=t+1}^{\infty}$). Function $\Phi_t(h_t(\omega))$ is a random variable dependent on the kernels $\pi^t = \{\mu_\theta\}_{\theta=1}^{t}$.

It is not difficult to show by means of Theorem 3 that in the semi-continuous model at P^0 fixed, for the optimal policy π^* the equality is valid:

$$\Phi_{t+\tau}(h_{t+\tau}) = V_t(\xi_t, a_{t+1}, \ldots, a_{t+\tau}) -$$

$$-\beta^{t+1} \cdot r(\xi_t, a_{t+1}) - \beta^{t+2} \cdot \int_X r(y_1, a_{t+2}) P_{a_{t+1}}(\xi_t, dy_1) -$$

$$- \ldots - \beta^{t+\tau} \cdot \int_X \ldots \int_X r(y_{\tau-1}, a_{t+\tau}) P_{a_{t+\tau-1}}(y_{\tau-2}, dy_{\tau-1}) \ldots$$

$$\ldots P_{a_{t+1}}(\xi_t, dy_1) \quad P^{\pi^*}-\text{a.s.}, \quad t=0, 1, \ldots$$

Theorem 6. If the reward function r is uniformly bounded from above, the following statements are equivalent:

1) function $\Phi_t(h_t)$, at each policy π, is a functional dependent (P^{π^t} – a.s.) on the conditional measure $P_t^{\pi^t}(\cdot, P^0, \omega)$ only, with

$$\Phi_t(h_t) = \tilde{\Phi}_t(P_t^{\pi^t}) = \beta^t \int_X F_{t\wedge\tau}(z) P_t^{\pi^t}(dz, P^0, \omega)$$

(P^{π^t} – a.s.), $t=0, 1, 2, \ldots$, where $\{F_i(z)\}_{i=0}^{\tau}$ is some set of A-functions from D that meets conditions C4 such that

$$\forall \pi \lim_{T\to\infty} M^\pi[\beta^T \cdot F_\tau(\xi_T)] = 0.$$

2) Model Z is the τ-model.

The proof is cited in Section 9.

If any finite collection of measures from B is linearly independent and $\Phi_t(h_t) = \tilde{\Phi}_t(P_t^{\pi^t})$ (P^{π^t} – a.s.) is a functional on B, some sufficient conditions may be found in [6] for which the functional $\tilde{\Phi}_t$ has an integral form. Therefore, the model being investigated will be the τ-model. Note in case of $\tau=0$, the measures $P_t^{\pi^t}$ are centered at points ξ_t (see Definition (1)).

5. Example

Let $X=\mathbf{R}^1$, $A=\mathbf{R}^1$; the process ξ_t is defined by the equality $\xi_t=B\xi_{t-1}+$ $+Ca_t+D\eta_t$ where $\{\eta_t\}_{t=1}^{\infty}$ is a sequence of independent random variables with distribution density $h(z)$, the first two moments of which being equal to zero and one, respectively. Then the transition probability for one step is absolutely continuous with respect to the Lebesque measure, and the distribution density is the following:

$$p_a(y, z)=\frac{1}{|D|}h\left(\frac{z-By-Ca}{D}\right).$$

Let $r(y, a)=Ka^2+My^2$. Here B, $C\neq 0$, $D\neq 0$, $K<0$, $M<0$ are arbitrary constants.

In the optimization tasks for linear dynamic systems with square criterion (without delay) Bellman function looks like a square polynomial. In our case functions F_i from hypotheses H_1, H_2 play the role of the Bellman function. So, it is natural to seek them in the form $F_i(y)=g_iy^2+S_iy+q_i$. Let the initial distribution P^0 of the random variable ξ_0 be fixed.

After verification of hypothesis H_1 we have:

$$g_i=g=\frac{\beta}{2C^2}[MC^2+KB^2-K/\beta-\sqrt{(MC^2+KB^2-K/\beta)^2+4C^2MK/\beta}];$$

$$S_i=0;$$

$$q_\tau=\frac{D^2\beta}{1-\beta}\left[g+(M-g/\beta+gB^2)\sum_{i=0}^{\tau-1}B^{2i}\right];$$

$$q_i=\beta\left\{q_{i+1}+gD^2+(M-g/\beta+gB^2)\cdot\left[D_{\xi_0}B^{2i}+D^2\sum_{j=0}^{i-1}B^{2j}\right]\right\},$$

$$i=\tau-1, \tau-2, \ldots, 0.$$

(Here and below M_{ξ_0}, D_{ξ_0} are expectation and variance of random variable ξ_0.)

At the same time policy φ^* is constructed. It looks like following:

$$\varphi^*(P)=\alpha^*\triangleq-\frac{gBC}{K+gC^2}\int_X yP(dy). \tag{4}$$

So, in this example hypothesis H_1 is true for φ^* and $\{F_i\}_{i=0}^{\tau}$. According to Theorem 4, the stationary Markov B-policy (4) is the uniformly optimal policy, and

$$W(\varphi^*)=\sup_{\pi\in\Pi}W(\pi)=\int_X F_0(y)P^0(dy).$$

If we call the admissible only those policies for which $\lim_{T\to\infty} M^{\pi}[\beta^T \cdot \xi_T^2]=0$, then policy (4) is the admissible one, and hypothesis H_2 is true for $\{F_i\}_{i=0}^{\tau}$ i.e. we deel with the τ-model. The uniformly optimal synthesis is given by (4).

One can calculate the value $\int_X yP_{t-1}^{\varphi^*}(dy, P^0, \omega)$ using the following formula:

$$\int_X yP_{t-1}^{\varphi^*}(dy, P^0, \omega)=\begin{cases}\xi_{t-\tau-1}B^{\tau}+C\sum_{s=t-\tau}^{t-1}a_s^*B^{t-s-1}, & t>\tau;\\ M_{\xi_0}B^{t-1}+C\sum_{s=1}^{t-1}a_s^*B^{t-s-1}, & 1\leqq t\leqq\tau.\end{cases}$$

Note that, in our example, the model is not semi-continuous.

If we neglect the delay in this example then we shall get the model studied in [5]. Using the dynamic programming method one can ascertain that the control must depend on the last observed state of the process ξ_t:

$$\hat{a}_t=\hat{\varphi}(P^0, h_{t-1})=-\frac{gBC}{K+gC^2}\xi_{t-\tau-1}\quad (t\geqq\tau+1).$$

In order to close the model correctly on the initial interval $t=1, 2, \ldots, \tau$ we shall assume that $M_{\xi_0}=0$; $\hat{a}_t=0 (t\leqq\tau)$. If $\tau=0$ then policy $\hat{\varphi}$ coincides with φ^* and is optimal.

Assume that $\tau>0$ and determine the prize that gives the optimal policy (4) in comparison with the "pseudooptimal" policy $\hat{\varphi}$ in the case of $B=1$. We shall use Lemma 3 by the functions $\{F_i\}_{i=0}^{\tau}$.

One can easily verify that

$$\lim_{T\to\infty} M^{\hat{\varphi}}[\beta^T \cdot F_{\tau}(\xi_T)]=0;$$

$$M^{\hat{\varphi}}\left[r(\xi_{\theta-1}, \hat{a}_{\theta})-\frac{1}{\beta}F_{(\theta-1)\wedge\tau}(\xi_{\theta-1})+\int_X F_{\theta\wedge\tau}(y)P_{\hat{a}_{\theta}}(\xi_{\theta-1}, dy)\,|\,\mathscr{F}_{\theta-\tau-1}\right]=$$

$$=\begin{cases}0, & \theta=1, 2, \ldots, \tau;\\ \frac{g^2C^4}{K+gC^2}\left(\sum_{i=0}^{\tau-1}\hat{a}_{\theta-i-1}\right)^2, & \theta\geqq\tau+1.\end{cases}$$

Hence,

$$W(\varphi^*)-W(\hat{\varphi})=-\frac{g^4C^6}{(K+gC^2)^3}\cdot\beta^{\tau+2}\sum_{t=0}^{\infty}\beta^t\cdot M\left[\left(\sum_{i=0}^{\tau-1}\xi_{t-i}\right)^2\right],$$

where

$$\hat{\xi}_{-\tau+1}=\hat{\xi}_{-\tau+2}=\ldots=\hat{\xi}_{-1}\triangleq 0;\quad \hat{\xi}_0\triangleq\xi_0;$$

$$\hat{\xi}_t\triangleq\begin{cases}\hat{\xi}_{t-1}+D\eta_t, & 1\leqq t\leqq\tau;\\ \hat{\xi}_{t-1}-\dfrac{gC^2}{K+gC^2}\hat{\xi}_{t-\tau-1}+D\eta_t, & t>\tau;\end{cases}$$

M is the expectation relative to the measure corresponding to the stochastic process $\hat{\xi}_t$.

$$M\left[\left(\sum_{i=0}^{\tau-1}\hat{\xi}_{t-i}\right)^2\right]=(t+1)^2D_{\xi_0}+D^2\times$$

$$\times\frac{t(t+1)(2t+1)}{6}\quad\text{for}\quad t=0,1,\ldots,\tau-1.$$

Remember that constants K, g are negative; so $W(\varphi^*)>W(\hat{\varphi})$.

6. Proof of Theorems 2, 3

Let us consider the model $\hat{Z}$ without delay (with the same discount factor $\beta\in(0,1)$) which is defined by the following elements:

$$\hat{X}\triangleq A^\tau\times X;\quad \hat{A}=A;\quad \hat{P}_{\hat{a}}(\hat{x},\hat{\Gamma})\triangleq P_{\hat{a}^1}(x,\Gamma_{(\hat{a}^2,\ldots,\hat{a}^\tau,\hat{a})});\quad \hat{r}(\hat{x},\hat{a})\triangleq r(x,\hat{a}^1).$$

Here and further

$$\hat{x}=(\hat{a}^1,\ldots,\hat{a}^\tau,x)\in\hat{X};\quad \hat{\Gamma}\in\mathscr{B}(\hat{X});\quad \Gamma_{y_1}\triangleq\{y_2\in Y_2\,|\,(y_1,y_2)\in\Gamma\}$$

is the section of the set $\Gamma\subseteq Y_1\times Y_2$. It is obvious that the model $\hat{Z}$ is semi-continuous if the initial model Z is semi-continuous.

The concepts of history $\hat{h}_t\in\hat{H}_t$, policy $\hat{\pi}$ and σ-algebra $\mathscr{F}_t$ in the space $\hat{\Omega}\triangleq\hat{H}_\infty$ are introduced by analogy of Section 2. (See also [1, 2].) Every policy $\hat{\pi}$ gives the unique probability measure $\hat{P}^{\hat{\pi}}$ on $\hat{\Omega}$ if some initial distribution $\hat{P}^0$ is fixed.

Definition. Let $\{\mu_t\}_{t=1}^\tau$ be a sequence of universally measurable stochastic kernels $\mu_t(\Gamma\,|\,P^0,h_{t-1})$ on $A(\mu_1:B\to P(A))$. The symbol $\varphi(P^0,\{\mu_t\}_{t=1}^\tau)$ will denote the probability measure $\hat{P}^0$ on $\hat{X}$ which is defined by the following values on measurable rectangles:

$$\hat{P}^0\{\hat{a}^1\in\Gamma_1^{\hat{A}},\ldots,\hat{a}^\tau\in\Gamma_\tau^{\hat{A}},\xi\in\Gamma^X\}\triangleq P^0(\Gamma^X)\int_{\Gamma_1^{\hat{A}}}\ldots \tag{5}$$

$$\ldots\int_{\Gamma_{\tau-1}^{\hat{A}}}\mu_\tau(\Gamma_\tau^{\hat{A}}\,|\,P^0,\hat{a}^1\ldots\hat{a}^{\tau-1})\mu_{\tau-1}(d\hat{a}^{\tau-1}\,|\,P^0,\hat{a}^1\ldots\hat{a}^{\tau-2})\ldots\mu_1(d\hat{a}^1\,|\,P^0).$$

Definition. Let $\{\mu_t\}_{t=\tau+1}^{\infty}$ be a sequence of universally measurable stochastic kernels $\mu_t(\Gamma | P^0, h_{t-1})$ on A. The symbol $\gamma(P^0, \{\mu_t\}_{t=\tau+1}^{\infty})$ will denote the policy $\hat{\pi}=\{\nu_t\}_{t=1}^{\infty}$ in model $\hat{Z}$ which is given (for each $P^0 \in B$) by the following equality:

$$\nu_t(\Gamma | \hat{\xi}_0 \hat{a}_1 \hat{\xi}_1 \dots \hat{a}_{t-1} \hat{\xi}_{t-1}) \triangleq$$
$$\triangleq \mu_{\tau+t}(\Gamma | P^0, \hat{a}_0^1 \hat{a}_0^2 \dots \hat{a}_0^\tau \xi_0 \hat{a}_1 \xi_1 \dots \hat{a}_{t-1} \xi_{t-1}). \tag{6}$$

Here and further $\hat{\xi}_i=(\hat{a}_i^1, \dots, \hat{a}_i^\tau, \xi_i)$.

Definition. The symbol $\hat{\Omega}' \subset \hat{\Omega}$ denotes the class of sequences $\hat{\xi}_0 \hat{a}_1 \hat{\xi}_1 \dots$, for which the elements $\hat{\xi}_t$, $\hat{a}_{t+1}$, $\hat{\xi}_{t+1}$ satisfy the following equalities:

$$\hat{a}_{t+1}^1 = \hat{a}_t^2; \quad \hat{a}_{t+1}^2 = \hat{a}_t^3; \dots;$$
$$\hat{a}_{t+1}^{\tau-1} = \hat{a}_t^\tau; \hat{a}_{t+1}^\tau = \hat{a}_{t+1} \quad (t=0, 1, 2, \dots).$$

The symbol ψ designates the one-to-one mapping from Ω into $\hat{\Omega}'$ which is defined by the formula:

$$\psi(a_1 a_2 \dots a_\tau \xi_0 a_{\tau+1} \xi_1 \dots) =$$
$$= \{(a_1 a_2 \dots a_\tau \xi_0) a_{\tau+1} (a_2 \dots a_\tau a_{\tau+1} \xi_1) \dots\}.$$

One can easily verify that measure $\hat{P}^{\hat{\pi}}$ is concentrated on the set $\hat{\Omega}'$ for every policy $\hat{\pi}$ with each initial distribution $\hat{P}^0 \in P(\hat{X})$.

Lemma 2. Let $P^0 \in \tilde{P}(X)$; $\pi=\{\mu_t\}_{t=1}^{\infty}$ be a policy in model Z;

$$\hat{P}^0 = \varphi(P^0, \{\mu_t\}_{t=1}^{\tau}); \quad \hat{\pi} = \gamma(P^0, \{\mu_t\}_{t=\tau+1}^{\infty}).$$

Then

1) the measure $\hat{P}^{\hat{\pi}}$ corresponding to $\hat{P}^0$, $\hat{\pi}$ and the measure P^π corresponding to P^0, π have the property $\hat{P}^{\hat{\pi}}(\hat{\Gamma}') = P^\pi(\psi^{-1}(\hat{\Gamma}'))$;

2)
$$M^\pi\left[\sum_{\theta=t+1}^{T} \beta^\theta \cdot r(\xi_{\theta-1}, a_\theta) | \psi^{-1}(\hat{\mathscr{F}}_t)\right] =$$
$$= M^{\hat{\pi}}\left[\sum_{\theta=t+1}^{I} \beta^\theta \cdot \hat{r}(\hat{\xi}_{\theta-1}, \hat{a}_\theta) | \hat{\mathscr{F}}_t\right] \quad \hat{P}^{\hat{\pi}}\text{-a.s., } t=0, 1, 2, \dots$$

Remark. In Section 2 $\psi^{-1}(\hat{\mathscr{F}}_t)=\mathscr{F}_t$ and the conditional mathematical expectation coincidence means the coincidence of two functions on Ω and on $\hat{\Omega}'$ after the substitution $\omega=\psi^{-1}(\hat{\omega}')$.

The proof follows from the uniqueness of the measure $\hat{P}^{\hat{\pi}}$ and from the properties of stochastic measure images [8].

The function

$$\hat{V}_t(\hat{h}_t) \triangleq P^{\hat{\pi}^t} - \operatorname{ess\,sup}_{\hat{\pi}_t} \lim_{T \to \infty} M^{\hat{\pi}}\left[\sum_{\theta=t+1}^{T} \beta^\theta \times \hat{r}(\hat{\xi}_{\theta-1}, \hat{a}_\theta) | \hat{\mathscr{F}}_t\right]$$

is called "the model's $\hat{Z}$ value". (The definition of operation $\hat{P}^{\hat{\pi}_t}$ — ess sup is similar to that presented in Section 2). The supremum is taken by kernels $\hat{\pi}_t = \{\hat{\nu}_\theta\}_{\theta=t+1}^{\infty}$. It is well known [1] that in a semi-continuous model the function $\hat{V}_t(h_t)$ depends only on t, $\hat{\xi}_t$ and is upper-semi-continuous and bounded from above. Besides, there exists an optimal stationary Markov policy $\hat{\varphi}^*(\hat{\xi}_{t-1})$ for which $\forall\, \hat{P}^0\ \forall\, t \geqq 0$;

$$\hat{V}(\hat{\xi}_t)\hat{P}^{\hat{\varphi}^*} - \text{a.s.} = \lim_{T\to\infty} M^{\hat{\varphi}^*}\left[\sum_{\theta=t+1}^{T} \beta^\theta \cdot \hat{r}(\hat{\xi}_{\theta-1}, \hat{a}_\theta) | \hat{\mathscr{F}}_t\right];\ \int_{\hat{X}} \hat{V}_0(\hat{\xi}_0) \times$$

$$\times\, \hat{P}^0(d\hat{\xi}_0) = \sup_{\hat{\pi}} \lim_{T\to\infty} M^{\hat{\pi}}\left[\sum_{\theta=1}^{T} \beta^\theta \cdot \hat{r}(\hat{\xi}_{\theta-1}, \hat{a}_\theta)\right].$$

It follows from Section 2) of Lemma 2 that

$$P^{\pi^{t+\tau}} - \operatorname*{ess\,sup}_{\pi_{t+\tau}} \lim_{T\to\infty} M^{\pi_{t+\tau}}\left[\sum_{\theta=t+1}^{T} \beta^\theta \times\right.$$

$$\left.\times\, r(\xi_{\theta-1}, a_\theta) | \mathscr{F}_t\right] P^{\pi^{t+\tau}} - \text{a.s.} = \hat{V}_t(\{\xi_{t+1}, \ldots, a_{t+\tau}, \xi_t\}). \quad (\forall\, P^0).$$

Proof of Theorem 2. Even in general case one may hold that the optimal stationary Markov policy in model $\hat{Z}$ has the form $\hat{\varphi}^* = \gamma(P^0, \varphi^*)$ where $\varphi^*(h_{t-1})$ is "the piece" of some deterministic B-policy in model Z $(t > \tau)$ which is independent on P^0. Function

$$\Phi(a_1, \ldots, a_\tau, P^0) \triangleq \int_X V_0(\xi_0, a_1, \ldots, a_\tau) P^0(d\xi_0)$$

is upper-semi-continuous and bounded from above [1]. Hence according to the Yankov–von Neumann lemma [1, 2] there exist Borel functions $\varphi_1^*(P^0), \ldots, \varphi_\tau^*(P^0)$ such that

$$\Phi(\varphi_1^*(P^0), \ldots, \varphi_\tau^*(P^0), P^0) = \sup_{(a_1,\ldots,a_\tau)\in A^\tau} \Phi(a_1, \ldots, a_\tau, P^0).$$

One can easily verify that the deterministic B-policy $\pi^* = \{\varphi_1^*, \ldots, \varphi_\tau^*, \varphi^*\}$ in the model Z is uniformly optimal according to the definition, and

$$W(\pi^*) = \sup_{(a_1,\ldots,a_\tau)\in A^\tau} \int_X V_0(\xi_0, a_1, \ldots, a_\tau) P^0(d\xi_0).$$

Theorem 2 is proved.

Proof of Theorem 3. The sufficiency follows from Theorem 2 and from the obvious equality:

$$W(\pi) = \sup_{(a_1,\ldots,a_\tau)\in A^\tau} \int_X V_0(\xi_0, a_1, \ldots, a_\tau) P^0(d\xi_0).$$

Let one of conditions 1), 2) be not satisfied. Then one can prove the inequality $W(\pi) < \sup_{(a_1,\ldots,a_\tau)\in A^\tau} \int_X V_0(\xi_0, a_1, \ldots, a_\tau) P^0(d\xi_0)$ using Lemma 2. So, according to Theorem 2, the policy π is not optimal.

7. Proof of Lemma 1

Let us fix an arbitrary $t (0 < t \leqq \tau)$.

The function $f_t(y, a) \triangleq r(y, a) + \int_X F_t(z) P_a(y, dz)$ is upper-semi-continuous and bounded from above [1]. Let us introduce mappings $\psi_1 : B \times A \to P(X \times A)$ and $\psi_2 : P(X \times A) \to \mathbf{R}^1$ by the following formulae: $\psi_1(P, a) \triangleq P(\cdot)\delta_a(\cdot)$ where $\delta_a(\cdot)$ is the stochastic measure concentrated in the point a; $\psi_2(P) \triangleq \int_{X\times A} f_t(y, a) P(d(y, a))$. Mapping ψ_1 is continuous; ψ_2 is upper-semi-continuous function bounded from above [1]. Hence, the function

$$\Phi_t(P, a) \triangleq \int_X f_t(y, a) P(dy) = \psi_2(\psi_1(P, a)) : B \times A \to \mathbf{R}^1$$

is upper-semi-continuous and bounded from above. Statements 1), 2) now follow from the Yankov–von Neumann's lemma [1, 2]. The truth of Section 3) follows from the equality $\Phi_t(P, \varphi_t(P)) = \sup_{a\in A} \Phi_t(P, a)$ which is true for all $P \in B$.

8. Proof of Theorem 4

Lemma 3. Let the initial distribution P^0 be fixed and a set $\{F_i\}_{i=0}^{\tau}$ of functions from class D satisfy conditions C4. Then for all $\pi \in \Pi$ the following equality is true:

$$\lim_{T\to\infty} M^\pi \left[\sum_{\theta=t+1}^{T} \beta^\theta \cdot r(\xi_{\theta-1}, a_\theta) | \mathscr{F}_{t-\tau} \right] P^\pi - \text{a.s.} = \beta^t \cdot \int_X F_{t\wedge\tau}(z) P_t^\pi(dz, P^0, \omega) +$$

$$+ \lim_{T\to\infty} M^\pi \left[\sum_{\theta=t+1}^{T} \beta^\theta \cdot M^\pi \left[r(_{\theta-1}, a_\theta) - \frac{1}{\beta} F_{(\theta-1)\wedge\tau}(\xi_{\theta-1}) + \int_X F_{\theta\wedge\tau}(y) \times \right. \right. \tag{7}$$

$$\left. \left. \times P_{a_\theta}(\xi_{\theta-1}, dy) | \mathscr{F}_{\theta-\tau-1} \right] | \mathscr{F}_{t-\tau} \right] - \lim_{T\to\infty} M^\pi[\beta^T \cdot F_\tau(\xi_T) | \mathscr{F}_{t-\tau}], \quad t = 0, 1, 2, \ldots$$

Proof. Let us consider the model on the finite horizon $\{0, 1, \ldots T\}$ without final reward in which $P_a(t, x, dy) \triangleq P_a(x, dy)$; $R(t, x, a) = \beta^t \cdot r(x, a)$. According to [6]

it follows from conditions C4-a, b) that $\forall\, \pi \in \Pi$

$$M^{\pi}\left[\sum_{\theta=t+1}^{T} \beta^{\theta} \cdot r(_{\theta-1}, a_{\theta}) | \mathscr{F}_{t-\tau}\right] P^{\pi} - \text{a.s.} = \beta^{t} \cdot \int_{X} F'_{t}(z) P^{\pi}_{t}(dz, P^{0}, \omega) +$$

$$+ M^{\pi}\left[\sum_{\theta=t+1}^{T} M^{\pi}[\beta^{\theta} \cdot r(\xi_{\theta-1}, a_{\theta}) - \beta^{\theta-1} \cdot F'_{\theta-1}(\xi_{\theta-1}) + \beta^{\theta} \cdot \int_{X} F'_{\theta}(y) \times\right.$$

$$\left.\times P_{a_{\theta}}(\xi_{\theta-1}, dy) | \mathscr{F}_{\theta-\tau-1}] - \beta^{T} \cdot F'_{T}(\xi_{T}) | \mathscr{F}_{t-\tau}\right], \quad t = 0, 1, \ldots, T.$$

Equality (7) is obtained by the limit transition using condition C4-c). Here $F'_{t}(y) \triangleq$ $\triangleq F_{t \wedge \tau}(y)$.

Remark. If conditions C4 are satisfied for all policies (not only for $\pi \in \Pi$), then the statement of Lemma 3 can be reversed in the following sense; if there exist finite conditional mathematical expectations and limits in the right part of (7) then $\pi \in \Pi$. For example, this statement is true for bounded functions F_i.

Proof of Theorem 4. According to Lemma 3 and the definition of policy value we have:

$$W(\pi^{*}) = \int_{X} F_{0}(y) P^{0}(dy) - \lim_{T \to \infty} M^{\pi^{*}}[\beta^{T} \cdot F_{\tau}(\xi_{T})].$$

On the other hand, according to (7), $\forall\, \pi \in \Pi$

$$W(\pi) = \int_{X} F_{0}(y) P^{0}(dy) + \lim_{T \to \infty} M^{\pi}\left[\sum_{\theta=1}^{T} \beta^{\theta} \cdot M^{\pi}\left[r(\xi_{\theta-1}, a_{\theta}) - \right.\right.$$

$$\left.\left. - \frac{1}{\beta} F_{(\theta-1) \wedge \tau}(\xi_{\theta-1}) + \int_{X} F_{\theta \wedge \tau}(y) P_{a_{\theta}}(\xi_{\theta-1}, dy) | \mathscr{F}_{\theta-\tau-1}\right]\right] -$$

$$- \lim_{T \to \infty} M^{\pi}[\beta^{T} \cdot F_{\tau}(\xi_{T})] \leqq \int_{X} F_{0}(y) P^{0}(dy) -$$

$$- \lim_{T \to \infty} M^{\pi^{*}}[\beta^{T} \cdot F_{\tau}(\xi_{T})] = W(\pi^{*}), \text{ q.e.d.}$$

9. Proof of Theorem 6

Note that in our case all policies are admissible.

Let statement 1) be true. We shall prove that hypothesis H_2 is true for the functions $\{F_i\}_{i=0}^{\tau}$. Really, $\forall\, \pi\ \forall\, t = 1, 2, \ldots$

$$\frac{1}{\beta} \int_{X} F_{(t-1) \wedge \tau}(z) P^{\pi^{t-1}}_{t-1}(dz, P^{0}, \omega) = \frac{1}{\beta^{t}} \tilde{\Phi}_{t-1}(P^{\pi^{t-1}}_{t-1}) =$$

$$= \frac{1}{\beta^t} P^{\pi^{t-1}} - \operatorname*{ess\,sup}_{\pi_{t-1}} \lim_{T\to\infty} M^{\pi_{t-1}} \left[\sum_{\theta=1}^{T} \beta^\theta \cdot r(\xi_{\theta-1}, a_\theta) | \mathscr{F}_{t-\tau-1} \right] =$$

$$= P^{\pi^{t-1}} - \operatorname*{ess\,sup}_{\pi_{t-1}} \left\{ M^{\pi_{t-1}}[r(\xi_{t-1}, a_t) | \mathscr{F}_{t-\tau-1}] + M^{\pi_{t-1}} \left[\frac{1}{\beta^t} \times \right. \right.$$

$$\left. \left. \times \lim_{T\to\infty} M^{\pi_t} \left[\sum_{\theta=t+1}^{T} \beta^\theta \cdot r(\xi_{\theta-1}, a_\theta) | \mathscr{F}_{t-\tau} \right] | \mathscr{F}_{t-\tau-1} \right] \right\} =$$

$$= P^{\pi^{t-1}} - \operatorname*{ess\,sup}_{a_t \in A} \left\{ M^{\pi}[r(\xi_{t-1}, a_t) | \mathscr{F}_{t-\tau-1}] + M^{\pi} \left[\frac{1}{\beta^t} \tilde{\Phi}_t(P_t^{<\pi^{t-1}, a_t>}) | \mathscr{F}_{t-\tau-1} \right] \right\} =$$

$$= P^{\pi} - \operatorname*{ess\,sup}_{a \in A} \{ M^{\pi}[r(\xi_{t-1}, a) | \mathscr{F}_{t-\tau-1}] + M^{\pi} \left[\int_X F_{t\wedge\tau}(z) P_a(\xi_{t-1}, dz) | \mathscr{F}_{t-\tau-1} \right] \}.$$

(All equalities are true P^π – a.s.)

Let statement 2) be true. According to Lemma 3

$$\lim_{T\to\infty} M^{\pi} \left[\sum_{\theta=t+1}^{I} \beta^\theta \cdot r(\xi_{\theta-1}, a_\theta) | \mathscr{F}_{t-\tau} \right] P^{\pi^t} - \text{a.s.} \leqq \beta^t \times \int_X F_{t\wedge\tau}(z) P_t^{\pi^t}(dz, P^0, h_t).$$

On the other hand it follows from the Yankov–von Neumann lemma that there exists a stationary Markov A-policy φ such that

$$\int_X \left[r(y, \varphi(\theta, P)) - \frac{1}{\beta} F_{(\theta-1)\wedge\tau}(y) + \int_X F_{\theta\wedge\tau}(z) \times P_{\varphi(\theta,P)}(y, dz) \right] P(dy) > -\varepsilon$$

(for $P^{\varphi\theta^{-1}}$ — almost every measure $P \in B$ and for each moment $\theta > 0$). Hence, according to Lemma 3,

$$P^{\pi^t} - \operatorname*{ess\,sup}_{\pi_t} \lim_{T\to\infty} M^{\pi_t} \left[\sum_{\theta=t+1}^{T} \beta^\theta \cdot r(\xi_{\theta-1}, a_\theta) | \mathscr{F}_{t-\tau} \right] \geqq$$

$$\geqq \beta^t \cdot \int_X F_{t\wedge\tau}(z) P_t^{\pi^t}(dz, P^0, h_t) = \tilde{\Phi}_t(P_t^{\pi^t}) \qquad P^{\pi^t} - \text{a.s.}$$

The theorem is proved.

References

1. *Bertsekas, D. P., Shreve, S. F.*, Stochastic optimal control. N.Y.–S. Francisco–London, Academic Press, 1978; Russian transl.: Moscow, Nauka, 1985.
2. *Juškevič, A. A., Chitashvili, R. Ya.*, Controlled stochastic sequences and Markov chains. Uspekhi mathem. nauk, 1982, **37**, 6, pp. 213–242.
3. *Lipčer, R. Sh., Shiryaev, A. N.*, Statistics of stochastic processes. Moscow, Nauka, 1974; English transl.: Springer-Verlag, 1978.

4. *Mine, H., Osaki, S.*, Markovian decision processes. N.Y., 1970; Russian transl.: Moscow, Nauka, 1977.
5. *Dynkin, E. B., Juškevič, A. A.*, The controlled Markov processes and their applications. Moscow, Nauka, 1975; English transl.: Springer-Verlag, 1979.
6. *Aliseenko, O. V., Khametov, V. M., Piunovski, A. B.*, The optimal control of stochastic sequences with delay. Problems of Control and Information. Theory, 1986, **15**, *1*, pp. 3–17.
7. *Vasuck, B. A., Shternberg, A. A.*, The structural properties of optimal linear stochastic systems with control's delay. Avtomatika i telemekhanika, 1983, *3*, pp. 65–75.
8. *Meyer, P. A.*, Probability and potentials. Toronto–London, 1966; Russian transl.: Moscow, Mir, 1973.

Дисконтированная управляемая цепь Маркова с запаздывающим управлением

А. Б. ПИУНОВСКИЙ

(Москва)

В статье исследуется управляемая цепь Маркова с запаздыванием по управлению, заданная на бесконечном интервале времени, при наличии дисконтирующего фактора. В подобных моделях марковское свойство может нарушаться даже для простейших стратегий управления, а метод динамического программирования в общем случае неприменим. Тем не менее, для полунепрерывных моделей доказана достаточность множества борелевских селекторов и получены необходимые и достаточные условия оптимальности стратегии управления.

Ряд более простых достаточных условий оптимальности приводится для τ-моделей, метод исследования которых аналогичен методу динамического программирования. В статье приводятся характеристические свойства и условия существования τ-моделей.

В качестве примера рассмотрена задача оптимального управления линейной системой с квадратичным критерием, для которой построен точный оптимальный синтез.

А. Б. Пиуновский
Институт физико-технических проблем
СССР, Москва 119034, Кропоткинская 13/7.

Problems of Control and Information Theory, Vol. 17 (5), pp. 285—295 (1988)

EXPERT SYSTEMS WITH NON-NUMERICAL BELIEF FUNCTIONS

I. KRAMOSIL

(Prague)

(Received September 4, 1987)

Some argumentation is presented in favour of the idea that the assumption concerning the numerical character of values of belief functions or weights in expert systems may be sometimes too restrictive and that also objects of another nature may be used for this purpose. Rule-based expert systems with rules equipped by weights taking their values in the space of subsets of a fixed set are discussed in more detail together with elementary combination and decision meta-rules. Connections with the "possible worlds" semantics for modal logics are also mentioned.

1. Introduction and motivation

Expert systems seem to belong to the most popular notions of contemporary theoretical and practical artificial intelligence and we believe this fact to make unnecessary any explanations of what expert systems may be as well as any apologization for our dealing with this subject. Just for the sake of this introduction, an expert system may be understood as consisting of two parts:

(a) A *data base*, i.e. a collection of formulas, very often written in the implicative form $A_1 \,\&\, A_2 \,\&\, \ldots \,\&\, A_n \rightarrow C(w)$, each formula being labeled by a weight w, as a rule, a real number from $\langle 0, 1 \rangle$. Every formula expresses some dependence between the validity of observable predicates $A_1, A_2, \ldots, A_n$ (e.g. symptoms or syndroms), and the validity of a hypothesis C (e.g. medical or technical diagnosis), and the corresponding weight quantifies the degree of validity of this implication or the degree of belief in its validity (very often w is understood as the probability, either in an objective or statistical sense, or as a subjective or personal probability).

(b) A *deduction mechanism* which, given a formula of the form as above, deduces the weight with which this implication is supported by the given data base. In more sophisticated cases, given some observations $A_1, A_2, \ldots, A_n$, the expert system finds the hypothesis C, for which the weight ascribed to $A_1 \,\&\, A_2 \,\&\, \ldots \, A_n \rightarrow C$ is the maximum one.

Considering the way in which this deduction device computes the weight of the tested formula, expert systems can be classified as *extensional* or *intensional* ones.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

In the extensional case, the weight of the tested formula is uniquely determined by the *weights* of those formulas in the data base, which the deduction mechanism takes as relevant to the tested formula. In other words, there is an a priori given arithmetic expression which operates over the weights of the relevant formulas as its arguments in order to obtain the resulting weight, but this arithmetic expression is independent of the relevant formulas themselves.

In the *intensional* expert systems the way of how the weights are combined depends on the data base as a whole, including not only the weights but even the particular forms of all the formulas contained inside. Since a few years quarrels between the defenders of extensional expert systems and the "intensionalists" have been heard at almost every seminar, lecture, or conference on expert systems. The argumentation of the first group of specialists is based, after all, on the proclaimed computational simplicity and relatively good practical results reached by extensional expert systems. On the other hand, supposing that an uncertainty, which can be described by the tools of the classical (Kolmogorov, set-theoretic) probability theory, is of ontological nature, the extensional approach is equivalent to certain rather strong and hard-to-accept assumptions concerning the kind of statistical dependences among the observational variables and hypotheses. In such a case of "ontological probabilistic indeterminism" the probabilistic version of intensional expert system, as founded and developed by *A.* Perez and some others [7], seems to be more adequate, as it preserves all the powers and flexibilities offered by the probability theory in order to describe and handle such structures.

It is beyond the intended scope of this paper to continue on this polemic discussion at this level. We should rather put into question another aspect of the expert systems which is accepted, so far, by both groups or schools mentioned above. Namely, we shall discuss the fact that the weights or degrees of belief ascribed to formulas in expert systems are supposed to be of *numerical* nature and we shall consider some possibilities of non-numerical belief functions. When returning to the problem "extensionality vs. intensionality" and considering it from the new, "non-numerical" viewpoint, this question can be seen in quite a new light and it seems to be worth a more detailed investigation.

2. Lights and shadows of numerical belief functions

The space of real numbers together with all the arithmetical operations defined on this space represents the oldest and one of the richest mathematical structures; the greatest part of other structures have been obtained from this one by appropriate abstractions and generalizations. Therefore, a mapping of an extra-mathematical structure, e.g. that of the degree of beliefs of experts, into reals is considered to be

the most perfect way of mathematization. As a matter of fact, both the most developed mathematical tools for describing and handling with uncertainties, the classical probability theory and the theory of fuzzy sets, despite of many rather deep philosophical differences, agree with each other in taking the most richest structure over reals, the unit interval, as the domain of values for their quantitative characteristics. But this approach is not free of some negative features which are worth mentioning in our context.

First, real numbers are considered as strictly separable and mutually exclusive objects so that, e.g., ascribing a value $y = f(x)$ to a function f and argument value x, we must also accept that $f(x) \neq y \pm \varepsilon$, no matter how small $\varepsilon > 0$ may be. This seemingly trivial fact, when applied to expert systems, forces an expert to accept, that his degree of belief (for an implication, say), is neither 0.65 nor 0.75 supposing he declared it to be 0.70. Moreover, in a particular situation, his or her declaring the degree of belief to be 0.70 may involve a decision quite different from those resulting when he or she said this degree to be 0.65 or 0.75. However, in the greatest part of practical problems the three assertions ("degree of belief is 0.65", "d. of b. is 0.70", "d. of b. is 0.75") are usually not understood as mutually fully exclusive and such an interpretation is often felt as very unnatural by the specialists of the domain in question (physicians, say) supposing they are not "intoxicated" too much by mathematics.

Second, the great variety of operations over real numbers, which enables to discover and prove many properties of numerical images of an extra-mathematical structure, involves a strong temptation to "ontologize" these operations and their results, i.e. to consider them also as "images" of relations and properties holding in the extra-mathematical structure in question and discovered thanks to mathematics. This approach may even lead to some dangerous illusions concerning the discovering powers of mathematics when applied to real world. For example, the obvious fact that each non-empty set of reals in the unit interval possesses also the supremum value does not imply that an operation like taking supremum is of sense or of use in the space the elements of which are investigated by projecting them into the unit interval.

Probability theory is rather conservative in its keeping the unit interval as the space of probability values, because of strong theoretical and practical connections between probability values and empirical relative frequencies, for which the numerical presentation is quite natural. However, the argumentation introduced above has been presented, in more details and in a more convincing form, in papers dealing with philosophical and methodological backgrounds of fuzzy sets, and has led to the notion of the so called *L-fuzzy sets*. In fact, an L-fuzzy set is a mapping from the basic space (universum of discourse) into an *abstract* set equipped by an appropriate structure. Historically, the first structure considered in this role was the *lattice*, hence L-fuzzy sets (cf. [2]). Whether this is just the optimal set of values for characteristic functions

of fuzzy sets, may be and should be a matter of discussion, but in every case this or other choice for the set of values must be justified somehow, and it is not "obvious" or "automatic" as in the case of real numbers in general or of the unit interval in particular.

In the rest of this paper we shall try to apply the same idea to rule-based expert systems, i.e. we shall admit that the degrees of belief or weights ascribed to implicative formulas are not reals, but elements of other appropriately structured abstract spaces. An extremal possibility would be to take, as the set of weight values, the set of formulas of the language in question (or the Lindenbaum–Tarski algebra over this language) and to identify the weight ascribed to a formula with this formula itself. This approach, which reduces combination of weights to logical deduction, is perhaps worth a more detailed investigation, but in this paper, also because of its limited scope, we shall limit ourselves to another particular case, when the values of weights are *subsets of a non-empty basic (or universal) set* of "worlds" equipped by all the usual set-theoretic operations. The relative similarity of this model with the probabilistic one goes so far that random events are also described by (or identified with) certain subsets of a basic space, however, in what follows, these subsets are not supposed to be endowed by their measures or by some numerical quantitative characteristics at all. Certain analogies with the Kripke semantics of modal logics based on the "possible worlds" should become more clear and will be discussed in more detail later.

3. Rule-based expert systems with set-valued degrees of beliefs or weights

Let us approach to a more detailed description of the model mentioned above with a possible interpretation and application given below.

Reference structure $\mathscr{R}$ is a triplet $\langle \Omega, \mathscr{I}, \mathscr{N} \rangle$ or $\langle \Omega, \mathscr{I}(\mathscr{R}), \mathscr{N}(\mathscr{R}) \rangle$, if the dependence of $\mathscr{I}$ and $\mathscr{N}$ on $\mathscr{R}$ is to be explicitly stated, where

Ω is a non-empty set,

$\mathscr{I} \subset \mathscr{P}(\Omega)$ is a system of *important* (or large) subsets of Ω,

$\mathscr{N} \subset \mathscr{P}(\Omega)$ is a system of *negligible* (or small) subsets of Ω.

Reference structure $(RS)\mathscr{R}$ is called *non-trivial*, if $\emptyset \neq \mathscr{I} \neq \mathscr{P}(\Omega)$ and $\emptyset \neq \mathscr{N} \neq \mathscr{P}(\Omega)$. $\mathscr{R}$ is called *directed*, if $\mathscr{I}$ is *top-directed* and $\mathscr{N}$ is *bottom-directed*, i.e. if $A \in \mathscr{I}$ and $B \supset A$ imply $B \in \mathscr{I}$, and $A \in \mathscr{N}$, $B \subset A$ imply $B \in \mathscr{N}$. $\mathscr{R}$ is called *regular*, if $\mathscr{I} \cap \mathscr{N} = \emptyset$, $\mathscr{R}$ is called *dual*, if $\mathscr{N} = \mathscr{I}^c = \{B : B = A^c = \Omega - A,\ A \in \mathscr{I}\}$. If $\mathscr{I}$ is top-directed, then $\mathscr{I}^c$ is bottom-directed, or, if $A \in \mathscr{I}^c$ and $B \subset A$, then $A^c \in \mathscr{I}$ and $A^c \subset B^c$, so that $B^c \in \mathscr{I}$ and $B \in {}^c$. If $\mathscr{R}$ is a directed dual RS, then $\mathscr{R}$ need not be regular. Or, take $\Omega = \{1, 2, 3, 4\}$, $\mathscr{I} = \{A:\ A \subset \Omega,\ \text{card}\,(A) \geqq 2\}$, then $\{1, 2\} \in \mathscr{I}$ and $\{3, 4\} \in \mathscr{I}$, but $\{1, 2\} = \{3, 4\}^c$, so that $\{1, 2\} \in \mathscr{I}^c$ and $\mathscr{I} \cap \mathscr{I}^c \neq \emptyset$ (in fact, $\mathscr{I} \cap \mathscr{N} = \{A:\ A \subset \Omega,\ \text{card}$

$(A)=2\}$). Reference structure $\mathscr{R}$ is called *normal*, if it is directed and $\mathscr{N} \subset \mathscr{I}^c - \mathscr{I}$; evidently, normal *RS*'s are regular.

Consider a propositional language $\mathscr{L}$ with an infinite countable set $\mathscr{V} = \{p_1, p_2, \ldots\}$ of propositional indeterminates. Let $\mathscr{O} \subset \mathscr{V}$, $\mathscr{H} \subset \mathscr{V}$. $\mathscr{O} \cap \mathscr{H} = \emptyset$ be two subsets of indeterminates, those from $\mathscr{o}$ being called *observations*, and those from $\mathscr{H}$ are *hypotheses*. Let $\mathscr{R} = \langle \Omega, \mathscr{I}, \mathscr{N} \rangle$ be a reference structure, then *$\mathscr{R}$-evaluation* of a set of formulas $\mathscr{L}_0 \subset \mathscr{L}$ is a mapping E: $\mathscr{L}_0 \to \mathscr{P}(\Omega)$. $\mathscr{R}$-evaluation E is called *consistent*, if $E(T) = \Omega$ for each propositional tautology $T \in \mathscr{L}_0$, and if $E(F_1) \subset E(F_2)$ for each $F_1, F_2 \in \mathscr{L}_0$ such that $F_1 \to F_2$ is a propositional tautology. *$\mathscr{R}$-evaluated data base* $\mathscr{B}$ is a finite set of formulas of the form $A_1 \,\&\, A_2 \,\&\, \ldots \,\&\, A_n \to C$, where $A_i \in \mathscr{O}$, $i = 1, 2, \ldots, n$, $C \in \mathscr{H}$, together with their consistent $\mathscr{R}$-evaluation E inside a reference structure $\mathscr{R}$.

An implication $B_1 \,\&\, \ldots \,\&\, B_n \to C$ is *relevant* with respect to a hypothesis C^* and observations $A_1, A_2, \ldots, A_m$, if $C = C^*$ and $\{B_1, \ldots, B_n\} \subset \{A_1, \ldots, A_m\}$, let $\mathrm{Rel}(\mathscr{B}, C, A_1, \ldots, A_m)$ denote the set of all implications from a data base $\mathscr{B}$ which are relevant with respect to $C, A_1, \ldots, A_m$. *Set-valued* (or Ω-valued, or $\mathscr{R}$-valued) *weight* $W(C/A_1, \ldots, A_m)$ *of the hypothesis* C *under observations* $A_1, A_2, \ldots, A_m$ is defined simply by

$$W(C/A_1, \ldots, A_m) = U\{E(F)\colon F \in \mathrm{Rel}(\mathscr{B}, C, A_1, \ldots, A_m)\}. \tag{1}$$

The notion of the acceptance of a hypothesis on the ground of a data base and some observations can be defined in several ways. Let us introduce some of them.

Hypothesis C is *i-accepted on the ground of an $\mathscr{R}$-evaluated database $\mathscr{B}$ and observations* $A_1, \ldots, A_m$, in symbols $D_i(\mathscr{B}, A_1, \ldots, A_m) = Acc$, for $i = 1, 2, 3$, if

$$D_1(\mathscr{B}, C, A_1, \ldots, A_m) = Acc, \quad \text{if} \quad (\exists F \in \mathrm{Rel}(\mathscr{B}, C, A_1, \ldots, A_m))\ (E(F) \in \mathscr{I}(\mathscr{R})), \tag{2}$$

$$D_2(\mathscr{B}, C, A_1, \ldots, A_m) = Acc, \quad \text{if} \quad W(C/A_1, \ldots, A_m) \in \mathscr{I}(\mathscr{R}), \tag{3}$$

$$D_3(\mathscr{B}, C, A_1, \ldots, A_m) = Acc, \quad \text{if} \quad (W(C/A_1, \ldots, A_m))^c \in \mathscr{N}(\mathscr{R}). \tag{4}$$

For the sake of simplicity we do not introduce the possible modifications of (2), (3) and (4) which differ from the presented ones only for non-directed *RS*'s (e.g., the right-hand side in (2) could be replaced by $(\exists F \in \mathrm{Rel}(\mathscr{B}, C, A_1, \ldots, A_m))\ (\exists \tilde{C} \in \mathscr{I}(\mathscr{R}))\ (E(F) \supset \tilde{C})$, or in (3) by $(\exists \tilde{C} \in \mathscr{I}(\mathscr{R}))\ (W(C/A_1, \ldots, A_m) \supset \tilde{C})$).

The intuition behind (2) is that the pieces of information presented by particular experts cannot be combined, and a hypothesis is accepted just in the case of at least one expert considers as important (large enough) the degree of validity of an implication the antecedent of which is known to hold due to the observations being at hand. This corresponds to the case when the maximum function is taken to combine numerical weights. In the case described by (3), a non-trivial combination of knowledge

is possible and its actual form depends on (and can be handled by) the choice of $\mathscr{I}$. This combination is of intensional character if compared with those used in numerically weighted extensional expert systems. Evidently, if F and G are relevant implications with the same set weights $E(F)=E(G)$ ascribed to them by the corresponding experts, with the sets taken as "sets of worlds when the implication in question hold", then their union is, again, $E(F)$ and the act of combination improves nothing (the case of deterministic mutual dependence between the validity of the two assertions in the probabilistic models). If $E(F)\neq E(G)$ then the expert which claims G and $E(G)$ says that there are still other worlds covered by the observed facts and assuring the validity of C, than those for which F holds, and this information strictly improves that one given by F and $E(F)$ (the case of at least partial statistical independence in the probabilistic models).

Almost self-evident is the following.

Assertion 1. (a) If the *RS* $\mathscr{R}$ is such that $\mathscr{I}(\mathscr{R})$ is top-directed, then 1-acceptance implies 2-acceptance.

(b) If $\mathscr{R}$ is dual, then 2-acceptance and 3-acceptance coincide. □

Proof. (a) If $D_1(\mathscr{B}, C, A_1, \ldots, A_m)=Acc$, then there is $F\in \mathrm{Rel}\,(\mathscr{B}, C, A_1, \ldots, A_m)$ such that $E(F)\in\mathscr{I}$, $\mathscr{I}$ is top-directed, hence,

$$E(F)\subset U\{E(G):\ G\in \mathrm{Rel}\,(\mathscr{B}, C, A_1, \ldots, A_m)\}=$$

$$=W(C/A_1, \ldots, A_m)\in\mathscr{I}(\mathscr{R}), \tag{5}$$

so that $D_2(\mathscr{B}, C, A_1, \ldots, A_m)=Acc$.

(b) If $\mathscr{R}$ is dual, then $\mathscr{N}(\mathscr{R})=(\mathscr{I}(\mathscr{R}))^c$, hence, $W(C/A_1, \ldots, A_m)\in\mathscr{I}(\mathscr{R})$ implies $\Omega-W(C/A_1, \ldots, A_m)\in\mathscr{N}(\mathscr{R})$ and vice versa. □

Considering various specific cases of reference structures we would be able to introduce and prove more particular assertions like the presented one, but we shall not follow this line now, keeping in the focus of our attention rather some more fundamental and methodological aspects of the chosen approach. Let us remark, first, that not only the acceptance of a hypothesis on the ground of an $\mathscr{R}$-evaluated data base and some observations, but also the choice of the "best" or "strongest" hypothesis can be defined in the same terms as follows.

Hypothesis C is called *a strong solution* given by the $\mathscr{R}$-valued data base $\mathscr{B}$ and observations $A_1, A_2, \ldots, A_m$, if for each hypothesis $C_1\in\mathscr{H}$, $W(C_1/A_1, \ldots, \ldots, A_m)\subset W(C/A_1, \ldots, A_m)$. C is called *a weak solution* given by $\mathscr{B}$, $A_1, \ldots, A_m$, if there is *no* $C_1\in\mathscr{H}$ such that $W(C/A_1, \ldots, A_m)\subset W(C_1/A_1, \ldots, A_m)$. Evidently, if $\mathscr{B}$, $A_1, \ldots, A_m$ give a strong solution, then it is also a weak solution and there are no other weak solutions. However, it is possible that there are weak solutions such that their weights are not comparable with respect to the set inclusion, hence, no strong solutions exist. The difference between the two kinds of solutions reflects the

higher degree of flexibility offered by the set-valued weights when compared with the numerical ones. A strong (weak, resp.) solution is called *an accepted strong solution* (*an accepted weak solution*, resp.), if it is 2-accepted, i.e. accepted in the sense of (3).

Assertion 2. The following two statements are equivalent:

(a) The reference structure $\mathscr{R}$ is such that $\mathscr{I}(\mathscr{R})$ is nested, i.e. either $A \subset B$ or $B \subset A$ for each $A, B \in \mathscr{I}$.

(b) The accepted strong solutions given by $\mathscr{B}, A_1, \ldots, A_m$ coincide with the accepted weak solutions given by $\mathscr{B}, A_1, \ldots, A_m$. □

Proof. If $\mathscr{I}$ is nested, then for every $C, C_1 \in \mathscr{H}$ such that $W(C/A_1, \ldots, A_m) \in \mathscr{I}$, $W(C_1/A_1, \ldots, A_m) \in \mathscr{I}$, either $W(C/A_1, \ldots, A_m) \subset W(C_1/A_1, \ldots, A_m)$ or vice versa. Hence, there are (in general, more) hypotheses $C_1, C_2, \ldots, C_k \in \mathscr{H}$ such that

$$W(C_1/A_1, \ldots, A_m) = W(C_2/A_1, \ldots, A_m) = \ldots = W(C_k/A_1, \ldots, A_m),$$
$$W(C'(A_1, \ldots, A_m) \nleqq$$
$$\nleqq W(C_1/A_1, \ldots, A_m) \text{ for all } C' \in \mathscr{H} - \{C_1, \ldots, C_k\}.$$

So, $C_1, \ldots, C_k$ are all the accepted strong as well as weak solutions given by $\mathscr{B}, A_1, \ldots, A_m$.

Contrary, let $I_1, I_2 \in \mathscr{I}$ be such that $I_1 - I_2 \neq \emptyset$, $I_2 - I_1 \neq \emptyset$. Consider such an $\mathscr{R}$-evaluation of $\mathscr{B}$ that $E(F) = I_1$ for all $F \in \mathrm{Rel}(\mathscr{R}, C_1, A_1, \ldots, A_m)$, $E(G) = I_2$ for all $G \in \mathrm{Rel}(\mathscr{B}, C_2, A_1, \ldots, A_m)$, where $C_1 \neq C_2$ are two hypotheses, $E(H) \nleqq I_1 \cup I_2$ for all $H \in U_{C^* \in \mathscr{H} - \{C_1, C_2\}} \mathrm{Rel}(\mathscr{B}, C', A_1, \ldots, A_m)$. Then $W(C_1/A_1, \ldots, A_m) = I_1$, $W(C_2/A_1, \ldots, A_m) = I_2$ and the two values are neither comparable nor dominated by some $W(C'/A_1, \ldots, A_m)$. Hence, C_1 and C_2 are accepted weak, but not strong solutions given by $\mathscr{B}, A_1, \ldots, A_m$. □

4. The possible-worlds interpretation of set-valued weights in expert systems

A possibility to take profit of modal logics when building expert systems has been already suggested by P. Hájek in [4]. We were inspired by this suggestion in our effort to apply some ideas of modal logics at the level of their semantic interpretations based on the conception of "possible worlds". Some more ideas in this direction will be presented in this chapter, however, for the sake of illustration, let us begin with the example promised above.

Consider an ecological investigation concerning a limited territory (region) and dealing with dependences among occurrences of particular species of plants or animals within this region. A specialist has investigated a quarter of this region and discovered the validity of an implication for 80% of the investigated area (i.e. for 20% of the whole territory). In the case of numerical weights he may express his observation

in two ways. First, he may use the minimax or "safety-first" principle and proclaim the implication in question to hold with the weight 0.2. It is what he may swear on, without any risk, as it is based on his personal experience and evidence. Second, he may extrapolate his observation to all the region and ascribe to the implication the weight 0.8. Evidently, there are several hidden assumptions behind such an extrapolation, the validity of which may be a matter of discussion (what are the reasons to believe the territory to be ecologically homogeneous?). In both the cases, however, a combination of pieces of knowledge of the type like this, offered by two or more experts, is practically impossible or useless supposing we have no more data about the way in which the weights have been obtained and about the validity of some more assumptions (e.g. the disjointness of the investigated sub-areas in the first case, ecological homogeneity of the territory in the second case).

When considering set-valued weight functions, the problem becomes much more simple and can be solved almost at the hardware level. Each expert is given a copy of a map of the investigated territory and he simply draws in the region where he observed the implication in question to hold. This region is nothing else than the value of the weight function he ascribes to this implication. An easy superprojection of all copies obtained from particular experts, together with an extra-copy of this map showing the important regions (sets) demonstrates, whether the union of all regions given by experts coincides with (or covers) an important region. In such a case the implication is accepted (2-accepted, in the terms from above, 1-acceptance or 3-acceptance could be investigated similarly) with all the possible extra-mathematical (i.e. ecological, in this case) consequences.

As a matter of fact, there is a degree of "true" randomness in the fact whether such and such biological species occur at this place. However, for the sake of this paper we may limit ourselves to a purely "deterministic ecology", where the places of occurrences of particular species are fully determined by the properties of these places and every uncertainty is caused by the incompleteness of our knowledge concerning these properties (conditions of occurrence, in other words). Hence, instead of propositional implication $A_1 \,\&\, A_2 \,\&\, \ldots \,\&\, A_n \rightarrow C$ saying "if species A_1 and ... and A_n occur at a place, then also species C occurs here", perhaps with some weight w ascribed to this sentence, we should rather consider a first-order implication $A_1(x) \,\&\, A_2(x) \,\&\, \ldots \,\&\, A_n(x) \rightarrow C(x)$, denoted by I in the sequel, with an indeterminate x ranging over the investigated territory (in ecological terms) or universe of discourse (in the logical terms). T. The interpretation is straightforward: "if species A_1 and ... and A_n occur at a place x in T, then also species C occurs at this place x". Setting $\tilde{W}(I) = \{x\colon x \in T,\ I(x) \text{ holds}\}$, we may either take $\tilde{W}(I)$ as the value of a set-valued weight function ascribed to I, or we may take a numerical characteristic of $\tilde{W}(I)$, its probability with respect to a probability measure defined on an appropriate measurable space over T, as the weight w of validity for the implication in question.

Let us remark that the way from I to $\tilde{W}(I)$ can also be seen as an interpretation of I, taken as a formula of a modal propositional calculus, in the semantical structure of "possible worlds" over T. Hence, formula I is valid simply if the actual world is in $\tilde{W}(I)$, formula NI ("necessary I", "I necessarily holds") is valid, if $\tilde{W}(I) = T$, or, more generally, if a probability measure of $\tilde{W}(I)$ equals to one, or, even more generally, if $\tilde{W}(I)$ is an important set (i.e. $\tilde{W}(I) \in \mathscr{I}$ for an appropriate reference structure $\mathscr{R} = \langle \Omega, \mathscr{I}, \mathscr{N} \rangle$). Dually, formula MI ("maybe I", "possibly I", "it is possible that I holds") is valid if $\tilde{W}(I) \neq \emptyset$, or, more generally, if probability of $\tilde{W}(I)$ is positive, or, even more generally, if $\tilde{W}(I)$ is not a negligible set (i.e. if $\tilde{W}(I) \in \mathscr{P}(\Omega) - \mathscr{N}$).

In the framework of this interpretation, the partial (incomplete) character of knowledge provided by particular experts can be easily expressed by set inclusion. Hence, the set-valued weight $E(F)$ ascribed by an expert to a formula F is a subset of $\tilde{W}(F)$, namely, the subset of worlds for which the expert knows (by his personal evidence or experience, say) the formula to be valid. Of course, "worlds" here need not only be "topological places" in the sense of the example presented above, but, in general, truth-valued combinations which generate, through appropriate formulas, an atomization of the universe of discourse such that the validity or non-validity of F is completely defined for each atom. Now, if more experts ascribe set-valued weights $E_1(F), E_2(F), \ldots, E_K(F)$ to the same formula F, the union $\bigcup_{i=1}^{K} E_i(F)$ ($\subset \tilde{W}(F)$) seems to be the most natural way of how to combine the pieces of knowledge offered by particular experts, at least under the interpretation in question.

Let $A_1, A_2, \ldots, A_m$ be observations, let C be a hypothesis, let $F_1, F_2, \ldots, F_K$ be all implications from a data base $\mathscr{B}$ which are relevant to $A_1, \ldots, A_m, C$, together with their set-valued $E(F_i) \subset \tilde{W}(I) \subset T$, $i = 1, 2, \ldots, K$, where T is an appropriate space of possible worlds. If $F_i = B_{i_1} \,\&\, B_{i_2} \,\&\, \ldots \,\&\, B_{i_{M(i)}} \to C$, $i \leqq K$, then $\{B_{i_1}, B_{i_2}, \ldots, B_{i_{M(i)}}\} \subset \{A_1, \ldots, A_m\}$ due to the definition of relevance, hence, the implication $F^* = A_1 \,\&\, \ldots \,\&\, A_m \to C$ is a logical consequence of F_i, so $\tilde{W}(F_i) \subset \tilde{W}(F^*)$. Due to the definition of $W(C/A_1, \ldots, A_m) = U\{E(F)\colon F \in \mathrm{Rel}(\mathscr{R}, C, A_1, \ldots, A_m)\}$, also $W(C/A_1, \ldots, A_m) \subset \tilde{W}(F^*)$. So, if $\mathscr{R} = \langle \Omega, \mathscr{I}, \mathscr{N} \rangle$, $\mathscr{I}$ is top-directed, and $W(C/A_1, \ldots, A_m) \in \mathscr{I}$, then also $\tilde{W}(F^*) \in \mathscr{I}$.

Given a set $S \subset \Omega$ and reference structure $\mathscr{R} = \langle \Omega, \mathscr{I}, \mathscr{N} \rangle$ another set S' may be called *S-important*, $S \cap S' \in \mathscr{I}_S = \{S \cap R\colon R \in \mathscr{I}\}$. Hypothesis C is called *observationally important* (σ-important) given $\mathscr{R}$, $A_1, \ldots, A_m$, if C is $\mathscr{A}$-important for $\mathscr{A} = \{x \in T\colon A_i(x) \text{ holds}, i = 1, 2, \ldots, m\}$, hence, if $\tilde{W}(F^*) \in \mathscr{I}$. So we have arrived at this

Assertion 3. If a hypothesis C is 2-accepted on the ground of an $\mathscr{R}$-evaluated database $\mathscr{B}$ and observations $A_1, \ldots, A_m$, and if $\mathscr{I}(\mathscr{R})$ is top-directed, then C is observationally important with respect to $\mathscr{B}, \mathscr{R}, A_1, \ldots, A_m$, i.e. C is $\mathscr{A}$-important with respect to the set $\mathscr{A}$ of worlds in which $A_1, A_2, \ldots, A_m$ hold. □

The assertion presented above and the foregoing argumentation proves the set-union operation, even if it is extensional with respect to the space of all subsets of Ω, to be quite reasonable and justifiable, at least under the interpretation in question, when applied to values of relevant implications in order to accumulate them. This fact is in narrow connection with another evident fact, according to which the set of all one-dimensional marginal set-valued distributions uniquely determines the corresponding simultaneous distribution (which is, of course, not true for the corresponding numerical-valued marginals). Or, let $X_1, X_2, \ldots, X_n$ be random variables taking an abstract probability space $\langle \Omega, \mathscr{F}, P \rangle$ into measurable spaces $\langle Z_i, \mathscr{Z}_i \rangle$, $i \leqq n$, then the system $\{\{\omega\colon \omega \in \Omega, X_i(\omega) \in A\}\colon A \in \mathscr{Z}_i\}_{i=1}^n$ evidently determines the system $\{\{\omega\colon \omega \in \Omega, \langle X_1(\omega), X_2(\omega), \ldots, X_n(\omega)\rangle \in X_{i=1}^n A_i\}\colon A_i \in \mathscr{Z}_i, i \leqq n\}$, simply by using the set-joint operation, but the system $\{P(\{\omega\colon \omega \in \Omega, X_i(\omega) \in A\})\colon A \in \mathscr{Z}_i\}_{i=1}^n$ does not determine the system $\{P(\{\omega\colon \omega \in \Omega, \langle X_1(\omega), X_2(\omega), \ldots, X_n(\omega)\rangle \in X_{i=1}^n A_i\})\colon A_i \in \mathscr{Z}_i, i \leqq n\}$. Denoting the support of the considered probability space by Ω, we would like to emphasize the common features of the "possible worlds" semantic behind modal logics as well as behind the classical probability theory.

Without any doubts, the scale of possible applications of modal logics and their semantics, when combining non-numerically weighted pieces of knowledge, is much more larger than to contain only the few simple ideas mentioned above. These problems, as well as those connected with non-numerical weights or degrees of belief in general, would deserve a more detailed investigation including many problems of more specific and technical (in the mathematical sense) nature arising from the conception briefly outlined above. Nevertheless, this paper has been conceived with the aim to discuss, before all, the motivations, meta-theoretical backgrounds and outcomes for the approach presented above, rather than to develop a detailed formal apparatus. Keeping this aim in mind, we postpone the latter purpose to another occasion.

References

1. *Feys, R.*, Modal Logics. Dunod, Paris, 1965 (Russian translation: Nauka, Moscow, 1974).
2. *Goguen, J., L*-fuzzy sets. Journal of Math. Analysis and Applications **18** (1967), pp. 147–174.
3. *Hájek, P.*, Combining functions for certainty degrees in consulting systems. Inter. J. Man-Machine Studies **22** (1985), pp. 59–76.
4. *Hájek, P., Valdes, J.*, Algebraic foundations of uncertainty processing in rule based consulting systems I. In: Sistemy obrabotki znanij i izobraženij (I. Plander, ed.), Smolenice 1986, vol. **1**, pp. 93–98.
5. *Hájek, P., Valdes, J.*, Algebraic foundations of uncertainty processing in rule based consulting systems II. In: Aplikace umělé inteligence AI87 (T. Chrz, ed.), ÚISK, Praha, 1987, pp. 9–14.
6. *Kripke, S. A.*, Semantical considerations on modal logic. Acta Phil. Fennica **16** (1963), pp. 83–94.
7. *Perez, A.*, Probabilistic approach to the integration of partial knowledge for medical decision making (in Czech). In: Souhrn přednášek 1, 1. celostátní vědecký sjezd biomedicínského inženýrství BMI83, Mariánské Lázně, pp. 221–226.

8. *Perez, A., Jiroušek, R.*, Constructing an intensional expert system (INES). In: Medical Decision Making: Diagnostic strategies and expert systems. Elsevier Science Publishers (North Holland), IFIP-IMIA, 1985, pp. 307–315.

Экспертные системы с ненумерическими функциями доверия

И. КРАМОСИЛ

(Прага)

Предлагаются некоторые аргументы, показывающие, что предположение о нумерической природе значений функций доверия или весов в экспертных системах может являться слишком ограничивающим и что для этих целей возможно применять также объекты другого характера. Более подробно изучаются экспертные системы, основанные на правилах, которым приписываются доверия или веса в виде подмножеств заданного множества; предлагаются также некоторые элементарные мэта-правила для комбинирования весов и для принятия решений. Изучаются некоторые связи с так называемой семантикой возможных миров для модальных логик.

I. Kramosil
Institute of Information Theory and Automation
Czechoslovak Academy of Sciences
Pod vodárenskou věží 4, 18208 Praha 8
Czechoslovakia

Problems of Control and Information Theory, Vol. 17 (5), pp. 297—310 (1988)

ON THE STABILITY OF SOLUTION OF THE INVERSE PROBLEMS IN CONTROL SYSTEM DYNAMICS

M. I. GUSEV

(Sverdlovsk)

(Received June 5, 1987)

This paper is concerned with the general scheme of solution of guaranteed estimation problems. A solution that is stable with respect to errors in available data are described. It is based on constructions from the theory of control and observation under uncertainty and the theory of ill-posed problems. Applications to the problem of estimating of disturbances in the input of control system and the problem of identification of system parameters are considered.

1. Introduction

The inverse problem of dynamics deals with the identification of forces that generate the motion of a mechanical system from the given properties of motion. The classical statements of the problem in analytical mechanics and the methods of their solutions are developed under assumptions that the trajectory of the system is smooth enough (as a function of time) and that it is exactly specified. However, these conditions are usually not fulfilled if one considers a control system: the unknown input of the system (the control forces, disturbances, and the initial conditions) generate a motion which may not be a smooth function and the measurements of the parameters of may be given with errors. Moreover, the number of parameters accessible for measurement may not be sufficient to obtain the unique solution. This leads to the necessity of describing the whole set of forces generating the given measurement.

The mentioned specificity of the problem makes natural its treatment as a dynamic problem of estimation under uncertainty conditions and the use of constructions developed in the corresponding theory [1, 2].

A characteristic feature of the problem is its instability with respect to the data errors which requires the application of special regularization procedures.

In the general solution scheme described here the inverse problem of control system dynamics is treated as a problem of estimation of the value of some operator in Banach space under given operator restrictions. The scheme may as well be applied to problems of estimation of the state space vector and parameters of a dynamic system. Stable procedures of solution, based on a combination of constructions of

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

the theory of guaranteed estimation [2] and the theory of ill-posed problems [3–4] are described in this paper.

The results represented in this paper were described in reports [5–7].

The inverse problems of dynamics for mechanical systems have been considered in various papers (see, e.g. a survey in [8]).

Some papers are concerned with a single-valued reconstruction of the input from precise measurements of the output (see, e.g. [9–11]). The results of these investigations show that even if such a reconstruction is possible, the corresponding operator output–input is, in general, discontinuous.

The problem of an a priori estimation [1, 2] of the input of a linear system was considered in [12, 13].

A stable reconstruction of the input for systems with completely observable state space vector, based on constructions from the theory of differential games, was considered in [14].

In this paper the description of the dynamics of estimates is not considered (see, e.g. [15, 16]).

Let the motion of a controlled plant on the interval $[t_0, t_1]$ be described by the differential equation

$$\dot{x} = f(t, x, u(t)), \quad x(t_0) = x^0, \quad x \in R^n, \quad u \in R^r, \tag{1.1}$$

where the initial state x^0 and the function $u(\cdot)$ representing the disturbance (control) are assumed to be unknown.

Let us assume that $f(t, x, u)$ is continuous and that the following standard conditions are satisfied: there exist some values $c_1, c_2 > 0$, such that

a) $\|f(t, x, u)\| \leqq c_1(1 + \|x\| + \|u\|)$,
b) $\|f(t, x, u) - f(t, x', u)\| \leqq c_2 \|x - x'\|(1 + \|u\|)$

for all $t \in [t_0, t_1]$, $x, x' \in R^n$, where $\|\cdot\|$ means the Euclidean norm. We shall say that a function from the space $L_2([t_0, t_1], R^r)$ (further also written as L_2^r) is an admissible control (disturbance) Assume $u(\cdot) \in L_2^r$. An absolutely continuous function $x(t)$ is said to be a solution of system (1.1) if

$$x(t) = x^0 + \int_{t_0}^{t} f(\tau, x(\tau), u(\tau))\, d\tau, \qquad t_0 \leqq t \leqq t_1.$$

For every $x^0 \in R^n$, $u(\cdot) \in L_2^r$, there exists a unique solution $x(t, x^0, u(\cdot))$.

Let the output of the plant be given by the equation

$$y(t) = h(t, x(t, x^0, u(\cdot)), u(t)), \qquad t \in [t_0, t_1], \tag{1.2}$$

where h: $[t_0, t_1] \times R^n \times R^r \to R^m$ is continuous.

The problem is to identify x_0 and $u(\cdot)$ (or $u(\cdot)$ or some of the coordinates of $u(\cdot)$ on a given basis) from an output $y(\cdot)$. Usually some additional a priori information is assumed to be known. We assume this information to be given by the inclusion $w=(x^0, u(\cdot)) \in W$, where W is a set in a corresponding functional space. Examples of a priori restrictions on $(x^0, u(\cdot))$ could be the following: geometrical

$$W=\{(x^0, u(\cdot)): x^0 \in X^0,\ u(t) \in U, \quad \text{a.e.} \quad t \in [t_0, t_1]\},$$

where $X \subset R^n$, $U \subset R^r$ or integral

$$W=\left\{(x^0, u(\cdot)): x^{0T} M x^0 + \int_{t_0}^{t_1} u^T(t) R(t) u(t) dt \leqq \mu^2\right\}$$

where M, $R(t)$ are positive definite matrices.

2. General scheme of solution of the problem of guaranteed estimation (static statement)

A wide range of problems of estimation for dynamic systems, when the measurements are performed on a fixed interval of time admit the following statement.

Let X, Y, Z be real Banach spaces, and operators $A: X \to Y$, $F: X \to Z$, a point $y \in Y$ and a set $W \subset X$ be given. It is necessary to find $z=Fw$ under the conditions $Aw=y$, $w \in W$. Here w may be treated as the unknown input of the system, W gives an a priori information on w, A is an input–output operator, and y is the result of measurements of the output.

Operator A is, as a rule, noninvertible, therefore, $z=Fw$ is not uniquely determined. Following [2] we shall consider the set

$$\hat{Z}=\{z=Fw,\ Aw=y,\ w \in W\} \tag{2.1}$$

which we shall call the informational domain associated with y. The set $\hat{Z}$ depends on W, A, F and y: $\hat{Z}=Z(W, A, F, y)$, we shall further consider that W, A, F are known exactly and will denote $\hat{Z}$ as $Z(y)$.

As a rule, the problem of identification of $Z(y)$ is instable with respect to errors in y. In other words, if y is given with an error that does not exceed $\delta > 0 (\|y-y_\delta\| < \delta)$ then $Z(y_\delta)$ may have an arbitrarily large deviation from $Z(y)$, regardless of how small δ is taken. Further, we shall describe a construction which ensures the stability (robustness) of the estimate. Instead of $Z(y)$ we consider its extension $Z_{\varepsilon,\alpha}(y)$, that depends on positive parameters ε, α (parameters of regularization) and approximates $Z(y)$ as ε, $\alpha \to 0$. We shall describe construction of $Z_{\varepsilon,\alpha}(y)$ which is applicable in the case of Hilbert spaces X, Y, Z. This practically does not restrict the variety of estimation problems for which this construction may be applied.

Let X, Y, Z be Hilbert spaces, let $\langle \cdot, \cdot \rangle_X$ represent the corresponding scalar product and let $W \subset X$ be weakly compact. Let $w^* \in X$, $\mu \in R$ be such that

$$W \subset \{w: \|w - w^*\| \leqq \mu^2\} \text{ (for example } \mu = \sup\{\|w - w^*\|: w \in W\}).$$

Definition. The set

$$Z_{\varepsilon,\alpha}(y) = \{z = Fw: Aw + \xi \equiv y,\ \langle w - w^*,\ w - w^* \rangle_X + 1/\varepsilon \langle \xi, \xi \rangle_Y \leqq \leqq \mu^2 + \alpha,\ w \in W\} \tag{2.2}$$

is called the regularized informational domain (or the regularization of $Z(y)$). For the case $W = \{w: \|w - w^*\| \leqq \mu^2\}$ let us define $Z_{\varepsilon,\alpha}(y)$ as follows

$$Z_{\varepsilon,\alpha}(y) = \{z = Fw: Aw + \xi \equiv y,\ \langle w - w^*,\ w - w^* \rangle_X + + 1/\varepsilon \langle \xi, \xi \rangle_Y \leqq \mu^2 + \alpha\}. \tag{2.3}$$

Here ε, α are positive parameters (the regularization parameters).

An operator A is said to be weakly closed (sequentially weakly closed) iff for every sequence $w_n \in X$ conditions $w_n \rightharpoonup w$, $Aw_n \rightharpoonup y$ imply that $Aw = y$. Here sign $\rightharpoonup$ means a weak convergence in the corresponding space.

Proposition 2.1. If F_a is bounded operator, then $Z(y)$, $Z_{\varepsilon,\alpha}(y)$ are bounded sets. If A, F are weakly closed then $Z(y)$, $Z_{\varepsilon,\alpha}(y)$ are closed in the weak topology.

The proof is obvious.

Proposition 2.2

a) $Z(y) \subset Z_{\varepsilon,}{}^{\alpha}(y_\delta)$ as $\delta^2/\varepsilon \leqq \alpha$,

b) if A, F are weakly closed and F is completely continuous, then

$$h(Z_{\varepsilon,\alpha}(y_\delta),\ Z(y)) \to 0 \text{ as } \varepsilon \to 0,\ \alpha \to 0,\ \delta^2/\varepsilon \leqq \alpha,$$

where h is the Hausdorff metric.

Proof. Let $z \in Z(y)$, that is $z = Fw$, $Aw = y$, $w \in W$. Let us rewrite the equality $Aw = y$ in the following way $Aw + \xi = y_\delta$, where $\xi \equiv y_\delta - y$. Obviously

$$\langle w - w^*,\ w - w^* \rangle_X + 1/\varepsilon \langle \xi, \xi \rangle_Y \leqq \mu^2 + 1/\varepsilon \|y - y_\delta\|^2 \leqq \mu^2 + \delta^2/\varepsilon \leqq \mu^2 + \alpha,$$

hence $z \in Z_{\varepsilon,\alpha}(y_\delta)$.

Let us assume that $h(Z_{\varepsilon,\alpha}(y_\delta), Z(y)) \to 0$ as $\varepsilon \to 0$, $\alpha \to 0$, $\delta^2/\varepsilon \leqq \alpha$. Then there exists a number $\beta > 0$ and some sequences

$$\varepsilon_n, \alpha_n \to 0,\ y_n;\ \|y_n - y\| \leqq \delta_n,\ \delta_n^2/\varepsilon_n \leqq \alpha_n,\ z_n \in Z_{\varepsilon_n,\alpha_n}(y_n),$$

such that in $f\{\|z_n - z\|: z \in Z(y)\} \geqq \beta > 0$. Condition $z_n \in Z_{\varepsilon_n,\alpha_n}(y_n)$ implies that $z_n = Fw_n$, $Aw_n + \xi_n = y_n$, $\langle w_n - w^*,\ w_n - w^* \rangle_X + 1/\varepsilon \langle \xi_n,\ \xi_n \rangle \leqq \mu^2 + \alpha_n$. As $z_n \in Z_{\varepsilon_n,\alpha_n}(y_n)$ it follows from boundedness of W and the complete continuity of F that $\{z_n\}$ contains a convergent subsequence. Let us suppose that $z_n \to z^*$. Obviously $\inf(\|z - z^*\|: z \in$

$\in Z(y)) \geqq \beta$. As $\langle \xi_n, \xi_n \rangle_Y \leqq \varepsilon_n(\mu^2 + \alpha_n)$, then $\xi_n \to 0$, $n \to \infty$. Sequence w_n contains a weakly convergent subsequence. Let us suppose that $w_n \to w^*$, $w^* \in W$. A weak closure of A, F implies that $Aw^* = y$, $z^* = Fw^*$, hence $z^* \in Z(y)$. This contradicts the assumption that condition b) is not true.

For the case when $Z_{\varepsilon,\alpha}(y)$ is defined by (2.3) the proof is similar.

Remark 2.1. Obviously $Z(y) \neq 0$ if and only if $y \in AW$, $Z_{\varepsilon,\alpha}(y_\delta) \neq 0$ if and only if $\min \{\|y_\delta - v\|^2 : v \in AW\} \leqq \varepsilon\alpha$.

Proposition 2.3. Let operator A be weakly closed and F be a linear continuous operator. Then for every $z^* \in Z$

$$\rho(z^* | Z_{\varepsilon,\alpha}(y_\delta)) \to \rho(z^* | Z(y))$$

as $\varepsilon \to 0$, $\alpha \to 0$, $\delta^2/\varepsilon \leqq \alpha$, where $\rho(\cdot | Z)$ is the support function of Z.

Proof. The operator $\hat{F} : X \to R$ defined by the equality $\hat{F}w = \langle z^*, Fw \rangle$ is completely continuous. The sets $\hat{Z}(y)$, $\hat{Z}_{\varepsilon,\alpha}(y)$ defined by (2.1), (2.2) with substitution of F into $\hat{F}$ are the intervals

$$\hat{Z}(y) = [a(y), b(y)], \quad Z_{\varepsilon,\alpha}(y) = [a(\varepsilon, \alpha, y), \ b(\varepsilon, \alpha, y)]$$

where

$$a(y) = -\rho(-z^* | Z(y)), \quad b(y) = \rho(z^* | Z(y))$$

$$a(\varepsilon, \alpha, y) = -\rho(-z^* | Z_{\varepsilon,\alpha}(y)), \quad b(\varepsilon, \alpha, y) = \rho(z^* | Z_{\varepsilon,\alpha}(y)).$$

Using proposition 2.2 we observe that $[a(\varepsilon, \alpha, y_\delta), b(\varepsilon, \alpha, y_\delta)] \to [a(y), b(y)]$ in Hausdorff metric, hence

$$b(\varepsilon, \alpha, y_\delta) = \rho(z^* | Z_{\varepsilon,\alpha}(y_\delta)) \to b(y) = \rho(z^* | Z(y)).$$

Let us consider examples of control systems for which operators A, F satisfy the conditions of propositions 2.1–2.3.

Let equations (1.1), (1.2) be as follows

$$dx/dt = A(t, x) + B(t, x)u(t), \quad x(t_0) = x^0, \quad t_0 \leqq t \leqq t_1, \tag{2.4}$$

$$y(t) = h_1(t, x(t)) + h_2(t, x(t))u(t), \tag{2.5}$$

where $x \in R^n$, $y \in R^m$. If matrices $A(t, x)$, $B(t, x)$ are continuous in t, x and Lipschitz in x, and satisfy conditions $\|A(t, x)\| \leqq C(1 + \|x\|)$, $\|B(t, x)\| \leqq C$ and if h_1, h_2 are continuous, then the operator A transforms $R^n \times L_2^r$ into L_2^m.

Lemma 2.1. The operator $A : R^n \times L_2^m$ defined by equations (2.4), (2.5) is weakly closed.

The proof is given in [17].

Let us consider some typical examples of operators F in estimation problems.

a) $Fw = x^0$ — the problem of estimation of the initial state;

b) $Fw = P_L u(\cdot)$, where P_L is the projection operator on subspace $L \subseteq L_2^r$;

c) $Fw = x(t^*, x^0, u(\cdot))$ where $x(t, x^0, u(\cdot))$ is the solution of (1.1), $t^* \in (t_0, t_1]$ is a given instant of time.

Operator F is linear continuous in cases a), b), c) (if $A(t, x)$, $B(t, x)$ does not depend on x); completely continuous for a), b) (if L is finite dimensional), c); weakly closed for a), b), c) (if the control system is described by equation (2.4)).

Remark 2.2. The convergence of $Z_{\varepsilon,\alpha}(y_\delta)$ to $Z(y)$ in Hausdorff metric takes place if F is completely continuous, for example, when the estimated parameter is finite dimensional. For the problem of the estimation of the input of the system this is not true in general.

As an example of the estimation problem with an infinite dimensional Z and a completely continuous F, let us consider the problem of estimation of the whole trajectory of system (2.4), (2.5). There $Fw = x(\cdot, u(\cdot), x^0)$ is completely continuous, hence $Z_{\varepsilon,\alpha}(y_\delta) \to Z(y)$ in Hausdorff metric in $L_2^n \times R^n$ (it is not difficult to show that convergence also takes place in the space $C \times R^n$).

Let us briefly describe the procedure of constructing the regularized informational domain for the case when the operator A is given with an error, assuming the linearity of the problem. Let X, Y, Z be Hilbert spaces, A, F be linear continuous operators, $W = \{w : \|w - w^*\| \leqq \mu^2\}$. Let A be given with an error not exceeding $\gamma : \|A_\gamma - A\| \leqq \gamma$. Let us define the regularized informational domain by equation (2.3) and denote it by $Z_{\varepsilon,\alpha}(A, y)$, the informational domain corresponding to the precise data is denoted as $Z(A, y)$.

Proposition 2.4

a) $Z(A, y) \subset Z_{\varepsilon,\alpha}(A_\gamma, y_\delta)$ for $(\delta + \gamma\mu)^2/\varepsilon \leqq \alpha$;

b) if F is completely continuous then $h(Z_{\varepsilon,\alpha}(A_\gamma, y_\delta), Z(A, y)) \to 0$, as $\varepsilon \to 0$, $\alpha \to 0$, $(\delta + \gamma\mu)^2/\varepsilon \leqq \alpha$.

The proof is similar to the proof of proposition 2.2.

3. An analytical description of informational domain. The linear-quadratic case

Let A, F be linear continuous operators, $W = \{w : \|w - w^*\| \leqq \mu^2\}$. Let us specify a support function of $Z_{\varepsilon,\alpha}(y)$ defined by (2.3). We shall suppose that $\min\{\|y - v\| : v \in AW\} \leqq \delta$ and $\delta^2/\varepsilon \leqq \alpha$.

It is necessary to find a solution of the problem

$$\langle z^*, z\rangle \to \max$$

$$z = Fw,\ Aw + \xi = y,\ \langle w - w^*,\ w - w^*\rangle_X + 1/\varepsilon\langle \xi, \xi\rangle_Y \leqq \mu^2 + \alpha. \tag{3.1}$$

Let us rewrite (3.1) as follows

$$\langle F^* z^*, w\rangle \to \max$$

$$\langle w - w^*, w - w^*\rangle_X + 1/\varepsilon\langle Aw - y, Aw - y\rangle_Y \leqq \mu^2 + \alpha. \tag{3.2}$$

We shall transform the last inequality, writing it as follows

$$\langle C(w-\hat{w}), w-\hat{w}\rangle_X \leqq \mu^2+\alpha-\sigma, \tag{3.3}$$

where C is a self-adjoint positive definite operator, $\hat{w}$ is an element of X, $\sigma \in R$.

From (3.2), (3.3) we obtain

$$C=I+1/\varepsilon A^*A=1/\varepsilon(A^*A+\varepsilon I),$$

where I an is identity operator from X to X, A^* stands for the operator adjoint to A. Transforming (3.2), (3.3) comes to the identity

$$\langle w, C^*\hat{w}\rangle_X=\langle w, w^*\rangle_X+\langle w, 1/\varepsilon A^*y\rangle_X,$$

whence it follows that $C^*\hat{w}=w^*+1/\varepsilon A^*y$. Operator $C^*=C$ is non-degenerate, since $\langle w, Cw\rangle_X \geqq \|w\|_X^2$, the inverse operator is determined by the equality

$$C^{-1}=\varepsilon(A^*A+\varepsilon I)^{-1}, \tag{3.4}$$

hence

$$\hat{w}=C^{-1}w^*+1/\varepsilon C^{-1}A^*y=\varepsilon(A^*A+\varepsilon I)^{-1}w^*+(A^*A+\varepsilon I)^{-1}A^*y.$$

The following equality is true

$$(A^*A+\varepsilon I)^{-1}A^*y=A^*(AA^*+\varepsilon I)^{-1}y, \tag{3.5}$$

where operator I in the right-hand side of (3.5) is an identity operator from Y to Y. Actually, let $v\equiv(AA^*+\varepsilon I)^{-1}y$, hence the right-hand side of (3.5) equals to A^*v. Substituting $y=(AA^*+\varepsilon I)v$ into the left-hand side of (3.5) we obtain

$$A^*y=A^*(AA^*+\varepsilon I)v=(A^*A+\varepsilon I)A^*v,$$

therefore $(A^*A+\varepsilon I)^{-1}A^*y=A^*v$ which proves (3.5).

The value σ in (3.3) is determined by the equality

$$\sigma=\langle w^*, w^*\rangle_X+1/\varepsilon\langle y, y\rangle_Y-\langle C\hat{w}, \hat{w}\rangle_X=\langle w^*, w^*\rangle_X+1/\varepsilon\langle y, y\rangle_Y-$$

$$-\langle w^*, \varepsilon(A^*A+\varepsilon I)^{-1}w^*\rangle_X-\langle w^*, (A^*A+\varepsilon I)^{-1}A^*y\rangle_X-\langle y, A(A^*A+$$

$$+\varepsilon I)^{-1}w^*\rangle_Y-\langle 1/\varepsilon A^*y, A^*(A^*A+\varepsilon I)^{-1}y\rangle_X.$$

Let us transform this equality, using the identities

$$I-AA^*(AA^*+\varepsilon I)^{-1}=\varepsilon(AA^*+\varepsilon I)^{-1},$$

$$I-A^*A(A^*A+\varepsilon I)^{-1}=\varepsilon(A^*A+\varepsilon I)^{-1}.$$

Obviously

$$1/\varepsilon\langle y, y\rangle_Y-\langle 1/\varepsilon A^*y, A^*(AA^*+\varepsilon I)^{-1}y\rangle_X=\langle y, (AA^*+\varepsilon I)^{-1}y\rangle_Y,$$

$$\langle w^*, w^*\rangle_X-\langle w^*, \varepsilon(A^*A+\varepsilon I)^{-1}w^*\rangle_X=\langle Aw^*, (AA^*+\varepsilon I)^{-1}Aw^*\rangle_Y \tag{3.6}$$

(here the identity $A(A^*A+\varepsilon I)^{-1}w^*=(AA^*+\varepsilon I)^{-1}Aw^*$ is used). From the equalities

$$\langle w^*, (A^*A+\varepsilon I)^{-1}A^*y\rangle_X=\langle Aw^*, (AA^*+\varepsilon I)^{-1}y\rangle_Y,$$

$$\langle y, A(A^*A+\varepsilon I)^{-1}w^*\rangle_Y=\langle y, (AA^*+\varepsilon I)^{-1}Aw^*\rangle_Y$$

it follows that

$$\sigma=\langle y-Aw^*, (AA^*+\varepsilon I)^{-1}(y-Aw^*)\rangle.$$

Identity (3.6) implies the following formula for C^{-1}

$$C^{-1}=\varepsilon(A^*A+\varepsilon I)^{-1}=I-A^*(AA^*+\varepsilon I)^{-1}A.$$

Thus, problem (3.4) takes the following form

$$\langle F^*z^*, w\rangle_X\to\max,$$

$$\langle C(w-\hat{w}), w-\hat{w}\rangle_X\leqq\mu^2+\alpha-\sigma, \tag{3.7}$$

where

$$\hat{w}=\varepsilon(A^*A+\varepsilon I)^{-1}w^*+A^*(AA^*+\varepsilon I)^{-1}y.$$

Operator C may be presented as $C=C_1C_1$, where $C_1\equiv C^{1/2}$ is a positive definite self-adjoint operator, hence, the restrictions for the problem (3.7) take the following form

$$\langle C_1(w-\hat{w}), C_1(w-\hat{w})\rangle_X\leqq\mu^2+\alpha-\sigma.$$

Let us denote $s=C_1(w-\hat{w})$, then $w=\hat{w}+C_1^{-1}s$ and problem (3.7) may be rewritten as follows

$$\langle F^*z^*, \hat{w}+C_1^{-1}s\rangle_X\to\max,$$

$$\langle s, s\rangle_X\leqq\mu^2+\alpha-\sigma. \tag{3.8}$$

This problem has an obvious solution $s^*=\Delta C_1^{-1}F^*z^*$, where Δ may be determined from the equality $\langle s^*, s^*\rangle_X=\mu^2+\alpha-\sigma$

$$\Delta=(\mu^2+\alpha-\sigma)^{1/2}(\langle z^*, FC_1^{-1}F^*z^*\rangle_Z)^{1/2}.$$

Substituting s^* into the functional of problem (3.8) we get

$$\max\langle F^*z^*, \hat{w}+C_1^{-1}s\rangle_X=\langle F^*z^*, \hat{w}+C_1^{-1}s^*\rangle_X=$$

$$=\langle z^*, F\hat{w}\rangle+(\mu^2+\alpha-\sigma)^{1/2}(\langle z^*, FC_1^{-1}F^*z^*\rangle)_Z^{1/2}.$$

As a result of this reasoning one may formulate the following theorem.

Theorem 3.1. Domain $Z_{\varepsilon,\alpha}(y)$ is an ellipsoidal set with support function

$$\rho(z^*|Z_{\varepsilon,\alpha}(y))=(\mu^2+\alpha-\sigma)^{1/2}(\langle z^*, FC_1^{-1}F^*z^*\rangle_Z)^{1/2}+\langle z^*, \hat{z}\rangle, \tag{3.9}$$

where the center of symmetry of $Z_{\varepsilon,\alpha}(y)$ — a point $\hat{z}$ — is determined by the equality

$$\hat{z} = F\hat{w}(\varepsilon, y, w^*),$$

$$\hat{w}(\varepsilon, y, w^*) = C^{-1}(w^* + 1/\varepsilon A^* y) = \varepsilon(A^*A + \varepsilon I)^{-1} w^* + (A^*A + \varepsilon I)^{-1} A^* y =$$

$$= \varepsilon(A^*A + \varepsilon I)^{-1} w^* + A^*(AA^* + \varepsilon I)^{-1} y,$$

$C^{-1} = \varepsilon(A^*A + \varepsilon I)^{-1} = I - A^*(AA^* + \varepsilon I)^{-1} A$ — is a positive definite self-adjoint operator

$$\sigma = \langle y - Aw^*, (AA^* + \varepsilon I)^{-1}(y - Aw^*)\rangle_Y.$$

In particular, under $w^* = 0$, $Fw \equiv w$, $Z_{\varepsilon,\alpha}(y)$ is an ellipsoid

$$Z_{\varepsilon,\alpha}(y) = \{w \in X : \langle w - \hat{w},\ 1/\varepsilon(A^*A + \varepsilon I)(w - \hat{w})\rangle_X \leqq \mu^2 + \alpha - 1/\varepsilon\langle y, y - A\hat{w}\rangle\}.$$

Proposition 2.4. The elements appearing in (3.9) may be specified by the solutions to the following variational problems

$$\hat{w}(\varepsilon, y, w^*) = \arg\min_w \{\|Aw - y\|^2 + \varepsilon\|w - w^*\|^2\}, \tag{3.10}$$

$$C^{-1}v = \arg\min_w \{\|Aw\|^2 + \varepsilon\|w - v\|^2\},\ \forall v \in X, \tag{3.11}$$

$$\bar{w} = \arg\min_w \{\|Aw - x\|^2 + \varepsilon\|w\|^2\},\ \forall x \in X. \tag{3.12}$$

For the proof of this proposition let us find

$$\min\{\|Aw - y\|^2 + \varepsilon\|w - v\|^2 : w \in X\}$$

for arbitrary $y \in Y$, $v \in X$. Equating to zero the Frechet derivative of the minimized functional, we get

$$A^*(Aw - y) + \varepsilon(w - v) = 0.$$

Hence the solution of the problem equals to

$$w(\varepsilon, y, v) = (A^*A + \varepsilon I)^{-1} A^* y + \varepsilon(A^*A + \varepsilon I)^{-1} v. \tag{3.13}$$

For $v = w^*$, obviously,

$$\hat{w}(\varepsilon, y, w^*) = w(\varepsilon, y, w^*),\ C^{-1}v = \varepsilon(A^*A + \varepsilon I)^{-1} v = w(\varepsilon, \sigma, v).$$

Finally, from the identity

$$(AA^* + \varepsilon I)(AA^* + \varepsilon I)^{-1} x = x$$

it follows that

$$(AA^*+\varepsilon I)^{-1}x = 1/\varepsilon(x - AA^*(AA^*+\varepsilon I)^{-1}x) =$$

$$= 1/\varepsilon(x - A(A^*A+\varepsilon I)^{-1}A^*x) = 1/\varepsilon(x - Aw(\varepsilon, x, 0)).$$

Thus the description of the regularized informational domain for the linear-quadratic problem may be obtained by solving the following variational problem of Tikhonov's regularization method

$$I(w) = \|Aw - y\|^2 + \varepsilon\|w - v\|^2 \to \min. \tag{3.14}$$

The center of symmetry $\hat{z}$ of the domain $Z_{\varepsilon,\alpha}(y)$ is the image of Tikhonov's regularizator under the mapping F. An element $\hat{z}$ may be considered to be the best regularized minimax estimate [2] of z.

For systems (1.1), (1.2) which are linear in x, u

$$dx/dt = A(t)x + B(t)u, \quad x(t^0) = x^0, \tag{3.15}$$

$$y = C(t)x(t) + D(t)u(t), \tag{3.16}$$

problem (3.14) is a standard linear-quadratic problem of optimal control

$$\int_{t_0}^{t_1} (y - Cx - Du)^T Q(y - Cx - Du)dt + \varepsilon \int_{t_0}^{t_1} (u - v)^T R(u - v)dt +$$

$$+ \varepsilon(x^0 - v^0)^T M(x^0 - v^0) \to \min \tag{3.17}$$

(here $w = (x^0, u(\cdot))$, $y = y(\cdot)$, $v = (v^0, v(\cdot))$). The solution of problem (3.17) is described, for example in [18]. For the problem of estimation of the input $Fw = u(\cdot)$, $z^* = z^*(\cdot) \in L_2^r$, $F^*z^* = (z^*(\cdot), 0) \in L_2^r \times R^n$ and for finding of the projection of $Z_{\varepsilon,\alpha}(y)$ on the direction $z^*(\cdot)$ it is necessary to solve problem (3.17) for $v(t) = z^*(t)$, $v^0 = 0$.

For the problem of estimating $x(t_1)$ ($Fw = x(t_1)$, $z^* \in R^n$) F^*z^* takes the form $F^*z^* = (s(z^*, \tau)B(\tau), s(z^*, t_0), t_0 \leqq t \leqq t_1$, where $s(z, \tau)$ is the solution to the adjoint system

$$ds/d\tau = -A^T(\tau)s, \quad s(z, t_1) = z.$$

To define a projection of $Z_{\varepsilon,\alpha}(y)$ on z^*, it is necessary to solve problem (3.17) for $v(t) = s(z^*, t)B(t)$, $v^0 = s(z^*, t_0)$. The solution of the problem is given in [2], where it is shown that description of $Z_{\varepsilon,\alpha}(y)$ may be reduced to the solution of the system of differential equations (equations of the minimax filter).

If Z is infinite dimensional, then $Z_{\varepsilon,\alpha}(y_\delta)$ does not converge to $Z(y)$ in the Hausdorff metric but contains elements converging to $Z(y)$ as $\varepsilon \to 0$, $\delta^2/\varepsilon \to 0$. These are the elements $\hat{z}(\varepsilon, y_\delta, w^*) = F\hat{w}(\varepsilon, y_\delta, w^*)$. Let at first $F = I$.

Theorem 3.2. Let system (1.1), (1.2) be of the form (2.4), (2.5) and the assumptions of Lemma 2.1 be satisfied. Let W be weakly closed on $L_2^r \times R^n$. Let $\hat{w}(\varepsilon, y, w^*)$ be the

solution of the variational problem

$$\| Aw - y \|^2_{L_2^m} + \varepsilon \| w - w^* \|^2_{L_2^r \times R^n} \to \min, \quad w \in W,$$

where $w^* \in L_2^r \times R^n$. Then

$$\hat{w}(\varepsilon, y_\delta, w^*) \to \{w : Aw = y, \ w \in W\}$$

$\varepsilon \to 0$, $\delta^2/\varepsilon \to 0$.

The proof is given in [17].

If F is uniformly continuous on every bounded set, then $\hat{z} = F\hat{w}(\varepsilon, y_\delta, w^*) \to Z(y)$ in the metric of Z. Operator F is, obviously, uniformly continuous, if $Fw = w$, $Fw = u(\cdot)$, $Fw = x^0$. For $Fw = x(\cdot, u(\cdot), x^0)$, $Fw = x(t_1, u(\cdot), x^0)$ F is uniformly continuous if the control system is of the form (2.4), and the assumptions of Lemma 2.1 are fulfilled.

4. Estimation of the value of a linear functional from given values of n other functionals

Let us consider as an illustrative example the following problem [19]. Estimate the value of a linear continuous functional $z = \langle g_0, w \rangle$, $w \in X$, assuming the values y_i of n other linear continuous functionals

$$\langle g_i, w \rangle = y_i, \quad i = 1, \ldots, n. \tag{4.1}$$

be given, and assuming that $\langle w, w \rangle \leqq \mu^2$ (here X is a Hilbert space).

This problem is obviously a special case of the problem from paragraph 2 with $Y = R^n$, $Z = R$, and operators $A : X \to R^n$, $F : X \to R$ defined by the equations

$$Aw = (\langle g_1, w \rangle, \ldots, \langle g_n, w \rangle)^T, \quad Fw = \langle g_0, w \rangle, \quad y = (y_1, \ldots, y_n)^T,$$

$$W = \{w : \langle w, w \rangle \leqq \mu^2\}.$$

Here set $Z(y)$ is an interval of R. We assume that y is given exactly, hence $\delta = 0$ and also $\alpha = 0$. Domain (interval) $Z_{\varepsilon,0}(y)$ defined by (2.3) ($w^* = 0$, $\alpha = 0$) approximates from above the interval $Z(y)$. Representing $Z_{\varepsilon,0}(y)$ as $\{z : |z - \hat{z}(\varepsilon)| \leqq R(\varepsilon)\}$, where $\hat{z}(\varepsilon)$ is the center of the interval, we may define $R(\varepsilon)$ from the equality

$$\rho(1 \mid Z_{\varepsilon,0}(y)) = \hat{z}(\varepsilon) + R(\varepsilon).$$

Values $\hat{z}(\varepsilon)$, $R(\varepsilon)$ may be obtained from Theorem 3.1. According to the theorem

$$\rho(1 \mid Z_{\varepsilon,0}(y)) = \hat{z}(\varepsilon) + (\mu^2 - \sigma)^{1/2} (\langle 1, FC^{-1}F^*1 \rangle)^{1/2}$$

where $F^*1 = g_0$. Let us represent C^{-1} by the formula $C^{-1} = I - A^*(AA^* + \varepsilon I)^{-1}$. Obviously, $AA^* = K$, where K is Gramm's matrix of the system

$$\{g_i, \quad i = 1, \ldots, n\}, \quad (K = \| \langle g_i, g_j \rangle \|), \quad Ag_0 = (\langle g_0, g_1 \rangle, \ldots, \langle g_0, g_1 \rangle)^T$$

hence $\langle F^*1, C^{-1}F^*1\rangle = k_{00} - k_0^T(K+\varepsilon I)^{-1}k_0$, where $k_0 = Ag_0$, $k_{00} = \langle g_0, g_0\rangle$.

It follows from Theorem 3.1 that

$$\hat{z}(\varepsilon) = F\hat{w} = \langle g_0, A^*(AA^*+\varepsilon I)^{-1}y\rangle = k_0^T(K+\varepsilon I)^{-1}y.$$

Finally

$$\sigma = \langle y, (AA^*+\varepsilon I)^{-1}y\rangle = y^T(K+\varepsilon I)^{-1}y.$$

Thus

$$R(\varepsilon) = (\mu^2 - y^T(K+\varepsilon I)^{-1}y)^{1/2}(k_{00} - k_0^T(K+\varepsilon I)^{-1}\cdot k_0)^{1/2},$$

and $Z_{\varepsilon,0}(y) = \{z : |z-\hat{z}(\varepsilon)| \leqq R(\varepsilon)\}$ approximates from above as $\varepsilon \to 0$, the interval $Z(y) = \{z : |z-\hat{z}| \leqq R\}$. This implies the statement of [19] (for the real space X).

5. Identification of parameters of linear control systems

Let us consider the problem of identification of the parameters of a stationary control system

$$dx/dt = Ax + bu, \quad x(0) = 0, \tag{5.1}$$

on the interval [0, 1] ($x \in R^n$, u is scalar control). Let us assume that vector $x(t)$ is accessible for measurements. It is necessary to reconstruct matrix A and vector b from the results of measurements of $x(t)$. It may be shown that, if $u \neq 0$ (as an element of L_2), then the necessary and sufficient condition of identifiability of A, b from the results of exact measurements of $x(t)$ is the complete controllability of system (5.1). Denoting by w vector $(a_{i1}, \ldots, a_{in}, b_i)$, $i \in \overline{1, n}$, we get the following infinite system of equations

$$x_i(t) = \sum_{j=1}^{n+1} w_j \rho_j(t), \quad 0 \leqq t \leqq 1, \tag{5.2}$$

where

$$\rho_j(t) = \int_0^t x_j(\tau)d\tau, \quad j = 1, \ldots, n, \ \rho_{n+1}(t) = \int_0^t u(\tau)d\tau.$$

System (5.2) is, in general, inconsistent, because of the errors of measurements. Let us define operator $A : R^{n+1} \to L_2$ as follows

$$(Aw)(t) = \sum_{j=1}^{n+1} w_j \rho_j(t), \quad 0 \leqq t \leqq 1.$$

Let us assume that $x_i(t)$ is measured with an error $\eta_i(t)$ where

$$\int \eta_i(t)dt \leqq \delta_i^2,$$

and $u(t)$ is known exactly. Then operator A will be given with an error non exceeding $\mu\|\delta\|$, where $\delta = (\delta_1, \ldots, \delta_n)^T$.

Consider the domain $Z_{\varepsilon,\alpha}(y_\delta)$ determined by (2.3). It contains the set of the exact values of w if $(\delta_i+\mu\|\delta\|)^2/\varepsilon \leqq \alpha$, and converges to $Z(y)$ as $\varepsilon, \alpha \to 0$. The analytical description of $Z_{\varepsilon,\alpha}(y)$ may be given by the following relation, arising from Theorem 3.1

$$Z_{\varepsilon,\alpha}(y)=\{w: \|w-\hat{w}\|^2+1/\varepsilon(w-\hat{w})^T G(w-\hat{w}) \leqq v^2(\varepsilon,\alpha)\},$$

where G is Gramm's matrix of the system $\rho_1(\cdot), \ldots, \rho_{n+1}(\cdot)$,

$$\hat{w}=(G+\varepsilon I)^{-1}v,\ v=(\int \rho_1 x_i dt, \ldots, \int \rho_{n+1} x_i dt)^T,$$

$$v^2(\varepsilon,\alpha)=\mu^2+\alpha-1/\varepsilon \int x_j^2(t)dt+1/\varepsilon(G+\varepsilon I)^{-1}v.$$

If system (5.1) is completely controllable, then $Z_{\varepsilon,\alpha}(y_\delta)$ converges to a unique point

$$\hat{w}=G^{-1}v|_{\eta_i(\cdot)=0},$$

as $\varepsilon \to 0$, $\alpha \to 0$, $(\delta_i+\mu\|\delta\|)^2/\varepsilon \leqq \alpha$.

A detailed description of the informational domains for the linear-quadratic identification problem (for discrete-time case) is presented in [20].

References

1. *Krasovskii, N. N.*, Theory of control of motion. M. Nauka, 1968, 476 pp.
2. *Kurzhanskii, A. B.*, Control and observation under conditions of uncertainty. M.: Nauka, 1977, 392 pp.
3. *Tikhonov, A. N., Arsenin, W. J.*, Methods of solving ill-posed problems. M.: Nauka, 1975, 224 pp.
4. *Ivanov, V. K., Vasin, V. V., Tanana, P. V.*, Theory of linear ill-posed problems and its applications. M.: Nauka, 1978, 206 pp.
5. *Gusev, M. I.*, On a certain class of inverse problems in control system dynamics. Lecture Notes in Control and Information Sciences. Proceedings of the international conference on stochastic optimisation. Kiev, 1984, S.–V., 1986, pp. 650–656.
6. *Gusev, M. I., Kurzhanskii, A. B.*, On the inverse problems of control system dynamics. 12th IFIP conference on system modelling and optimisation, Budapest, 1985, pp. 132–133.
7. *Gusev, M. I., Kurzhanskii, A. B.*, On the stable methods of solution of inverse problems in control system dynamics. Proceedings of 6th All Union Congress on theoretical and applied mechanics, Tashkent, 1986, pp. 234.
8. *Galiullin, A. S.*, Inverse problems of dynamics. M.: Nauka, 1981.
9. *Nikolskii, M. S.*, On perfectly observable systems. Different. Uravnen. *7*, 1971, pp. 631–638.
10. *Hirschorn, R. M.*, Invertibility of nonlinear control system. SIAM J. Control & Optim., 1979, *17*, pp. 289–297.
11. *Nijmeijer, H.*, Invertibility of affine nonlinear control systems: a geometric approach. Systems Control Let., 1982, **2**, *3*, pp. 163–186.
12. *Kirin, N. E.*, Sequential estimate methods for the optimization of control systems. Leningrad State University, 1975, 160 pp.
13. *Anan'ev, B. I.*, Minimax problems of filtering and correction of motion of control systems. Dissertation. Institute of math. and mech., Sverdlovsk, 1977.
14. *Krjazimskii, A. V., Osipov, Ju. S.*, On the positional modelling of control in dynamic systems. Izvestia Akademii Nauk. SSSR, Technicheskaya Kibernetika, 1983, *2*, pp. 51–60.

15. *Kurzhanskii, A. B.*, Dynamic control system estimation under uncertainty conditions. 1. Problems of Control and Information Theory, 1980, **9,** *6*, pp. 593–606.
16. *Kurzhanskii, A. B.*, Dynamic control system estimation under uncertainty conditions 2. Problems of Control and Information Theory, 1981, **10,** *1*, pp. 33–42.
17. *Anikin, S. A., Gusev, M. I.*, Estimation of disturbing forces from the results of measurements of motion. In "Guaranteed estimation and control problems" Sverdlovsk: Urals Cent. of Academy of Science, 1986, pp. 19–30.
18. *Athans, M., Falb, P. L.*, Optimal control. McGraw-Hill, 1964.
19. *Belov, Ju. A., Kasianjuk, V. S.*, On optimal estimate of linear functional from the values of *n* other functionals in Hilbert space. Doklady Acad. Nauk., 1986, **289,** *5*, pp. 1033–1035.
20. *Isakov, A. I.*, On some problems of quadratic identification for linear, discrete-time systems. Prikl. matem. i mech., 1981, **45,** *2*, pp. 108–113.

Об устойчивости решения обратных задач динамики управляемых систем

М. И. ГУСЕВ

(Свердловск)

Работа посвящена исследованию ряда вопросов, связанных с устойчивостью решения некоторых типичных обратных задач динамики. Рассмотрена общая схема решения, в рамках которой обратная задача трактуется как задача оценивания значения некоторого линейного оператора в Банаховом пространстве при операторных ограничениях. Описаны устойчивые относительно ошибок в исходных данных процедуры решения, основанные на использовании конструкций, развитых в теории управления и наблюдения в условиях неопределенности и теории некорректных задач. Даны приложения к задачам оценивания возмущений на входе управляемой системы и задаче идентификации параметров.

M. I. Gusev
Institute of Math. & Mech.
of the Ural Branch of
the USSR Academy of Sciences
S. Kovalevskoi st., 16
620066 Sverdlovsk
USSR

Problems of Control and Information Theory, Vol. 17 (5), pp. 311—317 (1988)

INCREASING THE LINEAR COMPLEXITY OF *m*-SEQUENCES USING PSEUDORANDOM EXPONENTIATION

I. VAJDA, J. LANDSMANN

(*Budapest*)

(Received September 1, 1987)

The linear complexity is calculated for sequences generated by discrete exponentiation of elements of *m*-sequences over $GF(q)$, $q=2^m$, $m>1$, where the consecutive exponents are generated by another *m*-sequence over $GF(\bar{q})$, $\bar{q}=2^{\bar{m}}$, $m>\bar{m}$. As a special case, the interleaving of exponentiated *m*-sequences using constant exponents is also considered.

1. Introduction

Pseudorandom sequences generated by linear feedback shift-registers have widespread applications. The maximal period sequences (*m*-sequences) produced by linear recursions with primitive characteristic polynomials have good pseudorandom properties, nevertheless, they are very bad from the linear complexity point of view. Pseudorandom periodical sequences used as running-key sequences over $GF(2)$ or signature sequences in slow frequency-hopped spread-spectrum systems over $GF(q)$, $q>2$ must have a very large period and a large linear complexity [4], [5].

For the sequence $\tilde{b}=b_0, b_1, b_2, \ldots$ of period N the linear complexity is defined as the smallest nonnegative integer L_b, for which there exist constants $c_1, c_2, \ldots, c_{L_b}$ such that

$$b_i+c_1b_{i-1}+\ldots+c_{L_b}b_{i-L_b}=0, \qquad i\geqq L_b. \tag{1}$$

Here, obviously, $L_b\leqq N$.

The generally used structures for producing such sequences are based on linear feedback shift-registers, generating *m*-sequences over $GF(q)$ (i.e. having a primitive connection polynomial of degree l_0), whose register content $r_1, r_2, \ldots, r_{l_0}$ is mapped into a sequence over $GF(q)$ by some Boolean function $f(r_1, r_2, \ldots, r_{l_0})$ [3], [5].

The *m*-sequence $\tilde{d}$ has period $N=q^{l_0}-1$. The output sequence $\tilde{b}$, after the nonlinear feed-forward mapping, can be represented uniquely by

$$b_i=\sum_{n=1}^{L^*} B_n(\gamma_n)^i, \qquad i\geqq 0 \tag{2}$$

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

where B_n and γ_n are nonzero elements from the field $GF(q^{l_0})$; γ_n's are different, $1 \leqq n \leqq L^*$, furthermore, $L^* = L_b$. It is also well known that the elements γ_n, $1 \leqq n \leqq L_b$ are the roots of the characteristic polynomial of the linear recursion corresponding to the sequence $\tilde{b}$. The m-sequence $\tilde{d}$ also can be given in the representation of type (2), namely

$$d_i = Tr_q^{q^{l_0}}(\delta \beta^i), \qquad i \geqq 0, \tag{3}$$

where the trace polynomial $Tr_q^{q^s}(z) \triangleq \sum_{n=0}^{s-1} z^{q^n} : GF(q^s) \to GF(q)$ is introduced and β is a primitive element of $GF(q^{l_0})$, furthermore, δ is a nonzero element of $GF(q^{l_0})$ which uniquely corresponds to the actual phase-shift of the m-sequence. In this paper we analyse a generation method differing from the one in Fig. 1 in the use of the nonlinear feed-forward mapping, while the power series (or trace) representation (2) [or (3)] is applied.

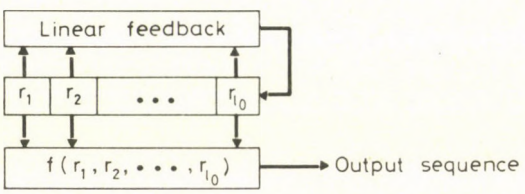


Fig. 1. The structure of the sequence generator

2. Discrete exponentiation as nonlinear mapping

We have considered the discrete exponentiation of the elements of an m-sequence as a possible nonlinear mapping, where the exponents are produced by another m-sequence.

The elements of the m-sequence $\tilde{d}$ over $GF(q)$, $q = 2^m$, $m > 1$,

$$\tilde{d} = d_0, d_1, \ldots, d_i, \ldots$$

are raised to the powers given by the corresponding elements of the sequence of exponents

$$\tilde{e} = e_0, e_1, \ldots, e_i, \ldots, \qquad 1 \leqq e_i \leqq q-2,$$

resulting in the output sequence

$$\tilde{b} = d_0^{e_0}, d_1^{e_1}, \ldots, d_i^{e_i}, \ldots. \tag{4}$$

The special case of $\tilde{e} = j, j, \ldots, j, \ldots, 1 \leqq j \leqq q-2$ has been considered in [4], namely the linear complexity of $\tilde{b}$ can be given by

$$L_b = l_0^{w(j)}, \qquad 1 \leqq j \leqq q-2, \tag{5}$$

where $w(j)$ denotes the binary weight of the integer j. Formula (5) shows that considerable increase in linear complexity can be reached using simple exponentiation.

As a plausible generalization one may take the case of constant exponents, $\tilde{e}=1, 2, \ldots, q-2, 1, 2, \ldots, q-2, \ldots$, which leads to a sequence $\tilde{b}$ of powers, which is an interleaved version of subsequences $\tilde{b}^{(j)}$, $j=1, 2, \ldots, q-2$, where elements of $\tilde{d}$ are raised to the same constant power, i.e.

$$b_k^{(j)}=b_{k(q-2)+j-1} \qquad k\geqq 0. \tag{6}$$

Taking $\delta=1$ in (3) we do not loose the generality of calculating the linear complexity, so we may suppose

$$b_k^{(j)}=[Tr_q^{q^{l_0}}(\beta^{j-1}\gamma^k)]^j \qquad k\geqq 0, \tag{7}$$

where $\gamma=\beta^{q-2}$ and $1\leqq j\leqq q-2$. The proper selection of q for reaching large period is to choose $q-2$ and $q^{l_0}-1$ as relative primes. Therefore, g.c.d. $(l_0, m-1)=1$ is supposed. It follows consequently, in this case, γ is also a primitive element of $GF(q^{l_0})$.

Let us denote the characteristic polynomials of the minimal degree linear recursions corresponding to sequences $\tilde{b}$ and $\tilde{b}^{(j)}$ by $g(x)$ and $g^{(j)}(x)$, respectively. It follows from (5) that deg $g^{(j)}(x)=l_0^{w(j)}$, $1\leqq j\leqq q-2$. Let us introduce the formal power series representation

$$\tilde{b}(x)=\sum_{i=0}^{\infty} b_i x^i=\sum_{i=0}^{\infty} x^j \tilde{b}^{(j)}(x^{q-2}). \tag{8}$$

The sequence corresponding to $\tilde{b}^{(j)}(x^{q-2})$ can be generated by $g^{(j)}(x^{q-2})$ with minimal degree and the same is true for $x^j b^{(j)}(x^{q-2})$. Consequently, the minimal degree monic polynomial $g(x)$ over $GF(q)$, generating the sequence $\tilde{b}$ can be given as follows

$$g(x)=\hat{g}(x^{q-2}), \tag{9}$$

where $\hat{g}(x)=$l.c.m. $\{g^{(1)}(x), g^{(2)}(x), \ldots, g^{(q-2)}(x)\}$. Below it is shown that the polynomials $g^{(1)}(x), g^{(2)}(x), \ldots, g^{(q-2)}(x)$ have no common factors, therefore

$$g(x)=\sum_{j=1}^{q-2} g^{(j)}(x). \tag{10}$$

Using the technique leading to result (5) in [4] one gets

$$[Tr_q^{q^{l_0}}(\gamma^k)]^j \prod_{n=1}^{w(j)} [Tr_q^{q^{l_0}}(\gamma^k)]^{2^{j_n}}= \\ =\sum_{h_1=0}^{l_0-1}\sum_{h_2=0}^{l_0-1}\cdots\sum_{h_{w(j)}=0}^{l_0-1}\gamma^{kc(h,j)},$$

where

$$c(h,j)=\sum_{n=1}^{w(j)} 2^{mh_n+j_n}, \tag{11}$$

$h=(h_1, h_2, \ldots, h_{w(j)})$, $0 \leqq j_1 < j_2 < \ldots < j_m < m$. It must be shown that for $j \neq j'$, $1 \leqq j, j' \leqq q-2$ there is no pair of vectors h, h', where $h=(h_1, h_2, \ldots, h_{w(j)})$, $h'=(h'_1, h'_2, \ldots, \ldots, h'_{w(j)})$, $0 \leqq h_n, h'_n \leqq l_0$, for which $c(h, j)=c(h', j')$, i.e. for which

$$\sum_{n=1}^{w(j)} 2^{mh_n+j_n} = \sum_{n=1}^{w(j')} 2^{mh'_n+j'_n} \tag{12}$$

would be true. The sums on both sides of (12) are such numbers in binary representation, where there are ones at positions mh_n+j_n and $mh'_n+j'_n$, respectively. Because $j \neq j'$, the corresponding vectors $j_1, j_2, \ldots, j_{wj}$ and $j'_1, j'_2, \ldots, j'_{w(j')}$ are also different, which contradicts (12). These considerations conclude into.

Theorem 1. When the period of the sequence of exponents is an arbitrary permutation of the integers $1, 2, \ldots, q-2$ the sequence of powers defined by formulae (3) and (4) has period $N=(q-2)(q^{l_0}-1)$ and linear complexity

$$L_b=(q-2)\sum_{j=1}^{q-2} l_0^{w(j)}, \tag{13}$$

if g.c.d. $(l_0, m-1)=1$, $q=2^m$.

Example. If $m=4$, $l_0=2$ then $N=3570$ and $L_b=896$, which result was checked by the Berlekamp–Massey shift-register synthesis algorithm.

The m-sequences over $GF(q)$ have interesting properties rarely emphasized in the literature. Namely, the zeroes have a periodicity with period $(q^{l_0}-1)/(q-1)$ and the indicator sequences of different characters in the sequence are shifted versions of each other. The exponentiation cannot hide the vulnerable structure of the periodically appearing zeroes. One way to circumvent this problem is the following. The one-character delayed version of the sequence b is added to itself, i.e.

$$b^*=\{b_i+b_{i+1}\}_{i=0}^{\infty} \tag{14}$$

is the modified sequence.

Lemma. $L_b^*=L_b$.

Proof. Using representation (2) we get

$$b_i^*=b_i+b_{i+1}=\sum_{n=1}^{L_b} B_n(1+\gamma_n)(\gamma_n)^i.$$

Observing that $\gamma_n \neq 1$, $1 \leqq n \leqq L_b$, by formula (11), the result follows.

3. The case of pseudorandom exponents

Let the sequence $\tilde{c}$ defined by

$$\tilde{c}=Tr_{\bar{q}}^{\bar{q}^{l_1}}(\alpha^i)\}_{i=0}^{\infty}$$

be another m-sequence over $GF(\bar{q})$, where α is a primitive element of $GF(\bar{q})$, $\bar{q}=2^{\bar{m}}$, $m>\bar{m}$. Suppose that g.c.d. $(ml_0, \bar{m}l_1)=1$. If this assumption is fulfilled, then the fields

$GF(q^{l_0})$ and $GF(\bar{q}^{l_1})$ have only $GF(2)$ as common subfield and it can be easily shown that $\delta_1\delta'_1 \neq \delta_2\delta'_2$, $\delta_1, \delta_2 \in GF(q^{l_0})$, δ'_1, $\delta'_2 \in GF(\bar{q}^{l_1})$ if

$$\delta_1 \neq \delta_2 \quad \text{and/or} \quad \delta'_1 \neq \delta'_2.$$

Let the functions $F_k: GF(q)$ be defined by $F_k(x) = x^{i_k}$, where $i_k \in \{1, 2, \ldots, q-2\}$ and $i_k \neq i_{k'}$ if $k \neq k'$, $1 \leqq k, k' \leqq \bar{q}$. The sequence of powers is generated as follows

$$b_i = \sum_{k=1}^{\bar{q}} S_{h(k)}(c_i) F_k(d_i) \tag{15}$$

where $h(k)$ is a one-to-one mapping from $\{1, 2, \ldots, \bar{q}\}$ to $GF(\bar{q})$ and the functions $S_{h(k)}: GF(\bar{q}) \to GF(2)$, $1 \leqq k \leqq \bar{q}$ are defined by

$$S_{h(k)}(x) = \begin{cases} 1, & \text{if } x = h(k) \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

This means that the functions S make selection from the terms in the sum (15), namely, $b_i = Fh^{-1}(k)(d_i)$. In this way we generate a sequence of powers corresponding to formula (4) using a pseudorandomly generated sequence of exponents.

Using the polynomials

$$f_\rho(x) = \begin{cases} \rho x^{N'-1} + \rho^2 x^{N'-2} + \ldots + \rho^{N'-1} x + 1, & \text{if } \rho \neq 0 \\ x^{N'} - 1 \quad , & \text{if } \rho = 0,\ N' = \bar{q} - 1 \end{cases} \tag{17}$$

we get the desired selection functions, because $(x-\rho) f_\rho(x) = \rho(x^{N'} - 1)$ for $\rho \neq 0$ and $f_\rho(\rho) = 1$ for arbitrary $\rho \in GF(\bar{q})$.

Applying the sum of power series representation of type (2) we get

$$\begin{aligned} S_{h(k)}(c_i) &= \sum_{n_1=1}^{L_{S_k}} A_{n_1}^{(k)} (\gamma_{n_1}^{(k)})^i \\ F_k(d_i) &= \sum_{n_2=1}^{L_{F_k}} B_{n_2}^{(k)} (\delta_{n_2}^{(k)})^i \end{aligned} \tag{18}$$

$1 \leqq k \leqq \bar{q}$, where L_{S_k} and L_{F_k} stand for the linear complexities of the corresponding sequences and $A_{n_1}^{(k)}$, $\gamma_{n_1}^{(k)} \in GF(q^{l_0})$, $B_{n_2}^{(k)}$, $\delta_{n_2}^{(k)} \in GF(\bar{q}^{l_1})$. Using the assumption g.c.d. $(ml_0, \bar{m}l_1) = 1$ we conclude that in

$$(H_i^{(k)} \triangleq) S_{h(k)}(c_i) F_k(d_i) = \sum_{n_1=1}^{L_{S_k}} \sum_{n_2=1}^{L_{F_k}} A_{n_1}^{(k)} B_{n_2}^{(k)} (\gamma_{n_1}^{(k)} \delta_{n_2}^{(k)})^i \tag{19}$$

the products $\gamma_{n_1}^{(k)} \delta_{n_2}^{(k)}$, $1 \leqq n_1 \leqq L_{S_k}$, $1 \leqq n_2 \leqq L_{F_k}$ are different elements of $GF(q^{l_0} \bar{q}^{l_1})$, so sequence $\tilde{H}^{(k)}$ has linear complexity $L_{S_k} L_{F_k}$.

We have seen in the derivation of Theorem 1 that the characteristic polynomials of sequences of powers have disjunct root sets when the elements of an m-sequence over $GF(q)$ are raised to a constant exponent j, $1 \leqq j \leqq q-2$, so this is the

case with the sequences $\{F_k(d_i)\}_{i=0}^{\infty}$, $1 \leqq k \leqq \bar{q}$, too. From this it follows that

$$\gamma_{n_1}^{(k)} \delta_{n_2}^{(k)} \neq \gamma_{n_1'}^{(k')} \delta_{n_2'}^{(k')}$$

if $k \neq k'$ and n_1, n_2, n_1', n_2' are arbitrary in their range. This results in

Theorem 2. Using the above notations, the sequence $\tilde{b}$ given by (16) has linear complexity

$$L_b = \sum_{k=1}^{\bar{q}} L_{S_k} L_{F_k}, \tag{20}$$

where

$$L_{F_k} = l_0^{w(i_k)},$$

and

$$L_{S_k} = \begin{cases} \sum_{j=1}^{N'-1} l_1^{w(j)} + 1 & \text{if} \quad h^{-1}(k) \neq 0 \\ \dfrac{\bar{q}^{l_1} - 1}{\bar{q} - 1} & , \quad \text{if} \quad h^{-1}(k) = 0. \end{cases} \tag{21}$$

In (21) we have used the fact that the binary sequence $\{c_i^{N'} - 1\}_{i=0}^{\infty}$ has ones only at zeroes of the sequence $\tilde{c}_1$ taking into account the above mentioned periodicity of the zeroes in an m-sequence over $GF(\bar{q})$.

Applying the same reasoning used in the proof of the Lemma it can be shown that the (14)-type modification can be used again without changing the linear complexity.

References

1. *Golomb, S. W.*, Shift-Register Sequences. San Francisco: Holden Day, 1967.
2. *Zierler, N.*, Linear recurring sequences, J. SIAM, **7,** *1*, March 1959.
3. *Key, E. L.*, An analysis of the structure and complexity of nonlinear binary sequence generators, IEEE Trans. Inform. Theory, **IT-22,** *6*, Nov. 1976.
4. *Simon, M. K., Omura, J. K., Scholtz, R. A., Levitt, B. K.*, Spread Spectrum Communications, **I,** Ch. 5, Computer Science Press 1984.
5. *Rueppel, R. A.*, New Approaches to Stream Ciphers. Doctoral Thesis. ETH Zürich, 1984.

Повышение линейной сложности $(p-n)$-последовательностей с возведением в степень

И. ВАЙДА, Й. ЛАНДСМАНН

(Будапешт)

Определяется линейная сложность для последовательностей степенями элементов $(p-n)$-последовательностей над полем $GF(q)$, $q=2^m$, $m>1$, где последовательные элементы генерируются элементами другой последовательности. Рассматривается частный случай чередование степеней последовательности с постоянным экспоненталии.

I. Vajda
J. Landsmann
Technical University of Budapest
H-1111 Budapest, Stoczek u. 2.
Hungary

Problems of Control and Information Theory, Vol. 17 (5), pp. 319—325 (1988)

AN ENTROPIC TEST OF UNIFORMITY AND OPTIMALITY OF CODEBOOKS

J. FEISTAUEROVÁ

(*Prague*)

(Received October 1, 1987)

In this paper we introduce a statistical test of uniformity of a discrete probability distribution based on the concept of entropy. The uniformity of the actual use of codewords in practical applications is quite a reasonable criterion for the optimality of a codebook. Our test is proposed mainly for this purpose and its application to a codebook for speech compression is demonstrated.

We consider a finite probability space $(\mathcal{X}, P)$, where $\mathcal{X}=\{1, 2, \ldots, m\}$ and P is a vector $(p_1, p_2, \ldots, p_m)$ of probabilities on $\mathcal{X}$. We consider an independent random sample $\mathbf{X}=(X_1, X_2, \ldots, X_n)$ from $(\mathcal{X}, P)$, and for every sample size $n=1, 2, \ldots$ we consider a random probability distribution $P_n=(p_{n1}, p_{n2}, \ldots, p_{nm})$ on $\mathcal{X}$ defined by

$$p_{ni}=\frac{\#\, j\in\{1, 2, \ldots, n\}: X_j=i}{n}, \qquad i=1, \ldots, m.$$

For each realization $\mathbf{x}=(x_1, x_2, \ldots, x_n)\in\mathcal{X}^n$ of the random vector $\mathbf{X}$ one obtains one possible realization of P_n.

We further assume $p_i>0$ for every $i=1, 2, \ldots, m$ and consider the I-divergences

$$I(P_n\|P)=\sum_{i=1}^{m} p_{ni}\ln\frac{p_{ni}}{p_i},$$

where the summand is assumed to be zero if $p_{ni}=0$, and the χ^2-divergences

$$\chi^2(P_n\|P)=\sum_{i=1}^{m}\frac{(p_{ni}-p_i)^2}{p_i}.$$

The expression

$$n\chi^2(P_n\|P)=\sum_{i=1}^{m}\frac{(np_{ni}-np_i)^2}{np_i}$$

is the well known χ^2-statistic for which it holds under the assumptions considered above (cf. [1], Chapter XXX).

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

Lemma 1. For $n \to \infty$ it holds

$$n\chi^2(P_n \| P) \xrightarrow{w} \chi^2(m-1),$$

where $\xrightarrow{w}$ denotes the weak convergence (the convergence in distribution) and $\chi^2(k)$, $k = 1, 2, \ldots$ denotes the random variable with χ^2-distribution and k degrees of freedom.

This result enables to use the χ^2-divergence in the well-known goodness-of-fit statistical tests.

The key role in our derivation of an entropic test of uniformity of the discrete probability model $P = (p_1, \ldots, p_m)$ plays the following result which is an analogue to Lemma 1 formulated for the I-statistic

$$nI(P_n \| P) = \sum_{i=1}^{m} np_{ni} \ln \frac{np_{ni}}{np_i}.$$

Lemma 2. For $n \to \infty$ it holds

$$2nI(P_n \| P) \xrightarrow{w} \chi^2(m-1),$$

where the meaning of $\xrightarrow{w}$ and $\chi^2(m-1)$ is the same as in Lemma 1.

This result was established in Theorem 16.3 of Vajda [3]. Its proof was reduced there to a demonstration of that the I-statistic in fact coincides with the generalized log-likelihood ratio and to a subsequent application of a known theorem about asymptotic distribution of the generalized log-likelihood ratio. Here we present an alternative proof which is based on Lemma 1.

Proof of Lemma 2. Let us consider the function $f(u) = u \ln u$, $u > 0$.

It holds

$$f'(u) = \ln u + 1$$

$$f''(u) = \frac{1}{u}$$

$$f'''(u) = -\frac{1}{u^2}$$

so that, by the Taylor theorem, it holds for every $u > 0$

$$f(u) = u - 1 + \frac{1}{2}(u-1)^2 - \frac{(u-1)^3}{6\xi^2}$$

where ξ is between 1 and u. Taking into account the obvious identity

$$I(P_n \| P) = \sum_{i=1}^{m} p_i f\left(\frac{p_{ni}}{p_i}\right)$$

we can write, substituting p_{ni}/p_i for u in the formula for $f(u)$ above,

$$2nI(P_n \| P) = 2n \sum_{i=1}^{m} (p_{ni} - p_i) + n\chi^2(P_n \| P) -$$

$$- \frac{n}{3} \sum_{i=1}^{m} \frac{(p_{ni} - p_i)^3}{\xi_{ni}^2 p_i^2} =$$

$$= n\chi^2(P_n \| P) - \frac{1}{3} \sum_{i=1}^{m} \frac{[n^{\frac{1}{3}}(p_{ni} - p_i)]^3}{\xi_{ni}^2 p_i^2}.$$

Since, by the law of large numbers, p_{ni} tends in probability to p_i for $i = 1, 2, \ldots, m$, ξ_{ni}^2 tends in probability to 1. Further, it can be easily deduced, e.g. from the central limit theorem, that $n^{1/3}(p_{ni} - p_i)$ tends weakly to the constant 0 for $i = 1, 2, \ldots, m$. Hence it follows from the last identity and from Lemma 1 that Lemma 2 holds. □

Let us consider the entropy function

$$H(P) = -\sum_{i=1}^{m} p_i \ln p_i \quad \text{(entropy in nats)}.$$

It is well known that this function is nonnegative and bounded above by $\ln m$ and that the equality

$$H(P) = \ln m$$

holds iff P is uniform, i.e. iff

$$p_1 = p_2 = \ldots = p_m = \frac{1}{m}.$$

The question is whether one can use the statistical evidence represented by the nonnegative difference

$$\ln m - H(P_n)$$

in statistical testing of the simple hypothesis that the information source $(\mathcal{X}, P)$ which generates the data $\mathbf{x} = (x_1, \ldots, x_n)$ leading to P_n is uniform. The difference $\ln m - H(P_n)$ is known as the redundancy of the information source $(\mathcal{X}, P_n)$. Sometimes the redundancy is defined rather in percents by the expression $100[\ln m - H(P_n)]$. The zero redundancy is a desired property of certain information sources, in particular of codebooks as it is discussed in the end of this paper.

Thus, the question is whether the sample redundancy $\ln m - H(P_n)$ enables us to decide (in accordance with the common statistical criteria of optimality) whether the messages from the source under observation are distributed uniformly. Our answer is positive and this answer follows from the next statement.

Theorem. Under the above conditions the statistic

$$R_n = 2n[\ln m - H(P_n)]$$

tends in distribution, as $n \to \infty$, to the random variable $\chi^2(m-1)$.

Proof. If P is uniform then it is easy to verify that it holds

$$I(P_n || P) = \ln m - H(P_n).$$

Hence Theorem 1 follows from Lemma 2. □

Using the theorem we can describe the asymptotically α-level test of uniformity as follows: Reject the uniformity hypothesis if $R_n > \chi^2_{(1-\alpha)}(m-1)$, where $\chi^2_{(1-\alpha)}(m-1)$ denotes the $(1-\alpha)$-quantile of the above defined random variable $\chi^2(m-1)$. These quantiles are tabulated for $\alpha = 0.001$, 0.01 and 0.05 and for $m = 2, \ldots, 101$ e.g. in [2]. For $m > 101$ (in fact for $m > 30$) we can use the formula

$$\chi^2_{(1-\alpha)}(m-1) \doteq \frac{1}{2}[\sqrt{2m-3} + \Phi_{(1-\alpha)}]^2,$$

where $\Phi_{(1-\alpha)}$ is the $(1-\alpha)$-quantile of the standard normal random variable. The restriction on the model

$$\min \{np_1, \ldots, np_m\} \geqq 5$$

considered in practical applications of the general goodness-of-fit χ^2-test (cf. [5]) reduces in our case to the inequality

$$n \geqq 5m.$$

This condition should be observed in practical applications of our test.

In the rest of this paper we consider practical applications of the redundancy as one possible criterium of optimality of a codebook for coding an arbitrary information source described by a probability space $(A, \mathscr{A}, Q)$. Let us suppose that there is a distance (not necessarily a metric) ρ in the source "alphabet" A and that there is given a finite subset $\mathscr{X} \subset A$ called a codebook. The elements of $\mathscr{X}$ can be enumerated by $1, 2, \ldots, m$ as above, since the topological properties of $\mathscr{X}$ are playing no explicite role. Let us suppose that the source is encoded onto $\mathscr{X}$ by means of the minimum distance rule. If the distance is measurable, then encoding plus Q define a probability distribution P on $\mathscr{X}$, i.e. the source and the encoder define a new source,

an output source $(\mathcal{X}, P)$. The optimality of the codebook may be viewed from different points (cf. [4]), but one point of view is the redundancy $\ln m - H(P)$ of the codebook. According to this point of view, a codebook is good if the burden to represent the source A uniformly shared by all codewords, the total number m of which is limited in advance, in accordance with the capacity of the transmission channel. If in this situation there are codewords the frequencies of which differ by an order from frequencies of other codewords, then the codebook bedly represents the source alphabet and it does not seem to be optimum from the point of view of minimization of the average distance as well. The asymptotic equipartition property (see [6]) is promising to be a justification of the redundancy optimality criterion considered in this paper.

In the forthcoming applications of our test of uniformity we consider the first-kind error level $\alpha = 0.01$. It is to be noted that, in cases when this test fails to reject the uniformity hypothesis, absolutely nothing is exactly guaranteed as to the acceptance of the hypothesis. If the uniformity cannot be rejected at level $\alpha = 0.01$, then there is still possible that it can be rejected at a higher probability of error level α, say $\alpha = 0.05$. However, there is a reasonable hope that the unknown distribution P is at least quite close to the uniform, if it is not ideally uniform. If the sample size is very large then the distribution P may be very close to the uniform even if the uniformity as such is rejected by our test (cf. the Euclidean distance in Tables 1, 2).

In this paper we present some results obtained by verifying the quality of a CODEBOOK 86-2 designed for speech compression below the level 1000 bit/s. This 10-bit codebook was calculated from a data base containing 7898 vectors. It is organized into two hierarchical levels. It means that the book is divided into 32 pages and each page contains 32 codewords (thus the book contains altogether 1024 words). This structure was chosen on the basis of results published in [7], where it is shown that the structure 32–32 is optimal from many points of view. The pages, in fact, represent small codebooks. The procedure of encoding is the following: a) for each speech signal segment we determine the page by means of a minimum distance rule, b) on this page we find the codeword by means of the minimum distance rule.

In Table 1 we present the results obtained for the pages of CODEBOOK 86-2. Here under Euclideen distance ρ $(P$, uniform) we understood the expression

$$\rho(P, \text{uniform}) = \sqrt{\sum_{i=1}^{m} \left(p_{ni} - \frac{1}{m} \right)^2} .$$

In Table 2 the values of the same quantities are presented as in the preceeding one, but for all CODEBOOK 86-2 (i.e. for 1024 codewords).

Table 1. Codebook 86-2

Page	Sample size n	Entropy	Statistic R_n	Rejection of uniformity	ρ (P, uniform)
1	219	4.88	36.4318		0.0721
2	256	4.86	49.6848		0.0779
3	172	4.87	30.9975		0.0750
4	190	4.88	31.6075		0.0721
5	149	4.81	39.2460		0.0908
6	284	4.79	82.6786	yes	0.0954
7	298	4.85	61.9673	yes	0.0806
8	289	4.89	44.0703		0.0692
9	160	4.77	51.0156		0.0998
10	251	4.76	83.5103	yes	0.1020
11	256	4.87	46.1359		0.0750
12	300	4.85	62.3832	yes	0.0806
13	312	4.88	51.9029	yes	0.0721
14	257	4.88	42.7533		0.0721
15	232	4.88	38.5944		0.0721
16	267	4.89	40.7155		0.0690
17	345	4.89	52.6099	yes	0.0690
18	229	4.87	41.2699		0.0750
19	147	4.88	24.4542		0.0721
20	309	4.79	89.9567	yes	0.0954
21	289	4.72	112.1789	yes	0.1101
22	241	4.78	73.5013	yes	0.0976
23	242	4.96	13.4193		0.0416
24	218	4.82	54.3982	yes	0.0883
25	271	4.87	86.4077	yes	0.0998
26	196	4.93	19.0199		0.0551
27	292	4.87	93.1035	yes	0.0998
28	236	4.80	65.4331	yes	0.0931
29	221	4.93	21.4460		0.0551
30	294	4.83	69.2869	yes	0.0858
31	202	4.85	42.0047		0.0806
32	274	4.93	26.5891		0.0551

quantile $\chi^2_{0.95}(31) = 51.3718$

The entropy is expressed in binary units. It holds $\ln n - H_{\text{nat}}(P) = \ln 2(\log_2 n - H_{\text{bin}}(P))$

Table 2

Sample size n	7898
Entropy H	9.825
Statistics R_n	1916.06
$\rho(P_n$, unif.)	0.0138

quantile $\chi^2_{0.99}(1023) = 1130.39$

Thus, our test rejects the uniformity hypothesis at the significance level 0.01. But the statistic R_n exceeds relatively little the critical value $\chi^2_{0.99}(1023)$. Thus we may

conclude that the codebook, while probably not optimum, is very close to the optimum. This conclusion is supported also by the very small Euclidean distance $\rho(P_n, \text{unif.})$.

Acknowledgement

The author is grateful to I. Vajda from our Institute for his helpful advices and for the interesting and inspiring discussions on the subject.

References

1. *Cramér, H.*, Mathematical Methods of Statistics. Princeton Univ. Press, 1946.
2. *van der Waerden*, Mathematical Statistics. Springer, Berlin, 1969.
3. *Vajda, I.*, Information Theory and Statistical Decision. (in Czech), Alfa, Bratislava 1982.
4. *Vajda, I.*, Alphabet-oriented and user-oriented source coding theories. Kybernetika (1987), *5*.
5. *Anděl, J.*, Mathematical Statistics (in Czech). SNTL, Praha 1978.
6. *Perez, A.*, Sur la théorie de l'information dans le cas d'un alphabet abstrait. Transactions 1st Prague Conf. on Inf. Theory 1956, Praha 1957.
7. *Wong, D. Y. et al.*, An 800 bit/s vector quantization LPC vocoder, IEEE Trans. on Acoustics, Speech and Signal Processing, October 1982, No. *5*, pp. 770–780.

Энтропическая проверка равномерности и оптимальность кодових книг

Я. ФЕЙСТАУЕРОВА

(Прага)

В этой статье вводится статистическая проверка равномерности дискретного распределения вероятностей основанная на понятии энтропии. Равномерность использования кодовых слов в практических применениях является разумным критерием оптимальности кодовой книги. Наша проверка задумана именно для этих приложений. Продемонстрировано одно такое приложение на кодовую книге для сжатия речевого сигнала.

J. Feistauerová
Institute of Information Theory and Automation
Czechoslovak Academy of Sciences
182 08 Prague
Pod vodárenskou věží 4
Czechoslovakia

Problems of Control and Information Theory, Vol. 17 (5), pp. 327—331 (1988)

A SIMPLE AND GENERAL APPROACH TO THE DECIMATION OF FEEDBACK SHIFT-REGISTER SEQUENCES

H. NIEDERREITER

(Vienna)

(Received October 1, 1987)

A matrix-theoretic approach is used to determine characteristic and minimal polynomials of decimated feedback shift-register sequences in the general context of commutative rings and modules over commutative rings.

If σ is a sequence (s_n), $n=0, 1, \ldots$, of elements of an arbitrary set, then for integers $d \geqq 1$ and $h \geqq 0$ the decimated sequence $\sigma_d^{(h)}$ is given by (s_{nd+h}), $n=0, 1, \ldots$. In information theory the operation of decimation has received particular attention in the case where σ is a feedback shift-register sequence of elements of a finite field. In this context, decimated sequences were already studied by Zierler [14], Selmer [12], and Golomb [3]. Typical applications of decimation in information theory are to the theory of pseudoandom sequences (see Golomb [3] and Lidl and Niederreiter [5, Ch. 7]), to the theory of stream ciphers (see Lidl and Niederreiter [5, Ch. 9] and Rueppel [11, Ch. 6]), and to the design of cryptosystems (see Niederreiter [7], [8]).

If R is an arbitrary commutative ring, then a sequence $\sigma=(s_n)$, $n=0, 1, \ldots$, of elements of R is called a kth-order (linear) feedback shift register sequence if it satisfies a recurrence relation

$$s_{n+k}=a_{k-1}s_{n+k-1}+\ldots+a_0s_n \qquad \text{for} \quad n=0, 1, \ldots \tag{1}$$

with constant coefficients $a_0, \ldots, a_{k-1} \in R$. Any sequence obtained from σ by decimation is again a kth-order feedback shift-register sequence, and so there arises the problem of finding a recurrence relation for $\sigma_d^{(h)}$. In the case where R is a finite field this problem was settled by Duvall and Mortick [2] who used a rather complicated method. Some special cases were dealt with earlier by Selmer [12]. A more convenient approach for finite fields is based on the use of generating functions (see Niederreiter [8] and Smeets [13]). The method in [8] works, in fact, for arbitrary fields and produces the result of Corollary 1 below. In this note we show that matrix theory yields an even simpler and more general approach to the problem of finding a recurrence relation for a decimated sequence.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

We first note that if R is a commutative ring without (multiplicative) identity, then R can be embedded into a ring with identity. Following Dorroh [1] we let Z be the ring of rational integers and endow the Cartesian product $Z \times R$ with the operations

$$\begin{aligned}(i, r)+(j, s)&=(i+j, r+s),\\ (i, r)(j, s)&=(ij, is+jr+rs)\end{aligned} \tag{2}$$

for $i, j \in Z$ and $r, s \in R$. This yields a commutative ring R' with identity $(1, 0)$ which contains the subring $\{(0, r): r \in R\}$ isomorphic to R. Therefore, we can view R as a subring of R'. If R is a commutative ring with identity, we put $R'=R$. In all cases we write e for the identity of R'.

For an arbitrary commutative ring R and a $k \times k$ matrix B over R we define the characteristic polynomial of B by $g(x)=\det(xI-B)$, where I is the $k \times k$ identity matrix over R'. If the sequence $\sigma=(s_n)$ of elements of R satisfies (1), then the polynomial

$$f(x)=ex^k-a_{k-1}x^{k-1}-\ldots-a_0 \in R'[x] \tag{3}$$

is called a characteristic polynomial of σ. We note that for both $g(x)$ and $f(x)$ all coefficients, with the possible exception of the leading coefficient e, belong to R.

In the special case where d is a power of 2, a characteristic polynomial of $\sigma_d^{(h)}$ was found by Pethő [9] by a method different from that in Theorem 1.

Theorem 1. Let R be an arbitrary commutative ring and let σ be a kth-order feedback shift-register sequence of elements of R satisfying (1). Then for any $d \geqq 1$ and $h \geqq 0$ a characteristic polynomial of the decimated sequence $\sigma_d^{(h)}$ is given by the characteristic polynomial of A^d, where A is the $k \times k$ matrix

$$A=\begin{pmatrix} 0 & e & 0 & \ldots & 0 \\ 0 & 0 & e & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & e \\ a_0 & a_1 & a_2 & \ldots & a_{k-1} \end{pmatrix}.$$

Proof. Fix $d \geqq 1$ and $h \geqq 0$ and let

$$g(x)=ex^k-b_{k-1}x^{k-1}-\ldots-b_0 \in R'[x]$$

be the characteristic polynomial of $B=A^d$. We first show that all b_j, $0 \leqq j \leqq k-1$, belong to R (this is, of course, trivial if R has an identity). For $1 \leqq d<k$ the first $k-d$ rows of B are equal to the rows $d, d+1, \ldots, k-1$ of A, and the last d rows of B only contain elements of R. From this the desired property of $g(x)=\det(xI-B)$

follows easily by expansion of the determinant along the last d rows. For $d \geqq k$ we see that B is a matrix over R, and then the desired property of $g(x)$ was already noted earlier.

If $\sigma = (s_n)$, $n = 0, 1, \ldots$, we consider the columns

$$\mathbf{s}_n = (s_n, s_{n+1}, \ldots, s_{n+k-1})^T \qquad \text{for} \quad n = 0, 1, \ldots .$$

From (1) we get $\mathbf{s}_{n+1} = A\mathbf{s}_n$ for $n = 0, 1, \ldots$, hence by induction $\mathbf{s}_n = A^n \mathbf{s}_0$ for $n = 0, 1, \ldots$. With $\mathbf{t}_n = \mathbf{s}_{nd+h}$ we obtain

$$\mathbf{t}_n = A^{nd+h}\mathbf{s}_0 = A^{nd}\mathbf{s}_h = B^n \mathbf{t}_0 \qquad \text{for} \quad n = 0, 1, \ldots .$$

By the Cayley–Hamilton theorem for the commutative ring R' with identity (see [10, Satz 173]) we have

$$B^k = b_{k-1}B^{k-1} + \ldots + b_0 I.$$

It follows that

$$\mathbf{t}_{n+k} = B^{n+k}\mathbf{t}_0 = (b_{k-1}B^{n+k-1} + \ldots + b_0 B^n)\mathbf{t}_0 =$$

$$= b_{k-1}\mathbf{t}_{n+k-1} + \ldots + b_0 \mathbf{t}_n \qquad \text{for} \quad n = 0, 1, \ldots .$$

Comparing first components in the identity above and putting $u_n = s_{nd+h}$, $n = 0, 1, \ldots$, for the terms of $\sigma_d^{(h)}$, we get

$$u_{n+k} = b_{k-1}u_{n+k-1} + \ldots + b_0 u_n \qquad \text{for} \quad n = 0, 1, \ldots . \qquad \square$$

Corollary 1. Let F be a field and let σ be a kth-order feedback shift-register sequence of elements of F with characteristic polynomial $f(x)$. If

$$f(x) = \prod_{j=1}^{k} (x - \alpha_j)$$

is the factorization of $f(x)$ in its splitting field over F, then for any $d \geqq 1$ and $h \geqq 0$ a characteristic polynomial of $\sigma_d^{(h)}$ is given by

$$f_d(x) = \prod_{j=1}^{k} (x - \alpha_j^d).$$

Proof. If $f(x)$ is given by (3), then $f(x)$ is the characteristic polynomial of the matrix A in Theorem 1. By a well-known result from linear algebra (see [4, Theorem 9.15.2]), the characteristic polynomial of A^d is equal to $f_d(x)$. The rest follows from Theorem 1. $\square$

The result of Theorem 1 can be extended further by considering feedback shift-register sequences of elements of a left R-module M as in Niederreiter [6]. A sequence $\sigma = (s_n)$, $n = 0, 1, \ldots$, of elements of M is called a kth-order feedback shift register sequence if it satisfies (1) with constant coefficients $a_0, \ldots, a_{k-1} \in R$. The polynomial $f(x)$ in (3) is again called a characteristic polynomial of σ.

Theorem 2. If R is an arbitrary commutative ring, M is an arbitrary left R-module, and σ is a kth-order feedback shift register sequence of elements of M satisfying (1), then the result of Theorem 1 holds.

Proof. Let R_1 be the ring obtained by endowing $Z \times R$ with the operations in (2). Then R_1 is a commutative ring with identity (1, 0) which contains R' and R as subrings. We turn M into a left R_1-module by defining

$$(i, r)m = im + rm \qquad \text{for all} \quad i \in Z,\ r \in R,\ m \in M.$$

Since $(0, r)m = rm$, the ring R operates on M as before. The R_1-module M is unitary, i.e. $(1, 0)m = m$ for all $m \in M$. Let A be the matrix in Theorem 1 with the entries e replaced by (1, 0) if R has an identity. One shows as in the first part of the proof of Theorem 1 that all coefficients of the characteristic polynomial of $B = A^d$, except possibly the leading coefficient, belong to R. Then we proceed as in the second part of the proof of Theorem 1, using the fact that $k \times k$ matrices over R_1 operate from the left in an obvious way on columns of k elements of M and applying the Cayley–Hamilton theorem for the ring R_1. □

For periodic feedback shift-register sequences we show a result for minimal polynomials in the general context of a left R-module M (note that a commutative ring R can also be viewed as a left R-module). By a minimal polynomial of a feedback shift-register sequence σ of elements of M we mean a characteristic polynomial of σ of least degree. The following theorem generalizes a result of the author [8].

Theorem 3. Let R be an arbitrary commutative ring, let M be an arbitrary left R-module, and let σ be a periodic feedback shift-register sequence of elements of M with a minimal polynomial given by (3) and with period p. Then for any $d \geqq 1$ with $\gcd(d, p) = 1$ and any $h \geqq 0$ a minimal polynomial of the decimated sequence $\sigma_d^{(h)}$ is given by the characteristic polynomial of A^d, where A is the matrix in Theorem 1.

Proof. By Theorem 2 a characteristic polynomial of $\tau = \sigma_d^{(h)}$ of degree k is given by the characteristic polynomial of A^d. Suppose τ has a characteristic polynomial of degree $< k$. Since $\gcd(d, p) = 1$, there exist integers $c \geqq 1$ and $j \geqq 0$ with $cd \equiv 1 \bmod p$ and $jd \equiv -h \bmod p$. Using the fact that σ has period p, it follows that $\tau_c^{(j)} = \sigma$. Applying Theorem 2 to τ, we conclude that σ has a characteristic polynomial of degree $< k$. This contradiction shows the desired result. □

We remark that Theorems 2 and 3 hold with obvious modifications for right R-modules as well.

References

1. *Dorroh, J. L.*, Concerning adjunctions to algebras, Bull. Amer. Math. Soc. **38,** 85–88 (1932).
2. *Duvall, P. F., Mortick, J. C.*, Decimation of periodic sequences, SIAM J. Appl. Math. **21,** 367–372 (1971).
3. *Golomb, S. W.*, Shift Register Sequences, Holden-Day, San Francisco, 1967.
4. *Hohn, F. E.*, Elementary Matrix Algebra, 3rd ed., Macmillan, New York, 1973.
5. *Lidl, R., Niederreiter, H.*, Introduction to Finite Fields and Their Applications, Cambridge Univ. Press, Cambridge, 1986.
6. *Niederreiter, H.*, On the cycle structure of linear recurring sequences, Math. Scand. **38,** 53–77 (1976).
7. *Niederreiter, H.*, A public-key cryptosystem based on shift register sequences, Advances in Cryptology — EUROCRYPT '85 (F. Pichler, ed.), pp. 35–39, Lecture Notes in Computer Science, Vol. **219,** Springer-Verlag, Berlin, 1986.
8. *Niederreiter, H.*, Some new cryptosystems based on feedback shift register sequences, Math. J. Okayama Univ., to appear.
9. *Pethő, A.*, personal communication.
10. *Rédei, L.*, Algebra, Geest& Portig, Leipzig, 1959.
11. *Rueppel, R. A.*, Analysis and Design of Stream Ciphers, Springer-Verlag, Berlin, 1986.
12. *Selmer, E. S.*, Linear Recurrence Relations over Finite Fields, Lecture Notes, Univ. of Bergen, 1966.
13. *Smeets, B.*, Some Results on Linear Recurring Sequences, Ph. D. Dissertation, Univ. of Lund, 1987.
14. *Zierler, N.*, Linear recurring sequences, J. Soc. Indust. Appl. Math. **7,** 31–48 (1959).

Простой и общий подход к децимации последовательностей регистров сдвига с обратной связью

Х. НИДЕРРАЙТЕР

(Вена)

Теория матриц используется для определения характеристических и примитивных многочленов последовательностей регистров сдвига с обратной связью в общей структуре коммутативных колец и модулей над коммутативными кольцами.

H. Niederreiter
Mathematical Institute
Austrian Academy of Sciences
Dr. Ignaz-Seipel-Platz 2
A-1010 Vienna
Austria

РУССКИЙ ПЕРЕВОД

Проблемы управления и теории информации, том **17**, *номер 5 (1988)*

ДИСКОНТИРОВАННАЯ УПРАВЛЯЕМАЯ ЦЕПЬ МАРКОВА С ЗАПАЗДЫВАЮЩИМ УПРАВЛЕНИЕМ

А. Б. Пиуновский

(*Москва*)

Исследуется управляемая цепь Маркова с запаздыванием по управлению. Приводятся необходимые и достаточные условия оптимальности стратегии управления для полунепрерывных моделей. Даётся определение класса τ-моделей, для которых развиваются методы, аналогичные методу динамического программирования. Приводятся характеристические свойства τ-моделей; рассматривается конкретный пример.

1. Введение

Управляемым марковским процессам без запаздывания посвящены работы [1–5], в которых для широкого класса моделей доказан принцип динамического программирования, выведено уравнение Беллмана и получены необходимые и достаточные условия оптимальности стратегий управления. Однако, во многих случаях управление на выходе регулятора имеет запаздывание, например, управление удаленными подвижными объектами, обслуживание сложных технических систем при наличии скрытых отказов и т. д. В отличие от указанных выше работ, в настоящей статье исследуются вопросы существования и построения оптимальных стратегий в однородных моделях с запаздывающим управлением на бесконечном временном интервале. Аналогичные модели с конечным горизонтом рассмотрены в [6, 7]. Исследование подобных моделей, как правило, приводит к существенным сложностям, связанным, в частности, с нарушением марковского свойства изучаемого процесса даже для простейших стратегий управления.

В п. 2 вводятся основные определения, которые в случае нулевого запаздывания переходят в общепринятые понятия [1–5]. В частности, вводится определение класса марковских стационарных селекторов, который в дальнейшем оказывается достаточным для решения проблемы оптимального синтеза. В статье даётся описание класса полунепрерывных моделей и τ-моделей. Для полунепрерывных моделей устанавливается существование равномерно оптимального B-селектора (теорема 2), а также необходимые и достаточные условия

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

оптимальности произвольной стратегии управления (теорема 3). В общих полунепрерывных моделях метод динамического программирования непосредственно применить нельзя, а оптимальный B-селектор может быть немарковским. Поэтому в п. 4 вводится новый класс «τ-моделей», для которых проблему синтеза можно решить с помощью методов, аналогичных методу динамического программирования. Ключевую роль в описании τ-моделей играют гипотезы H_1 и H_2. В основе указанных гипотез лежит предположение, что достаточной статистикой при выборе управления может служить «прогнозируемое» распределение вероятностей текущего состояния, а также предположение, что оптимальный функционал качества имеет интегральный вид (теорема 6).

Отметим, что τ-модель не обязательно является полунепрерывной. Однако, если это так, то существует оптимальный стационарный марковский B-селектор (лемма 1, теоремы 4, 5), который можно построить, решив «уравнение Беллмана» из формулировки гипотезы H_2.

В качестве примера рассмотрена задача оптимизации линейной системы с квадратичным критерием. Показано, что если пренебречь наличием ненулевого запаздывания, то качество управления значительно ухудшается.

Следует отметить, что в случае отсутствия запаздывания полученные результаты переходят в известные факты теории динамического программирования [1–5].

2. Определения и обозначения

Пусть заданы целое неотрицательное число τ и произвольные борелевские пространства $(X, \mathscr{B}(X))$ — пространство состояний; и $(A, \mathscr{B}(A))$ — пространство управлений. Прямое произведение

$$H_t = \begin{cases} A^t & \text{при} \quad 1 \leqq t < \tau, \\ A^\tau \times X \times (A \times X)^{t-\tau}, & \text{при} \quad t \geqq \tau \end{cases}$$

называется пространством историй. Очевидно H_t при всех $t = 1, 2, \ldots$, а также $\Omega \triangleq H_\infty$ и $H = \bigcup_{t=1}^{\infty} H_t$ являются борелевскими пространствами с σ-алгебрами борелевских множеств $\mathscr{B}(H_t)$, $\mathscr{F} = \mathscr{B}(H_\infty)$, $\mathscr{B}(H)$ соответственно.

В дальнейшем, если $(E, \mathscr{B}(E))$ — борелевское пространство, символами $\mathscr{A}(E), \mathscr{U}(E)$ обозначаются аналитическая и универсально измеримая σ-алгебры в E. Как известно [1], пространство вероятностных мер на борелевском пространстве E является борелевским; оно обозначается символом $P(E)$. В дальнейшем считаем, что в $P(E)$ зафиксирована слабая топология. Свойства борелевских пространств и вероятностных мер на них, а также A-функций и A-измеримых и универсально измеримых отображений подробно изложены в [1, 2].

Элементами пространства Ω являются траектории, то есть реализации последовательностей управлений и состояний: $\omega=\{a_1, \ldots, a_\tau, \xi_0, a_{\tau+1}, \xi_1, \ldots\}$, которые выписаны в порядке их появления перед наблюдателем. Подчеркнем, что состояние ξ_t формируется под влиянием управления a_2 и становится доступным для наблюдения только после выбора управления $a_{t+\tau}$. При $\tau=0$ получается классическая схема [1, 2]. Элементы ξ_t, a_t последовательности $\omega\in\Omega$ в дальнейшем обозначаются $\xi_t(\omega)$, $a_t(\omega)$, то есть рассматриваются как функции (проекции) из Ω в X и A. Аналогичным образом $h_t(\omega)=\{a_1, \ldots, a_\tau, \xi_0, a_{\tau+1}, \ldots, \xi_{t-\tau}\}$ — часть траектории, доступная для наблюдения после выбора управления a_t. Аргумент ω в дальнейшем, как правило, опускается. Символом $\mathscr{F}_t$ обозначается σ-алгебра в Ω, являющаяся прообразом $\mathscr{B}(H_{t+\tau})$ относительно проектирования $h_{t+\tau}:\Omega\to H_{t+\tau}$. Таким образом, $\{\xi_t\}_{t\geq 0}$ и $\{a_t\}_{t\geq 1}$ являются $\mathscr{F}_t$-согласованными случайными последовательностями. Как правило, параметр t принимает натуральные значения; однако иногда нам будет удобно обозначать σ-алгебру $\sigma\{a_s, S\leqq\tau-\theta\}$ в Ω символом $\mathscr{F}_{-\theta}$, который согласуется с принятыми выше обозначениями ($\theta=1, 2, \ldots, \tau$; $\mathscr{F}_{-\tau}=\{\varnothing, \Omega\}$ — тривиальная σ-алгебра).

Предположим, что заданы: 1) начальное распределение $P^0(\cdot)$ — вероятностная мера на X из некоторого класса $\tilde{P}(X)\subseteq P(X)$; 2) переходная вероятность за один шаг $P_a(x, \Gamma)$ — вероятностная мера на X (при произвольных фиксированных $a\in A$, $x\in X$), $\mathscr{B}(A\times X)$ — измеримая по (a, x) при любом фиксированном $\Gamma\in\mathscr{B}(X)$; 3) коэффициент дисконтирования $\beta\in(0, 1)$ и A — функция дохода $r(x, a)$.

Объект $Z=\{X, A, P, r\}$ называется моделью.

Определение. Стратегией управления $\pi=\{\mu_t\}_{t=1}^{\infty}$ называется последовательность универсально измеримых стохастических ядер $\mu_t(\Gamma|P^0, h_{t-1})$ на A. При $t=1$ μ_1 — вероятностная мера, которая может зависеть только от начального распределения $P^0(\cdot)$.

Теорема 1. Всякой стратегии $\pi=\{\mu_t\}_{t=1}^{\infty}$ при произвольном фиксированном $P^0\in P(X)$ соответствует единственная вероятностная мера P^π на Ω такая, что

$$1)\ P^\pi\{a_1\in\Gamma_1^A, \ldots, a_{t-1}\in\Gamma_{t-1}^A, a_t\in\Gamma_t^A\}=$$

$$=\int_{\Gamma_1^A}\ldots\int_{\Gamma_{t-1}^A}\mu_t(\Gamma_t^A|P^0, a_1, \ldots, a_{t-1})\mu_{t-1}(da_{t-1}|P^0, a_1, \ldots, a_{t-2})\times$$

$$\times\ldots\times\mu_1(da_1|P^0),\qquad 1\leqq t\leqq\tau;$$

$$2)\ P^\pi\{a_1\in\Gamma_1^A, \ldots, a_\tau\in\Gamma_\tau^A, \xi_0\in\Gamma_0^X, a_{\tau+1}\in\Gamma_{\tau+1}^A, \ldots, \xi_t\in\Gamma_t^X\}=$$

$$=\int_{\Gamma_1^A}\ldots\int_{\Gamma_\tau^A}\int_{\Gamma_0^X}\int_{\Gamma_{\tau+1}^A}\ldots\int_{\Gamma_{\tau+t}^A}P_{a_t}(\xi_{t-1}, \Gamma_t^X)\mu_{\tau+t}(da_{\tau+t}|P^0, a_1\ldots$$

$$\ldots a_\tau\xi_0 a_{\tau+1}\ldots\xi_{t-1})\ldots\mu_{\tau+1}(da_{\tau+1}|P^0, a_1\ldots a_\tau\xi_0)P^0(d\xi_0)\times$$

$$\times\mu_\tau(da_\tau|P^0, a_1\ldots a_{\tau-1})\ldots\mu_1(da_1|P^0),\qquad t=0, 1, 2, \ldots$$

3) $P^{\pi}\{a_1 \in \Gamma_1^A, \ldots, a_\tau \in \Gamma_\tau^A, \xi_0 \in \Gamma_0^X, a_{\tau+1} \in \Gamma_{\tau+1}^A, \ldots$

$\ldots \xi_t \in \Gamma_t^X, a_{\tau+t+1} \in \Gamma_{\tau+t+1}^A\} = \int\limits_{\Gamma_1^A} \ldots \int\limits_{\Gamma_\tau^A} \int\limits_{\Gamma_0^X} \int\limits_{\Gamma_{\tau+1}^A} \ldots \int\limits_{\Gamma_{\tau+t}^A} \int\limits_{\Gamma_t^X} \mu_{\tau+t+1} \times (\Gamma_{\tau+t+1}^A |$

$P^0, a_1 \ldots a_\tau \xi_0 a_{\tau+1} \ldots \xi_t) P_{a_\tau}(\xi_{t-1}, d\xi_t) \mu_{\tau+t}(da_{\tau+t} | P^0, a_1 \ldots$

$\ldots a_\tau \xi_0 a_{\tau-1} \ldots \xi_{t-1}) \ldots \mu_{\tau+1}(da_{\tau+1} | P^0, a_1 \ldots a_\tau \xi_0) P^0(d\xi_0) \times$

$\times \mu_\tau(da_\tau | P^0, a_1 \ldots a_{\tau-1}) \ldots \mu_1(da_1 | P^0), \qquad t = 0, 1, 2, \ldots$

Здесь $\Gamma_i^A \in \mathscr{B}(A)$, $\Gamma_i^X \in \mathscr{B}(X)$ — произвольные множества.

Доказательство непосредственно следует из теоремы Ионеску–Тульча [1, 2].

Рандомизированное управление a_t задается ядром $\mu_t(\cdot | P^0, h_{t-1})$, то есть при выборе управления a_t можно пользоваться только информацией о P^0 и значениях $a_1, a_2, \ldots, a_\tau, \xi_0, a_{\tau+1}, \xi_1, \ldots, a_{t-1}, \xi_{t-\tau-1} (t \geqq \tau + 1)$. Другими словами, управление a_t является $U_{t-\tau-1}$ — измеримым случайным элементом, где U_t — универсальное расширение σ-алгебры $\mathscr{F}_t$. Следовательно, величина τ является запаздыванием в управлении. При $\tau = 0$ получаем модель без запаздывания, изучению которой посвящены работы [1–5].

Согласно теореме 1, всякая стратегия τ при фиксированном P^0 определяет согласованный случайный процесс (a_t, ξ_t) на стохастическом базисе $(\Omega, \mathscr{F}, P^\pi, (\mathscr{F}_t)_{t \geqq 0})$. Интеграл по мере P^π обозначается символом M^π, σ-алгебры $\mathscr{F}_t$ пополнены множествами P^π-меры нуль.

Определение. Символы $\pi^t = \{\mu_\theta\}_{\theta=1}^t$, $\pi_t = \{\mu_\theta\}_{\theta=t+1}^\infty$ в дальнейшем обозначают «части» стратегии π.

Очевидно, что фиксированная π^t при заданном $P^0 \in P(X)$ по аналогии с теоремой 1 определяет единственную вероятностную меру $P_H^{\pi^t}$ на H_t, прообраз которой на Ω обозначается P^{π^t} $(t = 1, 2, \ldots)$. Пусть $P^0 \in P(X)$ фиксировано и $\{\psi_T(\omega)\}_{T=1}^\infty$ — последовательность случайных величин такая, что для всех стратегий π определён P^π — почти наверное (в дальнейшем: P^π — п. н.) конечный предел $\lim\limits_{T \to \infty} M^\pi[\psi_T(\omega) | \mathscr{F}_{t-\tau}]$, $t = 0, 1, 2, \ldots$. Предположим, что π^t фиксирована, а π_t меняется. Тогда $\mathscr{F}_{t-\tau}$ — измеримые случайные величины

$$\lim_{T \to \infty} M^\pi[\psi_T(\omega) | \mathscr{F}_{t-\tau}]$$

образуют семейство, для которого π_t выступает в роли параметра, и имеет смысл ввести операцию P^{π^t} — ess sup по множеству $\{\pi_t\}$. Описанная конструкция в дальнейшем обозначается P^{π^t} — $\operatorname{ess\,sup}\limits_{\pi_t} \lim\limits_{T \to \infty} M^{\pi_t}[\psi_T(\omega) | \mathscr{F}_{t-\tau}]$ (предполагается, что существенная верхняя грань определена.). Аналогичный смысл вкладывается в запись: P^{π^t} — $\operatorname{ess\,sup}\limits_{\pi_t} M^{\pi_t}[\psi(\omega) | \mathscr{F}_{t-\tau}]$.

Определение. При фиксированном P^0 стратегия π называется допустимой, если $\forall\, t \geq 0\ \forall\, \theta = t+1,\ t+2,\ \ldots$ определены P^π — п. н. конечные условные математические ожидания $M^\pi[r(\xi_{\theta-1},\ a_\theta)|\mathscr{F}_{t-\tau}]$ и существует конечный P^π — п. н. предел $\lim\limits_{T\to\infty} M^\pi\left[\sum\limits_{\theta=t+1}^{I} e^{-\alpha\theta}\cdot r(\xi_{\theta-1}, a_\theta)|\mathscr{F}_{t-\tau}\right]$. Класс допустимых стратегий обозначается символом Π.

Величина $W(\pi) \triangleq \lim\limits_{T\to\infty} M^\pi\left[\sum\limits_{t=1}^{T} e^{\alpha t}\cdot r(\xi_{t-1}, a_t)\right]$ при заданном P^0 называется оценкой стратегии $\pi \in \Pi$. Стратегия $\pi^* \in \Pi$ называется оптимальной (для начального распределения P^0), если $W(\pi^*) = \sup\limits_{\pi\in\Pi} W(\pi)$. Стратегия π^* называется равномерно оптимальной (ε-оптимальной), если она допустима при всех

$$P^0 \in \tilde{P}(X) \text{ и } W(\pi^*) = \sup_{\pi\in\Pi} W(\pi)\,(W(\pi^*) > \sup_{\pi\in\Pi} W(\pi) - \varepsilon)$$

одновременно при всевозможных начальных распределениях $P^0 \in \tilde{P}(X)$.

Стратегия $\pi = \{\mu_t\}_{t=1}^{\infty}$ называется селектором, если $\forall\, t > 0$ мера μ_t сосредоточена в точке $a_t = \varphi(P^0, h_{t-1})$. Селектор называется марковским, если

$$\varphi(P^0, h_{t-1}) = \varphi(t, P^\varphi_{t-1}).$$

Здесь и ниже

$$P^\pi_{t-1}(\Gamma, P^0, \omega) \triangleq P^\pi\{\xi_{t-1} \in \Gamma\,|\mathscr{F}_{t-\tau-1}\} \tag{1}$$

— условная вероятность; $\Gamma \in \mathscr{B}(X)$. Марковский селектор называется стационарным, если $\varphi(t,\ P^\varphi_{t-1}) = \varphi(t \wedge \tau + 1, P^\varphi_{t-1})$. Здесь и ниже $a \wedge b \triangleq \min\{a, b\}$. В соответствии с введенными обозначениями условная вероятность P^π_{t-1} полностью определяется «частью» стратегии π^{t-1}. Поэтому иногда используется обозначение $P^{\pi^{t-1}}_{t-1}$.

Символом B_{t-1} будем обозначать семейство $P^\pi_{t-1}(\cdot, P^0, \omega)$ при различных P^0, π, ω. В дальнейшем предполагается, что все семейства B_t совпадают: $B_t \equiv B$ и $\tilde{P}(X) \subseteq B$; $B \in \mathscr{B}(P(X))$ и $\forall\, P \in B\ \forall\, a \in A$ существует конечный интеграл

$$\int_X r(y, a)P(dy).$$

Определения. Если стратегия (селектор) задается борелевскими ядрами (функциями), то она называется B-стратегией (B-селектором). Марковский селектор φ называется A-селектором, если отображение $\varphi(t, \cdot): B \to A$ является $\mathscr{A}(B)$ — измеримым при всех $t = 1,\ 2,\ \ldots$. В дальнейшем множество всех допустимых стационарных марковских B-селекторов обозначается символом Π^m.

3. Полунепрерывные модели

Определение. Модель Z называется полунепрерывной, если выполнены следующие условия:

у1. Функция r ограничена сверху и полунепрерывна сверху.

у2. Переходная вероятность $P_a(x, \Gamma)$ задаёт непрерывное стохастическое ядро P: $A \times X \to P(X)$.

у3. Множество A-компакт.

Замечание. Для полунепрерывных сверху функций допускается значение «$-\infty$».

В полунепрерывной модели целесообразно не ограничивать класс допустимых стратегий, так как все условные математические ожидания и пределы можно корректно определить на расширенной прямой с помощью стандартной операции «урезания» функции дохода снизу.

Как показано в п. 6, $\forall\, t \geqq 0\ \forall\, P^0$ — случайная величина

$$P^{\pi^{t+\tau}}\text{—ess}\sup_{\pi_{t+\tau}} \lim_{T\to\infty} M^{\pi_{t+\tau}}\left[\sum_{\theta=t+1}^{I} e^{-\alpha\theta}\cdot r(\xi_{\theta-1}, a_\theta)\middle|\mathscr{F}_t\right] P^{\pi^{t+\tau}}\text{— п. н.} = V_t(\xi_t, a_{t+1}, \ldots, a_{t+\tau})$$

зависит только от перечисленных аргументов. Полное исследование полунепрерывных моделей при $\tau = 0$ содержится в [1, 2, 5].

Как отмечалось выше, в случае $\tau > 0$ в модели может не существовать ε-оптимального марковского селектора. Однако, имеют место следующие утверждения.

Теорема 2. В полунепрерывной модели существует равномерно оптимальный B-селектор π^*, причём

$$\forall\, P^0 \in \tilde{P}(X)\ W(\pi^*) = \sup_{\pi} W(\pi) = \sup_{(a_1,\ldots,a_\tau)\in A^\tau} \int_X V_0(\xi_0, a_1, \ldots, a_\tau) P^0(d\xi_0).$$

Теорема 3. При любом фиксированном начальном распределении $P^0 \in \tilde{P}(X)$ для оптимальности стратегии $\pi = \{\mu_t\}_{t=1}^{\infty}$ в полунепрерывной модели необходимо и достаточно выполнения условий:

1) $\forall\, t = 0, 1, \ldots\ \lim_{T\to\infty} M^{\pi}\left[\sum_{\theta=t+1}^{T} e^{-\alpha\theta}\cdot r(\xi\theta_{-1}, a_\theta)\middle|\mathscr{F}_t\right] P^{\pi}\text{— п. н.} = V_t(\xi_t, a_{t+1}, \ldots, a_{t+\tau})$;

2) $M^{\pi}[V_0(\xi_0, a_1, \ldots, a_\tau)] = \sup_{(a_1,\ldots,a_\tau)\in A^\tau} \int_X V_0(\xi_0, a_1, \ldots, a_\tau) P^0(d\xi_0)$.

Доказательства изложены в п. 6.

4. τ-модели

Обозначим символом D линейное пространство действительных универсально измеримых функций $F(\cdot)$ на X таких, что $\forall P \in B$ существует конечный интеграл $\int_X F(y)P(dy)$.

Предположим, что начальное распределение $P^0 \in \tilde{P}(X)$ фиксировано.

Пусть $\{F_i\}_{i=0}^{\tau}$ — набор функций из класса D. Нам потребуются следующие условия:

У4. а) $\forall a \in A \ \forall x \in X$ существует конечный интеграл $\int_X F_i(y)P_a(x, dy)$; $i=0, 1, \ldots, \tau$;

б) $\forall \pi \in \Pi \ \forall \theta = 1, 2, \ldots$ существуют конечные условные математические ожидания $M^{\pi}[F_{\theta \wedge \tau}(\xi_{\theta})|\mathscr{F}_{t-\tau}]$, $t=0, 1, \ldots, \theta$; $M^{\pi}[\int_X F_{\theta \wedge \tau}(y)P_{a_{\theta}}(\xi_{\theta-1}, dy)|\mathscr{F}_{t-\tau}]$, $t=0,1, \ldots, \theta-1$.

в) $\forall \pi \in \Pi \ \forall t = 0, 1, \ldots$ существует $\lim_{T\to\infty} M^{\pi}[e^{-\alpha T} \cdot F_{\tau}(\xi_T)|\mathscr{F}_{t-\tau}]$.

Например, У4 выполнены для ограниченных функций F_i.

Гипотеза H_1. Существуют стратегия $\pi^* \in \Pi$ и набор функций $\{F_i(y)\}_{i=0}^{\tau} \in D$, удовлетворяющий условиям У4, такие, что

1) $\forall t = 1, 2, \ldots$

$$M^{\pi^*}[r(\xi_{t-1}, a_t^*) - e^{\alpha} \cdot F_{(t-1)\wedge\tau}(\xi_{t-1}) + \int_X F_{t\wedge\tau}(y) \times$$

$$\times P_{a_t^*}(\xi_{t-1}, dy)|\mathscr{F}_{t-\tau-1}] = 0 \qquad P^{\pi^*} - \text{п. н.} \tag{2}$$

$$2) \ \forall \pi \in \Pi \quad P^{\pi} - \text{п. н.} \quad M^{\pi}[r(\xi_{t-1}, a_t) - e^{\alpha} \cdot F_{(t-1)\wedge\tau}(\xi_{t-1}) +$$

$$+ \int_X F_{t\wedge\tau}(y)P_{a_t}(\xi_{t-1}, dy)|\mathscr{F}_{t-\tau-1}] \leqq 0$$

$$3) \ \forall \pi \in \Pi \ \lim_{T\to\infty} M^{\pi}[e^{-\alpha T} \cdot F_{\tau}(\xi_T)] \geqq \lim_{T\to\infty} M^{\pi^*}[e^{-\alpha T} \cdot F_{\tau}(\xi_T)].$$

Гипотеза H_2. Существует набор A-функций $\{F_i(y)\}_{i=0}^{\tau}$, удовлетворяющий условиям У4, такой, что

1) $\forall \pi \in \Pi \quad \forall t = 1, 2, \ldots$

$$P^{\pi} - \operatorname{ess\,sup}_{a\in A} \{M^{\pi}[r(\xi_{t-1}, a) - e^{\alpha} \cdot F_{(t-1)\wedge\tau}(\xi_{t-1}) +$$

$$+ \int_X F_{t\wedge\tau}(y)P_a(\xi_{t-1}, dy)|\mathscr{F}_{t-\tau-1}]\} = 0 \quad P^{\pi} - \text{п. н.} \tag{3}$$

2) $\forall \pi \in \Pi \lim\limits_{T \to \infty} M^{\pi}[e^{-\alpha T} F_{\tau}(\xi_T)] = 0.$

Определение. Модели, в которых справедлива гипотеза H_2, будем называть τ-моделями. (Аналогичное определение дано в [6] для моделей с конечным горизонтом.)

Замечание. В основе гипотез H_1, H_2 лежат два предположения:

1) при выборе управления на каждом шаге t достаточной статистикой является прогнозируемое распределение вероятностей P^{π}_{t-1} (1);

2) оптимальный функционал качества $V_t(\cdot)$, введенный в п. 3, определенным образом выражается с помощью интеграла по мере $P^{\pi}_{t+\tau}$ (см. теорему 6), причем подынтегральные функции совпадают с функциями F_i из формулировки гипотез.

Отметим, что равенство (3) из гипотезы H_2 играет роль уравнения Беллмана; с помощью него можно определить функции F_i, которые естественно назвать функциями Беллмана. В случае нулевого запаздывания гипотезы H_1, H_2 справедливы, например, для полунепрерывных моделей с ограниченной функцией дохода; при этом функция F_0 совпадает с решением классического уравнения Беллмана [5].

Лемма 1. Пусть модель Z полунепрерывна и $\{F_i(y)\}_{i=0}^{\tau}$ — произвольные ограниченные сверху полунепрерывные сверху функции на X. Тогда

$$1)\ \forall \pi\ \forall t \geqq 1 \text{ достигается } P^{\pi} - \operatorname*{ess\,sup}_{a \in A} \{M^{\pi}[r(\xi_{t-1}, a) -$$

$$- e^{\alpha} F_{(t-1) \wedge \tau}(\xi_{t-1}) + \int_X F_{t \wedge \tau}(y) P_a(\xi_{t-1}, dy) | \mathscr{F}_{t-\tau-1}]\}$$

при некотором управлении $\varphi_t(\mathrm{P}^{\pi}_{t-1})$, измеримо (по Борелю), зависящем лишь от условной вероятностной меры $P^{\pi}_{t-1} \in B$;

2) значение супремума является ограниченной сверху полунепрерывной сверху функцией на B;

3) для стационарного марковского B-селектора $\varphi(t, P^{\varphi}_{t-1} \triangleq \varphi_{t \wedge \tau+1}(P^{\varphi}_{t-1})$ при всех $t \geqq 1$ выполнены равенства

$$M^{\varphi}[r(\xi_{t-1}, a_t) - e^{\alpha} F_{(t-1) \wedge \tau}(\xi_{t-1}) + \int_X F_{t \wedge \tau}(y) P_{a_t}(\xi_{t-1}, dy) \mathscr{F}_{t-\tau-1}] P^{\varphi} - \text{п.н.} =$$

$$P^{\varphi} - \operatorname*{ess\,sup}_{a \in A} \{M^{\varphi}[r(\xi_{t-1}, a) - e^{\alpha} F_{(t-1) \wedge \tau}(\xi_{t-1}) + \int_X F_{t \wedge \tau}(y) P_a(\xi_{t-1}, dy) | \mathscr{F}_{t-\tau-1}]$$

(существует «измеримый выбор»).

Доказательство приведено в п. 7.

Заметим, что если в τ-модели применима лемма 1, причем $\varphi \in \Pi$, то справедлива гипотеза H_1.

Теорема 4. Если в модели справедлива гипотеза H_1, то стратегия π^* оптимальна, причем

$$W(\pi^*) = \int_X F_0(y)P^0(dy) - \lim_{T\to\infty} M^{\pi^*}[e^{-\alpha T}F_\tau(\xi_T)].$$

Доказательство приведено в п. 8.

Теорема 5. Пусть в τ-модели для стратегии $\pi^* \in \Pi$ P^{π^*} – п. н. выполнено равенство (2) при $t = 1, 2, \ldots$. Тогда стратегия π^* оптимальна, причем $W(\pi^*) =$

$$= \int_X F_0(y)P^0(dy).$$

Доказательство следует из теоремы 4, если заметить, что для набора функций $\{F_i\}_{i=0}^{\tau}$ и стратегии π^* выполнена гипотеза H_1.

Пусть для τ-модели выполнены следующие условия:

У5. а) $\forall \pi \lim_{T\to\infty} M^{\pi}[e^{-\alpha T}F_\tau(\xi_T)] = 0$;

б) условия У4 выполнены для всех стратегий (а только для $\pi \in \Pi$);

в) для всех π, $t = 1, 2, \ldots$ P^{π} – п. н. выполнено равенство (3).

Тогда теорему 5 можно обратить: если стратегия π^* оптимальна, то для нее $P^{\pi N}$ – п. н. выполнено равенство (2). Доказательство проводится от противного с использованием леммы 3 и замечания к ней. При этом справедливы следующие утверждения:

1) для всякого $\varepsilon > 0$ существует ε-оптимальный стационарный марковский A-селектор;

2) если в п. 3) леммы 1 существует универсально измеримый (измеримый) выбор, то существует оптимальный стационарный марковский селектор (B-селектор).

Предположим, что допустимыми являются все стратегии. Например, модель Z полунепрерывна, либо функция дохода r равномерно ограничена.

Определение. При фиксированном начальном распределении P^0 положим

$$\Phi_t(h_t) \triangleq P^{\pi^t} - \operatorname{ess\,sup}_{\pi_t} \lim_{T\to\infty} M^{\pi_t}\left[\sum_{\theta=t+1}^{I} e^{-\alpha\theta} r(\xi_{\theta-1}, a_\theta) \mid \mathscr{F}_{t-\tau}\right], \qquad t = 0, 1, 2, \ldots$$

Функция $\Phi_t(h_t(\omega))$ является случайной величиной, зависящей от ядер $\pi^t = \{\mu_\theta\}_{H-1}^{t}$. С помощью теоремы 3 нетрудно показать, что в полунепрерывной модели при фиксированном P^0 для оптимальной стратегии π^* справедливо равенство

$$\Phi_{t+\tau}(h_{t+\tau}) = V_t(\xi_t, a_{t+1}, \ldots, a_{t+\tau}) -$$

$$- e^{-\alpha(t+1)} r(\xi_t, a_{t+1}) - e^{-\alpha(t+2)} \cdot \int_X r(y_1, a_{t+2}) P_{a_{t+1}}(\xi_t, dy_1) -$$

$$- \ldots - e^{-\alpha(t+\tau)} \int_X \ldots \int_X r(y_{\tau-1}, a_{t+\tau}) P_{a_{t+\tau-1}}(y_{\tau-2}, dy_{\tau-1}) \ldots$$

$$\ldots P_{a_{t+1}}(\xi_t, dy_1) \quad P^{\pi^*} - \text{п. н.} \qquad t = 0, 1, \ldots$$

Теорема 6. Если функция дохода r равномерно ограничена сверху, то следующие утверждения эквивалентны:

1) функция $\Phi_t(h_t)$ при любой стратегии π является функционалом, зависящим ($P^{\pi t}$ – п. н.) только от условной меры $P_t^{\pi t}(\cdot, P^0, \omega)$, причем

$$\Phi_t(h_t)P^{\pi t} - \text{п. н.} = \tilde{\Phi}_t(P_t^{\pi t})P^{\pi t} - \text{п. н.} = e^{-\alpha t}\int_X F_{t\wedge\tau}(z)P_t^{\pi t}(dz, P^0, \omega), \quad t=0, 1, 2, \ldots,$$

где $\{F_i(z)\}_{i=0}^{\tau}$ — набор A-функций из D, удовлетворяющий условиям У4 такой, что

$$\forall \pi \lim_{T\to\infty} M^{\pi}[e^{-\alpha T}\cdot F_{\tau}(\xi_T)] = 0,$$

2) модель Z является τ-моделью.

Доказательство приведено в п. 9.

Если любой конечный набор мер из B линейно независим и

$$\Phi_t(h_t)P^{\pi t} - \text{п. н.} = \tilde{\Phi}_t(P_t^{\pi t})$$

является функционалом на B, то некоторые достаточные условия, при которых функционал $\tilde{\Phi}_t$ имеет интегральный вид, можно найти в [6]. При этом изучаемая модель будет τ-моделью. Отметим, что в случае $\tau = 0$ меры $P_t^{\pi t}$ сосредоточены в точках ξ_t. (См. определение (1).)

5. Пример

Пусть $X = A = R^1$; процесс ξ_t допускает представление:

$$\xi_t = B\xi_{t-1} + Ca_t + D\eta_t,$$

где $\{\eta_t\}_{t=1}^{\infty}$ — последовательность независимых случайных величин с плотностью распределения $h(z)$ и первыми двумя моментами, равными нулю и единице соответственно. Тогда переходная вероятность за один шаг абсолютно непрерывна относительно меры Лебега, и плотность распределения имеет вид:

$$p_a(y, z) = \frac{1}{|D|}h\left(\frac{z - By - Ca}{D}\right).$$

Положим $r(y, a) = ka^2 + My^2$. Здесь B, $C \neq 0$, $D \neq 0$, $K < 0$, $M < 0$ — произвольные константы.

В задачах оптимизации линейных динамических систем с квадратичным критерием (без запаздывания) функция Беллмана имеет вид квадратичной

формы по состоянию. В нашем случае в роли функции Беллмана выступают функции F_i из гипотез H_1, H_2. Поэтому естественно искать их в виде

$$F_i(y)=g_iy^2+S_iy+q_i \qquad (i=0, 1, \ldots, \tau).$$

Пусть начальное распределение P^0 величины ξ_0 фиксировано.

Проверяя гипотезу H_1, получаем:

$$g_i\equiv g=\frac{1}{2C^2e^{\alpha}}[MC^2+KB^2-Ke^{\alpha}-\sqrt{(MC^2+KB^2-Ke^{\alpha})^2+4C^2e^{\alpha}MK}];$$

$$S_i=0,$$

$$q_{\tau}=\frac{D^2}{e^{\alpha}-1}\cdot\left[g+(M-e^{\alpha}g+gB^2)\cdot\sum_{i=0}^{\tau-1}B^{2i}\right].$$

$$q_i=e^{-\alpha}\left[q_{i+1}+gD^2+(M-e^{\alpha}g+gB^2)\cdot\left\{D_{\xi_0}\cdot B^{2i}+D^2\cdot\sum_{j=0}^{i-1}B^{2j}\right\}\right],$$

$$i=\tau-1, \tau-2, \ldots, 0.$$

(Здесь и ниже M_{ξ_0}, D_{ξ_0} — математическое ожидание и дисперсия случайной величины ξ_0.) Одновременно строится стратегия φ^*, которая имеет следующий вид:

$$\varphi^*(P)=a^*\triangleq-\frac{gBC}{K+gC^2}\int_X yP(dy). \tag{4}$$

Таким образом, в рассматриваемом примере справедлива гипотеза H_1 для φ^* и $\{F_i\}_{i=0}^{\tau}$. В соответствии с теоремой 4 стационарный марковский B-селектор (4) является равномерно оптимальной стратегией и

$$W(\varphi^*)=\sup_{\pi\in\Pi}W(\pi)=\int_X F_0(y)P^0(dy).$$

Если считать допустимыми только стратегии, для которых

$$\lim_{T\to\infty}M^{\pi}[e^{-\alpha T}\xi_T^2]=0,$$

то селектор (4) является допустимым, и для $\{F_i\}_{i=0}^{\tau}$ справедлива гипотеза H_2, то есть рассмотренная модель является τ-моделью. Равномерно оптимальный синтез задается равенством (4).

Величина $\int_X yP_{t-1}^{\varphi^*}(dy, P^0, \omega)$ вычисляется по формуле:

$$\int_X yP_{t-1}^{\varphi^*}(dy, P^0, \omega)=\begin{cases}\xi_{t-\tau-1}\cdot B^{\tau}+C\sum\limits_{s=t-\tau}^{t-1}a_s^*\cdot B^{t-s-1}, & t>\tau;\\ M_{\xi_0}\cdot B^{t-1}+C\sum\limits_{s=1}^{t-1}a_s^*\cdot B^{t-s-1}, & 1\leqq t\leqq\tau.\end{cases}$$

Отметим, что в рассмотренном примере модель не является полунепрерывной.

Если в данном примере пренебречь запаздыванием, то получится модель, исследованная, например, в [5]. С помощью метода динамического программирования устанавливается, что управления должно линейно зависеть от последнего наблюдаемого состояния процесса ξ_t:

$$\hat{a}_t = \hat{\varphi}(P^0, h_{t-1}) = -\frac{gBC}{K+gC^2} \cdot \xi_{t-\tau-1} \quad (t \geqq \tau+1).$$

Чтобы корректно замкнуть модель на начальном интервале $t = 1, 2, \ldots, \tau$, будем считать, что $M_{\xi_0} = 0$; $\hat{a}_t = 0$ $(t \leqq \tau)$. В случае $\tau = 0$ стратегия управления $\hat{\varphi}$ совпадает с φ^* и является оптимальной.

Предположим, что на самом деле $\tau > 0$ и определим выигрыш, который даёт оптимальная стратегия (4) по сравнению с «псевдооптимальной» стратегией $\hat{\varphi}$ в случае $B = 1$. Для этого воспользуемся леммой 3 (см. п. 8) при построенных функциях $\{F_i(y)\}_{i=0}^{\tau}$. Легко проверить, что

$$\lim_{T\to\infty} M^{\hat{\varphi}}[e^{-\alpha T} \cdot F_\tau(\xi_T)] = 0; \qquad M^{\hat{\varphi}}[r(\xi_{\theta-1}, \hat{a}_\theta) -$$

$$-e^{\alpha} \cdot F_{(\theta-1)\wedge\tau}(\xi_{\theta-1}) + \int_X F_{\theta\wedge\tau}(y) P_{\hat{a}_\theta}(\xi_{\theta-1}, dy) \mid \mathscr{F}_{\theta-\tau-1}] -$$

$$= \begin{cases} 0, & \theta = 1, 2, \ldots, \tau; \\ \dfrac{g^2C^4}{K+gC^2} \cdot \left(\displaystyle\sum_{i=0}^{\tau-1} \hat{a}_{\theta-i-1} \right)^2, & \theta \geqq \tau+1. \end{cases}$$

Следовательно,

$$W(\varphi^*) - W(\hat{\varphi}) = -\frac{g^4C^6}{(K+gC^2)^3} \cdot e^{-\alpha(\tau+2)} \times \sum_{t=0}^{\infty} e^{-\alpha t} \times M\left[\left(\sum_{i=0}^{\tau-1} \hat{\xi}_{t-i} \right)^2 \right],$$

где

$$\hat{\xi}_{-\tau+1} = \hat{\xi}_{-\tau+2} = \ldots = \hat{\xi}_{-1} \triangleq 0; \quad \hat{\xi}_0 \triangleq \xi_0;$$

$$\hat{\xi}_t \triangleq \hat{\xi}_{t-1} + D\eta_t, \qquad 1 \leqq t \leqq \tau;$$

$$\hat{\xi}_{t-1} - \frac{gC^2}{K+gC^2} \hat{\xi}_{t-\tau-1} + D\eta_t, \qquad t > \tau;$$

M — символ математического ожидания по мере, отвечающей случайному процессу $\hat{\xi}_t$. При $t = 0, 1, \ldots, \tau-1$

$$M\left[\left(\sum_{i=0}^{\tau-1} \hat{\xi}_{t-i} \right)^2 \right] = (t+1)^2 D_{\xi_0} + D^2 \cdot \frac{t(t+1)(2t+1)}{6}.$$

Напомним, что величины K, g — отрицательны, то есть $W(\varphi^*) > W(\hat{\varphi})$.

6. Доказательство теорем 2, 3

Рассмотрим дисконтированную модель $\hat{Z}$ без запаздывания (с тем же коэффициентом дисконтирования $\alpha\in(0, 1)$, которая задается элементами:

$$\hat{X}\triangleq A^\tau\times X;\quad \hat{A}\triangleq A;\quad \hat{P}_{\hat{a}}(\hat{x},\hat{\Gamma})\triangleq P_{\hat{a}^1}(x,\hat{\Gamma}_{(\hat{a}^2,\ldots,\hat{a}^\tau,\hat{a})});\quad \hat{r}(\hat{x},\hat{a})\triangleq r(x,\hat{a}^1).$$

Здесь и ниже

$$\hat{x}=(\hat{a}^1,\ldots,\hat{a}^\tau,x)\in\hat{X};\quad \hat{\Gamma}\in\mathscr{B}(\hat{X});\quad \Gamma_{y_1}\triangleq\{y_2\in Y_2\,|\,(y_1,y_2)\in\Gamma\}$$

сечение множества $\Gamma\subseteq Y_1\times Y_2$. Очевидно, что если исходная модель Z полунепрерывна, то построенная модель $\hat{Z}$ также будет полунепрерывной [1, 2]. Понятия истории $\hat{h}_t\in\hat{H}_t$, стратегии $\hat{\pi}$ и σ-алгебры $\hat{\mathscr{F}}_t$ в $\hat{\Omega}\triangleq\hat{H}_\infty$ вводятся по аналогии с п. 2. (См. [1, 2].) При фиксированном начальном распределении $\hat{P}^0$ всякая стратегия $\hat{\pi}$ однозначно определяет вероятностную меру $\hat{P}^{\hat{\pi}}$ на $\hat{\Omega}$.

Определение. Пусть $\{\mu_t\}_{t=1}^\tau$ — последовательность универсально измеримых стохастических ядер

$$\mu_t(\Gamma\,|\,P^0,h_{t-1})\text{ на }A(\mu_1:B\to P(A)).$$

Символом $\varphi(P^0,\{\mu_t\}_{t=1}^\tau$ обозначим вероятностную меру $\hat{P}^0$ на $\hat{X}$, значения которой на измеримых прямоугольниках определяются следующим образом:

$$\begin{aligned}\hat{P}^0\{\hat{a}^1\in\Gamma_1^{\hat{A}},\ldots,\hat{a}^\tau\in\Gamma_\tau^{\hat{A}},\xi\in\Gamma^X\}\triangleq P^0(\Gamma^X)\int\limits_{\Gamma_1^{\hat{A}}}\ldots\int\limits_{\Gamma_{\tau-1}^{\hat{A}}}\mu_\tau(\Gamma_\tau^{\hat{A}}\,|\,P^0,\\ \hat{a}^1\ldots\hat{a}^{\tau-1})\mu_{\tau-1}(d\hat{a}^{\tau-1}\,|\,P^0,\hat{a}^1\ldots\hat{a}^{\tau-2})\ldots\mu_1(d\hat{a}^1\,|\,P^0).\end{aligned}\tag{5}$$

Определение. Пусть $\{\mu_t\}_{t=\tau+1}^\infty$ — последовательность универсально измеримых ядер $\mu_t(\Gamma\,|\,P^0,h_{t-1})$ на A. Символом $\gamma(P^0,\{\mu_t\}_{t=\tau+1}^\infty)$ обозначим стратегию $\hat{\pi}=\{\nu_t\}_{t=1}^\infty$ в модели $\hat{Z}$, задаваемую (при каждом $P^0\in B$) следующим образом:

$$\begin{aligned}\nu_t(\Gamma\,|\,\hat{\xi}_0\hat{a}_1\hat{\xi}_1\ldots\hat{a}_{t-1}\hat{\xi}_{t-1})\triangleq\\ \triangleq\mu_{\tau+t}(\Gamma\,|\,P^0,\hat{a}_0^1\hat{a}_0^2\ldots\hat{a}_0^\tau\xi_0\hat{a}_1\xi_1\ldots\hat{a}_{t-1}\xi_{t-1}).\end{aligned}\tag{6}$$

Здесь и ниже $\hat{\xi}_i=(\hat{a}_i^1,\ldots,\hat{a}_i^\tau,\xi_i)$.

Определение. Символом $\hat{\Omega}'\subset\hat{\Omega}$ обозначается класс последовательностей $\hat{\xi}_0\hat{a}_1\hat{\xi}_1\ldots$, для которых элементы

$$\hat{\xi}_t=(\hat{a}_t^1,\ldots,\hat{a}_t^\tau,\xi_t),\ \hat{a}_{t+1},\ \hat{\xi}_{t+1}=(\hat{a}_{t+1}^1,\ldots,\hat{a}_{t+1}^\tau,\xi_{t+1})$$

удовлетворяют равенствам:

$$\hat{a}_{t+1}^1=\hat{a}_t^2;\ \hat{a}_{t+1}^2=\hat{a}_t^3;\ldots;$$

$$\hat{a}_{t+1}^{\tau-1}=\hat{a}_t^\tau;\quad \hat{a}_{t+1}^\tau=\hat{a}_{t+1}\qquad(t=0,1,2,\ldots).$$

Символом ψ обозначается биекция из Ω в $\hat{\Omega}'$, задаваемая формулой

$$\Psi(a_1 a_2 \ldots a_\tau \xi_0 a_{\tau+1} \xi_1 \ldots) =$$

$$= \{(a_1 a_2 \ldots a_\tau \xi_0) a_{\tau+1} (a_2 \ldots a_\tau a_{\tau+1} \xi_1) \ldots\}.$$

Легко проверить, что для всякой стратегии $\hat{\pi}$ при любом начальном распределении $\hat{P}^0 \in P(\hat{X})$ мера $\hat{P}^{\hat{\pi}}$ сосредоточена на $\hat{\Omega}'$.

Лемма 2. Пусть $P^0 \in \tilde{P}(X)$, $\pi = \{\mu_t\}_{t=1}^{\infty}$ — стратегия в Z;

$$\hat{P}^0 = \varphi(P^0, \{\mu_t\}_{t=1}^{\tau}); \quad \hat{\pi} = \gamma(P^0, \{\mu_t\}_{t=\tau+1}^{\infty}).$$

Тогда

1) мера $\hat{P}^{\hat{\pi}}$, отвечающая $\hat{P}^0$ и $\hat{\pi}$, и мера P^{π}, отвечающая P^0, π обладает свойством: $\hat{P}^{\hat{\pi}}(\hat{\Gamma}') = P^{\pi}(\Psi^{-1}(\hat{\Gamma}'))$;

2) $$\mathbf{M}^{\pi}\left[\sum_{\theta=t+1}^{T} e^{-\alpha\theta} \cdot r(\xi_{\theta-1}, a_\theta) \mid \Psi^{-1}(\hat{\mathscr{F}}_t)\right] =$$

$$= M^{\hat{\pi}}\left[\sum_{\theta=t+1}^{T} e^{-\alpha\theta} \cdot \hat{r}(\hat{\xi}_{\theta-1}, \hat{a}_\theta) \mid \hat{\mathscr{F}}_t\right]$$

$\hat{P}^{\hat{\pi}}$ – п. н.,

$t = 0, 1, 2, \ldots$

Замечание. В п. 2), очевидно, $\Psi^{-1}(\hat{\mathscr{F}}_t) = \mathscr{F}_t$, и совпадение условных математических ожиданий понимается как совпадение двух функций на Ω и на $\hat{\Omega}'$ при подстановке $\omega = \Psi^{-1}(\hat{\omega}')$.

Доказательство следует из единственности меры $\hat{P}^{\hat{\pi}}$ и свойств образов вероятностных мер [8].

Функция

$$\hat{V}_t(\hat{h}_t) \triangleq \hat{P}^{\hat{\pi}^t} - \operatorname*{ess\,sup}_{\hat{\pi}_t} \lim_{T\to\infty} M^{\hat{\pi}_t}\left[\sum_{\theta=t+1}^{I} e^{-\alpha\theta} \cdot \hat{r}(\hat{\xi}_{\theta-1}, \hat{a}_\theta) \mid \hat{\mathscr{F}}_t\right]$$

называется оценкой модели $\hat{Z}$. (Определение операции $\hat{P}^{\hat{\pi}^t}$ — ess sup аналогично приведенному в п. 2.) Супремум берется по ядрам $\hat{\pi}_t = \{\nu_\theta\}_{\theta=t+1}^{\infty}$, $\hat{\pi}_t$. Как известно [1], в полунепрерывной модели функция $\hat{V}_t(\hat{h}_t)$ зависит только от t, $\hat{\xi}_t$ и является ограниченной сверху и полунепрерывной сверху. Кроме того, существует оптимальный стационарный B-селектор $\hat{\varphi}^*(\hat{\xi}_{t-1})$, для которого $\forall\, \hat{P}^0\ \forall\, t \geqq 0$

$$\hat{V}(\hat{\xi}_t)\hat{P}^{\hat{\varphi}^*} - \text{п. н.} = \lim_{T\to\infty} M^{\hat{\varphi}^*}\left[\sum_{\theta=t+1}^{I} e^{-\alpha\theta} \cdot \hat{r}(\hat{\xi}_{\theta-1}, \hat{a}_\theta) \mid \hat{\mathscr{F}}_t\right];$$

$$\int_{\hat{X}} \hat{V}(\hat{\xi}_0)\hat{P}^0(d\hat{\xi}_0) = \sup_{\hat{\pi}}\left\{\lim_{T\to\infty} M^{\hat{\pi}}\left[\sum_{\theta=1}^{T} e^{-\alpha\theta} \cdot \hat{r}(\hat{\xi}_{\theta-1}, \hat{a}_\theta)\right]\right\}.$$

Из п. 2) леммы 2 следует, что

$$P^{\pi^{t+\tau}} - \operatorname*{ess\,sup}_{\pi_{t+\tau}} \lim_{T\to\infty} M^{\pi_{t+\tau}}\Bigg[\sum_{\theta=t+1}^{I} e^{-\alpha\theta}\times$$

$$\times r(\xi_{\theta-1}, a_\theta)|\mathscr{F}_t\Bigg]P^{\pi^{t+\tau}} - \text{п. н.} = \hat{V}_t(\{a_{t+1}, \ldots, a_{t+\tau}, \xi_t\}). \quad (\forall\, P^0)$$

Доказательство теоремы 2. Не теряя общности, можно считать, что оптимальный в $\hat{Z}$ стационарный B-селектор имеет вид $\hat{\varphi}^* = \gamma(P^0, \varphi^*)$, где $\varphi^*(h_{t-1})$ — «часть» некоторого B-селектора в $Z(t>\tau)$, не зависящего от P^0. Функция

$$\Phi(a_1, \ldots, a_\tau, P^0) \triangleq \int_X V_0(\xi_0, a_1, \ldots, a_\tau)P^0(d\xi_0)$$

полунепрерывна сверху и ограничена сверху [1]. Следовательно, согласно лемме Янкова–Фон Неймана [1, 2], существуют борелевские функции $\varphi_1^*(P^0), \ldots, \varphi_\tau^*(P^0)$ такие, что

$$\Phi(\varphi_1^*(P^0), \ldots, \varphi_\tau^*(P^0), P^0) = \sup_{(a_1,\ldots,a_\tau)\in A^\tau} \Phi(a_1, \ldots, a_\tau, P^0).$$

Легко проверить, что B-селектор $\pi^* = \{\varphi_1^*, \ldots, \varphi_\tau^*, \varphi^*\}$ в модели Z равномерно оптимален по определению, причем

$$W(\pi^*) = \sup_{(a_1,\ldots,a_\tau)\in A^\tau} \int_X V_0(\xi_0, a_1, \ldots, a_\tau)P^0(d\xi_0).$$

Теорема 2 доказана.

Доказательство теоремы 3. Достаточность вытекает из теоремы 2 и очевидного равенства

$$W(\pi) = \sup_{(a_1,\ldots,a_\tau)\in A^\tau} \int_X V_0(\xi_0, a_1, \ldots, a_\tau)P^0(d\xi_0).$$

Допустим, что одно из условий 1), 2) нарушено. Тогда с помощью леммы 2 можно получить соотношение

$$W(\pi) < \sup_{(a_1,\ldots,a_\tau)\in A^\tau} \int_X V_0(\xi_0, a_1, \ldots, a_\tau)P^0(d\xi_0),$$

и по теореме 2 стратегия π не оптимальна.

7. Доказательство леммы 1

Зафиксируем произвольное $t(0 \leqq t \leqq \tau)$.

Функция $f_t(y, a) \triangleq r(y, a) + \int\limits_X F_t(z) P_a(y, dz)$ полунепрерывна сверху и ограничена сверху [1]. Введем отображения

$$\psi_1 : B \times A \to P(X \times A) \text{ и } \psi_2 : P(X \times A) \to R^1$$

по следующим формулам: $\psi_1(P, a) \triangleq P(\cdot)\delta_a(\cdot)$, где $\delta_a(\cdot)$ — вероятностная мера, сосредоточенная в точке a; $\psi_2(P) \triangleq \int\limits_{X \times A} f_t(y, a) P(d(y, a))$. Отображение ψ_1 непрерывно; ψ_2 — ограниченная сверху, полунепрерывная сверху функция [1]. Следовательно, функция

$$\Phi_t(P, a) \triangleq \int\limits_X f_t(y, a) P(dy) = \psi_2(\psi_1(P, a)): \quad B \times A \to R^1$$

ограничена сверху и полунепрерывна сверху. Утверждения 1), 2) следуют из леммы Янкова–Фон Неймана [1, 2]. Справедливость п. 3) следует из равенства

$$\Phi_t(P, \varphi_t(P)) = \sup_{a \in A} \Phi_t(P, a),$$

которое имеет место при всех $P \in B$.

8. Доказательство теоремы 4

Лемма 3. Пусть начальное распределение P^0 фиксировано и набор $\{F_i\}_{i=0}^{\tau}$ функций из D удовлетворяет условиям У4.

Тогда $\forall \pi \in \Pi$. Справедливо равенство:

$$\lim_{T \to \infty} M^\pi \left[\sum_{\theta=t+1}^{T} e^{-\alpha\theta} \cdot r(\xi_{\theta-1}, a_\theta) | \mathscr{F}_{t-\tau} \right] P^\pi - \text{п. н.} = e^{-\alpha t} \int\limits_X F_{t \wedge \tau}(z) P_t^\pi(dz, P^0, \omega) +$$

$$+ \lim_{T \to \infty} M^\pi \left[\sum_{\theta=t+1}^{T} e^{-\alpha\theta} \cdot M^\pi [r(\xi_{\theta-1}, a_\theta) - e^\alpha F_{(\theta-1) \wedge \tau}(\xi_{\theta-1}) + \int\limits_X F_{\theta \wedge \tau}(y) \times \right. \tag{(7)}$$

$$\left. \times P_{a_\theta}(\xi_{\theta-1}, dy) | \mathscr{F}_{\theta-\tau-1} \right] \Big| \mathscr{F}_{t-\tau} \Big] - \lim_{T \to \infty} M^\pi [e^{-\alpha T} F_\tau(\xi_T) | \mathscr{F}_{t-\tau}], \quad t = 0, 1, 2, \ldots$$

Доказательство. Рассмотрим модель с конечным горизонтом $\{0, 1, \ldots, T\}$ и нулевым финальным доходом, в которой

$$P_a(t, x, dy) \triangleq P_a(x, dy); \quad R(t, x, a) \triangleq e^{-\alpha t} r(x, a).$$

В соответствии с [6] из условий У4-а, б) следует, что $\forall \pi \in \Pi$

$$M^{\pi}\left[\sum_{\theta=t+1}^{T} e^{-\alpha\theta} r(\xi_{\theta-1}, a_{\theta}) | \mathscr{F}_{t-\tau}\right] P^{\pi} - \text{п. н.} = e^{-\alpha t} \int_X F'_t(z) P^{\pi}_t(dz, P^0, \omega) + .$$

$$+ M^{\pi}\left[\sum_{\theta=t+1}^{T} M^{\pi}[e^{-\alpha\theta} r(\xi_{\theta-1}, a_{\theta}) - e^{-\alpha(\theta-1)} F'_{\theta-1}(\xi_{\theta-1}) + e^{-\alpha\theta} \int_X F'_{\theta}(y) \times \right.$$

$$\left. \times P_{a_{\theta}}(\xi_{\theta-1}, dy) | \mathscr{F}_{\theta-\tau-1}] - e^{-\alpha T} F'_T(\xi_T) | \mathscr{F}_{t-\tau}\right], \quad t = 0, 1, \ldots, T,$$

где $F'_t(y) \triangleq F_{t \wedge \tau}(y)$. Равенство (7) получается с помощью предельного перехода с использованием условия У4-в).

Замечание. Если условия У4 выполнены для всех стратегий (а не только для $\pi \in \Pi$), то утверждение леммы 3 в определенном смысле можно обратить: из существования конечных условных математических ожиданий и пределов в правой части (7) следует, что $\pi \in \Pi$. Например, это утверждение справедливо, если функции F_i ограничены.

Доказательство теоремы 4. В силу леммы 3 и определения оценки стратегии имеем:

$$W(\pi^*) = \int_X F_0(y) P^0(dy) - \lim_{T \to \infty} M^{\pi^*}[e^{-\alpha T} F_{\tau}(\xi_T)].$$

С другой стороны, $\forall \pi \in \Pi$ согласно (7)

$$W(\pi) = \int_X F_0(y) P^0(dy) + \lim_{T \to \infty} M^{\pi}\left[\sum_{\theta=1}^{T} e^{-\alpha\theta} M^{\pi}[r(\xi_{\theta-1}, a_{\theta}) - e^{\alpha} \cdot F_{(\theta-1) \wedge \tau}(\xi_{\theta-1}) + \right.$$

$$\left. + \int_X F_{\theta \wedge \tau}(y) P_{a_{\theta}}(\xi_{\theta-1}, dy) | \mathscr{F}_{\theta-\tau-1}]\right] - \lim_{T \to \infty} M^{\pi}[e^{-\alpha T} F_{\tau}(\xi_T)] \leqq$$

$$\leqq \int_X F_0(y) P^0(dy) - \lim_{T \to \infty} M^{\pi^*}[e^{-\alpha T} F_{\tau}(\xi_T)] = W(\pi^*),$$

что и требовалось.

9. Доказательство теоремы 6

Заметим, что в рассматриваемом случае допустимы все стратегии.

Пусть выполнено утверждение 1). Докажем, что для функций $\{F_i\}_{i=0}^{\tau}$ справедлива гипотеза H_2. Действительно,

$$\forall \pi \forall t = 1, 2, \ldots \quad e^{\alpha} \int_X F_{(t-1)\wedge\tau}(z) P_{t-1}^{\pi^{t-1}}(dz, P^0, \omega) = e^{\alpha t} \tilde{\Phi}_{t-1}(P_{t-1}^{\pi^{t-1}}) =$$

$$e^{\alpha t} \times P^{\pi^{t-1}} - \operatorname*{ess\,sup}_{\pi_{t-1}} \lim_{T\to\infty} M^{\pi_{t-1}} \left[\sum_{\theta=1}^{T} e^{-\alpha\theta} r(\xi_{\theta-1}, a_\theta) | \mathscr{F}_{t-\tau-1} \right] =$$

$$P^{\pi^{t-1}} - \operatorname*{ess\,sup}_{\pi_{t-1}} \left\{ M^{\pi_{t-1}}[r(\xi_{t-1}, a_t) | \mathscr{F}_{t-\tau-1}] + \right.$$

$$\left. + M^{\pi_{t-1}} \left[e^{\alpha t} \lim_{T\to\infty} M^{\pi_t} \left[\sum_{\theta=t+1}^{T} e^{-\alpha\theta} r(\xi_{\theta-1}, a_\theta) | \mathscr{F}_{t-\tau} \right] \middle| \mathscr{F}_{t-\tau-1} \right] \right\} =$$

$$= P^{\pi^{t-1}} - \operatorname*{ess\,sup}_{a_t \in A} \left\{ M^{\pi}[r(\xi_{t-1}, a_t) | \mathscr{F}_{t-\tau-1}] + M^{\pi}[e^{\alpha t} \Phi_t(P_t^{<\pi^{t-1}, a_t>}) | \mathscr{F}_{t-\tau-1}] \right\} =$$

$$= P^{\pi} - \operatorname*{ess\,sup}_{a\in A} \{ M^{\pi}[r(\xi_{t-1}, a) | \mathscr{F}_{t-\tau-1}] + M^{\pi}[\int_X F_{t\wedge\tau}(z) P_a(\xi_{t-1}, dz) | \mathscr{F}_{t-\tau-1}] \}.$$

(Все равенства выполнены P^{π} – п. н.)

Пусть выполнено утверждение 2). Докажем 1). В силу леммы 3

$$\lim_{T\to\infty} M^{\pi_t} \left[\sum_{\theta=t+1}^{I} e^{-\alpha\theta} \cdot r(\xi_{\theta-1}, a_\theta) | \mathscr{F}_{t-\tau} \right] P^{\pi^t} - \text{п. н.} \leqq e^{-\alpha t} \int_X F_{t\wedge\tau}(z) P_t^{\pi^t}(dz, P^0, \omega).$$

С другой стороны, из леммы Янкова–Фон Неймана следует, что найдется стационарный марковский A-селектор φ такой, что

$$\int_X [r(y, \varphi(\theta, P)) - e^{\alpha} \cdot F_{(\theta-1)\wedge\tau}(y) + \int_X F_{\theta\wedge\tau}(z) P_{\varphi(\theta,P)}(y, dz)] P(dy) > -\varepsilon.$$

(Для P^{φ} — почти всех мер $P \in B$ в любой момент $\theta > 0$). Следовательно, в соответствии с леммой 3,

$$P^{\pi^t} - \operatorname*{ess\,sup}_{\pi_t} \lim_{T\to\infty} M^{\pi_t} \left[\sum_{\theta=t+1}^{I} e^{-\alpha\theta} \cdot r(\xi_{\theta-1}, a_\theta) | \mathscr{F}_{t-\tau} \right] \geqq$$

$$\geqq e^{-\alpha t} \int_X F_{t\wedge\tau}(z) P_t^{\pi^t}(dz, P^0, \omega) = \tilde{\Phi}_t(P_t^{\pi^t}) \, P^{\pi^t} - \text{п. н.}$$

Теорема доказана.

Литература

1. *Бертсекас Д., Шрив С.* Стохастическое оптимальное управление. М., «Наука», 1985.
2. *Юшкевич А. А., Читашвили Р. Я.* Управляемые случайные последовательности и цепи Маркова. «Успехи матем. наук», 1982, **37**, вып. *6*, с. 213–242.
3. *Липцер Р. Ш., Ширяев А. Н.* Статистика случайных процессов. М., «Наука», 1974.
4. *Майн Х., Осаки С.* Марковские процессы принятия решений. М., «Наука», 1977.
5. *Дынкин Е. Б., Юшкевич А. А.* Управляемые марковские процессы и их приложения. М., «Наука», 1975.
6. *Aliseenko, O. V., Khametov, V. M., Piunovski, A. B.*, The optimal control of stochastic sequences with delay. Problems of Control and Information Theory, 1986, **15,** *1*, pp. 3–17.
7. *Власюк Б. А., Штернберг А. А.* Структурные свойства оптимальных линейных стохастических систем с запаздываниями в управлениях. «Автом. и телемех.», 1983, *3*, с. 65–75.
8. *Мейер А.-А.* Вероятность и потенциалы. М., «Мир», 1973.

PRINTED IN HUNGARY
Akadémiai Kiadó és Nyomda Vállalat, Budapest

NOTE TO CONTRIBUTORS

Two copies of the *manuscript* (each complete with figures, tables and references) are to be sent to

E.D. Teryaev coordinating editor
Department of Mechanics and Control Processes
Academy of Sciences of the USSR
Leninsky Prospect 14, Moscow V-71, USSR

or to V. Strejc
UTIA ČSAV
18208 Prague 8
Pod vodárenskou věží 4, Czechoslovakia

or to L. Györfi
Technical University of Budapest
H-1111 Budapest, Stoczek u. 2, Hungary

Authors are requested to retain a third copy of the submitted typescript to be able to check the proofs.

The papers, preferably in English or Russian, should be typed double spaced on one side of good-quality paper with wide margins (4–5 cm). The first page of the paper should carry the title, the author(s)' names and the name of the town where they are active. The name and address of the author to whom the proofs should be sent should be given at the end of the paper. An *abstract* should head the paper. English papers should also have a Russian abstract.

The papers should not exceed 15 pages (25 × 50 characters per page) including tables and references. The proper location of the tables and figures must be indicated on the margin.

Mathematical notations should follow up-to-date usage. Equations longer than half a line should not be incorporated in the text. In-text equations must be typed on a single line except that one level of subscripting and/or superscripting is permissible. Use / instead of horizontal bars. Displayed equations should be written so as to require the fewest possible lines. Therefore use "exp" for the exponential function whenever the exponent requires more than a single line. Matrices should, if possible, not be written in full. Use subscript notations instead such as $A = ||a_{ij}||$. Write diagonal matrices as diag $(d_1, d_2, \ldots d_n)$.

The authors will be sent galley proofs to be returned by next mail. Rejected manuscripts will be returned. Authors will receive 100 reprints free of charge. Additional reprints may be ordered.

К СВЕДЕНИЮ АВТОРОВ

Рукописи статей в трех экземплярах на русском языке и в трех на английском следует направлять по адресу: 117312 Москва В-312, просп. 60 летия Октября, 9, МНИИПУ. Редакция журнала «Проблемы управления и теории информации» (зав. редакцией Н. И. Родионова, тел. 208-60-19).

Объём статьи не должен превышать 15 печатных страниц (25 строк по 50 букв). Статье должна предшествовать аннотация объемом 50–100 слов и приложено резюме–реферат объемом не менее 10–15% объема статьи на русском языке в трех экземплярах, на котором напечатан служебный адрес автора (фамилия, название учреждения, адрес).

При написании статьи авторам надо строго придерживаться следующей формы: введение (постановка задачи), основное содержание, примеры практического использования, обсуждение результатов, выводы и литература.

Статьи должны быть отпечатаны с промежутком в два интервала, последовательность таблиц и рисунков должна быть отмечена на полях. Математические обозначения рекомендуется давать в соответствии с современными требованиями и традициями. Разметку букв следует производить только во втором экземпляре и русского, и английского варианта статьи.

Авторам высылается верстка, которую необходимо незамедлительно проверить и возвратить в редакцию.

После публикации авторам высылаются бесплатно 100 оттисков их статей.

Рукописи непринятых статей возвращаются авторам.

CONTENTS · СОДЕРЖАНИЕ

HU ISSN 0370-2529 Index 26 660

VOL. TOM 17 · NUMBER HOMEP 6

ACADEMY OF SCIENCES OF THE USSR
HUNGARIAN ACADEMY OF SCIENCES
CZECHOSLOVAK ACADEMY OF SCIENCES

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

АКАДЕМИЯ НАУК СССР
ВЕНГЕРСКАЯ АКАДЕМИЯ НАУК
ЧЕХОСЛОВАЦКАЯ АКАДЕМИЯ НАУК

1988

AKADÉMIAI KIADÓ, BUDAPEST
DISTRIBUTED OUTSIDE THE COMECON-COUNTRIES
BY PERGAMON PRESS, OXFORD

PROBLEMS OF CONTROL AND INFORMATION THEORY

An international bi-monthly sponsored jointly by the Presidium of the Academy of Sciences of the USSR, of the Hungarian Academy of Sciences and of the Czechoslovak Academy of Sciences. The six issues published per year make up a volume of some 480 pp. It offers publicity for original papers and short communication of the following topics:

— theory of control processes
— theory of adaptive systems
— theory of estimation and identification
— theory of controlling robot-technologic and flexible manufacturing systems
— information theory
— information-theoretic aspects of multiple access networks.

While this bi-monthly is mainly a publication forum of the research results achieved in the socialist countries, also papers of international interest from other countries are welcome.

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

Международный журнал Академии наук СССР, Венгерской Академии наук и Чехословацкой Академии наук выходит 6 раз в год общим объемом 480 печатных страниц.

В журнале публикуются оригинальные научные статьи и статьи обзорного характера по следующим проблемам управления и теории информации:

— теория процессов управления;
— теория адаптивных систем;
— теория оценивания и идентификации;
— теория управления робототехническими и гибкими производственными системами;
— теория информации;
— теория информации в области сетей с множественным доступом.

Целью журнала является ознакомление научной общественности различных стран с важнейшими проблемами, имеющими актуальный и перспективный характер, научными достижениями ученых социалистических и других стран.

Distributors

For the Soviet Union:

SOYUZPECHATY, Moscow 123 308 USSR

For Albania, Bulgaria, China, Cuba, Czechoslovakia, German Democratic Republic, Korean People's Republic, Mongolia, Poland, Rumania, Vietnam and Yugoslavia:

KULTURA Hungarian Foreign Trading Co.
P. O. Box 149 H-1389 Budapest, Hungary

For all other countries:

PERGAMON JOURNALS LTD, Headington Hill Hall, Oxford OX3 OBW, England

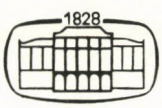
or

PERGAMON JOURNALS INC, Maxwell House, Fairview Park, Elmsford, NY 10523, USA

1988 Subscription Rate DM 535,— per annum including postage and insurance.

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ



AKADÉMIAI KIADÓ

PUBLISHING HOUSE OF THE HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST

Problems of Control and Information Theory, Vol. 17 (6), pp. 333–343 (1988)

ON THE PROGRAM SYNTHESIS OF A GUARANTEED CONTROL

N. N. KRASOVSKII, T. N. RESHETOVA

(*Sverdlovsk*)

(Received December 17, 1987)

In the paper the problem of forming a guaranteed control for a dynamical system [1] is considered. Forming of such a control according to the method of stochastic program synthesis [1–4] is reduced to recurrent procedures of designing concave hulls for auxiliary functions arising in the program construction. An example of computer simulation of a control process for a model problem is given.

1. Statement of the problem

Consider a system described by the differential equation

$$\dot{x} = A(t)x + B(t)u + C(t)v, \quad t_0 \leqq t \leqq \vartheta. \tag{1.1}$$

Here x is the n-dimensional state vector of the object, u is the r-dimensional control vector, v is the s-dimensional hindrance vector; $A(t)$, $B(t)$, and $C(t)$ are continuous matrix functions; values u and v are constrained by the conditions

$$u \in P, \quad v \in Q \tag{1.2}$$

where P and Q are compact.

A quality index

$$\gamma = |x(\vartheta)| + \int_{t_*}^{\vartheta} \chi(t, u[t], v[t]\mathrm{d}t \tag{1.3}$$

is given; here $|x|$ denotes some norm of vector x, function $\chi(t, u, v)$ is continuous.

According to [1], the problem of finding an optimal minimax strategy $u^0(t, x, \varepsilon)$ guaranteeing the optimal result $\rho^0(t_*, x_*)$ for any possible initial position (t_*, x_*), has a solution. The step-by-step control law U which forms control forces

$$u[t] = u^0(t_i, x[t_i], \varepsilon)\, t_i \leqq t < t_{i+1},$$

$$i = 1, \ldots, k,\ t_1 = t_*,\ t_{k+1} = \vartheta$$

ensures the inequality

$$\gamma \leqq \rho^0(t_*, x_*) + \xi \tag{1.4}$$

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

for any $\xi > 0$ chosen beforehand, provided that parameter $\varepsilon > 0$ and step $\delta = \max_i(t_{i+1} - t_i)$ are sufficiently small. The strategy $u^0(t, x, \varepsilon)$ is constructed as the extremal one to the function $\rho^0(t, x)$. Function $\rho^0(t, x)$ can be calculated by the method of stochastic program synthesis according to the equality

$$\rho^0(t, x) = \lim_{\delta \to 0} \sup_{\|l(\cdot)\| \leq 1} [\langle m_0 \cdot X[\vartheta, t]x \rangle + \tag{1.5}$$

$$+ M \left\{ \sum_{i=1}^{k} \int_{\tau_i}^{\tau_{i+1}} \min_{u \in P} \max_{v \in Q} [\langle m(\tau_i, \omega) \cdot X[\vartheta, \tau](B(\tau)u + C(\tau)v) \rangle + \chi(\tau, u, v)] d\tau \right\}.$$

Here points τ_i, $i = 1, \ldots, k$ form a partition Δ_δ of the segment $[t_*, \vartheta]$ with the step $\delta = \max_i(\tau_{i+1} - \tau_i)$; $l(\cdot) = \{l_1(\omega), \ldots, l_n(\omega), \omega \in \Omega\}$ is an n-dimensional random variable on the probability space $\{\Omega, F, P\}$ where an elementary event $\omega = \{\xi_1, \ldots, \xi_k\}$ is a collection of jointly independent random variables $0 \leqq \xi_i \leqq 1$ distributed uniformly and realized at time instants τ_i; $\|l(\cdot)\|$ denotes the norm

$$\|l(\cdot)\| = \text{vrai sup}_\omega |l(\omega)|^* \tag{1.6}$$

of a random variable $l(\cdot)$ where $|l|^*$ is a norm of vector l conjugate to the norm $|x|$ occurring in (1.3);

$$m_0 = M\{l(\cdot)\}, m(\tau_i, \omega) = M\{l(\cdot) | \xi_1(\omega), \ldots, \xi_i(\omega)\}, \omega \in \Omega, \qquad i = 1, \ldots, k$$

where symbols $M\{\ldots\}$ and $M\{\ldots|\ldots\}$ denote mathematical expectation and conditional mathematical expectation; $X[t, \tau]$ is the fundamental solution matrix for the equation $dx/dt = A(t)x$; symbol $\langle a \cdot b \rangle$ denotes the scalar product of vectors a and b. Equality (1.5) and the values occurring in it are commented in detail in [1].

So, in order to form the control $u^0(t_i, x[t_i], \varepsilon)$, optimal whith respect to the guarantee $\rho^0(t_*, x_*)$ it is sufficient to calculate efficiently the values standing in (1.5) under the limit sign for current values $t = t_i$. In the present paper a method of calculations is considered.

2. The recurrent estimation

Introduce vectors w and z where

$$w = \{w_1, \ldots, w_n\}, \; z = \{z_1, \ldots, z_n, z_{n+1}\} = \{w_1, \ldots, w_n, z_{n+1}\} = \{w, z_{n+1}\}.$$

Denote

$$\psi_i(m) = \int_{\tau_i}^{\tau_{i+1}} \min_{u \in P} \max_{v \in Q} [\langle m \cdot X[\vartheta, \tau](B(\tau)u + C(\tau)v) \rangle + \chi(\tau, u, v)] d\tau, \tag{2.1}$$

$$i = 1, \ldots, k.$$

According to the method of the program synthesis, one should calculate the program extremum [1]

$$e(t, z, \Delta_\delta) = \sup_{\|u(\cdot)\| \leq 1} \left[\langle m_0 \cdot X[\vartheta, t]w \rangle + z_{n+1} + M \left\{ \sum_{i=1}^{k} \psi_i(m(\tau_i, \cdot)) \right\} \right]. \tag{2.2}$$

We define the upper concave hull of a function $\xi(m)$, $|m|^* \leq 1$ to be the function $\varphi(m) = \{\xi(\cdot)\}_*$ satisfying the following conditions.

1. The function $\varphi(m)$ is concave for $|m|^* \leq 1$, i.e. for any $m^{(1)}$, $|m^{(1)}|^* \leq 1$, and $m^{(2)}$, $|m^{(2)}|^* \leq 1$ and any $\lambda \in [0, 1]$ the inequality

$$\varphi(\lambda m^{(1)} + (1-\lambda)m^{(2)}) \geq \lambda\varphi(m^{(1)}) + (1-\lambda)\varphi(m^{(2)}) \tag{2.3}$$

is true.

2. The inequality

$$\xi(m) \leq \varphi(m) \tag{2.4}$$

holds for any m, $|m|^* \leq 1$.

3. Let $r = \{r_1, \ldots, r_n\}$ be an n-dimensional vector. For any m, $|m|^* \leq 1$ there exists a probability measure $\mu(R|m)$ on the set $|r|^* \leq 1$ such that the equalities

$$\varphi(m) = \int_{|r|^* \leq 1} \xi(r)\mu(\mathrm{d}r|m), \tag{2.5}$$

$$m = \int_{|r|^* \leq 1} r\mu(\mathrm{d}r|m) \tag{2.6}$$

are true.

Actually, one can choose a measure $\mu(R|m)$ concentrated on a finite collection of points $r^{[j]}(m)$. This simplifies the practical constructions.

Let us form the recurrent sequence of functions

$$\varphi_k(m) = \{\psi_k(\cdot)\}_*, \tag{2.7}$$

$$\varphi_i(m) = \{\psi_i(\cdot) + \varphi_{i+1}(\cdot)\}_*, \quad |m|^* \leq 1, \qquad i = 1, \ldots, k-1. \tag{2.8}$$

The following equality holds:

$$\sup_{\|u(\cdot)\| \leq 1,\, M\{u(\cdot)\} = m_0} M \left\{ \sum_{i=1}^{k} \psi_i(m(\tau_i, \cdot)) \right\} = M\{\varphi_1(m_0)\} = \varphi_1(m_0). \tag{2.9}$$

From (2.2), (2.9), we get the equality

$$e(t, z, \Delta_\delta) = \sup_{|m_0|^* \leq 1} [\langle m_0 \cdot X[\vartheta, t]w \rangle + z_{n+1} + \varphi_1(m_0)]. \tag{2.10}$$

This equality together with equality (1.5) shows that the calculation of the value $e(t, z, \Delta_\delta)$, which approximates the value $\rho^0(t, x)$ is reduced to the procedure of forming the sequence of functions $\varphi_j(m)$.

Let us prove equality (2.9).

Fix a certain value m_0, $|m_0|^* \leqq 1$. Construct the random variable $m^{(1)}(\omega) = m^{(1)}[\xi_1]$ as follows. Its distribution is determined by the probability measure $\mu^{(1)}(R|m_0)$ which, according to conditions (2.5), (2.6), satisfies the equalities

$$\varphi_1(m_0) = \int\limits_{|r|^* \leqq 1} [\psi_1(r) + \varphi_2(r)]\mu^{(1)}(\mathrm{d}r|m_0), \tag{2.11}$$

$$m_0 = \int\limits_{|r|^* \leqq 1} r\mu^{(1)}(\mathrm{d}r|m_0). \tag{2.12}$$

Random variables $m^{(j)}$, $j = 2, \ldots, k$ are defined inductively. Let the random variable $m^{(j)}[\xi_1, \ldots, \xi_j]$, $j \leqq i$ be already constructed. We determine the random variable $m^{(i+1)}[\xi_1, \ldots, \xi_{i+1}]$ by its conditional probability measure $\mu^{(i+1)}(R|m^{(i)}[\xi_1, \ldots, \xi_i])$ which satisfies the following conditions

$$\varphi_{i+1}(m^{(i)}) = \int\limits_{|r|^* \leqq 1} [\psi_{i+1}(r) + \varphi_{i+2}(r)]\mu^{(i+1)}(\mathrm{d}r|m^{(i)}), \tag{2.13}$$

$$m^{(i)} = \int\limits_{|r|^* \leqq 1} r\mu^{(i+1)}(\mathrm{d}r|m^{(i)}). \tag{2.14}$$

This construction does not lead to principal difficulties since, for the cases of interest, each measure $\mu^{(j)}(R|m^{(j-1)})$ is concentrated on a finite collection of points $r_{[j]}^{[s]}(m)$. However, it is not necessary to take into account the finiteness of the set of points $r_{[j]}^{[s]}(m)$ in theoretical constructions. Indeed, if we apply in the usual way the theorems on measurable selectors [5], we can verify that needed measures $\mu^{(j)}(r|m^{(j-1)})$ can be taken measurable (in $m^{(j-1)}$, in the appropriate sense) without assuming measures $\mu^{(i)}$ to be concentrated on finite sets of points $r_{[j]}^{(s)}(m)$. Therefore, these measures can indeed play the role of conditional measures.

Thus, we may suppose that we have the sequence of random variables m_0, $m^{(i)}[\xi_1, \ldots, \xi_i]$, $i = 1, \ldots, k$ satisfying conditions (2.11)–(2.14). If we put

$$l(\omega) = l[\xi_1, \ldots, \xi_k] = m^{(k)}[\xi_1, \ldots, \xi_k], \tag{2.15}$$

then we deduce from (2.13), (2.14) by induction that

$$m^{(i)}[\xi_1, \ldots, \xi_i] = M\{m^{(i+1)}[\xi_1, \ldots, \xi_{i+1}]|m^{(i)}[\xi_1, \ldots, \xi_i]\},$$
$$i = k, \quad k-1, \ldots, 1, \tag{2.16}$$

$$m_0 = M\{m^{(1)}[\xi_1]\}, \tag{2.17}$$

$$\varphi_i(m^{(i-1)}[\xi_1, \ldots, \xi_{i-1}]) = \int_{|r^{(i)}|^* \leq 1} \psi_i(r^{(i)})\mu(\mathrm{d}r^{(i)}|m^{(i-1)}) +$$

$$+ \int_{|r^{(i+1)}|^* \leq 1} \int_{|r^{(i)}|^* \leq 1} \psi_{i+1}(r^{(i+1)})\mu^{(i+1)}(\mathrm{d}r^{(i+1)}|m^{(i)}\mu^{(i)}(\mathrm{d}r^{(i)}|m^{(i-1)}) +$$

$$+ \ldots + \int_{|r^{(k)}|^* \leq 1} \int_{|r^{(i-1)}|^* \leq 1} \psi_k(r^{(k)})\mu(\mathrm{d}r^{(k)}|m^{(k-1)}) \ldots \mu((\mathrm{d}r^{(i)}|m^{(i-1)}) =$$

$$= M\left\{\sum_{j=i}^{k} \psi_j[\xi_1, \ldots, \xi_j]|m^{(i-1)}[\xi_1, \ldots, \xi_{i-1}]\right\}, \qquad i=1, \ldots, k. \tag{2.18}$$

Using notation from Section 1 we rewrite relations (2.16)–(2.18) in the form

$$m(\tau_i, \omega) = M\{m(\tau_{i+1}, \cdot)|m(\tau_i, \omega)\}, \qquad i=k, k-1, \ldots, 1, \tag{2.19}$$

$$m_0 = M\{m(\tau_1, \cdot)\} = M\{l(\cdot)\}, \tag{2.20}$$

$$\varphi_i(m(\tau_{i-1}, \omega)) = M\left\{\sum_{j=i}^{k} \psi_j(\tau_j, \cdot)|m(\tau_{i-1}, \omega)\right\}, \qquad i=k,\ k-1, \ldots, 1. \tag{2.21}$$

For $i=1$ we have the equality

$$\varphi_1(m_0) = M\left\{\sum_{j=1}^{k} \psi_j(m(\tau_j, \cdot))\right\}. \tag{2.22}$$

Thus, we have constructed a random variable $l(\omega)$, $\|l(\cdot)\| \leq 1$ satisfying equality (2.22) where $m_0 = M\{l(\cdot)\}$.

Prove now that for any random variable $l(\omega)$, $\|l(\cdot)\| \leq 1$, satisfying condition (2.20) the inequality

$$M\left\{\sum_{i=1}^{k} \psi_j(m(\tau_j, \cdot))\right\} \leq \varphi_1(m_0) \tag{2.23}$$

is true. Then equality (2.9) will be proved.

Fix a random variable $l(\omega)$ and an integer i, $1 \leq i \leq k$. As follows from the definition of the function $\varphi_i(m)$ and from condition (2.4) (where $\xi(m) = \psi_i(m) + \varphi_{i+1}(m)$, $\varphi(m) = \varphi_i(m)$), we have

$$M\{[\psi_i(m[\tau_i, \xi_1, \ldots, \xi_i]) + \varphi_{i+1}(m[\tau_i, \xi_1, \ldots, \xi_i])]|m[\tau_{i-1}, \xi_1, \ldots, \xi_{i-1}]\} \leq$$

$$\leq M\{\varphi_i(m[\tau_i, \xi_1, \ldots, \xi_i])|m[\tau_{i-1}, \xi_1, \ldots, \xi_{i-1}]\}. \tag{2.24}$$

On the other hand, the concavity of the function $\varphi_i(m)$, implies the inequality

$$M\{\varphi_i(m[\tau_i, \xi_1, \ldots, \xi_i]\}|m[\tau_{i-1}, \xi_1, \ldots, \xi_{i-1}]\} \leq$$

$$\leq \varphi_i(m[\tau_{i-1}, \xi_1, \ldots, \xi_{i-1}]). \tag{2.25}$$

(2.24) and (2.25) yield the inequality

$$M\{[\psi_i(m[\tau_i, \xi_1, \ldots, \xi_i] + \varphi_{i+1}(m[\tau_i, \xi_1, \ldots, \xi_i])]| \\ |m[\tau_{i-1}, \xi_1, \ldots, \xi_{i-1}]\} \leqq \varphi_i(m[\tau_{i-1}, \xi_1, \ldots, \xi_{i-1}]). \tag{2.26}$$

Such inequalities are true for $u=1, \ldots, k$. Using induction in i from $i=k$ to $i=1$ and applying as above (see (2.18)) the formula of iterated mathematical expectations, we get the desired inequality (2.23).

Thus, equality (2.9) is proved. Consequently, equality (2.10) is proved too.

3. Construction of the control

According to the procedure indicated in [1], the desired control forces $u[t_i]=u^0(t_i, x[t_i], \varepsilon, \Delta_\delta)$ can be formed as extremal ones given by the condition

$$\langle m^0[t_i] \cdot B(t_i)u[t_i]\rangle = \min_{u \in P} \langle m^0[t_i] \cdot B(t_i)u\rangle. \tag{3.1}$$

Here

$$m^0[t_i] = x[t_i] - w[t_i], \tag{3.2}$$

where $w[t_i]$ is the n-dimensional component of the satellite point $z[t_i]$. This point is determined by condition [1] (pp. 207–210). If we use equalities (1.5) and (2.10), and approximate functions $\rho^0(t, x)$ and $\rho^0(t, x)+z_{n+1}$ by functions $e(t, \{x, 0, \}, \Delta_\delta)$ and $e(t, z, \Delta_\delta)$, respectively, we transform the conditions determining the point $z[t_i]$ into the following form

$$\min_z \max_{|m|^* \leqq 1} (z_{n+1} + \langle m \cdot X(\vartheta, t_i)w\rangle + \varphi_1(m)) = \\ = \max_{|m|^* \leqq 1} \min_z (z_{n+1} + \langle m \cdot X(\vartheta, t_i)w\rangle + \varphi_1(m)) \tag{3.3}$$

with

$$|z - \{x, 0\}|_e \leqq [\varepsilon + \varepsilon(t_i - t_0)] \exp(\mathscr{L}(t_i - t_0)), \tag{3.4}$$

where $\mathscr{L} = \max_{t_0 \leqq t \leqq \vartheta} (|A(t)|+1)$, and $|c|_e$ denotes the Euclidean norm of vector c, and $|A(t)| = \max_{|x| \leqq 1} |A(t)x|_e$. The transposition of min and max operations in (3.3) is ensured by the concavity of the function $\varphi_1(m)$.

It follows from (3.2) and (3.3) that the vector $m^0[t_i]$ is given by the equality

$$m^0[t_i] = [\varepsilon + \varepsilon(t_i - t_0)] \exp(\mathscr{L}(t_i - t_0))X'(\vartheta, t_i]m^0/(|X'(\vartheta, t_i)m^0|_e^2 + 1)^{1/2}. \tag{3.5}$$

Here the prime denotes transposition. Vector m^0 is a solution of the following problem:

$$\max_{|m|^* \leq 1} [-[\varepsilon + \varepsilon(t_i - t_0)] \exp(\mathscr{L}(t_i - t_0)) (|X'(\vartheta, t_i)m|_e^2 + 1)^{1/2} +$$

$$+ \langle m \cdot X(\vartheta, t_i)x[t_i] + \varphi_1(m)] = [-[\varepsilon + \varepsilon(t_i - t_0)] \exp(\mathscr{L}(t_i - t_0) *$$

$$* (|X'(\vartheta, t_i)m^0|_e^2 + 1)^{1/2} + \langle m^0 \cdot X(\vartheta, t_i)x[t_i]\rangle + \varphi_1(m^0)]. \tag{3.6}$$

Besides, m^0 is the gradient of the function $\rho^0(t, x, \varepsilon, \Delta_\delta)$ which approximates the function $\rho^0(t, x)$. The solution of problem (3.6) is unique since the maximized function is concave in m.

Thus, if we assume that the control chain contains a high-speed computer, we may organize the following control process. During a very short time interval $t_i \leqq t < t_i + \alpha$ following a time instant t_i problems (3.5) and (3.6) are solved for a known value $x[t_i]$, and the control force $u[t_i]$ is calculated from condition (3.1). This control force is put onto the object during the time interval $t_i + \alpha \leqq t < t_{i+1} + \alpha$. In some cases it is indeed possible to carry out such a control process using available computers.

4. A model problem

Let the controlled object be a material point moving in the three-dimensional space under the action of the force attracting the point to a center, the friction force, the reactive control force, and an irregular disturbance force. The control process goes within a given time interval $t_0 \leqq t \leqq \vartheta$. The control quality index is the distance between the controlled object at time ϑ and the attraction center which we let to be the origin. Let $r[t] = (r_1[t], r_2[t], r_3\{t\})$ be the radius-vector of the point; $a[t]$ be the mass of the object varying according to the chosen law (continuous function); $f_1 = \beta(t)r$ be the force attracting the point to the origin; $f_2 = \alpha(t)\dot{r}$ be the friction force (the point over the letter denotes differentiation in time); $\alpha(t)$, $\beta(t)$ be given functions of time; $f_3 = \dot{a}u$ be the reactive force (u be the vector of the relative speed of the reactive mass); $f_4 = v$ be the irregular disturbance force.

The motion of the object described by the Mestcherckii differential equation

$$\ddot{r} = (\alpha(t)/a)\dot{r} + (\beta(t)/a)r + (\dot{a}/a)u + v/a, \qquad t_0 \leqq t \leqq \vartheta. \tag{4.1}$$

Let the control force $u = \{u_1, u_2, u_3\}$ and the disturbance $v = \{v_1, v_2, v_3\}$ be constrained by the conditions

$$u_1^2 + v_1^2 u_2^2 + v_2^2 u_3^2 \leqq 1 \tag{4.2}$$

$$v_1^2 + v_1^2 v_2^2 + v_2^2 v_3^2 \leqq 1. \tag{4.3}$$

The quality index is

$$\gamma=(r_1^2[\vartheta]+r_2^2[\vartheta]+r_3^2[\vartheta])^{1/2}. \tag{4.4}$$

Introduce the phase vector

$$y=\{y_1,\ldots,y_6\}=\{r_1,\dot{r}_1,r_2,\dot{r}_2,r_3,\dot{r}_3\}. \tag{4.5}$$

Vector y satisfies the differential equation

$$\dot{y}=F(t)y+G(t)u+H(t)v, \tag{4.6}$$

where matrices $F(t)$, $G(t)$, $H(t)$ are expressed through the variables indicated above. Assume that the mass a of the object varies according to the law

$$a(t)=a_0\exp(-\lambda[t-t_0]). \tag{4.7}$$

Using the transformation of the variable y into variable x, similar to that given in [1] (p. 52), pass to the differential equation for the 3-dimensional vector x:

$$\dot{x}=-\lambda y_{12}(\vartheta,t)u+\exp(\lambda[t-t_0])y_{12}(\vartheta,t)v/a_0, \tag{4.8}$$

where $y_{12}(\vartheta,t)$ is the element of the fundamental matrix $Y(\tau,t)$ for the equation $dy/d\tau=F(\tau)y$. Vectors u and v are constrained by conditions (4.2) and (4.3) as above. The index γ has the form

$$\gamma=(x_1^2[\vartheta]+x_2^2[\vartheta\}+x_3^2[\vartheta])^{1/2}. \tag{4.9}$$

According to Section 2 functions $\psi_i(m)$ have the form

$$\psi_i(m)=\int_{\tau_i}^{\tau_{i+1}}\min_{u\in P}\max_{v\in Q}\langle m\cdot(-\lambda y_{12}(\vartheta,\tau)u+$$
$$+\exp(\lambda[\tau-t_0])*y_{12}(\vartheta,\tau)v/a_0)\rangle d\tau, \tag{4.10}$$

where $m=\{m_1,m_2,m_3\}$.

To be particular, assume that $\nu_1>1$ and $\nu_2>1$. The construction of upper concave hulls $\varphi_i(m)$ depends on the sign of the expression

$$\eta(t)=\exp(t-t_0))/(\lambda a_0)-1. \tag{4.11}$$

Let us include the root τ_* of the equation $\eta(t)=0$ into the collection of points τ_i of the partition Δ_δ. Then two cases are possible.

1. If $\eta(\tau)<0$ for $\tau_i<\tau<\tau_{i+1}$, then function $\psi_i(m)$ is concave in m. Consequently,

$$\varphi_i(m)=\psi_i(m)+\varphi_{i+1}(m). \tag{4.12}$$

2. If $\eta(\tau)>0$ for $\tau_i<\tau<\tau_{i+1}$, then function $\psi_i(m)$ is not concave in m. Consequently,

$$\varphi_i(m)=\{\psi_i(m)\}_*+\varphi_{i+1}(m). \tag{4.13}$$

Taking into account that the function $\eta(t)$ is monotonous, we pass to the following result:

$$\varphi(m)=\lambda\int_{\tau_1}^{\tau_*}|y_{12}(\vartheta,\tau)|\eta(\tau)\mathrm{d}\tau\cdot(m_1^2+m_2^2/v_1^2+m_3^2/v_2^2)^{1/2}+$$

$$+\lambda\int_{\tau_*}^{\tau_{k+1}}|y_{12}(\vartheta,\tau)|\eta(\tau)\mathrm{d}\tau\cdot(1+(1/v_1^2-1)m_2^2+(1/v_2^2-1)m_3^2)^{1/2}. \tag{4.14}$$

Thus, vector $m^0=\{m_1^0, m_2^0, m_3^0\}$ which determines vector $m^0[t_i]$ determining the optimal control $u^0[t_i]$ is the solution of the following problem:

$$\max_{|m|^*\leq 1}\{-[\varepsilon+\varepsilon(t_i-t_0)](|m|^2+1)^{1/2}+\langle m\cdot x[t_i]\rangle+\varphi_1(m)\}=$$

$$=\{-[\varepsilon+\varepsilon(t_i-t_0)](|m^0|^2+1)^{1/2}+\langle m^0\cdot x[t_i]\rangle+\varphi_1(m^0)\}. \tag{4.15}$$

Vector $m^0[t_i]$ takes the form

$$m^0[t_i]=[\varepsilon+\varepsilon(t_i-t_0)]m^0/(|m^0|^2+1)^{1/2}. \tag{4.16}$$

The control force $u^0[t_i]$ is determined now by the formulae

$$u_1^0[t_i]=\operatorname{sign}(y_{12}(\vartheta,t_i))m_1^0[t_i]/((m_1^0[t_i]))^2+(m_2^0[t_i])^2/v_1^2+$$
$$+(m_3^0[t_i])^2v_2^2)^{1/2},$$
$$u_2^0[t_i]=\operatorname{sign}(y_{12}(\vartheta,t_i))m_2^0[t_i]/((m_1^0[t_i]))^2+(m_2^0[t_i])^2/v_1^2+$$
$$+(m_3^0[t_i])^2v_2^2)^{1/2})/v_1^2,$$
$$u_3^0[t_i]=\operatorname{sign}(y_{12}(\vartheta,t_i))m_3^0[t_i]/((m_1^0[t_i]))^2+(m_2^0[t_i])^2/v_1^2+$$
$$+(m_3^0[t_i])^2v_2^2)^{1/2}/v_2^2. \tag{4.17}$$

The control process for the considered object was simulated on a computer for the following values of the parameters

$$t_0=0,\quad \vartheta=4,\quad y(t_0)=\{0,1,1,0,1,0\},$$
$$a_0=10,\quad \lambda=0.6,$$
$$\alpha(t)=-4\exp(-0.6t),\quad \beta(t)=0.5\exp(-0.6t)$$
$$v_1=1.225,\quad v_2=1.265. \tag{4.18}$$

Consider three situations.

For the first situation, the optimal control acts together with the most unfavourable disturbance v which was constructed in accordance with the procedures described in [1]. The motion for this case is shown at Fig. 1 with the continuous curve. For the second case, the optimal control strategy acts together with the disturbance of the form

$$v_1[t] = \sin(10t)\cos(10t), \qquad v_2[t] = (\sin^2(10t))/v_1,$$
$$v_3[t] = (\cos(10t))/v_2, \tag{4.19}$$

which is not counter-optimal. The corresponding motion is shown at Fig. 1 with dotted line. For the third case a non-optimal strategy $u(t, r)$ was chosen. For each time instant t it fixes vector $u[t]$:

$$u_1[t] = -r_1[t]/(r_1^2[t] + v_1^2 r_2^2[t] + v_2^2 r_3^2[t]),$$
$$u_2[t] = -r_2[t]/(r_1^2[t] + v_1^2 r_2^2[t] + v_2^2 r_3^2[t]),$$
$$u_3[t] = -r_3[t]/(r_1^2[t] + v_1^2 r_2^2[t] + v_2^2 r_3^2[t]) \tag{4.20}$$

which is collinear to the current radius-vector $r[t]$. The disturbance was formed here in a way most unfavourable for us. The touched-and-dotted line at Fig. 1 corresponds to this case. The experiment gave three values of results: $\gamma_1 = 1.3593$, $\gamma_2 = 0.5199$, $\gamma_3 = 1.9864$. As it could be expected, these values satisfy the inequalities

$$\gamma_2 < \gamma_1 < \gamma_3. \tag{4.21}$$

For the taken initial position $\{t_0, x_0\}$ the optimal ensured result is $\rho^0(t_0, x_0) = 1.3587$. As it could be expected, the experiment gave the values γ_1 close to $\rho^0(t_0, x_0)$:

$$\gamma_1 \approx \rho^0(t_0, x_0). \tag{4.22}$$

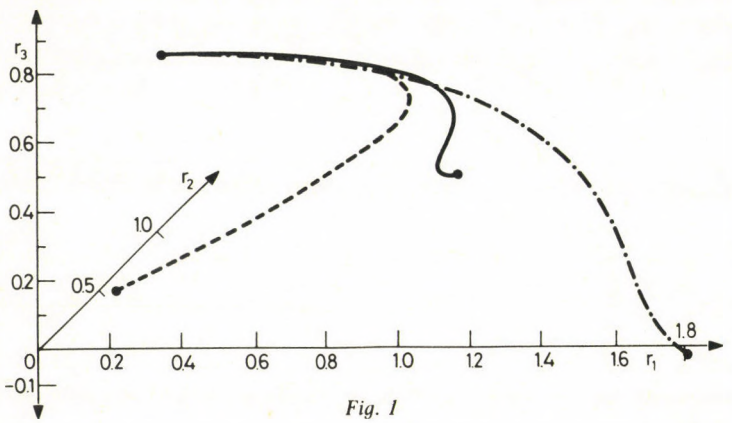


Fig. 1

References

1. *Krasovskii, N. N.*, Controlling of a dynamical system. Moscow, Nauka, 1985, 520 pp.
2. *Krasovskii, N. N.*, Tret'jakov, V. E., Stochastic program synthesis of an ensuring control. Probl. Cont. Inf. Theory, 1983, **12***(2)*, pp. 70–95.
3. *Krasovskii, A. N.*, Construction of mixed strategies on the basis of stochastic programs. Priklad. mat. i meh., **51,** *2* (1987), pp. 186–192.
4. *Tret'jakov, V. E.*, Synthesis of optimal guaranteeing control (to appear).
5. *Arkin, V. I., Levin, V. L.*, Convexity of the values of vector integrals, theorems of measurable selection and variational problems. Uspekhi matem. nauk, 1972, **12,** *3*.

О программном синтезе гарантирующего управления

Н. Н. КРАСОВСКИЙ, Т. Н. РЕШЕТОВА

(Свердловск)

Исследуется задача об оптимальной гарантирующей стратегии управления динамической системой, эволюция которой описывается обыкновенными дифференциальными уравнениями. Система подвержена воздействию неизвестной, но ограниченной помехи. Управление формируется по принципу обратной связи. Оптимальная стратегия должна минимизировать верхнюю оценку значений функционала качества, который есть сумма расстояния от начала координат фазовой точки в момент окончания управления и интегральных затрат реализовавшихся управления и помехи.

Известно, что решение этой задачи можно определить в форме стратегии, экстремальной к функции цены $\rho^0(t, x)$ соответствующей дифференциальной игры. Разъясняется рекуррентная процедура построения вогнутых оболочек функций, возникающих в рамках вспомогательной стохастической программной конструкции. Такая процедура вытекает из строения стохастического программного максимина, включающего пошаговое усреднение названных функций. Это позволяет аппроксимировать величину $\rho^0(t, x)$ для позиций, реализующихся вдоль движений системы. Тем самым открывается возможность эффективного вычисления управляющих воздействий по ходу реализации процесса в истинном масштабе времени.

Предложенная процедура проиллюстрирована примером. Приведены симуляции модельной задачи на ЭВМ.

Н. Н. Красовский
Институт математики и механики Уральского отделения АН СССР
СССР 620219 Свердловск ГСП-384,
ул. С. Ковалесвской, 16

Problems of Control and Information Theory, Vol. 17 (6), pp. 345–356 (1988)

BOUNDS ON FREE EUCLIDEAN DISTANCE FOR VARIOUS SYSTEMS OF MODULATION AND CODING

V. V. ZYABLOV, S. A. SHAVGULIDZE, M. G. SHANIDZE

(*Moscow, Tbilisi*)

(Received October 15, 1987)

Systems of modulation and coding on the basis of signals of phase-shift keying modulation and various trellis codes with unit memory are constructed and investigated. The asymptotic bounds on free Euclidean distance are given. The comparison of error-correcting capabilities are carried out for some interesting practical constructions for different rates of transmission.

1. Introduction

To quantify the error-correcting capability of a code some scientists often use its minimum Hamming distance. However, this criterion does not always reflect the real effectiveness of systems of modulation and coding, for example, for the multipositional signals of phase-shift keying modulation or amplitude-phase-shift keying modulation. The error-control capability of these or other codes is defined to a greater extent by their minimum Euclidean distance in such systems. Codes which are optimal within Hamming metric are not always optimal within Euclidean metric and vice versa.

Today there exist several efficient methods of synthesis of codes with good Hamming distance and multipositional signals of phase-shift keying or amplitude-phase-shift keying modulation. The coordination of coding and modulation with the help of Gray codes as a mapping function and by decomposition of the original system of signals onto embedding subsystems are mostly widespread [1]. The combination of these two methods which ensures good system characteristics both at finite code length and at the asymptotic case is also very perspective [2, 3].

In this paper another approach, first proposed in [4], is used to construct the systems of modulation and coding. Namely, the error-correcting codes that are optimal in the sense of minimum Euclidean distance are directly constructed. q-positional phase-shift keying (q-PSK) is used as the system of signals. This system may be considered as the set $D(q)$ of q different signals that may be represented by equally

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

spaced points on the unit circle. We use two classes of q-ary error-control codes: nonlinear trellis codes with unit memory (TUM-codes) and nonlinear trellis concatenated codes of order m with unit memory (TCUM(m)-codes) as the coding system. The asymptotic lower bounds on the squared free Euclidean distance are obtained for these systems.

The content of the paper is as follows. In Section 2 we introduce some notation and definitions. In Section 3 the asymptotic lower bounds on the squared free Euclidean distance are obtained for TUM-codes and TCUM(m)-codes. In Section 4 some results of computation and main conclusions are given.

2. Main notations and definitions

The plane in which the signals are represented is considered as the complex plane C. Any signal point on the unit circle is associated with the complex number $\exp(i\varphi)$ for φ in $[0, 2\pi]$ and $i^2 = -1$. Let us define $\varepsilon = \exp(2\pi i/q)$ — a primitive q-th root of one, where $q \geqq 2$ is an integer.

The alphabet $D(q)$ is defined to be

$$D(q) = \{\varepsilon^{\eta+\mu} : \eta = 0, 1, \ldots, q-1\} \tag{1}$$

with μ a real number in $\left[-\frac{1}{2}, \frac{1}{2}\right]$. For example, Fig. 1 represents $D(8)$ for $\mu = \frac{1}{2}$.

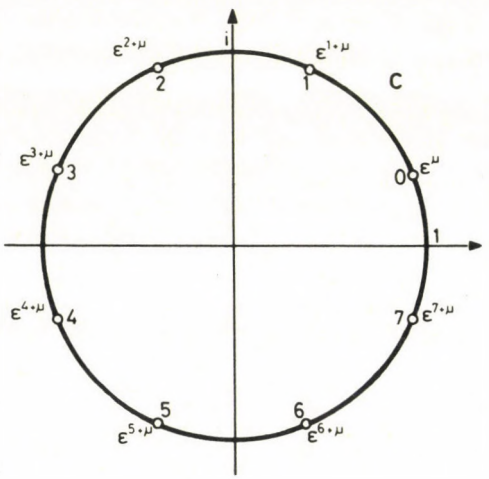


Fig. 1

The choice of μ has no influence on the distance properties of the code and in the sequel we assume $\mu=0$. Below we shall substitute $D(q)$ simply by the D. The Euclidean distance $d_E(\varepsilon^\eta, \varepsilon^\xi)$ between the ε^η and ε^ξ is defined to be

$$d_E(\varepsilon^\eta, \varepsilon^\xi)=|\varepsilon^\eta-\varepsilon^\xi| \tag{2}$$

where $|l|$ denotes the norm of the complex number l. This distance satisfies

$$d_E^2(\varepsilon^\eta, \varepsilon^\xi)=4\sin^2[(\eta-\xi)\pi/q]. \tag{3}$$

We extend the definition of distance to include distances between nT-tuples over D. Given two elements of D^{nT}

$$y=(y_1, \ldots, y_n, y_{n+1}, \ldots, y_{2n}, \ldots, y_{(T-1)n+1}, \ldots, y_{Tn})$$

and

$$y'=(y'_1, \ldots, y'_n, y'_{n+1}, \ldots, y'_{2n}, \ldots, y'_{(T-1)n+1}, \ldots, y'_{Tn})$$

we define

$$d(y, y')=\left[\sum_{j=1}^{nT} d_E^2(y_j, y'_j)\right]^{1/2}. \tag{4}$$

We consider the case when $T\to\infty$. It is easy to check that this definition of distance satisfies all the axioms of a distance over D^{nT}.

We denote TUM-codes as $A[n_a, R_a, d_a^2, q]$, where n_a is the code constraint length (CCL), $R_a=\frac{k_a}{n_a}\ln q$ is the transmission rate (k_a is the number of information symbols which falls on a CCL), d_a^2 is the squared free Euclidean distance, and q is the base of a code.

Let us represent these codes with the help of trellis diagram [5]. The branching points at the trellis are called nodes and the segments which connect the nodes are called branches. The trellis contains q^{k_a} nodes and q^{k_a} branches enter and come out in each node on the tier $t\geqq 2$. n_a q-ary symbols are assigned to each branch. Suppose that every code symbol on every branch is chosen independently at random according to some probability distribution $P=\{p_l, 1\leqq l\leqq q\}$. We limit ourselves by the case when $p_l=\frac{1}{q}$, $l=\overline{1, q}$. Below we shall consider the ensemble (n_a, R_a, d_a^2, q) of the codes A, where each code is chosen in accordance with the distribution $\{p_l\}$ of the individual code symbols. We call this ensemble as the ensemble of random TUM-codes.

Let y and y' be two arbitrary code sequences belonging to A. Then [5]

$$d_a^2=\min_{y, y'\in A} d^2(y, y'). \tag{5}$$

We define the normalized squared free Euclidean distance of a code A as follows:

$$\delta_a = \frac{d_a^2}{n_a}. \tag{6}$$

Let us consider TCUM(m)-codes. They are constructed on the base of outer Reed–Solomon (RS) codes $B_i[n_b, r_{b,i}, d_{b,i}, q^{a_i}]$, $i=\overline{1,m}$, over the field $GF(q^{a_i})$, with length — n_b, transmission rate — $r_{b,i}$, and Hamming distance — $d_{b,i}$ and inner TUM-codes $A_{1,j}[n_a, R_{a,1}, d_{a,1}^2, q]$, $j=\overline{1,n_b}$, over the field $GF(q)$, with CCL — n_a, transmission rate in nats — $R_{a,1}=\dfrac{\sum_{i=1}^{m} a_i}{n_a}\ln q$ ($\sum_{i=1}^{m} a_i$ is the number of information symbols that falls on the CCL), and squared free Euclidean distance — $d_{a,1}^2$. We consider the case when all codes $A_{1,j}$, $j=\overline{1,n_b}$, are the same and later index j will be omitted. Code A_1 may be decomposed into the embedding system of the inner codes

$$A_1[n_a, R_{a,1}, d_{a,1}^2, q] \supset A_2[n_a, R_{a,2}, d_{a,2}^2, q] \supset \ldots \supset A_m[n_a, R_{a,m}, d_{a,m}^2, q]$$

where $R_{a,i_0}=\sum_{i=i_0}^{m} a_i/n_a$,

$$R_{a,1} > R_{a,2} > \ldots > R_{a,m},$$
$$d_{a,1}^2 \leqq d_{a,2}^2 \leqq \ldots \leqq d_{a,m}^2. \tag{7}$$

The q-ary TCUM(m)-code constructed on the base of given inner and outer codes, have the CCL

$$n_{ab} = n_a n_b, \tag{8}$$

the transmission rate in nats

$$R_{ab} = \sum_{i=1}^{m} (R_{a,i} - R_{a,i+1}) r_{b,i} \qquad \text{where} \quad R_{a,m+1}=0 \tag{9}$$

and the squared free Euclidean distance

$$d_{ab} \geqq \min_i \{d_{a,i}^2 \cdot d_{b,i}\}. \tag{10}$$

Analogically to TUM-codes, we define the normalized squared free Euclidean distance by the following expression

$$\delta_{ab} = \frac{d_{ab}^2}{n_{ab}}. \tag{11}$$

We shall consider the asymptotic case, that is $n_a \to \infty$, $n_b \to \infty$ and, therefore, $n_{ab} \to \infty$. The concatenated codes of infinite order ($m \to \infty$), which we call the TGUM(∞)-codes of infinite order are of great interest. We limit ourselves to the case when $\lim\limits_{m\to\infty} a_i = \infty$ and $\lim\limits_{m\to\infty} \left(\frac{a_i}{n_a}\right) = 0$, $i = \overline{1, m}$. We suppose that all quantities a_i, $i = \overline{1, m}$, are equal.

The aim of our investigation is to obtain lower bounds (bounds of existence) for quantities $\delta_a(R_a)$ and $\delta_{ab}(R_{ab}, m)$, accordingly, when $n_a \to \infty$ and $n_{ab} \to \infty$.

3. The lower bounds on the normalized squared free Euclidean distance

We shall consider TUM-codes at first. Let us choose the arbitrary code sequence (the path) on the trellis of these codes, which we call the correct path. All other paths form the set of incorrect paths. The subset of incorrect paths y' are denoted by S_t. It consists of all incorrect paths which diverge with the correct path at the tier t. The configuration of any possible incorrect path and the correct path is completely described by the two times t and $t+\tau$ in which they diverge and remerge in the trellis. We let $S_{t,\tau}$ stand for the set of all incorrect paths, which diverge with the correct path at time t and remerge at time $t+\tau$ first. Analogically to [5], it is possible to show that the cardinality of this subset is bounded above by the expression:

$$|S_{t,\tau}| \leqq \exp_q \{k_a(\tau - 1)\} = \exp \{n_a R_a(\tau - 1)\}. \tag{12}$$

Let us denote $d^2_{a,t}$ the squared free Euclidean distance between the correct path and the nearest path (in the sense of Euclidean distance) from the subset S_t.

Assertion 1. There exist TUM-codes $A[n_a, R_a, d^2_{a,t}, q]$ with parameters

$$R_a = \frac{k_a}{n_a} \ln q, \quad d^2_{a,t} = \delta_a n_a$$

such that when they are used in q-PSK systems then the following equations

$$R_a < \ln q - \ln \sum_{j=1}^{q} \exp \{-4\alpha \sin^2(j\pi/q)\};$$

$$\delta_a < \frac{2\left(\ln q - \ln \sum\limits_{j=1}^{q} \exp \{-4\alpha \sin^2 (j\pi/q)\}\right) - R_a}{\alpha} - \frac{1}{\alpha n_a} \ln \beta n_a \tag{13}$$

are satisfied where $\alpha \geqq 0$ and $\beta \to 1$ for $n_a \to \infty$.

Proof. Clearly, for any path $y' \in S_t$ the squared Euclidean distance — d^2 between y' and the correct path y is the sum of independent increments:

$$d^2(y, y') = \sum_j d_E^2(y_j, y'_j)$$

where

$$d_E^2(y_j, y'_j) = \begin{cases} 0, & \text{if} \quad y_j = y'_j; \\ 4\sin^2 [\pi(\eta - \xi)/q], & \text{if} \quad \varepsilon^\eta = y_j \neq y'_j = \varepsilon^\xi. \end{cases} \tag{14}$$

We now seek to bound the probability that for a given random TUM-code $A[n_a, R_a, d^2_{a,t,q}]$ $d^2_{a,t}$ will be less than a parameter d_0^2.

Consider the quantity

$$Z(\alpha) = \sum_{y' \in S_t} \exp\{\alpha d^2(y, y')\}, \quad \alpha \geqq 0. \tag{15}$$

If any $y' \in S_t$ has $d^2(y, y') \leqq d_0^2$ then $Z(\alpha) \geqq \exp\{-\alpha d_0^2\}$, or $\exp\{\alpha d_0^2\} \cdot Z(\alpha) \geqq 1$. Hence,

$$\overline{\Pr(d^2_{a,t} \leqq d_0^2)} \leqq \exp\{-\alpha d_0^2\} \overline{Z(\alpha)} \tag{16}$$

where the overbar is an average over all codes in the ensemble $(n_a, R_a, d^2_{a,t}, q)$.

We divide the S_t into configurations $S_{t,\tau}$. It follows from (15)

$$\overline{Z(\alpha)} \leqq \sum_{\tau=2}^{\infty} \sum_{y' \in S_{t,\tau}} \overline{\exp\{-\alpha d^2(y, y')\}}.$$

Then we obtain

$$\overline{Z(\alpha)} \leqq \sum_{\tau=2}^{\infty} \exp_q\{k_a(\tau - 1)\} \exp\{-n_a \tau E(\alpha)\} \tag{18}$$

where

$$E(\alpha) \triangleq -\ln\left[\sum_{j=1}^{q} p_j \sum_{s=1}^{q} p_s \exp\{-\alpha d^2(y_j, y'_s)\}\right] =$$

$$= \ln q - \ln \sum_{j=1}^{q} \exp\{-4\alpha \sin^2(j\pi/q)\} \tag{19}$$

from (12), (14) and Lemma A3 [5].

Consequently, if $R_a < E(\alpha)$, we get

$$\overline{\Pr(d^2_{a,t} \leqq d_0^2)} \leqq \exp\{\alpha d_0^2\} \exp\{-n_a R_a\} \sum_{\tau=2}^{\infty} \exp\{-n_a \tau [E(\alpha) - R_a]\} =$$

$$= \beta \exp\{-n_a [2E(\alpha) - \alpha\delta_a - R_a]\} \tag{20}$$

where $\delta_a = \dfrac{d_0^2}{n_a}$ and $\beta \triangleq \dfrac{1}{1 - \exp\{-n_a [E(\alpha) - R_a]\}} \to 1$ for $n_a \to \infty$.

Thus, if we have simultaneously

$$R_a < E(\alpha); \qquad \delta_a < \frac{2E(\alpha) - R_a}{\alpha} - \frac{1}{\alpha n_a} \ln \beta n_a$$

and $n_a \to \infty$, then $\Pr(d_{a,t}^2 \leqq d_0^2) = 0$. Hence, there exist TUM-codes which have $d_{a,t}^2 > d_0^2$ and we have proved Assertion 1.

Before starting the investigation of the asymptotic characteristics of TCUM(m)-codes, it is necessary to show that there exist the embedding systems of TUM-codes in which the main code and all their subcodes have the error-correcting capabilities given by Assertion 1.

Assertion 2. Let A_i ($i = 2, 3, \ldots, m$) be $m-1$ subcodes of a code A_1 where $A_1 \supset A_2 \supset \ldots \supset A_m$. Then there exists a code $A_1[n_a, R_{a,1}, d_{a,t,1}^2, q]$ in the ensemble $(n_a, R_{a,1}, d_{a,t,1}^2, q)$ such that the following equations

$$R_{a,i} < \ln q - \ln \sum_{j=1}^{q} \exp\{-4\alpha_i \sin^2(j\pi/q)\};$$

$$\delta_{a,i} = \frac{d_{a,t,i}^2}{n_a} < \frac{2\left(\ln q - \ln \sum_{j=1}^{q} \{-4\alpha_i \sin^2(j\pi/q)\}\right) - R_{a,i}}{\alpha_i} - \frac{1}{\alpha_i n_a} \ln \beta_i n_a \qquad (21)$$

are simultaneously satisfied for this code and all their subcodes $A_i[n_a, R_{a,i}, d_{a,t,i}^2, q]$ in q-PSK systems, where $i = \overline{1, m}$, $\alpha_i \geqq 0$ and $\beta_i \to 1$ for $n_a \to \infty$.

Proof. For each code from the ensemble $(n_a, R_{a,1}, d_{a,t,1}^2, q)$ and real numbers $d_{0,i}^2$, $i = \overline{1, m}$, we introduce the indicator functions

$$P_{\text{ind}}^{(i)} = \begin{cases} 0, & \text{if } d_{a,t,i}^2 > d_{0,i}^2; \\ 1, & \text{if } d_{a,t,i}^2 \leqq d_{0,i}^2 \end{cases}.$$

Then, if $R_{a,1} < E(\alpha_1)$ we get

$$P_{\text{ind}}^{(1)} \leqq \beta_1 \exp\{-n_a[2E(\alpha_1) - \alpha_1 \delta_{a,1} - R_{a,1}]\} \qquad (22)$$

from (20), where

$$\delta_{a,1} \triangleq \frac{d_{0,1}^2}{n_a}, \quad E(\alpha_1) \triangleq \ln q - \ln \sum_{j=1}^{q} \exp\left\{-4\alpha_1 \sin^2\left(\frac{j\pi}{q}\right)\right\}$$

and $\beta_1 \triangleq \dfrac{1}{1 - \exp\{-n_a[E(\alpha_1) - R_{a,1}]\}} \to 1$ for $n_a \to \infty$.

Now we consider the random ensemble of TUM-codes $(n_a, R_{a,i}, d_{a,t,i}^2, q)$ over the field $GF(q)$ with CCL — n_a, transmission rate — $R_{a,i}$ and squared free Euclidean distance — $d_{a,t,i}^2$. The suggested method of choosing of a subcode of a code A_1 ensures

that each code from the ensemble $(n_a, R_{a,i}, d^2_{a,t,i}, q)$ corresponds to a subcode A_i of some code A_1 and vice versa. Therefore, all estimations obtained for the ensemble of codes $(n_a, R_{a,i}, d^2_{a,t,i}, q)$ are true for the ensemble of subcodes $A_i[n_a, R_{a,i}, d^2_{a,t,i}, q]$ as they are identical. Taking into consideration that these arguments are true for any $i=\overline{2,m}$, we get for real numbers $d^2_{0,i}$ the expressions

$$\bar{P}^{(i)}_{\text{ind}} \leqq \beta_i \exp\{-n_a[2E(\alpha_i)-\alpha_i\delta_{a,i}-R_{a,i}]\}, \tag{23}$$

if $R_{a,i} < E(\alpha_i)$ where

$$\delta_{a,i} \triangleq \frac{d^2_{0,i}}{n_a}, \quad E(\alpha_i) \triangleq \ln q - \ln \sum_{j=1}^{q} \exp\left\{-4\alpha_i \sin^2\left(\frac{j\pi}{q}\right)\right\}$$

and $\beta_i \triangleq \dfrac{1}{1-\exp\{-n_a[E(\alpha_i)-R_{a,i}]\}} \to 1$ for $n_a \to \infty$.

Thus the average of function $\bar{P}^{(1)}_{\text{ind}}$ for codes A_1 through the ensemble of TUM-codes is bounded by expression (22) above, and function $\bar{P}^{(i)}_{\text{ind}}$ for codes $A_i (A_1 \supset A_i)$ — by expression (23). In order to prove that these equations are true simultaneously for some code A_1 and all its subcodes A_i, $i=\overline{2,m}$, we introduce the indicator function

$$\lambda = \begin{cases} 0, & \text{if } P^{(i)}_{\text{ind}} = 0, \quad \text{for each } i; \\ 1, & \text{if } P^{(i)}_{\text{ind}} = 1, \quad \text{at least for one } i' \end{cases}$$

where $i=\overline{1,m}$.

It is evident that

$$\lambda \leqq \sum_{i=1}^{m} P^{(i)}_{\text{ind}}.$$

Therefore

$$\bar{\lambda} \leqq \sum_{i=1}^{m} \bar{P}^{(i)}_{\text{ind}} \leqq \sum_{i=1}^{m} \beta_i \exp\{-n_a[2E(\alpha_i)-\alpha_i\delta_{a,i}-R_{a,i}]\}$$

if $R_{a,i} < E(\alpha_i)$ for each $i=\overline{1,m}$.

Hence, if the conditions

$$R_{a,i} < E(\alpha_i) \quad \text{and} \quad \delta_{a,i} < \frac{2E(\alpha_i)-R_{a,i}}{\alpha_i} - \frac{1}{\alpha_i n_a} \ln \beta_i n_a$$

are simultaneously satisfied for each $i=\overline{1,m}$, then $\bar{\lambda}=0$ for $n_a \to \infty$. Thus, there exists the embedding system of TUM-codes for which the conditions $d^2_{a,t,i} > d^2_{0,i}$ are simultaneously satisfied for each $i=\overline{1,m}$. Q.E.D.

Asymptotically ($n_a \to \infty$) we may neglect terms $\frac{1}{\alpha n_a} \ln \beta n_a$ in (13) and $\frac{1}{\alpha_i n_a} \ln \beta_i n_a$, $i = \overline{1, m}$, in (21). Therefore the functions $\delta_a(R_a)$ and $\delta_{a,i}(R_{a,i})$, $i = \overline{1, m}$, may be written parametrically as follows

$$\delta_a \triangleq \frac{d_{a,t}^2}{n_a} = \sup_{\alpha \geq 0} \frac{2\left(\ln q - \ln \sum_{j=1}^{q} \exp\{-4\alpha \sin^2(j\pi/q)\}\right) - R_a}{\alpha}, \qquad R_a < \ln q - \ln \sum_{j=1}^{q} \exp\{-4\alpha \sin^2(j\pi/q)\} \tag{24}$$

in the case of TUM-codes and

$$\delta_{a,i} \triangleq \frac{d_{a,t,i}^2}{n_a} = \sup_{\alpha_i \geq 0} \frac{2\left(\ln q - \ln \sum_{j=1}^{q} \exp\{-4\alpha_i \sin^2(j\pi/q)\}\right) - R_{a,i}}{\alpha_i}, \qquad R_{a,i} < \ln q - \ln \sum_{j=1}^{q} \exp\{-4\alpha_i \sin^2(j\pi/q)\} \tag{25}$$

in the case of embedding system of TUM-codes.

Then it is possible to show analogically [6] that the lower bound on quantity $\delta_{ab}(R_{ab}, m)$ for TCUM(m)-codes in q-PSK systems is given by expressions

$$\delta_{ab}(R_{ab}, m) = \max_{R_{a,1}} \left\{ m(R_{a,1} - R_{ab}) \Big/ \left[R_{a,1} \sum_{i=1}^{m} (\delta_{a,i}((m-i+1)R_{a,1}/m))^{-1} \right] \right\} \text{ for } m = \text{const};$$

$$\delta_{ab}(R_{ab}, \infty) = \max_{R_{a,1}} \left\{ (R_{a,1} - R_{ab}) \Big/ \int_0^{R_{a,1}} \frac{du}{\delta_a(u)} \right\} \text{ for } m \to \infty. \tag{26}$$

4. Summary

Functions $\delta_a(R_a)$ and $\delta_{ab}(R_{ab}, m)$ in q-PSK systems are represented in Figs 2–5. Curves 1^1, 1^2, 1^3, 1^4 correspond to TUM-code, curves 2^1, 2^2, 2^3, 2^4 correspond to TCUM(1)-code, curves 3^1, 3^2, 3^3, 3^4 correspond to TCUM(5)-code and curves 4^1, 4^2, 4^3, 4^4 correspond to the TCUM(∞)-code for $q = 2, 4, 8, 16$ accordingly.

The results of the computations show that when identical methods of coding are used, the systems with $q = 4, 8, 16$ have approximately the same error-correctiong capabilities at low rates of transmission. With the increasing of rates systems with $q = 16$ become the best, then systems with $q = 8$ and then systems with $q = 4$. It should

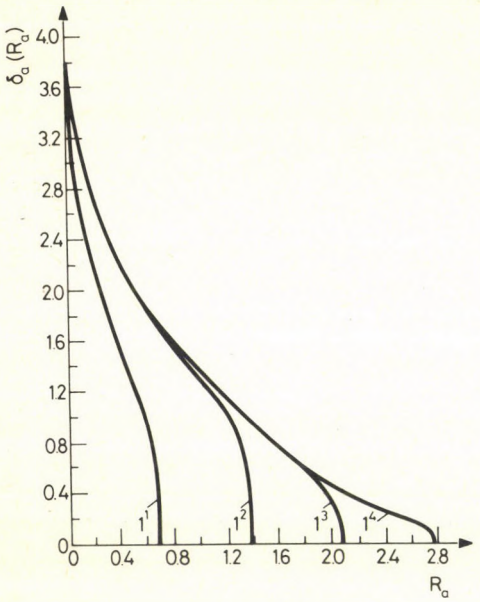


Fig. 2

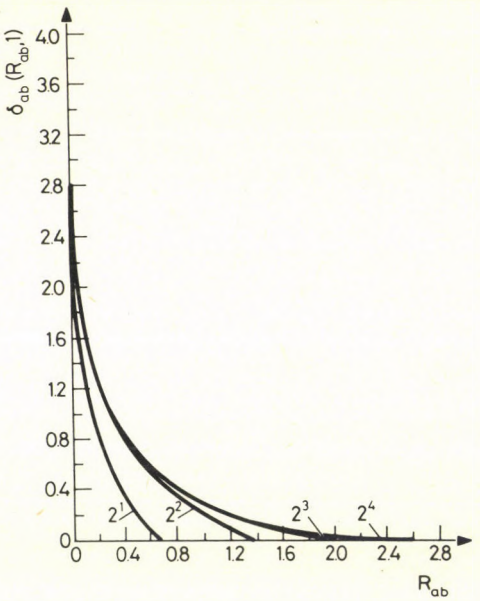


Fig. 3

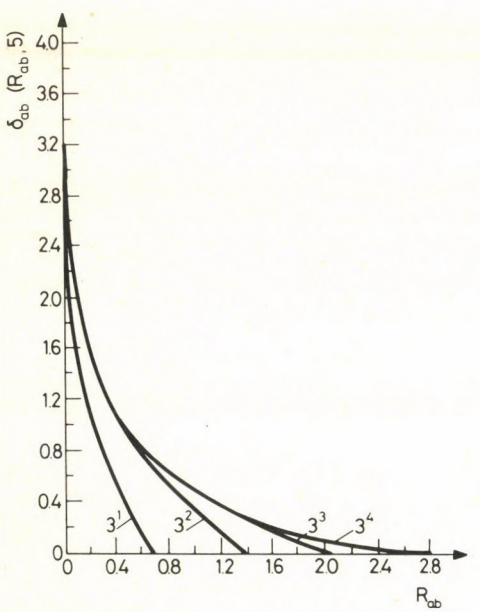


Fig. 4

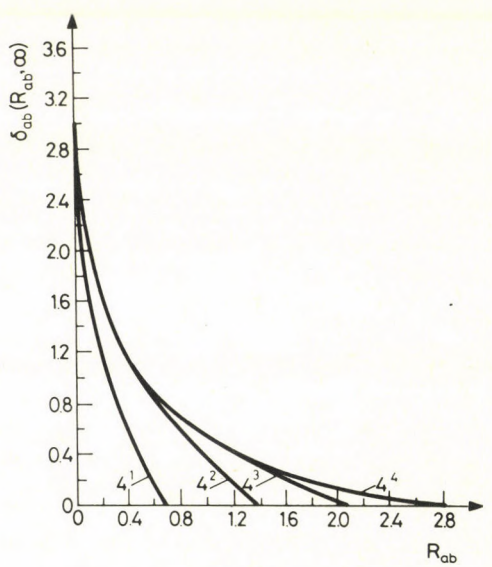


Fig. 5

be especially noted the significant loss of systems with $q=2$ in comparison with other systems at any rate of transmission. The increase of concatenated codes' order leads to amelioration of error-correcting capabilities and though in this respect TCUM(m)-codes are inferior in comparison with TUM-codes, we should take into consideration that the latter have substantially more realization complexity [6–8]. This fact makes the usage of TCUM(m)-codes more expedient in communication systems.

References

1. *Nemirovsky, E. E., Portnoy, S. L.*, Band-effective coding and modulation of Gaussian channel. Zarubezhnaya Radioelectronika, 1984, **8,** pp. 3–18 (in Russian).
2. *Portnoy, S. L.*, Characteristics of system of modulation and coding from the point of view of cascade codes. Problemy Peredachi Informatsii, 1985, **21,** *3*, pp. 14–27 (in Russian).
3. *Portnoy, S. L., Shavgulidze, S. A.*, The asymptotic behaviour of a system of modulation and coding. In book: XI Soviet Union workshop on computation networks. Part 2, Abstracts, Moscow–Riga, 1986, pp. 126–131.
4. *Piret, P.*, Bounds for codes over the unit circle. IEEE Trans. Inform. Theory, 1986, **32,** *6*, pp. 760–767.
5. *Forney, G. D.*, Convolutional codes II. Maximum-likelyhood decoding. Information and Control, 1974, **25,** *3*, pp. 222–266.
6. *Zyablov, V. V., Shavgulidze, S. A.*, The estimation of code distance of convolutional generalized concatenated codes with unit memory. Problemy Peredachi Informatsii, 1987, **23,** *2*, pp. 17–27 (in Russian).
7. *Zyablov, V. V., Shavgulidze, S. A.*, Generalized convolutional concatenated codes with unit memory. Problemy Peredachi Informatsii, 1986, **22,** *4*, pp. 9–28 (in Russian).
8. *Ginzburg, V. V.*, Multidimensional signals for continuous channel. Problemy Peredachi Informatsii, 1984, **20,** *1*, pp. 28–46 (in Russian).

Границы свободного евклидова расстояния для различных систем модуляции и кодирования

В. В. ЗЯБЛОВ, С. А. ШАВГУЛИДЗЕ, М. Г. ШАНИДЗЕ

(Москва, Тбилиси)

Для оценивания корректирующих свойств помехоустойчивых кодов часто пользуются их минимальным хемминговым расстоянием. Однако этот критерий не всегда отражает реальную эффективность систем модуляции и кодирования, например, в случае использования многопозиционных сигналов фазовой или амплитудно-фазовой модуляции. В таких системах помехоустойчивость тех или иных кодов в большей степени определяется их минимальным евклидовым расстоянием.

В настоящей работе строятся системы модуляции и кодирования — оптимальные в смысле евклидова расстояния. В качестве системы сигналов применяется q-позиционная фазовая модуляция, а в качестве системы кодирования используются два класса q-ичных помехоустойчивых кодов — нелинейные решетчатые коды с единичной памятью (РЕП-коды) и нелинейные решетчатые

каскадные коды с единичной памятью (РКЕП(m)-коды). Для указанных систем модуляции и кодирования получены асимптотические нижние границы квадрата свободного евклидова расстояния. Приведены некоторые примеры расчета по полученным формулам для случаев $q=2, 4, 8, 16$.

Результаты расчетов показывают, что в случае использования идентичных способов кодирования системы с $q=4, 8, 16$ обладают примерно одинаковыми корректирующими свойствами при малых скоростях передачи. С увеличением скорости наилучшими становятся системы с $q=16$, далее идут системы с $q=8$, а затем — системы с $q=4$. Следует особо отметить значительный проигрыш систем с $q=2$ по сравнению с остальными системами при любых скоростях передачи. С ростом порядка каскадного кода улучшаются его корректирующие свойства, и хотя РКЕП(m)-коды проигрывают в этом отношении РЕП-кодам, следует учитывать, что последние имеют существенно большую сложность реализации. Это обстоятельство делает более целесообразным использование РКЕП(m)-кодов в системах связи.

В. В. Зяблов
Институт проблем передачи информации АН СССР
СССР 101447 Москва,
ГСП-4, ул. Ермоловой, 19

С. А. Шавгулидзе, М. Г. Шанидзе
Грузинский политехнический институт им. В. И. Ленина
СССР 380075 Тбилиси,
ул. Ленина, 77

Problems of Control and Information Theory, Vol. 17 (6), pp. 357–369 (1988)

AN APPROACH OF APPROXIMATING THE ε-NULL-CONTROLLABILITY SET OF NONLINEAR NONAUTONOMOUS SYSTEMS

A. G. PLOCHOV, P. BURGMEIER

(*Kursk*) (*Köthen*)

(Received September 4, 1987)

For general nonlinear sampled data systems approaches are given for approximating the ε-0, or the 0-controllability sets by use of Lyapunov-functions. They are illustrated by examples.

1.Introduction

Problems of controllability are undoubtedly of great practical importance. This also includes special investigations of the so-called controllability set (CS), the results of which can be utilized in different directions.

Of course, the statements on CS for nonlinear systems are not of such a generality as for linear systems. This leads to many investigations of special systems up to the detailed consideration of separate examples as the equation of Van der Pol with controls (for a summary look [1]). Because the exact determination of the CS is obviously complicated, approximations of it are obtaining justified importance. So, the general investigation of the synthesis problem by Korobov [2] leads to a variant of the estimation of the CS from the inside, whereby so-called controllability-functions are used, in special cases these are Lyapunov-functions. With the help of such functions direct approximations from the inside are recommended in [3] and from the in- and the outside in [4]. In [5, 6, 7] other ways are shown opening practically the possibility for the estimation of the appropriate ranges. All that, however, refers to autonomous systems.

Results for nonautonomous systems are very rare. Here, the dependence of the CS on the initial time t_0 is to be considered additionally. Controllability conditions for such systems were obtained by linearization in [8, 9] and by the use of controllability functions in [10]. Statements on CS itself are unknown to us.

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

2. Notation and formulation of the problem

Consider the system

$$\dot{x}(t) = f(x(t), u(t), t) \tag{1}$$

with $f(\cdot): X * U * T \to R^n$, $X \subseteq R^n$, $0 \in \operatorname{int} X$, $U \subset R^m$, U compact, $m \leqq n$ and $T = [t_0, \infty]$. Let $f(\cdot)$ be constrained on $X * U * T$ by $M \in R_+ = \{z \in R : z > 0\}$, e.g. $\| f(x, u, t) \| \leqq M$ for all $(x, u, t) \in X * U * T$, $\| \cdot \|$ denotes the Euclidean norm.

As admissible controls we consider the set $D^0(T; U)$ of piecewise continuous functions from T into U and assume further that for each admissible control (1) with the initial condition $x(t_0) = x^0 \in X$ has a unique solution

$$x(t) = x(t; t_0, x^0, u(\cdot)), \quad t \geqq t_0.$$

This is the case, for example, if f and f_x with respect to x, u, are continuous, and f, with respect to t, is piecewise continuous [11].

Definition. System (1) is called ε-0-controllable from $x^0 \in X$ iff

$$\underset{\varepsilon > 0}{\forall} \; \underset{t_1 \in T}{\exists} \; \underset{u(\cdot) \in D^0(T; U)}{\exists} \; x(t_1; t_0, x^0, u(\cdot)) \in S(\varepsilon)$$

with $S(\varepsilon) = \{z \in R^n : \| z \| < \varepsilon\}$.

The set of all these points $x^0 \in X$ is called ε-Null-controllability set (ε-0-CS) of system (1). □

There are a lot of papers dealing with this kind of controllability in different variants and for different systems. Most of them contain derivations of controllability criteria (necessary or sufficient or both of these) for various linear systems, for instance for autonomous systems without control-restraints in [12], with restrained controls in [13], with time lag in [14], for multipass systems described by ordinary differential equations in [15], for systems in Hilbert spaces in [16] and [17], for uncertain systems in [18] and for these with incomplete information in [19].

The relation between the 0-CS and the ε-0-CS of linear autonomous systems by special restraints of control is discussed in [20].

For bilinear [21], [22] and general autonomous nonlinear systems [23], [24] often the global ε-0-controllability, i.e. ε-0-CS $= R^n$, is considered. For this in many cases Lyapunov-functions are used, the existence of which in connection with these questions is discussed in [24].

In this paper we deal with the approximation of the ε-0-CS, when it is not the whole space, also with the help of such functions.

Hereby let be a Lyapunov-function such a positive definite function $V(\cdot)$ for which $V(\cdot) \in C^0(X * T; R)$, $V(\cdot) \in C^1(X \setminus \{0\} * T; R)$, and $V(0, t) = 0$ on T holds.

A necessary and sufficient condition for $V(\cdot)$ to be positive definite is that there exists a function $v(\cdot) \in C^0(\bar{X}; R)$ ($\bar{X}$ is the closure of X) with $v(0)=0$ and $v(x)>0$ for $x \in \bar{X}\backslash\{0\}$ such that

$$V(x,t) \geqq v(x) \quad \text{for all } (x,t) \in X * T \quad \text{(see [25])}. \tag{2}$$

$\mathscr{L}$ denotes the set of all Lyapunov-functions. Further the following notations are used:

$$Q(t,c)=\{x \in X: V(x,t)<c\}, \quad c \in R_+,$$

$$D(c)=\{x \in X: v(x)<c\}$$

$$D(c,\varepsilon)=D(c)\backslash S(\varepsilon)$$

$$B(E,\eta)=\{x \in X: \inf_{y \in E} \|x-y\|<\eta\} \text{ where } E \subset X,$$

$$G(c,E,\varepsilon,\eta)=D(c,\varepsilon) \cap B(E,\eta)$$

$$W(x,u,t)=V_t(x,t)+\sum_{i=1}^{n} V_{x_i}(x,t) f_i(x,u,t).$$

Below we consider such $c \in R_+$ for which $\overline{D(c)} \subseteq X$ and further let $D(c)$ be constrained for all c. A sufficient condition for this constraint is $\lim_{\|x\| \to \infty} v(x)=\infty$ (see [26]).

3. Main Result

Theorem. Suppose, there exist a number $c \in R_+$ and functions $V(\cdot) \in \mathscr{L}$, $W^*(\cdot) \in C^0(X * U; R)$, and $\Phi(\cdot) \in C^1(X * T; R)$ such that the following conditions hold:

(V1) $W(x,u,t) \leqq W^*(x,u)$ for all $(x,u,t) \in X * U * T$;

(V2) $\min_{u \in U} W^*(x,u) \leqq 0$ for all $x \in D(c)$,

$$(E=\{x \in X: \min_{u \in U} W^*(x,u)=0\});$$

(V3) $|\Phi(x,t)| \leqq b$ for a $b \in R_+$ and all $(x,t) \in X * T$;

(V4) for each $\varepsilon \in R_+$ (with $S(\varepsilon) \subset D(c)$) there exist numbers $\eta, \zeta \in R_+$ and a decomposition of $G=G(c,E,\varepsilon,\eta)$ in G_1 and G_2 with the distance $d(G_1,G_2)>0$ such that for each $\bar{x} \in G$ there is a $\bar{u} \in U$ satisfying the inequalities

$$W^*(\bar{x},\bar{u}) \leqq 0 \quad \text{and} \quad \begin{cases} \Psi(\bar{x},\bar{u},t) > \zeta, & \text{for } \bar{x} \in G_1 \\ \Psi(\bar{x},\bar{u},t) < -\zeta, & \text{for } \bar{x} \in G_2 \end{cases},$$

here is

$$\Psi(x, u, t) = \Phi_t(x, t) + \sum_{i=1}^{n} \Phi_{x_i}(x, t) f_i(x, u, t);$$

(V5) for a function $w(\cdot) \in C^0(R; R)$ with $w(0) = 0$ holds

$$|\Psi(x^1, u, t) - \Psi(x^2, u, t)| \leqq w(\|x^1 - x^2\|)$$

for all $x^1, x^2 \in X$ and $(u, t) \in U * T$.

Then the ε-0-CS of system (1) contains the set $Q(t_0, c)$.

Proof. Obviously $0 \in \operatorname{int} Q(t_0, c)$ and because of (2) $Q(t_0, c) \subset D(c)$. Hence from now we can consider such $\varepsilon \in R_+$ for which $S(\varepsilon) \subset Q(t_0, c) \subset D(c)$ holds. For these we introduce the sets

$$H = H(c, \varepsilon, \eta) = \overline{D(c, \varepsilon)} \setminus B(E, \eta/2)$$

and

$$G = G(c, E, \varepsilon, \eta) = D(c, \varepsilon) \cap B(E, \eta)$$

and remark that H is compact and that $D(c, \varepsilon) \subseteq H \cup G$. We shall show that each point $x^0 \in Q(t_0, c)$ for arbitrary given $\varepsilon \in R_+$ can be steered in $S(\varepsilon)$ in finite time. For this an admissible control is constructed depending of $x^0 \in H$ or $x^0 \in G$ that the appropriate trajectory remains in $D(c)$ but not in H or G, respectively, and finally always leeds into $S(\varepsilon)$.

A. Let us consider the set H.

A1. First it is shown, that there is a $\beta \in R_+$ with $\sup_{x \in H} \min_{u \in U} W^*(x, u) = -\beta$.

Assume that $\beta = 0$. Then there exists a sequence $\{(x^j, u^j)\} \subset H * U$ with $\lim_{j \to \infty} W^*(x^j, u^j) = 0$, where u^j satisfies the condition $W^*(x^j, u^j) = \min_{u \in U} W^*(x^j, u)$. From this one can select a partial sequence $\{(x^k, u^k)\}$ converging against $(x^*, u^*) \in H * U$, for which in consequence of the continuity $W^*(x^*, u^*) = \lim_{k \to \infty} W^*(x^k, u^k) = 0$ holds. Hence, there are $x^* \in E$ which contradicts the definition of H.

A2. Suppose that $\hat{x} \in H$, $\hat{t} \geqq t_0$ and $\hat{u} \in U$ such that $W^*(\hat{x}, \hat{u}) = \min_{u \in U} W^*(\hat{x}, u)$. The trajectory $x(t) = x(t; \hat{t}, \hat{x}, \hat{u})$ starting from $\hat{x}$ at the time $\hat{t}$ by the given control $u(\cdot) = \hat{u}$ satisfies, for $t_2 \geqq t_1 \geqq \hat{t}$, the inequality

$$\|x(t_2) - x(t_1)\| \leqq \int_{t_1}^{t_2} \|f(x(t), \hat{u}, t)\| dt \leqq M(t_2 - t_1) \tag{3}$$

while $x(t)$ remains in X. For $W^*(\cdot)$ it is uniformly continuous on the compact $H * U$, for all $\gamma \in R_+$ there is a $\Delta \in R_+$ such that $|W^*(x^2, \hat{u}) - W^*(x^1, \hat{u})| < \gamma$, if $\|x^2 - x^1\| < \Delta$

for arbitrary x^1, $x^2 \in H$, and $u \in U$. Hence,

$$W^*(x(t; \hat{t}, \hat{x}, \hat{u}), \hat{u}) < -\beta/2, \tag{4}$$

when $\gamma = \beta/2$, $x^1 = \hat{x}$, $t_1 = \hat{t}$ and $\|x(t) - \hat{x}\| < \Delta$, which is satisfied because of (3) for $0 < t - \hat{t} < \delta = \Delta/M$. Therefore, there is a $\delta \in R_+$ independent of $\hat{t} \in T$, and $\hat{x} \in H(c, \varepsilon, \eta)$ such that (4) holds while $0 < t - \hat{t} < \delta$.

A3. Suppose $x^0 \in Q(t_0, c) \cap H(c, \varepsilon, \eta)$. Then the control $u(\cdot)$ with

$$u(t) = \begin{cases} u^0 & \text{for } t \in [t_0, t_0 + \delta] \\ u^1 & \text{for } t \in [t_0 + \delta, t_0 + 2\delta] \\ \dots, \end{cases}$$

where u^i is chosen from $W^*(x^i, u^i) = \min_{u \in U} W^*(x^i, u)$ as long as

$$x^i = x(t_0 + i\delta; t_0 + (i-1)\delta, x^{i-1}, u^{i-1}) \in H(c, \varepsilon, \eta), \quad i > 0,$$

has the following effect: Because of the appropriate trajectory for the sake of (V1) and (4)

$$v(x(t)) \leqq V(x(t), t) = V(x^0, t_0) + \int_{t_0}^{t} W(x(s), u(s), s)\, ds \leqq$$

$$\leqq V(x^0, t_0) + \int_{t_0}^{t} W^*(x(s), u(s))\, ds \leqq$$

$$\leqq V(x^0, t_0) - (\beta/2)(t - t_0) < c - (\beta/2)(t - t_0)$$

holds, on the one hand it cannot leave $D(c)$, on the other hand, however, it not remains in $H(c, \varepsilon, \eta)$ longer than $2c/\beta$. Therefore the control constructed in such a way steeres the state x^0 either in $S(\varepsilon)$, thus proving the theorem, or in $G(c, E, \varepsilon, \eta)$.

B. Let us take up G.

B1. Assume $\bar{x} \in G(c, E, \varepsilon, \eta)$, $\bar{t} \in T$ and $\bar{u}$ are chosen according to (V4). Then there is again by reason of the uniform continuity of W^* on the compact $\bar{G} * U$ and of (V5), for each $\gamma \in R_+$, $\zeta_1 \in R_+$ a $\Delta_1 \in R_+$ such that for all $x \in \bar{G}$ as well

$$|W^*(x, \bar{u}) - W^*(\bar{x}, \bar{u})| < \gamma \quad \text{as} \quad |\Psi(x, \bar{u}, \bar{t}) - \Psi(\bar{x}, \bar{u}, \bar{t})| < \zeta_1$$

are satisfied, if $\|x - \bar{x}\| < \Delta_1$. Suppose $\Delta_1 < d(G_1, G_2)$. With (3), in analogy to A2, there exists a $\delta' = \Delta_1/M$ such that for the trajectory $x(t) = x(t; \bar{t}, \bar{x}, \bar{u}) \subset \bar{G}(c, E, \varepsilon, \eta)$ holds

$$W^*(x(t), \bar{u}) < \gamma \tag{5}$$

and

$$|\Psi(x(t), \bar{u}, \bar{t}) - \Psi(\bar{x}, \bar{u}, \bar{t})| < \zeta_1, \tag{6}$$

while $0 < t - \bar{t} < \delta'$ and $x(t)$ does not leave $\bar{G}$.

Because $\Psi(\cdot)$, with respect to t, is also uniformly continuous on each compact $T' \subseteq T$ for each $\zeta_2 \in R_+$ there is a $\delta'' \in R_+$ for which $|\Psi(x, u, t_1) - \Psi(x, u, t_2)| < \zeta_2$ holds for arbitrary $(x, u) \in X * U$ and $t_1, t_2 \in T'$, while $|t_1 - t_2| < \delta''$. With this from (6) it follows that $|\Psi(x(t), \bar{u}, t) - \Psi(\bar{x}, \bar{u}, \bar{t}) < \zeta_1 + \zeta_2$ which implies by $\zeta_1 + \zeta_2 = \zeta/2$

$$\Psi(x(t), \bar{u}, t) < -\zeta/2, \quad \text{for} \quad \bar{x} \in G_1(c, E, \varepsilon, \eta)$$

$$\Psi(x(t), \bar{u}, t) > \zeta/2, \quad \text{for} \quad \bar{x} \in G_2(c, E, \varepsilon, \eta) \tag{7}$$

if $0 < t - \bar{t} < \delta_1 = \min\{\delta', \delta''\}$ and $x(t)$ remains in $\bar{G}$.

B2. As in A2, δ_1 is independent of $\bar{x}$ and $\bar{t}$, and, therefore, it always allows the construction of an admissible control in the same way. Assume the initial point $x^0 \in Q(t_0, c) \cap G(c, E, \varepsilon, \eta)$ at the time t_0. Then one constructs $u(\cdot)$ such that

$$u(t) = \begin{cases} \bar{u}^0 & \text{for} \quad t \in [t_0, t_0 + \delta_1] \\ \bar{u}^1 & \text{for} \quad t \in [t_0 + \delta_1, t_0 + 2\delta_1] \\ \dots, \end{cases}$$

where $\bar{u}^i$ is this time chosen according to (V4) as long as the points

$$\bar{x}^i = x(t_0 + i\delta_1; t_0 + (i-1)\delta_1, \bar{x}^{i-1}, \bar{u}^{i-1}), \quad i \geqq 0, \quad \bar{x}^0 = x^0,$$

belong to $G_1(c, E, \varepsilon, \eta)$ and $G_2(c, E, \varepsilon, \eta)$, respectively. For the appropriate trajectory $x(t) = x(t; t_0, x^0, u(\cdot))$ then with (V3) and (7) holding

$$2b \geqq |\Phi(x(t), t) - \Phi(x^0, t_0)| = \left| \int_{t_0}^{t} \Psi(x(s), u(s), s) ds \right| =$$

$$= \int_{t_0}^{t} |\Psi(x(s), u(s), s)| ds > (t - t_0)\zeta/2.$$

Hence $x(t)$ cannot remain in G_1 and G_2, respectively longer than $4b/\zeta$.

On the other hand, this trajectory cannot leave $D(c)$. This is why, with (5), the following is true:

$$v(x(t)) \leqq V(x(t), t) = V(x^0, t_0) + \int_{t_0}^{t} W(x(s), u(s), s) ds \leqq$$

$$\leqq V(x^0, t_0) + \int_{t_0}^{t} W^*(x(s), u(s)) ds <$$

$$< V(x^0, t_0) + \gamma(t - t_0) < V(x^0, t_0) + \gamma 4b/\zeta.$$

Since $V(x^0, t_0) < c$, there is an $\alpha \in R_+$ with $V(x^0, t_0) = c - \alpha$, and that is why the above statement is true by choosing $\gamma < \alpha\zeta/4b$. Therefore, $x(t)$ leads from $G(c, E, \varepsilon, \eta)$ into $S(\varepsilon)$ the theorem is proved) or into $H(c, \varepsilon, \eta)$.

C. To be done is the case that a trajectory according to A3 and B2 changes several times from G into H and back without reaching $S(\varepsilon)$. In this situation there are three times $t_1 < t_2 < t_3$ at which the change from G into H, from H into G and, finally, back to H takes place. Hence, $d(x(t_1), E) = d(x(t_3), E) = \eta$ and $d(x(t_2), E) = \eta/2$. So from

$$\eta/2 \leqq \|x(t_2) - x(t_1)\| \leqq \int_{t_1}^{t_2} \|f(x(t), u(t), t)\| dt \leqq M(t_2 - t_1)$$

follows that $t_2 - t_1 \geqq \eta/2M$. In $[t_1, t_2]$ the control is constructed according to A3, in $[t_2, t_3]$ according to B2. By this we obtain

$$\begin{aligned} V(x(t_3), t_3) &= V(x(t_1), t_1) + \int_{t_1}^{t_3} W(x(t), u(t), t) dt \leqq \\ &\leqq V(x(t_1), t_1) + \int_{t_1}^{t_3} W^*(x(t), u(t)) dt < \\ &< V(x(t_1), t_1) - (t_2 - t_1)\beta/2 + (t_3 - t_2)\gamma < \\ &< V(x(t_1), t_1) - \beta\eta/8M, \end{aligned}$$

if $\gamma < \min\{\alpha\zeta/4b, \zeta\beta\eta/32bM\}$ is chosen.

Hence only a finite number of changes of the trajectory from G into H and back can occur, and the trajectory must lead into $S(\varepsilon)$ in finite time. The theorem is completely proved. □

Remark. A similar method based on the application of several support functions (so-called vector-Lyapunov-functions) was used by Matrosov [27] in order to investigate the properties of solutions of systems of differential equations.

4. Examples and corollaries

Example 1. Suppose the system is described by

$$\begin{aligned} \dot{x}_1(t) &= p(t)x_2(t) \\ \dot{x}_2(t) &= x_2(t) - q(t)x_1(t) + g(x_2(t)) + u(t) \end{aligned} \tag{8}$$

with $X = \{x \in R^2 : |x_1| < a, |x_2| < 1\}$, $a \in R_+$, $U = [-1, 1]$, $T = [t_0, \infty]$,

$$p(\cdot) \in C^1(T; R), \quad p(t) \geqq p > 0, \quad p_1 \leqq \dot{p}(t) \leqq 0, \quad t \in T,$$

$$q(\cdot) \in C^1(T; R), \quad q(t) \geqq q > 0, \quad \dot{q}(t) \leqq 0, \quad t \in T,$$

$$g(\cdot) \in C^1(]-1, 1[; R), \quad g(0) = 0, \quad x_2 g(x_2) \leqq 0 \quad \text{for all } |x_2| < 1.$$

Let us consider the Lyapunov-function $V(\cdot)$ with $V(x,t)=0.5(q(t)x_1^2+p(t)x_2^2)$, which is positive definite by (2) with $v(x)=0.5(qx_1^2+px_2^2)$. Then it is

$$W(x,u,t)=0.5(\dot{q}(t)x_1^2+\dot{p}(t)x_2^2)+p(t)\,(x_2^2+x_2g(x_2)+x_2u)$$

and one can choose as $W^*(\cdot)$ a function $W^*(x,u)=p(t_0)\,(x_2^2+x_2u)$ to satisfy (V1). Obviously at the same time (V2) is satisfied because of

$$\min_{u\in U} W^*(x,u)=p(t_0)\,(x_2^2-|x_2|)\leqq 0$$

for all $x\in X$, and it is $E=\{x\in X\colon x_2=0\}$.

The three other assumptions of the theorem can be satisfied e.g. by a function $\Phi(\cdot)$ of the form $\Phi(x,u,t)=x_1-p(t)x_2$, where (V3) holds since $p(t_0)\geqq p(t)\geqq p$. With $\Psi(x,u,t)=-\dot{p}(t)x_2+p(t)\,(q(t)x_1-g(x_2)-u)$ the inequality

$$\begin{aligned}|\Psi(x^1,u,t)-\Psi(x^2,u,t)|&\leqq|-\dot{p}(t)||x_2^1-x_2^2|+\\&+|p(t)q(t)||x_1^1-x_1^2|+|p(t)||g(x_2^1)-g(x_2^2)|\leqq\\&\leqq(-p_1+p(t_0)q(t_0)+p(t_0)L)\|x^1-x^2\|\end{aligned}$$

implies (V5), where $L\geqq|g'(x_2)|$ for all $|x_2|<1$.

Still to be proved is (V4). For this purpose assume $\varepsilon\in R_+$ and $c\in R_+$ such that $\overline{D(c)}=\{x\in R^2\colon qx_1^2+px_2^2\leqq 2c\}\subset X$. Furthermore, suppose only $\eta<\varepsilon$, then

$$G(c,E,\varepsilon,\eta)=\{x\in R^2\colon\ |x_1|\geqq\sqrt{\varepsilon^2-\eta^2},\ \ |x_2|<\eta,\ \ x\in D(c,\varepsilon)\}$$

can be broken up into sets $G_1(c,E,\varepsilon,\eta)$ with $x_1\geqq\sqrt{\varepsilon^2-\eta^2}$ and $G_2(c,E,\varepsilon,\eta)$ with $x_1\leqq-\sqrt{\varepsilon^2-\eta^2}$ in the required sense. If one chooses for a given $\bar{x}\in G(c,E,\varepsilon,\eta)$ an admissible control $\bar{u}(\cdot)$ of the form $\bar{u}=-\bar{x}_2$, then $W^*(\bar{x},\bar{u})=0$ holds. Under the assumptions on $g(\cdot)$ we have $|g(x_2)|<pq\varepsilon/16p_0$, while $|x_2|<\eta_1=\min\{pq\varepsilon/16p_0,\ pq\varepsilon/16p_0L\}$, where $p_0=\min\{p(t_0),|p_1|\}$, and $|g(x_2)-x_2|<pq\varepsilon/8p_0$, when $|x_2|<\eta_1$.

Setting $\eta=\min\{\eta_1,\varepsilon/4\}$ we obtain for $\bar{x}\in G_1(c,E,\varepsilon,\eta)$,

$$\Psi(\bar{x},\bar{u},t)\geqq-|p_1|\,|\bar{x}_2|+pq\sqrt{\varepsilon^2-\eta^2}-p(t_0)\,|g(\bar{x}_2)-\bar{x}_2|>pq\varepsilon/2$$

and for $\bar{x}\in G_2(c,E,\varepsilon,\eta)$,

$$\Psi(\bar{x},\bar{u},t)\leqq|p_1|\,|\bar{x}_2|-pq\sqrt{\varepsilon^2-\eta^2}+p(t_0)\,|g(\bar{x}_2)-\bar{x}_2|<-pq\varepsilon/2.$$

With this also (V4) is true and, therefore, system (8) is ε-0-controllable from each point of the set

$$Q(t_0,c)=\{x\in X\colon q(t_0)x_1^2+p(t_0)x_2^2<2c\}.$$

Remark. The simple linear autonomous system, which can be derived from the above example by setting $p(t) \equiv 1$, $q(t) \equiv 1$ and $g(x_2) \equiv 0$ for all x, t, is considered in [28].

All further examples are to demonstrate the applicability of the theorem and at the same time to indicate some variants of the choice of the Lyapunov-function $V(\cdot)$ without carrying out the approximations in detail.

Example 2. Suppose the system

$$\begin{aligned} \dot{x}_1 &= p(t)\,[f(x_1(t), x_2(t)) + u(t)] \\ \dot{x}_2 &= h(x_1(t)) \end{aligned} \tag{9}$$

satisfies the following conditions:

$$t_0 > 0, \quad U = [-1, 1], \quad p(\cdot) \in C^1(T; R), \quad 0 < p(t_0) \leqq p(t) \leqq p_1, \quad \dot{p}(t) \geqq 0, \quad t \in T,$$

$$f(\cdot) \in C^1(R^2; R), \quad f(0,0) = 0, \quad h(\cdot) \in C^1(R; R), \quad x_1 h(x_1) > 0 \quad \text{for}$$

$$x_1 \neq 0, \quad \lim_{|x_1| \to \infty} \int_0^{x_1} h(s)ds = \infty .$$

$V(\cdot)$ with $V(x,t) = 0.5x_2^2 + (1/p(t)) \int_0^{x_1} h(s)ds$ proves to be a suitable Lyapunov-function.

For it we have

$$W(x, u, t) = -(\dot{p}(t)/p^2(t)) \int_0^{x_1} h(s)ds + h(x_1)\,[f(x_1, x_2) + u] + x_2 h(x_1)$$

such that with $W^*(x, u) = h(x_1)\,[f(x_1, x_2) + x_2 + u]$ (V1) is satisfied. Based on the assumptions on $f(\cdot)$ and $h(\cdot)$ there is a set $Q = \{x \in R^2 : |f(x_1, x_2) + x_2| < 1\}$ in which $W^*(x, u) \leqq 0$, when

$$u = \begin{cases} -1 & \text{for } x_1 \geqq 0 \\ 1 & \text{for } x_1 < 0 \end{cases}$$

and in which is $E = \{x \in R^2 : x_1 = 0\}$.

The function $\Phi(\cdot)$ can simply be chosen as $\Phi(x, t) = x_1$, so that (V3) (assume X is accordingly fixed in a suitable way) and with $\Psi(x, u, t) = p(t)\,[f(x_1, x_2) + u]$ also (V5) are satisfied.

Let $\varepsilon \in R_+$ and $c \in R_+$ be given such that $D(c, \varepsilon) \subset Q$, then $G(c, E, \varepsilon, \eta) = \{x \in D(c, \varepsilon) : |x_1| < \eta, |x_2| \geqq \sqrt{\varepsilon^2 - \eta^2}\}$ can be broken up into G_1 and G_2 analogously to example 1, when $\eta < \varepsilon$. By choosing $\bar{u} = -f(\bar{x}, \bar{x}_2) - \bar{x}_2$, for $\bar{x} \in G(c, E, \varepsilon, \eta)$ is $\bar{u} \in U$ (since $\bar{x} \in Q$) and $W^*(\bar{x}, \bar{u}) = 0$. In addition to that

$$\Psi(\bar{x}, \bar{u}, t) = -\bar{x}_2 p(t) \leqq -\sqrt{\varepsilon^2 - \eta^2}\, p(t_0) \quad \text{when} \quad \bar{x}_2 \geqq \sqrt{\varepsilon^2 - \eta^2}$$

and

$$\Psi(\bar{x}, \bar{u}, t) = -\bar{x}_2 p(t) \geqq \sqrt{\varepsilon^2 - \eta^2}\, p(t_0) \quad \text{when} \quad \bar{x}_2 \leqq -\sqrt{\varepsilon^2 - \eta^2}.$$

Hence, (V4) also holds.

Therefore, all conditions of the theorem are satisfied in $Q(t_0, c) \subset Q$ and (9) is ε-0-controllable from each point of $Q(t_0, c)$.

Some corollaries can be deduced from the theorem and its proof respectively.

Corollary 1. If

$$AS(x^0, t) = \{x(t; t_0, x^0, u(\cdot)) \colon u(\cdot) \in D^0([t_0, t]; U)\}$$

denotes the attainable set of system (1) from x^0 at the time t, then the set $\{x^0 \in R^n \colon \exists_{t \geqq t_0} AS(x^0, t) \cap Q(t, c) \neq \varnothing\}$ belongs to ε-0-CS of the system. □

The approximation of the range looked for is essentially simplified if one succeeds in finding a Lyapunov-function with negative "derivative".

Corollary 2. If the two functions $V(\cdot) \in \mathscr{L}$ and $W^*(\cdot) \in C^0(X * U; R)$ satisfy conditions (V1) and

(V2′) $\min_{u \in U} W^*(x, u) < 0$ for all $x \in D(c) \backslash \{0\}$, $c \in R_+$, then $Q(t_0, c)$ belongs to the ε-0-CS of system (1).

Proof. In analogy to A1 (in the proof of the theorem) there is a $\beta_1 \in R_+$ with

$$\sup_{x \in \overline{D(c, \varepsilon)}} \min_{u \in U} W^*(x, u) = -\beta_1.$$

With A2 and A3 the trajectory from $x^0 \in Q(t_0, c)$ constructed correspondingly cannot leave the range $D(c)$. However, it cannot remain in $D(c, \varepsilon)$ longer than $2c/\beta_1$, either. Therefore, it must lead into $S(\varepsilon)$. □

Example 3. Consider the system

$$\begin{aligned} \dot{x}_1(t) &= x_1(t) f(x_1(t), x_2(t)) \\ \dot{x}_2(t) &= x_2(t) g(x_1(t), x_2(t)) + u(t) \end{aligned} \tag{10}$$

with the following assumptions: $f(\cdot)$, $g(\cdot) \in C^1(R^2; R)$, $f(0, 0) = g(0, 0) = 0$, $f(x) < 0$ for all $x \in \mathrm{R}^2 \backslash \{0\}$.

Then with the Lyapunov-function $V(x) = \sqrt{x_1^2 + x_2^2}$ we obtain

$$\begin{aligned} W^*(x, u) = W(x, u) &= \frac{x_1^2 f(x)}{\sqrt{x_1^2 + x_2^2}} + \frac{x_2(x_2 g(x) + u)}{\sqrt{x_1^2 + x_2^2}} = \\ &= \sqrt{x_1^2 + x_2^2}\, f(x) + (x_2 g(x) - x_2 f(x) + u) x_2 / \sqrt{x_1^2 + x_2^2} < \\ &< (x_2(g(x) - f(x)) + u) x_2 / \sqrt{x_1^2 + x_2^2} \leqq 0 \end{aligned}$$

when

$$u = \begin{cases} -1, & \text{for } x_2 \geqq 0 \\ 1, & \text{for } x_2 < 0 \end{cases}$$

in a range $Q = \{x \in R^2\colon |x_2|\,|g(x) - f(x)| < 1\}$. Hence, each point of $D(c) \subset Q$ is ε-0-controllable.

If instead of (1) an autonomous system

$$\dot{x}(t) = f(t), u(t)) \tag{1'}$$

is given with the same or appropriate assumptions and if $\mathscr{L}'$ denotes the set of positive definite Lyapunov-functions $V(\cdot)\colon R^n \to R$, and $W(x,u) = \sum_{i=1}^{n} V_{x_i}(x) f_i(x,u)$, then we have

Corollary 3. Suppose $c \in R_+$, $V(\cdot) \in \mathscr{L}'$ and $\Phi(\cdot) \in C^1(R^n; R)$ such that

(i) $\min_{u \in U} W(x,u) \leqq 0$ for all $x \in D(c)$;

(ii) for each $\varepsilon \in R_+$ (with $S(\varepsilon) \subset D(c)$) there are such $\eta, \zeta \in R_+$ and a decomposition of $G(c, E, \varepsilon, \eta)$ in $G_1(c, E, \varepsilon, \eta)$ and $G_2(c, E, \varepsilon, \eta)$ with the distance $d(G_1, G_2) > 0$, that for each $\bar{x} \in G$ there exists a $\bar{u} \in U$ satisfying the inequalities

$$W(\bar{x}, \bar{u}) \leqq 0 \quad \text{and} \quad \begin{cases} \Psi(\bar{x}, \bar{u}) > \zeta, & \text{for } x \in G_1(c, E, \varepsilon, \eta) \\ \Psi(\bar{x}, \bar{u}) < -\zeta, & \text{for } \bar{x} \in G_2(c, E, \varepsilon, \eta) \end{cases}.$$

Then the set $D(c)$ belongs to the ε-0-CS of system (1'). □

Remark: Essentially this corresponds to Theorem 1 in [4]. If for (1') in addition we assume local 0-controllability, then with $D(c)$ one obtains an approximation of the 0-CS of (1') from the inside.

5. Conclusion

The method recommended above is, as far as we know, the first approach for approximating the ε-0-CS or the 0-CS of such general nonlinear systems under the given assumptions. Of course, in concrete examples to prove the validity of the conditions of the theorem (especially (V4)) can be cumbersome. Therefore, the question arises whether, on the one hand, simplifications of the method or, on the other, its numerical realization are possible. The authors continue their work in both directions.

Acknowledgement

The authors wish to thank Prof. J. Bergmann for the valuable comments during the preparation of this paper and StR. D. Gerth for his help by its translation into English.

References

1. *Плисс П. В.* О кривой разрывов множества управляемости уравнения Ван-дер-Поля с управлением. Вестн. ЛГУ (1981) *2*, 59–62.
2. *Коробов В. И.* Общий подход к решению задачи синтеса ограниченных управлений в задаче управляемости. Матем. сборн., **109** (1979), *4*, 582–606.
3. *Блинов А. П.* Об оценке области управляемости в нелинейных системах. ПММ **48** (1984) *4*, 593–600.
4. *Plochov, A. G., Burgmeier, P.*, Zur Abschätzung des Nullsteuerbarkeitsbereiches nichtlinearer Systeme mit konzentrierten Parametern. Optimization **16** (1985), *6*, 863–868.
5. *Байтман М. М.* Синтез оптимальных траекторий в плоскости, СУНАТНЕ, Рига 1971.
6. *Брандин В. Н.* Достаточное условие управляемости нелинейных систем с ограничениями на управление. Изв. АН СССР Техн. киберн. (1977) **6**, 164–166.
7. *Grantham, W. J., Vincent, T. L.*, A controllability minimum principle. JOTA **17** (1975), *1/2* 93–114.
8. *Габасов Р., Кириллова Ф.* Качественная теория оптимальных процессов. „Наука", Москва 1971.
9. *Тонков Е. Л.* Равномерная локальная управляемость и стабилизация нелинейных рекуррентных систем. ДУ **18** (1982), 908–910.
10. *Овечкина Х. А.* Об условиях управляемости и стабилизации движения нелинейных систем, Вопросы качеств. теории дифф. уравн. Чебоксары, 1982, 72–82.
11. *Габасов Р., Кириллова Ф.* Принцип максимума в теории оптимальных управлений. Изд. Наука и техника, Минск 1974.
12. *Габелая А. Г., Иваненко В. И., Одарич О. Н.* Стабилизируемость автономных систем управления. Кибернетика (1975) **3**, 69–73.
13. *Коробов В. И. Нгуен Хоа Шон* ε-управляемость линейных автономных систем при наличии ограничений на управление. ДУ **14** (1980) *3*, 95–104.
14. *Куперман Л. И.* К теории ε-управляемости линейных систем с запаздыванием. ДУ **14** (1978) *10*, 1791–1975.
15. *Collins, W. D.*, Approximate controllability of multipass systems described by linear ordinary differential systems. Control Theory, Proc. 3rd. IMA Conf., Sheffield 1980 (1981), 685–707.
16. *Levan, N.*, Controllability, *-controllability and stabilizability. J. of Diff. Equ., **38** (1980) *1*, 61–79.
17. *Germani, A., Monaco, S.*, Functional output ε-controllability for linear systems in Hilbert spaces. System & Control Letters **2** (1983), 313–320.
18. *Barmish, B. R.*, Stabilization of uncertain systems via linear control. IEEE Trans. Autom. Control, **AC-28** (1983) *8*, 848–850.
19. *Габелая А. Г.* Критерии стабилизируемости двумерных линейных систем с неполной информацией. Сообщ. АН ГССР **112** (1983), *1*, 49–51.
20. *Луценко А. В., Синицын П. В.* О стабилизации линейной системы финитными управлениями. ДУ **10** (1974) *5*, 799–803.
21. *Derese, I., Noldus, E.*, Stabilization of bilinear systems. Lect. Notes Contr. and Inf. Sci., **44** (1982), 974–987.
22. *Ryan, E. P., Buckingham, N. J.*, On asymptotically stabilizing feedback control of bilinear systems. IEEE Trans. Autom. Contr. **AC-28** (1983) *8*, 863–865.
23. *Artstein, Z.*, Stabilization with relaxed controls. Nonlin. Anal., Theory, Methods and Appl., **7** (1983) *11*, 1163–1173.
24. *Sontag, E. D.*, A Lyapunov-like charakterization of asymptotic controllability. SIAM J. Contr. Opt., **21** (1983) *3*, 462–471.
25. *Руш Н., Абетс П., Лалуа М.* Прямой метод Ляпунова в теории устойчивости. „Мир", Москва 1980.
26. *Барбашин Е. А.* Функции Ляпунова. Изд. „Наука", Москва 1970.
27. *Матросов В. М.* Об устойчивости движения. ПММ **26** (1962) *6*, 863–868.
28. *Формальский А. М.* Управляемость и устойчивость систем с огрениченными ресурсами. „Наука", Москва, 1974.

Метод аппроксимации множества ε-0-управляемости нелинейных неавтономных систем

А. Г. ПЛОХОВ, П. БУРГМЕЙЕР

(Курск) (Кётен)

Для нелинейных, неавтономных систем с сосредоточенными параметрами оценивается область ε-0-управляемости с помощью функции Ляпунова.

А. Г. Плохов
Гос. Пед. Инст. Курск
мат. фак.
305416 Курск
Радищева 33,
СССР

P. Burgmeier
Pedagogische Hochschule, Köthen
Sektion Math.
Lehmannstr. 23
Köthen 4370
DDR

Problems of Control and Information Theory, Vol. 17 (6), pp. 371–380 (1988)

EQUILIBRIUM EXISTENCE FOR SOME HEATING CONTROL SYSTEMS

S. A. BRYKALOV

(Sverdlovsk)

(Received August 5, 1987)

The paper deals with the steady state existence problem for the feedback controlled heating of a rod or a ring. It is supposed that parameters of a distributed heat source and a finite number of point sources depend upon the temperature distribution because of the feedback. We use nonlinear boundary value problems for functional-differential equations to characterize time-independent temperature distributions.

Introduction

Various feedback controlled systems are widely spread in many branches of technology. A control law being chosen and fixed, the feedback control device governs the process taking into account its parameters at the present moment. One of the problems naturally arising here is the question whether there exists an equilibrium. That is, the question whether the parameters of the process guided by a given control law can be time-independent.

The present paper deals with the above mentioned problem in the case of heating control, in which a steady state of the system corresponds to a time-independent temperature distribution over a body being heated. We suppose that the body is one-dimensional (e.g. a rod, a ring), or at least that the considered temperature distributions are one-dimensional (e.g. due to radial symmetry). The heating is carried out by a distributed heat source and a finite number of point sources. The point sources of three types are of interest to us: those enabling the feedback device to control 1) the released heat flux, 2) the temperature, 3) the process of heat exchange in accordance with Newton's cooling law. The parameters of the sources, including the coordinates of the point ones, are appointed by the control device, which takes into account the distribution of temperature. It is supposed that these controlling parameters depend on the temperature distribution continuously.

In Section 1 we consider one special case concerned with rod heating, the problem statement is given, and the result is formulated. Ring heating is dealt with

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

in a similar way in Section 2. For the sake of simplicity in these two sections we restrict ourselves to some particular cases and the question in the terms of heating control is discussed. Section 3 is devoted to the general case of this problem considered as a mathematical one and sufficient conditions for the existence of solution for some nonlinear boundary value problems for functional-differential equations are dealt with.

All the functions mentioned in the present paper are scalar. As usual, $C^m[a, b]$, $m \geqq 0$, is the space of m-times continuously differentiable functions, and $L_q[a, b]$, $q \geqq 1$, is the space of Lebesgue measurable $x(\cdot)$ with integrable $|x(\cdot)|^q$. We denote by $CL_q^m[a, b]$, $m \geqq 1$, the subspace of such $x(\cdot) \in C^{m-1}[a, b]$ that $x^{(m-1)}(\cdot)$ is absolutely continuous and $x^{(m)}(\cdot) \in L_q[a, b]$. The norm can be given by

$$\|x(\cdot)\|_{CL_q^m} = \|x(\cdot)\|_{C^{m-1}} + \|x^{(m)}(\cdot)\|_{L_q}.$$

1. Problem statement. Case of rod heating

Let us discuss one particular case of the problem concerned with rod heating. Consider the axis s directed along the rod. Denote the coordinates of the rod ends by a, b, $a < b$. Let $x(s)$ denote the temperature distribution over the rod. The thermal conductivity $k(s)$ depends upon the point s. The density of a distributed heat source is $-h(s)x(s) + f(s)$. Three point sources are placed at the ends a, b and at a variable point s_1. The temperature of the ends a, b is kept equal to T_1, T_2, respectively. A heat flux q is released at the point s_1. It is supposed that the parameters characterizing the heating process take values in bounded sets. The change of $s_1 \in [a, b]$ corresponds to the motion of the point source along the rod. It is natural to suppose that in cases $s_1 = a$, $s_1 = b$ the temperature of the rod ends is still kept equal to T_1, T_2. In other words, the point sources at the rod ends are powerful enough to compensate the heat flux q.

The process of rod heating is controlled by a feedback device. The input of the device is the temperature distribution $x(\cdot) \in C^0[a, b]$. The output is the collection of the parameters $k(\cdot) \in C^1[a, b]$, $h(\cdot)$, $f(\cdot) \in C^0[a, b]$, T_1, T_2, q, $s_1 \in \mathbf{R}$ satisfying inequalities

$$0 < \varepsilon_k \leqq k(s) \leqq N_k, \quad \left|\frac{dk}{ds}(s)\right| \leqq N_k, \quad 0 \leqq h(s) \leqq N_h, \quad |f(s)| \leqq N_f, \tag{1}$$

$$|T_i| \leqq N_T \quad (i = 1, 2), \qquad |q| \leqq N_q, \quad a \leqq s_1 \leqq b \tag{2}$$

where ε_k, N_k, N_h, N_f, N_T, N_q are fixed numbers. The parameters corresponding to a temperature distribution $x(\cdot)$ are denoted by $k(x(\cdot), s)$, $h(x(\cdot), s)$, $f(x(\cdot), s)$, $T_1(x(\cdot))$, $T_2(x(\cdot))$, $q(x(\cdot))$, $s_1(x(\cdot))$. The dependence of the output parameters upon the input data is given by a single-valued continuous mapping. In the present paper it is

supposed that this mapping is fixed, and so the question of control law choice is not touched upon. It should be mentioned that heating control law choice problems are investigated by many authors from the point of view of various formalizations [1–7]. A problem concerned with heating of a rod by means of a movable point source is considered in [4] within the framework of differential game theory.

It is clear on physical grounds that the temperature distribution $x(s, t)$ in general is time-dependent. So are the parameters describing the heat sources. Because of the control device the parameters of the heat sources depend on the temperature distribution, which in its turn is effected by the heat sources. In this paper we confine ourselves to considering time-independent temperature distributions $x(s)$ only. It can be shown that there always exists such an equilibrium of the above described rod heating control system.

In order to characterize the time-independent temperature distributions consider the boundary value problem for the functional-differential equation

$$\frac{d}{ds}(k(x(\cdot), s)\frac{dx}{ds}(s)) - h(x(\cdot), s)x(s) + f(x(\cdot), s) = 0, \tag{3}$$

$$x(a) = T_1(x(\cdot)), \qquad x(b) = T_2(x(\cdot)), \tag{4}$$

$$\frac{dx}{ds}(s_1(x(\cdot)) - 0) - \frac{dx}{ds}(s_1(x(\cdot)) + 0) = Q(x(\cdot)), \qquad \text{if} \quad s_1(x(\cdot)) \neq a, b \tag{5}$$

where $Q(x(\cdot)) = (k(x(\cdot), s_1(x(\cdot))))^{-1} q(x(\cdot))$. Let $x(\cdot)$ be C^0-function on $[a, b]$ and C^2-function both on $[a, s_1(x(\cdot))]$ and $[s_1(x(\cdot)), b]$. Equation (3) should be fulfilled for $s \neq s_1(x(\cdot))$. Boundary condition (5) contains one-side derivatives at the variable point $s_1(x(\cdot))$ dependent on the solution. If $s_1(x(\cdot)) = a$ or $s_1(x(\cdot)) = b$, we suppose that (5) is valid. This agreement can easily be interpreted in terms of point sources.

Thus the solutions $x(\cdot)$ satisfying $s_1(x(\cdot)) = a$ or $s_1(x(\cdot)) = b$ are subject to the two boundary conditions (4). In case $a < s_1(x(\cdot)) < b$ the additional boundary condition (5) is necessary because of the discontinuity of the first derivative. According to Theorem 1 stated in Section 3 below, there exists a solution to (3), (4), (5) showing the existence of a time-independent temperature distribution.

2. Case of ring heating

In the case presented in Section 1 the temperature of the rod ends and the heat flux released by the movable source are controlled. It is clear, however, that this problem statement is not the only possible one. We can combine boundary conditions chosen for each of the points independently. Besides of dealing with the temperature and the heat flux we can as well consider the heat exchange in accordance with

Newton's cooling law. All these problems can be dealt with simultaneously. For this purpose boundary conditions of a somewhat more general type should be used. In the present section we give an example of such a problem for the case of ring heating.

As in the previous section, s takes values between a and b. Thus, we identify the ring with the half-open interval $[a, b)$ shaped so that its ends coincide. In addition to the distance $|\beta-\alpha|$ on $[a, b)$ we need the metric $\rho(\alpha, \beta)=\min\{|\beta-\alpha|, b-a-|\beta-\alpha|\}$ equal to the length of the shortest of the two arcs connecting the points α, β. The mappings $k(x(\cdot), s)$, $h(x(\cdot), s)$, $f(x(\cdot), s)$ satisfy all the above imposed conditions including the inequalities (1). Let $k(x(\cdot), a)\equiv k(x(\cdot), b)$ for all $x(\cdot)$. For the sake of convenience the functions x, k, h, f are supposed to be defined on the closed interval $[a, b]$. There are two movable point sources. Their coordinates $s_1(x(\cdot))$, $s_2(x(\cdot))$, depend on $x(\cdot)$ because of the feedback. Mappings $C^0 \ni x(\cdot)\to s_i(x(\cdot))\in[a, b)$ are continuous in the sense of the metric ρ.

The heat fluxes q_1, q_2 released by the point sources satisfy equalities

$$\mu_i(x(\cdot))q_i=-\lambda_i(x(\cdot))x(s_i(x(\cdot)))+\nu_i(x(\cdot)), \qquad i=1, 2. \tag{6}$$

Coefficients $\mu_i\in[0, N_\mu]$, $\lambda_i\in[0, N_\lambda]$, $\nu_i\in[-N_\nu, N_\nu]$ are continuous in $x(\cdot)\in C^0$. They depend on $x(\cdot)$ due to the feedback control device. The constants N_μ, N_λ, N_ν are fixed.

In case $\mu_i(x(\cdot))\equiv 0$, $\lambda_i(x(\cdot))\equiv 1$ equality (6) takes the form $x(s_i(x(\cdot)))=\nu_i(x(\cdot))$ corresponding to the control of the temperature at the point $s_i(x(\cdot))$. If $\mu_i(x(\cdot))\equiv 1$, $\lambda_i(x(\cdot))\equiv 0$ we have $q_i=\nu_i(x(\cdot))$, and the heat flux released at the point $s_i(x(\cdot))$ is controlled. The equality

$$q_i=-\alpha_i(x(\cdot))(x(s_i(x(\cdot)))-\vartheta_i(x(\cdot))) \tag{7}$$

is also a special case of (6). Here $\alpha_i\geqq 0$, $\vartheta_i\in[-N_\vartheta, N_\vartheta]$ are continuous in $x(\cdot)\in C^0$. Equality (7) corresponds to the heat exchange with the heat flux in proportion to the temperature difference. The temperature ϑ_i and the constant of proportionality α_i are regulated by the control device. The connection between the coefficients of (6) and (7) can be given in the form

$$\mu_i(x(\cdot))=(1+\alpha_i(x(\cdot)))^{-1},$$

$$\lambda_i(x(\cdot))=\alpha_i(x(\cdot))(1+\alpha_i(x(\cdot)))^{-1},$$

$$\nu_i(x(\cdot))=\vartheta_i(x(\cdot))\alpha_i(x(\cdot))(1+\alpha_i(x(\cdot)))^{-1}.$$

These formulae guarantee the boundedness of μ_i, λ_i, ν_i and fulfilment of (11) imposed below.

By χ denote the function such that $\chi(0)=1$, $\chi(z)=0$ for $z\neq 0$. Thus

$$q_1\chi(s-s_1(x(\cdot)))+q_2\chi(s-s_2(x(\cdot)))$$

equals to the heat flux released at a point s by the point sources. This expression

takes one of the values $0, q_1, q_2, q_1+q_2$. On the other hand, the heat flux in question can be written down with the help of the derivatives of x on the left and on the right. Equating the two expressions we obtain

$$k(x(\cdot), s)\left(\frac{dx}{ds}(s-0)-\frac{dx}{ds}(s+0)\right)=$$

$$=q_1\chi(s-s_1(x(\cdot)))+q_2\chi(s-s_2(x(\cdot))) \tag{8}$$

in case $s\in(a, b)$, and

$$k(x(\cdot), a)\left(\frac{dx}{ds}(b-0)-\frac{dx}{ds}(a+0)\right)=$$

$$=q_1\chi(a-s_1(x(\cdot)))+q_2\chi(a-s_2(x(\cdot))) \tag{9}$$

for the point a. Continuity of the temperature distribution at the point a gives

$$x(a)=x(b). \tag{10}$$

A solution to the boundary value problem (3), (6), (8), (9), (10) is the triplet $x(\cdot)$, q_1, q_2. The function $x(\cdot)$ is continuous. Its first and second derivatives are piecewise continuous. Equation (3) should be satisfied for every s except for the points of discontinuity of the first and second derivatives. Condition (8) should be valid for $s\in(a, b)$.

Let us suppose that for some $\varepsilon>0$ and all $x(\cdot)\in C^0$ the following inequalities hold

$$\mu_i(x(\cdot))+\lambda_i(x(\cdot))\geqq\varepsilon, \qquad i=1, 2, \tag{11}$$

$$\mu_1(x(\cdot))+\mu_2(x(\cdot))+\rho(s_1(x(\cdot)), s_2(x(\cdot)))\geqq\varepsilon, \tag{12}$$

$$\lambda_1(x(\cdot))+\lambda_2(x(\cdot))+\int_a^b h(x(\cdot), s)ds\geqq\varepsilon. \tag{13}$$

Then Theorem 2 of Section 3 below implies the existence of a solution to the boundary value problem (3), (6), (8), (9), (10) and thus implies the existence of an equilibrium of the control system under investigation.

Assumptions (11), (12), (13) prevent the boundary value problem from degeneration. Inequality (11) forbids μ_i, λ_i to vanish simultaneously. Thus (11) makes it impossible for (6) to take the form $0=1$. Inequality (12) excludes the case

$$\mu_1=\mu_2=\nu_1=0, \quad s_1=s_2=a, \quad \lambda_1=\lambda_2=\nu_2=1$$

in which the two boundary conditions (6) take the form $x(a)=0$, $x(a)=1$.

Simultaneous vanishing of all the three terms in (13) means that neither distributed nor point heat exchange takes place. Thus (13) excepts the case of the

thermally insulated ring effected by distributed and point heaters. It is clear on physical grounds that in this case the temperature should increase in the course of time because of the absence of heat sinks, and so there are no time-independent temperature distributions.

It is natural to try to replace (11), (12), (13) by the assumptions that the corresponding expressions are not equal to zero. It can be shown, however, that neither of (11), (12), (13) can be relaxed that way.

3. The general case

The equilibrium existence results for feedback heating control systems stated in Sections 1 and 2 can be generalized in some respects, the main of which are the following. Any fixed finite number of point sources can be considered. The distributed heat source density may be discontinuous (Lebesgue measurable). The control device takes into account not only the temperature distribution $x(\cdot) \in C^0$ itself but also some information concerning its derivative (which is connected with the heat flux along the rod or the ring). In the present section we deal with the mathematical aspects of the problem in general case.

Consider a boundary value problem generalizing (3), (4), (5). Fix some $r, q \in [1,\infty)$. Let continuous mappings be given transforming a function $x(\cdot) \in CL^1_r[a, b]$ into functions

$$k(x(\cdot), \cdot) \in CL^1_q[a, b], \quad h(x(\cdot), \cdot) \in L_q[a, b], \quad f(x(\cdot), \cdot) \in L_q[a, b]$$

and numbers

$$\mu_i(x(\cdot)) \in [0, N_\mu], \quad \lambda_i(x(\cdot)) \in [0, N_\lambda], \quad v_i(x(\cdot)) \in [-N_v, N_v], \quad s_i(x(\cdot)) \in [a, b]$$

where $i = 1, \ldots, n$. It is supposed that

$$\left.\begin{aligned} &0 < \varepsilon_k \leqq k(x(\cdot), s) \leqq N_k, \qquad \left\| \frac{d}{ds} k(x(\cdot), \cdot) \right\|_{L_q} \leqq N_k, \\ &h(x(\cdot), s) \geqq 0, \qquad \|h(x(\cdot), \cdot)\|_{L_q} \leqq N_h, \qquad \|f(x(\cdot), \cdot)\|_{L_q} \leqq N_f. \end{aligned}\right\} \tag{14}$$

The numbers N_μ, N_λ, N_v, ε_k, N_k, N_h, N_f are fixed.

For instance, the conditions imposed on f are fulfilled if

$$f(x(\cdot), s) = \varphi(s, x(s), \frac{dx}{ds}(s)),$$

where φ: $[a, b] \times \mathbf{R}^2 \to \mathbf{R}$ is a bounded Carathéodory function. The following can be mentioned. The considered mappings are supposed to be continuous in the sense of

the norms CL_r^1, CL_q^1, L_q. In Sections 1 and 2 the norms C^0, C^1, C^0 are used instead. Thus, in the present section we impose less restrictive continuity assumptions than in Sections 1 and 2.

Let S denote the set of all the functions $x(\cdot) \in C^0[a, b]$ such that there exists some finite partition $a = \xi_0 < \xi_1 < \ldots < \xi_m = b$ (dependent on $x(\cdot)$) with the following property. For each i the function $x(\cdot)$ considered on $[\xi_i, \xi_{i+1}]$ is CL_q^2-function. Thus a function belonging to S is continuous, its first derivative is piecewise continuous, its second derivative to the power q is integrable.

Consider the problem

$$\frac{d}{ds}(k(x(\cdot), s)\frac{dx}{ds}(s)) - h(x(\cdot), s)x(s) + f(x(\cdot), s) = 0, \tag{15}$$

$$k(x(\cdot), s)\left(\frac{dx}{ds}(s-0) - \frac{dx}{ds}(s+0)\right) =$$

$$= \sum_{i=1}^{n} q_i \chi(s - s_i(x(\cdot))), \quad s \in [a, b], \tag{16}$$

$$\mu_i(x(\cdot))q_i = -\lambda_i(x(\cdot))x(s_i(x(\cdot))) + v_i(x(\cdot)). \tag{17}$$

As above, we assume that $\chi(0) = 1$, $\chi(z) = 0$ for $z \neq 0$. We say that $x(\cdot) \in S$ and real numbers $q_1, \ldots, q_n$ satisfy the boundary value problem in question, if (15) holds almost everywhere, (16) is valid everywhere on $[a, b]$, and (17) is satisfied for each i.

In case $s = a$ or $s = b$ condition (16) contains the derivatives $\frac{dx}{ds}(a-0)$, $\frac{dx}{ds}(b+0)$ whose values must be defined specially. We suppose them to be equal to zero. According to this agreement, equality (16) implies

$$-k(x(\cdot), a)\frac{dx}{ds}(a+0) = \sum_{i=1}^{n} q_i \chi(a - s_i(x(\cdot))), \tag{18}$$

$$k(x(\cdot), b)\frac{dx}{ds}(b-0) = \sum_{i=1}^{n} q_i \chi(b - s_i(x(\cdot))). \tag{19}$$

Theorem 1. Suppose that

$$\inf(\mu_i(x(\cdot)) + \lambda_i(x(\cdot))) > 0 \qquad \text{for each} \quad i, \tag{20}$$

$$\inf(\mu_i(x(\cdot)) + \mu_j(x(\cdot)) + |s_i(x(\cdot)) - s_j(x(\cdot))|) > 0 \qquad \text{for} \quad i \neq j, \tag{21}$$

$$\inf\left(\sum_{i=1}^{n} \lambda_i(x(\cdot)) + \|h(x(\cdot), \cdot)\|_{L_1}\right) > 0 \tag{22}$$

where inf is taken over all $x(\cdot) \in CL_r^1[a, b]$. Then the boundary value problem (15), (16), (17) has at least one solution.

In neither of (20), (21), (22) the sign of the infimum can be omitted (i.e. an inequality of the form $\inf \Psi(x(\cdot)) > 0$ can not be replaced by the less restrictive assumption $\Psi(x(\cdot)) > 0$). This is shown by the following counter-examples.

Counter-example 1. For a fixed i suppose that

$$\mu_i(x(\cdot)) \equiv 0, \quad \lambda_i(x(\cdot)) = (1 + x^2(a))^{-1}, \quad \nu_i(x(\cdot)) \equiv 1, \quad s_i(x(\cdot)) \equiv a.$$

Then (17) is equivalent to a quadratic equation having no real roots $x(a)$. So the boundary value problem has no solutions. Here (20) does not hold, though $\mu_i(x(\cdot)) + \lambda_i(x(\cdot)) > 0$.

Counter-example 2. Let $a = 0$, $b = 2$, $n = 3$. Consider the three conditions (17) of the form $x(0) = 0$, $x(\psi(\dot{x}(\cdot))) = 1$, $x(2) = 0$, where the continuous non-linear functional $\psi: L_1[0, 2] \to [0, 1]$ satisfies inequalities $\psi(u(\cdot)) > 0$, $\int_0^{\psi(u(\cdot))} u(s)ds < 1$ for all $u(\cdot)$. E.g. it suffices to define $\psi = \psi(u(\cdot))$ as the unique root of the equation $\psi + \int_0^{\psi} |u(l)| dl = 1$. Condition (21) failes to hold for $i = 1, j = 2$, but the corresponding expression is positive for all $x(\cdot)$. If $x(\cdot)$ satisfies the boundary conditions, then

$$1 = x(\psi(\dot{x}(\cdot))) - x(0) = \int_0^{\psi(\dot{x}(\cdot))} \dot{x}(s)ds < 1.$$

This contradiction proves that there are no solutions.

Counter-example 3. Consider a particular case of the problem (15), (16), (17)

$$\frac{d^2x}{ds^2} = 0, \quad \frac{dx}{ds}(0) = (1 + x^2(0))^{-1} x(0) - 1, \quad \frac{dx}{ds}(2) = 0$$

where $x(\cdot) \in C^2[0, 2]$. (The numbers q_1, q_2 are already eliminated). Supposition (22) is not valid, but the expression in (22) is positive for all $x(\cdot)$. The equation and the boundary condition at the point $s = 2$ imply that x is constant. Then the boundary condition at the point $s = 0$ turns out to be a quadratic equation without real roots.

Now we pass on to a boundary value problem generalizing (3), (6), (8), (9), (10). Fix r, $q \in [1, \infty)$. Consider the subspace of functions $x(\cdot) \in CL_r^1[a, b]$ satisfying $x(a) = x(b)$, and some continuous mappings k, h, f, μ_i, λ_i, ν_i $(i = 1, \ldots, n)$ from the mentioned subspace into $CL_q^1[a, b]$, $L_q[a, b]$, $L_q[a, b]$, $[0, N_\mu]$, $[0, N_\lambda]$, $[-N_\nu, N_\nu]$. We suppose that $k(x(\cdot), a) \equiv k(x(\cdot), b)$ and the inequalities (14) hold. Mappings s_i $(i = 1, \ldots, n)$ from the mentioned subspace into $[a, b)$ are continuous in the sense of the metric on $[a, b)$ given by $\rho(\alpha, \beta) = \min \{|\beta - \alpha|, b - a - |\beta - \alpha|\}$.

Consider the differential equation (15) subject to the boundary conditions

$$k(x(\cdot), s)\left(\frac{dx}{ds}(s-0)-\frac{dx}{ds}(s+0)\right)=\sum_{i=1}^{n} q_i\chi(s-s_i(x(\cdot))), \qquad s\in[a, b), \tag{23}$$

$$\mu_i(x(\cdot))q_i=-\lambda_i(x(\cdot))x(s_i(x(\cdot)))+\nu_i(x(\cdot)), \tag{24}$$

$$x(a)=x(b). \tag{25}$$

We suppose that $(x(\cdot), q_1, \ldots, q_n)\in S\times \mathbf{R}^n$, equation (15) should be satisfied almost everywhere, the condition (23) should be valid everywhere on $[a, b)$. Note that in case $s=a$ condition (23) contains the derivative $\frac{dx}{ds}(a-0)$, whose value is not defined. It is supposed that this derivative is to be replaced by $\frac{dx}{ds}(b-0)$. Thus (23) implies

$$k(x(\cdot), a)\left(\frac{dx}{ds}(b-0)-\frac{dx}{ds}(a+0)\right)=\sum_{i=1}^{n} q_i\chi(a-s_i(x(\cdot))). \tag{26}$$

Comparing (18) with (26) we see that conditions (16), (23) in case $s=a$ are interpreted in different ways, though they are of the same form. The following can be mentioned. If a function $x(\cdot)$ is such that $s_i(x(\cdot))\neq a$ for all $i=1, \ldots, n$, then (25), (26) turn out to be the periodic boundary conditions.

Theorem 2. Suppose that (20), (21), (22) and

$$\inf(\mu_i(x(\cdot))+\mu_j(x(\cdot))+s_i(x(\cdot))-a+b-s_j(x(\cdot))>0 \qquad \text{for} \quad i\neq j \tag{27}$$

are satisfied. Let inf in (20), (21), (22), (27) be taken over all $x(\cdot)\in CL_r^1[a, b]$ such that $x(a)=x(b)$. Then there exists at least one solution to the boundary value problem (15), (23), (24), (25).

Theorems 1, 2 are proved together. The detailed proof was given by the author in [8]. For some other existence theorems applicable to boundary value problems with solution dependent points in boundary conditions see [9, 10].

References

1. *Butkovskii, A. G., Pustyl'nikov, L. M.*, Theory of moving control of systems with distributed parameters. Nauka, Moscow, 1980 (in Russian).
2. *Egorov, A. I.*, Optimal control of thermal and diffusion processes. Nauka, Moscow, 1978 (in Russian).
3. *Krasovskii, N. N.*, On differential evolution systems. Prikl. Mat. Mekh. **41** (1977), *5* (in Russian).
4. *Krasovskii, N. N., Tret'jakov, V. E.*, Control solution with optimal ensured result. Problems of Control and Information Theory **11** (1982), *4*.
5. *Lions, J. L.*, Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles, Dunod, Paris, 1968.

6. *Osipov, Yu. S.*, Position control in parabolic systems. Prikl. Mat. Mekh. **41** (1977), *2* (in Russian).
7. *Sirazetdinov, T. K.*, Optimization of systems with distributed parameters. Nauka, Moscow, 1977 (in Russian).
8. *Brykalov, S. A.*, Boundary value problems for functional-differential equations of second order with discontinuity of the first derivative in variable points. Manuscript 6503-V87, deposited at VINITI, 1987 (in Russian).
9. *Brykalov, S. A.*, Some boundary value problems for differential equations and inclusions. Dokl. Akad. Nauk SSSR **279** (1984), *2* (in Russian).
10. *Brykalov, S. A.*, Some criteria for the existence of solutions of nonlinear boundary value problems. Dokl. Akad. Nauk SSSR **284** (1985), *6* (in Russian).

Существование равновесия в некоторых системах управления нагреванием

С. А. БРЫКАЛОВ

(Свердловск)

Рассматривается существование решений некоторых краевых задач для обыкновенных функционально-дифференциальных уравнений. Полученные результаты могут быть использованы для проверки наличия стационарных состояний в системах управления нагреванием стержня или кольца. Поясним содержательный смысл результатов в одном частном случае. Пусть ось s направлена вдоль стержня, a, b — координаты концов, $x(s)$, $k(s)$ — температура и коэффициент теплопроводности в точке s. Имеется распределенный источник тепла с плотностью $-h(s)x(s)+f(s)$ и три сосредоточенных источника в точках a, b, s_1. На концах a, b поддерживается температура T_1, T_2, в точку s_1 поступает поток тепла q. В каждый момент времени $x(\cdot)\in C^0$ подается на вход управляющего устройства. На выходе получаем параметры $k(\cdot)\in C^1$, $h(\cdot)f(\cdot)\in C^0$, $T_1, T_2, q, s_1 \in \mathbf{R}$, удовлетворяющие неравенствам

$$0<\varepsilon\leqq k\leqq N,\quad \left|\frac{dk}{ds}\right|\leqq N,\quad 0\leqq h\leqq N,\quad |f|\leqq N,\quad |T_i|\leqq N,\quad |q|\leqq N,\quad a\leqq s_1\leqq b,$$

где ε, N — константы. Выход управляющего устройства определяется входом однозначно и зависит от него непрерывно. В общем случае распределение температуры $x(\cdot)$ и другие параметры системы будут меняться со временем. (Изменение координаты s_1 соответствует перемещению одного из точечных источников тепла вдоль стержня.) Но некоторые распределения температуры могут оказаться стационарными. Из полученных результатов следует, что описанная выше система всегда имеет такие положения равновесия.

С. А. Брыкалов
Институт математики и механики
Уральского отделения АН СССР
СССР, 620066, Свердловск,
ул. С. Ковалевской, 16

Problems of Control and Information Theory, Vol. 17 (6), pp. 381–396 (1988)

NEW MULTIPLE ACCESS ALGORITHMS FOR INTEGRATING VOICE AND DATA ON LOCAL AREA NETWORKS

V. GARAI, CS. SZABÓ, T. SZÉP

(Budapest)

(Received October 4, 1987)

One of the major approaches to voice and data integration in local environment is based on local area network technologies, already well established in the field of computer communication networks. This could be a good alternative if, at least, the following two requirements are met: (i) the local network is connected to the public telephone network, and (ii) the integrated system is appropriately compatible with existing standards. The prerequisite for (i) is that the access algorithm (or, as it is frequently called, the medium access protocol) should guarantee appropriately bounded voice packet delay. Compatibility is achieved in a certain sense if the integrated service stations can coexist with the standard ones.

This paper deals with voice and data integration on a standard CSMA/CD bus (on the so-called "Ethernet" system).

Two classes of protocols, namely, the one based on transmitting voice packets in strings, and the other based on transmitting voice packets in individual "time slots" (but without central timing and control) are considered, the first one in a more detailed fashion.

Two new protocol versions are proposed. Voice throughput calculations and the impact of silence detection on voice throughput are given, taking into account finite and different propagation delays. Simulation results are presented to explore the data throughput–delay performance of these protocols.

1. Introduction

One of the two basic solutions for integrated local communication (office automation etc.) is based on the well-established LAN technologies. (The other is the PBX approach.)

When implementing a real-time telephone service on a LAN, and thinking of connecting the LAN to the public telephone network, one faces the PTT requirements, placed upon the delay. Without going into details, let us briefly quote from the corresponding CCITT recommendation, that delays above 300 ms are not acceptable. This value limits the total one-way subscriber-to-subscriber propagation delay. The worst (but not unusual) case is when a satellite channel is used as the international part of the connection; in this case the maximum permissible delay for a LAN is 15 ms for the 300 ms total value. The delay of the LAN consists of two parts: the

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

packetization delay (the time required for collecting voice samples to form a packet) and the transmission delay of the LAN (determined by the access protocol etc.). In packet speech networks, the packetization time is usually chosen from the range of 10 . . . 30 ms, so as a rough estimation, we can say that the delay caused by the protocol itself must be bounded by a value of the order of the packetization time.

One possible solution is to allow packet loss and to establish a tradeoff between delay and loss, and throwing away voice packets after a limited (and small) number of attempts (see e.g. [1]).

Unfortunately, as it turned out from simulation studies, the CSMA/CD delay strongly fluctuates around its mean (see e.g. [2]), and it is difficult to fix exactly the maximum number of conversations.

There is a strong motivation for elaborating specific "integrated" medium access protocols (which will be implemented in specific "integrated" controllers) to ensure a specified delay limit with no packet loss. An additional requirement is also reasonable: the new integrated protocols must be "compatible" with the standard data protocol in the sense that the integrated controllers can coexist with the standard ones on the same bus.

2. Two approaches to integrating periodic and aperiodic traffic in a CSMA/CD compatible way

Basically, there are two approaches to integrating aperiodic (data) and periodic (e.g. voice) traffic in a CSMA/CD compatible way:

(i) *Organizing contiguous strings* of packets from periodic sources which occupy a part (or the total duration) of the period of a periodic source (e.g. the packetization period when digitizing voice sources). The aperiodic sources, following the standard CSMA/CD protocol, will sense carrier during the string and, therefore, will not disturb it.

The basic idea of this method is due to Chlamtac [3], [4]. Our first protocol described in Section 3, is based upon this idea but it is less complex in practical implementation and more efficient.

(ii) *Organizing individual time slots* for periodic packets. The term "time slot" means simply a place anywhere within the packetization period, which — once acquired by a given periodic source, using the corresponding rules of the protocol — is granted for this source periodically (not by a central controller facility, of course, but by means of the appropriate rules of the distributed protocol).

This is the idea proposed by Maxemchuk. He worked out a protocol, based upon this idea, in details and gave a performance analysis [5]. Our second protocol of Section 3 belongs to this class but differs from the aforementioned one in several aspects.

3. New protocol versions

Two types of stations are defined.

The Type "S" station is a source of aperiodic traffic (e.g. interactive data traffic) and is called standard station. The protocol for these stations is the carrier sense multiple access with collision detection (CSMA/CD), as defined in the IEEE standard No. 802.3. Some minor modifications may be necessary, but the details are omitted here.

The Type "I" station is a source of both aperiodic and periodic (e.g. voice) traffic and is called *integrated station.* The protocols are specific for this integrated type of stations and they are called *integrated protocols.*

The integrated protocols are described below according to the following scheme:

(a) packet formats

(b) the rules of the protocol

(c) remarks and references.

3.1. The PROT.1 protocol

This protocol belongs to the family of string-type protocols.

(a) Packet formats

The packet of the string leader is shown in Fig. 1, all other packets are of standard format.

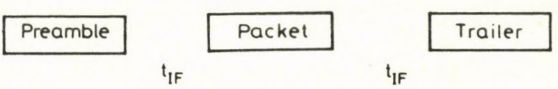


Fig. 1

(b) The rules of the protocol

(i) The system is either in *data mode* or in *speech mode.* The system, after turning on at least two stations, enters the data mode. It switches over to the speech mode after the first voice connection has been built up. The system remains in the speech mode if at least one voice connection exists.

(ii) The protocol uses, in addition to the carrier sense (CS) and collision detect (CD) information, some address information as well. This information can be represented by a table which contains the addresses of off-hook stations.

(iii) The rules to be followed are different for the *leader of the string* and the *members of the string*. A station becomes the leader in any of the following cases:

— The station initiates a string after the system has turned over from the data mode to the speech mode;

— The station takes over the function of the leader from the previously leading station because the voice connection of the latter one has been disconnected.

The task of the leader is to initiate periodically the voice string. The packet format of the leader makes possible detecting the "start of the string" event by other stations.

(iv) Every member of the string prepares for the transmission $t_{IF}+(i-1)\cdot\Delta$ seconds after the next CS→$\overline{\text{CS}}$ transition. Here Δ is an appropriately chosen time unit, and "i" is the position number of the station within the string which corresponds to the physical location of the station on the cable.

The condition $N\cdot\Delta<\tau_s$ must be fulfilled so that the standard stations will not disturb the voice string. (Here N is the maximum number of simultaneous voice channels that are allowed in the system, and τ_s is the standard slot time.) In this case, the standard stations, the linear priority of which is programmed to be greater than or equal to 2, will not disturb the string.

(c) Remarks and references

This protocol differs from the RT-V protocol proposed by Chlamtac [3] in the way the voice strings are formed. According to the RT-V protocol, the "handshaking" between succeeding packets in the voice string is accompished physically, causing intentional collisions between packet trailers and headers.

According to our version, an appropriate timing ensures the correct order of the packets within a string and prevents it from being disturbed by aperiodic packets.

The advantage of the presented version is that its implementation is simpler. Also, this version is more efficient than the original RT-V because of arranging the voice packets within the string according to the physical order of the stations on the bus (see the performance analysis of Section 4).

3.2. The PROT.2 protocol

(a) Packet formats

The format of the first voice packet is the same as that of a standard data packet. The second and succeeding voice packets consist of an extra preamble of length t_{PR} and of a standard packet (the latter is constructed according to the rules

of the standard CSMA/CD protocol). The length of the extra preamble corresponds to the two-way end-to-end propagation delay, and, therefore, is sufficiently long to "kill" all standard data packets transmitted at any place of the cable. The length of the standard part depends on the packetization time and the digitization rate. The two parts are separated by an interframe time t_{IFV} specific for the "I" stations (and less than the standard interframe time, t_{IF}), see Fig. 2.

The packets of "S" stations are standard and there is no limitation to their length.

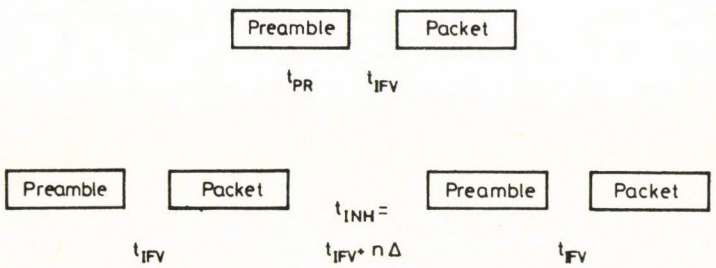


Fig. 2

(b) The rules of the protocol

The integrated medium access (MAC) protocol consists of the rules for the periodical transmission of voice packets belonging to the connections which have been already built up and have not been disconnected.

Silence detection is implemented in the "I" stations. That is, periodic packets are transmitted, in general, only during the active periods of the voice source.

The rules of the protocol are as follows:

(i) A station at the beginning of a new voice connection starts searching for a free "time slot" (for a free place, among the already existing virtual voice channels). To make sure that all existing channels, e.g. all "off-hook" stations are taken into account, the new station sends a broadcast message requesting that all "off-hook" stations send voice packets in the succeeding k cycles, irrespective of the fact that they have active periods or silences. This way, the new station will be able to "see" all existing time slots.

(ii) The next step is an attempt to finding a free place, sending a standard packet according to the rules of the standard protocol. If this attempt is successful within any of the k cycles, the second and all succeeding voice packets will be transmitted according to the rules (iii) and (iv), given below. If not, the new station should ask for a new series of cycles sending a new broadcast packet.

(iii) The second and all succeeding voice packets — which can be called regular packets — are transmitted, in general, in every t_p seconds, where t_p is the packetization

time. (The exceptions are given below.) While transmitting the preamble part of the regular packet, possible collisions will not be taken into account (note that a collision can occur only between a voice packet and a data packet, or, more exactly, the collision between two voice packets is an event with very small probability). Thus, a voice packet will always "kill" data packets, so that the second — information — part of the voice packet will be transmitted undisturbed.

(iv) A station will modify the period (which is in general t_p) in the case when it observes that the distance between the end of its own transmission and the start of its successor's transmission is less than a properly chosen "inhibited interval", t_{INH}. In this case the next transmission will take place in $t_p - \Delta$. (After that, the normal period t_p will again be used.) The practical values of t_{INH} and Δ are chosen taking clock instabilities into account. The above time parameters must satisfy the following conditions:

$$t_{INH} = t_{IFV} + n \cdot \Delta; \qquad t_{IF} > t_{INH}. \tag{3.1}$$

This method for correcting timing instabilities requires that each "off-hook" station sends a packet regularly in every r cycle (r is a design parameter), irrespective of whether it is active or is in silence.

If, after a series of corrections, the time difference reaches the value of the sampling interval, a voice sample will be thrown away. This event takes place very rarely (provided the system parameters are properly chosen) and has no effect on the quality of the reproduced voice.

(v) The periodic (voice) packets of the "I" type stations will be granted higher priority than data packets, in the following sense:

— If, at the time instant, when the next voice packet is to be transmitted, a data packet is being transmitted, or, a data packet is waiting for a new transmission attempt after an unsuccessful one, it will be thrown away;

— The higher protocol layers will be advised that no more data packet can be accepted;

— The voice packet is transmitted according to the above rules, and, after having the transmission completed, the higher layers will be informed that the MAC layer is again able to receive a data packet transmission request.

(c) Remarks and references

The main differences between this protocol and Maxemchuk's (described in [5]) are as follows.

— A preemptive priority is given to the voice packets. (Advantages: there is no need to organize an overflow area in periodic packets for transmitting voice samples collected during delays, and a simpler hardware and software. There is no need to

restrict the length of aperiodic packets, either. (Disadvantage: sensing the channel busy, a periodic packet will always kill the aperiodic packet.)

— Silence periods are utilized and the corresponding protocol rules are worked out.

— Timing instabilities are corrected and the corresponding protocol rules are worked out.

The disadvantage of this version is that the channel is not used efficiently for data when the throughput required by periodic traffic is more than half of the total throughput. (In this region, it is difficult for a new periodic source to find a place among the existing channels.) In our opinion, however, this situation will not occur in a practical integrated service network where voice packets occupy only a fraction of the total channel time which is less than the half of the total capacity.

The exact definitions of the protocols using flow-charts and state diagrams are given in the Appendix.

4. Voice throughput analysis of the string-type protocol

In this section, we shall establish a correspondence between the string length and the number of voice channels for the PROT1 algorithm. As a special case, we shall determine the voice throughput in terms of the maximum number of simultaneous virtual voice channels when the string occupies the whole packetization period.

A similar analysis was carried out in [6] for the RT-V protocols and its modified version.

Figure 3 illustrates how the string is formed in a simple time diagram, and a more exact illustration using the so-called time-space diagram technique is shown in Fig. 4.

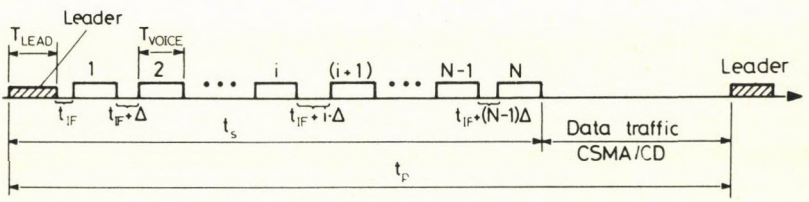


Fig. 3

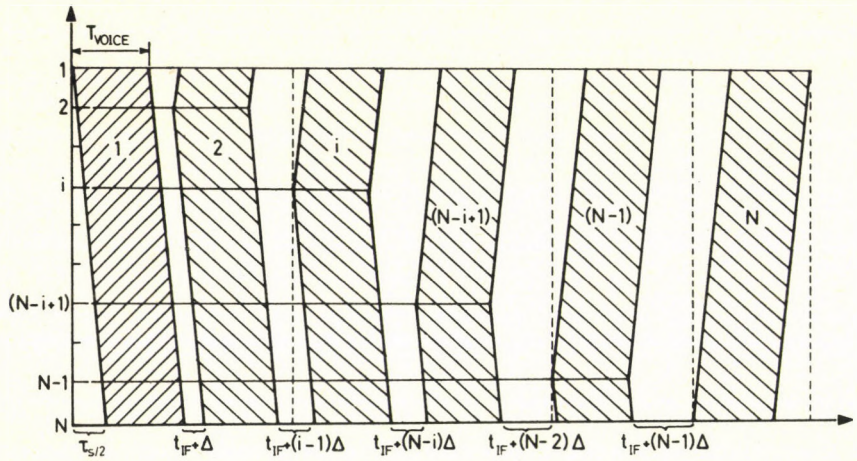


Fig. 4

The string length, t_s, obviously equals:

$$t_s = N \cdot T_{\text{VOICE}} + T_{\text{LEAD}} + N \cdot t_{IF} + \sum_{i=1}^{N} (i-1) \cdot \Delta + \tau_s. \tag{4.1}$$

Using equation (4.1), it is easy to calculate the number of virtual voice channels, N:

$$\frac{\Delta}{2} \cdot N^2 + \left(T_{\text{VOICE}} + t_{IF} - \frac{\Delta}{2} \right) \cdot N + (T_{\text{LEAD}} + \tau_s - t_s) = \emptyset. \tag{4.2}$$

Some numerical results are shown in Table 1 below.

Table 1. The number of virtual voice channels

t_s/t_p	$t_p = 30$ ms	$t_p = 10$ ms
0.2	41	27
0.4	75	49
0.6	104	66
0.8	129	81
1.0	152	95

5. Simulation results

In Section 4, *voice throughput* analysis was carried out for the string-type (type (i) as defined in Section 2) protocol. (The throughput was defined as the maximum number of allowable voice channels for the worst combination of propagation delays between adjacent stations of the string.) A similar calculation can be carried out for the type (ii) protocols.

Voice delay considerations has not been given yet. It is obvious, however, that the voice packet delay, caused by the medium access protocol, is upper bounded by the packetization time in the case of type (i) protocols and this delay is zero for our protocol PROT2 belonging to the class (ii).

The questions related to the *data throughput and delay* performance have not been answered yet. The formulation of the problem itself is simple: how much is the delay of data packets for a given input intensity when only a fraction of the total channel time can be used which is left free by voice packets. The latter are either grouped to form a string thus occupying the first part of the packetization time continuously, or are transmitted separately at random time instants within the packetization time (see the illustration of Fig. 5). However, it seemed to be difficult to answer the above question analytically and simulation studies were carried out. Some characteristic results are summed up below.

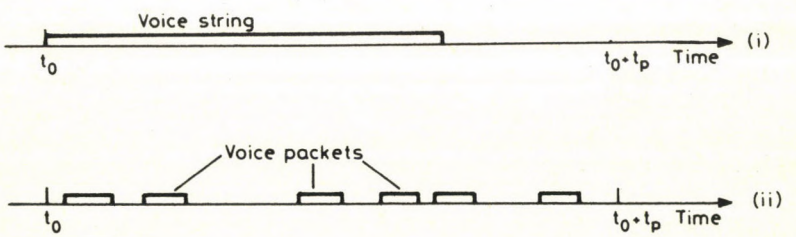


Fig. 5

For the string-type protocol, a simple simulation model was used, where a special station was included which occupied periodically a given part of the packetization time for a given duration, thus modelling the voice string. The average data packet delay versus the data throughput is shown in Fig. 6 for $t_s/t_p = .3, .4, .5,$ and .6.

A disadvantage of this protocol for data stations is obvious: if the latter follow the standard CSMA/CD protocol, then, immediately after the end of the period occupied by the voice string, a conflict occurs almost surely if the string is long enough. This effect can be weakened by introducing different priorities for different groups of data stations. (This prioritizing means the following: if a priority number

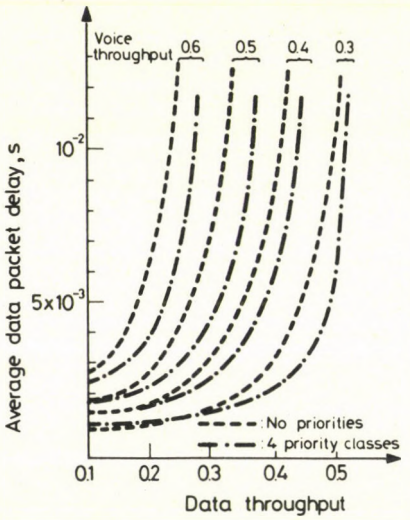


Fig. 6

n is assigned to a station then it will transmit $t_{IF}+n\tau_s$ seconds after the CS→$\overline{CS}$ transition, where, according to our previous notations, t_{IF} the interframe time, the standard value of which is 9.6 μs and τ_s the standard slot time which equals 51.2 μs. The programming capabilities of the Intel 586 Ethernet controller allow to vary this number between 0 — this case corresponds to the Ethernet standard — and 7.) We examined the effect of assigning different priorities on delay-throughput performance of data packets, using four priority classes ($n=0$, 1, 2, 3). A collection of results is shown in Fig. 6.

The average packet delay itself is not a sufficient measure; its variance and higher moments are also of importance. The packet delay can be characterized completely by means of its probability distribution function. Our simulation program provided the histograms of the corresponding density function. Some examples are shown in Fig. 7.

The same performance measures are important for the class (ii) protocols. We investigated the protocol version PROT2 to illustrate the performance of this protocol class, and, at the same time, to forecast the performance of a practical system that can be built implementing this protocol. An example of results is shown in Fig. 8, which illustrates the data and voice throughput versus the number of data sources for a given number of voice channels. The average data packet delay is shown in Fig. 9. The initial voice channel establishing delay is also of importance, it is shown in Fig. 10, again as a function of the number of data sources.

In the above examples, 32 kbps voice and 30 ms packetization time was used.

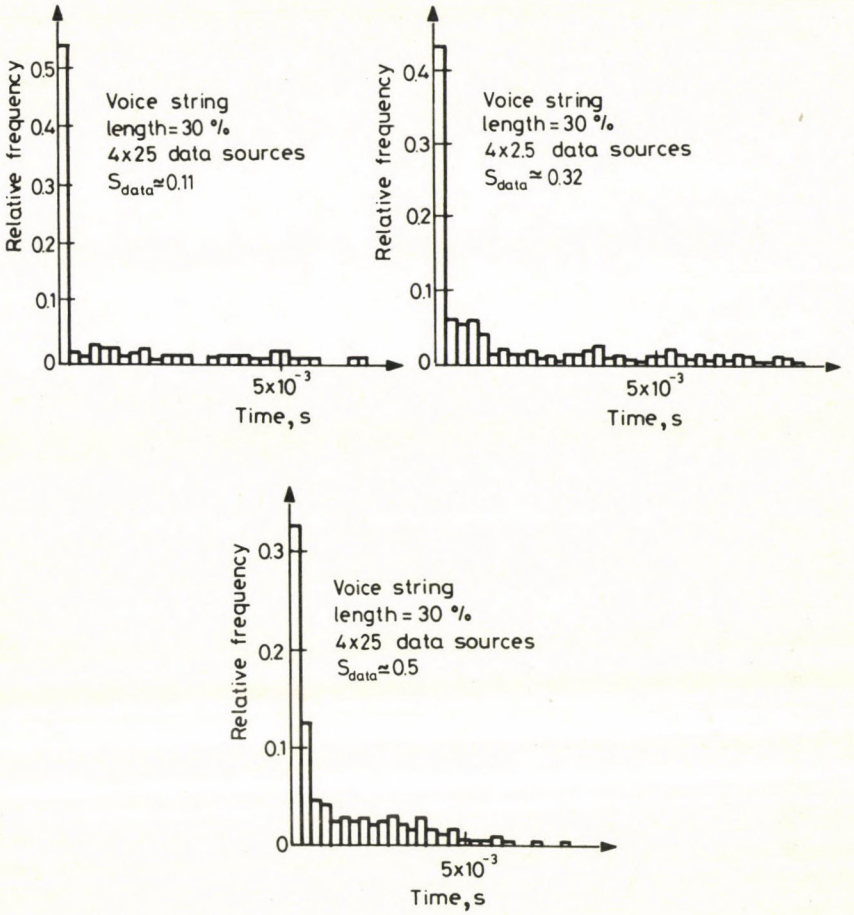

Relative frequency
0.5
0.4
0.3
0.2
0.1
0
Voice string
length = 30 %
4x25 data sources
$S_{data} \simeq 0.11$
$5x10^{-3}$
Time, s
Relative frequency
0.4
0.3
0.2
0.1
0
Voice string
length = 30 %
4x2.5 data sources
$S_{data} \simeq 0.32$
$5x10^{-3}$
Time, s
Relative frequency
0.3
0.2
0.1
0
Voice string
length = 30 %
4x25 data sources
$S_{data} \simeq 0.5$
$5x10^{-3}$
Time, s

Fig. 7

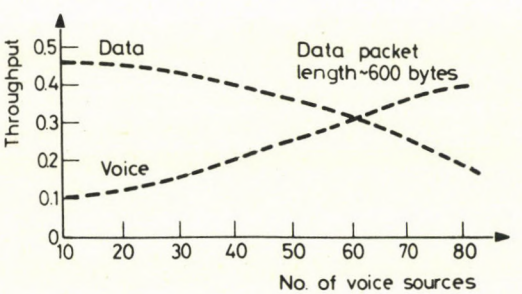

Throughput
0.5
0.4
0.3
0.2
0.1
0
Data
Data packet
length~600 bytes
Voice
10
20
30
40
50
60
70
80
No. of voice sources

Fig. 8

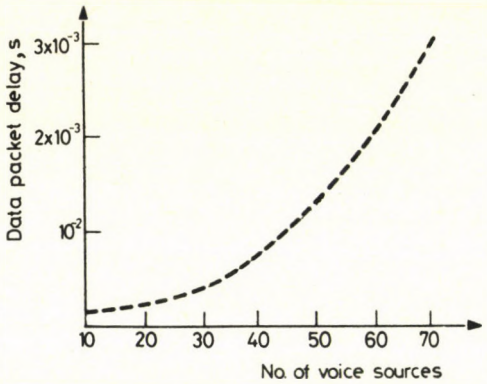


Fig. 9

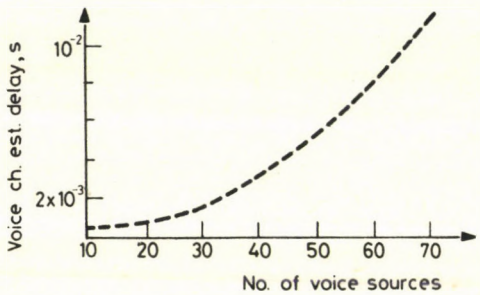


Fig. 10

6. Conclusions

There is no doubt that a LAN-based system consisting of integrated workstations and connected to the PSTN via an appropriate gateway is a possible solution for integrated local voice/data communication.

The work presented in this paper provides a good basis for working out a prototype of an integrated service LAN.

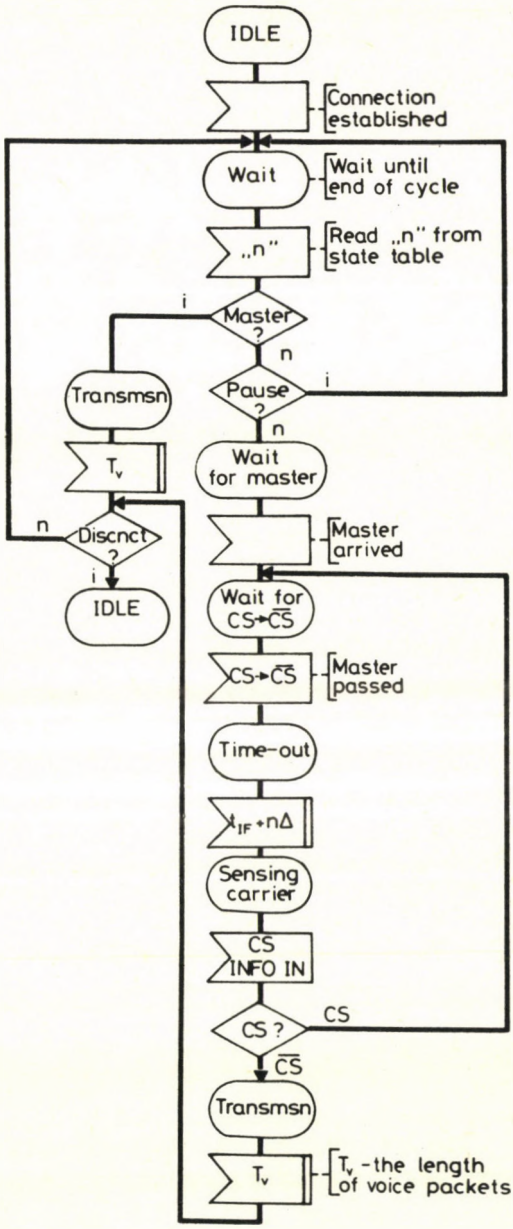

IDLE
Connection established
Wait
Wait until end of cycle
„n"
Read „n" from state table
Master ?
i
n
Pause ?
i
n
Transmsn
T_v
Wait for master
n
Discnct ?
i
IDLE
Master arrived
Wait for CS→$\overline{CS}$
CS→$\overline{CS}$
Master passed
Time-out
$t_{IF}+n\Delta$
Sensing carrier
CS INFO IN
CS ?
CS
$\overline{CS}$
Transmsn
T_v
T_v - the length of voice packets

Fig. A1

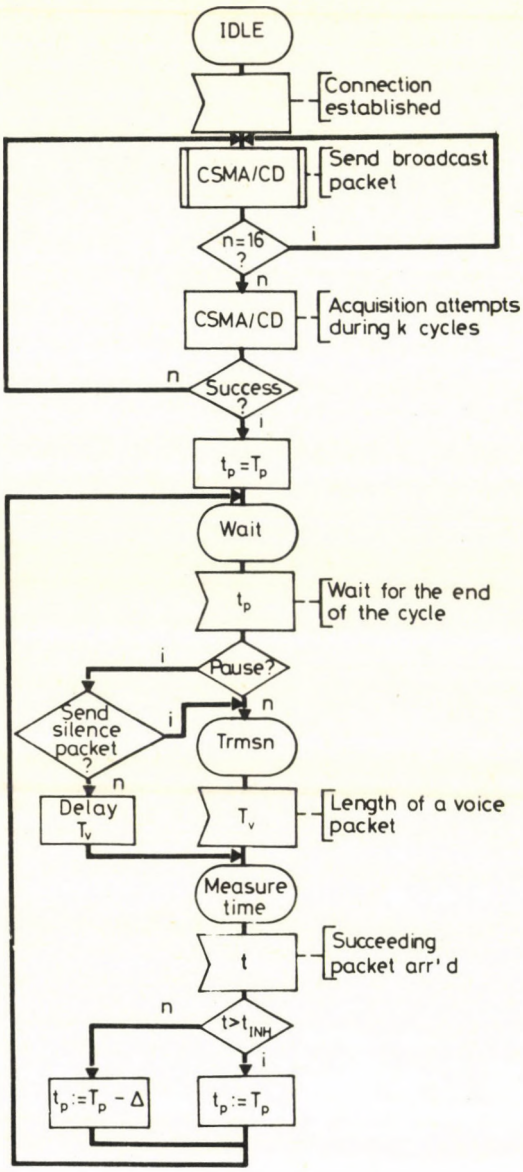

IDLE
Connection established
CSMA/CD
Send broadcast packet
n=16 ?
i
n
CSMA/CD
Acquisition attempts during k cycles
n
Success ?
i
$t_p = T_p$
Wait
t_p
Wait for the end of the cycle
i
Pause?
n
Send silence packet ?
i
Trmsn
n
Delay T_v
T_v
Length of a voice packet
Measure time
t
Succeeding packet arr'd
n
$t > t_{INH}$
i
$t_p := T_p - \Delta$
$t_p := T_p$

Fig. A2

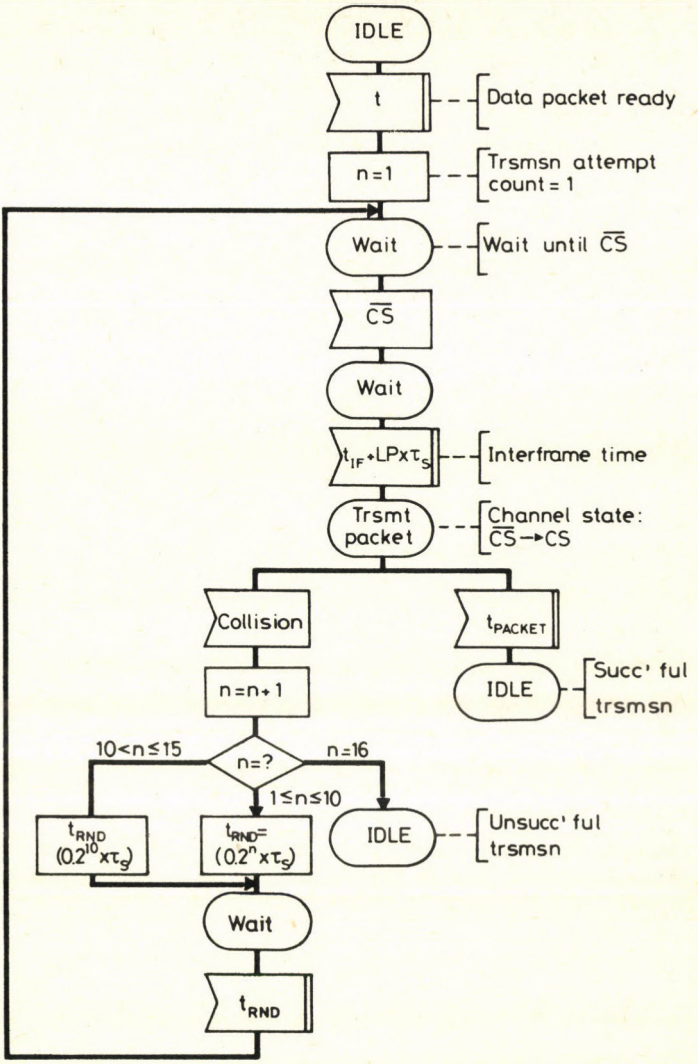

IDLE
t
Data packet ready
n=1
Trsmsn attempt count = 1
Wait
Wait until $\overline{CS}$
$\overline{CS}$
Wait
$t_{IF}+LPx\tau_s$
Interframe time
Trsmt packet
Channel state: $\overline{CS}\rightarrow CS$
Collision
t_{PACKET}
n=n+1
IDLE
Succ' ful trsmsn
10<n≤15
n=?
n=16
1≤n≤10
t_{RND} $(0.2^{10}x\tau_s)$
t_{RND}= $(0.2^{n}x\tau_s)$
IDLE
Unsucc' ful trsmsn
Wait
t_{RND}

Fig. A3

References

1. *Dunlop, J., Rashid, M. A.*, Improving the delay characteristics of the standard Ethernet for speech transmission. Journal IERE, **56,** *5*, pp. 184–186 (May 1986).
2. *DeTreville, J. D.*, A simulation-based comparison of voice transmission on CSMA/CD networks and on token-buses. AT&T Bell Labs Technical Journal, **63,** *1*, pp. 33–35 (1984).
3. *Chlamtac, I.*, An Ethernet-compatible protocol for real-time voice/data integration. Computer Networks and ISDN Systems, *10*, pp. 81–96, 1985.
4. *Chlamtac, I., Eisinger, M.*, Voice and data integration on Ethernet: protocols and performance analysis. Preprint, Comp. Sci. Dept., Technion, Haifa, Israel, 1986.
5. *Maxemchuk, N.*, A variation on CSMA/CD that yields movable TDM slots in integrated voice/data local networks. BSTJ, **61,** *7*, September 1982, pp. 1527–1550.
6. *Fratta, L., Szabó, Cs.*, CSMA/CD compatible protocols to support real-time voice transmission. Proc. MELECON'87, Rome, March 24–27, 1987, 87 CH 2425-7, pp. 203–206.

Appendix

The formal specification of the protocols, described qualitatively in Section 3, is given in Figs A1–A3. The SDL language, recommended by CCITT, turned out to be the most useful means for the symbolic description.

We restricted ourselves to defining the access algorithms of voice and data packets only, i.e. the protocols of connecting and disconnecting two conversating parties are not given here.

According to the PROT1 protocol, the voice packets follow the rules given in Fig. A1. The protocol for data packets is the standard one, with the only modification that a linear priority $LP = 2$ is assigned to data stations (see Fig. A3). The PROT2 protocol is given in Fig. A2, for voice packets. The "CSMA/CD" protocol, used by the first voice packet is shown in Fig. A3. (In this case, $LP = 0$.)

Новые алгоритмы множественного доступа для интегрирования речи и данных в локальных сетях

В. ГАРАИ, Ч. САБО, Т. СЕП

(Будапешт)

Один из подходов к интегрированию речи и данных в локальных сетях базируется на широко применяемых в практике технологий и протоколов локальных вычислительных сетей. В статье рассматриваются возможности передачи речи и данных в стандартной системе »Этернет«.

В статье предложены два новых протокола, для которых вычисляется пропускная способность при передаче речи и рассматривается влияние детекции пауз с учетом конечного времени распространения. Задержка пакетов данных в зависимости от нагрузки определяется путем моделирования на ЭВМ.

Cs. Szabó
Institute of Communication Electronics
Technical University of Budapest
H-1111 Stoczek u. 2, Budapest, Hungary

РУССКИЙ ПЕРЕВОД

Проблемы управления и теории информации, том **17**, *номер 6 (1988)*

О ПРОГРАММНОМ СИНТЕЗЕ ГАРАНТИРУЮЩЕГО УПРАВЛЕНИЯ

Н. Н. Красовский, Т. Н. Решетова

(*Свердловск*)

В статье рассматривается задача о формировании гарантирующего управления динамической системой [1]. Формирование такого управления методом программного стохастического синтеза [1–4] сводится к рекуррентным построениям вогнутых оболочек для вспомогательных функций, которые возникают в программной конструкции. Приводится пример симуляции на ЭВМ процесса управления для модельной задачи.

1. Постановка задачи

Рассмотрим систему, описываемую дифференциальным уравнением

$$\dot{x} = A(t)x + B(t)u + C(t)v, \qquad t_0 \leqq t \leqq \vartheta. \tag{1.1}$$

Здесь $x - n$-мерный фазовый вектор объекта, $u - v$-мерный вектор управления, $v - s$-мерный вектор помехи; $A(t)$, $B(t)$, $C(t)$ — непрерывные матрицы-функции, значения u и v стеснены условиями

$$u \in P, \quad v \in Q, \tag{1.2}$$

где P и Q-компакты.

Задан показатель качества

$$\gamma = |x(\vartheta)| + \int_{t_*}^{\vartheta} \chi(t, u[t], \quad v[t])\,dt, \tag{1.3}$$

где символ $|x|$ обозначает какую-либо норму вектора x, функция $\chi(t, u, v)$ непрерывна.

Согласно [1], задача об оптимальной минимаксной стратегии $u^0(t, x, \varepsilon)$, которая гарантирует оптимальный результат $\rho^0(t_*, x_*)$ для всякой возможной исходной позиции $\{t_*, x_*\}$, имеет решение. Пошаговый закон управления U, который формирует воздействия $u[t] = u^0(t_i, x[t_i], \varepsilon)$, $t_i \leqq t < t_{i+1}$, $i = 1, \ldots, k$, $t_1 = t_*$, $t_{k+1} = \vartheta$, гарантирует неравенство

$$\gamma \leqq \rho^0(t_*, x_*) + \zeta \tag{1.4}$$

Akadémiai Kiadó, Budapest
Pergamon Press, Oxford

при любом наперед выбранном значении $\zeta > 0$, если только параметр $\varepsilon > 0$ и шаг $\delta = \max_i(t_{i+1} - t_i)$ выбраны достаточно малыми, стратегия $u^0(t, x, \varepsilon)$ строится как экстремальная к функции $\rho^0(t, x)$; функция $\rho^0(t, x)$ может вычисляться методом программного стохастического синтеза согласно равенству

$$\rho^0(t, x) = \lim_{\delta \to 0} \sup_{\|l(\cdot)\| \leq 1} [\langle m_0 \cdot X[\vartheta, t]x \rangle + \tag{1.5}$$

$$+ M \left\{ \sum_{i=1}^{k} \int_{\tau_i}^{\tau_{i+1}} \min_{u \in P} \max_{v \in Q} \langle m(\tau_i, \omega) \cdot X[\vartheta, \tau](B(\tau)u + C(\tau)v) \rangle + \chi(\tau, u, v)] d\tau \right\}.$$

Здесь значения τ_i, $i = 1, \ldots, k$ образуют разбиение Δ_δ отрезка $[t_*, \vartheta]$ с шагом $\delta = \max_i(\tau_{i+1} - \tau_i)$; $l(\cdot) = \{l_1(\omega), \ldots, l_n(\omega), \omega \in \Omega\}$ — n-мерная случайная величина на вероятностном пространстве $\{\Omega, F, P\}$, где элементарное событие $\omega = \{\xi_1, \ldots, \xi_k\}$ — набор значений независимых в совокупности случайных величин $0 \leqq \xi_i \leqq 1$, распределенных равномерно и реализующихся в моменты τ_i; символ $\|l(\cdot\|$ обозначает норму

$$\|l(\cdot)\| = \text{vrai sup}_\omega |l(\omega)|^* \tag{1.6}$$

случайной величины $l(\cdot)$, где $|l|^*$ — норма вектора l, сопряженная к норме вектора $|x|$, которая фигурирует в (1.3);

$$m_0 = M\{l(\cdot)\}, \quad m(\tau_i, \omega) = M\{l(\cdot) \mid \xi_1(\omega), \ldots, \xi_i(\omega)\}, \quad \omega \in \Omega, \quad i = 1, \ldots, k,$$

где символы $M\{\ldots\}$ и $M\{\ldots|\ldots\}$ обозначают математическое ожидание и условное математическое ожидание; $X[t, \tau]$ — фундаментальная матрица решений для уравнения $dx/dt = A(t)x$; символ $\langle a \cdot b \rangle$ обозначает скалярное произведение векторов a и b. Равенство (1.5) и участвующие в нем величины пояснены подробно в [1].

Итак, для формирования оптимального по гарантии $\rho^0(t_*, x_*)$ управления $u^0(t_i, x[t_i], \varepsilon)$ достаточно уметь вычислять эффективно для текущих значений $t = t_i$ величины, стоящие под знаком предела в (1.5). В данной статье рассматривается один путь таких вычислений.

2. Рекуррентная оценка

Введем векторы w и z, где

$$w = \{w_1, \ldots, w_n\}, \quad z = \{z_1, \ldots, z_n, z_{n+1}\} = \{w_1, \ldots, w_n, z_{n+1}\} = \{w, z_{n+1}\}.$$

обозначим

$$\psi_i(m) = \int_{\tau_i}^{\tau_{i+1}} \min_{u \in P} \max_{v \in Q} [\langle m \cdot X[\vartheta, \tau](B(\tau)u + C(\tau)v)\rangle + \chi(\tau, u, v)]d\tau, \qquad i = 1, \ldots, k. \tag{2.1}$$

В согласии с методом программного синтеза надлежит вычислять программный экстремум [1]

$$e(t, z, \Delta_\delta) = \sup_{\|u(\cdot)\| \leqq 1} \left[\langle m_0 \cdot X[\vartheta, t]w\rangle + z_{n+1} + M \left\{ \sum_{i=1}^{k} \psi_i(m(\tau_i, \cdot)) \right\} \right]. \tag{2.2}$$

Будем называть верхней вогнутой оболочкой функции $\xi(m)$, $|m|^* \leqq 1$ и обозначать символом $\varphi(m) = \{\zeta(\cdot)\}_*$ функцию, которая удовлетворяет следующим условиям.

1. Функция $\varphi(m)$ вогнута при $|m|^* \leqq 1$, т. е. при всяких $m^{(1)}$, $|m^{(1)}|^* \leqq 1$ и $m^{(2)}$, $|m^{(2)}|^* \leqq 1$ и всяком числе $\lambda \in [0, 1]$ справедливо неравенство

$$\varphi(\lambda m^{(1)} + (1-\lambda)m^{(2)}) \geqq \lambda\varphi(m^{(1)}) + (1-\lambda)\varphi(m^{(2)}). \tag{2.3}$$

2. Справедливо неравенство

$$\zeta(m) \leqq_* \varphi(m) \tag{2.4}$$

при всяком m, $|m|^* \leqq 1$.

3. Пусть $r = \{r_1, \ldots, r_n\}$ — n-мерный вектор. Для всякого значения m, $|m|^* \leqq 1$ найдется вероятностная мера $\mu(R|m)$ на множестве $|r|^* \leqq 1$, такая, что справедливы равенства

$$\varphi(m) = \int_{|r|^* \leqq 1} \zeta(r)\mu(dr|m), \tag{2.5}$$

$$m = \int_{|r|^* \leqq 1} r\mu(dr|m). \tag{2.6}$$

На самом деле меру $\mu(\mathrm{R}|m)$ можно выбрать сосредоточенной в конечном количестве точек $r^{[j]}(m)$. Это облегчает практические построения.

Построим рекуррентную последовательность функций

$$\varphi_k(m) = \{\psi_k(\cdot)\}_*, \tag{2.7}$$

$$\varphi_i(m) = \{\psi_i(\cdot) + \varphi_{i+1}(\cdot)\}_*, \quad |m|^* \leqq 1, \qquad i = 1, \ldots, k-1. \tag{2.8}$$

Справедливо следующее равенство

$$\sup_{\|u(\cdot)\| \leqq 1,\, M\{u(\cdot)\} = m_0} M \left\{ \sum_{i=1}^{k} \psi_i(m(\tau_i, \cdot)) \right\} = M\{\varphi_1(m_0)\} = \varphi_0(m_0). \tag{2.9}$$

Из (2.2), (2.9) получаем равенство

$$e(t, z, \Delta_\delta) = \sup_{|m_0|^* \leq 1} [\langle m_0 \cdot X[\vartheta, t]w\rangle + z_{n+1} + \varphi_1(m_0)]. \tag{2.10}$$

Это равенство вместе с равенством (1.5) показывает, что вычисление величины $e(t, z, \Delta_\delta)$, которая аппроксимирует величину $\varphi^0(t, x)$, сводится к построению последовательности функций $\varphi_j(m)$.

Докажем равенство (2.9).

Зафиксируем некоторое значение m_0, $|m_0|^* \leq 1$. Построим случайную величину $m^{(1)}(\omega) = m^{(1)}[\xi_1]$ следующим образом. Положим, что ее распределение определяется вероятностной мерой $\mu^{(1)}(R|m_0)$, которая, согласно условиям (2.5), (2.6), удовлетворяет равенствам

$$\rho_1(m_0) = \int_{|r|^* \leq 1} [\psi_1(r) + \varphi_2(r)]\mu^{(1)}(dr|m_0), \tag{2.11}$$

$$m_0 = \int_{|r|^* \leq 1} r\mu^{(1)}(dr|m_0). \tag{2.12}$$

Дальнейшее построение случайных величин $m^{(j)}$, $j = 2, \ldots, k$ осуществляется по индукции. Пусть уже построены случайные величины $m^{(j)}[\xi_1, \ldots, \xi], j \leq i$. Случайную величину $m^{(i+1)}[\xi_1, \ldots, \xi_{i+1}]$ определим ее условной вероятностной мерой $\mu^{(i+1)}(R|m^{(i)}[\xi_1, \ldots, \xi_i])$, которая удовлетворяет условиям

$$\varphi_{i+1}(m^{(i)}) = \int_{|r|^* \leq 1} [\psi_{i+1}(r) + \varphi_{i+2}(r)]\mu^{(i+1)}(dr|m^{(i)}), \tag{2.13}$$

$$m^{(i)} = \int_{|r|^* \leq 1} r\mu^{(i+1)}(dr|m^{(i)}). \tag{2.14}$$

Согласно предыдущему, при таком построении не возникает принципиальных затруднений, так как в рассматриваемых нами случаях все используемые меры $\mu^{(j)}(R|m^{(j-1)})$ сосредоточены каждая в конечном количестве точек $r_{[j]}^{[s]}(m)$. Однако конечность множества точек $r_{[j]}^{[s]}(m)$ учитывать в теоретических построениях не обязательно. В самом деле, опираясь известным образом на теоремы об измеримом выборе [5], можно проверить, что и без предположения о конечности множества точек $r_{[j]}^{[s]}(m)$ сосредоточения для мер $\mu^{(i)}$, нужные меры $\mu^{(j)}(r|m^{(j-1)})$ можно выбрать так, что они будут измеримы (должным образом по $m^{(j-1)}$), т. е. эти меры действительно могут играть роль условных вероятностных мер.

Итак, мы можем полагать, что построена последовательность случайных величин m_0, $m^{(i)}[\xi_1, \ldots, \xi_i]$, $i = 1, \ldots, k$, которая удовлетворяет условиям

(2.11)–(2.14). Если положить

$$l(\omega)=l[\xi_1,\ldots,\xi_k]=m^{(k)}[\xi_1,\ldots,\xi_k], \tag{2.15}$$

то из (2.13), (2.14) по индукции вытекают соотношения

$$m^{(i)}[\xi_1,\ldots,\xi_i]=M\{m^{(i+1)}[\xi_1,\ldots,\xi_{i+1}]|m^{(i)}[\xi_1,\ldots,\xi_i]\},$$

$$i=k,k-1,\ldots,1, \tag{2.16}$$

$$m_0=M\{m^{(1)}[\xi_1]\}, \tag{2.17}$$

$$\varphi_i(m^{(i-1)}[\xi_1,\ldots,\xi_{i-1}])=\int_{|r^{(i)}|^*\leqq 1}\psi_i(r^{(i)})\mu(dr^{(i)}|m^{(i-1)})+$$

$$+\int_{|r^{(i+1)}|^*\leqq 1}\int_{|r^{(i)}|^*\leqq 1}\psi_{i+1}(r^{(i+1)})\mu^{(i+1)}(dr^{(i+1)}|m^{(i)})*\mu^{(i)}(dr^{(i)}|m^{(i-1)})+$$

$$+\ldots+\int_{|r^{(k)}|^*\leqq 1}\int_{|r^{(i-1)}|^*\leqq 1}\psi_k(r^{(k)})\mu(dr^{(k)}|m^{(k-1)})\ldots\mu(dr^{(i)}|m^{(i-1)})=$$

$$=M\left\{\sum_{j=i}^{k}\psi_j[\xi_1,\ldots,\xi_j]|m^{(i-1)}[\xi_1,\ldots,\xi_{i-1}]\right\},\qquad i=1,\ldots,k. \tag{2.18}$$

В предыдущих обозначениях из раздела 1 соотношения (2.16)–(2.18) принимают вид

$$m(\tau_i,\omega)=M\{m(\tau_{i+1},\cdot)|m(\tau_i,\omega)\},\qquad i=k,k-1,\ldots,1, \tag{2.19}$$

$$m_0=M\{m(\tau_1,\cdot)\}=M\{l(\cdot)\}, \tag{2.20}$$

$$\varphi_i(m(\tau_{i-1},\omega))=M\left\{\sum_{j=i}^{k}\psi_j(\tau_j,\cdot)|m(\tau_{i-1},\omega)\right\},\qquad i=k,k-1,\ldots,1. \tag{2.21}$$

При $i=1$ имеем равенство

$$\varphi_1(m_0)=M\left\{\sum_{j=1}^{k}\psi_j(m(\tau_j,\cdot))\right\}. \tag{2.22}$$

Таким образом, мы построили случайную величину $l(\omega)$, $\|l(\cdot)\|\leqq 1$, для которой справедливо равенство (2.22), где $m_0=M\{l(\cdot)\}$.

Проверим теперь, что для всякой случайной величины $l(\omega)$, $\|l(\cdot)\|\leqq 1$, которая удовлетворяет условию (2.20), справедливо неравенство

$$M\left\{\sum_{j=1}^{k}\psi_j(m(\tau_j,\cdot))\right\}\leqq\varphi_1(m_0). \tag{2.23}$$

Тогда будет доказано равенство (2.9).

Зафиксируем случайную величину $l(\omega)$ и некоторое значение l, $1 \leqq i \leqq k$. По определению функции $\varphi_i(m)$ вследствие условия (2.4) (где $\zeta(m) = \psi_i(m) + \varphi_{i+1}(m)$, $\varphi(m) = \varphi_i(m)$ справедливо неравенство

$$M\{[\psi_i(m[\tau_i, \xi_1, \ldots, \xi_i]) + \varphi_{i+1}(m[\tau_i, \xi_1, \ldots, \xi_i])]|m[\tau_{i-1}, \xi_1, \ldots, \xi_{i-1}]\} \leqq$$

$$M\{\varphi_i(m[\tau_i, \xi_1, \ldots, \xi_i])|m[\tau_{i-1}, \xi_1, \ldots, \xi_{i-1}]\}. \tag{2.24}$$

В то же время, вследствие вогнутости функции $\varphi_i(m)$, справедливо неравенство

$$M\{\varphi_i(m[\tau_i, \xi_1, \ldots, \xi_i]\}|m[\tau_{i-1}, \ldots, \xi_{i-1}]\} \leqq$$

$$\leqq \varphi_i(m[\tau_{i-1}, \xi_1, \ldots, \xi_{i-1}]). \tag{2.25}$$

Из (2.24), (2.25) следует неравенство

$$M\{[\psi_i(m[\tau_i, \xi_1, \ldots, \xi_i] + \varphi_{i+1}(m[\tau_i, \xi_1, \ldots, \xi_i])]|$$

$$|m[\tau_{i-1}, \xi_1, \ldots, \xi_{i-1}]\} \leqq \varphi_i(m[\tau_{i-1}, \xi_1, \ldots, \xi_{i-1}]). \tag{2.26}$$

Из таких неравенств, верных для $i = 1, \ldots, k$, индукцией по $i = k$ от до $i = 1$ и притом, используя формулы повторных математических ожиданий, подобно тому, как это сделано выше (см. (2.18)), получим нужное нам неравенство (2.23).

Таким образом, равенство (2.9) доказано, стало быть, доказано и равенство (2.10).

3. Построение управления

В соответствии с процедурой, указанной в [1], искомые управляющие воздействия $u[t_i] = u^0(t_i, x[t_i], \varepsilon, \Delta_\delta)$ можно строить, как экстремальные воздействия, вычисляя их из условия

$$\langle m^0[t_i] \cdot B(t_i)u[t_i]\rangle = \min_{u \in P} \langle m^0[t_i] \cdot B(t_i)u\rangle. \tag{3.1}$$

Здесь

$$m^0[t_i] = x[t_i] - w[t_i], \tag{3.2}$$

где $w[t_i]$ — n-мерная компонента сопутствующей точки $z[t_i]$. Эта точка определяется из условий [1] (стр. 207–210). Опираясь на равенства (1.5) и (2.10) и аппроксимируя функцию $\rho^0(t, x)$ функцией $e(t, \{x, 0,\}, \Delta_\delta)$ и функцию $\rho^0(t, x) + z_{n+1}$ функцией $e(t, z, \Delta_\delta)$, условия для определения точки $z[t_i]$ приводим к следующему виду

$$\min_z \max_{|m|^* \leqq 1} (z_{n+1} + \langle m \cdot X(\vartheta, t_i)w\rangle + \varphi_1(m)) =$$

$$\max_{|m|^* \leqq 1} \min_z (z_{n+1} + \langle m \cdot X(\vartheta, t_i)w\rangle + \varphi_1(m)). \tag{3.3}$$

при

$$|z-\{x,0\}|_e \leqq [\varepsilon+\varepsilon(t_i-t_0)]\exp(\mathscr{L}(t_i-t_0)), \tag{3.4}$$

где $\mathscr{L}=\max\limits_{t_0\leqq t\leqq\vartheta}(|A(t)|+1)$, символ $|c|_e$ обозначает евклидову норму вектора c, $|A(t)|=\max\limits_{|x|\leqq 1}|A(t)x|_e$. Перестановка операций минимума и максимума в (3.3) определяется тем обстоятельством, что функция $\varphi_1(m)$ является вогнутой.

Из (3.2), (3.3) следует, что вектор $m^0[t_i]$ определяется равенством

$$m^0[t_i]=[\varepsilon+\varepsilon(t_i-t_0)]\exp(\mathscr{L}(t_i-t_0))X'(\vartheta,t_i]m^0/$$
$$/(|X'(\vartheta,t_i)m^0|_e^2+1)^{1/2}. \tag{3.5}$$

Здесь верхний индекс штрих означает транспонирование.Вектор m^0 является решением следующей задачи на максимум

$$\max_{|m|\leqq 1}[-[\varepsilon+\varepsilon(t_i-t_0)]\exp(\mathscr{L}(t_i-t_0))(|X'(\vartheta,t_i)m|_e^2+1)^{1/2}+$$
$$+\langle m\cdot X(\vartheta,t_i)x[t_i]+\varphi_1(m)]=[-[\varepsilon+\varepsilon(t_i-t_0)]\exp(\mathscr{L}(t_i-t_0)*$$
$$*(|X'(\vartheta,t_i)m^0|_e^2+1)^{1/2}+\langle m^0\cdot X(\vartheta,t_i)x[t_i]\rangle+\varphi_1(m^0)]. \tag{3.6}$$

При этом вектор m^0 является вектором градиентом для функции $\rho^0(t,x,\varepsilon,\Delta_\vartheta)$, которая аппроксимирует функцию $\rho^0(t,x)$. Решение задачи (3.6) является единственным, так как максимизируемая функция вогнута по m.

Таким образом, полагая, что в цепи управления имеется быстродействующая ЭВМ, можно организовать следующий процесс управления. В течение очень малого отрезка времени $t_i\leqq t<t_i+\alpha$ вслед за моментом t_i при известном значении $x[t_i]$ решается задача (3.5), (3.6) и из условия (3.1) вычисляется воздействие $u[t_i]$. Это воздействие в течение времени $t_i+\alpha\leqq t<t_{i+1}+\alpha$ подается на объект. В ряде случаев организация такого процесса управления вполне возможна с использованием доступных ЭВМ.

4. Модельная задача

Пусть управляемый объект представляет из себя материальную точку, движущуюся в трехмерном пространстве под действием центральной силы, силы трения, реактивной управляющей силы и силы нерегулярной помехи. Процесс управления осуществляется в пределах заданного отрезка времен $t_0\leqq t\leqq\vartheta$. Обозначим: $r[t]=\{r_1[t], r_2[t], r_3\{t\}\}$ — радиус-вектор точки; $a[t]$ — масса объекта, меняющаяся по выбираемому закону, непрерывная функция;

$f_1 = \beta(t)r$ — центральная сила; $f_2 = \alpha(t)\dot{r}$ — сила трения, точка над буквой означает дифференцирование по времени; $\alpha(t)$, $\beta(t)$ — заданные функции времени; $f_3 = \dot{a}u$ — реактивная сила, u — вектор относительной скорости реактивной массы; $f_4 = v$ — сила нерегулярной помехи. За показатель качества управления выбрано расстояние в момент ϑ от управляемого объекта до начала координат.

Движение описывается дифференциальным уравнением Мещерского

$$\ddot{r} = (\alpha(t)/a)\dot{r} + (\beta(t)/a)r + (\dot{a}/a)u + v/a, \qquad t_0 \leqq t \leqq \vartheta. \tag{4.1}$$

Примем, что управляющее воздействие $u = \{u_1, u_2, u_3\}$ и помеха $v = \{v_1, v_2, v_3\}$ стеснены условиями

$$u_1^2 + v_1^2 u_2^2 + v_2^2 u_3^2 \leqq 1 \tag{4.2}$$

$$v_1^2 + v_1^2 v_2^2 + v_2^2 v_3^2 \leqq 1. \tag{4.3}$$

Показатель качества

$$\gamma = (r_1^2[\vartheta] + r_2^2[\vartheta] + r_3^2[\vartheta])^{1/2}. \tag{4.4}$$

Введем фазовый вектор

$$y = \{y_1, \ldots, y_6\} = \{r_1, \dot{r}_1, r_2, \dot{r}_2, r_3, \dot{r}_3\}. \tag{4.5}$$

Вектор y удовлетворяет дифференциальному уравнению

$$\dot{y} = F(t)y + G(t)u + H(t)v, \tag{4.6}$$

где матрицы $F(t)$, $G(t)$, $H(t)$ выражаются известным образом через переменные, указанные выше. При этом примем, что масса a объекта меняется по закону

$$a(t) = a_0 \exp(-\lambda[t - t_0]). \tag{4.7}$$

Подходящим преобразованием переменной y в переменную x, подобным тому, которое приведено в [1] (стр. 52), придем к дифференциальному уравнению для трехмерного x:

$$x = -\lambda y_{12}(\vartheta, t)u + \exp(\lambda[t - t_0])y_{12}(\vartheta, t)v/a_0, \tag{4.8}$$

где $y_{12}(\vartheta, t)$ — элемент фундаментальной матрицы $Y(\tau, t)$ для уравнения

$$dy/d\tau = F(\tau)y.$$

Векторы u и v ограничены по-прежнему условиями (4.2), (4.3). Показатель γ имеет вид

$$\gamma = (x_1^2[\vartheta] + x_2^2[\vartheta] + x_3^2[\vartheta])^{1/2} \tag{4.9}$$

Согласно раздела 2 функции $\psi_i(m)$ имеют вид

$$\psi_i(m) = \int_{\tau_i}^{\tau_{i+1}} \min_{u\in P} \max_{v\in Q} \langle m \cdot (-\lambda y_{12}(\vartheta, \tau)u + \\ + \exp(\lambda[\tau - t_0])y_{12}(\vartheta, \tau)v/a_0)\rangle d\tau, \tag{4.10}$$

где $m = \{m_1, m_2, m_3\}$.

Примем для определенности, что $\nu_1 > 1$ и $\nu_2 > 1$. Построение верхних вогнутых оболочек $\varphi_i(m)$ зависит от знака выражения

$$\eta(t) = \exp(\lambda(t - t_0))/(\lambda a_0) - 1. \tag{4.11}$$

Включим корень τ_* уравнения $\eta(t) = 0$ в число точек τ_i разбиения Δ_δ. Тогда возможны два случая.

1. В случае $\eta(\tau) < 0$ при $\tau_i < \tau < \tau_{i+1}$ функция $\psi_i(m)$ является вогнутой функцией от m. Тогда

$$\varphi_i(m) = \psi_i(m) + \varphi_{i+1}(m). \tag{4.12}$$

2. В случае $\eta(\tau) > 0$ при $\tau_i < \tau < \tau_{i+1}$ функция $\psi_i(m)$ вогнутой функцией от m не является. Тогда

$$\varphi_i(m) = \{\psi_i(m)\}_* + \varphi_{i+1}(m). \tag{4.13}$$

Учитывая монотонность функции $\eta(t)$ и соотношения (4.12), (4.13), придем к следующему результату

$$\varphi_1(m) = \lambda \int_{\tau_1}^{\tau_*} |y_{12}(\vartheta, \tau)|\eta(\tau)d\tau \cdot (m_1^2 + m_2^2/\nu_1^2 + m_3^2/\nu_2^2)^{1/2} + \\ + \lambda \int_{\tau_*}^{\tau_{k+1}} |y_{12}(\vartheta, \tau)|\eta(\tau)d\tau \cdot (1 + (1/\nu_1^2 - 1)m_2^2 + (1/\nu_2^2 - 1)m_3^2)^{1/2}. \tag{4.14}$$

Таким образом, вектор $m^0 = \{m_1^0, m_2^0, m_3^0\}$, определяющий вектор $m^0[t_i]$, который определяет оптимальное управление $u^0[t_i]$, является решением следующей задачи

$$\max_{|m|^* \leq 1} \{-[\varepsilon + \varepsilon(t_i - t_0)](|m|^2 + 1)^{1/2} + \langle m \cdot x[t_i]\rangle + \varphi_1(m)\} = \\ = \{-[\varepsilon + \varepsilon(t_i - t_0)](|m^0|^2 + 1)^{1/2} + \langle m^0 \cdot x[t_i]\rangle + \varphi_1(m^0)\}. \tag{4.15}$$

Вектор $m^0[t_i]$ принимает вид

$$m^0[t_i] = [\varepsilon + \varepsilon(t_i - t_0)]m^0/(|m^0|^2 + 1)^{1/2}. \tag{4.16}$$

Управляющее воздействие $u^0[t_i]$ определяется теперь формулами

$$u_1^0[t_i] = \operatorname{sign}(y_{12}(\vartheta, t_i))m_1^0[t_i]/((m_1^0[t_i])^2 + (m_2^0[t_i])^2/v_1^2 + \\ + (m_3^0[t_i])^2 v_2^2)^{1/2},$$

$$u_2^0[t_i] = \operatorname{sign}(y_{12}(\vartheta, t_i))m_2^0[t_i]/((m_1^0[t_i])^2 + (m_2^0[t_i])^2/v_1^2 + \\ + (m_3^0[t_i])^2 v_2^2)^{1/2})/v_1^2,$$

$$u_3^0[t_i] = \operatorname{sign}(y_{12}(\vartheta, t_i))m_3^0[t_i]/((m_1^0[t_i])^2 + (m_2^0[t_i])^2/v_1^2 + \\ + (m_3^0[t_i])^2 v_2^2)^{1/2}/v_2^2. \tag{4.17}$$

Процесс управления для рассматриваемого объекта был симулирован на ЭВМ для следующих значений параметров

$$t_0 = 0, \quad \vartheta = 4, \quad y(t_0) = \{0, 1, 1, 0, 1, 0\},$$

$$a_0 = 10, \quad \lambda = 0{,}6,$$

$$\alpha(t) = -4\exp(-0{,}6t), \quad \beta(t) = 0{,}5\exp(-0{,}6t)$$

$$v_1 = 1{,}225, \quad v_2 = 1{,}265 \tag{4.18}$$

Рассмотрены три ситуации.

В первом случае оптимальное управление работает в паре с самой неблагоприятной помехой v, которая конструировалась в соответствии с процедурами, описанными в [1]. На рис. 1 движение в этом случае представлено сплошной линией. Во втором случае оптимальная стратегия управления работает в паре с помехой вида

$$v_1[t] = \sin(10t)\cos(10t), \qquad v_2[t] = (\sin^2(10t))/v_1,$$

$$v_3[t] = (\cos(10t))/v_2, \tag{4.19}$$

которая не является оптимальной. Соответствующее движение показано на рис. 1 пунктирной линией. В третьем случае была выбрана неоптимальная стратегия $u(t, r)$. Она для каждого момента t назначает вектор $u[t]$:

$$u_1[t] = -r_1[t]/(r_1^2[t] + v_1^2 r_2^2[t] + v_2^2 r_3^2[t]),$$

$$u_2[t] = -r_2[t]/(r_1^2[t] + v_1^2 r_2^2[t] + v_2^2 r_3^2[t]),$$

$$u_3[t] = -r_3[t]/(r_1^2[t] + v_1^2 r_2^2[t] + v_2^2 r_3^2[t]) \tag{4.20}$$

— коллинеарный текущему радиусу-вектору $r[t]$. Помеха формировалась при этом самым неблагоприятным для нас образом. На рис. 1 этому случаю соответствует штрих-пунктирная кривая. В эксперименте получились три зна-

чения для расстояния от точки до начала координат в момент $\vartheta = 4$: $\gamma_1 = 1{,}3593$, $\gamma_2 = 0{,}5199$, $\gamma_3 = 1{,}9864$. Как и следовало ожидать, значения γ_1, γ_2, γ_3 удовлетворяют неравенствам

$$\gamma_2 < \gamma_1 < \gamma_3. \tag{4.21}$$

Для взятой исходной позиции $\{t_0, x_0\}$ оптимальный гарантированный результат есть $\rho^0(t_0, x_0) = 1{,}3587$. Как и следовало ожидать, в эксперименте получилось значение γ_1, близкое к числу $\rho^0(t_0, x_0)$:

$$\gamma_1 \approx \rho^0(t_0, x_0). \tag{4.22}$$

Литература

1. *Красовский Н. Н.* Управление динамической системой. М., Наука, 1985, 520 с.
2. *Красовский Н. Н., Третьяков В. Е.* Стохастический программный синтез одного гарантирующего управления. Проблемы управления и теории информации, 1983, **12***(2)*, 70–95.
3. *Красовский А. Н.* Построение смешанных стратегий на основе стохастических программ. Прикл. мат. и мех., **51,** *2* (1987), 186–192.
4. *Третьяков В. Е.* Синтез гарантирующего управления. Проблемы управления и теории информации, (в печати).
5. *Аркин В. И., Левин В. Л.* Выпуклость значений векторных интегралов, теоремы измеримого выбора и вариационные задачи. Успехи матем. наук. **27,** *3* (1972).

PRINTED IN HUNGARY

Akadémiai Kiadó és Nyomda Vállalat, Budapest

NOTE TO CONTRIBUTORS

Two copies of the *manuscript* (each complete with figures, tables and references) are to be sent to

E.D. TERYAEV coordinating editor
Department of Mechanics and Control Processes
Academy of Sciences of the USSR
Leninsky Prospect 14, Moscow V-71, USSR

or to V. STREJC
UTIA ČSAV
182 08 Prague 8
Pod vodárenskou věží 4, Czechoslovakia

or to L. GYÖRFI
Technical University of Budapest
H-1111 Budapest, Stoczek u. 2, Hungary

Authors are requested to retain a third copy of the submitted typescript to be able to check the proofs.

The papers, preferably in English or Russian, should be typed double spaced on one side of good-quality paper with wide margins (4–5 cm). The first page of the paper should carry the title, the author(s)' names and the name of the town where they are active. The name and address of the author to whom the proofs should be sent should be given at the end of the paper. An *abstract* should head the paper. English papers should also have a Russian abstract.

The papers should not exceed 15 pages (25 × 50 characters per page) including tables and references. The proper location of the tables and figures must be indicated on the margin.

Mathematical notations should follow up-to-date usage. Equations longer than half a line should not be incorporated in the text. In-text equations must be typed on a single line except that one level of subscripting and/or superscripting is permissible. Use / instead of horizontal bars. Displayed equations should be written so as to require the fewest possible lines. Therefore use "exp" for the exponential function whenever the exponent requires more than a single line. Matrices should, if possible, not be written in full. Use subscript notations instead such as $A = ||a_{ij}||$. Write diagonal matrices as diag $(d_1, d_2, \ldots d_n)$.

The authors will be sent galley proofs to be returned by next mail. Rejected manuscripts will be returned. Authors will receive 100 reprints free of charge. Additional reprints may be ordered.

К СВЕДЕНИЮ АВТОРОВ

Рукописи статей в трех экземплярах на русском языке и в трех на английском следует направлять по адресу: 117312 Москва В-312, просп. 60 летия Октября, 9, МНИИПУ. Редакция журнала «Проблемы управления и теории информации» (зав. редакцией Н. И. Родионова).

Объём статьи не должен превышать 15 печатных страниц (25 строк по 50 букв). Статье должна предшествовать аннотация объемом 50–100 слов и приложено резюме–реферат объемом не менее 10–15% объема статьи на русском языке в трех экземплярах, на котором напечатан служебный адрес автора (фамилия, название учреждения, адрес).

При написании статьи авторам надо строго придерживаться следующей формы: введение (постановка задачи), основное содержание, примеры практического использования, обсуждение результатов, выводы и литература.

Статьи должны быть отпечатаны с промежутком в два интервала, последовательность таблиц и рисунков должна быть отмечена на полях. Математические обозначения рекомендуется давать в соответствии с современными требованиями и традициями. Разметку букв следует производить только во втором экземпляре и русского, и английского варианта статьи.

Авторам высылается верстка, которую необходимо незамедлительно проверить и возвратить в редакцию.

После публикации авторам высылаются бесплатно 100 оттисков их статей.

Рукописи непринятых статей возвращаются авторам.

CONTENTS · СОДЕРЖАНИЕ

HU ISSN 0370-2529 Index 26 660