VOL. 1 / TOM 1 * NUMBER 1 / НОМЕР 1

ACADEMY OF SCIENCES OF THE USSR
HUNGARIAN ACADEMY OF SCIENCES

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

АКАДЕМИЯ НАУК СССР
АКАДЕМИЯ НАУК ВЕНГРИИ

1972

AKADÉMIAI KIADÓ, BUDAPEST

PROBLEMS OF CONTROL AND INFORMATION THEORY

is an international quarterly sponsored jointly by the Presidium of the Academy of Sciences of the U. S. S. R. and of the Hungarian Academy of Sciences. It offers publicity for original papers and short communications on the following topics:

- general theory of control systems and system theory
- theory of automata
- information theory
- operation research; theory of complex systems
- theory of economic control; system modelling
- theory and methods of adaptation, learning, identification and pattern recognition
- methods of information processing; application of digital computers in control and communication systems
- new physical principles in constructing technical devices for automation, control and information processing

The four issues published per year make up a volume of some 320 pp.

While this quarterly is mainly a publication forum of the research results achieved within the U. S. S. R. and Hungary, also papers of international interest from other countries are welcome.

Distributor:

KULTURA

Hungarian Trading Co. for Books and Newspapers

Budapest 62, P.O. Box 149, Hungary

The quarterly is published by

AKADÉMIAI KIADÓ

Publishing House of the Hungarian Academy of Sciences
Budapest 502, P.O. BOX 24, Hungary

Subscription price per volume: $16.00 DM 64.— £6.75

Президиумами Академии Наук СССР и Академии Наук ВНР было принято решение о совместном издании журнала

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

Журнал создан для обеспечения быстрейшей публикации материалов о новейших результатах научных исследований, разработок и кратких сообщений о достижениях в следующих областях науки:

- общая теория процессов управления и теории систем;
- теория автоматов
- теория информации
- исследование операций; теория сложных систем
- теория управления экономическими системами; модели систем
- теория и методы адаптации, обучения, идентификации и распознавания образов
- методы обработки информации; применение вычислительных машин в системах управления и передачи информации
- новые физические принципы создания технических средств автоматики, управления и передачи информации

Журнал издается четыре раза в год, общим объёмом около 320 печатных страниц.

Журнал в первую очередь будет публиковать научные достижения обеих стран, а также статьи из других стран, представляющие международный интерес.

Распространитель:

KULTURA

Венгерское Общество по распространению книг и журналов

Будапешт 62, п. о. 149, Венгрия

Издатель журнала:

AKADÉMIAI KIADÓ

Издательство Академии Наук Венгрии

Будапешт 502, п. о. 24, Венгрия

Подписная цена: $16.00 за том

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

AKADÉMIAI KIADÓ

PUBLISHING HOUSE OF THE HUNGARIAN ACADEMY OF SCIENCES

BUDAPEST

Printed in Hungary
ACADEMY PRESS, BUDAPEST

PROBLEMS OF CONTROL AND INFORMATION THEORY

Volume 1, 1972

PAPERS

AUTHORS

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

Том 1, 1972

СТАТЬИ

АВТОРЫ

TO THE READER

This is the first issue of *Problems of Control and Information Theory,* a new quarterly edited jointly by the Academy of Sciences of the U. S. S. R. and of the Hungarian Academy of Sciences.

The main purpose of this journal is to endeavour exchange of new ideas in the field of control processes and information theory.

Original contributions, survey papers and short reports on new advances in these very fields are to be published.

It is hoped that this journal will be of use to a large international community of the relevant specialists. It is kindly asked to send any critical comments or suggestions to the Editorial Board.

B. N. PETROV, F. CSÁKI
EDITORS

УВАЖАЕМЫЕ ЧИТАТЕЛИ!

Мы начинаем выпуск нового международного советско-венгерского журнала *Проблемы управления и теории информации*.

Основная задача нашего журнала — обеспечить быстрый обмен информацией о новых результатах научных исследований и разработок в области процессов управления и теории информации.

В журнале будут публиковаться оригинальные научные статьи, статьи обзорного характера по важнейшим проблемам и краткие сообщения о новых достижениях в данных областях науки и техники.

Мы надеемся, что журнал будет полезен широкому кругу мировой научной общественности и будем благодарны за Ваши пожелания и критические замечания в адрес журнала.

Б. Н. ПЕТРОВ, Ф. ЧАКИ.
РЕДАКТОРЫ ЖУРНАЛА

Problems of Control and Information Theory, Vol. 1 (1), pp. 5–28 (1972)

ТЕОРЕТИЧЕСКИЕ ПРОБЛЕМЫ ПРАКТИЧЕСКОГО ПРИЛОЖЕНИЯ ТЕОРИИ КОНЕЧНЫХ АВТОМАТОВ

М. А. ГАВРИЛОВ

Москва

(Поступила в редакцию 23 февраля 1971 г.)

В статье описывается состояние и проблемы, возникающие в теории конечных автоматов в связи с практическими требованиями, возникающими при синтезе структур реальных дискретных управляющих устройств. Рассматриваются проблемы создания языка для описания условий работы релейных устройств, близкого к естественному, а также проблемы структурного и абстрактного синтеза устройств большой размерности.

Вопрос о практическом приложении теории конечных автоматов имеет две стороны: одна из них заключается в рассмотрении того, что требуется от теории конечных автоматов в этом отношении, а вторая — в том, что эта теория может в настоящее время дать.

Настоящая статья посвящена рассмотрению обеих этих сторон и рассмотрению возникающих при этом проблем, относящихся к самой теории конечных автоматов.

Рассмотрим первоначально, что в принципе хотел бы получить от теории конечных автоматов проектировщик дискретных устройств. Можно сформулировать это, вероятно, в следующем виде:

а) методы синтеза структуры дискретных устройств должны быть приспособлены к вводу условий их работы на языке, возможно более близком к естественному языку, обычно употребляемому в проектных разработках;

б) они должны учитывать все ограничения, присущие реальным релейным элементам, и ограничения, связанные с монтажем их в реальных условиях;

в) они должны указывать проектировщику на ошибки в отношении полноты и противоречивости сформулированных условий и позволять легко корректировать последние;

г) они должны давать возможность учитывать дополнительные требования к структуре устройства в отношении ее быстродействия, надежности, ремонтопригодности и т. п;

д) необходимо, чтобы методы синтеза давали структуры дискретных устройств, близкие к оптимальным, и позволяли проектировщику получать

альтернативные решения как в отношении параметров получаемых структур, так и в отношении условий работы синтезируемых дискретных устройств;

е) в предлагаемых методах синтеза должна быть предусмотрена возможность анализа получаемых результатов;

ж) размерность решаемых задач должны составлять: 200—300 входных переменных, 200—300 выходов, 300—400 внутренних состояний (для комбинационных структур 1000—1500 интервалов).

Несколько лет тому назад эти требования рассматривались бы как утопические. Я постараюсь показать, что сейчас они являются реальными, хотя и требуют существенного развития теории конечных автоматов и существенного изменения самой методологии синтеза.

1. Языки записи условий работы дискретных устройств

В настоящее время существует достаточно большое количество языков для записи условий работы дискретных устройств. К ним можно отнести: языки таблиц переходов и таблиц состояний, графы переходов, язык логических схем алгоритмов (ЛСА), язык блочного синтеза, язык описания алгоритмических структур вычислительных машин (АЛОС) и т. д.

Однако подавляющее число из них имеют в своей основе некоторые математические модели дискретных устройств и слабо приспособлены для решения практических задач. Достоинством их является определенная формализация. Однако именно это и служит причиной трудности их практического использования и недостаточной размерности задач, для которых они могут применяться.

Классическим примером этому может служить язык таблиц переходов. Он достаточно исчерпывающе теоретически изучен и гарантирует, уже в силу характера записи, полноту и непротиворечивость условий. Для него разработаны формализованные методы преобразований условия, включая минимизацию записи («сжатие» таблиц переходов). Для него разработаны методы кодирования с учетом устранения недопустимых состязаний, обеспечения заданной безотказности и т. д.

Однако в его чистом виде, даже при использовании УВМ, он пригоден для решения задач размерностью до 10—11 входных переменных в связи с громоздкостью самой записи и громоздкостью преобразований и требует опыта и определенных интуитивных приемов при записи условий работы даже наиболее простых дискретных устройств с последовательностями «конкретного типа».[1]

[1] Для последовательностных дискретных устройств, с точки зрения их методов синтеза, полезно различать следующие их классы:

Поэтому возникает необходимость в разработке так называемых «первичных» языков записи условий работы дискретных устройств. Требования к таким языкам могут быть сформулированы следующим образом:

1. Близость языка к естественному. (Под языком естественного типа подразумевается язык, буквенный состав которого соответствует словарному составу, употребляемому в языках, применяемых конструкторскими и проектными организациями при разработке проектов дискретных устройств, а также наиболее часто встречающимся фрагментам этого языка).

2. Учет особенностей различных классов дискретных управляющих устройств, а именно:

а) особенностей записи условий и операций синтеза для комбинационных и последовательностных устройств, в том числе для одношаговых и многошаговых;

б) особенностей записи условий работы дискретных устройств и операций синтеза с заранее заданными исполнительными элементами, выполняющими также функции элементов памяти;

в) особенности записи условий работы и операций синтеза для дискретных устройств с потенциальными и импульсными воздействиями на входах и выходах.

3. Язык должен содержать операции по определению полноты и непротиворечивости записи условий работы или должен быть предусмотрен интерпретатор, осуществляющий эти операции на другом, более формализованном языке.

4. Желательно иметь в составе языка операции, которые позволяли бы упрощать выражения[1] языка или проводить также операции с помощью интерпретатора в другом, более формализованном языке.

5. Желательно иметь в составе языка операции, позволяющие получать альтернативы технических условий, в том числе при изменении характера воздействий по обратным связям от объекта.

Естественными требованиями являются: простота записи, не требующая от проектировщиков специальных математическах знаний и затраты большого времени.

а) Устройства с последовательностями «конкретного» типа, для которых можно перечислить все заданные последовательности смены состояний на входах устройства и соответствующие им последовательности смены состояний на выходах и *б)* устройства с последовательностями «общего» типа, для которых можно перечислить лишь признаки, характеризующие некоторый класс смены состояний на входах устройства и соответствующие им смены состояний на выходах.

[1] Понятие «упрощать выражение» определяется в зависимости от целей преобразования: в ряде случаев это определяется необходимостью уменьшения числа символов в выражении, в других случаях это может быть необходимость приспособления выражения к структурным особенностям релейных элементов или блоков, на которых предполагается реализовать устройство.

Близость к естественному языку естественно уменьшает формализацию языка и увеличивает требования к процедурам обнаружения противоречивости условий и обнаружения неполноты их.

Как показывают исследования, проведенные в Институте проблем управления, словарный состав естественного языка записи условий работы дискретных устройств немногочислен и содержит достаточно простые понятия, хотя и должен включать для избежания большого перебора символы, характеризующие некоторые общие фрагменты. Примером таких фрагментов может служить достаточно важный фрагмент: «все остальное», вводимый для обеспечения полноты условий и служащий для выражения всех элементов, не входящих в сформулированные условия, например для выражения совокупности последовательностей воздействий на входах или совокупности состояний, которые не могут иметь место, или которые приводят к определенным одинаковым воздействиям на выходах. К таким фрагментам относятся также выражения: «возврат в исходное состояние», «счет числа циклов» и т. п.

Запись указанных выражений в виде перечня последовательностей воздействий и переходов или состояний естественно невозможна. Поэтому необходима разработка некоторого нового исчисления, оперирующего в экономной форме с комплексами последовательностей, состояний и т. п., в числе операций которого имелась бы операция «определение дополнения», предусматривающая например определение «всего остального» как разницы между «всем возможным» и «всем заданным».

Начало такому рассмотрению было положено в работах [1] и [2]. Первая из них касается операции между некоторыми категориями подмножеств состояний рабочей (на выходе должна быть единица) и нерабочей (на выходе должен быть нуль) частями таблицы состояний. Вторая рассматривает таблицы переходов и в ней задача получения «дополнения» сводится к декомпозиции автоматов.

Очень часто из трех категорий состояний: рабочих, нерабочих и условных (на выходе может быть нуль или единица) проектировщик знает только одну категорию состояний, большей частью только категорию рабочих состояний. Определение остальных состояний, что бывает необходимым для получения оптимальных или близких к оптимальным структур, требует также применения при записи условий символа «все остальное», а при определении содержания этого символа — применения операции «дополнения».

Определение полноты и непротиворечивости сформулированных условий работы дискретных устройств является весьма громоздкой и требующей большого времени процедурой, даже при использовании УВМ. Следует полагать, что действенность этого этапа синтеза для задач большой размерности может

быть обеспечена только при разработке эффективных методов общения человека с вычислительной машиной, т. е. при разработке некоторых человеко-машинных процедур. Можно представить себе это в виде разработки некоторой машинной программы, при которой условия работы дискретного устройства вводятся поэтапно (в пределе каждое условие в отдельности), а вычислительная машина, в которую введены также сведения о синтезируемом классе дискретных устройств, сведения об известных неиспользуемых состояниях и т. п., выдает проектировщику данные о необходимой корректировке условий с точки зрения их полноты и непротиворечивости. Такая работа проектировщика с машиной должна продолжаться до тех пор, пока все известные проектировщику условия не будут введены и машина не подтвердит их полноту и непротиворечивость.

Разработка соответствующих методов и машинных программ явится, по мнению автора доклада, существенным шагом на пути внедрения научно обоснованных методов синтеза дискретных устройств в практику.

Существенной особенностью формулировки условий работы дискретных устройств в практических задачах, является их укрупненное представление, а именно в виде разбиения общих условий на блоки и представления входных воздействий и воздействий между блоками в интервальной форме, т. е. в виде произведений переменных или скобочных форм.

При больших размерностях задач эта форма представления является, вероятно, единственно возможной.

В виде блоков, иногда достаточно сложных, могут задаваться также и элементы, на которых необходимо реализовать структуру дискретного устройства.

Эти обстоятельства выдвигают задачу создания языков для описания и преобразования блочных структур, а также оперирования при синтезе не с состояниями переменных, а с интервалами или скобочными формами, что также является блочным представлением условий работы дискретных устройств.

В настоящее время в СССР и в других странах имеется ряд работ, посвященных рассмотрению задач блочного синтеза. Укажем, например, на язык логических схем алгоритмов ЛСА, предложенный А. А. Ляпуновым и развитый применительно к задачам теории дискретных устройств в Институте Проблем передачи информации АН СССР, специальный язык для описаний алгоритмических структур вычислительных машин и устройств (АЛОС), разработанный в Институте кибернетики АН УССР, язык блочного синтеза, разработанный в Институте проблем управления (автоматики и телемеханики) и др.

Рассмотрим проблему преобразования блочных структур дискретных устройств на примере языка блочного синтеза (рис. 1).

Алфавит этого языка составляют:

а) блоки $q_1, q_2, \ldots, q_n$, условия работы которых записаны на каком-либо формализованном языке;

б) внешние входы и выходы блочной структуры, к которым относятся входы и выходы блоков, не связанные с входами и выходами других блоков,

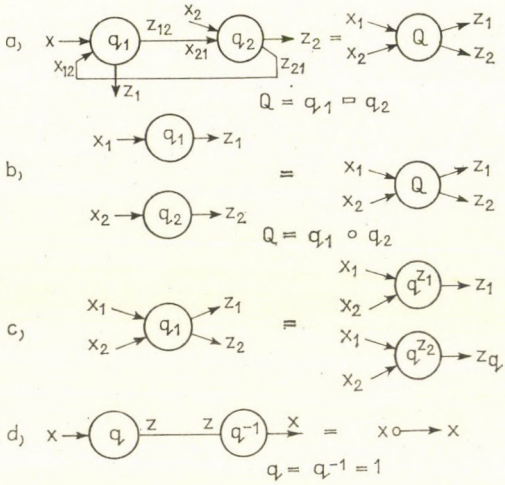


Рис. 1

входящих в данную структуру (x_i и z_i на рис. 1, где индекс i означает номер блока, к которому принадлежит вход и выход) и в) внутренние входы и выходы, к которым принадлежат входы и выходы блоков, связанные с входами и выходами других блоков, входящих в данную структуру (x_{ij} и z_{ij} на рис. 1, где первый индекс означает номер блока, к которому принадлежит данный вход или выход, а второй индекс — номер блока, от которого или к которому он идет).

Операциями языка блочного синтеза являются следующие:

а) «умножение» — замена последовательно включенных блоков одним эквивалентным (рис. 1а) — обозначается символом $\sqsupset$

б) «объединение» — замена параллельно включенных блоков одним эквивалентным (рис. 1б) — обозначается символом $\circ$

в) «разделение» (рис 1в) — разделение блока по внутренним выходам (рис. 1в) — обозначается символом блока с верхним индексом выхода, по которому произведено разделение;

г) «обращение» — взаимная замена в таблице переходов символов входов и выходов с соответствующим преобразованием таблицы переходов (рис. 1г) — обозначается верхним индексом — 1. Умножение первоначального и обращенного блока дает прямую связь между выходами и входами блочной структуры, содержащей эти блоки.

На рис. 2 дан пример преобразования блочной структуры, состоящей из трех блоков, к одному эквивалентному с помощью приведенных выше операций и даны формулы преобразований.

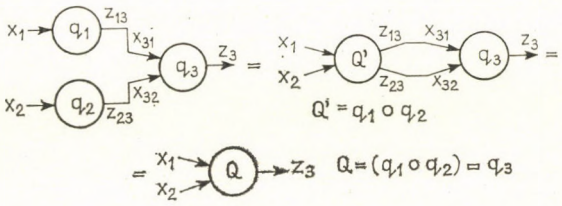


Рис. 2

В [2, 3, 4, 5] показано, что с помощью указанных операций любая блочная структура может быть проведена к одному эквивалентному блоку или разбита на заданное число блоков. Разработаны процедуры определения условий работы эквивалентного блока при композиции блоков на основе условий работы первоначальных блоков, а также условия работы блоков, получаемых при декомпозиции блоков. Получены оценки числа элементов памяти и сложности реализации логической части блочной структуры, позволяющие судить о сравнительной сложности реализации, различных блочных структур.

Рассмотрим теперь вопрос о получении альтернатив условий работы дискретных управляющих устройств при наличии заданных условий работы дискретной системы в целом и условий работы управляемого объекта. В принципе, если задан в виде автоматного отображения алгоритм функционирования дискретной системы и управляющее устройство имеет дискретный характер, т. е. дискретными являются как воздействия, получаемые им извне и по обратным связям от управляемого объекта, так и воздействия, оказываемые им на этот объект по управляющим связям, то управляемый объект, вне зависимости от его физической природы, можно представить также, как дискретное устройство. Однако тогда задачу получения всех альтернатив условий работы управляющего устройства можно представить себе, как задачу выделения из дискретного устройства, представляющего собой всю систему в целом, дискретного устройства, представляющего собой управляемый объект. Остатком этого выделения будет являться, очевидно, управляющее устройство во всех его модификациях.

Такое выделение представляет собой, по существу, операцию «дополнения», о которой упоминалось выше. Оно может быть представлено также как задача декомпозиции автоматов (рис. 3).

Действительно, пусть задана дискретная система ДС, состоящая из управляющего устройства УУ и управляемого объекта УО (рис. 3). Выделим управляемый объект из дискретной системы. Для этого применим операцию

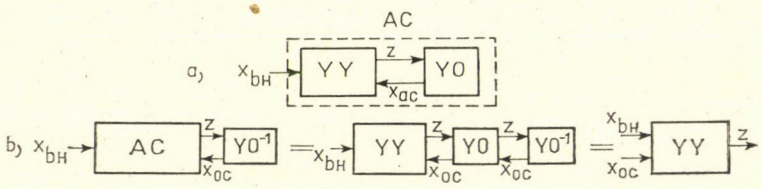


Рис. 3

обращения к объекту УО и операцию блочного умножения к устройству ДС и обращенному объекту $УО^{-1}$, (рис. 3б). Произведя затем операцию блочного умножения с УО и $УО^{-1}$ получим в остатке управляющее устройство УУ. Операция обращения блоков, как показано в [2], дает недетерминированный автомат. Поэтому управляющее устройство, полученное в результате описанной процедуры, будет содержать все возможные условия управляющего устройства работы в формализованном виде (если они существуют), которые удовлетворяют условиям, заданным для дискретной системы ДС. Следует указать, что описанный метод получения альтернатив условий работы управляющих устройств, дает решение этой задачи лишь в принципе. Для практического применения этого метода необходимо разработать процедуры получения из множества возможных альтернатив, которое может быть весьма большим, подмножества, дающие наиболее простую реализацию на заданном наборе релейных элементов или удовлетворяющие каким-либо другим условиям.

Нужно отметить, что указанная процедура дает также возможность нахождения наиболее оптимального взаимодействия между управляющим устройством и управляемым объектом, поскольку в ряде случаев проектировщик может воздействовать на структуру самого объекта (во всяком случае в отношении характера его обратных воздействий на управляющее устройство) с целью получения наиболее оптимальной реализации всей дискретной системы в целом. Вероятно решение этой задачи будет наиболее целесообразным также с помощью разработки человеко-машинных процедур, подобных описанной выше.

2. Проблемы структурного синтеза

Структурный синтез остается пока непременным этапом синтеза структуры многотактных (последовательностных) релейных устройств и останется основным этапом синтеза однотактных (комбинационных) релейных устройств.

Требования, предъявляемые к методам структурного синтеза со стороны практических задач, могут быть сформулированы следующим образом:

а) практическая пригодность при задании условий работы релейного устройства в полной или интервальной форме, для полностью определенных и недоопределенных условий и при размерности задач: до 200—300 входных переменных, 200—300 выходов и до 1000—1500 заданных интервалов;

б) пригодность для широкого класса элементов в том числе избыточных наборов элементов[2] и произвольных элементов,[3] при учете реальных ограничений: на число входов элементов, на число разветвлений, на характер соединений элементов друг с другом, на глубину структуры (число уровней) и т. п.;

в) наличие процедур для направленного поиска реализаций, близких к оптимальным, с учетом структурных свойств элементов;

г) наличие процедур, учитывающих в случае необходимости ограничения, связанные с подачей на входы структуры лишь неинверсных значений переменных и предусматривающих получение для этого случая структур, близких к оптимальным, с учетом числа инверторов на входах;

д) наличие процедур, предусматривающих в случае необходимости устранение состязаний между входными переменными и обеспечение заданной безотказности структуры при отказах логических элементов;

е) наличие процедур, обеспечивающих получение неизбыточных реализаций структур.[4]

Существующие методы синтеза, основанные на классических работах Квайна и Маккласски (определение так называемых «минимальных членов» и

[2] При избыточности в наборе элементов сверх числа, необходимого для обеспечения функциональной полноты, возникает дополнительная задача выбора на каждом этапе синтеза из множеств всех элементов элемента, обеспечивающего наиболее оптимальную реализацию структуры.

[3] Примером произвольного элемента может служить элемент, для которого рабочие и нерабочие состояния (или интервалы) определяются с помощью генератора случайных чисел. Пригодность метода синтеза для таких элементов делает его в принципе пригодным для любых элементов.

[4] Одним из возможных определений неизбыточной структуры является следующее: многовыходная структура типа связного дерева неизбыточна, если недопустимы (приводят к противоречивой реализации) следующие преобразования: *а)* замена переменной на каком-либо входе какого-либо элемента на константу, *б)* непосредственное соединение выхода какого-либо внутреннего элемента структуры с каким-либо выходом всей структуры в целом и *в)* непосредственное соединение выхода какого-либо внутреннего элемента структуры с каким-либо входом какого-либо другого внутреннего элемента структуры.

«минимальных покрытий»), приспособлены, как правило, к реализации структур на простейших логических элементах. Они дают в результате процедуры синтеза описание структур в нормальной форме булевых функций, не учитывают реальных ограничений элементов, слабо приспособлены к реализации на элементах с усложненными структурными свойствами и характеризуются резким возрастанием числа необходимых операций с ростом размерности решаемых задач. Получение абсолютно минимальной реализации требует в этих методах полного перебора решений. Применение различных эвристических процедур для получения приближенных решений, хотя и дает возможность существенного расширения размерности решаемых задач, но, как правило, не устраняет недостатков, связанных с неучетом реальных ограничений элементов и их структурных особенностей.

Более эффективными являются развитые в последние годы методы, основанные на так называемой «функциональной декомпозиции». Общая постановка задачи состоит здесь в следующем.

Заданы условия работы многовыходного недоопределенного релейного устройства в виде набора k функций от n переменных: $F_{n,k}=\{F_1, F_2, \ldots, F_k\}$. Каждая из функций $F_i\,(1 \leq i \leq k)$ задана наборами конституентов (или интервалов) для рабочих и нерабочих состояний: $M_1^i = \{\alpha_1^i\, \alpha_2^i, \ldots, \alpha_m^i\}$ и $M_0^i = \{\beta_1^i, \beta_2^i, \ldots, \beta_p^i\}$, где $\alpha_1^i\, \alpha_2^i, \ldots, \alpha_m^i$ — конституенты (или интервалы), дающие на i-м выходе единицу и $\beta_1^i, \beta_2^i, \ldots, \beta_p^i$ — конституенты (или интервалы), дающие на i-м выходе нуль. Задан функционально полный или избыточный набор элементов, на которых должна быть реализована структура $\Phi = \{\varphi_1, \varphi_2, \ldots, \varphi_t\}$. Каждый элемент φ_j задан наборами состояний его входов: $\pi_1^j = \{\zeta_1^j, \zeta_2^j, \ldots, \zeta_r^j\}$ и $\pi_0^j = \{\eta_1^j, \eta_2^j, \ldots, \eta_s^j\}$, где $\zeta_1^j, \zeta_2^j, \ldots \zeta_r^j$ — состояния входов, дающие на выходе элемента единицу и $\eta_1^j, \eta_2^j, \ldots, \eta_s^j$ — состояния входов, дающие на выходе элемента нуль.

Требуется построить неизбыточную структуру, реализующую заданную функцию $F(n,k)$ на заданном наборе элементов Φ, близкую к оптимальной по числу элементов с учетом присущих им ограничений, ограничений по реализации структуры и т. п.

Общая идея решения этой задачи заключается в том, чтобы представить функцию $F(n, k)$ в виде некоторой композиции элементов из набора Φ, т. е. представить ее зависящей не от первичных переменных $x_1, x_2, \ldots, x_n$, а от некоторых подфункций, реализуемых выбранными элементами из набора Φ, путем соответствующего выбора как самих элементов, так и переменных, подаваемых на них, обеспечивающего оптимальную или близкую к оптимальной реализацию.

Требования неизбыточности реализации и близости ее к оптимальной существенно осложняют эту задачу. Укажем, например, что в случае реализа-

ции многовыходной структуры для получения реализации, близкой к оптимальной, нужно уметь находить:

а) оптимальную последовательность реализации выходных функций структуры;

б) для каждого i-го выбранного выхода – оптимальный выходной элемент;

в) для каждого выбранного выходного элемента – оптимальный набор переменных или функций, подаваемых на входы этого элемента и обеспечивающих оптимальную реализацию всей структуры в целом.

Для указанных выше выборов применяются различные критерии, основанные на различных эвристических соображениях. Основной гипотезой при этом, используемой в подавляющем числе случаев, является гипотеза о том, что реализация структуры в целом получается тем проще, чем «ближе» функция, реализуемая на данном выходе структуры (или части ее), выбранным выходным элементом. С учетом максимального достижения этой близости выбираются переменные или функции, подаваемые на входы этого элемента.

Различные предложенные методы можно разделить на две основные группы, отличающиеся друг от друга методами определения наиболее «близких» функций, реализуемых на выходном элементе.

В первой из них рассматриваются или так называемые «декомпозиционные таблицы [6], дающие перечень функций, на которые можно разложить заданную функцию безотносительно к структурным данным заданных элементов, или все возможные функции, которые можно получить с помощью подачи на входы заданных элементов всех возможных комбинаций входных переменных и констант. Из множества полученных при этом функций выбирается с помощью соответствующих критериев наиболее близкая к заданной для данного выхода [7, 8]. Эти методы, основанные на *выборе* подходящей функции, требуют естественно полного перебора, что, в связи с чрезвычайно быстрым ростом всего множества функций при росте числа входных переменных, резко ограничивает размерность решаемых задач.

Во второй группе методов функция, наиболее близкая к заданной, определяется с помощью *конструирования* ее путем определения на основе соответствующих критериев набора переменных и констант, подаваемых на входы выходного элемента. Перебор приходится здесь применять только в пределах числа заданных элементов при выборе наиболее оптимального из них.

По мнению автора метод конструирования оптимальных выходных функций позволяет наиболее полно удовлетворить сформулированным выше условиям для структурного синтеза [9]. Рассмотрим коротко его идейную сторону.

Пусть условия работы релейного устройства заданы в виде так называемой «обобщенной таблицы состояний», в которой каждому состоянию на

входах, заданному в полной или интервальной форме, сопоставлено состояние выходов (рис. 4). Условия работы каждого из выходов могут быть представлены при этом в виде некоторого множества определенных состояний $M = M_1 \cup M_0$.

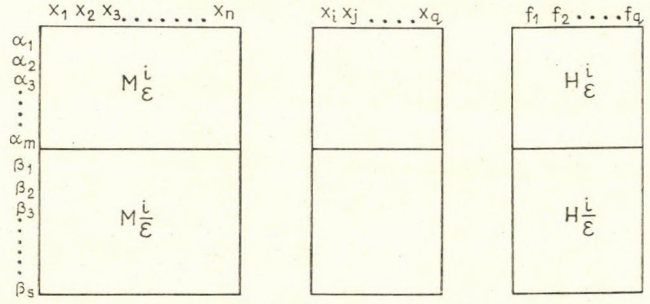
Так как реализация заданной функции может быть осуществлена как по множеству состояний M_1, так и по множеству состояний M_0, то для общности будем обозначать реализуемую сторону таблицы состояний для каждого i-го выхода через M_ε^i, а нереализуемую — через $M_{\bar{\varepsilon}}^i$.

$x_1\ x_2\ x_3\ x_4 \ldots\ x_n$	$z_1\ z_2\ z_3 \ldots\ z_k$
1 0 1 – ⋯⋯ 0	0 1 – ⋯ 1
– – 0 1 ⋯⋯ 1	1 – 0 ⋯ 0
⋯⋯⋯⋯⋯⋯	⋯⋯⋯⋯
⋯⋯⋯⋯⋯⋯	⋯⋯⋯⋯

Рис. 4

Очевидно, что реализация на выходном элементе φ_i функции F_i, заданной этими множествами, будет решена, если на его входы будет подан такой набор входных переменных и констант или функций от них или и тех и других одновременно, что для каждого из состояний, принадлежащих подмножеству M_ε^i на выходе этого элемента будет появляться воздействие, равное ε, а для каждого из состояний, принадлежащего к подмножеству $M_{\bar{\varepsilon}}^i$ — воздействие, равное $\bar{\varepsilon}$.

Обозначим совокупность значений переменных или функций, поданных на входы элемента φ_i, соответствующую какому-либо состоянию из M_ε^i, через h_ε^i и соответствующую какому-либо состоянию из $M_{\bar{\varepsilon}}^i$ — через $h_{\bar{\varepsilon}}^i$. Назовем «приведенной таблицей состояний» функции F_i по элементу φ_i таблицу, содержащую множество $H_\varepsilon^i = \{h_{\varepsilon,1}^i, h_{\bar{\varepsilon},2}^i, \ldots, h_{\varepsilon,r}^i\}$ строк, соответствующих всем состояниям из M_ε и множество $H_{\bar{\varepsilon}}^i = \{h_{\bar{\varepsilon},1}^i, h_{\bar{\varepsilon},2}^i, \ldots, h_{\varepsilon,s}^i\}$ строк, соответствующих всем состояниям из $M_{\bar{\varepsilon}}^i$. Эта таблица будет очевидно, содержать M строк и $q(\varphi_i)$ столбцов, где $q(\varphi_i)$ — число входов элемента φ_i (рис. 5в).



а) Рис. 5 б) в)

Будем характеризовать элементы следующими параметрами:

а) «характеристическим числом» ($[\tau_\varphi]$), представляющим собой минимальное число входов, подача на которые определенных воздействий дает на выходе элемента также определенное воздействие независимо от воздействий, подаваемых на остальные входы;

б) «индексом вхождения» (γ) – значением воздействий, подача которых на число входов элемента, равное $[\tau_\varphi]$, дает определенное значение воздействия на его выходе;

в) «индексом реализации» (ε) – значением воздействия на выходе элемента, которое имеет место при подаче его $[\tau_\varphi]$ входов воздействий, равных γ.[5]

Назовем рангом r состояния из множества $M = M_\varepsilon \cup M_{\bar{\varepsilon}}$ число переменных, принимающих в этом состоянии значение γ. Сформулированное выше условие реализации на выходном элементе заданной множеством M^i функции F^i будет, очевидно, выполнено, если для всех состояний из множества $M^i_{\bar{\varepsilon}}$ ранг $r \leq [\tau_\varphi]-1$ и для всех множеств из состояний M^i_ε ранг $r \geq [\tau_\varphi]$.

Первой операцией алгоритма является определение последовательности реализации выходов структуры. С помощью специального критерия, дающего оценку сложности реализации структуры по мощности множеств M^i_ε и $M^i_{\bar{\varepsilon}}$, выходы структуры располагаются в порядке возрастания сложности их реализации и, в первую очередь, реализуется выход, имеющий наименьшую оценку сложности реализации.

Второй операцией является определение оптимального набора переменных и констант, подаваемых на входы выходного элемента и построение так называемой переходной «таблицы» (рис. 5б). Выбор таких переменных производится с помощью критериев, определяющих для каждой переменной близость ее неинверсного или инверсного значения к заданной функции. Если выбор неинверсного или инверсного значения безразличен (на входах структуры могут быть получены как те, так и другие), то оценка переменных производится по наибольшему значению из критериев: s_γ — для неинверсного значения и $s_{\bar{\gamma}}$ для инверсного, $s = \max\{s_\gamma, s_{\bar{\gamma}}\}$. Если ставится задача получения структуры с наименьшим числом инверторов на входах, что выбор переменных производится по критерию s_γ[6]. Структурные свойства элементов учитываются тем, что определение критериев s_γ и $s_{\bar{\gamma}}$ для переменных после выбора каждой переменной производится с учетом только тех состояний, для которых ранги, полученные в результате выбора предыдущих пере-

[5] Указанные параметры характеризуют элементы с симметричными входами. Однако при некотором расширении их они могут применяться и для элементов с несимметричными входами.

[6] Применение алгоритма показывает, что это дает в некоторых случаях до 20—25% экономии элементов с учетом числа необходимых инверторов.

менных $r_\varepsilon < [\tau_\varphi]$ и $r_{\bar{\varepsilon}} + l \geqq [\tau_\varphi]$, где l – число незаполненных столбцов переходной таблицы.

Критерии s_γ и $s_{\bar{\gamma}}$ имеют следующий вид:

$$S_\gamma = \frac{n_\varepsilon^\gamma \cdot n_{\bar{\varepsilon}}^{\bar{\gamma}}}{n_\varepsilon^\gamma + n_{\bar{\varepsilon}}^{\bar{\gamma}}} \quad \text{и} \quad S_{\bar{\gamma}} = \frac{n_\varepsilon^{\bar{\gamma}} \cdot n_{\bar{\varepsilon}}^{\gamma}}{n_\varepsilon^{\bar{\gamma}} + n_{\bar{\varepsilon}}^{\gamma}}.$$

Здесь $n_\varepsilon^\gamma(n_\varepsilon^{\bar{\gamma}})$ — число значений $\gamma(\bar{\gamma})$ переменной в состояниях M_ε, соответствующих состояниям переходной таблицы, для которых ранг $r < [\tau_\varphi]$, и $n_{\bar{\varepsilon}}^\gamma(n_{\bar{\varepsilon}}^{\bar{\gamma}})$ — число значений $\gamma(\bar{\gamma})$ переменной в состояниях $M_{\bar{\varepsilon}}$, соответствующих состояниям переходной таблицы, для которых $r_{\bar{\varepsilon}} + l \geq [\tau_\varphi]$.

Если после заполнения столбцов переходной таблицы значениями выбранных переменных окажется, что указанные выше условия реализации заданной функции выполнены, то это будет означать, что реализация ее может быть выполнена на одном выходном элементе, поскольку переходная таблица будет совпадать с приведенной. Если же окажется, что для некоторых состояний из M_ε ранг $r < [\tau_\varphi]$ и для некоторых состояний из $M_{\bar{\varepsilon}}$ ранг $r \geq [\tau_\varphi]$, то это будет означать, что некоторые или все переменные должны быть заменены функциями, устраняющими противоречивую реализацию заданной функции, т. е. для реализации ее помимо выходного элемента требуются также и другие.

Переходные таблицы составляются для всех элементов из набора, имеющегося в распоряжении проектировщика. Критерием для выбора из них наиболее оптимального является число состояний, которые реализуются противоречиво, т. е. для которых $r_\varepsilon < [\tau_\varphi]$ и $\tau_{\bar{\varepsilon}} \geq [\tau_\varphi]$. Выбирается элемент, для которого это число является наименьшим.

Реализация заданной функции на выходном элементе на основе характеристического числа $[\tau_\varphi]$ позволяет легко выполнить требование обеспечения заданной безотказности структуры. Пусть, например, задано, что структура устройства должна обеспечивать точное выполнение заданной функции при одновременном отказе типа $\bar{\gamma} \to \gamma$ для d_γ элементов и типа $\gamma \to \bar{\gamma}$ для $d_{\bar{\gamma}}$ элементов. Это требование будет, очевидно, выполнено, если выбрать при реализации стркутуры ранги $r_\varepsilon = [\tau_\varphi] + d_{\bar{\gamma}}$ и $r_{\bar{\varepsilon}} = [\tau_\varphi] - 1 - d_{\bar{\gamma}}$. В том случае, если это не может быть обеспечено полностью на выходном элементе, то обеспечение их переносится на элементы, расположенные в следующих уровнях.

Если условия реализации заданной функции не выполняются на выходном элементе, то следующей операцией после выбора оптимального выходного элемента является доопределение выбранных переменных на входах выходного элемента до функций, непротиворечиво реализующих заданную

функцию, т. е. позволяющих перейти от переходной таблицы рис. 5б к приведенной таблице 5в.

Переходная таблица содержит все сведения, необходимые для определения этих функций. Очевидно, что для некоторой переменной x^i таблица состояний должна содержать в подмножестве M_ε^i все состояния, для которых $r \leq [\tau_\varphi]$ и в подмножестве $M_{\bar{\varepsilon}}^i$ все состояния, для которых $r \geq [\tau_\varphi]-1$.

Реализация доопределяющих функций может рассматриваться как реализация новых заданных функций, т. е. по отношению к ним применяются все перечисленные выше операции. Однако при их реализации возникает весьма важная практическая задача такого доопределения переменных, поданных на входы выходного элемента, при котором глубина структуры была бы примерно одинакова по всем входам.

Возможность этого заключается в том, что при доопределении переменных до некоторых функций f существует много альтернатив построения таблиц состояний последних. Действительно, пусть для какого-либо состояния из M_ε ранг $r_\varepsilon = [\tau_\varphi]-1$. Это состояние будет очевидно реализовано, если на каком-либо (любом) из входов доопределяющая функция будет такова, что в этом состоянии значение переменной $\bar{\gamma}$ будет заменено на γ. Это дает возможность построения некоторого адаптивного процесса, позволяющего получить решение указанной выше задачи. Один из вариантов этого процесса состоит в том, что доопределяющая функция, выбранная первой для реализации, реализуется на одном элементе. Реализация всех состояний, для которых при этом ранги r_ε и $r_{\bar{\varepsilon}}$ получаются не соответствующими условиям реализации, переносятся на доопределяющую функцию следующего выхода путем расширения ее подмножеств M_ε и $M_{\bar{\varepsilon}}$. Такой процесс переноса нереализованных состояний продолжается, если это необходимо, до доопределяющей функции последнего выхода. Если эта функция окажется такой, что она не может быть реализована на одном элементе, то осуществляется попытка реализации ее на следующем уровне. Если такая организация второго уровня на последнем входе входного элемента также не приведет к полной реализации заданной функции, то переходят к организации второго уровня на предпоследнем входе выходного элемента и т. д., пока не будет заполнен второй уровень для всех входов. Затем переходят, если это окажется необходимым, к организации третьего уровня и т. д.

После полного построения структуры первого выхода многовыходной структуры выходы всех составляющих ее элементов принимаются за новые переменные и их значения вводятся в первоначальную таблицу состояний многовыходной функции $F_{(n,k)}$. Это делается для того, чтобы при реализации других выходов последней выходы элементов уже построенной структуры,

если это будет выгодно, могли быть использованы для реализации функций других выходов путем введения связности между ними. Таким же путем реализуется связность и при реализации доопределяющих функций какого-либо одного выхода.

Описанный алгоритм был запрограммирован на машине М-220 для предельной размерности задач на 200 входных переменных и 200 выходов и набора элементов: «И», «ИЛИ», «НЕ», «ИЛИ-НЕ», «И-НЕ» и мажоритарные. Для задач достаточно большой размерности при быстродействии машины в 10 000 операций в секунду алгоритм требует в среднем 15—20 *сек* на элемент структуры. Статистическая проверка алгоритма на потоке задач, образуемом с помощью генератора псевдослучайных чисел, показала, что при наборе элементов «И», «ИЛИ», «НЕ» точно в 50% случаев на выходе структуры выбирался элемент «И» и в 50% — элемент «ИЛИ», а при наборе элементов «И», «ИЛИ», «НЕ», «3-х входовой мажоритарный с отрицанием» и «5-ти входовой мажоритарный с отрицанием» во всех 100% на выходе структуры выбирался 5-ти входовой мажоритарный элемент, а на втором уровне трехвходовой или элемент «ИЛИ». Это полностью соответствует статистической природе потока задач.[7]

Несмотря на достаточно большую размерность задач, решаемых с помощью рассмотренного выше алгоритма, он имеет ограничения, связанные с самой формой задания многовыходной функции $F_{(n,k)}$. Запись условий работы дискретного автомата в виде таблицы рис. 4 является недостаточно экономной, т. к. требует заполнения всех клеток таблицы. Более экономной является в этом отношении запись в виде аналитического выражения. Это особенно существенно при записи условий работы полностью определенных автоматов, задаваемых лишь множеством M_ε, ввиду громоздкости в ряде случаев множества $M_{\bar{\varepsilon}}$ и большой трудоемкости его определения. Существенные затруднения вызывает также запись условий работы элементов со сложными структурными свойствами, число входо-выходов которых для современных микроэлектронных элементов может составлять порядка 15—20.

Перевод операций любого алгоритма с табличной формы на буквенную не представляет принципиальных трудностей, однако в связи с большой громоздкостью операций в этой форме, практическое применение таких операций требует разработки специальных приемов для перехода, например, от множества M_ε к множеству $M_{\bar{\varepsilon}}$, дополнительных переменных, определения доопределяющих функций и т. п.

Автор не утверждает, что описанный алгоритм является наилучшим из возможных. Его эффективность, а также эффективность ряда дополнитель-

[7] Статистическая проверка проводилась для одновыходных структур с разномерностью задач в 45 входных переменных. Время решения задач составляло 2—3 сек.

ных операций, которые здесь не были описаны, может быть определена только при статистическом сравнении на потоке задач с другими аналогичными алгоритмами, если таковые существуют. Однако приведенное описание позволяет утверждать, что создание алгоритмов структурного синтеза, удовлетворяющего приведенным в начале настоящего раздела комплексным требованиям, является вполне реальным, в особенности, если выполнение их будет органически следовать, как в описанном алгоритме, из основных, заложенных в нем принципов.

3. Проблемы абстрактного синтеза

Общеупотребительным представлением дискретных устройств на этапах абстрактного синтеза является математическая модель конечного автомата, предложенная еще в 1951 году г. Клини. Не будет ошибкой сказать, что в настоящее время теория конечных автоматов разработана исчерпывающе. Давая общетеоретическую базу для рассмотрения конечных дискретных автоматов, она, тем не менее, весьма далека от практического применения в связи с рядом идеализаций, имеющихся в ее основе.

Основной задачей теории конечных автоматов является рассмотрение преобразования условий работы дискретного устройства, направленного к минимизации числа внутренних состояний и определение числа внутренних элементов и переходов их из одного состояния в другое. Дискретное устройство представляется при этом в виде взаимодействующих между собой логического блока и блока «памяти», содержащего задержки или элементы с фиксацией воздействий (рис. 6).

Модель конечного автомата описывается некоторым множеством состояний его элементов.

$$A = \{X, Y, Z, \delta, \lambda\},$$

где X — подмножество состояний входов, У — подмножество состояний внутренних элементов, Z — подмножество состояний выходов, δ — подмножество функций переходов внутренних состояний: $Y_{t+1} = \delta(X_t, Y_t)$ и λ-подмножество функций выходов $Z_t = \lambda(X_t, Y_t)$.

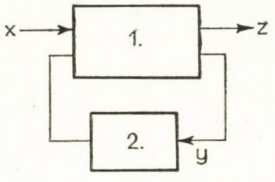


Рис. 6

Если описать конечный автомат в терминах так называемых «полных состояний» $S = \{X, Y\}$, то действие конечного автомата можно представить себе как движение по ряду непересекающихся траекторий в дискретном многомерном пространстве, в котором каждой точке его, совпадающей с какой-либо траекторией, сопоставлено заданное состояние выходов.

Основной задачей синтеза дискретного управляющего устройства в его абстрактном представлении является размещение указанных траекторий в многомерном пространстве наименьшей возможной размерности при наибольшей простоте функций δ и λ и выполнения требований к «устойчивости» устройства. Эти требования в общем смысле заключаются в отсутствии (или гарантированной вероятности, не превышающей заданной) переходов от одной заданной траектории к другой при изменении (в заданных пределах) временных и других параметров элементов и т. п.

Модель конечного автомата обладает, с точки зрения практических требований, рядом существенных недостатков, к числу которых можно отнести:

а) сложность реализации дискретного устройства определяется как числом элементов блока памяти, так и числом элементов логического блока. Между тем, все алгоритмы, разработанные для этапов абстрактного синтеза, преследуют цель минимизации лишь числа элементов в блоке памяти

б) в алгоритмах абстрактного синтеза учитываются некоторые общие свойства элементов, а именно неодинаковость задержек, наличие или отсутствие отказов и т. п. Конкретные структурные свойства элементов не учитываются несмотря на то, что они существенно влияют на сложность реализации дискретного устройства

в) модель конечного автомата требует детального перечисления элементов множеств X, Y и Z. Это делает весьма затруднительным (даже на УВМ) решение задач с числом переменных более 10—15 ($\{X\} = 2^{10}—2^{15}$).

Требования, предъявляемые к этапам абстрактного синтеза практическими задачами, можно сформулировать следующим образом:

а) наличие процедур для преобразования структур на этапах абстрактного синтеза с оценкой простоты реализации и учетом структурных свойств элементов и требований к быстродействию

б) в случае задания дискретного устройства в виде блочной структуры — наличие процедур для преобразования блочных структур с оценкой простоты реализации с учетом структурных свойств элементов и требований к быстродействию

в) наличие процедур для направленного поиска реализаций, а также кодирования состояний, близких к оптимальным по сложности реализации,

с учетом, в случае необходимости, устранения состязаний, обеспечения безотказности и т. п.

Рассмотрим эти требования несколько подробнее.

Процедуры равносильного преобразования условий работы дискретных устройств как при блочном, так и неблочном их построении, разработаны достаточно подробно, однако без оценки сложности получаемых структур. Оценка числа элементов памяти не представляет сложности. По вопросу об оценке сложности логического блока известна лишь предложенная А. К. Григорян ассимптотическая оценка для записи условий работы дискретных устройств на языке таблиц переходов, имеющая вид [2,3]:

$$R = N\left(-\sum_{i=1}^{k} p_i \log_2 p_i\right),$$

где $p_i = \frac{\omega_i}{N}$, N — общее число клеток, заполненных в таблице переходов, ω_i — число вхождений i-го состояния внутренних элементов как в устойчивых, так и неустойчивых состояниях. Методов направленного преобразования структур, ведущих к получению их оптимальной сложности реализации, пока не существует. Следует указать, что сложность реализации может существенно изменяться в зависимости от способа реализации элементов памяти (например, с применением задержек или элементов с фиксацией воздействий триггеров).

Упомянем также, что для одношаговых дискретных устройств возможна реализация структуры без задержек в цепях обратной связи, что также может представить один из вариантов реализации. Сравнительная оценка этих вариантов пока отсутствует.

Что касается учета структурных свойств элементов, то это весьма важная задача пока почти никем не исследуется. Имеются два аспекта ее: учет структурных свойств элементов при оценке сложности реализации дискретных устройств и непосредственный синтез структуры на элементах с заданными структурными свойствами.

Учет структурных свойств элементов для целей оценки сложности реализации окажется, очевидно, полезным при разработке методов направленного поиска оптимальных или близких к оптимальным структур на этапах абстрактного синтеза.

Разработка методов непосредственного синтеза является исключительно важной задачей в связи с тем, что, во-первых, наличие таких методов позволит устранить достаточно громоздкие методы структурного синтеза и, во-вторых, в связи с тем, что это будет служить непосредственным переходом к синтезу дискретных управляющих устройств на субблоках. Последняя по-

становка непосредственно связана с развитием микроэлектроники, в том числе больших интегральных схем.

В принципе задача синтеза структуры дискретного устройства на абстрактном уровне может быть сведена к задаче декомпозиции автоматов. Действительно, структурные свойства элементов могут быть записаны на том же формализованном языке, что и условия работы всего дискретного устройства в целом. Таким образом, так же как и для этапов структурного синтеза, процесс абстрактного синтеза может быть представлен как выбор оптимального выходного элемента (или одновременно нескольких элементов), и выделение его («вычитания») из структуры в целом с определением остатка, с тем, чтобы с этим остатком (или остатками) поступать таким же образом, пока реализация структуры не будет завершена (остаток равен нулю). Основные операции декомпозиции, а именно, выделение в различных вариантах подавтомата q из заданного автомата Q (параллельно от входов и выходов, последовательно от входов и выходов, с введением обратных связей и т. п.) были рассмотрены в [2, 3] и составляют принципиальное решение этой задачи. Однако, как и для структурного синтеза, для практического применения декомпозиции здесь должны быть разработаны критерии для оптимального или близкого к оптимальному выбору входных переменных и функций, подаваемых на входы элементов, и сама процедура синтеза, обеспечивающего выбор из большего числа возможных вариантов неизбыточной и близкой к оптимальной реализации.

Естественно, что эта процедура существенно отличается от рассмотренной выше для структурного синтеза, однако в методологическом отношении в них, вероятно, можно будет найти много общего.

Другим направлением решения задачи синтеза структур на этапах абстрактного синтеза является синтез на типовых блоках. Существенным здесь является классификация дискретных устройств, позволяющая выяснить их наиболее важные свойства, влияющие на характер реализации.

Решение этой задачи требует детального обследования широкого класса дискретных устройств, применяемых на практике. Предварительная классификация, разработанная в СССР, дана на рис. 7.

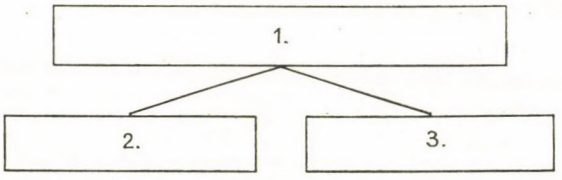


Рис. 7. 1 — многотактные (последовательностные) устройства. 2 — с последовательностями конкретного типа. 3 — с последовательностями общего типа.

К классу устройств с последовательностями общего типа относятся такие устройства, для которых можно перечислить все заданные последовательности смены состояний на входах устройства и соответствующие им смены состояний на выходах.[8] Ко вторым относятся такие, для которых можно лишь перечислить признаки, характеризующие определенный класс этих последовательностей.

Рассмотрение различных реализаций условий работы первого типа показывает, что одним из типовых блоков для них является сдвиговый регистр. Общее исследование методов реализации на сдвиговых регистрах проведено достаточно исчерпывающе (см. напр. [10]). Разработаны специальные методы кодирования и методы определений оптимальных размеров сдвиговых регистров при их различном взаимодействии, а также процедуры введения дополнительных внутренних переменных, обеспечивающих выполнение необходимых процедур кодирования. Следует отметить, что применение здесь существующих общих методов кодирования дает существенно не оптимальную реализацию структур.

Для последовательностей общего типа существенной является задача определения субблоков, обеспечивающих оптимальную их реализацию. Постановка задачи может быть сформулирована здесь следующим образом. Заданы условия работы некоторого множества модификаций устройств. Требуется определить совокупность типовых блоков, оптимально реализующих заданное множество модификаций с точки зрения, например, необходимого числа субблоков, их избыточности и т. п., при заданных ограничениях в числе входов-выходов субблока, числе его внутренних элементов и т. п. Эта весьма важная задача находится еще в самом начале своего исследования.

Одной из существенных задач здесь будет являться определение «близости» выбираемых субблоков, к заданным типовым условиям работы или частям их.

В заключение остановимся на проблеме кодирования внутренних состояний с обеспечением устойчивости (устранения недопустимых состязаний) и безотказности дискретных устройств. Решению этих задач в отдельности посвящено весьма большое число работ. Детальный обзор их с характеристикой различных существующих направлений дан в [11]. Можно считать, что задача кодирования состояний с учетом обеспечения необходимой устойчи-

[8] В принципе перечисление всех конкретных последовательностей на входах и выходах невозможно. Оно, однако, становится возможным, если число заданных последовательностей перечислимо, а относительно всех остальных известно, что они не могут быть, или что они все приводят к вполне определенному состоянию на выходах. Однако при этом существенным является, как уже указывалось выше, знание приемов для формальной записи выражения «все остальное». Следует отметить, что при автоматизации промышленных процессов указанный выше частный случай является типичным для подавляющего большинства встречающихся задач.

вости решена достаточно полно. Найдены эффективные алгоритмы кодирований и разработаны необходимые машинные программы.

В задаче кодирования состояний с учетом обеспечения заданной безотказности значительное число принципиальных вопросов также решено на основе распространения на эти задачи основных результатов, полученных в теории корректирующих кодов. Однако оказалось, что наиболее эффективные коды требуют весьма сложной реализации. Задача нахождения кодов, обладающих достаточно простой реализацией и одновременно достаточно эффективных в отношении необходимой длины кодовых слов, что соответствует минимуму числа внутренних элементов, еще не решена.

Практически надежность выполнения заданных условий работы зависит как от обеспечения устойчивости, так и от обеспечения безотказности работы дискретных устройств. Это значительно усложняет задачу кодирования состояний. Если, например, требуется обеспечить точное вычисление алгоритма функционирования (при одновременном отказе d внутренних элементов устройства (так называемая d — безотказность), то для каждой из заданных траекторий переходов в модели конечного автомата каждую вершину в многомерном пространстве нужно заменить некоторой областью этого пространства, границы которой отстояли бы от «основной» вершины на d единичных переходов. Задача размещения единичных траекторий переходов заменяется при этом задачей размещения без пересечений пучков траекторий, соответствующих указанным областям.

Впервые эти задача была исследована в работах Ю. Л. Сагаловича [12], который предложил типовую таблицу размещений, подобную таблице Д. Хафмена. Требуемое число элементов составляет $(2d+1)(N-1) \leq n \leq (2d+1)(N+1)-1$, где $N = 2^k \geq R$ и R — число состояний таблицы переходов. Также предлагается метод, заключающийся в повторении $(2d+1)$ раз кода, обеспечивающего устранение недопустимых состязаний.

В работе [13] аналогичная задача решена в предположении, что отказы элементов могут возникать только в устойчивых состояниях устройства. Получены типовые размещения, требующие $3[\log_2 R]$ элементов памяти при $d = 1$. Соответствующая таблица для $R = 4$ показана на рис. 8. Аналогичные размещения могут быть получены также и для $d > 1$. В той же работе показано, что если граф переходов для устранения состязаний закодирован соседним кодом длины m, то при $d = 1$ достаточно этот код удвоить и прибавить еще одну внутреннюю переменную так, чтобы смежным вершинам графа перехода были приписаны различные значения дополнительной переменной.

Приведенные выше соображения показывают, что решение требований, выдвигаемых практическими задачами, на этапах абстрактного синтеза еще

далеки от своего решения. Наиболее важными задачами здесь являются: учет на этапах абстрактного синтеза структурных свойств элементов, решение задачи определения оптимальных субблоков для отдельных классов последовательностей, разработка методов определения «близости» субблоков и структуры устройства в целом и разработка методов синтеза сложных структур на субблоках с обеспечением заданной устойчивости и безотказности.

$y_6y_5y_4$ \ $y_1y_2y_3$	000	001	010	011	100	101	110	111
000	⓪	0	0	0	0	1	2	2
001	0	1	0	0	0	1	3	3
010	0	1	1	1	0	1	2	2
011	1	①	1	1	0	1	3	3
100	0	0	2	3	2	2	②	2
101	1	1	2	3	2	2	2	3
110	0	0	2	3	3	3	2	3
111	1	1	2	3	3	3	3	③

Рис. 8

Цитированная литература

1. *Гаврилов М. А.:* Проблемы поиска минимальных решений при синтезе структуры релейных устройств. Труды III Всес. совещания по автоматич. управл. (технической кибернетики). Самонастраивающиеся системы. Распознавание образов. Релейные устройства и конечные автоматы. Изд. «Наука», Москва, 1967.
2. *Григорян А. К.:* Метод декомпозиции конечных автоматов. Автоматика и телемеханика *5* (1968) 101.
3. *Григорян А. К.:* Метод декомпозиции конечных автоматов с выделением выходного и входного автоматов. Автоматика и телемеханика *10* (1968) 113.
4. *Гаврилов М. А.:* Построение релейных устройств и конечных автоматов из блоков. Известия АН СССР, Техническая кибернетика *3* (1963) 13—27.
5. *Гаврилов М. А.:* Структурная теория релейных устройств. часть. IV. разд. «Блочное построение релейных устройств». Изд. Всесоюзн. заочн. энергетич. инст., 1964.
7. *Пархоменко П. П.:* Синтез структур релейных устройств методом замены входных переменных. Автоматика и телемеханика *1* (1967) 100.
8. *Горовой В. Р.:* Синтез релейных структур методом замены выходных переменных. Автоматика и телемеханика *1* (1967) 112.
9. *Гаврилов М. А., Копыленко В. М.:* Метод переходных таблиц синтеза многовыходных комбинационных структур на произвольных элементах. Изд. Ордена Ленина института проблем управления, 1970 г.
10. *Девятков В. В.:* Алгоритмы синтеза устройства на сдвиговых регистрах. Изд. Ордена Ленина института Автоматики и Телемеханики, 1969 г.
11. *Гаврилов М. А., Остиану, В. М., Потехин А. И.:* Надежность дискретных систем. Итоги науки «Теория вероятности, Математическая статистика. Теоретическая кибернетика». Изд. ВИНИТИ, 1967 г.
12. *Сагалович Ю. Л.:* Метод повышения надежности конечного автомата. Проблемы передачи информации *2* (1965) 27.
13. *Потехин А. И.:* Методы обеспечения устойчивости и *d*-безотказности работы конечного автомата. Автоматика и телемеханика *12* (1970).

Theoretical problems of practical application of finite automaton theory

M. A. GAVRILOV
(Moscow)

Summary

Modern discrete systems are distinguished, in some cases, by great complexity and, on the other hand, their practical realization, require to take real restrictions inherent to logical elements into consideration. In the present paper the requirements and the theoretical problems are investigated which arise in discrete system synthesis with the following characteristics: about 200—300 inputs, about 200—300 outputs and about 300—400 inside states. For systems of such complexity the operational conditions of discrete devices, in some cases, are given only for blocks and not always can be formulated in a non-contradiction manner. Our task is to create a language for the work conditions of discrete systems, and to create a man-machine system which automatizes the elimination of non-completeness and contradictionary work conditions.

The languages for the block structures of discrete systems are investigated. For example, the block synthesis language, elaborated in the Institute of Control Problems (USSR) is presented. Block structure transformations are described (Fig. 1). The alternative conditions of discrete control systems are also considered applying decompositions if the controller and controlled plant are given in the form of finite automata (Fig. 3).

The real properties of the logical elements are taken into consideration, mainly in structural synthesis. Therefore, in these stages corresponding procedures demanding the consideration of restrictions on the number of inputs and on the number of bifurcations of outputs, on the connections, on the number of structural levels, and the like, must be taken into account.

The technical requirements to the stages of structural synthesis are formulated. Directed construction methods of structures are considered: expedient choice is carried out for the subsequent direction of operations by applying appropriate criteria. The so-called state-transition-table method is described (Fig. 5) in which the structural synthesis is realized beginning from the output elements. The following expedient choice is foreseen:

a) realization of the output function structure similar to the optimal sequence;

b) for every given output — an appropriate choice of output element type similar to the optimal;

c) for every given output element — an optimal choice of input variables or functions. The method is applicable on wide class of logical elements including arbitrary elements.

In abstract synthesis the consideration of structural properties and logical elements restrictions is of great importance. These problems are solved by means of finite automaton decomposition techniques. The coding of states, connecting the abstract and structural synthesis, becomes essentially complicated if it is necessary to ensure specified reliability of performance (stability and efficiency). By way an example the standard coding is also presented considering both the stability and reliability (Fig. 8).

проф. М. А. Гаврилов, чл. корр. АН СССР
Институт проблем управления (автоматики и телемеханики)
СССР Москва В-485
Профсоюзная ул.81

Problems of Control and Information Theory, Vol. 1 (1), pp. 29–36 (1972)

A DISPERSION EQUATION OF NONLINEAR PLANT IDENTIFICATION

A. L. BUNICH, N. S. RAJBMAN

Moscow

(Received February 23, 1971)

General identification equation for a nonlinear plant by random function dispersion methods is obtained. The optimal weighting function of a linearized nonlinear plant is given by dispersion identification equation. Some practical examples are applied.

The rapidly advancing identification theory is now applied in biology, medicine, agriculture [1] etc. as well as in traditionally "technological" fields. Recently, effective methods for linear plant identification have been developed based on the correlation theory of random functions treated in sufficient detail in [2]. Unfortunately, most actual plants are nonlinear. This paper deals with the problem of obtaining an identification equation for a nonlinear plant by random function dispersion methods discussed in [3]. As a result, a dispersion identification equation is obtained that gives the optimal weighting function of a linearized nonlinear plant. The use of dispersion methods of random functions for the estimate of the nonlinearity degree and the correspondence between the model and the actual plant was the subject of [3, 4].

1. *Basic definitions.* Let us first introduce certain definitions of random function dispersion characteristics (see also [3]).

A function of three variables

$$\Theta_{yzx}(t, s, \tau) = M\{[M(y_t/x_\tau)] - M(y_t)] \cdot [M(z_s/x_\tau) - M(z_s)]\}, \qquad (1.1)$$

where M is the symbol of mathematical expectation will be termed a mutual cross dispersion function of random functions $Y(t)$ and $Z(s)$ with respect to the random function $X(t)$.

For simplification all random functions are assumed centered. In particular, we will not differ the terms of the correlation and the covariant function of two random functions. Consequently, a cross dispersion function can be defined as a covariant of the appropriate mathematical expectations

$$\Theta_{yzx}(t, s, \tau) = \operatorname{cov}\{M(y_t/x_\tau);\ M(z_s/x_\tau)\}. \qquad (1.2)$$

This definition is recommended for practical measurements of mutual generalized dispersion functions of actual plants or models. The function $\Theta_{yyx}(t, t, \tau) = M\{M(y_t/x_\tau)\}^2$ of two variables is termed a mutual dispersion function of the processes y_t and x_τ, while the function $\Theta_{xxx}(t, t, \tau) = M\{M(x_t/x_\tau)\}^2$ is a mutual self-dispersion function of the process x_t [3, 4]. For brevity these functions will be denoted $\Theta_{yx}(t, \tau)$ and $\Theta_{xx}(t, \tau)$ respectively. These functions can also be defined in the following way $\Theta_{yx}(t, \tau) = D\{V_\tau(t)\}$, $\Theta_{xx}(t, \tau) = D\{U_\tau(t)\}$, where $U_\tau(t) = M(x_t/x_\tau)$, $V_\tau(t) = M(y_t/x_\tau)$; D — is the symbol of dispersion. From the definition of dispersion functions and the Cauchy inequality follows the important estimate

$$| \Theta_{yzx}(t, s, \tau |^2 \leq \Theta_{yx}(t, \tau)\, \Theta_{zx}(s, \tau) . \tag{1.3}$$

In particular, for non-dispersion cross-sections, i.e. when $\Theta_{yz}(t, \tau) = 0$ or $\Theta_{zx}(s, \tau) = 0$, the cross generalized dispersion function with the appropriate values of the argument vanishes. The random function x_t is termed stationary in the dispersion sense if the following conditions are met [5]:

a) the distribution function x_t is constant, i.e.

$$F_{x_t}(u) = F_{x_s}(u) \qquad \text{for all } t, s, u, \tag{1.4}$$

b) the regression function depends only on the difference between the arguments t and s, i.e.

$$M(x_t/x_s) = \varphi_{t-s}(x_s),\ M(x_s/x_t) = \varphi_{s-t}(x_t) \quad \text{for all } t, x_t, x_s, s.$$

This property is an amplification of the concept of stationarity in a broad sense in the correlation theory of random functions.

Two functions, y_t and x_s, stationary in the dispersion sense are called stationarily bounded in terms of dispersion if for any t and s the regression function depends only on the difference of the integrals t and s and is independent of their position on the time axis; in other words, for all t and s the following condition is met

$$M(y_t/x_s) = \psi_{t-s}(x_s)\ ; \quad M(x_s/y_t) = \sigma_{s-t}(y_t) . \tag{1.5}$$

2. *Dispersion equation of identification.* In the problems of linear identification the optimality criterion depends on a function of the difference between the signal at the output to the converter and the desired value. For stationary linear systems the solution of the problem is equivalent to the minimization of the functional

$$I(g) = D\Big\{y_t - \int_{-\infty}^{t} g(t - \tau)\, x_\tau\, d\tau\Big\} \tag{2.1}$$

and leads to the Wiener-Hopf equation for the weighting function of the system.

In order to obtain a more general optimality criterion, let us formulate a problem of searching for a linear identifier (or a linear integral operator whose kernel is the model weghting function) that would relate the conditional mathematical expectations of the signals rather than the signals x_t and y_t themselves.

The quality, I of the plant identification by a linear operator L can be estimated by the quantity

$$I(L) = \Theta_{\xi x}(t, \tau) , \tag{2.2}$$

where $\xi = y - Lx$ the linearization error.

The latter expression can also be represented in the form

$$I(L) = D\{V_\tau(t) - \int_{-\infty}^{t} g_\tau(t, s) U_\tau(s)\, ds\} , \tag{2.2'}$$

where $g_\tau(t, s)$ is the kernel of the operator L. The identification quality improves with decreasing $I(L)$ and we come to the following criterion for finding an optimal L:

$$I(L) = \min. \tag{2.3}$$

The necessary extremum condition has the form

$$\delta I(L; x, y) = 0 . \tag{2.4}$$

We have therefore come to a general problem of the optimization the linear system at fixed $\tau = \tau_0$ by the minimal r.m.s. error for the input signal $u_\tau(t)$ and the required output signal $v_\tau(t)$. For a stochastic plant this statement of the problem optimization involving the consideration of the signals $u_\tau(t)$ and $v_\tau(t)$ is evidently quite natural.

Consequently, the final result can be written as

$$\int_{-\infty}^{t} g_\tau(t, s)\, R_{uu}(s, s')\, ds' = R_{vu}(t, s), (-\infty < s < t) , \tag{2.5}$$

where R_{uu} is a self-correlation function $u_\tau(t)$ while R_{vu} is a cross-correlation function of $u_\tau(t)$ and $v_\tau(t)$. By using the dispersion functions we will represent eq. (2.5) in the form

$$\int_{-\infty}^{t} g_\tau(t, s')\, \Theta_{xxx}(s', s, \tau)\, ds' = \Theta_{yxx}(t, s, \tau) . \tag{2.6}$$

This dispersion equation (2.6) is considerably simplified when the processes x_t, y_t are stationarily related in terms of dispersion. In this case the mutual cross-correlation dispersion functions do not change at a simultaneous time

shift of the variables, therefore the notation $\Theta_{xxx}(p, s) = \Theta_{xxx}(p, s, \tau)$, $\Theta_{yxx}(t, s) = \Theta_{yxx}(t, s, \tau)$ would be useful. Furthermore, since the time τ can be assumed fixed (to select another time for the fixation of the cross-section x_τ would be equivalent to a time shift), then a new designation for the weighting function of the plant could be introduced $g(t, p) = g_\tau(t + \tau, p + \tau)$. Then the dispersion equation (2.5) could be represented as

$$\int_{-\infty}^{t} g(t, p)\, \Theta_{xxx}(p, s)\, dp = \Theta_{yxx}(t, s)\,. \tag{2.7}$$

For a linear, physically feasible, model where the random process y_t^* at the output and a random process x_τ at the input are related by the linear integral operator

$$y_t^* = \int_{-\infty}^{t} g(t, \tau)\, x_\tau\, d\tau + \eta_t, \tag{2.8}$$

where η_t is a noise stochastically independent of x_τ

$$M(y_t^*/x_s) = \int_{-\infty}^{t} g(t, \tau)\, M(x_\tau/x_s)\, d\tau. \tag{2.9}$$

By equating the dispersion of both parts of the latter equation to dispersion we will have

$$\Theta_{y^* x}(t, s) = \int_{-\infty}^{t} \int_{-\infty}^{t} g(t, \tau)\, g(t, \tau')\, \Theta_{xxx}(\tau, \tau', s)\, d\tau\, d\tau'\,. \tag{2.10}$$

Therefore, if we know the mutual dispersion function $\Theta_{yx}(t, s)$ of the plant, then the difference between $\Theta_{yx}(t, s)$ and the right-hand part of eq. (2.10) may be assumed to be the nonlinearity measure. In the case where the generalized mutual dispersion function of the plant is known, the identification quality $I(L)$ can be taken as such a measure.

From the identity

$$D\left\{y_t - \int_{-\infty}^{t} g_\tau(t, s)\, x_s\, ds\right\} = DM\left\{\left[y_t - \int_{-\infty}^{t} g_t(t, s)\, x_s\, ds\right]/x_\tau\right\} + \tag{2.11}$$

$$+ MD\left\{\left[y_t - \int_{-\infty}^{t} g_\tau(t, s)\, x_s\, ds\right]/x_\tau\right\}$$

follows that the quality $I(L)$ of identification by using the functional (2.3) is at least as good as that by the criterion (2.1). For systems that are irreducible to linear the conditional dispersion in the right-hand part of eq. (2.11) is strictly

positive, therefore the quality of identification by the criterion (2.3) is better than by the criterion (2.1).

A dispersion equation for a plant with n inputs and m outputs can be obtained in a similar way from the optimal criterion

$$I_{ij}(L_{ij}) = \Theta_{\xi_i x_i}(t, \tau) = \min , \tag{2.12}$$

where

$$\xi_i = y_i - \sum_{j=1}^{n} L_{ij} x_j , \quad i = 1, \ldots, m; \quad j = 1, \ldots, n.$$

The corresponding set of linear integral equations has the form

$$\sum_{r=1}^{n} \int_{-\infty}^{t} g_\tau^{ir}(t, s') \Theta_{x_\tau x_j x_j}(s', s, \tau)\, ds' = \Theta_{y_i x_j x_j}(t, s, \tau) , \quad (i = 1, 2, \ldots, m) \tag{2.13}$$

In computer solution it is preferable to use directly the optimal criterion (2.12) by using direct methods of extremalization. Where high accuracy is not required, the system may be solved by statistical tests. Where the number of channels is not high and the computer storage is sufficient for the system coefficients, the solution can be obtained by conventional algorithms [6]. In this case integration should be replaced by summation over a discrete lattice with a constant step.

3. *On methods to solve the dispersion equation.* If a self-dispersion function of a random process at the input to the plant decreases sufficiently quickly on the infinity, then the kernel of the integral operator $\Theta_{xxx}(s', s, \tau)$ is quadratically integrable and in this case the dispersion identification equation (2.6) is an integral Fredholm equation of the first kind. This equation is known to belong to the class of improperly stated problems and is solvable by a regularizing algorithm described in [7].

On the other hand, unlike a correlation function of an ergodic process, stationary in a wide sense, the self-dispersion function of a random process should not necessarily tend to zero on the infinity.

Therefore, the integral operator in the left-hand part of the dispersion equation (2.6) is generally speaking unbounded and the functional $I(g)$ should not necessarily be continous. Then the functional may achieve the exact lower edge at no eigen-elements of the space and generalized solutions to eq. (2.6) should be considered.

When solving eq. (2.6) on a computer at a fixed value of the variable τ we will have the following matrix equation for the weighting function

$$\sum_{s'=t} \Theta_{xxx}(s', s, \tau)\, g_\tau(t, s') = \Theta_{yxx}(t, s, \tau)$$

The system of integral equations (2.13) is solved in a similar way.

4. *Comparison with the Wiener-Hopf equation.* Let us show now that an integral equation from the correlation theory of identification can be regarded as a first approximation to a dispersion equation. If to use series expansion of the conditional mathematical expectation and eliminate the terms which contain x_τ in powers above one, then $M(x_t/x_\tau) = f(s', \tau)\, x_\tau + 0(x_\tau)$, where $0(x_\tau)$ — is an infinitesimal of the second order with respect to x_τ. Consequently, in a linear approximation the process regression function at the input $B(s', \tau) = f(s', \tau) x_\tau$. By multiplying the identity $M(x_\tau\, y_t/x_\tau) = x_\tau\, M(y_t/x_\tau)$ by $f(s, \tau)$ we will have $f(s, \tau)\, M[x_\tau y_t/x_\tau] = M(x_s/x_\tau)\, M(y_t/x_\tau)$ or, in terms of mathematical expectation and equating both parts we will have

$$\Theta_{yxx}(t, s, \tau) = f(s, \tau)\, R_{yx}(t, \tau). \tag{4.1}$$

In particular, at $y_t = x_t$ we have

$$f(s, \tau)\, R_{xx}(t, \tau) = \Theta_{xxx}(t, s, \tau). \tag{4.2}$$

Following the substitution of (4.1) and (4.2) into eq. (2.6) and dividing by $f(s, \tau)$ we will have

$$\int_{-\infty}^{t} g_\tau(t, s')\, R_{xx}(s', \tau)\, ds = R_{yx}(t, \tau). \tag{4.3}$$

Similar considerations applied to the dispersion equation (2.7) for processes stationarily related in terms of dispersion lead to the Wiener-Hopf equation.

For processes with linear regression this approximation is accurate and in this case the dispersion equation is equivalent to its correlational analog. Gaussian processes may be an example.

In the general case a dispersion equation contains more information than the corresponding correlational equation. Indeed, from the identity

$$R_{yx}(t, \tau) = M(y_t\, x_\tau) = M\{M\,[x_\tau\, y_t/x_\tau]\} = M\{x_\tau\, M[y_t/x_\tau]\} = M\{M(y_t/x_\tau)\, M(x_\tau/x_\tau)\}$$

and

$$R_{xx}(t, \tau) = \Theta_{xxx}(t, \tau, \tau),$$

which is obtained from the preceding one by replacing y by x it follows that the dispersion equation (2.6) becomes the correlation equation (4.3) if the arguments s and τ coincide. Generally speaking, the cross-correlation function of the processes x_τ and y_s can vanish and the correlational equation will lead to a zero solution. But to conclude that there is no statistical relation between the signals whose cross-correlation function vanishes would be erroneous. A quadratic detector with a Gaussian process at the input is an example. In that case the cross-correlation function $R_{yx}(t, s) = R_{x^2x}(t, s) \equiv 0$ but at the same time the mutual dispersion function $\Theta_{yx}(t, s) = 2R^4_{xx}(t, s)\, D^2\{x_t\}$ is non-zero [3].

5. *Examples.* 1) The random process $x(t)$ for which the conditional mathematical expectation $M(x_t/x_s)$ is a white noise will be termed white noise stationary in terms of dispersion. The dispersion function of such a process is a delta-function of the difference between the arguments t and s: $\Theta_{xxx}(t, s, \tau) = \delta(t - s)$. The substitution of this expression into the dispersion equation (2.7) gives

$$g_\tau(t, s) = \Theta_{yxx}(t, s, \tau) , \tag{5.1}$$

in other words the weighting function of the plant coincides with its mutual generalized dispersion function.

2) Let us solve the dispersion equation explicitly for the case of an inertialess plant described by the functional equation $y_t = f(x_t)$. Since a cross-section of y_t depends uniquely on the cross-section of x_t, the weighting function should naturally be sought in the form $g_\tau(t, s') = \varphi(t, \tau)\,\delta(t - s')$. The substitution of this expression into the dispersion equation will give

$$\varphi(t, \tau)\,\Theta_{xxx}(t, s, \tau) = \Theta_{yxx}(t, s, \tau) . \tag{5.2}$$

Assuming $t = s$ we will have

$$g_\tau(t, s') = \frac{\Theta_{yx}(t, \tau)}{\Theta_{xx}(t, \tau)}\,\delta(t - s') . \tag{5.3}$$

3) Let us consider a dynamic plant where the relation between the input process x_t and the output process y_t can be described by the equation

$$Y = \hat{Q}x + W, \tag{5.4}$$

where $\hat{Q}$ is a linear integral operator, while W is a process andispersed with x. Then $I(Q) = \Theta_{wx}(t, \tau) = 0$ and since the functional I is non-negative, the identification by the operator Q is optimal. This example is the only case where the left-hand part of eq. (2.6) vanishes since $I(Q) = 0$ follows from. (5.4).

References

1. *Astrom, K. J., Eykhoff, P.:* System identification. IFAC symposium on "Identification and Process-Parameter Estimation", Prague, 1970.
2. *Пугачев, В. С.:* Теория случайных функций и ее применение к задачам автоматического управления. Физматгиз, 1962.
3. *Райбман, Н. С., Терехин, А. Т.:* Дисперсионные методы случайных функций и их применение для исследования нелинейных объектов управления. Автоматика и телемеханика *26* (1965).
4. *Райбман, Н. С., Чадеев, В. М.:* Адаптивные модели в системах управления. Изд-во «Советское радио», 1966.
5. *Райбман, С. Н., Ханш О.:* Дисперсионные методы идентификации многомерных нелинейных объектов управления. Автоматика и телемеханика *5* (1967).
6. *Березин, И. С., Жидков, Н. П.:* Методы вычислений т. 1, П. М. Наука, 1966, издание 3-е.
7. *Тихонов, А. Н.:* О регуляризации некорректно поставленных задач. ДАН, т. 153, № 1.

Уравнение дисперсии идентификации нелинейных объектов

А. Л. БУНИЧ–Н. С. РАЙБМАН
(Москва)

Резюме

Непараметрическая идентификация линейных стационарных объектов связана с с решением интегрального уравнения Винера—Хопфа с целью определения весовой функции объекта по корреляционной функции входа и взаимной корреляционной функции входа и выхода. В настоящей статье получено аналогичное уравнение для идентификации нелинейных объектов. Обычные методы идентификации, основанные на определении весовой функции линейной модели объекта по их корреляционным характеристикам не учитывают структуру случайного процесса на входе объекта. Так, например, поскольку любая автокорреляционная функция является автокорреляционной функцией некоторого нормального процесса, то случайный процесс на входе объекта можно считать нормальным и такое предположение нисколько не влияет на построение оптимальной линейной модели. Кроме того взаимная корреляционная функция процессов на входе и на выходе объекта не является достаточно полной характеристикой стохастической зависимости. При исследовании объектов с нелинейной регрессией взаимная корреляционная функция может тождественно обратиться в нуль или дать заниженное значение силы связи между процессами и в этих случаях весовая функция оптимальной линейной модели не является удовлетворительной характеристикой идентифицируемого объекта. Таким образом возникает необходимость введения более полных характеристик случайных процессов и построение на основе этих характеристик оптимальной модели, учитывающей структуру входного сигнала. Это уравнение базируется на дисперсионных методах случайных функций, которые рассмотрены в [3]. В первой части статьи вводятся определение обобщенной дисперсионной функции (1.1), понятие стационарной случайной функции (1.4) и стационарно связанных (1.5) случайных функций в дисперсионном смысле. Вторая часть посвящена получению дисперсионного уравнения идентификации по критерию минимума функционала (2.3). В результате получено уравнение идентификации (2.6), а для стационарного в дисперсионном смысле случая — уравнение (2.7). Приводится также обобщение на многомерный случай — уравнение (2.13).

В заключение обсуждается вопрос о корректности дисперсионного уравнения идентификации, сравнение последнего с корреляционным уравнением и приводятся примеры.

A. L. Bunich—N. S. Rajbman
Institute of Control Problems
Profsoyuznaya ul. 81
Moscow V—485, USSR

Problems of Control and Information Theory, Vol. 1 (1), pp. 37–50 (1972)

ON ITERATION RULES WITH MEMORY IN MACHINE LEARNING

S. CSIBI

Budapest

(Received February 23, 1971)

Constraints are studied under which a broad class of machine learning procedures, governed by iteration rules with memory, converge. More distinctly, iteration rules with weak (or just finite) memory with respect to previously proposed discrimination functions are assumed. Procedures of this sort offer rich possibilities to find convergent learning algorithms with more powerful processing steps and reduce, in this way, the time necessary for learning. Relations of interest in devising such sort of iteration rules are proved.

1. Introduction

Assume a teacher, drawing samples ω_t, $t = 0, 1, \ldots$ afteranother from, say, a finite dimensional Euclidean space Ω, and presenting, together with each ω_t, also either a real $\Theta(\omega_t)$ or, simply, a binary label (to indicate which of two assumed hypotheses H_0 and H_1 holds).

When $\{\omega_t, \Theta(\omega_t);\ t = 0, 1, \ldots\}$ is directly taught, our problem obviously is to devise, by observing this sequence, an estimate for the unknown function $\Theta = \{\Theta(\omega), \omega \in \Omega\}$.

However, also for binary labels, one may appropriately pose, at least as an initial step, the problem of learning as a problem of estimating some appropriate discrimination function Θ. We may take for this purpose the a posteriori probability function with respect to say H_1, which is, in typical learning problems, of course, also unknown. ($\Theta(\omega)$ stands for the conditional probability that, e.g., H_1 holds, given ω.) Specifically, for unambiguous decision problems, we may take, as a Θ, any real valued function for which

$$\operatorname{sign} \Theta(\omega) = \begin{cases} +1, & \text{if } \omega \in A\,, \\ -1, & \text{if } \omega \in B\,, \end{cases}$$

provided H_0 and H_1 hold iff $x \in A$ and $x \in B$, respectively, and $\omega_t \in A \cup B$, for all t.

It does not make much matter, within the scope of this study, which way the teacher exactly works, and how is Θ, specifically, introduced. Neverthe-

less, in order to avoid vagueness, one may simply think of a Θ which is defined, just in one of the previous ways. Let our problem be estimating Θ recursively. Let us denote by $Z_t = \{Z_t(\omega), \omega \in \Omega\}$ the estimate we propose, having observed ω_t.

Assume Θ as well as any Z_t to be members of some separable Hilbert space $\mathcal{H}$, the inner product of which is, for any $\xi, \eta \in \mathcal{H}$, $\langle \xi, \eta \rangle$. ($||\xi||^2 = \langle \xi, \xi \rangle$).

One may, of course, also specify the function Θ and any Z_t, simply, by a sequence of coordinates with respect to either some base or other complete linearly independent set of functions in $\mathcal{H}$ [5, 6]. (This is just another way to describe a class of functions defined in Ω.) There are, however, also learning problems in which some Θ is to be estimated in a space $\mathcal{H}$, which is not, specifically, a space of functions defined in Ω. Non-supervised learning, governed by costs and assuming for each class prototypes, is a well known example for such learning problems [5].

Let us propose, as an estimate of Θ at t, $Z_t = X_t$ ($t \in T = (0, 1, \ldots)$); $X = \{X_t, t \in T\}$ being an iteration process, defined successively, starting at some arbitrary $X_0 \in \mathcal{A} \subset \mathcal{H}(\mathbf{E}\, ||X_0||^2 < \infty)$, by

$$X_{t+1} = \Phi(X_t + \alpha_t W_t) . \tag{1}$$

(E stands for the expectation.)

While the distinction, at t, between an arbitrary estimate Z_t and the one X_t, generated by previous estimates and labelled samples according, specifically, to the iteration rule (1), may seem subtle, it turns out in the sequel, that we actually need this distinction.

$\alpha = \{\alpha_t\}_{t=0}^{\infty}$ stands for a sequence of positive constants (to be specified in the sequel). Φ: $\mathcal{H} \to \mathcal{A}$ denotes a truncation (also to be specified in more detail) by which the range of X may be kept, for all $t \in T$, within some appropriately defined $\mathcal{A} \subset \mathcal{H}$.

The labelled samples given by the teacher (or the evaluation in unsupervised learning) governs, at t, the iteration through $W_t \cdot W_t$ is assumed to be a function within $\mathcal{H}$ and measurable with respect to $\mathcal{F}(X^t) \times \mathcal{B}^t$. ($\mathcal{F}(X^t)$ denotes the σ-algebra generated by $X^t = \{X_\vartheta, \vartheta \leq t\}$ and $\mathcal{B}^t$ the Borel sets in $X_{\vartheta=0}^t \Omega_\vartheta$, $\Omega_\vartheta = \{\omega_\vartheta\}$). We term, in the sequel W regulator.

There are well known ways how to arrive at such regulators.

Let, e.g., dim $\mathcal{H} = n$ and $\Sigma_{i=1}^n c_i \varphi_i$ denote the least square estimate of a real valued function $F = \{F(\omega), \omega \in \Omega\}$ being taught sequentially by a teacher, with respect to some set $\{\varphi_i\}_{i=1}^n$ of linearly independent functions. (Let ω_t, $t = 0, 1, \ldots$ be drawn completely independently, according to some probability distribution Q. Assume F and φ_i to be in $\mathfrak{L}_2(Q)$, and take, simply,

$\Phi = \mathbf{I}$. (Denote by $\Sigma_{i=1}^{n} c_i^{(t)} \varphi_i$ the approximation, we propose having observed ω_t. It is well known [5] that one may, in this case, set up a learning algorithm in terms of $\{\varphi_i\}_{i=1}^{n}$ and $X_t = \{c_i^{(t)}\}_{i=1}^{n}$, taking as a regulator $W_t = \{W_t^{(i)}\}_{i=1}^{n}$. ($W_t^{(i)} = 2(F(\omega) - \Sigma_{i=1}^{n} c_i^{(t)} \varphi_i(\omega))\, \varphi_i(\omega)$, $i = \overline{1, n}$).

We may take as another example for W any of the well known correction terms in potential function type learning algorithms, as defined by Aĭzerman, Braverman and Rozonoer [1, 2]. One may also use, as a regulator, correction terms, according to Parzen estimators and kernel type estimates of a similar sort [13].

In these well known cases one usually adopts such regulators W_t which either depend, at any t, on the collection $Z^t = \{Z_\vartheta, \vartheta \leq t\}$ of previously proposed estimates just through the most recent estimate Z_t, or assume specific expression for the dependence on Z^t [12]. In contrast to this, we admit in this paper regulators W_t, having almost arbitrary dependence on Z^t, apart from the usually met constraint that all previous estimates, proposed sufficiently long ago, are assumed to have, on W_t, either no or just negligible influence. We reduce in the sequel, by means of this and some other properties, the conditions under which a broad class of such procedures converge, to well known criteria by Braverman and Rozonoer [2] on the stability of random processes. By this we get a useful insight (i) how to guarantee convergence for such sort of learning procedures (or retain their convergence when modifying the algorithm) and (ii) extend the scope of random processes, for which stability may be proved, essentially.

2. Regulators with weak memory

Next let us define the properties, we assume for W, precisely. (We impose, in what follows, conditions (C.1) through (C.3) for any $t \in T$ almost surely.)

Let

$$\mathbf{E}(||W_t||^2 \mid Z^t) \leq C_0\,, \tag{C.1}$$

($\mathbf{E}(.|.)$ stands for the conditional expectation and C_i, $i = 0, 1, \ldots$ for some given positive real.) We also assume weak memory with respect to previous estimates, in the sense that for some $\mu > 0$:

$$||\mathbf{E}(W_t \mid Z^t) - \mathbf{E}(W_t \mid Z^{\overline{t-\tau,t}})|| = 0(\tau^{-\mu}), \tag{C.2}$$

as $\tau \to \infty$.

($Z^{\overline{t-\tau,t}} = \{\tilde{Z}_\vartheta\,;\, \vartheta \leq t\}$; $\tilde{Z}_\vartheta = Z_\vartheta$ for $t - \tau < \vartheta \leq t$, and $\tilde{Z}_\vartheta = 0$ for $\vartheta \leq t - \tau$.)

If, specifically, there exists some $\tau_0 > 0$, for which

$$||\mathbf{E}(W^t \mid Z^t) - \mathbf{E}(W_t \mid Z^{\overline{t-\tau,t}})|| = 0\,, \tag{C.2a}$$

for any $\tau > \tau_0$ we say W_t exhibits, with respect to Z^t, finite memory.

Let

$$G_t(Z^t) = \mathbf{E}\left(\langle Z_t - \Theta,\, W_t \rangle \mid Z^t\right).$$

We assume for this functional the following property:

$$|G_t(Z^t) - G_t(\tilde{Z}^t)| \leq C_{14}\,\delta_t \tag{C.3}$$

for any $Z_t, \tilde{Z}_t \in \mathfrak{A}$. Here $\delta_t = \sup_{\vartheta<t} ||Z_\vartheta - \tilde{Z}_\vartheta||$.

In addition to these restrictions (which are loose in the sense that meeting them scarcely makes any difficulty) let us impose on the regulator W also the following three significant constraints: Let:

$$\mathbf{E}(W_t \mid Z^t)_{Z_\vartheta = \xi,\, \vartheta \leq t} = m(\xi), \tag{C.4}$$

for any $\xi \in A$ independently of t (which means a sort of conditional stationarity), and assume that there exists such a $\Theta \in \mathfrak{A}$, for which

$$\langle \xi - \Theta, m(\xi) \rangle \leq 0\,, \tag{C.5}$$

for all $\xi \in A$, and let

$$\lim p_{k\to\infty} ||Z_{t_k} - \Theta|| = 0\,, \tag{C.6}$$

for any subsequence $\{t_k\}_{k=0}^{\infty}$ for which $\mathbf{P}(\lim_{k\to\infty} V_{t_k} = 0) = 1$, provided that $Z_{t_k} \in \mathfrak{A}$, for all t_k. ($V_t = \langle Z_t - \Theta, m(Z_t) \rangle$, lim p denotes limit in probability. Here, and in what follows $\mathbf{E}(\tilde{Z} \mid Z^t)_{Z_\vartheta = \xi,\, \vartheta \leq t}$ denotes, for any random variable $\tilde{Z}$, the value $\mathbf{E}(\tilde{Z} \mid Z^t)$ takes, provided Z^t takes the value $Z_\vartheta = \xi, \vartheta \leq t$).

There are learning problems for which we can not guarantee that condition (C.6) holds, however, instead of this, we have an evidence for the following two specific properties. Viz., (i) the objects $\omega_0, \omega_1, \ldots$ are drawn at $t = 0, 1, \ldots$ completely independently, according to some probability distribution Q, and (ii) there exists a $\Theta \in \mathfrak{A}$, for which

$$\langle \xi - \Theta, m(\xi) \rangle = -\,C_1 \int_{\Omega} |\, \xi(\omega) - \Theta(\omega) \,|\, Q(d\omega). \tag{C.7}$$

Here, and in what follows, we specifically assume that $\mathfrak{A} = \{a_i \leq \leq \langle \eta, e_i \rangle \leq b_i,\ i = 1, 2, \ldots\}$ (i.e., we take as $\mathfrak{A}$ a parallelepiped, spanned by

some base $\{e_i\}_{i=1}^{\infty}$ in $\mathcal{H}$, $-\infty \leq a_i \leq b_i \leq \infty$ being a priori fixed reals) and define the truncation operator Φ by

$$\Phi(\eta) = \eta, \text{ if } \eta \in \mathcal{A},$$

$$\langle \Phi(\eta), e_i \rangle = \begin{cases} b_i, & \text{if } \langle \eta, e_i \rangle > b_i, \\ a_i, & \text{if } \langle \eta, e_i \rangle < a_i. \end{cases} \tag{2}$$

Theorem 1. If (i) for the regulator W (C.1)—(C.6) hold, (ii) we generate the process X by iteration, according to (1) and (2), and (iii) impose on α the following constraints

$$\sum_{t=0}^{\infty} \alpha_t = \infty, \quad \sum_{t=0}^{\infty} \alpha_t^2 < \infty, \quad \sum_{t=0}^{\infty} \alpha_t t^{-\varepsilon} < \infty, \quad \sum_{t=0}^{\infty} t^{2\eta} \alpha_t \left(\max_{t-t^{\eta} < \vartheta \leq t} \alpha_{\vartheta} \right) < \infty, \tag{3}$$

for any $\varepsilon > 0$ and some $0 < \eta < 1$, then $\mathbf{P}(\lim_{t \to \infty} || X_t - \Theta || = 0) = 1$.

Theorem 2. If (i) the samples $\omega_0, \omega_1, \ldots$ are drawn completely independently according to some probability distribution Q, (ii) the regulator W is uniformly bounded (viz., $||W_t|| < C_2$, $|W_t(\omega)| < C_3$, for all t and ω) and also meets (C.1) (C.2a) (C.3) (C.4) and (C.7) and (iii) the process X is generated as given in Theorem 1, then $\mathbf{P}(\lim_{t \to \infty} \int_{\Omega} |X_t(\omega) - \Theta(\omega)| Q(d\omega) = 0) = 1$.

Remarks. (I) Observe that the essential constraints imposed on regulator W in Theorems 1 and 2 (viz. (C.4)—(C.6), and (C.4) and (C.7), respectively) constrain the regulator W only under that specific configuration of previous estimates Z^t for which all Z_ϑ, $\vartheta \leq t$ are identical. For any other Z^t we have no essential constraint how to devise W. This offers a considerable freedom to adopt even heuristic ideas when devising algorithms with guaranteed convergence [4]. — (II) Condition (3) holds, e.g., if $\alpha_t = t^{-1}$ for $t \geq 1$ and $0 < \eta < 1/2$. — (III) One may also extend a broad class of well known temporary continuous iteration procedures [9—11, 3] in an almost similar way to regulators with memory.

One may specialize (C.6) by

Lemma 1. (C.6) holds if, in addition to (C.5), we have

$$\inf_{\varepsilon < ||\xi - \Theta||} \langle \xi - \Theta, m(\xi) \rangle < 0, \tag{C.8}$$

for any $\xi \in \mathcal{A}$ and $\varepsilon > 0$.

3. Cost functions

Assume we may, at any instant $t \in T$, evaluate the performance of the learning procedure by some appropriate cost function $\tilde{Y}_t$ (i.e., an appropriate function taking non-negative values) which, however, may have memory with respect to the previous estimates Z^t.

Let us assume the use of costs in statistics as well as, specifically, in machine learning well known [5, 6], provided $\widetilde{Y}_t$ depends on Z^t merely through the most recently proposed estimate Z_t. We, therefore, immediately turn to the more general form of this sort of dependence.

Observe that, within a wide scope of actual learning problems, the final form of the estimate has to be developed within some finite training period, and one has actually to use the classifier only at such instants t at which the after-effects of this training period have for long disappeared; and at such instants we may well replace Z^t by $\{Z_\vartheta = Z_t, \vartheta \leq t\}$.

Assume that $\widetilde{Y}$ is conditionally stationary, at least in the sense that $\mathbf{E}(\widetilde{Y}_t \mid Z^t)_{Z_\vartheta = \xi, \vartheta \leq t} = R(\xi)$ is independent of t. If $\widetilde{Y}$ is such and the classifier is really to be used only long after the training has ended, we obviously have, in the course of the learning procedure, to search for such a function Θ, for which $R(\xi) = \min$, if $\xi = \Theta$. (I.e. Θ is a minimum of the regression function R.)

In machine learning we may in many cases, introduce $\widetilde{Y}$ ourselves, and may, therefore, compute, at any t, either its gradient defined in an appropriate sense (if such exists) or look for some other function which may take, in finding Θ, a similar role. (Next, for the sake of simplicity, we assume that, for the $\widetilde{Y}$ that is given, $\operatorname{grad}_\xi \widetilde{Y}_t$ exists).

In this case it is a well known and efficient way to use, for finding Θ, some gradientlike function instead of $\widetilde{Y}_t$. One need not, however, insist in so doing. In this section we show that any Y_t may be used instead of $\widetilde{Y}_t$ for this purpose, provided $\mathbf{E}(Y_t \mid Z_t)_{Z_\vartheta = \xi, \vartheta \leq t}$ behaves, for all $\xi \in A$, like $\operatorname{grad}_\xi R$.

More distinctly, assume that Y is such that (i) $\mathbf{E}(Y_t \mid Z^t)_{Z_\vartheta = \xi, \vartheta \leq t} = r(\xi)$ is independent of t, and (ii) $r(\xi) = \operatorname{grad}_\xi R$ for any $\xi \in A$. (We term any such Y_t a quasi-gradient of $\widetilde{Y}_t$. Obviously, great many functions may serve as quasi-gradients, given $\widetilde{Y}$.)

The first of these relations means a sort of conditional stationarity. The second enables one, this time again, to introduce also heuristic ideas when associating a quasi-gradient Y with some given $\widetilde{Y}$. (Observe that in all of these cases one may find Θ among the roots of the regression function r.)

That any such quasi-gradient may play, in finding Θ, the role of a gradient, follows from

Theorem 3. If (i) for the specific choice of the regulator $W_t = -Y_t$ (C.1), (C.2)—(C.4) and (C.8) hold, and (ii) we generate the process X by iteration, as given by (1) and (2) adopting coefficients α according to (3), then $\mathbf{P}(\lim_{t \to \infty} ||X_t - \Theta|| = 0) = 1$.

Remarks (I) For the specific choice of the regulator $W_t = -Y_t$ (C.8) means that the regression function R has just a single minimum within $\mathfrak{K}$ and does not approach to such at the boundaries. — (II) Confining X to

some bounded $\mathcal{A}$ is particularly relevant if we can not a priori specify any $\mathcal{A}$, for which (C.8) holds, and R may also be multimodal. (For such cases presently just some fundamental theoretical results are known [9, 10].) The truncation of X to some bounded set enables one to embed (1) into some Monte Carlo search, in a sense that we draw $\mathcal{A}$ at each time from some subdivision of $\mathcal{H}$, and then adopt iterations with coefficients α and $-\alpha$ anotherwise.

4. Potential functions

Assume that a teacher presents, together with each sample ω_t, specifically, the value $\Theta(\omega_t)$ that the discrimination function Θ takes at ω_t. Let us adopt, when defining the regulator W, a potential function [1]; more distinctly, an extention of this to a more general class of positive definite functions [7]. Accordingly, let us define the regulator W by means of a positive definite function $K = \{K(\omega, \tilde{\omega}), (\omega, \tilde{\omega}) \in \Omega \times \Omega\}$, for which $\Theta \in \mathcal{H}(K) = \mathcal{H}$ holds. ($\mathcal{H}(K)$ stands for the reproducing kernel Hilbert space, generated by K. We assume that $\mathcal{A} \subset \mathcal{H}(K)$).

Some possibilities of regulators with memory, generated by means of a potential function, are illustrated by

Theorem 4. If (i) the samples ω_t, $t = 0, 1, \ldots$ are drawn completely independently according to some probability distribution Q, (ii) $|K(\omega, \tilde{\omega})| < C_4$ for any $(\omega, \tilde{\omega}) \in \Omega \times \Omega$ (iii) we generate the process X by iteration according to (1) and (2), taking a sequence α of coefficients according to (3), and (iv) adopt, at any $t \in T$, as a regulator $W_t = (\Sigma_{\vartheta \in \Delta} r_\vartheta(\omega_t)) K(., \omega_t)$ $(r_\vartheta(\omega_t) = \operatorname{sign}((\Theta(\omega_t) - Z_\vartheta(\omega_t))$, $\Delta \subset [t - \tau_0, t]$, τ_0 denoting some a priori fixed integer) then $\mathbf{P}(\lim_{t \to \infty} \int_\Omega |X_t(\omega) - \Theta(\omega)| Q(d\omega) = 0) = 1$.

Remarks (I). We have chosen a regulator, specifically, as given by (iv) in Theorem 4, just for illustration. However, this specific choice already points out some obvious possibilities how to include heuristic steps into convergent iteration procedures. — (II) One may completely define initial conditions, e.g., by setting $Z_\vartheta = 0$, for all $\vartheta < 0$. — (III) Condition (i) is adopted just for the sake of simplicity. For results including also dependent observations see Ref. [4a].

Appendix — Proofs

The problem in proving Theorems 1 and 2 is how to relate the conditional expectations $\mathbf{E}(||X_t||^2 | X^t)$ and $\mathbf{E}(W_t | Z^t)_{Z_\vartheta = X_\vartheta, \vartheta \leq t}$ (the former being of interest in convergence studies, and the latter within the properties (C.4)—C.8)) although

the conditioning is in these two cases different. One may remove this difficulty by means of (C.2) and (C.3), as we show it, precisely, next in Lemma 3.

The main point is, therefore, the proof of Lemma 3. Having done this, Theorems 1 and 2 may readily be reduced to well known convergence theorems by Braverman and Rozonoer [2, 8] concerning certain functionals of X^t.

Theorems 3 and 4 are corollaries for specific choices of the regulator W.

Let $\widetilde{X}_t = X_t - \Theta$ and

$$\widehat{X}_{t+1} = X_t + \alpha_t W_t . \tag{A.1}$$

From (A.1) and the definition of Φ (viz. (2)) obviously follows

Lemma 2

$$||X_t - \eta|| \leq ||\widehat{X}_t - \eta|| , \text{ for any } \eta \in \mathfrak{N} .$$

I.e., Φ is, with respect to $\mathfrak{N}$, uniformly norm-reducing.

Remarks. (I) The specific form of the set $\mathfrak{N}$ and the truncation Φ enables us to give upper bounds on $||\widetilde{X}_{t+1}||$ and $||X_{t+1} - X_t||$, simply, by means of the untruncated iteration (A.1). — (II) (2) is the sort of truncation one usually adopts when computations are to be confined to finite intervals. (2) also is, specifically in one dimension, the truncation usually adopted for confining probability estimates to [0, 1] [1].

Lemma 3. For X the following inequality holds:

$$\mathbf{E}(||X_{t+1}||^2 \mid X^t) \leq ||X_t||^2 + 2\alpha_t \langle \widetilde{X}_t, m(X_t) \rangle + \tilde{S}_t ,$$

where $\tilde{S}_t$ is a non-negative valued $\mathfrak{F}(X^t)$-measurable function, and $\Sigma_{t=0}^{\infty} \mathbf{E}\, \tilde{S}_t < \infty$.

Remark. The inequality, in Lemma 3, together with the constraints on the regulator W and the iteration rule (1), says that, as $t \to \infty$, X approaches to a super-martingale, with probability one.

Proof of Lemma 3. From (1) and Lemma 2:

$$\mathbf{E}(||\widetilde{X}_{t+1}||^2 \mid X^t) \leq ||\widetilde{X}_t||^2 + 2\alpha_t \mathbf{E}(\langle \widetilde{X}_t, W_t \rangle \mid X_t) + \alpha_t^2 \mathbf{E}(||W_t||^2 \mid X^t) . \tag{A.2}$$

Let us consider the following approximations:

$$\mathbf{E}(\langle X^t, W_t \rangle \mid X_t) = \langle \widetilde{X}_t, m(X_t) \rangle + \langle \widetilde{X}_t, h_t(\tau) + \tilde{h}_t(\tau) \rangle + g_t . \tag{A.3}$$

for all $t \in T$, where

$$h_t(\tau) = \mathbf{E}(W_t \mid X^t) - \mathbf{E}(W_t \mid X^{\overline{t-\tau,t}}),$$

$$g_t = G_t(X^{\overline{t-\tau,t}}) - G_t(f_{x_t}^{\overline{t-\tau,t}}),$$

$$\tilde{h}_t(\tau) = \mathbf{E}(W_t \mid Z^t)_{Z^t = f_{x_t}^{\overline{t-\tau,t}}} - m(X_t).$$

(Here $f_{x_t}^{\overline{t-\tau,t}} = \{\tilde{f}_\vartheta, \vartheta \leq t\}$; $\tilde{f}_\vartheta = X_t$ for $t - \tau < \vartheta \leq t$, $\tilde{f}_\vartheta = 0$ for $\vartheta \leq t - \tau$. $X^{\overline{t-\tau,t}}$ is defined as $Z^{\overline{t-\tau,t}}$, replacing Z by X. Here, an in what follows, e.g., $\mathbf{E}(\tilde{Z} \mid Z^t)_{Z_\vartheta = Z_t, \vartheta \leq t}$ denotes, for any random variable $\tilde{Z}$, an $\mathcal{F}(Z_t)$-measurable function, which takes the value $\mathbf{E}(\tilde{Z} \mid Z^t)_{Z_\vartheta = \xi, \vartheta \leq t}$, provided Z_t takes the value ξ. In (A.3) we also realize that $\mathcal{H}$ is separable, and, therefore, $\mathbf{E}(\langle Z_t, W_t \rangle \mid Z^t) = \langle Z_t, \mathbf{E}(W_t \mid Z^t) \rangle$ is obvious.)

Let

$$q(\tau) = \begin{cases} h_t(\tau) \\ \tilde{h}_t(\tau) \end{cases}$$

and $\tau = \lceil t^\eta \rceil$. $(0 < \eta < 1/2$. $\lceil t^\eta \rceil$ stands for $1 +$ ent $t^\eta)$.

From this and (C.2):

$$||q(\lceil t^\eta \rceil)|| \leq C_5 t^{-\eta\mu} \tag{A.4}$$

almost surely for any $t > t_0 > 0$.

From the triangle inequality and Lemma 2

$$||X_t - X_\vartheta|| \leq \sum_{s=\vartheta}^{t-1} ||X_{s+1} - X_s|| \leq \sum_{s=\vartheta}^{t-1} \alpha_s ||W_s||. \tag{A.5}$$

Observe that for $Z^t = X^{\overline{t-\tau,t}}$ as well as $Z^t = f_{x_t}^{\overline{t-\tau,t}}$ we have $Z_\vartheta = 0$ for $\vartheta \leq t - \tau$. The deviation between these two sequences is, therefore (in the sense given in (C.3))

$$\delta_t(\lceil t^\eta \rceil) = \max_{t-\tau < \vartheta \leq t} ||X_t - X_\vartheta|| \leq \sum_{\vartheta=t-\tau}^{t-1} ||X_t - X_\vartheta|| \tag{A.6}$$

From (A.5) and (A.6):

$$\delta_t(\lceil t^\eta \rceil) \leq \sum_{\vartheta=t-\tau}^{t-1} \sum_{s=\vartheta}^{t-1} \alpha_s ||W_s||. \tag{A.7}$$

Observe that $\delta_t(\lceil t^\eta \rceil)$ and the property that $\delta_t(\lceil t^\eta \rceil)$ is a $\mathcal{F}(X^t)$-measurable function, for all t.

From (A.4) and Schwartz's inequality:

$$| \langle \widetilde{X}_t, h_t([t^\eta]) + \tilde{h}_t([t^\eta]) \rangle | \leq C_6 \, || \, \widetilde{X}_t \, || \, t^{-\eta\mu}, \tag{A.8}$$

for all $t > t_0$, almost surely.

From (A.3), (A.8), the definition of $\delta_t([t^\eta])$ and (C.3)

$$2\alpha_t \, \mathbf{E}(\langle X^t, W_t \rangle \mid X^t) \leq 2\alpha_t \langle \widetilde{X}_t, m(X_t) \rangle + S_t. \tag{A.9}$$

Here:

$$| \, S_t \, | = 2\alpha_t \big(|| \, \widetilde{X}_t \, || \, C_6 \, t^{-\eta\mu} + C_7 \, \delta_t([t^\eta]) \big) . \tag{A.10}$$

From (A.2) and (A.9)

$$\mathbf{E}(|| \, \widetilde{X}_{t+1} \, ||^2 \mid X^t) < || \, \widetilde{X}_t \, ||^2 + 2\alpha_t \langle \widetilde{X}_t, m(X_t) \rangle + \tilde{S}_t \tag{A.11}$$

$\big(\tilde{S}_t = S_t + \alpha_t^2 \, E(\| W_t \|^2 \, | \, X^t)\big)$.

From $\mathbf{E} \, || \, X_0 \, ||^2 < \infty$, (A. 10), (A. 11), (C. 5) and (C.1):

$$\mathbf{E} \, || \, \widetilde{X}_t \, ||^2 < C_8 \tag{A.12}$$

From (A.10), (A.12), (A.7) and (C. 1)

$$\mathbf{E} \, | \tilde{S}_t | < \alpha_t (C_9 \, t^{-\eta\mu} + C_{10} \, t^{2\eta} \max_{t - t^\eta < \vartheta \leq t} \alpha_\vartheta). \tag{A.13}$$

Since $\delta_t([t^\eta])$ is a non-negative valued $\mathfrak{F}(X^t)$-measurable function, and all right side terms in (A.13) are, for given t, fixed positive reals, $\tilde{S}_t$ is also a non-negative valued $\mathfrak{F}(X^t)$-measurable function.

It follows from (A.13) and (3) that there exists such a $0 < \eta < 1$, for which

$$\sum_{t=0}^{\infty} \mathbf{E} \, \tilde{S}_t < \infty \, ,$$

by which we proved that all conditions imposed for $\tilde{S}_t$ hold. This completes the proof of Lemma 3.

We refer to the following two results on the convergence of random processes:

Theorem 5 (Braverman and Rozonoer [2, 8], Theorem III). Let, for any $t \in T$, U_t be a non-negative valued $\mathfrak{F}(X^t)$-measurable function, and $\mathbf{E} U_0 < \infty$. In addition, let $\widetilde{V} = \{\widetilde{V}_t, t \in T\}$ be a non-negativ valued sequence of random variables, related to $U = \{U_t, t \in T\}$ as follows:

$$\mathbf{E}(U_{t+1} \mid X^t) \leq (1 + \beta_t) \, U_t - \gamma_t \, \widetilde{V}_t + S_t^{(0)}, \tag{A.14}$$

for any $t > t_1 > 0$ where $\{\beta_t\}_{t=0}^{\infty}$ and $\{\gamma_t\}_{t=0}^{\infty}$ are a priori fixed real valued sequences $\beta_t \geq 0$, $\gamma_t > 0$,

$$\lim_{t\to\infty} \gamma_t = 0, \quad \sum_{t=0}^{\infty} \gamma_t = \infty, \quad \sum_{t=0}^{\infty} \beta_t < \infty.$$

$S_t^{(0)}$ is a non-negative valued $\mathcal{F}(X^t)$-measurable function, and $\Sigma_{t=0}^{\infty} \mathbf{E} S_t^{(0)} < \infty$. In addition assume that

$$\lim \mathrm{p}_{k\to\infty}\, U_{t_k} = 0 \tag{A.15}$$

for any such subsequence $\{t_k\}_{k=0}^{\infty}$ for which $\mathbf{P}(\lim_{k\to\infty} \tilde{V}_{t_k} = 0) = 1$. Then: $\mathbf{P}(\lim_{t\to\infty} U_t = 0) = 1$.

Theorem 6 (Braverman and Rozonoer [2, 8]. Theorem IV). Let, for any $t \in T$, U_t be a non-negative valued $\mathcal{F}(X_t)$-measurable function, and $\tilde{V}_t$ a non-negative valued sequence of random variables, meeting condition (A.14), for which

$$\tilde{V}_{t+1} \leq (1 + C_{11}\gamma_t)\tilde{V}_t + C_{12}\gamma_t + S_t^{(1)} \tag{A.16}$$

for any $t > t_2 > 0$, where $S_t^{(1)}$ is a non-negative valued $\mathcal{F}(X^t)$-measurable function and $\Sigma_{t=0}^{\infty} \mathbf{E}\, S_t^{(1)} < \infty$. Then $\mathbf{P}(\lim_{t\to\infty} \tilde{V}_t = 0) = 1$.

Remark. We refer, in the present context, with some slight extensions [2, 8] to Braverman's and Rozonoer's Theorem III, and to Theorem IV in the somewhat restricted form that (A.16) holds almost surely.

Proof of Theorem 1. Let us adopt the following substitutions: $U_t = ||X_t||^2$ $\tilde{V}_t = -V_t$, $V_t = \langle \tilde{X}_t, m(X_t)\rangle$, $\gamma_t = \alpha_t$, $S_t^{(0)} = \tilde{S}_t$., $\beta_t = 0$. Then (A.14) follows from Lemma 3, and (A.15) from (C.6). Thus the functionals $U_t = ||X_t||$, and $\tilde{V}_t = -V_t$ meet the conditions given in Theorem 5, from which $\mathbf{P}(\lim_{t\to\infty} ||\tilde{X}_t|| = 0) = 1$ follows. This completes the proof of Theorem 1.

Proof of Theorem 2. Let us follow the substitutions adopted in the proof of Theorem 1, and let $S_t^{(1)} \equiv 0$. Then (A.14) follows from Lemma 3.

From the property of conditional expectations, we already referred to, and (C.4), follows

$$\mathbf{E}(\langle \tilde{X}_t, W_t\rangle \mid Z^t)_{Z_\vartheta = X_t, \vartheta \leq t} = \langle \tilde{X}_t, m(X_t)\rangle.$$

Thus

$$G_t(f_{x_{t+1}}^t) = \langle \tilde{X}_{t+1}, m(X_{t+1})\rangle = G_{t+1}(f_{x_{t+1}}^{t+1}). \tag{A.17}$$

$(f_{x_t}^t = \{Z_\vartheta = X_t, \vartheta \leq t\})$.

Observe that $V_t = G_t(f_{x_t}^t)$, from which and (A.17):

$$|V_{t+1} - V_t| = |G_t(f_{x_{t+1}}^t) - G_t(f_{x_t}^t)|.$$

From this, (C.2a) and (C.3):

$$|V_{t+1} - V_t| \leq C_{13} \, ||X_{t+1} - X_t|| \, . \tag{A.18}$$

By (A.18), Lemma 2 and $||W_t|| < C_3$:

$$|V_{t+1} - V_t| < C_{14} \, \alpha_t$$

for any $t > t_3 > 0$. Thus, for the same t,

$$V_{t+1} \leq V_t + C_{14} \, \alpha_t \, ,$$

from which (and the aforementioned substitutions) (A.16) follows.

Thus Theorem 6 holds, and

$$\mathbf{P}(\lim_{t \to \infty} \int_{\Omega} |X_t(\omega) - \Theta(\omega)| \, Q(d\omega) = 0) = 1.$$

This completes the proof of Theorem 2.

Proof of Lemma 1. Considerations well known for Robbins Monro processes [8] may be carried over also to this case, provided we replace the regression function by $-m$. (We reproduce, here the proof only to provide a complete overview.)

We prove more then Lemma 1, viz., show that $\lim_{k \to \infty} ||\widetilde{X}_{t_k}(\lambda)|| = 0$ for any sequence $\{t_k\}_{k=0}^{\infty}$ and sample function $X(\lambda)$, for which $\lim_{k \to \infty} V_{t_k} = 0$.

Assume that, on the contrary of our assertion, there is some sequence $\{t_k\}_{k=0}^{\infty}$ and an elementary event λ, for which $\lim_{k \to \infty} V_{t_k}(\lambda) = 0$, howerever $\lim_{k \to \infty} || \widetilde{X}_{t_k}(\lambda) || \neq 0$. In this case, there exists a real $\tilde{\varepsilon} > 0$ and a sequence $\{t_{k_i}\}_{i=0}^{\infty} \subset \{t_k\}_{k=0}^{\infty}$ for which $||X_{t_{k_i}}(\lambda)|| > \tilde{\varepsilon}$ for all t_{k_i}. Then, however, it follows from (C.8) that $|V_{t_{k_i}}(\lambda)| \geq \tilde{q}(\tilde{\varepsilon}) \cdot (\tilde{q}(\tilde{\varepsilon})$ being for, any given $\tilde{\varepsilon}$, some fixed positive real) which contradicts the initial assumption. By this, the considered assertion and, therefore, also Lemma 1 is proved.

Proof of Theorems 3 and 4. Theorem 3 readily follows from Theorem 1 and Lemma 1, if one takes $m = -r$.

In Theorem 4, (C.1), (C.2a), (C.3), (C.4) follow from the definition of the regulator W. Concerning (C.7) we refer to the complete independence of ω_t, $t = 0, 1, \ldots$ and the notion of the conditional expectation. From this and for the specific form of W,

$$m(Z_t) = |\Delta| \int_{\Omega} r_t(\omega) \, K(., \omega) \, Q(d\omega). \tag{A.19}$$

($|\Delta|$ stands for the number of elements in Δ.) From (A.19), $r_t(\omega) = \operatorname{sign}(\Theta(\omega) - \xi(\omega))$, and the reproducing property of the kernel K, we obtain

$$\langle \xi - \Theta, m(\xi) \rangle = - |\Delta| \int_{\Omega} |\xi(\omega) - \Theta(\omega)| \, Q(d\omega),$$

from which (C.7) follows.

Thus, by adopting $\mathcal{H} = \mathcal{H}(K)$ and the chosen specific form for regulator W, all conditions imposed in Theorem 2 are met, and, therefore, $\mathbf{P}\left(\lim_{t \to \infty} \int_{\Omega} |X_t(\omega) - \Theta(\omega)| \, Q(d\omega) = 0\right) = 1$. This completes the proof of Theorem 4.

References

1. *Aĭzerman, M. A.—Braverman, E. M.—Rozonoer, L. I.:* Teoretičeskie osnovy metoda potencial'nyh funkcii v zadače ob obučenii avtomatov razdelniju vhodnyh situacii na klassy. Avtomatika i Telemehanika *6* (1964) 917—936.
2. *Braverman, É. M.—Rozonoer, L. I.:* Shodimost' slučainyh processov v teorii obučenija masin. I. Avtomatika i Telemehanika *1* (1969) 57—77 [see ibid. *2* (1970) 182].
3. *Csibi, S.:* On continuous stochastic approximation. Proc. Coll. Inform. Theory, Bolyai Math. Soc., Debrecen, (1967) pp. 89—100.
4. *Csibi, S.:* On embedding heuristics and including complexity constraints into convergent learning algorithms. To appear in Watanabe, M. S. (ed.) Frontiers of pattern recognition. Academic Press.

4a. *Csibi, S.*: Simple and compound processes in machine learning. Lecture Notes. International Center for Mechanical Sciences, Udine, Italy (to appear)

5. *Cypkin, Ja. Z.:* Osnovy teorii obučajuščihsja sistem. Nauka, (1970) Gl. 3, 6.
6. *Fu, K. S.:* Sequential methods in pattern recognition and machine learning. Academic Press, 1968.
7. *Gulyás, O.:* On extended potential function type learning algorithms and their convergence rate. Problems of Control and Information Theory, *1* (1972).
8. *Györfi, L.:* Iteration rules without memory in machine learning. Colloquia Series (Preprint), Telecommunication Research Institute, Budapest, 1970 (in Hungarian).
9. *Has'minskiĭ, R. Z.:* Ustoĭčivost' sistem differencial'nyh uravnenii pri slučainih vozmuščenijah ih parametrov. Nauka, 1969, Gl. VII., §6
10. *Has'minskiĭ, R. Z.—Nevel'son, M. B.:* O nepreryvnyh procedurah stohastičeskoĭ approksimacii. Problemy Peredači Informacii (to be published).
11. *Sakrison, D. J.*: A continuous Kiefer—Wolfowitz procedure for random processes. Ann. Math. Statist. *35* (1964) 590—599.
12. *Saridis, G. N.:* Learning applied to successive approximation algorithms. IEEE Trans. on System Science and Cybernetics, SSC-*6*, (1970) 97—103.
13. *Wolverton, Ch. T.—Wagner, T. J.:* Asymptotically optimal discriminant functions for pattern classification. IEEE Trans. on Information Theory, *IT-15*, (1969) 258—265.

О машинном обучении при итерационных правилах с памятью

Щ. ЧИБИ

Будапешт

Резюме

В работе даются ограничения, при которых широкий класс управляемых машинных методов обучения сходится в случае применения итерационных правил с памятью. Точнее, рассматриваются итерационные правила, обладающие слабой (или даже конечной во времени) памятью относительно функций дискриминации, предложенных в прежних моментах. Такие методы предоставляют богатые возможности для применения более эффек-

тивных шагов обработки в сходящихся алгоритмах и тем самым для сокращения времени, необходимого для обучения. Доказываются соотношения, способствующие конструированию итерационных правил такого рода. Применение сообщаемых результатов, например для сохраняющей сходимость модификации сходящихся обучающихся алгоритмов, и другие аналогичные вопросы будут рассмотрены в дальнейшей работе.[4]

Пусть будет $\Omega = \{\omega\}$ пространство представленных симптомов и $\Theta = \{\Theta(\omega), \omega \in \Omega\}$ — функция дискриминации, для которой надо дать оценку, напр. на основании произведенного в моментах времени $t = 0,1, \ldots$ обучения. Пусть будет $Z_t = \{Z_t(\omega), \omega \in \Omega\}$ оценка, предлагаемая в моменте t. Предположим, что искомая функция дискриминации Θ и возможные оценки Z_t являются элементами сепарабельного Гильбертова пространства $\mathcal{H}$. Точнее предположим, что $\Theta \in \mathcal{A} \subset \mathcal{H}$, где $\mathcal{A}$ не является обязательно подпространством.

Рассматривается следующий класс итеративных обучающихся алгоритмов. Пусть будет в любом моменте $t = 0,1, \ldots$ $Z_t = X_t$, где итерационный процесс $X = \{X_t, t = 0,1,\ldots\}$, исходя из произвольной начальной оценки $X_0 \in \mathcal{A}$ ($\mathbf{E}\,\|X_0\|^2 < \infty$), создается с помощью итерационного правила (1) $X_{t+1} = \Phi(X_t + \alpha_t W_t)$. (В данном случае существенно отличать созданную произвольно оценку Z_t от выработанной специально по (1) оценки). Здесь α представляет собой по существу последовательность коэффициентов, соответствующую обычным ограничениям случайного приближения. Φ: $\mathcal{H} \to \mathcal{A}$ — усечение, с помощью которого X можно удерживать в пределах выбранного надлежащим образом множества $\mathcal{A} \subset \mathcal{H}$. Множество $\mathcal{A}$ может быть любым параллелепипедом в $\mathcal{H}$. Φ — равномерно уменьшающее норму отображение, имея в виду следующий шаг оценки и любой элемент множества $\mathcal{A}$. Обучение в моменте t влияет на итерацию через регулятор W_t.

Сначала даются хорошо известные примеры для регулятора W. Затем устанавливаются ограничения, которым регулятор W должен удовлетворять для обеспечения сходимости.

На регулятор W наложены слабые и сильные ограничения. Слабые ограничения являются существенными с точки зрения сходимости процесса обучения, но их выполнение не связано с особыми трудностями. Такие ограничения: (С. 2) — слабая (или конечная во времени) память и (С. 3) — чувствительность $\mathbf{E}(\langle Z_t - \Theta, W_t\rangle \mid Z^t)$, не превышающая определенную границу в отношении всей совокупности $Z^t = \{Z_\vartheta, \vartheta \leq t\}$ ранее предложенных оценок. Здесь $\langle \cdot \rangle$ означает внутреннее произведение. В то же время сильные ограничения или суживают круг исследуемых проблем, или являются ограничениями по конструкции. (Таковы: (С. 4) — независимость от времени условного математического ожидания m регулятора W_t, взятого в предположении тождественных прежних оценок (что означает условную стационарность своего рода) и ограничения по конструкции более общего или более специфичного характера (С. 5 . . . С. 7).

Для регуляторов W такого рода доказываются теоремы сходимости. Теоремы 1 и 2 относятся к регуляторам общего вида, теорема 3 относится к итерационным процессам, управляемым расходами, а теорема 4 — к итерационным процессам, создаваемым на основании потенциальных функций. Из этих теорем хорошо видно, что, оставаясь в пределах регуляторов, удовлетворяющих нашим требованиям, сходимость итерации зависит по существу только от свойств момента m, т. е. от тех свойств регулятора W, которые при тождественных прежних оценках появляются, так же и в случае $Z_\vartheta = \text{const}$, $\vartheta \leq t$. Это распознавание дает весьма широкие возможности для расширения круга сходящихся процедур обучения и даже для внедрения эвристических шагов обработки информации.

Существо наших доказательств заключается в том, что проблемы обучения, обладающие памятью такого рода, с помощью леммы 3 сводятся к теореме Бравермана и Розоноэра, относящейся к случайным процессам.

Это приводит к значительному расширению гарантированно абсолютно стабильных процессов, создавая новые возможности для отыскания функционалов U и V типа Ляпунова.

Sándor Csibi
Telecommunication Research Institute
Budapest 2, Gábor Áron út 65, Hungary

Problems of Control and Information Theory, Vol. 1 (1), pp. 51–64 (1972)

ON EXTENDED POTENTIAL FUNCTION TYPE LEARNING ALGORITHMS AND THEIR CONVERGENCE RATE

O. GULYÁS

Budapest

(Received February 23, 1971)

In this paper the potential function type learning algorithm is generalized. As potential function, an arbitrary positive definite function is chosen. Concerning the discrimination function to be produced by the algorithm it is assumed that this is an element of Hilbert space with a reproducing kernel generated by the potential function. It is demonstrated that the convergence theorems of the algorithms introduced in [1–6] (which are special cases of this) remain true also in the general case. Two theorems for the convergence rate of the algorithms are proved and some problems of the choice of potential functions are investigated.

The problem

In this paper we are concerned with learning discrimination functions with a teacher.

Definition 1. Given a space X of points $\mathbf{x}$, and a pair of disjoint subsets A and B in X. Viz.,

$$A \cup B \subset X \tag{1}$$

$$A \cap B = 0 \tag{2}$$

We call the real-valued function $f = \{f(\mathbf{x}), \mathbf{x} \in X\}$ discrimination function with respect to $\{A, B\}$ provided

$$\operatorname{sign} f(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in A, \\ -1 & \text{if } \mathbf{x} \in B \end{cases} \tag{3}$$

Definition 2. We call, in the present paper, the sequence $\Pi = \{\mathbf{x}^1, \mathbf{x}^2, \ldots \mathbf{x}^n \ldots\}$ of points in X sequence of learning samples if

a) Π is a totally independent sequence of points in $A \cup B$

b) the probability density function of every $\mathbf{x}^i$ is $p(\mathbf{x})$ for all i, $[p(\mathbf{x}) = 0$ if $\mathbf{x} \in (A \cup B)^c]$, (where superscript C stands for the complement.)

Definition 3. We call the sequence $\hat{\Pi} = \{\hat{\mathbf{x}}^1, \hat{\mathbf{x}}^2, \ldots \hat{\mathbf{x}}^n \ldots\}$ of random variables $\hat{\mathbf{x}}^i = f(\hat{\mathbf{x}}^i)$ the sequence of labels. Here

$\mathbf{x}^i$ is the i-th learning sample

$f(\mathbf{x})$ is the value the discrimination function f takes at $\mathbf{x}$.

Let Π_n and $\hat{\Pi}_n$ be the first n members of the sample and label sequences, respectively. Viz.,

$$\Pi_n = \{\mathbf{x}^1\, \mathbf{x}^2, \ldots \mathbf{x}^n\} \tag{4}$$

$$\hat{\Pi}_n = \{\hat{\mathbf{x}}^1, \hat{\mathbf{x}}^2, \ldots \hat{\mathbf{x}}^n\} \tag{5}$$

The purpose of learning to devise a function $f_n = \{f_n(\mathbf{x}),\ \mathbf{x} \in X_n\}$, which depends on x, Π_n, $\hat{\Pi}_n$, and tends as $n \to \infty$, is some appropriate sense (to be specified in the sequel) to a discrimination function f in accordance of what has been taught us to n by the teacher.
Viz.,

$$f_n(\mathbf{x}) = f_n(\mathbf{x}; \Pi_n, \hat{\Pi}_n) \simeq f(\mathbf{x})$$

for any $\mathbf{x} \in A \cup B$.

Prerequisites in the theory of potential function type algorithms ([1—6])

Given $X, \{A, B\}, p(\mathbf{x})$; Π, $\hat{\Pi}$, assume there exists a partition function $f(\mathbf{x})$ of the form

$$f(\mathbf{x}) = \sum_{i=1}^{N} C_i\, \varphi_i(\mathbf{x})\,. \tag{6}$$

Here $\{\varphi_i(\mathbf{x}),\ \mathbf{x} \in X\}$ denotes a system of linearly independent functions, N an integer and

$$\mathbf{C} = \{C_1, C_2, \ldots C_n\} \text{ and } N\text{-tuple of reals.}$$

Let us consider learning algorithms of the following form:

$$f_0(\mathbf{x}) \equiv 0$$

and

$$f_n(\mathbf{x}) = \sum_{k=1}^{n} r_{k-1}\,(\mathbf{x}^k)\; K(\mathbf{x}, \mathbf{x}^k)\,. \tag{7}$$

In Aizerman's — Braverman's and Rozonoer's fundamental theory [1] the function $K = \{K(\mathbf{x}, \mathbf{y}),\ \mathbf{x}, \mathbf{y} \in X\}$ is defined by

$$K(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{N} \varphi_i(\mathbf{x})\; \varphi_i(\mathbf{y}) \tag{8}$$

and is called potential function. It is also assumed, that $|\, K(\mathbf{x}, \mathbf{y})\, | < L$

Specifically, for unambiguous teaching, the following choices of the coefficient r_k are well known:

$$r_k^{\Theta}(\mathbf{x}) = \Theta[f(\mathbf{x}) - f_k(\mathbf{x})] \text{ where } 0 < \Theta < 2/L \tag{9}$$

$$r_k^{\gamma}(\mathbf{x}) = \gamma_k \operatorname{sign}[f(\mathbf{x}) - f_k(\mathbf{x})] \quad \text{where}$$

$$\gamma_k > 0; \sum_{k=1}^{\infty} \gamma_k = \infty; \sum_{k=1}^{\infty} \gamma_k^2 < \infty, \tag{10}$$

$$r_k^{\mu}(\mathbf{x}) = \operatorname{sign} f(\mathbf{x}) - \operatorname{sign} f_k(\mathbf{x}), \tag{11}$$

Observe that, specifically, for r_k^{μ} the teacher has, of course, not to present (together with $\mathbf{x}^i/f(\mathbf{x}^i)$) itself but only sign $f(\mathbf{x}^i)$.

For (10)—(12) the most simple convergence theorems are as follows ([1]—[6]): As $n \to \infty$,

$$M_x \mid f(\mathbf{x}) - f_n^{\Theta}(\mathbf{x}; \Pi_n, \hat{\Pi}_n) \mid^2 \to 0 \qquad \text{in probability} \tag{12a}$$

$$M_x \mid f(\mathbf{x}) - f_n^{\gamma}(\mathbf{x}; \Pi_n, \hat{\Pi}_n) \mid \to 0 \qquad \text{in probability} \tag{12b}$$

$$M_x \mid \operatorname{sign} f(\mathbf{x}) - \operatorname{sign} f_n^{\mu}(\mathbf{x}; \hat{\Pi}_n, \Pi_n) \mid^2 \to 0 \quad \text{in probability} \tag{12c}$$

(Here M_x denotes the expectation with respect to $\mathbf{x}$, in (12c) $|f(\mathbf{x})| > \varepsilon > 0$ is also assumed for some $\varepsilon > 0$ and any $\mathbf{x} \in A \cup B$.)

Prerequisites in the theory of Reproducing Kernel Hilbert Spaces [7—11]

The results to be presented are based on the theory of reproducing kernel Hilbert-spaces. We recollect here, only the notions and theorems we directly refer to in the sequel.

Definition 4. Hilbert-space of functions $\{f(s), s \in S\}$ is a reproducing kernel Hilbert-space (RKHS), if there exists a function $\{K(s, t); (s, t) \in S \times S\}$, such that

$$K(\cdot, t) \in H \qquad \forall t \in S, \tag{13}$$

$$(f(\cdot), K(\cdot, t)) = f(t) \quad \forall f \in H. \tag{14}$$

$K(s, t)$ is called the reproducing kernel of the space.

The reproducing kernel defines the RKHS unambiguously in the sense of the following Lemmas 1—3.

Lemma 1 [9]

An RKHS may process only one reproducing kernel

Proof: Assume, that $K(s, t)$ and $K'(s, t)$ are two reproducing kernels. Then, for any $f \in H$:

$$(f(s), K(s, t)) = (f(s), K'(s, t)) = f(t)$$

Thus: $$(f(s), K(s, t) - K'(s, t)) = 0$$

and, therefore: $K(s, t) - K'(s, t) \perp f(s) \quad \forall f \in H$

However, this may hold only if $K(s, t) = K'(s, t)$, which completes the proof.

Lemma 2 [9]

The function-set $\{K(s, t), t \in S\}$ *is a basis in the RKHS generated by* K.

Proof: — If $f(s) \perp K(s, t)$ for some $f \in H$ and t, then:

$$(f(s), K(s, t)) = f(t)$$

$$(f(s), K(s, t)) \equiv 0$$

Thus: $$f(t) \equiv 0$$

Lemma 3 [9]. — *A reproducing kernel can be associated only with single RKHS*

Proof: — Let us suppose, that H_1 and H_2 are two RKHS-s having the same kernel $K(s, t)$. According to Lemma 2, H_1 and H_2 are the closures of convex linear manifolds with respect K

However

$$(K(s, t), K(s, u))_1 = (K(s, t), K(s, u))_2 = K(s, t),$$

from which

$$|| \Sigma c_i K(s, t_i) ||_1 = || \Sigma c_i K(s, t_i) ||_2,$$

from which Lemma 3 follows. (The subscripts refer to H_1 and H_2, respectively.)

Next we show how are reproducing kernels related to positive definite functions.

Definition — 5. A function $K = K(s, t)$; $(s, t) \in S \times S$ of two variables in *positive definite*, if for any positive integer n, any n-tuple $(z_1, \ldots z_n)$ *of complex numbers* and any n-tuple $(t_1, t_2, \ldots t_n)$ of reals

$$\sum_{i=1}^{n} \sum_{j=1}^{n} K(t_i, t_j)\, z_i \bar{z}_j \geq 0$$

holds and (15) equals 0 only if $z_i = 0$ for all $i = 1, 2, \ldots n$

Lemma 4 [9]. — The kernel $K = \{K(s, t); (s, t) \in S \times S\}$ of any *RKHS is positive definite*

Proof: — For any positive integer n, $\{t_i\}_{i=1}^n$, $t_i \in S$ and $\{z_i\}_{i=1}^n$

$$0 \leq \left\| \sum_{i=1}^{n} K(s, t_i)\, z_i \right\|^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \big(K(s, t_i),\ K(s, t_j)\big)\, z_i \bar{z}_j = \sum_{i=1}^{n} \sum_{j=1}^{n} K(t_j, t_i)\, z_i \bar{z}_j$$

holds, from which Lemma 4 follows.

Moore's Theorem — [9]. For any positive definite function K there exists one and only one RKHS, the kernel of which is K. (For the proof of Lemma 4 see [9].)

Consider functions of the following form:

$$f(s) = \sum_{i=1}^{n} c_i\, K(s, t_i) \tag{15}$$

(n denotes a positive integer, $\{c_i\}$ a sequence of reals, $\{t_i\}$ a tuple of n elements from S.)

We obtain in this way a linear space. Let us generate an inner product by

$$(f, g) = \sum_{i=1}^{n} \sum_{j=1}^{m} c_i\, d_j\, K(t'_j, t_i),$$

having

$$g(s) = \sum_{j=1}^{m} d_j\, K(s, t'_j)\,.$$

(The bar stands for the complex conjugate.) Let us take the closure of the convex linear manifold of (15). The space, obtained in this way is an RKHS with kernel $K(s, t)$.

We will denote, in the sequel, by $H(K)$ the reproducing kernel Hilbert space generated by a positive definite function K.

How are convergence properties related in an RKHS is shown by

Lemma 5 [9]. — in RKHS convergence in norm implies pointwise convergence. If K is uniformly bounded, pointwise uniform convergence holds.

Proof: — Obviously

$$| f(\mathbf{s}) - f_n(\mathbf{s}) | = | \big(f(\mathbf{u}), K(\mathbf{u}, \mathbf{s})\big) - \big(f_n(\mathbf{u}), K(\mathbf{u}, \mathbf{s})\big) | =$$

$$= | \big(f(\mathbf{u}) - f_n(\mathbf{u}), K(\mathbf{u}, \mathbf{s})\big) | \leq || f(\mathbf{u}) - f_n(\mathbf{u}) ||\, || K(\mathbf{u}, \mathbf{s}) ||\,,$$

which completes the proof.

Example. Any finite dimensional H Hilbert space of complex valued functions is an RKHS. Let $\{\varphi_i(\mathbf{x})\}_{i=1}^N$ be a *complete* orthonormal system in H. Then only the functions of the form $f(\mathbf{x}) = \sum_{i=1}^{N} c_i\, \varphi_i(\mathbf{x})$ belong to H.

Given y, consider

$$K(\cdot, \mathbf{y}) = \sum_{i=1}^{N} \varphi_i(\mathbf{x})\, \varphi_i(\mathbf{y}),\ \mathbf{x} \in X.$$

Obviously:

$$K(\mathbf{x}, \mathbf{y}) \in H \quad \forall\ \mathbf{y} \in X$$

and

$$\big(f(\mathbf{x}),\ K(\mathbf{x}, \mathbf{y})\big) = \sum_{i=1}^{N} \sum_{j=1}^{N} c_i\, \varphi_j(\mathbf{y})\, \big(\varphi_i(\mathbf{x}),\ \varphi_j(\mathbf{x})\big) = f(\mathbf{y})\,.$$

Further significant examples for RKHS-s may be found in Hajek [13], Parzen [7—8] and Kailath [11].)

Extending potential function type learning algorithms

We (i) extend, in what follows, the algorithms given by (9)—(11) and prove, that the convergence theorem [12] hold even under more general condition, (ii) prove, that the convergence bounds, given in [5], hold even under more general condition (iii) and make some remarks concerning the choice of potential functions.

Theorem 1. Given a positive definite function $K = \{K(\mathbf{x}, \mathbf{y}),\ (\mathbf{x}, \mathbf{y}) \in X \times X\}$.

Assume, that $| K(\mathbf{x}, \mathbf{y}) | \leq L$ and $f(\mathbf{x}) \in H(K)$. Then the estimate $f_n = \{f_n(\mathbf{x}),\ \mathbf{x} \in X\}$ defined by (9)—(11), with

$$f_0(\mathbf{x}) \equiv 0 \tag{16}$$

$$f_n(\mathbf{x}) = \sum_{k=1}^{n} r_{k-1}(\mathbf{x}^k)\, K(\mathbf{x}, \mathbf{x}^k) \tag{17}$$

tends, as $n \to \infty$, to $f(\mathbf{x})$, in the following sense

$$M_x\, | f(\mathbf{x}) - f_n^{\Theta}(\mathbf{x}) |^2 \to 0, \qquad \text{in probability} \tag{18}$$

$$M_x\, | f(\mathbf{x}) - f_n^{\gamma}(\mathbf{x}) | \to 0, \qquad \text{in probability} \tag{19}$$

$$M_x\, | \operatorname{sign} f(\mathbf{x}) - \operatorname{sign} f_n^{\mu}(\mathbf{x}) |^2 \to 0, \qquad \text{in probability} \tag{20}$$

((18)—(20) refer of the cases when f equals f_n^{Θ}, f_n^{γ} and f_n^{μ} respectively, the definition of which is given by (9)—(11).)

Proof: (See Appendix 1.)

Remark 1. Theorem 1 obviously extends the scope of the potential function type algorithms defined in [1]. It may be readily shown that

$$K = \Big\{K\,(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{N} \varphi_i(\mathbf{x})\, \varphi_i(\mathbf{y}),\ (\mathbf{x}, \mathbf{y}) \in X \times X\Big\}$$

is a positive definite function and $H(K)$ is a function space of the form

$$f(\mathbf{x}) = \sum_{i=1}^{N} c_i \varphi_i(\mathbf{x}).$$

Theorem 2. Assume, that a real $r > 0$ exists, for wich

$$r\, K(\mathbf{u}, \mathbf{v}) \leqq M_x\, K(\mathbf{x}, \mathbf{u})\, K(\mathbf{x}, \mathbf{v}) \tag{21}$$

for all $(\mathbf{u}, \mathbf{v}) \in X \times X$.

If, in addition to (21), conditions in Theorem 1 also hold, we have (for any $1 - \mathrm{ar} < 1$)

$$\frac{M \| f(\mathbf{x}) - f_n^{\ominus}(\mathbf{x}) \|^2}{\| f(\mathbf{x}) \|^2} \leq (1 - ra)^n, \tag{22}$$

$$\frac{M | f(\mathbf{x}) - f_n^{\ominus}(\mathbf{x}) |^2}{M | f(\mathbf{x}) |^2} \leq \frac{L}{r} (1 - ra)^n, \tag{23}$$

$$f_n(\mathbf{x}) \to f(\mathbf{x}) \quad \text{if } n \to \infty \text{ (almost everywhere)} \tag{24}$$

Proof: (See Appendix 2)

Theorem 3. If in addition to the conditions in Theorem 2 we have, for some $\lambda > 0$ and $1\, n > n_o$

$$\left(\frac{\gamma_n}{\gamma_{n+1}}\right)^{\lambda} (1 - r\, \gamma_{n+1}) \leq 1, \tag{25}$$

$$\sum_{k=1}^{\infty} \gamma_k^{2-\lambda} < \infty, \tag{26}$$

then

$$M_x \| f(\mathbf{x}) - f_n^{\gamma}(\mathbf{x}) \|^2 \leq C_1\, \gamma_n^{\lambda}, \tag{27}$$

$$M_x \ | f(\mathbf{x}) - f_n^{\gamma}(\mathbf{x}) | \leq C_2\, \gamma_n^{\lambda/2} \tag{28}$$

(C_1, C_2 denote given positive constants)

Proof: (See Appendix 3)

Remark 2. Obviously from (26) the property $\Sigma\, \gamma_k^2 < \infty$ also follows and (25) implies $\Sigma\, \gamma_k = \infty$.

Remark 3. Theorems 2 and 3 on error bounds are the extensions of the corresponding theorems in [5]. Further theorems in [5], concerning the algorithms, given by (9)—(11), may also be extended in a similar way.

Remark 4. Define the relation $\preccurlyeq$ between two positive definite functions as follows:

$$K_1(\mathbf{x}, \mathbf{y}) \preccurlyeq K_2(\mathbf{x}, \mathbf{y}) , \tag{29}$$

if

$$H(K_1) \subset H(K_2) . \tag{30}$$

Obviously if the potential function is K_2 and if $f(x) \in H(K_1)$ then Theorems 1—3 hold and

$$f_n(\mathbf{x}) = \sum_{k=1}^{n} r_{k-1}(\mathbf{x}^k)\, K(\mathbf{x}, \mathbf{x}^k) \tag{31}$$

converges, as $n \to \infty$, to $f(\mathbf{x})$.

Obviously if the potential function is K_1 and $f(\mathbf{x}) \in H(K)$, and the orthogonal decomposition of f with respect to $H(K_1)$ is

$$f(\mathbf{x}) = P_1 f(\mathbf{x}) + P_2 f(\mathbf{x}) .$$

$$P_1 f(\mathbf{x}) \in H(K_1) \quad \text{and} \quad P_1 f(\mathbf{x}) \perp P_2 f(\mathbf{x}) , \text{ then} \tag{32}$$

$$|| P_2 f(\mathbf{x}) || \leq || f(\mathbf{x}) - f_n(\mathbf{x}) || . \tag{33}$$

(Observe that $f_n(\mathbf{x}) \in H(K_1)$ per definition.)

Remark 5. Theorem 1 offers further possibilities for finding appropriate potential functions. E.g., let X be the real interval [a, b]. Assume that A resp. B may be separated by an almost everywhere continuously differentiable function (e. g. by a polynomial).

Observe that in this very case a discrimination function $f \in H(K)$ exists, provided

$$K(t, s) = \frac{1}{2\beta} e^{-\beta |t-s|} \quad \text{and} \quad f_n \to f \tag{34}$$

if we replace in (17) $K(\mathbf{x}, \mathbf{x}^k)$ by a potential function according to (34).

The inner product specifically for this $H(K)$ may be written as

$$(h, g) = \int_a^b [h'(t) + \beta\, h(t)]\, [g'(t) + \beta\, g(t)]\, dt + 2\beta\, h(a)\, g(a) \tag{35}$$

which obviously follows from the fact that $H(K)$ is now, specifically, the space of almost everywhere continuously differentiable functions [7]—[8].

Appendix 1 — Proof of Theorem 1

We adopt the ideas given in [4] to our more general context. Let us introduce the following notation

$$\alpha_n = || f(\mathbf{x}) - f_n(\mathbf{x}; \Pi_n, \hat{\Pi}_n) ||^2 \tag{A1.1}$$

(Observe that f_n as well as f is a member of $H(K)$.)

Because of the reproducing property:

$$\big(K(\mathbf{x}, \mathbf{x}^{n+1}),\ K(\mathbf{x}, \mathbf{x}^{n+1})\big) = K(\mathbf{x}^{n+1}, \mathbf{x}^{n+1}) , \tag{A1.2}$$

and

$$\big(f(\mathbf{x}) - f_n(\mathbf{x}), K(\mathbf{x}, \mathbf{x}^{n+1})\big) = f(\mathbf{x}^{n+1}) - f_n(\mathbf{x}^{n+1}) , \tag{A1.3}$$

Therefore:

$$\alpha_{n+1} = || f(\mathbf{x}) - f_{n+1}(\mathbf{x}) ||^2 = || f(\mathbf{x}) - f_n(\mathbf{x}) - r_{n+1}(\mathbf{x}^{n+1}) K(\mathbf{x}, \mathbf{x}^{n+1}) ||^2 =$$

$$= || f(\mathbf{x}) - f_n(\mathbf{x}) ||^2 - 2r_{n+1}(\mathbf{x}^{n+1}) \big(f(\mathbf{x}) - f_n(\mathbf{x}), K(\mathbf{x}, \mathbf{x}^{n+1})\big) +$$

$$+ r_{n+1}^2(\mathbf{x}^{n+1}) \big(K(\mathbf{x}, \mathbf{x}^{n+1}), K(\mathbf{x}, \mathbf{x}^{n+1})\big).$$

From which we obtain:

$$\alpha_{n+1} = \alpha_n - 2r_{n+1}(\mathbf{x}^{n+1}) [f(\mathbf{x}^{n+1}) - f_n(\mathbf{x}^{n+1})] + r_{n+1}^2(\mathbf{x}^{n+1}) K(\mathbf{x}^{n+1}, \mathbf{x}^{n+1}) \tag{A1.4}$$

This relation is the same as (20) in [4] and from here on the proof exactly follows that of [4].

Appendix 2 — Proof of Theorem 2

2a. We prove, that

$$r || f ||^2 \leq M_x | f(\mathbf{x}) |^2 \tag{A2.1}$$

for any $f \in H$.

Consider the set of functions in H-space which are of the form

$$f(\mathbf{x}) = \sum_{i=1}^{n} c_i K(\mathbf{x}, \mathbf{x}^i) . \tag{A2.2}$$

For these

$$|| f(\mathbf{x}) ||^2 = (f, f) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \big(K(\mathbf{x}, \mathbf{x}^i), K(\mathbf{x}, \mathbf{x}^j)\big) = \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j K(\mathbf{x}^j, \mathbf{x}^i) \tag{A2.3}$$

and

$$M_x |f(\mathbf{x})|^2 = M_x \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j K(\mathbf{x}, \mathbf{x}^i) K(\mathbf{x}, \mathbf{x}^j) =$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j M_x K(\mathbf{x}, \mathbf{x}^i) K(\mathbf{x}, \mathbf{x}^j) . \tag{A2.4}$$

(A2.1) follows, for such f, from (2.1) by comparing the right sides in (A2.3) and (A2.4)

Next let us expand (A2.1) for any f in H. Observe that any $f \in H$ is the limit of functions of the form (A2.2). Since, by Lemma 5, from convergence in RKHS — norm pointwise convergence follows, we get, as $n \to \infty$,

$$f_n(\mathbf{x}) \to f(\mathbf{x}) \tag{A2.5}$$

for all $\mathbf{x} \in H$.

By (A2.5) and Lemma 5

$$|f_n(\mathbf{x})| \leq \sqrt{K(\mathbf{x}, \mathbf{x})} \, \| f_n \| \leq C \tag{A2.6}$$

that's why thus we may use Lebesgue's theorem and write

$$M_x \, | \, f_n(\mathbf{x}) \, |^2 \to M_x \, | \, f(\mathbf{x}) \, |^2 \tag{A2.7}$$

Observe that

$$|| \, f_n(\mathbf{x}) \, || \to || \, f(\mathbf{x}) \, || \tag{A2.8}$$

From (A2.5) and (A2.8) we finally get:

$$r \, || \, f \, ||^2 \leq M_x \, | \, f(\mathbf{x}) \, |^2 \tag{A2.9}$$

2b. For the sake of brevity shall adopt, in what follows, the following notations

$$\beta_n = M_x \, | \, f(\mathbf{x}) - f_n(\mathbf{x}; \Pi_n, \hat{\Pi}_n) \, |^2 \tag{A2.10}$$

Then we may replace (A2.9) by:

$$r \, M(\alpha_n) \leq M(\beta_n) \tag{A2.11}$$

2c. Let, specifically, $r_n = r_n^{\Theta}$. Then

$$\alpha_{n+1}^{\Theta} = \alpha_{\Theta}^{n} - 2\Theta \, [f(\mathbf{x}^{n+1}) - f_n^{\Theta}(\mathbf{x}^{n+1})]^2 + \Theta^2 \, [f(\mathbf{x}^{n+1}) -$$

$$- f_n(\mathbf{x}^{n+1})]^2 \, K(\mathbf{x}^{n+1}, \mathbf{x}^{n+1}) \tag{A2.12}$$

from which we obtain:

$$M_{x^{n+1}}(\alpha_{n+1}^{\Theta}) \leq \alpha_n^{\Theta} - 2\Theta\,\beta_n + \Theta^2 L\beta_n, \tag{A2.13}$$

$$M(\alpha_{n+1}^{\Theta}) \leq M(\alpha_n^{\Theta}) - a\,M(\beta_n)\,. \tag{A2.14}$$

Here

$$a = 2\Theta - \Theta^2 L\,. \tag{A2.15}$$

2d. Iterating (A2.13) n-time, we get:

$$M(\alpha_{n+1}^{\Theta}) \leq M(\alpha_n^{\Theta})\,(1 - ar) \leq M(\alpha_o^{\Theta})\,(1 - ar)^n \tag{A2.16}$$

observe that

$$M(\alpha_0^{\Theta}) = M\,||\,f(\mathbf{x}) - f_0(\mathbf{x})\,||^2 = M||\,f(\mathbf{x})\,||^2 \tag{A2.17}$$

From which we get

$$\frac{M\,\|f(\mathbf{x}) - f_n(\mathbf{x})\|^2}{M\,\|f(\mathbf{x})\|^2} \leq (1 - ar)^n\,, \tag{A2.18}$$

by which (23) in Theorem 2 is proved.

2e. The proof of (24) directly follows from (23), since

$$|\,f(\mathbf{x}) - f_n(\mathbf{x})\,|^2 \leq L\,||\,f(\mathbf{x}) - f_n(\mathbf{x})\,||^2 \tag{A2.19}$$

and therefore

$$M(\beta_n) < L\,M(\alpha_n^{\Theta}) \tag{A2.20}$$

From (A.2.18) and (A.2.20) we get:

$$\frac{M\,|f(\mathbf{x}) - f_n(\mathbf{x})\,|^2}{M\,|f(\mathbf{x})\,|^2} \leq L\,\frac{M(\alpha_n^{\Theta})}{rM\,||f(\mathbf{x})\,||^2} \leq \frac{L}{r}\,(1 - ar)^n \tag{A2.21}$$

2f. — In order to prove (24) we write, by using (A2.21):

$$\sum_{k=1}^{\infty} M(\beta_k) = \sum_{k=1}^{\infty} M\,|f(\mathbf{x}) - f_n|^2 \leq \frac{LM\,|f|^2}{r} \sum_{k=1}^{\infty} (1 - ar)^k = \frac{LM\,|f|^2}{ar^2} < \infty. \tag{A2.22}$$

By Markov's inequality

$$\sum_{k=1}^{\infty} P\big(|f(\mathbf{x}) - f_k(\mathbf{x})\,|^2 > \varepsilon\big) < \infty\,, \tag{A2.23}$$

for any $\varepsilon > 0$. Thus, by the Borel-Cantelli lemma, we get

$$|\,f(\mathbf{x}) - f_n(\mathbf{x})\,| \to 0\,, \tag{A2.24}$$

as $n \to \infty$, almost surely, which is (24). This completes the proof of Theorem 2).

Appendix 3 — Proof of Theorem 3

3a. — First we prove, that

$$| f(\mathbf{x}) - f_n^{\gamma}(\mathbf{x}) | < A \,, \tag{A3.1}$$

(A being, for any n the same constant). It is sufficient to show, that

$$|| f(\mathbf{x}) - f_n^{\gamma}(\mathbf{x}) || \leq A \,, \tag{A3.2}$$

since

$$| f(\mathbf{x}) - f_n^{\gamma}(\mathbf{x}) | \leq L \, || f(\mathbf{x}) - f_n^{\gamma}(\mathbf{x}) || \,, \tag{A3.3}$$

In order to prove this, we replace in (A.1.4.) r_n by r_n^{γ}. I.e.,

$$\alpha_{n+1}^{\gamma} = \alpha_n^{\gamma} - 2\gamma_{n+1} | f(\mathbf{x}^{n+1}) - f_n^{\gamma}(\mathbf{x}^{n+1}) | + \gamma_{n+1}^2 K(\mathbf{x}^{n+1}, \mathbf{x}^{n+1}) \,. \tag{A3.4}$$

By overbounding (A3.4) and rearrangement:

$$\alpha_{n+1}^{\gamma} - \alpha_n^{\gamma} \leq L\gamma_{n+1}^2 \,.$$

Let $\Gamma = \sum_{i=1}^{\infty} \gamma_i^2$ By this:

$$\sum_{i=0}^{s} (\alpha_{i+1}^{\gamma} - \alpha_i^{\gamma}) = \alpha_{s+1}^{\gamma} - \alpha_o^{\gamma} = L \sum_{i=1}^{s} \gamma_i^2 \leq L\Gamma, \tag{A3.5}$$

and

$$\alpha_{s+1}^{\gamma} \leq L\Gamma + \alpha_o^{\gamma} \leq C_1, \tag{A3.6}$$

from which we obtain

$$\alpha_{n+1}^{\gamma} = || f(\mathbf{x}) - f_{n+1}^{\gamma}(\mathbf{x}) ||^2 < A \,, \tag{A3.7}$$

and

$$| f(\mathbf{x}) - f_n^{\gamma}(\mathbf{x}) | < A \tag{A3.8}$$

Let

$$\delta_n = M_x | f(\mathbf{x}) - f_n(\mathbf{x}, \Pi_n, \hat{\Pi}_n) | \tag{A3.9}$$

3b. — From (A3.8) and (A3.9)

$$\beta_n = M_x | f(\mathbf{x}) - f_n^{\gamma}(\mathbf{x}) |^2 \leq AM_x | f(\mathbf{x}) - f_n(\mathbf{x}) | = A\delta_n \tag{A3.10}$$

and from this and (A2.11)

$$r M(\alpha_n) \leq M(\beta_n) \leq AM(\delta_n) \tag{A3.11}$$

from which we obtain

$$\frac{r}{A} M(\alpha_n) \leq M(\delta_n) \,. \tag{A3.12}$$

3c. — From (A3.11) and (A3.4)

$$M(\alpha_{n+1}^{\gamma}) \leq M(\alpha_n^{\gamma}) - 2\gamma_{n+1} M(\delta_n) + L\gamma_{n+1}^2 \leq$$

$$\leq M(\alpha_n^{\gamma}) - 2\frac{r}{A}\gamma_{n+1} M(\alpha_n^{\gamma}) + L\gamma_{n+1}^2 \leq \tag{A3.13}$$

$$\leq M(\alpha_n^{\gamma})\left(1 - \frac{2r}{A}\gamma_{n+1}\right) + L\gamma_{n+1}^2$$

by which we arrive exactly at the relation obtained in [4], from which it follows, in the same way as in [4] that

$$M(\alpha_n^{\gamma}) \to 0\,, \qquad \text{as} \quad n \to \infty \tag{A3.14}$$

$$M(\alpha_n^{\gamma}) < C\gamma_n^{\lambda}, \qquad \text{if} \quad n > n_0 \tag{A3.15}$$

This proves (27) and (28) in Theorem 3, viz. that

$$M\,\|f(\mathbf{x}) - f_n^{\gamma}(\mathbf{x})\,\|^2 \leq C\,\gamma_n^{\lambda} \quad n > n_0 \tag{A3.16}$$

and

$$M\,|f(\mathbf{x}) - f_n^{\gamma}(\mathbf{x})|^2 \leq LM\,\|f(\mathbf{x}) - f_n(\mathbf{x})\,\|^2 \leq LC\,\gamma_n^{\lambda} \text{ if } n > n_0\,. \tag{A3.17}$$

3d. — (28) in Theorem 3 follows from (27). Viz.,

$$\delta_n^2 = [M_x\,|f(\mathbf{x}) - f_n(\mathbf{x})\,|\,]^2 \leq M_x\,|f(\mathbf{x}) - f_n(\mathbf{x})\,|^2 = \beta_n\,. \tag{A3.18}$$

Observe that

$$\Big[\int_x |f(\mathbf{x}) - f_n(\mathbf{x})\,|\,p(\mathbf{x})\,d\mathbf{x}\Big]^2 \leq \int_x |f(\mathbf{x}) - f_n(\mathbf{x})\,|^2\,p(\mathbf{x})\,d\mathbf{x} \cdot \int p(\mathbf{x})\,d\mathbf{x}\,;$$

From this

$$\delta_n \leq \sqrt{\beta_n}\,,$$

and

$$M(\delta_n) \leq \sqrt{M(\beta_n)}\,.$$

Obviously:

$$\Big[\int \delta_n(\mathbf{x})\,p(\mathbf{x})\,d\mathbf{x}\Big]^2 \leq \int_x \sqrt{\beta_n(\mathbf{x})}\,p(\mathbf{x})\,d\mathbf{x} \leq \int_x \beta_n(\mathbf{x})\,p(\mathbf{x})\,d\mathbf{x} \int_x p(\mathbf{x})\,d\mathbf{x} = M(\beta_n)\,,$$

Therefore, (observing that $M(\beta_n) < C\,\gamma_n^{\lambda}$ if $n > n_0$) we obtain

$$M(\delta_n) \leq C_2\,\gamma_n^{\lambda/2} \tag{A3.19}$$

from which and (A3.9) follows (28) by which Theorem 3 is proved.

References

1. *Айзерман, М. А.—Браверман, Э. М.:* Теоретические основы метода потенциальных функций в задаче об обучении автоматов разделению ситуаций на классы. Автоматика и телемеханика *25*, (1964), 917—936.
2. *Айзерман, М. А.—Браверман, Э. М.—Розоноэр, Л. И.:* Вероятная задача об обучении автоматов распознаванию классов и метод потенциальных функций. Автоматика и телемеханика *25* (1964), 1307—1323.

3. *Айзерман, М. А.—Браверман, Э. М.—Розоноэр, Л. И.:* Метод потенциальных функций в задаче о восстановлении характеристики функционального преобразователя по случайно наблюдаемым точкам. Автоматика и телемеханика *25* (1964).
4. *Браверман, Э. М.:* О методе потенциальных функций. Автоматика и Телемеханика *26* (1965). 2205—2213.
5. *Браверман, Э. М.—Пятницкий, Е. С.:* Оценки скорости сходимости алгоритмов, основанных на методе потенциальных функций. Автоматика и телемеханика *25* (1966).
6. *Петерсен, И. Ф.:* Восстановление функции из класса с воспроизводящим ядром. Автоматика и телемеханика *28* (1969), 95—98.
7. *Parzen, E.:* Regression analysis of continuous parameter time series. Proc. of the Fourth Berkeley Symposium on Prob. and Math. Stat., Univ. of Calif, Press. *1* (1961).
8. *Parzen, E.:* Extraction and detection problems and reproducing kernel Hilbert Spaces. J. SIAM Control *1* (1962) 35—62.
9. *Aronszajn, N.:* Theory of reproducing kernels. Trans. of the Math. Soc. *68* (1950) 337—404.
10. *Yao, K.:* Reproducing kernels and bandlimited signals (manuscript).
11. *Kailath, T.:* Some results on singular detection. Techn. Rept. No. 7050—2, Stanford El Lab., Stanford Univ., 1965. June.
12. *Писаренко, В. Ф.—Розанов, Ю. А.:* О некоторых задачах для стационарных процессов, приводящих к уравнениям, родственным уравнению Винера — Хопфа. Проблемы передачи информации, *14* (1963) 113—135.
13. *Hajek, J.:* On linear statistical problems in stochastic processes. Czhech. Mat. J. *12* (1962) 404—444.

Об обобщении алгоритма обучения потенциальных функций и о скорости сходимости

О. ГУЛЬЯШ

(Будапешт)

Резюме

В работе обобщается метод потенциальных функций.

Пусть $K(\mathbf{x}, \mathbf{y})$ будет положительно-определенная, конечная функция на множестве $X \times X$. Пусть $H(K)$ обозначает Гильбертово пространство с воспроизводящим ядром, образованное из функции $K(\mathbf{x}, \mathbf{y})$.

Обозначим $\{x_i\}_{i=1}^{\infty}$ последовательность независимых случайных величин с одним и тем же распределением, определенную на X-множестве.

Теорема 1

Если $f(\mathbf{x}) \in H(K)$, тогда алгоритм $f_n(\mathbf{x}) = \sum_{k=1}^{n}{}' r_{k-1}(\mathbf{x}^k) K(\mathbf{x}, \mathbf{x}^k)$ сходится к $f(\mathbf{x})$ в смысле (12), где $r_k(\mathbf{x})$ задан в формулах (9), (10), (11).

Теорема 2

Если добавим к условиям теоремы 1 условие (21) и $r_k(\mathbf{x})$ выберем соответственно (9), тогда существуют критерии (22), 23), (24) относящиеся к скорости сходимости алгоритма.

Теорема 3

Если добавим к теоремам 1 и 2 условия (25), (26) и $r_k(\mathbf{x})$ выберем соответственно (10), тогда существуют критерии (27), (28) относящиеся к скорости сходимости алгоритма.

Ottó Gulyás
Research Institute for Telecommunication
Budapest II, Gábor Áron u. 65. Hungary

Problems of Control and Information Theory, Vol. 1 (1), pp. 65–75 (1972)

STOCHASTIC SYSTEMS AND THEIR CONNECTIONS

V. S. PUGACHEV

Moscow

(Received February 23, 1971)

The class of stochastic systems is defined and its closeness with respect to all possible connections by which large-scale systems are composed is proved. The concept of a decision function as the main characteristic of a stochastic system is introduced. The relations between the decision functions of the main types of connections of stochastic systems and those of connected systems are established yielding the means to find the decision functions of any composed stochastic system given the decision functions of the components. The stochastic system can serve as a model of a system with indeterministic behaviour in general statistical theory of control processes.

1. The main problem of modern control theory is the control of the large-scale systems usually containing men. The behaviour of such a system is somewhat indeterministic. Performing repeatedly in the same situation the behaviour of such a system is not the same each time. The result is that the outcome of the action of such a system can not be predicted with absolute accuracy.

The natural tendency is to apply the powerful techniques of probability theory to study indeterministic systems. Hence it is necessary to determine the probabilistic models of such systems. In other words it is necessary to specify a class of systems (i.e. mathematical models of systems) which admit a statistical description.

On the other hand, the main achievements of recent automatic control theory must be used. In particular, the structural theory being so fruitful in automatic control theory [1], it is natural to generalize this theory in such a manner that it be applicable to systems with indeterministic behaviour. Hence it is necessary to extend the definition of the main types of connections of automatic systems to more general systems and to define a statistical model of a system with indeterministic behaviour in such a way that the class of such models be closed with respect to all possible connections by which large-scale systems are composed of simpler ones.

The definitions of a stochastic system and of the main types of connections of such systems are given in this paper. The closeness of the class of stochastic

systems with respect to all possible connections is proved, and the relations between the characteristics of connections and those of the connected systems are derived [2].

2. We call *the system* any ensemble of interacting subjects of any nature.

The set of all external influences on a system, including the influences of surrounding medium, is called *the input* of the system.

The set of all essential features of the behaviour of a system is called its *output.*

Various features of phenomena in a system and surrounding medium can serve as the inputs and outputs in control problems. For instance, the inputs and outputs of automatic systems represent scalar or vector functions, those of finite automata are logical variables, those of service (queueing) systems are flows of events, those of recognition systems are images, sounds of speach, situations etc. It is thus necessary to consider inputs and outputs of systems as elements of any abstract spaces in general control theory.

A relation between elements of two spaces X and Y called respectively *the space of inputs* and *the space of outputs* can serve as a mathematical model of a system. Given an input $x \in X$, the system elaborates such an output $y \in Y$ that the pair (x, y) belongs to the relation characterizing the system.

To apply statistical methods it is necessary to consider only systems having the input-output relation of a statistical nature.

3. We call *stochastic* a system whose input-output relation is a family of probability distributions in the space of outputs Y depending upon the input $x \in X$. Such a system elaborates its output $y \in Y$ in accordance with the probability distribution in Y corresponding to a given $x \in X$.

To develop stochastic system theory it is necessary to suppose that the spaces of inputs and outputs are measurable. We shall thus suppose that a σ-algebra $\mathcal{A}$ of sets is defined in the space of inputs X, and a σ-algebra $\mathcal{B}$ is defined in the space of outputs Y.

The probability distribution in the space of outputs Y of a stochastic system for a given input $x \in X$ is determined by a conditional probability measure $\mu(E \mid x)$ representing the probability of an output belonging to a set $E \in \mathcal{B}$ for a given input $x \in X$. The measure $\mu(E \mid x)$ is a non-negative σ-additive function of a set E defined on the σ-algebra $\mathcal{B}$ for each $x \in X$ and a function of point x measurable with respect to the σ-algebra $\mathcal{A}$ for each set $E \in \mathcal{B}$ [3].

It is natural to take conditional probability measure $\mu(E \mid x)$ as the main characteristic of a stochastic system. We shall call this measure *the decision function* of a stochastic system. Considering several stochastic systems we add corresponding subscripts to μ. For instance, the decision function of the system $\mathcal{A}$ we denote μ_A.

4. The main types of connections of systems are the parallel connection, the connection in cascade and the feedback. Any complex systems can be composed using these typical connections.

The parallel connection of systems A and B is such a connection in which A and B have the same input x, and the pair (y, x) of their outputs serves as the output of the composed system (Fig. 1). The parallel connections of the

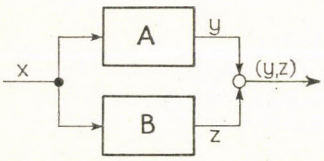


Fig. 1

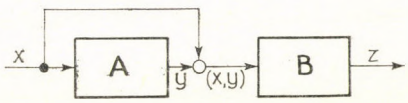


Fig. 2

systems A and B we denote $A \oplus B$. It is evident that the parallel connection is commutative $B \oplus A = A \oplus B$.

The case in which only some of the components of the input of the system A coincide with some components of the input of the system B is evidently covered by above definition the common input x of A and B representing the set of all components of the inputs of A and B. This case is thus a special case in which the decision functions of A and B are independent of some of the components of x.

The connection of systems A and B in cascade is such a connection in which the input of B contains the output of A (Fig. 2). The connection of the systems A and B in cascade we denote $A \otimes B$. It is clear that the connection in cascade is not commutative in general $B \otimes A \neq A \otimes B$.

The usual connection in cascade considered in automatic control theory in which the input of the second system is the output of the first one represents, evidently, the special case of the general connection in cascade with the input-output relation of the second system independent of the input x of the first system. On the other hand the general case can be reduced to this special case the pair (x, y) being considered as the output of the parallel connection of the system A and the identity system I (Fig. 3), $A \otimes B = (A \oplus I) \otimes B$.

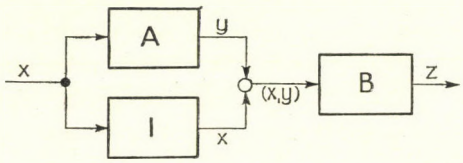


Fig. 3

The elementary feedback of a system A is such a connection of A with itself in which the input of A contains its output. Fig. 4a shows a system A whose input represents an element (x, y) of the Cartesian product of spaces $X \times Y$ and output is the element z of the space Y. Fig. 4b shows the elementary feedback of the system A. The elementary feedback of the system A we denote by $[A]$.

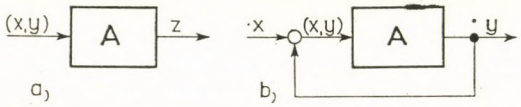


Fig. 4

The connection of a system A with itself in which its input contains only some of the components of its output is, evidently, the special case of the general elementary feedback in which the input-output relation of A is independent of some of the components of y.

The feedback of a system A via a system B is such a connection in which the input of the system A contains its output transformed by the system B (Fig. 5). This connection we denote $[A]_B$.

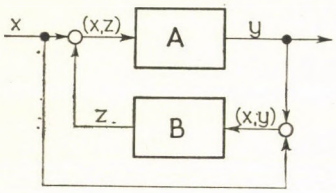


Fig. 5

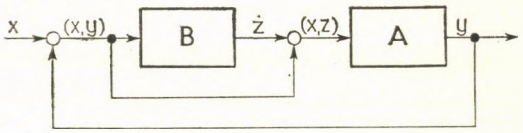


Fig. 6

It is clear that the feedback of the system A via the system B is equivalent to the elementary feedback of the connection of the systems B and A in cascade, $[A]_B = [B \otimes A]$ (Fig. 6).

It is easily understood that all possible connections of systems represent various combinations of parallel connections, connections in cascade and feedbacks. For instance the system N shown in Fig. 7 is $[(G \oplus H) \otimes ([D \otimes \otimes ([C \otimes (A \oplus B \oplus I_u)] \oplus I_v)] \otimes F)]$, I_u and I_v representing the systems with the input (x, u, v, p, q) and outputs u and v respectively (i.e. identity systems for the signals u and v non-passing other components of the inputs). Figs 8, 9, and 10 show consequent steps by which the system N is composed.

5. The main feature of connections of stochastic systems is the possibility of random breaking and arising of some connections. The connections in a stochastic system are thus *stochastic* themselves. The random variations of the

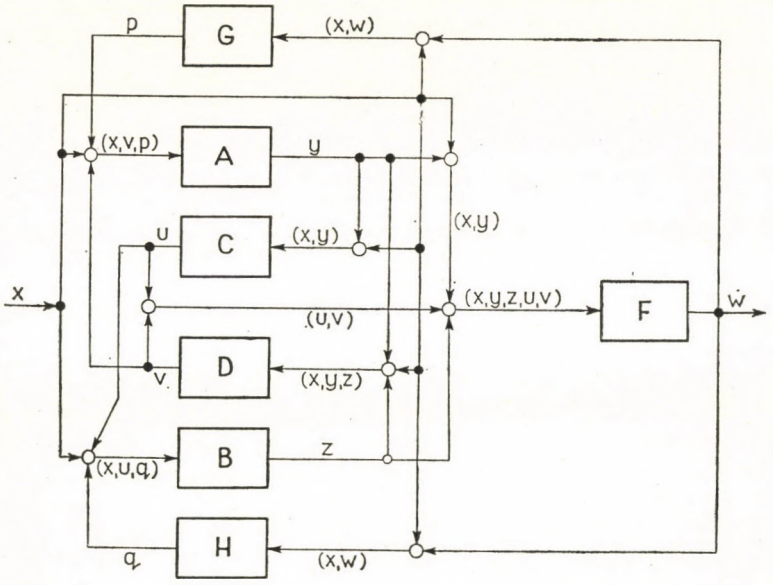


Fig. 7. System N

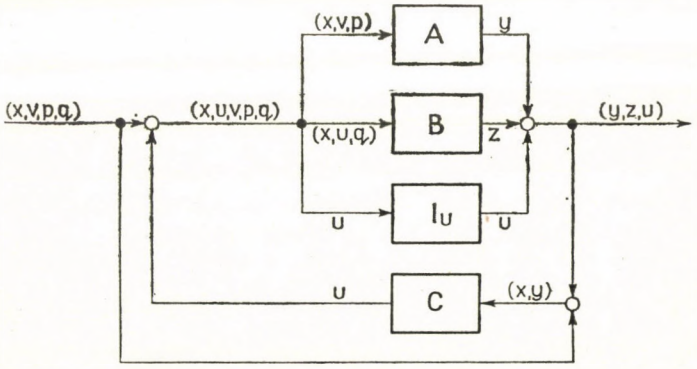


Fig. 8. System $L = [L \otimes (A \oplus B \oplus I_u)]$

connections in the stochastic system cause the corresponding random variations of its decision function. The random variations of the decision function of a stochastic system may also be caused by any other random variations of its inner state.

It is clear that the flow of events causing random variations of the state of a stochastic system A, in particular, of the connections among its components, can be considered as the output of some other stochastic system B and at the

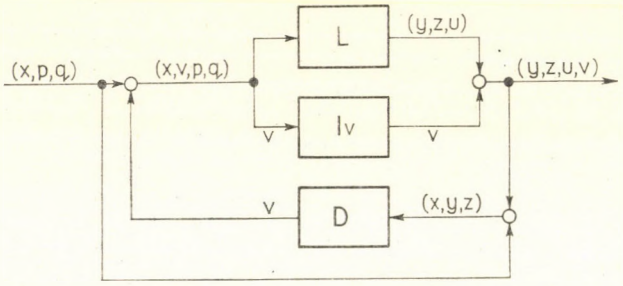


Fig. 9. System $M = [D \otimes (L \oplus I_v)]$

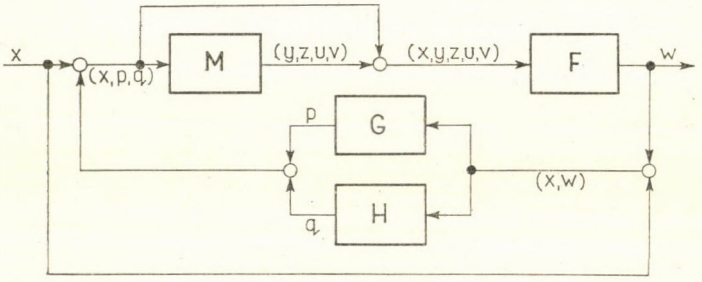


Fig. 10. System $N = [(G \oplus H) \otimes (M \otimes F)]$

same time as the additional component of the input of A. Thus the study of the system A with random variations of its decision function can be reduced to the study of the connection in cascade of two stochastic systems with definite decision functions. One is the system B generating the random flow of events causing the variations of the decision function of the system A, and other is the system A.

6. Consider the parallel connection of stochastic systems A and B (Fig. 1). The decision functions μ_A and μ_B of these systems define the conditional probability distribution in Cartesian product $Y \times Z$ of the spaces of outputs of these systems for a given input $x \in X$. The corresponding conditional probability measure is given by

$$\mu_c(E \mid x) = \iint\limits_E \mu_A(dy \mid x)\, \mu_B(dz \mid x) = \int\limits_{E_0} \mu_B(E_y \mid x)\, \mu_A(dy \mid x), \quad E \in B \times C, \quad (1)$$

E_y being the section of the set E in the point y, $E_0 = \{y : (y, z) \in E\}$ (i.e. the set of y for which the pair (y, z) belongs to E), $\mathcal{C}$ the σ-algebra of sets of the space Z, and $\mathcal{B} \times \mathcal{C}$ the product of the σ-algebras $\mathcal{B}$ and $\mathcal{C}$, i.e. the least σ-algebra containing all measurable rectangles $B \times C$, $B \in \mathcal{B}$, $C \in \mathcal{C}$ [3].

Thus *the parallel connection of stochastic systems A and B represents the stochastic system C whose decision function* μ_c *is determined by (1).*

7. Consider the connection of stochastic systems A and B in cascade (Fig. 2). The decision functions μ_A and μ_B of these systems define the conditional probability distribution in the Cartesian product space $Y \times Z$ of the outputs of these systems. The corresponding conditional probability measure is uniquely determined by

$$\mu_D(F \mid x) = \int_{F_0} \mu_B(F_y \mid x, y)\mu_A(dy \mid x), \quad F \in \mathfrak{B} \times \mathcal{C}, \tag{2}$$

F_y being the section of the set F in the point y, and $F_0 = \{y : (y, z) \in F\}$. To find the conditional probability distribution in the space Z for a given $x \in X$ we must restrict the measure μ_D to the σ-algebra $\mathcal{C}$. To do this it is sufficient to put $F = Y \times E$, $E \in \mathcal{C}$, in (2) yielding the conditional probability measure μ_c in Z:

$$\mu_C(E \mid x) = \int \mu_B(E \mid x, y)\mu_A(dy \mid x), \quad E \in \mathcal{C}, \tag{3}$$

the integral being extended over the whole space Y.

Thus *the connection of stochastic systems A and B in cascade represents the stochastic system C with the decision function* μ_c *determined by (3).*

8. Consider the elementary feedback of a stochastic system A (Fig. 4). Define a one-to-one measurable transform T_y of the space Y onto itself, $u = T_y z$, $(z, u \in Y)$, transforming the measure $\mu_A(E \mid x, y)$ in such a way that the resulting measure $\mu(T_y E \mid x)$ be independent of y*):

$$\mu(F \mid x) = \mu_A(T_y^{-1} F \mid x, y), \quad F \in \mathfrak{B}. \tag{4}$$

(* Such a transform always exists. In particular, it is easily determined, if the output of the system A may be represented by a finite-dimensional random vector, or separable scalar or finite-dimensional vector random function, taking $\mu(F \mid x) = \mu_A(T_{y_0}^{-1} F \mid x, y_0)$ for an arbitrary y_0.)

Then the equation $z = y$, i.e. $u = T_y y = Sy$ determines the output y of the elementary feedback of the system A as the function of the element u of the space Y with the probability measure $\mu(F \mid x)$ for a given $x \in X$. In other words, the output of the elementary feedback of the system A represents the random variable Y in the space Y which is a function of the random variable U in the same space Y.

The equation $u = Sy$ has always an unique solution with respect to y because only in this case the closed-loop system $[A]$ can perform (i.e. elaborate a single realization y of the random output Y for a given input x). Hence the

conditional probability measure $\mu_c(E \mid x)$ of the output Y of the close-loop system $[A]$ for a given input $x \in X$ is determined by

$$\mu_c(E \mid x) = \mu(SE \mid x), \quad E \in \mathfrak{B}. \tag{5}$$

Putting $y = y_0$, $F = SE$ in (4) relation (5) takes the form

$$\mu_c(E \mid x) = \mu_A(T_{y_0}^{-1} SE \mid x, y_0), \quad E \in \mathfrak{B}. \tag{6}$$

The right hand member of (6) is independent of y_0, since for any y

$$\mu_A(T_y^{-1}F \mid x, y) = \mu_A(T_{y_0}^{-1} F \mid x, y_0) \tag{7}$$

in virtue of (4). Furthermore, putting $F = T_y E$ we obtain from (7)

$$\mu_A(E \mid x, y) = \mu_A(T_{y_0}^{-1} T_y E \mid x, y_0). \tag{8}$$

This relation shows that the operator $V_{y_0}^{y} = T_{y_0}^{-1} T_y$ is independent of the specific choice of the measure μ. Therefore the equation $u = V_{y_0}^{y}\, y = T_{y_0}^{-1} T_y\, y = T_{y_0}^{-1} Sy$ is also independent of μ. Hence the right hand member of (6) is independent of μ. Taking $\mu(F \mid x) = \mu_A(F \mid x, y_0)$ formula (5) coincides with (6).

Thus *the elementary feedback of a stochastic system A represents the stochastic system C with the decision function μ_c determined by (6).*

To find the decision function of the feedback of the stochastic system A via the stochastic system B it is sufficient to substitute into (6)

$$\mu_D(F \mid x, y) = \int \mu_A(F \mid x, z)\mu_B(dz \mid x, y), \quad F \in \mathfrak{B},$$

instead of $\mu_A(F \mid x, y)$.

9. The three theorems proved establish the closeness of the class of stochastic systems with respect to all possible connections. In virtue of these theorems any connections of stochastic systems represent stochastic systems. Relations (1), (3) and (6) enable us in principle to determine the decision function of any compound stochastic system given those of its components.

The properties of the class of stochastic systems established enable us to apply structural approach to study complex systems. Representing a complex system as a connection of suitable subsystems and determining the statistical properties of each subsystem we may then study the behaviour of the overall system. This principle may be used, in particular, to simulate the behaviour of

large-scale systems using electronic computers. Determining first available statistical characteristics of the subsystems we may then use these characteristics instead of the subsystems themselves in simulating the behaviour of the overall system. As a result we obtain a reduction of the amount of necessary computations for simulating a given system and hence the possibility to enlarge the scope of systems which may be studied by a given computer.

However, the decision functions of components of large-scale systems are usually unknown or partially known (i.e. only incomplete information is available concerning their statistical characteristics). Thus the main problems of stochastic system control and optimization have to be posed and solved supposing that only such information about the decision functions involved is given which is practically available. This has to be taken into account in further developments of stochastic system theory.

References

1. *Pugachev, V. S.* (ed.): The foundations of automatic control (in Russian). Publishing House "Nauka", 1963 and 1968.
2. Pugachev V. S.: Stochastičeskie systemi i ich soyedineniya. Doklady Academii Nauk SSSR, 197, (1971), 6, 1288–1290.
3. *Loeve, M.:* Probability theory. D. Van Nostrand Co., 1955 and 1960.

Стохастические системы и их соединения

В. С. ПУГАЧЕВ

Москва

Резюме

1. Основной проблемой современной теории управления является управление сложными системами, как правило, включающими людей. Для таких систем характерна некоторая неопределенность поведения.

Естественно стремление применить для исследования систем с неопределенным поведением мощный аппарат статистических методов.

С другой стороны, при построении общей статистической теории процессов управления целесообразно использовать структурный подход, оказавшийся таким плодотворным для исследования сложных автоматических систем [1]. Поэтому необходимо определить статистическую модель системы с неопределенным поведением так, чтобы класс этих моделей был замкнут относительно всех возможных соединений, с помощью которых строятся сложные системы.

В данной работе определяются понятие стохастической системы и основные типы соединений таких систем. Доказывается, что класс стохастических систем замкнут относительно всех возможных соединений, и выводятся соотношения между характеристиками соединений и соединяемых систем [2].

2. Будем называть *системой* любую совокупность взаимодействующих предметов любой природы.

Совокупность всех внешних воздействий, которым подвергается система, включая воздействия среды, в которой она работает, будем называть *входным сигналом* системы.

Совокупность всех интересующих нас характеристик поведения системы будем называть *выходным сигналом* системы.

Входными и выходными сигналами систем, с которыми приходится встречаться в задачах управления, могут служить самые разнообразные характеристики явлений, происходящих в системе и в окружающей ее среде. Так, например, входные и выходные сигналы в автоматических системах представляют собой скалярные или векторные функции, в конечных автоматах — логические переменные, в системах массового обслуживания — потоки событий, в системах распознавания — изображения, звуки речи, ситуации и другие образы, как их принято называть. Поэтому в общей теории необходимо рассматривать входные и выходные сигналы систем как элементы произвольных абстрактных пространств.

Математической моделью системы при таком общем подходе может служить некоторое отношение между элементами двух пространств — *пространства входных сигналов* X и *пространства выходных сигналов* Y. При данном входном сигнале $x \in X$ система вырабатывает такой выходной сигнал $y \in Y$, что пара (x, y) принадлежит характеризующему систему отношению. Чтобы можно было применить статистические методы, необходимо ограничиться только такими системами, у которых отношение между входными и выходными сигналами имеет статистическую природу.

3. Будем называть *стохастической системой* такую систему, которая ставит в соответствие данному входному сигналу $x \in Y$ определенное распределение вероятностей в пространстве выходных сигналов Y.

Для построения теории стохастических систем необходимо рассматривать пространства входных и выходных сигналов как измеримые пространства. В соответствии с этим будем считать, что в пространстве входных сигналов X определена σ-алгебра множеств $\mathfrak{A}$, а в пространстве выходных сигналов Y — σ-алгебра множеств $\mathfrak{B}$.

Распределение вероятностей в пространстве выходных сигналов Y стохастической системы при данном входном сигнале $x \in X$ определяется условной вероятностной мерой $\mu(E|x)$ — переходной вероятностью, которая представляет собой вероятность того, что при данном входном сигнале $x \in X$ система выработает выходной сигнал y, принадлежащий множеству $E \in \mathfrak{B}$. Мера $\mu(E|x)$ при каждом y представляет собой неотрицательную σ-аддитивную функцию множества E, определенную на σ-алгебре $\mathfrak{B}$, и при каждом множестве $E \in \mathfrak{B}$ функцию переменной x, измеримую относительно σ-алгебры $\mathfrak{A}$ [2].

Меру $\mu(E|x)$ естественно принять за основную характеристику стохастической системы. Мы будем называть эту характеристику *решающей функцией* стохастической системы. Рассматривая несколько систем, будем отмечать их решающие функции соответствующими индексами. Например, решающую функцию системы A будем обозначать символом μ_A.

4. Основными типами соединений систем являются параллельное соединение, последовательное соединение и замыкание обратной связью.

Параллельным соединением $A \oplus B$ систем A и B будем называть такое их соединение, при котором на них подается один и тот же входной сигнал x, а выходным сигналом служит совокупность (y, z) выходных сигналов этих систем (рис. 1). Очевидно, что параллельное соединение коммутативно $B \oplus A = A \oplus B$.

Последовательным соединением $A \otimes B$ систем A и B будем называть такое их соединение, при котором выходной сигнал первой системы входит в состав входного сигнала второй (рис. 2). Очевидно, что последовательное соединение в общем случае некоммутативно, $B \otimes A \neq A \otimes B$.

Замыканием $[A]$ *системы* A *элементарной обратной связью* будем называть ввод выходного сигнала системы A в состав ее входного сигнала. На рис. 4*a* показана система A, у которой входным сигналом служит элемент (x, y) произведения пространств $X \times Y$, а выходным сигналом — элемент z пространства Y. На рис. 4*b* показано замыкание системы A элементарной обратной связью.

Замыканием $[A]_B$ *системы* A *обратной связью через систему* B будем называть ввод выходного сигнала системы A, преобразованного системой B, в состав входного сигнала системы A (рис. 5). Легко видеть, что $[A]_B = [B \otimes A]$ (рис. 6).

Легко понять, что все возможные соединения систем представляют собой комбинации параллельных и последовательных соединений и замыканий обратными связями. Так, например, система N, показанная на рис. 7, представляет собой

$$[(G \oplus H) \otimes ([D \otimes ([C \otimes (A \oplus B \oplus I_u)] \oplus I_v)] \otimes F)],$$

где I_u и I_v — системы с выходными сигналами u и v соответственно при входном сигнале (x, u, v, p, q) (системы тождественного преобразования сигналов u и v, не пропускающие других компонент входных сигналов).

5. Характерной особенностью соединений стохастических систем является возможность случайного возникновения и нарушения связей в процессе работы системы. Таким образом, связи в стохастической системе тоже являются *стохастическими*. Случайное изменение связей в стохастической системе приводит к случайным изменениям ее решающей функции.

Очевидно, что поток событий, определяющий случайные изменения решающей функции стохастической системы A, можно рассматривать как выходной сигнал некоторой другой стохастической системы B и в то же время как дополнительную компоненту входного сигнала данной системы A. Тогда исследование системы A со случайными изменениями решающей функции сведется к изучению последовательного соединения двух стохастических систем — системы B, генерирующей поток событий, управляющий случайными изменениями решающей функции системы A, и данной системы A.

6. Решающие функции μ_A и μ_B систем A и B, соединенных параллельно, определяют условную вероятностную меру в прямом произведении пространств $Y \times Z$ выходных сигналов этих систем. Эта мера выражается формулой (1), где E_y — сечение множества E в точке y, $E_0 = \{y : (y, z) \in E\}$, а $\mathcal{C}$ — σ-алгебра множеств пространства Z.

Таким образом, *параллельное соединение стохастических систем A и B представляет собой стохастическую систему C, решающая функция которой μ_c определяется формулой* (1).

7. Аналогично решающие функции μ_A и μ_B систем A и B, соединенных последовательно, определяют условное распределение вероятностей при данном $x \in X$ в прямом произведении пространств $Y \times Z$ их выходных сигналов (формула (2)). Чтобы найти условное распределение вероятностей в пространстве Z при данном $x \in X$, достаточно ограничиться в формуле (2) множествами F вида $Y \times E$, $E \in \mathcal{C}$. В результате получаем формулу (3).

Таким образом, *последовательное соединение стохастических систем A и B представляет собой стохастическую систему C, решающая функция которой μ_c определяется формулой* (3).

8. Рассмотрим замыкание стохастической системы A элементарной обратной связью. Определим взаимно однозначное измеримое отображение T_y пространства Y в себя, $u = T_y z$, переводящее меру $\mu_A(E|x, y)$ в данную меру $\mu(T_y E|x)$, не зависящую от y (формула (4)).*

(*) Такое отображение всегда существует при соответствующем выборе меры μ. В частности, такое отображение легко находится в случаях, когда выходной сигнал системы A представляет собой конечномерный случайный вектор или сепарабельную скалярную или конечномерную векторную случайную функцию, если принять $\mu(F|x) = \mu_A(F|x, y_0)$ при произвольном y_0.)

Тогда уравнение $z = y$, т. е. $u = T_y y = Sy$, определит выходной сигнал y замкнутой системы $[A]$ как однозначную функцию элемента u пространства Y, в котором распределение вероятностей при данном $x \in X$ определяется мерой $\mu(F \in x)$. Однозначность этой функции необходима для того, чтобы замкнутая система могла работать (т. е. сопоставлять каждому входному сигналу $x \in X$ единственную реализацию y случайного выходного сигнала Y). Поэтому условная вероятностная мера $\mu_C(E|x)$ выходного сигнала Y замкнутой системы $[A]$ при данном $x \in X$ определяется формулой (5). Эта формула приводится к виду (6) вследствие (4). На основании формул (7) и (8), вытекающих из (4), правая часть формулы (6) не зависит от выбора значения y_0 и меры μ. Если принять $\mu(F|x) = \mu_A(T_{y_0}^{-1} F|x, y_0)$, то формула (5) совпадет с (6).

Таким образом, *замыкание стохастической системы A элементарной обратной связью представляет собой стохастическую систему C, решающая функция которой μ_c определяется формулой* (6).

9. Доказанные три теоремы устанавливают замкнутость класса стохастических систем относительно всех возможных соединений. Формулы (1), (3) и (6) дают принципиальную возможность находить решающие функции любых соединений стохастических систем по данным решающим функциями соединяемых систем. Установленные свойства класса стохастических систем дают возможность применить для изучения сложных систем структурные методы. Расчленив сложную систему на подсистемы, можно исследовать характе-

ристики отдельных подсистем, а затем изучить поведение всей системы в целом. Это может существенно упростить моделирование поведения сложных систем на ЭВМ, сократить объем вычислений и увеличить масштаб систем, которые можно моделировать с помощью данной ЭВМ.

При исследовании сложных систем решающие функции входящих в их состав систем, как правило, неизвестны или известны лишь частично (т. е. имеется лишь неполная информация о них). Поэтому основные задачи управления стохастическими системами и их оптимизации необходимо ставить и решать, предполагая, что об их решающих функциях имеется только та информация, которая может быть практически получена. Эту характерную особенность задач исследования стохастических систем необходимо учитывать в дальнейшем развитии теории стохастических систем.

V. S. Pugachev
Institute of Control Problems
Profsoyuznaya ul. 81
Moscow V—485 USSR

Problems of Control and Information Theory, Vol. 1 (1), pp. 77–101 (1972)

ПРИБЛИЖЕННЫЕ МЕТОДЫ ОПИСАНИЯ РАБОТЫ ОДНОЛИНЕЙНОЙ СИСТЕМЫ МАССОВОГО ОБСЛУЖИВАНИЯ

А. А. ВОРОНОВ, Ю. В. ЧИСТЯКОВ

Москва

(Поступила в редакцию 23 февраля 1971 г.)

Для однолинейной системы массового обслуживания, у которой функция распределения входящего потока и обслуживания не являются одновременно решетчатыми, обладают конечными, не равными нулю, средними значениями получены выражения для преобразований Лапласа-Стилтьеса от функций распределения времени ожидания начала обслуживания, простоя прибора и выходящего потока. Для их практической реализации разработаны приближенные методы, использующие понятия частных характеристик и логарифмических корневых годографов.

Теория массового обслуживания находит широкое применение при разработке больших систем, в частности, в решении вопросов обоснованного выбора технических средств для автоматизированных систем управления [1, 2]. Среди моделей теории массового обслуживания особое место занимает однолинейная система. К ней может быть сведено математическое описание большого числа различных режимов работы реальных технических устройств.

В настоящей работе предлагаются приближенные методы для исследования работы однолинейной системы с достаточно произвольными функциями распределения входящего потока и обслуживания. Впервые эта задача была решена в работе [3]. В настоящей работе дается другое решение этой проблемы. Разработанные при этом приближенные методы снабжены таблицами и просты в использовании.

1. Вероятностная модель

Рассматриваемая система массового обслуживания состоит из одного прибора, на который поступает рекуррентный поток заявок, заданный функцией распределения A(t):

$$A(t) = P\{a_k \leq t\}, \quad k \geq 1,$$

$a_k = t_k - t_{k-1}$, t_k — момент поступления в систему k-й заявки.

Времена обслуживания отдельных заявок представляют собой независимые, одинаково распределенные случайные величины, заданные функцией распределения $B(t)$:

$$B(t) = P\{b_k \leq t\}, \quad k \geq 1,$$

$b_k = \tau_k - T_k, T_k$ — момент начала обслуживания k-й заявки; τ_k — момент окончания ее обслуживания; b_k — время обслуживания k-й заявки.

Если в момент поступления очередной заявки прибор оказался занятым обслуживанием одной из ранее поступивших заявок, вновь поступившая заявка становится в очередь. Длина очереди предполагается неограниченной. Обслуживание заявок производится в порядке поступления в систему.

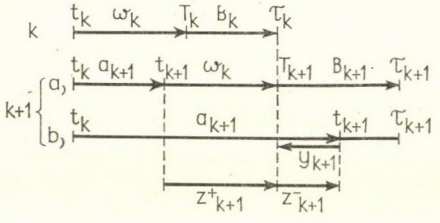


Рис. 1. Временная диаграмма

На рис. 1 показана временная диаграмма прохождения через систему k-ой и $(k+1)$-ой заявок. Приняты следующие обозначения:

$$c_k = \tau_{k+1} - \tau_k,$$

$$w_{k+1} = \begin{cases} \tau_k - t_{k+1}, & \text{если } \tau_k - t_{k+1} \geq 0, \\ 0, & \text{если } \tau_k - t_{k+1} < 0, \end{cases}$$

w_k — время ожидания начала обслуживания k-й заявки;

$$y_{k+1} = \begin{cases} t_{k+1} - \tau_k, & \text{если } t_{k+1} - \tau_k \geq 0, \\ 0, & \text{если } t_{k+1} - \tau_k < 0, \end{cases}$$

y_k — время простоя прибора перед началом обслуживания k-й заявки;

z_k — сдвиг по фазе [4] для k-й заявки, равный w_k, если $z_k \geq 0$, или $-y_k$, если $z_k < 0$;

$$z_{k+1} = \tau_k - t_{k+1} = \begin{cases} z^+_{k+1}, & \text{если } \tau_k - t_{k+1} \geq 0, \\ z^-_{k+1}, & \text{если } \tau_k - t_{k+1} < 0. \end{cases}$$

Непосредственно из временной диаграммы получаем следующие зависимости между случайными величинами:

$$z^+_k + b_k - a_{k+1} = z^-_{k+1} + z^+_{k+1}, \tag{1.1}$$

$$w_k = z^+_k, \tag{1.2}$$

$$y_k = -z_k^-, \tag{1.3}$$

$$c_k = y_k + b_k. \tag{1.4}$$

Рассматриваемые случайные величины обладают следующими свойствами:

1) $$z_k^+ \geq 0,\ b_k \geq 0,\ a_k \geq 0,\ y_k \geq 0,\ c_k \geq 0,\ z_k^- < 0, \tag{1.5}$$

2) Случайные величины w_k, b_k и a_{k+1}, а также y_k и b_k между собой независимы для любого $k \geq 1$.

Перейдем в выражениях (1.1—1.4) от случайных величин к их функциям распределения, которые будем обозначать соответствующими прописными буквами с указанием аргумента.

Формулам (1.1)—(1.4) соответствуют выражения:

$$\int_{-\infty}^{t} Z_k^+(t-x)\, d_x \int_0^{\infty} B(x+y)\, dA(y) = \begin{cases} Z_{k+1}^+(t), & \text{если } t \geq 0, \\ Z_{k+1}^-(t), & \text{если } t < 0, \end{cases} \tag{1.6}$$

где

$$Z_k^+(t) = P\{z_k^+ \leq t\} = \begin{cases} P\{z_k \leq t\}, & \text{если } t \geq 0, \\ 0, & \text{если } t < 0, \end{cases} \tag{1.7}$$

$$Z_k^-(t) = P\{z_k^- \leq t\} = \begin{cases} 1, & \text{если } t \geq 0, \\ P\{z_k \leq t\}, & \text{если } t < 0; \end{cases} \tag{1.8}$$

$$W_k(t) = Z_k^+(t); \tag{1.9}$$

$$Y_k(t) = 1 - Z_k^-(-t); \tag{1.10}$$

$$C_k(t) = \int_{-\infty}^{t} Y_k(t-x)\, dB(x). \tag{1.11}$$

Аналогично тому, как в работе [5] показано существование предела $\lim\limits_{k\to\infty} W_k(t) = W(t)$, можно показать, что существуют пределы $\lim\limits_{k\to\infty} Y_k(t) = Y(t)$ и $\lim\limits_{k\to\infty} Z_k(t) = Z(t)$, в силу чего при рассмотрении стационарного режима, соответствующего $k \to \infty$, индексы в выражениях (1.6—1.11) могут быть опущены.

В дальнейшем будем рассматривать только этот стационарный режим.

2. Основные соотношения

Перейдем в выражениях (1.6—1.11) к обобщенным преобразованиям Лапласа-Стилтьеса [5], которые будем обозначать соответствующими строчными буквами с указанием аргумента.

Формуле (1.6) соответствует выражение

$$z^+(s)\,b(s)\,a(-s) = z^+(s) - z(s)\,. \tag{2.1}$$

Действительно, раскладывая показатель при экспоненте на соответствующие слагаемые, производя замену переменных $u = x + y$ и $v = t - x$, а также принимая во внимание свойства (1.5), левая часть выражения (1.6) может быть преобразована следующим образом:

$$\int_{-\infty}^{\infty} e^{-st}\,d_t \int_{-\infty}^{t} Z^+(t-x)\,d_x \int_0^{\infty} B(x+y)\,dA(y) =$$

$$= \int_{-\infty}^{\infty} e^{-s(t-x)}\,d_t \int_{-\infty}^{t} Z^+(t-x)\,e^{-s(x+y)}\,d_x \int_0^{\infty} B(x+y)\,e^{-s(-y)}\,dA(y) =$$

$$= \int_{-\infty}^{\infty} e^{-sv}\,dZ^+(v) \int_{-\infty}^{\infty} e^{-su}\,dB(u) \int_0^{\infty} e^{sy}\,dA(y) =$$

$$= \int_0^{\infty} e^{-sv}\,dZ^+(v) \int_0^{\infty} e^{-su}\,dB(u) \int_0^{\infty} e^{sy}\,dA(y) = z^+(s)\,b(s)\,a(-s)\,.$$

Для правой части выражения (1.6), в соответствии с (1.5), получим

$$\int_{-\infty}^{\infty} e^{-st}\,dZ^+(t) = \int_{-\infty}^{\infty} e^{-st}\,dZ^+(t) = z^+(s);$$

$$\int_{-\infty}^{\infty} e^{-st}\,dZ^-(t) = \int_{-\infty}^{0} e^{-st}\,dZ^-(t) = -\int_0^{\infty} e^{st}\,dZ(-t) = -z^-(s)\,.$$

Аналогично получаются соотношения:

$$y(s) = 1 - z^-(s); \tag{2.2}$$

$$w(s) = z^+(s); \tag{2.3}$$

$$c(s) = y(s)\,b(s)\,. \tag{2.4}$$

При определенных ограничениях, наложенных на функции распределения $A(t)$ и $B(t)$, используя свойства аналитичности функций, входящих в уравнения (2.1—2.4), с помощью краевой задачи Римана теории функций комплексного переменного [6] эти уравнения могут быть решены. Первое применение задачи Римана к решению подобного рода проблем было дано в [7].

Полученный в настоящей работе результат сформулируем ниже в виде теоремы, доказательство которой приведено в п. 4.

Теорема: Если функция распределения входящего потока $A(t)$ и времени обслуживания $B(t)$ не является одновременно решетчатыми и, кроме того,

$$1)\quad A(0)<1,\quad B(0)<1; \tag{2.5}$$

$$2)\quad \bar{a}=\int_0^\infty t\,dA(t)<\infty,\quad \bar{b}=\int_0^\infty t\,dB(t)<\infty; \tag{2.6}$$

$$3)\quad \bar{a}>\bar{b}, \tag{2.7}$$

то для стационарного режима, соответствующего $k\to\infty$, для преобразований Лапласа-Стилтьеса от функций распределения времени ожидания $w(s)$ и простоя прибора $y(s)$ справедливы следующие выражения:

$$w(s)=k^{\frac{1}{2}}(j0)\exp\left\{-\frac{s}{2\pi j}\int_{j\infty}^{-j\infty}\frac{\ln k(\sigma)}{\sigma(\sigma-s)}\,d\sigma\right\}\quad\text{для}\quad \operatorname{Re}s>0; \tag{2.8}$$

$$y(s)=1-k^{\frac{1}{2}}(j0)\,s\exp\left\{\frac{s}{2\pi j}\int_{j\infty}^{-j\infty}\frac{\ln k(\sigma)}{\sigma(\sigma+s)}\,d\sigma\right\}\quad\text{для}\quad \operatorname{Re}s>0, \tag{2.9}$$

где

$$k(s)=\frac{1-a(-s)\,b(s)}{-s}.$$

Подставляя выражение (2.9) в выражение (2.4), получим также выражение для преобразования Лапласа-Стилтьеса от функции распределения выходящего потока $c(s)$.

3. Частотные методы расчета

В выражениях (2.8), (2.9) и (2.4) перейдем к пределу при $s\to j\omega$, где $\omega\in(-\infty,\infty)$, воспользовавшись при этом формулой Сохоцкого [6]:

$$w(j\omega)=\frac{k^{\frac{1}{2}}(j0)}{k^{\frac{1}{2}}(j\omega)}\exp\left\{\frac{\omega}{2\pi j}\int_{-\infty}^{\infty}\frac{\ln k(j\sigma)}{\sigma(\sigma-\omega)}\,d\sigma\right\}; \tag{3.1}$$

$$y(j\omega)=1+j\omega k^{\frac{1}{2}}(j0)\,k^{\frac{1}{2}}(j\omega)\exp\left\{-\frac{\omega}{2\pi j}\int_{-\infty}^{\infty}\frac{\ln k(j\sigma)}{\sigma(\sigma+\omega)}\,d\sigma\right\}; \tag{3.2}$$

$$c(j\omega)=y(j\omega)\,b(j\omega). \tag{3.3}$$

Прологарифмируем полученные выражения и разложим тождество для комплексных функций на два тождества, соответственно, для вещественной и мнимой части:

$$\ln|w(j\omega)| = \frac{1}{2}\ln|k(j0)| - \frac{1}{2}\ln|k(j\omega)| + I_1(\omega); \tag{3.4}$$

$$\arg w(j\omega) = -\frac{1}{2}\arg k(j\omega) - I_2(\omega); \tag{3.5}$$

$$\operatorname{Re} y(j\omega) = 1 + \operatorname{Re} g(j\omega); \tag{3.6}$$

$$\operatorname{Im} y(j\omega) = \operatorname{Im} g(j\omega), \tag{3.7}$$

где

$$\arg g(j\omega) = \frac{\pi}{2} - \frac{1}{2}\arg k(j\omega) - I_2(-\omega), \tag{3.8}$$

$$\ln|g(j\omega)| = \ln\omega + \frac{1}{2}\ln|k(j0)| + \frac{1}{2}\ln|k(j\omega)| + I_1(-\omega), \tag{3.9}$$

$$\ln|c(j\omega)| = \ln|y(j\omega)| + \ln|b(j\omega)|, \tag{3.10}$$

$$\arg c(j\omega) = \arg y(j\omega) + \arg b(j\omega), \tag{3.11}$$

$$I_1(\omega) = \frac{\omega}{2\pi}\int_{-\infty}^{\infty}\frac{\arg k(j\sigma)}{\sigma(\sigma-\omega)}\,d\sigma, \tag{3.12}$$

$$I_2(\omega) = \frac{\omega}{2\pi}\int_{-\infty}^{\infty}\frac{\ln|k(j\sigma)|}{\sigma(\sigma-\omega)}\,d\sigma. \tag{3.13}$$

Отметим, что выражение (3.2) удобно сначала разложить на вещественную и мнимую часть.

По аналогии с [8] введем понятие частных характеристик, понимая под ними графическое изображение соответствующих функций от переменной ω, называемой частотной.

Построение $I_1(\omega)$

Выполним ряд эквивалентных преобразований выражения (3.12):

$$I_1(\omega) = \frac{\omega}{2\pi}\int_{-\infty}^{\infty}\frac{\arg k(j\sigma)}{\sigma(\sigma-\omega)}\,d\sigma = \frac{\omega^2}{\pi}\int_{0}^{\infty}\frac{\arg k(j\sigma)}{\sigma(\sigma^2-\omega^2)}\,d\sigma = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\arg k(j\omega e^u)}{e^{2u}-1}\,du =$$

$$= \arg k(j\omega e^u)\left[-u + \frac{1}{2}\ln|e^{2u}-1|\right]\Bigg|_{-\infty}^{\infty} -$$

$$-\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{d}{du}\arg k(j\omega e^{u})\left[-u+\frac{1}{2}\ln|e^{2u}-1|\right]du=$$

$$=-\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{d}{du}\arg k(j\omega e^{u})\,d\left[-\frac{u^2}{2}+\frac{1}{4}\int_{0}^{e^{2u}}\frac{\ln|x-1|}{x}dx\right]. \quad (3.14)$$

В процессе сделанных преобразований, используя свойство нечетности функции $\arg k(j\sigma)$, прежде всего, были изменены пределы интегрирования; далее была сделана замена переменного, использующая подстановку $\frac{\sigma}{\omega}=e^{u}$; была использована формула интегрирования по частям ([9], формула 569.1), а также то обстоятельство, что

$$\lim_{\substack{u\to\pm\infty\\ u\to 0}}\arg k(j\omega e^{u})\left[-u+\frac{1}{2}\ln|e^{2u}-1|\right]=0,$$

справедливость чего может быть доказана по правилу Лопиталя, если дополнительно учесть, что при малых значениях аргумента $\arg k(j\sigma)\sim\sigma$. На заключительном этапе преобразований была сделана замена переменного $e^{2u}=x$.

Принимая во внимание, что

$$\frac{d}{du}\arg k(j\omega e^{u})=\frac{d}{d\ln\sigma}\arg k(j\sigma), \quad (3.15)$$

аппроксимируем функцию $\arg k(j\sigma)$ кусочно-линейными функциями при логарифмическом масштабе по оси σ:

$$\arg k(j\sigma)\simeq\sum_{i=1}^{N_1}F_{1,i}(\sigma), \quad (3.16)$$

причем, производная от функции $F_{1,i}(\sigma)$ имеет вид:

$$\frac{dF_{1,i}(\sigma)}{d\ln\sigma}=\begin{cases}0, & \text{если}\quad \ln|\sigma|<\ln\sigma_i-\dfrac{\varphi_i}{2k_i},\\ [\operatorname{sign}\sigma]\,k_i, & \text{если}\quad \ln\sigma_i-\dfrac{\varphi_i}{2k_i}\le\ln|\sigma|\le\ln\sigma_i+\dfrac{\varphi_i}{2k_i},\\ 0, & \text{если}\quad \ln|\sigma|>\ln\sigma_i+\dfrac{\varphi_i}{2k_i},\end{cases} \quad (3.17)$$

где

$$k_i = \left.\frac{d \arg k(j\sigma)}{d \ln \sigma}\right|_{\sigma=\sigma_i}.$$

В этом случае

$$I_1(\omega) \simeq \sum_{i=1}^{N_1} I_1(\omega, k_i, \varphi_i, \sigma_i), \tag{3.18}$$

где

$$I_1(\omega, k_i, \varphi_i, \sigma_i) = \frac{\omega}{2\pi} \int_{-\infty}^{\infty} \frac{F_i(\sigma)}{\sigma(\sigma - \omega)} d\sigma = \frac{k_i}{\pi}\left[\frac{\bar{u}^2 - u^2}{2} - \frac{1}{4}\int_{e^{2u}}^{e^{2\bar{u}}} \frac{\ln|x-1|}{x} dx\right],$$

где

$$\bar{u} = \ln\sigma - \ln\omega = \ln\sigma_i + \frac{\varphi_i}{2k_i} - \ln\omega,$$

$$u = \ln\sigma_i - \frac{\varphi_i}{2k_i} - \ln\omega,$$

откуда

$$I_1(\omega, k_i, \varphi_i, \sigma_i) = \frac{1}{\pi}\left[\varphi_i \ln\frac{\sigma_i}{\omega} - \frac{k_i}{4}\int_{\left(\frac{\sigma_i}{\omega}\right)^2 \exp\left\{-\frac{\varphi_i}{k_i}\right\}}^{\left(\frac{\sigma_i}{\omega}\right)^2 \exp\left\{\frac{\varphi_i}{k_i}\right\}} \frac{\ln|x-1|}{x} dx\right]. \tag{3.19}$$

Поскольку $I_1(\omega, k_i, \varphi_i, \sigma_i)$ зависит от величин $\frac{\sigma_i}{\omega}$, $\frac{\varphi_i}{k_i}$ и φ_i, то для получения $I_1(\omega, k_i, \varphi_i, \sigma_i)$ необходимо функцию $I_1\left(\omega, k_i, \frac{\pi}{2}, 1\right)$ передвинуть параллельно самой себе на величину $\ln\sigma_i$ и изменить масштаб по вертикали в $\frac{2\varphi_i}{\pi}$ раз. При больших значениях ω имеет место

$$I_1(\omega, k_i, \varphi_i, \sigma_i) \simeq \frac{\varphi_i}{\pi} \ln\frac{\sigma_i}{\omega}. \tag{3.20}$$

Производная от функции $I_1\left(\omega, k_i, \frac{\pi}{2}, 1\right)$ по $\ln\omega$ равна

$$\frac{d}{d\ln\omega} I_1\left(\omega, k_i, \frac{\pi}{2}, 1\right) = \frac{1}{\pi}\left\{\frac{\pi}{2}\ln\frac{1}{\omega} - \frac{k_i}{4}\int_{\left(\frac{1}{\omega}\right)^2 \exp\left\{-\frac{\pi}{2k_i}\right\}}^{\left(\frac{1}{\omega}\right)^2 \exp\left\{\frac{\pi}{2k_i}\right\}} \frac{\ln|x-1|}{x} dx\right\}'_{\ln\omega} =$$

$$= -\frac{1}{2} - \frac{d}{d\ln\omega} I_1\left(\frac{1}{\omega}, k_i, \frac{\pi}{2}, 1\right). \tag{3.21}$$

Таким образом, достаточно иметь значения функции $I_1\left(\omega, k_i, \frac{\pi}{2}, 1\right)$ для $0 < \omega \leq 1$. Значение же этой функции для $1 \leq \omega < \infty$ могут быть построены, как сумма $I_1\left(\frac{1}{\omega}, k_i, \frac{\pi}{2}, 1\right)$ и $-\frac{1}{2}\ln\omega$, причем последняя функция при логарифмическом масштабе по оси ω представляет наклонную прямую, проходящую через точку (1.1) с наклоном $-\frac{1}{2}$.

В таблице 1 приведены значения функции $I_1\left(\omega, k_i, \frac{\pi}{2}, 1\right)$ для ряда значений k_i.

Интеграл $I_1(\omega)$ согласно (3.18) получается как сумма соответствующих интегралов $I_1(\omega, k_i, \varphi_i, \sigma_i)$ для функций, аппроксимирующих $\arg k(j\sigma)$.

Построение $I_2(\omega)$

Выполнив ряд эквивалентных преобразований выражения (3.13), получим

$$I_2(\omega) = \frac{\omega}{2\pi}\int_{-\infty}^{\infty}\frac{\ln|k(j\sigma)|}{\sigma(\sigma-\omega)}\,d\sigma = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{d\ln|k(j\omega e^u)|}{du}\,d_u\int_0^{e_u}\ln\left|\frac{1+x}{1-x}\right|d\ln x. \quad (3.22)$$

В процессе сделанных преобразований, используя свойство четности функции $\ln|k(j\sigma)|$, прежде всего интеграл по оси $(-\infty, \infty)$ был заменен интегралом по лучу $[0, \infty)$; далее была сделана замена переменного, используя подстановку $\frac{\sigma}{\omega} = e^u$, и была использована формула интегрирования по частям ([9], формула 140.02), а также то обстоятельство, что

$$\lim_{u\to\pm\infty}\ln\left|\frac{1+e^u}{1-e^u}\right| = 0,$$

справедливость чего может быть проверена по правилу Лопиталя. На заключительном этапе преобразований была сделана замена переменного:

$$e^u = x.$$

Принимая во внимание, что

$$\frac{d}{du}\ln|k(j\omega e^u)| = \frac{d}{d\ln\sigma}\ln|k(j\sigma)|, \quad (3.23)$$

Таблица I Значения* интеграла $I_1\left(\omega, k_i, \frac{\pi}{2} \cdot 1\right)$

ω, $\partial\delta$ \ k_i	1,0	0,9	0,8	0,7	0,6	0,5	0,4	0,3
0	1,117	0,871	0,592	0,268	0,118	—0,598	—1,233	—2,162
0,05	1,339	1,098	0,822	0,502	0,119	—0,358	—0,990	—1,917
0,1	1,504	1,277	1,014	0,704	0,331	—0,138	—0,763	—1,683
0,15	1,609	1,401	1,164	0,874	0,516	0,062	—0,551	—1,461
0,2	1,646	1,480	1,272	1,009	0,676	0,241	—0,355	—1,250
0,25	1,601	1,494	1,332	1,108	0,807	0,399	—0,174	—1,051
0,3	1,442	1,430	1,338	1,168	0,909	0,536	—0,009	—0,862
0,35	0,988	1,248	1,274	1,182	0,980	0,650	0,140	—0,685
0,4	0,669	0,803	1,103	1,139	1,016	0,741	0,273	—0,518
0,45	0,490	0,562	0,704	1,015	1,010	0,807	0,388	—0,364
0,5	0,369	0,416	0,496	0,681	0,950	0,844	0,487	—0,220
0,55	0,283	0,315	0,368	0,469	0,801	0,849	0,567	—0,089
0,6	0,219	0,242	0,279	0,345	0,501	0,813	0,628	0,031
0,65	0,170	0,188	0,215	0,261	0,358	0,714	0,667	0,139
0,7	0,133	0,146	0,166	0,200	0,267	0,468	0,682	0,234
0,75	0,105	0,155	0,130	0,155	0,203	0,326	0,667	0,317
0,8	0,083	0,090	0,102	0,121	0,157	0,241	0,612	0,387
0,85	0,065	0,071	0,080	0,095	0,122	0,183	0,472	0,442
0,9	0,051	0,056	0,063	0,074	0,095	0,140	0,299	0,482
0,95	0,041	0,044	0,050	0,059	0,075	0,109	0,218	0,506
1,0	0,032	0,035	0,039	0,046	0,059	0,085	0,164	0,511
1,05	0,026	0,028	0,031	0,037	0,046	0,067	0,125	0,492
1,1	0,020	0,022	0,025	0,029	0,036	0,052	0,097	0,487
1,15	0,016	0,017	0,020	0,023	0,029	0,041	0,075	0,293
1,2	0,013	0,014	0,016	0,018	0,023	0,033	0,059	0,202
1,25	0,010	0,011	0,012	0,014	0,018	0,026	0,046	0,149
1,3	0,008	0,009	0,010	0,011	0,014	0,020	0,036	0,113
1,35	0,006	0,007	0,008	0,009	0,011	0,016	0,029	0,087
1,4	0,005	0,006	0,006	0,007	0,009	0,013	0,023	0,067
1,45	0,004	0,004	0,005	0,005	0,007	0,010	0,018	0,052
1,5	0,003	0,003	0,004	0,005	0,006	0,008	0,014	0,041
1,55	0,003	0,003	0,003	0,004	0,005	0,006	0,011	0,032
1,6	0,002	0,002	0,002	0,003	0,004	0,005	0,009	0,025
1,65	0,001	0,002	0,002	0,002	0,003	0,004	0,007	0,020
1,7	0,001	0,001	0,002	0,002	0,002	0,003	0,006	0,016
1,75	0,001	0,001	0,001	0,001	0,002	0,003	0,004	0,013
1,8	0,001	0,001	0,001	0,001	0,001	0,002	0,004	0,010
1,85	0,001	0,001	0,001	0,001	0,001	0,002	0,003	0,008
1,90	0,001	0,001	0,001	0,001	0,001	0,001	0,002	0,006
1,95	0,000	0,000	0,001	0,001	0,001	0,001	0,002	0,005
2,00	0,000	0,000	0,000	0,000	0,001	0,001	0,001	0,004

* Значения даны в децибеллах [$\partial\delta$].

Таблица I
(продолжение)

ω, $\partial\delta \diagdown k_i$	1,0	2,0	3,0	4,0	5,0	6,0	7,0	8,0
0	1,117	2,677	3,568	4,196	4,682	5,079	5,414	5,704
0,05	1,339	2,827	3,595	4,044	4,275	4,313	4,002	3,789
0,1	1,504	2,760	3,037	2,517	2,360	2,292	2,255	2,232
0,15	1,609	2,373	1,761	1,636	1,587	1,562	1,548	1,539
0,2	1,646	1,442	1,224	1,167	1,114	1,130	1,122	1,118
0,25	1,601	1,007	0,897	0,864	0,850	0,842	0,838	0,835
0,3	1,442	0,741	0,675	0,654	0,644	0,639	0,636	0,634
0,35	0,989	0,559	0,515	0,501	0,494	0,490	0,489	0,487
0,4	0,669	0,428	0,398	0,387	0,382	0,290	0,379	0,378
0,45	0,490	0,331	0,309	0,301	0,298	0,232	0,295	0,294
0,5	0,369	0,258	0,241	0,236	0,233	0,181	0,231	0,230
0,55	0,283	0,202	0,189	0,185	0,183	0,143	0,181	0,180
0,6	0,219	0,158	0,149	0,145	0,144	0,113	0,144	0,142
0,65	0,170	0,125	0,117	0,115	0,114	0,089	0,113	0,112
0,7	0,133	0,098	0,093	0,091	0,090	0,071	0,089	0,089
0,75	0,105	0,078	0,073	0,072	0,071	0,056	0,070	0,070
0,8	0,083	0,061	0,058	0,057	0,056	0,044	0,056	0,056
0,85	0,065	0,049	0,046	0,045	0,044	0,035	0,044	0,044
0,9	0,051	0,039	0,036	0,036	0,035	0,028	0,035	0,035
0,95	0,041	0,030	0,029	0,028	0,028	0,022	0,028	0,028
1,0	0,032	0,024	0,023	0,022	0,022	0,018	0,022	0,022
1,05	0,026	0,019	0,018	0,018	0,018	0,014	0,017	0,017
1,1	0,020	0,015	0,014	0,014	0,014	0,011	0,014	0,014
1,15	0,016	0,012	0,011	0,011	0,011	0,009	0,011	0,011
1,2	0,013	0,010	0,009	0,009	0,009	0,007	0,009	0,009
1,25	0,010	0,008	0,007	0,007	0,007	0,005	0,007	0,007
1,3	0,008	0,006	0,006	0,006	0,006	0,004	0,005	0,005
1,35	0,006	0,005	0,005	0,004	0,004	0,004	0,004	0,004
1,4	0,005	0,004	0,004	0,004	0,003	0,003	0,003	0,003
1,45	0,004	0,003	0,003	0,003	0,003	0,002	0,003	0,003
1,5	0,003	0,002	0,002	0,002	0,002	0,002	0,002	0,002
1,55	0,003	0,002	0,002	0,002	0,002	0,001	0,002	0,002
1,6	0,002	0,002	0,001	0,001	0,001	0,001	0,001	0,001
1,65	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
1,7	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
1,75	0,001	0,001	0,001	0,001	0,001	0,001	0,001	0,001
1,8	0,001	0,001	0,001	0,001	0,001	0,000	0,000	0,000
1,85	0,001	0,001	0,000	0,000	0,000	0,000	0,000	0,000
1,9	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
1,95	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
2,0	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

* Значения даны в децибеллах [$\partial\delta$].

аппроксимируем $\ln|k(j\sigma)|$ полубесконечными прямыми при логарифмическом масштабе по оси σ:

$$\ln|k(j\sigma)| \simeq \sum_{i=1}^{N_2} F_{2,i}(\sigma), \tag{3.24}$$

причем производная от функции $F_{2,i}(\sigma)$ имеет вид:

$$\frac{dF_{2,i}(\sigma)}{d\ln\sigma} = \begin{cases} 0 & \text{для } \ln|\sigma| < \ln\sigma_i, \\ l_i & \text{для } \ln|\sigma| \geq \ln\sigma_i. \end{cases} \tag{3.25}$$

В этом случае

$$I_2(\omega) \simeq \sum_{i=1}^{N_2} I_2(\omega, l_i, \sigma_i), \tag{3.26}$$

где

$$I_2(\omega, l_i, \sigma_i) = \frac{l_i}{\pi} \int_{\sigma_i/\omega}^{\infty} \ln\left|\frac{1+x}{1-x}\right| d\ln x. \tag{3.27}$$

Из полученного выражения следует, что $I_2(\omega, l_i, \sigma_i)$ зависят от величины $\frac{\sigma_i}{\omega}$ и l_i. Таким образом, если известно значение $I_2(\omega, 1, 1)$, то для получения $I_2(\omega, l_i, \sigma_i)$ необходимо эту функцию передвинуть параллельно самой себе на величину $\ln\sigma_i$ и изменить масштаб по вертикали в l_i раз.

При больших значениях ω имеем

$$\lim_{\omega\to\infty} I_2(\omega, l_i, \sigma_i) = \frac{l_i}{2\pi} \lim_{\omega\to\infty} \int_{\sigma_i/\omega}^{\infty} \ln\left|\frac{1+x}{1-x}\right| d\ln x = l_i \frac{\pi}{4}. \tag{3.28}$$

Производная от функции $I_2(\omega, 1, 1)$ по $\ln\omega$ равна

$$\frac{d}{d\ln\omega} I_2(\omega, 1, 1) = \frac{d}{d\ln\omega} I_2\left(\frac{1}{\omega}, 1, 1\right). \tag{3.29}$$

Таким образом достаточно иметь значения функции $I_2(\omega, 1,1)$ для $0 < \omega \leq 1$. Значения же этой функции для $1 \leq \omega < \infty$ могут быть построены, как $I_2\left(\frac{1}{\omega}, 1, 1\right)$. В работе [10] приведены таблицы значений функции

$$\varphi(\omega) = \frac{2}{\pi} \int_{1}^{\infty} \ln \operatorname{cth}\left|\frac{\lambda}{2}\right| d\lambda.$$

Очевидно имеет место соотношение

$$I_2(\omega, 1, 1) = 0{,}5\,\varphi(\omega), \tag{3.30}$$

поэтому для построения $I_2(\omega, 1, 1)$ можно воспользоваться приведенной в работе [10] таблицей, умножая на 0,5 соответствующие значения.

Интеграл $I_2(\omega)$, согласно (3.24), получается как сумма соответствующих интегралов $I_2(\omega, l_i, \sigma_i)$ для функций, аппроксимирующих $\ln|k(j\omega)|$.

Для удобства построения могут быть изготовлены шаблоны функций $I_1\left(\omega, k_i, \frac{\pi}{2}, 1\right)$ и $I_2(\omega, 1, 1)$.

4. Приложение

Доказательство теоремы

Функции $b(s)$, $z^+(s)$, $w(s)$, $y(s)$ и $c(s)$ являются аналитическими в полуплоскости $\mathrm{Re}\, s \geq 0$.

Действительно, в этой полуплоскости эти функции являются ограниченными, так как, например,

$$|b(s)| = \left| \int_0^\infty e^{-st}\, dB(t) \right| \leq \int_0^\infty |e^{-st}|\, |dB(t)| = \int_0^\infty |dB(t) = 1\,, \qquad (4.1)$$

в силу того, что в полуплоскости $\mathrm{Re}\, s \geq 0$ имеет место $e^{-st} = e^{-at} + e^{-j\omega t}$, где $a \geq 0$, $\omega \geq 0$ — вещественные числа, и, кроме того, в силу свойства предельных соотношений для преобразований Лапласа-Стилтьеса справедливо:

$$\lim_{s \to 0} b(s) = \lim_{t \to \infty} B(t) = 1\,. \qquad (4.2)$$

Аналогично показывается справедливость (4.1) и (4.2) для функций $z^+(s)$, $w(s)$, $y(s)$ и $c(s)$.

Точно так же можно показать, что функция $a(-s)$ является аналитической в полуплоскости $\mathrm{Re}\, s \leq 0$. Относительно же функции $z^-(s)$ можно лишь доказать справедливость в полуплоскости $\mathrm{Re}\, s \leq 0$ соотношения (4.1). Соотношение (4.2) для этой функции не выполняется и, более того, справедливы соотношения:

$$\lim_{s \to -0} z^-(s) = \lim_{t \to -\infty} Z^-(t) = 0, \qquad (4.3)$$

$$\lim_{s \to 0} \frac{dz^-(s)}{ds} = \lim_{s \to 0} \frac{d}{ds}\left[z^+(s) - Z^+(s)\, b(s)\, a(-s)\right] =$$

$$= -b'(0) + a'(0) = \bar{b} - \bar{a} < 0\,, \qquad (4.4)$$

в силу предельных соотношений для преобразований Лапласа-Стилтьеса. Из соотношений (4.1), (4.3) и (4.4) следует, что функция $z^-(s)$ является аналитической в полуплоскости $\mathrm{Re}\, s \leq 0$ за исключением бесконечно удаленной точки, где она имеет особенность типа полюс первого порядка.

Введем вспомогательные функции $z_{*}^{\pm}(s)$ и $z_{**}^{\pm}(s)$, связанные следующими соотношениями с функциями, входящими в выражения (2.1—2.4):

$$z^{\pm}(s) = z_{*}^{\pm}(s)\, z_{**}^{\pm}(s)\,; \tag{4.5}$$

$$z_{*}^{-}(s) = z_{*}^{+}(s)\, k(s); \tag{4.6}$$

$$-\frac{1}{s}\, z_{**}^{-}(s) = z_{**}^{+}(s)\,, \tag{4.7}$$

где

$$k(s) = \frac{1 - a(-s)\, b(s)}{-s}\,,$$

причем функции $z_{*}^{+}(s)$ и $z_{**}^{+}(s)$ — аналитические во всей правой полуплоскости $\operatorname{Re} s \geq 0$, функция $z_{*}^{-}(s)$ — аналитическая во всей левой полуплоскости $\operatorname{Re} s \leq 0$, а функция $z_{**}^{-}(s)$ — аналитическая во всей левой полуплоскости $\operatorname{Re} s \leq 0$ за исключением бесконечно удаленной точки, где она имеет полюс первого порядка.

Рассмотрим свойства функции $k(s)$ на мнимой оси $\operatorname{Re} s = 0$ и на контуре $\bar{C}_R$, состоящем из отрезка мнимой оси $[-jR, jR]$ и полуокружности радиуса R, лежащей в правой полуплоскости.

Свойства функции $k(s)$ на мнимой оси $\operatorname{Re} s = 0$

Аргумент s на мнимой оси будем обозначать через $j\omega, \omega \in (-\infty, \infty)$.

1. Если функции распределения $A(t)$ и $B(t)$ обладают конечными значениями математических ожиданий $\bar{a} < \infty$ и $\bar{b} < \infty$, соответственно, то функция $k(j\omega)$ удовлетворяет условию Гельдера [6].

Действительно, при сформулированных условиях имеем:

$$\left|\frac{da(j\omega)}{d\omega}\right| = \left|-j\int_0^{\infty} te^{-j\omega t}\, dA(t)\right| \leq \int_0^{\infty} |t|\, |e^{-j\omega t}|\, |dA(t)| = \int_0^{\infty} t\, dA(t) = \bar{a};$$

$$\left|\frac{db(j\omega)}{d\omega}\right| \leq \bar{b}\,,$$

то есть, функции $a(s)$ и $b(s)$ удовлетворяют условию Гельдера на мнимой оси $\operatorname{Re} s = 0$. Сделанное выше утверждение справедливо, так как согласно [6] сумма и произведение функций, удовлетворяющих условию Гельдера, также удовлетворяют этому условию.

2. Если функции распределения $A(t)$ и $B(t)$ не являются одновременно решетчатыми [11] и, кроме того, $\bar{a} > \bar{b}$; $A(0) < 1$ и $B(0) < 1$, то функция $k(j\omega) = 0$ только при $\omega = \infty$.

Действительно,

$$\lim_{\omega \to \infty} k(j\omega) = \lim_{\omega \to \infty} \frac{1 - a(-j\omega)\, b(j\omega)}{-j\omega} = 0,$$

так как на основании предельных соотношений для преобразований Лапласа-Стилтьеса:

$$\lim_{\omega \to \infty} a(j\omega) = \lim_{t \to 0} A(t) = A(0) < 1;$$

$$\lim_{\omega \to \infty} b(j\omega) = \lim_{t \to t} B(t) = B(0) < 1.$$

При $\omega = 0$ получим значение $k(j\omega)$ путем раскрытия неопределенности по правилу Лопиталя:

$$\lim_{\omega \to 0} k(j\omega) = \lim_{\omega \to 0} \frac{1 - a(-j\omega)\, b(j\omega)}{-j\omega} = \frac{ja'(-j\omega)\, b(j\omega) - ja(j\omega) b'(j\omega)}{-j}\Bigg|_{\omega=0} =$$

$$= \bar{a} - \bar{b} > 0,$$

причем здесь мы воспользовались предельными соотношениями для преобразований Лапласа-Стилтьеса.

Для значений $\omega \neq 0, \pm \infty$ справедливо

$$|a(-j\omega)\, b(j\omega)| < 1,$$

так как если бы при каком-либо значении $\omega \neq 0, \pm \infty$ одновременно выполнялось, что $a(j\omega) = 1$ и $b(j\omega) = 1$, то это бы означало согласно [11], что $A(t)$ и $B(t)$ одновременно являются решетчатыми, что противоречит сделанным на эти функции ограничениям. Следовательно, при $0 < |\omega| < \infty$ для функции $k(j\omega)$ справедливо

$$0 < |k(j\omega) = \left| \frac{1 - a(-j\omega)\, b(j\omega)}{-j\omega} \right|.$$

3. При сделанных выше на функции $A(t)$ и $B(t)$ ограничениях, изменение аргумента функции $k(j\omega)$ при изменении ω от $-\infty$ до $+\infty$ равно нулю. Действительно, все значения функции $a(-j\omega)\, b(j\omega)$ лежат в круге $|s| \leq 1$. Годограф функции $1 - a(-j\omega)\, b(j\omega)$ лежит в круге $|1 - s| \leq 1$ и лишь при $\omega = 0$ касается начала координат, причем

$$\frac{d}{d\omega} [1 - a(-j\omega) b(j\omega)] \Big|_{\omega=0} = \bar{a} - \bar{b} > 0,$$

то есть изменение аргумента функции $1-a(-j\omega)\,b(j\omega)$ при изменении аргумента ω от $-\infty$ до $+\infty$ равно -2π. Аргумент функции $\dfrac{1}{-j\omega}$ равен 2π. Сделанное утверждение справедливо, так как аргумент произведения равен сумме аргументов сомножителей.

Свойство функции $k(s)$ на контуре $\bar{C}_R$

Свойства функции $k(s)$ на отрезке $[-jR,\ jR]$ совпадают с рассмотренным ранее для мнимой оси $\operatorname{Re} s=0$. Аргумент s на дуге C_R будем обозначать $\operatorname{Re}^{j\varphi}$, $\varphi\in(-\pi,\ \pi)$. Будем считать, что для функций $A(t)$ и $B(t)$ выполняются все ограничения, сделанные выше.

1. На дуге C_R функция $k(s)$ удовлетворяет условию Гельдера, так как для точек $\operatorname{Re} s>0$ справедливо:

$$\left|\frac{da(s)}{ds}\right|=\left|-\int_0^\infty te^{-(a+j\omega)t}\,dA(t)\right|\le\int_0^\infty|t|\,|e^{-at}|\,|e^{-j\omega t}|\,|dA(t)|\le\bar{a};$$

$\left|\dfrac{db(s)}{ds}\right|\le\bar{b}$, так как $|e^{-at}|\le 1$ для $a\ge 0$.

2. На дуге C_R функция $k(s)$ в ноль не обращается, так как для $\operatorname{Re} s>0$ $|a(s)|<|a(j\omega)|$; $|b(s)|<|b(j\omega)|$, а $k(j\omega)\neq 0$ при $\omega\in[-R,R]$.

3. При обходе контура $\bar{C}_R$ изменение аргумента функции $k(s)$ равно нулю в силу соображений, высказанных ранее при доказательстве аналогичного утверждения при обходе мнимой оси $\operatorname{Re} s=0$.

Выражение (4.6) справедливо как на мнимой оси $\operatorname{Re} s=0$, так и на контуре $\bar{C}_R$, так как там определены функции, входящие в это выражение.

На контуре $\bar{C}_R$ выполняются условия задачи Римана [6] (на мнимой оси $\operatorname{Re} s=0$ эти условия не выполняются, так как функция $k(s)$ обращается в нуль в точке $s=j\infty$, принадлежащей этой оси), воспользовавшись решением которой получим

$$z_*^+(s)=A_1\exp\left\{-\frac{1}{2\pi j}\int_{\bar{C}_R}\frac{\ln k(\sigma)}{\sigma-s}\,d\sigma\right\}\quad\text{для } s\in\bar{C}_R,\tag{4.8}$$

$$z_*^-(s)=A_1\exp\left\{-\frac{1}{2\pi j}\int_{\bar{C}_R}\frac{\ln k(\sigma)}{\sigma-s}\,d\sigma\right\}\quad\text{для } s\notin\bar{C}_R,\tag{4.9}$$

где $s\in\bar{C}_R$ обозначает, что точка принадлежит внутренней области контура $\bar{C}_R$, а обход этого контура производится так, что внутренняя область остается слева; A_1 — произвольная постоянная.

Определим теперь функции $z^{\pm}_{**}(s)$ так, чтобы на контуре $\overline{C}_R$ выполнялось условие (4.7), причем напомним, что функция $z^{+}_{**}(s)$ — аналитическая в полуплоскости $\operatorname{Re} s \geqq 0$, а функция $z^{-}_{**}(s)$ — в полуплоскости $\operatorname{Re} s \leqq 0$ за исключением бесконечно удаленной точки, где она имеет полюс первого порядка. Если такие функции $z^{\pm}_{**}(s)$ найдены, то в силу условия (4.7), функции $z^{+}_{**}(s)$ и $-sz^{-}_{**}(s)$ на контуре $\overline{C}_R$ совпадают и, следовательно, образуют одну аналитическую во всей конечной плоскости функцию. Эта функция имеет в бесконечности полюс первого порядка, и, следовательно, согласно теореме Лиувиля [6]

$$z^{+}_{**}(s) = A_2, \tag{4.10}$$

$$z^{-}_{**}(s) = -sA_2. \tag{4.11}$$

Объединяя (4.8—4.11) — получим

$$z^{+}(s) = A \exp\left\{-\frac{1}{2\pi j}\int\limits_{\overline{C}_R}\frac{\ln k(\sigma)}{\sigma - s}\,d\sigma\right\} \quad \text{для } s \in \overline{C}_R, \tag{4.12}$$

$$z^{-}(s) = -sA \exp\left\{-\frac{1}{2\pi j}\int\limits_{\overline{C}_R}\frac{\ln k(\sigma)}{\sigma - s}\,d\sigma\right\} \quad \text{для } s \notin \overline{C}_R, \tag{4.13}$$

где

$$A = A_1 A_2,$$

откуда

$$w(s) = A \exp\left\{-\frac{1}{2\pi j}\int\limits_{\overline{C}_R}\frac{\ln k(\sigma)}{\sigma - s}\,d\sigma\right\} \quad \text{для } s \in \overline{C}_R, \tag{4.14}$$

$$y(s) = 1 - sA \exp\left\{-\frac{1}{2\pi j}\int\limits_{\overline{C}_R}\frac{\ln k(\sigma)}{\sigma + s}\,d\sigma\right\} \quad \text{для } s \in \overline{C}_R. \tag{4.15}$$

Постоянную A определим из условия

$$\lim_{s\to 0} w(s) = \lim_{t\to\infty} W(t) = 1.$$

С этой целью, воспользовавшись формулой Сохоцкого [6], рассмотрим значение функции $w(s)$ на контуре $\overline{C}_R$:

$$w(u) = Ak^{-\frac{1}{2}}(u) \exp\left\{-\frac{1}{2\pi j}\int\limits_{\overline{C}_R}\frac{\ln k(\sigma)}{\sigma - u}\,d\sigma\right\},$$

откуда

$$A = k^{\frac{1}{2}}(j0) \exp\left\{\frac{1}{2\pi j}\int\limits_{\overline{C}_R}\frac{\ln k(\sigma)}{\sigma}\,d\sigma\right\}. \tag{4.16}$$

Подставляя полученное значение A в выражения (4.14) и (4.15), получим

$$w(s) = k^{\frac{1}{2}}(j0)\exp\left\{-\frac{s}{2\pi j}\int\limits_{\overline{C}_R}\frac{\ln k(\sigma)}{\sigma(\sigma - s)}\,d\sigma\right\} \quad \text{для} \quad s \in \overline{C}_R, \tag{4.17}$$

$$y(s) = 1 - k^{\frac{1}{2}}(j0)\,s\exp\left\{\frac{s}{2\pi j}\int\limits_{\overline{C}_R}\frac{\ln k(\sigma)}{\sigma(\sigma + s)}\,d\sigma\right\} \quad \text{для} \quad s \in \overline{C}_R. \tag{4.18}$$

Переходя в полученных выражениях к пределу при $R \to \infty$, окончательно получим выражения (2.8) и (2.9), что и требовалось показать.

Замечание. Теорема, сформулированная в п. 2, остается справедливой, если функции распределения $A(t)$ и $B(t)$ являются решетчатыми, но интервалы их дискретности кратны некоторому числу Δ.

Действительно, в этом случае функция $k(j\sigma)$ является периодической с периодом $\frac{2\pi}{\Delta}$. Поэтому подстановкой $z = e^{\frac{2\pi}{\Delta}\sigma}$ можно эту задачу свести к рассмотренной в настоящем разделе.

5. Пример

В качестве примера рассмотрим важный частный случай, когда $A(t)$ и $B(t)$ представляют собой дробнорациональные функции:

$$a(s) = \frac{\prod\limits_{l=1}^{L}\left(1 + \frac{1}{a_{1,l}}s\right)}{\prod\limits_{k=1}^{K}\left(1 + \frac{1}{a_{2,k}}s\right)}, \qquad b(s) = \frac{\prod\limits_{n=1}^{N}\left(1 + \frac{1}{b_{1,n}}s\right)}{\prod\limits_{m=1}^{M}\left(1 + \frac{1}{b_{2,m}}s\right)}, \tag{5.1}$$

где $a_{i,r}$ и $b_{i,r}$; $i = 1, 2$, — вещественные положительные числа;

$$\overline{a} = \sum_{l=1}^{L}\frac{1}{a_{1,l}} - \sum_{k=1}^{K}\frac{1}{a_{2,k}}, \quad b = \sum_{n=1}^{N}\frac{1}{b_{1,n}} - \sum_{m=1}^{M}\frac{1}{b_{2,m}}.$$

Заметим, что, так как любая функция распределения представляет собой монотонно-возрастающую функцию, ее преобразование Лапласа-Стилтьеса не должно содержать комплексных нулей и полюсов. Это в одинаковой мере относится как к исходным функциям $a(s)$ и $b(s)$, так и к производным от них, искомым функциям $w(s)$, $y(s)$ и $c(s)$.

Из выражения (5.1) следует, что функции $A(t)$ и $B(t)$ удовлетворяют условиям теоремы, сформулированной в разделе 2.

Функция $k(s)$ в рассматриваемом случае имеет вид:

$$k(s) = \frac{1 - a(-s)\,b(s)}{-s} = c\,\frac{\prod_{p=1}^{P}\left(1+\frac{1}{d_{1,p}}s\right)\prod_{q=1}^{Q}\left(1-\frac{1}{d_{2,q}}s\right)}{\prod_{m=1}^{M}\left(1+\frac{1}{b_{2,m}}s\right)\prod_{k=1}^{K}\left(1-\frac{1}{a_{2,k}}s\right)}, \tag{5.2}$$

причем

$$P+Q=K+M-1; \tag{5.3}$$

$d_{i,j}$; $i=1,2$ — вещественные положительные числа, являющиеся, соответственно, положительными и отрицательными корнями многочлена

$$P(s)=\frac{1}{s}\left[\prod_{m=1}^{M}\left(1+\frac{1}{b_{2,m}}s\right)\prod_{k=1}^{K}\left(1-\frac{1}{a_{2,k}}s\right)-\right.$$
$$\left.-\prod_{p=1}^{P}\left(1+\frac{1}{d_{1,p}}s\right)\prod_{q=1}^{Q}\left(1-\frac{1}{d_{2,q}}s\right)\right], \tag{5.4}$$

которые могут быть определены, например, методом логарифмических корневых годографов [8];

коэффициент c определяется из условия

$$c = P(0) = \sum_{m=1}^{M}\frac{1}{b_{2,m}} - \sum_{k=1}^{K}\frac{1}{a_{2,k}} - \sum_{p=1}^{P}\frac{1}{d_{1,p}} + \sum_{q=1}^{Q}\frac{1}{d_{2,q}}. \tag{5.5}$$

Подставляя выражение (5.2) в формулу (4.17), получим:

$$w(s) = c^{\frac{1}{2}}\exp\left\{-\frac{s}{2\pi j}\int_{\overline{C}_R}\frac{\ln c}{\sigma(\sigma-s)}\,d\sigma-\right.$$
$$-\sum_{p=1}^{P}\frac{s}{2\pi j}\int_{\overline{C}_R}\frac{\ln\left(1+\frac{1}{d_{1,p}}\sigma\right)}{\sigma(\sigma-s)}\,d\sigma-\sum_{q=1}^{Q}\frac{s}{2\pi j}\int_{\overline{C}_R}\frac{\ln\left(1-\frac{1}{d_{2,q}}\sigma\right)}{\sigma(\sigma-s)}\,d\sigma+$$
$$\left.+\sum_{k=1}^{K}\frac{s}{2\pi j}\int_{\overline{C}_R}\frac{\ln\left(1-\frac{1}{a_{2,k}}\sigma\right)}{\sigma(\sigma-s)}\,d\sigma+\sum_{m=1}^{M}\frac{s}{2\pi j}\int_{\overline{C}_R}\frac{\ln\left(1+\frac{1}{b_{2,m}}\sigma\right)}{\sigma(\sigma-s)}\,d\sigma\right\} \tag{5.6}$$

для $s\in\overline{C}_R$.

Для определения первого интеграла, входящего в выражение (5.6), рассмотрим интеграл по контуру, показанному на рис. 2а. Согласно теореме о вычетах [12], получим:

$$\int_{\bar{C}_R} = \lim_{r\to 0}\left\{\int_{\bar{C}_R} + \int_{jR}^{jr} + \int_{-jr}^{-jR}\right\} = \lim_{r\to 0}\left\{\sum \operatorname{res}\frac{\ln c}{\sigma(\sigma - s)} - \int_{\bar{C}r}\right\};$$

$$\sum \operatorname{res}\frac{\ln c}{\sigma(\sigma - s)} = 2\pi j \lim_{\sigma\to s}\frac{\ln c}{\sigma} = \frac{2\pi j}{s}\ln c,$$

$$\int_{C_r}\frac{\ln c}{\sigma(\sigma - s)}d\sigma = \int_{\pi}^{-\pi}\frac{\ln c}{re^{j\varphi}(re^{j\varphi} - s)}re^{j\varphi}\,j\,d\varphi \xrightarrow[r\to 0]{} \frac{\ln c}{-s}j(-\pi) = \frac{\pi j}{s}\ln c,$$

$$\int_{\bar{C}_R} = \frac{2\pi j}{s}\ln c - \frac{\pi j}{s}\ln c = \frac{\pi j}{s}\ln c,$$

Используя полученные результаты, для первого интеграла из выражения (5.6) получим:

$$\exp\left\{-\frac{s}{2\pi j}\int_{\bar{C}_R}\frac{\ln c}{\sigma(\sigma - s)}d\sigma\right\} = c^{-\frac{1}{2}}. \tag{5.7}$$

Для определения второго (пятого) интеграла рассмотрим интеграл по контуру, показанному на рис. 2б. Внутри этого контура функция логарифма допускает выделение однозначной ветви. Будем рассматривать ветвь, удовлетворяющую на контуре $\bar{C}_R$ неравенству

$$-2\pi \le \arg\left(1 + \frac{1}{d_{1,p}}\sigma\right) \le 0.$$

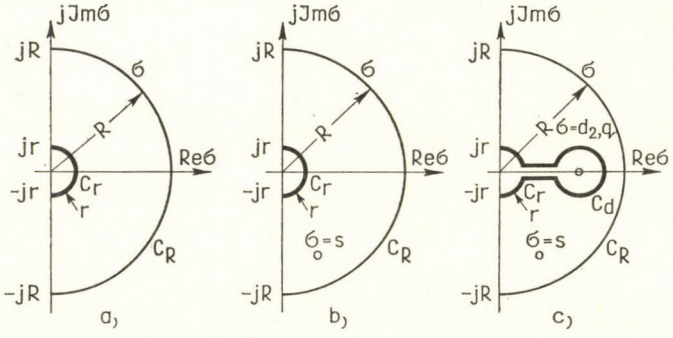


Рис. 2. Контуры интегрирования

Согласно теореме о вычетах, получим:

$$\int_{\overline{C}_R} = \lim_{r\to 0}\left\{\int_{C_R} + \int_{jR}^{jr} + \int_{-jr}^{-jR}\right\} = \lim_{r\to 0}\left\{\sum \operatorname*{res}_{\overline{C}_R} \frac{\ln\left(1+\frac{1}{d_{1,p}}\sigma\right)}{\sigma(\sigma-s)} - \int_{C_r}\right\};$$

$$\sum \operatorname*{res}_{\overline{C}_R} \frac{\ln\left(1+\frac{2}{d_{1,p}}\sigma\right)}{\sigma(\sigma-s)} = \frac{2\pi j}{s}\ln\left(1+\frac{1}{d_{1,p}}s\right),$$

$$\lim_{r\to 0}\int_{C_r} \frac{\ln\left(1+\frac{1}{d_{1,p}}\sigma\right)}{\sigma(\sigma-s)}\,d\sigma = \lim_{r\to 0}\int_{C_r} \frac{\ln\left(1+\frac{1}{d_{1,p}}re^{j\varphi}\right)}{re^{j\varphi}(re^{j\varphi}-s)}\,re^{j\varphi}j\,d\varphi = \frac{\ln 1}{-s}j(-\pi)=0,$$

$$\int_{\overline{C}_R} = \frac{2\pi j}{s}\ln\left(1+\frac{1}{d_{1,p}}s\right).$$

Используя полученные результаты, для второго (пятого) интеграла из выражения (5.6) получим:

$$\exp\left\{-\frac{s}{2\pi j}\int_{\overline{C}_R} \frac{\ln\left(1+\frac{1}{d_{1,p}}\sigma\right)}{\sigma(\sigma-s)}\,d\sigma\right\} = \frac{1}{1+\frac{1}{d_{1,p}}s}. \qquad (5.8)$$

Для определения третьего (четвертого) интеграла рассмотрим интеграл по контуру, показанному на рис. 2в. Внутри этого контура функция логарифма допускает выделение однозначной ветви. Будем рассматривать ветвь, удовлетворяющую на контуре $\overline{C}_R$ неравенству

$$-2\pi \le \arg\left(1-\frac{1}{d_{2,q}}\sigma\right) \le 0.$$

Согласно теореме о вычетах, получим

$$\int_{\overline{C}_R} = \lim_{r\to 0}\left\{\int_{C_R} + \int_{jR}^{jr} + \int_{-jr}^{-jR}\right\} = \lim_{r\to 0}\sum \operatorname*{res}_{\overline{C}_R} \frac{\ln\left(1-\frac{1}{d_{2\,q}}\sigma\right)}{\sigma(\sigma-s)} - \int_{C_r} - \int_{C_d} - \int_{r}^{d_{2,q}} -$$

$$- \int_{d_{2,q}}^{r} - \int_{d_{2,q}}^{r} \frac{-2\pi j}{\sigma(\sigma-s)}\,d\sigma - \int_{-jr}^{-jR} \frac{-2\pi j}{\sigma(\sigma-s)}\,d\sigma - \int_{\substack{r\\ \text{на } C_r}}^{-jr} \frac{-2\pi j}{\sigma(\sigma-s)}\,d\sigma - \int_{C_R} \frac{-2\pi j}{\sigma(\sigma-s)}\,d\sigma.$$

Последние четыре интеграла учитывают тот факт, что при обходе точки $\sigma = d_{2,q}$ имеет место изменение аргумента функции, стоящей под знаком логарифма, на величину -2π.

$$\sum \operatorname*{res}_{C_R} \frac{\ln\left(1-\frac{1}{d_{2,q}}\sigma\right)}{\sigma(\sigma-s)} = \frac{2\pi j}{s}\ln\left(1-\frac{1}{d_{2,q}}s\right),$$

$$\left|\int\limits_{C_r}\right| = \left|\int\limits_{C_r} \frac{\ln\left(1-\frac{1}{d_{2,q}}re^{j\varphi}\right)}{re^{j\varphi}(re^{j\varphi}-s)}\, re^{j\varphi}\, j\, d\varphi\right| \xrightarrow[r\to 0]{} \frac{\ln 1}{-s}\, j(-\pi) = 0,$$

$$\left|\int\limits_{C_d}\right| = \left|\int\limits_{C_d} \frac{\ln\left[1-\frac{1}{d_{2,q}}(d_{2,q}+re^{j\varphi})\right]}{(d_{2,q}+re^{j\varphi})(d_{2,q}+re^{j\varphi}-s)}\, re^{j\varphi}\, j\, d\varphi\right| \leq r\ln r(-2\pi) \xrightarrow[r\to 0]{} 0,$$

$$\int\limits_{r}^{d_{2,q}} + \int\limits_{d_{2,q}}^{r} = 0,$$

$$\int\limits_{d_{2,q}}^{r} \frac{-2\pi j}{\sigma(\sigma-s)}\, d\sigma = \frac{2\pi j}{s}\left[\ln r - \ln d_{2,q} - \ln(r-s) + \ln(d_{2,q}-s)\right],$$

$$\int\limits_{-jr}^{-jR} \frac{-2\pi j}{\sigma(\sigma-s)}\, d\sigma = \frac{2\pi j}{s}\left[\ln(-jR) - \ln(-jr) - \ln(-jR-s) + \ln(-jr-s)\right],$$

$$\int\limits_{\text{на } C_r}^{-jr} \frac{-2\pi j}{\sigma(\sigma-s)}\, d\sigma = \int\limits_{0}^{-\pi/2} \frac{-2\pi j}{re^{j\varphi}(re^{j\varphi}-s)}\, re^{j\varphi}\, j\, d\varphi \xrightarrow[r\to 0]{} -\frac{\pi^2 j}{s},$$

$$\left|\int\limits_{C_R} \frac{-2\pi j}{\sigma(\sigma-s)}\, d\sigma\right| = \left|\int\limits_{C_R} \frac{-2\pi j}{Re^{j\varphi}(Re^{j\varphi}-s)}\, Re^{j\varphi}\, j\, d\varphi\right| \leq \frac{2\pi^2}{R}.$$

Точные значения интегралов при конечном значении R получить не удается, поэтому рассмотрим предельный случай при $R \to \infty$.

$$\left|\int\limits_{\overline{C}_R}\right| \leq \left|\frac{2\pi j}{s}\ln\left(1-\frac{1}{d_{2,q}}s\right) + \frac{2\pi j}{s}\ln\frac{d_{2,q}}{d_{2,q}-s} + \frac{\pi j}{s}\ln\frac{jR}{jR+s} - \frac{2\pi^2}{R}\right|_{R\to\infty} \to 0.$$

Используя полученные оценки, для третьего (пятого) интеграла из выражения (5.6) получим:

$$\exp\left\{-\frac{s}{2\pi j}\int_{j\infty}^{-j\infty}\frac{\ln\left(1+\dfrac{1}{d_{1,p}}\sigma\right)}{\sigma(\sigma-s)}\,d\sigma\right\}=1\,. \tag{5.9}$$

Подставляя в выражение (5.6) результаты, полученные в (5.7—5.9), и переходя в этих выражениях к пределу при $R\to\infty$, получим

$$w(s)=\frac{\prod\limits_{m=1}^{M}\left(1+\dfrac{1}{b_{2,m}}s\right)}{\prod\limits_{p=1}^{P}\left(1+\dfrac{1}{d_{1,p}}s\right)}, \tag{5.10}$$

что совпадает с результатом, полученным для рассматриваемого случая в работе [5]. Аналогично, воспользовавшись соотношениями (4.18) и (2.4), получим:

$$y(s)=1-s\,\frac{\prod\limits_{q=1}^{Q}\left(1+\dfrac{1}{d_{2,q}}s\right)}{\prod\limits_{k=1}^{K}\left(1+\dfrac{1}{a_{2,k}}s\right)}, \tag{5.11}$$

$$c(s)=\left[1-s\,\frac{\prod\limits_{q=1}^{Q}\left(1+\dfrac{1}{d_{2,q}}s\right)}{\prod\limits_{k=1}^{K}\left(1+\dfrac{1}{a_{2,k}}s\right)}\right]\frac{\prod\limits_{n=1}^{N}\left(1+\dfrac{1}{b_{1,n}}s\right)}{\prod\limits_{m=1}^{M}\left(1+\dfrac{1}{b_{2,m}}s\right)}\,. \tag{5.12}$$

Выводы

1. Доказана теорема, расширяющая класс функций, для которых до настоящего времени были получены [5] аналитические выражения для преобразований Лапласа-Стилтьеса от функций распределения времени ожидания заявкой начала обработки, простоя прибора и выходящего потока заявок.

2. Разработан ряд приближенных методов расчета, основанных на использовании логарифмических корневых годографов и логарифмических частотных характеристик, являющихся более удобными в вычислительном отношении, чем существующие до настоящего времени [3].

Литература

1. *Воронов А. А., Чистяков, Ю. В.:* О выборе технических средств для автоматизированной системы управления материально-техническим снабжением. Автоматика и телемеханика *11* (1969).
2. *Воронов А. А., Чистяков Ю. В.:* Аналитические методы определения количества и типа аппаратуры в АСОЭИ. В сб. «Вопросы создания автоматизированных систем управления», Изд. МЭСИ, М., 1969.
3. *Кузин Л. Г.:* Частотные методы расчета систем массового обслуживания. В сб. «Современные методы проекти рования системавтоматического управления», Изд. «Машиностроение», М., 1967.
4. *Кофман А., Крюон Р.:* Массовое обслуживание. Теория и приложение, Изд. «Мир», М., 1965.
5. *Климов Г. Е.:* Стохастические системы массового обслуживания Изд. «Наука», М., 1967.
6. *Гахов Ф. Д.:* Краевые задачи. Физматгиз. М., 1958.
7. *Петров П. Г.:* Распределение времени ожидания в стационарном режиме для однолинейной системы обслуживания с «разогревом». В. сб. «Вычислительные методы и программирование». Вып. *IX*, Изд. МГУ, И., 1967.
8. Теория автоматического регулирования. кн. 1, Изд. «Машиностроение», М., 1967.
9. *Двайт Г. Б.:* Таблицы интегралов и другие математические формулы. Изд. «Наука», М., 1969.
10. *Солодовников В. В.:* Статистическая динамика линейных систем автоматического управления. Физматгиз, М., 1960.
11. *Ибрагимов И. А., Линник Ю. В.:* Независимые и стационарно связанные величины. Изд. «Наука», М., 1965.
12. *Лаврентьев М. А., Шабат Б. Е.:* Методы теории функций комплексного переменного. Изд. «Наука», М., 1965.

Approximate methods of description of one-channel queueing system

A. VORONOV—YU. CHISTYAKOV

Moscow

Summary

One-channel queueing system fed by recurrent flow of demands determined by the distribution function $A(t)$ is considered. Servicing times of each demand are mutually independent and identically distributed and determined by the distribution function $B(t)$. Demands arriving in the system are serviced immediately if a server is free, and get in a waiting line if it is busy servicing an earlier arrival. The length of the waiting line is supposed to be unlimited, demands are being serviced in order of their arrival.

For the above mentioned system the following theorem is proved.

Theorem: If the distribution functions $A(t)$ and $B(t)$ are not latticed at one and the same time and conditions (2.5)—(2.7) are satisfied then in a stationary process for the Laplace-Stielties transformation from the distribution function $W(t)$ — the time the demand waits for the beginning of the servicing and $Y(t)$ — the time the device waits for the demand, the equations (2.8) and (2.9) are fulfilled.

Restrictions emposed by the conditions of the theorem may be easily checked.

For the functions $A(t)$ and $B(t)$ the calculation method based on the frequency characteristic of determining the functions $W(t)$ is worked out. The determining of the functions $W(t)$ and $Y(t)$ is supposed to be carried out by way of the trapeze frequency characteristics. To determine them the well-known tables and nomogramms can be used.

To calculate singular integrals (3.12) and (3.13) a piece linear approximation of the under integral expressions are suggested to be used.

The approximating functions (3.17) and (3.19) are selected in such a way that their graphs at a random value of argument can be determined by way of changing their scale and parallel displacement of their graphs of the functions having certain value of arguments for which tables are given.

As an example an important case is considered where the Laplace-Stielties transformation from the functions $A(t)$ and $B(t)$ is a ratio of rational functions. The obtained equations (5.10) and (5.11) correspond to the equations obtained for a similar case of [5], which confirms the correctness of our results.

А. А. Воронов, Ю. В Чистяков
Институт проблем управления (автоматики и телемеханики)
СССР Москва В-485
Профсоюзная ул. 81

NOTE TO CONTRIBUTORS

Two copies of the manusripts (each duly completed by figures, tables and references) are to be sent either 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

J. Kocsis coordinating editor
Department of Automation
Technical University
Budapest XI, Garami Ernő tér 3, Hungary

The authors are requested to retain a third copy of the submitted typescript to be able to check the proofs against it.

The papers, preferably in English or Russian, should be typed double-spaced on one side of good-quality paper, with wide margins (c. 4–5 cm) should carry the title of the contribution, the author(s)' name, and the name of the country. At the end of the typescript the name of that author who manages the proof-reading should also be given.

An abstract of about 50 to 100 words should head the paper.

The authors are encouraged to use the following headings: Introduction, (outlining the problem), Methods and results, Discussion, Conclusions, References. The entire material should not exceed 15 pages including tables and references. The proper location of the tables and figures must be indicated on the margins. Mathematical notations should follow up-to-date usage.

The summary – possibly in Russian if the paper is in English and *vice-versa* – should contain a brief account of the proposition and indications of the formulas used and figures shown in the paper. The summary is not supposed to exceed 10–15 per cent of the paper.

The authors will be sent sheet-proofs which they are to return by next mail to the sender Regional Editorial Board.

Authors are entitled to 100 reprints free of charge. Rejected manuscripts will be returned to the authors.

К СВЕДЕНИЮ АВТОРОВ

Рукописи в двух экземплярах (каждый из которых должен содержать рисунки, таблицы и литературу) направляются

Е. Д. *Теряеву* — Научный секретарь журнала
Отделение механики и процессов управления
Академия Наук СССР
Ленинский Проспект, 14, Москва В-71, СССР

или

Я. Кочишу — Научный секретарь журнала
Кафедра Автоматизации
Будапештского Технического
Университета, Будапешт XI, площадь
Гарами Эрнё, 3, Венгрия

Авторам рекомендуется оставлять у себя копию всех представленных ими материалов для справок при корректуре.

Статьи, желательно на русском или английском языках, отпечатанные на бумаге хорошего качества, с промежутком в два интервала и широкими (4–5 см) полями должны содержать наименование статьи, фамилию автора (авторов), название страны. В конце статьи необходимо также указать фамилию автора, ответственного за корректуру гранок.

Статье должна предшествовать аннотация объемом до 50—100 слов.

Авторы при написании статьи должны придерживаться следующей формы: введение (постановка задачи), основное содержание и результаты, обсуждение, выводы и литература. Обьем статьи не должен превышать 15 печатных страниц, включая таблицы и ссылки. Последовательность таблиц и рисунков должна быть отмечена на полях. Математические обозначения рекомендуется давать в соответствии с современными требованиями и традициями.

К статье обязательно должно быть приложено резюме-реферат. Резюме — на русском языке, если статья написана на английском, и наоборот — должно содержать краткое изложение текста статьи со ссылками на необходимые формулы и графики, имеющиеся в основном тексте. Объем резюме не должен превышать 10—15% объема статьи.

Авторам высылаются гранки статьи, которые они должны незамедлительно возвратить в Региональную секцию Редколлегии журнала.

Авторам обеспечивается бесплатно 100 оттисков их статей. Рукописи непринятых статей возвращаются авторам.

CONTENTS • СОДЕРЖАНИЕ

VOL. 1 * NUMBER 2
ТОМ 1 * НОМЕР 2

ACADEMY OF SCIENCES OF THE USSR
HUNGARIAN ACADEMY OF SCIENCES

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

АКАДЕМИЯ НАУК СССР
АКАДЕМИЯ НАУК ВЕНГРИИ

1972

AKADÉMIAI KIADÓ, BUDAPEST

PROBLEMS OF CONTROL AND INFORMATION THEORY

is an international quarterly sponsored jointly by the Presidium of the Academy of Sciences of the U. S. S. R. and of the Hungarian Academy of Sciences. It offers publicity for original papers and short communications on the following topics:

- general theorz of control systems and system theory
- theory of automata
- information theory
- operation research; theory of complex systems
- theory of economic control; system modelling
- theory and methods of adaptation, learning, identification and pattern recognition
- methods of information processing; application of digital computers in control and communication systems

new physical principles in constructing technical devices for automation, control and information processing

The four issues published per year make up a volume of some 320 pp.

While this quarterly is mainly a publication forum of the research results achieved within the U. S. S. R. and Hungary, also papers of international interest from other countries are welcome.

Distributor:

KULTURA

Hungarian Trading Co. for Books and Newspapers

Budapest 62, P.O. Box 149, Hungary

The quarterly is published by

AKADÉMIAI KIADÓ

Publishing House of the Hungarian Academy of Sciences

Budapest 502, P.O. BOX 24, Hungary

Президиумами Академии Наук СССР и Академии Наук ВНР было принято решение о совместном издании журнала

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

Журнал создан для обеспечения быстрейшей публикации материалов о новейших результатах научных исследований, разработок и кратких сообщений о достижениях в следующих областях науки:

- общая теория процессов управления и теория систем
- теория автоматов
- теория информации
- исследование операций; теория сложных систем
- теория управления экономическими системами; модели систем
- теория и методы адаптации, обучения, идентификации и распознавания образов
- методы обработки информации; применение вычислительных машин в системах управления и передачи информации
- новые физические принципы создания технических средств автоматики, управления и передачи информации

Журнал издается четыре раза в год общим объёмом около 320 печатных страниц.

Журнал в первую очередь будет публиковать научные достижения обеих стран, а также статьи из других стран, представляющие международный интерес.

Распространитель:

KULTURA

Венгерское Общество по распространению книг и журналов

Будапешт 62, п. о. 149, Венгрия

Издатель журнала:

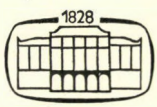
AKADÉMIAI KIADÓ

Издательство Академии Наук Венгрии

Будапешт 502, п. о. 24, Венгрия

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ



AKADÉMIAI KIADÓ

PUBLISHING HOUSE OF THE HUNGARIAN ACADEMY OF SCIENCES
BUDAPEST

Printed in Hungary
AKADÉMIAI NYOMDA, BUDAPEST

Problems of Control and Information Theory, Vol. 1 (2), pp. 103—115 (1972)

ИЗМЕРЕНИЯ И ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНЫЕ СИСТЕМЫ

Б. С. СОТСКОВ

(Москва)

Измерения являются одной из важнейших составляющих научных исследований и совершенствования технологических процессов. В статье рассматривается структура современных измерительных устройств, вопросы развития новых физических принципов их построения, методов и средств хранения и обработки результатов измерения, применения вычислительных устройств и машин.

Определяется ряд основных задач для современного этапа развития средств измерительной техники: изыскание новых физических принципов построения, развитие средств передачи измерительной информации, развитие средств хранения и обработки информации, автоматизация основных и вспомогательных процессов при измерении. Одной из основных задач ставится стандартизация параметров сигналов, параметров источников питания, форм и параметров носителей информации при записи и способов записи и воспроизведения: унификация конструктивных форм и мест соединения.

Измерение — процесс получения количественных данных (количественной информации) об изучаемой величине или параметре — является одной из основных составляющих научных исследований.

Эффективность научных исследований определяется не только наличием квалифицированных научных кадров, но и наличием достаточно полной и своевременной научной информации и вооруженностью современными средствами исследований, измерений и обработки результатов измерений. Только те области науки, где соблюдаются эти условия, успешно развиваются и приносят новые, часто исключительно важные научные результаты. Достаточно указать на развитие космических исследований, исследований в ядерной физике, радиоастрономии и ряде других областей науки.

Любое измерительное устройство (прибор) можно разделить на три основные функциональные части: *первую* («датчик»), воспринимающую воздействие измеряемой величины (параметра) — «x» и создающую на выходе изменение другой величины — «y» («естественный» или «первичный» сигнал), однозначно связанной со значением измеряемой величины; *вторую* — производящую сравнение величины «y» с некоторым единичным значением («мерой») $y = y_0$ и формирующим выходной сигнал z; *третью* («индикатор»)

— служащую для превращения выходного сигнала z в другой сигнал z^* путем линейного или углового перемещения стрелки (указателя) по градуированной шкале или на экране осциллографа, имеющего градуировку.

Измерительное устройство может состоять из указанных трех отдельных частей либо они могут быть конструктивно объединены в одно целое («измерительный прибор»).

В случае, если между датчиком и измерительной частью прибора, где производится сравнение сигнала датчика с единичным значением («мерой»), либо между измерительной частью и индикатором имеется большое расстояние, то приходится вводить четвертую часть — элементы для передачи сигнала от датчика или измерительной части по тем или иным каналам связи. В этом случае измерительное устройство называется телеметрическим.

Развитие современной измерительной техники определяется резким увеличением числа величин (параметров), подлежащих измерению. Только для прикладных измерений (в технике, сельском хозяйстве, медицине) в настоящее время требуется измерять около 2 000 величин (параметров), а существующие методы и средства позволяют измерить всего лишь около 450—500 величин (параметров).

Все это вызывает необходимость коренного и опережающего развития теории, методов и средств измерений.

В технике измерения используются как прямые измерения интересующей величины или параметра, так и более сложные — косвенные и совокупные измерения, когда измеряемая величина определяется путем обработки результатов измерений других, связанных с ней, величин.

I

Первая и наиболее трудная задача при создании новых средств измерения состоит в использовании уже известных и в изыскании новых еще не использованных физических эффектов и явлений для построения частей (органов), воспринимающих воздействие измеряемой величины (параметра) и создающих на выходе «естественный» (или «первичный») сигнал.

1. Обычная («классическая») структурная схема построения измерительного прибора была

$$x \to l \to l^*$$

или

$$x \to f \to l \to l^*,$$

где f — усилие, l — перемещение, l^* — перемещение стрелки или указателя по градуированной шкале.

Эта структурная схема продолжает применяться и развиваться. Здесь, кроме метода непосредственной оценки, широко используются компенсационные методы, при которых производится компенсация перемещения (реже скорости) или усилия. Компенсационные измерительные приборы обычно обладают меньшей погрешностью, что определяет их широкое использование.

Однако возможности построения измерительных приборов по данной структурной схеме относительно невелики. Кроме того, выходной сигнал в виде перемещения стрелки или указателя неудобен для передачи на расстояние или для дальнейшей автоматической обработки. В связи с этим широкое применение за последние 25—30 лет получили электрические методы измерения неэлектрических величин.

2. В электрических методах измерения неэлектрических величин изменение любых измеряемых (механических, акустических, тепловых, оптических, электрических, магнитных) величин приводит к изменению какого-либо из электрических параметров: сопротивления — ΔR, емкости — ΔC, индуктивности или взаимоиндуктивности — ΔL или ΔM, ЭДС — Δe и т.д.

Для этого используются различные физические явления и эффекты, связанные с изменением электрических (ϱ, ε, e) или магнитных (μ) свойств веществ. Начинают использовать двойные эффекты (эффект Холла, эффект Кикоина и т. п.), когда изменение выходного электрического параметра (обычно Δe) определяется совместным воздействием двух физических величин.

Полученное изменение электрического параметра ΔR, ΔC, ΔL, ΔM или Δe используется в измерительной схеме для сравнения с некоторым нормальным («образцовым», «эталонным») значением того же параметра и формирования выходного сигнала.

При использовании постоянного тока происходит формирование выходного сигнала напряжения или тока, значение которого пропорционально значению изменения электрического параметра (ΔR, ΔC, ΔL и ΔM или Δe) и, следовательно, измеряемой величины (параметра) «x».

При использовании импульсных измерительных схем выходной сигнал может формироваться путем изменения амплитуды, длительности, фазы, частоты или числа импульсов.

Широко используются измерительные схемы переменного тока, которые могут формировать выходные сигналы с изменяющейся в зависимости от значения измеряемой величины (параметра) «x» амплитудой, частотой или фазой напряжения или тока.

Используются измерительные схемы переменного тока с частотами питающего тока от десятков герц до десятков и сотен мегагерц. Приме-

нение токов высокой частоты при резонансных и фазовых методах формирования выходного сигнала позволяет достигать весьма высокой чувствительности измерительных устройств и приборов для измерения многих физических величин (параметров).

Дальнейшим естественным шагом является переход к миллиметровому и оптическому диапазону электромагнитных колебаний, что позволит ввести ряд новых методов формирования выходного сигнала.

Однако, несмотря на чрезвычайно широкие возможности, даваемые электрическими методами измерения неэлектрических величин, в ряде случаев они не могут быть непосредственно использованы и приходится применять вспомогательные физические процессы или химические реакции.

Среди вспомогательных физических процессов первое место занимает использование эффектов интегрального или селективного поглощения или отражения потоков различных излучений:

а) акустических (инфразвуковых, звуковых, ультразвуковых);

б) электромагнитных (различных радиоволновых, инфракрасных, видимых, ультрафиолетовых, рентгеновских, γ-излучений);

в) α, β и нейтронных.

Изменение интегрального или селективного потока излучений воспринимается одним из приемников излучения с изменяющимся R или e, а затем происходит дальнейшее формирование сигнала, как и в случае электрических измерений неэлектрических величин.

Кроме излучений, в случае измерений свойств и состава веществ, применяются вспомогательные физические процессы, связанные:

1. с преобразованием фазового состояния пробы:
   (1.1. — с конденсацией пробы,
   1.2. — с испарением пробы);
2. с сорбционными процессами:
   (2.1. — с абсорбционными процессами,
   2.2. — с адсорбционными процессами,
   2.3. — с хроматографическим разделением смеси).

Вспомогательные химические реакции используются для того, чтобы получить:

1. Изменение объема пробы.
2. Изменение тепловых свойств или возникновение тепловых эффектов.
3. Изменение электропроводности жидкости или ионизированного газа.
4. Изменение оптических свойств (цвета, прозрачности и т. п.).

Особую, очень важную и быстро развивающуюся группу составляют спектроскопические и резонансные методы.

Наибольшее применение они получили для определения состава и структуры веществ, контроля содержания H_2O и НО, хода химических процессов, измерения ряда физических величин. Применяются методы:

1. Оптической спектроскопии (инфракрасной, видимой, ультрафиолетовой).
2. Рентгено- и γ-спектроскопии.
3. Масс-спектроскопии.

Все более широкое применение находят методы, использующие оптический и ядерный резонанс:

1. Электронный парамагнитный резонанс (ЭПГ).
2. Ядерный магнитный резонанс (ЯМР).
3. Ядерный квадрупольный резонанс (ЯКР).
4. Ядерный γ — резонанс (эффект Мессбауэра).
5. Методы, использующие газовое микроволновое резонансное поглощение.

Сокращаются сроки между моментом открытия нового физического эффекта или явления и временем его применения для построения средств получения первичной информации в измерительной технике или автоматике. Если для эффекта Холла этот срок был около 60 лет, то для эффекта Мессбауэра он составил всего 4—6 лет. Дальнейший успех развития измерительной техники в значительной мере будет определяться привлечением новых, еще неиспользованных физических эффектов и явлений.

II

Другим важным изменением в современной измерительной технике является быстрое и разнообразное развитие методов и средств обработки измерительной информации.

До последнего времени измерительное устройство (измерительный прибор) имело, кроме воспринимающей части («датчика») и измерительной части, *индикатор* в виде отдельного прибора или электронного осциллографа и *регистратор* в виде самопишущего прибора или шлейфного осциллографа (рис. 1).

В настоящее время эта часть измерительного устройства коренным образом изменилась. Кроме индикатора и регистратора (стрелочного прибора, электронного осциллографа, самописца и шлейфного осциллографа) для обработки результата измерений (выходного сигнала измерительной части) в аналоговой форме (рис. 1) используются:

а) различные анализаторы:

а.а) амплитудные,

а.б) частотные,

а.в) для статической обработки данных измерений (включая корреляграфы),

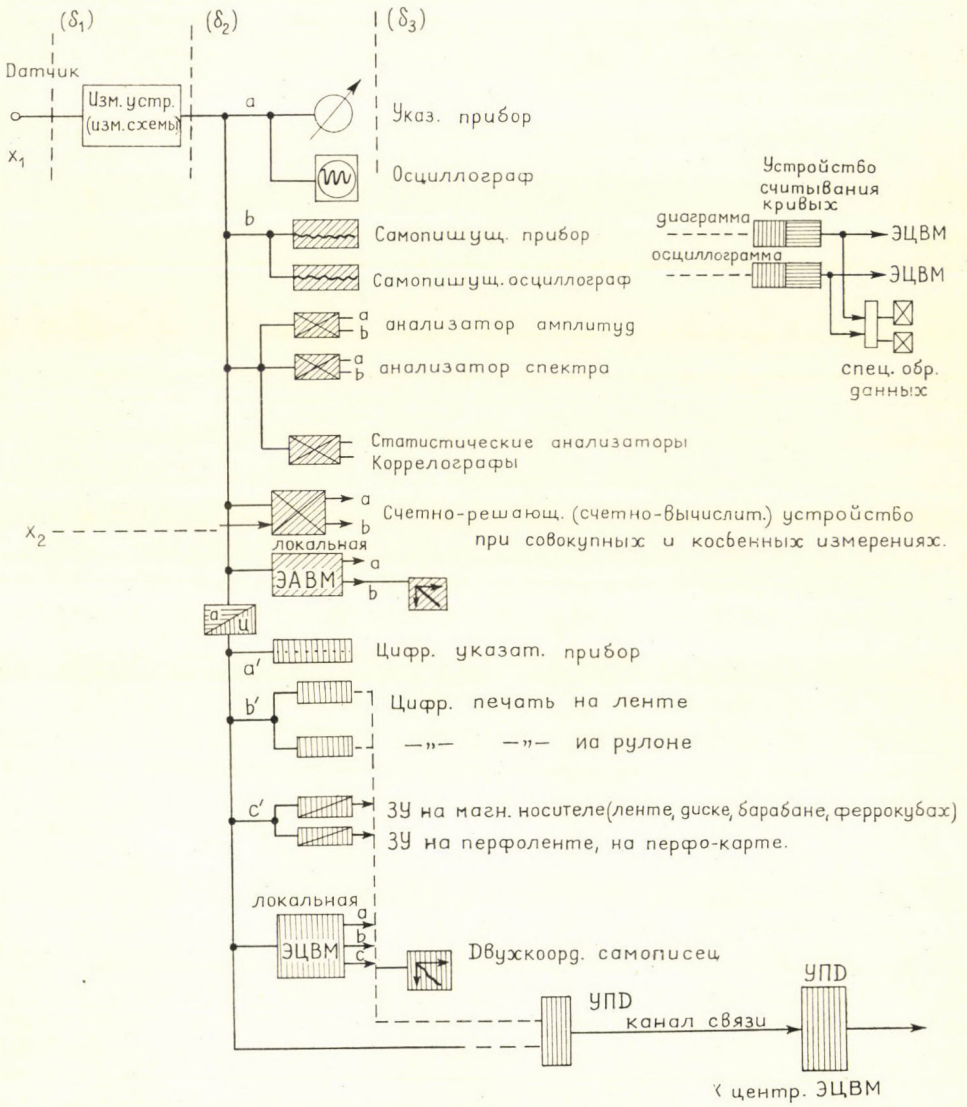


Рис. 1

б) аналоговые счетно-вычислительные устройства и машины, необходимые:
   б.а) для косвенных и совокупных измерений,
   б.б) для коррекции динамических погрешностей измерительного устройства,

в) аналоговые вычислительные машины (АВМ) — для обработки результатов в соответствии с заданной зависимостью и представления результата в виде графиков.

Появились и все шире используются автоматы для считывания и обработки записей осциллограмм и диаграмм и передачи результатов обработки в ЭЦВМ.

Кроме того, через преобразователь «аналог-цифра» и на выходе измерительного устройства могут быть включены:

а) цифровой указатель (цифровой индикатор);

б) устройства памяти — на перфолентах и перфокартах;
на магнитных лентах, барабанах, дисках, ферромагнитных кубах;

в) местные ЭЦВМ («мини-ЭЦВМ») — для цифровой обработки результатов измерений на месте:

г) устройства передачи данных через каналы связи (провода, радио и т. п.) в центральные ЭЦВМ.

Важнейшей задачей в настоящее время является установление общих технических принципов построения средств измерительной техники: стандартизации параметров выходных сигналов для измерительной части и для выходных сигналов АВМ, устройств памяти и ЭЦВМ, стандартизации форм и параметров носителей, параметров источников сигналов, унификации конструктивных форм и соединительных элементов. Это позволяет строить необходимые измерительные системы и комплексы из типовых составных частей, что значительно удешевит и ускорит создание средств измерительной техники. В настоящее время вошла в практику система УРС (ГСП) — для промышленных устройств контроля и управления, разработаны и находят применение системы типовых конструктивных приборных модулей: NIM (Nuclear Instrument Module); DIM (Digital Instrument Module); ADIOS (Analog Digital Input-Output System) и, особенно, CAMAC (Computer Application to Measuring and Control). Целесообразно установление одной из этих систем (например, CAMAC) в качестве типовой для измерительных приборов и устройств для научных исследований.

Другим важнейшим вопросом является унификация носителей и способов записи информации и ее воспроизведения. Быстрое, но некоордини-

рованное развитие средств и способов записи привело к тому, что большое количество научной и технической информации, накопленное за последние годы, скоро (через 3—5 лет) не сможет быть использовано вследствие выхода из употребления тех частных методов и средств, с помощью которых она может записываться и воспроизводиться. Необходимо установление, например, с помощью ISO, стандартных форм и параметров носителей и способов записи и воспроизведения. Это позволит создать единые язык и средства для регистрации и хранения измерительной информации, что

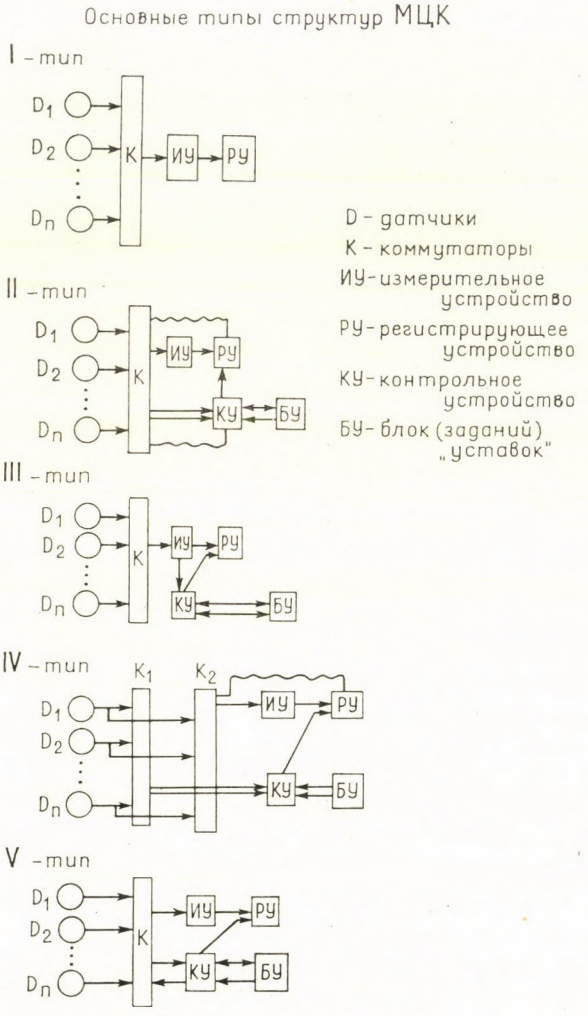


Рис. 2

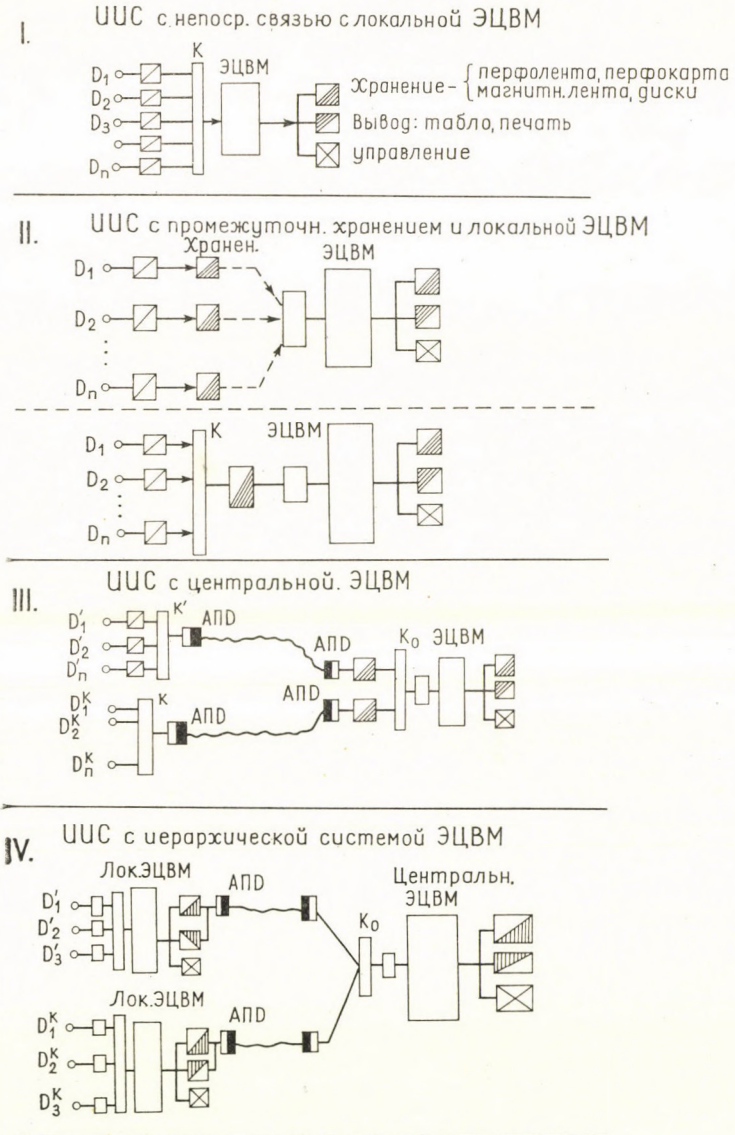

I.
ИИС с непоср. связью с локальной ЭЦВМ
K
ЭЦВМ
D1
D2
D3
Dn
Хранение – { перфолента, перфокарта / магнитн. лента, диски
Вывод: табло, печать
управление
II.
ИИС с промежуточн. хранением и локальной ЭЦВМ
Хранен.
ЭЦВМ
D1
D2
Dn
K
ЭЦВМ
III.
ИИС с центральной. ЭЦВМ
K'
АПД
АПД
K0
ЭЦВМ
АПД
K
АПД
IV.
ИИС с иерархической системой ЭЦВМ
Лок.ЭЦВМ
АПД
Центральн. ЭЦВМ
K0
Лок.ЭЦВМ
АПД

Рис. 3

облегчит обмен научной информацией и проведение совместных исследований, которые все шире и шире развиваются в различных областях науки (космические и океанологические исследования, исследования по программе Международного геофизического года и т. п.).

III

В случае необходимости концентрации результатов измерения в одном месте широко применяются так называемые машины централизованного контроля (МЦК) (рис. 2). В них производится периодический или с определенным приоритетом сбор первичных сигналов от датчиков, централизованное проведение операций измерения и централизованная регистрация их результатов. Часто, параллельно с измерением производятся операции контроля, т. е. сравнение с заданными для данной величины (параметра) значениями—«уставками». При выходе величины (параметра) за заданные верхние или нижние допустимые пределы производится регистрация их значения, а также подаются предупреждающие сигналы персоналу или вводятся регулирующие воздействия в контролируемый процесс. В этом случае МЦК обладают некоторыми начальными функциями управляющих вычислительных машин.

Применение МЦК целесообразно вследствие относительно небольшой стоимости для контроля многих технологических процессов и вспомогательного контроля работоспособности мощных исследовательских установок.

Другим методом концентрации результатов измерений, более широким по своим возможностям, является применение центральных ЭЦВМ, связанных с измерительными устройствами посредством проводных линий связи или специальных каналов передачи данных, включающих аппаратуру, позволяющую устранять искажения сигналов (АПД).

Обработка результатов измерений происходит непосредственно или после временного накопления данных в буферных устройствах памяти (рис. 3): перфокартах, перфолентах, магнитных носителях (лента, диски), магнитных кубах. Оптимальное планирование и организация передачи данных измерения и их обработки является сложной задачей, определяемой особенностями объекта измерений и возможными режимами его работы.

Одной из существенных задач является передача данных измерения к месту их обработки и регистрации, что стимулирует развитие различных средств телеметрии для передачи данных в аналоговой и цифровой форме по различным каналам связи: проводным, радио, акустическим и оптическим.

Трудности создания помехоустойчивых каналов связи и необходимость лучшего их использования привели к появлению и развитию методов и

средств для сжатия передаваемой информации и повышения помехоустойчивости передачи данных.

Изучение необычных сред: космоса, глубин океана, глубин земли — требует разработки новых специализированных средств телеметрической передачи данных. Эта работа успешно проводится в ряде стран.

IV

Наконец, важнейшей особенностью современной измерительной техники является автоматизация процессов измерения, включая все подготовительные и вспомогательные операции.

Наиболее ярким примером является применение разнообразных автоматических космических станций и спутников. Вообще, при исследованиях в необычных средах: космосе, глубинах океана и глубинах земли — роль автоматических станций является первостепенной и их развитию должно уделяться непрерывное и самое большое внимание. Развитие гидрометеорологической службы связано с применением автоматических радио-метеостанций (АРМС) и радио-зондов. Современные океанологические исследования включают широкое использование автоматических якорных и дрейфующих станций («буев») и автоматические и полуавтоматические буксируемые станции. Все это объединяется центрами обработки информации на исследовательских судах.

Аналогично строятся службы для геофизических исследований, с той разницей, что часто обработка всей информации производится в одном общем ВЦ.

Многие современные физические исследования, в первую очередь исследования в ядерной физике, требуют многочисленных одновременных измерений, быстрой совместной обработки результатов измерений и, в случае необходимости, вмешательства в проводимый эксперимент. Для обслуживания сложных экспериментов приходится использовать ряд современных ЭЦВМ, часть которых работает по предварительной обработке данных и подготавливает входные данные в быстродействующие центральные ЭЦВМ.

Построение таких сложных информационно-измерительных систем с иерархической структурой ставит ряд новых проблем.

Проведение массовых физических, химических, биологических и медицинских лабораторных исследований становится невозможным без применения типовых или специализированных автоматических измерительных установок и даже без создания автоматических или автоматизированных измерительных лабораторий. За последние 5—8 лет во всех странах мира приборостроительная промышленность разработала и выпустила боль-

шое число автоматических установок и лабораторий различного назначения. Многие из этих установок могут работать как с местными малогабаритными ЭЦВМ («мини-ЭЦВМ»), так и в режиме разделения времени с мощными центральными ЭЦВМ.

V

Выше были изложены наиболее характерные черты современного развития техники измерений, которые вызвали ее коренные изменения за последние 10—15 лет. Есть еще много важных задач, связанных, в первую очередь, с повышением точности и быстродействия средств измерения, созданием эталонов и образцов, развитием теории, принципов построения и методов расчета различных видов измерительных приборов, устройств и т. п., которые должны получить свое дальнейшее развитие.

Однако основными задачами являются: коренное расширение возможностей измерительной техники, путей изыскания все новых и новых физических принципов построения измерительных приборов и устройств, развитие методов и средств хранения, передачи и обработки результатов измерения с применением различных ВМ, полная или частичная автоматизация исследований. Решение этих задач требует создания современной теории измерений и информационно-измерительных систем.

Measurements and information-measurement systems

B. S. SOTSKOV

(Moscow)

Measurements is one of the most important fields of scientific research. Only those branches of science advance successfuly that are equipped with modern hardware for data acquisition, transmission, storage and processing.

Fig. 1 represents a block-diagram of a modern meter. The rapid increase in the number of quantities to be measured dictates a search for new physical principles for the construction of elements that would convert the effect of the given quantity (or a parameter) into a signal. Various physical phenomena and effects inside substances cause a change in an electrical or magnetic parameter of the substance under the action of the quantity measured. Methods using various radiations such as acoustic, radio, optical, X-ray, radioactive are widespread. Auxiliary physical processes such as the transition from one state into another, sorption, etc. or auxiliary chemical reactions cause changes in electrical, optical, heat, or mechanical parameters leading to the appearance of a signal that characterizes the quantity to be measured. Scientific instrumentation also makes full use of various spectral methods and electronic and nuclear resonance methods.

The comparison against a unit reference signal leads to the formation of an output signal which characterizes the quantitative value of the variable to be measured in the accepted measurement units. Frequency and phase methods using high and very high frequencies are very important in this connection.

A modern meter may contain at its output: the right-hand part of Fig. 1: *a)* an indicator, as a dial meter or oscilloscope; *b)* a recorder, as an automatic recorder or an oscilloscope whose records can be read and processed by and transmitted to a digital

computer; *c)* an analyzer, for amplitude, frequency or probabilistic estimates; *d)* analog computing units, employed to compute the result in indirect or combined measurements or to introduce corrections for dynamic responses of the meter.

The application of analog-to-digital converters enables to use: *a)* digital indicators; *b)* different units to record and store digital information, punched card, punched tape, magnetic tape, discs or drums, or prints; *c)* desk-size (mini) computers employed to process the measurement results or prepare the data for the central digital computer; *d)* central digital computers which receive the measurement data for final processing via communication channels and hardware.

Extensive use of digital computers, for on-line processing of samples taken from a multitude of parameters, is a characteristic feature of modern instrumentation.

In many cases digital computers are replaced by central monitoring processors that collect and estimate the measurement results. Fig. 2 represents the main types of such equipments.

One major problem in the development of instrumentation hardware is to work out general principles underlying their structure, including: the standardization of signal parameters for meters, analog computers, digital computers, and memory units, standardization of supply source parameters, of structural units and interfaces.

Another major problem is the automation of both basic and auxiliary operations in the process of measurements.

Б. С. Сотсков, чл.-корр. АН СССР
Институт проблем управления
СССР Москва В—485
Профсоюзная ул. 81

Problems of Control and Information Theory, Vol. 1 (2), pp. 117—121 (1972)

ON A GENERALIZATION OF THE STATISTICAL LINEARIZATION METHOD

I. N. SINITSIN

(Moscow)

The method of studying the accuracy of non-linear non-stationary stochastic systems based on the statistical linearization of non-linear functions by means of canonical expansions of random functions is given.

The suggested scheme of statistical linearization preserves the coefficients of the canonical expansions. The mathematical expectation and co-ordinate functions are obtained by minimizing the mean square error. The study of the accuracy of non-linear non-stationary stochastic systems is reduced to the investigation of the non-linear non-stationary deterministic systems for mathematical expectations and co-ordinate functions.

1. Introduction

The method of studying the accuracy of non-linear non-stationary stochastic systems based on canonical expansions (canonical expansions and integral canonical expansions) of random functions is given in [1]. In practice this method needs in addition to the continuity of the non-linear function also the existence of its derivatives up to a given order. It is often necessary to deal with the stochastic systems which contain discontinuous non-linear functions. The proposed method of studying the accuracy of such systems generalizes the well known statistical linearization method [1, 2]. The suggested scheme of statistical linearization preserves the random coefficients of the canonical expansions. The mathematical expectation and co-ordinate functions are obtained by minimizing the mean square error.

It is important to mention that the non-correlated random coefficients are not necessary Gaussian. The study of the accuracy of the non-linear non-stationary stochastic systems is reduced to the investigation of the non-linear non-stationary deterministic systems for mathematical expectations and co-ordinate functions.

2. Statistical linearization of non-linear functions by means of canonical expansions

Let two real variables X and Z linked by functional relationship

$$Z = \varphi(X) . \tag{2.1}$$

Expressing the random function X by the canonical expansion [1]

$$X = m^x + \sum_\nu V_\nu x_\nu, \qquad t \in (0, T), \tag{2.2}$$

where

$$m^x = M[X] , \quad M[V_\nu] = 0,$$
$$D[V_\nu] = D_\nu , \quad M[V_\nu \overline{V}_\mu] = 0 \quad (\nu \neq \mu) ,$$

we shall approximate (2.1) by the following canonical expansion

$$Z \approx \varphi_0 + \sum_\nu V_\nu z_\nu \qquad t \in (0, T) . \tag{2.3}$$

φ_0 and z_ν are obtained by minimizing the mean square error and are defined by

$$\varphi_0 = M[\varphi], \quad z_\nu = \frac{1}{D_\nu} M[\varphi \overline{V}_\nu] . \tag{2.4}$$

Concerning the integral canonical expansions we have the following formulae

$$X = m^x + \int_{-\infty}^{\infty} V(\lambda)\, x(t, \lambda)\, d\lambda \tag{2.5}$$

$$Z \approx \varphi_0 + \int_{-\infty}^{\infty} V(\lambda)\, z(t, \lambda)\, d\lambda , \tag{2.6}$$

where

$$m^x = M[X] , \quad M[V] = 0 , \quad M[V(\lambda)\, \overline{V(\lambda')}] = G(\lambda)\, \delta(\lambda - \lambda')$$

and

$$\varphi_0 = M[\varphi], \quad z(t, \lambda) = \frac{1}{G(\lambda)} M[\varphi\, \overline{V(\lambda)}] . \tag{2.7}$$

Let us consider the multi-dimensional non-linear function

$$Z = \varphi(X_1, \ldots, X_n) . \tag{2.8}$$

Expressing X_h $(h = 1, \ldots, n)$ by means of joint canonical expansions and joint integral canonical expansions [1], respectively

$$X_h = m_h^x + \sum_\nu V_\nu x_{\nu h}, \tag{2.9}$$

$$X_h = m_h^x + \int_{-\infty}^{\infty} V(\lambda)\, x_h(t, \lambda)\, d\lambda, \tag{2.10}$$

and Z by means (2.3) and (2.6), having (2.4) and (2.7) resp.

Concerning several non-linear functions

$$Z_j = \varphi_j (X_1, \ldots, X_n) \quad (j = 1, \ldots, l) \tag{2.11}$$

we approximate (2.11) by

$$Z_j \approx \varphi_{j0} + \sum_\nu V_\nu z_{j\nu}, \tag{2.12}$$

$$Z_j \approx \varphi_{j0} + \int_{-\infty}^{\infty} V(\lambda)\, z_j(t, \lambda)\, d\lambda, \tag{2.13}$$

where

$$\varphi_{j0} = M[\varphi_j], \quad z_{j\nu} = \frac{1}{D_\nu} M[\varphi_j \overline{V_\nu}], \tag{2.14}$$

$$\varphi_{j0} = M[\varphi_j], \quad z_j = \frac{1}{G} M[\varphi_j \overline{V}]. \tag{2.15}$$

When actually computing canonical expansions one has, in (2.2) or (2.9), to preserve the terms up to a given number $(\nu = 1 \ldots, N)$. In such cases formulae (2.4) for specific non-linear functions gives the exact expressions for φ_0 and z_ν. It is to be noticed that the probability distributions of uncorrelated V_ν are not necessarily Gaussian. For Gaussian V_ν one may, for typical non-linear functions, utilize the known formulae for statistical transfer constants [2]. In this case one may take into account the following relations:

$$z_\nu = k x_\nu, \quad k = \frac{\partial \varphi_0}{\partial m^x}, \tag{2.16}$$

$$z_\nu = \sum_{h=1}^{\infty} k_h x_{\nu h}, \quad k_h = \frac{\partial \varphi_0}{\partial m_h^x} \quad (h = 1, \ldots, n), \tag{2.17}$$

$$z_{j\nu} = \sum_{h=1}^{n} k_{jh} x_{\nu h}, \quad k_{jh} = \frac{\partial \varphi_{j0}}{\partial m_h^x} \quad (h = 1, \ldots, n;\ j = 1, \ldots, l). \tag{2.18}$$

3. Accuracy of non-linear non-stationary stochastic systems

Let us consider the set of ordinary non-linear differential equations

$$\dot{Y}_k = f_k(t, Y_q, X_s), \quad (k, q = 1, \ldots, h;\ s = 1, \ldots, n), \tag{3.1}$$

where the inputs $X_s = X_s(t, Y_q)$ are random functions of t and Y_q. We represent the random functions $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_h$ by their canonical expansions

$$\begin{aligned} X_s &= m_s^x + \sum_{\nu=1}^{N} V_\nu x_{\nu s}, \\ Y_k &= m_k^y + \sum_{\nu=1}^{N} V_\nu y_{\nu k}. \end{aligned} \tag{3.2}$$

Then we replace the non-linear functions by statistically linearized functions

$$\begin{aligned} Z_k = f_k(t, Y_q, X_s) \approx Z_{0k}(t, m_q^y, m_s^x, y_{\nu q}, x_{\nu s}, D_\nu) + \\ + \sum_{\nu=1}^{N} V_\nu z_{\nu k}(t, m_q^y, m_s^x, y_{\nu q}, x_{\nu s}, D_\nu). \end{aligned} \tag{3.3}$$

Substituting (3.2), (3.3) in (3.1) we get two deterministic system of equations for m_k^y and $y_{\nu k}$

$$\dot{m}_k^y = Z_{0k}, \dot{y}_{\nu k} = z_{\nu k}. \tag{3.4}$$

Integrating (3.4) we find m_k^y and $y_{\nu k}$. Accuracy may be checked by computing the covariance and cross-covariance functions

$$K_{pq}^y(t, t') = \sum_{\nu=1}^{N} D_\nu y_{\nu p}(t) \overline{y_{\nu q}(t')}, \tag{3.5}$$

$$(p, q = 1, \ldots, h).$$

Similarly the accuracy problem can be solved by employing the integral canonical expansions. Then we have the following formulae

$$\begin{aligned} X_s &= m_s^x + \int_{-\infty}^{\infty} V(\lambda)\, x_s(t, \lambda)\, d\lambda, \\ Y_k &= m_k^y + \int_{-\infty}^{\infty} V(\lambda)\, y_k(t, \lambda)\, d\lambda, \end{aligned} \tag{3.6}$$

$$Z_k = f_k(t, Y_q, X_s) \approx Z_{0k} + \int_{-\infty}^{\infty} V(\lambda)\, z_k(t, \lambda)\, d\lambda, \tag{3.7}$$

where

$$\begin{aligned} Z_{0k} &= Z_{0k}\big(t, m_q^y, m_s^x, y_q(t, \lambda), x_s(t, \lambda), G(\lambda)\big), \\ z_k(t, \lambda) &= z_k\big(t, m_q^y, m_s^x, y_q(t, \lambda), x_s(t, \lambda), G(\lambda)\big). \end{aligned}$$

Substituting (3.6), (3.7) in (3.1), we get

$$\dot{m}_k^y = Z_{0k}, \dot{y}_k = z_k. \tag{3.8}$$

After integrating (3.8) we get m_k^y and y_k, and then the covariance and cross covariance functions

$$K_{pq}^y(t,t') = \int_{-\infty}^{\infty} G(\lambda)\,\overline{y_p(t,\lambda)}\,y_q(t',\lambda)\,d\lambda\,, \qquad (p,q=1,\ldots,h)\,. \tag{3.9}$$

References

1. *Pugachev, V. S.:* Theory of random functions and application to control problems. 3rd edition, Fizmatgiz, Moscow 1962.
2. *Kazakov, I. Ye.—Dostupov, B. G.:* Statistical dynamics of non-linear control systems. Fizmatgiz, Moscow 1962.

Об одном обобщении метода статистической линеаризации

И. Н. СИНИЦИН

(Москва)

Известно, что канонические представления случайных функций (канонические разложения и интегральные канонические представления) служат основой ряда статистических методов исследования точности линейных систем. Для применения канонических представлений в нелинейных системах при описанной в [1] схеме аппроксимации нелинейностей (2.1), (2.8) или (2.11) достаточно непрерывности нелинейной функции. Практически аппроксимация нелинейностей проще всего получается в том случае, когда можно допустить существование производных до определенного порядка.

Предлагаемая ниже схема статистической линеаризации нелинейностей (2.3), например в случае использования канонических разложений, состоит в том, что коэффициенты канонического разложения сохраняются, а координатные функции определяются из условия минимума средней квадратической ошибки. В таком случае формулы (2.4) или (2.14) для конкретных нелинейностей позволяют определить математические ожидания и координатные функции. При этом, что особенно важно, законы распределения случайных коэффициентов могут быть отличными от гауссовых. Использование интегральных канонических представлений приводит к формулам (2.7). Поэтому такое обобщение известного метода статистической линеаризации [1,2] полезно в первую очередь для статистической линеаризации существенных нелинейностей. Расчет точности процессов управления в нелинейных нестационарных системах при использовании упомянутой схемы статистической линеаризации сводится к совместному интегрированию нелинейных детерминированных систем уравнений как для математических ожиданий, так и для координатных функций.

И. Н. Синицин
Институт проблем управления
СССР Москва В—485
Профсоюзная 81

Problems of Control and Information Theory, Vol. 1 (2), pp. 123—134 (1972)

THE SEQUENTIAL EVALUATION OF LINEAR SIMPLEX DESIGN

L. KEVICZKY

(Budapest)

The sequential evaluation of a simplex based linear design suggested by the author is discussed. The design is analysed in detail and conclusions concerning the practical role of the statistical properties are drawn. Then the application of the design procedure is shown on a few numerical examples.

1. The method of sequential evaluation

Let us assume that at N points in the range of the variables $x_1, x_2, \ldots, x_n$ we have observed the value of the function $y=y(\mathbf{x}, \boldsymbol{\xi})$. Here $\mathbf{x}=[x_1, x_2, \ldots, x_n]^T$ is the vector of the independent variables, $\boldsymbol{\xi}$ is a zero mean random vector of the disturbing variables causing the statistical variation of the function, and T stands for the transposition.

In the following — assuming active experiments — the set of the points $\mathbf{x}_u$ $(u = 1, 2, \ldots, N)$ will be called an experimental design.

Let our problem be the determination of the function

$$\hat{y} = \mathbf{f}^T(\mathbf{x})\, \mathbf{b}_0 \tag{1.1}$$

approximating measured data in the least quadratic sense. Here

$\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_q(\mathbf{x})]_T$ — is the vector of the component functions and

$\mathbf{b}_0 = [b_1, b_2, \ldots, b_q]^T$ — is the vector of the unknown coefficients.

We shall call the matrix with elements $f_i(\mathbf{x}_u)$ $(i = 1, 2, \ldots, q;\ u = 1, 2, \ldots, N)$

$$\mathbf{X} = \| f_i(\mathbf{x}_u) \| = \| f_{ui} \| \tag{1.2}$$

as design-matrix.

We wish to determine the coefficients of the approximating function (1.1) under the following condition

$$\sum_{u=1}^{N} \varepsilon_u^2 = \sum_{u=1}^{N} (y_u - \hat{y}_u)^2 = (\mathbf{y}_0 - \hat{\mathbf{y}}_0)^T (\mathbf{y}_0 - \hat{\mathbf{y}}_0) = \text{minimum} \tag{1.3}$$

where $\mathbf{y}_0$ is the column-vector containing N number of measured values of the function and

$$\hat{\mathbf{y}}_0 = \mathbf{X}_0 \mathbf{b}_0 \tag{1.4}$$

denotes the value the approximating function (1.1) takes at N points in the form of a column vector.

If we assume that the errors ε_u are uncorrelated zero mean random variables of identical dispersion, then the minimum of (1.3) is obtained by solving what is called normal system of equations: [1]

$$\mathbf{X}_0^T \mathbf{X}_0 \mathbf{b}_0 = \mathbf{Y}_0^T \mathbf{y}_0 . \tag{1.5}$$

By solving this system of equations (1.5) for the unknown coefficients $\mathbf{b}_0$ we obtain:

$$\mathbf{b}_0 = (\mathbf{X}_0^T \mathbf{X}_0)^{-1} \mathbf{X}_0^T \mathbf{y}_0 . \tag{1.6}$$

(The matrix $(\mathbf{X}_0^T \mathbf{X}_0)^{-1} \sigma^2\{y\}$ is called a covariance, or a dispersion matrix; $\sigma^2\{y\}$ is the deviation characterizing the error of the measured data.)

In practical cases it may happen that the result obtained by solving the equation system (1.6) is in some respect still unsatisfactory and therefore further measurements are to be made. If vector $\mathbf{b}_0$ contains q different coefficients, then a matrix of the dimension $q \times q$ has to be inverted in order to solve (1.6). If a matrix $\mathbf{R}$ containing k rows (measurements) is coupled to the matrix $\mathbf{X}_0$, then the design matrix of dimension $(N+k) \times q$ containing the values of the independent variables will have the form of

$$\mathbf{X} = \left[\frac{\mathbf{X}_0}{\mathbf{R}}\right] \tag{1.7}$$

and the new normal equation system will be

$$(\mathbf{X}^T \mathbf{X})\, \mathbf{b}^* = \mathbf{X}^T \mathbf{y} \tag{1.8}$$

where

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_0 \\ \mathbf{y}_k \end{bmatrix} \tag{1.9}$$

is a matrix of dimension $(N + k) \times 1$. Here $\mathbf{y}_k$ contains the supplementary k measurements. If we want to solve (1.8) in the classical way, i.e. by carrying out the operations

$$\mathbf{b}^* = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \tag{1.10}$$

then we have to invert a matrix of dimension $(q \times q)$. Here $\mathbf{b}^*$ is the column vector containing the corrected values of the coefficients.

Plackett and Burman [1] reduced the solution of (1.10) to the inversion of a $k \times k$ dimensioned matrix implying a considerable simplification in the computations. Their result (see the Appendix) is:

$$(\mathbf{X}^T \mathbf{X})^{-1} = (\mathbf{X}_0^T \mathbf{X}_0)^{-1} - \mathbf{H}^T \mathbf{G} \mathbf{H} \tag{1.11}$$

where

$$\mathbf{H} = \mathbf{R}(\mathbf{X}_0^T \mathbf{X}_0)^{-1} \tag{1.12}$$

and

$$\mathbf{G} = (\mathbf{E} + \mathbf{H}\mathbf{R}^T)^{-1} \tag{1.13}$$

Here $\mathbf{H}$ is a $k \times q$ and $\mathbf{G}$ is a $k \times k$ matrix, $\mathbf{E}$ denotes the unit matrix.

With the help of relation (1.11) we obtain (see Appendix, Theorem 3.):

$$\mathbf{b}^* = \mathbf{b}_0 + \mathbf{H}^T \mathbf{G} \mathbf{d} \tag{1.14}$$

where

$$\mathbf{d} = \mathbf{y}_k - \hat{\mathbf{y}}_k \tag{1.15}$$

Here $\hat{\mathbf{y}}_k = \mathbf{R}\mathbf{b}_0$.

Obtaining the regression coefficients in this way is called sequential evaluation, with respect to the fact that the supplementary measurements may be considered subsequently.

The value of the quadratic sum of the residuals for the statistical investigations may also be calculated by a similar correction:

$$S_{ek}^2 = S_{e_0}^2 + \mathbf{d}^T \mathbf{G} \mathbf{d} \tag{1.16}$$

where

$$S_{e_0}^2 = \sum_{u=1}^{N} (y_u - \hat{y}_u)^2 = (\mathbf{y}_0 - \hat{\mathbf{y}}_0)^T (\mathbf{y}_0 - \hat{\mathbf{y}}) = \mathbf{y}_0^T \mathbf{y}_0 - \mathbf{b}_0^T \mathbf{X}_0^T \mathbf{y}_0 . \tag{1.17}$$

These results in sequential processing were utilized by Hunter [2] for the subsequent evaluation of the two-level active 2^n-type experimental designs.

He assumed that the coefficients $\mathbf{b}_0$ from the orthogonal block containing N measurements each size m have been determined and they are to be corrected by k supplementary measurements. The calculations can be carried out very simply if the orthogonality by rows of the blocks is simultaneously ensured as well. These designs are called sequential factorial designs and denoted by the letters **SFD**.

From the orthogonality of the columns it follows that

$$\mathbf{X}_0^T \mathbf{X}_0 = N\mathbf{E} . \tag{1.18}$$

Thus equation (1.6) is reduced to

$$\mathbf{b}_0' = (\mathbf{X}_0^T \mathbf{X}_0)^{-1} \mathbf{X}_0^T \mathbf{y}_0' = \frac{1}{N} \mathbf{X}_0^T \mathbf{y}_0 .$$

If we want to estimate q coefficients, i.e. the matrix $\mathbf{X}$ consists of number q columns, then the orthogonality by rows of the matrix $\mathbf{R}$ containing the supplementary measurements means the following:

$$\mathbf{R}^T\mathbf{R} = q\mathbf{E} \tag{1.19}$$

(It has to be emphasized that the orthogonality of the columns and rows of matrices is not perfectly identical with the usual notion of orthogonality. In the theory of designing experiments the rows or columns of matrices are considered to be orthogonal when the quadratic sums of their elements are identical in value and the linear combinations formed with each other result in zero.)

Adopting these auxiliary conditions the relation (1.14) for the sequential calculation of the coefficients will run as follows:

$$\mathbf{b}^* = \mathbf{b}_0 + \frac{1}{mN+q}\mathbf{R}^T(\mathbf{y}_k - \mathbf{R}\mathbf{b}_0) \tag{1.20}$$

The correction of the matrix $(\mathbf{X}^T\mathbf{X})^{-1}$ and that of the quadratic sum of the residuals is effected in the following way:

$$(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{mN}\left[\mathbf{E} - \frac{1}{mN+q}\mathbf{R}^T\mathbf{R}\right] \tag{1.21}$$

and

$$S^2_{e_k} = S^2_{e_0} + \Delta S^2_{e_k} = S^2_{e_0} + \frac{1}{mN+q}(\mathbf{y}_k - \mathbf{R}\mathbf{b}_0)^T(\mathbf{y}_k - \mathbf{R}\mathbf{b}_0) \tag{1.22}$$

The dispersion belonging to the quadratic sum of the residuals is:

$$S^2_{e_k} = \frac{S^2_{e_k}}{f_e} = \frac{S^2_{e_0} + \Delta S^2_{e_k}}{mN+k-q} \tag{1.23}$$

On the basis of (1.21) the covariance matrix is:

$$(\mathbf{X}^T\mathbf{X})^{-1}\sigma^2\{y\} = \frac{1}{mN}\left[\mathbf{E} - \frac{1}{mN+q}\right]\sigma^2\{y\} \tag{1.24}$$

As the course of $\mathbf{R}^T\mathbf{R}$ depends strongly on the construction of the designs, therefore nothing can be said about the cov $\{b_i b_j\}$ quantities. But we know the diagonal elements of $\mathbf{R}^T\mathbf{R}$, whose value is:

$$\sum_{u=mN+1}^{mN+k} x_{iu}^2 = k \tag{1.25}$$

as two-level design are involved. Thus

$$\sigma^2\{b_i\} = \frac{1}{mN}\left[1 - \frac{k}{mN+q}\right]\sigma^2\{y\} \tag{1.26}$$

We note that the dispersion of the measurement results may be estimated also sequentially:

$$s_u^2\{y\} = \left[s_{u-1}^2\{y\} + \frac{s^2\{y_u\}}{u-1}\right]\frac{u-1}{u} \tag{1.27}$$

Here $s^2\{y_u\}$ is the dispersion of the u^{th} experimental result and $s_{u-1}^2\{y\}$ is the average dispersion of $u-1$ measurements.

If we use two-level design for the sequential evaluations, then orthogonality by rows can be ensured only for the quadratic blocks. Such arrangements are given by the fractional factorial design (**FFD**-s) of the resolution degree III (saturated) and **FFD**-s of a resolution degree higher than III, if they are completed for the sequential evaluation by some means (by a certain group of interrelations) to a quadratic design [3]. These methods of solution raise always the problem of how to design mixing. Let us consider how these questions are forming when the method of the sequential evaluation is applied to a linear simplex design of the construction as suggested in [4].

2. The sequential simplex factorial design (SSFD)

Let us construct the design-matrix in the following way:

$$\mathbf{X} = \begin{bmatrix} \mathbf{e}, & \mathbf{D}_1 \\ \mathbf{e}, & -\mathbf{D}_1 \end{bmatrix} \tag{2.1}$$

where $\mathbf{e}$ is a column vector consisting of ones, $\mathbf{D}_1$ is the linear orthogonal simplex design, whose construction is [4,6] the following:

$$\mathbf{D}_1 = \begin{bmatrix} -r_1 & -r_2 & -r_3 & \dots & -r_n \\ r_1 & -r_2 & -r_3 & \dots & -r_n \\ 0 & 2r_2 & -r_3 & \dots & -r_n \\ 0 & 0 & 3r_3 & \dots & -r_n \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & nr_n \end{bmatrix} \tag{2.2}$$

where

$$r_i = \sqrt{\frac{n+1}{i(i+1)}} \tag{2.3}$$

The row- and column vectors of the matrix $\mathbf{X}_0 = [\mathbf{e}, \mathbf{D}_1]$ form an orthonormed vector system. Naturally the same applies to the matrix $[\mathbf{e}, -\mathbf{D}_1]$ as well [4].

Let us determine the $n + 1$ coefficients of a linear polynomial with the help of an $\mathbf{X}_0$ design consisting of $N = n + 1$ rows and columns. The elements of the required vector $\mathbf{b}_0$ may be determined by the well known formula

$$b_i = \frac{\sum_{u=1}^{N} x_{iu} y_u}{N} \tag{2.4}$$

(because of the normality $\sum_{u=1}^{N} x_{iu}^2 = N$ [5]).

If we wish to determine the influence of additional k supplementary measurements on the values of the coefficients, this determination can be carried out according to (1.20), as the second block of design matrix $\mathbf{X}$ is orthogonal both by rows and by columns, as explained above.

So let $\mathbf{R}$ contain k rows (e.g. the first k (this piece) of the matrix $[\mathbf{e}, -\mathbf{D}_1]$ and let us correct by this the values of $q = n + 1 = N$ coefficients. Observe that the number of the evaluated blocks ($\mathbf{X}_0$) is $m = 1$, the corrector equation will be:

$$\mathbf{b}^* = \mathbf{b}_0 + \frac{1}{2N} \mathbf{R}^T (\mathbf{y}_k - \mathbf{R}\mathbf{b}_0) \; ; \quad (k \leq N) \tag{2.5}$$

The design matrix $\mathbf{X} = \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{R} \end{bmatrix}$ used for the evaluation will be called in the following a sequential simplex factorial design, or in short **SSFD**.

With substitution similar to the preceding the correction of the residual quadratic sum at the **SSFD** is:

$$S_{ek}^2 = \frac{1}{2} (\mathbf{y}_k - \mathbf{R}\mathbf{b}_0)^T (\mathbf{y}_k - \mathbf{R}\mathbf{b}_0) \, . \tag{2.6}$$

We see that the correction is made also here like in the **SFD** case. The statistical properties of the **SSFD** are also determined by the covariance matrix, therefore we have to study the course of the inverse matrix $(\mathbf{X}^T\mathbf{X})^{-1}$.

As already mentioned:

$$(\mathbf{X}^T\mathbf{X})^{-1} = (\mathbf{X}_0^T\mathbf{X}_0 + \mathbf{R}^T\mathbf{R})^{-1} = \frac{1}{mN} \left[\mathbf{E} - \frac{1}{mN + q} \mathbf{R}^T\mathbf{R} \right] . \tag{2.7}$$

With the conditions of $q = N$ and $m = 1$ taken into consideration, we obtain:

$$(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{N} \left[\mathbf{E} - \frac{1}{2N} \mathbf{R}^T\mathbf{R} \right] . \tag{2.8}$$

In the following we shall investigate the construction of the matrix $\mathbf{R}^T\mathbf{R}$. As $\mathbf{R}$ contains the first k rows of the matrix $-\mathbf{D}_1$, we can write up, that:

$$\mathbf{R}^T\mathbf{R} = \left[\begin{array}{c|cccccc|cccccc} k & 0 & 0 & 0 & \dots & 0 & kr_k & \dots & kr_i & \dots & kr_n \\ \hline 0 & N & 0 & 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & N & 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & 0 & 0 & N & \dots & 0 & 0 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 0 & \dots & N & 0 & \dots & 0 & \dots & 0 \\ \hline kr_k & 0 & 0 & 0 & \dots & 0 & kr_k^2 & & & & \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \vdots & & kr_ir_j & \\ kr_i & 0 & 0 & 0 & \dots & 0 & & & kr_i^2 & & \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & & kr_ir_j & & \vdots & \\ kr_n & 0 & 0 & 0 & \dots & 0 & & & & & kr_n^2 \end{array}\right] \tag{2.9}$$

Where the marked separation line is shifted depending on k. In a simplex form:

$$\mathbf{R}^T\mathbf{R} = \left[\begin{array}{c|c|c} k & \mathbf{0} & k\mathbf{a}_k^T \\ \hline \mathbf{0} & N\cdot\mathbf{E} & \mathbf{0} \\ \hline k\mathbf{a}_k & \mathbf{0} & k\mathbf{A} \end{array}\right]. \tag{2.10}$$

Here

$$\mathbf{A} = \begin{bmatrix} r_r^2 & r_k r_{k+1} & \dots & r_k r_n \\ r_{k+1} r_k & r_{k+1}^2 & \dots & r_{k+1} r_n \\ \vdots & \vdots & \dots & \vdots \\ r_n r_k & r_n r_{k+1} & \dots & r_n^2 \end{bmatrix} \tag{2.11}$$

and

$$\mathbf{a}_k = [r_k, r_{k+1}, \dots, r_n]^T. \tag{2.12}$$

If we substitute these result into (2.8), we obtain

$$(\mathbf{X}^T\mathbf{X})^{-1} = \left[\begin{array}{c|c|c} \dfrac{2N-k}{2N^2} & \mathbf{0} & \dfrac{-k}{2N^2}\mathbf{a}_k^T \\ \hline \mathbf{0} & \dfrac{1}{2N}\mathbf{E} & \mathbf{0} \\ \hline \dfrac{-k}{2N^2}\mathbf{a}_k & \mathbf{0} & \dfrac{k}{2N^2}\mathbf{B} \end{array}\right] \tag{2.13}$$

where

$$\mathbf{B} = \frac{2N}{k}\mathbf{E} - \mathbf{A}. \tag{2.14}$$

Thus the auxiliary quantities characterizing the statistical properties of the **SSFD** may be already determined, i.e.

$$\operatorname{cov}\{b_i\, b_j\} = \begin{cases} 0 & , \quad \text{when} \quad i,j < k \\ \dfrac{k}{2N^2}\, r_i\, r_j\, \sigma^2\{y\} & , \quad \text{when} \quad i,j \geq k \end{cases} \tag{2.15}$$

and

$$\operatorname{cov}\{b_0\, b_i\} = \begin{cases} 0 & , \quad \text{when} \quad i < k \\ \dfrac{-k}{2N^2}\, r_i\, \sigma^2\{y\} & , \quad \text{when} \quad i \geq k \end{cases} \tag{2.16}$$

For the dispersions of the coefficients the following relations hold:

$$\sigma^2\{b_i\} = \begin{cases} \dfrac{1}{2N}\, \sigma^2\{y\} & , \quad \text{when} \quad i < k \\ \dfrac{2N - k r_i^2}{2N^2}\, \sigma^2\{y\}\,, & \quad \text{when} \quad i \geq k \end{cases} \tag{2.17}$$

$$\sigma^2\{b_0\} = \frac{2N - k}{2N^2}\, \sigma^2\{y\}\,. \tag{2.18}$$

One may see from the relation obtained for the covariances and the deviations that their course in the case of the **SSFD** is well defined and in the individual cases the statistical property given by the grouping (indexing, i.e. the rotation of the design) of the variables may also be attained.

E.g.: if we want to keep some factor statistically independent of the rest during the correction, it is advised to define it as first variable, respectively the corresponding column of the matrix should be ordered to it.

We have seen in [4] that $\mathbf{X}_0$ has good statistical properties only concerning the coefficients of the linear approximation. The interrelations cannot be determined statistically independently of each single linear effect. In consequence of what has been said here and the disadvantageous properties discussed in detail in [4], the **SSFD** can also be applied only for the sequential calculation of the coefficients of the linear approximation (e.g., when using method of the steepest rise). Yet the advantage of the **SSFD** is that the design problem of mixing the various effects does not arise with it.

3. Model tests

Tests on a quadratic and a linear three-variable model have been carried out for deterministic and stochastic cases to test the **SSFD**.

The quadratic model was of the form:

$$y_1 = 800 - [(x_1 - 10)^2 + 2(x_2 - 20)^2 + 3(x_3 - 30)^2] =$$
$$= -2800 + 20x_1 + 80x_2 + 180x_3 - x_1^2 - 2x_2^2 - 3x_3^2 \qquad (3.1)$$

and a linear model:

$$y_2 = 800 + x_1 - 2x_2 + 3x_3 \qquad (3.2)$$

The modelling was carried out on a digital computer (type ODRA 1013) as well as the loading with the experimental error of the function surface, in the stochastic case, by means of a random number generator with normal distribution. The obtained "measurements" were performed with a program prepared for the evaluation of the **SSFD**.

The results may be summarized as follows:

3.1. The deterministic case

The measurements for the quadratic model were carried out at the points

$$x_{10} = 5\ ; \quad x_{20} = 15\ ; \quad x_{30} = 25 \qquad (3.3)$$

with testing steps

$$\lambda_1 = \lambda_2 = \lambda_3 = 1 \qquad (3.4)$$

The results obtained from the first block were:

$$b_0 = 644;\ b_1 = 9.9999;\ b_2 = 19.1835;\ b_3 = 28.2679$$

The results of the corrections were:

k	b_0	b_1	b_2	b_3
1	644.4167	10.5893	19.5237	28.5085
2	644.8333	9.9999	19.8639	28.7491
3	644.7500	9.9999	20.0000	28.7010
4	644.0000	9.9999	20.0000	29.9999

From these results it may be seen that the value of b_i does not vary during the correction, when $k > i$ already, which follows from the construction of the design (0 elements).

Let us investigate the accuracy of the investigation:

$$\frac{\partial y_1}{\partial x_1} = 20 - 2x_1\Big|_5 = 10\,; \quad \frac{\partial y_1}{\partial x_2} = 80 - 4x_2\Big|_{15} = 20\,; \quad \frac{\partial y_1}{\partial x_3} = 180 - 6x_3\Big|_{25} = 30.$$

As we had $\lambda_1 = \lambda_2 = \lambda_3 = 1$, the b_i coefficients give obviously a very good estimation for these partial derivates. The variations of the coefficients during the corrections may have two reasons: the statistical variation (fluctuation) and the variation of the measuring points. In the present case only this latter variation can play a role as the surface was not loaded with a statistical error. This means that in the case when the surface is not linear, the coefficients obtained as a result of the correction may vary also because of the subsequent points of the design belonging to other points of the surface. The statement concerning the independence of the coefficients stays naturally valid in this case too, as we have seen before.

The above considerations are supported also by the fact that the correction carried out for the y_2 linear model in the deterministic case at the points

$$x_{10} = 10\ ; \quad x_{20} = 15\ ; \quad x_{30} = 20 \tag{3.5}$$

with steps of

$$\lambda_1 = \lambda_2 = \lambda_3 = 1 \tag{3.6}$$

did not change the values of the coefficients $b_0 = 840.0000$; $b_1 = 0.9999$; $b_2 = -2.0000$; $b_3 = 2.9999$ obtained from the first block. The accuracy of the approximation can be verified very easily in this case as well.

(An interesting result was also the following: the difference between the value of function y_1 at the working point and the coefficient b_0 agreed exactly with the sum of the quadratic coefficients, but this phenomenon did not follow directly from the construction of the **SSFD**).

3.2. The stochastic case

In these investigations the correction was performed at the same points and with the same testing steps. The size of the scattering band (dispersion band) of the surfaces was also varied in the individual tests. The numerical results obtained for the values of the coefficients disclose nothing much, but the statements made concerning the statistical properties of the **SSFD** were verified also by these experiments, which offered valuable informations for the relationship between the experimental error and the accuracy of the correction as well.

Appendix

Theorem 1. Be **A** a matrix of dimension $p \times q$, **B** a matrix of dimension $q \times p$ in this case the identity

$$(\mathbf{E}_p + \mathbf{AB})^{-1} = \mathbf{E}_p - \mathbf{A}\,(\mathbf{E}_q + \mathbf{BA})^{-1}\,\mathbf{B},$$

holds, where $\mathbf{E}_p$ is a unit matrix of dimension $p \times p$,
$\mathbf{E}_q$ is a unit matrix of dimension $q \times q$.

Proof

Let us multiply both sides of the expression by $(\mathbf{E}_p + \mathbf{AB})$:

$$\mathbf{E}_p = \mathbf{E}_p(\mathbf{E}_p + \mathbf{AB}) - \mathbf{A}(\mathbf{E}_q + \mathbf{BA})^{-1}\mathbf{B}(\mathbf{E}_p + \mathbf{AB}) =$$
$$= \mathbf{E}_p + \mathbf{AB} - \mathbf{A}(\mathbf{E}_q + \mathbf{BA})^{-1}(\mathbf{E}_q + \mathbf{BA})\mathbf{B} = \mathbf{E}_p + \mathbf{AB} - \mathbf{AB} = \mathbf{E}_p$$

Theorem 2. Be $\mathbf{X}_0^T\mathbf{X}_0 = \mathbf{C}_0$ the matrix of the coefficients of the initial, and

$\mathbf{X}^T\mathbf{X} = \mathbf{C}$ the matrix of the coefficients of the new normal equation system.

With consideration to the relationship (1.7) we may write up:

$$\mathbf{C}^{-1} = [\mathbf{C}_0 + \mathbf{R}^T\mathbf{R}]^{-1} = \mathbf{C}_0^{-1}[\mathbf{E} + \mathbf{R}^T\mathbf{R}\mathbf{C}_0^{-1}]^{-1}$$

as $\mathbf{C}_0^{-1}$ and $\mathbf{R}^T\mathbf{R}$ are symmetrical quadratic matrices.

By applying Theorem 1 we obtain:

$$\mathbf{C}^{-1} = \mathbf{C}_0^{-1}[\mathbf{E} - \mathbf{R}^T(\mathbf{E} + \mathbf{R}\,\mathbf{C}_0^{-1}\mathbf{R}^T)^{-1}\mathbf{R}\mathbf{C}_0^{-1}] =$$
$$= \mathbf{C}_0^{-1} - (\mathbf{R}\mathbf{C}_0^{-1})^T(\mathbf{E} + \mathbf{R}\mathbf{C}_0^{-1}\mathbf{R}^T)^{-1}\mathbf{R}\mathbf{C}_0^{-1}$$

Be $\mathbf{H} = \mathbf{R}\mathbf{C}_0^{-1}$ and $\mathbf{G} = (\mathbf{E} + \mathbf{H}\mathbf{R}^T)^{-1}$, then

$$\mathbf{C}^{-1} = \mathbf{C}_0^{-1} - \mathbf{H}^T\mathbf{G}\mathbf{H}$$

Theorem 3. With the relationships (1.6), (1.7), (1.9) and (1.10) taken into consideration the solution of the normal equation system containing the supplementary measurements as well, runs as follows:

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = (\mathbf{X}_0^T\mathbf{X}_0 + \mathbf{R}^T\mathbf{R})^{-1}(\mathbf{X}_0^T\mathbf{y}_0 + \mathbf{R}^T\mathbf{y}_k) =$$
$$= (\mathbf{C}_0^{-1} - \mathbf{H}^T\mathbf{G}\mathbf{H})(\mathbf{X}_0^T\mathbf{y}_0 + \mathbf{R}^T\mathbf{y}_k) = \mathbf{C}_0^{-1}\mathbf{X}_0^T\mathbf{y}_0 - \mathbf{H}^T\mathbf{G}\mathbf{R}\mathbf{C}_0^{-1}\mathbf{X}_0^T\mathbf{y}_0 +$$
$$+ \mathbf{H}^T\mathbf{y}_k - \mathbf{H}^T\mathbf{G}\mathbf{H}\mathbf{R}^T\mathbf{y}_k = \mathbf{b}_0 - \mathbf{H}^T\mathbf{G}\mathbf{R}\mathbf{b}_0 + \mathbf{H}^T\mathbf{y}_k - \mathbf{H}^T\mathbf{G}\mathbf{H}\mathbf{R}^T\mathbf{y}_k =$$
$$= \mathbf{b}_0 - \mathbf{H}^T\mathbf{G}\hat{\mathbf{y}}_k + \mathbf{H}^T\mathbf{y}_k - \mathbf{H}^T\mathbf{y}_k + \mathbf{H}\,\mathbf{G}^T\hat{\mathbf{y}}_k$$

Where $\hat{\mathbf{y}}_k = \mathbf{R}\mathbf{b}_0$. By introducing the notation $\mathbf{d} = \mathbf{y}_k - \hat{\mathbf{y}}_k$, the corrector equation has the following form:

$$\mathbf{b} = \mathbf{b}_0 + \mathbf{H}^T\mathbf{G}\mathbf{d}$$

References

1. *Plackett, R. L.:* Some theorems in least squares. Biometrika **1** (1950).
2. *Hunter, J. S.:* Sequential factorial estimation. Technometrics. **6** (1964).
3. *Box, G. E. P.—Hunter, J. S.:* The 2^{n-q} fractional factorial design. Part I, Technometrics **3** (1961), Part II, Technometrics **4** (1961)
4. *Keviczky, L.:* The analysis of first and second order simplex design. Periodica Polytechnica, Electrical Engineering-Elektrotechnik **14** (1970).
5. *Налимов, В. В.—Чернова, Н. А.:* Статистические методы планирования экстремальных экспериментов. Изд. «Наука», 1965, Москва.
6. *Ермуратский, П. В.:* Симплексный метод оптимизации. Труды МЭИ, Выпуск 67, Изд. «Наука», 1966, Москва.
7. *Лисенков, А. Н.—Круг, Г. К.—Коршунов, М. А.—Лазарянц, В. Э.:* Оптимизация одного химического процесса методом последовательного планирования эксперимента. Доклады научно-технической конференции МЭИ (подсекция Автоматики и телемеханики), 1967, Москва.
8. *Albert, A.—Sittler, R. W.:* A method for computing least squares estimators that keep up with data. J. SIAM Control Ser. A. **3** (1966).
9. *Kalman, R. E.—Bucy, R. S.:* New results in linear filtering and prediction theory. Trans ASME (J. Basic Engrg.), Ser. D, vol **83** (1961).

Последовательная оценка линейных симплексных планов

Л. КЕВИЦКИ

(Будапешт)

Рассматривается один из вариантов последовательной оценки линейного плана на основе симплексной базы. В первой части приведен метод последовательной оценки. В многомерном пространстве переменных $x_1, x_2, \ldots, x_n$ наблюдались ординаты функции $y = y(x, \xi)$ в N разных точках. Задача состоит в определении приближенной функции по формуле (1.1) таким образом, чтобы условия наименьших квадратов были выполнены. Приведены результаты, известные из [1] на случай, когда после первых N активных экспериментов из-за неточности приближения следует проводить еще несколько экспериментов и затем решать модифицированную систему уравнений (1.8). Решение системы уравнений представлено в формулах (1.11—1.13). Путем последовательной коррекции можно также рассчитывать величину квадратной суммы разностей по формулам (1.16—1.17).

Во второй части предлагается последовательный симплексный факторный план на основе достигнутых ранее автором результатов [4]. Из полученных формул на ковариацию (2.15—2.16) и дисперсию (2.17—2.18) можно заключить, что при последовательных симплексных факторных планах эти величины однозначно определены и далее при помощи соответствующей перегруппировки переменных можно добиться определенных статистических свойств. Например, если один из факторов рассматривается независимым от других факторов в ходе коррекции, целесообразно определить этот фактор первым.

Конкретные эксперименты были проведены на квадратичной и линейной трехмерных моделях для детерминистического и стохастического случаев. Структура и параметры моделей представлены в формулах (3.1—3.2). В приложении дается доказательство трех теорем.

László Keviczky
Department of Automation
Technical University
Budapest 11, Garami E. t. 3, Hungary

Prvblems of Control and Information Theory, Vol. 2 (1), pp. 135—152 (1972)

ИНВАРИАНТНЫЕ ОЦЕНКИ В СТАТИСТИЧЕСКОЙ ТЕОРИИ ОПТИМАЛЬНЫХ СИСТЕМ

М. Е. ШАЙКИН

(Москва)

Для решения задач теории адаптивных систем, действующих в условиях неопределённости, используются теоретико-групповые методы. Последние применимы в случае, когда задача инвариантна относительно подходящей группы преобразований входящих в неё переменных. Рассмотрены свойства инвариантных решений, методы их нахождения. Доказана теорема о характеризации оптимального инвариантного решения как обобщенного бейесова, отвечающего правоинвариантной (обычно ненормируемой) мере Хаара в качестве априорного распределения на пространстве параметров. Метод инвариантной оптимизации иллюстрируется на примере задачи идентификации линейного объекта с неизвестной дисперсией помехи.

Статистическая теория оптимальных систем обнаружения сигнала, фильтрации, предсказания, идентификации и управления базируется на общей теории решающих функций и исходит из предположения, что известны вероятностные характеристики входных сигналов и объекта управления и что целью функционирования системы является минимизация некоторого заданного показателя качества. Все наиболее важные результаты статистической теории были получены на основе этого допущения о полной определенности возникающих здесь математических задач.

Современная тенденция в статистической теории систем управления является более реалистической. Развиваемые в настоящее время методы для определения самонастраивающихся и адаптивных фильтров, обучающихся и самообучающихся систем управления, индентификации и распознавания имеют целью отыскание систем, использующих существенно меньшую информацию о свойствах входных сигналов и характеристиках объекта управления, чем их классические предшественники. Естественно, что для решения новых задач прежние методы оказываются недостаточными, и в теорию и практику синтеза оптимальных систем внедряются новые идеи и методы: метод стохастической аппроксимации, эмпирический бейесов метод и методы общей статистической теории, основанные на асимптотической независимости бейесовых рисков от априорного распределения ненаблюдаемых параметров сигнала или объекта управления. Всеми этими

методами обеспечивается лишь асимптотическая оптимальность рекомендуемых ими систем.

В статье излагается круг идей и методов для определения структуры систем, оптимальных в условиях заданной неопределенной ситуации. Кроме интереса, который представляют такие системы, их нахождение полезно с точки зрения сопоставления их с различными субоптимальными системами. Основным математическим аппаратом является здесь аппарат теории групп. Теоретико-групповой метод применяется в случае, когда задача определения оптимальной системы инвариантна относительно некоторой группы преобразований входящих в нее переменных. Из существа задачи часто можно видеть, что предполагаемая инвариантность действительно имеет место. Если задача инвариантна относительно некоторой группы преобразований, то и решение ее должно быть инвариантным относительно той же группы преобразований. Это утверждение составляет содержание принципа инвариантности, формулируемого ниже. Ограничение инвариантными решениями следует той же идее сужения класса возможных решений, которая широко используется в статистической теории при отыскании оптимальных решений в классе линейных, квадратичных, несмещённых и т. п. оценок. Инвариантные решения обладают рядом интересных свойств, которые будут изложены ниже.

В § 1 дается постановка задачи отыскания оптимальных систем в условиях неопределенности, формулируются достаточно общие условия инвариантности статистических задач, приводятся некоторые свойства инвариантных решений. В § 2 изложен метод нахождения оптимальных инвариантных оценок, который иллюстрируется на примере инвариантного оценивания коэффициентов линейной регрессии. Наконец, в § 3 формулируется теорема о связи между бейесовыми и инвариантными оценками, доказательство которой вынесено в приложение.

§ 1. Постановка задачи. Принцип инвариантности

С точки зрения общей теории статистических систем структура любой системы управления, фильтрации, предсказания или обнаружения сигнала определяется: 1) ансамблем ее входных сигналов $\mathbf{z}$ вместе с заданным на нем семейством вероятностных распределений $\mathbf{P}_\theta$, $\theta \in \Omega$, параметризованных элементами θ параметрического множества Ω, 2) множеством выходных сигналов $\mathbf{d}$ системы, 3) функцией потерь $\mathbf{L}(\theta, \mathbf{d})$, определяющей потери для различных комбинаций параметра θ распределения входного сигнала $\mathbf{z}$ и значений $\mathbf{d}$ выходного сигнала. Структура оптимальной системы считается известной, если найдена решающая функция $\hat{\delta}(\mathbf{z})$ системы, отображающая

множество входных сигналов $\mathbf{z}$ в множество выходных сигналов $\mathbf{d}$ и такая что средний риск

$$R(\theta, \delta) = \int \mathbf{L}(\theta, \delta(\mathbf{z}))\, d\mathbf{P}_\theta(\mathbf{z}) \tag{1.1}$$

достигает на $\hat{\delta}$ минимального значения в классе D допустимых решающих функций.

В такой формулировке задача определения оптимального решения остается математически неопределенной. Функция $\hat{\delta}$, минимизирующая риск (1.1), в общем случае будет зависеть от параметра θ, значение которого обычно бывает неизвестным. В самом деле, задание семейства $\mathbf{P}_\theta$, $\theta \in \Omega$ распределений входного сигнала вместо указания одного единственного распределения $\mathbf{P}_{\theta_0}$ с $\theta = \theta_0$ есть формальное выражение неопределенности, в которой действует система. Последняя должна быть оптимальной (или, по крайней мере, удовлетворительно функционирующей) при любой входной ситуации, т. е. при любом значении параметра θ. Ясно, что если $\hat{\delta}$ зависит от θ, а значение последнего неизвестно заранее, то правилом $\hat{\delta}$ нельзя будет воспользоваться. В действительности необходима решающая функция, оптимальная равномерно по всем $\theta \in \Omega$.

Заметим, что при бейесовой постановке указанной трудности не возникает, поскольку параметр θ считается случайной величиной с известным априорным распределением $\nu(\theta)$, а оптимальное правило находится из условия минимума средних потерь

$$\varrho(\delta) = \int R(\theta, \delta)\, d\nu(\theta)$$

и, очевидно, от θ не зависит [1]. Но в нашем случае предполагается, что θ — неизвестный параметр, а информация о виде распределения ν не является ни достаточно надежной, ни достаточно точной, чтобы оправдать бейесов подход.

В бейесовом методе неопределенность задачи снимается введением априорного распределения ν. Другой способ достижения той же цели состоит в ограничении класса решающих функций некоторым подклассом их, внутри которого могут найтись оптимальные функции с равномерно (по θ) наименьшим риском. Важный подкласс с таким свойством образуют инвариантные решающие функции; он возникает в случае задач, инвариантных относительно какой-либо группы преобразований [2].

Задачу определения оптимальной системы при заданных $\mathbf{P}_\theta$, $\theta \in \Omega$ и $\mathbf{L}(\theta, \mathbf{d})$ будем называть инвариантной, если существует группа преобразований $(\mathbf{z}, \theta, \mathbf{d}) \to (g\,\mathbf{z}, \bar{g}\,\theta, g^*\,\mathbf{d})$ пространства входных сигналов, параметрического множества Ω и множества выходных сигналов D соответственно,

такая, что для всех элементов $(g, \bar{g}, g^*)$ группы

$$\begin{aligned} \mathbf{P}\,\bar{g}\theta(g\,\mathbf{A}) &= \mathbf{P}_\theta(\mathbf{A}), \\ \mathbf{L}(\bar{g}\theta, g^*\,\mathbf{d}) &= L(\theta, \mathbf{d}) \end{aligned} \tag{1.2}$$

при любых $\mathbf{A}$, θ, $\mathbf{d}$, где $\mathbf{A}$ — произвольное (измеримое) множество значений входного сигнала. Решающая функция δ преобразуется при этом по закону $\delta \to \delta_g = g^*\delta\, g^{-1}$, а для риска справедливо соотношение $R(\theta, \delta) = R(\bar{g}\theta, \delta_g)$.

Согласно принципу инвариантности, если задача инвариантна, то ее решение также должно быть инвариантным, так что $\delta_g = \delta$ или, подробнее,

$$g^*\,\delta g^{-1} = \delta \tag{1.3}$$

для любых элементов группы. Совокупность всех δ, удовлетворяющих (1.3), образует класс инвариантных решающих функций. В этом классе $R(\theta, \delta) = R(\bar{g}\,\theta, \delta)$, откуда следует независимость $R(\theta, \delta)$ от θ, если группа действует транзитивно на Ω. Ясно, что и оптимальная решающая функция в этом случае не зависит от θ, а отвечающий ей риск минимален сразу для всех $\theta \in \Omega$.

Как видно из (1.2), инвариантность статистической задачи складывается из инвариантности семейства распределений и инвариантности функции потерь. Достаточно широкий класс входных сигналов с инвариантным семейством распределений получается из следующих соображений. Пусть входной сигнал $\mathbf{z} = (z_1, z_2, \ldots, z_n)$ системы имеет вид

$$\mathbf{z} = f(\mathbf{x}\,;\theta) \tag{1.4}$$

$\big(\theta = (\theta_1, \theta_2, \ldots, \theta_r), \mathbf{f} = (f_1, f_2, \ldots, f_n)\big)$, причем преобразования $g_c : z' = f(z\,; c)$ образуют группу: $g_a\, g_b\, z = g_c\, z$, $c = \varphi(a, b)$, $g a_0\, z = z$, $g_a^{-1}\, z = g_{a^{-1}}\, z$ или, подробнее,

$$\begin{aligned} f\big(f(z\,;a)\,;b\big) &\equiv f\big(z\,;\varphi(a,b)\big), \\ f(z\,;a_0) &\equiv z, \\ f^{-1}(z\,;a) &\equiv f(z\,;a^{-1}) \end{aligned}$$

(необходимые условия того, чтобы преобразования g_c определяли группу, даются первой основной теоремой теории непрерывных групп Ли [3]). Если случайный вектор X имеет некоторое фиксированное распределение P_0, тогда семейство P_θ распределений вектора $Z = f(X\,;\,\theta)$ инвариантно относительно групп G и $\bar{G}$ преобразований вида

$$g_a : z' = f(z\,;a) \;\text{ и }\; \bar{g}_a : \theta' = \varphi(\theta\,;a) \tag{1.5}$$

соответственно. В самом деле,

$$P_\theta\{Z \in A\} = P_0\{g_\theta X \in A\} = P_0\{X \in g_\theta^{-1} A\},$$
$$P_{\bar{g}_a\theta}\{Z \in g_a A\} = P_{\theta'}\{Z \in g_a A\} = P_0\{X \in g_{\theta'}^{-1} g_a A\}. \tag{1.6}$$

Но поскольку

$$z' = f(z; a) = f(f(x; \theta); a) = f(x; \varphi(\theta; a)) = f(x; \theta')$$

или, символически,

$$z' = g_a z = g_a g_\theta x = g_{\theta'} x \quad \text{для всех } x,$$

то

$$g_{\theta'}^{-1} g_a = g_\theta^{-1}$$

и, следовательно, в силу (1.6),

$$P_{\bar{g}_a\theta}(g_a A) = P_\theta(A).$$

Если заданы произвольное семейство плотностей $P_\theta(z)$ и группы G, $\bar{G}$ преобразований вида (1.5), то для проверки инвариантности семейства $P_\theta(z)$ относительно G, $\bar{G}$, т. е. справедливости равенства

$$P_\theta(z) = P_{\varphi(\theta; a)}(f(z; a)) \operatorname{Det}\left(\frac{\partial f_i(z; a)}{\partial z_j}\right)$$

тождественно по z, θ, a, привлекаются инфинитезимальные операторы

$$\sum_{k=1}^{n} \xi_\alpha^k \frac{\partial}{\partial z_k}, \quad \sum_{k=1}^{r} \eta_\beta^k \frac{\partial}{\partial \theta_k},$$

где

$$\xi_\alpha^k = \xi_\alpha^k(z) = \left(\frac{\partial f_k}{\partial a_\alpha}\right)_{a=a_0}, \quad \begin{matrix} k = 1, 2, \ldots, n, \\ \alpha = 1, 2, \ldots, r, \end{matrix}$$
$$\eta_\beta^k = \eta_\beta^k(\theta) = \left(\frac{\partial \varphi_k}{\partial a_\beta}\right)_{a=a_0}, \quad \begin{matrix} k = 1, 2, \ldots, r \\ \beta = 1, 2, \ldots, r. \end{matrix}$$

Необходимое и достаточное условие инвариантности $p_\theta(z)$ относительно (связных) групп G и $\bar{G}$ имеет вид [4]

$$\sum_{k=1}^{n} \xi_\alpha^k(z) \frac{\partial \log p_\theta(z)}{\partial z_k} + \sum_{k=1}^{r} \eta_\beta^k(\theta) \frac{\partial \log p_\theta(z)}{\partial \theta_k} = -\sum_{k=1}^{n} \frac{\partial \xi_\alpha^k(z)}{\partial z_k}, \tag{1.7}$$

Аналогичным образом, необходимым и достаточным условием инвариантности функции потерь $L(\theta, d)$ относительно групп $\bar{G}$ и G^* (для случая, когда

элементы G^* имеют вид $g_a^*: d' = \psi(d; a)$) является равенство

$$\sum_{k=1}^{r} \eta_\beta^k(\theta) \frac{\partial L(\theta, d)}{\partial \theta_k} + \sum_{k=1}^{s} \xi_\beta^k(d) \frac{\partial L(\theta, d)}{\partial d_k} = 0, \quad \beta = 1, 2, \ldots, r, \qquad (1.8)$$

где

$$\xi_\gamma^k = \xi_\gamma^k(d) = \left(\frac{\partial \psi_k}{\partial a_\gamma}\right)_{a=a_0}, \quad \begin{matrix} k = 1,2, \ldots, s \\ \gamma = 1,2, \ldots, r \end{matrix}$$

— инфинитезимальные преобразования группы G^*(s — число измерений пространства решений). Соотношения (1.7) и (1.8) полезны не только при проверке условий инвариантности; они могут служить также для отыскания семейств распределений и функций потерь, инвариантных относительно групп, заданных своими инфинитезимальными операторами. Для этого достаточно разрешить уравнения (1.7) и (1.8) относительно неизвестных функций $p_\theta(z)$ и $L(\theta, d)$.

Приведем некоторые свойства инвариантных оценок.

1. Класс $\mathfrak{I}$ инвариантных оценок замкнут относительно группы G: если $\delta \in \mathfrak{I}$ (т. е. $g^*\delta g^{-1} = \delta$ для всех $g \in G$), то $h^*\delta h^{-1} \in \mathfrak{I}$ для любого $h \in G$. Этим свойством замкнутости обладают и некоторые другие классы оценок, например класс несмещенных [2] и, очевидно, класс всех возможных оценок. Однако существование равномерно наилучшей оценки в любом из таких классов не может быть гарантировано, а в классе инвариантных оценок равномерно наилучшее решение обычно существует (для случая транзитивной группы оно существует всегда). Тем не менее, если в произвольном классе $\mathfrak{C}$ процедур, замкнутом относительно G, существует единственная равномерно наилучшая процедура δ_0, то δ_0 инвариантна [2]. В самом деле, из $R(\theta, \delta_0) \leq R(\theta, \delta)$ для всех $\theta \in \Omega$ и $\delta \in \mathfrak{C}$ вытекает $R(\bar{g}\theta, g^*\delta_0 g^{-1}) \leq$ $\leq R(\bar{g}\theta, g^*\delta g^{-1})$ для всех $\bar{g}\theta = \theta' \in \Omega$ и $g^*\delta g^{-1} = \delta' \in \mathfrak{C}$, так что $g^*\delta_0 g^{-1}$ — тоже равномерно наилучшая процедура. В силу единственности последней должно быть $g^*\delta_0 g^{-1} = \delta_0$, т. е. $\delta_0 \in \mathfrak{I}$.

2. Из тождества $R(\theta, \delta) = R(\bar{g}\theta, g^*\delta g^{-1})$, справедливого для любой инвариантной задачи, вытекает, что класс $\mathfrak{M}$ минимаксных решений δ_0 (т. е. таких, что $\sup_\theta R(\theta, \delta_0) \leq \sup_\theta R(\theta, \delta)$ для любой δ из класса $\mathfrak{D}$ всех возможных оценок) замкнут относительно группы G. В самом деле, если $\delta_0 \in \mathfrak{M}$, то $\sup_\theta R(\bar{g}\theta, g^*\delta_0 g^{-1}) \leq \sup_\theta R(\bar{\theta}, g^*\delta g^{-1})$ для любой $g^*\delta g^{-1} =$ $= \delta' \in D$ ($\bar{g}\Omega = \Omega$ и $g^*Dg^{-1} = \mathfrak{D}$), так что $g^*\delta_0 g^{-1} \in \mathfrak{M}$. Отсюда вытекает, что если минимаксное правило единственно, то оно инвариантно. Из предыдущего пункта следует, что если в $\mathfrak{M}$ имеется единственная процедура с равномерно наименьшим риском, то она инвариантна.

3. Пусть λ — относительно инвариантная мера на параметрическом пространстве Ω, т. е. $\lambda(\bar{g}A) = \Delta(\bar{g})\lambda(A)$ для любых $\bar{g} \in \bar{G}$ и $A \subset \Omega$ ($\Delta(\bar{g})$ — вещественнозначная функция на группе $\bar{G}$, $\Delta(\bar{g}) > 0$, $\Delta(\bar{g}_1)\Delta(\bar{g}_2) = \Delta(\bar{g}_1\,\bar{g}_2)$. Если существует интеграл $\int_{\Omega} R(\theta, \delta)\,\lambda(d\theta) = r(\lambda, \delta)$, то для симметричной задачи он подчиняется правилу $r(\lambda, \delta) = r(\lambda, g^* \delta g^{-1})/\Delta(\bar{g})$. Класс $\mathfrak{B}$ бейесовых решений δ_0 (т. е. таких, что $r(\lambda, \delta_0) \leq r(\lambda, \delta)$ для любой $\delta \in \mathfrak{D}$) замкнут относительно G. В самом деле, если $\delta_0 \in D$, то $r(\lambda, g^* \delta_0 g^{-1}) \leq r(\lambda, g^* \delta g^{-1})$ (после сокращения на $1/\Delta(\bar{g})$), так что $g^* \delta_0 g^{-1} \in \mathfrak{B}$. Если бейесова оценка единственна, то она инвариантна.

Для транзитивной группы $\bar{G}$ $R(\theta, \delta) = \varrho(\delta)$ не зависит от $\theta \in \Omega$, если $\delta \in \mathfrak{I}$ [2]. Если δ_0 — единственная инвариантная оптимальная оценка, то она является оптимальной минимаксной и оптимальной бейесовой, причем

$$r(\lambda, \delta_0) = \sup_{\theta} R(\theta, \delta_0) = \inf_{\delta \in I} \varrho(\delta)\,.$$

4. При определенных условиях из каждого правила δ можно сконструировать инвариантное правило $\tilde{\delta}$ такое, что $\sup_{\theta} R(\theta, \tilde{\delta}) \leq \sup_{\theta} R(\theta, \delta)$. В самом деле, пусть δ — рандомизированное правило ($\delta(A/z)$ — вероятность принять решение $d \in A$ при заданном z) и существуют интегралы $\int_{G} R(\bar{g}\,\theta, \delta)\,d\nu(g)$ и

$$\tilde{\delta}(A \mid z) = \int_{G} \delta(g^* A \mid gz)\,d\nu(g)$$

по правоинвариантной мере ν на компактной группе G. Правило $\tilde{\delta}$ инвариантно в том смысле, что для любого $h \in G$

$$\tilde{\delta}(h^* A \mid hz) = \int_{G} \delta(g^* h^* A \mid ghz)\,d\nu(g) = \int_{G} \delta(k^* A \mid kz)\,d\nu(kh^{-1}) = \tilde{\delta}(A \mid z)\,,$$

поскольку $gh = k \in G$ и $d\nu(kh^{-1}) = d\nu(k)$. Далее,

$$R(\theta, \tilde{\delta}) = \int_{G} R(\bar{g}\,\theta, \delta)\,d\nu(g) \leq \sup_{\theta} R(\theta, \delta) \int_{G} d\nu(g)\,,$$

откуда при нормировке $\nu(G) = 1$ вытекает

$$\sup_{\theta} R(\theta, \tilde{\delta}) \leq \sup_{\theta} R(\theta, \delta)\,.$$

5. Для инвариантного семейства распределений P_θ оценки максимального правдоподобия являются инвариантными. Пусть $d = \delta(z)$ — оценка максимального правдоподобия,

$$\max_{\theta \in \Omega} p_\theta(z) = p_d(z)$$

($p_\theta(z) = dP_\theta/d\mu$ для некоторой меры μ). Поскольку

$$p_\theta(z) = p_{\bar{g}\theta}(gz) \cdot D(g, z)$$

и якобиан D не зависит от θ, максимум по θ функции $p_{\bar{g}\theta}(gz)$ достигается в той же точке d, так что максимум по θ' функции $p_{\theta'}(z')$ (где $\theta' = \bar{g}\,\theta$, $z' = gz$) достигается в точке $g^* d = \delta(z')$, откуда вытекает инвариантность δ относительно G:

$$g^*\,\delta(z) = \delta\, g(z)\,.$$

Таким образом, класс инвариантных оценок достаточно широк. Для симметричных задач он содержит оценки максимального правдоподобия, минимаксные оценки, бейесовы оценки для специального априорного распределения параметра θ, а также, как показано в [5], для случая линейной регрессии, оценки метода наименьших квадратов.

§ 2. Метод определения оптимальных инвариантных оценок

В ряде частных случаев отыскание оптимальных инвариантных оценок не вызывает затруднений. Пусть пространства входных сигналов, параметров и решений конечномерны, а группы их преобразований — конечнопараметрические группы Ли с элементами вида (1.5). Будем считать, что существует взаимнооднозначное и биизмеримое преобразование $(z_1 z_2, \ldots, z_n) \to (J_1, \ldots, J_k, T_1, \ldots, T_{n-k})$, где $\mathbf{J}=(J_1, \ldots, J_k)$, $\mathbf{T}=(T_1, \ldots, T_{n-k})$ — векторные инвариант и достаточная статистика задачи соответственно, причем на пространстве значений $\mathbf{T}$ индуцируется однотранзитивная группа, так что для любой точки $\mathbf{T}$ найдется единственное преобразование $g_{a(T)}$ с параметром a, зависящим от $\mathbf{T}$, для которого $g^{-1}_{a(T)}\,\mathbf{T} = \mathbf{T}_0$, где $\mathbf{T}_0$ — произвольная фиксированная точка. Если $\delta_1(\mathbf{z})$ — инвариантное решение, тогда функция $\delta(\mathbf{T}, \mathbf{J}) = \delta_1(\mathbf{z}(\mathbf{T}, \mathbf{J}))$ удовлетворяет соотношению $\delta(\mathbf{T}, \mathbf{J}) = g^*_a\,\delta\, g_a^{-1}$ $(\mathbf{T}, \mathbf{J})$. Выбирая здесь параметр $a = a(\mathbf{T})$ так, чтобы $g^{-1}_{a(T)}\,\mathbf{T} = \mathbf{T}_0$ (тогда $g^{-1}_{a(T)}(\mathbf{T}, \mathbf{J}) = (g^{-1}_{a(T)}\,\mathbf{T}, g^{-1}_{a(T)}\,\mathbf{J}) = (\mathbf{T}_0, \mathbf{J})$), и обозначая $\delta(\mathbf{T}_0, \mathbf{J}) = f(\mathbf{J})$, находим $\delta(\mathbf{T}, \mathbf{J}) = g^*_{a(T)}\, f(\mathbf{J})$. Таким образом, инвариантное правило «факторизовано» относительно переменных $\mathbf{T}$, $\mathbf{J}$, и его значение в произвольной точке $\mathbf{T}$ инвариантного многообразия $\mathbf{J}(\mathbf{z}) = \mathbf{J}$ получается применением группового преобразования $g^*_{a(T)}$ с параметром $a = a(\mathbf{T})$ к постоянному на заданном многообразии вектору $\mathbf{f} = \mathbf{f}(\mathbf{J})$.

Преобразование $\mathbf{z} \to (\mathbf{T}, \mathbf{J})$ переводит $dP_\theta(\mathbf{z})$ в вероятностный элемент $p_\theta(\mathbf{T}\,|\,\mathbf{J})\,d\mu(\mathbf{T})\,d\lambda_\theta(\mathbf{J})$, где λ_θ — мера на множестве значений $\mathbf{J}$, $p_\theta(\mathbf{T}\,|\,\mathbf{J})\,d\mu(\mathbf{T})$ — элемент условного распределения с плотностью $p_\theta(\mathbf{T}\,|\,\mathbf{J})$ по мере μ, получаемой перенесением на множество значений $\mathbf{T}$ левоинвариантной меры дейст-

вующей на нем группы. Выражение для риска $R(\theta, \delta)$ инвариантного правила $\delta(\mathbf{T}, \mathbf{J}) = g^*_{a(T)} f(\mathbf{J})$ принимает вид

$$R(\theta, \delta) = \int L\big(\theta, g^*_{a(T)} f(\mathbf{J})\big)\, p_\theta(\mathbf{T} \mid \mathbf{J})\, d\mu(\mathbf{T})\, d\lambda_\theta(\mathbf{J}). \tag{2.1}$$

Будем считать, что $\bar{G}$ действует на Ω однотранзитивно, тогда λ_θ не зависит от θ [2]. Для нахождения $\hat{f}(\mathbf{J})$, минимизирующего (2.1), достаточно минимизировать при каждом $\mathbf{J}$ условный риск

$$\varrho_\theta(f \mid \mathbf{J}) = \int L(\theta, g^*_{a(T)} f)\, p_\theta(\mathbf{T} \mid \mathbf{J})\, d\mu(\mathbf{T}).$$

В силу предположенной однотранзитивности $\bar{G}$ на Ω для любой точки $\theta \in \Omega$ найдется единственное преобразование $\bar{g}_{a(\theta)}$ с параметром $a = a(\theta)$, для которого $\theta = \bar{g}_{a(\theta)} \theta_0$, где θ_0 — некоторая фиксированная точка, $\theta_0 \in \Omega$. Выполним в выражении для $\varrho_\theta(f/\mathbf{J})$ замену переменных $\mathbf{T} = g_{a(\theta)} \tau$ (θ фиксировано), используя инвариантность μ, p_θ и L (т. е. $p_\theta(\mathbf{T} \mid \mathbf{J})\, d\mu(\mathbf{T}) = p_{\theta_0}(\tau \mid \mathbf{J})\, d\mu(\tau)$, $L(\bar{g}_{a(\theta)} \theta_0,\ g^*_{a(T)} f) = L(\bar{g}^{-1}_{a(T)} \bar{g}_{a(\theta)} \theta_0, f)$) и замечая, что $\bar{g}^{-1}_{a(T)}\, \bar{g}_{a(\theta)} = \bar{g}^{-1}_{a(\tau)}$ (это вытекает из равенств $g_{a(\theta)} \tau = \mathbf{T} = g_{a(T)}\, \mathbf{T}_0$, $\tau = g_{a(\tau)}\, \mathbf{T}_0$ и предполагаемого нами изоморфизма групп $\bar{G}$ и G). В результате указанной замены переменных $\varrho_\theta(f \mid \mathbf{J})$ преобразуется к виду

$$\varrho_\theta(f \mid \mathbf{J}) = \varrho_{\theta_0}(f \mid \mathbf{J}) = \int L(\bar{g}^{-1}_{a(\tau)} \theta_0, f)\, p_{\theta_0}(\tau \mid \mathbf{J})\, d\mu(\tau). \tag{2.2}$$

Если семейство распределений $\mathbf{T}$ ограниченно-полно, тогда $\mathbf{T}$ и $\mathbf{J}$ статистически независимы [6], и $\varrho_{\theta_0}(f \mid \mathbf{J})$, а следовательно и $\hat{f}(\mathbf{J})$ не зависят от $\mathbf{J}$. Но даже и в общем случае произвольной статистики $\mathbf{T}$ задача определения оптимальной инвариантной функции $\hat{\delta}(\mathbf{T}, \mathbf{J})$ не намного сложнее бейесовой задачи (2.2) оценки параметра $\bar{g}^{-1}_{a(T)} \theta_0$ по выборке $\mathbf{J}$.

В § 3 будут использованы обозначения, принятые в [7], в соответствии с которыми, например, $\bar{g}^{-1}_{a(T)} \theta$ представляется в виде $\mathbf{T}^{-1} \circ \theta$, $g^{-1}_{a(T)} \mathbf{T}$ — в виде $\theta^{-1} \circ \mathbf{T}$, $g^*_{a(T)} f$ — в виде $\mathbf{T} \circ f$ и т. д. В силу предположенного изоморфизма групп $G, \bar{G}, G^*$ и пространств $\mathbf{J}(\mathbf{z}) = \mathbf{J}$, Ω, D, на которых они действуют, эти обозначения корректны.

Изложенный метод можно применить и в случае бесконечномерного пространства входных сигналов, если возможна предварительная редукция задачи к конечномерному случаю (например, с помощью принципа достаточности).

В качестве примера отыскания оптимальных инвариантных оценок рассмотрим задачу оценки неизвестных параметров σ^2 и $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_r)$ в схеме

$$Z(\tau) = \sum_{k=1}^{r} \theta_k \varphi_k(\tau) + \sigma X(\tau), \quad s - T \leq \tau \leq s. \tag{2.3}$$

(s — правый конец интервала наблюдения — предполагается фиксированным), где $\varphi_k(t)$, $k = 1, 2, \ldots, r$ — известные функции, $X(t)$ — гауссовский случайный процесс с нулевым средним и известной корреляционной функцией $k_x(t_1, t_2)$ (частный случай $\delta = 1$ модели (2.3) рассмотрен в [5]). Функционалы

$$\eta_k = \int_{s-T}^{s} g_k(s, \tau)\, Z(\tau)\, d\tau, \qquad k = 1, 2, \ldots, r,$$

где $g_k(s, \tau)$ — решение интегрального уравнения

$$\int_{s-T}^{s} k_x(t, \tau)\, g(s, t)\, dt = \varphi_k(s), \qquad s - T \leq \tau \leq s$$

с ядром $k_x(t, \tau)$ и функцией $\varphi_k(s)$ в правой части, являются достаточными статистиками семейства $P_{\theta,\sigma}$, $-\infty < \theta < \infty$ распределений процесса $Z(\tau)$ при любом фиксированном $\sigma > 0$. Вводя обозначение $\mathrm{B} = (b_{ij})$ для матрицы с элементами

$$b_{ij} = \int_{s-T}^{s} g_i(s, \tau)\, \varphi_j(\tau)\, d\tau, \qquad i, j = 1, 2, \ldots, \tau,$$

и обозначения $\boldsymbol{\eta}^T = (\eta_1, \eta_2, \ldots, \eta_r)$, $\boldsymbol{\xi}^T = (\xi_1, \xi_2, \ldots \xi_r)$, где

$$\xi_k = \int_{s-T}^{s} g_k(s, \tau)\, X(\tau)\, d\tau$$

(T — знак транспонирования), модель (2.3) можно преобразовать к матричному виду

$$\boldsymbol{\eta} = \mathbf{B}\,\theta + \boldsymbol{\sigma}\boldsymbol{\xi}. \tag{2.4}$$

Семейство распределений вектора $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_r)$ с компонентами вида (2.4) допускает статистически независимые достаточные статистики

$$\mathbf{T}_1 = (\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1}\boldsymbol{\Phi}^T\boldsymbol{\eta}, \quad \mathbf{T}_2 = \boldsymbol{\eta}^T\boldsymbol{\eta} - T_1^T\boldsymbol{\Phi}^T\boldsymbol{\eta} \quad (\boldsymbol{\Phi}^T = (\mathbf{B}, \mathbf{B}, \ldots \mathbf{B})\,(n \text{ раз}))$$

и инвариантно относительно преобразований

$$\mathbf{T}_1' = k\mathbf{T}_1 + \mathbf{c}, \mathbf{T}_2' = k^2\mathbf{T}_2, \ \theta' = k\theta + \mathbf{c}, \ \boldsymbol{\sigma}' = k\boldsymbol{\sigma} \quad \begin{pmatrix} k > 0 \\ -\infty < \mathbf{c} < \infty \end{pmatrix}.$$

Если $d = (d_1, \ldots, d_r)$ оценивает r-мерный вектор θ, а d_{r+1} — скаляр σ^2, тогда при $\bar{g} = g^*$ произвольная инвариантная функция потерь зависит от $(\theta - d)/\sigma$ и d_{r+1}/σ^2. Для определенности примем ее равной

$$\frac{1}{\sigma^2}(\theta - d)^T(\theta - d) + \left(1 - \frac{d_{r+1}}{\sigma^2}\right)^2. \tag{2.5}$$

Векторный параметр $\mathbf{a}(\mathbf{T}) = (\mathbf{c}(\mathbf{T}),\ k(\mathbf{T}))$ преобразования $g^*_{a(T)}$ определяется из условий $\mathbf{T}_1 = k\,\mathbf{T}_1^0 + \mathbf{c}$, $\mathbf{T}_2 = k^2\,\mathbf{T}_2^0$, так что $\mathbf{c} = \mathbf{T}_1$, $k^2 = \mathbf{T}_2$, если выбрать $\mathbf{T}^0 = (0,\ 0, \ldots,\ 0,\ 1)$. Поэтому оптимальные инвариантные оценки равны

$$\hat{\delta}_1(\mathbf{T}) = \mathbf{T}_1 + \hat{f}_1\sqrt{\mathbf{T}_2},\quad \hat{\delta}_2(\mathbf{T}) = \hat{f}_2 \cdot \mathbf{T}_2\,.$$

Поскольку здесь $\mathbf{T}$ и $\mathbf{J}$ статистически независимы, $\hat{f}_1$ и $\hat{f}_2$ не зависят от $\mathbf{J}$.

В выражение для риска входит переменная $\boldsymbol{\alpha} = \bar{g}^{-1}_{a(T)}\,\theta_0$. Найдем ее, положив $\boldsymbol{\theta}_0 = (0, \ldots,\ 0,\ 1)$. Из условия $\tau = g_{a(\tau)}\,\mathbf{T}_0$, определяющего $\mathbf{a}(\tau)$, находим, что $\mathbf{a}(\tau) = (\mathbf{c}(\tau),\ k(\tau)) = (\tau_1, \tau_2)$, а из $\boldsymbol{\alpha} = g^{-1}_{a(\tau)}\,\boldsymbol{\theta}_0$ получаем $\alpha = (\alpha_1, \alpha_2) = \left(-\dfrac{\tau_1}{\sqrt{\tau_2}},\ \dfrac{1}{\tau_2}\right)$. Таким образом, $\hat{f}_1$, $\hat{f}_2$ определяются из условия

$$\min_{f_1, f_2} E_0\left\{(\tau_1 + \sqrt{\tau_2}\,f_1)^T(\tau_1 + \sqrt{\tau_2}\,f_1) + (1 - \tau_2 f_2)^2\right\},$$

где E_0 — операция осреднения по распределению случайных величин τ_1, τ_2, отвечающему значениям $\boldsymbol{\theta} = 0$, $\boldsymbol{\sigma} = 1$ параметров θ, σ. Нетрудно видеть, что $\hat{f}_1 = 0$ (τ_1 и τ_2 статистически независимы) и $\hat{f}_2 = 1/(n - r + 2)$, так что окончательно

$$\hat{\delta}_1(\eta) = (\boldsymbol{\Phi}\boldsymbol{\Phi}^T)^{-1}\boldsymbol{\Phi}^T\eta,\quad \hat{\delta}_2(\eta) = \frac{1}{n - r + 2}(\eta - \boldsymbol{\Phi}\hat{\delta}_1)^T \times (\eta - \boldsymbol{\Phi}\hat{\delta}_1)\,.$$

Первая оценка совпадает с обычной оценкой метода наименьших квадратов, вторая оказывается смещенной: вместо множителя $\dfrac{1}{n - r}$, обеспечивающего несмещенность $\hat{\delta}_2$, здесь фигурирует $1/(n - r + 2)$.

Если вместо σ^2 требуется оценить σ при функции потерь $(1 - d_{r+1}/\delta)^2$, тогда, как нетрудно показать, оценка σ равна

$$\hat{\delta}_3(\eta) = \hat{f}_3 \cdot \mathbf{T}_2,\quad \hat{f}_3 = \frac{1}{\sqrt{2}} \cdot \frac{\Gamma\left(\dfrac{n - r + 1}{2}\right)}{\Gamma\left(\dfrac{n - r + 2}{2}\right)}\,.$$

При больших n получается приближение $f_3 \approx 1/\sqrt{n - r + 2}$, так что квадрат оценки близок к оценке квадрата σ.

§ 3. Бейесов метод и инвариантность.

Предположим, что статистическая задача инвариантна относительно группы, допускающей существование левоинвариантной (μ) и правоинвариантной (ν) мер. При допущениях § 2 об изоморфизме групп $G, \bar{G}, G^*$ и про-

странств $\mathbf{J}(\mathbf{z}) = \mathbf{J}$, Ω, D, на которых они действуют, меры μ и ν можно с группы G перенести на любой из изоморфных ей объектов. Если мера ν нормирована, тогда, принимая ее в качестве априорной на Ω, для апостериорного бейесового риска $r(d\,|\,\mathbf{T}, \mathbf{J})$ получим выражение

$$r(d\,|\,\mathbf{T}, \mathbf{J}) = \int_{\Omega} L(\theta, d)\, p_0(\theta^{-1}\circ\mathbf{T}\,|\,\mathbf{J})\, d\nu(\theta) \Big/ \int_{\Omega} p_0(\theta^{-1}\circ\mathbf{T}/\mathbf{J})\, d\nu(\theta). \tag{3.1}$$

Выполняя здесь замену переменных по формуле $\theta^{-1}\circ\mathbf{T} = \tau$ ($\mathbf{T}$ фиксировано), учитывая, что $L(\mathbf{T}\circ\tau^{-1}, d) = L(\tau^{-1}, \mathbf{T}^{-1}\circ d)$ и что $d\nu(\mathbf{T}\circ\tau^{-1}) = \Delta^{-1}(\mathbf{T})\, d\mu(\tau)$ ($\Delta(\cdot)$ — модулярная функция), а также вводя обозначение $d = \mathbf{T}\circ f$, найдем

$$r(\mathbf{T}\circ f\,|\,\mathbf{T}, \mathbf{J}) = \varrho(f\,|\,\mathbf{J}) \tag{3.2}$$

(здесь для краткости τ и τ^{-1} обозначают $\tau\circ\mathbf{T}^0$ и $\tau^{-1}\cdot\theta_0$ соответственно; аналогичные сокращения используются без оговорок и в других сходных случаях). Отсюда заключаем, что минимум условного бейесового риска $r(d\,|\,\mathbf{T}, \mathbf{J})$ достигается при $\hat{d} = \mathbf{T}\circ\hat{f}(\tau)$, совпадающем со значением $\mathbf{T}\circ\hat{f}(\mathbf{J})$ оптимального инвариантного решения при заданных $\mathbf{T}$ и $\mathbf{J}$.

Если мера ν не нормируема, $r(d\,|\,\mathbf{T}, \mathbf{J})$ нельзя интерпретировать как условный бейесов риск. Однако формула (3.2) сохраняет свое значение в качестве вспомогательного средства при отыскании оптимальных инвариантных решений, поскольку вычисление $r(\mathbf{T}\circ f\,|\,\mathbf{T}, \mathbf{J})$, может оказаться значительно более простой задачей, чем вычисление $\varrho(f\,|\,\mathbf{J})$. Для примера, разобранного в § 2, $d\nu(\theta, \sigma) = d\theta_1 \ldots d\theta_r\, d\sigma/\sigma$; используя (3.2), можно снова получить приведенные там оценки.

Пусть

$$d\nu_n(\theta) = p_n(\theta)\, d\nu(\theta) = \begin{cases} \dfrac{q(\theta)\, d\nu(\theta)}{\int\limits_{\theta_n} q(\theta)\, d\nu(\theta)}, & \theta \in \theta_n \\ 0, & \theta \overline{\in} \theta_n \end{cases} \tag{3.3}$$

— последовательность вероятностных мер, где Θ_n — множества в Ω, $n = 1, 2, \ldots$, а $q(\theta)$ — произвольная плотность из класса относительно инвариантных, удовлетворяющих условию $q(\theta_1\circ\theta_2) = q(\theta_1)\, q(\theta_2)$ для любых θ_1, θ_2. (Причина, по которой, следуя [8], мы ограничиваемся этим классом, состоит в том, что бейесова задача с априорной мерой (3.3) имеет в этом случае инвариантное решение). Отвечающая (3.3) последовательность бейесовых рисков

$$r_n(\mathbf{T}\cdot f\,|\,\mathbf{T}, \mathbf{J}) = \frac{\int\limits_{\theta_n^{-1}\circ T} L(\tau^{-1}, f)\, q(\tau^{-1})\, p_0(\tau\,|\,\mathbf{J})\, d\mu(\tau)}{\int\limits_{\theta_n^{-1}\circ T} q(\tau^{-1})\, p_0(\tau\,|\,\mathbf{J})\, d\mu(\tau)} \tag{3.4}$$

при заданном $\mathbf{J}$ является последовательностью функций случайного аргумента, имеющего распределение (условное, при заданном $\mathbf{J}$) $P_n(\mathbf{T})\,d\mu(\mathbf{T})$, где

$$p_n(\mathbf{T}) = \int\limits_{\Theta_n} p_0(\theta^{-1}\circ\mathbf{T}\,|\,\mathbf{J})\,d\nu_n(\theta)$$

($d\nu_n$ определена в (3.3)). Можно показать, что

$$p_n(\mathbf{T})\,d\mu(\mathbf{T}) = \frac{\int\limits_{\Theta_n^{-1}\circ T} q(\tau^{-1})\,p_0(\tau\,|\,\mathbf{J})\,d\mu(\tau)}{\int\limits_{\Theta_n} q(\theta)\,d\nu(\theta)}\,q(\mathbf{T})\,d\nu(\mathbf{T})\,. \tag{3.5}$$

Вероятностное распределение с плотностью $p_n(\mathrm{T})$ по мере μ будем обозначать через P_n. Справедлива следующая теорема о байесовом характере инвариантных задач.

Условие $q(\theta)\equiv 1$ необходимо и достаточно для сходимости $r_n(\mathbf{T}\circ f\,|\,\mathbf{T},\mathbf{J})$ вида (3.4) к $\varrho(f\,|\,\mathbf{J})$ в смысле

$$\lim_{n\to\infty} P_n\{\mathbf{T}: |\,r_n(\mathbf{T}\circ f\,|\,\mathbf{T},\mathbf{J}) - \varrho(f\,|\,\mathbf{J})\,| > \varepsilon\} = 0\,. \tag{3.6}$$

Таким образом, инвариантный риск является пределом бейесовых. Отсюда при определенных не рассматриваемых здесь условиях можно заключить, что оптимальное инвариантное решение является пределом, в некотором смысле, оптимальных бейесовых решений.

Приложение. Доказательство теоремы

Лемма. Пусть V_n и $W_n (n = 1, 2, \ldots)$ — две последовательности множеств, на которых интегрируемы случайные величины $|X_n(\omega) - X(\omega)|^2$, и п. н. $\sup\limits_{\omega\in W_n} |X_n(\omega) - X(\omega)| > 0$. Тогда

$$\frac{\int\limits_{W_n} |X_n - X|^2\,dP - \varepsilon^2}{\text{п. н. } \sup\limits_{\omega\in W_n} |X_n - X|^2} \le P\{|X_n - X| \ge \varepsilon\} \le \frac{1}{\varepsilon^2}\int\limits_{V_n} |X_n - X|^2\,dP + P(V_n^c)$$

(V_n^c — дополнение множества V_n в пространстве элементарных событий).

Левое неравенство доказывается как в [9], правое следует из замечания, что $\dfrac{|X_n - X|^2}{\varepsilon^2} \ge 1$ на множестве $A_n = \{\omega : |X_n(\omega) - X(\omega)| > \varepsilon\}$, так что

$$P(A_n) = P(A_n \cap V_n) + P(A_n \cap V_n^c) \le \int\limits_{A_n\cap V_n} \frac{|X_n - X|^2}{\varepsilon^2}\,dP + P(V_n^c)\,.$$

Из леммы вытекает полезное *следствие*. Пусть для некоторой последовательности множеств V_n, $n = 1, 2, \ldots$

а) $\lim\limits_{n\to\infty} P(V_n) = 1$,

б) $\lim\limits_{n\to\infty} \int\limits_{V_n} |X_n - X|^2 dP = 0$.

Тогда $X_n \xrightarrow{P} X$, так что а) и б) достаточны для сходимости по вероятности.

Лемма и следствие остаются справедливыми и для последовательности мер P_n, т. е. для случая, когда везде P заменяется на P_n, а $X_n \xrightarrow{P_n} X$ означает

$$\lim_{n\to\infty} P_n\{|X_n - X| > \varepsilon\} = 0.$$

На группу G наложим следующее ограничение: для любой последовательности C_n множеств точек τ существуют последовательности Θ_n и Ω и W_n в пространстве значений T такие, что а) $\Theta_n^{-1} \circ T \supset C_n$ для любого $T \in W_n$, б) $\lim \nu(W_n)/\nu(\Theta_n) = 1$. Это условие не слишком ограничительно и выполняется для большинства групп, возникающих в статистических приложениях [8] ($\Theta_n^{-1} \circ \mathbf{T}$ означает множество точек τ вида $\tau = g_{a(\theta)}^{-1}\mathbf{T}$, где $\Theta \in \Theta_n$).

Приступим к доказательству теоремы.

1) *необходимость условия* $q(\theta) \equiv 1$.

Выберем последовательность ε_n, $n = 1, 2, \ldots$, $\varepsilon_n \to 0$ и возрастающую последовательность множеств $C_1 \subset C_2 \subset \ldots$ такую, что

$$\frac{\int\limits_{C_n} q(\tau^{-1})\, p_0(\tau \mid \tau)\, d\mu(\tau)}{\int q(\tau^{-1})\, p_0(\tau \mid J)\, d\mu(\tau)} > 1 - \varepsilon_n. \qquad (\text{П. 1})$$

(здесь и далее отсутствие указания на область интегрирования означает интегрирование по всему пространству). Оценим каждый из трех интегралов в правой части тождества

$$\int\limits_{W_n} (r_n - \varrho)^2\, dP_n = \int\limits_{W_n} r_n^2\, dP_n - 2\varrho \int\limits_{W_n} r_n\, dP_n + \varrho^2 \int\limits_{W_n} dP_n, \qquad (\text{П. 2})$$

где для краткости обозначено $r_n = r_n(\mathbf{T} \circ f/\mathbf{T},\ \mathbf{J})$, $\varrho = \varrho(f/\mathbf{J})$, $dP_n = p_n(\mathbf{T})\, d\mu(\mathbf{T})$. Третий интеграл не превосходит 1, второй сходится к ϱ (последнее вытекает из явных выражений (3.3) и (3.5) для r_n и dP_n), а первый преобразуется к виду

$$\iint H_n(\tau_1, \tau_2)\, q(\tau_1^{-1})\, q(\tau_2^{-1})\, L(\tau_1^{-1}, f)\, L(\tau_2^{-1}, f)\, p_0(\tau_1 \mid J)\, p_0(\tau_2 \mid J)\, d\mu(\tau_1)\, d\mu(\tau_2), \qquad (\text{П. 3})$$

где (χ_n — индикатор множества $\Theta_n^{-1}\circ\mathbf{T}$)

$$H_n(\tau_1,\tau_2)=\frac{1}{\int\limits_{\Theta n} q(\theta)\,d\nu(\theta)}\int\limits_{W_n}\frac{q(T)\,\chi_n(\tau_1)\,\chi_n(\tau_2)}{\int q(\sigma^{-1})\,\chi_n(\sigma)\,p_0(\sigma\,|\,J)\,d\mu(\sigma)}\,d\nu(T) \qquad (\text{П. 4})$$

(предполагается допустимость изменения порядка интегрирования). Поскольку интегрирование в (П. 4) осуществляется по $T\in W_n$, для которых $\Theta_n^{-1}\circ\mathbf{T}\supset C_n$, то в силу (П. 1)

$$H_n(\tau_1,\tau_2)\le\frac{(1-\varepsilon_n)^{-1}}{\int q(\sigma^{-1})\,p_0(\sigma\,|\,J)\,d\mu(\sigma)}\cdot\frac{\int q(T)\,\chi_n(\tau_1)\,\chi_n(\tau_2)\,d\nu(T)}{\int\limits_{\Theta n} q(\theta)\,d\nu(\theta)},$$

где χ_n — индикатор множества $\theta_n^{-1}\circ\mathbf{T}$. Так как $\chi_n(\tau_1)\chi_n(\tau_2)=\min\limits_{j=1,2}\{\chi_n(\tau_j)\}$, а $\int q(\mathbf{T})x_n(\tau_j)\,d\nu(\mathbf{T})$ после замены переменных $\mathbf{T}=\theta\circ\tau_j$ (τ_j фиксировано) приводится к $q(\tau_i)\int\limits_{\Theta n} q(\theta)\,d\nu(\theta)$, то

$$H_n(\tau_1,\tau_2)\le\frac{(1-\varepsilon_n)^{-1}}{\int q(\sigma^{-1})\,p_0(\sigma\,|\,J)\,d\mu(\sigma)}\min_{j=1,2}\{q(\tau_j)\}.$$

Замечая, что

$$\min_{j=1,2}\{q(\tau_j)\}=\frac{1}{2}\,(q(\tau_1)+q(\tau_2))-\frac{1}{2}\,|\,q(\tau_1)-q(\tau_2)\,|\,,$$

для интеграла в (П. 3) получаем оценку сверху вида $(r\varrho-K)/(1-\varepsilon_n)$, где обозначено

$$r=\frac{\int g(\tau)^{-1}\,L(\tau^{-1},f)\,p_0(\tau\,|\,J)\,d\mu(\tau)}{\int q(\tau^{-1})\,p_0(\tau\,|\,J)\,d\mu(\tau)},$$

$$K=\frac{1}{2}\,\frac{\iint|\,q^{-1}(\tau_1)-q^{-1}(\tau_2)\,|\,L(\tau_1^{-1},f)\,L(\tau_2^{-1},f)\,p_0(\tau_1\,|\,J)\,p_0(\tau_2\,|\,J)\,d\mu(\tau_1)\,d\mu(\tau_2)}{\int q(\tau^{-1})\,p_0(\tau/J)\,d\mu(\tau)}.$$

Таким образом

$$\int\limits_{W_n}|\,r_n-\varrho\,|^2\,dP_n\le\frac{r\varrho-K}{1-\varepsilon_n}-2\varrho^2+\varrho\,. \qquad (\text{П. 5})$$

Нетрудно показать, что если r_n сходится к ϱ, то *а*) $|r_n-\varrho|$ ограничена на W_n для всех n, б) $r=\varrho$. В самом деле, $|r_n-\varrho|\le|r-r_n|+|r-\varrho|$ причем для $T\in W_n$, при достаточно больших n, $r_n>r-\varepsilon$. Далее, P_n — мера W_n положительна, а P_n — мера множества $R_n=\{T:|r_n-\varrho|<\varepsilon\}$ при $n\to\infty$ стремится к 1, поэтому пересечение W_n и R_n не пусто. На этом

пересечении $|r-\varrho|<|r_n-\varrho|+|r-r_n|<2\varepsilon$. Но r и ϱ не зависят от T. Значит $|r-\varrho|<2\varepsilon$ всюду, так что $r=\varrho$.

Доказательство необходимости завершается так. Если r_n сходится к ϱ, в смысле (3.6), и выполнено условие a), тогда из леммы вытекает, что интеграл слева в (П. 5) стремится к нулю и $K\leq\varrho(r-\varrho)$. Отсюда, в силу б), следует $K\leq 0$. Но при $q(\theta)\not\equiv\text{const}$ имеем $K>0$. Полученное противоречие завершает доказательство необходимости, поскольку из $q(\theta)=\text{const}$ следует $q(\theta)\equiv 1$.

2) *Достаточность условия* $q(\theta)\equiv 1$.

Пусть $q(\theta)\equiv 1$. Снова по C_n построим последовательности множеств V_n и Θ_n (не совпадающих с прежними W_n и Θ_n), но обеспечим дополнительно (переходя, в случае необходимости, к подпоследовательностям W_n и Θ_n), чтобы

$$1-\varepsilon_n<\frac{\nu(V_n)}{\nu(\Theta_n)}<\frac{1}{1-\varepsilon_n}\,. \tag{П. 6}$$

Как и выше, можно показать, что в (П. 2) предел (при $n\to\infty$) первого интеграла справа не превосходит ϱ^2, предел второго не превосходит ϱ. Можно записать, что

$$\lim_{n\to\infty}\int\limits_{V_n} p_n(T)\,d\mu(T)=1\,. \tag{П. 7}$$

Отсюда следует

$$\lim\int\limits_{V_n}|r_n-\varrho|^2\,dP_n=0\,. \tag{П. 8}$$

Тогда доказательство достаточности завершается использованием следствия из леммы.

Для доказательства (П. 7) заметим, что

$$P_n(V_n)=\frac{1}{\nu(\Theta_n)}\int\limits_{V_n}\Big\{\int\chi_n(\tau)\,p_0(\tau\,|\,J)\,d\mu(\tau)\Big\}\,d\nu(T)\,.$$

А поскольку $T\in V_n$, то $\chi_n(\tau)\geq\chi_{C_n}(\tau)$, так что

$$P_n(V_n)>(1-\varepsilon_n)\frac{\nu(V_n)}{\nu(\Theta_n)}>(1-\varepsilon_n)^2$$

(последнее неравенство — в силу (П. 6)). Это завершает доказательство теоремы.

Заключение

Бейесов метод определения оптимальной, в статистическом смысле, системы наталкивается на трудности, связанные с заданием априорного распределения, если последнее неизвестно или вообще не существует. Выбор какого-либо определённого априорного распределения не может быть однозначным и приводит к неконтролируемому отклонению риска от риска истинного распределения. В этой ситуации предпочтительнее использовать инвариантное правило, отыскание которого не связано с заданием априорного распределения и риск которого, по крайней мере в транзитивном случае, не зависит от оцениваемого параметра. Заметим, что оцениваемая величина может быть не только неизвестным постоянным, но и случайным параметром [5].

К сожалению, в классе традиционно используемых функций потерь и моделей образования наблюдаемого сигнала из полезного сигнала и помехи лишь немногие оказываются инвариантными, а процесс отыскания группы симметрии задачи, даже если она существует, не всегда может быть формализован. Выявление класса сигналов и функций потерь с инвариантными свойствами могло бы оказаться полезным для развития теории адаптивных систем.

Метод отыскания оптимальных инвариантных правил, изложенный в § 2, справедлив для однотранзитивных групп преобразований и конечномерных сигналов. Нетранзитивный и бесконечномерный случаи требуют специального рассмотрения.

Теорема о бейесовом характере инвариантного правила, сформулированная в § 3, аналогична известной теореме Ханта—Стейна [2]. Для её доказательства оказалась достаточной небольшая модификация метода, рассмотренного в [8].

Литература

1. *Пугачев, В. С.:* Теория случайных функций. Физматгиз, 1962.
2. *Леман, Э.:* Проверка статистических гипотез. Изд-во «Наука», М. 1964.
3. *Эйзенхарт, Л. П.:* Непрерывные группы преобразований. Изд-во иностр. лит., М. 1947.
4. *Brillinger, D. R.:* Necessary and sufficient conditions for a statistical problem to be invariant under a Lie group. The Annals of Math. Stat. **34** (1963) 492—500.
5. *Шайкин, М. Е.:* Инвариантное оценивание коэффициентов линейной регрессии. Доклады 2-го Всесоюзного совещания по статистическим методам теории управления, том «Идентификация», Изд-во «Наука», 1970.
6. *Berk, R. H.—Bickel, P. T.:* On invariance and almost invariance. The Annals of Math. Stat. **39** (1968) 1573—1576.
7. *Fraser, D. A. S.:* The fiducial method and invariance. **48** (1961) 261—280.
8. *Stone, M.:* Right Haar measure for convergence in probability to quasi posterior distributions. The Annals of Math. Stat. **36** 440—451 (1965).
9. *Лоэв, М.:* Теория вероятностей. Изд-во ин. лит., М. 1962.

Invariant estimates in the statistical theory of optimal systems

M. E. SHAYKIN

(Moscow)

The theory of statistical optimal systems for signal detection, filtering, prediction, identification and control is based on a general theory of decision functions with the assumption that statistical characteristics of input signals and plants are known and that the goal of the system is to minimize some given criterion.

The modern trends in the statistical control theory are more realistic. Methods of selfadjusting and adaptive filters determination, supervised and nonsupervised learning systems of, identification and pattern recognition require much less information on the properties of input signals and plant characteristics than their classical forerunners did.

In this article an attempt is made to apply the methods of the theory of groups for the determination of optimal systems. These methods can be used only if a problem is invariant under some transformation group G such that (1.2) are fulfilled, where P_θ, $\theta \in \Omega$ is a family of probability measures on the values of input signals Z, L is a loss function and $g \in G$. In this case it is natural to confine to invariant estimates δ of a parameter, θ, for which (1.3) is true for every $g \in G$. A rather large class of invariant vector-valued signals is described by the model (1.4), if transformations g_c: $z' = f(z\,;c)$ are elements of the group G. Necessary and sufficient conditions for P_θ and L to be invariant under the transformations (1.5), are well-known and are given in (1.7), (1.8), where ξ_α^k, η_β^l, ζ_γ^m are infinitesimal transformations of G.

The method to determine optimal invariant estimates is described in Section 2 for vector-valued Z, Q, d and a transitive Lee group G ($d \in D$ — a range of a decision function δ). Let $(Z_1, Z_2, \ldots Z_n) \to (J_1, \ldots, J_k, T_1, \ldots, T_{n-k})$ be a one-to-one correspondence, $J = (J_1, \ldots, J_k)$ and $T = (T_1, T_2, \ldots, T_{n-k})$ are a group invariant of G and sufficient statistics of the family P_θ, $\theta \in \Omega$ respectively. If the range of T is a homogeneous space under induced group of transformations on range T and $T \leftrightarrow g_a(T)$ (where $T = g_{a(T)}T_0$, T_0 is fixed within to range of T) is an isomorphism, then $\delta(z) \equiv \delta(T(z), J(z)) = g_a(T(z))$, $f(J(z))$ for the arbitrary function f of the invariant J. The optimal invariant estimate $\hat{\delta}$ is $\hat{\delta} = g_{a(T)}\hat{f}(J)$, where $\hat{f}$ minimizes $\varrho_\theta(f/J)$ in (2.2) for every given J (the notations of [6] are used). The method to obtain optimal invariant estimates is illustrated with an identification problem for the linear system (2.3) with an unknown noise dispersion.

If the problem is invariant and $\nu(\,\cdot\,)$ is an a priori measure on Ω, then the conditional Bayes risk (under given T, J) (3.1) coincides formally with $\varrho_\theta(f/J)$ in (2.2). This gives us another method for determination of optimal invariant estimates which can be more prefereable. If $\nu(\,\cdot\,)$ fails to satisfy the assumption $\nu(\Omega) < \infty$, then $r(d/T, J)$ in (3.1) cannot be interpreted as the conditional Bayes risk. But $d\nu_n(\,\cdot\,)$ in (3.3) is a probabilistic measure, so we may select a sequence (3.4) of the conditional Bayes risks. Then the following theorem is true: if $\nu(\,\cdot\,)$ is a right-invariant Haar measure on Ω and P_{r_n}, the probability measure with the probability element (3.5), then $q(\theta) \equiv 1$ ($q(\,\cdot\,)$ is an element from the set of relative-invariant measures on Ω [8]) gives a necessary and sufficient condition for the convergence r_n (3.4) to $\varrho(f/J)$ "in probability"),

$$\lim_{n \to \infty} P_n \{T : | r_n - \varrho | > \varepsilon\} = 0 .$$

The proof is based on the lemma and closely related to the proof of the convergence to quasi-posterior distributions in [8].

M. E. Шайкин
Институт проблем управления
СССР Москва В—485
Профсоюзная ул. 81

Problems of Control and Information Theory, Vol. 1 (2), pp. 153—162 (1972)

A POSSIBLE USE OF ADAPTIVE PROGRAMMING

J. KOCSIS

(Budapest)

Generalized adaptive programming algorithms are presented for process optimization of multivariable continuous plants, according to optimization of some economical performance index (minimum costs, maximum production effectiveness etc.). The term "adaptive programming", points out somewhat vaguely the adaptation suggested by Tsypkin on one hand, and the necessity of digital computer for realisation — on the other. The actual algorithmic scheme is a fourfold generalized perception scheme devised by a multivariable stochastic approximation method.

Introduction

Adaptive programming is an algorithm realized on a digital computer for automatic changes of input signals, on the base of continuously accumulated information, in such a way that the optimal economical system state will be achieved, in spite of initial indefinity and variably operating conditions. The general notion of adaptation in automatic control systems was suggested by Tsypkin [1, 2].

The problem

The plant in Fig. 1 is considered. The notations are as follows:

$x(t)$ — performance index

$\mathbf{u}(t)$ — vector of input control variables

$\mathbf{z}(t)$ — vector of measured but, because of the plant technological conditions noncontrolable, input variables

$\mathbf{h}(t)$ — vector of the constrained variables.

The plant operation is described by the following equation:

$$x(t) = A\{\mathbf{v}[\mathbf{u}(t), \mathbf{z}(t)]\} + \xi(t) \tag{1}$$

where A — is some absolute convex operator, connecting the input system variables with the performance index

ξ — noise, when measuring the performance index, with zero mean and finite variance.

Let us assume that the performance index is continuously measured but the exact form of the equation (1) is unknown. Hence the problem consists in finding the optimal control variable vector thus, that the performance index will be aspiring to the extremum value while the changes of noncontrolable variables are unknown beforehand and the constraints are as follows:

$$J_1(\mathbf{u}) = M\{F_1[x(\mathbf{u}, \mathbf{z})]\} \to \min_{\mathbf{u}} \tag{2}$$

(Here and in what follows M stands for the expectation)

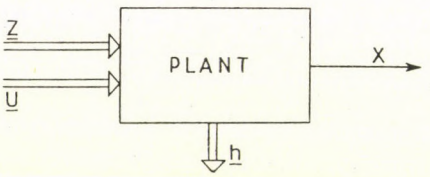


Fig. 1

Simultaneously the next inequality type constraint are introduced

$$M\{h[x(\mathbf{u}, \mathbf{z})]\} \leq 0 \tag{3}$$

The possibility of approximating the static relation between performance index and input variables by a polynomial of second order (by a whole quadratic form) and the exact dynamic connections in the system by simple dynamic elements is assumed, as shown in Fig. 2. In Fig. 2 the transfer function of a simple lag is presented by $W_i(S)$, and the approximate performance index — by $\hat{x}$.

Reality of such an assumption is not discussed here but we refer to the results of simple examples, published previously elsewhere [3, 4], from which

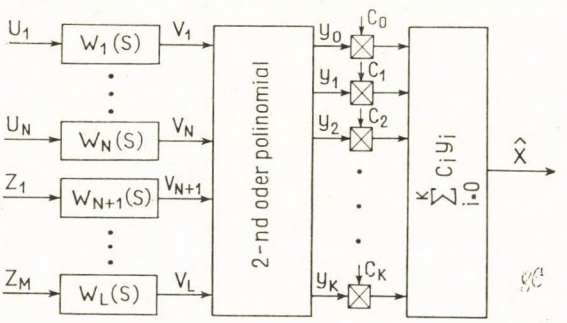


Fig. 2

on the basis of preliminary plant analysis, not only the possibility of the above-mentioned approximation (construction of the plant model) but of the observability and controlability can also be decided.

Solution of the problem

The above formulated has been solved in [5] for a quadratic performance criterion. In the present paper a solution is given for a generalized performance criterion and also the method of preliminary evaluation of unknown parameters of the plant model (identification) are treated. More distinctly, the task is solved in two stages. For studying statical and dynamical characteristic only the identification problem, that is evaluation of unknown parameters of the plant model is considered at first, without interference to the normal plant operation, and then the control variable vector $\mathbf{u}(t)$ will be defined on the base of identification results and simultaneously the statical plant model coefficients will be precised further.

1. Determination of the unknown parameters of the plant model

Identification algorithms of the plant, described by equation (1), are presented in [6, 7]. Here the generalization of such an identification procedure is given. The aim of the identification process is to determine minimum of the following performance criterion:

$$J_2(\mathbf{c}, \mathbf{d}) = M\{F_2[x - \hat{x}(\mathbf{c}, \mathbf{d})]\} \to \min \tag{4}$$

where $\hat{x}$ — approximate performance index

$\mathbf{c}$ — vector of unknown statical model parameters, coefficients of second order polynomial

$\mathbf{d}$ — vector of unknown dynamical model parameters, transfer function coefficients (see Fig. 2).

At the minima of the functional (4) the gradient with respect to parameters equals 0:

$$\frac{\partial J_2(\mathbf{c}, \mathbf{d})}{\partial \mathbf{c}} = -M\left\{\frac{dF_2[x - \hat{x}(\mathbf{c}, \mathbf{d})]}{d\hat{x}} \cdot \frac{\partial \hat{x}(\mathbf{c}, \mathbf{d})}{\partial \mathbf{c}}\right\} = 0 \tag{5}$$

$$\frac{\partial J_2(\mathbf{c}, \mathbf{d})}{\partial \mathbf{d}} = -M\left\{\frac{dF_2[x - \hat{x}(\mathbf{c}, \mathbf{d})]}{d\hat{x}} \cdot \frac{\partial \hat{x}(\mathbf{c}, \mathbf{d})}{\partial \mathbf{d}}\right\} = 0 \tag{6}$$

For determining the root of regression equations (5, 6) the well known stochastic approximation method is used:

$$\mathbf{c}[n] = \mathbf{c}[n-1] + \mathbf{R}_2[n] \frac{dF_2\{x[n] - \hat{x}(\mathbf{c}[n-1], \mathbf{d}[n-1])\}}{d\hat{x}[n]} \times$$

$$\times \frac{\partial \hat{x}(\mathbf{c}[n-1], \mathbf{d}[n-1])}{\partial \mathbf{c}[n-1]} \tag{7}$$

$$\mathbf{d}[n] = \mathbf{d}[n-1] + \mathbf{R}_3[n] \frac{dF_2\{x \times [n] - \hat{x}(\mathbf{c}[n-1], \mathbf{d}[n-1])\}}{d\hat{x}[n]} \times$$

$$\times \frac{\partial \hat{x}(\mathbf{c}[n-1], \mathbf{d}[n-1])}{\partial \mathbf{d}[n-1]} \tag{8}$$

During the identification process, algorithms (7, 8) determine the unknown dynamical (**d**) and statical (**c**) model parameters adopting some convex function (F_2), and minimizing the difference of the plant and model outputs.

The convergence criteria of learning algorithms in [1] have to be fulfilled.

Generally $\mathbf{R}_2$, $\mathbf{R}_3$ are diagonal matrices and the possible optimal and quasi-optimal convergence matrices **R** are given in [1].

2. Adaptive programming algorithms

Approximate dynamical and model characteristics determined by algorithm (8) are not changed when the optimal control vector $\mathbf{u}^*(t)$ has to be found. The aim of the control process consists in minimizing functional (2) under constraints (3). The gradient equals 0 at the minimum point of equation (2):

$$\frac{dJ_1(\mathbf{u})}{d\mathbf{u}} M \left\{ \frac{\partial F_1[x(\mathbf{u}, \mathbf{z})]}{\partial x} \cdot \frac{dx}{d\mathbf{u}} \right\} = 0 \tag{9}$$

In the fortunate cases, when functional J_1 can be determined analytically (deterministic case or the preliminary known distributions), the task may be reduced to the simple search for an extremum point. But in the general case — by the very features of the problem — when input and output plant variables are stationary stochastic signals with apriori unknown distributions, we have to apply the adaptive approach. An approximative performance index function $\hat{x}$ has to be given in a differentiable form for determining gradient $dx/d\mathbf{u}$. On the base of the plant model shown in Fig. 2 the approximate performance index function has the following form:

$$\hat{x} = \mathbf{c}^T \mathbf{y}[\mathbf{v}(\mathbf{u}, \mathbf{z})] \tag{10}$$

and if the model — by the preliminary plant identification results — simulates the real plant processes with the sufficient precision, then

$$\frac{dx}{d\mathbf{u}} \simeq \frac{d\hat{x}}{d\mathbf{u}} \tag{11}$$

can be substituted in the regression equation (9), that is, it is possible to take the approximate gradient instead of the real one.

By the help of Lagrangean multiplicator techniques constraint (3) can easily met. In such a case let us take a new functional instead of (2):

$$J(\mathbf{u}, \boldsymbol{\lambda}) = M\{F_1[x(\mathbf{u}, \mathbf{z})] + \boldsymbol{\lambda}^T \mathbf{h}[x(\mathbf{u}, \mathbf{z})]\} \to \min \tag{12}$$

from which the necessary condition of extremum (applicating incidentally the change by (11) is):

$$\frac{\partial J(\mathbf{u}, \boldsymbol{\lambda})}{\partial \mathbf{u}} = M\left\{\frac{\partial F_1[x(\mathbf{u}, \mathbf{z})]}{\partial x} \cdot \frac{d\hat{x}}{d\mathbf{u}} + \left[\frac{d\mathbf{h}}{d\mathbf{u}}\right]^T \boldsymbol{\lambda}\right\} = 0 \tag{13}$$

and furthermore

$$\frac{\partial J(\mathbf{u}, \boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = M\{\mathbf{h}[x(\mathbf{u}, \mathbf{z})]\} \leqq 0 \tag{14}$$

and the iteration rules according to stochaistic approximation method to regression equations (13, 14), have the following form:

$$\mathbf{u}[n] = \mathbf{u}[n-1] - \mathbf{R}_1[n]\left(\frac{\partial F_1\{x(\mathbf{u}[n-1], \mathbf{z}[n])\}}{\partial x[n]} \cdot \frac{d\hat{x}[n]}{d\mathbf{u}[n-1]} + \right.$$
$$\left. + \left[\frac{d\mathbf{h}[n]}{d\mathbf{u}[n-1]}\right]^T \boldsymbol{\lambda}[n-1]\right) \tag{15}$$

and

$$\boldsymbol{\lambda}[n] = \max\left(0; \boldsymbol{\lambda}[n-1] - \mathbf{R}[n]\mathbf{h}\{x(\mathbf{u}[n-1], \mathbf{z}[n]\}\right) \tag{16}$$

Finally, if the model parameters could not be changed at the time of adaptive control, then the optimal control algorithms are presented by Eqs. (15, 16). But when the control signal is changed then the model parameters (**c**) are changed too. Therefore one has to correct the plant model by algorithm (7), the form of which shows some changes because of the connection $\mathbf{c} = \mathbf{c}(\mathbf{u})$:

$$\mathbf{c}[n] = \mathbf{c}[n-1] + \mathbf{R}_2[n]\frac{dF_2}{dx}[n]\left(\frac{\partial \hat{x}[n]}{\partial \mathbf{c}[-1]} + \left[\frac{\partial \mathbf{u}[n-1]}{\partial \mathbf{c}[n-1]}\right]^T \frac{\partial \hat{}[n]}{\partial \mathbf{u}[n-1]}\right) \tag{17}$$

where the sensitivity matrix

$$\mathbf{S}[n-1] = \frac{\partial \mathbf{u}[n-1]}{\partial \mathbf{c}[n-1]} \tag{18}$$

can be given without any special difficulties by forming the gradient of algorithm (15) with respect to the vector of unknown plant model parameters c:

$$\mathbf{S}[n-1] = \mathbf{S}[n-1] - \mathbf{R}_1[n] \frac{dF_1}{dx}[n] \frac{\partial}{\partial \mathbf{c}[n-1]} \left(\frac{\partial \hat{x}[n]}{\partial \mathbf{u}[n-1]} \right) \tag{19}$$

Adaptive programming algorithms are presented in eqs. (15—16), (17—19). For the sake of the flow diagrams concerning the adaptive optimization realization of the continuous industrial plant are presented in Fig. 3.

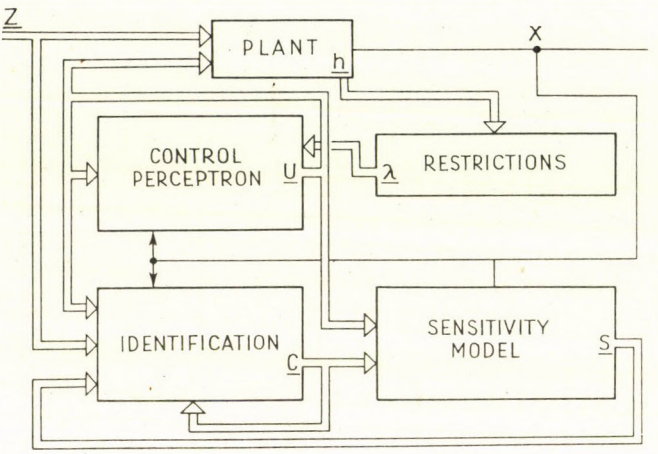


Fig. 3

A particular case

As it was presented in [5], when the quadratic performance criterion is given as:

$$F_1(\cdot) = (\cdot)^2$$

$$F_2(\cdot) = (\cdot)^2$$

and the different gradients in adaptive programming algorithms (15—16), (18—19) are presented in the following form:

$$\frac{d\hat{x}[n]}{d\mathbf{u}[n-1]} = \mathbf{D}(\mathbf{W}[1])\mathbf{C}[n-1]\mathbf{v}[n]$$

$$\frac{d\hat{x}[n]}{d\mathbf{c}[n-1]} = \mathbf{y}[n] + \mathbf{S}[n-1]^T \mathbf{D}(\mathbf{W}[1])\mathbf{C}[n-1]\mathbf{v}[n]$$

$$\frac{d}{d\mathbf{c}[n-1]}\left(\frac{d\hat{x}[n]}{du[n-1]}\right)=\begin{bmatrix}\mathbf{C}_0^*[n-1]\mathbf{D}(\mathbf{W}[1])\mathbf{s}_0[n-1]\\ \mathbf{C}_1^*[n-1]\mathbf{D}(\mathbf{W}[1])\mathbf{s}_1[n-1]\\ \cdot\\ \cdot\\ \cdot\\ \mathbf{C}_k^*[n-1]\mathbf{D}(\mathbf{W}[1])\mathbf{s}_k[n-1]\end{bmatrix}$$

then

$$\frac{\partial F_1}{\partial x}=\frac{\partial x^2}{\partial x}=2x$$

$$\frac{\partial F_2}{\partial \hat{x}}=\frac{\partial (x-\hat{x})^2}{\partial \hat{x}}=-2(x-\hat{x})$$

where

$$\mathbf{u}=\begin{bmatrix}u_1\\ u_2\\ \cdot\\ \cdot\\ \cdot\\ u_N\end{bmatrix};\quad \mathbf{z}=\begin{bmatrix}z_1\\ z_2\\ \cdot\\ \cdot\\ \cdot\\ z_M\end{bmatrix};\quad \mathbf{v}=\begin{bmatrix}1\\ v_1\\ \cdot\\ \cdot\\ \cdot\\ v_L\end{bmatrix};\quad \mathbf{c}=\begin{bmatrix}c_0\\ c_1\\ \cdot\\ \cdot\\ \cdot\\ c_K\end{bmatrix};\quad \mathbf{y}=\begin{bmatrix}1\\ y_1\\ \cdot\\ \cdot\\ \cdot\\ y_K\end{bmatrix}$$

$$L=M+N;\qquad K=\frac{(L+1)(L+2)}{2}-1$$

$$\mathbf{D}(\mathbf{w}[1])=\begin{bmatrix}w_1[1] & 0 & \cdot & \cdot & \cdot & 0\\ 0 & w_2[1] & \cdot & \cdot & \cdot & 0\\ \cdot & \cdot & \cdot & & & \cdot\\ \cdot & \cdot & & \cdot & & \cdot\\ \cdot & \cdot & & & \cdot & \cdot\\ 0 & 0 & \cdot & \cdot & \cdot & w_N[1]\end{bmatrix}\qquad \mathbf{C}=\begin{bmatrix}c_1 & 2c_{L+1} & \cdot & \cdot & \cdot & c_{2L}\\ c_2 & c_{L+1} & \cdot & \cdot & \cdot & c_{3L-1}\\ \cdot & \cdot & \cdot & & & \cdot\\ \cdot & \cdot & & \cdot & & \cdot\\ \cdot & \cdot & & & \cdot & \cdot\\ c_N & c_{L+N} & & & & c_K\end{bmatrix}$$

$$\mathbf{S}=\begin{bmatrix}s_{10} & s_{11} & \cdot & \cdot & \cdot & s_{1K}\\ s_{20} & s_{21} & \cdot & \cdot & \cdot & s_{2K}\\ \cdot & \cdot & \cdot & & & \cdot\\ \cdot & \cdot & \cdot & \cdot & & \cdot\\ \cdot & \cdot & & & \cdot & \cdot\\ s_{NO} & s_{N1} & \cdot & \cdot & \cdot & s_{NK}\end{bmatrix}\qquad \left(s_{ij}=\frac{du_i}{dc_j}\right)$$

$$(j=0,1,2,\ldots,K)$$

$$s_j^T = [1 \; s_{1j} \; \ldots \; s_{Nj}]$$

$$\mathbf{C}_j^* = \left[\mathbf{a}_j \; \vdots \; \mathbf{C}^*\right]$$

$$\mathbf{C}^* = \begin{bmatrix} 2c_{L+1} & c_{L+2} & \cdot \cdot \cdot & c_{L+N} \\ c_{L+2} & 2c_{2L+1} & \cdot \cdot \cdot & c_{2L+N-1} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ c_{L+N} & c_{2L} & \cdot \cdot \cdot & 2c_{\frac{N(2L-N+3)}{2}} \end{bmatrix}$$

$$\mathbf{A} = ||a_{ij}|| =$$

$$= \begin{bmatrix} 0\,1\,0\,0 \ldots 0\,0 \ldots 0 & 2v_1\, v_2 \ldots v_N & v_{N+1} \ldots v_L & 0\;0 \ldots 0 & 0 \ldots 0 & 0 \ldots 0 & 0 \ldots 0 & 0 \ldots 0 \\ 0\,0\,1\,0 \ldots 0\,0 \ldots 0 & 0 & v_1 \ldots 0 & 0 \ldots 0 & 2v_2\, v_3 \ldots v_N & v_{N+1} \ldots v_L & 0 \ldots 0 & 0 \ldots 0 \\ 0\,0\,0\,1 \ldots 0\,0 \ldots 0 & 0\;0 & 0 \ldots 0 & 0 \ldots 0 & 0 & v_2 \ldots 0 & 0 \ldots 0 & 2v_3 \ldots 0 \\ \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot & \cdot \cdot \cdot \\ \underbrace{0\,0\,0\,0 \ldots 1}_{N} & \underbrace{0 \ldots 0}_{M} & \underbrace{0\;0 \ldots v_1}_{N} & \underbrace{0 \ldots 0}_{M} & \underbrace{0\;0 \ldots v_2}_{N-1} & \underbrace{0 \ldots 0}_{M} & 0 \ldots 2v_N\, v_{N+1} \ldots v_L & 0 \ldots 0 \end{bmatrix}$$

In this case dimensions of the adaptive algorithms may be checked in the following way: if N-dimensional vector $\mathbf{u}$, M-dimensional vector $\mathbf{z}$ and P-dimensional vector $\mathbf{h}$ are measured, the number of recursive equations will be:

control equations: N
identification equations: $K + 1$
(where: $K + 1 = (L + 1)(L + 2)/2$; $L = M + N$)
restriction equations: P
sensitivity coefficient equations: $N \times K$
Viz.,: $P + N + 1 + (M + N + 1)(M + N + 2)/2$ equations are necessary and sufficient.

For example, in a complex plant, for which the processes are characterized by the following number of measurable variables:

$$P = 10 \qquad N = 10, \qquad M = 20$$

the total number of equations to be solved continuously is 5466. But if the model is not changed after the identification process, then this number is $P + N = 20$ (at the preliminary identification process $(M+N+1)(M+N+1)/2 = 496$ recursive equations have to be solved continuously.

References

1. *Цыпкин, Я. З.:* Основы теории обучающихся систем. Изд-во «Наука», Москва, 1970.
2. *Цыпкин, Я. З.:* Адаптация и обучение в автоматических системах. Изд-во «Наука», Москва, 1968.
3. *Кочиш, Я.:* Экспериментальное исследование устойчивости адаптивной экстремальной системы. Periodica Polytechnica El. **14** (1970).
4. *Кочиш, Я.:* Определение оптимальных параметров адаптивного управления. Periodica Polytechnica El. **14** (1970).
5. *Kocsis, J.:* Process optimization by adaptive programming. IFAC Symposium on Identification and Process Parameter Estimation, Prague, 1970. Periodica Polytechnica El. **14** (1970).
6. *Кочиш, Я.—Алмаши, Г.:* К вопросу идентификации промышленных объектов. Periodica Polytechica El. **14** (1970)
7. *Kocsis, J.:* Tanuló eljárás dinamikus rendszer vizsgálati identifikációjára. VI. Magyar Aut. Konf., A/2 szekció, Budapest, 1970.

Возможное применение адаптивного программирования

Я. КОЧИШ

(Будапешт)

В настоящей работе представлены обобщенные алгоритмы адаптивного программирования для решения задачи оптимизации непрерывных многомерных промышленных процессов по некоторому экономическому показателю, по экономической функции цели (минимальная себестоимость, максимальная эффективность производства, максимальная эксплуатация производственных емкостей или др.). Произвольно избранный термин «адаптивное программирование» указывает, с одной стороны, на применение адаптивного подхода, изложенного подробно Цыпкиным [1, 2], а с другой — на необходимость присутствия цифровой вычислительной машины при реализации подобного типа системы. Таким образом, под названием «адаптивное программирование» подразумевается реализуемый с помощью ЦВМ алгоритм автоматического определения процесса изменения входных параметров системы на основе обработки текущей информации с целью достижения с экономической точки зрения оптимального состояния системы при начальной неопределенности и изменяющихся условиях работы.

Данный принцип постановки задачи представляет собой возможную область применения быстродействующих управляющих ЦВМ на промышленных объектах с современной непрерывной технологией. Для такой цели преимущественно применяются управляющие ЦВМ третьего поколения (типа французской CII 10010, американской — ИБМ 1800).

Рассматривается объект, представленный на рис. 1, описываемый уравнением (1). Задача состоит в определении оптимальных управляющих параметров таким образом, чтобы функция цели при этом стремилась к минимальному значению при любых изменениях неуправляемых входных параметров системы. При определении вектора управляющих параметров следует соблюдать ограничения, наложенные на некоторые параметры объекта [3]. Предполагается, что статическое соотношение между функцией цели и входными параметрами системы можно приближенно описать полиномом второго порядка, а динамическую связь — простыми динамическими звеньями, как это представлено на рис. 2.

Вышеуказанная задача была решена в [5] на случай применения квадратичных критериев качества. В настоящей работе приводится решение задачи на обобщенные критерии качества, а также представлен метод предварительного определения неизвестных параметров модели объекта (идентификация). Таким образом, задача решается в два этапа. С целью изучения статического и динамического поведения объекта сначала проводится только его идентификация, т. е. определение неизвестных параметров модели объекта, а затем на основе результатов идентификации определяется вектор управляющих параметров и одновременно уточняются статические коэффициенты модели объекта. Алгоритмы идентификации представлены в уравнениях (7—8).

Определенные по алгоритму (8) приближенные динамические свойства модели зафиксируются в ходе поиска оптимального вектора управляющих параметров. Целью управления является нахождение минимума функционала (2) при соблюдении ограничений типа (3). В точке минимума уравнения (2) градиент (9) должен равняться нулю. Если по результатам предварительной идентификации объекта модель с достаточной точностью имитирует поведение действительных процессов в объекте, то в уравнение регрессии (9) можно записать вместо действительного — приближенный градиент (11). Решение уравнений регрессии (13, 14) представлено в алгоритмах (15, 16), которые представляют собой две первые рекуррентные формулы адаптивного программирования для определения вектора управляющих параметров и вектора множителей Лагранжа. Если параметры модели не меняются в ходе адаптивного управления, то алгоритмы (15, 16) представляют собой алгоритм определения оптимального управления. Но в то же время, если меняется управляющий сигнал, то меняются и параметры модели также. Поэтому необходимо проводить коррекцию модели объекта по уравнению (17), где определение матрицы чувствительности не представляет особую трудность, если образовать градиент алгоритма (15) по вектору неизвестных параметров модели объекта.

Алгоритмы адаптивного программирования представлены в уравнениях (15, 16, 17, 19). Схема прохода информации при реализации адаптивной оптимизации непрерывного промышленного объекта представлена на рис. 3.

Далее представлен пример для применения адаптивного программирования.

János Kocsis
Department of Automation
Technical University
Budapest 11, Garami E. t. 3, Hungary

Problems of Control and Information Theory, Vol. 1 (2), pp. 163—176 (1972)

САМООБУЧАЮЩИЙСЯ АЛГОРИТМ АСИМПТОТИЧЕСКИ ОПТИМАЛЬНОЙ ФИЛЬТРАЦИИ СЛУЧАЙНЫХ СИГНАЛОВ С НЕИЗВЕСТНЫМ АПРИОРНЫМ РАСПРЕДЕЛЕНИЕМ

А. В. ДОБРОВИДОВ

(Москва)

Предлагается алгоритм непараметрической оценки полиномиальных функций случайных сигналов в условиях, когда априорное распределение случайных сигналов неизвестно и условная (при фиксированных значениях случайных сигналов) плотность наблюдаемых случайных величин принадлежит экспонентному семейству плотностей. Доказана сходимость непараметрических оценок к оптимальной оценке и сходимость соответствующих средних рисков к оптимальному бейесову риску.

Классические методы построения оптимальных автоматических систем обработки информации [1, 2] требуют задания всех статистических характеристик сигналов и шумов, действующих на систему. Однако на практике часто оказывается, что статистические характеристики сигналов либо неизвестны, либо требуют для своего определения слишком больших затрат. В то же время при достаточно общих условиях можно построить самообучающиеся автоматические системы, алгоритмы которых приближаются к оптимальным алгоритмам, полученным при полностью известных статистических характеристиках, и при этом используется информация, поступающая только в процессе нормальной работы этих систем. Развитию методов построения таких автоматических систем посвящена настоящая работа.

Рассмотрим один важный класс автоматических систем обработки информации — системы оптимальной фильтрации сигналов. В приложениях часто возникает необходимость оценивать не сами полезные сигналы, а некоторые функции от них. Если такие функции являются нелинейными, то необходимо искать оптимальную оценку сразу всей функции полезных сигналов.

§ 1. Постановка задачи

Пусть ставится задача оценить некоторую в общем случае векторную функцию $\vec{\varphi}$ многомерного случайного сигнала $\vec{\Theta} = (\Theta_1, \ldots, \Theta_m)$[1] с априорным распределением $G(\vec{\theta})$, определённым на пространстве Ω, по наблюде-

[1] Везде ниже прописные буквы $\Theta, \mathbf{X}$ означают случайные величины, а строчные θ, $\mathbf{x}$ — реализации случайных величин.

ниям векторной случайной величины $\vec{\mathbf{X}} = (X_1, \ldots, X_r)$, условная плотность распределения $f(\vec{\mathbf{x}}/\vec{\theta})$ которой принадлежит общему экспонентному семейству плотностей $\{f(\vec{\mathbf{x}}/\vec{\theta}),\ \theta \in \Omega\}$,

$$f(\vec{\mathbf{x}}/\vec{\theta}) = C(\vec{\theta})\, h(\vec{\mathbf{x}}) \exp\Big\{ \sum_{j=1}^{m} \theta_j T_j(\vec{\mathbf{x}}) \Big\}, \tag{1}$$

где $T_j(\vec{\mathbf{x}})$ и $h(\vec{\mathbf{x}})$ — измеримые функции, определенные на выборочном пространстве x, а $C(\vec{\theta})$ — нормирующий множитель.

Оптимальная оценка $\vec{\delta}_G(\vec{\mathbf{x}})$ векторной функции $\vec{\varphi}(\vec{\theta})$, минимизирующая средний риск

$$R(\vec{\delta}, G) = \int_x \int_\Omega \left(\vec{\varphi}^T(\vec{\theta}) - \vec{\delta}^T(\vec{\mathbf{x}})\right)\left(\vec{\varphi}(\theta) - \vec{\delta}(\vec{\mathbf{x}})\right) f(\vec{\mathbf{x}}/\vec{\theta})\, dG(\vec{\theta})\, d\vec{\mathbf{x}}^2, \tag{2}$$

представляет собой апостериорное среднее

$$\vec{\delta}_G(\vec{\mathbf{x}}) = \frac{\int_\Omega \vec{\varphi}(\vec{\theta})\, f(\vec{\mathbf{x}}/\vec{\theta})\, G(\vec{\theta})}{f(\vec{\mathbf{x}})}, \tag{3}$$

где

$$f(\vec{\mathbf{x}}) = \int_\Omega f(\vec{\mathbf{x}}/\vec{\theta})\, dG(\vec{\theta}) \tag{4}$$

— безусловная плотность вероятности случайной величины $\vec{\mathbf{X}}$.

Оптимальную оценку (3) можно вычислить лишь тогда, когда известно априорное распределение $G(\vec{\theta})$ случайного сигнала $\vec{\mathbf{\Theta}}$. Однако в приложениях в подавляющем большинстве случаев неизвестно даже возможное параметрическое семейство априорных распределений, не говоря уже о точном знании $G(\vec{\theta})$. Поэтому в настоящей работе предполагается, что $G(\vec{\theta})$ полностью неизвестно. В этом случае точных оптимальных оценок $\vec{\delta}_G(\vec{\mathbf{x}})$ найти нельзя и приходится строить приближенные оценки $\vec{\delta}_G(\vec{\mathbf{x}})$, вычисленные на основе реализаций $\vec{x}_1, \ldots, \vec{x}_n$ наблюдаемой векторной случайной величины $\vec{\mathbf{X}}$.

Если такие сходятся к оптимальным оценкам (3) с полной статистической информацией, то они называются эмпирическими бейесовыми оценками [3].

§ 2. Непараметрическая оценка попарных произведений компонент полезного сигнала

Рассмотрим класс функций $\vec{\varphi}$, представляющий собой всевозможные попарные произведения $\Theta_j \Theta_k$ и квадраты Θ_j^2 компонент полезного случайного сигнала $\vec{\mathbf{\Theta}}$, $j = 1, \ldots, m$, $j \leq k \leq m$. В этом случае оптимальная оценка

[2] „T" означает транспонирование.

(3) будет иметь вид:

$$\delta_G^{(jk)}(\vec{\mathbf{x}}) = \frac{1}{f(\vec{\mathbf{x}})} \int_{\Omega} \theta_j \theta_k f(\vec{\mathbf{x}}/\vec{\theta})\, dG(\vec{\theta}), \qquad \begin{matrix} j = 1, \ldots, m \\ j \le k \le m \end{matrix}. \tag{5}$$

Для вычисления приближенной оценки, сходящейся к (5), воспользуемся условием (1) и выражением для оптимальной оценки $\vec{\delta}_G(\vec{\mathbf{x}})$ многомерного полезного сигнала $\vec{\Theta}$, полученным в [4]:

$$\vec{\delta}_G(\vec{\mathbf{x}}) = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \vec{\mathbf{b}}, \tag{6}$$

где $\vec{\delta}_G$ — вектор с компонентами $(\delta_G^{(1)}, \ldots, \delta_G^{(m)})$, $\mathbf{T}$ — прямоугольная $(r \times m)$ — матрица ранга m с элементами

$$\tau_{ij} = \frac{\partial T_j(\vec{\mathbf{x}})}{\partial x_i}, \qquad \begin{matrix} i = 1, \ldots, r \\ j = 1, \ldots, m \end{matrix}, \tag{7}$$

$\vec{\mathbf{b}} = (b_1, \ldots, b_r)$ — вектор с компонентами

$$b_k = \frac{\partial}{\partial x_k} \ln \left(\frac{f(\vec{\mathbf{x}})}{h(\vec{\mathbf{x}})} \right), \qquad k = 1, \ldots, r. \tag{8}$$

Так как, с другой стороны, $\vec{\delta}_G(\vec{\mathbf{x}})$ удовлетворяет соотношению (3), с $\vec{\varphi}(\vec{\theta}) = \vec{\theta}$, то можно записать

$$\int_{\Omega} \vec{\theta} C(\vec{\theta}) \exp\{\vec{\theta}^T \mathbf{T}(\vec{\mathbf{x}})\}\, dG(\vec{\theta}) = \frac{f(\vec{\mathbf{x}})}{h(\vec{\mathbf{x}})} (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \vec{\mathbf{b}}. \tag{9}$$

Продифференцируем это векторное равенство по каждому компоненту x_i вектора $\vec{\mathbf{x}}$. Получим r векторных уравнений

$$\int_{\Omega} \vec{\theta}\, \vec{\theta}^T f(\vec{\mathbf{x}}/\vec{\theta})\, dG(\vec{\theta}) \cdot \frac{\partial \vec{\mathbf{T}}(\vec{\mathbf{x}})}{\partial x_i} = h(\vec{\mathbf{x}}) \frac{\partial}{\partial x_i} \frac{f(\vec{\mathbf{x}})}{h(\vec{\mathbf{x}})} (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \vec{\mathbf{b}}, \quad i = 1, \ldots, r, \tag{10}$$

которые удобно записать в матричной форме

$$\mathbf{A} \cdot \mathbf{T}^T = \mathbf{H}, \tag{11}$$

где $\mathbf{A} = \| a_{jk} \|$ — матрица размера $(m \times m)$ с элементами

$$a_{jk} = \int_{\Omega} \theta_j \theta_k f(\vec{\mathbf{x}}/\theta)\, dG(\theta), \quad k, j = 1, \ldots m, \tag{12}$$

а $\mathbf{H} = \| \eta_{ji} \|$ — матрица размера $(m \times r)$ с элементами

$$\eta_{ji} = h(\vec{\mathbf{x}}) \frac{\partial}{\partial x_i} \left(\frac{f(\vec{\mathbf{x}})}{h(\vec{\mathbf{x}})} \sum_{k=1}^{r} t_{jk} b_k \right), \qquad \begin{matrix} j = 1, \ldots, m, \\ i = 1, \ldots, r, \end{matrix} \tag{13}$$

где t_{jk} — элементы матрицы $(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T$. Так как по определению (12) $\mathbf{A}^T=\mathbf{A}$, то уравнение (11) можно переписать в виде:

$$\mathbf{T}\cdot\mathbf{A}=\mathbf{H}^T. \tag{14}$$

Решение его

$$\mathbf{A}=(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{H}^T \tag{15}$$

выражается через ту же матрицу $(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T$, что и оптимальная оценка (6).

Из сравнения (12) и (5) нетрудно видеть, что

$$a_{jk}=f(\vec{\mathbf{x}})\,\delta_G^{(jk)}, \tag{16}$$

откуда с учетом (15) получаем выражение для оптимальной оценки произведения $\Theta_j\Theta_k$, $j=1,\ldots,m$, $j\le k\le m$,

$$\delta_G^{(jk)}(\vec{\mathbf{x}})=\frac{1}{f(\vec{\mathbf{x}})}\sum_{i=1}^{r}t_{jk}\eta_{ik}, \qquad \begin{matrix} j=1,\ldots,m \\ j\le k\le m \end{matrix}, \tag{17}$$

которое не зависит явно от априорного распределения $G(\vec{\theta})$.

Для того, чтобы вычислить оценку (17), необходимо, во-первых, определить элементы t_{ji} матрицы $(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T$, а во-вторых, оценить элементы η_{ik}, $i=1,\ldots,r$, $k=1,\ldots,m$, матрицы $\mathbf{H}^T$. Матрица $\mathbf{T}$ по условию считается заданной и нахождение элементов t_{ji} производится так же, как и в задаче оценки одного полезного сигнала $\vec{\Theta}$ [4], несмотря на то, что здесь оценивается нелинейная функция от компонент $\vec{\Theta}$; в этой части задача не усложняется. Что же касается элементов матрицы $\mathbf{H}^T$, то, как нетрудно видеть после подстановки (8) в (13), η_{ik} будут зависеть как от самой безусловной плотности $f(\vec{\mathbf{x}})$, так и от ее первых и вторых частных производных. Подставляя теперь (13) в (17), получим следующее выражение для оптимальной оценки:

$$\delta_G^{(jk)}(\vec{\mathbf{x}})=\sum_{i,s=1}^{r}\left[t_{ji}t_{ks}\cdot\left(\frac{f''_{si}}{f}-\frac{f'_ih'_s}{fh}-\frac{f'_sh'_i}{fh}+2\frac{h'_sh'_i}{h^2}-\frac{h''_{si}}{h}\right)+\right.$$
$$\left.+\,t_{ji}t'_{ksi}\left(\frac{f'_s}{f}-\frac{h'_s}{h}\right)\right], \qquad \begin{matrix} j=1,\ldots,m \\ j\le k\le m \end{matrix}, \tag{18}$$

где для краткости введены следующие обозначения:

$$f=f(\vec{\mathbf{x}}),\ f''_{si}=\frac{\partial^2}{\partial x_i\,\partial x_s}f(\vec{\mathbf{x}}),\ f'_i=\frac{\partial}{\partial x_i}f(\vec{\mathbf{x}}),\ h'_i=\frac{\partial}{\partial x_i}h(\vec{\mathbf{x}}).$$
$$h''_{si}=\frac{\partial^2}{\partial x_i\,\partial x_s}h(\vec{\mathbf{x}}),\ t'_{ksi}=\frac{\partial}{\partial x_i}t_{ks}(\vec{\mathbf{x}}). \tag{19}$$

В соответствии с условием $G(\vec{\theta})$ может быть произвольной функцией распределения, поэтому безусловная плотность $f(\vec{\mathbf{x}})$ с учетом (1) может

быть также достаточно произвольной. Во всяком случае класс функций $f(\vec{\mathbf{x}})$, описываемый (4), намного шире любого конечно-параметрического семейства плотностей. Поэтому для $f(\vec{\mathbf{x}})$ и ее производных естественно строить непараметрические оценки. Такие оценки для самой плотности и ее первых частных производных были построены в [4]. При некоторых условиях регулярности, накладываемых на $f(\mathbf{x})$ и ее первые производные, там было показано, что непараметрические оценки сходятся в среднеквадратическом к $f(\vec{\mathbf{x}})$ и соответственно к $\partial f(\vec{\mathbf{x}})/\partial x_i$, $i=1,\ldots,r$, в каждой точке $\mathbf{x}$. При более сильных предположениях была доказана сходимость оценок в среднеквадратическом равномерно по $\vec{\mathbf{x}}$.

В настоящей работе, кроме перечисленных оценок, потребовалось также строить оценки вторых частных производных $f(\vec{\mathbf{x}})$. Как показано в приложении, непараметрическая оценка вторых частных производных неизвестной плотности вероятности $f(\vec{\mathbf{x}})$ имеет вид:

$$q_n^{(\mu\nu)}(\vec{\mathbf{x}}) = \frac{1}{n}\sum_{k=1}^{n}\frac{2}{h^{2+r}}\left(\frac{(x_\mu - X_{k,\mu})(x_\nu - X_{k.\nu})}{h_n^2} - d_{\mu\nu}\right)K\left(\frac{\vec{\mathbf{x}}-\vec{\mathbf{X}}_k}{h_n}\right), \quad (20)$$

где $d_{\mu\nu} = \begin{cases} 1 \text{ при } \mu = \nu \\ 0 \text{ при } \mu \neq \nu. \end{cases}$

Эта оценка сходится к $\dfrac{\partial^2}{\partial x_\nu\,\partial x_\mu} f(\mathbf{x})$ в среднеквадратическом в каждой точке $\vec{\mathbf{x}}$, если вторые частные производные $f(\vec{\mathbf{x}})$ непрерывны и $\int\limits_x |f''_{sk}(\vec{\mathbf{x}})|d\vec{\mathbf{x}} < \infty$, $s, k = 1, \ldots, r$. При более жестких условиях на $f(\vec{\mathbf{x}})$, требующих равномерную непрерывность вторых частных производных $f(\vec{\mathbf{x}})$ и их ограниченность, оценка (2) сходится к $\dfrac{\partial^2}{\partial x_\nu\,\partial x_\mu} f(\vec{\mathbf{x}})$ в среднеквадратическом равномерно по $\vec{\mathbf{x}}$. Для наших целей из всех этих результатов потребуется только самый слабый, т. е. сходимость по вероятности в каждой точке $\vec{\mathbf{x}}$, которая следует из сходимости в среднеквадратическом. При этом на $f(\vec{\mathbf{x}})$ и ее первые и вторые частные производные будут наложены только условия непрерывности и абсолютной интегрируемости.

Подставляя теперь в оптимальную оценку $\delta_G^{(jk)}(\vec{\mathbf{x}})$ вместо $f(\vec{\mathbf{x}})$ и ее первых и вторых частных производных соответственно непараметрические оценки $f_n(\vec{\mathbf{x}})$, $g_n^{(\mu)}(\vec{\mathbf{x}})$ и $g_n^{(\mu\nu)}(\vec{\mathbf{x}})$, построенные по реализациям $\vec{x}_1, \ldots, \vec{x}_n$ векторной случайной величины $\vec{\mathbf{X}}$, получим эмпирическую оценку произведения $\Theta_j\Theta_k$

$$\delta_n^{(jk)}(\vec{\mathbf{x}}) = \sum_{i,s=1}^{r}\left[t_{ji}t_{ks}\left(\frac{q_n^{(si)}}{f_n} - \frac{g_n^{(i)}h'_s}{f_n h} - \frac{g_n^{(s)}h'_k}{f_n h} + 2\frac{h'_s h'_i}{h^2} - \frac{h''_{si}}{k}\right) + \right.$$
$$\left. + t_{ji}t'_{ksi}\left(\frac{g_n^{(s)}}{f_n} - \frac{h'_s}{h}\right)\right], \qquad \begin{matrix} j=1,\ldots,m \\ j\leq k\leq m \end{matrix}, \quad (21)$$

где для краткости опущены аргументы функций. Эмпирическая оценка (21) сходится по вероятности к оптимальной оценке (18) для п. в. $\vec{\mathbf{x}}$ в силу теоремы о сходимости борелевских функций [5, стр. 12], т. к. вероятность множества точек разрыва предельной функции (18) равна 0. Действительно, разрывы предельной функции обусловлены только нулями функций $f(\vec{\mathbf{x}})$ и $h(\vec{\mathbf{x}})$. Но нули функций $f(\vec{\mathbf{x}})$ и $h(\vec{\mathbf{x}})$ совпадают, в силу условия (1), а вероятность выпасть такому $\vec{\mathbf{x}}$, для которого $f(\vec{\mathbf{x}}) = 0$, равна нулю.

§ 3. Условия асимптотической оптимальности эмпирических оценок

Итак, построены эмпирические оценки (21), сходящиеся по вероятности для почти всех $\vec{\mathbf{x}}$ к оптимальным оценкам (18). Но это еще не есть полное решение задачи, поскольку нас интересуют не любые сходящиеся оценки, а только такие, для которых соответствующие средние риски сходятся к риску оптимальной бейесовой оценки с полной статистической информацией. Эмпирические оценки, обладающие таким свойством, были названы Роббинсом [6] асимптотически оптимальными. Понятно, что не для всех последовательностей сходящихся эмпирических оценок соответствующие средние риски сходятся к оптимальному бейесову риску, так же как не для любой последовательности сходящихся по вероятности случайных величин последовательность их математических ожиданий сходится к математическому ожиданию предельной случайной величины. Поэтому выясним достаточные условия асимптотической оптимальности эмпирических оценок.

Рассмотрим случай, когда значения векторной функции $\vec{\varphi}(\vec{\theta})$ принадлежат k-мерному замкнутому брусу $\mathbf{B} = \{a_i \leq \varphi_i \leq b_i,\ i = 1, \ldots, k\}$. Тогда по теореме о среднем значения векторной бейесовой оценки $\vec{\delta}_G(\vec{\mathbf{x}})$ также принадлежат брусу $\mathbf{B}$. Это свойство, однако, может не сохраняться для векторной эмпирической оценки $\vec{\delta}_n(\vec{\mathbf{x}}) = (\delta_{n,1}, \ldots, \delta_{n,k})$. Поэтому рассматривается усеченная векторная оценка $\vec{\delta}^*_n(\vec{\mathbf{x}}) = (\delta^*_{n,1}, \ldots, \delta^*_{n,k})$, компоненты которой определяются следующим образом:

$$\delta^*_{n,i} = \begin{cases} a_i\ , & \text{если} \quad \delta_{n,i} \leq a_i \\ \delta_{n,i}\ , & \text{если} \quad a_i < \delta_{n,i} < b_i \\ b_i\ , & \text{если} \quad \delta_{n,i} \geq b_i \end{cases}. \tag{22}$$

Ясно, что если эмпирическая оценка $\vec{\delta}_n(\vec{\mathbf{x}})$ сходится по вероятности к оптимальной оценке $\vec{\delta}_G(\vec{\mathbf{x}})$, то и усеченная оценка $\vec{\delta}^*_n(\vec{\mathbf{x}})$ сходится к $\vec{\delta}_G(\vec{\mathbf{x}})$. Достаточные условия асимптотической оптимальности выясняет следующая теорема.

Теорема 1. Если векторная функция $\vec{\varphi}(\vec{\theta})$ является ограниченной на Ω и эмпирическая оценка $\vec{\delta}_n(\vec{\mathbf{x}})$ сходится по вероятности к оптимальной бейе-

совой оценке $\vec{\delta}_G(\vec{\mathbf{x}})$ для п. в. $\vec{\mathbf{x}}$, то усеченная оценка $\vec{\delta}_n^*(\vec{\mathbf{x}})$ является асимптотически оптимальной.

Доказательство этой теоремы незначительно отличается от доказательства теоремы 1 работы [4] и поэтому здесь не приводится.

Применяя эту теорему к оценкам функций, представляющих собой всевозможные попарные произведения $\Theta_i\Theta_j$ и квадраты Θ_j^2, получаем, что для асимптотической оптимальности эмпирических оценок достаточно ограничить пространство Ω замкнутым брусом $\mathbf{B} = \{a_i \leq \theta_i \leq b_i,\ i = 1, \ldots, m\}$ и вместо оценок (21) рассматривать усеченные оценки $\delta_n^{(ij)^*}(\vec{\mathbf{x}})$, которые с учетом (22) получаются очевидным образом. Практически всегда можно выбрать такие значения a_i и b_i, $i = 1, \ldots, m$, что все значения исследуемых сигналов попадут в брус $\mathbf{B}$.

§ 4. Эмпирическая оценка полиномиальной функции одномерного случайного сигнала

Рассмотрим задачу оценки полиномиальной функции

$$\varphi(\Theta) \equiv a_1\Theta + a_2\Theta^2 + \ldots + a_N\Theta^N \tag{23}$$

неизвестного одномерного случайного сигнала Θ, где $a_1, a_2, \ldots, a_N$ — известные коэффициенты. В одномерном случае экспонентная плотность $f(\mathbf{x}/\theta)$ принимает вид

$$f(\mathbf{x} \mid \theta) = C(\theta)\, h(\mathbf{x})\, e^{\theta \mathbf{T}(\mathbf{x})}. \tag{24}$$

Так как оптимальная оценка линейной комбинации функций равняется линейной комбинации оптимальных оценок этих же функций, то в формуле (23) достаточно отдельно оценить каждый член суммы. Оптимальная оценка $\hat{\theta}$ первого члена получается из (6) при $r = m = 1$

$$\hat{\theta} = \frac{h}{f\mathbf{T}'} \frac{d}{d\mathbf{x}} \left(\frac{f}{h}\right), \tag{25}$$

где для краткости введены следующие обозначения $\mathbf{T}' = \dfrac{d}{dx}\mathbf{T}(\mathbf{x})$, $h = h(\mathbf{x})$, $f = f(\mathbf{x})$. Оптимальная оценка второго члена определяется формулой (18):

$$\hat{\theta}^2 = \frac{h}{f\mathbf{T}'} \frac{d}{d\mathbf{x}} \left[\frac{1}{\mathbf{T}'} \frac{d}{d\mathbf{x}} \left(\frac{f}{h}\right)\right]. \tag{26}$$

Для получения оптимальной оценки N-го члена (23) необходимо продифференцировать соотношение (9) N раз и при этом каждый раз делить левую

и правую часть на $\mathbf{T}'$. Тогда получается следующая оценка

$$\hat{\theta}^N = \frac{h}{f\mathbf{T}'} \frac{d}{d\mathbf{x}} \left\{ \frac{1}{\mathbf{T}'} \dots \frac{d}{d\mathbf{x}} \left[\frac{1}{\mathbf{T}'} \frac{d}{d\mathbf{x}} \left(\frac{f}{h} \right) \right] \dots \right\}. \tag{27}$$

Теперь можно записать оптимальную оценку $\hat{\varphi}(\theta)$ всей полиномиальной функции (23):

$$\hat{\varphi}(\vec{\theta}) = \frac{h}{f\mathbf{T}'} \frac{d}{d\mathbf{x}} \left\{ a_1 \frac{f}{h} + \frac{1}{\mathbf{T}'} \frac{d}{d\mathbf{x}} \left[a_2 \frac{f}{h} + \dots + \frac{1}{\mathbf{T}'} \frac{d}{d\mathbf{x}} \left(a_N \frac{f}{h} \right) \dots \right] \right\}. \tag{28}$$

Из (28) следует, что оценка $\hat{\varphi}(\theta)$ зависит от N первых производных безусловной плотности $f(\mathbf{x})$, так что для вычисления $\hat{\varphi}(\theta)$ необходимо уметь строить непараметрические оценки всех N производных плотности $f(\mathbf{x})$. Такие оценки для самой плотности $f(\mathbf{x})$ и ее первых двух производных были построены в [4] и приложении настоящей работы. Метод, который используется для построения оценок первых двух производных, позволяет в принципе строить оценки и более высоких производных, однако это сопряжено с достаточно длинными и утомительными выкладками, особенно при обобщении на многомерный случай. Поэтому в настоящей работе оценка более высоких производных плотности не рассматривается.

Приложение

Непараметрическая оценка вторых смешанных производных многомерной плотности вероятности

Необходимость построения оценок вторых производных некоторой плотности вероятности возникла в связи с задачей фильтрации полиномиальных функций случайных сигналов (см. § 2, (21)). Обозначим через $f''_{\mu\nu}(\vec{\mathbf{x}})$ вторую смешанную производную $\dfrac{\partial^2}{\partial x_\nu \partial x_\mu} f(\vec{\mathbf{x}})$ многомерной плотности $f(\vec{\mathbf{x}})$. Для этой производной предлагается следующая непараметрическая оценка:

$$q_n^{(\mu\nu)}(\vec{\mathbf{x}}) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{h_n^{2+r}} \left\{ \frac{(x_\mu - X_{k,\mu})(x_\nu - X_{k,\nu})}{h_n^2} - d_{\mu\nu} \right\} K\left(\frac{\vec{\mathbf{x}} - \vec{\mathbf{X}}_k}{h_n} \right), \tag{29}$$

где последовательность положительных констант $\{h_n\}$ обладает тем свойством, что

$$\lim_{n \to \infty} h_n = 0, \tag{30}$$

$$\lim_{n \to \infty} n h_n^{s+2r} = \infty, \tag{31}$$

а ядро $K(\vec{z})$ удовлетворяет следующим условиям:

1. $$K(\vec{z}) \geq 0\,, \tag{32.а}$$

2. $$K(\vec{z}) = K(-\vec{z})\,, \tag{32.б}$$

3. $$\sup_{\vec{z}\in x} K(\vec{z}) = C < \infty\,, \tag{32.в}$$

4. $$\int_x (z_\mu z_\nu - d_{\mu\nu})\, K(\vec{z})\, d\vec{z} = 0, \quad \mu, \nu = 1, \ldots, r, \tag{32.г}$$

5. $$\int_x z_i (z_\mu z_\nu - d_{\mu\nu})\, K(\vec{z})\, d\vec{z} = 0, \quad i, \mu, \nu = 1, \ldots, r, \tag{32.д}$$

6. $$\frac{1}{2}\int_x z_i z_j (z_\mu z_\nu - d_{\mu\nu})\, K(\vec{z})\, d\vec{z} = \begin{cases} 1 & \text{при } i=j=\mu=\nu, \\ 1/2 & \text{при } i=\mu,\ j=\nu, \\ 0 & \text{в противном случае} \end{cases} \tag{32.е}$$

7. $$\int_x |z_i z_j (z_\mu z_\nu - d_{\mu\nu})|\, K(\vec{z})\, d\vec{z} < \infty\,, \quad i, j, \mu, \nu = 1, \ldots r, \tag{32.ж}$$

8. $$\lim_{\substack{|z_k|\to\infty \\ k=1,\ldots,r}} |z_i z_j z_\mu z_\nu z_1 \ldots z_r\, K(\vec{z})| = 0, \quad i, j, \mu, \nu = 1, \ldots, r\,. \tag{32.и}$$

В качестве примера ядра $K(\vec{z})$, удовлетворяющего всем свойствам (32), можно указать многомерную нормальную плотность.

Вычислим математическое ожидание оценки (29). Учитывая условие (32.б), можно записать:

$$M_n q_n^{(\mu\nu)}(\vec{x}) = \frac{1}{n}\sum_{k=1}^{n} \frac{1}{h_n^{4+r}} \int_X (x_\mu - y_\mu)(x_\nu - y_\nu)\, K\left(\frac{\vec{x}-\vec{y}}{h_n}\right) f(\vec{y})\, d\vec{y} -$$

$$-\frac{1}{n}\sum_{k=1}^{n} \frac{d_{\mu\nu}}{h_n^{2+r}} \int_X K\left(\frac{\vec{x}-\vec{y}}{h_n}\right) f(\vec{y})\, d\vec{y} = \tag{33}$$

$$= \frac{1}{h_n^2}\int_X z_\mu z_\nu\, K(\vec{z})\, f(\vec{x} + h_n\vec{z})\, d\vec{z} - \frac{d_{\mu\nu}}{h_n^2}\int_X K(\vec{z})\, f(\vec{x} + h_n\vec{z})\, d\vec{z}.$$

По формуле конечных приращений

$$f(\vec{x} + h_n\vec{z}) = f(\vec{x}) + \sum_{i=1}^{r} h_n z_i f_i'(\vec{x}) + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} h_n^2 z_i z_j f_{ij}''(\vec{x} + \eta\, h_n\vec{z}), \tag{34}$$

где $0 < \eta < 1$.

Подставляя (34) в (33) и используя свойства (32.г), (32.д) и (32.е), получим для математического ожидания $q_n^{(\mu\nu)}(\vec{\mathbf{x}})$ следующее выражение:

$$M_n\, q_n^{(\mu\nu)}(\vec{\mathbf{x}}) = \frac{1}{2} \sum_{i,j=1}^{r} \int_X z_i z_j (z_\mu z_\nu - d_{\mu\nu})\, K(\vec{\mathbf{z}})\, f''_{ij}(\vec{\mathbf{x}} + \eta\, h_n \vec{\mathbf{z}})\, d\vec{\mathbf{z}}\,. \tag{35}$$

С другой стороны, производную $f''_{\mu\nu}(\vec{\mathbf{x}})$ с учетом (32.е) можно представить в виде

$$f''_{\mu\nu}(\vec{\mathbf{x}}) = \frac{1}{2} \sum_{i,j=1}^{r} \int_X z_i z_j (z_\mu z_\nu - d_{\mu\nu})\, K(\vec{\mathbf{z}})\, f''_{ij}(\vec{\mathbf{x}})\, d\vec{\mathbf{z}}. \tag{36}$$

Докажем теперь, что с ростом числа выборок n среднеквадратическое отклонение оценки от истинной производной

$$M_n [q_n^{(\mu\nu)}(\vec{\mathbf{x}}) - f''_{\mu\nu}(\vec{\mathbf{x}})]^2 \tag{37}$$

стремится к нулю.

Теорема. Пусть $K(\vec{\mathbf{z}})$ — действительная функция на R^r, удовлетворяющая условиям (32), а $\{h_n\}$ — последовательность положительных констант, обладающая свойствами (30) и (31). Пусть $\vec{\mathbf{X}}_1, \ldots, \vec{\mathbf{X}}_n$ последовательность независимых и одинаково распределенных случайных величин с плотностью $f(\vec{\mathbf{x}})$. Тогда, если

1^0 $f''_{\mu\nu}(\vec{\mathbf{x}})$, $\mu, \nu = 1, \ldots, r$, непрерывны,

2^0 $\int_X |f''_{\mu\nu}(\vec{\mathbf{x}})|\, d\vec{\mathbf{x}} < \infty$, $i, j = 1, \ldots, r$, то

(а) оценка (29) сходится в среднеквадратическом в каждой точке $\vec{\mathbf{x}}$ к $f''_{\mu\nu}(\vec{\mathbf{x}})$. Если же, кроме того,

3^0 $|f''_{\mu\nu}(\vec{\mathbf{x}})| \leq C_1$, $\mu, \nu = 1, \ldots, r$,

4^0 $f''_{\mu\nu}(\vec{\mathbf{x}})$, $\mu, \nu = 1, \ldots, r$, равномерно непрерывны на x, то

(б) оценка (29) сходится в среднеквадратическом равномерно по $\vec{\mathbf{x}}$ к $f''_{\mu\nu}(\vec{\mathbf{x}})$.

Доказательство. Подставляя в элементарное неравенство $(a-b)^2 \leq 2|a|^2 + 2|b|^2$ $q_n^{(\mu\nu)}(\vec{\mathbf{x}})$ вместо a и $f''_{\mu\nu}(\vec{\mathbf{x}})$ вместо b и беря от обеих частей математическое ожидание, получим неравенство

$$M_n [q_n^{(\mu\nu)}(\vec{\mathbf{x}}) - f''_{\mu\nu}(\vec{\mathbf{x}})]^2 \leq 2 M_n [q_n^{(\mu\nu)}(\vec{\mathbf{x}}) - M_n\, q_n^{(\mu\nu)}(\vec{\mathbf{x}})]^2 + 2 [M_n\, q_n^{(\mu\nu)}(\vec{\mathbf{x}}) - f''_{\mu\nu}(\vec{\mathbf{x}})]^2. \tag{38}$$

Оценив каждое из двух слагаемых правой части (38), мы сможем оценить (37) и тем самым доказать сходимость оценки (29). Рассмотрим сначала второе слагаемое правой части (38), которое с учетом (35) и (36) можно за-

писать в виде:

$$[M_n q_n^{(\mu\nu)}(\vec{\mathbf{x}}) - f''_{\mu\nu}(\vec{\mathbf{x}})]^2 = \left[\frac{1}{2}\sum_{i,j=1}^{r}\int_X z_i z_j(z_\mu z_\nu - d_{\mu\nu})\,K(\vec{\mathbf{z}})\cdot\right.$$

$$\left.\cdot\,(f''_{ij}(\vec{\mathbf{x}} + \eta\, h_n\vec{\mathbf{z}}) - f''_{ij}(\vec{\mathbf{x}}))\,d\vec{\mathbf{z}}\right].$$

Разбивая пространство X на области $|z_k| < \delta_k/h_n$ и $|z_k| \geq \delta_k/h_n$, $k = 1, \ldots, r$, где $\delta_1, \ldots, \delta_r$ — некоторые положительные постоянные, можно записать цепочку неравенств:

$$|M_n q_n^{(\mu\nu)}(\vec{\mathbf{x}}) - f''_{\mu\nu}(\vec{\mathbf{x}})|^2 \leq \frac{1}{2}\sum_{i,j=1}^{r}\int\limits_{\substack{|z_k|<\delta_k/h_n\\ k=1,\ldots,r}} z_i z_j(z_\mu z_\nu - d_{\mu\nu})\,K(\vec{\mathbf{z}})\,[f''_{ij}(\vec{\mathbf{x}} + \eta h_n\vec{\mathbf{z}}) -$$

$$-\,f''_{ij}(\vec{\mathbf{x}})]\,d\vec{\mathbf{z}}\Big| + \frac{1}{2}\Big|\sum_{i,j=1}^{r}\int\limits_{\substack{|z_k|\geq\delta_k/h_n\\ k=1,\ldots,r}} z_i z_j(z_\mu z_\nu - d_{\mu\nu})\,K(\vec{\mathbf{z}})\,f''_{ij}(\vec{\mathbf{x}} + \eta\, h_n\vec{\mathbf{z}})\,d\vec{\mathbf{z}}\Big| +$$

$$+\frac{1}{2}\Big|\sum_{i,j=1}^{r}\int\limits_{\substack{|z_k|\geq\delta_k/h_n\\ k=1,\ldots,r}} z_i z_j(z_\mu z_\nu - d_{\mu\nu})\,K(\vec{\mathbf{z}})\,f''_{ij}(\vec{\mathbf{z}})\,d\vec{\mathbf{z}}\Big| \leq \tag{39}$$

$$\leq \frac{1}{2}\sum_{i,j=1}^{r}\max_{\substack{|h_n z_k|<\delta_k\\ k=1,\ldots,r}}|f''_{ij}(\vec{\mathbf{x}} + \eta h_n\vec{\mathbf{z}}) - f''_{ij}(\vec{\mathbf{x}})|\int_X |z_i z_j(z_\mu z_\nu - d_{\mu\nu})|\,K(\vec{\mathbf{z}})\,d\vec{\mathbf{z}} +$$

$$+\frac{1}{2}\sum_{i,j=1}^{r}\sup_{\substack{|z_k|\geq\delta_k/h_n\\ k=1,\ldots,r}}|z_i z_j(z_\mu z_\nu - d_{\mu\nu})\,z_1\ldots z_r\,K(\vec{\mathbf{z}})|\,\frac{1}{\eta^r\,\delta_1\ldots\delta_r}\int_X |f''_{ij}(\vec{\mathbf{y}})|\,d\vec{\mathbf{y}} +$$

$$+\frac{1}{2}\sum_{i,j=1}^{r}|f''_{ij}(\vec{\mathbf{x}})|\int\limits_{\substack{|z_k|\geq\delta_k/h_n\\ k=1,\ldots,r}} |z_i z_j(z_\mu z_\nu - d_{\mu\nu})|\,K(\vec{\mathbf{z}})\,d\vec{\mathbf{z}}.$$

Пусть $n \to \infty$, тогда второе слагаемое последнего выражения стремится к нулю в силу (32.и) и условия 2° теоремы, а третье слагаемое стремится к нулю в силу (32.ж). Если затем все $\delta_k \to 0$, $k = 1, \ldots, r$, при выполнении условия 1° и (32.ж) получим сходимость к нулю первого слагаемого.

Если же, кроме того, выполнены условия 3° и 4° равномерной непрерывности и ограниченности производных плотности $f(\vec{\mathbf{x}})$, то

$$\sup_{\vec{\mathbf{x}}\in X}[M_n q_n^{(\mu\nu)}(\vec{\mathbf{x}}) - f''_{\mu\nu}(\vec{\mathbf{x}})]^2 \to 0. \tag{40}$$

Теперь оценим первое слагаемое правой части (38).

$$M_n[q_n^{(\mu\nu)}(\vec{\mathbf{x}}) - M_n q_n^{(\mu\nu)}(\vec{\mathbf{x}})]^2 = M_n\left[\frac{1}{n}\sum_{k=1}^{n}\frac{1}{h_n^{2+r}}\left\{\left(\frac{(x_\mu - X_{k,\mu})(x_\mu - X_{k,\nu})}{h_n^2} - d_{\mu\nu}\right)\cdot\right.\right.$$

$$\left.\left.\cdot K\left(\frac{\vec{\mathbf{x}} - \vec{\mathbf{X}}_k}{h_n}\right) - M_n\left(\frac{(x_\mu - X_{k,\mu})(x_\nu - X_{k,\nu})}{h_n^2} - d_{\mu\nu}\right) K\left(\frac{\vec{\mathbf{x}} - \vec{\mathbf{X}}_k}{h_n}\right)\right\}\right]^2 \leq$$

$$\leq \frac{1}{n^2 h^{4+2r}}\sum_{k=1}^{n} M_n\left[\left(\frac{(x_\mu - X_{k,\mu})(x_\nu - X_{k,\nu})}{h_n^2} - d_{\mu\nu}\right) K\left(\frac{\vec{\mathbf{x}} - \vec{\mathbf{X}}_k}{h_n}\right)\right]^2 =$$

$$\leq \frac{1}{nh^{4+r}}\int_X (z_\mu z_\nu - d_{\mu\nu})^2 K^2(\vec{\mathbf{z}})\, f(\vec{\mathbf{x}} + h_n\vec{\mathbf{z}})\, d\vec{\mathbf{z}} \leq$$

$$\leq \frac{C}{nh^{4+r}}\left[\int_{\substack{|z_k|<\delta_k/h_n \\ k=1,\ldots,r}} (z_\mu z_\nu - d_{\mu\nu})^2 K(\vec{\mathbf{z}})\, f(\vec{\mathbf{x}} + h_n\vec{\mathbf{z}})\, d\vec{\mathbf{z}} + \right.$$

$$\left. + \int_{\substack{|z_k|\geq\delta_n/h_n \\ k=1,\ldots,r}} (z_\mu z_\nu - d_{\mu\nu})^2 K(\vec{\mathbf{z}})\, f(\vec{\mathbf{x}} + h_n\vec{\mathbf{z}})\, d\vec{\mathbf{z}}\right] \leq$$

$$\leq \frac{C}{nh^{4+r}}\left[\frac{(\delta_\mu\delta_\nu + h_n^2 d_{\mu\nu})^2 C}{h_n^{4+r}} + \frac{1}{\delta_1\ldots\delta_r}\sup_{\substack{|z_k|\geq\delta_k/h_n \\ k=1,\ldots,r}} |(z_\mu z_\nu - d_{\mu\nu})\, z_1\ldots z_r\, K(\vec{\mathbf{z}})|\right].$$

Отсюда видно, что второе слагаемое последнего выражения стремится к нулю по предположению (32.и), а первое слагаемое — по условию (31). Заметим, что последнее выражение вообще не зависит от $\vec{\mathbf{x}}$, поэтому имеет место равномерная сходимость

$$\sup_{\vec{\mathbf{x}}\in X} M_n[q_n^{(\mu\nu)}(\vec{\mathbf{x}}) - M_n q_n^{(\mu\nu)}(\vec{\mathbf{x}})]^2 \to 0. \tag{42}$$

Из (39) и (41) в силу неравенства (38) следует доказательство утверждения (а), а из (40) и (42) — утверждение (б) настоящей теоремы.

Заключение

Методы оценки полиномиальных функций случайных параметров, рассмотренные в настоящей работе, являются развитием и обобщением эмпирического бейесового подхода Роббинса к оценкам случайных параметров с неизвестным априорным распределением. К сожалению, этот подход можно применять далеко не всегда. Как показывают примеры, успех достигается всякий раз, когда оптимальную оценку, зависящую от неиз-

вестного априорного распределения, удаётся выразить через характеристики безусловного распределения. Такими характеристиками могут быть, например, моменты распределения, сама функция распределения или её частные производные до некоторого порядка включительно. Поскольку каждая из таких характеристик в силу соотношения (4) тем или иным образом связана с неизвестным априорным распределением, то она, в свою очередь, неизвестна. Возникает математическая задача построения непараметрических приближений этих характеристик (см., например, приложение настоящей работы, где приводятся непараметрические оценки вторых частных производных безусловной плотности вероятности), которая является формализацией процесса самообучения в теории систем автоматического управления.

Однако возможность представления оптимальных оценок через характеристики безусловного распределения ограничена либо определённым классом оценок, либо определённым семейством условных распределений. В данной работе мы ограничились экспонентным семейством распределений (1), которое содержит большое число часто используемых на практике распределений, таких как многомерное нормальное χ^2-распределение, весь класс β-распределений и др. Было бы полезно выявить новые семейства условных распределений, позволяющих расширить область применения эмпирического бейесова подхода.

Литература

1. *Пугачев, В. С.:* Теория случайных функций и её применение к задачам автоматического управления. Физматгиз, 1962.
2. *Лэнинг, Дж. Х.—Бэттин, Р. Г.:* Случайные процессы в задачах автоматического управления. ИИЛ, 1958.
3. *Роббинс, Г.:* Эмпирический бейесов подход к статистике. Математика, сб. переводов **8**, 2 (1964).
4. *Добровидов, А. В.:* Об одном алгоритме непараметрической оценки случайных многомерных сигналов. Автоматика и телемеханика, *2* (1971).
5. *Скороход, А. В.:* Случайные процессы с независимыми приращениями. Физматгиз, 1963.
6. *Роббинс, Г.:* Эмпирический бейесов подход к задачам теории статистических решений. Математика, сб. переводов, **10**, 5 1966.

A nonsupervised algorithm of asymptotically optimal filtering of random signals with an unknown a priori distribution

A. V. DOBROVIDOV

(Moscow)

The classical methods of design of optimal information processing systems require the knowledge of all input signal and noise statistical properties. However, in practice we face the fact that the signal statistical properties are either unknown at all or require tremendous expenses for their determination. At the same time under sufficiently general conditions nonsupervised automatic systems may be constructed whose algorithms approach the optimal algorithms, resulting from completely known statistical properties of the problem. Furthermore, the information used for this algorithms is received only in on-line operation. This paper is concerned to the development of such automatic systems.

One important class of information processing systems, the optimal filtering systems are considered. In applications the necessity often arises to estimate not the useful signals themselves, but some of their functions. If such functions are non-linear, then optimal estimates are determined for all functions of useful signals at once.

Let it be required to estimate some vector-valued function of random multi-dimensional signal $\vec{\Theta} = (\Theta_1, \ldots, \Theta_m)$ with a priori distribution $G(\vec{\theta})$ on the basis of samplings of the vector-valued random variable $\vec{\mathbf{X}} = (X_1, \ldots, X_r)$ whose conditional density function $f(\vec{\mathbf{x}}/\theta)$ belongs to the general exponential family (1) where $T_j(\vec{\mathbf{x}})$ and $h(\vec{\mathbf{x}})$ are measured functions and $C(\vec{\theta})$ is a normalizing multiplier.

A solution of this problem minimizing the average risk (2) with squared loss function is an a posteriori mean (3) where $f(\vec{\mathbf{x}})$ is the marginal density function of random variable $\vec{\mathbf{X}}$.

The a priori distribution $G(\vec{\theta})$ of signal $\vec{\Theta}$ is as assumed completely unknown. Therefore the optimal estimate $\delta_G(\vec{\mathbf{x}})$ depending of $G(\vec{\theta})$ cannot be calculated. However, with the assumption (1) one can construct empirical Bayes estimates [3] converging to the optimal one with increasing number of samplings for a class of functions $\vec{\varphi}$ that are all kinds of paired products, $\Theta_j \Theta_k$, and squares, Θ_j^2, of useful random signal components. These empirical estimates are described by the expression (21), where t_{ji}, are elements of some well-determined matrix, f, h, h_i', h_{si}'' and t_{ksi}' are found in (19) and the remaining elements $f_n(\vec{\mathbf{x}})$, $g_n^{(i)}(\vec{\mathbf{x}})$ and $q_n^{(si)}(\vec{\mathbf{x}})$ are nonparametric estimates of the marginal density function $f(\vec{\mathbf{x}})$ and its first and second partial derivatives respectively. The necessity of using nonparametric approximation arises since we wish not to restrict the class of distributions $G(\vec{\theta})$ to a certain functional family. The nonparametric estimate of $f(\vec{\mathbf{x}})$ is a generalization of Parzen estimate. The nonparametric estimates of the first partial derivatives of $f(\vec{\mathbf{x}})$ are suggested in [4]. The nonparametric estimates of the second partial derivatives of $f(\vec{\mathbf{x}})$, necessary for (21), are defined by the expression (29) of this paper, where constants h_n and kernels $K(\cdot)$ satisfy the conditions (30)–(32).

Apart from the construction of converging succession of empirical estimates, the conditions are found whereby the average risks will converge to the optimal Bayes risk.

А. В. Добровидов
Институт проблем управления
СССР Москва В—485
Профсоюзная ул. 81

Problems of Control and Information Theory, Vol. 1 (2), pp. 177—191 (1912)

ON THE CHARACTERISTIC PROPERTIES OF GENERALIZED ENTROPY

J. FRITZ

(Budapest)

It is proved that generalized entropy — which is a common generalization of the most important information theoretical quantities including Shannon's entropy — can be characterized by a relatively simple set of postulates. Essentially, the mixing property, and the upper semicontinuity of entropy should be postulated. The proof is based on the fact that $\lim I(P_n \mid P) = 0$ implies convergence of P_n to P in the variation distance, P and P_n are probability measures.

The aim of this paper is to deduce generalized entropy from an intuitively plausible set of postulates. To compare this problem with the well known characterizations of Shannon's entropy [1—4], let us point out an essential difference in the formulation of the problems. Namely, Shannon's entropy is defined on the set of all finite discrete probability distributions, while generalized entropy can be defined for probability measures on a fixed measure space, so that isomorphic probability measures may have different generalized entropies. Especially, though Shannon's entropy can be considered as a special case of generalized entropy, the generalized entropy of a strictly atomic probability measure may differ from its Shannon entropy. Moreover, generalized entropy need not be additive for independent distributions like Shannon's entropy. Further difficulties follow from the fact that the continuity properties of generalized entropy are rather intricate.

These investigations are motivated by the following physical considerations. A possible starting point of statistical physics may be the principle that any state of a thermodynamical system can be represented by a probability measure on a suitable measure space, thus thermodynamical entropy will be a function on the set of these probability measures. Assumed that this function satisfies our postulates, we obtain that thermodynamical entropy can be interpreted as generalized entropy, the concrete form of which can be derived from some further postulates.

1. The elementary properties of generalized entropy

Let Σ denote the set of all probability measures P defined on the measurable space $(\Omega, \mathfrak{A})$; if we are given a σ-finite measure M on $\mathfrak{A}$, then the generalized entropy (M-entropy) of a probability measure $P \in \Sigma$ can be defined in the following way. Let f_P denote the density function (Radon—Nikodym derivative) of the absolutely continuous part of P with respect to M, and let Σ_M be the set of those probability measures $P \in \Sigma$, for which

$$\int f_P \ln^+ \frac{1}{f_P} dM < +\infty.$$

(Throughout this paper $\ln x$ means logarithm with respect to the base e with the additional convention that $0 \ln \frac{a}{0} = 0 \ln \frac{0}{a} = 0$, further $b \ln \frac{0}{b} = -\infty$, $b \ln \frac{b}{0} = +\infty$ if $a \geq 0$, $b > 0$.)

Definition 1. The generalized entropy (M-entropy) of a probability measure $P \in \Sigma_M$ is defined as

$$H_M(P) = \begin{cases} E(-\ln f_P) & \text{if } P \ll M, \\ -\infty & \text{if } P \not\ll M. \end{cases}$$

If $P \in \Sigma_M$, then $H_M(P)$ is not defined.

A systematic investigation of the properties of generalized entropy has been published by Csiszár [5], where $H_M(P)$ is defined for a bit wider class of probability measures than here, the difference however is not important.

Observe, that in case of a finite dominating measure M the M-entropy $H_M(P)$ is defined for every $P \in \Sigma$. If M is not finite, then $\Sigma_M \neq \Sigma$, and the description of Σ_M is very difficult. Fortunately, Σ_M has two simple algebraic properties. Let $P \perp Q$ denote that P and Q are orthogonal; i.e. there exists a set $A \in \mathfrak{A}$ such that $P(A) = 1$, $Q(A) = 0$; further, for a $P \in \Sigma$ and $A \in \mathfrak{A}$ let $P(\cdot \mid A)$ denote the conditional probability measure defined as $P(B \mid A) = \frac{P(AB)}{P(A)}$, provided that $P(A) > 0$. Then $AB = \emptyset$ implies $P(\cdot \mid A) \perp P(\cdot \mid B)$, and conversely, if $P_1 \perp P_2$ and $P_1(A) = P_2(B) = 1$, then $P_1 = P(\cdot \mid A)$, $P_2 = P(\cdot \mid B)$, where $P = wP_1 + (1 - w)P_2$, $0 < w < 1$. Since

$$\frac{dP(\cdot \mid A)}{dM} = \begin{cases} \frac{1}{P(A)} \frac{dP}{dM} & \text{on } A \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

If $P \ll M$ on the measurable subsets of A, this means that $P(\cdot \mid A) \in \Sigma_M$ if $P \in \Sigma_M$, further $P_1, P_2 \in \Sigma_M$, $P_1 \perp P_2$ imply that $P = wP_1 + (1 - w)P_2 \in \Sigma_M$ for an arbitrary w between 0 and 1.

Definition 2. A nonempty set $\Sigma_0 \subset \Sigma$ will be called a regular family of probability measure, if

(i) $P \in \Sigma_0$ implies $P(\cdot \mid A) \in \Sigma_0$ if $P(A) > 0$,

(ii) If $P \perp Q$, $0 \leq w \leq 1$ and $P, Q \in \Sigma_0$, then $wP + (1 - w)Q \in \Sigma_0$.

In view of the above statement, generalized entropy is defined on a regular family of probability measures. Of course, Σ_M has other significant properties, too, but (i) and (ii) will be sufficient for our investigations. Some further elementary properties of generalized entropy are summarized as follows.

A) $H_M(P) < +\infty$ for any $P \in \Sigma_M$.

B) If $\mathfrak{B} = \{A_i\}$ is a partition of Ω into measurable sets such that $\Sigma P(A_i) \ln P(A_i) > -\infty$, and $P(\cdot \mid A_i) \in \Sigma_M$ if $P(A_i) > 0$, then $P \in \Sigma_M$ and

$$H_M(P) = \Sigma P(A_i)\ H_M\big(P(\cdot \mid A_i)\big) + \Sigma P(A_i) \ln \frac{1}{(PA_i)},$$

provided that the right-hand side is less than $+\infty$.

C) If $M(A) < +\infty$ and $P(A) = 1$, then

$$H_M(P) \leq \ln M(A),$$

where equality holds if, and only if P is uniformly distributed on A with respect to M.

D) If $\mathfrak{B} = \{A_i\}$ is a partition of Ω and $M(A_i) < +\infty$ for each i, then $P \in \Sigma_M$ and

$$H_M(P) \leq \Sigma P(A_i) \ln \frac{M(A_i)}{P(A_i)},$$

provided that the right-hand side is less than $+\infty$.

Relation A) is an obvious consequence of the definition of Σ_M, while B) follows immediately from (1). To show C) and D) we may assume that $P \ll M$, thus $x \ln \frac{a}{x} \leq a - x$ implies

$$\int_A f_P \ln \frac{1}{f_P} dM - P(A) \ln \frac{M(A)}{P(A)} = \int_A f_P \ln \frac{P(A)}{M(A) f_P} dM \leq$$

$$\leq \int_A \left(\frac{P(A)}{M(A)} - f_P \right) dM = 0,$$

which proves C) and even D) if we apply it to the elements of the partition $\mathfrak{B}$.

The following simple special case of C) will be frequently used: If $\{p_i\}$ is a finite or countable probability distribution, and $a_i > 0$ for each i, then

$$\Sigma p_i \ln \frac{a_i}{p_i} \leq \ln \Sigma a_i , \tag{2}$$

where the condition of equality is $\frac{p_i}{a_i} = \text{const.}$

The above properties of generalized entropy suggest the assumptions of the following theorem.

Theorem 1. Let $H(P) < +\infty$ be a real valued function, defined on a regular family $\Sigma_0 \subset \Sigma$. ($H(P)$ may take on the value $-\infty$.) If $H(P)$ satisfies the postulates

(i) $$H\big(wP + (1-w)Q\big) = wH(P) + (1-w)H(Q) + w \ln \frac{1}{w} + (1-w) \ln \frac{1}{1-w},$$
$$\text{if } P \perp Q,\ P, Q \in \Sigma_0 \text{ and } 0 < w < 1 ,$$

(ii) $$\sup H(P) < +\infty ,$$

then

$$M(A) = \begin{cases} \sup e^{H(P)};\ P(A) = 1,\ P \in \Sigma_0 \\ 0 \text{ if } P(A) = 0 \text{ for each } P \in \Sigma_0 \end{cases} \tag{3}$$

defines such a finite measure M on $\mathfrak{R}$ that

$$H(P) \leq H_M(P)$$

holds on Σ_0. This measure M is unique in the sense that if $H(P) \leq H_M, (P)$ on Σ_0 with another measure M', then $M(A) \leq M'(A)$ for each $A \in \mathfrak{R}$.

Proof. First we show that

$$M(\Omega) \geq M(A) + M(B) \text{ if } A + B = \Omega,\ AB = \varnothing . \tag{4}$$

This is trivially true if $M(A) = 0$ or $M(B) = 0$, while in the opposite case, since Σ_0 is a regular family, we have the decomposition

$$\Sigma_0 = \{wP + (1-w)Q;\ P \in \Sigma_1, Q \in \Sigma_2, 0 \leq w \leq 1\} ,$$

where $\Sigma_1 = \{P; P(A) = 1, P \in \Sigma_0\}$, $\Sigma_2 = \{Q; Q(B) = 1, Q \in \Sigma_0\}$. Observe that Σ_1 and Σ_2 are disjoint sets, thus (ii) and (i) imply

$$+\infty > \ln M(\Omega) = \sup \{H(wP + (1-w)Q; wP + (1-w)Q \in \Sigma_0\} =$$

$$= \sup \left\{ w \ln \frac{e^{H(P)}}{w} + (1-w) \ln \frac{e^{H(Q)}}{1-w}; \; P \in \Sigma_1, \; Q \in \Sigma_2, \; 0 \leq w \leq 1 \right\} = .$$

$$= \sup_{0 \leq w \leq 1} \left\{ \sup_{P \in \Sigma_1} w \ln \frac{e^{H(P)}}{w} + \sup_{Q \in \Sigma_2} (1-w) \ln \frac{e^{H(Q)}}{1-w} \right\} =$$

$$= \sup_{0 \leq w \leq 1} \left\{ w \ln \frac{M(A)}{w} + (1-w) \ln \frac{M(B)}{1-w} \right\}$$

in view of the definition of M. Putting $w = \dfrac{M(A)}{M(A) + M(B)}$ we obtain $\ln M(\Omega) \geq \ln \big(M(A) + M(B)\big)$, which proves (4).

The following step is to show that

$$H(P) \leq \Sigma P(A_i) \ln \frac{M(A_i)}{P(A_i)} \leq \ln \Sigma M(A_i), \tag{5}$$

for any denumerable partition $\mathfrak{B} = \{A_i\}$ of Ω, and for any $P \in \Sigma_0$. Set $B_n = \sum_{i=n+1}^{\infty} A_i$, then $P(\cdot \mid A_i) - P(\cdot \mid A_j)$ if $i \neq j$, $P(\cdot \mid A_i) \perp P(\cdot \mid B_n)$ if $i \leq n$, further $H\big(P(\cdot \mid A_i)\big) \leq \ln M(A_i)$, $H\big(P(\cdot \mid B_n)\big) \leq \ln M(\Omega) < +\infty$. Since Σ_0 is a regular family, (i) can be extended for a finite number of mutually orthogonal probability measures by induction without any difficulty, thus

$$H(P) = \sum_{i=1}^{n} P(A_i) H\big(P(\cdot \mid A_i)\big) + \sum_{i=1}^{n} P(A_i) \ln \frac{1}{P(A_i)} +$$

$$+ P(B_n) H\big(P(\cdot \mid B_n)\big) + P(B_n) \ln \frac{1}{P(B_n)} \leq$$

$$\leq \sum_{i=1}^{n} (PA_i) \ln \frac{M(A_i)}{P(A_i)} + P(B_n) \ln \frac{M(Q)}{P(B_n)},$$

which proves the first part of (5), since $\lim P(B_n) \ln \dfrac{M(Q)}{P(B_n)} = 0$. The second part of (5) is a simple consequence of (2). Now, on account the definition of M, the inequality $H(P) \leq \ln \Sigma M(A_i)$ implies

$$M(\Omega) \leq \Sigma M(A_i), \tag{6}$$

which, compared with (4), shows that M is indeed a measure.

The relation $H(P) \leq H_M(P)$ follows immediately from the first part of (5), if $P \not\ll M$, since $P(A) \ln \frac{M(A)}{P(A)} = -\infty$ if $M(A) = 0$ but $P(A) > 0$. On the other hand, if $P \ll M$ and $f = \frac{dP}{dM}$, further

$$A_i = \{\omega;\ e^{-i\varepsilon} \geq f > e^{-(i+1)\varepsilon}\}\ ,\quad i \text{ integer},\ \varepsilon > 0\ ,$$

then from

$$e^{-i\varepsilon}\ M(A_i) \geq P(A_i) \geq e^{-(i+1)\varepsilon}\ M(A_i)$$

we obtain that

$$\frac{1}{f} \geq e^{i\varepsilon} \geq e^{-\varepsilon}\ \frac{M(A_i)}{P(A_i)} \quad \text{on } A_i, \text{ thus}$$

$$\varepsilon + \ln \frac{1}{f} \geq \ln \frac{M(A_i)}{P(A_i)} \quad \text{on } A_i, \text{ whence}$$

$$\varepsilon + H_M(P) \geq \sum_{i=-\infty}^{+\infty} P(A_i) \ln \frac{M(A_i)}{P(A_i)}\ . \tag{7}$$

However, $P(B) \ln \frac{M(B)}{P(B)} = 0$ for $B = \{\omega;\ f(\omega) = 0\}$, and the sets $\{B, A_i\}$ form a partition of Ω, thus the comparison of (5) and (7) yields

$$\varepsilon + H_M(P) \geq H(P)\ ,$$

which proves $H(P) \leq H_M(P)$, since ε can be choosen arbitrarily small.

The final statement of the theorem follows from

$$\ln M'(A) \geq$$

$$\geq \sup\{H_M,(P);\ P(A) = 1, P \in \Sigma_0\} \geq \sup\{H(P);\ P(A) = 1,\ P \in \Sigma_0\} = \ln M(A).$$

It is not too difficult to give examples, where the conditions of Theorem 1. are satisfied, but $H(P) < H_M(P)$ for some $P \in \Sigma_0$. The construction of such counter-examples is based on the following theorem.

Theorem 2. If $\Sigma_1 \subset \Sigma$ and $\Sigma_2 \subset \Sigma$ are disjoint regular families, then

$$\Sigma_0 = \{wP + (1 - w)Q;\ P \in \Sigma_1,\ Q \in \Sigma_2,\ P \perp Q,\ 0 \leq w \leq 1\} = \Sigma_1 \times \Sigma_2$$

will also be a regular family, and the above representation of the elements of Σ_0 is unique. (Except the trivial case $w = 0$ or $w = 1$.)

Proof: Assume that

$$w_1 P_1 + (1 - w_1) Q_1 = w_2 P_2 + (1 - w_2) Q_2 , \tag{8}$$

where $P_i \in \Sigma_1$, $Q_i \in \Sigma_2$, $P_i \perp Q_i$, $0 \leq w_i \leq 1$ for $i = 1, 2$. If $P_1(A) = Q_2(B) = 1$, $P_2(B) = Q_1(A) = 0$, then

$$w_1 P_1(ABC) = (1 - w_2) Q_2(ABC) \text{ for each } C \in \mathfrak{A} ,$$

thus

$$w_1 P_1(AB) = (1 - w_2) Q_2(AB) ,$$

whence the contradiction $P_1(\cdot \mid AB) = Q_2(\cdot \mid AB)$ follows if $w_1 P_1(AB) > 0$; that is our assumption implies that $w_1 P_1(AB) = (1 - w_2) Q_2(AB) = 0$. Similarly, putting $\bar{A}\bar{B}C$ in (8) we obtain that

$$(1 - w_1) Q_1(\bar{A}\bar{B}C) = w_2 P_2(\bar{A}\bar{B}C) \text{ for each } C \in \mathfrak{A} ,$$

whence $(1 - w_1) Q_1(\bar{A}\bar{B}) = w_2 P_2(\bar{A}\bar{B}) = 0$ follows in the same way as above. However, then either $P_1 \perp Q_2$ or $P_2 \perp Q_1$, and both of them are true if $0 < w_1 < 1$ or $0 < w_2 < 1$. In the latter case $A = \bar{B}$ may be supposed, and the uniqueness follows immediately. On the other hand, $w_1 = 1 - w_2 = 0$ or 1 is impossible, while $w_1 = w_2 = 0$ or 1 means that $Q_1 = Q_2$ or $P_1 = P_2$, respectively, thus the statement of the theorem is completely proved.

This result has the following interesting consequence: If we are given the functions $H_1(P)$ and $H_2(Q)$ on the disjoint regular families Σ_1 and Σ_2, respectively, then

$$H(R) = H\big(wP + (1 - w)Q\big) = wH_1(P) + (1 - w) H_2(Q) + w \ln \frac{1}{w} + \\ + (1 - w) \ln \frac{1}{1 - w}$$

is uniquely defined on $\Sigma_0 = \Sigma_1 \times \Sigma_2$, since the decomposition $R = wP + (1 - w)Q$, $P \in \Sigma_1$, $Q \in \Sigma_2$, $P \perp Q$, $0 \leq w \leq 1$ of the elements of Σ_0 is unique. Further, if H_1 and H_2 satisfy the assumptions (i) and (ii) of Theorem 1, then so does H, too. Namely, if $R_1 \in \Sigma_0$, $R_2 \in \Sigma_0$; $R_i = w_i P_i + (1 - w_i) Q_i$, $P_i \perp Q_i$, $P_i \in \Sigma_1$, $Q_i \in \Sigma_2$, $0 \leq w_i \leq 1$ for $i = 1, 2$, then $R_1 \perp R_2$ implies that $P_1 \perp P_2$ and $Q_1 \perp Q_2$, therefore, with the notation $\lambda = ww_1 + (1 - w)w_2$ we have $R = \lambda P + (1 - \lambda)Q$, $P \in \Sigma_1$, $Q \in \Sigma_2$ and $P \perp Q$, where

$$P = \frac{1}{\lambda} ww_1 P_1 + \frac{1}{\lambda} (1 - w)w_2 P_2$$

$$Q = \frac{1}{1 - \lambda} w(1 - w_1) Q_1 + \frac{1}{1 - \lambda} (1 - w)(1 - w_2) Q_2 .$$

Consequently, from the definition of H, and from the regularity of Σ_1 and Σ_2 we obtain that

$$H(R) = \lambda H_1(P) + (1-\lambda) H_2(Q) + \lambda \ln \frac{1}{\lambda} + (1-\lambda) \ln \frac{1}{1-\lambda} =$$

$$= ww_1 H_1(P_1) + (1-w) w_2 H_1(P_2) + ww_1 \ln \frac{\lambda}{ww_1} + (1-w)w_2 \ln \frac{\lambda}{(1-w)w_2} +$$

$$+ w(1-w_1) H_2(Q_1) + (1-w)(1-w_2) H_2(Q_2) + w(1-w_1) \ln \frac{1-\lambda}{w(1-w_1)} +$$

$$+ (1-w)(1-w_2) \ln \frac{1-\lambda}{(1-w)(1-w_2)} + \lambda \ln \frac{1}{\lambda} + (1-\lambda) \ln \frac{1}{1-\lambda} =$$

$$= wH(R_1) + (1-w) H(R_2) + w \ln \frac{1}{w} + (1-w) \ln \frac{1}{1-w}.$$

This proves the validity of (i) for H, while that of (ii) is an obvious consequence of the definition of H.

Since $\Sigma_1 \subset \Sigma_0$, $\Sigma_2 \subset \Sigma_0$, these investigations show that the values of H on Σ_1 can be choosen independently of its values on Σ_2. For example, if $(\Omega, \mathfrak{A}, M)$ is the $(0, 1)$ interval with the Lebesque-measure on the Borel subsets, further Σ_1 is the minimal regular family including the uniform distribution, Σ_2 is the minimal regular family including the exponential distribution of parameter 1; $H_1(P) = H_M(P)$ if $P \in \Sigma_1$, $H_2(Q) = H_M(Q) - 1$ if $Q \in \Sigma_2$, then H_1, H_2, and their common extension H on $\Sigma_0 = \Sigma_1 \times \Sigma_2$ all satisfy the assumptions (i) and (ii) of Theorem 1, but obviously,

$$\sup \{e^{H(R)};\ R(A) = 1, R \in \Sigma_0\} = M(A),$$

consequently, $H_M(R) > H(R) = H_M(R) - 1$ if $R \in \Sigma_2$.

These results show that to deduce $H(P) = H_M(P)$ we need some further postulates. Of course, the most natural idea is to postulate some continuity properties of generalized entropy.

2. The continuity properties of generalized entropy

We shall investigate convergence of generalized entropy with respect to the variation distance, defined as

$$| P - Q | = \sup \{2 \mid P(A) - Q(A) \mid ;\ A \in \mathfrak{A}\}.$$

Observe, that

$$| P - Q | = \int | f - g | \, dM$$

If $P \ll M, Q \ll M$ and $f = \frac{dP}{dM}$, $g = \frac{dQ}{dM}$, that is variation convergence means L_1 convergence of the density functions if they exist, and the limit of absolutely continuous probability measures will be also an absolutely continuous probability measure.

We shall use the following continuity properties of generalized entropy to characterize it.

E) If $P \in \Sigma_M$, A_i is a decreasing sequence of events, and $\lim P(A_i) = 0$, then

$$\overline{\lim}\, P(A_i)\, H_M\big(P(\cdot \mid A_i)\big) \leq 0$$

F) If M is finite, then $\lim \mid P - P_n \mid = 0$ implies

$$\overline{\lim}\, H_M(P_n) \leq H_M(P).$$

G) If $\lim \mid P - P_n \mid = 0$, $P, P_n \in \Sigma_M$, and A_i is such a decreasing sequence that $\lim P(A_i) = 0$, then

$$\overline{\lim_i}\, \overline{\sup_n}\, P_n(A_i)\, H_M\big(P_n(\cdot \mid A_i)\big) \leq 0,$$

then

$$\lim H_M(P_n) \leq H_M(P)\,.$$

The proof of these seemingly difficult statements needs only very simple analytical arguments. E) follows simply from

$$P(A)\, H_M\big(P(\cdot \mid A)\big) = \int_A f_P \ln \frac{1}{f_P}\, dM + P(A) \ln \frac{1}{P(A)}\,,$$

since the case $P \not\ll M$ is trivial. On account of $x \ln \frac{1}{x} < 1$, F) is a direct consequence of the Fatou—Lebesque Theorem in the only nontrivial case, when $\overline{\lim}\, H_M(P_n) > -\infty$, as then the elements of the corresponding subsequence, thus P, too, are absolutely continuous.

Finally, to prove G) let $\mathcal{C} = \{C_M\}$ be such a partition of Ω that $M(C_M) < +\infty$ and $P(C_M) > 0$ for each m, and set $A_i = \sum_{m=i+1}^{\infty} C_M$, $B_i = \sum_{m=1}^{i} C_M$. Then $\lim_n \mid P(\cdot \mid B_i) - P_n(\cdot \mid B_i) \mid = 0$ for each fixed i as $\lim_n P_n(B_i) = P(B_i) > 0$, therefore, since $\lim_n P(A_i) = 0$; E), F) and the assumption of G) imply

$$\overline{\lim}\, H_M(P_n) \leq \overline{\lim_i}\, \overline{\lim_n}\, P_n(B_i)\, H_M\big(P_n(\cdot \mid B_i)\big) +$$

$$+ \overline{\lim_i}\, \sup_n P_n(A_i) H_M\big(P_n(\cdot \mid A_i)\big) + \overline{\lim_i}\, \overline{\lim_n}\, \big(-P_n(B_i) \ln P_n(B_i)\big) +$$

$$+ \overline{\lim_i}\, \overline{\lim_n}\, \big(-P_n(A_i) \ln P_n(A_i)\big) \leq \overline{\lim}\, P(B_i)\, H_M\big(P(\cdot \mid B_i)\big) =$$

$$= \lim \int_{B_i} f_P \ln \frac{1}{f_P}\, dM = H_M(P)$$

if $\overline{\lim}\, H_M(P_n) > -\infty$, (then $P \ll M$), while in the opposite case we have $\lim H_M(P_n) = -\infty \leq H_M(P)$.

Let us remark that the assumption of G) is always satisfied if M is a finite measure, thus F) is a special case of G).

Theorem 3. If the function $H(P)$ is defined on a regular family $\Sigma_0 \subset \Sigma$, Σ_0 is closed with respect to the variation convergence, and

(i) $H(P) < +\infty$ on Σ_0,

(ii) If $P \perp Q$, $P, Q \in \Sigma_0$ and $0 \leq w \leq 1$, then

$$H\big(wP + (1-w)Q\big) = w\, H(P) + (1-w)\, H(Q) + w \ln \frac{1}{w} + (1-w) \ln \frac{1}{1-w},$$

(iii) $\lim \mid P - P_n \mid = 0$ implies $\overline{\lim}\, H(P_n) \leq H(P)$ on Σ_0,

then (3) defines such a finite measure M on $\mathfrak{A}$ that

$$H(P) = H_M(P) \text{ on } \Sigma_0 .$$

Proof: To show that

$$M(\Omega) = \sup_{P \in \Sigma_0} e^{H(P)} < +\infty, \tag{9}$$

we may assume that $H(P) > -\infty$ at least for two different elements P' and Q' of Σ_0. Then there exists an $A \in \mathfrak{A}$ such that $P'(A) \neq Q'(A)$; thus, for example, $P'(A) > 0$, $Q'(\bar{A}) > 0$, that is $P = P'(\cdot \mid A) \in \Sigma_0$, $Q = Q'(\cdot \mid \bar{A}) \in \Sigma_0$, $P(A) = Q(\bar{A}) = 1$, and from (ii) we know that $H(P) > -\infty$, $H(Q) > -\infty$. Let now $Q_n \in \Sigma_0$ be such a sequence, that $Q_n(\bar{A}) = 1$, further $\lim e^{H(Q_n)} = M(\bar{A})$, and set $P_n = (1 - w_n) P + w_n Q_n$, where $0 < w_n < 1$, and $\lim w_n = 0$. Observe, that for an arbitrary $B \in \mathfrak{A}$ we have

$$\mid P(B) - P_n(B) \mid = w_n \mid Q_n(B) - P(B) \mid \leq 2w_n ,$$

thus $\lim | P - P_n | = 0$, and (ii), (iii) imply, that

$$H(P) \geq \overline{\lim}\, H(P_n) \geq \underline{\lim}\, (1 - w_n)\left(H(P) - \ln (1 - w_n)\right) +$$
$$+ \underline{\lim}\, w_n \left(H(Q_n) - \ln w_n\right) = H(P) + \underline{\lim}\, w_n H_n(Q_n) ,$$

whence $\underline{\lim}\, w_n H(Q_n) \leq 0$ follows as $H(P)$ is finite. However, this is possible for an arbitrary sequence w_n with $\lim w_n = 0$ only if $\lim H(Q_n) < +\infty$, that is $M(\bar{A}) < +\infty$, and similarly, $M(A) < +\infty$. On the other hand, from (ii) and from (2) we have for any $P \in \Sigma_0$ that

$$H(P) = P(A) \ln \frac{e^{H(P(\cdot \mid A))}}{P(A)} + P(\bar{A}) \ln \frac{e^{H(P(\cdot \mid \bar{A}))}}{P(\bar{A})} \leq$$

$$\leq P(A) \ln \frac{M(A)}{P(A)} + P(\bar{A}) \ln \frac{M(\bar{A})}{P(\bar{A})} \leq \ln \left(M(A) + M(\bar{A})\right) < +\infty,$$

which implies (9), therefore Theorem 1 can be applied and it yields that M is a finite measure, and

$$H(P) \leq H_M(P) \text{ for each } P \in \Sigma_0 . \tag{10}$$

The second step of the proof is to show that if U denotes the uniform distribution with respect to M $\left[\text{i.e. } U(A) = \frac{M(A)}{M(\Omega)}\right]$, then $U \in \Sigma_0$, and

$$\sup_{P \in \Sigma_0} H(P) = \ln M(\Omega) = H(U) = H_M(U). \tag{11}$$

The basic idea of this argument is that $\lim H(P_n) = \ln M(\Omega)$ implies $\lim | U - P_n | = 0$, whence (11) follows by the assumptions of the theorem.

Let $P \in \Sigma_0$, $A \in \mathfrak{R}$; since we investigate P if $H(P)$ is close to its least upper bound $\ln M(\Omega) > -\infty$, we may assume that $H(P) > -\infty$. ($M(\Omega) = 0$ is a trivial case, when $H(P) = -\infty$ for each $P \in \Sigma_0$.) With the notations $e^{H(P(\cdot \mid A))} = a$, $e^{H(P(\cdot \mid \bar{A}))} = b$, $P(A) = w$ we have from (ii) and from (2) that

$$H(P) = w \ln \frac{a}{w} + (1 - w) \ln \frac{b}{1 - w} \leq \ln (a + b); \tag{12}$$

observe, that this is true even if $w = 0$ or $w = 1$. Further, as

$$a \leq M(A) , \quad b \leq M(\bar{A}) , \quad a + b \leq M(\Omega) , \tag{13}$$

it follows that

$$+\infty > \delta = \ln M(\Omega) - H(P) \geq (w + 1 - w) \ln (a + b) - H(P) =$$
$$= w \ln \frac{w}{\frac{a}{a + b}} + (1 - w) \ln \frac{1 - w}{\frac{b}{a + b}} \geq 2 \left(\frac{a}{a + b} - w\right)^2 ,$$

where the last estimation is a special case of Theorem 4.1 of Csiszár [6]. On the other hand, the relations of (13) imply

$$\frac{a}{M(\Omega)} - \frac{M(\Omega)-b}{M(\Omega)} \leq \frac{a}{a+b} - \frac{M(A)}{M(\Omega)} = 1 - \frac{b}{a+b} - \frac{M(A)}{M(\Omega)} \leq 1 - \frac{b}{M(\Omega)} - \frac{a}{M(\Omega)},$$

whence we obtain that

$$\left| \frac{a}{a+b} - \frac{M(A)}{M(\Omega)} \right| \leq 1 - \frac{a+b}{M(\Omega)} \leq \ln \frac{M(\Omega)}{a+b} = \ln M(\Omega) - \ln(a+b) \leq \delta,$$

since $1 - x \leq \ln \frac{1}{x}$, and $\ln M(\Omega) - \ln (a+b) \leq \ln M(\Omega) - H(P)$ from (12). The comparison of this estimation with the above one yields

$$\left| w - \frac{M(A)}{M(\Omega)} \right| \leq \left| w - \frac{a}{a+b} \right| + \left| \frac{a}{a+b} - \frac{M(A)}{M(\Omega)} \right| \leq \sigma + \sqrt{\frac{\sigma}{2}},$$

that is

$$| P(A) - U(A) | = 0 \left(\ln M(\Omega) - H(P)\right)$$

uniformly in $A \in \mathfrak{A}$, which proves that $\lim H(P_n) = \ln M(\Omega)$ implies that $\lim | U - P_n | = 0$; thus the basic relation (11) is completely proved.

We are now in a position to show that equality holds in (10). First we remark that putting $P = U$ in (12), we obtain from (11) and from (13) that

$$H\big(U(\,\cdot\,|\,A)\big) = H_M\big(U(\,\cdot\,|\,A)\big) = \ln M(A) \text{ if } M(A) > 0\,. \tag{14}$$

To prove $H(P) = H_M(P)$, we may assume that $P \ll M$, since $H(P) = H_M(P) = -\infty$ from (10) otherwise. Let $f = \frac{dP}{dM}$ and set

$$C_{ni} = \left\{\omega; \frac{i}{n} \leq f < \frac{i+1}{n}\right\}, i = 0, 1, 2, \ldots, \quad n = 1, 2, \ldots,$$

$$P_n = \sum_{i=0}^{\infty} P(C_{ni})\, U(\,\cdot\,|\,C_{ni}),$$

$$P_{nk} = \frac{1}{P(B_{nk})} \sum_{i=0}^{k} P(C_{ni})\, U(\,\cdot\,|\,C_{ni}),$$

where $B_{nk} = \sum_{i=0}^{k} C_{ni}$. As $U \in \Sigma_0$ and Σ_0 is a regular family, $P_{nk} \in \Sigma_0$, further

$$f_n = \frac{dP_n}{dM} = \frac{P(C_{ni})}{M(C_{ni})} \text{ on } C_{ni} \text{ and } f_{nk} = \frac{dP_{nk}}{dM} = \frac{P(C_{ni})}{P(B_{nk})\, M(C_{ni})} \text{ on } C_{ni},$$

thus we have $\lim_k |P_n - P_{nk}| = \lim_k \int |f_n - f_{nk}|\, dM = \lim_k \frac{1}{P(B_{nk})} - 1 = 0$, whence $P_n \in \Sigma_0$ follows as Σ_0 is closed with respect to the variation convergence, further, in view of (iii), we have

$$H(P_n) \geq \overline{\lim_k} H(P_{nk})$$

However, on account of (ii), we see from (14) that $H(P_{nk}) = H_M(P_{nk})$, and it is easy to verify that $\lim_k H_M(P_{nk}) = H_M(P_n)$, therefore

$$H(P_n) \geq H_M(P_n)\,. \tag{15}$$

On the other hand, as $\frac{i}{n} \leq \frac{P(C_{ni})}{M(C_{ni})} \leq \frac{i+1}{n}$, we have $|f - f_n| \leq \frac{1}{n}$; that is $\lim |P - P_n| = 0$, whence

$$H(P) \leq \overline{\lim} H(P_n) \tag{16}$$

follows by (iii). Finally, from D) we know that

$$H_M(P_n) \geq H_M(P)\,, \tag{17}$$

and the comparison of (15), (16) and (17) yields $H(P) \geq H_M(P)$, which in view of (10) proves the statement of the theorem.

A similar theorem holds for the case when the dominating measure is not necessarily finite; the proof is mainly based on Theorem 3.

Theorem 4. Let $H(P)$ be defined on a regular family $\Sigma_0 \subset \Sigma$. Assume that $H(P)$ satisfies the conditions (i), (ii) of Theorem 3, and E) and G) if we replace H_M by H and Σ_M by Σ; further there exists a partition $\mathcal{C} = \{C_i\}_{i=1}^{\infty}$ such that

H) $P(C_i) = 1$ implies $P \in \Sigma_0$,

I) If $P_n(C_i) = 1$ and $\lim |P_n - P| = 0$, then $\overline{\lim} H(P_n) \leq H(P)$.

Then (3) defines such a σ-finite measure M on $\mathfrak{A}$ that $H(P) = H_M(P)$ on Σ_0.

Proof: First we prove that

$$H(P) = \sum_{i=1}^{\infty} P(C_i)\,\big(H(P(\cdot\,|C_i)) - \ln P(C_i)\big). \tag{18}$$

Set $A_k = \sum_{i=k+1}^{\infty} C_i$, then from (*ii*) we have

$$H(P) = \sum_{i=1}^{k} P(C_i)\,\big(H(P(\cdot\,|C_i)) - \ln P(C_i)\big) + P(A_k)\,\big(H(P(\cdot\,|A_k)) - \ln P(A_k)\big),$$

but $\overline{\lim} P(A_k) H(P(\cdot \mid A_k)) \leq 0$ follows from E), which means that the left-hand side of (18) is not greater than the right-hand side. The opposite inequality is a simple consequence of G) and (ii), since $\lim \mid P - P(\cdot \mid \bar{A}_k) \mid = 0$, and the condition of G) is obviously satisfied.

Since Theorem 3 can be applied if H is restricted to $\Sigma_i = \{P;\ P(C_i) = P \in \Sigma_0\}$, the statement of the theorem follows from (18) if we know that M is a measure on $\mathfrak{R}$, that is

$$M(\Sigma B_n) = \Sigma M(B_n) \tag{19}$$

holds for any finite or countable subset $\{B_n\}$ of $\mathfrak{C}$. However, if $M(\Sigma B_n) < < +\infty$, then Theorem 1 can be applied to $\Sigma^+ = \{P;\ P(\Sigma B_n) = 1,\ P \in \Sigma_0\}$, and it yields (19), while if $\Sigma M(B_n) < +\infty$, then from (18) by (2) and (3) we have

$$H(P) \leq \Sigma P(B_n) \ln \frac{M(B_n)}{P(B_n)} \leq \ln \Sigma M(B_n) \quad \text{if} \quad P \in \Sigma^+,$$

whence $M(\Sigma B_n) \leq \Sigma M(B_n) < +\infty$; thus (19) follows again by Theorem 1, and the proof is complete.

References

1. *Shannon, C. E. — Weaver, W.:* The mathematical theory of communication. Urbana, University of Illinois Press, 1949.
2. *Fadeev, D. K.:* Zum Begriff der Entropie einer endlichen Wahrscheinlichkeitsschemas. Arbeiten zur Informationstheorie I. Berlin DVW (1957) 85—90.
3. *Rényi, A.:* On measures of entropy and information. Proc. Fourth Berkeley Symp. Math. Stat. Probability **1**. Berkeley 1961, 541—561.
4. *Aczél, J.:* On different characterizations of entropies. Proc. Int. Symp. McMaster Univ. (1968), Canada, Lecture Notes in Math. **89**.
5. *Csiszár, I.:* On generalized entropy. Studia Sci. Math. Hungar. **4** (1969) 401—419.
6. *Csiszár, I.:* Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hungar. **2** (1967) 299—318.
7. *Fritz, J.:* An approach to the entropy of point processes (in the press).

Об аксиоматическом описании обобщенной энтропии.

Й. ФРИЦ

(Будапешт)

Пусть Σ — совокупность вероятностных мер на измеримом пространстве (Ω, A), M — σ-конечная мера на $\mathfrak{A}$, и f_p — плотность абсолютно непрерывной части вероятности $P \in \Sigma$. Обобщенная энтропия $H_M(P)$ определена тогда и только тогда, если $\int f_p \ln^+ \frac{1}{f_p} dM < +\infty$, и в этом случае $H_M(P) = -\infty$, если $P \not\ll M$ и $H_M(P) = E(-\ln f_p)$, если $P \ll M$. Непустое множество $\Sigma_0 \subset \Sigma$ называется регулярным классом, если смесь его ортогональных элементов тоже принадлежит Σ_0. Далее, если условная вероятность $P(\cdot|A) \in \Sigma_0$ для $P \in \Sigma_0$, $\mathfrak{A} \in A$, $P(A) > 0$. $|P - Q|$ обозначит вариационное расстояние мер P и Q.

Основной результат работы – теорема 3.: Если $H(P)$ — действительная функция на регулярном классе, $\Sigma_0 \subset \Sigma$, Σ_0 замкнуто в норме $|P - Q|$, далее

(*i*) $H(P) < +\infty$ для $P \in \Sigma_0$,

(*ii*) $H\big(wP + (1-w)Q\big) = w\,H(P) + (1-w)\,H(Q) + w \ln \frac{1}{w} + (1-w) \ln \frac{1}{1-w}$

для $P, Q \in \Sigma_0$, $0 \leq w \leq 1$,

(*iii*) $\overline{\lim}\, H(P_n) \leq H(P)$ на Σ_0, если $\lim |P - P_n| = 0$,

тогда $M(A) = \sup \{e^{P(H)}; P(A) = 1\}$ определяет конечную меру M на $\mathfrak{A}$, и $H(P) = H_M(P)$ на Σ_0.

Решающий аргумент доказательства заключается в том, что из $\lim H(P_n) = \sup_{P \in \Sigma_0} H(P)$ вытекает $\lim |P_n - U| = 0$, где U — равномерное распределение. Это предположение вытекает из $|P - Q| \leq \sqrt{2I(P|Q)}$.

József Fritz
Mathematical Research Institute
Budapest V. Reáltanoda u. 13—15. Hungary

NOTE TO CONTRIBUTORS

Two copies of the manusripts (each duly completed by figures, tables and references) are to be sent either 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

J. Kocsis coordinating editor

Department of Automation
Technical University

Budapest XI, Garami Ernő tér 3, Hungary

The authors are requested to retain a third copy of the submitted typescript to be able to check the proofs against it.

The papers, preferably in English or Russian, should be typed double-spaced on one side of good-quality paper, with wide margins (c. 4 – 5 cm) should carry the title of the contribution, the author(s)' name, and the name of the country. At the end of the typescript the name of that author who manages the proof-reading should also be given.

An abstract of about 50 to 100 words should head the paper.

The authors are encouraged to use the following headings: Introduction, (outlining the problem), Methods and results, Discussion, Conclusions, References. The entire material should not exceed 15 pages including tables and references. The proper location of the tables and figures must be indicated on the margins. Mathematical notations should follow up-to-date usage.

The summary – possibly in Russian if the paper is in English and *vice-versa* – should contain a brief account of the proposition and indications of the formulas used and figures shown in the paper. The summary is not supposed to exceed 10 – 15 per cent of the paper.

The authors will be sent sheet-proofs which they are to return by next mail to the sender Regional Editorial Board.

Authors are entitled to 100 reprints free of charge. Rejected manuscripts will be returned to the authors.

К СВЕДЕНИЮ АВТОРОВ

Рукописи в двух экземплярах (каждый из которых должен содержать рисунки, таблицы и литературу) направляются

Е. Д. *Теряеву* — Научный секретарь журнала

Отделение механики и процессов управления
Академия Наук СССР

Ленинский Проспект, 14, Москва В-71, СССР

или

Я. Кочишу — Научный секретарь журнала

Кафедра Автоматизации
Будапештского Технического
Университета, Будапешт XI, площадь
Гарами Эрнё, 3, Венгрия

Авторам рекомендуется оставлять у себя копию всех представленных ими материалов для справок при корректуре.

Статьи, желательно на русском или английском языках, отпечатанные на бумаге хорошего качества, с промежутком в два интервала и широкими (4—5 см) полями должны содержать наименование статьи, фамилию автора (авторов), название страны. В конце статьи необходимо также указать фамилию автора, ответственного за корректуру гранок.

Статье должна предшествовать аннотация объемом до 50—100 слов.

Авторы при написании статьи должны придерживаться следующей формы: введение (постановка задачи), основное содержание и результаты, обсуждение, выводы и литература. Объем статьи не должен превышать 15 печатных страниц, включая таблицы и ссылки. Последовательность таблиц и рисунков должна быть отмечена на полях. Математические обозначения рекомендуется давать в соответствии с современными требованиями и традициями.

К статье обязательно должно быть приложено резюме-реферат. Резюме — на русском языке, если статья написана на английском, и наоборот — должно содержать краткое изложение текста статьи со ссылками на необходимые формулы и графики, имеющиеся в основном тексте. Объем резюме не должен превышать 10—15% объема статьи.

Авторам высылаются гранки статьи, которые они должны незамедлительно возвратить в Региональную секцию Редколлегии журнала.

Авторам обеспечивается бесплатно 100 оттисков их статей. Рукописи непринятых статей возвращаются авторам.

CONTENTS • СОДЕРЖАНИЕ

VOL. 1 * NUMBER 3–4
TOM ТОМ НОМЕР

ACADEMY OF SCIENCES OF THE USSR
HUNGARIAN ACADEMY OF SCIENCES

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ

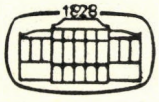
АКАДЕМИЯ НАУК СССР
АКАДЕМИЯ НАУК ВЕНГРИИ

1972

AKADÉMIAI KIADÓ, BUDAPEST

PROBLEMS OF CONTROL AND INFORMATION THEORY

ПРОБЛЕМЫ УПРАВЛЕНИЯ И ТЕОРИИ ИНФОРМАЦИИ



AKADÉMIAI KIADÓ

PUBLISHING HOUSE OF THE HUNGARIAN ACADEMY OF SCIENCES

BUDAPEST

Printed in Hungary

НЕКРОЛОГ Б. С. СОТСКОВА

4 ноября 1972 года скоропостижно скончался член редколлегии журнала «Проблемы управления и теории информации» член-корреспондент АН СССР, доктор технических наук профессор *Борис Степанович Сотсков*.

Советская наука потеряла крупнейшего ученого, педагога и общественного деятеля, создателя и руководителя отечественной школы теории элементов технических средств автоматики и теории надежности в приборостроении, организатора и руководителя ряда перспективных научных направлений в измерительной и информационной технике.

Борисом Степановичем опубликовано свыше 160 научных трудов, в том числе 7 книг по теории и расчетам элементов технических средств автоматики. Эти книги, переведенные на многие иностранные языки, служат основным учебником по подготовке специалистов по техническим средствам автоматики в СССР и в других социалистических странах.

Борис Степанович Сотсков — один из основоположников научных основ создания Государственной системы промышленных приборов и средств автоматики СССР.

Находясь постоянно на передовой линии современной науки, *Борис Степанович* успешно вел в последние годы важнейшие исследования в области отказов технических средств автоматики и методов прогнозирования надежности с целью построения новых высоконадежных технических средств автоматики и измерительной техники, занимался поисками новых путей построения технических средств автоматики, в частности на основе так называемых двойных физических эффектов.

Много сил отдал *Борис Степанович* организации и развитию научного сотрудничества с социалистическими странами в области приборостроения как при разработке научных основ создания Международной системы средств автоматического контроля и регулирования (УРС), так и по многим другим направлениям.

Борис Степанович был одним из организаторов журнала «Проблемы управления и теории информации» и вел раздел связанный с техническими средствами.

Огромная эрудиция, душевная щедрость и исключительная работоспособность позволили *Борису Степановичу Сотскову*, наряду с его многогранной научной деятельностью, иметь постоянные тесные связи с приборостроительной промышленностью, внести непосредственный вклад в технический прогресс в отечественном приборостроении.

Светлый образ замечательного человека, выдающегося советского ученого, организатора науки будет всегда жить в памяти его соратников и учеников.

TO THE MEMORY OF BORIS STEPANOVICH SOTSKOV

Professor Boris Stepanovich Sotskov, Doctor of Technical Sciences, corresponding member of the USSR Academy of Sciences, member of the Editorial Board of the journal "Problems of Control and Information Theory", has unexpectedly died on the 4th of November 1972.

The Soviet Science as well, as many scientific communities in touch with him outside of the USSR have lost a great scientist, teacher and man of public life, the founder and head of the Soviet scientific school in the theory of automation components, instruments and reliability, the initiator and soul of many long-lasting trends within Measurement- and Information Science and Technology.

Professor Sotskov is the author of more than 160 scientific publications, including seven books, on the theory and design of automation devices. These books, translated to several languages, are foundamental texts within the USSR as well as other socialist countries.

Professor Sotskov was one of the founders of the All-Union System of Industrial Automation Equipments and Instrumentation within the USSR.

He has made successful scientific investigations in the field of the failures within automation equipments and in reliability prediction techniques; and enabled in this way the design of high reliability automation equipments and instrumentation. He investigated also new ways of design of the elements of automation eguipments considering specifically what are called the double physical effects.

He was also one of the most active members of the scientific co-work with the Socialist Countries in the field of measuring equipment design as well as the foundation of the Unified Control System.

Professor Sotskov was one of the initiators of the journal „Problems of Control and Information Theory" and the chairman of the section dealing with devices.

His outstanding scientific competence, vivid interest and extreme working ability enabled also Professor Sotskov to contribute, in addition to his manysided scientific activity, also regularly to the design of specific measuring equipments.

His name will be remembered in the Scientific Community for many years.

Problems of Control and Information Theory, Vol. 1 (3—4), pp. 197—215 (1972)

СЛОЖНОСТЬ РЕАЛИЗАЦИИ АСИМПТОТИЧЕСКИ ОПТИМАЛЬНЫХ КОДОВ СХЕМАМИ ПОСТОЯННОЙ ГЛУБИНЫ

С. И. ГЕЛЬФАНД, Р. Л. ДОБРУШИН

(Москва)

(Поступила в редакцию 3 ноября 1971 г.)

Изучается сложность схем, осуществляющих линейное кодирование. Основной результат статьи состоит в следующем. Можно построить схему постоянной глубины, реализующую блоковый линейный код длины n со скоростью передачи R и кодовым расстоянием не менее dn, где $d < d_{\text{в.г}}$ и $d_{\text{в.г}}$ n – асимптотическая граница Варшамова–Гильберта, число элементов в которой имеет порядок $c_1 n \log n$, а глубина — порядок $c_2 \log n$, где c_1 и c_2 - константы, не зависящие от n.

В теории кодирования при сравнении методов кодирования и декодирования принято ориентироваться не только на параметры, характеризующие помехоустойчивость кода (например, кодовое расстояние), но и на сложность реализации алгоритмов кодирования и декодирования, причем эта сложность трактуется обычно как число элементарных действий, проводимых при реализации алгоритма. Поэтому представляется интересной задача об оптимизации кодов по параметру сложности при заданном значении параметра помехоустойчивости, которую можно надеяться исследовать лишь в асимптотической постановке. Постановка этой задачи требует, конечно, уточнения понятия сложности, что может быть сделано различными способами. В этой статье рассматривается реализация алгоритмов схемами из функциональных элементов (например, двувходовых сумматоров по модулю 2), причем предполагается, что схема не содержит обратных связей и элементов памяти. Такое предположение адекватно ситуации, когда на входе кодирующего устройства одновременно возникают все информационные символы сообщения. В качестве одного из исследуемых параметров сложности схемы выбирается ее глубина, определяемая как наибольшее число функциональных элементов, через которые проходит сигнал при его переработке схемой. Глубина схемы характеризует ее быстродействие. Вторым естественным параметром сложности является число функциональных элементов, образующих схему. Этот параметр исследуется в дополнительном предположении, что схема является схемой постоянной глубины, т. е. что на каждом пути сигнала от входа схемы к выходу он проходит одно и то же число элементов. Подобное ограничение естественно, если ориентироваться на функциональные элементы импульсного типа.

В этой работе мы ограничиваемся вопросом о сложности алгоритмов кодирования, оставляя в стороне более важный практически, но зато и более сложный вопрос о сложности алгоритма декодирования. Кодирование, которое интерпретируется как булевская функция, отображающая информационные символы сообщения в символы кодового слова, можно, как и любую булевскую функцию, реализовать многими разными схемами. Поэтому под сложностью кодирования будет пониматься минимальная сложность реализующей его схемы. Основной исследуемой в этой статье величиной является минимум сложностей кодирований в классе всех кодирований, задающих код с данной скоростью передачи и кодовым расстоянием, не меньшим данной константы.

В теории схем из функциональных элементов обычно используется функция Шеннона, равная максимуму сложности функций из некоторого класса функций. Таким образом, главное отличие состоит в том, что мы интересуемся самой легко реализуемой функцией из данного класса функций. Отметим, что подобные постановки задачи были одновременно предложены в недавних работах Севиджа [1], [2], который получил также оценки снизу, сходные с оценками раздела 3 этой статьи.

Основной результат этой статьи состоит в следующем. Можно построить схему, реализующую блоковый код длины n, со скоростью передачи R и кодовым расстоянием, не менее $d\,n$, где $d < d_{\text{в.г}}$ и $d_{\text{в.г}}\,n$ — асимптотическая граница Варшамова—Гильберта, число элементов в которой имеет порядок $c_1\,n \log n$, а глубина — порядок $c_2 \log n$, где c_1, c_2 — константы, зависящие от d и R. Оценки снизу, доказывающие невозможность дальнейших упрощений схемы, несложны. Оценка сверху основана на введении специального ансамбля групповых кодов, который с вероятностью, близкой к 1, состоит из кодов, обладающих нужными свойствами. Отметим для сравнения, что в ансамбле всех групповых кодов для почти всех кодов число элементов в минимальной схеме реализации имеет порядок $\dfrac{n^2}{\log n^2}$ [3]. Ситуация для ансамбля всех сверточных кодов (трактуемых как блоковые коды) такая же (если не допускать в схеме элементов памяти).

2. Основные определения

Определение 1. Двоичным сумматором называется логический элемент с двумя входами и произвольным числом выходов, реализующий функцию $x_1 \oplus x_2$ ($0 \oplus 0 = 1 \oplus 1 = 0, \quad 1 \oplus 0 = 0 \oplus 1 = 1$).

Определение 2. Схемой на сумматорах G называется направленный граф без циклов, в каждую вершину которого, кроме n отмеченных вершин $a_1, \ldots, a_n$,

входит два ребра; в вершины $a_1, \ldots, a_n$ не входит ни одного ребра. Кроме того в G выделены n вершин $b_1, \ldots, b_n$, из которых не выходит ни одного ребра. Вершины a_i называются входами G, вершины b_i — выходами G.

Определим то, что мы называем преобразованием, задаваемым схемой G. Пусть задан двоичный вектор $x = (x_1, \ldots, x_n)$, $x_i = 0,1$. Под состоянием схемы G, соответствующим входному вектору x, будем понимать задание для каждой вершины a схемы G числа $f(a)$, $f(a) = 0$, или 1, так, что выполнены следующие условия:

1. $f(a_i) = x_i$.

2. Если в вершину a входят ребра r' и r'', причем r' выходит из вершины a', а r'' — из вершины a'', то $f(a) = f(a') \oplus f(a'')$.

Легко видеть, что для каждого вектора $\mathbf{x}$ существует только одно состояние схемы G, соответствующее $\mathbf{x}$. Положим $\mathbf{y}_i = f(b_i)$ и $\mathbf{y} = (y_1, \ldots, y_n)$. Будем считать, что схема G преобразует входной вектор $\mathbf{x}$ в выходной вектор $\mathbf{y}$, и введем для этого обозначение $\mathbf{y} = G\,\mathbf{x}$.

Рассмотрим теперь следующее множество входных векторов $\mathbf{x}$. Зададим $m < n$ и рассмотрим множество $\mathfrak{X}_0$ из всех 2^m векторов, у которых $x_{m+1} = x_{m+2} = \ldots = x_n = 0$. Легко видеть, что множество векторов $\mathbf{y} = G\mathbf{x}$, $\mathbf{x} \in \mathfrak{X}_0$ является множеством кодовых векторов некоторого линейного двоичного кода $\mathfrak{K}$, имеющего скорость передачи, не превосходящую m/n. В такой ситуации можно сказать, что схема G реализует код $\mathfrak{K}$. Отметим, что различные схемы могут реализовать один и тот же код $\mathfrak{K}$.

Определение 3. $G(\mathfrak{K})$ — это множество всех схем, реализующих код $\mathfrak{K}$.

Определение 4. а) путем в схеме G называется направленный путь по графу, ведущий от входа к выходу. Длина пути — число составляющих его ребер;

б) схема G называется схемой постоянной глубины, если все пути в G имеют одну и ту же длину;

в) подмножество схем из $G(\mathfrak{K})$, состоящее из схем постоянной глубины обозначается через $G'(\mathfrak{K})$.

г) глубиной схемы называется длина самого длинного пути в G.

Определение 5. а) числом элементов $h(G)$ в схеме G называется число вершин в G.

3. Формулировка теорем и оценки снизу

Пусть $R = m/n$, где m и n — целые числа. Введем энтропию[1] $H(x) = -x \log x - (1-x) \log (1-x)$ и определим $d_{\text{в.г}}$, $0 < d_{\text{в.г}} < 1$ как решение уравнения $H(d_{\text{в.г}}) = 1-R$.

[1] Всюду в дальнейшем рассматриваются логарифмы по основанию 2.

Пусть $w(\mathbf{x})$ — число тех индексов i, $1 \leq i \leq n$, для которых $x_i = 1$. Напомним, что кодовым расстоянием $d(\mathfrak{A})$ линейного кода $\mathfrak{A}$ называется $\min w(\mathbf{y})$, где минимум берется по всем кодовым векторам $\mathbf{y}$, отличным от нулевого вектора. Тогда хорошо известно [4], что асимптотическая граница Варшамова—Гильберта $d_{\text{в.г}}\, n$ является асимптотической нижней границей для кодового расстояния $d(\mathfrak{A})$, оптимальных по кодовому расстоянию кодов $\mathfrak{A}$ длины n, имеющих скорость передачи R. Точнее, для любого $d < d_{\text{в.г}}$ и любого достаточно большого n существует код $\mathfrak{A}$ со скоростью передачи R, для которого $d(\mathfrak{A}) \geq dn$.

Определение 6. а) $h'(\mathfrak{A}) = \min\limits_{G \in G'(\mathfrak{A})} h(G)$,

б) $l(\mathfrak{A}) = \min\limits_{G \in G(\mathfrak{A})} l(G)$.

Выберем произвольное $d < d_{\text{в.г.}}$

Определение 7. а) $h'(R, dn) = \min h'(\mathfrak{A})$,

б) $l(R, dn) = \min l(\mathfrak{A})$;

минимум берется по всем кодам $\mathfrak{A}$ с $d(\mathfrak{A}) \geq dn$, имеющим скорость передачи R. Основным результатом статьи является следующая теорема.

Теорема 1. Существуют c_1, c_2 и c_1', c_2' такие, что при всех достаточно больших целых n

$$c_1 > \frac{h'(R, dn)}{n \log n} > c_1' \tag{3.1}$$

и

$$c_2 > \frac{l(R, dn)}{\log n} > c_2'. \tag{3.2}$$

Важное уточнение теоремы 1 состоит в том, что оценки сверху в (3.1) и (3.2) достигаются на одних и тех же схемах G. А именно, справедлива теорема 2.

Теорема 2. Для любого достаточно большого целого n существует схема G постоянной глубины, реализующая код $\mathfrak{A}$ длины n со скоростью передачи R $d(\mathfrak{A}) \geq dn$, для которого

$$\begin{aligned} h(G) &< c_1 n \log n, \\ l(G) &< c_2 \log n. \end{aligned} \tag{3.3}$$

Прежде, чем перейти к доказательству теорем, сделаем ряд замечаний.

Замечание 1. Не приводя точного значения констант c_1, c_2, c_1', c_2', отметим, что при доказательстве оценок снизу получаем, что $c_1' \geq R$, $c_2' \geq 1$.

С другой стороны, уточнение доказательства теоремы 2 позволяет увидеть, что $c_1 < 4$, $c_2 < 8$.

Замечание 2. Схемы, которые будут построены при доказательстве теоремы 2, обладают рядом полезных дополнительных свойств. В частности, они в некотором смысле однородны. Более подробно об этом говорится в замечании 5 в конце раздела 4.

Замечание 3. По-видимому, для лучшего приближения к технически реальной ситуации нужно ввести предположение, что число ребер, выходящих из каждой вершины схемы G тоже было ограничено. В этом случае можно доказать теорему, аналогичную теореме 1. Точнее, легко увидеть, что схемы, которые строятся для доказательства верхних оценок в теореме 1, можно видоизменить так, чтобы из каждой вершины выходило не больше двух ребер. При этом константы c_1, c_2 увеличатся не больше чем вдвое.

Замечание 4. Можно изучать схемы G на функциональных элементах в произвольном базисе [5], которые реализуют произвольный блоковый (не обязательно линейный) код $\mathcal{A}$. В этом случае справедлива теорема, аналогичная теореме 1. При этом доказательство верхней оценки не изменяется (если построить двоичный сумматор из элементов базиса). Доказательство нижней оценки изменяется незначительно.

Приведем доказательство нижних оценок в (3.1) и (3.2). Докажем сначала оценку в (3.2).

Будем говорить, что вход a_i и выход b_j схемы G связаны между собой, если в G существует путь, ведущий из a_i в b_j. Если кодовое расстояние кода $\mathcal{A}$, реализованного G, не меньше dn, то, рассматривая входные вектора веса 1, видим, что каждый вход a_i, $1 \leq i \leq m$ должен быть связан не меньше чем с dn выходами. Поэтому общее число пар (a_i, b_j), связанных между собой, не меньше $m\,dn = dRn^2$. Значит существует хотя бы один выход b_j, связанный больше чем с dRn входами. С другой стороны, число различных путей длины, не превосходящей l, ведущих из входов в этот выход b_j, не превосходит 2^l. Поэтому $2^l > dRn$, т. е. $l > \log(Rdn) > c_2' \log n$ для некоторого c_2'.

Докажем теперь нижнюю оценку в (3.1). Можно без ограничения общности предположить, что в схеме G единственными вершинами, из которых не выходит ни одного ребра, являются выходы b_j. После этого разобьем все вершины схемы на этажи, считая, что входы принадлежат нулевому этажу, и вершина a графа G принадлежит $i+1$-му этажу, если в нее приходит ребро из какой-нибудь вершины i-го этажа. Приведенное правило не приведет к противоречиям, поскольку рассматриваются схемы постоянной глубины. Общее число этажей, по доказанному ранее, не меньше $c_2' \log n$.

Пусть $G^{(i)} = (a_1^{(i)}, \ldots, a_{s_i}^{(i)})$ — множество вершин i-го этажа в схеме G, s_i — число элементов в $G^{(i)}$. Для каждого входного слова $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}_0$,

положим $x_j^{(i)} = f(a_j^{(i)})$ и $\mathbf{x}^{(i)} = (x_1^{(j)}, \ldots, x_{s_i}^{(i)})$, где f — состояние G, отвечающее входному вектору $\mathbf{x}$. Наше определение вершин i-го этажа показывает, что $\mathbf{x}^{(i+1)}$ зависит только от $\mathbf{x}^{(i)}$. Поэтому, если для двух различных входных векторов $\mathbf{x}$ и $\mathbf{x}'$ имеем $\mathbf{x}^{(i)} = \mathbf{x}'^{(i)}$ при каком-нибудь i, то $\mathbf{x}$ и $\mathbf{x}'$ переходят в один и тот же выходной вектор $\mathbf{y}$. Но это невозможно, поскольку $d(\mathcal{A}) > 0$. Так как общее число входных векторов равно 2^{Rn}, то $s_i \geq Rn$ при всех i. Поэтому $h(G) = \Sigma s_i \geq Rc_2' \, n \log n$. Таким образом, доказана нижняя оценка в (3.1) с $c_1' = c_2' \, R$.

4. Верхние оценки

Здесь приводится доказательство верхних оценок в формулах (3.1) и (3.2). Для этого необходимо построить простые схемы, реализующие хорошие коды. Используем вероятностный метод построения таких схем, построив множество $\mathfrak{F}$ схем, каждая из которых проста в смысле (3.1) и (3.2), и зададим на этом множестве $\mathfrak{F}$ распределение вероятностей таким образом, что среднее по этому множеству кодовое расстояние кодов, реализуемых построенными схемами, достаточно велико. Множество схем с заданным на нем распределением вероятностей будем называть ансамблем схем.

В дальнейшем для построения схем будет удобно использовать два вспомогательных понятия.

Двоичным t-сумматором назовем логический элемент с t входами и произвольным числом выходов, который реализует функцию $x_1 \oplus x_2 \oplus \ldots \oplus x_t$, где x_i — значения на входах t-сумматора. В этом смысле сумматор, введенный в определении 1, является 2-сумматором. Ясно, что t-сумматор можно получить соединением $(t-1)$ 2-сумматора (рис. 1). При этом глубина t-сумматора не превосходит[2] $[\log t]+1$. Если же $t = 2^s$, где s — целое, то t-сумматор является схемой постоянной глубины, глубина которой равна s. В дальнейшем под t-сумматором будем понимать схему, составленную из 2-сумматоров.

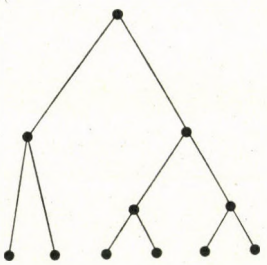


Рис. 1. Схема t-сумматора $(t = 6)$

[2] $[x]$ — целая часть x.

Вторым полезным понятием является понятие композиции схем. Пусть G_1 — схема с n входами и n выходами $b_1, \ldots, b_n$, G_2 — схема с n входами $a_1', \ldots, a_n'$ и n выходами. Тогда можно отождествить a_i' с b_i и получить новую схему G с n входами и n выходами, которую будем называть композицией G_1 и G_2, и обозначать $G_1 \circ G_2$. Ясно, что $h(G_1 \circ G_2) = h(G_1) + h(G_2) - n$, $l(G_1 \circ G_2) \leq \leq l(G_1) + l(G_2)$ и если G_1, G_2 — схемы постоянной глубины, то $G_1 \circ G_2$ — тоже схема постоянной глубины.

Перейдем теперь к построению ансамбля $\mathfrak{F}$. Для этого построим вспомогательный ансамбль $\mathfrak{E}$, свойства которого описываются теоремой 3.

Теорема 3. Существуют такие константы $0 < \alpha < \min(d, 1-R)$, $a > 1$, $\beta > 1$, $b > 0$, $\varepsilon > 0$, и целое s, что для всех достаточно больших n можно построить ансамбль $\mathfrak{E} = \mathfrak{E}_n(\alpha, a, \beta, b, \varepsilon, s)$ схем с n входами и n выходами, обладающий следующими свойствами:

1) $h(G) \leq bn$ для всех $G \in \mathfrak{E}$;
2) все схемы $G \in \mathfrak{E}$ имеют одну и ту же постоянную глубину s;
3) для любого вектора $\mathbf{x}$ такого, что $w(\mathbf{x}) = w \leq \alpha n$, вероятность $Pr\{w(G\mathbf{x}) < \beta w\} \leq (C_n^w)^{-1} n^{-1/2} a^{-w}$;
4) для любого вектора $\mathbf{x}$ такого, что $(1-\alpha)\, n \geq w(\mathbf{x}) \geq \alpha n$, вероятность $Pr\{w(G\mathbf{x}) < dn\} < 2^{-n(R+\varepsilon)}$.
5) для любого вектора $\mathbf{x}$ такого, что $w(\mathbf{x}) \leq (1-\alpha)\, n$, вероятность $Pr\{w(G\mathbf{x}) > (1-d)\, n\} < 2^{-n(R+\varepsilon)}$.

Покажем, как из теоремы 3 вывести теорему 2, а значит и верхнюю оценку в теореме 1. Зафиксируем некоторое множество[3] $\mathfrak{X}$, $|\mathfrak{X}| \leq 2^{nR}$, состоящее из векторов, вес каждого из которых больше αn, и меньше $(1-\alpha)n$ и обозначим через $T(\mathfrak{X}) = T$ следующее событие

$$T = \{w(G\mathbf{x}) < \beta w(\mathbf{x}) \quad \text{для некоторого } \mathbf{x} \text{ такого, что } w(\mathbf{x}) \leq d\, n\} \cup,$$
$$\cup = \{w(G\mathbf{x}) < dn \quad \text{для некоторого } \mathbf{x} \in \mathfrak{X}\} \cup,$$
$$\cup = \{w(G\mathbf{x}) > (1-d)n \text{ для некоторого } \mathbf{x} \in \mathfrak{X}\}.$$

Поскольку число векторов веса w равно C_n^w, пункты 3)—5) теоремы 3 показывают, что

$$Pr\{T\} < (a-1)^{-1} n^{-1/2} + 2 \cdot 2^{-n\varepsilon} < cn^{-1/2} \tag{4.1}$$

для некоторого c, не зависящего от n.

Пусть задано целое число k. Построим ансамбль схем $\mathfrak{F}_k = \mathfrak{E}^k$ следующим образом. Каждая схема $G \in \mathfrak{F}_k$ является композицией $G_1 \circ G_2 \circ \ldots \circ G_k$ k схем из $\mathfrak{E}$. Распределение вероятностей на $\mathfrak{F}_k$ определяется тем, что схемы G_i

[3] $|\mathfrak{X}|$ — число элементов конечного множества $\mathfrak{X}$.

выбираются статистически независимо друг от друга, и каждая из них имеет распределение, задаваемое ансамблем $\mathscr{E}$. Изучим такой ансамбль $\mathscr{F}_k$. Обозначим через T_k следующее событие в ансамбле $\mathscr{F}_k$.

$$T_k = \{w(G\mathbf{x}) < \min [\beta^k w(\mathbf{x}), dn] \text{ для некоторого } \mathbf{x} \in \mathscr{X}_0\}.$$

Для каждого i, $1 \leq i \leq k$ обозначим через $\mathscr{X}_i$ множество векторов $\mathbf{y} = G_1 \circ \ldots \circ G_i\, \mathbf{x}$, $\mathbf{x} \in \mathscr{X}_0$ таких, что $(1-\alpha)n \geq w(\mathbf{y}) \geq \alpha n$. Тогда мы можем применить к каждому множеству $\mathscr{X}_i$ формулу (4.1) и получить, что $Pr(T_k) \leq kcn^{-1/2}$. Далее при $k \geq k_0 = \left[\dfrac{\log dn}{\log \beta}\right] + 1$ имеем $\beta^k\, w(\mathbf{x}) > dn$ при $\mathbf{x} \neq 0$, поскольку $w(\mathbf{x}) \geq 1$. Поэтому при $k \geq k_0$ событие T_k означает, что $w(G\mathbf{x}) < dn$ для некоторого $\mathbf{x} \in \mathscr{X}_0$, $\mathbf{x} \neq 0$. При этом $Pr\{T_{k_0}\} \leq ck_0 n^{-1/2} < 1/2$ при достаточно больших n. Таким образом при достаточно больших n ансамбль $\mathscr{F} = \mathscr{F}_{k_0}$ таков, что в нем с вероятностью большей $1/2$ (и даже с вероятностью, стремящейся к 1 при $n \to \infty$) найдется схема G, реализующая код $\mathscr{A}$ с $d(\mathscr{A}) \geq dn$.

С другой стороны ясно, что все схемы $G \in \mathscr{F}$ имеют постоянную глубину и $h(G) < bnk_0 < c_1\, n \log n$, $l(G) < sk_0 < c_2 \log n$. Таким образом доказательство теоремы 2 и окончание доказательства теоремы 1 свелись к теореме 3.

Замечание 5. Приведенное рассуждение показывает, что схема G, которую мы строим для доказательства теоремы 2, является композицией k_0 схем $G_1, \ldots, G_{k_0}$, выбираемых из ансамбля $\mathscr{E}$. Легко показать, что схемы $G_1, \ldots, G_{k_0}$ можно считать совпадающими. Точнее, можно найти такую схему $G_0 \in \mathscr{E}$, что схема $G_0 \circ G_0 \circ \ldots \circ G_0$ (k_0 раз) удовлетворяет теореме 2.

5. Построение ансамбля $\mathscr{E}$

Построим теперь нужный ансамбль $\mathscr{E}$. Все схемы G из $\mathscr{E}$ имеют n входов $a_1, \ldots, a_n$ и n выходов $b_1, \ldots, b_n$. Из каждого входа a_i схемы G выходит по t ребер $r_i^{(1)}, \ldots, r_i^{(t)}$. В каждом выходе b_i схемы G находится t-сумматор Σ_i. Пусть $\bar{r}_i^{(1)}, \ldots, \bar{r}_t^{(t)}$ — ребра, входящие в Σ_i. Занумеруем все ребра $r_i^{(j)}$ и $\bar{r}_i^{(j)}$ числами от 1 до nt, полагая $r_i^{(j)} = r(t(i-1)+j)$, $\bar{r}_j^{(i)} = \bar{r}(t(i-1)+j)$. Пусть σ — подстановка чисел $1, \ldots, nt$. Отождествляя ребро $\bar{r}(k)$ с ребром $r(\sigma(k))$ получим схему $G = G(\sigma)$. Ансамбль $\mathscr{E}$ состоит из схем $G(\sigma)$ для всех σ, причем все подстановки σ считаются равновероятными.

Покажем, что построенный ансамбль $\mathscr{E}$ при некотором t удовлетворяет условиям теоремы 3 при некоторых $\alpha, \beta, a, b, \varepsilon, s$. Ясно, что $h(G) = n(t-1)$ для всех $G \in \mathscr{E}$ и если $t = 2^s$, то $l(G) = s$, и любая схема $G \in \mathscr{E}$ имеет постоянную глубину.

В дальнейшем будем использовать следующую теорему.

Теорема 4. Пусть $\xi_1, \ldots, \xi_n$ — случайные величины с конечным числом значений, $S_k = \sum_{i=1}^{k} \xi_i$, A — некоторое событие. Пусть далее условная вероятность $Pr\{\xi_{k+1} > v \mid \xi_1 = v_1, \ldots, \xi_k = v_k, A\} \geq \Phi(v)$ для всех $0 \leq k \leq n-l$, v_k и v, где $\Phi(v) > 0$ монотонно невозрастающая функция. Тогда при всех k

$$Pr\{S_k > v\} \geq \Phi^{*(k)}(v)\, Pr\{A\},$$

где

$$\Phi^{*(1)}(v) = \Phi(v),\ \Phi^{*(k+1)}(v) = \int \Phi^{*(k)}(w)\, d\Phi(v-w).$$

Доказательство. Заметим, что

$$Pr\{S_k > v\} \geq Pr\{S_k > v \mid A\}\, Pr\{A\}.$$

Поэтому для доказательства теоремы 4 достаточно доказать, что при всех k

$$Pr\{S_k > v\} \geq \Phi^{*(k)}(v). \tag{5.1}$$

Формула (5.1) будет доказываться индукцией по k. При $k = 1$ она очевидна. Пусть формула (5.1) справедлива для некоторого k. Тогда

$$Pr\{S_{k+1} > v \mid A\} = \sum_{v_1, \ldots, v_k} Pr\{\xi_{k+1} > v - v_1 - \ldots - v_k \mid \xi_1 = v_1, \ldots, \xi_k = v_k, A\}.$$

$$Pr\{\xi_1 = v_1, \ldots, \xi_k = v_k \mid A\} \geq \sum_{w} \Phi(v-w)\, Pr\{S_k = w \mid A\} =$$

$$= \int Pr\{S_k > w \mid A\}\, d\Phi(v-w) \geq \int \Phi^{*(k)}(w)\, d\Phi(v-w) = \Phi^{*(k+1)}(v).$$

Поэтому формула 5.1 верна для $k+1$ и теорема 4 доказана.

Пусть задано входное слово $\mathbf{x} = (x_1, \ldots, x_n)$, $w(\mathbf{x}) = w$. Назовем ребра, выходящие из вершин a_i, для которых $x_i = 1$, отмеченными. Ясно, что их число равно wt. Введем на $\mathcal{E}$ следующие случайные величины:

$\xi_i(\mathbf{x}) = y_i$ — состояние выхода b_i схемы G,

$\eta_i(\mathbf{x})$ — число отмеченных ребер среди ребер $\bar{r}(it+1), \ldots \bar{r}(it+t)$

(Ясно, что $\xi_i(\mathbf{x}) \equiv \eta_i(\mathbf{x}) \pmod 2$),

$$S_k(\mathbf{x}) = \sum_{i=1}^{k} \xi_i(\mathbf{x}),$$

$$v_k(\mathbf{x}) = \sum_{i=1}^{k} \eta_i(\mathbf{x})$$

(Иногда будем опускать аргумент $\mathbf{x}$ у величин ξ, η, S, v).

Одной из основных теорем, позволяющих установить свойства 3—5 ансамбля $\mathcal{E}$, является теорема 5.

Положим

$$H(\lambda, p) = H(\lambda) + \lambda \log p + (1 - \lambda) \log (1 - p),$$

$$P_t(\omega) = \frac{1}{2} - \frac{1}{2}(1 - 2\omega)^t.$$

Теорема 5. Для любого λ_1, $0 < \omega_1 < 1/2$, любого $\varepsilon_1 > 0$ и $\mathbf{x}$, $w(\mathbf{x}) = \omega n$ существует t_1 такое что [4]

$$Pr\{S_n(\mathbf{x}) \leq l\} \leq \exp_2 \{n [H(\lambda, p) + \varepsilon_1]\},$$

где $\lambda_1 < \lambda = l/n < p = p_t(\omega)$, при всех $t > t_1$, достаточно больших n и $1 - \omega_1 > > \omega > \omega_1$.

Теорема 5 опирается на ряд лемм. Доказательство этих лемм будет приведено в разделе 5.

Зафиксируем $\gamma > 0$ и введем события $\tilde{B}_i^\gamma$, $\tilde{\tilde{B}}_i^\gamma$, B_i^γ $(1 \leq i \leq n)$ на $\mathcal{E}$:

$$\tilde{B}_i^\gamma = \left\{\nu_i \geq \left(\frac{tw}{n} - t\gamma\right) i\right\}; \quad \tilde{\tilde{B}}_i^\gamma = \left\{\nu_i \leq \left(\frac{tw}{n} + t\gamma\right) i\right\}; \quad B_i^\gamma = \tilde{\tilde{B}}_i^\gamma \cap \tilde{B}_i^\gamma.$$

Если $\varepsilon' > 0$, то положим $A^\gamma = A^{\gamma, \varepsilon'} = \bigcap_{i=\varepsilon' n}^{(1-\varepsilon')n} B$.

Лемма 1. Пусть $\omega = w/n$. Тогда для любых $\varepsilon_2 > 0$ и $\varepsilon' > 0$ существуют γ и t_0 такие, что при всех достаточно больших n

$$|Pr\{\xi_{k+1}(\mathbf{x}) = 1 \mid \xi_1(\mathbf{x}) = \nu_1, \ldots, \xi_k(\mathbf{x}) = \nu_k, A^{\gamma, \varepsilon'}\} - p_t(\omega)| < \varepsilon_2$$

при всех $\nu_1, \ldots, \nu_k$, всех $t > t_0$ и $(1 - \varepsilon')n > k > \varepsilon' n$.

Лемма 1 позволяет применить к случайным величинам $\xi_{\varepsilon' n}, \ldots, \xi_{(1-\varepsilon')n}$ теорему 4. Используя ее, получаем следующую формулу:

$$Pr\left\{\sum_{i=\varepsilon' n}^{(1-\varepsilon')n} \xi_i \leq l\right\} \leq 1 - Pr\{A^{\gamma, \varepsilon'}\} + \\ + \sum_{i=0}^{l} C_{(1-2\varepsilon')n}^i (p_t(\omega) - \varepsilon_2)^i (1 - p_t(\omega) + \varepsilon_2)^{n(1-2\varepsilon')-i}. \qquad (5.1)$$

Лемма 2. Для любого $\theta < 0$, любого ε' и любого γ существует такое t_2, что при $t > t_2$ и всех достаточно больших n

$$Pr\{A^{\gamma, \varepsilon'}\} \geq 1 - \exp_2 n\theta.$$

[4] $\exp_2 x = 2^x$

При $\lambda = l/k < p$ имеет место хорошо известная формула [4], [6]

$$\sum_{i=0}^{l} C_k^i p^i (1-p)^{k-i} \leq \exp_2 \{kH(\lambda, p)\}. \tag{5.2}$$

Ясно, далее, что $S_n(\mathbf{x}) \geq \sum_{i=\varepsilon' n}^{(1-\varepsilon')n} \xi_i$. Кроме того, величина $H(\lambda, p)$ ограничена снизу при $\lambda < \lambda_1$, $p > p_t(\omega_1) > \omega_1$. Поэтому теорема 5 легко следует из формул (5.1), (5.2) и леммы 2.

Аналогично теореме 5 доказывается теорема 6.

Теорема 6. Для любых $1/2 > \omega_1 > 0$, λ_1, $\varepsilon_3 > 0$ и $\mathbf{x}$, $w(\mathbf{x}) = \omega n$, найдется t_3 такое, что

$$Pr\{S_n \geq l\} \leq \exp_2 \{n[H(\lambda, p) + \varepsilon_3]\},$$

где $\lambda_1 > \lambda = l/n > p_t(\omega)$ при $t > t_3$, $1 - w_1 > \omega$ и достаточно больших n.

Напомним теперь, что задано число $d < d_{\text{в.г}}$. Тогда существуют такие $\tilde{p} < 1/2$ и $\varepsilon > 0$, что

$$H(\lambda, p) < -R - \varepsilon \text{ при } \lambda < d \text{ и всех } p, \text{ для которых } |\tilde{p} - 1/2| \geq |p - 1/2|.$$

Выберем произвольное α такое, что $\alpha < d$ и $\alpha < 1 - R$. Поскольку $p_t(\omega) \to 1/2$ при $t \to \infty$ равномерно по ω, $1 - \alpha > \omega > \alpha$, существует такое $t = 2^s$, что $p_t(\omega) > \tilde{p}$ при всех ω, $1 - \alpha > \omega > \alpha$.

Если рассмотреть теперь ансамбль $\mathscr{E} = \mathscr{E}_t$, то теорема 5 обеспечивает выполнение свойства 4, а теорема 6—выполнение свойства 5 для тех $\mathbf{x}$, для которых $w(\mathbf{x}) = \omega n$, $1 - \alpha > \omega > \alpha$. Заметим, что $p_t(\omega)$ убывает, если $\omega < 1/2$ и ω убывает. Поэтому теорема 6 обеспечивает выполнение свойства 5 и при $\frac{1-d}{t} < \omega < \alpha$. Заметим, что всегда $w(G\mathbf{x}) < tw(\mathbf{x})$. Поэтому свойство 5 выполнено.

В дальнейшем выбранное t зафиксируем. Можно предположить, что $t > 12$.

Перейдем теперь к проверке свойства 3. Выполнение свойства 3 ансамбля при малых весах w обеспечит следующая лемма.

Лемма 3. Существует такое ω_1, что для любого $w < \omega_1 n$ и для каждого $\mathbf{x}$, $w(\mathbf{x}) = w$ с вероятностью большей $1 - C_n^w a^{-w} n^{-1/2}$ при достаточно больших n выполнено следующее событие

$$V = \left\{\text{существует не меньше } \frac{1}{2} wt \text{ значений } i, \text{ для которых } \eta_i(\mathbf{x}) = 1\right\}.$$

Для проверки свойства 3 нужно оценить $C_n^w Pr\{S_k(\mathbf{x}) < l\}$, где $w = w(\mathbf{x})$. Для этого отметим, что если $w = \omega n$, то при $\omega_1 < \omega < \alpha$ имеет место формула

$$C_n^w < c_3 \exp_2 \{n H(\omega)\}, \tag{5.3}$$

где c_3 не зависит от w и n. Из теоремы 5 и формулы (5.3) получаем

$$C_n^w Pr\{S_n(\mathbf{x}) < l\} \leq \exp_2 \{n [H(\omega) + H(\lambda, p_t(\omega)) + \varepsilon_4\},$$

где $\lambda = l/n$, $\omega = w/n$, ε_4^{ψ} стремится к 0 при $n \to \infty$ равномерно по ω при $\omega_1 < \omega < \alpha$.

Лемма 4. Пусть $\lambda = \lambda_t(\omega)$ — наименьший положительный корень уравнения $F(\lambda, \omega) = H(\omega) + H(\lambda, p_t(\omega)) = 0$. Тогда

1) $F(\lambda, \omega) < 0$ при $\lambda < \lambda_t(\omega)$.

2) $\lambda_t(\omega) \geq \tau\omega$ для некоторого $\tau > 1$ при $\omega_1 < \omega < \alpha$.

Доказательство леммы 4 сводится к непосредственной проверке.

Если теперь положить $\beta = \min(\tau, t/2)$, то выполнение свойства 3 сразу следует из лемм 3 и 4. Таким образом теорема 3 полностью доказана.

6. Доказательства лемм

Доказательство леммы 1. Пусть задан входной вектор $\mathbf{x}$, $w(\mathbf{x}) = w = \omega n$. Введем для упрощения записи следующие обозначения. Для каждого k положим

$$\boldsymbol{\xi}^{(k)} = (\xi_1, \ldots, \xi_k), \ \boldsymbol{\eta}^{(k)} = (\eta_1, \ldots \eta_k), \ \mathbf{v}^{(k)} = (v_1, \ldots, v_k), \ v_i = 0,1,$$

$\mathrm{y}^{(k)} = (y_1, \ldots, y_k)$, y_i — целое число, $0 \leq y_i \leq t$.

Нужно найти $Pr\{\xi_{k+1} = 1 \,|\, \boldsymbol{\xi}^{(k)} = \mathbf{v}^{(k)}, A^{\gamma}\}$. Положим

$$A^{\gamma}_{\leq k} = \bigcap_{i=\varepsilon' n}^{k} B_i^{\gamma}, \quad A^{\gamma}_{>k} = \bigcap_{i=k+1}^{(1-\varepsilon')n} B_i^{\gamma}.$$

Докажем следующую лемму.

Лемма 1а. При заданных ε', ω_1 для любого $\varepsilon_5 > 0$ существует такое γ (не зависящее от t), что

$$|Pr\{\xi_{k+1} \,|\, \boldsymbol{\xi}^{(k)} = \mathbf{v}^{(k)}, A^{\gamma}_{\leq k}\} - p_t(\omega)| < \varepsilon_5$$

при всех достаточно больших n равномерно по t, k, $(1 - \varepsilon')\, n \geq k \geq \varepsilon' n$, $1 - \omega_1 > \omega > \omega_1$.

Применение формулы полной вероятности показывает, что для доказательства леммы 1 [a] достаточно оценить условную вероятность

$$Pr\{\xi_{k+1} = 1 \mid \sigma(1) = g_1, \ldots, \sigma(kt) = g_{kt}\} = p(\mathbf{g})$$

при всех $\mathbf{g} = (g_1, \ldots, g_{kt})$.

Пусть $\mathfrak{R}$ — множество ребер $r(j)$, «оставшихся свободными», т.е. множество тех ребер $r(j)$, что $j \neq g_i$ для $i = 1, \ldots, kt$. Тогда $|\mathfrak{R}| = (n-k)\,t$, и если λ_k — доля отмеченных ребер в $\mathfrak{R}$, то из условия $A^{\gamma}_{\leq k}$ и из того, что $(1-\varepsilon')\,n > k > \varepsilon' n$ следует, что

$$|\lambda_k - \omega| < \gamma/\varepsilon'.$$

Пусть $p_l(\mathbf{g}) = Pr\{\eta_{k+1} = l \mid \sigma(1) = g_1, \ldots, \sigma(kt) = g_{kt}, A^{\gamma}_{\geq k}\}$. Покажем, что при любом $\mathbf{g} = (g_1, \ldots, g_{kt})$

$$p_l(\mathbf{g}) - C_t^l \lambda_k^l (1-\lambda_k)^{t-l} \leq 1 - \frac{[(n-k)\,t]!}{[(n-k-1)\,t]!\,n^t}. \tag{5.4}$$

В самом деле, для получения η_{k+1} нужно найти число отмеченных ребер среди t ребер, выбираемых без возвращения из множества $\mathfrak{R}$. Пусть $\bar{p}_l(\mathbf{g})$ — аналогичная $\bar{p}_l(\mathbf{g})$ вероятность для выбора с возвращением, и θ — вероятность того, что при выборе с возвращением хотя бы один элемент из $\mathfrak{R}$ будет выбран больше одного раза. Тогда $p_l(\mathbf{g}) \leq \bar{p}_l(\mathbf{g})\,(1-\theta)$, $\bar{p}_l(\mathbf{g}) = C_t^l\,\lambda_k^l\,(1-\lambda_k)^{t-l}$ и нетрудно показать, что

$$\theta \leq 1 - \frac{[(n-k)\,t]\,!}{[(n-k-1)\,t]\,!\,n^t}.$$

Отсюда следует формула (5.4). Поэтому при $\varepsilon' n < k < (1-\varepsilon')\,n$

$$p_l(\mathbf{g}) - C_t^l \lambda_k^l (1-\lambda_k)^{t-l} = o(n)$$

равномерно по k. Далее,

$$p(\mathbf{g}) = \sum_{j=1}^{t/2} p_{2j}(\mathbf{g}) = \sum_{j=0}^{t/2} C_t^{2j}\,\lambda_t^{2j}(1-\lambda_k)^{t-2j} + o_1(n) = \frac{1}{2} - \frac{1}{2}(1-2\lambda_k)^t + o_1(n),$$

где $o_1(n) \to 0$ при $n \to \infty$. Выбирая γ таким, чтобы правая часть отличалась от $p_t(\omega)$ на $\varepsilon_5/2$ при всех ω, $1 - \omega_1 > \omega > \omega_1$, получаем, что при достаточно n выполнено неравенство леммы 1[a].

Продолжим доказательство леммы 1. Для каждого $\mathbf{y}^{(k)}$ обозначим через $D(\mathbf{y}^{(k)})$ событие

$$D(\mathbf{y}^{(k)}) = \{\boldsymbol{\eta}^{(k)} = \mathbf{y}^{(k)}\}.$$

По формуле полной вероятности для доказательства леммы 1 достаточно оценить

$$Pr\{\xi_{k+1} = 1 \mid \xi^{(k)} = \mathbf{v}^{(k)},\ A^{\gamma},\ D(\mathbf{y}^{(k)})\}$$

для тех векторов $\mathbf{y}^{(k)}$, которые совместимы с $\mathbf{x}^{(k)}$ и $A^{\gamma}_{\leq k}$, т. е. для которых $v_i = y_i \pmod 2$ и $v_i = \sum_{j=i}^{i} y_j$ удовлетворяют неравенству

$$\left(\frac{tw}{n} - t\gamma\right) i \leq v_i \leq \left(\frac{tw}{n} - t\gamma\right) i$$

при $k \geq i \geq \varepsilon' n$. Для таких векторов $\mathbf{y}^{(k)}$ имеем, очевидно,

$$Pr\{\xi_{k+1} = 1 \mid \xi^{(k)} = \mathbf{v}^{(k)},\ A^{\gamma},\ D(\mathbf{y}^{(k)}) = Pr\{\xi_{k+1} = 1 \mid D(\mathbf{y}^{(k)}),\ A^{\gamma}_{>k}\}.$$

Лемма 1[a] показывает, что

$$|\, Pr\{\xi_{k+1} = 1 \mid D(\mathbf{y}^{(k)})\} - p_t(\omega)\,| < \varepsilon_5\,, \tag{5.5}$$

если $\mathbf{y}^{(k)}$ совместимо с $A^{\gamma}_{\leq k}$. Поэтому нужно оценить

$$\Lambda(\mathbf{y}^{(k)}) = \frac{Pr\{\xi_{k+1} = 1 \mid D(\mathbf{y}^{(k)}),\ A^{\gamma}_{>k}\}}{Pr\{\xi_{k+1} = 1 \mid D(\mathbf{y}^{(k)})\}}$$

для тех $\mathbf{y}^{(k)}$, которые совместимы с A^{γ}.

Надо доказать, что для каждого γ и каждого ε_6 можно выбрать такое t_4, что для всех $t > t_4$ в ансамбле $\mathscr{E} = \mathscr{E}_t$ выполнено неравенство

$$|\, 1 - \Lambda(\mathbf{y}^{(k)}\,| < \varepsilon_6$$

при всех $\mathbf{y}^{(k)}$, совместимых с $A^{\gamma}_{\leq k}$, и всех достаточно больших n. Обозначим $Pr\{A^{\gamma}_{>k} \mid \xi_{k+1} = v,\ D(\mathbf{y}^{(k)})\} = q_v,\ v = 0,1$. Тогда

$$\Lambda(\mathbf{y}^{(k)}) = \frac{q_1}{Pr\{A^{\gamma}_{>k} \mid D(\mathbf{y}^{(k)})\}} =$$

$$= \frac{q_1}{q_1\, Pr\{\xi_{k+1} = 1 \mid D(\mathbf{y}^{(k)})\} + q_0\,[1 - Pr\{\xi_{k+1} = 0 \mid D(\mathbf{y}^{(k)})\}}\,. \tag{5.6}$$

Оценим $q_1 - q_0$ Для этого воспользуемся тем, что

$$\{\xi_{k+1} = 0\} = \bigcup_{j=0}^{t/2} \{\eta_{k+1} = 2j\}\,,$$

$$\{\xi_{k+1} = 1\} = \bigcup_{j=1}^{t/2} \{\eta_{k+1} = 2j - 1\}$$

и тем, что для любых событий U, V и V_i таких, что $\bigcup_i V_i = V$ имеет место формула

$$Pr\{U \mid V\} = \sum_i Pr\{U \mid V_i\} Pr\{V_i \mid V\}.$$

Получим

$$q_1 - q_0 = \sum_{j=1}^{t/2} (p_{2j} \cdot a_{2j} - p_{2j-1} a_{2j-1}) + p_0 a_0,$$

где

$$a_{2j} = Pr\{\eta_{k+1} = 2j \mid \xi_{k+1} = 0, D(\mathbf{y}^{(k)})\};\ a_{2j-1} =$$

$$= Pr\{\eta_{k+1} = 2j - 1 \mid \xi_{k+1} = 1, D(\mathbf{y}^{(k)})\};\ p_l = Pr\{A^{\gamma}_{>k} \mid \eta_{k+1} = l, D(\mathbf{y}^{(k)})\}.$$

Изучим отдельно величины p_l и a_l.

Лемма 1^{б}. Для любого набора $\mathbf{y}^{(k)}$ при достаточно больших n имеем

$$\sum_{l=0}^{t-1} |p_{l+1} - p_l| < 2.$$

Для доказательства леммы 1^{б} положим

$$\tilde{A}^{\gamma}_{>k} = \bigcap_{i=k+1}^{(1-\varepsilon')n} \tilde{B}^{\gamma}_i,\ \tilde{\tilde{A}}^{\gamma}_{>k} = \bigcap_{i=k+1}^{(1-\varepsilon')n} \tilde{\tilde{B}}^{\gamma}_i$$

и определим $\tilde{p}_l$, $\tilde{\tilde{p}}_l$ аналогично p_l с заменой $A^{\gamma}_{>k}$ на $\tilde{A}^{\gamma}_{>k}$ и $\tilde{\tilde{A}}^{\gamma}_{>k}$, соответственно. Тогда легко видеть, что

$$\tilde{p}_{l+1} \leq \tilde{p}_l;\ \tilde{\tilde{p}}_{l+1} \geq \tilde{\tilde{p}}_l.$$

Кроме того, рассуждения, аналогичные проводимым при доказательстве леммы 2, приводят к следующему.

Если $v_k \geq \frac{wt}{n} k$, то $\tilde{\tilde{p}}_l \to 1$ при $n \to \infty$ равномерно по l.

Если $v_k \leq \frac{wt}{n} k$, то $\tilde{p}_l \to 1$ при $n \to \infty$ равномерно по l.

Далее, поскольку $A^{\gamma}_{>k} = \tilde{A}^{\gamma}_{>k} \cap \tilde{\tilde{A}}^{\gamma}_{>k}$, то

$$\min(\tilde{p}_l, \tilde{\tilde{p}}_l) \geq p_l \geq \tilde{p}_l \tilde{\tilde{p}}_l.$$

Поэтому лемма 1б сразу следует из указанных выше свойств $\tilde{p}_l$, $\tilde{\tilde{p}}_l$.

Лемма 1^{в}. $a_l \to 0$ при $t \to \infty$ равномерно по l и n при достаточно больших n.

Доказательство леммы 1^{в} проводится непосредственным вычислением нужных вероятностей аналогично тому, как это было сделано в лемме 1^{а}.

Наконец, воспользуемся следующим тривиальным замечанием. Если последовательности p_l, a_l таковы, что $\sum_l |p_{l+1} - p_l| < 2$, $a_l < \delta$ для всех l, то

$$\Sigma |p_{2j} a_{2j} - p_{2j-1} a_{2j-1}| < 2\delta .$$

Используя это замечание, получаем, что

$$q_1 - q_0 \to 0 \text{ при } t \to \infty . \tag{5.7}$$

Далее, ввиду формулы (5.5) и поскольку $p_t(\omega)$ возрастает вместе с t, существует такая константа c_4, что

$$1 - c_4 > Pr_{\mathbf{y}} \{\xi_{k+1} = 1 \mid D(\mathbf{y}^{(k)})\} < c_4 . \tag{5.8}$$

Поэтому формулы (5.6), (5.7) и (5.8) дают

$$|1 - \Lambda(\mathbf{y}^{(k)})| \to 0 \text{ при } t \to \infty \text{ равномерно по } \mathbf{y}^{(k)} \tag{5.9}$$

для всех $\mathbf{y}^{(k)}$, совместимых с A^γ и всех достаточно больших n. Выбирая t достаточно большим, выводим лемму 1 из леммы 1[a] и неравенства (5.9).

Доказательства леммы 2. Нужно найти $p'_\gamma = Pr\{A^{\gamma,\varepsilon'}\}$. Напомним, что $A^{\varphi,\varepsilon'} = \bigcap_{i=\varepsilon' n}^{(1-\varepsilon')n} B_i^\gamma$. Поэтому[5] $1 - p'_\gamma \leq \sum_{i=\varepsilon' n}^{(1-\varepsilon')n} Pr\{\bar{B}_i^\gamma\}$. Событие $\bar{B}_i^\gamma$ состоит в том, что в первые i выходов попало больше $\left(\frac{wt}{n} + \gamma t\right) i$ или меньше $\left(\frac{wt}{n} - \gamma t\right) i$ отмеченных отрезков, т. е. число отмеченных ребер среди ребер $r(1), \ldots, r(it)$ больше $\left(\frac{wt}{n} + \gamma t\right) i$ или меньше $\left(\frac{wt}{n} - \gamma t\right) i$.

Пусть событие F_i^k состоит в том, что число отмеченных ребер среди $r(1), \ldots, r(it)$ равно k. Тогда $\bar{B}_\gamma^i = \left(\bigcup_{k=0}^{\left(\frac{wt}{n} - \gamma t\right) i} F_i^k\right) \cup \left(\bigcup_{k=\left(\frac{wt}{n} + \gamma t\right) i}^{wt} F_i^k\right)$, и значит

$$Pr\{B_\gamma^i\} = \sum_{k=0}^{\left(\frac{wt}{n} - \gamma t\right) i} Pr\{F_i^k\} + \sum_{k=\left(\frac{wt}{n} + \gamma t\right) i}^{wt} Pr\{F_i^k\} = P_1 + P_2 .$$

Далее очевидно, что

$$Pr\{F_i^k\} = \frac{C_{wt}^k C_{(n-w)t}^{it-k}}{C_{nt}^{it}} .$$

[5] $\bar{T}$ — отрицание события T.

Оценим, например, P_1. Полагая $w/n = \omega$, $k/n = \varkappa$, $i/n = \varrho$, имеем

$$P_1 \leq c_5 \exp_2 \{nt\, F_t(\omega, \varkappa, \varrho)\},$$

где c_5 не зависит от n, t, w, k, i и

$$F_t(\omega, \varkappa, \varrho) = \omega H\left(\frac{\varkappa}{\omega t}\right) + (1-\omega) H\left(\frac{\varrho t - \varkappa}{(1-\omega) t}\right) - H(\varrho).$$

Для нужной оценки P_1 достаточно показать, что при $1-\omega_1 > \omega > \omega_1$, $\varrho > \varrho_0$ и $\varkappa > \varrho(\omega+\gamma)t$ имеем $F_t < A < 0$, где A не зависит от t. Для этого заметим, что при $\varkappa = \varrho\omega\, t$ имеем $F_t = 0$, $\frac{\partial F_t}{\partial \varkappa} = 0$ и $\frac{\partial^2 F_t}{\partial \varkappa^2} < 0$. Поэтому при каждых ω и ϱ точка $\varkappa = \varrho\omega\, t$ — точка максимума F_t как функции $\varkappa$, причем этот максимум равен нулю. Кроме того, $\frac{\partial F_t}{\partial \varkappa} < 0$ при всех $\varkappa > \varrho\omega\, t$. Поэтому достаточно найти $F_t(\omega, \varrho(\omega+\gamma)t, \varrho)$. Но

$$F_t(\omega, \varrho(\omega+\gamma)t, \varrho) = \omega H\left(\varrho + \frac{\varrho\gamma}{\omega}\right) + (1-\omega) H\left(\varrho - \frac{\varrho\gamma}{1-\omega}\right) - H(\varrho) < 0$$

и не зависит от t. Аналогично оценивается P_2. Лемма 2 доказана.

Доказательство леммы 3. Без ограничения общности можно, очевидно, предположить, что вектор $\mathbf{x}$ таков, что $x_1 = x_2 = \ldots = x_w = 1$, $x_{w+1} = \ldots = x_n = 0$. Тогда отмеченными ребрами будут ребра $r(1), \ldots, r(wt)$.

Определим случайные величины ζ_i, $1 \leq i \leq wt$, на $\mathcal{E}$ следующим образом. Пусть $i = \sigma(i')$ и $k = \left[\frac{i'-1}{t}\right]$ (т. е. ребро $r(i)$ попало в t-сумматор Σ_k). Положим $\zeta_i = 1$, если ни одно из ребер $r(j)$, $j < i$ не попало в Σ_k, $\zeta_i = -1$, если ровно одно ребро $r(j)$, $j < i$ попало в Σ_k и $\zeta_i = 0$ в остальных случаях. Ясно, что $\sum_{i=1}^{wt} \zeta_i$ — число тех i, что $v_i(x) = 1$, и событие V состоит в том, что $\sum_{i=1}^{wt} \zeta_i > \frac{wt}{2}$.

Лемма 3[a]. $Pr\{\zeta_{k+1} > 0 \,|\, \zeta_1 = v_1, \ldots, \zeta_k = v_k\} > p' = 1 - \frac{wt}{n}$ при $i \leq wt$.

Доказательство. Пусть $\varrho_k^{(i)}$ — случайная величина равная числу ребер $r(j)$, $j < i$, попавших в сумматор Σ_k. Тогда $\lambda_k^{(i)} = Pr\{$ребро $r(i)$ попало в $\Sigma_k \,|\, \varrho_k^{(i)} = g_{(k)}^i$ для $1 \leq k \leq n\} = \frac{t - g_k^{(i)}}{t - i + 1}$.

Поэтому $\lambda_k^{(i)} \geq \lambda_{k'}^{(i)}$, если $g_k^{(i)} = 0$. Но число тех k, что $g_k^{(i)} > 0$ не превосходит $i \leq wt$ и значит $\lambda_k^{(i)} > p'$, если $g_k^{(i)} = 0$. Применение формулы полной вероятности заканчивает доказательство леммы 3[а].

Лемма 3[б]. Пусть θ — случайная величина, принимающая 3 значения: $-1,0,1$ с вероятностями p_{-1}, p_0, p_1, соответственно, $\Phi(v) = Pr\{\theta > v\}$. Тогда

$$\Phi^{*(k)}\left(\frac{k}{2}\right) \geq C_k^{k/4}\, p_1^{3k/4}(1-p_1)^{k/4}\,.$$

Доказательство. Пусть $\Xi_k = \sum_{i=0}^{k} \theta_i$, где θ_i, — взаимо независимые случайные величины, распределенные так же, как θ. Тогда

$$\Phi^{*(k)}\left(\frac{k}{2}\right) = Pr\left\{\Xi_k > \frac{k}{2}\right\}$$

и лемма 3[б] следует из того, что имеет место включение $\left\{\Xi_k > \frac{k}{2}\right\} \subset \{\theta_i = 1$ больше, чем для $3k/4$ числа индексов i, $1 \leq i \leq k\}$.

Лемма 3[а] и 3[б] позволяют применить теорему 4 (когда A —достоверное событие) к случайным величинам $\zeta_1, \ldots, \zeta_{wt}$ и свести лемму 3 к оценке величины

$$a^w C_n^w C_{wt}^{\frac{wt}{4}} \left(1 - \frac{wt}{n}\right)^{\frac{3wt}{4}} \left(\frac{wt}{n}\right)^{\frac{wt}{4}} = B$$

при $w < \omega_1\, n$. Положим ω_0 таким, что $\omega_0\, t < 1$. Тогда при $w/n < \omega_0$ имеем $C_{wt}^{\frac{wt}{4}} < 2^{wt}$, $C_n^w < c_6\, 2^{nH\left(\frac{w}{n}\right)}$ и $1 - \frac{wt}{n} > 1 - \omega_0$, где c_6 — некоторая постоянная. Поэтому

$$\log B < wc_7 + nH\left(\frac{w}{n}\right) + w\frac{t}{4}\log\frac{w}{n}\,,$$

где c_7 — некоторая постоянная. Положим $\omega = \frac{w}{n}$. Тогда

$$\log B < n\left\{\omega c_7 + H(\omega) + \omega\frac{t}{4}\log\omega\right\}.$$

Легко проверить, что $H(\omega) < 2\omega \log \omega$ при $\omega < \omega_0$. Поэтому

$$\log B < n\omega\left\{c_7 + \left(\frac{t}{4} - 2\right)\log\omega\right\}.$$

Выбирая ω_1 так, чтобы $-\log \omega > \frac{4c_7}{t-12}$ выполнялось при всех ω, $0<\omega<\omega_1$ (что возможно, поскольку $t>12$) получаем, что при таких ω

$$\log B < n\,\omega \log \omega,$$

т. е.

$$B < 2^{w \log \frac{w}{n}}.$$

Ясно, что найдется такое n_0, что $2^{w \log \frac{w}{n}} < n^{-1/2}$ для всех $n > n_0$ и $1 \leq w < \omega_1 n$. Лемма 3 доказана.

Литература

1. *Savage, J. E.:* The complexity of decoders. I, II IEEE Trans. Inform. Theory, **IT-15,** **6** (1969), **IT-17,** 1 (1971).
2. *Savage, J. E.:* The complexity of deterministic source encoding with a fidelity criterion. Preprint, Brown Univ. (1971).
3. *Кузнецов А. В.:* Сложность кодирующих устройств для блоковых линейных кодов. Труды МФТИ, 1969, стр. 84—91.
4. *Питерсон У.:* Коды, исправляющие ошибки. М., изд-во «Иностр. лит.» (1962).
5. *Лупанов О. Б.:* О синтезе некоторых классов управляющих систем. Сб. «Проблемы кибернетики», *10* (1963).
6. *Возенкрафт Дж. — Рейффен Б.:* Последовательное декодирование. М., изд-во «Иностр. лит.» (1963).

The complexity of asymptotically optimal code realization by constant depth schemes

S. I. GELFAND, R. L. DOBRUSHIN

(Moscow)

Summary

The complexity of schemes realizing linear codes was studied. The main result is the following. Let R, $0 < R < 1$ be the rate of transmission, and $d_{\text{в.г}}\, n$ the asymptotical Varshamov—Gilbert bound for the code distance of the block code of the length n with rate R. Let us have any $d < d_{\text{в.г}}$. Then the scheme G of constant depth can be constructed, which realizes a linear code $\mathcal{K}$, with rate R and code distance greater than dn, whose complexity is not greater than $c_1 n \log n$ and depth is not greater, than $c_2 \log n$. Here c_1 and c_2 do not depend on n. Similar lower bounds are also given. These bounds show, that it is impossible to receive further essential simplifications of such schemes.

The upper bounds, which are more complicated, are received by method of random schemes. More precisely, we construct the set of schemes and define on this set the probability assignment in such a way, that a scheme from this set satisfies the required condition with probability p, which is great enough and even near to 1, when n is large. For the calculation of this probability we use methods, similar to those of constructing random codes with good properties.

С. И. Гельфанд, Р. Л. Добрушин

Институт проблем передачи информации

СССР, Москва Е-24, Авиамоторная ул. 8а, корп. 2

Problems of Control and Information Theory, Vol. 1 (3—4), pp. 217—238 (1972)

УСЛОВНАЯ НАБЛЮДАЕМОСТЬ ЛИНЕЙНЫХ СИСТЕМ

Р. Ф. ГАБАСОВ, Р. М. ЖЕВНЯК, Ф. М. КИРИЛЛОВА, Т. Б. КОПЕЙКИНА

(Минск)

(Поступила в редакцию 3 ноября 1971 г.)

Строится теория наблюдения линейных динамических систем, описываемых обыкновенными дифференциальными уравнениями, дифференциальными уравнениями с запаздывающим аргументом. Изучается наблюдаемость композитных систем. Строится теория наблюдения линейных динамических систем, описываемых обыкновенными дифференциальными уравнениями, дифференциальными уравнениями с запаздывающим аргументом, изучается наблюдаемость композитных систем.

Введение

Проблема наблюдаемости динамических систем является одной из центральных проблем теории автоматического регулирования. Восстановление состояний системы по доступным измерениям выходных координат важно при решении многих задач управления объектами. Особенно остро эта проблема встала после появления теории оптимальных процессов. Известно, что разработанные алгоритмы оптимизации основаны, как правило, на полном знании состояния системы, в то время как во многих реальных ситуациях имеющаяся информация о состоянии ограничена измерениями небольшого числа выходных переменных.

Математическая теория наблюдаемости стала развиваться с 1961 г., когда Р. Калман [1] сформулировал и дал решение следующей задачи наблюдения. Имеется система

$$\dot{\mathbf{x}} = A\mathbf{x}, \ \mathbf{x} = (x_1, \ldots, x_n), \tag{0.1}$$

где A — $n \times n$-постоянная матрица, характеризующая динамические свойства объекта. Состояние системы $\mathbf{x}(t)$, $t_0 \leq t \leq t_1$ недоступно непосредственному измерению; измеряется лишь m-векторная функция $\mathbf{y}(t)$, связанная с $x(t)$ соотношением:

$$\mathbf{y}(t) = C\,\mathbf{x}(t), \ t_0 \leq t \leq t_1, \tag{0.2}$$

где C — заданная, постоянная $m \times n$-матрица.

При каких условиях на известные матрицы A, C по функции[1] $\mathbf{y}(t)$, $t_0 \leq t \leq t_1$ можно восстановить вектор $\mathbf{x}_0 = \mathbf{x}(t_0)$. Оказывается, что эта задача имеет решение тогда и только тогда, когда[2]

$$\text{ранг} \begin{Vmatrix} C \\ CA \\ \cdot \\ \cdot \\ \cdot \\ CA^{n-1} \end{Vmatrix} = n, \tag{0.3}$$

или, как будем писать в дальнейшем, ранг $\begin{Vmatrix} CA^k \\ 0 \leq k \leq n-1 \end{Vmatrix} = n.$

В последующих работах теория наблюдения обогатилась новыми взглядами, подходами и интерпретациями. Но в опубликованных работах результат (0.3) по существу не обобщался и не подвергался анализу. Хотя обыкновенные линейные системы наблюдения можно представить в виде (0.1), (0.2), недостаток такого подхода, очевидно, состоит в том, что в данном случае приходится существенно повышать размерности матриц A, C, которые участвуют в формулировке критерия (0.3). С другой стороны, сведение любой динамической системы к виду (0.1), (0.2) приводит к матрицам A, C больших размерностей со специальными структурами, когда многие их элементы являются нулевыми. При таком положении естественна попытка не прибегать к преобразованию исходных уравнений к виду (0.1), (0.2), а формулировать результаты непосредственно в терминах заданных систем.

Данная работа посвящена построению теории наблюдения, учитывающей конкретную структуру систем. В отличие от известных результатов, здесь не только предельно упрощаются критерии наблюдаемости, но и выявляется физическая сущность явления. Преимущества предлагаемой теории особенно заметны при исследовании сложных систем, составленных из многих звеньев, когда по небольшому числу измерений требуется восстановить некоторые неизвестные координаты.

В качестве эталона для теории наблюдения линейных динамических систем нами взята классическая теория устойчивости движения линейных систем. Как известно, последняя полностью сводится к алгебраической задаче вычисления корней характеристического уравнения — фундаментального понятия теории устойчивости. Элементарность процедуры составления харак-

[1] Для стационарных систем можно положить $t_0 = 0$, что и будет делаться в дальнейшем.

[2] В [1] и последующих работах вместо матрицы, стоящей слева, рассматривалась матрица, ей сопряженная, т. е. условие (0.3) имело вид: ранг $(C^*, A^*C^*, \ldots, (A^*)^{n-1}C^*) = n$. В развиваемой ниже теории форма (0.3) естественнее.

теристического уравнения непосредственно по дифференциальным уравнениям движения без предварительного их решения является основным аргументом, в силу которого эта теория широко используется в приложениях. Введение частотных критериев, исходящих из характеристик, доступных экспериментальному вычислению и не требующих обязательного знания дифференциальных уравнений движения, еще более повысило инженерную ценность теории устойчивости линейных систем. В развиваемой ниже теории делается попытка перенести отмеченные черты теории устойчивости на проблему наблюдаемости.

1. Условная наблюдаемость

Рассмотрим динамический объект, движение которого можно описать дифференциальным уравнением (0.1). Будем считать, что состояния $\mathbf{x}(t)$ на отрезке $T = (0, t_1]$, $t_1 > 0$ недоступны непосредственному наблюдению. В распоряжении наблюдателя имеется лишь m-мерная вектор-функция $\mathbf{y}(t)$, $t \in T$, связанная с состоянием $\mathbf{x}(t)$ объекта соотношением (0.2).

Пусть, далее, интересующие наблюдателя начальные состояния $\mathbf{x}_0$ представимы в виде

$$\mathbf{x}_0 = H\mathbf{z}\,, \tag{1.1}$$

где H — постоянная $n \times n$-матрица, $\mathbf{z}$ — n-вектор. При изменении вектора $\mathbf{z}$ в n-мерном пространстве вектор $\mathbf{x}_0$ в силу (1.1) изменяется в некотором подпространстве, натянутом на векторы-столбцы матрицы H. Если H — неособая матрица то, очевидно, вектор $\mathbf{x}_0$ описывает все n-мерное пространство состояний.

Задача. По наблюдениям $\mathbf{y}(t)$, $t \in T$, порожденным некоторым (неизвестным) состоянием $\mathbf{x}_0$ вида (1.1), восстановить это начальное состояние.[3]

Введем определение. Систему (0.1) с начальными состояниями вида (1.1) и выходом (0.2) назовем условно наблюдаемой, если любое неизвестное начальное состояние указанного вида можно восстановить по измерениям $\mathbf{y}(t)$, $t \in T$.

Учитывая линейность задачи наблюдения (0.1), (0.2), (1.1), ограничимся частной математической формулировкой задачи [1], соответствующей рассматриваемому случаю.

Будем говорить, что направление[4] $\mathbf{p} = (p_1, \ldots, p_n)$ системы (0.1) условно наблюдаемо на отрезке $T = (0, t_1]$ по выходу (0.2), если существует изме-

[3] Для упрощения последующих рассуждений будем предполагать, что вся информация об объекте может обрабатываться несколькими однотипными измерительными устройствами, или, что формально то же самое, многократно одним устройством (0.2).

[4] В [1] вместо термина «направление» употребляется термин «кососостояние».

римая m-вектор-функция $\mathbf{r}(t)$, $t \in T$, такая, что линейный интегральный оператор $\int_0^{t_1} \mathbf{r}'(t)\,\mathbf{y}(t)\,dt$ восстанавливает значение проекции $\mathbf{p}'\,\mathbf{x}_0$ для любого начального состояния $\mathbf{x}_0$ вида $\mathbf{x}_0 = H\mathbf{z}$, т. е.

$$\int_0^{t_1} \mathbf{r}'(t)\,C\mathbf{x}(t)\,dt = \mathbf{p}'\mathbf{x}_0\,. \tag{1.2}$$

Система (0.1), (0.2) называется условно наблюдаемой, если каждое ее направление условно наблюдаемо.

Эквивалентность двух определений условной наблюдаемости нетрудно получить из [1].

По условию задачи решение $\mathbf{x}(t)$ можно записать в виде:

$$\mathbf{x}(t) = F(t)\,\mathbf{x}_0,\ \mathbf{x}_0 = H\mathbf{z}\,.$$

Здесь $F(t)$ — фундаментальная матрица решений системы (0.1). Подставив $\mathbf{x}(t)$ в (1.2) получим, что для условной наблюдаемости системы (0.1), (0.2) необходимо и достаточно, чтобы для каждого n-вектора $\mathbf{p}$ существовала измеримая m-вектор-функция $\mathbf{r}(t)$ такая, что

$$\int_0^{t_1} \mathbf{r}'(t)\,CF(t)\,Hdt = \mathbf{p}'H\,.$$

Условие разрешимости последней задачи следует из лемм 1,2 (доказательство проводится по схеме [2], стр. 279—282).

Лемма 1. Для того чтобы при каждом n-векторе $\mathbf{p}$ существовала измеримая m-векторная функция $\mathbf{r}(t)$ такая, что

$$\int_0^{t_0} \mathbf{r}'(t)\,CF(t)\,Hdt = \mathbf{p}'H\,, \tag{1.3}$$

необходимо и достаточно, чтобы

$$CF(t)\,H\mathbf{g} \not\equiv 0 \tag{1.4}$$

для всех n-векторов $\mathbf{g}$ таких, что $H\mathbf{g} \neq 0$.

Лемма 2. Чтобы для любого n-вектора $\mathbf{g}$, такого, что $H\mathbf{g} \neq 0$, выполнялось условие

$$CF(t)\,H\mathbf{g} \not\equiv 0,\quad t \in (0, t_1],\ t_1 > 0\,, \tag{1.5}$$

необходимо и достаточно, чтобы

$$\text{ранг}\left\{\begin{matrix} CA^kH \\ 0 \leq k \leq n-1 \end{matrix}\right\} = \text{ранг}\,H\,. \tag{1.6}$$

Теорема 1. Для того чтобы система (0.1), (0.2) была условно наблюдаемой, необходимо и достаточно, чтобы

$$\text{ранг}\begin{Vmatrix} CA^k H \\ 0 \leq k \leq n-1 \end{Vmatrix} = \text{ранг}\, H\,. \tag{1.7}$$

Введение понятия «условно наблюдаемые системы» естественно, например, при изучении объектов, описываемых уравнением

$$\mathbf{X}^{(\alpha)} + A_1 \mathbf{X}^{(\alpha-1)} + \ldots + A_{\alpha-1}\dot{\mathbf{X}} + A_\alpha \mathbf{X} = 0,\ \ \mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n), \tag{1.8}$$

когда по измерениям $\mathbf{Y}(t) = C_1,\ \mathbf{X}(t),\ t \in T$, требуется восстановить только начальное положение $\mathbf{X}_0$ траектории $\mathbf{X}(t)$. В этом случае, полагая $\mathbf{X} = \mathbf{X}_1$, $\dot{\mathbf{X}}_1 = \mathbf{X}_2, \ldots, \dot{\mathbf{X}}_{\alpha-1} = \mathbf{X}_\alpha$ и вводя обозначение $\mathbf{x} = (\mathbf{X}_1, \ldots, \mathbf{X}_\alpha)$, приходим к системе (0.1), (0.2) с $n\alpha \times n\alpha$-матрицей $\bar{A}$:

$$\bar{A} = \begin{bmatrix} O_n & E_n & O_n & \ldots & O_n \\ O_n & O_n & E_n & \ldots & O_n \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ O_n & O_n & O_n & \ldots & E_n \\ -A_\alpha & -A_{\alpha-1} & -A_{\alpha-2} & \ldots & -A_1 \end{bmatrix},$$

где O_n — прямоугольная $m \times n$-матрица с нулевыми элементами, и $m \times n\alpha$-матрицей $\bar{C} = (C_1, O_{mn}, \ldots, O_{mn})$. В новых обозначениях задача восстановления вектора $\mathbf{X}_0$ эквивалентна задаче условной наблюдаемости с $n\alpha \times n$-матрицей $\bar{H}$ вида:

$$\bar{H} = \begin{bmatrix} E_n \\ O_n \\ \vdots \\ O_n \end{bmatrix}.$$

Из критерия (1.7) следует: вектор $\mathbf{X}_0 = \mathbf{X}(0)$ в системе (1.8) можно восстановить по измерениям $\mathbf{Y}(t) = C_1\,\mathbf{X}(t),\ t \in T$, тогда и только тогда, когда

$$\text{ранг}\begin{Vmatrix} \bar{C}\bar{A}^k \bar{H} \\ 0 \leq k \leq n\alpha - 1 \end{Vmatrix} = \text{ранг}\,\bar{H} = n\,. \tag{1.9}$$

Аналогично исследуется условная наблюдаемость систем с инерционным измерительным устройством, когда объект описывается уравнением (1.8), а выходные переменные $\mathbf{Y}(t)$ удовлетворяют дифференциальному уравнению

$$\mathbf{Y}^{(\beta)} + B_1 \mathbf{Y}^{(\beta-1)} + \ldots + B_{\beta-1}\dot{\mathbf{Y}} + B_\beta \mathbf{Y} = C_1 \mathbf{X},\ \ \mathbf{Y} = (Y_1, \ldots, Y_m). \tag{1.10}$$

В этом случае при переходе к (0.1), (0.2) получаем систему $n\alpha + m\beta$-порядка с выходом $\mathbf{y} = \bar{C}\mathbf{x}$, где $m \times (n\alpha + m\beta)$-матрица $\bar{C}$ имеет вид $(O_{mn}, \ldots, O_{mn}, E_m, O_m, \ldots, O_n)$, E_m стоит на $\alpha+1$ месте, а в критерии типа (1.9) индекс κ будет меняться от 0 до $n\alpha + m\beta - 1$.

Недостатки такого («прямого») подхода к исследованию условной наблюдаемости систем (1.8) с выходом $\mathbf{Y}(t)$, удовлетворяющим уравнению (1.10), очевидны. Отметим основные из них: 1) критерий наблюдаемости (1.9) записывается не в терминах заданных параметров (матриц $A_1, \ldots, A_\alpha, B_1, \ldots, B_\beta, C_1$), а через косвенные параметры $\bar{A}, \bar{C}, \bar{H}$, 2) переход от заданных уравнений объектов (1.8) и измерительных устройств (1.10) к уравнениям (0.1), (0.2) резко повышает размеры матриц, подлежащих рассмотрению, что усложняет проверку критерия (1.9). Поэтому цель последующих рассмотрений состоит в том, чтобы обосновать новый подход к изучению наблюдения сложных систем, при котором можно избежать отмеченных недостатков. Главным в предлагаемом подходе является понятие определяющего уравнения системы.

2. Определяющее уравнение системы наблюдения

Прежде всего заметим, что критерий (1.7) можно записать в виде

$$\text{ранг} \left\{ \begin{matrix} Y_k H \\ 0 \leq k \leq n-1 \end{matrix} \right\} = \text{ранг}\, H\,, \tag{2.1}$$

где Y_k — последовательность $m \times n$-матриц, вычисленных по рекуррентным соотношениям:

$$X_{k+1} = AX_k,\ Y_k = CX_k,\ X_0 = E_n\,. \tag{2.2}$$

Форма записи (2.1) критерия (1.7) в некотором смысле предпочтительнее старой, так как соотношения (2.2), по которым строится критерий (2.1), непосредственно связаны с уравнением объекта (0.1), уравнением измерительного устройства (0.2) и подпространством начальных состояний. Первые два уравнения в (2.2) получены, очевидно, из (0.1), (0.2) с помощью элементарного правила (соответствия):

$$\mathbf{x} \to X_k,\ \dot{\mathbf{x}} \to X_{k+1},\ \mathbf{y} \to Y_k\,, \tag{2.3}$$

(X_k— $n \times n$-матрица, Y_k — $m \times n$-матрица).

Чтобы еще более упростить формулировку критериев условной наблюдаемости и приблизить их к исходным уравнениям задачи наблюдения, условимся в дальнейшем уравнения движения (0.1) записывать в виде

$$\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{z}(t),\ \mathbf{z}(t) = \mathbf{x}_0\,\delta(t)\,, \tag{2.4}$$

где $\delta(t)$ — δ-функция Дирака, и считать $\mathbf{x}(-0) = \mathbf{0}$. Понятно, что при $t \geq +0$ решения системы (0.1) с $\mathbf{x}(0) = \mathbf{x}_0$ и системы (2.4) с $\mathbf{x}(-0) = \mathbf{0}$ совпадают. Форма записи (2.4) с использованием δ-функции и ее производных довольно распространена в операционном исчислении.

Дополним соответствие (2.3) новым, положив

$$\mathbf{z}(t) \to Z_k \qquad (Z_k - n \times n\text{-матрица}). \tag{2.5}$$

Тогда из (2.4), (0.2) получаются рекуррентные соотношения:

$$X_{k+1} = AX_k + Z_k, \quad Y_k = CX_k, \tag{2.6}$$

При исследовании условной наблюдаемости обыкновенных динамических систем зададим раз и навсегда начальные условия для соотношений (2.6) в следующем виде:

$$X_k = 0_n, \; k < 0; \; Y_k = 0_{mn}, \; k \leq \beta, \; Z_k = 0_n, \; k \neq -1, \; Z_{-1} = E_n. \tag{2.7}$$

Здесь β — порядок собственного оператора измерительного устройства (для (0.2) $\beta = 0$). Нетрудно подсчитать, что последовательность Y_k, $k = 0, 1, \ldots, n-1$, участвующая в критерии (2.1), есть решение уравнений (2.6) с начальными условиями (2.7).

Рекуррентные уравнения (2.6) назовем определяющим уравнением системы наблюдения (0.1), (0.2).

Как покажут дальнейшие исследования, определяющее уравнение играет в задаче условной наблюдаемости такую же роль, как характеристическое уравнение в теории устойчивости системы (0.1). Уже в операционном правиле (2.3) составления определяющего уравнения можно усмотреть аналогию с известным правилом составления характеристического уравнения $\det|\lambda E_n - A| = 0$, когда полагают:

$$\mathbf{x} \to E_n, \; \dot{\mathbf{x}} \to \lambda E_n. \tag{2.8}$$

Как известно, соответствие (2.8) легко расширяется: $\mathbf{x}^{(i)} \to \lambda^i E_n$, после чего характеристическое уравнение для систем с собственным оператором общего вида (1.8) принимает форму

$$\det|\lambda^\alpha E_n + A_1 \lambda^{\alpha-1} + \ldots + A_{\alpha-1}\lambda + A_\alpha| = 0.$$

Аналогично можно расширить соответствие (2.3), (2.5). Но предварительно условимся о некоторых обозначениях. Через $\mathbf{x}$ будем всегда обозначать выход объекта наблюдения (для (0.1) вектор выходных координат совпадает с вектором состояния), под символом $\mathbf{y}$ будем понимать вектор выходных координат измерительного устройства. Далее, будем считать (если не

оговаривается другое), что восстановлению подлежат начальные значения $\mathbf{x}_0$ вектора $(\mathbf{x}_0, \dot{\mathbf{x}}_0, \ldots, \mathbf{x}_0^{(\alpha-1)})$.

При таких предположениях дифференциальное уравнение движения объекта

$$\mathbf{x}^{(\alpha)} + A_1 \mathbf{x}^{(\alpha-1)} + \ldots + A_{\alpha-1} \dot{\mathbf{x}} + A_\alpha \mathbf{x} = 0 \tag{2.9}$$

с начальными условиями: $\mathbf{x}_0 = H\mathbf{z}$, $\dot{\mathbf{x}}_0 = 0, \ldots, \mathbf{x}_0^{(\alpha-1)} = 0$, можно записать в виде

$$\mathbf{x}^{(\alpha)} + A_1 \mathbf{x}^{(\alpha-i)} + \ldots + A_{\alpha-1} \dot{\mathbf{x}} + A_\alpha \mathbf{x} = \mathbf{z}(t)^{(\alpha-1)} + A_1 \mathbf{z}(t)^{(\alpha-2)} + \ldots + A_{\alpha-1}\mathbf{z}(t),$$

или, с помощью оператора

$$D_{\alpha-i}(p) = Ep^{\alpha-i} + A_1 p^{\alpha-i-1} + \ldots + A_{\alpha-i-1} p + A_{\alpha-i}, \quad p = d/dt, \tag{2.10}$$

в более компактной форме

$$D_\alpha(p)\,\mathbf{x} = D_{\alpha-1}(p)\,\mathbf{z}(t). \tag{2.11}$$

(при этом считается, что объект до момента $t = 0$ находился в покое).

Расширение правила (2.3), (2.5), о котором шла речь выше, осуществим по схеме

$$\mathbf{x}^{(i)} \to X_{k+i}, \quad \mathbf{y}^{(i)} \to Y_{k+i}, \quad \mathbf{z}(t)^{(i)} \to Z_{k+i}. \tag{2.12}$$

Это позволяет для объекта (2.9) с измерительным устройством (0.2) записать соотношения:

$$\begin{gathered} X_{k+\alpha} + A_1 X_{k+\alpha-1} + \ldots + A_{\alpha-1} X_{k+1} + A_\alpha X_k = Z_{k+\alpha-1} + \\ A_1 Z_{k+\alpha-2} + \ldots + A_\alpha Z_k, \\ Y_k = CX_k, \end{gathered} \tag{2.13}$$

которые будем называть определяющим уравнением системы наблюдения (2.9), (0.2).

Запись уравнения (2.13) упрощается, если по аналогии с (2.10) ввести оператор Δ:

$$\Delta^i Z_k \equiv Z_{k+i}, \quad \Delta^i X_k \equiv X_{k+i}.$$

Тогда (2.13) принимает вид

$$D_\alpha(\Delta)\, X_k = D_{\alpha-1}(\Delta)\, Z_k, \quad Y_k = CX_k,$$

получаемый из (2.11) простой заменой $p \to \Delta$.

При изучении систем наблюдения с инерционным измерительным устройством будем предполагать, что до начала наблюдения устройство находилось в покое, т. е. $\mathbf{y}^{(i)}(-0) = 0$, $i = 0, 1, \ldots, \beta - 1$. Собственный оператор измерительного устройства обозначим через $N_\beta(p) = Ep^\beta + B_1 p^{\beta-1} + \ldots + B_{\beta-1}p + B_\beta$. Для оператора воздействий измерительного устройства (от объекта наблюдения) введем обозначение $M_\gamma(p) = C_0 p^\gamma + C_1 p^{\gamma-1} + \ldots + C_{\gamma-1} p + C_\gamma$.

При этих предположениях уравнение $N_\beta(p)\,\mathbf{y} = M_\gamma(p)\,\mathbf{x}$ полностью описывает поведение инерционного измерительного устройства. Объединяя это уравнение с (2.11), получим общее уравнение обыкновенной системы наблюдения:

$$D_\alpha(p)\,\mathbf{x} = D_{\alpha-1}(p)\,\mathbf{z}(t), \quad N_\beta(p)\,\mathbf{y} = M_\gamma(p)\,\mathbf{x}.$$

Ей соответствует следующее определяющее уравнение:

$$D_\alpha(\Delta)\,X_k = D_{\alpha-1}(\Delta)\,Z_k, \quad N_\beta(\Delta)\,Y_k = M_\gamma(\Delta)\,X_k. \tag{2.14}$$

Напомним, что начальные условия для определяющего уравнения (2.14) имеют вид (2.7).

Остальная часть данной работы, относящаяся к обыкновенным динамическим системам, посвящена доказательству критерия условной наблюдаемости

$$\text{ранг}\left\{\begin{matrix} Y_k H \\ \tau \leq k \leq s \end{matrix}\right\} = \text{ранг}\, H. \tag{2.15}$$

Это необходимое и достаточное условие наблюдаемости (условной) будет доказано для различных систем, встречающихся в приложениях. Рассматриваемые случаи отличаются друг от друга лишь тем, чему равны числа τ и s в (2.15) (s — максимальное число линейно независимых компонент Y_k, которые можно получить из определяющего уравнения, τ — номер первой ненулевой компоненты последовательности Y_k).

3. Условная наблюдаемость объектов с собственным оператором общего вида

Рассмотрим задачу наблюдения объекта

$$D_\alpha(p)\,\mathbf{x} = D_{\alpha-1}(p)\,\mathbf{z}(t), \tag{3.1}$$

когда измерительное устройство описывается уравнением

$$\mathbf{y} = C\,\mathbf{x}. \tag{3.2}$$

В соответствии с соглашениями, принятыми в 2, требуется по измерениям $\mathbf{y}(t)$, $t \in T$ восстановить начальное значение $\mathbf{x}_0$ вектора $\mathbf{x}(t)$, причем известно, что вектор $\mathbf{x}_0$ может иметь вид $\mathbf{x}_0 = H\mathbf{z}$.

Теорема 2. Для того, чтобы объект (3.1) с измерительным устройством (3.2) был условно наблюдаем, необходимо и достаточно, чтобы

$$\text{ранг} \left\{ \begin{matrix} Y_k H \\ 0 \leq k \leq n\alpha - 1 \end{matrix} \right\} = \text{ранг}\, H. \tag{3.3}$$

Здесь Y_k — решение определяющего уравнения системы наблюдения (3.1), (3.2):

$$D_\alpha(\Delta) X_k = D_{\alpha-1}(\Delta) Z_k, \quad Y_k = CX_x. \tag{3.4}$$

Как видно из (3.3), сформулированный критерий явно записывается через заданные параметры системы (3.1), (3.2) и оперирует с исходными матрицами. Доказательство этого факта сводится к проверке тождественности условий (3.3) и

$$\text{ранг} \left\{ \begin{matrix} \overline{C}\overline{A}^k \overline{H} \\ 0 \leq k \leq n\alpha - 1 \end{matrix} \right\} = \text{ранг}\, \overline{H}.$$

Введем в рассмотрение $n\alpha$-вектор $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_\alpha)$, $m \times n\alpha$-матрицу $\overline{C} = (C, O_{mn}, \ldots, O_{mn})$ и $n\alpha \times n\alpha$-матрицу $\overline{A}$, $n\alpha \times n$-матрицу $\overline{H}$:

$$\overline{A} = \begin{bmatrix} O_n & E_n & O_n & \ldots & O_n \\ O_n & O_n & E_n & \ldots & O_n \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ O_n & O_n & O_n & \ldots & E_n \\ -A_\alpha & -A_{\alpha-1} & -A_{\alpha-2} & \ldots & -A_1 \end{bmatrix}, \quad \overline{H} = \begin{bmatrix} H \\ O_n \\ \cdot \\ \cdot \\ \cdot \\ O_n \\ O_n \end{bmatrix}.$$

Лемма 3. Для любого $k \geq 0$ справедливо равенство

$$\overline{A}^k \overline{H} = \begin{bmatrix} X_k H \\ X_{k+1} H \\ \vdots \\ X_{k+\alpha-1} H \end{bmatrix}, \tag{3.5}$$

где X_i, $i = k, \ldots, k + \alpha - 1$ — решение[5] определяющего уравнения (3.4).

[5] В действительности X_k — компонента решения (X_k, Y_k) уравнения (3.4). Подобная условность встречается и в других местах, но в каждом случае ясно, о чем идет речь.

Доказательство проведем методом математической индукции. При $k = 0$ равенство (3.5) справедливо, ибо, как следует из (3.4), $X_0 = E_n$, $X_l = 0_{mn}$, $l = 1, \ldots, \alpha - 1$. Пусть (3.5) выполняется для $k = \gamma$. Докажем, что оно справедливо и для $k = \gamma + 1$. Имеем

$$\overline{A}^{\gamma+1}\overline{H} = \overline{A}(\overline{A}^{\gamma}\overline{H}) = \begin{bmatrix} O_n & E_n & O_n & \ldots & O_n \\ O_n & O_n & E_n & \ldots & O_n \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ O_n & O_n & O_n & \ldots & E_n \\ -A_\alpha & -A_{\alpha-1} & -A_{\alpha-2} & \ldots & -A_1 \end{bmatrix} \times$$

$$\times \begin{bmatrix} X_\gamma H \\ X_{\gamma+1} H \\ \cdot \\ \cdot \\ \cdot \\ X_{\gamma+\alpha-2} H \\ X_{\gamma+\alpha-1} H \end{bmatrix} = \begin{bmatrix} X_{\gamma+1} H \\ X_{\gamma+2} H \\ \cdot \\ \cdot \\ \cdot \\ X_{\gamma+\alpha-1} H \\ -A_\alpha X_\gamma H - A_{\alpha-1} X_{\gamma+1} H - \ldots - A_1 X_{\gamma+\alpha-1} H \end{bmatrix}.$$

Из (3.4) следует, что

$$-A_\alpha X_\gamma H - A_{\alpha-1} X_{\gamma+1} H - \ldots - A_1 X_{\gamma+\alpha-1} H = X_{\gamma+\alpha} H,$$

поскольку все $Z_i = 0_n$, $i = \gamma, \gamma+1, \ldots, \gamma+\alpha-1$. Утверждение (3.5) доказано.

Теперь докажем теорему 2. Полагая $\mathbf{x} = \mathbf{x}_1$, $\dot{\mathbf{x}}_1 = \mathbf{x}_2, \ldots, \dot{\mathbf{x}}_{\alpha-1} = \mathbf{x}_\alpha$, перейдем от уравнения (3.1) к эквивалентной системе:

$$\dot{\mathbf{x}}_1 = \mathbf{x}_2,\ \dot{\mathbf{x}}_2 = \mathbf{x}_3, \ldots, \dot{\mathbf{x}}_{\alpha-1} = \mathbf{x}_\alpha,$$

$$\dot{\mathbf{x}}_\alpha = -A_\alpha \mathbf{x}_1 - A_{\alpha-1}\mathbf{x}_2 - \ldots - A_2 \mathbf{x}_{\alpha-1} - A_1 \mathbf{x}_\alpha.$$

Применив введенные выше обозначения, получим уравнения:

$$\dot{X} = \overline{A}X,\quad Y = \overline{C}X,$$

для которых критерий условной наблюдаемости имеет вид

$$\text{ранг} \left\{ \begin{matrix} \overline{C}\overline{A}^k\overline{H} \\ 0 \le k \le n\alpha - 1 \end{matrix} \right\} = \text{ранг}\,\overline{H}. \tag{3.6}$$

Докажем тождественность условий (3.3) и (3.6), т. е. покажем, что $Y_k H = \overline{C}\overline{A}^k\overline{H}$.

Действительно из (3.4) следует, что $Y_k H = CX_k H$, а

$$\bar{C}\bar{A}^k H = (C, O_{mn}, \ldots, O_{mn}) \begin{bmatrix} X_k H \\ X_{k+1} H \\ \vdots \\ X_{k+\alpha-1} H \end{bmatrix} = CX_k H .$$

Теорема 2 доказана.

4. Системы наблюдения с инерционным измерительным устройством

Рассмотрим сначала простейший случай:

1. Объект наблюдения описывается уравнением

$$\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{z}(t) ,$$

а измерительное устройство

$$N_\beta(p)\,\mathbf{y} = C\mathbf{x} .$$

Определяющее уравнение такой системы имеет вид:

$$X_{k+1} = AX_k + Z_k, \quad N_\beta(\Delta)\, Y_k = CX_k .$$

Критерий условной наблюдаемости:

$$\text{ранг} \left\{ \begin{matrix} Y_k H \\ \beta \leq k \leq n\alpha + m\beta - 1 \end{matrix} \right\} = \text{ранг}\, H .$$

2. В качестве следующего шага рассмотрим систему, объект наблюдения которой имеет собственный оператор общего вида

$$D_\alpha(p)\,\mathbf{x} = D_{\alpha-1}(p)\,\mathbf{z}(t), \quad N_\beta(p)\,\mathbf{y} = C\mathbf{x} . \tag{4.1}$$

Определяющее уравнение такой системы:

$$D_\alpha(\Delta)\, X_k = D_{\alpha-1}(\Delta)\, Z_k, \quad N_\beta(\Delta)\, Y_k = CX_k . \tag{4.2}$$

Система (4.1) условно наблюдаема тогда и только тогда, когда

$$\text{ранг} \left\{ \begin{matrix} Y_k H \\ \beta \leq k \leq n\alpha + m\beta - 1 \end{matrix} \right\} = \text{ранг}\, H . \tag{4.3}$$

3. Не вызывает принципиальных затруднений и рассмотрение общего случая (рис. 1):

$$D_\alpha(p)\,\mathbf{x} = D_{\alpha-1}(p)\,\mathbf{z}(t), \quad N_\beta(p)\,\mathbf{y} = M_\gamma(p)\,\mathbf{x} . \tag{4.4}$$

Для того чтобы задача условной наблюдаемости имела решение в общем случае (4.4), необходимо и достаточно, чтобы

$$\text{ранг}\left\{\begin{matrix} Y_k H \\ \beta \leq k \leq n\alpha + m\beta - 1 \end{matrix}\right\} = \text{ранг}\, H\,. \tag{4.5}$$

Здесь Y_k — решение определяющего уравнения:

$$D_\alpha(\Delta)\, X_k = D_{\alpha-1}(\Delta)\, Z_k,\quad N_\beta(\Delta)\, Y_k = M_\gamma(\Delta)\, X_k\,. \tag{4.6}$$

Доказательства утверждений этого пункта строятся по схеме доказательств утверждений из 3.

5. Наблюдаемость объектов, содержащих запаздывание

Изучение систем наблюдения с последействием начнем с простейшего случая, когда объект наблюдения описывается дифференциальным уравнением с запаздывающим аргументом вида

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + A_1\,\mathbf{x}(t-h)\,. \tag{5.1}$$

Здесь $\mathbf{x} = (x_1, \ldots, x_n)$ — n-вектор, A, A_1 — постоянные $n \times n$-матрицы, h — постоянное число — запаздывание, $h \geq 0$.

Движение $\mathbf{x}(t)$, $t \geq 0$, будет однозначно определенным в силу (5.1), если заданы начальные условия

$$\mathbf{x}(t) = \begin{cases} \mathbf{x}_0, & t = 0 \\ \varphi(t), & -h \leq t < 0\,. \end{cases} \tag{5.2}$$

В отличие от случая обыкновенных динамических систем начальные условия для системы (5.1) задаются начальным вектором $\mathbf{x}_0$ и начальной кусочно-непрерывной функцией $\boldsymbol{\varphi}(t)$, $-h \leq t < 0$. В соответствии с этим при изучении задачи наблюдения для (5.1) можно различать два случая: в первом — задача состоит в восстановлении начального вектора $\mathbf{x}_0$ или его компонент, во втором случае нужно восстановить как $\mathbf{x}_0$, так и функцию $\varphi(t)$, $-h \leq t < 0$. Ниже изучается лишь первый случай.

Для начала ограничимся простейшим измерительным устройством, описываемым соотношением

$$\mathbf{y}(t) = C\mathbf{x}(t)\,, \tag{5.3}$$

где $\mathbf{y} = (y_1, \ldots, y_m)$ — m-вектор выходных переменных измерительного устройства, C — постоянная $m \times n$-матрица.

Задача. По измерениям $\mathbf{y}(t)$, $0 \leq t \leq t_1$, и известной начальной функции $\boldsymbol{\varphi}(t) = 0$, $-h \leq t < 0$, вычислить начальный вектор $\mathbf{x}_0$ вида:

$$\mathbf{x}_0 = H\mathbf{z}\,, \tag{5.4}$$

где H — постоянная $n \times n$-матрица с рангом $l \leq n$, $\mathbf{z}$ — произвольный n-вектор.

Системы наблюдения с объектом наблюдения (5.1) и измерительным устройством (5.3) назовем условно наблюдаемыми, если для них при некотором $t_1 < \infty$ разрешима сформулированная задача.

Прежде всего заменим уравнение (5.1) с начальным условием (5.2), где $\mathbf{x}_0$ имеет вид (5.4), уравнением

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + A_1\mathbf{x}(t-h) + \mathbf{z}(t), \quad \mathbf{z}(t) = \mathbf{x}_0\,\delta(t), \tag{5.5}$$

с начальным условием

$$\mathbf{x}(t) \equiv 0, \quad -h \leq t \leq 0.$$

Определяющим уравнением системы наблюдения (5.5), (5.3), (5.4) назовем рекуррентные соотношения

$$X_{k+1}(t) = AX_k(t) + A_1X_k(t-h) + Z_k(t), \quad Y_k(t) = CX_k(t), \tag{5.6}$$

с начальными условиями

$$X_k(t) = 0_n, \; Y_k(t) = 0_{mn}, \; k < 0 \text{ или } t \leq 0; \; X_k\big(t + (k+1)h\big) = 0_n,$$
$$k = 0, 1, \ldots; \; Z_k(t) = 0_n, \; k \neq -1; \; Z_{-1}(t) = E_n. \tag{5.7}$$

Нетрудно видеть, что определяющее уравнение (5.6) получено из (5.5) и (5.3) с помощью соответствия

$$\mathbf{x}^{(i)}(t) \to X_{k+i}(t), \quad \mathbf{y}^{(i)}(t) \to Y_{k+i}(t). \tag{5.8}$$

В отличие от обыкновенных систем здесь $X_k(t)$, $X_k(t)$, $Z_k(t)$ — функции от двух аргументов, что отражает различие между обыкновенными дифференциальными уравнениями и и дифференциальными уравнениями с запаздывающим аргументом. Напомним, что размерности $X_k(t)$, $Y_k(t)$, $Z_k(t)$ таковы: $X_k(t)$ — $n \times n$-матрица, $Y_k(t)$ — $m \times n$-матрица, $Z_k(t)$ — $n \times n$-матрица.

Рассмотрим решение уравнения (5.6) с начальными условиями при $k \geq 0$, $t \geq 0$. В первом квадранте линейно независимыми будут лишь элементы, лежащие в ограниченной окрестности начала координат плоскости $\{k, t\}$. Подсчитывать $Y_k(t)$ следует лишь в узлах решетки с параметрами 1 и h по оси k и t, так как в остальных точках плоскости $Y_k(t) = O_{mn}$. Поэтому введем следующее обозначение: $Y_k(l) \equiv Y_k(t)\big|_{t=lh}$, где $l = 0, 1, \ldots, k-1$.

Определяющее уравнение назовем невырожденным, если ранг последовательности $Y_k H$, $k \geq 0$, построенной в колонку, равен рангу H.

Как следует из [3], для проверки невырожденности уравнения (5.6) достаточно вычислить $Y_k(l)$ при $l = 0, 1, \ldots, k-1$; $k = 0, 1, \ldots, n-1$.

Аналогичные определения сохраним и при исследовании других систем наблюдения. Соответствие (5.8) и начальные условия (5.7) всюду остаются одинаковыми. Каждый новый случай будет отличаться от других лишь множеством узлов, лежащих на плоскости $\{k, t\}$ на которых сосредоточены линейно независимые элементы $Y_k(t)$. Явное указание этого множества в каждом конкретном случае может упростить вычисления, но не влияет на общее определение невырожденности определяющего уравнения.

Введенные понятия позволяют следующим образом сформулировать общий результат: система условно наблюдаема тогда и только тогда, когда ее определяющее уравнение невырождено.

Докажем это утверждение для системы наблюдения (5.1), (5.3), (5.4) Решение $\mathbf{x}(t)$ уравнения (5.1) имеет вид [4]

$$\mathbf{x}(t) = F(t, 0)\,\mathbf{x}_0 + \int_{-h}^{0} F(t, s+h)\,A_1\,\boldsymbol{\varphi}(s)\,ds, \tag{5.9}$$

где $F(t, s)$ — $n \times n$-матрица, являющаяся решением уравнения

$$\frac{\partial F(t,s)}{\partial s} = -F(t,s)\,A - F(t, s+h)\,A_1;\ F(s,s) = E_n,\ F(t,s) = 0,\ s > t. \tag{5.10}$$

Поэтому измеряемый сигнал $\mathbf{y}(t)$, $0 \leq t \leq t_1$ равен

$$\mathbf{y}(t) = CF(t, 0)\,\mathbf{x}_0 + C \int_{-h}^{0} F(t, s+h)\,A_1\,\boldsymbol{\varphi}(s)\,ds. \tag{5.11}$$

Направление $\mathbf{p} = (p_1, \ldots, p_n)$ назовем условно наблюдаемым, если найдутся суммируемая с квадратом функция $\mathbf{w}(t)$, $0 \leq t \leq t_1$ и n-мерная функция $\boldsymbol{\omega}(s)$, $-h \leq s < 0$, такие что

$$\int_0^{t_1} \mathbf{w}'(t)\,\mathbf{y}(t)\,dt + \int_{-h}^{0} \boldsymbol{\omega}'(s)\,\boldsymbol{\varphi}(s)\,ds = \mathbf{p}'\mathbf{x}_0, \tag{5.12}$$

для всех $\mathbf{x}_0 = H\mathbf{z}$.

Если все n-мерные направления $\mathbf{p}$ являются условно наблюдаемыми, то систему (5.1), (5.3), (5.4) назовем условно наблюдаемой.

Можно доказать эквивалентность данного определения с введенным ранее.

Физический смысл равенства (5.12) состоит в следующем. Ищутся такие линейные операции, порожденные функциями $\mathbf{w}(t)$ и $\boldsymbol{\omega}(s)$, которые, будучи примененными к известным (по условию задачи) функциям $\mathbf{y}(t)$, $0 \leq t \leq t_1$, и $\boldsymbol{\varphi}(s)$, $-h \leq s < 0$, восстанавливали бы значения проекции любого вектора $\mathbf{x}_0$ на заданное направление $\mathbf{p}$.

Подставив (5.11) в (5.12) получим

$$\int_0^{t_1} \mathbf{w}'(t)\, CF(t,0)\, \mathbf{x}_0\, dt + \int_0^{t_1} \mathbf{w}'(t)\, C \int_{-h}^{0} F(t, s+h)\, A_1\, \boldsymbol{\varphi}(s)\, ds\, dt + \\ + \int_{-h}^{0} \boldsymbol{\omega}'(s)\, \varphi(s)\, ds = \mathbf{p}'\, \mathbf{x}_0 . \tag{5.13}$$

Это равенство при всех $\mathbf{x}_0 = H\mathbf{z}$ и любых кусочно-непрерывных функциях $\boldsymbol{\varphi}(s)$, $-h \leq s < 0$ имеет место тогда и только тогда, когда

$$\int_0^{t_1} \mathbf{w}'(t)\, CF(t,0)\, H\, dt = \mathbf{p}'\, H , \tag{5.14}$$

$$\int_0^{t_1} \mathbf{w}'(t)\, CF(t, s+h)\, A_1\, dt = -\,\boldsymbol{\omega}'(s) , \quad -h \leq s < 0 . \tag{5.15}$$

Поскольку операция $\boldsymbol{\omega}(s)$ однозначно находится из последнего равенства при известной $\mathbf{w}(t)$, то задача наблюдения сводится к разрешимости относительно $\mathbf{w}(t)$ уравнения (5.14). Подобная задача в связи с другой проблемой решена в [2, 3]. Очевидно обобщение техники [2, 3] приводит к следующему необходимому и достаточному условию разрешимости уравнения (5.14)

$$\text{ранг} \begin{Bmatrix} CQ_k(l)\, H \\ 0 \leq k \leq n-1 \\ 0 \leq l \leq k-1 \end{Bmatrix} = \text{ранг}\, H ,$$

где

$$Q_{k+1}(l) = Q_k(l) A + Q_k(l-h)\, A_1 ;\ Q_0(0) = E_n ,\ l \geq 0,\ k = 1, 2, \ldots, Q_k(-h) = O_n .$$

Нетрудно поверить, что решение $Y_k(t)$ определяющего уравнения (5.6) совпадает с $Q_k(t)$ в точках определения последнего, в остальных точках $Y_k(t) = O_{mn}$.

Теорема 3. Система (5.1), (5.3), (5.4) условно наблюдаема на отрезке $[0, t_1]$ тогда и только тогда, когда определяющее уравнение (5.6) невырождено, т. е.

$$\text{ранг} \begin{Bmatrix} Y_k(l)\, H \\ 0 \leq k \leq n-1 \\ 0 \leq l \leq k-1 \end{Bmatrix} = \text{ранг}\, H .$$

Этот результат без принципиальных изменений переносится на задачу наблюдения объектов со многими запаздываниями

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + A_1\, \mathbf{x}(t-h_1) + \ldots + A_\alpha\, \mathbf{x}(t-h_\alpha) + \mathbf{z}(t) . \tag{5.16}$$

Здесь h_i, $i = 1, 2, \ldots, \alpha$ — постоянные неотрицательные числа.

Определяющее уравнение для системы наблюдения, состоящей из (5.16), (5.3), (5.4), имеет вид

$$X_{k+1}(t) = AX_k(t) + A_1 X_k(t - h_1) + \ldots + A_\alpha X_k(t - h_\alpha) + Z_k(t), \quad (5.17)$$
$$Y_k(t) = CX_k(t),$$

при начальных условиях (5.7).

Для проверки невырожденности уравнения (5.17) достаточно вычислить $Y_k(t)$ при $k = 0, 1, \ldots, n-1$; $0 \leq t \leq (k-1)\,h$, $h = \max(h_1, \ldots, h_\alpha)$.

Замечание. Исследование систем, содержащих запаздывание как в системе, так и в измерительном устройстве, проводится аналогично.

Дальнейшее обобщение системы наблюдения по линии введения общих дифференциальных операторов в объект наблюдения и измерительное устройство не сказывается на справедливости общего критерия наблюдаемости (теорема 3), а меняет лишь множество узлов, на которых достаточно вычислить $Y_k(t)$.

6. Физический смысл определяющего уравнения

Цель настоящего параграфа указать экспериментальный способ проверки системы на наблюдаемость. При этом опять будем следовать теории устойчивости. Как известно [5], при исследовании устойчивости не обязательно знать дифференциальное уравнение движения. Частотные критерии основаны на характеристиках, доступных непосредственному измерению на объекте. Такой способ тесно связан с физической интерпретацией характеристического полинома системы, по которой его модуль и аргумент характеризуют величину изменения модуля и амплитуды гармонического сигнала при прохождении его через объект.

Пусть имеется система наблюдения (5.1), (5.3), (5.4). Выпишем выход $\mathbf{y}(t, k+1, i)$ в момент t измерительного устройства, соответствующий входному сигналу объекта $\mathbf{u}(t, k+1, i)$ вида

$$\mathbf{u}(t, k+1, i) = \mathbf{e}^i H^{(k+1)}(t), \quad k = 0, 1, \ldots, \quad (6.1)$$

где $\mathbf{e}$ — i-тый столбец матрицы E_n, $H^{(k+1)}(t)$ — $(k+1)$-ая производная единичной функции (функции Хевисайда) [5]. Иначе говоря, решим следующую систему уравнений

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + A_1\mathbf{x}(t-h) + \mathbf{e}^i H^{(k+1)}(t), \quad \mathbf{y}(t) = C\mathbf{x}(t),$$
$$\mathbf{x}(t) = \begin{cases} \mathbf{x}_0, & t = 0 \\ \boldsymbol{\varphi}(t), & -h \leq t < 0. \end{cases} \qquad H(t) = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0. \end{cases} \quad (6.2)$$

Под решением будем понимать кусочно-аналитическую функцию со скачками, вызванными действием импульсов $\mathbf{e}^i H^{(k+1)}(t)$. Хотя последующие вычисления допускают строгое обоснование, ограничимся формальной стороной вопроса.

Рассмотрим первое уравнение системы (6.2) как неоднородное. Получим

$$\mathbf{x}(t) = F(t, 0)\,\mathbf{x}_0 + \int_{-h}^{0} F(t, s+h)\,A_1\,\varphi(s)\,ds + \int_{0}^{t} F(t, s)\,\mathbf{e}^i H^{(k+i)}(s)\,ds\,. \qquad (6.3)$$

Решение однородной части уравнения (6.2) обозначим через $\mathbf{x}^0(t)$:

$$\mathbf{x}^0(t) = F(t, 0)\,\mathbf{x}_0 + \int_{-h}^{0} F(t, s+h)\,A_1\,\varphi(s)\,ds\,.$$

Тогда равенство (6.3) примет вид

$$\mathbf{x}(t) = \mathbf{x}^0(t) + \int_{0}^{t} F(t, s)\,\mathbf{e}^i H^{(k+1)}(s)\,ds\,. \qquad (6.4)$$

В силу свойств (5.10), функция $F(t, s) \equiv 0$, если $s > t$. Тогда равенство (6.4) можно переписать следующим образом

$$\mathbf{x}(t) = \mathbf{x}^0(t) + \int_{0}^{\infty} F(t, s)\,\mathbf{e}^i H^{(k+1)}(s)\,ds\,.$$

Подинтегральное выражение в правой части можно рассматривать как k-тую производную от обобщенной δ-функции. Согласно правилу дифференцирования обобщенных функций [6], имеем из последнего равенства

$$\mathbf{x}(t) = \mathbf{x}^0(t) + (-1)^k \frac{\partial^k F(t, s)}{\partial s^k}\,\mathbf{e}^i\Big|_{s=0}, \quad k = 1, 2, \ldots\,. \qquad (6.5)$$

Покажем, что имеет место соотношение

$$\frac{\partial^k F(t, s)}{\partial s^k} = (-1)^k \sum_{j=0}^{k} F(t, s + jh)\,X_{k+1}(s + jh)\,. \qquad (6.6)$$

Действительно, пусть $k = 1$. Из уравнения (5.10) имеем

$$\frac{\partial F(t, s)}{\partial s} = (-1)\,[F(t, s)\,A + F(t, s+h)\,A_1] = (-1) \sum_{j=0}^{1} F(t, s + jh)\,X_2(s + jh)\,.$$

Предположим, что выполнено равенство

$$\frac{\partial^p F(t, s)}{\partial s^p} = (-1)^p \sum_{j=0}^{p} F(t, s + jh)\,X_{p+1}(s + jh)\,. \qquad (6.7)$$

В силу соглашения $X_{p+1}(jh) = X_{p+1}(j) = X_{p+1}(s + jh)\big|_{s=0}$ и члены $X_p(s + jh)$ — суть постоянные матрицы.

Пусть $s \neq jh$, $j = 0, 1, \ldots, p$; продифференцируем равенство (6.7). Имеем

$$\frac{\partial^{p+1} F(t, s)}{\partial s^{p+1}} = (-1)^p \sum_{j=0}^{p} \frac{\partial F(t, s + jh)}{\partial s} X_{p+1}(s + jh) =$$

$$= (-1)^{p+1} \sum_{j=0}^{p+1} F(t, s + jh) X_{p+2}(s + jh).$$

Таким образом, равенство (6.6) установлено по индукции. Из соотношений (6.5), (6.6) получим

$$\mathbf{x}(t) = \mathbf{x}^0(t) + \sum_{j=0}^{k} F(t, jh) X_{k+1}(jh) \mathbf{e}^i, \quad k = 1, 2, \ldots .$$

В силу (6.2) имеем

$$\mathbf{y}(t, k, i) = C\mathbf{x}^0(t) + \sum_{j=0}^{k} CF(t, jh) X_{k+1}(jh) \mathbf{e}^i .$$

По определению функция $F(t, s)$ непрерывна всюду, кроме точек $s = t$, где $F(t + 0, t) - F(t - 0, t) = E_n$. Поэтому скачок функции $\mathbf{y}(t, k, i)$ в точке $t = qh$ равен

$$\mathbf{y}(qh + 0, k, i) - \mathbf{y}(qh - 0, k, i) = CX_{k+1}(qh) \mathbf{e}^i = Y_{k+1}(q) \mathbf{e}^i, \quad q = 1, 2, \ldots .$$

Это дает i-тый столбец матрицы $Y_{k+1}(q)$ — решения определяющего уравнения.

Таким образом, i-тый столбец $m \times n$-матрицы $Y_{k+1}(q)$ — решения определяющего уравнения (5.6) представляет собой скачок в момент $t = qh$ решения $\mathbf{y}(t)$ уравнения (5.3), который порожден $(k + 1)$-ой производной H-функции (6.1), поданной в момент $t = 0$ на выход i-того уравнения системы (5.1).

7. Наблюдаемость композитных систем

Реальные динамические системы и измерительные устройства, как правило, являются композитными, т. е. составленными по определенным правилам из простых стандартных звеньев. Поэтому часто динамические системы задаются структурной схемой, показывающей элементы системы и способ их соединения. Распространенными способами являются: 1) последовательное, 2) параллельное, 3) соединение в обратную связь. Комбинированное применение этих способов к простейшим звеньям и целым комплексам приводит к системам с довольно сложной структурной схемой. Конечно, всю систему в целом можно изобразить в виде одного звена с соответствующим оператором.

Можно также, вводя надлежащим образом дополнительные переменные, уравнение движения записать в нормальной форме (0.1) или (5.1). Иногда такой путь целесообразен. Но во многих случаях, например, в теории устойчивости, он неестественен. Рассмотрение систем в первонально заданном виде зачастую помогает более ясному и понятному решению задачи. С этой целью в теории устойчивости разработаны специальные способы составления характеристических уравнений непосредственно по структурной схеме системы. В данном пункте аналогичные правила указываются для задачи наблюдения.

Рассмотрим в отдельности каждый из упомянутых выше способов соединения, причем ограничимся лишь соединениями двух звеньев и случаем, когда звенья описываются простейшими уравнениями.

1. Последовательное соединение объектов наблюдения.

В соответствии со схемой, приведенной на рис. 1, уравнение объекта наблюдения имеет вид

$$\begin{aligned} &\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{z}(t), \quad \dot{\mathbf{w}} = A_1\mathbf{w} + \mathbf{x}(t), \ \mathbf{z}(t) = \mathbf{x}_0\,\delta(t), \\ &\mathbf{x} = (x_1, \ldots, x_n) \ \ \mathbf{w} = (w_1, \ldots, w_n), \end{aligned} \tag{7.1}$$

уравнение измерительного устройства

$$\mathbf{y} = C\mathbf{w}, \ \ \mathbf{y} = (y_1, \ldots, y_m), \ \ C - m \times n\text{-матрица}. \tag{7.2}$$

По данному $\mathbf{y}(t), 0 \leq t \leq t_1$ требуется восстановить начальное состояние $\mathbf{x}_0 = H\mathbf{z}$.

Следуя правилу (5.6), составим определяющее уравнение

$$X_{k+1} = AX_k + Z_k, \ \ W_{k+1} = A_1 W_k + X_k, \ \ Y_k = CW_k, \tag{7. 3}$$

и к начальным условиям (5.7) присоединим $W_k = O_n, k \leq 0$.

Система (7.1), (7.2), (5.4) условно наблюдаема в том и только в том случае, когда определяющее уравнение (7.3) невырожденно, т. е.

$$\text{ранг} \begin{Bmatrix} Y_k H \\ 0 \leq k \leq 2 \end{Bmatrix} = \text{ранг}\, H. \tag{7.4}$$

Схема вычисления Y_k изображена схематично на рис. 2, где исследован оператор $\Delta : \Delta X_k = X_{k+1}$.

Сравнение двух рисунков (рис. 1 и рис. 2) показывает, что они элементарно получаются друг из друга.

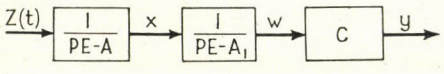


Рис. 1

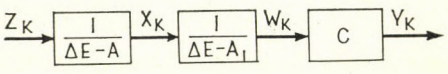


Рис. 2

2. Параллельное соединение объектов наблюдения (рис. 3).

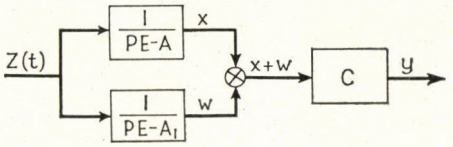


Рис. 3

Соответствующее уравнение системы наблюдения имеет вид

$$\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{z}(t),\ \dot{\mathbf{w}} = A_1\mathbf{w} + \mathbf{z}(t),\ \mathbf{y} = C(\mathbf{x} + \mathbf{w}). \tag{7.5}$$

Применение правила (5.6) приводит к определяющему уравнению

$$X_{k+1} = AX_k + Z_k,\ W_{k+1} = A_1 W_k + Z_k,\ Y_k = C(X_k + W_k), \tag{7.6}$$

со структурной схемой, изображенной на рис. 4, причем $W_k = 0_n$, $k \leq 0$.

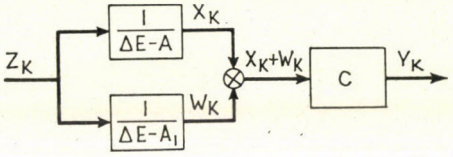


Рис. 4

Критерий условной наблюдаемости системы (7.5), (5.4):

$$\text{ранг}\left\{\begin{matrix} Y_k H \\ 0 \leq k \leq 2n-1 \end{matrix}\right\} = \text{ранг}\, H\,.$$

3. Соединение объектов наблюдения в обратную связь (рис. 5).

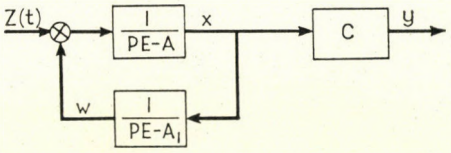


Рис. 5

Система наблюдения описывается уравнением

$$\dot{\mathbf{x}} = A\mathbf{x} + \mathbf{z}(t) + \mathbf{w},\ \dot{\mathbf{w}} = A_1\mathbf{w} + \mathbf{x},\ \mathbf{y} = C\mathbf{x}\,. \tag{7.7}$$

Ей соответствует определяющее уравнение

$$X_{k+1} = AX_k + Z_k + W_k,\ W_{k+1} = A_1 W_k + X_k,\ Y_k = CX_k\,, \tag{7.8}$$

со структурной схемой, изображенной на рис. 6. Здесь тоже $W_k = 0$, $k \leq 0$.

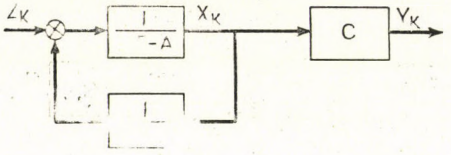


Рис. 6

Необходимое и достаточное условие условной наблюдаемости системы (7.7), (5.4):

$$\text{ранг}\left\{\begin{matrix} Y_k H \\ 0 \leq k \leq 2n-1 \end{matrix}\right\} = \text{ранг}\, H\,,$$

т. е. определяющее уравнение (7.8) должно быть невырожденным.

Исследование систем наблюдения, в которых составными являются не объекты наблюдения, а измерительные устройства, проводится аналогично.

В ы в о д. Таким образом с помощью понятия определяющего уравнения установлен единый критерий наблюдаемости динамических систем. Выяснен физический смысл определяющего уравнения.

Литература

1. *Калман Р.:* Об общей теории систем управления. Труды I конгресса ИФАК, Изд-во АН СССР, 521—547, 1961.
2. *Габасов Р. Ф.— Кириллова Ф. М.:* Качественная теория оптимальных процессов, «Наука», 1971.
3. *Кириллова Ф. М.—Чуракова С. В.:* К проблеме управляемости линейных систем с последействием. Дифференциальные уравнения, **3**, 3, 1967.
4. *Беллман Р.— Кук К.:* Дифференциально-разностные уравнения. «Мир», М., 1967.
5. *Айзерман М. А.:* Лекции по теории автоматического регулирования. ГИФМЛ, М., 1958.
6. *Шварц Л.:* Математические методы для физических наук, «Мир», М., 1965.

Conditional observability of linear systems

R. F. GABASOV, R. M. ZHEVNIAK, F. M. KIRILLOVA, T. B. KOPEIKINA

(Minsk)

Summary

The theory of observation of linear dynamic systems, described by ordinary differential equations were constructed. Conditional observability of systems with general proper differencial operator, with general input differential operator of measuring device, observational system with inertial measuring device were considered. The investigation of observation systems with time-lags both in observation object and in measuring device were carried out on the basis of a new concept of determining equation, which plays the same role in the optimal control problem as the characteristic equation in the theory of stability. The concept of nondegenerative determining equation was introduced and it was proved that the system is relatively observable if the determining equation is nondegenerate. Physical meaning of the determining equation was cleared up and conditional observability of composite system deduced.

Problems of Control and Information Theory, Vol. 1 (3—4), pp. 239—246 (1972)

БИНОИДНЫЕ ПОМЕХОУСТОЙЧИВЫЕ КОДЫ

С. И. САМОЙЛЕНКО

(Москва)

(Поступила в редакцию 22 января 1971 г.)

Вводится определение биноидов, используемых в качестве математического аппарата для построения помехоустойчивых кодов. Рассматриваются методы построения биноидных кодов, корректирующих пакетные и независимые ошибки, пригодные для кодирования как дискретных, так и аналоговых сообщений.

1. Введение

Подавляющее большинство известных помехоустойчивых кодов базируется на математическом аппарате теории конечных полей [1, 2]. Этот математический аппарат является мощным средством для построения корректирующих кодов, однако ограничения, которым должны удовлетворять кодируемые сообщения и используемые операции, являются достаточно сильными и ограничивают применение помехоустойчивого кодирования в некоторых специфичных условиях. Примерами таких условий являются: передача информации между ЭВМ или между ЭВМ и внешним ЗУ при алгоритмической реализации процедур кодирования и декодирования, передача аналоговых сообщений, передача дискретных сообщений с основанием, не равным степени простого числа, и др.

В связи с этим было бы желательным найти методы построения кодов и указать процедуры кодирования и декодирования для кодов с менее жесткими ограничениями на кодируемые элементы и используемые при кодировании операции.

В настоящей работе для построения кодов используется математический аппарат, названный биноидами. Он позволяет существенно ослабить требования к кодируемым элементам и используемым операциям.

На базе введенного аппарата указываются некоторые методы построения кодов, корректирующих пакетные и независимые ошибки.

2. Определение биноида

Пусть A есть множество элементов a, и M — множество операторов m. Пусть $\oplus$ — аддитивная операция, определенная над элементами из A, и $\otimes$ — мультипликативная операция, определенная между элементами A и M. Тогда можно ввести определение биноида.

Пара множеств $\langle A, M\rangle$ с операциями $\oplus$ и $\otimes$ является *биноидом*, если выполняются следующие аксиомы:

B1. Замкнутость множества A по аддитивной операции

$$(\forall a_i)(\forall a_j)[a_i \oplus a_j \in A].$$

B2. Ассоциативный закон для аддитивной операции

$$(\forall a_i)(\forall a_j)(\forall a_k)[(a_i \oplus a_j) \oplus a_k = a_i \oplus (a_j \oplus a_k)].$$

B3. Существование в A нулевого элемента O

$$(\forall a)[a \oplus 0 = 0 \oplus a = a].$$

B4. Существование в A обратных элементов по сложению

$$(\forall a)(\exists \bar{a})[a \oplus \bar{a} = \bar{a} \oplus a = 0].$$

B5. Замкнутость по мультипликативной операции

$$(\forall a)(\forall m)[a \otimes m \in A].$$

Легко убедиться, что множество A с операцией $\oplus$ является группой.

Если биноид удовлетворяет условию

$$(\forall a_i)(\forall a_j)(\forall m)[(a_i \oplus a_j) \otimes m = (a_i \otimes m) \oplus (a_j \otimes m)],$$

то биноид будем называть дистрибутивным. Очевидно, что дистрибутивный биноид является группой с операторами.

Биноид будем называть коммутативным, если он удовлетворяет условию

$$(\forall a_i)(\forall a_j)[a_i \oplus a_j = a_j \oplus a_i].$$

3. Биноидные коды, корректирующие пакетные ошибки

Теорема 1

Код K_D над коммутативным биноидом $\langle A, M\rangle$, определяемый проверочной матрицей

$$H_D = \left[\begin{array}{c|c|c|c|c|c} I_D & I_D & \dots & I_D & I_D & 0 \\ \hline m_1 I_D & m_2 I_D & \dots & m_N I_D & 0 & I_D \end{array}\right]_{m_i, m_j \in M}^{m_i \neq m_j, \text{ если } i \neq j}.$$

корректирует все пакеты ошибок длины D символов или менее с компонентами e_i, удовлетворяющими условиям:

$$(a_r \oplus e_i) \otimes m_j \oplus \overline{(a_r \otimes m_j)} \neq (a_s \oplus e_i) \otimes m_k \oplus \overline{(a_s \otimes m_k)};$$

$$(a_r \oplus e_r) \otimes m_j \oplus \overline{(a_r \otimes m_j)} \neq 0, \quad a_r, a_s \in A \tag{3.1}$$

где I_D — единичная $D \times D$ матрица.

Доказательство

Если ошибки лежат в пределах пакета длины D, то каждая проверка, соответствующая некоторой строке матрицы H_D, содержит не более одного искаженного элемента. При этом каждая отдельная компонента вектора-ошибки искажает пару проверок, если ошибка, расположенная на K-ой позиции, и K не превышает ND, или одну проверку, если $ND < K \leq (N+2)D$. Номера проверок, искаженных данной компонентой ошибки, отличаются на величину D.

Искажения, вносимые отдельными компонентами вектор-ошибки в проверки, являются независимыми и поэтому для доказательства теоремы достаточно показать возможность коррекции одной компоненты, расположенной на произвольной позиции.

Пусть рассматриваемая компонента ошибки e_K располагается на K-ой позиции кодовой комбинации, K не превышает ND, и ошибка искажает проверки h_i и h_{i+D}. Тогда синдром

$$S = \alpha' \otimes H_D^T = \langle s_1, s_2, \ldots, s_{2D} \rangle,$$

(где α' — принятая кодовая комбинация) имеет следующие компоненты s_i и s_{i+D}:

$$s_i = (a_k \oplus e_k) \oplus \bar{a}_k = e_k$$

$$s_{i+D} = [(a_k \oplus e_k) \otimes m_l] \oplus \overline{(a_k \otimes m_l)}, \quad 1 \leq i \leq D$$

где $l = \left\lfloor \frac{K}{D} \right\rfloor x \left[$ — наименьшее целое число, равное или большее x.

Для коррекции ошибки e_K необходимо найти значение l. Если l будет определено, то номер искаженного символа может быть вычислен по соотношению $K = (l-1)D + i$, а истинное значение K-ой компоненты кодовой комбинации может быть найдено из условия, что $\alpha \otimes h_i^T = 0$, $1 \leq i \leq D$, где α — неискаженная кодовая комбинация, h_i — i-я строка H. Для поиска l (найдем значение m_x, удовлетворяющее условию

$$(a'_{i+D(x-1)} \otimes m_x) \oplus [\overline{a'_{i+D(x-1)} \oplus \bar{s}_i)} \otimes m_x] = s_{i+D}, \quad i = 1, 2, \ldots, D. \quad \text{(a}$$

Покажем, что такое значение m_x существует в числе элементов M и что оно является единственным. Факт существования доказывается подстановкой $m_x = m_l$. В этом случае левая часть уравнения (а) приобретает следующий вид:

$$(a'_{i+D(l-1)} \otimes m_l) \oplus [\overline{(a'_{i+D(l-1)} \oplus \bar{e}_k) \otimes m_l}] = (a'_k \otimes m_l) \oplus [\overline{(a'_k \oplus \bar{e}_k) \otimes m_l}] =$$

$$= [(a_k \oplus e_k) \otimes m_l] \oplus \overline{(a_k \otimes m_l)} = s_{i+D}.$$

Следовательно (a) всегда имеет решение. Единственность этого решения вытекает из неравенств, записанных в формулировке теоремы.

Следовательно, если $1 \leq K \leq ND$, то ошибка может быть исправлена.

Если $ND < K \leq (N+2)D$, то из пары проверок s_i и s_{i+D} ненулевое значение будет иметь только одна. Этот факт является свидетельством того, что ошибка расположена на K-ой позиции, где

$$K = \begin{cases} ND + i, & \text{если } s_i \neq 0 \\ (N+1)D + i, & \text{если } s_{i+D} \neq 0, \ 1 \leq i \leq D. \end{cases}$$

Истинное значение a_k может быть найдено из условий $\alpha \otimes h_i^T = 0$, если $s_i \neq 0$, или $\alpha \otimes h_{i+D}^T = 0$, если $s_{i+D} \neq 0$, $1 \leq i \leq D$.

Настоящая теорема показывает возможность построения кодов над биоидами $\langle A, M \rangle$, в которых A является абелевой группой, а множество M и операция умножения являются в широких пределах произвольными. Единственное условие, которому они должны удовлетворять, состоит в том, что

$$a \otimes m \in A, \text{ если } a \in A \text{ и } m \in M.$$

Примером возможного использования недистрибутивных кодов может явиться кодирование элементов, принадлежащих полуинтервалу [0, 1).

В этом случае построение кодов может базироваться на биоиде $\langle B_1, B_N \rangle$, где B_1 — множество элементов, принадлежащих полуинтервалу [0, 1); B_N — конечное подмножество, включающее $N-1$ чисел ($N > 2$) из множества B_1 и единичный элемент, например, $B_N = \left\{0, \frac{1}{N-1}, \frac{2}{N-2}, \ldots, \frac{N-2}{N-1}, 1\right\}$.

В качестве аддитивной операции выберем сложение по модулю 1, а в качестве мультипликативной — обычное умножение.

Тогда пара множеств $\langle B_1, B_N \rangle$ с указанными операциями является коммутативным, но не дистрибутивным биоидом, и для кодирования элементов из B_1 могут быть применены вышеизложенные результаты.

Отметим, что для построения кодов и реализации процедуры декодирования в этом случае не требуется условие дистрибутивности. Благодаря

этому элементы биноида могут удовлетворять достаточно легким условиям (даже более легким, чем элементы группы с операторами), и это позволяет существенно расширить класс кодируемых сообщений и используемых при кодировании и декодировании преобразований.

Вместе с тем следует отметить, что конкретный анализ корректирующих способностей кодов, строящихся на столь широкой основе в некоторых условиях может оказаться достаточно громоздким. Хотя корректирующие способности кодов и определяются условиями теоремы, вычисление конкретных множеств корректируемых ошибок, особенно для кодов с высоким основанием, может оказаться достаточно трудоемким и потребовать использования ЭВМ.

Упрощение анализа кодов можно достигнуть за счет наложения на биноид дополнительных ограничений, в частности, условия дистрибутивности.

Для дистрибутивных биноидов справедлива теорема 2, приводимая ниже без доказательства.

Теорема 2

Код K_D над дистрибутивным коммутативным биноидом $\langle A, M\rangle$, определяемый проверочной матрицей H_D (см. теорему 1), корректирует все пакетные ошибки длины D символов, или менее, и часть пакетов большей длины с компонентами e_i, удовлетворяющими условиям

$$e_i \otimes m_j \neq e_i \otimes m_k,\ m_j \neq m_k,\ e_i \in A;\ m_i,\ m_j \in M.$$
$$e_i \otimes m_j \neq 0.$$

Примером сообщений, для которых удобно использовать коды, описываемые теоремой 2, является множество машинных n-разрядных слов с операциями сложения и умножения по модулю $2^n - 1$.

В этом случае пара множеств $\langle A_n, M_N\rangle$, где A_n — множество всех n-разрядных двоичных чисел, исключая $2^n - 1$, M_N — множество N целых чисел, меньших $2^n - 1$, с указанными операциями, является дистрибутивным коммутативным биноидом и для кодирования таких сообщений могут использоваться результаты теоремы 2.

4. Коды, корректирующие независимые ошибки

В этом разделе будет рассмотрен метод построения биноидных помехоустойчивых кодов на базе матриц инцидентности некоторых комбинаторных схем, например проективных плоскостей или ортогональных латинских квадратов.

Основная особенность таких методов состоит в том, что проверочные матрицы включают только нулевые и единичные элементы и это свойство может быть использовано для упрощения процедур кодирования и декодирования, а также для облегчения требований к множеству кодируемых сообщений.

Для построения кодов, рассматриваемых в настоящем разделе, будут использоваться биноиды $\langle A, M\rangle$, в которых A является произвольной конечной или бесконечной группой, а M — множеством из двух элементов 0 и 1. Такие биноиды будем обозначать $\langle A, M_2\rangle$.

Определим множество многократных корректируемых ошибок как множество $E_K = \{\mathcal{E}_1, \mathcal{E}_2, \ldots,\}$, $\mathcal{E} = \langle e_1, e_2, \ldots, e_n\rangle$, каждый элемент которого удовлетворяет условию:

если $\mathcal{E}$ имеет $\mathfrak{M}$ ненулевых элементов $e_{i_1}, e_{i_2}, \ldots, e_{i\mu}$, $\mathfrak{M} \leq \Delta$, $\Delta \geq 2$ и принадлежит E_k, то

$$(i)\, e_{j_1} \oplus e_{j_2} \oplus \ldots \oplus e_{j_x} \neq 0, \quad x \leq \mathfrak{M},$$

(i, i), если $\mathfrak{M} \geq 3$, то для каждой пары проверок, например s_i и s_j, которые содержат одну ошибку на определенной информационной позиции и ошибки на проверочных позициях $s_i \neq s_j$.

Теорема 3

Код K_Δ над биноидом $\langle A, M_2\rangle$, определяемый проверочной матрицей

$$H_\Delta = [T_\Delta \mid I_r],$$

где T_Δ — матрица инцидентности проектной плоскости порядка Δ;
I_r — единичная матрица, $r = \Delta^2 + \Delta + 1$, корректирует все ошибки кратности Δ, или менее, из множества E_k.

Доказательство

Пусть $\beta' = \langle b'_1, b'_2, \ldots, b'_n\rangle$ есть принятый кодовый вектор, содержащий Δ нли менее ошибок. При отсутствии ошибок

$$S = \beta' \otimes H_\Delta^T = \beta \otimes H_\Delta^T = \langle s_1, s_2, \ldots, s_r\rangle = \langle 0, 0, \ldots, 0\rangle.$$

Если Δ символов вектора β' искажены, то все проверки можно подразделить на два подмножества. К первому из них отнесем все проверки, которые дают нулевой результат, т. е. не обнаруживают ошибок

$$S_0 = \{s_{0_1}, s_{0_2}, \ldots, s_{0_k}\}.$$

Ко второму множеству отнесем все оставшиеся проверки

$$S_\Delta = \{s_{\Delta_1}, s_{\Delta_2}, \ldots, s_{\Delta_{r-k}}\}.$$

Каждая проверка содержит в точности $\Delta + 2$ ненулевых символов. Покажем, что, если выполняется условие (i), то проверки S_0 определяют все правильно принятые информационные символы.

Действительно, по условиям, которым удовлетворяют матрицы инцидентности проектной плоскости, каждый столбец и каждая строка матрицы $[T_\Delta]$ содержит в точности $\Delta + 1$ ненулевых (равных 1) символов, причем каждая пара строк (столбцов) имеет общие единицы только на одной позиции.

Вследствие этого каждый информационный символ кодовой комбинации входит в $\Delta + 1$ проверку, из которых не более Δ могут быть искажены. (Так как все пары столбцов имеют не более одной строки с общими единицами). Таким образом, всегда найдется хотя бы одна неискаженная проверка, в которую входит данный неискаженный символ, и все неискаженные информационные символы будет определены.

Покажем теперь, что если все принятые символы определены, и их значения подставлены в проверки из S_Δ, в которых остальные символы рассматриваются как неизвестные величины, то все неизвестные однозначно определяются системой S_x, полученной из S_Δ указанной заменой символов.

Доказательство этого положения основывается на том, что среди проверок S_x всегда найдется по крайней мере пара проверок, которые содержат только один неизвестный элемент.

Действительно, каждый искаженный символ b'_i, $1 \leq i \leq \Delta^2 + \Delta + 1$, входит в точности в $\Delta + 1$ проверок, из которых не более $\Delta - 1$ могут содержать кроме данного, другой искаженной символ. В соответствии с условием (i i) проверочные соотношения с одной ошибкой на информационной позиции будут иметь одинаковое значение, только при условии, что проверочные символы в этих соотношениях не содержат ошибок.

Следовательно, данный символ может быть определен и его значение можно подставить в оставшиеся проверки S_x. Таким образом, все искаженные символы, расположенные в первой (левой) части кодовой комбинации, будут определены, а если это выполнено, то и символы b'_i, $\Delta^2 + \Delta + 1 \leq i \leq n$ будут однозначно вычислены.

Отметим, что коды K_Δ с коррекцией ошибок позволяют обнаруживать часть ошибок большей, чем Δ кратности. Возможность обнаружения основывается на том, что при большем, чем Δ числе ошибок система S_x может оказаться неразрешимой, если S_0 оказывается пустым множеством, или неоднозначно разрешимой.

Методы построения корректирующих кодов, базирующихся на системе ортогональных латинских квадратов, были рассмотрены в книге [3] (коды с переменным параметром).

Литература

1. *Питерсон, У.:* Коды, исправляющие ошибки. М., «Мир», 1964.
2. *Berlecamp, E. R.:* Algebraic Coding Theory. McGraw-Hill Book Company, 1968.
3. *Самойленко, С. И.:* Помехоустойчивое кодирование. М., «Наука», 1966.

Binoidal error-correcting codes

S. I. SAMOILENKO

(Moscow)

Summary

Some very efficient codes described in the literature are based on group theory; commutative algebra, and finite field arithmetic. Codes best suited for implementation, namely cyclic codes, are based on finite field arithmetic. When their base and length are relatively small, such codes can be easily implemented by means of special-purpose feedback shift register equipment.

In a real burst noise channel, coding can be used efficiently only if the code block is much longer than the average error burst. Implementation of long codes using general-purpose computers is very desirable. However, cyclic codes require a prohibitively high proportion of the computer processing time when realized on general purpose computers. This is primarily due to the complexity of performing multiplication and division of field elements when the order of the field is high.

Therefore the need has arisen for error-correcting codes based on analytical relations amenable to easy computer implementation.

Another problem is the extension of error control techniques to a greater variety of messages including digital messages with an arbitrary base and continuous or discrete time analog messages.

In this paper techniques are described for the construction of error-correcting codes which satisfy the above conditions.

Problems of Control and Information Theory, Vol. 1 (3—4), pp. 247—265 (1972)

CONVENGENCE OF POTENTIAL FUNCTION TYPE LEARNING ALGORITHMS

L. GYŐRFI

(Budapest)

(Received February 23, 1971)

Convergence theorems for some learning algorithms generated by a potential function are dealt with. The kernel function of a reproducing kernel Hilbert space (RKHS) as a potential function is used. In Part I we examine, how may one use a "teaching" several times during the algorithm, and then we turn to the case of noisy "teaching". A question of particular interest, we also investigate, how to choose the potential function, if one only knows, that the function being taught is an element of a Hilbert space.

In Part II we examine algorithms, which turned out to be particularly efficient in learning in unambigouos models by potential function type learning algorithms, and besides this Part gives extentions to Part I, in the sense that the samples are assumed neither to be independent nor identically distributed.

Part I

Prerequisites

What we are next concerned with, is mainly related to the following *Theorem 1* (*Braverman, Rozonoer* [2]). Let $x_1, x_2 \ldots x_n, \ldots$ be a sequence of random variables defined on the probability field $(\Omega, \mathfrak{A}, P)$. Let $U_n \geq 0$, $V_n \geq 0$ $(n = 1, 2 \ldots)$ be a sequence of random variables, U_n being measurable on $x_1 \ldots x_n$, and $M(U_1) < +\infty$, furthermore

$$M(U_{n+1}/x_1 \ldots x_n) \leq (1 + \mu_n) U_n - \gamma_n V_n + \eta_n . \tag{i}$$

Here $$\sum_{n=1}^{\infty} |\mu_n| < +\infty, \quad \gamma_n \geq 0, \quad \lim_{n \to \infty} \gamma_n = 0, \quad \sum_{n=1}^{\infty} \gamma_n = +\infty , \tag{ii}$$

$$\eta_n \geq 0 \text{ are measurable on } x_1 \ldots x_n, \text{ and } \sum_{n=1}^{\infty} M(\eta_n) < +\infty. \tag{iii}$$

Assume that for any sequence $\{\eta_n\}_{k=1}^{\infty}$ for which we have $\lim_{n_k \to \infty} V_{n_k} = 0$ with probability 1, $\lim_{n_k \to \infty} U_{n_k} = 0$ in probability (iv). Then $\lim_{n \to \infty} U_n = 0$ with probability 1.

Let $(\Omega, \mathfrak{A}, P)$ be a probability field, X an arbitrary set, Z a δ-field of subsets of X. Let H be some Hilbert space of functions defined on X with a real

valued inner product. Assume a sequence of random variables $x_1, x_2 \ldots x_n \ldots$ taking values in X and the reals $f(x_1) \ldots f(x_n) \ldots$ to be given ($f \in H$). Our problem is to estimate $f = \{f(x), x \in X\}$ by using these data. We are, specifically, concerned with such algorithms, for which $f_n = \{f_n(x), x \in X\}$ is the approximation of f at the n-th step, f_n is the function of the random variables $x_1 \ldots x_n$.

Let $\alpha_n = ||f(x) - f_n(x)||^2$.

We call $x_1, x_2 \ldots x_n \ldots$ samples, $f(x_1), f(x_2) \ldots$ labels, and the sequence $\{x_i, f(x_i),\ i = 1, 2, \ldots\}$ teaching.

Results

Theorem 2. Let H be a RKHS ([1], [3]) with a kernel $K = \{K(x,y);\ (x,y) \in (X \times X)$ such that

$$\sup_n M(K(x_n, x_n)) = L_2 < +\infty. \tag{1}$$

Let

$$f_{n+1}(x) = f_n(x) + \gamma_n \operatorname{sign}\,(f(x_{n+1}) - f_n(x_{n+1}))\, K(x, x_{n+1}) \tag{2}$$

where

$$\gamma_n \geq 0,\ \sum_{n=0}^{\infty} \gamma_n = +\infty,\ \sum_{n=0}^{\infty} \gamma_n^2 < +\infty,\ f_0(x) \equiv 0\,. \tag{3}$$

If we have, for any $\varepsilon > 0$,

$$\operatorname*{ess\,inf}_{\varepsilon \leq ||f - f_n||^2} M(|f(x_{n+1}) - f_n(x_{n+1})|/\alpha_1 \ldots \alpha_n) \geq q(\varepsilon) \geqslant 0 \tag{4}$$

then $\lim_{n \to \infty} ||f(x) - f_n(x)||^2 = 0$, with probability 1.

Proof. Theorem 1 involves Theorem 2. Using (2) and the fact, that K is a kernel function, we get:

$$\begin{aligned} \alpha_{n+1} &= \alpha_n - 2\,\gamma_n\, |f(x_{n+1}) - f_n(x_{n+1})| + \gamma_n^2\, K(x_{n+1}, x_{n+1}) \\ M(\alpha_{n+1}/\alpha_1 \ldots \alpha_n) &= \alpha_n - 2\gamma_n\, M(|f(x_{n+1}) - f_n(x_{n+1})|\, \alpha_1 \ldots \alpha_n) + \\ &\quad + \gamma_n^2\, M(K(x_{n+1}, x_{n+1})/\alpha_1 \ldots \alpha_n). \end{aligned} \tag{5}$$

Let

$$\begin{aligned} U_n &\sim \alpha_n \\ V_n &\sim M(|f(x_{n+1}) - f_n(x_{n+1})|/\alpha_1 \ldots \alpha_n) \\ \gamma_n &\sim 2\gamma_n \\ \mu_n &\sim 0 \\ \eta_n &\sim \gamma_n^2 M(K(x_{n+1}, x_{n+1})/\alpha_1 \ldots \alpha_n). \end{aligned}$$

Then the assumptions of Theorem 1 are met, thus (5) implies (i), (1) and (3) imply (ii) and (iii). We prove, instead of (iv), an assertion which implies (iv). We prove that if $\lim_{n_k \to \infty} V_{n_k} = 0$ with probability 1, then $\lim_{n_k \to \infty} U_{n_k} = 0$ with probability 1. Let $\Omega' \subset \Omega$ be the set of those $\omega \in \Omega$, for which $\lim_{n_k \to \infty} V_{n_k}(\omega) = 0$ and, for all n_k, (4) also holds. I.e., for every $\omega \in \Omega'$, $V_{n_k}(\omega) \geq q(\varepsilon)$ if $U_{n_k}(\omega) \geq \varepsilon$. We prove that for every member of $\Omega'(P(\Omega') = 1)$:

$$\lim_{n_k \to \infty} U_{n_k}(\omega) = 0 .$$

We suppose an $\omega \in \Omega'$ to exist, such that $\lim_{n_k \to \infty} V_{n_k}(\omega) = 0$, however $U_{n_k} \not\to 0$. Then we have, for any $\varepsilon > 0$, some $\{n_{k_i}\} \subset \{n_k\}$, such that $U_{n_{k_i}}(\omega) \geq \varepsilon$ for all n_{k_i}. From (4) and from the definition of Ω' it follows, that

$$V_{n_{k_i}}(\omega) \geq q(\varepsilon) > 0$$

for all n_{k_i}, i.e. $V_{n_{k_i}} \not\to 0$ as $n_{k_i} \to \infty$. This contradicts the assumption $\lim_{n_{k_i} \to \infty} V_{n_{k_i}}(\omega) = 0$.

Thus all assumptions of the Theorem 1 are fulfilled. Therefore $\lim_{n \to \infty} U_n = \lim_{n \to \infty} ||f(x) - f_n(x)||^2 = 0$, with probability 1.

Remark. From $\lim_{n \to \infty} ||f(x) - f_n(x)||^2 = 0$ with probability 1, it follows that, for any $x \in X$, $\lim_{n \to \infty} f_n(x) = f(x)$ with probability 1. This is obvious, since:

$$|f_n(x) - f(x)| = |(f_n(s) - f(s), K(s), x))| \leq ||f_n(s) - f(s)|| \cdot ||K(s), x)|| =$$

$$= ||f_n(s) - f(s)|| \cdot \sqrt{K(x, x)} .$$

In algorithm (2) any label $f(x_n)$ is used only once during the iteration. In the following two theorems we deal with such algorithms, which may use any label $f(x_n)$ several times.

This fact of particular interest when adopting algorithm (2), since in this very case one utilizes, at any step, just $\operatorname{sign}(f(x_{n+1}) - f_n(x_{n+1}))$.

Theorem 3. Let H be a RKHS with K as a kernel. Let $k \geq 0$ be an arbitrary non-negative integer, for which

$$\sup_n M\left(\left(\sum_{l=n-k}^{n} \sqrt{K(x_l, x_l)}\right)^2\right) = L_3 < +\infty . \tag{7}$$

Let the algorithm be

$$f_{n+1}(x) = f_n(x) + \gamma_n \sum_{l=n+1-k}^{n+1} \operatorname{sign}\left(f(x_l) - f_n(x_l)\right) K(x, x_l) , \tag{8}$$

where $\gamma_n \geq 0$,

$$\sum_{n=0}^{\infty} \gamma_n = +\infty, \quad \sum_{n=0}^{\infty} \gamma_n^2 < +\infty \quad \text{and} \quad f_k(x) \equiv 0. \tag{9}$$

If for any $\varepsilon > 0$

$$\operatorname*{ess\,inf}_{\varepsilon \leq |f-f_n|} M\left(\sum_{l=n+1-k}^{n+1} |f(x_l) - f_n(x_l)| / \alpha_1 \ldots \alpha_n\right) \geq q(\varepsilon) > 0 \tag{10}$$

then $\lim_{n\to\infty} ||f(x) - f_n(x)|| = 0$, with probability 1.

Proof. It follows from (8), by using K as a kernel, that

$$\alpha_{n+1} = \alpha_n - 2\gamma_n \sum_{l=n+1-k}^{n+1} |f(x_l) - f_n(x_l)| + \\ + \gamma_n^2 \left\| \sum_{l=n+1-k}^{n+1} \sin n\big(f(x_l) - f_n(x_l)\big) K(x, x_l) \right\|^2,$$

and

$$M(\alpha_{n+1}/\alpha_1 \ldots \alpha_n) \leq \alpha_n - 2\gamma_n M\left(\sum_{l=n+1-k}^{n+1} |f(x_l) - f_n(x_l)| / \alpha_1 \ldots \alpha_n\right) + \\ + \gamma_n^2 M\left[\left(\sum_{l=n+1-k}^{n+1} \sqrt{K(x_l, x_l)}\right)^2 \Big/ \alpha_1 \ldots \alpha_n\right]. \tag{11}$$

Let us adopt the following substitions

$$U_n \sim \alpha_n$$

$$V_n \sim M\left(\sum_{l=n+1-k}^{n+1} |f(x_l) - f_n(x_l)| / \alpha_1 \ldots \alpha_n\right)$$

$$\gamma_n \sim 2\,\gamma_n$$

$$\mu_n \sim 0$$

$$\eta_n \sim \gamma_n^2 M\left[\left(\sum_{l=n+1-k}^{n=1} \sqrt{K(x_l, x_l)}\right)^2 \Big/ \alpha_1 \ldots \alpha_n\right].$$

All assumptions of the Theorem 1 are met. I.e. (11), (7) and (9) imply (i), (ii) and (iii), resp. Assumption (iv) is proved by using the inequality (10), in the same way as in Theorem 2. Since all conditions of Theorem 1 are met, we have $\lim_{n\to\infty} U_n = \lim_{n\to\infty} ||f(x) - f_n(x)||^2 = 0$ with probability 1.

Theorem 4. Let H be a RKHS with K as a kernel and a_n a sequence of integers, for which $1 \leq a_n \leq n$ and

$$\sup_n M\left(\left(\frac{1}{a_n}\sum_{l=n+1-a_n}^{n+1}\sqrt{K(x_l,x_l)}\right)^2\right) = L_4 < +\infty\,. \tag{12}$$

Let the algorithm be

$$f_{n+1}(x) = f_n(x) + \gamma_n \frac{1}{a_{n+1}} \sum_{l=n+2-a_{n+1}}^{n+1} \operatorname{sign}\left(f(x_l) - f_n(x_l)\right) K(x, x_l)\,. \tag{13}$$

Here $\gamma_n \geq 0$,

$$\sum_{n=0}^{\infty}\gamma_n = +\infty\,, \quad \sum_{n=0}^{\infty}\gamma_n^2 < +\infty\,, \quad f_0(x) \equiv 0\,. \tag{14}$$

If, for any $\varepsilon > 0$, $\operatorname*{ess\,inf}_{\varepsilon \leq \|f-f_n\|^2} M\left(\frac{1}{a_{n+1}}\sum_{l=n+2-a_{n+1}}^{n+1}|f(x_l) - f_n(x_l)|/\alpha_1 \ldots \alpha_n\right) \geq$

$$\geq q(\varepsilon) > 0 \tag{15}$$

then $\lim_{n\to\infty} \|f(x) - f_n(x)\|^2 = 0$, with probability 1.

Proof. (13) implies that

$$\alpha_{n+1} = \alpha_n - 2\gamma_n \frac{1}{a_{n+1}} \sum_{l=n+2-a_{n+1}}^{n+1} |f(x_l) - f_n(x_l)|$$

$$+ \gamma_n^2 \left\| \frac{1}{a_{n+1}} \sum_{l=n+2-a_{n+1}}^{n+1} \operatorname{sign}\left(f(x_l) - f_n(x_l)\right) K(x, x_l) \right\|^2$$

and

$$M(\alpha_{n+1}/\alpha_1 \ldots \alpha_n) \leq \alpha_n - 2\gamma_n M\left(\frac{1}{a_{n+1}} \sum_{l=n+2-a_{n+1}}^{n+1} \|f(x_l) - f_n(x_l)\|/\alpha_1 \ldots \alpha_n\right) +$$

$$+ \gamma_n^2 M\left[\left(\frac{1}{a_{n+1}} \sum_{l=n+2-a_{n+1}}^{n+1} \sqrt{K(x_l, x_l)}\right)^2 \Big/ \alpha_1 \ldots \alpha_n\right]. \tag{16}$$

Adopting the following substitutions:

$$U_n \sim \alpha_n$$

$$\gamma_n \sim 2\gamma_n$$

$$\mu_n \sim 0$$

$$V_n \sim M\left(\frac{1}{a_{n+1}} \sum_{l=n+2-a_{n+1}}^{n+1} |f(x_l) - f_n(0_l)|/\alpha_1 \ldots \alpha_n\right)$$

$$\eta_n \sim \gamma_n^2 M\left[\left(\frac{1}{a_{n+1}} \sum_{l=n+2-a_{n+1}}^{n+1} \sqrt{K(x_l, x_l)}\right)^2 \Big/ \alpha_1 \ldots \alpha_n\right].$$

Proving that all assumption of the Theorem 1 hold follows the same line as described in Theorems 2 and 3 [using (12), (14), (15), (16)].

For the use of Theorems 3 and 4 we refer to the following specific problem. One wishes to use the teaching several times, but only N points may be stored. One may take $k = N$ according to Theorem 3. In this case, for $1 \leq i \leq N$, $f(x_i)$ is used i-times, and, for $i > N$, N-times.
In the algorithm of Theorem 4 taking, for $n \leq N$, $n_n = n$ and $a_n = N$ otherwise, we use each teaching just N-times.

Theorem 5 deals with such cases when, at the n-th step, we teach, instead of f, the value of some $\hat{f}_n = \{\hat{f}_n(x), x \circ X\}$. Theorem 5 answers the question, under which conditions, for $\hat{f}_n$, does the algorithm tend to f. ($\hat{f}_n$ may depend on the random variables $x_1 \ldots x_n$).

Example. Let $f(x) = \sum_{i=1}^{N} c_i F_i(x)$, where $c_i \geq 0$, $\sum_{i=1}^{N} c_i = 1$ and $F_i = \{F_i(x), x \in X\}$ are linearly independent distribution functions.

Let us estimate the coefficients c_i, if a sequence $x_1 \ldots x_n \ldots$ of independent and identically distributed random variables with distribution $f(x)$ is given. At the n-th step we compute the values of the function $F_n(x)$ from the random variables $x_1 \ldots x_n$.

Theorem 5. H is a RKHS with K as a kernel, for which

$$\sup_l M\big(K(x_l, x_l)\big) = L_5 < +\infty. \tag{17}$$

Let the algorithm be

$$f_{n+1}(x) = f_n(x) + \gamma_n \operatorname{sign}\left(\hat{f}_n(x_{n+1}) - f_n(x_{n+1})\right) K(x, x_{n+1}), \tag{18}$$

where $\gamma_n \geq 0$

$$\sum_{n=0}^{\infty} \gamma_n = +\infty, \quad \sum_{n=0}^{\infty} \gamma_n^2 < +\infty, \quad f_0(x) \equiv 0 \tag{19}$$

$\hat{f}_n$ may depend on the random variables $x_1 \ldots x_n$. In addition,

$$\sum_{n=0}^{\infty} \gamma_n M\big(|\hat{f}_n(x_{n+1}) - f(x_{n+1})|\big) < +\infty. \tag{20}$$

If, for any $\varepsilon > 0$,

$$\operatorname*{ess\,inf}_{\varepsilon \leq \|f - f_n\|} M\big(|\hat{f}_n(x_{n+1}) - f(x_{n+1})| \,/\, \alpha_1 \ldots \alpha_n\big) \geq q(\varepsilon) > 0 \tag{21}$$

then $\lim_{n \to \infty} \|f(x) - f_n(x)\|^2 = 0$, with probability 1.

Proof. (18) implies that

$$\alpha_{n+1} = \alpha_n - 2\gamma_n \operatorname{sign}\left(\hat{f}_n(x_{n+1}) - f_n(x_{n+1})\right)\left(f(x_{n+1}) - f_n(x_{n+1})\right) + \\ + \gamma_n^2 K(x_{n+1}, x_{n+1}) .$$

Thus

$$\alpha_{n+1} = \alpha_n - 2\,\gamma_n |\hat{f}_n(x_{n+1}) - f_n(x_{n+1})| - 2\gamma_n \operatorname{sign}\left(\hat{f}_n(x_{n+1}) - \right. \\ \left. - f_n(x_{n+1})\right)\left(f(x_{n+1}) - \hat{f}_n(x_{n+1})\right) + \gamma_n^2 K(x_{n+1}, x_{n+1})$$

and

$$M(\alpha_{n+1}/\alpha \ldots \alpha_n) \leq \alpha_n - 2\,\gamma_n M(|\hat{f}_n(x_{n+1}) - f_n(x_{n+1})|/\alpha_1 \ldots \alpha_n) + \\ + M(2\,\gamma_n |f(x_{n+1}) - \hat{f}_n(x_{n+1})| + \gamma_n^2 K(x_{n+1})/\alpha_1 \ldots \alpha_n) . \qquad (22)$$

Adopting again

$$U_n \sim \alpha_n$$

$$V_n \sim M(|f_n(x_{n+1}) - f_n(x_{n+1})|/(\alpha_1 \ldots \alpha_n)$$

$$\gamma_n \sim 2\,\gamma_n$$

$$\mu_n \sim 0$$

$$\eta_n \sim M(2\,\gamma_n |f(x_{n+1}) - \hat{f}_n(x_{n+1})| + \gamma^2 K(x_{n+1}, x_{n+1})/\alpha_1 \ldots \alpha_n) .$$

Here (22) and (19) imply (i) and (ii) resp., (17), (19) and (20) imply (iii), and (iv) holds because of (21). In the previous theorems f was an element of an RKHS.

Next f is a member of any arbitrary Hilbert-space of functions.

Theorem 6. Let $K_n = K_n(x, y)$, $(x, y) \in X \times X (n = 1, 2 \ldots)$ be a sequence of functions for which $K_n \in H$, if $y \in X$ and n are arbitrary. Let the algorithm be:

$$f_{n+1}(x) = f_n(x) + \gamma_n \operatorname{sign}\left(f(k_{n+1}) - f_n(x_{n+1})\right) K_n(x, x_{n+1}) . \qquad (23)$$

Here

$$\gamma_n \geq 0 , \quad \sum_{n=0}^{\infty} \gamma_n = +\infty , \quad \lim_{n \to \infty} \gamma_n = 0 , \quad f_0(x) \equiv 0 , \qquad (24)$$

$$\sum_{n=0}^{\infty} \gamma_n^2 M\left(\| K_n(x, y) \|^2\right) < +\infty . \qquad (25)$$

In addition, let us assume that

$$\sum_{n=0}^{\infty} \gamma_n M\left(|(f(x)-f_n(x), K_n(x, x_{n+1})) - (f(x_{n+1}) - f_n(x_{n+1}))|\right) < +\infty . \quad (26)$$

Assume that, for any $\varepsilon > 0$,

$$\operatorname*{ess\,inf}_{\varepsilon \leq ||f-f_n||^2} M(|f(x_{n+1}) - f_n(x_{n+1})|/\alpha_1 \ldots \alpha_n) \geq q(\varepsilon) > 0 \quad (27)$$

then $\lim_{n\to\infty} ||f(x) - f_n(x)||^2 = 0$ with probability 1.

Proof. (23) implies that

$$\alpha_{n+1} = \alpha_n - 2\gamma_n \operatorname{sign}\left(f(x_{n+1}) - f_n(x_{n+1})\right)\left(f(x) - f_n(x), K_n(x, x_{n+1})\right) + \\ + \gamma_n^2 ||K_n(x, x_{n+1})||^2$$

Thus

$$\alpha_{n+1} \leq \alpha_n - 2\gamma_n |f(x_{n+1}) - f_n(x_{n+1})| + \gamma_n^2 ||K_n(x_1 x_{n+1})||^2 + \\ + 2\gamma_n\left(|(f(x) - f_n(x), K_n(x_1 x_{n+1})) - (f(x_{n+1}) - f_n(x_{n+1}))|\right)$$

and

$$M(\alpha_{n+1}/\alpha_1 \ldots \alpha_n) \leq \alpha_n - 2\gamma_n M(|f(k_{n+1}) - f_n(x_{n+1})|/\alpha_1 \ldots \alpha_n) + \\ + 2\gamma_n M\left(|f((x) - f_n(x), K_n(x, x_{n+1})) - f(x_{n+1}) - (f_n(x_{n+1}))|/\alpha_1 \ldots \alpha_n\right) + \\ + \gamma_n^2 M(||K_n(x, x_{n+1})||^2/\alpha_1 \ldots \alpha_n). \quad (28)$$

Adopt

$$U_n \sim \alpha_n$$

$$V_n \sim M\left(|(f_{n+1}) - f_n(x_{n+1})|/\alpha_1 \ldots \alpha_n\right)$$

$$\gamma_n \sim 2\gamma_n$$

$$\mu_n \sim 0$$

$$\eta_n \sim M\{2\gamma_n |(f(x) - f_n(x), K(x, x_{n+1})) - (f(x_{n+1} - f_n(x_{n+1}))| + \\ + \gamma_n^2 ||K_n(x, x_{n+1})||^2/\alpha_1 \ldots \alpha_n\}.$$

All conditions of Theorem 1 hold, viz.:

$$(28) \rightarrow \text{(i)}$$

$$(24) \rightarrow \text{(ii)}$$

$$(25),\ (26) \rightarrow \text{(iii)}$$

$$(27) \rightarrow \text{(iv)}$$

An application of Theorem 6

Let H be a separable Hilbert space, and $\varphi_1 \dots \varphi_n \dots$ be a base in H. Let $K_n(x, y) = \sum_{i=1}^{n} \varphi_i(x)\, \varphi_i(y)$.

Then (25) holds, if

$$\sum_{n=0}^{\infty} \gamma_n^2 M\left(\sum_{i=1}^{n} \varphi_i^2(x_{n+1})\right) < +\infty$$

or

$$\sum_{n=0}^{\infty} \gamma_n^2 \sum_{i=1}^{n} M\left(\varphi_i^2(x_{n+1})\right) < +\infty$$

From (26), and $\quad f(x) = \sum_{i=1}^{\infty} c_i\, \varphi_i(x) \quad (f \in H)$

we have

$$\sum_{n=1}^{\infty} \gamma_n M\left(\left|\sum_{i=n+1}^{\infty} c_i \varphi_i(x_{n+1})\right|\right) < +\infty. \tag{29}$$

(Let $f_n(x) = \sum_{i=1}^{n} d_i\, \varphi_i(x)$). Observe, that (29) is independent of f_n, thus (in this specific case) (26) is independent of the algorithm. Let H' denote the set of those $f \in H$ for which (29) is met.

H' obviously depends, on the series γ_n. It is easy to see from (29) that, if $f \in H'$, (29) is met, then, for any real α, $\alpha f \in H'$, and $f, g \in H'$ implies $f + g \in H'$. Hence H' is a linear space.

If $f(x) = \sum_{i=1}^{N} c_i \varphi_i(x)$, then

$$\sum_{n=1}^{\infty} \gamma_n M\left(\left|\sum_{i=n+1}^{\infty} c_i \varphi_i(x_{n+1})\right|\right) = \sum_{n=1}^{N-1} \gamma_n M\left|\sum_{i=n+1}^{N} c_i \varphi_i(x_{n+1})\right| < +\infty.$$

Thus, if $M(|\varphi_i(x_j)|) < +\infty$, for any $1 \le i \le N$, $1 \le j \le N$, then (30) holds.

Part II

Let $(\Omega, \mathcal{A}, P)$ be a probability space, and (X, Z) be an arbitrary measurable space. Let H denote some reproducing kernel Hilbert space (RKHS) (see [1]) of real valued functions defined on X by the kernel function $K = \{K(x, y)\, x, y \in X\}$. Assume a sequence of random variables $x_1, x_2 \dots x_n \dots$ taking values in X and the reals $f(x_1), f(x_2) \dots f(x_n) \dots$ to be given (f denotes some given member of H). Our problem is to estimate f by using these data.

We are concerned with such algorithms, for which f_n is an approximation of f at the n-th step, and f_n is measurable with respect to the random variables $x_1, x_2 \dots x_n$.

Let $\alpha_n = ||f - f_n||^2$.

We call $x_1, x_2 \dots x_n \dots$ samples, and $f(x_1), f(x_2) \dots f(x_n) \dots$ labels.

Theorem 1. Let us suppose, that

$$\operatorname*{ess\,sup}_{\substack{x_n \\ n=1,2\dots}} K(x_n, x_n) = L_1 < +\infty . \tag{1}$$

Let $f_{n+1}(x) = f_n(x) + \Theta\big(f(x_{n+1}) - f_n(x_{n+1})\big) K(x, x_{n+1})$, $f_0(x) \equiv 0$. (2)

$$0 < \Theta < \frac{2}{L_1} \quad \text{stands for an arbitrary constant.} \tag{3}$$

Then
$$\sum_{n=0}^{\infty} \big(f(x_{n+1}) - f_n(x_{n+1})\big)^2 \leqq \frac{||f||^2}{2\Theta - \Theta^2 L_1} \text{ a.s.} \tag{4}$$

Proof. (2) implies that

$$||f - f_{n+1}||^2 = ||f - f_n||^2 - 2\big(f(x) - f_n(x), \Theta\big(f(x_{n+1}) - f_n(x_{n+1}) K(x, x_{n+1})\big)\big) +$$
$$+ \Theta^2 \big(f(x_{n+1}) - f_n(x_{n+1})\big)^2 \, ||K(x, x_{n+1})||^2 . \tag{5}$$

Since K is the kernel function of H, we have from (5)

$$\alpha_{n+1} = \alpha_n - 2\,\Theta\big(f(x_{n+1}) - f_n(x_{n+1})\big)^2 + \Theta^2 \big(f(x_{n+1}) - f_{n+1})\big)^2 K(x_{n+1}, x_{n+1})$$

Let us use (1)

$$\alpha_{n+1} \leqq \alpha_0 - (2\Theta - \Theta^2 L_1)\big(f(x_{n+}) - f_n(x_{n+1})\big)^2 \qquad \text{a.s.}$$

$$\alpha_{n+1} \leqq \alpha_0 - (2\Theta - \Theta^2 L_1) \sum_{i=o}^{n} \big(f(x_{i+1}) - f_{i(i+1)}\big)^2 \qquad \text{a.s.}$$

$$\sum_{=0}^{n} \big(f(x_{i+1}) - f_i(x_{i+1})\big)^2 \leqq \frac{\alpha_0}{2\Theta - \Theta^2 L_1} \qquad \text{a.s. for all } n.$$

Since the sequence $\sum_{i=0}^{n} (f(x_{i+1}) - f_i(x_{i+1}))^2$ is monotone increasing and a.s. bounded, it is convergent and

$$\sum_{i=0}^{\infty} \big(f(x_{i+1}) - f_i(x_{i+1})\big)^2 \leqq \frac{||f||^2}{2\Theta - \Theta^2 L_1} \quad \text{with probability 1.}$$

Remark 1.1. If the assumptions in Theorem 1 are met and $x_1, x_2 \ldots$ are independent and identically distributed with the distribution function Q, then

$$\sum_{n=0}^{\infty} \int_X (f(x) - f_n(x))^2 Q(dx) < +\infty \text{ a.s.}$$

Observe that (4) implies

$$M\left(\sum_{n=0}^{\infty} (f(x_{n+1}) - f_n(x_{n+1}))^2\right) \leq \frac{||f||^2}{2\Theta - \Theta^2 L_1}.$$

Using Lebesque's Theorem

$$\sum_{n=0}^{\infty} M\left[(f(x_{n+1}) - f_n(x_{n+1}))^2\right] \leq \frac{||f||^2}{2\Theta - \Theta^2 L_1}$$

$$\sum_{n=0}^{\infty} M\left[M\left((f(x_{n+1}) - f_n(x_{n+1}))^2 / x_1 \ldots x_n\right)\right] \leq \frac{||f||^2}{2\Theta - \Theta^2 L_1}.$$

Since $x_1, x_2, \ldots x_n \ldots$ are independent and identically distributed

$$\sum_{n=0}^{\infty} M\left(\int_X (f(x) - f_n(x))^2 Q(dx)\right) \leq \frac{||f||^2}{2\Theta - \Theta^2 L_1}$$

and

$$M\left(\sum_{n=0}^{\infty} \int_X (f(x) - f_n(x))^2 Q(dx)\right) \leq \frac{||f||^2}{2\Theta - \Theta^2 L_1}.$$

Since the random variable $\sum_{n=0}^{\infty} \int_X (f(x) - f_n(x))^2 Q(dx)$ has finite expectation, we have

$$\sum_{n=0}^{\infty} \int_X (f(x) - f_n(x))^2 Q(dx) < +\infty \text{ a.s.} \tag{6}$$

In the application of this theorem the question is how to choose the value of Θ, since for doing so either the value L_1 or an upper bound of this is to be known. The next theorem settles this problem.

Theorem 2. Assume, that

$$K(x_n, x_n) \neq 0 \quad \text{a.s.} \tag{7}$$

for all n, and

$$f_{n+1}(x) = \begin{cases} f_n(x) + \dfrac{f(x_{n+1}) - f_n(x_{n+1})}{K(x_{n+1}, x_{n+1})} K(x, x_{n+1}), & \text{if } K(x_{n+1}, x_{n+1}) \neq 0, \\ f_n(x) & \text{otherwise,} \end{cases} \tag{8}$$

and $f_0(x) \equiv 0$.

Then

$$\lim_{n\to\infty} \|f - f_n\|^2 = \|f\|^2 - \sum_{n=0}^{\infty} \frac{(f(x_{n+1}) - f_n(x_{n+1}))^2}{K(x_{n+1}, x_{n+1})} \text{ a.s.} \tag{9}$$

Proof. (7) and (8) imply that

$$f_{n+1}(x) = f_n(x) + \frac{f(x_{n+1}) - f_n(x_{n+1})}{K(x_{n+1}, x_{n+1})} K(x, x_{n+1}) \text{ a.s.}$$

thus

$$\alpha_{n+1} = \alpha_n - 2\frac{(f(x_{n+1}) - f_n(x_{n+1}))^2}{K(x_{n+1}, x_{n+1})} + \frac{(f(x_{n+1}) - f_n(x_{n+1}))^2}{K^2(x_{n+1}, x_{n+1})} K(x_{n+1}, x_{n+1}) \quad \text{a.s.}$$

$$\alpha_{n+1} = \alpha_n - \frac{(f(x_{n+1}) - f_n(x_{n+1}))^2}{K(x_{n+1}, x_{n+1})} \quad \text{a.s.}$$

$$\alpha_{n+1} = \alpha_0 - \sum_{i=0}^{\infty} \frac{(f(x_{i+1}) - f_i(x_{i+1}))^2}{K(x_{i+1}, x_{i+1})} \quad \text{a.s.} \tag{10}$$

Since $\{\alpha_n\}_{n=0}^{\infty}$ is monotone decreasing and non-negative, it is convergent, too,

$$\lim_{n\to\infty} \alpha_n = \lim_{n\to\infty} \|f - f_n\|^2 = \|f\|^2 - \sum_{i=0}^{\infty} \frac{(f(x_{i+1}) - f_i(x_{i+1}))^2}{K(x_{i+1}, x_{i+1})} \quad \text{a.s.} \tag{11}$$

Remark 2.1. From (11) we have

$$\sum_{n=0}^{\infty} \frac{(f(x_{n+1}) - f_n(x_{n+1}))^2}{K(x_{n+1}, x_{x+1})} \leq \|f\|^2 \quad \text{a.s.} \tag{12}$$

Remark 2.2. If $\operatorname{ess\,sup}_{\substack{x_n \\ R=1,2\ldots}} K(x_n, x_n) = L_2 < +\infty$

then (12) implies

$$\frac{1}{L_2} \sum_{n=0}^{\infty} (f(x_{n+1}) - f_n(x_{n+1}))^2 \leq \|f\|^2 \quad \text{a.s.}$$

$$\sum_{n=0}^{\infty} (f(x_{n+1}) - f_n(x_{n+1}))^2 \leq L_2 \|f\|^2 \quad \text{a.s.}$$

Remark 2.3. If the conditions of Theorem 2 hold and $x_1, x_2 \ldots$ are independently distributed according to Q, then

$$\sum_{n=1}^{\infty} \int_X \frac{(f(x) - f_n(x))^2}{K(x, x)} Q(dx) < +\infty \quad \text{a.s.}$$

(The proof of Remark 2.3 is the same, as that of Remark 1.1.)

By the next theorem we examine the rate of the convergence of the algorithm adopted in Theorem 2.

First of all we prove

Lemma 1. Let Q be a probability measure on (X, Z). Let us suppose, that $Q(K(x, x) \neq 0) = 1$.

Let E denote the set $\{f : f \in H \, \|f\| = 1\}$, and $F \subset E$ some set, which is dense everywhere in E. If

$$\inf_{f \in F} \int_X \frac{f^2(x)}{K(x, x)} Q(dx) \overset{\text{Def.}}{=\!=} r > 0 , \tag{13}$$

then for every $f \in H$

$$r \|f\|^2 \leq \int_X \frac{f^2(x)}{K(x, x)} Q(dx) \overset{\text{Def.}}{=\!=} M_x \left(\frac{f^2(x)}{K(x, x)} \right) . \tag{14}$$

Proof. It may be readily proved, that the set $F = \{\lambda f : f \in F, \lambda \in R^1\}$ is dense everywhere in H and for all $f \in F'$ (14) holds, so we need only to prove that if

$$\lim_{n \to \infty} f_n = f \tag{15}$$

and

$$r \|f\|^2 \leq \int_X \frac{|f_n(x)|^2}{K(x, x)} Q(dx) \quad \text{for each } n, \tag{16}$$

then

$$r \|f\|^2 \leq \int_X \frac{|f(x)|^2}{K(x, x)} Q(dx) .$$

If $\{f_n\}_{n=0}^{\infty}$ is convergent in the RKHS H, then

$$\lim_{n \to \infty} \|f_n\|^2 = \|f\|^2 \tag{17}$$

and for all $x \in X$

$$\lim_{n \to \infty} f_n(x) = f(x) . \qquad \text{(See [1])} \tag{18}$$

Since $\|f_n\|$ is convergent, it is also bounded and

$$\frac{|f_n(x)|^2}{K(x, x)} = \frac{|(f_n(s), K(s, x))|^2}{K(x, x)} \leq \|f_n\|^2 .$$

Thus the sequence of the functions $\frac{|f_n(x)|}{K(x, x)}$ is convergent for every $x \in X$, and

bounded, and, therefore,

$$\lim_{n\to\infty} M_x\left(\frac{|f_n(x)|^2}{K(x,x)}\right) = M_x\left(\lim_{n\to\infty}\frac{|f_n(x)|^2}{K(x,x)}\right) = M_x\left(\frac{|f(x)|^2}{K(x,x)}\right). \tag{19}$$

Using (17) and (19) at the inequality (16) we have

$$r\,\|f\|^2 \le \int_X \frac{|f(x)|^2}{K(x,x)}\,Q(dx) \qquad \text{for all } f\in H. \tag{20}$$

Remark. (20) implies

$$r\,\|f\|^2 \le M_x\left(\frac{|f(x)|^2}{K(x,x)}\right) = M_x\left(\frac{|(f(s),K(s,x))|^2}{K(x,x)}\right) \le M_x\left(\frac{\|f\|^2\,K(x,x)}{K(x,x)}\right) = \|f\|^2. \tag{21}$$

Thus: $0 < r \le 1$.

Theorem 3. Let $x_1, x_2 \ldots x_n \ldots$ be independent and identically distributed random variables with distribution function Q, for which

$$Q(K(x,x) \neq 0) = 1. \tag{22}$$

Assume, that

$$\inf_{f\in F}\int_X \frac{f^2(x)}{K(x,x)}\,Q(dx) = r > 0, \tag{23}$$

here $F \subset E = \{f : f \in H,\ \|f\| = 1\}$ and F is dense everywhere in E.

In the algorithm is defined as

$$f_{n+1}(x) = \begin{cases} f_x(x) + \dfrac{f(x_{n+1}) - f_n(x_{n+1})}{K(x_{n+1},x_{n+1})}\,K(x,x_{n+1}), & \text{if } K(x_{n+1},x_{n+1}) \neq 0 \\ f_n(x) & \text{, otherwise,} \end{cases} \tag{24}$$

then

$$M(\|f - f_n\|) \le \|f\|^2\,(1-r)^n. \tag{25}$$

Proof. From (22) and (24) we have

$$\alpha_{n+1} = \alpha_n - \frac{(f(x_{n+1}) - f_n(x_{n+1}))^2}{K(x_{n+1},x_{n+1})}$$

a.s. (See Theorem 2)

$$M(\alpha_{n+1}) = M(\alpha_n) - M\left[M\left(\frac{(f(x_{n+1}) - f_n(x_{n+1}))^2}{K(x_{x+1},x_{n+1})}\,\middle/\,x_1 \ldots x_n\right)\right]. \tag{26}$$

Since $x_1, x_2 \ldots x_n \ldots$ are independent and Q distributed, therefore

$$M\left(\frac{(f(x_{n+1}) - f_n(x_{n+1}))^2}{K(x_{n+1}, x_{n+1})} \Big/ x_1 \ldots x_n\right) = \int_X \frac{(f(x) - f_n(x))^2}{K(x, x)} Q(dx). \tag{27}$$

For the measure Q (23) holds, we may use, therefore, Lemma 1

$$\int_X \frac{|f(x) - f_n(x)|^2}{K(x, x)} Q(dx) \geq r \|f - f_n\|^2 = r\alpha_n. \tag{28}$$

From (27) and (28) at (26)

$$M(\alpha_{n+1}) \leq M(\alpha_n) - rM(\alpha_n)$$

$$M(\alpha_{n+1}) \leq M(\alpha_n)(1 - r)$$

$$M(\alpha_{n+1}) \leq M(\alpha_0)(1 - r)^{n+1} = \|f\|^2(1 - r)^{n+1}. \tag{29}$$

Remark 3.1. (29) implies

$$\lim_{n \to \infty} \|f - f_n\|^2 = 0$$

a.s., since from (29) we have

$$\sum_{n=0}^{\infty} M(\alpha_n) \leq \|f\|^2 \frac{1}{r}$$

$$M\left(\sum_{n=1}^{\infty} \alpha_n\right) \leq \|f\|^2 \frac{1}{r}$$

therefore

$$\sum_{n=0}^{\infty} \alpha_n < +\infty \qquad \text{a.s.}$$

In the following theorems we examine algorithm devised specifically for pattern recognition. Let $A, B \subset X$ be disjoint sets and $f(x)$ is positive on A and negative on B. f is to be estimated. In [2] and [3] we may find such algorithms. If $f \in H$ and H is a RKHS, then

$$f_{n+1}(x) = f_n(x) + (\operatorname{sign} f(x_{n+1}) - \operatorname{sign} f_n(x_{n+1}))\, K(x, x_{n+1}).$$

This algorithm proceeds only if

$$\operatorname{sign} f(x_{n+1}) \neq \operatorname{sign} f_n(x_{n+1}).$$

Let us have in what follows

$$A_n = \{x;\ \operatorname{sign} f(x) \neq \operatorname{sign} f_n(x)\} \tag{30}$$

and call A_n the error-set in the n-th step.

Theorem 4. Assume, that $K(x_n, x_n) \neq 0$ a.s. and (31)

$$f_{n+1}(x) = \begin{cases} f_n(x) + \mathcal{X}_{A_n}(x_{n+1}) \dfrac{f(x_{n+1}) - f_n(x_{n+1})}{K(x_{n+1}, x_{n+1})} K(x, x_{n+1}), & \text{if } K(x_{n+1}, x_{n+1}) \neq 0, \\ f_n(x), & \text{ortherwise,} \end{cases} \tag{32}$$

$f_0(x) \equiv 0$ ($\mathcal{X}_{A_n}(x)$ is the indicator of A_n).

Then

$$\sum_{n=0}^{\infty} \mathcal{X}_{A_n}(x_{n+1}) \frac{(f(x_{n+1}) - f_n(x_{n+1}))^2}{K(x_{n+1}, x_{n+1})} < \|f\|^2 \quad \text{a.s.} \tag{33}$$

Proof. From (31) and (32) we have, that

$$\alpha_{n+1} = \alpha_n - \mathcal{X}_{A_n}(x_{n+1}) \frac{(f(x_{n+1}) - f_n(x_{n+1}))^2}{K(x_{n+1}, x_{n+1})} \quad \text{a.s.}$$

$$\alpha_{n+1} = \alpha_0 - \sum_{i=0}^{n} \mathcal{X}_{A_i}(x_{i+1}) \frac{(f(x_{i+1}) - f_i(x_{i+1}))^2}{K(x_{i+1}, x_{i+1})} \quad \text{a.s.}$$

$$\sum_{i=0}^{\infty} \mathcal{X}_{A_i}(x_{i+1}) \frac{(f(x_{i+1}) - f_i(x_{i+1}))^2}{K(x_{i+1}, x_{i+1})} \leq \|f\|^2 \quad \text{a.s.} \tag{34}$$

Theorem 5. Let us suppose, that the conditions in Theorem 4 hold and

$$\operatorname*{ess\,inf}_{\substack{x_n \\ n=1,2\ldots}} \frac{f^2(x_n)}{K(x_n, x_n)} = b > 0\,. \tag{35}$$

Then the number of actually correcting steps is less then entier $\left(\dfrac{\|f\|^2}{b}\right)$ and

$$\sum_{n=0}^{\infty} P(x_{n+1} \in A_n) \leq \frac{\|f\|}{b}\,. \tag{36}$$

Proof. From Theorem 4 we have

$$\sum_{n=0}^{\infty} \mathcal{X}_{A_n}(x_{n+1}) \frac{(f(x_{n+1}) - f_n(x_{n+1}))^2}{K(x_{n+1}, x_{n+1})} \leq \|f\|^2 \quad \text{a.s.}$$

Using the definition of A_n (30) we have

$$\mathcal{X}_{A_n}(x_{n+1})\,(f(x_{n+1}) - f_n(x_{n+1}))^2 = \mathcal{X}_{A_n}(x_{n+1})\,(|f(x_{n+1})| + |f_n(x_{n+1}|)^2 \geq$$
$$\geq \mathcal{X}_{A_n}(x_{n+1})\,|f(x_{n+1})|^2$$

and therefore

$$\sum_{n=0}^{\infty} \mathcal{X}_{A_n}(x_{n+1}) \frac{|f(x_{n+1})|^2}{K(x_{n+1}, x_{n+1})} \leq \|f\|^2 \qquad \text{a.s.}$$

By (35)

$$b \sum_{n=0}^{\infty} \mathcal{X}_{A_n}(x_{n+1}) \leq \|f\|^2 \qquad \text{a.s.}$$

$$\sum_{n=0}^{\infty} \mathcal{X}_{A_n}(x_{n+1}) \leq \frac{\|f\|^2}{b}. \qquad \text{a.s.} \qquad (37)$$

From (37) the first assertion of Theorem 5 follows, and

$$M\left(\sum_{n=0}^{\infty} \mathcal{X}_{A_n}(x_{n+1})\right) \leq \frac{\|f\|^2}{b}$$

$$\sum_{n=0}^{\infty} M\left(\mathcal{X}_{A_n}(x_{n+1})\right) \frac{\|f\|^2}{b}$$

$$\sum_{n=0}^{\infty} P(x_{n+1} \in A_n) \leq \frac{\|f\|^2}{b}.$$

Lemma 2. If $f(x)$ is a random variable defined on (X, Z, Q) such that $Q(f(x) > 0) = 1$, then from

$$\sum_{n=0}^{\infty} \int_{A_n} f(x)\, Q(dx) < +\infty \qquad (38)$$

follows:

$$\lim_{n \to \infty} Q\left(\sum_{i=n}^{\infty} A_i\right) = 0 \qquad (A_n \in Z \;\; n = 1, 2 \ldots) \qquad (39)$$

Theorem 6. If the conditions in Theorem 4 hold, $x_1, x_2 \ldots$ are independent and Q-distributed random variables and

$$Q\left(\frac{f^2(x)}{K(x, x)} > 0\right) = 1$$

then

$$\lim_{n \to \infty} Q\left(\sum_{i=n}^{\infty} A_i\right) = 0 \qquad \text{a.s.}$$

Proof. In Theorem 5 we have

$$\sum_{n=0}^{\infty} \mathcal{X}_{A_n}(x_{n+1}) \frac{f^2(x_{n+1})}{K(x_{n+1}, x_{n+1})} \leq \|f\|^2 \qquad \text{a.s.}$$

$$\sum_{n=0}^{\infty} M\left(\mathcal{X}_{A_n}(x_{n+1}) \frac{f^2(x_{n+1})}{K(x_{n+1}, x_{n+1})}\right) \leq \|f\|^2 . \tag{40}$$

Since $x_1, x_2 \ldots x_n \ldots$ are independent and Q-distributed random variable

$$M\left(\mathcal{X}_{A_n}(x_{n+1}) \frac{f^2(x_{n+1})}{K(x_{n+1}, x_{n+1})}\right) = M\left(M\left(\mathcal{X}_{A_n}(x_{n-1}) \frac{f^2(x_{n+1})}{K(x_{n+1}, x_{n+1})} \Big/ x_1 \ldots x_n\right)\right) =$$

$$= M\left(\int_X \mathcal{X}_{A_n}(x_{n+1}) \frac{f^2(x_{n+1})}{K(x_{n+1}, x_{n+1})} Q(dx_{n+1})\right) = M\left(\int_{A_n} \frac{f^2(x)}{K(x, x)} Q(dx)\right) \tag{41}$$

Using (40) and (41) we have

$$\sum_{n=0}^{\infty} M\left(\int_{A_n} \frac{f^2(x)}{K(x, x)} Q(dx)\right) \leq \|f\|^2$$

$$M\left(\sum_{n=0}^{\infty} \int_{A_u} \frac{f^2(x)}{K(x, x)} Q(dx)\right) \leq \|f\|^2$$

$$\sum_{n=0}^{\infty} \int_{A_n} \frac{f^2(x)}{K(x, x)} Q(dx) < +\infty \qquad \text{a.s.} \tag{42}$$

For all $\omega \in \Omega$, for which

$$\sum_{n=0}^{\infty} \int_{A_n} \frac{f^2(x)}{K(x, x)} Q(dx) < +\infty,$$

we may use Lemma 2, because of (42), a.s. Thus

$$\lim_{n \to \infty} Q\left(\sum_{i=n}^{\infty} A_i\right) = 0 \qquad \text{a.s.}$$

References

1. *Aronszajn, N.:* Theory of reproducing kernels. Trans. Amer. Math. Soc., 68 (1950).
2. *Браверман Э. М.:* О методе потенциальных функций. Автоматика и телемеханика, М., 12 (1965).
3. *Gulyás, O.:* On extended potential learning algorithms and their convergence rate. Problems of Control and Information Theory, **1**, 1, (1971).

О сходимости алгоритмов обучения потенциальных функций

Л. ДЁРФИ

(Будапешт)

Резюме

Приводится ряд теорем, посвященных сходимости алгоритмов, основанных на методе потенциальных функций [2]. Рассматривается многократное обучение, а также обучение с помехами. Особое внимание уделяется выбору потенциальных функций в случае, когда известно только, что потенциальная функция является элементом произвольного гильбертова пространства.

L. Győrfi
Telecommunication Research Institute
Budapest 2, Gábor Á. u. 65, Hungary

Problems of Control and Information Theory, Vol. 1 (3—4), pp. 267—276 (1972)

ЗАДАЧА ИДЕНТИФИКАЦИИ ДИНАМИЧЕСКИХ СИСТЕМ В ФАКТОРПРОСТРАНСТВЕ

Ю. П. ЛЕОНОВ

(Москва)

(Поступила в редакцию 1 сентября 1971 г.)

Рассматривается задача оценки весовой функции линейной системы по квадратичным критериям близости. Доказывается существование обобщенных решений в гильбертовом пространстве с энергетической нормой. Показывается, что некорректность задачи связана с обобщенным характером решения. При этом приближение решения равносильно регуляризации задачи.

Задачи идентификации динамических систем в последнее время привлекают все больше внимания [1]. Причиной этого является, во-первых, большой научно-познавательный интерес к этим задачам как обратным задачам математической физики, и, во-вторых, чисто вычислительные трудности, вызванные некорректностью этих задач [2]. Последние связаны с обобщенным характером решения задачи и большой чувствительностью решения к вариациям экспериментальных данных.

Несмотря на большое количество работ, появившихся за последнее время, до сих пор не было получено решения задачи идентификации.

В [3] предложено приближенное решение основного уравнения теории оптимальных статистических систем в пространстве с энергетической нормой. В работах американских исследователей, появившихся позднее, была сделана попытка получить решение задачи идентификации таким же методом. Однако эту попытку нельзя признать удачной, так как не удалось получить решения задачи минимума квадратичного функционала с положительным вполне непрерывным оператором [4]. Развивая метод энергетических пространств, в [3] получено решение задачи о минимуме квадратичного функционала в факторпространстве и, таким образом, найдено решение задачи идентификации для квадратичных критериев ошибки. Ниже это решение излагается.

1. Задача о минимуме квадратичного функционала

Применение квадратичных критериев в задаче идентификации приводит к проблеме минимума квадратичного функционала. Это имеет место как для статистического, так и для детерминированного сигналов.

Действительно, пусть необходимо определить характеристику линейной системы $\omega(\tau)$ с конечной памятью T_1, если наблюдаются сигналы $x(t)$, $u(t)$; $0 \leq t \leq T$, где $x(t)$ — возмущение и $u(t)$ — искаженная шумом реакция системы. Если предположить, что $x(t)$ и $u(t)$ — стационарные случайные процессы и определить характеристику из условия минимума математического ожидания квадрата ошибки

$$\varepsilon^2 = M\,[u(t) - L\omega]^2\,,$$

$$L\omega = \int_0^{T_1} \omega(\tau)\, x(t-\tau)\, d\tau\,, \tag{1}$$

то (1) простым преобразованием приводится к квадратичному функционалу вида [1]

$$I(\omega) = (A\omega, \omega) - 2(\omega, \chi)\,,$$

$$A\omega = \int_0^{T_1} k(\tau - \tau_1)\, \omega(\tau)\, d\tau\,, \tag{2}$$

где

$$\chi(\tau) = M\,[x(t-\tau)\, u(t)]\,,$$

$$k(\tau) = M\,[x(t-\tau)\, x(t)]\,.$$

В работе будут использоваться скалярные произведения и нормы в двух пространствах $L_2\,(0,\, T_1)$ и $L_2\,(0, T)$ соответственно с носителями для весовых функций $\omega(\cdot)$ и сигналов $u(\cdot)\; x(\cdot)$. Если не будет специального примечания, то скалярное произведение следует понимать в $L_2(0,\, T_1)$.

Однако критерий (1) вряд ли можно считать достаточно реалистичным для идентификации системы. Действительно, для вычисления по (1) требуется знать точные статистические характеристики процессов $x(t)$ и $u(t)$. В реальном эксперименте известны лишь оценки соответствующих характеристик. Однако использование оценок требует осторожности, т. к. задача определения функции, минимизирующей (2), некорректна. Поэтому вместо критерия (1) часто удобно использовать квадратичный критерий близости, считая $x(\cdot)$ и $u(\cdot)$ неслучайными функциями. Тогда можно записать

$$\varepsilon^2 = \|\, u - L\omega\,\|^2\,, \tag{3}$$

где

$$L\omega = \int_0^{t} \omega(\tau)\, x(t-\tau)\, d\tau\,,$$

$$\omega(\tau) \equiv 0;\;\; \tau < 0;\;\; \tau > T_1\,,$$

[1] $I(\omega)$ отличается от ε^2 на величину $\|u\|^2$, не зависящую от ω.

$u(t)$ и $x(t)$ — реакция с ошибкой и возмущение, принадлежащие пространству $L_2(0, T)$. Система имеет конечную память T_1.

Замечание. Предположение о конечной памяти при использовании квадратичных критериев является существенным. В случае (1) фильтр с конечной памятью позволяет получить стационарный процесс $y(t)$ на выходе. Кроме того, процессы $x(t)$ и $y(t)$ оказываются стационарно связанными.

Предположение о конечной памяти в (3) приводит к избыточным данным для $T > T_1$. Если положить $T = T_1$, то использование критерия ε^2 становится затруднительным, т. к. для широкого класса функций $x(\cdot)$, $u(\cdot)$, удовлетворяющих условию $u'(t)$, $x'(t) \in L_2$, существует решение уравнения

$$L\omega = u\,.$$

В этом случае $\varepsilon^2 = 0$, и, следовательно, нельзя отличить по критерию ε^2 плохую систему от хорошей.

Квадратичный критерий (3) приводит также к квадратичному функционалу

$$I(\omega) = (B\omega, \omega) - 2(\omega, \chi)\,,$$

$$B\omega = \int_0^{T_1} k(\tau, \tau_1)\,\omega(\tau)\,d\tau\,, \tag{4}$$

где

$$k(\tau_1, \tau_1) = \int_{\tau_1}^{T} x(t-\tau_1)\,x(t-\tau_2)\,dt\,, \qquad \tau_1 \geq \tau_2,$$

$$k(\tau_1, \tau_2) = \int_{\tau_2}^{T} x(t-\tau_1)\,x(t-\tau_2)\,dt\,, \qquad \tau_2 > \tau_1, \tag{5}$$

$$\chi(\tau) = \int_{\tau}^{T} x(t-\tau)\,u(t)\,dt\,. \tag{6}$$

Оператор B в (4) является положительным и вполне непрерывным.

Положительность B очевидна, а его полная непрерывность является следствием предположения u; $x \in L_2$, т. к. это имеет следствием

$$\int_0^{T_1}\int_0^{T_1} k^2(\tau_1, \tau_2)\,d\tau_1\,d\tau_2 < \infty\,.$$

Эти свойства оператора B являются причиной трудностей при решении задачи о минимуме $I(\omega)$ (4). Действительно, если бы оператор был положительно определен, т. е.

$$p(\omega, \omega) \leq (B\omega, \omega) \leq P(\omega, \omega)\,,$$

где $0 < p \leq P < \infty$, то решение ω^* задачи (4), как показано в [6], всегда существует и притом $\omega^* \in L_2(0, T_1)$. Однако, если $p = 0$, то может не выполняться необходимое условие существования решения уравнения Фредгольма первого рода

$$B\omega^* = \chi, \tag{7}$$

которому должна удовлетворять функция, минимизирующая функционал $I(\omega)$, если последняя существует.

Отсутствие решения (7) в пространствах C или L_2 является следствием полной непрерывности B, т. к. точкой сгущения спектра оператора B является 0 [7], [8].

Решение задачи о минимуме функционала (4) тем не менее существует. Однако элемент, минимизирующий функционал, в общем случае, не является функцией точки в обычном понимании анализа. Винер и Хопф доказали существование изображения Фурье решения уравнения (7), когда левая часть является сверткой [9].

Решению уравнения (7) были посвящены также другие работы, например [4], [10].

Нерешенным остался вопрос о характере решений уравнения (7) и, как следствие, вопрос о приближенных решениях. Однако именно решение двух последних задач оказывается существенным для идентификации.

Основным для их решения является определение множества элементов, на которых функционал (4) достигает минимума.[2] Оказывается таким множеством является нормированное факторпространство. Оно определяется следующим образом.

Пусть областью определения оператора B является L_2 $(0, T_1)$. Тогда можно факторизовать [8] пространство L_2 $(0, T_1)$ относительно подпространства $\overline{S}(B)$ нулей оператора B так, что оператор B, индуцированный на факторпространстве $L_2/\overline{S}$, будет строго положительным. То есть для любого Z имеет место

$$(Bz, z) > 0, \qquad z \in L_2/\overline{S},$$

где Z — элемент факторпространства $L_2/\overline{S}$. Тогда на $L_2/\overline{S}$ можно ввести скалярное произведение и норму формулами

$$(z, \vartheta)_B = (Bz, \vartheta),$$

$$\|z\|_B^2 = (Bz, z).$$

[2] Если такое множество не определено, то задача о минимуме функционала не может быть поставлена вообще.

Норма $\|Z\|_B^2$ сохраняет свое значение для любого представителя элемента Ω факторпространства $L_2/\overline{S}$. В силу эрмитовости оператора B введенное в $L_2/\overline{S}$ скалярное произведение и норма удовлетворяют всем необходимым для них аксиомам [8]. Наконец факторпространство $L_2/\overline{S}$ пополняется, в смысле нормы $\|\cdot\|_B$ [8]. Полное нормированное факторпространство будет обозначаться $H/\overline{S}$.

Относительно элемента, минимизирующего функционал (4), имеет место следующая

Теорема 1

Пусть выполнены условия $x \in L_2$ и $u \in L_2$, и уравнение (7) определено для $\omega \in H/\overline{S}$. Тогда (7) имеет единственное решение.

Доказательство

Для того, чтобы уравнение (7) имело решение, необходимо выполнение условия

$$\chi \in S(B), \tag{8}$$

где $S(B)$ — подпространство значений оператора B.

Это необходимое условие здесь выполняется. Действительно, пусть $\psi \in \overline{S}(B)$. Тогда, меняя порядок интегрирования, скалярное произведение (ψ, χ) можно представить в виде

$$(\psi, \chi) = \int_0^T \int_0^t \psi(\tau)\, x(t-\tau)\, u(t)\, d\tau\, dt\,.$$

Из неравенства Коши для правой части после простого преобразования получается неравенство

$$|(\psi, \chi)| \leq \|u\|\, \|\psi\|_B\,,$$

откуда следует (8), т. к. для $\psi \in \overline{S}(B)$; $B\psi = 0$.

Теперь можно определить элемент ω^*, минимизирующий функционал (4). Для этого вначале определяется функция

$$\omega_n = \sum_{k=1}^{n} \frac{(\chi, \varphi_k)}{\lambda_k}\, \varphi_k\,, \tag{9}$$

где φ_k и λ_k удовлетворяют уравнению

$$B\varphi_k = \lambda_k \varphi_k\,.$$

Значение функционала $I(\omega_n)$ равно

$$I(\omega_n) = -\sum_{k=1}^{n} \frac{(\chi, \varphi_k)^2}{\lambda_k} \tag{9a}$$

и для приращения функционала можно записать

$$I(\omega_n) - I(\omega_{n+1}) = \frac{(\chi, \varphi_n)^2}{\lambda_n}.$$

В силу положительности и условия (8) число в правой части равенства положительно, а значит и последовательность $I(\omega_n)$ — монотонно убывающая. С другой стороны, функционал $I(\omega)$ ограничен снизу. Монотонно убывающая ограниченная последовательность сходится к своей нижней грани

$$\inf I(\omega_n) = I(\omega^*) = -\sum_{k=1}^{\infty} \frac{(\chi, \varphi_k)^2}{\lambda_k}. \tag{10}$$

На основании (9) и (9а) функционал достигает минимума для

$$\omega^* = \sum_{k=1}^{\infty} \frac{(\chi, \varphi_k)}{\lambda_k} \varphi_k. \tag{11}$$

Последовательность ω_n сходится к ω^* по энергетической норме, т. к.

$$\lim_{n\to\infty} \|\omega_n - \omega^*\|_B^2 = \lim_{n\to\infty} [I(\omega_n) - I(\omega^*)] = 0.$$

Норма ω^* на основании (11) равна

$$\|\omega^*\|_B^2 = \sum_{k=1}^{\infty} \frac{(\chi, \varphi_k)^2}{\lambda_k} = -I(\omega^*). \tag{12}$$

Таким образом, функционал достигает минимума для элемента ω^*, для которого существует энергетическая норма. Однако вместе с ω^* функционал (4) сохраняет свое значение также на всех элементах множества Ω^*, полученных сдвигами ω^* на нули оператора B.

Из определения факторпространства $H/\overline{S}$ следует, что $\Omega^* \in H/\overline{S}$.

Теорема доказана.

Минимальное значение функционала определяется на основании (3), (7) и (12)

$$\varepsilon_{\min}^2 = \|u\|^2 - \|\omega^*\|_B^2, \tag{13}$$

где первая норма вычисляется в $L_2(0, T)$.

2. Свойства решения

Теорема 1 позволяет понять особенности полученного решения.

Во-первых, так как элементом факторпространства является множество, Ω^*, то любой представитель фактор-группы может быть решением. Это имеет место, когда $\lambda = 0$ является собственным числом оператора B. Однако ясно, что смещение решения на любую функцию $\psi \in \overline{S}(B)$ не меняет стационарного значения функционала $I(\omega^*)$. Если $\lambda = 0$ не является собственным числом оператора B, то решение задачи единственно. Во-вторых, решение не является функцией точки в обычном смысле. Но, как показано, оно является элементом гильбертова пространства $H/\overline{S}$. В пространстве $H/\overline{S}$ действует оператор L, определяющий реакцию идентифицированной системы. Действительно

$$\begin{aligned} y &= L\omega, \\ y^* &= L\omega^*, \end{aligned} \tag{15}$$

где y и ω — реакция и весовая функция истинной системы. Откуда следует

$$\begin{gathered} \|y^* - y\|^2 = (B\Delta\omega, \Delta\omega) = \|\Delta\omega\|_B^2, \\ \Delta\omega = \omega^* - \omega. \end{gathered} \tag{16}$$

Из последнего равенства можно заключить, что энергетическая норма пространства $H/\overline{S}$ естественным образом порождается квадратичной метрикой для реакций системы. Имеет место следующее важное правило: расстояние между реакциями y^* и y равно расстоянию между характеристиками ω^* и ω в энергетическом пространстве.

Это означает, что малость $\|y^* - y\|^2$ вовсе не влечет малость $\Delta\omega$ в этом же пространстве. Более того, величина $\Delta\omega$ может иметь точки роста типа дельта, но на множестве точек меры нуль. Таким образом, устраняется источник многочисленных недоразумений, существующих в ряде работ. Именно, если система идентифицируется из условия близости реакций, то при этом не следует ожидать близости характеристик модели и системы в «хорошем» пространстве типа C или L_2, т. е. в «обычном» смысле. Напротив, это требование допускает сколь угодно большое отличие в обычном смысле характеристики модели от характеристики истинной системы. Наконец, решение в факторпространстве позволяет иначе взглянуть на некорректность уравнения (7) Фредгольма первого рода.

Задача корректна, если *а)* решение единственно *б)* решение непрерывно зависит от исходных данных [2] (в данном случае от функции $u(t)$). Ясно, что задача идентификации некорректна как в смысле *а)*, так и в смысле *б)*.

При решении происходит как бы необратимая потеря «естественного» пространства, в котором содержится характеристика ω истинной системы. Таким «естественным» пространством может быть, например, пространство C или L_2.

3. Приближенное решение и естественная регуляризация задачи

Теорема 1 устанавливает существование решения для более широкого класса интегральных уравнений (7) по сравнению с уравнением типа свертки, рассмотренным в [9].

Кроме того, теорема позволяет получать приближенные решения уравнения (7), необходимые для задач идентификации, т. к. $x(t)$ и $u(t)$ определяются из экспериментов.

Для приближенного решения (7) можно строить минимизирующие последовательности функционала $I(\omega)$. Если использовать процедуру наискорейшего спуска, то приближения задаются рекуррентными соотношениями

$$\omega_{n+1} = \omega_n - \mu_n V_n \qquad (n = 0, 1, 2, \ldots,), \quad (17)$$

где

$$V_n = B\omega_n - \chi,$$

$$\mu_n = \frac{\| V_n \|^2}{\| V_n \|_B^2}.$$

Нулевое приближение ω_0 выбирается произвольным в L_2. Относительно сходимости последовательности ω_n к решению имеет место

Теорема 2

Пусть x, $u \in L_2(0, T)$. Тогда последовательность ω_n сходится в пространстве $H/\overline{S}$ к элементу ω^*, минимизирующему функционал $I(\omega)$.

Доказательство

Для приращения функционала $I(\omega)$, используя (17), можно записать

$$I(\omega_n) - I(\omega_{n+1}) = \frac{\| V_n \|^2}{\| V_n \|_B^2}.$$

Числа, стоящие в правой части этого равенства, для любого n и любой функции χ — неотрицательны.

Таким образом, последовательность $I(\omega_n)$ — монотонно убывающая, и в силу ограниченности $I(\omega_n)$ снизу сходится к своей нижней грани.

Следует отметить отличие результата, содержащегося в этой теореме, от результата, относящегося к положительно-определенному оператору [6].

Хотя в обоих случаях используется одна и та же минимизирующая последовательность, результат получается различный. В случае положительно-определенного оператора B последовательность (17) сходится к функции $\omega^* \in L_2\,(0, T_1)$.

В случае положительного оператора B последовательность (17) сходится к ω^* в метрике пространства $H/\overline{S}$. При этом в L_2 последовательность, как правило, является расходящейся.

Если остановить вычисление на n-ом шаге, то ω_n, полученная на основании (17), будет принадлежать $L_2(0, T_1)$ (если нулевое приближение выбрано в $L_2\,(0, T_1)$). Таким образом, остановкой вычисления на конечном n можно добиться гладкости приближенного решения, т. е. такого же эффекта, как и регуляризацией функционала $I(\omega)$ [2].

Как и следовало ожидать, такая регуляризация приводит к увеличению $\varepsilon^2_{\min}$. Поэтому необходимо знать, каким увеличением $\varepsilon^2_{\min}$ можно «заплатить» за гладкие решения. Степень гладкости решения должна быть при этом задана. В [2] эта степень задается константами регуляризующих функционалов.

Таким образом, в работе решается задача идентификации динамических систем по квадратичным критериям в нормированном факторпространстве. Доказывается существование решения и определяются приближенные решения. Предложенное решение задачи идентификации в нормированном факторпространстве определяет другой подход к решению некоторых задач.

Литература

1. System Identification Problem. The IFAC Symposium, Prague GSSR, June, 1970.
2. *Тихонов А. Н.:* Решение некорректно поставленных задач. Доклады АН СССР, **151**, 3, 1963.
3. *Леонов Ю. П.:* О приближенном методе синтеза оптимальных линейных систем. Автоматика и телемеханика. **20**, 8, 1959.
4. *Hsieh, H. C.:* The least squares estimation of linear and nonlinear system weighting. Information and Control Journ., **7**, 84—115, 1964.
5. *Леонов Ю. П.:* Обобщенные решения метода наименьших квадратов. Доклады АН СССР, **206**, 1, 1972.
6. *Канторович Л. В.:* Функциональный анализ и прикладная математика. Успехи математических наук, 3, **6**, 1948.
7. *Михлин С. Г.:* Прямые методы в математической физике. Гостеоретиздат, М., 1950.
8. *Колмогоров А. Н.—Фомин С. В.:* Элементы теории функций и функционального анализа. «Наука», 1968.
9. *Wiener, W.:* Extrapolation interpolation and smoothing of stationary times series. New York, 1949.
10. *Пугачев В. С.:* Теория случайных функций и ее применение к задачам автоматического управления. Физматгиз, М., 1962.

The problem of dynamic systems identification in factor space

Yu. P. LEONOV

(Moscow)

Summary

The problem of dynamic systems identification over quadratic criteria of closeness is solved. The theorem of existence of generalized solutions beins elements of normed factor space is proved. "Energy" norm is used in the factor space which permits one to find solutions of a rather general form. Similar solutions may not be, for instance, point functions in a conventional sense. An approximate solution of the problem is defined in the normed space. Connection between generalized solutions and incorrectness of the problem is established.

The suggested solution of the problem determines another approach for the solution of incorrect problems.

Ю. П. Леонов
Институт проблем управления (автоматики и телемеханики)
СССР, Москва В—485, Профсоюзная ул. 81

Problems of Control and Information Theory, Vol. 1 (3—4), pp. 277—286 (1972)

О КОДАХ, ЛОКАЛИЗУЮЩИХ ОШИБКИ

И. М. БОЯРИНОВ

(Москва)

(Поступила в редакцию 5 января 1970 г.)

Рассматриваются линейные над GF (2) коды, локализующие ошибки, проверочные матрицы которых являются кронекеровскими произведениями проверочных матриц циклических кодов с символами из GF (2). Описываются процедуры декодирования построенных кодов.

Линейный (n, k)-код C, проверочная матрица H которого равна кронекеровскому произведению [1] проверочной матрицы H_1 линейного (n_1, κ_1)-кода C_1, исправляющего класс ошибок E_c, и проверочной матрицы H_2 линейного (n_2, k_2)-кода C_2, обнаруживающего класс ошибок E_d, обладает свойствами, промежуточными между кодами обнаруживающими и исправляющими ошибки. Если матрицы H_1 и H_2 удовлетворяют определенным условиям, то избыточность кода С, имеющего длину $n = n_1 n_2$, дает возможность определить любое подмножество, принадлежащее классу E_c подблоков длины n_2, в каждом из которых имели место ошибки класса E_d.

Под классом ошибок E_c (E_d) понимается множество комбинаций ошибок (например, независимые ошибки или пачки ошибок), исправляемых (обнаруживаемых) кодом. Коды с такими свойствами были названы Вулфом и Элспасом [2] кодами, локализующими ошибки (EL-кодами).

Использовав в качестве кода $C_2 — (n_2, n_2 — \varrho)$-код над $GF(2)$, обнаруживающий класс ошибок E_d, в качестве кода C_1 — $(n_1, n_1 — m)$-код с символами из $GF(2^\varrho)$, исправляющий класс ошибок E_c, Вулф [3] построил $(n, n — r)$-код над $GF(2)$ с $n = n_1 n_2$ и $r = m \varrho$, который локализует класс E_c подблоков, искаженных ошибками из класса E_d.

Гёталс [4] рассмотрел случай, когда коды C_2 с символами из $GF(q)$ и C_1 с символами из $GF(q^m)$[2] циклические (q — степень простого, m-зависит от свойств кода C_2).

Ниже при самых общих предположениях изучаются линейные коды, локализующие ошибки, и процедуры их декодирования, а также рассматриваются два класса $E\overline{L}$-кодов над $GF(2)$, построенных с помощью циклических кодов C_1 и C_2 с символами из $GF(2)$.

Т е о р е м а 1. *Пусть C_1 — линейный (n_1, n_1-s)-код над $GF(q_1)$, исправляющий класс ошибок E_c. Пусть C_2 — линейный (n_2, n_2-r)-код над произвольным подполем $GF(q_2)$ поля $GF(q_1)$, обнаруживающий класс ошибок E_d. Тогда существует линейный код C над $GF(q_2)$ длины $n = n_1 n_2$, имеющий не более чем rs проверочных символов и локализующий класс подблоков E_c длины n_2, искаженных ошибками класса E_d.*

Д о к а з а т е л ь с т в о. Пусть

$$H_1 = \begin{pmatrix} h_{11}^{(1)} & h_{12}^{(1)} & \dots & h_{1n_1}^{(1)} \\ h_{21}^{(1)} & h_{22}^{(1)} & \dots & h_{2n_1}^{(1)} \\ \dots & \dots & \dots & \dots \\ h_{s1}^{(1)} & h_{s2}^{(1)} & \dots & h_{sn_1}^{(1)} \end{pmatrix}$$

— проверочная матрица кода C_1, а

$$H_2 = \begin{pmatrix} h_{11}^{(2)} & h_{12}^{(2)} & \dots & h_{1n_2}^{(2)} \\ h_{21}^{(2)} & h_{22}^{(2)} & \dots & h_{2n_2}^{(2)} \\ \dots & \dots & \dots & \dots \\ h_{r1}^{(2)} & h_{r2}^{(2)} & \dots & h_{rn_2}^{(2)} \end{pmatrix}$$

— проверочная матрица кода C_2. Покажем, что код C, проверочная матрица H которого равна кронекеровскому произведению матриц H_1 и H_2

$$H = H_1 \otimes H_2 = \begin{pmatrix} h_{11}^{(1)} H_2 & h_{12}^{(1)} H_2 \dots h_{1n_1}^{(1)} H_2 \\ h_{21}^{(1)} H_2 & h_{22}^{(1)} H_2 \dots h_{2n_1}^{(1)} H_2 \\ \dots & \dots \\ h_{s1}^{(1)} H_2 & h_{s2}^{(1)} H_2 \dots h_{sn_1}^{(1)} H_2 \end{pmatrix} =$$

$$= \begin{Vmatrix} h_{11}^{(1)} h_{11}^{(2)} \dots h_{11}^{(1)} h_{1n_2}^{(2)} \dots h_{1n_1}^{(1)} h_{11}^{(2)} \dots h_{1n_1}^{(1)} h_{1n_2}^{(2)} \\ \dots \\ h_{11}^{(1)} h_{r1}^{(2)} \dots h_{11}^{(1)} h_{rn_2}^{(2)} \dots h_{1n_1}^{(1)} h_{r1}^{(2)} \dots h_{1n_1}^{(1)} h_{rn_2}^{(2)} \\ \dots \\ h_{s1}^{(1)} h_{11}^{(2)} \dots h_{s1}^{(1)} h_{1n_2}^{(2)} \dots h_{sn_1}^{(1)} h_{11}^{(2)} \dots h_{sn_1}^{(1)} h_{1n_2}^{(2)} \\ \dots \\ h_{s1}^{(1)} h_{r1}^{(2)} \dots h_{s1}^{(1)} h_{rn_2}^{(2)} \dots h_{sn_1}^{(1)} h_{r_1}^{(2)} \dots h_{sn_1}^{(1)} h_{rn_2}^{(2)} \end{Vmatrix},$$

удовлетворяет требованиям теоремы.

Для того чтобы код C локализовал класс подблоков E_c, искаженных ошибками класса E_d, необходимо и достаточно, чтобы синдром кодового вектора, в котором произвольное множество (два множества) подблоков класса E_c икажено ошибками класса E_d, не был равен нулю.

Предположим противное. Пусть кодовый вектор

$$a = \left(a_1^{(1)}, a_2^{(1)}, \ldots, a_{n_2}^{(1)}, \ldots, a_1^{(n_1)}, a_2^{(n_1)}, \ldots, a_{n_2}^{(n_1)}\right), \tag{1}$$

в котором некоторое множество подблоков класса E_c искажено ошибками класса E_d, перешел в вектор

$$b = \left(b_1^{(1)}, b_2^{(1)}, \ldots b_{n_2}^{(1)}, \ldots, b_1^{(n_1)}, b_2^{(n_1)}, \ldots, b_{n_2}^{(n_1)}\right), \tag{2}$$

синдром S которого равен нулю

$$\sum_{j=1}^{n_1} \sum_{i=1}^{n_2} b_i^{(j)} h_{kj}^{(1)} h_{li}^{(2)} = 0, \tag{3}$$

$$k = 1, 2, \ldots, s; \quad i = 1, 2, \ldots, r.$$

Преобразуем (3) к виду

$$\sum_{j=1}^{n_1} \left(\sum_{i=1}^{n_2} b_i^{(j)} h_{li}^{(2)} \right) h_{kj}^{(1)} = 0, \tag{4}$$

$$k = 1, 2, \ldots s; \quad l = 1, 2, \ldots, r.$$

Представим таким же образом синдром кодового вектора a

$$\sum_{j=1}^{n_1} \left(\sum_{i=1}^{n_2} a_i^{(j)} h_{li}^{(2)} \right) h_{kj}^{(1)} = 0, \tag{5}$$

$$k = 1, 2, \ldots, s, \qquad = 1, 2 \ , \ldots, r$$

Так как $a_i^{(j)}$, $b_i^{(j)}$, $h_{li}^{(2)}$ являются по условию элементами поля $GF(q_2)$, то суммы

$$a_l^{(j)} = \sum_{i=1}^{h_2} a_i^{(j)} h_{li}^{(2)} \quad \text{и} \quad b_l^{(j)} = \sum_{i=1}^{n_2} b_i^{(j)} h_{li}^{(2)} \tag{6}$$

также принадлежат полю $GF(q_2)$ и, естественно, полю $GF(q_1)$, являющемуся расширением $GF(q_2)$ (в случае $q_1 = q_2$ оба поля, очевидно, совпадают).

При каждом фиксированном l ($l = 1, 2, \ldots, r$) системы равенств (4) и (5) можно рассматривать как синдромы векторов

$$b_l = \left(b_l^{(1)}, b_l^{(2)}, \ldots, b_l^{(n_1)}\right) \quad \text{и} \quad a_l = \left(a_l^{(1)}, a_l^{(2)}, a_l^{(2)}, \ldots, a_l^{(n_1)}\right) \tag{7}$$

принадлежащих коду C_1. В силу линейности кода C_1 векторы $c_l = b_l - a_l = (c_l^{(1)}, c_l^{(2)}, \ldots, c_l^{(n_1)})$ будут являться кодовыми векторами.

Покажем, что ненулевыми компонентами векторов c_l ($l = 1, 2, \ldots, r$) могут быть только те, которые соответствуют искаженным подблокам (j-му подблоку соответствует $c_l^{(j)}$) и, обратно, каждому j-му искаженному подблоку соответствует по крайней мере один вектор c_l с ненулевой компонентой $c_l^{(j)}$.

Действительно, по построению,

$$c_l^{(j)} = \sum_{i=1}^{n_2} (b_i^{(j)} - a_i^{(j)}) h_{li}^{(2)} = \sum_{i=1}^{n_2} c_i^{(j)} h_{li}^{(2)}.$$

Если j-й подблок не искажен, то $b_i^{(j)} = a_i^{(j)}$, а следовательно, и $c_i^{(j)} = 0$ для всех $i = 1, 2, \ldots, n_2$; если же j-й подблок искажен, то синдром вектора-ошибки $c^{(j)} = (c_1^{(j)} c_2^{(j)}, \ldots, c_{n_2}^{(j)})$ кода C_2, принадлежащего по предположению классу ошибок E_d, не равен нулю и имеет поэтому по крайней мере одну ненулевую компоненту, например,

$$c_l^{(j)} = \sum_{i=1}^{n_2} c_i^{(j)} h_{li}^{(2)}.$$

Таким образом, в каждом из векторов c_l ($l = 1, 2, \ldots, r$) ненулевыми компонентами могут быть только те, которые соответствуют искаженным подблокам (не обязательно всем). Каждый вектор c_l является вследствие этого вектором-ошибкой, принадлежащим классу E_c ошибок, исправляемых кодом C_1, и, очевидно, не может быть кодовым вектором. Получили противоречие. Теорема доказана.

Число проверочных символов кода C равно числу линейно независимых строк проверочной матрицы H и, очевидно, не превосходит rs.

С помощью теоремы 1, используя алгебраическую структуру исходных кодов, строятся EL-коды, обладающие требуемыми свойствами. Например, непосредственным следствием теоремы 1 является результат [3].

Как это показано ниже, в связи с тем, что декодирование EL-кода C сводится к декодированию кода C_1, желательно по возможности выбирать в качестве кодов C_1 такие, которые допускают простые методы декодирования. Одним из таких путей при построении EL-кода над $GF(2)$ является выбор циклических кодов над $GF(2)$ в качестве кодов C_1 и C_2.

Теорема 2. *Пусть C_1-двоичный укороченный циклический код длины n_1 и $H_1 = (1, \beta, \beta^2, \ldots, \beta^{n_1-1})$ — проверочная матрица кода, где β — примитивный элемент поля $GF(2^{m_1})$. Пусть C_2 — двоичный циклический код длины n_2, обнаруживающий класс ошибок E_d, имеет проверочную матрицу*

$$H_2 = \begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \ldots & \alpha_1^{n_2-1} \\ 1 & \alpha_2 & \alpha_2^2 & \ldots & \alpha_2^{n_2-1} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & \alpha_r & \alpha_r^2 & \ldots & \alpha_r^{n_2-1} \end{pmatrix},$$

где $\alpha_1, \alpha_2, \ldots, \alpha_r$ — корни r различных неприводимых многочленов $g_i(x)$ ($g_i(\alpha_i) = 0$, $i = 1, 2, \ldots, r$), на которые разлагается порождающий многочлен $g(x)$ кода, $GF(2^{m_2})$ — наименьшее расширение $GF(2)$, содержащее все $\alpha_1, \alpha_2, \ldots, \alpha_r$.

Пусть $d = (2^{m_1}-1, 2^{m_2}-1)$, $n_1 \leq (2^{m_1}-1)/d$ и $GF(2^m)$ — *наименьшее расширение* $GF(2)$, *содержащее* β, $\alpha_i (i = 1, 2, \ldots, r)$. *Тогда существует код* C *длины* $n = n_1, n_2$, *имеющий не более чем* rm *проверочных символов и локализующий одиночный подблок длины* n_2, *в котором имели место ошибки класса* E_d.

Доказательство. На основании теоремы 1 естественно предположить, что проверочная матрица H EL-кода C равна кронекеровскому произведению матриц H_1 и H_2

$$H = H_1 \otimes H_2 = \begin{pmatrix} 1 & \alpha_1 & \ldots & \alpha_1^{n_2-1} & \ldots & \beta^{n_1-1} & \beta^{n_1-1}\alpha_1 & \ldots & \beta^{n_1-1}\alpha_1^{n_2-1} \\ 1 & \alpha_2 & \ldots & \alpha_2^{n_2-1} & \ldots & \beta^{n_1-1} & \beta^{n_1-1}\alpha_2 & \ldots & \beta^{n_1-1}\alpha_2^{n_2-1} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & \alpha_r & \ldots & \alpha_r^{n_2-1} & \ldots & \beta^{n_1-1} & \beta^{n_1-1}\alpha_r & \ldots & \beta^{n_1-1}\alpha_r^{n_2-1} \end{pmatrix}.$$

Для того чтобы код локализовал одиночный подблок, необходимо и достаточно, чтобы синдром кодового вектора, в котором два произвольных подблока искажены ошибками класса E_d, не был равен нулю.

Предположим противное. Пусть l-й и j-й подблоки вектора a (*1*) кода C подверглись искажениям из класса ошибок E_d так, что вектор a перешел в вектор b (*2*), синдром S которого равен нулю

$$S_k = \beta^{j-1} \sum_{i=1}^{n_2} b_i^{(j)} \alpha_k^{i-1} + \beta^{l-1} \sum_{i=1}^{n_2} b_i^{(l)} \alpha_k^{i-1} = 0, \qquad (8)$$

$$k = 1, 2, \ldots, r.$$

Так как код C_2 обнаруживает ошибки класса E_d, то найдется хотя бы одно k такое, что по крайней мере одно из двух слагаемых (например, первое) S_k отлично от нуля.

Преобразуем выражение (8) к виду

$$\beta^{j-l} = \frac{\sum_{i=1}^{n_2} b_i^{(l)} \alpha_k^{i-1}}{\sum_{i=1}^{n_2} b_i^{(j)} \alpha_k^{i-1}}. \qquad (9)$$

Обозначим правую часть равенства (9) через α^b, где α — примитивный элемент подполя $GF(2^{m_1})$, b — целое. Элементы β^{j-l} и α^b по условию принадлежат полю $GF(2^m)$.

Покажем, что при всех допустимых значениях l и j элементы β^{j-l} и α^b не принадлежат пересечению подполей $GF(2^{m_1})$ и $GF(2^{m_2})$, поэтому равенство (9) не выполняется при тех же значениях l и j. Действительно, для того

чтобы β^{l-j} и α^b принадлежали пересечению подполей $GF(2^{m_1})$ и $GF(2^{m_2})$, необходимо и достаточно, чтобы выполнялось равенство

$$|l-j|\,\frac{2^m-1}{2^{m_1}-1}=bc\,\frac{2^m-1}{2^{m_2}-1},$$

причем $(c, 2^{m_2}-1)=1$ или

$$|l-j|\,(2^{m_2}-1)=bc(2^{m_1}-1). \tag{10}$$

Если

$$|l-j|<\frac{2^{m_1}-1}{d}, \quad \text{где} \quad d=(2^{m_1}-1, 2^{m_2}-1),$$

что непосредственно следует из условия теоремы, то равенство (10), а значит и равенство (9), невозможно ни при одном $b \neq 0$. Следовательно, предположение о равенстве нулю синдрома S вектора b неверно. Число проверочных символов кода C, очевидно, не превосходит rm. Теорема доказана.

Следствие. *Положим* $m_1=m>m_2$, $r=1$, $n_2=2^{m_2}-1$. *Код* C *имеет длину* $n=2^m-1$, *число проверочных символов* m *и локализует одиночный подблок, в котором произошло не более двух ошибок. Код* C, *как это следует из* [2], *является в этом случае оптимальным.*

Пример. Пусть $C_1^{(1)}$ — циклический (255, 247)-код Хэмминга, β-корень примитивного многочлена

$$g_1^{(1)}(x)=x^8+x^4+x^3+x^2+1 \quad \text{и} \quad H_1^{(1)}=(1, \beta, \ldots, \beta^{254})$$

— проверочная матрица кода $C_1^{(1)}$.

Пусть C_2 — циклический (15,7)-код Боуза-Чоудхури, исправляющий две или обнаруживающий четыре ошибки, порожденный многочленом

$$g_2(x)=(x^4+x+1)(x^4+x^3+x^2+x+1);$$

α, α^3 — корни $g_2(x)$ и

$$H_2=\begin{pmatrix}1 & \alpha & \alpha^2 & \ldots & \alpha^{14}\\ 1 & \alpha^3 & \alpha^6 & \ldots & \alpha^{12}\end{pmatrix}$$

— проверочная матрица кода C_2.

В соответствии с теоремой 2 код C_1 получим укорочением кода $C_1^{(1)}$. Так как

$$d=(2^{m_1}-1, 2^{m_2}-1)=(255, 15)=15 \quad \text{и} \quad n_1=\frac{2^{m_1}-1}{d}=17,$$

то C_1 — укороченный циклический (17,9)-код и $H_1=(1, \beta, \beta^2, \ldots, \beta^{16})$ — проверочная матрица кода C_1.

Код C будет иметь длину $n = n_1 n_2 = 255$, число проверочных символов $m = 16$ и локализовывать одиночный подблок длины 15, в котором произошло не более четырех ошибок. Проверочная матрица кода C равна

$$H = \begin{pmatrix} 1 & \alpha & \dots & \alpha^{14} & \dots & \beta^{16} & \beta^{16}\alpha & \dots & \beta^{16}\alpha^{14} \\ 1 & \alpha^3 & \dots & \alpha^{12} & \dots & \beta^{16} & \beta^{16}\alpha^3 & \dots & \beta^{16}\alpha^{12} \end{pmatrix}.$$

Для случая локализации многократных искаженных подблоков имеет место следующая теорема.

Т е о р е м а 3. *Пусть C_1 — двоичный циклический код длины n_1, исправляющий класс ошибок E_c, порожден многочленом $g^{(1)}(x) = g_1^{(1)}(x)\, g_2^{(1)}(x) \dots g_s^{(1)}(x)$, где $g_i^{(1)}(x)$ — неприводимый над $GF(2)$ многочлен, $i = 1, 2, \dots, s$, и $GF(2^{m_1})$ — наименьшее расширение $GF(2)$, содержащее все корни $g^{(1)}(x)$.*

Пусть C_2 — двоичный циклический код длины n_2, обнаруживающий класс ошибок E_d, порожден многочленом $g^{(2)}(x) = g_1^{(2)}(x)\, g_2^{(2)}(x) \dots g_r^{(2)}(x)$, где $g_i^{(2)}(x)$ — неприводимый над $GF(2)$ многочлен, $i = 1, 2, \dots r$, и $GF(2^{m_2})$ — наименьшее расширение $GF(2)$, содержащее все корни $g^{(2)}(x)$. Если $(m_1, m_2) = 1$ и $GF(2^m)$ — наименьшее расширение $GF(2)$, содержащее все корни многочлена $g^{(1)}(x)\, g^{(2)}(x)$, то существует EL-код C длины $n = n_1 n_2$, имеющий не более чем rsm проверочных символов и локализующий класс подблоков E_c, в каждом из которых имели место ошибки класса E_d. Код C эквивалентен циклическому коду.

Не останавливаясь на доказательстве теоремы, в целом аналогичном доказательству предыдущей теоремы, заметим лишь по поводу последнего утверждения, что код C, как легко показать, эквивалентен циклическому коду, порождающий многочлен которого равен

$$g(x) = g_{11}(x)\, g_{12}(x) \dots g_{ij}(x) \dots g_{rs}(x), \quad i = 1, 2, \dots, s;\ j = 1, 2, \dots r,$$

где $g_{ij}(x)$ — минимальный многочлен элемента $\alpha_i \beta_j$ поля $GF(2^m)$, $g_i^{(1)}(\alpha_i) = 0$, $g_j^2(\beta_j) = 0$.

Следует заметить, что число проверочных символов EL-кодов C, удовлетворяющих теоремам 2 и 3, пропорционально $m = m_1 m_2/(m_1, m_2)$.

Наименьшее значение m, равное m_1, принимает при $(m_1, m_2) = m_2$, а наибольшее значение, равное $m_1 m_2$, m принимает при $(m_1, m_2) = 1$. Поэтому теорема 2 позволяет строить локализующие одиночный подблок коды с меньшей избыточностью, чем теорема 3, при помощи которой, однако, строятся коды, локализующие многократные искаженные подблоки, которые эквивалентны циклическим кодам.

Рассмотрим декодирование EL-кодов, удовлетворяющих теореме 1, и покажем что оно сводится к декодированию кодов C_1.

Если декодирование кода C_1 включает в себя вычисление синдромов (например, поэтапное декодирование кодов Боуза—Чоудхури), то процедура состоит в следующем. Вычисляются компоненты синдрома S вектора b (2), полученного на выходе

$$\sum_{j=1}^{n_1} \sum_{i=1}^{n_2} b_i^{(j)} h_{kj}^{(1)} h_{li}^{(2)}, \quad k = 1, 2, \ldots, s \quad l = 1, 2, \ldots, r.$$

Если все компоненты синдрома S равны нулю, ошибок не произошло и декодирование на этом заканчивается. В противном случае для каждого l ($l = 1, 2, \ldots, r$) вычисляется синдром S_l вектора

$$b_l = (b_l^{(1)}, b_l^{(2)}, \ldots, b_l^{(n_1)}), \tag{11}$$

отличающегося от вектора a_l (7), принадлежащего коду C_1, в позициях (не обязательно во всех), соответствующих искаженным подблокам (j-му подблоку соответствует компонента $b_l^{(j)}$), по следующему правилу: p_l-я компонента ($p_l = l + pr$, $p = 0, 1, \ldots, s - 1$) синдрома S является $(p + 1)$-й компонентой синдрома S_l. Доказательство последнего и других приводимых здесь утверждений непосредственно вытекает из доказательства теоремы 1.

По синдрому S_l в соответствии с процедурой декодирования кода C_1, находим искаженные символы вектора b_l, а значит, и определяем искаженные подблоки кода C. Повторяя процедуру для каждого l и объединяя результаты вычислений, определяем все искаженные подблоки.

Необходимость последних операций вытекает из того, что в каждом из векторов b_l могут быть искажены символы, соответствующие лишь некоторым искаженным подблокам, но для каждого искаженного подблока найдется по крайней мере один вектор b_l, в котором будет искажен соответствующий этому подблоку символ.

Если метод декодирования кода C_1 не включает в себя вычисление синдромов (например, мажоритарное декодирование), процедура декодирования сводится к построению r векторов b_l (11).

Способ построения векторов b_l, вытекающий из доказательства теоремы 1, заключается в следующем. Первая компонента вектора b_l равна

$$b_l^{(1)} = \sum_{i=1}^{n_2} b_i^{(1)} h_{li}^{(2)},$$

вторая компонента равна $b_l^{(2)} = \sum_{i=1}^{n_2} b_i^{(2)} h_{li}^{(2)}$, j-я компонента вектора b_l равна $b_l^{(j)} = \sum_{i=1}^{n_2} b_i^{(j)} h_{li}^{(2)}$, n_1-я компонента вектора равна $b_l^{(n_1)} = \sum_{i=1}^{n_2} b_i^{(n_1)} h_{li}^{(2)}$.

В каждом из векторов b_l, используя процедуру декодирования кода C_1, находятся искаженные символы и, следовательно, соответствующие им искаженные подблоки. Объединяя результаты вычислений по всем $l (l = 1, 2, \ldots, r)$, определим все искаженные подблоки кода C.

Приведенные процедуры декодирования пригодны, в принципе, для всех кодов, локализующих ошибки, проверочные матрицы которых равны кронекеровскому произведению проверочных матриц исходных кодов, однако в каждом конкретном случае необходимо или усложнять или упрощать рассмотренные процедуры в зависимости от алгебраической структуры выбранных кодов.

Литература

1. *Мишина А. П.—Проскуряков И. В.:* Высшая алгебра. «Наука», М., 1965.
2. *Wolf, J. K.—Elspas, B.:* Error-locating codes — a new concept in error control. IEEE Trans. Inform. Theory, **9**, 2, 1963.
3. *Wolf, J. K.:* On an extended class of error-locating codes. Inform. and Control, **8**, 2, 1965.
4. *Goethals, J. M.:* Cyclic error-locating codes. Inform. and Contol, **10**, 4, 1967.

On the linear error-locating codes

I. M. BOYARINOV

(Moscow)

Summary

Linear error-locating codes over $GF(2)$ whose check matrices are the Kronecker products of the check matrices of cyclic codes with elements from $GF(2)$ are discussed.

The following theorem is proved for codes which locate single error sub-blocks in a code word.

Let C_1 be a shortened cyclic code of length n_1. Also assume that β is a primitive element of the field $GF(2^{m_1})$ and $H_1 = (1, \beta, \beta^2, \ldots, \beta^{n_1-1})$ is the check matrix of the code C_1.

Let C_2 be a cyclic code of length n_2 which detections a class of errors E_d. The check matrix of C_2 is defined as

$$H_2 = \begin{pmatrix} (1\ \alpha_1\ \alpha_1^2\ \ldots\ \alpha_2^{n_2-1}) \\ (1\ \alpha_2\ \alpha_2^2\ \ldots\ \alpha_2^{n_2-1}) \\ \ldots\ldots\ldots\ldots\ldots \\ (1\ \alpha_r\ \alpha_r^2\ \ldots\ \alpha_r^{n_2-1}) \end{pmatrix}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_r$ are roots of r different irreducible polynomials $g_i(x)$ $(g_i(\alpha_i) = 0$, $i = 1, 2, \ldots, r)$ and $GF(2^{m_2})$ is the least extension field over $GF(2)$ which contains all $\alpha_1, \alpha_2, \ldots, \alpha_r$.

If $d = (2^{m_1} - 1, 2^{m_2} - 1)$, $n_1 \leq \frac{2^{m_1} - 1}{d}$ and $GF(2^m)$ is the least extension over $GF(2)$ which contains β, $\alpha_i (i = 1, 2, \ldots, r)$, then, there exists a code C of a length $n = n_1 n_2$ with not more than rm check bits, which locates a single sub-block of length n_2 with errors of the class E_d.

An analogous theorem is proved for a code which locates multiple error sub-blocks.

Coding procedures for the codes discussed are also described.

И. М. Бояринов

Научный совет по комплексной проблеме «Кибернетика»

СССР, Москва В-333, ул. Вавилова 40

Problems of Control and Information Theory, Vol. 1 (3—4), pp. 287—296 (1972)

ON TWO METHODS OF A DISCRETE SYSTEM IDENTIFICATION

CS. BÁNYÁSZ, J. GERTLER

(Budapest)

(Received April 5, 1971)

The maximum-likelihood method and the generalized least-squares method for the discrete identification are compared. On the basis of algorithms for computers conclusions are drawn about the possibility of the application of these methods, about their numerical claim, in the case of different noise-structures. The actual importance of this study is shown by the fact that in our time the spreading of the process control by computer requires more and more discrete identifications of linear dynamical systems. The reason of this is the data processing and the fact that the informations on the signals of processes are supplied by data collectors in discrete moments. If necessary, the discrete model can be easily set into a continuous form.

1. The maximum-likelihood method

Åström and Bohlin [1] suggested the following system-model in their maximum likelihood parameter estimation method (Fig. 1);

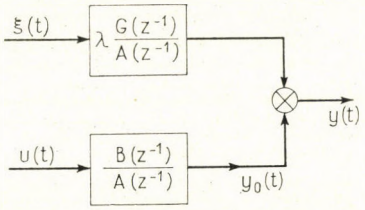


Fig. 1

where

$\frac{B(z^{-1})}{A(z^{-1})}$ is the model of the process

$u(t)$ is the input signal (a random signal with zero mean and 1 variance)

$\xi(t)$ random noise, non-autocorrelated with zero mean and 1 variance

$y(t)$ output signal

$\lambda \frac{G(z^{-1})}{A(z^{-1})}$ noise-model.

The equation of the system is:

$$y(t) = \frac{B(z^{-1})}{A(z^{-1})} u(t) + \lambda \frac{G(z^{-1})}{A(z^{-1})} \xi(t) \tag{1}$$

The noise is considered as an equivalent output noise. Let us define the following vectors:

$$\mathbf{q} = [1, a_1, a_2, \ldots, a_n, -b_0, -b_1, \ldots, -b_m, -g_1, -g_2, \ldots, -g_k]^T$$

$$\mathbf{x}(t) = [y(t), y(t-1), \ldots, y(t-n), u(t), u(t-1), \ldots, u(t-m),$$

$$E(t-1), \ldots, E(t-k)]^T$$

(where $E(t)$ is a non-autocorrelated random noise with zero mean and variance λ). Reducing the vectorial form of the system equation to $E(t)$ we get:

$$E(t) = \mathbf{x}^T(t) \cdot \mathbf{q} \tag{2}$$

In the case of N samplings Eq. (2) becomes:

$$\mathbf{E} = \mathbf{X} \cdot \mathbf{q} \tag{3}$$

where

$$\mathbf{X} = \{\mathbf{x}^T(t)\} \tag{4}$$

$$\mathbf{E} = \{E(t)\} \quad t = 1, 2, \ldots, N \tag{5}$$

The common density function for N samplings of the signal $E(t)$ is:

$$f^N(\mathbf{E}, O, \lambda) = \left(\frac{1}{\sqrt{2\pi}\,\lambda}\right)^N e^{-\frac{1}{2\lambda^2}\mathbf{E}^T\mathbf{E}} \tag{6}$$

On the basis of this the likelihood function has the following form:

$$L = -\frac{N}{2} \ln 2\pi - N \ln \lambda - \frac{1}{2\lambda^2} \mathbf{E}^T\mathbf{E} \tag{7}$$

Let the so called error-function be the following:

$$V(\boldsymbol{\Theta}) = \frac{1}{2} \mathbf{E}^T\mathbf{E} \tag{8}$$

where $\boldsymbol{\Theta} = [a_1, a_2, \ldots, a_n, b_0, b_1, \ldots, b_m, g_1, g_2, \ldots, g_k]$ is the vector of the unknown parameters.

The likelihood function is to be maximized in $\boldsymbol{\Theta}$. This is equivalent to the minimization of the function $V(\boldsymbol{\Theta})$. Applying a general algorithmical description to determine the minimum of the cost-function

$$F(\mathbf{c}) = E_x\{f(\mathbf{x}, \mathbf{c})\}$$

of its parameter c, the following algorithm can be used [4]:

$$\mathbf{c}[n] = \mathbf{c}[n-1] - \mathbf{\Gamma}[n] \nabla_{\mathbf{c}} f(\mathbf{x}[n], \mathbf{c}[n-1]) \tag{9}$$

The identification tasks with off-line aspects can be considered to be deterministical ($F(\mathbf{c}) = f(\{O, \mathbf{c}) = f(\mathbf{c})$) from the point of view of $\mathbf{x}[n]$, thus the optimal weight-function can be determined from the second-order Taylor-series of the function:

$$\mathbf{\Gamma}[n] = \left[\mathbf{H}\left(f(\mathbf{c}[n-])\right)\right]^{-1} \tag{10}$$

(where $\mathbf{H}$ is the matrix of the second-order partial derivates of the function $f(\mathbf{\Theta}, \mathbf{c})$, the so called Hessian).

Applying to the error-function to be minimized in the Åström—Bohlin method we get:

$$\nabla_{\Theta} V(\mathbf{\Theta}) = \frac{\partial V(\mathbf{\Theta})}{\partial \mathbf{\Theta}} = \frac{\partial \mathbf{E}^T}{\partial \mathbf{\Theta}} \mathbf{E} = \mathbf{J}^T(\mathbf{E}, \mathbf{\Theta}) \cdot \mathbf{E} \tag{11}$$

(where $\mathbf{J}$ is the Jacobian, containing the first-order derivates of $\mathbf{E}(\mathbf{\Theta})$).

$$\mathbf{H}\left(V(\mathbf{\Theta})\right) = \frac{\partial}{\partial \mathbf{\Theta}} \left(\frac{\partial V(\mathbf{\Theta})}{\partial \mathbf{\Theta}}\right) = \frac{\partial}{\partial \mathbf{\Theta}} \cdot \mathbf{J}^T(\mathbf{E}, \mathbf{\Theta})\, \mathbf{E} + \mathbf{J}^T(\mathbf{E}, \mathbf{\Theta}) \frac{\partial \mathbf{E}}{\partial \mathbf{\Theta}^T} =$$

$$= \left\{\left(\frac{\partial}{\partial \Theta_1} \mathbf{J}^T(E, \mathbf{\Theta})\right) \mathbf{E}; \ldots; \left(\frac{\partial}{\partial \Theta_{n+m+k}} \mathbf{J}^T(\mathbf{E}, \mathbf{\Theta})\right) \mathbf{E}\right\} + \mathbf{J}^T(\mathbf{E}, \mathbf{\Theta})\, \mathbf{J}(\mathbf{E}, \mathbf{\Theta}) \tag{12}$$

(12) is the matrix $\mathbf{V}_{\Theta\Theta}(\mathbf{\Theta})$ in [1].)

All these become more complicated, since the roots of the polynomial G are restricted to the interior of the circle of radius 1.

2. The generalized least-squares method

The method [3] starts from the noise-model, reduced to an output, which is more general than that of the model on Fig. 1. The equation of the system:

$$y(t) = \frac{B(z^{-1})}{A(z^{-1})} u(t) + e(t)$$

wh ere

$$e(t) = \lambda \frac{G(z^{-1})}{F(z^{-1})} \xi(t) \tag{13}$$

Let us define the following vectors:

$$\mathbf{Q} = [a_1, a_2, \ldots, a_n, b_0, b_1, \ldots, b_m]^T$$

$$\mathbf{x}(t) = [-y(t-1), \ldots, -y(t-n), u(t), u(t-1), \ldots, u(t-m)]^T$$

The equation of the system in vectorial form is:

$$y(t) = \mathbf{x}^T(t)\,\mathbf{Q} + \Delta(t) \tag{14}$$

where

$$\Delta(t) = A(z^{-1}) \cdot e(t)$$

The form of (14) in the case of N samplings is:

$$\mathbf{y} = \mathbf{XQ} + \boldsymbol{\Delta} \tag{15}$$

where

$$\mathbf{X} = \{\mathbf{x}^T(t)\}$$

$$\mathbf{y} = \{y(t)\} \qquad t = 1, \ldots, N$$

$$\boldsymbol{\Delta} = \{\Delta(t)\}$$

The method minimizes the error-function

$$V(\boldsymbol{\Theta}) = \boldsymbol{\Delta}^T\boldsymbol{\Delta} = (\mathbf{y} - \mathbf{XQ})^T(\mathbf{y} - \mathbf{XQ}) \tag{16}$$

Since (16) is a quadratic function of $\mathbf{Q}$, the minimum of the error-function can be obtained in one step:

$$\hat{\mathbf{Q}}_{LS} = [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{y} \tag{17}$$

Since $\mathbf{y} = \mathbf{YX} + \boldsymbol{\Delta}$, the expression (17) is a biased estimation of $\mathbf{Q}$, because

$$\begin{aligned}\hat{\mathbf{Q}}_{LS} &= [\mathbf{X}^T\mathbf{X}]\,\mathbf{X}^T[\mathbf{XQ} + \boldsymbol{\Delta}] = \\ &= [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{XQ} + [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\boldsymbol{\Delta} = \\ &= \mathbf{Q} + [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\boldsymbol{\Delta}\end{aligned} \tag{18}$$

and the expected value of the second part of the sum (18) does not equal to zero.

The bias can be ceased by the proper filtering of the input and output signals.

The method presumes that the noise can be approached as an auto-regressional one, i.e.:

$$\Delta(t) = \frac{1}{C(z^{-1})}\,\xi(t)$$

$$\left(C(z^{-1} = 1 + c_1z^{-1} + \cdots + c_kz^{-k};\right. \tag{19}$$

$\xi(t)$ is uncorrelated with zero mean and variance 1.

Defining the vectors

$$\mathbf{C} = [c_1, c_2, \ldots, c_k]^T$$

$$\boldsymbol{\omega}(t) = [-\Delta(t-1), \ldots, -\Delta(t-k)]^T \tag{20}$$

we get

$$\Delta(t) = \boldsymbol{\omega}(t) \cdot \mathbf{C} + \xi(t) \tag{21}$$

Eq. (21) for N samplings has the following form:

$$\begin{aligned} &\mathbf{\Delta} = \mathbf{\Omega C} + \mathbf{\xi} \\ &\mathbf{\Omega} = \{\boldsymbol{\omega}^T(t)\} \\ &\mathbf{\Delta} = \{\Delta(t)\} \qquad t = 1, \ldots, N \\ &\mathbf{\xi} = \{\xi(t)\} \end{aligned} \tag{22}$$

On this way an other least-squares estimation can be given to the noise-coefficients:

$$\hat{\mathbf{C}}_{LS} = [\mathbf{\Omega}^T\mathbf{\Omega}]^{-1}\mathbf{\Omega\Delta} \tag{23}$$

(the estimation is unbiased).

Applying the estimation (23) we filter the input and output signals:

$$\begin{aligned} u^F(t) &= (1 + c_1 z^{-1} + \ldots + c_k z^{-k})\, u(t) \\ y^F(t) &= (1 + c_1 z^{-1} + \ldots + c_k z^{-k})\, y(t) \end{aligned} \tag{24}$$

More and more correct estimations of $\hat{Q}$ and $\hat{C}$ can be obtained on a recursive way from the transformed data. For the detailed description of the method we refer to [3].

Since during the estimation the noise is considered to be autoregressional, certainly the following equation must hold, for the estimated coefficients, taking (13) into consideration:

$$\lambda \frac{A(z^{-1})\, G(z^{-1})}{F(z^{-1})} = \lambda \frac{1}{C(z^{-1})} \tag{25}$$

If the coefficients of the polynominals A, C are identificated, then the coefficients of the polynomials G and F cannot be uniquely restored. Counting with the Åstrom-model of Fig. 1 (where $A(z^{-1}) = F(z^{-1})$) (25) will have the following form:

$$G(z^{-1}) \cdot C(z^{-1}) = 1 \tag{26}$$

3. The numerical study of the two methods

Tables I and II show the results of the identifications for the same model by the two methods.

Table I belongs to the model of Fig. 1. Obviously, exact results are given by both of the methods in the case of small noises. In the cases of high noises more exact results can be obtained for the coefficients a_i and b_i by the Åström—Bohlin method and by the generalized least-squares method respectively. The disadvantage of the generalized least-squares method occurs in

Table I
(N = 500)

Parameters	Principal values	λ = 0.4 (10% noise)		
		L.S.E.	Åström	G.L.S.E.
a_1	−1.5	−1.45	−1.5	−1.5
a_2	0.7	0.65	0.7	0.703
b_1	1	1.05	1.02	1.03
b_2	0.5	0.496	0.474	0.47
g_1 c_1	−1 1	0.746*	−1.01	0.9
g_2 c_2	0.2 0.8	0.397*	0.209	0.585
HFV	—	83.02**	41.65	89.8
λ	.....	—	0.408	—

Table II
(N = 250); ($\lambda_0 = 0$)

Parameters	Principal values	$\lambda_i = 0.2$			$\lambda_i = 1$		
		L.S.E.	Åstrom	G.L.S.E.	L.S.E.	Åström	G.L.S.E.
a_1	−0.8	−0.801	−0.798	−0.796	−0.829	−0.783	−0.756
b_0	0.6	0.584	0.582	0.582	0.338	0.33	0.33
b_1	0.2	0.192	0.194	0.195	0.093	0.11	0.115
g_1	−0.8						
c_1	0.33	−0.35*	0.37	−0.364	−.342	0.391	−0.419
	−0.33						
f_1	−0.8						
HFV	—	1.8**	1.56	3.14	26.07	22.41	45.14

* The data have been gained from the G.L.S.E. estimation.
** The L.S.E. data are the initial data of the Åström-identification.

cases where the exact identification of the noise-model is also among the aims, since the method is not able to estimate λ, and since the determining of the coefficients g_i is difficult (because autoregressional noises are presumed) and the coefficients obtained on this way are moreover less exact, than the ones, obtained from the Åström–Bohlin method.

The numerical algorithm of the Åström–Bohlin method is more difficult, than the one of the generalized least-squares method, but only an

$\lambda = 1.8$ (50% noise)			$\lambda = 7.2$ (200% noise)		
L.S.E	Åström	G.L.S.E.	L.S.E.	Åström	G.L.S.E.
—1	—1.5	—1.47	—0.61	—1.5	—1.34
0.259	0.69	0.684	0.053	0.67	0.611
1.16	1.11	1.09	1.46	1.44	1.33
0.879	0.37	0.482	0.99	—0.08	0.476
0.22*	—1.01	0.889	0.026*	—1.01	0.83
—0.005*	0.204	0.561	—0.003*	0.167	0.439
1283**	843.1	1807.5	15110**	13440	28038
—	1.83	—	—	7.33	—

* The data have been gained from the G.L.S.E. estimation.
** The L.S.E. data are the initial data of the Åström-identification.

unsignificant quantity of time is gained because of the filters and the remaining matrix inversion in the latter one.

The noises, which may occur at the measure of the input signal is not taken into consideration at any of the two methods compared above.

In this case the structure of the system is as follows on Fig. 2:

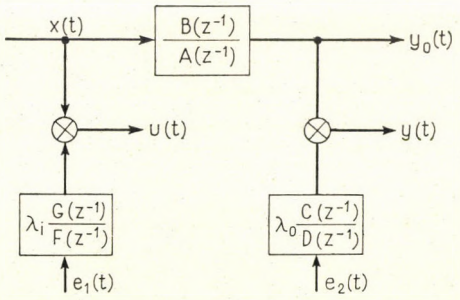


Fig. 2

The equation of the system is:

$$Y(z) = \frac{B(z^{-1})}{A(z^{-1})}\left[U(z) - \lambda_i \frac{G(z^{-1})}{F(z^{-1})} E_1(z)\right] + \lambda_0 \frac{C(z^{-1})}{D(z^{-1})} E_2(z) \tag{27}$$

If the following equation can be satisfied:

$$\lambda_0 \frac{C(z^{-1})}{D(z^{-1})} E_2(z) - \lambda_i \frac{B(z^{-1})\, G(z^{-1})}{F(z^{-1})\, A(z^{-1})} E_1(z) = \lambda \frac{J(z^{-1})}{H(z^{-1})} E(z) \tag{28}$$

then the input noise can be identificated as an equivalent output noise, but only formally, since the noise reduced on this way cannot be considered independently from the system. The exact identification of the input noise is complicated [5], and supposes, that either the standard deviation or the covariance-matrix of the input noise is known.

Table II shows the results if estimated identification is made for a non-autocorrelated input noise (where the noise is formally reduced to the output according to (28)). (The differences, arising at the two different methods where the signals and the numerical values of the noise-coefficients are estimated, come from the different noise-structure properties of the two methods.)

In this case we do not get as exact values for the process-parameters as in the case of the input noise — as it can be seen from Table II. Its reason is the input noise identification, having an approximating character. Even in the case of medium noise-level the results are good.

4. The study of the on-line aspects of the methods

The optimal matrix $\mathbf{\Gamma}[n]$ for the on-line application can be generally given for the two methods in the algorithmical description [4].

Using the notations of Eq. (9), (10):

$$\mathbf{\Gamma}[n] = \left[\sum_{m=1}^{n} \mathbf{H}(f(\mathbf{x}[m], \mathbf{c}[n-1]))\right]^{-1} \tag{29}$$

The optimal $\mathbf{\Gamma}[n]$ cannot be counted, however, recursively in the on-line version, since, in the Åstrom—Bohlin method the $\mathbf{H}$ matrix is the function of the system-parameters, as it could be seen above.

Nevertheless, the on-line version can be realised, but a suboptimal $\mathbf{\Gamma}[n]$ is wanted instead of the optimal $\mathbf{\Gamma}[n]$.

The on-line version of the generalized least-squares method can be realised without any difficulty, moreover, the matrix can be inverted on a recursive way.

The estimation algorithm is as follows:

$$\hat{\mathbf{Q}}[n] = \mathbf{Q}[n-1] + \frac{[\mathbf{X}^{FT}[n-1]\mathbf{X}^{F}[n-1]]^{-1}\mathbf{x}^{F}[n](y^{F}[n] - \mathbf{x}^{FT}[n]\hat{\mathbf{Q}}[n-1])}{d[n]};$$

$$d[n] = \varrho + \mathbf{x}^{FT}[n][\mathbf{X}^{FT}[n-1]\mathbf{X}^{F}[n-1]]^{-1}\mathbf{x}^{F}[n];$$

$$\hat{\mathbf{c}}[n] = \hat{\mathbf{c}}[n-1] - \frac{[\mathbf{\Omega}^{T}[n-1]\mathbf{\Omega}[n-1]]^{-1}\mathbf{\Delta}[n](\mathbf{\Delta}[n] - \mathbf{\Delta}^{T}[n]\hat{\mathbf{c}}[n-1])}{r[n]};$$

$$r[n] = \delta + \mathbf{\Delta}^{T}[n][\mathbf{\Omega}^{T}[n-1]\mathbf{\Omega}[n-1]]^{-1}\mathbf{\Delta}[n];$$

$$[\mathbf{X}^{FT}[n]\mathbf{X}^{F}[n]]^{-1} = \frac{1}{\varrho}\left\{[\mathbf{X}^{FT}[n-1]\mathbf{X}^{F}[n-1]]^{-1} - \right.$$

$$\left. - \frac{[\mathbf{X}^{FT}[n-1]\mathbf{X}^{F}[n-1]]^{-1}\mathbf{x}^{F}[n]\mathbf{x}^{FT}[n][\mathbf{X}^{FT}[n-1]\mathbf{X}^{F}[n-1]]^{-1}}{d[n]}\right\}$$

$$[\mathbf{\Omega}^{T}[n]\mathbf{\Omega}[n]]^{-1} = \frac{1}{\delta}\left\{[\mathbf{\Omega}^{T}[n-1]\mathbf{\Omega}[n-1]]^{-1} - \right.$$

$$\left. - \frac{[\mathbf{\Omega}^{T}[n-1]\mathbf{\Omega}[n-1]]^{-1}\mathbf{\Delta}[n]\mathbf{\Delta}^{T}[n][\mathbf{\Omega}^{T}[n-1]\mathbf{\Omega}[n-1]]^{-1}}{r[n]}\right\} \quad (30)$$

(ϱ and δ are the learning coefficients)

The defence against the singularity of the matrix $\mathbf{X}^T\mathbf{X}$ is not contained in the algorithm (it does not frequently occur at discrete identifications), for its solution we refer to [6].

The method is quick and easy [3].

Results, gained in the case of a first-ordes lag, can be seen on Fig. 3. The coefficients are satisfactorily exact already at 500 samplings. By the method the change of the coefficients can be adaptively taken into consideration.

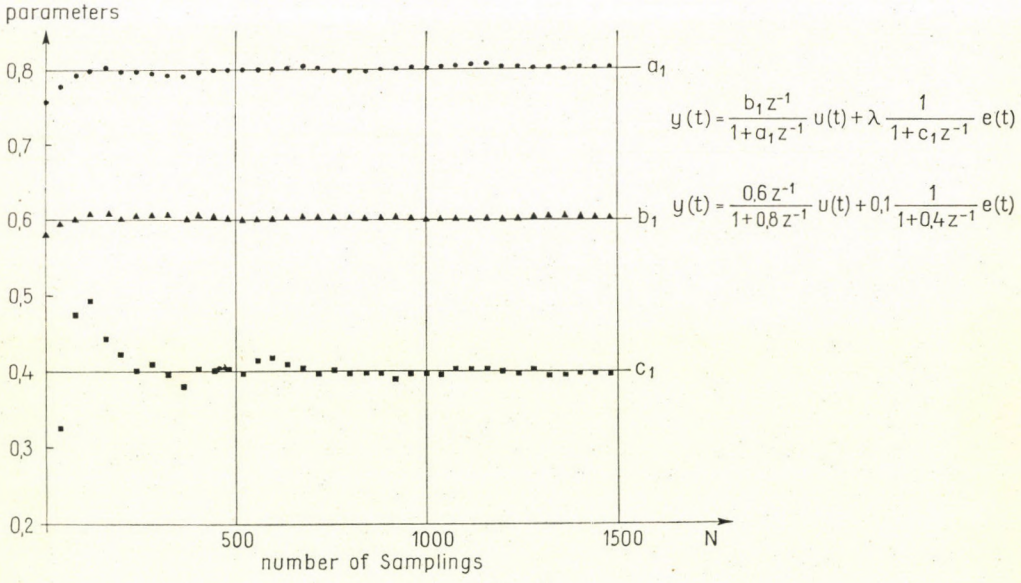


Fig. 3

References

Aström, K. J., Bohlin, T. (1966), Numerical identification of linear dynamic systems from normal operating records, *Proc. IFAC Symposium on Self-Adaptive Control Systems*. Teddington Sept. 1966.

Bohlin, T. (1970), On the maximum likelihood method of identification, *IBM J. Res. develop.* January 7970.

Hastings, R. J., Sage, M. W. (1969), Recursive generalised-least-squares procedure for online identification of process parameters. *Proc. IEE*, **116,** *12*. December.

Цыпкин Я. З. (1968), *Адаптация и обучение в автоматических системах.* «Наука», Москва

Rogers, A. E., Steiglitz, K. (1970), On system identification from noise-obscured input and output measurements, *Int. J. Control*, **12,** *4*.

Albert, A., Sittler, R. W. (1965), A method for computing least-squares estimators that keep up with the data. *Journal of SIAM eontrol*, Series A, **3,** *3*.

О двух методах дискретной идентификации систем

Ч. БАНЯС–Я. ГЕРТЛЕР

(Будапешт)

Резюме

В статье проведено сопоставление двух методов дискретной идентификации: метода максимального правдоподобия и обобщенного метода наименьших квадратов. Приведен алгоритм методов идентификации и показаны численные результаты.

Для разных случаев отношений помех и для различных структур помех сделаны проверки. На основе алгоритмов, разработанных на ЦВМ, сделаны выводы о возможности применения и вычислительной потребности данных методов при модификации методов типа «оф-лайн» и «он-лайн».

Актуальность этих исследований обоснована тем, что при нынешнем распространении управления процессами с помощью ЦВМ станет всё более необходимым дискретная идентификация линейных динамических систем. Причиной этого является обработка данных с помощью вычислительных машин и информация, полученная о параметрах процессов в дискретных моментах времени. Но дискретная модель в случае наборности может быть преобразована в непрерывную.

Cs. Bányász—dr. J. Gertler

Research Institute for Automation

Hungarian Academy of Sciences

Budapest II., Kende u. 13—17, Hungary

Problems of Control and Information Theory, Vol. 1 (3—4), pp. 297—305 (1972)

ОБ ОДНОЙ УПРОЩЕННОЙ МОДЕЛИ ВЗАИМОДЕЙСТВИЯ В КОЛЛЕКТИВЕ АВТОМАТОВ

В. Л. СТЕФАНЮК, С. Б. КОТЛЯР

(Москва)

(Поступила в редакцию 28 июня 1971 г.)

В статье рассматривается простейшая модель итеративного формирования мнения. Особенностью модели является случайный характер взаимодействия ее элементов. Удалось получить достаточно простое описание системы и исследовать ее поведение в зависимости от параметров.

Для многих задач технического, биологического, социологического характера существенным является обмен информацией или наличие взаимодействий между отдельными частями и наличие процесса, приводящего к некоторому равновесному состоянию. В статье рассматривается простейшая модель процесса формирования «мнения»[1] коллектива после поступления в коллектив определенной «начальной» информации. Это «мнение» формируется в результате обмена информацией внутри некоторого множества автоматов, которые обладают в некотором смысле целесообразным поведением. Предполагается, что в течение ряда последовательных тактов весь коллектив автоматов разбивается на группы и такой обмен информацией протекает независимо в каждой из групп. Заметим, что подобная ситуация итеративного формирования мнения на основе знания мнений остальных участков характерна, в частности, для метода Дальфи, применяемого для получения экспертных оценок [1, 2].

Будем рассматривать случай, когда каждый член коллектива представляет собой простейший автомат, выходная величина которого (или состояние) $\varphi(t)$ может принимать одно из двух значений, т. е. $\varphi(t) = 0, 1$. Поведение автомата во времени задается уравнением

$$\varphi(t+1) = F(\varphi(t), s(t), \xi(t)), \qquad (t = 0, 1, 2, \ldots),$$

где $s(t)$ — входная переменная, которая принимает два значения: 0 — нештраф, 1 — штраф, $\xi(t)$ — последовательность случайных независимых величин, так-

[1] Термин «мнение» коллектива понимается как, в некотором смысле, стационарное состояние всего коллектива. В случае коллектива людей — это набор мнений, которых придерживается каждый из членов коллектива.

же принимающих два значения 0, 1 с вероятностью $1-\varepsilon$, ε соответственно. Предполагается, что функция F задана следующим образом:

$$F(\varphi(t), 0, 0) = F(\varphi(t)\, 1, 1) = \varphi(t),$$

$$F(\varphi(t), 1, 0) = F(\varphi(t)\, 0, 1) = \overline{\varphi}(t) = 1 - \varphi(t).$$

При $\xi(t) = 0$ автомат меняет выходную величину только при штрафе, поэтому $\xi(t) = 1$ можно понимать как проявление «упрямства» автомата (с вероятностью ε).

По-другому можно представлять $\xi(t)$ как «ошибку» или «сбой» автомата.

Будем рассматривать коллектив из N таких автоматов, считая, что $\xi_i(t)$ $(i = 1, 2, \ldots, N)$ являются независимыми случайными величинами для всех членов коллектива. Ниже мы будем считать, что в каждый момент времени $t = 0, 1, 2 \ldots$ автоматы разбиваются на группы, причем величина входной переменной для каждого автомата в состоянии $\varphi(t)$ зависит от средней доли $X(t)$ автоматов в этой группе, выбравших выходную величину, равную $\varphi(t)$ (иными словами, находящихся в том же состоянии): если $X(t) > \frac{1}{2}$, то $s(t) = 1$, если $X(t) < \frac{1}{2}$, то $s(t) = 0$.

Таким образом, в каждой группе происходит «голосование» по большинству. Можно также рассматривать случай, когда таким образом «голосует» лишь один случайно выбранный автомат группы (т. е. все входные переменные всех остальных автоматов полагаются равными 0). Ниже всюду мы будем придерживаться второго способа голосования, когда случайно выделяется один автомат. Кроме того, везде, кроме приложения, предполагается, что в группе имеется ровно три автомата (N — кратно 3). Разбиение на группы коллектива автоматов происходит случайным образом. Предполагаем также, что величина ε для всех автоматов одна и та же (в дальнейшем это предположение будет несколько ослаблено).

Сделанные ограничения упрощают рассмотрение, но не являются обязательными. Из сказанного видно, что взаимодействие элементов имеет много общего с задачей «голосования с ошибкой», рассмотренной в ряде работ, например [3], в которых, однако, предполагалось, что все автоматы расположены на неограниченной прямой (или плоскости), соседи фиксировались раз и навсегда, и где целью исследования являлось обнаружение случаев неэргодичности системы.

На самом деле, легко показать, что при описанном выше групповом взаимодействии коллектив оказывается заведомо эргодической системой (но ниже мы будем рассматривать систему на достаточно малом отрезке времени,

на котором состояние всей системы с большой вероятностью определяется начальными условиями).

Рассмотрим функцию $\mu(t)$ — долю автоматов находящихся в момент времени t в состоянии 1.

Получаем приближенное уравнение, описывающее процесс (см. приложение).

$$\mu(t+1) = -\frac{2}{3}(1-2\varepsilon)\,\mu^3(t) + (1-2\varepsilon)\,\mu^2(t) + \frac{2}{3}\,\mu(t) + \frac{\varepsilon}{3}. \qquad \text{(I)}$$

Стационарные точки μ удовлетворяют уравнению

$$-\frac{2}{3}(1-2\varepsilon)\,\mu^3 + (1-2\varepsilon)\,\mu^2 - \frac{1}{3}\,\mu + \frac{\varepsilon}{3} = 0.$$

При $\varepsilon < \frac{1}{6}$ имеются три стационарные точки:

$$\mu^0_{1,3} = \frac{1}{2} \mp \frac{1}{2}\sqrt{\frac{1-6\varepsilon}{1-2\varepsilon}},$$

$$\mu^0_2 = \frac{1}{2}.$$

При $\frac{1}{6} \leq \varepsilon \leq \frac{1}{2}$ получаем стационарную точку $\mu = \frac{1}{2}$.

Как легко видеть, при $\varepsilon < \frac{1}{6}$ точки μ_1 и μ_3 являются устойчивыми, причем область устойчивости корня μ_1 есть $\left[0, \frac{1}{2}\right)$, а область устойчивости корня μ_3 есть $\left(\frac{1}{2}, 1\right]$ (при $\varepsilon > \frac{1}{2}$ существует единственная стационарная точка $\mu = \frac{1}{2}$)[2].

Полученный результат противоречит тому, что исходная система является эргодической. Это противоречие означает, что полученное описание может быть верным только на достаточно «малом» промежутке времени.

[2] Как отмечалось выше, факт разбиения всего коллектива на тройки не является обязательным для данного рассмотрения. Область для ε, при которых существуют три стационарные точки при разбиении, например, на пятерки, определяется неравенством $\varepsilon < \frac{7}{30}$ (см. приложение).

На некотором «малом» интервале времени происходит формирование и сохранение средней доли «1» в коллективе, величина которой зависит от начальных условий.

Для оценки времени пребывания в стационарных состояниях и вообще правомочности такого приближенного анализа было проведено моделирование описанного процесса на ЦВМ. Моделирование проводилось для $N = 1024$, $N = 510$, $N = 225$ автоматов. Как показало моделирование, приближенный анализ является достаточно точным.

Пример. Для $N = 510$, $\varepsilon = 0{,}05$, начального состояния $M_{\text{нач.}} = 200$ (т. е. 200 автоматов в начальный момент времени находятся в состоянии $\varphi(0) = 1$) система приходит в стационарную точку за 26 тактов. Система наблюдалась в течение 40 тыс. тактов и все это время находилась в окрестности стационарной точки. Таким образом, «малый» отрезок времени, на котором справедливо наше описание на самом деле может быть достаточно большим.

При $\varepsilon = 0$, 1 окрестность шире. Время формирования стационарной средней доли «1» системы примерно такое же. Интересно, что полученный результат значительно отличается от результатов, полученных в работе [3], где соседи фиксировались.

Можно говорить, что при ε достаточно малых такая система обладает памятью, так как состояние коллектива с достаточно большой вероятностью определено начальными условиями. (Память системы здесь понимаем в том же смысле, что и память каждого члена коллектива, который в интервале времени порядка $\frac{1}{\varepsilon}$ сохраняет свое состояние, если на вход не подается «штраф»).

Тогда оказывается, что коллектив в целом «помнит» свое состояние гораздо дольше, чем каждый член этого коллектива. Фактически речь идет о повышении надежности работы одного элемента способом близким к [4].

Можно было бы говорить также и о надежности операции «голосования» в такой системе.

Основываясь на том же приближенном описании, рассмотрим случай, когда коллектив N автоматов неоднороден.

Взаимодействие в этом коллективе такое же, но пусть коллектив распадается на n подколлективов с численностью соответственно $K_1, K_2, \ldots, K_n$, причем каждый такой подколлектив характеризуется своим параметром ε — вероятностью «ошибки» и предполагается, что $K_1, K_2, \ldots, K_n \gg 1$.

Рассмотрим, как свойства отдельных частей сказываются на поведении целого. Пусть $\alpha_i = \frac{K_i}{N}$ $(i = 1, 2, \ldots, n)$, тогда в уравнении для μ следует за-

менить параметр ε на $\bar{\varepsilon} = \sum_{i=1}^{n} \alpha_i \varepsilon_i$, где ε_i — вероятность «ошибки» для i-го подколлектива с численностью K_i.

Тогда, очевидно, что условием существования трех стационарных точек является $\bar{\varepsilon} < \frac{1}{6}$.

Из этого условия видно, что, если, например, таких подколлективов два и $\alpha_1 = \frac{7}{8}$ (т. е. I подколлектив существенно больше II-го), а $\varepsilon_1 = \frac{1}{7}$, то для выполнения условия $\bar{\varepsilon} < \frac{1}{6}$ достаточно, чтобы $\varepsilon_2 < \frac{1}{3}$.

На рис. 1—5 показаны некоторые характерные траектории $\mu(t)$ — доли единиц в коллективе из N автоматов, в котором происходит процесс формирования «мнения». Приведены средние значения $\mu(t)$ по 100 тактам времени ($t = 100, 200, \ldots$). (На рис. 1—3 и 4—5 траектории $\mu(t)$ для соответственно однородных и неоднородных коллективов автоматов).

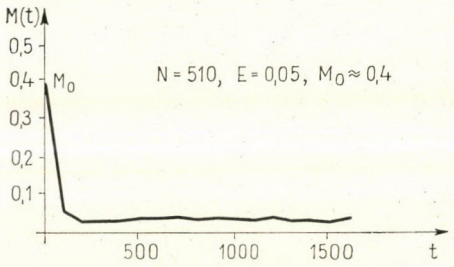


Рис. 1

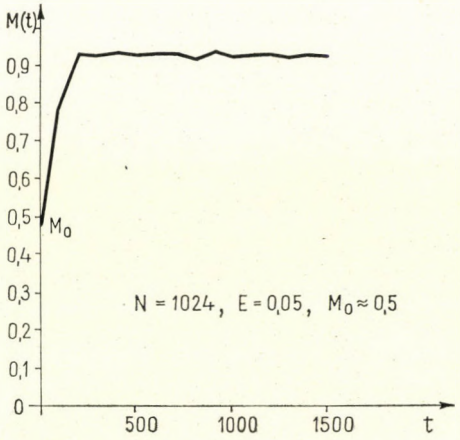


Рис. 2

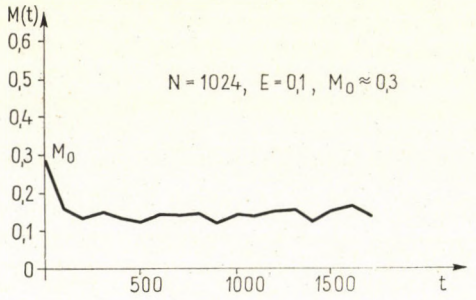


Рис. 3

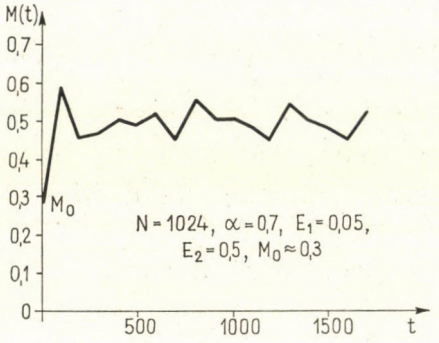


Рис. 4

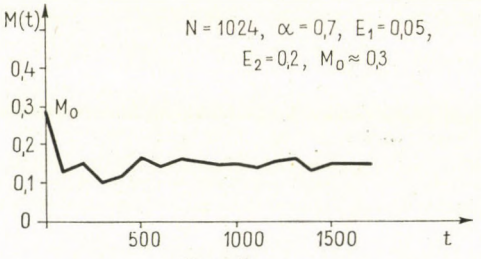


Рис. 5

Приложение

1. Вывод уравнения процесса формирования мнения

Ниже предполагается, что коллектив из N автоматов в каждый момент времени t случайным образом разбивается на группы из K автоматов (N кратно K), и в каждой группе «голосует» один случайно выбранный автомат.

Пусть в начальный момент времени $t = 0$ доля автоматов, находящихся в состоянии 1, задана и равна $\mu(0) = \mu_0$. Вычислим среднее значение $\mu(1)$ —

доли автоматов, находящихся в состоянии 1 в следующий момент времени $t = 1$. Заметим, что после разбиения на группы, вероятность того, что данная группа из K автоматов содержит ровно S автоматов в состоянии 1, равна $C_K^S \mu_0^S (1-\mu_0)^{K-S}$. Выбирая теперь случайно «голосующий» автомат и производя выбор мнения «по большинству», как описано в основном тексте, получаем следующее выражение:

$$\mu(1) = \Phi(\mu_0) = \mu_0 + \frac{1}{K}\left[(1-\varepsilon)\sum_{S>\frac{K}{2}} C_K^S \mu_0^S (1-\mu_0)^{K-S}\left(1-\frac{S}{K}\right) + \right.$$

$$+\varepsilon \sum_{S\le\frac{K}{2}} C_K^S \mu_0^S (1-\mu_0)^{K-S}\left(1-\frac{S}{K}\right) - (1-\varepsilon)\sum_{S<\frac{K}{S}} C_K^S \mu_0^S (1-\mu_0)^{K-S}\cdot\frac{S}{K} -$$

$$\left. -\varepsilon \sum_{S\ge\frac{K}{2}} C_K^S \mu_0^S (1-\mu_0)^{K-S}\cdot\frac{S}{K}\right]. \qquad (\text{П. 1})$$

Положительные члены в квадратных скобках дают среднюю суммарную долю автоматов, которые в силу процесса голосования переходят из состояния 0 в состояние 1.

Множитель $\frac{1}{K}$ отражает тот факт, что в группе меняет мнение лишь один случайно выбранный автомат.

Рассмотрим теперь рекуррентное соотношение:

$$\mu(t+1) = \Phi(\mu(t)), \qquad \mu(0) = \mu_0. \qquad (\text{П. 2})$$

Предположение, которого мы придерживаемся в настоящей работе, состоит в том, что: 1) средняя доля автоматов в состоянии 1 в каждый момент времени является представительной величиной, 2) эта средняя доля в момент t равна $\mu(t)$, вычисляемой по соотношению (П .2).

При $K = 3$ из (П. 2) получаем уравнение (1).

2. Исследование приближенного уравнения процесса при нечетных

$$K = 2l + 1$$

Будет показано, что при условии

$$\varepsilon < \frac{1}{2} - \frac{2^{2l-1}}{(2l+1)\,C_{2l}^{l}} \qquad (\text{П. 3})$$

у уравнения (П.1), как и в случае $K = 3$, имеются три стационарные точки: μ_1, $\mu_2 = \frac{1}{2}$, μ_3, причем μ_1, μ_2 расположены симметрично относительно точки $\frac{1}{2}$.

Из (П. 3) видно, что при $l \to \infty$ величина, ограничивающая ε, стремится к $\frac{1}{2}$. Получим условие (П. 3). В стационарном случае согласно (П. 2) получаем

$$\mu = \varphi(\mu)\,. \tag{П. 4}$$

Обозначив $\nu = \frac{\mu}{1-\mu}$ и учитывая $K = 2l + 1$ из (П. 2), (П. 4), получаем следующее уравнение для ν:

$$\begin{aligned} &\varepsilon K \nu^K + C_K^1(\varepsilon K - 1)\,\nu^{K-1} + C_K^2(\varepsilon K - 2)\,\nu^{K-2} + \ldots + \\ &+ C_K^l(\varepsilon K - l)\,\nu^{l+1} - C_K^l(\varepsilon K - l)\,\nu^l - C_K^{l-1}(\varepsilon K - l + 1)\,\nu^{l-1} - \ldots - \\ &- \ldots - C_k^1(\varepsilon K - 1)\,\nu - \varepsilon K = \psi(\nu) = 0\,. \end{aligned} \tag{П. 5}$$

Нетрудно видеть, что это уравнение имеет решение $\nu = 1$ и, кроме того, если оно имеет положительное решение ν^*, то $\frac{1}{\nu^*}$ также является его решением.

Также нетрудно заключить, используя правило Декарта о числе перемен знака в ряде коэффициентов, что число положительных решений (П. 5) при всех ε не более трех. Поскольку $\psi(\nu) \to +\infty$ при $\nu \to \infty$, то условие $\psi'(\nu)|_{\nu=1} < 0$ является необходимым и достаточным для того, чтобы (П. 5) имело три различных положительных корня (при $\psi'(1) = 0$ (П. 5) имеет корень $\nu = 1$ кратности 3).

Вычисляя $\psi'(\nu)|_{\nu=1}$ получаем, что для этого необходимо и достаточно, чтобы

$$\varepsilon < \frac{\sum\limits_{i=0}^{l-1} (2i+1)\,(l-i)\,C_K^{l-i}}{(2l+1)\sum\limits_{i=0}^{l-1} (2i+1)\,C_K^{l-i}}\,. \tag{П. 6}$$

Остается упростить (П. 6), чтобы получить (П. 1), использовав для этого очевидные соотношения:

$$\sum_{s=0}^{l-1} C_{2l-1}^{l+s} = 2^{2l-2}, \quad \sum_{s=0}^{l-1} C_{2l}^{l+s+1} = 2^{2l-1} - \frac{1}{2} C_{2l}^l\,.$$

Литература

1. *Янч Э.:* Научно-техническое прогнозирование в перспективе. М., «Прогресс», 1970.
2. *Глушков В. М.:* Прогнозирование на основе экспертных оценок. Кибернетика, 2, К., 1969.
3. *Васильев Н. Б.—Петровская М. Б.—Пятецкий-Шапиро И. И.:* Моделирование голосования со случайной ошибкой. Автоматика и телемеханика, 10, 1969.
4. *Нейман Дж.:* Вероятностная логика и синтез надежных организмов из ненадежных компонент. «Автоматы». Сб. статей под редакцией К. Э. Шеннона и Дж. Маккати, «Иностр. лит.», 1956.

On a simple scheme of the interaction of the collective members

V. L. STEFANIUK—S. B. KOTLIAR

(Moscow)

Summary

The question of yielding common "opinion" in the collective of some somponents plays an important role in many problems: technology, biology, social science and so on. This question was not yet answered in general, though some results are well known.

One simple scheme of the interaction of the members of the collective is considered, which will lead to the formation of certain collective "opinion" through the time.

The "opinion" of a member is "I" or "O". The "opinion" can be changed by the member for another one during the contacts with some other members of the collective. The parameter ε, named the "passivity" (or "obstinacy") of the member, is introduced. The collective is studied on a sufficiently small time interval. In spite of the system ergodicity the definite mean value of "ones" is formed during this time interval, the mean value fully depending on the initial conditions.

It was considered a simple description of the system, which allowed to study the role of ε in the variety of situations.

The description was checked by the computer simulation and the results of simulation appeared to be in a good agreement with the theoretical conception.

В. Л. Стефанюк — С. Б. Котляр
Институт проблем управления (автоматики и телемеханики)
СССР, Москва В—485, Профсоюзная ул, 81

Problems of Control and Information Theory, Vol. 1 (3—4), pp. 307—323 (1972)

ОПТИМАЛЬНОЕ ЛИНЕЙНОЕ ПРЕДЫСКАЖЕНИЕ И КОРРЕКТИРОВАНИЕ В СИСТЕМЕ ПЕРЕДАЧИ ИНФОРМАЦИИ С ДОПОЛНИТЕЛЬНЫМ ШУМОМ

Н. К. МИЛЕНИН

(Рязань)

(Поступила в редакцию 10 марта 1971 г.)

Рассматривается задача синтеза оптимальных линейных предыскажающих и корректирующих фильтров в стационарной системе передачи информации с двумя дополнительными источниками шумов.

В качестве критерия верности воспроизведения сообщений используется предложенный критерий минимума дисперсии общей взвешенной ошибки, равной сумме динамической и случайной ошибок, взвешенных по разным законам.

Приведена сравнительная оценка эффективности предыскажения сигналов и непосредственно самих сообщений.

1. Введение

Анализу и синтезу линейных предыскажающих (кодирующих) и корректирующих (декодирующих) устройств при передаче различных сообщений по стационарному каналу связи посвящены работы [1—12]. В этих работах рассматривается случай, когда в канале связи имеется только один источник флуктуационного шума, который непосредственно воздействует на вход корректирующего фильтра. Однако в реальных системах передачи информации кодированию часто подвергается не первоначальное сообщение, интересующее получателя, а лишь предварительно искаженный и зашумленный сигнал [13]. Дополнительные шумы могут также возникать в приемном устройстве как до, так и после декодирующего фильтра [13].

В настоящей работе проведен синтез оптимальных линейных предыскажающих и корректирующих фильтров, когда в системе передачи информации наряду с шумами канала связи воздействуют еще два источника дополнительных флуктуационных шумов. Одним из этих источников дополнительных шумов являются некоторые звенья передающего устройства, а другим — звенья приемного устройства.

При синтезе учтены линейные искажения сообщений и сигналов, возникающие в звеньях системы передачи информации до предыскажающего и после корректирующего фильтров.

Источниками дополнительных шумов и линейных искажений являются, в частности, фотографические и магнитные фонограммы [12], а также преобразователи сообщений в электрические сигналы, например телевизионные передающие камеры.

2. Блок-схема исследуемой системы передачи информации

На рис. 1. представлена блок-схема системы передачи информации, которая исследуется в данной работе. Стационарные гауссовские сообщения (одномерные или многомерные), математическое ожидание которых предполагается равным нулю, преобразуется без шумов и линейных искажений идеальным преобразователем сообщений в одномерный электрический сигнал $S(t)$. Полезный сигнал $S(t)$, представляющий стационарный случайный процесс с энергетическим спектром $S(\omega)$, искажается в реальном передающем устройстве (в том числе в реальном преобразователе сообщений в сигнал) с эквивалентной передаточной функцией $\Phi(j\omega)$. К искаженному сигналу $S_1(t)$ в передающем устройстве добавляется некоррелированный (независимый) с сигналом аддитивный шум $n_1(t)$ с пересчитанным ко входу предыскажающего фильтра энергетическим спектром $G(\omega)$. Далее суммарный случайный процесс $x(t) = S_1(t) + n_1(t)$ поступает на вход предыскажающего фильтра с

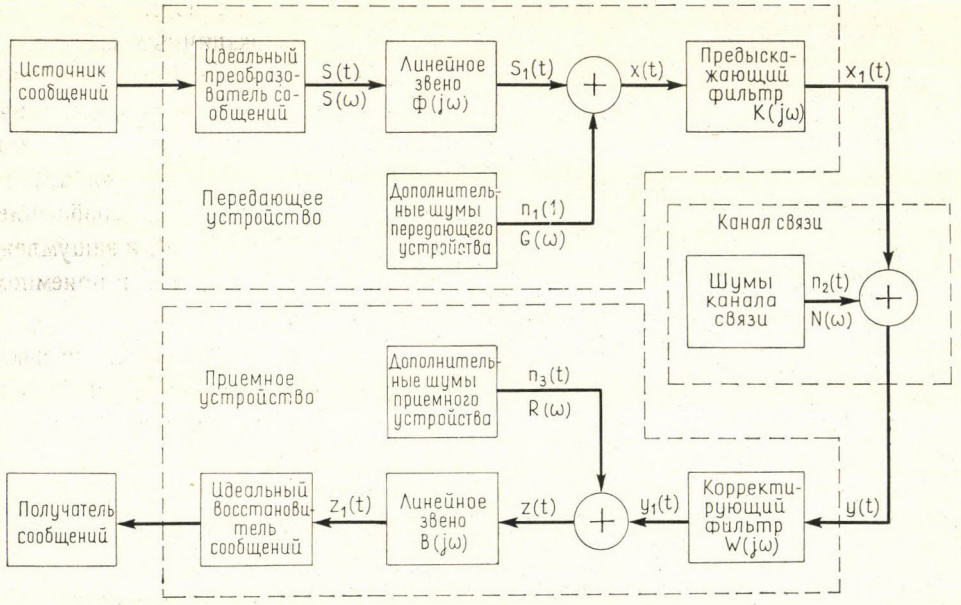


Рис. 1. Блок-схема исследуемой системы передачи информации

передаточной функцией $K(j\omega)$ к сигналу $x_1(t)$, на выходе которого добавляется независимый шум канала связи $n_2(t)$ с энергетическим спектром $N(\omega)$. Причем, к шумам канала связи отнесены все шумы, возникающие между предыскажающим и корректирующим фильтром.

После прохождения канала связи суммарный стационарный случайный процесс $y(t) = x_1(t) + n_2(t)$ поступает на вход корректирующего фильтра с передаточной функцией $W(j\omega)$. Выходной случайный процесс $y_1(t)$ складывается с независимым шумом $n_3(t)$, возникающим в звеньях приемного устройства, расположенных после корректирующего фильтра. Пересчитанный к выходу корректирующего фильтра энергетический спектр шума $n_3(t)$ обозначен $R(\omega)$. Суммарный случайный процесс $z(t) = y_1(t) + n_3(t)$ воздействует на вход линейного звена с передаточной функцией $B(j\omega)$. Это звено позволяет учесть линейные искажения процесса $z(t)$ во всех звеньях реального приемного устройства (в том числе в восстановителе сообщений), расположенных после корректирующего фильтра. Выходной стационарный случайный процесс $z_1(t)$, стационарно связанный с входным сигналом $S(t)$, преобразуется в принятое сообщение с помощью идеального восстановителя сообщений, который не вносит дополнительных искажений и шумов в рабочей полосе частот.

3. Критерий верности воспроизведения сообщений

В линейной системе передачи информации принятые сообщения отличаются от переданных за счет динамических и случайных ошибок. Коррелированные с передаваемым сигналом $S(t)$ динамические ошибки $\varepsilon_д(t) = S_{вых}(t) - S(t)$ возникают за счет отличия выходного полезного сигнала $S_{вых}(t)$ от входного $S(t)$ (рис. 2). Случайные ошибки $\varepsilon_ш(t)$ обусловлены воздействием флук-

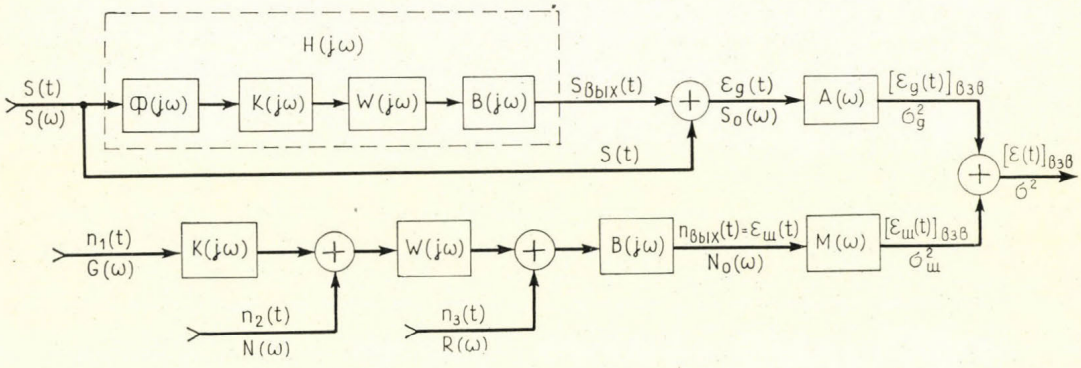


Рис. 2. Блок-схема для вычисления дисперсий ($\sigma^2_д, \sigma^2_ш$ и σ^2) взвешенных динамической $[\varepsilon_д(t)]_{взв}$, случайной $[\varepsilon_ш(t)]_{взв}$ и суммарной $[\varepsilon(t)]_{взв}$ ошибок в исследуемой системе передачи информации

туационных шумов $n_1(t)$, $n_2(t)$ и $n_3(t)$. Эти ошибки некоррелированы с сигналом $S(t)$ и определяются уровнем шумов на выходе системы (рис. 2), т. е. $\varepsilon_{ш}(t) = n_{вых}(t)$.

Мешающее действие динамических и случайных ошибок при одинаковой их средней мощности различно и зависит от свойств получателя сообщений. Поэтому верность воспроизведения сообщений определяется взвешенными (по разным законам) динамической $[\varepsilon_{д}(t)]_{взв}$ и случайной $[\varepsilon_{ш}(t)]_{взв}$ ошибками.

В качестве критерия верности воспроизведения сообщений будем использовать критерий минимума дисперсии σ^2 суммарной взвешенной ошибки $[\varepsilon(t)]_{взв}$, равной сумме дисперсий $\sigma^2_{д}$ и $\sigma^2_{ш}$ независимых между собой взвешенных динамической и случайной ошибок, т. е.

$$\left.\begin{aligned} [\varepsilon(t)]_{взв} &= [\varepsilon_{д}(t)]_{взв} + [\varepsilon_{ш}(t)]_{взв} \\ \sigma^2 &= \sigma^2_{д} + \sigma^2_{ш} = \min \end{aligned}\right\}. \tag{3.1}$$

Для наглядности расчетов взвешенных дисперсий $\sigma^2_{д}$ и $\sigma^2_{ш}$ применяется принцип суперпозиции, т. е. прохождение полезного сигнала $S(t)$ и шумов через линейную систему рассматривается раздельно, как это показано на рис. 2. При этом вычисление дисперсий $\sigma^2_{д}$ и $\sigma^2_{ш}$ взвешенных динамической и случайной ошибок производится в частотной области с помощью разных весовых функций $A(\omega)$ и $M(\omega)$ соответственно.

Величина времени задержки корректирующих и предыскажающих фильтров, как и в работах [1—12], не ограничивается, т. е. характеристики этих фильтров синтезируются без учета условия физической осуществимости.

4. Синтез оптимальной передаточной функции корректирующего фильтра

Передаточные функции дополнительных линейных звеньев $\Phi(j\omega)$ и $B(j\omega)$ будем все время считать заданными. Зафиксируем сначала передаточную функцию предыскажающего фильтра $K(j\omega)$ и подберем передаточную функцию корректирующего фильтра $W(j\omega)$ так, чтобы величина дисперсии σ^2 суммарной взвешенной ошибки была минимальна..

Взвешенную дисперсию σ^2 (3.1) выразим через энергетические спектры $S_0(\omega)$ и $N_0(\omega)$ динамической и случайной ошибок, а также весовые функции $A(\omega)$ и $M(\omega)$, т. е. (рис. 2)

$$\sigma^2 = \frac{1}{2\pi}\int\limits_{\omega_н}^{\omega_в} S_0(\omega)\, A^2(\omega)\, d\omega + \frac{1}{2\pi}\int\limits_{\omega_н}^{\omega_в} N_0(\omega)\, M^2(\omega)\, d\omega, \tag{4.1}$$

где $\omega_н = 2\pi f_н$; $\omega_в = 2\pi f_в$; $f_н$ и $f_в$ — нижняя и верхняя границы полосы пропускания, за пределами которой весовая функция $A(\omega)$ равна нулю.[1]

Из рис. 2 следует, что энергетические спектры $S_0(\omega)$ и $N_0(\omega)$ динамической и случайной ошибок соответственно равны:

$$S_0(\omega) = S(\omega)\,|H(j\omega) - 1|^2 = S(\omega)\,|\Phi(j\omega)\,K(j\omega)\,W(j\omega)\,B(j\omega) - 1|^2, \quad (4.2)$$

$$N_0(\omega) = \{[G(\omega)\,|K(j\omega)|^2 + N(\omega)]\,|W(j\omega)|^2 + R(\omega)\}\,|B(j\omega)|^2, \quad (4.3)$$

где $H(j\omega)$ — передаточная функция линейной системы в целом.

Подставляя соотношения (4.2) и (4.3) в (4.1), получаем

$$\sigma^2 = \frac{1}{2\pi}\int_{\omega_н}^{\omega_в} \{S(\omega)\,A^2(\omega)\,|\Phi(j\omega)\,K(j\omega)\,W(j\omega)\,B(j\omega) - 1|^2 +$$

$$+ M^2(\omega)\,|B(j\omega)|^2\{[G(\omega)\,|K(j\omega)|^2 + N(\omega)]\,|W(j\omega)|^2 + R(\omega)\}\}\,d\omega\,. \quad (4.4)$$

В формулу (4.1) входит положительный множитель

$$|H(j\omega) - 1| = |\Phi(j\omega)\,K(j\omega)\,W(j\omega)\,B(j\omega) - 1|\,, \quad (4.5)$$

поэтому величина дисперсии σ^2 суммарной взвешенной ошибки зависит как от модуля, так и от аргумента передаточной функции корректирующего фильтра $W(j\omega)$. Дисперсия σ^2 (4.4) может только уменьшиться, если подбором $\arg W(j\omega)$ сделать передаточную функцию всей системы $H(j\omega)$ действительным числом

$$H(j\omega) = |H(j\omega)|\,, \quad (4.6)$$

когда величина модуля разности (4.5) становится наименьшей.

Оптимальную передаточную функцию корректирующего фильтра $W_0(j\omega)$, обеспечивающего при фиксированном значении $K(j\omega)$ минимум дисперсии σ^2 (4.4), можно найти, например, если решить известную вариационную задачу [14]. Составив уравнение Эйлера для подинтегральной функции $F[W(j\omega)]$ в выражении (4.4) $\partial F[W(j\omega)]/\partial W(j\omega) = 0$ и решив его при соблюдении условия (4.6), будем иметь

$$W_0(j\omega)\,B(j\omega) = \frac{\Phi^*(j\omega)\,K^*(j\omega)\,S(\omega)}{|\Phi(j\omega)|^2\,|K(j\omega)|^2\,S(\omega) + L^2(\omega)\,[|K(j\omega)|^2\,G(\omega) + N(\omega)]}\,, \quad (4.7)$$

где $\Phi^*(j\omega)$ и $K^*(j\omega)$ — функции, сопряженные передаточным функциям $\Phi(j\omega)$ и $K(j\omega)$ соответственно, а $L(\omega) = M(\omega)/A(\omega)$.

[1] Если полоса пропускания системы выбрана из других соображений, то искажения, обусловленные устранением частотных компонент полезного сигнала, лежащих вне этой полосы, следует также относить к ошибкам воспроизведения сообщения.

Из соотношений (4.4) и (4.7) следует, что минимальное значение дисперсии σ_0^2 суммарной взвешенной ошибки при фиксированном значении передаточной функции предыскажающего фильтра $K(j\omega)$ равно:

$$\sigma_0^2 = \frac{1}{2\pi} \int_{\omega_H}^{\omega_B} \left\{ \frac{S(\omega)\, A^2(\omega)}{1 + \dfrac{|\Phi(j\omega)|^2\, |K(j\omega)|^2\, S(\omega)\, A^2(\omega)}{[|K(j\omega)|^2\, G(\omega) + N(\omega)]\, M^2(\omega)}} + R(\omega)\, M^2(\omega)\, |B(j\omega)|^2 \right\} d\omega . \tag{4.8}$$

5. Нахождение оптимальной передаточной функции предыскажающего фильтра

Вычислим передаточную функцию предыскажающего фильтра $K(j\omega)$, обеспечивающего минимум дисперсии σ_0^2 (4.8), при оптимальном значении передаточной функции $W_0(j\omega)$ (4.7) корректирующего фильтра и ограниченной средней мощности Q сигнала $x_1(t)$ на входе канала связи (рис. 1)

$$Q = \frac{1}{2\pi} \int_{\omega_H}^{\omega_B} [|\Phi(j\omega)|^2\, S(\omega) + G(\omega)]\, |K(j\omega)|^2\, d\omega . \tag{5.1}$$

Из формул (4.8) и (5.1) видно, что фазовая характеристика предыскажающего фильтра $\arg\,(K(j\omega))$ не оказывает влияния на величины взвешенной дисперсии σ_0^2 (4.8) и средней мощности передающего устройства Q (5.1), а поэтому может быть выбрана произвольной. Однако выбор $\arg K(j\omega)$ предопределяет собой оптимальное значение $\arg W_0(j\omega)$ (4.7). На основании выражений (4.6) и (4.7) можно заключить, что оптимальные фазовые характеристики передающего и приемного устройств должны быть взаимно обратными.

Оптимальное значение частотной характеристики предыскажающего фильтра, обеспечивающего минимум дисперсии σ_0^2 (4.8), при выполнении условия (5.1) и очевидного условия $|K(j\omega)| \geq 0$ можно найти, если решить изопериметрическую задачу вариационного исчисления [2]. Здесь уравнение Эйлера имеет вид

$$\frac{\partial}{\partial |K(j\omega)|^2} \left\{ \frac{S(\omega)\, M^2(\omega)\, [G(\omega)\, |K(j\omega)|^2 + N(\omega)]}{|\Phi(j\omega|^2\, |K(j\omega)^2|\, S(\omega) + L^2(\omega)\, [|K(j\omega)|^2\, G(\omega) + N(\omega)]} + \right.$$

$$\left. + R(\omega)\, M^2(\omega)\, |B(j\omega)|^2 + \lambda\, [|\Phi(j\omega)|^2\, S(\omega) + G(\omega)]\, |K(j\omega)|^2 \right\} = 0 , \tag{5.2}$$

где λ — некоторый постоянный множитель.

После дифференцирования из уравнения (5.2) находим, что квадрат оптимальной частотной характеристики предыскажающего фильтра $|K_0(j\omega)|^2$ равен:

$$|K_0(j\omega)|^2 = \begin{cases} \dfrac{S(\omega)\sqrt{N(\omega)\,M^2(\omega)\,|\Phi(j\omega)|^2}}{\sqrt{\lambda}\,[|\Phi(j\omega)|^2 S(\omega) + G(\omega)]} - \dfrac{N(\omega)\,L^2(\omega)}{|\Phi(j\omega)|^2 S(\omega) + G(\omega)\,L^2(\omega)} & \text{при } h(\omega) \geq 1, \\ 0 & \text{при } h(\omega) < 1, \end{cases} \tag{5.3}$$

где

$$h(\omega) = \frac{S(\omega)\,A^2(\omega)\,|\Phi(j\omega)|}{\sqrt{\lambda N(\omega)\,M^2(\omega)\,[|\Phi(j\omega)|^2 S(\omega) + G(\omega)]}}. \tag{5.4}$$

Из формул (4.7) и (5.3) следует, что оптимальные предыскажающий и корректирующий фильтры полностью исключают из передачи те участки полосы пропускания системы, для которых справедливо неравенство $h(\omega) < 1$.

Подставляя выражение (5.3) в (4.8), получаем наименьшее значение дисперсии $\sigma^2_{\text{мин}}$ суммарной взвешенной ошибки при оптимальных передаточных функциях корректирующего $W_0(j\omega)$ (4.7) и предыскажающего $K_0(j\omega)$ (5.3) фильтров:

$$\sigma^2_{\text{мин}} = \frac{1}{2\pi}\int_{\omega_\text{н}}^{\omega_\text{в}} R(\omega)\,M^2(\omega)\,|B(j\omega)|^2\,d\omega + \frac{1}{2\pi}\int_{\Delta\omega_2} S(\omega)\,A^2(\omega)\,d\omega + \tag{5.5}$$

$$+\frac{1}{2\pi}\int_{\Delta\omega_1}\frac{S(\omega)\{\sqrt{\lambda N(\omega)\,M^2(\omega)\,|\Phi(j\omega)|^2\,[S(\omega)\,|\Phi(j\omega)|^2 + G(\omega)]} + G(\omega)\,M^2(\omega)\}}{S(\omega)\,|\Phi(j\omega)|^2 + G(\omega)\,L^2(\omega)}\,d\omega,$$

где $\Delta\omega_1$ и $\Delta\omega_2$ — области частот, расположенные в полосе пропускания системы $\Delta\omega = \omega_\text{в} - \omega_\text{н} = \Delta\omega_1 + \Delta\omega_2$, каждая из которых определяется неравенством $h(\omega) \geq 1$ и $h(\omega) < 1$ соответственно.

Постоянную λ легко определить, если подставить выражение (5.3) в (5.1). При этом

$$\sqrt{\lambda} = \frac{\displaystyle\int_{\Delta\omega_1}\frac{S(\omega)\sqrt{N(\omega)\,M^2(\omega)\,|\Phi(j\omega)|^2\,[S(\omega)\,|\Phi(j\omega)|^2 + G(\omega)]}}{S(\omega)\,|\Phi(j\omega)|^2 + G(\omega)\,L^2(\omega)}\,d\omega}{\displaystyle\int_{\Delta\omega_1}\frac{N(\omega)\,L^2(\omega)\,[S(\omega)\,|\Phi(j\omega)|^2 + G(\omega)]}{S(\omega)\,|\Phi(j\omega)|^2 + G(\omega)\,L^2(\omega)}\,d\omega + \int_{\omega_\text{н}}^{\omega_\text{в}}[S(\omega)\,|\Phi(j\omega)|^2 + G(\omega)]\,d\omega}. \tag{5.6}$$

6. Некоторые частные критерии верности

В предыдущих разделах сформулирован и использовался наиболее общий (в рамках корреляционной теории случайных процессов) частотно-взвешенный среднеквадратичный критерий верности воспроизведения сообщений с неодинаковыми весовыми функциями $A(\omega)$ и $M(\omega)$ для динамических и случайных ошибок. Накладывая определенные ограничения на выбор весовых функций $A(\omega)$ и $M(\omega)$, можно получить несколько частных критериев.

1. Случай, когда $A(\omega) = M(\omega) = 1$ соответствует обычному среднеквадратичному критерию верности воспроизведения сообщений,[2] который не учитывает особенностей восприятия динамических и случайных ошибок получателем сообщений.

2. Часто весовые функции $A(\omega)$ и $M(\omega)$, зависящие от частоты ω, отличаются друг от друга только постоянным множителем k, т. е.

$$L(\omega) = \frac{M(\omega)}{A(\omega)} = k\,, \tag{6.1}$$

где обычно $0 \leq k \leq 1$. При выполнении условия (6.1) критерий верности воспроизведения сообщений можно назвать частотно-взвешенным среднеквадратичным критерием с неодинаковыми весами (стоимостями) динамических и случайных ошибок.

3. Если в формуле (6.1) положить $k = 1$, то получим частотновзвешенный среднеквадратичный критерий, при использовании которого вес динамических и случайных ошибок считается одинаковым, что не всегда справедливо.

4. Когда $\Phi(j\omega) = B(j\omega) = 1$, а коэффициент k (6.1) принят равным нулю ($k = 0$), то критерий верности воспроизведения сообщений будет совпадать с критерием максимума отношения мощности полезного сигнала к взвешенной мощности шума при взаимно обратных передаточных функциях корректирующего и предыскажающего фильтров. При этом эффективность применения оптимальных предыскажений сигнала можно оценить с помощью безразмерного коэффициента

$$\beta_1 = \sigma_0^2/\sigma_{\text{мин}}^2\,, \tag{6.2}$$

где при вычислении дисперсии σ_0^2 в формуле (4.8) нужно положить $K(j\omega) = 1$. После подстановки соотношений (4.8), (5.5) и (5.6) в формулу (6.2), получаем (при $L(\omega) = 0$, $\Phi(j\omega) = B(j\omega) = 1$)

[2] По этому критерию задача линейного кодирования двумерных изображений в ситуации с дополнительным шумом впервые рассмотрена Б. С. Цыбаковым в 1962 г.

$$\beta_1 = \int_{\omega_н}^{\omega_в} [G(\omega) + N(\omega) + R(\omega)]\, M^2(\omega)\, d\omega \int_{\omega_н}^{\omega_в} [S(\omega) + G(\omega)]\, d\omega \times$$
$$\times \left(\left\{ \int_{\omega_н}^{\omega_в} \sqrt{N(\omega)\, M^2(\omega)\, [S(\omega) + G(\omega)]}\, d\omega \right\}^2 + \right.$$
$$\left. + \int_{\omega_н}^{\omega_в} M^2(\omega)\, [G(\omega) + R(\omega)]\, d\omega \int_{\omega_н}^{\omega_в} [S(\omega) + G(\omega)]\, d\omega \right)^{-1}. \quad (6.3)$$

Коэффициент β_1 в (6.3) характеризует выигрыш в отношении сигнал—шум при использовании предыскажений. Выигрыш отсутствует ($\beta_1 = 1$), когда

$$N(\omega)\, M^2(\omega) \equiv S(\omega) + G(\omega). \quad (6.4)$$

5. Если $|\Phi(j\omega)| \neq 1$, $|B(j\omega)| \neq 1$, а $G(\omega) \neq 0$ и $R(\omega) \neq 0$, то выбор $k = 0$ в большинстве случаев не может быть оправданным, так как такой выбор предписывает производить полную коррекцию линейных искажений полезного сигнала при любом уровне дополнительных шумов, что приводит к чрезмерному увеличению уровня случайных ошибок (шумов). Поэтому здесь наиболее целесообразно считать, что $0 < k < 1$. Конкретное значение коэффициента k в той или иной системе передачи информации можно установить экспериментально.

7. Оптимальное линейное предыскажение сообщений

В ряде случаев оказывается возможным подвергать предыскажению и корректированию непосредственно сами сообщения, т. е. производить предыскажение сообщений до их преобразования в видеосигнал, а корректирование осуществлять после восстановления сообщений из видеосигнала [5, 6, 10]. Блок-схема такой системы передачи информации показана на рис. 3, где

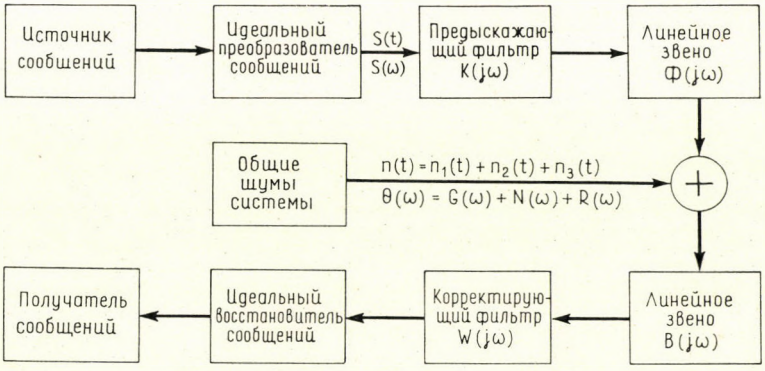


Рис. 3. Блок-схема системы передачи информации, в которой предыскажению и корректированию подвергаются непосредственно сами сообщения

линейные предыскажающий и корректирующий фильтры (с целью представления их характеристик в одномерном виде) переставлены местами с идеальными преобразователем и восстановителем сообщений соответственно.

Нетрудно показать, что в последнем случае при ограниченной средней мощности передатчика

$$Q = \frac{1}{2\pi} \int_{\omega_н}^{\omega_в} [S(\omega) \, |\Phi(j\omega)|^2 \, |K(j\omega)|^2 + G(\omega)] \, d\omega$$

можно также использовать все ранее полученные соотношения, если положить в них $G(\omega) = R(\omega) = 0$, а вместо $N(\omega)$ подставить значение суммарного энергетического спектра шумов системы $\Theta(\omega) = G(\omega) + N(\omega) + R(\omega)$ (рис. 3). В частности, при $L(\omega) = 0$ и $\Phi(j\omega) = B(j\omega) = 1$ из формулы (6.3) следует, что выигрыш в отношении сигнала к шуму за счет предыскажения сообщений β_2 равен:

$$\beta_2 = \frac{\int_{\omega_н}^{\omega_в} \Theta(\omega) \, M^2(\omega) \, d\omega \int_{\omega_н}^{\omega_в} S(\omega) \, d\omega}{\left[\int_{\omega_н}^{\omega_в} \sqrt{S(\omega) \, \Theta(\omega) \, M^2(\omega)} \, d\omega\right]^2} . \tag{6.5}$$

Выигрыш в отношении сигнал-шум при предыскажении сообщений отсутствует ($\beta_2 = 1$), если

$$S(\omega) \equiv M^2(\omega) \, [G(\omega) + N(\omega) + R(\omega)] . \tag{6.6}$$

Сравнивая между собой выражения (6.4) и (6.6) приходим к выводу, что при выполнении условия (6.4) эффективными могут оказываться только предыскажения самих сообщений (до их преобразования в сигнал), а при выполнении условия (6.6) — предыскажения сигналов (после преобразования сообщений в сигнал).

8. Пример

С целью иллюстрации полученных соотношений рассчитаем эффективность предыскажения телевизионного сигнала по критерию максимума отношения сигнал—шум при взаимно обратных характеристиках предыскажающего и корректирующего фильтров, когда $L(\omega) = 0$, $R(\omega) = 0$, $\Phi(j\omega) = B(j\omega) = 1$ и $\omega_н = 0$. Для упрощения расчетов ограничимся рассмотрением

только огибающих (грубой структуры) энергетического спектра полезного сигнала и весовой функции помех, которые имеют вид [11]:

$$S(\omega) = \frac{S_0}{1 + \alpha^2 \omega^2}, \tag{8.1}$$

$$M^2(\omega) = \frac{1}{1 + \tau^2 \omega^2}, \tag{8.2}$$

где S_0, α и τ — постоянные величины.

Энергетический спектр дополнительных (внутренних) шумов передающего устройства будем считать равномерным

$$G(\omega) = G_0, \tag{8.3}$$

что справедливо для передающих камер на суперортиконе, а спектр внешних шумов (канала связи) представим выражением

$$N(\omega) = N_0(m^2 + \gamma^2 \omega^2), \tag{8.4}$$

где G_0, N_0, m, γ — постоянные величины. При $m = 1$ и $\gamma = 0$ спектр внешних шумов становится равномерным, а при $m = 0$ квадратичным.

Подставляя выражения (8.1—8.4) в (5.3) и (6.3), получаем

$$|K_0(j\omega)|_1^2 = \sqrt{\frac{N_0(m^2 + \gamma^2 \omega^2)(1 + \alpha^2 \omega^2)}{\lambda(1 + \tau^2 \omega^2)[S_0 + G_0(1 + \alpha^2 \omega^2)]}}, \tag{8.5}$$

$$\beta_1 = \frac{1 + \dfrac{N_0}{G^0}\left[m^2 + \dfrac{\gamma^2}{\tau^2}\left(\dfrac{\omega_в \tau}{\operatorname{arc\,tg} \omega_в \tau} - 1\right)\right]}{1 + \dfrac{N_0}{G_0} \dfrac{\alpha\tau R_0^2(\alpha, \tau, \gamma, m, \omega_в, S_0, G_0)}{(G_0 \omega_в \alpha + S_0 \operatorname{arc\,tg} \omega_в \alpha) \operatorname{arc\,tg} \omega_в \tau}}, \tag{8.6}$$

где

$$R_0(\alpha, \tau, \gamma, m, \omega_в, S_0, G_0) = \int_0^{\omega_в} \sqrt{\frac{(m^2 + \gamma^2 \omega^2)[S_0 + G_0(1 + \alpha^2 \omega^2)]}{(1 + \alpha^2 \omega^2)(1 + \tau^2 \omega^2)}}\, d\omega. \tag{8.7}$$

Форма квадрата оптимальной частотной характеристики предыскажающего фильтра (8.5) для разных отношений S_0/G_0 при условии, что $N(\omega) = N_0$ и $\alpha/\tau = 10$, изображена на рис. 4.

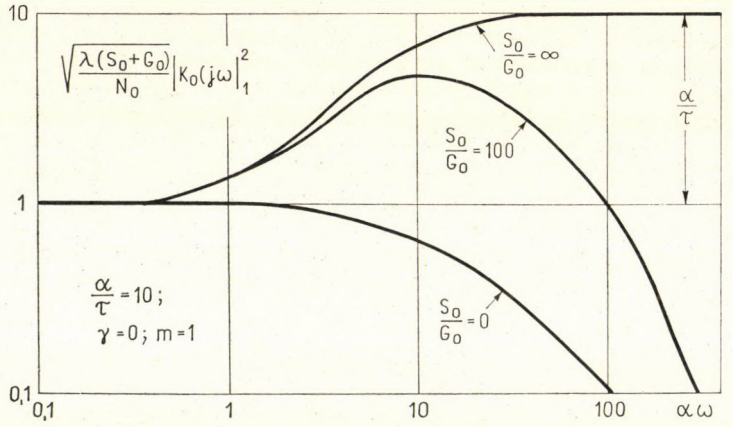


Рис. 4. Форма квадрата частотной характеристики оптимального предыскажающего фильтра при наличии дополнительных шумов в передающем устройстве

Выигрыш в отношении сигнал—шум за счет предыскажения сигнала вычислим по формулам (8.6) и (8.7) для нескольких случаев.

1. Пусть $N(\omega)=N_0\omega^2\gamma^2$ и $S_0/G_0=\alpha^2/\tau^2-1$. Тогда

$$\beta_1=\frac{1+(\omega_в\tau/\operatorname{arc\,tg}\omega_в\tau-1)N_0\gamma^2/G_0\tau^2}{1+\dfrac{N_0\gamma^2}{G_0\tau^2}\dfrac{[\sqrt{1+\alpha^2\omega_в^2}-1]^2}{\left[\omega_в\tau+\left(\dfrac{\alpha}{\tau}-\dfrac{\tau}{\alpha}\right)\operatorname{arc\,tg}\omega_в\alpha\right]\dfrac{\alpha^2}{\tau^2}\operatorname{arc\,tg}\omega_в\tau}}. \tag{8.8}$$

2. Если $N(\omega)=N_0\gamma^2\omega^2$, а $S_0/G_0\to\infty$, то

$$\beta_1=\frac{1+(\omega_в\tau/\operatorname{arc\,tg}\omega_в\tau-1)N_0\gamma^2/G_0\tau^2}{1+\dfrac{N_0\gamma^2}{G_0\tau^2}\dfrac{\dfrac{\tau}{\alpha}\left[\ln\dfrac{\alpha\sqrt{1+\omega_в^2\tau^2}+\tau\sqrt{1+\omega_в^2\alpha^2}}{\alpha+\tau}\right]^2}{\operatorname{arc\,tg}\omega_в\tau\operatorname{arc\,tg}\omega_в\alpha}}. \tag{8.9}$$

3. Когда $N(\omega)=N_0$, а $S_0/G_0=\alpha^2/\tau^2-1$, то

$$\beta_1=\frac{1+N_0/G_0}{1+\dfrac{N_0}{G_0}\dfrac{\ln^2(\omega_в\alpha+\sqrt{1+\omega_в^2\alpha^2})}{[\omega_в\tau+(\alpha/\tau-\tau/\alpha)\operatorname{arc\,tg}\omega_в\alpha]\operatorname{arc\,tg}\omega_в\tau}}, \tag{8.10}$$

4. Пусть $N(\omega)=N_0$, а $S_0/G_0\to\infty$. Тогда

$$\beta_1=\frac{1+N_0/G_0}{1+\dfrac{N_0}{G_0}\dfrac{\tau F^2(\varphi,k_0)}{\alpha\operatorname{arc\,tg}\omega_в\tau\operatorname{arc\,tg}\omega_в\alpha}}, \tag{8.11}$$

где $F(\varphi, k_0)$ — эллиптический интеграл первого рода,

$$\varphi = \operatorname{arc\,tg}\,\omega_в \alpha, \qquad k_0 = \sqrt{1 - \frac{\tau^2}{\alpha^2}}, \qquad \tau \leq \alpha.$$

На рис. 5а, б представлены кривые зависимостей выигрыша в отношении сигнал—шум за счет предыскажения сигнала β_1 (8.8—8.11) от отношения величина α/τ в случае квадратичного (рис. 5а) и равномерного (рис. 5б) энергетического спектра шумов канала связи. Вычисление коэфициента β_1 проведено по формулам (8.8—8.11) для различных отношений S_0/G_0 и S_0/G_0 при условии, что $\omega_в\tau = 12$ (для отечественного стандарта $\tau = 0{,}3 \div 0{,}33$ мксек, $f_в = \omega_в/2\pi = 6$ Мгц)[3].

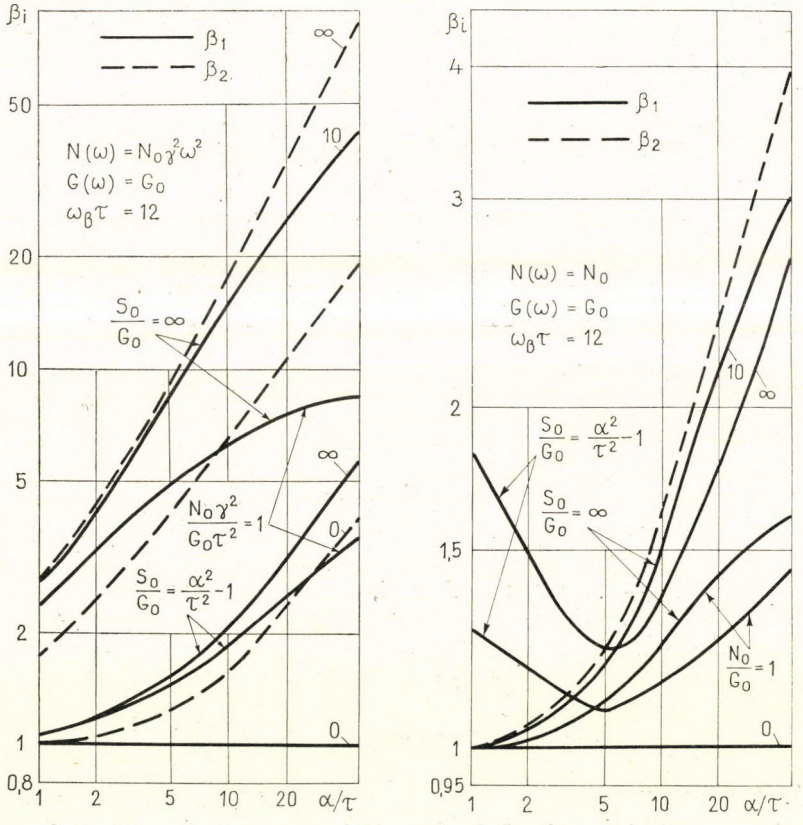


Рис. 5. Зависимость выигрыша в отношении сигнал—шум при использовании предыскажений сигнала (β_1) и сообщений (β_2) от отношения α/τ

[3] В разделе 8 считается, что искажения, вызванные устранением частотных компонент сигнала, лежащих выше $\omega_в$, зрителем не замечаются, т. е. $\omega_в$ выбрана так, что $A(\omega) = 0$ при $\omega_в < \omega \leq \infty$.

Для сравнения оценим также выигрыш в отношении сигнал—шум за счет одномерного предыскажения непосредственно телевизионных изображений, когда $L(\omega) = 0$; $B(j\omega) = \Phi(j\omega) = 1$, $R(\omega) = 0$, $\omega_в = 0$. Здесь выражения (8.1—8.4) необходимо подставить в формулу (6.5). При этом будем иметь

$$\beta_2 = \frac{[\tau^2(G_0 + m^2 N_0)\operatorname{arc tg}\omega_в\tau + N_0\gamma^2(\omega_в\tau - \operatorname{arc tg}\omega_в\tau)\operatorname{arc tg}\omega_в\alpha]}{\alpha\tau^3\left[\int\limits_0^{\omega_в}\sqrt{\frac{G_0 + N_0(m^2+\gamma^2\omega^2)}{(1+\tau^2\omega^2)(1+\alpha^2\omega^2)}}\,d\omega\right]^2}. \quad (8.12)$$

По формуле (8.12) находим:

1) При

$$\frac{N_0\gamma^2}{G_0\tau^2} = \frac{m^2 N_0}{G_0} + 1$$

$$\beta_2 = \frac{\omega_в\alpha\operatorname{arc tg}\omega_в\alpha}{\ln^2\left[\omega_в\alpha + \sqrt{1+\omega_в^2\alpha^2}\right]}. \quad (8.13)$$

2) Когда $N(\omega) = N_0\gamma^2\omega^2$ и $N_0\gamma^2/G_0\tau^2 \to \infty$, то

$$\beta_2 = \frac{\alpha(\omega_в\tau - \operatorname{arc tg}\omega_в)\operatorname{arc tg}\omega_в\alpha}{\tau\ln^2\left[\frac{\alpha\sqrt{1+\omega_в^2\tau^2} + \tau\sqrt{1+\omega_в^2\alpha^2}}{\alpha+\tau}\right]}. \quad (8.14)$$

3) Если $N(\omega) = N_0$, то

$$\beta_2 = \frac{\alpha\operatorname{arc tg}\omega_в\alpha\operatorname{arc tg}\omega_в\tau}{\tau F^2\left(\operatorname{arc tg}\omega_в\alpha;\sqrt{1-\tau^2/\alpha^2}\right)}. \quad (8.15)$$

Кривые зависимостей коэффициента β_2 от отношения α/τ, рассчитанные по формулам (8.13—8.15) при $\omega_в\tau = 12$, также представлены на рис. 5а, б. Из этого рисунка видно, что коэффициент β_2 может быть как бóльшим, так и меньшим коэффициента β_1.

9. Заключение

1. В работе проведен синтез оптимальных линейных предыскажающих и корректирующих фильтров в системе передачи информации с двумя дополнительными источниками шумов. При этом сформулирован и использовался критерий минимума дисперсии общей взвешенной ошибки, равной сумме динамической и случайной ошибок, взвешенных по разным законам (в спектральной области с помощью весовых функций $A(\omega)$ и $M(\omega)$ соответственно).

2. Критерии максимума отношения сигнал—шум и минимума среднего квадрата ошибки, ранее использовавшиеся в работах [1—12], а также ряд других критериев, являются частными случаями названного критерия верности воспроизведения сообщений. Причем весьма важным является то, что полностью отпадает необходимость каждый раз заново проводить синтез оптимальных предыскажающих и корректирующих фильтров по частным критериям. Достаточно лишь во все полученные соотношения подставить выбранные значения весовых функций $A(\omega)$ и $M(\omega)$, отражающие свойства конкретного получателя сообщений.

3. Проведен также сравнительный анализ эффективности применения оптимальных предыскажений сигналов и непосредственно самих сообщений (до их преобразования в сигнал). Показано, что при наличии дополнительных шумов предыскажения сообщений могут быть как более, так и менее эффективными, чем предыскажения сигналов (в зависимости от величины мощностей сигнала, шумов канала связи и дополнительных шумов).

4. В работе [13] указано, что в ряде случаев кодирование и декодирование в ситуации с дополнительным шумом сводится к последовательности оптимальной фильтрации зашумленного сообщения и оптимального кодирования и декодирования. Как показывает проведенный в приложении анализ, с точки зрения общего критерия верности воспроизведения сообщений, принятого в данной работе, описанный в [13] способ передачи информации не является оптимальным, если $L(\omega) \neq 0$ и $L(\omega) \neq 1$, но оптимален при $L(\omega) = 0$ и $L(\omega) = 1$.

5. Следует отметить, что полученные в данной работе выражения будут справедливы и для многомерных каналов связи, если в этих выражениях заменить одномерные спектральные плотности многомерными, а однократные интегралы — многократными.

Приложение

Об одном способе передачи информации

На рис. 6 представлена блок-схема одного способа передачи информации при наличии дополнительных шумов, который описан в [13]. Здесь передаваемый сигнал $x(t)$ сначала выделяется из шумов оптимальным фильтром

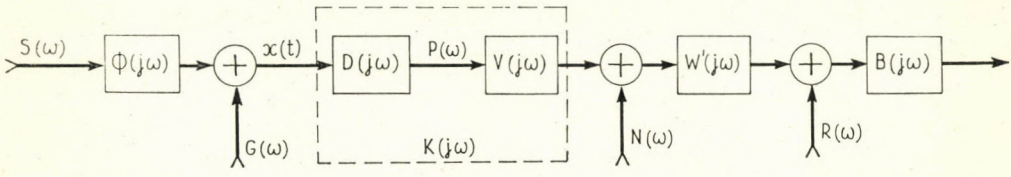


Рис. 6. Блок-схема одного из возможных способов передачи информации

с передаточной функцией $D(j\omega)$. Выделенный сигнал с энергетическим спектром $P(\omega)$ считается полезным и затем подвергается оптимальному линейному кодированию и декодированию уже без непосредственного учета дополнительных шумов и искажений передающего устройства. Передаточные функции кодирующего и декодирующего фильтра обозначены $V(j\omega)$ и $W'(j\omega)$ соответственно. Остальные обозначения имеют прежний смысл.

Выясним условия, при которых рассмотренный способ передачи информации является оптимальным в рамках принятого в данной работе общего критерия верности воспроизведения сообщений.

Из соотношений (4.7), (5.3), (5.4), (5.6) и рис. 6 следует, что при ограниченной средней мощности передатчика $Q = \frac{1}{2\pi}\int\limits_{\omega_н}^{\omega_в} P(\omega)\,d\omega$ оптимальные передаточные функции $D_0(j\omega)$, $W_0'(j\omega)$ и $V_0(j\omega)$ в последнем случае таковы:

$$D_0(j\omega) = \frac{\Phi^*(j\omega)\,S(\omega)}{|\Phi(j\omega|^2\,S(\omega) + G(\omega)\,L^2(\omega)}\,, \tag{П. 1}$$

$$W_0^1(j\omega)\,B(j\omega) = \frac{V^*(j\omega)\,P(\omega)}{|\,V(j\omega)|^2\,P(\omega) + N(\omega)\,L^2(\omega)}\,, \tag{П. 2}$$

$$P(\omega) = [S(\omega)\,|\,\Phi(j\omega)|^2 + G(\omega)]\,|\,D_0(j\omega)|^2\,, \tag{П. 3}$$

$$V_0(j\omega) = \begin{cases} \sqrt{\dfrac{N(\omega)\,M^2(\omega)}{\lambda_1\,P(\omega)}} - \dfrac{N(\omega)\,L^2(\omega)}{P(\omega)} & \text{при } h_1(\omega) \geq 1\,, \\ 0 & \text{при } h_1(\omega) < 1\,, \end{cases} \tag{П. 4}$$

где

$$h_1(\omega) = \frac{P(\omega)\,M^2(\omega)}{\lambda_1\,N(\omega)\,L^4(\omega)}\,;\quad \sqrt{\lambda_1} = \frac{\int\limits_{\Delta\omega_1} \sqrt{P(\omega)\,N(\omega)\,M^2(\omega)}\,d\omega}{\int\limits_{\Delta\omega_1} N(\omega)\,L^2(\omega)\,d\omega + \int\limits_{\omega_н}^{\omega_в} P(\omega)\,d\omega}\,.$$

Используя соотношения (П. 1—П. 4), получаем

$$|\,K(j\omega)|^2 = \begin{cases} \dfrac{\sqrt{\dfrac{S^2(\omega)\,N(\omega)\,M^2(\omega)\,|\,\Phi(j\omega)|^2}{\lambda_1\,[S(\omega)\,|\,\Phi(j\omega)|^2 + G(\omega)]}}}{|\,\Phi(l\omega)|^2\,S(\omega) + G(\omega)\,L^2(\omega)} - \dfrac{N(\omega)\,L^2(\omega)}{S(\omega)\,|\Phi(j\omega)|^2 + G(\omega)} & \text{при } h_1(\omega) \geq 1, \\ 0 & \text{при } h_1(\omega) < 1, \end{cases} \tag{П. 5}$$

$$W_0'(j\omega)\,B(j\omega) = \frac{K^*(j\omega)\,\Phi^*(j\omega)\,S(\omega)\,\dfrac{S(\omega)\,|\,\Phi(j\omega)\,|^2 + G(\omega)}{S(\omega)\,|\,\Phi(j\omega)\,|^2 + G(\omega)\,L^2(\omega)}}{[\,|\,\Phi(j\omega\,|^2\,S(\omega) + G(\omega)]\,|\,K(j\omega)\,|^2 + N(\omega)\,L^2(\omega)}\,, \tag{П. 6}$$

где $K(j\omega) = D_0(j\omega)\cdot V_0(j\omega)$.

При наличии дополнительных шумов ($G(\omega) \neq 0$) выражения (5.3) и (П.5), (4.7) и (П.6) совпадают только тогда, когда $L(\omega) = 0$ или $L(\omega) = 1$. Следовательно, исследуемый способ передачи информации (рис. 6) является оптимальным лишь при этих условиях (а также в случае, когда $G(\omega) = 0$).

Литература

1. *Costas, K. P.:* Coding with linear Systems. Proc. IRE *40*, 9, 1101—1103 (1952).
2. *Синай Я. Г.:* Наименьшая ошибка и наилучший способ передачи стационарных сообщений при линейном кодировании и декодировании в случае гауссовских каналов связи. Сб. «Проблемы передачи информации», 2, М., Изд. АН СССР 40—48 (1959).
3. *Овсеевич И. А.—Пинскер М. С.:* Оптимальное линейное предыскажение и корректирование. Изв. АН СССР, Техническая кибернетика 3, 54—61 (1963).
4. *Цыбаков Б. С.:* Линейное кодирование сообщений. Радиотехника и электроника, **7, 1**, 25—38 (1962).
5. *Цыбаков Б. С.:* Линейное кодирование изображений. Радиотехника и электроника, **7, 3**, 375—385 (1962).
6. *Цыбаков Б. С.:* Вопросы линейного кодирования движущихся изображений. Сб. «Кибернетику — на службу коммунизму», 3, М., «Энергия» 134—148 (1966).
7. *Berkowitz, M. O.:* Optimum linear shaping and Filtering networks. Proc. IRE, 41, 4, 532—537 (1953).
8. *Штейн В. М.:* О расчете предыскажающих и корректирующих устройств. Радиотехника, **11, 2**, 60—63 (1956).
9. *Franks, L. E.:* A model for the Random Video Process. Bell Syst. Techn. J. **45,** 4, 609—630 (1966).
10. *Лебедев Д. С.:* Линейные двумерные преобразования изображений, увеличивающие помехоустойчивость передачи. Сб. «Иконика». I. М., «Наука», 1968, 15—27.
11. *Миленин Н. К.:* Об эффективности применения предыскажающих устройств в телевизионных системах. Труды РРТИ, 17, Рязань, 1969, 59—66.
12. *Раковский В. В.:* Частотные характеристики, сигнал—шум и нелинейные искажения фотографических фонограмм. Техника кино и телевидения **11**, 31—36 (1970).
13. *Добрушин Р. Л.—Цыбаков Б. С.:* Передача информации с дополнительным шумом. Сб. «Проблемы передачи информации», 14. М., Изд. АН СССР 21—42 (1963).
14. *Харкевич А. А.:* Борьба с помехами. М., «Наука», 1965.

Optimal linear predistorting and correcting in the information transmission system with additional noise

N. K. MILENIN

(Riazan')

Summary

The synthesis of the optimal linear predistorting and correcting filters is carried out in the information transmission system with two additional noise sources. The criterion of the minimun weighing dispersion error is offered to use. This error equals the sum of dynamic and random errors, which have been weighed on different laws. The comparative estimation of the efficiency of preemphasis of the signals and direct information itself carried out.

Н. К. Миленин
Рязанский радиотехнический институт
СССР, Рязань, ул. Гагарина, 59/1

Problems of Control and Information Theory, Vol. 1 (3—4), pp. 325—336 (1972)

A SEQUENTIAL TREATMENT OF INFORMATION TRANSMISSION

P. WHITTLE

(Cambridge)

(Received March 1, 1972)

A situation is considered in which the sender of a message over a noisy channel is aware of what signals have actually been received at the other end. The sender and receiver agree sequentially on a coding which is optimal in that it minimizes the expected number of signals required to bring the receiver to a prescribed degree of certainty as to what message is intended. This process is termed *interrogation* of sender by receiver, and is a situation occurring in sequential experimentation. The solution to this sequential optimisation problem yields automatically the concepts of information measure and of channel capacity, also a certain form of the coding theorems, and a sequential construction for the optimal coding.

1. Introduction

The conventional information transmission situation is the following. One wishes to send a number of messages along a channel, which is in general noisy. In order to exploit the increased possibility of efficient coding inherent in a long message, one groups a long sequence of basic messages to form a compound message. A block of signals is then sent which is long enough and informative enough to enable the receiver to identify all but the relatively improbable compound messages with an assigned small probability of error, but so encoded that the number of signals in the block in near-minimal.

Roughly, the basic theorems of information theory state that the number of binary signals required to achieve this aim is, with efficient encoding,

$$N = H/C + \Delta. \tag{1}$$

Here H is the entropy of the compound message source, C the ergodic capacity of the channel, and Δ a term which becomes small relative to H/C as the length of the compound message increases. The entropy H is defined by

$$H = H(P) = -\Sigma P_i \log P_i, \tag{2}$$

where i indexes the possible compound messages, and P_i is the probability of message i. All logarithms are to base 2, unless the contrary is indicated. We shall consider the definition of C in Sections 3, 4.

A rigorous proof of assertion (1) is a substantial and lengthy matter. Proofs have been improved and shortened over the years (Shannon's random coding device (1957) being especially elegant and effective) but they must incorporate careful treatment of questions of ergodicity and stochastic convergence. Moreover, the key quantities, H and C, do not emerge from these treatments in a particularly inevitable manner, although inevitable they must be.

Consider now the following sequential modification of the original transmission problem. After sending an individual signal, the sender is able to observe what signal has in fact been received at the other end, and so how the probabilities P_i that the receiver attaches to the various compound messages, conditional on information available to him, have changed. In the light of this, the sender chooses his next signal optimally (so as to minimize the expected number of signals needed). Transmission ceases when the receiver has identified the intended message with an assigned level of probability. The number of signals N will then be a random variable. However, what we shall prove, by fairly direct means, is that, corresponding to (1).

$$E(N) = H/C + \Delta', \tag{3}$$

where H and C have the same definitions as before for the case of a memoryless channel, (that of C changes if the channel has memory: see Section 5), and the remainder Δ' the same character as Δ.

As an example of information transmission following these rules, consider a scientist carrying out a sequence of experiments so as to determine which of a number of hypotheses $\mathcal{H}_i$ may be true. The hypotheses are supposed exhaustive and mutually exclusive, and experimentation continues until the Bayes probability of any one hypothesis exceeds a prescribed value, $1 - \lambda$. At each stage the investigator chooses his experiment in such a way as to minimize the expected number of experiments needed, conditional on the available information. The choice of experiment then corresponds to choice of signal.

From this point of view the situation appears as one of *interrogation* rather than of transmission. The scientist arranges for the experiment (or the receiver calls for the signal from the sender, with a mutually understood coding convention), which, conditional on his current knowledge, minimizes the expected future number of experiments (signals) needed to reach a state of sufficient conclusiveness.

Communication theorists have also considered this situation, under the name "signal transmission with noiseless feedback" (see Shannon (1956), Elias (1961), Horstein (1963), Schalkwijk (1968), Zigangirov (1968)). However, they

have not used the sequential decision approach adopted here, in which the current posterior distribution over messages P is regarded as a state variable, with an associated expected signal-length $F(P)$ which obeys a dynamic programming (D.P.) equation, this equation determining $F(P)$ and the optimal coding simultaneously. The advantage of this approach is that the definitions and roles of H and C, information measure and channel capacity, fall out immediately, in that we find that $F(P) = H(P)/C$ is approximately a solution of the D.P. equation (this is assertion (3)). So, rather than appearing as constructions, they crystallise inevitably from the D.P. formulation of the problem. The optimal coding, or at least an approximation to it, also falls out. We shall not be able to avoid questions of ergodicity altogether, but the sequential approach, yielding an assertion such as (3), requires less in the way of convergence arguments than the non-sequential approach associated with (1).

D. P. methods have been used before for optimal coding, but in other senses. For example, Huffman (1962) derived an optimal encoding procedure for noiseless channels, by D. P. methods. However, this is an optimisation of a deterministic mapping by recursive methods, rather than a real-time sequential procedure, and it will not deal with the case of noisy channels.

In order to separate the basic ideas we shall treat three cases of increasing generality: perfect channels, noisy channels without memory, and noisy channels with memory. That is, in the context of experimentation: error-free experiments, experiments with error which are statistically independent, and sequences of statistically dependent error-laden experiments.

2. Interrogation over a perfect channel

Consider a situation in which each signal can take one of m values, and is received faithfully. That is, one is working with an m-letter alphabet through a perfect channel.

We shall use P_i to denote the receivers posterior probability (i.e. probability conditional on the available evidence) that the compound message being sent is the i^{th}. The vector with elements P_i will be denoted by P. The value of P changes as transmission proceeds. In the case of the perfect channel one can consider continuing until the message has been identified with certainty, so the stopping rule is "cease interrogation when the distribution P is degenerate — i.e. concentrated on a single i-value".

The coding or interrogation rule will take the form of a direction from the receiver: "transmit signal $x^i(P)$ if $\mathcal{R}_i$ is true". Here x takes one of the signal values 1, 2, . . ., m and we have included the P argument to emphasise

the fact that the coding the receiver, or interrogator, will choose at any stage will depend upon the current P value. The problem is then one of choosing the function $x^i(P)$ optimally, and we shall take as "optimal" that coding rule which minimises the expected number of signals required to reach certainty.

From the experimental point of view, one has an experiment which is capable of m responses, or outcomes, and the experiment is so designed that it will have outcome $x^i(P)$ if $\mathcal{H}_i$ is true: P being the vector of posterior probabilities of the respective hypothesis at the stage immediately before this particular experiment.

We shall say that P describes the "state" of the interrogator, or experimenter.

Let $F(P)$ denote the expected number of signals needed, starting from a state P, if optimal coding is used. We have then

$$F(P) = 0. \qquad (P \text{ degenerate}) \tag{4}$$

If P is not degenerate then at least one more signal must be transmitted (experiment performed). Signal x will be received with probability

$$\alpha(x) = \sum_{A(x)} P_i, \tag{5}$$

Here $A(x)$ is the set of i for which $x^i(P) = x$ (so that both $\alpha(x)$ and $A(x)$ are P-dependent, but we shall leave this dependence to be understood, for notational simplicity). After the observation of signal x, the posterior probabilities P_i will suffer the transformation

$$P_i \to P_i^* = \begin{cases} 0, & (i \notin A(x)) \\ \dfrac{P_i}{\alpha(x)}, & (i \in A(x)), \end{cases} \tag{6}$$

o that $F(P)$ then suffers the transformation

$$F(P) = A \min \left[\sum_x \alpha(x) F(P^*) \right]. \tag{7}$$

Here the minimum is taken over the possible decompositions $\{A(x)\}$ of the set of i values into disjoint, exhaustive sets $A(x)$; alternatively over choices of the function x^i. Because the coding is optimal at each stage, $F(P)$ will satisfy the dynamic programming equation (7). This equation can be simplified in one respect. If we write the unit in the righthand member of (7) as ΣP_i, then one establishes recursively that $F(P)$ can consistently be extended to non-negative arguments P, not necessarily satisfying the normalisation

$$\Sigma P_i = 1, \tag{8}$$

by the convention that $F(P)$ be taken as homogenous of degree of one in P (see Whittle, 1964, 1965, 1969). For this extended function the recursion (7) simplifies to

$$F(P) = \Sigma P_i + \min_x [\sum F(\delta(x)P)], \tag{9}$$

where the transformed vector $\delta(x)\,P$ has its i^{th} argument defined by

$$(\delta(x)\,P)_i = \begin{cases} 0, & (i \notin A(x)) \\ P_i, & (i \in A(x)). \end{cases} \tag{10}$$

Equations (4) and (9) determine F; the minimisation in (9) simultaneously determines the optimal decomposition $\{A(x)\}$ i.e., the optimal choice of the function $x^i(P)$.

Theorem 1. The solution $F(P)$ of (4), (9) satisfies

$$F(P) = \frac{H(P)}{\log m} + \Delta, \tag{11}$$

where

$$H(P) = -\Sigma P_i \log P_i + (\Sigma P_i) \log (\Sigma P_i), \tag{12}$$

and Δ is a non-negative quantity, which is zero iff the decomposition $\{A(x)\}$ can be chosen at all stages so that $\alpha(1) = \alpha(2) = \ldots = \alpha(m)$

In (12) we have given an extended definition of the Shannon entropy function, which is homogeneous of degree one in P and which reduces, when the normalisation condition (8) is satisfied, to the previous definition (2). Thus (11) provides a statement that $E(N) \sim H/C$, with

$$C = \log_2 m \tag{13}$$

the capacity (in bits per signal) of a perfect m-letter channel.

The lower bound H/C is attained when all signal values (experimental responses) can be made equally probable at each stage: the optimal coding is presumably that bringing one as near as possible to equi-probability in some sense. The smaller the largest of the P; the nearer one can come to this goal. The remainder therefore represents a "discreteness" effect. We know the exact conditions under which Δ is zero; what is also required is a determination of conditions on P which ensure that Δ is of smaller order than H/C. This is the stage at which ergodicity enters in this treatment. If the elements of P are sufficiently numerous and uniform and if, in particular, they represent the realisation probabilities of a long segment of an ergodic process, then the point can presumably be proved, at least with a sufficiently careful treatment of the code optimisation. We shall leave this point to a later paper.

The proof goes by induction. We know from (4) that (11) holds (with $\Delta = 0$) for one-point distributions. We shall assume it true for distributions on less than r points and deduce its validity for r-point distribution from(9). Note that, if P is an r-point distribution in (9), then all the $\delta(x)P/\alpha(x)$ are distributions on less than r points.

Suppose then that (11) holds for distributions on less than r points. We see from (9) that

$$F(P) \geq H(P)/C + \Sigma P_i - (\Sigma P_i) \log_m (\Sigma P_i) + \min_A [\sum_x \alpha(x) \log_m \alpha(x)], \quad (14)$$

equality holding of $\Delta = 0$ for distributions on less than r points. The minimum in (14) is over decompositions $\{A(x)\}$ We shall have

$$\sum_x \alpha(x) = \Sigma P_i = S, \quad (15)$$

say. Suppose we replace the minimisation by the freer one of minimising with respect to the $\alpha's$, subject only to $\alpha \geq 0$ and (15). If the P_i are numerous and small enough, the two procedures are approximately equivalent. The free minimum is achieved when all the $\alpha(x)$ are equal, when

$$\min_A [\sum_x \alpha(x) \log_m \alpha(x)] \geq \min_\alpha [\sum_x \alpha(x) \log_m \alpha(x)] = S \log_m S - S. \quad (16)$$

Substituting (16) into (14), we see that we have extended (11) to r point distributions, with the condition for $\Delta = 0$ plainly being that given in the theorem.

3. Interrogation over a noisy memoryless channel

Suppose that the sending of signal x may produce signal y at the receiving end with probability $P(y|x)$, independently of signals previously sent or received. If signal y is in fact received, then transformation (6) now becomes

$$P_i \to P_i^* = \frac{P_i P(y|x^i)}{\sum_i P_i P(y|x^i)}, \quad (17)$$

where x^i is the signal sent at this stage if $\mathcal{H}_i$ is true (we suppress the P-dependence for the moment). Correspondingly, the recursion (9) now becomes

$$F(P) = \Sigma P_i + \min_A \sum_y F(\{P_i P(y|x^i)\}), \quad (18)$$

where we use the notation $F\{P_i\}$ instead of $F(P)$ if we wish to write out the i^{th} element of P explicity.

In the case of the perfect channel, interrogation continued until the compound message i had been identified with certainty. This state will in general be unattainable in the present case, and one must give thought to the formulation of an explicit stopping rule. The usual rule is: stop when the message has been identified with probability greater than $1 - \lambda$, where λ is a prescribed small quantity. The stopping region in P space is then $\max_i P_i \geq \geq 1 - \lambda$, or

$$\max_i P_i \geq (1 - \lambda) \Sigma P_i \tag{19}$$

if the P-distribution is unnormalised. The point of the following lemma (due to $F_{an\,0}$) lies in the second sentence.

Lemma. On the surface

$$\max_i P_i = (1 - \lambda) \Sigma P_i \tag{20}$$

in P-space we have

$$- \Sigma P_i \log P_i = [- \lambda \log \lambda + 0(\lambda)] \Sigma P_i. \tag{21}$$

Thus, for λ small and a normalised P-distribution the stopping rules $\max P_i \geq 1 - \lambda$ and $H(P) \leq - \lambda \log \lambda$ are effectively equivalent.

The lemma follows easily from the fact that, for a given values of $\max P_i$ (distribution normalised) $H(P)$ attains its extreme values when the remaining P_i values are either equal, or concentrated on a single second value of i.

Suppose we modify the stopping rule, then, to

$$H(P) \leq \delta, \tag{22}$$

where small δ corresponds to small λ. For non-normalised P distributions the rule will be

$$H(P) \leq \delta \Sigma P_i . \tag{23}$$

Theorem 2. Let $F(P)$ denote the minimal expected number of signals needed, starting from P, until the receivers posterior distribution $\{P_i\}$ enters the stopping region (23). Then

$$F(P) = H/C + (\Delta_1 + \Delta_2) (\Sigma P_i) , \tag{24}$$

where H is defined by (12), and C a constant defined by

$$C = \max_p \sum_x \sum_y p(x) P(y \mid x) \log \left[\frac{P(y \mid x)}{\sum_x p(x) P(y \mid x)} \right], \tag{25}$$

the maximum being taken over distributions $p(x)$ on the set of signal values which can be transmitted. The term Δ_1 is non-negative, and represents a P_i-discreteness effect in that it is zero if at each stage the coding x^i can be chosen so that

$$\sum_{A(x)} P_i = p(x) \sum P_i, \tag{26}$$

where $A(x)$ is the set of i for which $x^i = x$, i.e. for which x is transmitted, and $p(x)$ is the maximising distribution determined in (25). The term Δ_2 is bounded by δ/C.

Thus, the entropy H again makes a natural appearance, as does the channel capacity C, whose definition (25) coincides with the Shannon definition for a memoryless channel. Relation (26) embodies the optimal coding rule, or, rather, the ideal rule to which the optimal rule will approximate as closely as possible.

Let the termination region in P space determined by (23) be denoted by D, and its complement by $\bar{D}$. Then we wish to solve (18) in $\bar{D}$, subject to $F(P) = 0$ in D.

The proof of the theorem depends on the fact that $F(P) = H(P)/C$ is a solution of (18) if the signal sets $A(x)$ can be chosen by prescription (26). For, we find that

$$\sum P_i + \min_A \sum_y H\left(\{P_i P(y \mid x^i)\}\right)/C - H(P)/C = \tag{27}$$

$$= \sum P_i - \frac{1}{C} \max_A \left[\sum_x \sum_y \alpha(x)\, P(y \mid x) \log \left[\frac{P(y \mid x)}{\sum_x \alpha(x)\, P(y \mid x)} \right] + (\sum P_i) \log (\sum P_i) \right],$$

where $\alpha(x)$ is defined by (4). Setting

$$\alpha(x) = \varrho(x) \sum P_i, \tag{28}$$

so that $\varrho(x)$ is a distribution on the signal set, and replacing the A maximisation in (27) by the the freer maximisation with respect to the distribution $\varrho(x)$, we see that expression (27) exceeds

$$\sum P_i - \frac{1}{C} \max_\varrho \left[\sum_x \sum_y \varrho(x)\, P(y \mid x) \log \left[\frac{P(y \mid x)}{\sum_x \varrho(x) P(y \mid x)} \right] \right] (\sum P_i) = 0, \tag{29}$$

the excess being zero if the A-maximisation can achieve the same upper bound as the freer p-maximisation. Thus, $F = H/C$ is a solution of (18), apart from discreteness effects. What we have now to show is that it at least approximates the solution appropriate to the boundary condition, $F = 0$ in D.

The solution we have obtained is undoubtedly that appropriate to the terminal condition, $F = H/C$ in D, discreteness effects apart. We are thus giving F a terminal value which is incorrect by an amount H/C, or at most $(\delta/c)(\Sigma P_i)$. This will affect F for all P by a term with the same bound, which accounts for the presence of the term Λ_2 in (24).

4. Interrogation over a general non-anticipating channel

If a signal sequence X_t up to time t has been sent, and a signal sequence Y_{t-1} received up to time $t-1$, then we can specify the conditional probability $P(y_t/X_t, Y_{t-1})$ of the signal y_t received at t from a knowledge of the channel's statistical characteristics.

Let P_i denote, as ever, the receiver's posterior probability that the compound message i is intended. Then, after reception of y_t, this will be modified to

$$P_i^* = \frac{P_i P(y_t \mid X_t^i, Y_{t-1})}{\sum P_i P(y_t \mid X_t^i, Y_{t-1})} = \frac{P_i P(y_t \mid X_t^i, Y_{t-1})}{\sum_{x_t} \alpha(X_t) P(y_t \mid X_t, Y_{t-1})}. \tag{30}$$

Here X_t^i is the sequence which will have been transmitted, under the agreed coding, if message i is intended, and $\alpha(X_t)$ is the sum of P_i over all i such that $X_t^i = X_t$: the probability (unnormalised) that some i is intended for which the sequence X_t would be transmitted.

In this situation with memory the minimal expected further number of signals required from time t will be a function of Y_t as well as of P, $F(P, Y_t)$, because the effect of past observations upon future ones will in general not be adequately reflected by the changing P distribution alone. If interrogation does not terminate at time $t-1$ then F will obey the recursion

$$F(P, Y_{t-1}) = \sum P_i + \min_{x_t^i} \Im F, \tag{31}$$

where

$$\Im F = E[F(P^*, Y_t) \mid Y_{t-1}] = \sum_{I_t} \sum_i P; P(y_t \mid X_t^i, Y_{t-1}) F(P^*, Y_t). \tag{32}$$

The minimisation in (31) is over the t^{th} stage of coding: over choice of the signal x_t^i which will be sont at time t if message i is intended.

Theorem 3. Suppose that the loss function F is written in the form

$$F(P, Y_t) = H(P)/C + \mu(Y_t) \sum P_i, \tag{33}$$

where H is defined by (12), and C is a constant, to be determined. Then μ obeys a

recursion

$$\mu(Y_{t-1}) = 1 + \alpha_t + \min_{P(x_t \mid X_{t-1}, Y_{t-1})} E[\mu(Y_t) - \xi_t/C \mid Y_{t-1}], \tag{34}$$

where

$$\xi_t = \log \frac{P(y_t \mid X_t, Y_{t-1})}{P(y_t \mid Y_{t-1})}. \tag{35}$$

The non-negative term α_t represents a discreteness effect; in the P_i — distribution: it would be zero if the coding x_t could be chosen such that

$$\alpha(X_t)/\alpha(X_{t-1}) = P(x_t \mid X_{t-1}, Y_{t-1}) \tag{36}$$

for all X_t, where $P(x_t/X_{t-1}, Y_{t-1})$ is the minimising value in (39). Such a coding would be optimal.

Note that, to write F in the form (38) implies no loss of generality, since F is a function only of Y_t and of the fixed initial value of the distribution P. The conditional expectation in (39) is to be calculated on the basis that

$$P(X_n, Y_n) = \prod_{r \le n} P(x_t \mid X_{t-1}, Y_{t-1})\, P(y_t \mid X_t, Y_{t-1}), \tag{37}$$

the final factor in (31) being determined by channel characteristics, the one before it by choice of coding.

Recursion (34) is established very much on the same lines as before, and we shall omit the detailed proof. From it we deduce

Theorem 4. Suppose we adopt the termination region (23) Then the minimal expected number of signals required from time t is

$$F(P, Y_t) = H(P)/C + \frac{F}{N} E\left[N - \frac{1}{C} \sum_{t+1}^{t+N} \xi_s \mid Y_t\right] + (\Delta_1 + \Delta_2)(\Sigma P_i). \tag{38}$$

Here $t + N$ is the instant at which interrogation terminates, Δ_1 is a non-negative term corresponding to discreteness effects in the P_i — distribution, and Δ_2 is a term of order.

The value of the constant C has not yet been assigned, we shall attempt to choose it so that the squarebracketed term in (38) is of lower order than $H(P)/C$, by setting

$$C = \lim_{N \to \infty} \frac{1}{N} \log \left[\prod_{t+1}^{t+N} \xi_S\right]. \tag{39}$$

With a little rewriting one sees that expression (39) agrees with the definition of capacity for a channel with feedback given by Shannon (1956). One has, of course, still to show that, under suitable conditions, the limit (39) exists, and that, with this definition of C, the term $H(P)/C$ is the term of dominant order

in expression (38). These are the points at which the ergodic properties of source and channel enter rather more critically than in the memoryless case. The sequential approach cannot entirely evade these issues, but it does, as we have shown, generate the notions of information and capacity measures in a relatively direct fashion.

I am grateful to Professor P. Elias for very helpful comments on an earlier draft of this paper.

References

Elias, P. (1961), "Channel Capacity without Coding" in *Lectures on Communication System Theory*, Baghdady (ed.) McGwar Hill, N. Y.

Gallager, R. (1965), "A Simple Derivations of the Coding Theorem and some Applications", *IEEE Transactions on Information Theory, IT*-11, pp. 3—18.

Gallager, R. (1968), *Information Theory and Reliable Communication*, John Wiley & Son, N. Y.

Horstein, M. (1963), "Sequential Transmission Using Noiseless Feedback", *IEEE Transactions on Information Theory, IT*-9, No. 3. pp. 136—143.

Huffman, (1962), A method for the construction of minimum redundancy codes, *Proc IRE*, 40, 1098—1101.

Rényi, A. (1965), On the foundations of information theory, *Rev. Int. Statist. Inst.* 33, 1—14.

Schalkwijk, R. P. M. (1968), "Center of Gravity Information Feedback", *IEEE Transactions on Information Theory, IT*-14, pp. 324—331.

Shannon, C. (1948), *The Mathematical Theory of Communication*, U. of Illinois Press, Urbana.

Shannon, C. (1965), "The Zero Error Capacity of a Noisy Channel", *IEEE Transactions on Information Theory, IT*-3, No. 3, pp. 8—19.

Shannon, C.(1957), „Certain Results in Coding Theory for Noisy Channels", *Information and Control*, 1, pp. 6—25.

Whittle, P. (1964), Some general results in sequential analysis *Biometrika*, 51, 123—141.

Whittle, P. (1965), Some general results in sequential design. *J. R. Statist. Soc. B.*, 27, 371—387.

Whittle, P. (1969), Sequential decision process with essential unobservables *Adv. Appl. Prob.* 1, 271—287.

Zigangirov, K. (1968) articles in issue number 3 of *Problems of Information Transmission.*

Последовательная обработка передачи информации

П. УАЙТТЛ

(Кембридж)

Резюме

Рассматривается случай передачи сообщений по зашумленному каналу, при котором на передающем конце неизвестно, какой сигнал фактически будет принят. Передатчик и приемник последовательно подвергаются кодированию, которое является оптимальным, в том смысле, что минимизирует среднее число сигналов, требующихся для приема нужного сообщения с заданной достоверностью. Этот процесс называется переспросом передатчика со стороны приемника и соответствует ситуации, возникающей при последовательном экспериментировании. Решение рассматриваемой задачи последовательной оптимизации автоматически связано с понятиями меры информации и пропускной способности канала, а также с определенной формой теорем кодирования и последовательным конструированием оптимального кодирования.

Prof. P. Whittle

Faculty of Mathematics

University of Cambridge

England

Problems of Control and Information Theory, Vol. 1 (3—4), pp. 337—344 (1972)

ON THE SECOND INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY

The Second International Symposium on Information Theory was held at Tsakhadsor, Armenia, September 2—8, 1971. It was arranged by the Council for Cybernetics, the Institute for Problems of Information Transmission (both of the USSR Academy of Science) and the Computing Center of the Armenian Academy of Sciences in cooperation with the International Union of Radio Science (URSI).

This Symposium, as well as the first one held at Dubna, USSR, in June 1959, was devoted to mathematical problems of information theory and its application. The program included general topics in information theory, probabilistic and algebraic coding techniques, statistical methods in information theory, communication channel studies and quantum aspects of information theory.

Scientists from Bulgaria, Canada, Czechoslovakia, Finland, German Federal Republic, Hungary, Iran, Italy, Japan, Poland, Romania, Sweden, Switzerland, USA and USSR participated.

Plenary meetings were held at the opening and closing sessions. Technical sessions on information theory, algebraic codes, probabilistic coding, communication system, quantum channels, statistical methods, feedback channels, random processes, source encoding, error correcting codes and communication channels made up the rest of the program. The complete list of the papers presented at the Symposium is given in the Appendix.

We do not enter here into the details of the presented results, just briefly review the topics of the meeting. In so doing we point out some of the papers to give a general idea of what kind of problems have been treated.

General methods of information theory. This area is represented by a large group of papers on "classical" information theory. In the papers by E. Posner (USA), J. Csiszár (Hungary) and J. Kulikowski (Poland) various approaches to entropy and information definitions are investigated. A series of papers consider proofs of coding theorems for various kinds of channels. Thus,

the paper by T. Kadota and A. Wyner (USA), delivered by T. Kailath (USA), presents a proof of Shannon's theorem for stationary, asymptotically memoryless, continuous channels. The paper by R. Ahlswede (USA) investigates the capacity of multy-way communication channels.

These particular papers as well as the others show continuous progress in the development of Shannon's methods in Information Theory.

Coding theory and complexity estimates of encoders and decoders. A number of results concerning decoding error probability estimation, code distance estimation and sequential decoding have been presented. Papers on encoding and decoding schemes for practical implementation appeared to be of particular interest. Some results recently obtained show that near-optima decoders can be implemented, the complexity of which (measured by the number of the logic elements involved) increases only slightly faster than the blocklength. These results are presented in the paper by S. Gelfand, R. Dobrushin and M. Pinsker (USSR) "Asymptotically optimal encoding using simple schemes". The paper by J. Savage (USA) "The complexity of deterministic source encoding with a fidelity criterion" proves the existence of easily implementable encoding schemes. Some specific decoding schemes and according complexity estimates are given by V. Zyablov and M. Pinsker (USSR), J. Stiffler (USA) and Y. Iwadare (Japan). It is expected that the techniques of implementing encoding and decoding processes will further develop in the neare future. Among other results on probabilistic encoding one should mention the improvement of the lower reliability bound, obtained by A. Sheverdiaev (USSR) for low-rate communication over a memoryless Gaussian channel.

Feedback systems. A survey of some work done by scientists in the USA on feedback communication systems was given by G. Turin (USA). His paper summarizes a number of suggestions for designing communication systems for fast as well as slow fading channels. P. Shalkwijk (USA) introduced an original algorithm for communication over memoryless binary symmetric channels, using a feedback. More accurate upper and lower bounds, to error probability for communication over a discrete, memoryless and asymmetric channel are given in the paper by E. Arutyunian (USSR).

Feedback system are really at the first stage of their study and this branch of Information Theory offers a vast field of research for designing transmission algorithms and estimating their efficiency. One should also notice the progress achieved by American scientists in the field of feedback systems operating over real communication channels.

Algebraic methods of encoding. A significant part of the papers are devoted to problems in this field. E. Berlekamp (USA) and Y. Jönsson (Sweden) improve estimates for code distances of BCH codes. Results on further investigations of

Reed—Muller codes are reported by T. Kasami (Japan) and S. Berman (USSR). E. Weldon presents methods of implementing error-correcting codes on a computer. In an interesting paper by V. Goppa (USSR) is proposing a new class of algebraic codes.

The discussion of papers presented at the Section on Algebraic Codes well illustrated how present trends in error correcting codes are more and more defined by requirements of computing technology.

Source encoding. Though the basic theorem on source encoding has been formulated by Shannon together with the coding theorem for noisy channels, the problems of data compression and epsilon-approximation of information sources are again attracting mathematicians as well as engineers.

This can be explained by the discrepancy arising in these days between the relatively low capacity of some of the channels and the growing amount of information to be transmitted through these channels. A lively discussion was provoked by papers read by T. Nemetz (Hungary), G. Longo (Italy), E. Posner (USA), L. Davisson (USA), P. Shalkwijk (USA) and others.

P. Stucki (Switzerland) was concerned with computer simulations of different digital picture compression techniques.

Lack of accurate statistical data on real information source is known to be one of the major difficulties arising in source encoding. Therefore the paper by Yu. Shtar'kov and V. Babkin (USSR) presenting a method for encoding sources with unknown statistics and also other papers dealing with problems of this sort were of a great interest.

Statistical methods and learning systems. A series of new results in the theory of random processes were presented by T. Kailath (USA), A. Yaglom (USSR), A. Shiryaev (USSR) and others. A. Perez (Czechoslovakia) gave information-theoretical bounds on the error probability for hypothesis testing in Markovian processes.

Studies of learning processes under a finite memory constraint are of interest in mathematical statistics as well as in learning systems. T. Cover and M. Hellman (USA) showed in their review that constraining the memory results is an exponentially decaying probability of error. By using randomized algorithms fairly good results can be achieved is testing between two hypotheses, within a relatively small amount of memory.

S. Csibi (Hungary) considered a class of sequential learning algorithms which assymptotically tend to potential-function-type procedures. Various approaches to the problem of learning algorithms have been treated by O. Gulyás, L. Győrfi, G. Katona an L. Molnár (Hungary).

A new theoretical approach to modeling communication channels with memory and some experimental results were presented by L. Kanal (USA).

H. Ohnsorge (FRG) described new result achieved in realizing high capacity information media (fiberglass guides and so like).

The Section on Quantum Channels organized by R. Kazaian and other scientists of the Erevan Research Institute of Physics has been particularly active. R. Kennedy (USA), R. Ingarden (Poland) and also scientists from Finland, USSR and other countries contributed to this section. A series of interesting papers have been read on the use of quantum channels for data transmission and the application of laser technology for sounding of the upper atmosphere.

As a matter of fact, three main trends may be distinguished among the papers presented at this Symposium.

First, there are the papers concerned with traditional information-theoretic problems: proving coding theorems, estimating channel capacity, investigating new encoding and decoding techniques, etc.

The second kind of papers use and develop information-theoretical ideas in statistics, learning systems theory, random process theory and quantum physics.

The third direction consists of such papers on applied aspects of communications theory, signal detection and computing technology which do not use directly information-theoretical approaches, but suggest new ideas, and broaden in this way the methods of Information Theory.

All these papers indicate that Information Theory might have a growing influence on communication technology, control theory, computing technology and information retrieval systems in the near future.

The Symposia at Dubna and Tsakhadsor had confirmed the timeliness off the treated subjects. There was a highly representative and competent attendance at both of these meetings.

It was generally suggested to continue this series of Symposia also in the future. Accordingly the Council for Cybernetics and the Institute for Problems of Information Transmission schedule to arrange a next (third) Symposium in Summer (1973).

The Proceedings of the Second International Symposium on Information Theory appears now, in the form of this book, as a special supplement to the 1972 Volume of the "Problems of Control and Information Theory".

M. S. Pinsker, V. N. Koshelev

Appendix

List of papers

Adomian, G. (USA), Signal processing in a randomly time-varying system.
Ahlswede, R. (USA), Multi-way communication channels.
Ahlswede, R., Wolfowitz, G. (USA), Channels without synchronization.
Akaike, H. (Japan), Information theory and an extension of the maximum likelihood principle.
Aripov, M. (USSR), Evaluation of request system transmission rate decrease due to channel multiplexing.
Arutyunian, E. (USSR), Error probability lower bound for data transmission over channels with a feedback.
Berman, S. (USSR), Geometry of least-weight elements in second order Reed–Muller codes.
Benedetto, S., Biglieri, E. (Italy), A spectral analysis technique for digital nonstationary random processes.
Berlekamp, E. (USA), Long primitive binary BCH codes have distance $d \sim \frac{2n \ln R^{-1}}{\log n} \ldots$
Blokh, E., Popov, O., Turin, V. (USSR), On evaluation of probability distributions of error patterns in a fixed-length sequence.
Borodin, L. (USSR), Resolution of composite signals in correlated Gaussian noise.
Boyarinov, I. (USSR), Linear codes capable of correcting limited density error bursts.
Breusov, V. (USSR), Multiconvolutional and blockmulticonvolutional error-protecting encoding.
Burnashev, M. (USSR), Block technique for weak signal transmission over a memoryless channel.
Campbell, L. (Canada), Characterization of entropy of probability distribution on the real line.
Conte, E., Corti, E., Esposito, R., Pescatory, L. (Italy), Error probabilities due to a mixture of impulsive and Gaussian noise in digital communication.
Csibi, S. (Hungary), Learning optimal decision functions recursively.
Csiszár, I. (Hungary), A class of informativity measures for observation channels.
Davisson, L. (USA), Error threshold crossings for adaptive algorithms.
Deryugin, I., Vishnevsky, A., Kurashov, V. (USSR), On potential accuracy of phase measurements in quantum channels.
Győrfi, L. (Hungary), On convergence of potential-function-type learning algorithms.
Dolainsky, F., Dorsch, B. (G.F.R.), Transmission limits for the Gaussian channel with finite rate codes.
Dorsch, B. (G.F.R.), Optimum quantization for the Gaussian channel.
Dyachkov, A., Pinsker, M. (USSR), On block transmission over a discrete Gaussian channel with a feedback.
Dyn'kin, V., Kimmelfeld, B. (USSR), Designing nonbinary arithmetic codes capable of correcting single errors.
Endres, W. (G.F.R.), A comparison of the redundancy in the written and spoken language.
Ene, D., Györfi, E. (Romania), On Hamming bounds for random and burst error detecting and correcting codes.
Fitingoff, B. (USSR), On the encoding for a class of stationary sources.
Gabidulin, E. (USSR), Combinatorial metrics in coding theory.
Gelfand, S., Dobrushin, R., Pinsker, M. (USSR), Asymptotically optimal encoding using simple schemes.
Gindikin, S. (USSR), On a class of problems of probabilistic logics.
Goppa, V. (USSR), Lengthened binary (L, g)-codes.
Guarguaglini, P., Marcoz, F., Minguzzi, B. (Italy), A queuing problem for an intermittent data transmission system.
Gulyas, O. (Hungary), On generalization and convergence rate of potential-function-type learning algorithms.
Halme, S. (Finland), Optical communications in a turbulent atmosphere.

Hammer, C. (USA), Detection of synchronisation in pulse trains.
Hellman, M., Cover, T. (USA), A review of recent results on learning with finite memory.
Ibragimov, I., Has'minsky, R. (USSR), Asymptotic behavior of some statistical estimates and weak signal transmission over a memoryless channel.
Ingarden, R. (Poland), Information-theoretical irreversible thermodynamics and its application to lasers.
Iwadare, Y. (Japan), A high speed decodable burst correcting code.
Jönsson, I. (Sweeden), Some codes with greater Hamming distances than those given by the BCH bounds.
Kadotta, T., Wyner, A. (USA), A coding theorem for stationary, asymtotically memoryless, continuous channels.
Kagan, A. (USSR), On an approach to optimal detection of a constant signal in additive noise.
Kailath, T. (USA), Some new applications of the innovation method.
Kanal, L. (USA), Modeling communication channels with memory.
Kasami, T. (Japan), Some results on the weight structure of Reed-Muller codes.
Katona, G. (Hungary),), Code-learning.
Kennedy, R. (USA), Performance limitations for communication through optical scattering channels.
Khaikin, B., Avetisov, E. (USSR), On a technique for redundant encoding using principles of coherent optics.
Kharitonenko, A. (USSR), Study of a linear-quadratic scheme for estimating the arrival time of a rapidly fluctuating optical signal.
Khodak, G. (USSR), Low-redundancy encoding for Markov sources.
Klovsky, D., Kirillov, N., Soifer, V. (USSR), Time-space communication channels (statistical model and optimal signal processing).
Kolesnik, V., Mironchikov, E. (USSR), Probability distributions on Abelian groups and decoding algorithm analysis.
Korzhik, V., Gabidulin, E. (USSR), A class of two-dimentional codes, capable of correcting lattice-patterned errors.
Krasovsky, R. (USSR), An adaptive energy-sensitive reception of optical signals.
Krieger, W. (USA), On the sphere packing bound in information theory.
Kulikowski, J. (Poland), Remarks on the value of information transmitted through a channel.
Kursky, V. (USSR), Evaluation of detecting and correcting capabilities of group codes given by their spectra.
Kuznetsov, A., Tsibakov, B. (USSR), On designing reliable storage with unreliable elements.
Lebedev, D. (USSR), Autoregressive model of a random field and its application to filtering problems.
Levenshtein, V. (USSR), On a method of designing quasilinear codes with a given "overlap distance".
Levin, B., Kushnir, A., Shinakov, Yu. (USSR), On asymptotic methods in statistical synthesis theory.
Levit, B. (USSR), On the use of generalized Bayes estimates for feedback transmission.
Levitin, L. (USSR), The amount of information and quantummechanical irreversibility of a measurement.
Lin'kov Yu. (USSR), Some information properties of discrete parameter estimates.
Linnik, I. (USSR), On the probability of great deviations of signal power.
Liptser, R., Shiryaev, A. (USSR), On the random processes with measures equivalent to the Wiener measure.
Longo, G. (Italy), On the evaluation of error probability for a finite Markov source.
Margulis, G. (USSR), On a combinatorial problem of network and coding theory.
Marinov, Yu., Gechev, B. (Bulgaria), On an adaptive technique for optimal classification of information signals.
Marinov, Yu., Gechev, B., Yanakiev, B. (Bulgaria), On the noise stability of a communication system with a two-ensamble coding alphabet.
Markov, A. (USSR), On the codes permitting nonunique decoding.
Markosian, A. (USSR), The inner stability number of Cartesian graph products and its application to information theory.

Massey, J. (USA), The theory of error propagation in convolutional codes.
Meulen van der, E. (USA), The discrete memoryless channel with two senders and one receiver.
Milenin, N. (USSR), Linear encoding in a data transmission system with additional noise sources.
Mityugov, V. (USSR), Quantum noise in coherent signals amplifiers.
Molnár, L. (Hungary), A comparison of several learning algorithms based on the probability distribution estimation method.
Morozov, V., Vilkova, L. (USSR), Sequential decoding for burstnoise communication channels.
Nemetz, T. (Hungary), An iterative model in noiseless coding.
Nevel'son, M., Khas'minsky, R. (USSR), On the stability of solution of one-dimensional stochastic equations.
Oganesian, S., Yagdjan, V. (USSR), On some classes of cyclic codes.
Ohnsorge, H. (F.R.G.), Communication without band-width economy.
Ovchinnikov, V. (USSR), Sequential decoding in channels defined by renewal processes.
Panov, T. (Bulgaria), On possibilities of improving the performance indices of technical information systems.
Panova, M. (Bulgaria), Some aspects of synthesizing technical data handling systems.
Perez, A. (ČSSR), Asymptotic discernability of two stationary Markov chains.
Pitskel, B., Styopin, A. (USSR), On the individual ergodic theorem of information theory for stationary stochastic fields.
Poddubny, V., Trivozhenko, B. (USSR), On the capacity of quantum angle-measuring channel.
Posner, E. (USA), Coverings and mutual information.
Pougeman, J. (ČSSR), Optimal codes for request transmission.
Prelov, V. (USSR), On the asymtotic capacity of continuous channel with highl-level nonadditive noise.
Prosin, A., Levshin, J. (USSR), Elements of statistical theory for randomly time-varying radio-channels and their computer simulation techniques.
Pyatoshin, Yu. (USSR), Some asymptotic properties of m-ary communication systems with coding in a semicontinuous channel.
Rabinovich, M., Yaroslavsky, L. (USSR). The results of noise statistical performance measurements in FM receivers.
Rázga, T. (Hungary), A combined decoding-demodulating method to increase the capacity of the communication channel.
Róna, P. (Hungary), On approximations used in the nonlinear analysis of angle-modulation channels.
Sagalovich, Yu. (USSR), Automata codes.
Samoilenko, S. (USSR), Some results of binoide codes simulation.
Sapple, E. (G.F.R.), Efficient encoding of sources with incompletely known probabilities.
Savage, J. (USA), The complexity of deterministic source encoding with a fidelity criterion.
Schalkwijk, P. (USA), An efficient source coding algorithm based on Pascal's triangle.
Semakov, N., Zinoviev, V., Zaitsev, G. (USSR), On relations between Hamming, Preparata and Golay codes and on extensions of Hamming codes.
Sheverdiaev, A. (USSR), An inequality for the error probability in a Gaussian memoryless channel.
Shtar'kov, Yu., Babkin, V. (USSR), A method of encoding discrete stationary sources with unknown statistics.
Shtein, V. (USSR), Legibility-optimal preemphasizing for quantized speech transmission.
Stiffler, J. (USA), Concatenated coding for correction of symbol insertions and deletions.
Stucki, P. (Switzerland), Picture processing by computer.
Tarasenko, F. (USSR), On the properties of D-structure statistics.
Taraskin, A. (USSR), On using the maximum likelihood method for signal detection in the presence of noise.
Tempel'man, A. (USSR), Information-theoretical ergodic theorems for stochastic fields.

Tenengol'ts, G. (USSR), Codes correcting random and burst errors in information exchange between computers.
Tourzan, G. (Iran), Optimum radar signal design and processing for randomly time-varying linear channels.
Turin, G. (USA), A summary of recent work on feedback communication.
Voronov, E., Sidorenko, V. (USSR), Selection of the group synchronization signal for data transmission using a fixed-length convolutional code.
Weldon, E. (USA), Encoding error-correcting codes on a general-purpose computer.
Yaglom, A. (USSR), Information and canonical correlations for Gaussian stochastic processes.
Zaidman, R. (USSR), On nonprobabilistic information theory.
Zigangirov, K. (USSR), Code distance lower bounds for convolutional codes.
Zyablov, V. (USSR), The complexity of iterasive and concatenated codes decoding.
Pinsker, M., Zyablov, V. (USSR), Correcting properties and decoding complexity of codes with small number of ones in parity check-matrices.
Shalkwijk, P. (USA), A class of simple and optimal strategies for block coding in a binary symmetric channel with a noiseless feedback.
Ebenau, K. V. (GFR), Some aspects of non-ground transmitting binary communication networks for future telecommunication service.
Hajan, A., Ito, E., Kakutani, S. (USA), Invariant measures and commuting transformations.
Geruny, R. (USSR), Information processing using methods of radioholography and coherent optics.
Petrov, V., Uskov, A. (USSR), Capacity of automatical systems.
Kladov, T. (USSR), On majority decoding of linear codes.
Varshamov, R. (USSR), Theory of irreducible polynomial synthesis.
Winkelbauer, K. (CSSR), On the problem of capacity for decomposible channels.

CONTENTS • СОДЕРЖАНИЕ