Intelligent Autonomous Primary 3D Feature Extraction in Vehicle System Dynamics' Analysis: Theory and Application

Annamária R. Várkonyi-Kóczy

Dept. of Measurement and Information Systems
Budapest University of Technology and Economics
Integrated Intelligent Systems Japanese-Hungarian Laboratory
Magyar tudósok krt. 2, H-1521 Budapest, Hungary
e-mail: koczy@mit.bme.hu

Abstract: 3D model reconstruction plays a very important role in computer vision as well as in different engineering applications. The determination of the 3D model from multiple images is of key importance. One of the most important difficulties in autonomous 3D reconstruction is the (automatic) selection of the 'significant' points which carry information about the shape of the 3D bodies i.e. are characteristic from the model point of view. Another problem to be solved is the point correspondence matching in different images.

In this paper a 3D reconstruction technique is introduced, which is capable to determine the 3D model of a scene without any external (human) intervention. The method is based on recent results of image processing, epipolar geometry, and intelligent and soft techniques.

Possible applications of the presented algorithm in vehicle system dynamics are also presented. The results can be applied advantageously at other engineering fields, like car-crash analysis, robot guiding, object recognition, supervision of 3D scenes, etc,. as well.

Keywords: 3D reconstruction, perspective geometry, point correspondence matching, epipolar geometry, fuzzy image processing, features extraction, information enhancement, image understanding, car-body deformation modeling, crash analysis

1 Introduction

3D reconstruction from images is a common issue of several research domains. In recent time the interest in 3D models has dramatically increased [1] [2]. More and more applications are using computer generated models. The main difficulty lies with the model acquisition. Although, more tools are at hand to ease the generation of models, it is still a time consuming and expensive process. In many cases models of existing scenes or objects are desired. Traditional solutions

include the use of stereo rigs, laser range scanners, and other 3D digitizing devices. These devices are often very expensive and require careful handling, and complex calibration procedures.

Creating photorealistic 3D models of a scene from multiple photographs is a fundamental problem in computer vision and in image based modeling. The emphasis for most computer vision algorithms is on automatic reconstruction of the scene with little or no user interaction [3].

In this paper an alternative approach is proposed which avoids most of the problems mentioned above. The object which has to be modeled is recorded from different viewpoints by a camera. The relative position and orientation of the camera and its calibration parameters will automatically be retrieved from image data. For the reconstruction we use characteristic features, like edges and corner points of the objects. The complexity of the technique is kept low on one hand by filtering out the points and edges carrying non-primary information (i.e. the so-called texture edges and points) while on the other by applying recent methods of digital image processing (see e.g. [4]-[7]) combined with intelligent and soft (e.g. fuzzy) techniques. This makes possible e.g. autonomous point correspondence matching which is the hardest step in 3D reconstruction and the biggest difficulty towards the automation of the procedure.

The introduced autonomous 3D reconstruction and its algorithms can be applied advantageously at many fields of engineering. In the second part of this paper, we will show a possible application in vehicle system dynamics: the usage in car-crash analysis.

Crash and catastrophe analysis has been a rather seldom discussed field of traditional engineering in the past. In recent time, both the research and theoretical analyses have become the part of the everyday planning work (see e.g. [8]). The most interesting point in crash analysis is that even though the crash situations are random probability variables, the deterministic view plays an important role in them. The stochastic view, statistical analysis, and frequency testing all concern past accidents. Crash situations, which occur the most frequently (e.g. the characteristic features of the crash partner, the direction of the impact, the before-crash speed, etc.) are chosen from these statistics and are used as initial parameters of crash tests. These tests are quite expensive, thus only some hundred tests per factory are realized annually, which is not a sufficient amount for accident safety. For the construction of optimal car-body structures, more crash-tests were needed. Therefore, real-life tests are supplemented by computer-based simulations, which increases the number of analyzed cases to 1-2 thousands. The computer-based simulations – like the tests – are limited to precisely defined deterministic cases. The statistics are used for the strategy planning of the analysis. The above mentioned example clearly shows that the stochastic view doesn't exclude the deterministic methods.

Crash analysis is very helpful for experts of road vehicle accidents, as well, since their work requires simulations and data, which are as close to the reality as possible. By introducing intelligent methods and algorithms, we can make the simulations more precise and so contribute towards the determination of the factors causing the accident.

The results of the analysis of crashed cars, among which the energy absorbed by the deformed car-body is one of the most important, are of significance at other fields, as well. They also carry information about the deformation process itself and may have a direct effect on the safety of the persons sitting in the car. Thus, through the analysis of traffic accidents and crash tests we can obtain information concerning the vehicle which can be of help in modifying the structure/parameters to improve its future safety. The ever-increasing need for more correct techniques, which use less computational time and can widely be applied results in the demand and acceptance of new modeling and calculating methods.

The techniques of deformation energy estimation used up till now can be classified into two main groups: The first one applies the method of finite elements [9]. This procedure is enough accurate and is suitable for simulating the deformation process, but this kind of simulation requires very detailed knowledge about the parameters of the car-body and its energy absorbing properties, which in most of the cases are not available. Furthermore, if we want to get enough accurate results, its complexity can be very high.

The other group covers the so called energy grid based methods, which starts from known crash test data and from the shape of the deformation or from the maximum car-body deformation [10]. The distribution of the energy, which can be absorbed by the cells, is considered just in 2D and the shape of the deformation is described also by a 2D curve which equals the border of the deformation visible from the top view of the car-body. The accuracy of this technique is not acceptable: In many of the cases, the shape of deformation can not be described in 2D and furthermore, the energy absorbing properties of the car-body change along the vertical axis as well causing serious impreciseness in the results.

In this paper new methods are introduced which avoid the above discussed disadvantages of the recently used techniques. First, the energy distribution is considered in 3D. Secondly, for the description of the shape of deformation spline surfaces are used, which are very suitable for modeling complex deformation surfaces. Third, the computational time and cost need is significantly decreased while the accuracy is increased by the application of intelligent techniques. Last, the deformation surface is obtained by a new 3D reconstruction method using only digital photos of the crashed car-body as input.

The methods presented in this paper can be applied at different fields of engineering. In this paper, we will show how can we construct a system capable to automatically build the 3D model of a crashed car as well as to determine the energy absorbed by the car-body deformation and the speed of the crash.

The paper is organized as follows: In Section II the primary edge extraction method is summarized. Section III is devoted to the 3D model estimation from multiple images, while Section IV is devoted to the conclusions. Intelligent applications in vehicle system dynamics and examples of the presented methods can be followed in the second part of the paper.

2 Primary Edge Extraction

Images usually contain a lot of different edges, among which there are texture edges and object contour edges, as well. From the point of view of scene reconstruction and image retrieval the latter ones are important because they carry the primary information about the shape of the objects. In we considered all of the possible edges during the model building/ searching/comparison, it would cause that the complexity/ time need of the procedure might grow to a possibly intolerable degree and furthermore, the (probably high number of) non-important details (edges) might lead to false decisions and increased the uncertainty of the modeling or caused that we disregarded recognizing an object. As a consequence, the separation of the 'significant' and 'unimportant' subsets of the edges, i.e. the enhancement of those ones which correspond to the object boundaries and thus carry primary information and the filtering out of the others which represent information of minor importance, not only significantly decreases the computational complexity of the processing but is of key importance from interpretation point of view.

2.1 Surface Smoothing

Let S_t be the surface describing the image to be processed, i.e. $S_t = \{(x, y, z); z = I(x,y,t)\}$, where the variables x and y represent the horizontal and vertical coordinates of the pixel, z stands for the luminance value, which is the function of the pixel coordinates and of the time t. The smoothing is performed by image surface deformation. Such a process preserves the main edges (contours) in the image. The surface deformation process satisfies the following differential equation [11]:

$$\frac{\partial I_t}{\partial t} = k\mathbf{n}, \tag{1}$$

where k corresponds to the 'speed' of the deformation along the normal direction $\mathbf{n}$. In our case, this value k is represented by the mean curvature of the surface at location $[x, y]$, i.e. the speed of the deformation at a point will be the function of the mean curvature at that point. The mean curvature is defined as:

$$k = \frac{k_1 + k_2}{2}, \tag{2}$$

where k_1 and k_2 stand for the principal curvatures. Starting from equation (2), the following partial differential equation can be derived (for details, see [12]):

$$k = \frac{(1+I_y^2)I_{xx} - 2I_x I_y I_{xy} + (1+I_x^2)I_{yy}}{2\left(1+I_x^2+I_y^2\right)^{3/2}}. \tag{3}$$

Here I_x, I_y, I_{xx}, I_{xy}, I_{yy} stand for the partial derivatives with respect to the variables indicated as lower indices. Starting from equation (1) the surface at time $t+\Delta t$ (for small Δt) can be calculated as follows [11]:

$$\begin{aligned} &I(x,y,t+\Delta t) = \\ &\quad I(x,y,t) + k\sqrt{1+I_x^2(x,y,t)+I_y^2(x,y,t)}\Delta t + o(\Delta t) \end{aligned} \tag{4}$$

where $o(\Delta t)$ represents the error of the approximation.

2.2 Edge Detection

Let $z_{0,x,y}$ be the pixel luminance at location $[x,y]$ in the original image. Let us consider the group of neighboring pixels which belong to a 3x3 window centered on $z_{0,x,y}$.

The output of the edge detector is yielded by the following equation:

$$\begin{aligned} z^p_{x,y} &= (L-1)MAX\left\{m_{LA}(\Delta v_1), m_{LA}(\Delta v_2)\right\} \\ \Delta v_1 &= \left|z_{0,x-1,y} - z_{0,x,y}\right| \\ \Delta v_2 &= \left|z_{0,x,y-1} - z_{0,x,y}\right| \end{aligned} \tag{5}$$

where $z^p_{x,y}$ denotes the pixel luminance in the edge detected image and m_{LA} stands for the used membership function (see Figure 1). $z_{0,x-1,y}$ and $z_{0,x,y-1}$ correspond to the luminance values of the left and upper neighbors of the processed pixel at location $[x,y]$. L-1 equals to the maximum luminence value (e.g. 255). For more details see [5].

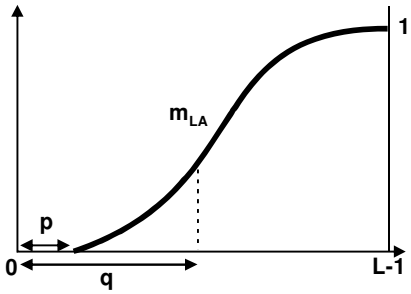


Figure 1
Fuzzy membership function m_{LA} of 'edge'. L-1 equals to the maximum intensity value, p and q are tuning parameters

2.3 Edge Separation

The obtained smoothed image is used for extracting the most characteristic edges of the objects. The procedure is performed as follows:

For each edge point taken from the edge map of the original image, the environment of the point is analyzed in the smoothed image. The analysis is realized by calculating the mean squared deviation of the color components (in case of grayscale images the gray-level component) in the environment of the selected edge point.

Let $\mathbf{p}=[p_x, p_y]$ be an edge point in the original image and let $\mathbf{M}$ denote a rectangular environment of $\mathbf{p}$ with width w and height h. The mean squared deviation is calculated as follows:

$$d = \frac{\sum_{i=p_x-w/2}^{p_x+w/2} \sum_{j=p_y-h/2}^{p_y+h/2} (\mu - I(i,j,t_{stop}))^2}{hw}, \tag{6}$$

where t_{stop} represents the duration of the surface deformation. In case of grayscale images, μ denotes the average gray level inside the environment $\mathbf{M}$. For color images, the whole process should be done for each component separately and in this case μ corresponds to the average level of this color component inside the environment $\mathbf{M}$.

If the so calculated deviation exceeds a predefined threshold value, then the edge point is considered as useful edge. As result, an image containing only the most characteristic edges is obtained.

3 3D Model Estimation from Multiple Images

The topic of building 3D models from images is a relatively new research area in computer vision and, especially when the objects are irregular, not finished at all. In the field of computer vision, the main work is done at one hand on the automation of the reconstruction while on the other on the implementation of an intelligent human-like system, which is capable to extract relevant information from image data and not by all means on building a detailed and accurate 3D model like usually in photogrammetry is. For this purpose, i.e. to get the 3D model of scenes, to limit/delimit the objects in the picture from each other is a key importance [13].

3.1 Noise Smoothing

As the first step, the pictures, used in the 3D object reconstruction are preprocessed. As a result of the preprocessing procedure the noise is eliminated. For this purpose we use a special fuzzy system characterized by an IF-THEN-ELSE structure and a specific inference mechanism proposed by Russo [4], [6]. Different noise statistics can be addressed by adopting different combinations of fuzzy sets and rules.

Let $I(\mathbf{r})$ be the pixel luminance at location $\mathbf{r}=[x,y]$ in the noisy image, where x is the horizontal and y the vertical coordinate of the pixel. Let $I_0 = I(\mathbf{r}_0)$ denote the luminance of the input sample having position $\mathbf{r}_0$ and being smoothed by a fuzzy filter. The input variables of the fuzzy filter are the amplitude differences defined by:

$$\Delta I_j = I_j - I_0, j = 1,...,8 \tag{7}$$

where the $I_j=I(\mathbf{r}_j)$, $j=1,\ldots,8$ values are the luminance values of the neighboring pixels of the actually processed pixel $\mathbf{r}_0$ (see Figure 2a). Let K_0 be the luminance of the pixel having the same position as $\mathbf{r}_0$ in the output image. This value is determined by the following relationship:

$$K_0 = I_0 + \Delta I \tag{8}$$

where ΔI is determined later (see (11)).

Let $W = \bigcup_{i=1}^{9} W_i$ be defined by a subset of the eight neighboring pixels around $\mathbf{r}_0$. Let the rule base deal with the pixel patterns $W_1,\ldots,W_9$ (see Figure 2b). The value K_0 can be calculated, as follows:

$$\lambda = MAX\left\{MIN\left\{m_{LP}(\Delta I_j) : r_j \in W_i\right\}; i = 1,...,9\right\} \tag{9}$$

$$\lambda^* = MAX\left\{MIN\left\{m_{LN}(\Delta I_j) : r_j \in W_i\right\}; i = 1,...,9\right\} \tag{10}$$

$$\Delta I = (L-1)\Delta\lambda$$
$$K_0 = I_0 + \Delta I \tag{11}$$

where $\Delta\lambda=\lambda-\lambda^*$, L is the maximum of the gray level intensity, m_{LP} and m_{LN} correspond to the membership functions and $m_{LP}(I)=m_{LN}(-I)$ (see Figure 2c). The filter is recursively applied to the input data.

r_1	r_2	r_3
r_4	r_0	r_5
r_6	r_7	r_8

Figure 2a
The neighboring pixels of the actually processed pixel r_0

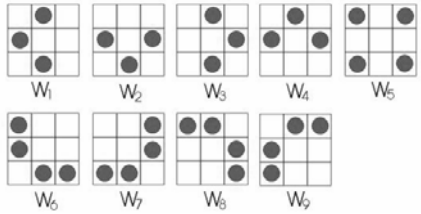


Figure 2b
Pixel Patterns

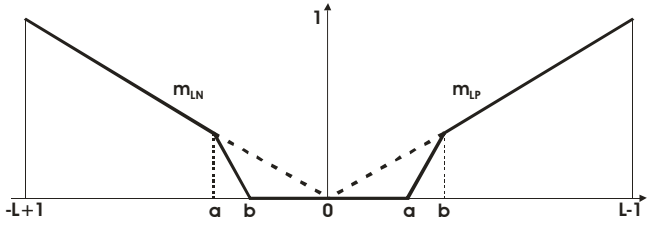


Figure 2c
Membership functions m_{LN} (large negative) and $m_{LP,}$ (large positive), a and b are parameters for the tuning of the sensitivity to noise of the filtering

3.2 Corner Detection

The edge and corner points are the most characteristic feature points in an image. For our modeling system the determination of the corners are very important. The applied corner detection algorithm utilizes the principles of the fuzzy filters and

edge detection algorithms of Russo. Besides fuzzy reasoning it uses a local structure matrix composed of the partial derivatives of the gray level intensity of the pixels. As input, we consider the noiseless and smoothed image, while as output the corners are got. A corner is indicated by two strong edges [14].

Most of the corner detection algorithms are derived from a so called local structure matrix, which has the form of

$$\mathbf{L}(x,y) = G(x,y) * \begin{bmatrix} \left(\frac{\partial I}{\partial x}\right)^2 & \left(\frac{\partial I}{\partial x}\frac{\partial I}{\partial y}\right) \\ \left(\frac{\partial I}{\partial x}\frac{\partial I}{\partial y}\right) & \left(\frac{\partial I}{\partial y}\right)^2 \end{bmatrix}, \tag{13}$$

where $G(x,y)$ represents the so called 2D Gaussian hump and * stands for the convolution. One of the corner detection algorithms, which uses the above local structure matrix is the Förstner's one. Förstner determines corners as local maxima of function $H(x,y)$ [15].

$$H(x,y) = \frac{\left(\frac{\partial I}{\partial x}\right)^2\left(\frac{\partial I}{\partial y}\right)^2 - \left(\frac{\partial I}{\partial x}\frac{\partial I}{\partial y}\right)^2}{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2} \tag{14}$$

In most of the cases we can not unambiguously determine that the analyzed image point is a corner or not with only the help of a certain concrete threshold value. Therefore in the proposed algorithm, fuzzy techniques are applied for the calculation of the values (corners) which increases the rate of correct corner detection. As higher the calculated H value as higher the membership value, that the analyzed pixel is a corner. Fuzzifying the H values into fuzzy sets and applying a fuzzy rulebase we can evaluate the 'cornerness' of an analyzed pixel. This property of the pixel can advantageously be used also at the searching for the corresponding corner points in stereo image pairs (point correspondence matching), which is an indefinite step of the automatic 3D reconstruction (see also [16] and [17]).

3.3 Point Correspondence Matching and Determination of the 3D Coordinates of the Corner Points

For increasing the efficiency of the process, before starting with the actual model building, we can filter out the non-significant (texture type) edges and corners of the pre-processed images. The next step is the determination of the 3D coordinates of the remaining, primary edge points of the object. First the corner point correspondences are determined which is followed by the determination of the

edge correspondences. If the angle between the camera positions of the different images is relatively small then after the evaluation of the projection matrices of the images the corresponding points can be calculated automatically with high reliability in each image. The problem to overcome is that a point determines not another point but a line (the so called epipolar line) in the other images. To decrease the number of candidate points, first, we search the corner or edge points of the epipolar line, i.e. those points which belong to a corner or edge of the image and then the fuzzy measure of the differences of the environment of the points are minimized. The similarity of the above mentioned 'cornerness' is also considered. Having the most probable point correspondences we can calculate the 3D position of the image points and in the knowledge of the 3D coordinates of the significant points the spatial model of the object can easily be built.

3.4 Epipolar Geometry

Epipolar geometry exists between a two camera system. An important practical application of epipolar geometry is to aid the search for corresponding points, reducing it from the entire second image to a single epipolar line. The epipolar geometry can easily be found from a few point correspondences. Consider the case of two perspective images of a scene illustrated by Figure 3. The 3D point **M** is projected to point $\mathbf{m_1}$ in the left image and $\mathbf{m_2}$ in the right one. Let $\mathbf{C_1}$ and $\mathbf{C_2}$ be the centers of projection of the left and right cameras, respectively. Points $\mathbf{m_1}$ in the first image and $\mathbf{m_2}$ in the second image are the imaged points of the point **M** of the 3D space. The epipolar constraint can be written as

$$\mathbf{m}_2^T \mathbf{F} \mathbf{m}_1 = 0 . \tag{15}$$

F is known as the fundamental matrix, which defines a bilinear constraint between the coordinates of the corresponding image points. If $\mathbf{m_2}$ is the point in the second image corresponding to $\mathbf{m_1}$, it must lie on the epipolar line $\mathbf{l_{m1}}$ (see Figure 3).

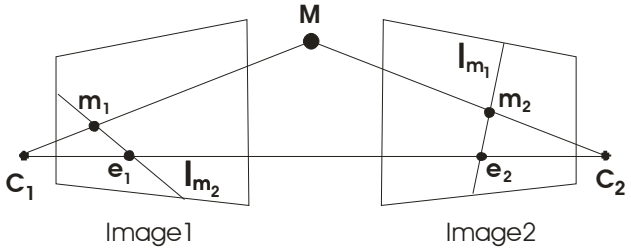


Figure 3

Illustration of epipolar geometry ($\mathbf{e_1}$ and $\mathbf{e_2}$ are the epipoles, $\mathbf{m_1}$ and $\mathbf{m_2}$ the corresponding image points, $\mathbf{l_{m1}}$ and $\mathbf{l_{m2}}$ are the epipolar lines, $\mathbf{C_1}$ and $\mathbf{C_2}$ are the camera positions. **M** is the projected 3D point)

3.5 Image Point Matching

First we have to find the most characteristic image points. These points are the corners of the analyzed object. Corners can effectively be detected with the help of the fuzzy based corner detector. Then, for each detected corner we have to determine the corresponding epipolar line. We know that the corresponding point will lay (in fuzzy sense) on this epipolar line and is also a corner point (see Fig. 4). Thereinafter the fuzzy measure of the differences of the environments of the point to be matched and the so got candidate points are minimized by a fuzzy based searching algorithm (see Figs. 4 and 5) which determines the most probably corresponding point. The same procedure is applied to edge points. For details see [13], [16], and [18].

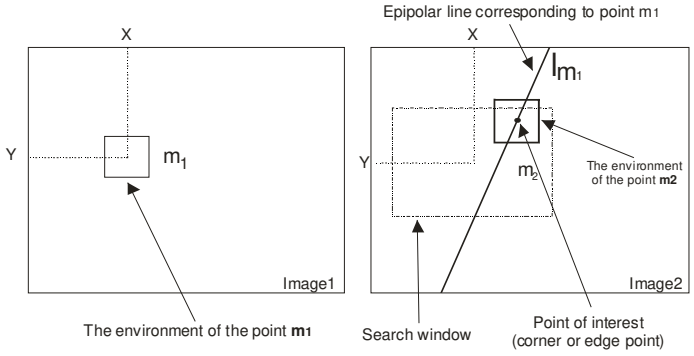


Figure 4
Illustration of the matching algorithm

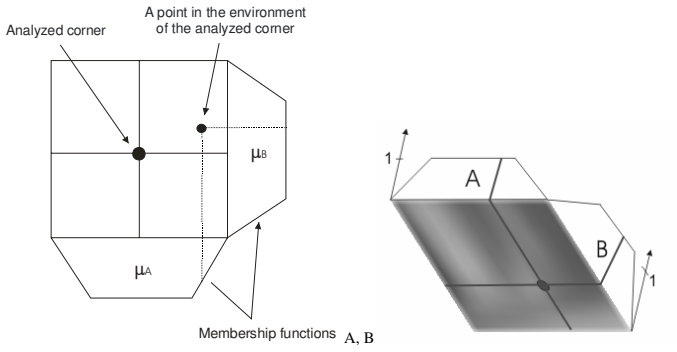


Figure 5
Illustration of an image point from the environment of the point $\mathbf{m_2}$ (see Figure 4) and the corresponding values of the membership functions of the fuzzy sets A and B)

3.6 Camera Calibration by Estimation of the Perspective Projection Matrix

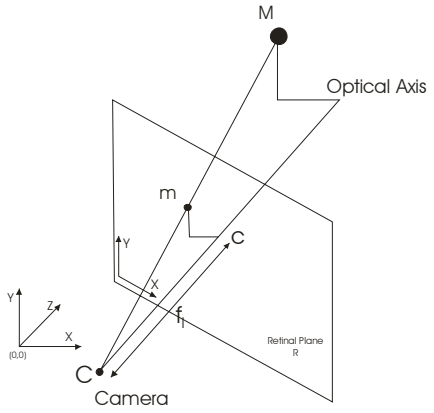


Figure 6
Perspective projection – illustration of points **M**=[*X,Y,Z*] and its projection **m**=[*x,y*] in the retinal plane *R*

There exists a collineation, which maps the projective space to the camera's retinal plane: 3D to 2D. Then the coordinates of a 3D point $\mathbf{M} = [M_X, M_Y, M_Z]^T$ (determined in an Euclidean world coordinate system) and the retinal image coordinates $\mathbf{m} = [m_x, m_y]^T$ (see Figure 6) are related by the following equations:

$$W\mathbf{m} = \mathbf{PM} \tag{16}$$

$$\begin{bmatrix} m_x W \\ m_y W \\ W \end{bmatrix} = \begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & 1 \end{bmatrix} \begin{bmatrix} M_X \\ M_Y \\ M_Z \\ 1 \end{bmatrix} \tag{17}$$

where W is a scale factor, $\mathbf{m} = [m_x, m_y, 1]^T$ and $\mathbf{M} = [M_X, M_Y, M_Z, 1]^T$ are the homogeneous coordinates of points **m** and **M**, and **P** is a 3 × 4 matrix representing the collineation 3D to 2D. One parameter of **P** can be fixed ($l = 1$). **P** is called perspective projection matrix. Values *a, b, c, d, e, f, g, h, i, j, k* are the elements of the projection matrix **P**. It is clear that

$$W = iM_X + jM_Y + kM_Z + 1 \tag{18}$$

From (17) we can calculate the coordinates of point **m** (m_x, m_y), as follows:

$$m_x = \frac{aM_X + bM_Y + cM_Z + d}{W} \tag{19}$$

$$m_y = \frac{eM_X + fM_Y + gM_Z + h}{W} \tag{20}$$

All together we have eleven unknowns (the elements of the projection matrix) that means that we need six points to determine the projection matrix. For more details see [19].

4 Primary Edge Extraction

The basic concept of the primary edge extraction method described in the first part includes the following steps: Consider that we have an image and we want to extract the edges corresponding to the object contours.

As the first step, it is necessary to remove the unimportant details from the image. The smoothing procedure used for this purpose is based on surface deformation. After smoothing the image, only the most characteristic contours are kept.

Next, the edge map of the original image is constructed using the fuzzy-based edge detection method described in [24]. Such an edge map contains all the possible edges.

After this step, the two processed images – the smoothed one and the edge map of the original image – are analyzed simultaneously in the following way: In case of each of the edge points a small environment of the point is taken in the original image and using the smoothed image the variance of the color components inside of this environment is analyzed. If the variance is below a predefined threshold value then the edge point is removed while otherwise it is considered as a useful, primary edge point.

The effectivity of the above information enhancement method detailed in Section II of the first part of this paper is illustrated by two simple examples. In all of the examples color images are used. The figures allow the comparison of the original edge map and the edges after applying the proposed method.

Figs. 7 and 8 are illustrations for the virtual process of changing of an image surface along the time. The two examples are fine fragments of the next example (surface of the car) at time 0 and at time t_{stop}, respectively.

The next example analyses a photo taken of a car. In Figure 9 the original image can be seen, while Figure 10 shows the smoothed image using the discussed surface deformation. Figs. 11 and 12 represent the edge maps before and after the processing, respectively. As you can see in Figure 12 many of the details disappear after the processing and only the charac-teristic edges of the car are left. This helps filtering out the non-important details and enhancing the most significant features/objects in images thus making easier image retrieval, object recognition, reconstruction of scenes, etc.

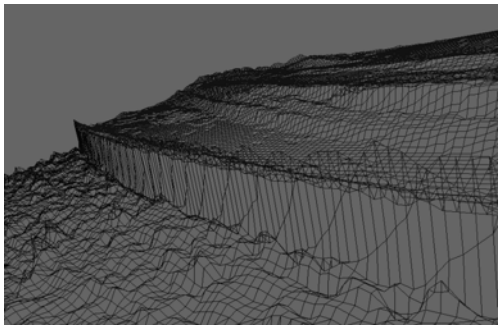


Figure 7
Illustration of an image surface before the deformation

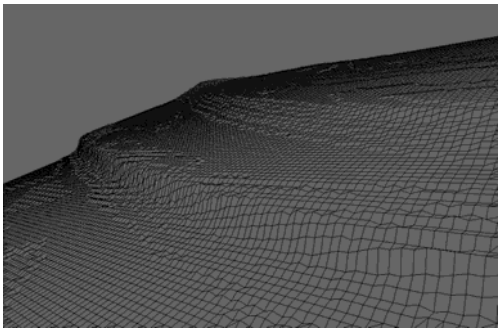


Figure 8
Illustration of an image surface after the deformation

Figure 9
Original image taken of a car

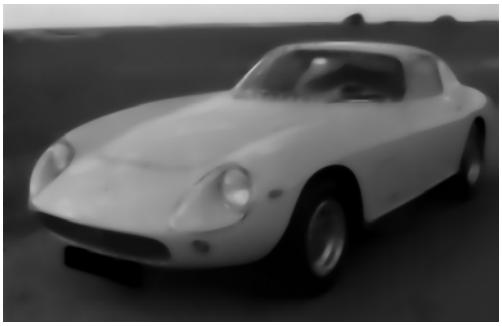


Figure 10
Smoothed image using surface deformation based on mean curvature

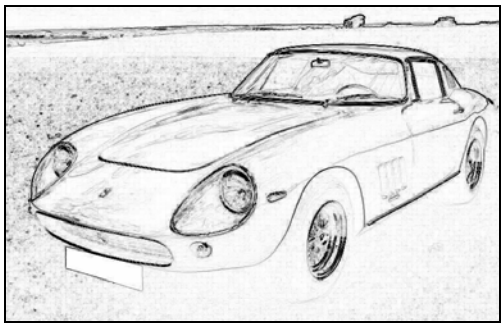


Figure 11
Edge map of the original image

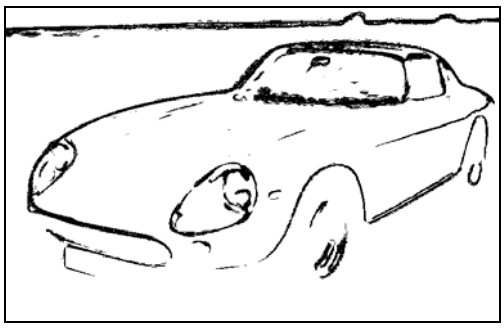


Figure 12
Edges after applying the proposed information enhancement method

5 3D Model Estimation from Multiple Images

The basic concept of the 3D model estimation method described in the first part can be summarized as follows: As the first step, the pictures, used in the 3D-object reconstruction are preprocessed, which starts with noise elimination and edge detection by applying the fuzzy filters and fuzzy edge detection algorithm described in [24], [25]. This is usually followed by the primary edge extraction method (see Section V) [38].

For the modeling the determination of the primary edges and corners are very important because they carry the most characteristic information about the shape of the objects to be modeled. The applied corner detection method utilizes that a corner is indicated by two strong edges. It also applies fuzzy reasoning and the used local structure matrix composed of the partial derivatives of the gray level intensity of the pixels is extended by fuzzy decision making. The algorithm assigns also a new attribute, the fuzzy measure of being a corner, to the analyzed pixel. This property of the corners can advantageously be used at the searching for the corresponding corner points in stereo image pairs.

The next step is the determination of the 3D coordinates of the extracted edge points. First the corner point correspondences are determined which is followed by the determination of the edge correspondences in the different images. If the angle between the camera positions is relatively small then after the estimation of the projection matrices of the images (necessary for the calibration) the corresponding points can be calculated automatically with high reliability in each image. We search for the characteristic corner or edge points lying (in fuzzy sense) on the epipolar line and then the point correspondence matching is done by minimizing the fuzzy measure of the differences of the environment of the points with the help of a fuzzy supported searching algorithm [36]. The similarity of the above mentioned 'cornerness' is also considered. (The corresponding corner points keep their 'cornerness' property in the pictures near to each other with high reliability). Having the point correspondences we can calculate the 3D position of the image points (the camera calibration is solved by the determination of the Perspective Projection Matrix [32]) and in the knowledge of the 3D coordinates and the correspondences of the significant points the spatial model of the car body can easily be built.

The effectivity of the above 3D reconstruction method detailed in Section III of this paper is illustrated by a simple example.

Figure 13a shows the original photo of the crashed car corrupted by noise. In Figure 13b the fuzzy filtered image while in Figs. 13c and 13d the images after fuzzy based edge and corner detection can be followed. Figs. 13e-13h illustrate a different camera position of the car. The 3D model of the deformed part of the car-body is shown in Figure 14.

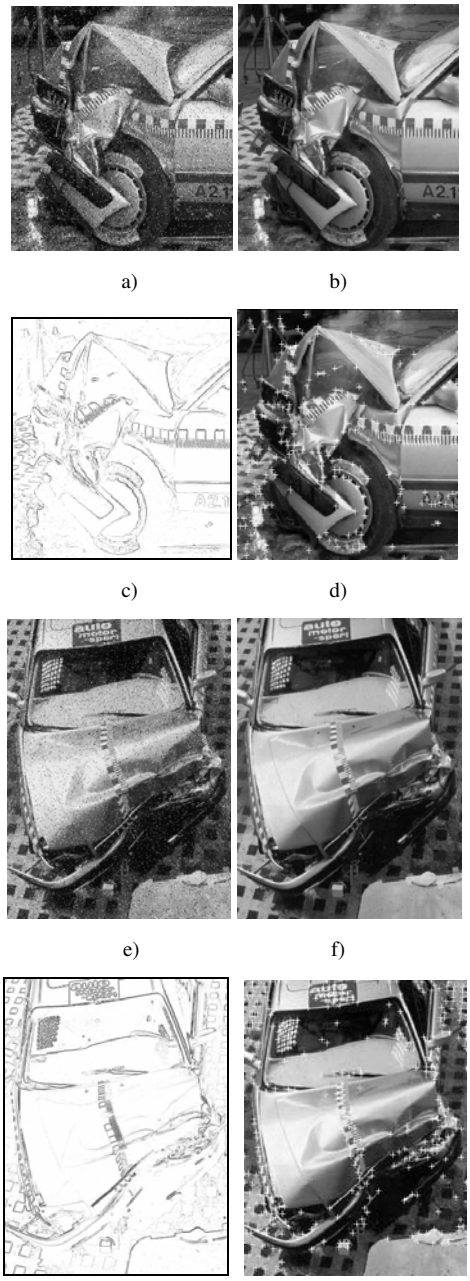
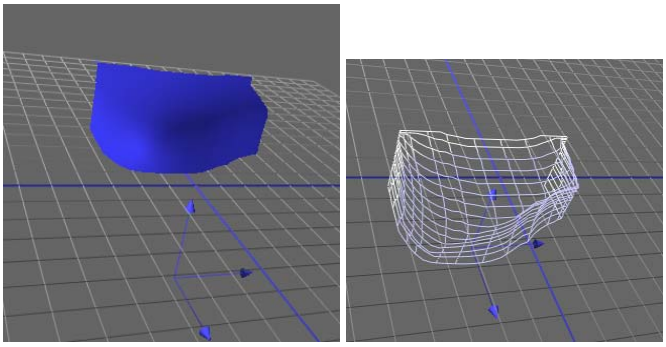


Figure 13
2 examples of the (a), (e): original photos, (b), (f): fuzzy filtered images, (c), (g): results after edge and (d), (h): corner detection of a crashed Audi 100



a) b)

Figure 14
3D model of the deformed part of the car body

6 Car Crash Analysis

In this Section a possible application of the introduced methods taken from vehicle system dynamics will be presented. The system aims the intelligent analysis of crashed cars and is able to determine the 3D model, the amount of the energy absorbed by the deformation and further important information, e.g. the energy equivalent speed and the direction of impact of the crash.

The block structure of the proposed new car crash analysis system can be followed in Figure 9. It contains four well defined sub-blocks. The first (image processing) is responsible for the pre-processing of the digital photos (noise elimination/filtering, edge detection, corner detection) and for the 3D modeling (including the point correspondence matching and the 3D model building). The second part of the system (comparison of models) calculates the volumetric change of the car body from the deformed and the original 3D models of the car. Parallel with it an expert system (Expert system) determines the direction of the impact. Based on the direction of impact and volumetric change a hierarchical fuzzy-neural network system (Fuzzy-Neural Network) determines the absorbed energy and the energy equivalent speed of the car. In the followings we will briefly outline the steps of the analysis not discussed previously.

After constructing the 3D model of the deformed car body (see Figure 8) we have to determine the volume of the detoriated car body which means that it is necessary to compare the deformed and the undamaged 3D car-bodies. This calculation is performed by the module named ‘Comparison of models’ (see Figure 15). The inputs of this module are the spatial models of the damaged and undamaged car-bodies. As result, we obtain the volumetric difference between the two models.

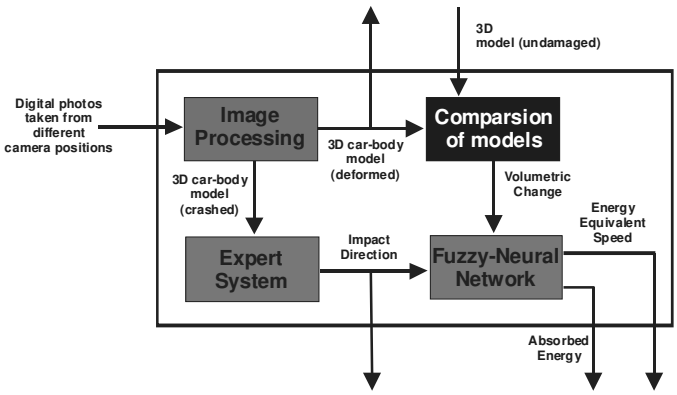


Figure 15

Block-structure of the intelligent car crash analysis system

The spatial model of the deformed car-body serves as input of another module, as well. This module applies an expert system and produces the direction of impact. For this we use the so called 'energy-centers' of the undamaged and deformed car-body parts and the direction is estimated from the direction of movement of the energy-center. (During the deformation the different 3D cells of the car-body absorb a certain amount of energy. The energy-center can be determined by weighting the cells by the corresponding energy values.)

From the volumetric difference and from the direction of impact an intelligent hierarchical fuzzy-neural network system evaluates the energy absorbed by the deformation and the equivalent energy equivalent speed (EES). For the training of this part of the system simulation and crash test data can be used. The training data include the volumetric change, the direction of the impact (input data) and the corresponding deformation energy (output data).

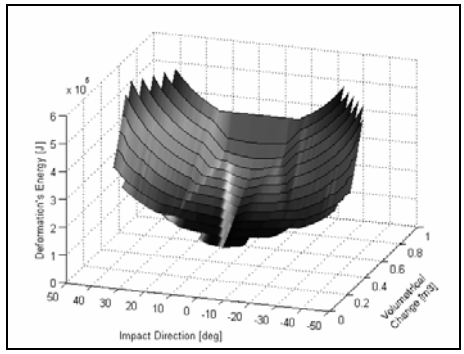


Figure 16

Relation among the direction of impact, volumetric change, and the deformation energy based on simulation data (Mercedes 290)

The relation among the direction of impact, volumetric change, and the deformation energy is illustrated by Figure 16. If, as usually is the case, this surface is symmetric (to the longitudinal axes of the vehicle), it is enough to deal with its half part. The mapping is approximated by a hierarchical fuzzy-NN system (subsystem 'Fuzzy-Neural Network' in Figure 15). The surface is divided into domains, which can 'easily' be modeled. Each domain is modeled separately by a small NN. Because of the uncertainties in the transitions among the domain, a fuzzy system is applied for the determination of the fired domain(s). The mapping in Figure 16 needs only to be divided into two domains according to the impact direction (see Fig. 17), thus in this very simple case the fuzzy rulebase 'above' the NN system contains only two rules (The input fuzzy sets are shown in Fig. 18):

IF the direction *IS* D1 *THEN* use NN1

IF the direction *IS* D2 *THEN* use NN2

Here we would like to remark two things:

1 In general the mapping is more complex and it can be advantageous to define more domains using both inputs to keep the complexity of the used NNs low.

2 The module responsible for the determination of the absorbed energy applies a pre-classification step according to a hierarchical decision-tree (Figure 19), because for choosing the correct set of neural networks we have to pre-determine the category and the type of the analyzed vehicle and the main character of the crash (frontal full impact, frontal offset impact, side impact, corner impact, rear impact). Cars are categorized into car types according to their weights. (In this paper as example a crashed Audi is shown. Although, the analysis is based on the NNs taught by the simulation data of a similar, but Mercedes car). Side impact means that neither the front nor the rear of the vehicle is touched.

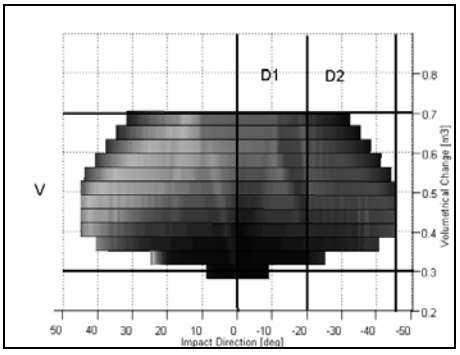


Figure 17
Segmentation of the surface in Figure 10

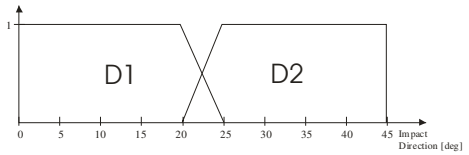


Figure 18
Membership functions defined on the universe of impact direction

For approximating domains D1 and D2 we applied simple feed-forward backpropagation NNs with one hidden layer and three hidden neurons. The NNs are used to determine the deformation's energy and EES. During the tuning (teaching period) of the system, the determined EES values were compared to known test results and the parameters of the expert system were modified to minimize the LMS error.

The operation of the introduced intelligent crash analysis system is illustrated on a crashed car. The parameters of the car are as follows:

Vehicle/Mass of the vehicle: Audi 100/1325 kg

Volumetric change (evaluated): 0.62 m^3

Absorbed deformation energy (evaluated): 171960 Joule

The resulted 3D model is shown in Figure 14. The results of the analysis are summarized in Table 1 (see also [39]). The error of the analysis depends on the resolution of the model (i.e. the distance between two layers in the 3D model, Figure 14b) and also on the accuracy of the crash test data.

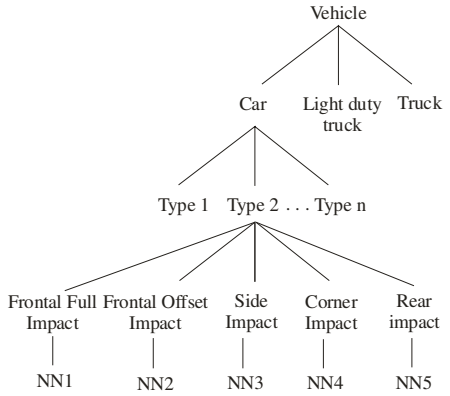


Figure 19
Hierarchical structure of the pre-classification in the EES determination

Table 1
The direction of impact and the energy equivalent speed of the crashed car

	Direction of impact [deg]	EES of the vehicle [km/h]
Real Data	0	55
Proposed method	2	58
2D method	2	59,5

Conclusions

In this paper intelligent methods are introduced which open a way for autonomous 3D model reconstruction. The 3D model reconstruction uses as input only digital images taken from different camera positions. The technique combines recent results of epipolar geometry, intelligent methods of image processing, and different fuzzy techniques. It applies a new edge information extraction procedure, as well, which is able to separate the edges carrying primary information and those representing only information of minor importance. The methods presented in the paper can advantageous be used in many 2D and 3D applications, in computer vision, in sketch based image retrieval methods, in vehicle system dynamics, etc.

As a possible new application taken of the field of vehicle system dynamics, an intelligent expert system is also presented which includes significant steps towards the autonomous analysis of car-crashes. It makes easy to determine the special shape of crashed cars (or other objects), the amount of the energy absorbed by the deformation, and further important information, like the energy equivalent speed (EES).

Acknowledgment

This work was sponsored by the Hungarian Fund for Scientific Research (OTKA T049519) and the Structural Fund for Supporting Innovation in New Knowledge and Technology Intensive Micro- and Spin-off Enterprises (GVOP-3.3.1-05/1.2005-05-0160/3.0)

References

[1] Taylor, C., P. Debevec, and J. Malik, “Reconstructing Polyhedral Models of Architectural Scenes from Photographs,” *Computer Vision - ECCV'96*, Lecture Notes in Computer Science, Vol. 1065/II, pp. 659-668, 1996

[2] Hartley, R., “Euclidean Reconstruction from Uncalibrated Views,” in J. L. Mundy, A. Zisserman, and D. Forsyth (eds.), *Applications of Invariance in Computer Vision*, Lecture Notes in Computer Science, Vol. 825, Springer-Verlag, pp. 237-256, 1994

[3] Hartley, R. and A. Zisserman, “*Multiple View Geometry in Computer Vision*,” Cambridge University Press, 2000

[4] Russo, F., "Fuzzy Filtering of Noisy Sensor Data," *In Proc. of the IEEE Instrumentation and Measurement Technology Conference*, Brussels, Belgium, 4-6 June 1996, pp. 1281-1285

[5] Russo, F., "Edge Detection in Noisy Images Using Fuzzy Reasoning," IEEE Transactions on Instrumentation and Measurement, Vol. 47, No. 5, Oct. 1998, pp. 1102-1105

[6] Russo, F., "Recent Advances in Fuzzy Techniques for Image Enhancement," *IEEE Transactions on Instrumentation and Measurement*, Vol. 47, No. 6, Dec. 1998, pp. 1428-1434

[7] Rogers D. F., *Procedural Elements for Computer Graphics,* McGraw Hill, New York, 1985

[8] Happer A., M. Araszewski, Practical Analysis Technique for Quantifying Sideswipe Collisions, 1999

[9] Melander, A., "Finite Element Simulation of Crash Testing of Laser Welded Joints," Research report, Swedish Institute for Metals, no: IM-2000-062, 2000

[10] Chen, H. F., C. B. Tanner, N. J. Durisek, and D. A. Guenther, "Pole Impact Speeds Derived from Bilinear Estimations of Maximum Crush for Body-On-Frame Constructed Vehicles," Paper no. 2004-01-1615, Society of Automotive Engineers, Warrendale, Pennsylvania, 2004

[11] C. Lu, Y. Cao, D. Mumford, "Surface Evolution under Curvature Flows", Submitted for the special issue on Partial Differential Equations (PDE's) in Image Processing," Computer Vision, and Computer Graphics, p. 19, 2002

[12] Gray, A. "The Gaussian and Mean Curvatures" and "Surfaces of Constant Gaussian Curvature," §16.5 and Ch. 21 in Modern Differential Geometry of Curves and Surfaces with Mathematica, 2nd ed. Boca Raton, FL: CRC Press, pp. 373-380 and 481-500, 1997

[13] Pollefeys, M., *Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences*, PhD thesis, ESAT-PSI, K.U. Leuven, 1999

[14] Várkonyi-Kóczy, A. R., A. Rövid, "Fuzzy Logic Supported Corner Detection," *Journal of Intelligent and Fuzzy Systems*, to be published in 2007

[15] W. Förstner, "A Feature Based Correspondence Algorithm for Image Matching," *Int. Arch. Photogramm. Remote Sensing*, Vol. 26, pp. 150-166, 1986

[16] Rövid, A., A. R. Várkonyi-Kóczy, "Corner Detection in Digital Images Using Fuzzy Reasoning," *In Proc. of the 2nd IEEE International Conference on Computational Cybernetics*, August 30-Sept. 1, 2004, Vienna, Austria, pp. 95-99

[17] Várkonyi-Kóczy, A. R., A. Rövid, "Improved Fuzzy Based Corner Detection Method," *In Proc. of the IEEE Int. Workshop on Soft Computing Appl., SOFA'2005*, Szeged-Arad, Hungary-Romania, Aug. 27-30, 2005, pp. 237-242

[18] Várkonyi-Kóczy, A. R., A. Rövid, "Point Correspondence Matching for 3D Reconstruction Using Fuzzy Reasoning," *In Proc. of the 3rd IEEE Int. Conf. on Computational Cybernetics, ICCC 2005*, Mauritius, Apr. 13-16, 2005, pp. 87-92

[19] Rövid, A., A. R. Várkonyi-Kóczy, P. Várlaki, "3D Model Estimation from Multiple Images," *In Proc. of the IEEE International Conference on Fuzzy Systems, FUZZ-IEEE'2004*, July 25-29, 2004, Budapest, Hungary, Vol. 3, pp. 1661-1666

[20] Taylor, C., P. Debevec, J. Malik, "Reconstructing Polyhedral Models of Architectural Scenes from Photographs," *Computer Vision - ECCV'96*, Lecture Notes in Computer Science, Vol. 1065/II, pp. 659-668, 1996

[21] Hartley, R., "Euclidean Reconstruction from Uncalibrated Views," in J. L. Mundy, A. Zisserman, and D. Forsyth (eds.), *Applications of Invariance in Computer Vision*, Lecture Notes in Computer Science, Vol. 825, Springer-Verlag, pp. 237-256, 1994

[22] Hartley, R., A. Zisserman, "*Multiple View Geometry in Computer Vision*," Cambridge University Press, 2000

[23] Russo, F., "Fuzzy Filtering of Noisy Sensor Data," *In Proc. of the IEEE Instrumentation and Measurement Technology Conference*, Brussels, Belgium, 4-6 June 1996, pp. 1281-1285

[24] Russo, F., "Edge Detection in Noisy Images Using Fuzzy Reasoning," IEEE Transactions on Instrumentation and Measurement, Vol. 47, No. 5, Oct. 1998, pp. 1102-1105

[25] Russo, F., "Recent Advances in Fuzzy Techniques for Image Enhancement," *IEEE Transactions on Instrumentation and Measurement*, Vol. 47, No. 6, Dec. 1998, pp. 1428-1434

[26] Rogers D. F., *Procedural elements for Computer Graphics,* McGraw Hill, New York, 1985

[27] Happer A., M. Araszewski, Practical Analysis Technique for Quantifying Sideswipe Collisions, 1999

[28] Melander, A., "Finite Element Simulation of Crash Testing of Laser Welded Joints," Research report, Swedish Institute for Metals, no: IM-2000-062, 2000

[29] Chen, H. F., C. B. Tanner, N. J. Durisek, and D. A. Guenther, "Pole Impact Speeds Derived from Bilinear Estimations of Maximum Crush for Body-

On-Frame Constructed Vehicles," Paper no. 2004-01-1615, Society of Automotive Engineers, Warrendale, Pennsylvania, 2004

[30] C. Lu, Y. Cao, D. Mumford, "Surface Evolution under Curvature Flows", Submitted for the special issue on Partial Differential Equations (PDE's) in Image Processing," Computer Vision, and Computer Graphics, p. 19, 2002

[31] Gray, A. "The Gaussian and Mean Curvatures" and "Surfaces of Constant Gaussian Curvature," §16.5 and Ch. 21 in Modern Differential Geometry of Curves and Surfaces with Mathematica, 2nd ed. Boca Raton, FL: CRC Press, pp. 373-380 and 481-500, 1997

[32] Pollefeys, M., *Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences*, PhD thesis, ESAT-PSI, K.U. Leuven, 1999

[33] W. Förstner, "A Feature Based Correspondence Algorithm for Image Matching," *Int. Arch. Photogramm. Remote Sensing*, Vol. 26, pp. 150-166, 1986

[34] Rövid, A., A. R. Várkonyi-Kóczy, "Corner Detection in Digital Images Using Fuzzy Reasoning," *In Proc. of the 2nd IEEE International Conference on Computational Cybernetics*, August 30-Sept. 1, 2004, Vienna, Austria, pp. 95-99

[35] Várkonyi-Kóczy, A. R., A. Rövid, "Improved Fuzzy Based Corner Detection Method," *In Proc. of the IEEE Int. Workshop on Soft Computing Appl., SOFA'2005*, Szeged-Arad, Hungary-Romania, Aug. 27-30, 2005, pp. 237-242

[36] Várkonyi-Kóczy, A. R., A. Rövid, "Point Correspondence Matching for 3D Reconstruction Using Fuzzy Reasoning," *In Proc. of the 3rd IEEE Int. Conf. on Computational Cybernetics, ICCC 2005*, Mauritius, Apr. 13-16, 2005, pp. 87-92

[37] Rövid, A., A. R. Várkonyi-Kóczy, P. Várlaki, "3D Model Estimation from Multiple Images," In Proc. of the IEEE International Conference on Fuzzy Systems, FUZZ-IEEE'2004, July 25-29, 2004, Budapest, Hungary, Vol. 3, pp. 1661-1666

[38] Rövid, A., T. Hashimoto, A. R. Várkonyi-Kóczy, Y. Shimodaira, "Information Enhancement Method for Image Retrieval and Object Recognition," In Proc. of the 3rd Int. Symposium on Computational Intelligence and Intelligent Informatics, ISCIII 2007, Agadir, Morocco, March 28-30, pp. 25-29

[39] Várkonyi-Kóczy, A. R., A. Rövid, M. G. Ruano, "Soft Computing Based Car Body Deformation and EES Determination for Car Crash Analysis Systems," IEEE Trans. on Instrumentation and Measurement, Vol. 55, No. 4, August 2006

Pseudo-analysis approach to nonlinear partial differential equations

Endre Pap

Department of Mathematics and Informatics, University of Novi Sad
Trg Dositeja Obradovića 4, 21000 Novi Sad, Serbia
e-mail: pape@eunet.yu

Abstract: An overview of methods of pseudo-analysis in applications on important classes of nonlinear partial differential equations, occurring in different fields, is given. Hamilton-Jacobi equations, specially important in the control theory, are for important models usually with non-linear Hamiltonian H which is also not smooth, e.g., the absolute value, min *or* max *operations, where it can not apply the classical mathematical analysis. Using the pseudo-analysis with generalized pseudo-convolution it is possible to obtain solutions which can be interpreted in the mentioned classical way. Another important classes of nonlinear equations, where there are applied the pseudo-analysis, are the Burgers type equations and Black and Shole equation in option pricing. Very recent applications of pseudo-analysis are obtained on equations which model fluid mechanics (Navier-Stokes equation) and image processing (Perona and Malik equation).*

Keywords: Pseudo-analysis, nonlinear partial differential equation, Hamilton-Jacobi equation, Burgers type equation, Bellman differential equation, Navier-Stokes equation, Perona and Malik equation.

1 Introduction

The pseudo-analysis, see [12, 16, 17, 20, 21, 22], is based, instead of the usual field of real numbers, on a semiring acting on the real interval $[a, b] \subset [-\infty, \infty]$, denoting the corresponding operations as $\oplus$ (pseudo-addition) and $\odot$ (pseudo-multiplication), see Section 2. It is applied, as universal mathematical theory, successfully in many fields, e.g., fuzzy systems, decision making, optimization theory, differential equations, etc. This structure is applied for solving nonlinear equations (ODE, PDE, difference equations, etc.) using the pseudo linear principle, which means that if u_1 and u_2 are solutions of the considered nonlinear equation, than also $a_1 \odot u_1 \oplus a_2 \odot u_2$ is a solution for any constants a_1 and a_2 from $[a, b]$. Based on the semiring structure (see [13]) it is developed in

[17, 18, 19, 20, 21, 22, 24] the so called pseudo-analysis in an analogous way as classical analysis, introduced $\oplus$-measure, pseudo-integral, pseudo-convolution, pseudo-Laplace transform, etc. There is so called "viscosity solution" method (see [14]) which gives upper and lower solutions but not a solution in the classical sense, i.e., that its substitution into the equation reduces the equation to the identity. There is given an overview of methods of pseudo-analysis in applications on important classes of nonlinear partial differential equations occurring in different fields, see [7, 8, 12, 16, 18, 19, 20, 21, 22, 24].

First we will show in Section 3 the pseudo linear superposition principle on the Burgers equation and in the limit case on a Hamilton-Jacobi equation. Pseudo-analysis was applied for finding weak solution of Hamilton-Jacobi equation with non-smooth Hamiltonian, [16, 22, 24], see Section 4. Another important class of nonlinear equations, where it is applied the pseudo-analysis, is the Black and Shole equation in option pricing, see Section 6. Very recent applications of pseudo-analysis are obtained on equations which model fluid mechanics, see Section 7. In the section 8 it is presented a general form of PDE for image restoration and there is given a connection with Gaussian linear filtering. The starting PDE in image restoration is the heat equation. Because of its oversmoothing property (edges get smeared), it is necessary to introduce some nonlinearity. Framework to study this equation is nonlinear semigroup theory ([1, 2, 4]). It is proved that Perona and Malik equation satisfy the pseudo linear superposition.

2 Pseudo-analysis

Let $[a,b]$ be closed (in some cases semiclosed) subinterval of $[-\infty,+\infty]$. We consider here a total order $\leq$ on $[a,b]$. The operation $\oplus$ (pseudo-addition) is function $\oplus : [a,b]\times[a,b]\to[a,b]$ which is continuous, commutative, nondecreasing, associative and has a zero element, denoted by $\mathbf{0}$. Let $[a,b]_+ = \{x : x\in[a,b], x\geq \mathbf{0}\}$. The operation $\odot$ (pseudo-multiplication) is a function $\odot : [a,b]\times[a,b]\to[a,b]$ which is continuous, commutative, positively nondecreasing, i.e., $x\leq y$ implies $x\odot z\leq y\odot z, z\in[a,b]_+$,associative and for which there exist a unit element $\mathbf{1}\in[a,b]$, i.e., for each $x\in[a,b], 1\odot x = x$. We suppose $\mathbf{0}\odot x = \mathbf{0}$ and that $\odot$ is a distributive pseudo-multiplication with respect to $\oplus$, i.e.,

$$x\odot(y\oplus z) = (x\odot y)\oplus(x\odot z)$$

The structure $([a,b],\oplus,\odot)$ is called a *semiring* (see [13, 20]). We consider here two special important cases $([0,\infty),\min,+)$ and the g-calculus, i.e., there exists a bijection $g : [a,b]\to[0,\infty]$ such that $x\oplus y = g^{-1}(g(x)+g(y))$ and $x\odot y = g^{-1}(g(x)g(y))$.

There is introduced $\oplus$-measure $m : \mathcal{A}\to[a,b]$ on a σ-algebra $\mathcal{A}$ of subsets of a given set X, and the corresponding pseudo-integral, see [20]. Important cases are $([0,\infty),\min,+)$ and g-calculus, where the corresponding integrals are

given, for $m_\varphi(A) = \inf_x \varphi(x)$ by

$$\int^{\min} f(x)\,dx = \inf_x (f(x) + \varphi(x)),$$

and by

$$\int^{g} f(x)\,dx = g^{-1}\left(\int g(f(x))\,dx\right),$$

respectively.

The *pseudo-character* of group $(G,+)$, $G \subset \mathbb{R}^n$, is a continuous (with respect to the usual topology of reals) map $\xi : G \to [a,b]$, of the group $(G,+)$ into the semiring $([a,b], \oplus, \odot)$, with the property

$$\xi(x+y) = \xi(x) \odot \xi(y), \quad x, y \in G.$$

The map $\xi \equiv \mathbf{0}$ is the trivial pseudo-character. The forms of the pseudo-character in the special cases can be found in [9, 24], where for important cases $([0,\infty), \max, +)$ and g-calculus we have $\xi(x,c) = c \cdot x$ and $\xi(x,c) = g^{-1}(e^{cx})$, respectively, for each $c \in \mathbb{R}$.

Definition 2.1 *The* pseudo-Laplace transform $\mathcal{L}^{\oplus}(f)$ *of a function* $f \in B(G,[a,b])$ *is defined by*

$$(\mathcal{L}^{\oplus} f)(\xi)(z) = \int_{G\cap[0,\infty)^n}^{\oplus} \xi(x,-z) \odot\, dm_f(x),$$

where ξ *is the pseudo-character.*

When at least pseudo-addition is idempotent operation we can consider the second type of pseudo-Laplace transform:

$$(\mathcal{L}^{\oplus} f)(\xi)(z) = \int_{G}^{\oplus} \xi(x,-z) \odot\, dm_f(x),$$

i.e., pseudo-integral has been taken over the whole G.

For the special important cases $([0,\infty), \max, +)$ and g-calculus, we have that the pseudo-Laplace transform has the following form

$$(\mathcal{L}^{\min} f)(z) = \inf_x (-xz + f(x)),$$

and

$$(\mathcal{L}^{g} f)(z) = g^{-1}\left(\int_0^{\infty} e^{-xz} g(f(x))\,dx\right),$$

respectively.

3 Two simpe examples of nonlinear PDE

We start with two examples to illustrate how can be applied the pseudo-linear superposition principle on some non-linear partial differential equations.

An important nonlinear partial differential equation is the Burgers equation for a function $u = u(x,t)$. Burgers (1948), Hopf (1950) and Cole (1951) investigated as a model of turbulence the following equation

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = \frac{c}{2}\frac{\partial^2 v}{\partial x^2}, \tag{1}$$

where c is a parameter. Putting $v = \frac{\partial u}{\partial x}$ in (1) and integrating with respect to x we obtain the equation

$$\frac{\partial u}{\partial t} + \frac{1}{2}\left(\frac{\partial u}{\partial x}\right)^2 - \frac{c}{2}\frac{\partial^2 u}{\partial x^2} = 0, \tag{2}$$

for $x \in \mathbb{R}$ and $t > 0$, with the initial condition $u(x,0) = u_0(x)$, where c is the given positive constant, and which models the burning of a gas in a rocket. We shall apply on this equation the g-calculus, with the generator $g(u) = e^{-u/c}$. Then, the corresponding pseudo-addition is $u \oplus v = -c\ln(e^{-u/c} + e^{-v/c})$, and the distributive pseudo-multiplication $u \odot v = u + v$. Then for solutions u_1 and u_2 of (2) the function $(\lambda_1 \odot u_1) \oplus (\lambda_2 \odot u_2)$ is also a solution of Burgers equation (2). The solution of the given initial problem is

$$u(x,t) = \frac{c}{2}\ln(2\pi ct) \odot \int^{\oplus} \frac{(x-s)^2}{2t} \odot u_0(s)\, ds.$$

Taking $c \to 0$ in the Burgers equation (2) we obtain Hamilton-Jacobi equation

$$\frac{\partial u}{\partial t} + \frac{1}{2}\left(\frac{\partial u}{\partial x}\right)^2 = 0.$$

Then for solutions u_1 and u_2 the function $(\lambda_1 \odot u_1) \oplus (\lambda_2 \odot u_2)$, where

$$u \oplus v = \min(u,v) \text{ and } u \odot v = u + v,$$

is also a solution of the preceding Hamilton-Jacobi equation.

4 Hamilton-Jacobi equation with non-smooth Hamiltonian

We consider here the nonlinear PDE, so called Hamilton-Jacobi-Bellman equation

$$\frac{\partial u(x,t)}{\partial t} + H\left(\frac{\partial u}{\partial x}, x, t\right) = 0, \tag{3}$$

see [12, 16, 20, 21, 22, 24]. Hamilton-Jacobi equations are specially important in the control theory. Unfortunately, usually the interesting models are represented by Hamilton-Jacobi equations in which the non-linear Hamiltonian H is not smooth, for example the absolute value, min or max operations. Hence we can not apply on such cases the classical mathematical analysis. There is so called "viscosity solution" method (see [14]) which gives upper and lower solutions but not a solution in the classical sense, i.e., that its substitution into the equation reduces the equation to the identity. Using the pseudo-analysis with generalized pseudo-convolution it is possible to obtain solutions which can be interpreted in the mentioned classical way.

We extend now the pseudo-superposition principle to a more general case, see [12, 21, 22].

Theorem 4.1 *If u_1 and u_2 are solutions of the Hamilton-Jacobi equation (3), where $H \in C(\mathbb{R}^{n+2})$ and $\frac{\partial u}{\partial x}$ is the gradient of u, then $(\lambda_1 \odot u_1) \oplus (\lambda_2 \odot u_2)$ is also a solution of the Hamilton-Jacobi equation (3), with respect to the operations $\oplus = \min$ and $\odot = +$.*

Let $C_{\mathbf{0}}(\mathbb{R}^n)$ be the space of continuous functions $f : \mathbb{R}^n \to P$ (P is of type $(\min, +)$ or $(\min, \max)$) with the property that for each $\varepsilon > 0$ there exists a compact subset $K \subset \mathbb{R}^n$ such that $d(\mathbf{0}, \inf_{x \in \mathbb{R}^n \setminus K} f(x)) < \varepsilon$, with the metric $D(f, g) = \sup_x d(f(x), g(x))$. Let $C_{\mathbf{0}}^{cs}(\mathbb{R}^n)$ be the subspace of $C_{\mathbf{0}}(\mathbb{R}^n)$ of functions f with compact support $\operatorname{supp}_{\mathbf{0}} = \overline{\{x |\ f(x) \neq \mathbf{0}\}}$. The dual semimodul $(C_{\mathbf{0}}(\mathbb{R}^n))^*$ is the semimodul of continuous pseudo-linear P-valued functionals on $C_{\mathbf{0}}(\mathbb{R}^n)$ (with respect to pointwise operations). Analogously the dual semimodul $(C_{\mathbf{0}}^{cs}(\mathbb{R}^n))^*$ is the semimodul of continuous pseudo-linear P-valued functionals on $C_{\mathbf{0}}^{cs}(\mathbb{R}^n)$ (with respect to pointwise operations). We shall need the following representation theorem, see [12].

Theorem 4.2 *Let f be a function defined on $\mathbb{R}^n$ and with values in the semiring P of type $(\min, +)$ or $(\min, \max)$, and a functional $m_f : C_{\mathbf{0}}^{cs}(\mathbb{R}^n) \to P$ is given by*

$$m_f(h) = \int^{\oplus} f \odot dm_h = \inf_x (f(x) \odot h(x)).$$

Then

1) *The mapping $f \mapsto m_f$ is a pseudo-isomorphism of the semimodule of lower semicontinuous functions onto the semimodule $(C_{\mathbf{0}}^{cs}(\mathbb{R}^n))^*$.*

2) *The space $C_{\mathbf{0}}^*(\mathbb{R}^n)$ is isometrically isomorphic with the space of bounded functions, i.e., for every $m_{f_1}, m_{f_2} \in C_{\mathbf{0}}^*(\mathbb{R}^n)$ we have*

$$\sup_x d(f_1(x), f_2(x))$$

$$= \sup\{d(m_{f_1}(h), m_{f_2}(h)) : h \in C_{\mathbf{0}}(\mathbb{R}^n), D(h, \mathbf{0}) \leq 1\}.$$

3) The functionals m_{f_1} and m_{f_2} are equal if and only if $Clf_1 = Clf_2$, where

$$Clf(x) = \sup\{\psi(x) : \psi \in C(\mathbb{R}^n), \psi \leq f\}.$$

We consider now the following *Cauchy problem for Hamilton-Jacobi(- Bellman) equation*

$$\frac{\partial u}{\partial t} + H\left(\frac{\partial u}{\partial x}\right) = 0, \quad u(x,0) = u_0(x), \tag{4}$$

where $x \in \mathbb{R}^n$, and the function $H : \mathbb{R}^n \to \mathbb{R}$ is convex (by boundedness of H it is also continuous). For control theory the important examples of the Hamiltonian H are non-smooth functions, e.g., max and $|.|$. The approach with pseudo-analysis avoids the use of the so called "viscosity solution" method, which does not give the exact solution of (4) (see [14]). We apply now the methods of pseudo-analysis. For that purpose we define the family of operators $\{R_t\}_{t>0}$, for a function $u_0(x)$ bounded from below in the following way

$$u(t,x) = (R_t u_0)(x) = \inf_{z \in \mathbb{R}^n} (u_0(z) - t\mathcal{L}^{\min}(H)(\frac{x-z}{t})), \tag{5}$$

where $\mathcal{L}$ is considered on the whole $\mathbb{R}^n$. The operator R_t is pseudo-linear with respect to $\oplus = \min$ and $\odot = +$, where $\mathcal{L}^{\oplus}(H)(q) = \inf_{p \in \mathbb{R}^n} (-pq + H(p))$.

First we suppose that u_0 is smooth and strongly convex. We shall use the notations $< x, y >$ and $\|x\|$ for the scalar product and Euclidean norm in $\mathbb{R}^n$, respectively. For a function $F : \mathbb{R}^n \to [-\infty, +\infty]$ its *subgradient* at a point $u \in \mathbb{R}^n$ is a point $w \in \mathbb{R}^n$ such that $F(u)$ is finite and

$$< w, v - u > + F(u) \leq F(v)$$

for all $v \in \mathbb{R}^n$. Then we have by [12].

Lemma 4.3 *Let $u_0(x)$ be smooth and strongly convex and there exists $\delta > 0$ such that for all x the eigenvalues of the matrix $u_0''(x)$ of all second derivatives are not less than δ. Then*

1) *For every $x \in \mathbb{R}^n$, $t > 0$,there exists a unique $\xi(t,x) \in \mathbb{R}^n$ such that $\frac{x-\xi(t,x)}{t}$ is a subgradient of the function H at the point $u_0'(\xi(t,x))$ and*

$$(R_t u_0)(x) = u_0(\xi(t,x)) - t\mathcal{L}^{\min}(H)(\frac{x-\xi(t,x)}{t}).$$

2) *The function $\xi(t,x)$ for $t > 0$ satisfies the Lipschitz condition on compact sets, and $\lim_{t \to 0} \xi(t,x) = x$.*

3) *The Cauchy problem (4) has a unique C^1 solution given by (4.3), and*

$$\frac{\partial u}{\partial x}(t,x) = u_0'(\xi(t,x)).$$

The Cauchy problem

$$\frac{\partial u}{\partial t} + H\left(-\frac{\partial u}{\partial x}\right) = 0, \qquad (6)$$
$$u(0,x) = u_0(x),$$

is the adjoint problem of the Cauchy problem (4). The classical resolving operator R_t^* of the Cauchy problem (6) on the smooth convex functions by Lemma 4.3 is given by

$$(R_t^* u_0)(x) = \inf_{\xi}(u_0(\xi) - t\mathcal{L}^{\min}(H)(\frac{\xi - x}{t})).$$

We note that R_t^* is the *adjoint of the resolving operator* R_t with respect to bipseudo-linear functional

$$\int_{\mathbb{R}^n}^{\oplus} f \odot h \, dm.$$

Then we can introduce, as in the theory of linear equation, the notion of generalized weak solution (using Theorem 4.2), see [12].

Definition 4.4 *Let u_0 be a bounded from below function $u_0 : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and m_{u_0} the corresponding functional from $C_{\mathbf{0}}^*(\mathbb{R}^n)$. The generalized weak pseudo solution of Cauchy problem (4) is a continuous function from below $(R_t u_0)(x)$ which is defined uniquely by*

$$m_{R_t u_0}(\varphi) = m_{u_0}(R_t^* \varphi)$$

for all smooth convex functions φ.

We can construct the solution for the case when u_0 is a smooth strictly convex function by Lemma 4.3. Then it follows by Theorem 4.2 and Definition 4.4.

Theorem 4.5 *For an arbitrary function $u_0(x)$ bounded from below the weak pseudo-solution of the Cauchy problem (4) is given by*

$$(R_t u_0)(x) = (R_t C l u_0)(x) = \inf_z (C l u_0(z) + t\mathcal{L}^{\min}(H)(\frac{x - z}{t})),$$

where

$$Clf(x) = \sup\{\psi(x) : \psi \in C(\mathbb{R}^n), \psi \leq f\}.$$

5 Bellman differential equation for multicriteria optimization problems

We present results from [12] obtained for the controlled process in $\mathbb{R}^n$ specified by a controlled differential equation $\dot{x} = f(x, v)$ (where v belongs to a metric

control space V) and by a continuous function $\varphi \in B(\mathbb{R}^n \times V, \mathbb{R}^k)$, which determines a vector-valued integral criterion

$$\Phi(x(\,\cdot\,)) = \int_0^t \varphi(x(\tau), u(\tau))\, d\tau$$

on the trajectories. Let us pose the problem of finding the Pareto set $\omega_t(x)$ for a process of duration t issuing from x with terminal set determined by some function $\omega_0 \in B(\mathbb{R}^n, \mathbb{R}^k)$, that is,

$$\omega_t(x) = \operatorname{Min} \bigcup_{x(\cdot)} (\Phi(x(\,\cdot\,)) \odot \omega_0(x(t))), \tag{7}$$

where $x(\,\cdot\,)$ ranges over all admissible trajectories issuing from x. We can encode the functions $\omega_t \in B(\mathbb{R}^n, P\mathbb{R}^k)$ by the functions

$$u(t, x, a)\colon \mathbb{R}_+ \times \mathbb{R}^n \times L \to \mathbb{R}.$$

The optimality principle permits us to write out the following equation, which is valid modulo $O(\tau^2)$ for small τ:

$$u(t, x, a) = \operatorname{Min}_v (h_{\tau\varphi(x,v)} \star u(t - \tau, x + \Delta x(v)))(a).$$

It follows from the representation of $h_{\tau\varphi(x,v)}$ and from the fact that n is, by definition, the multiplicative unit in $CS_n(L)$ that

$$u(t, x, a) = \min_v (\tau\overline{\varphi}(x, v) + u(t - \tau, x + \Delta x(v), a - \tau\varphi_L(x, v))).$$

Let us substitute $\Delta x = \tau f(x, v)$ into this equation, expand S in a series modulo $O(\tau^2)$, and collect similar terms. Then we obtain the equation

$$\frac{\partial u}{\partial t} + \max_v \left(\varphi_L(x, v) \frac{\partial u}{\partial a} - f(x, v) \frac{\partial u}{\partial x} - \overline{\varphi}(x, v) \right) = 0. \tag{8}$$

Although the presence of a vector criterion has resulted in a larger dimension, this equation coincides in form with the usual Bellman differential equation. Consequently, the generalized solutions can be defined on the basis of the idempotent superposition principle, as Section 4. We have the following result by [12].

Theorem 5.1 *The Pareto set $\omega_t(x)$ (7) is determined by a generalized solution $u_t \in B(\mathbb{R}^n, CS_n(L))$ of (8) with the initial condition $u_0(x) = h_{\omega_0(x)} \in B(\mathbb{R}^n, CS_n(L))$. The mapping $R_{CS}\colon u_0 \mapsto u_t$ is a linear operator on $B(\mathbb{R}^n, CS_n(L))$.*

6 Option pricing

Black-Sholes and Cox-Ross-Rubinstein formulas are basic results in the modern theory of option pricing in financial mathematics. They are usually deduced

by means of stochastic analysis; various generalizations of these formulas were proposed using more sophisticated stochastic models for common stocks pricing evolution. The systematic deterministic approach to the option pricing leads to a different type of generalizations of Black-Sholes and Cox-Ross-Rubinstein formulas characterized by more rough assumptions on common stocks evolution (which are therefore easier to verify). This approach reduces the analysis of the option pricing to the study of certain homogeneous nonexpansive maps, which however, unlike the situations described in previous subsections, are "strongly" infinite dimensional: they act on the spaces of functions defined on sets, which are not (even locally) compact.

In the paper of [11] it was shown what type of generalizations of the standard Cox-Ross-Rubinstein and Black-Sholes formulas can be obtained using the deterministic (actually game-theoretic) approach to option pricing and what class of homogeneous nonexpansive maps appear in these formulas, considering first a simplest model of financial market with only two securities in discrete time, then its generalization to the case of several common stocks, and then the continuous limit. One of the objective was to show that the infinite dimensional generalization of the theory of homogeneous nonexpansive maps (which does not exists at the moment) would have direct applications to the analysis of derivative securities pricing. On the other hand, this approach, which uses neither martingales nor stochastic equations, makes the whole apparatus of the standard game theory appropriate for the study of option pricing.

7 Navier-Stokes and Stokes equations

Pseudo liner superposition principle was applied also on important equations of fluid mechanics [27]. We consider an incompressible homogeneous viscous flow: that means that div $\mathbf{u} = 0$, for the density $\rho = 1$, ν is the coefficient of viscosity, for the forces $\mathrm{f} = 0$. The equations of motion of this flow are the *Navier-Stokes equations*, see [6]:

$$\rho \frac{D\mathbf{u}}{Dt} = - \text{ grad p} + \nu\Delta\mathbf{u}$$
$$\text{div } \mathbf{u} = 0$$
$$\mathbf{u} = 0 \quad \text{on } \partial D$$

where $\Delta\mathbf{u}$ is the Laplacian of the velocity u, defined in this way: $\Delta\mathbf{u} = (\partial_{xx} + \partial_{yy})\mathbf{u} = (\partial_{xx}u + \partial_{yy}v)$, as $\mathbf{u}(\mathbf{x}, t) = (u(x, y, t), v(x, y, t))$.

We consider two-dimensional incompressible flow in the upper half plane $y > 0$; so the projections of the Navier-Stokes equations on axes x and y are the following:

$$\partial_t u + u\partial_x u + v\partial_y u + \partial_x p + \nu(\partial_{xx}u + \partial_{yy}u) = 0 \tag{9}$$

$$\partial_t v + u\partial_x v + v\partial_y v + \partial_y p + \nu(\partial_{xx}v + \partial_{yy}v) = 0 \tag{10}$$

$$\partial_x u + \partial_y v = 0 \tag{11}$$

$$u = v = 0 \qquad \text{on} \qquad \partial D. \tag{12}$$

We have proved in [27] the following two theorems.

Theorem 7.1 *Let $\mathbf{s}_{i,p} = (u_i, v_i, p), i = 1, 2$, be two solutions of (9) - (12) and a_1, a_2 two real numbers. Then the pseudo-linear combination*

$$(a_1 \odot \mathbf{s}_{1,p}) \oplus (a_2 \odot \mathbf{s}_{2,p}) = \mathbf{min}(\mathbf{max}(a_1, \mathbf{s}_{1,p)}, \mathbf{max}(a_2, \mathbf{s}_{2,p}))$$

is again a solution of (9) - (12).

Theorem 7.2 *Let $\mathbf{s}_{i,p} = (u_i, v_i, p), i = 1, 2$, be two solutions of (9) - (12) which satisfy*

$$\partial_y u_i = \partial_y v_i \qquad i = 1, 2.$$

Then the pseudo-linear combination $(a_1 \odot \mathbf{s}_{1,p}) \oplus (a_2 \odot \mathbf{s}_{2,p})$, for two real numbers a_1, a_2 where $\odot$ is given by

$$\lambda \odot \mathbf{s} = \lambda \odot (u, v, p) = (\lambda + u, \lambda + v, \lambda + p),$$

is again a solution of (9) - (12).

The *Stokes equations* approximate equations for incompressible flow ([5]):

$$\partial_t \mathbf{u} + \text{ grad p} + \nu \Delta \mathbf{u} = 0 \tag{13}$$

$$\text{div } \mathbf{u} = 0 \tag{14}$$

We have proved in [27] the following theorem.

Theorem 7.3 *Let $\mathbf{s}_i(t) = (u_i(t), v_i(t), p_i(t))$, $i = 1, 2$ be solutions of (13) and (14). Then the pseudo-linear combination $(a_1 \odot \mathbf{s}_1) \oplus (a_2 \odot \mathbf{s}_2)$, for two real numbers a_1, a_2 where $\oplus$ is given by $(\mathbf{s}_1 \oplus \mathbf{s}_2)$*

$$= (g^{-1}(g(u_1) + g(u_2), g^{-1}(g(v_1) + (g(v_2), g^{-1}(g(p_1) + g(p_2)),$$

and

$$a \odot \mathbf{s} = ((g^{-1}(g(a) \cdot g(u), g^{-1}(g(a) \cdot g(v), (g^{-1}(g(a) \cdot g(p))$$
$$= (a + u, a + v, a + p)$$

with g defined by $g(a) = e^{-\ c\ a}, c > 0$ and $g^{-1}(b) = -\frac{1}{c}\ \log\ b$, is again solution of (13) - (14).

8 Pseudo-linear superposition principle for Perona and Malik equation

Partial differential equations are applied for image processing ([1, 3, 28]). In that method a restored image can be seen as a version of the initial image at a special scale. An image u is embedded in an evolution process, denoted by $u(t,\cdot)$. The original image is taken at time $t=0$, $u(0,\cdot)=u_0(\cdot)$. The original image is then transformed, and this process can be written in the form $\frac{\partial u}{\partial t}(t,x)+F\left(x,u(t,x),\nabla u(t,x),\nabla^2 u(t,x)\right)=0$ in Ω. Some possibilities for F to restore an image are considered in [1]. PDE-methods for restoration is in general form:

$$\begin{cases} \frac{\partial u}{\partial t}(t,x)+F\left(x,u(t,x),\nabla u(t,x),\nabla^2 u(t,x)\right)=0 \text{ in } (0,T)\times\Omega, \\ \frac{\partial u}{\partial N}(t,x)=0 \text{ on } (0,T)\times\partial\Omega, \quad u(0,x)=u_0(x), \end{cases} \tag{15}$$

where $u(t,x)$ is the restored version of the initial degraded image $u_0(x)$. The idea is to construct a family of functions $\{u(t,x)\}_{t>0}$ representing successive versions of $u_0(x)$. As t increases $u(t,x)$ changes into a more and more simplified image. We would like to attain two goals. The first is that $u(t,x)$ should represent a smooth version of $u_0(x)$, where the noise has been removed. The second, is to be able to preserve some features such as edges, corners, which may be viewed as singularitis. The basic PDE in image restoration is the heat equation:

$$\begin{cases} \frac{\partial u}{\partial t}(t,x)-\Delta u(t,x)=0, t\geq 0, x\in\mathbb{R}^2, \\ u(0,x)=u_0(x). \end{cases} \tag{16}$$

We consider that $u_0(x)$ is primarily defined on the square $[0,1]^2$. We extend it by symmetry to $C=[-1,1]^2$, and then on all $\mathbb{R}^2$, by periodicity. This way of extending $u_0(x)$ is classical in image processing. If $u_0(x)$ is extended in this way and satisfies in addition $\int_C |u_0(x)|\,dx<+\infty$, we will say that $u_0\in L^1_{\#}(C)$ (see [1]). Solving (16) is equivalent to carrying out a Gaussian linear filtering, which was widely used in signal processing. If $u_0\in L^1_{\#}(C)$, then the explicit solution of (16) is given by

$$u(t,x)=\int_{\mathbb{R}^2} G_{\sqrt{2t}}(x-y)\,u_0(y)\,dy=\left(G_{\sqrt{2t}}*u_0\right)(x),$$

where $G_\sigma(x)$ denotes the two-dimensional Gaussian kernel

$$G_\sigma(x)=\frac{1}{2\pi\sigma}e^{-\frac{|x|^2}{2\sigma^2}}$$

The heat equation has been (and is) successfully applied in image processing but it has some drawback. It is too smoothing and because of that edges can be lost or severely blurred. In [1] authors consider models that are generalizations of the heat equation. The domain image will be a bounded open set Ω of $\mathbb{R}^2$.

The following equation is initially proposed by Perona and Malik [28]:

$$\begin{cases} \frac{\partial u}{\partial t} = \operatorname{div}\left(c\left(|\nabla u|^2\right)\nabla u\right) & \text{in } (0,T)\times\Omega, \\ \frac{\partial u}{\partial N} = 0 \text{ on } (0,T)\times\partial\Omega, \\ u(0,x) = u_0(x) \text{ in } \Omega \end{cases} \tag{17}$$

where $c:[0,\infty)\to(0,\infty)$. If we choose $c\equiv 1$, then it is reduced on the heat equation. If we assume that $c(s)$ is a decreasing function satisfying $c(0)=1$ and $\lim_{s\to\infty} c(s)=0$, then inside the regions where the magnitude of the gradient of u is weak, equation (17) acts like the heat equation and the edges are preserved. For each point x where $|\nabla u|\neq 0$ we can define the vectors $N=\frac{\nabla u}{|\nabla u|}$ and T with $T\cdot N=0$, $|T|=1$. For the first and second partial derivatives of u we use the usual notation u_{x_1}, u_{x_2}, $u_{x_1x_1}$,... We denote by u_{NN} and u_{TT} the second derivatives of u in the T-direction and N-direction, respectively:

$$\begin{aligned} u_{TT} &= T^t\nabla^2 uT = \frac{1}{|\nabla u|^2}\left(u_x^2u_{yy}+u_y^2u_{xx}-2u_xu_yu_{xy}\right), \\ u_{NN} &= N^t\nabla^2 uN = \frac{1}{|\nabla u|^2}(u_x^2u_{xx}+u_y^2u_{yy}+2u_xu_yu_{xy}). \end{aligned}$$

The first equation in (17) can be written as

$$\frac{\partial u}{\partial t}(t,x) = c\left(|\nabla u(t,x)|^2\right)u_{TT} + b\left(|\nabla u(t,x)|^2\right)u_{NN}, \tag{18}$$

where $b(s)=c(s)+2sc'(s)$. Therefore, (18) is a sum of a diffusion in the T-direction and a diffusion in the N-direction. The function c and b act as weighting coefficients. Since N is normal to the edges, it would be preferable to smooth more in the tangential direction T than in the normal direction. Because of that we impose

$$\lim_{s\to\infty}\frac{b(s)}{c(s)}=0 \quad \text{or} \quad \lim_{s\to\infty}\frac{sc'(s)}{c(s)}=-\frac{1}{2} \tag{19}$$

If $c(s)>0$ with power growth, then (19) implies that $c(s)\approx 1/\sqrt{s}$ as $s\to\infty$. The equation (17) is parabolic if $b(s)>0$. The assumptions imposed on $c(s)$ are

$$\begin{cases} c:[0,\infty)\to(0,\infty) \text{ decreasing}, \\ c(0)=1,\ c(s)\approx\frac{1}{\sqrt{s}} \text{ as } s\to\infty, \\ b(s)=c(s)+2sc'(s)>0. \end{cases} \tag{20}$$

Often used function $c(s)$ satisfying (20) is $c(s)=\frac{1}{\sqrt{1+s}}$. Because of the behavior $c(s)\approx 1/\sqrt{s}$ as $s\to\infty$, it is not possible to apply general results from parabolic equations theory. Framework to study this equation is nonlinear semigroup theory (see [1, 2, 4]).

We have proved in [25] that the pseudo-linear superposition principle holds for Perona and Malik equation.

Theorem 8.1 *If* $u_1 = u_1(t,x)$ *and* $u_2 = u_2(t,x)$ *are solutions of the equation*

$$\frac{\partial u}{\partial t} - div\left(c\left(|\nabla u|^2\right)\nabla u\right) = 0, \tag{21}$$

then $u_1 \oplus u_2$ *is also a solution of (21) on the set*

$$D = \{(t,x)\,|t \in (0,T)\,, x \in \mathbb{R}^2, u_1(t,x) \neq u_2(t,x)\},$$

with respect to the operation $\oplus = \min$.

The obtained results will serve for further investigation of the weak solutions of the equation (21) in the sense of Maslov [10, 12, 22, 23] and Gondran [7, 8], as well as their important applications.

9 Conclusion

The pseudo-linear superposition principle, as it was shown, allows us to transfer the methods of linear equations to many important nonlinear partial differential equations. Some further developments related more general pseudo-operations with applications on nonlinear partial differential equations were obtain in [22, 23, 26].

Acknowledgment

The author would like to thank for the support in part by the project MNZŽSS 144012, grant of MTA of HTMT, French-Serbian project "Pavle Savić", and by the project "Mathematical Models for Decision Making under Uncertain Conditions and Their Applications" of Academy of Sciences and Arts of Vojvodina supported by Provincial Secretariat for Science and Technological Development of Vojvodina.

References

[1] G. Aubert, P. Kornprobst, *Mathematical Problems in Image Processing*, Springer-Verlag, 2002.

[2] H. Breyis, *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert*, North-Holland Publishing Comp, Amsterdam-London, 1973.

[3] F. Catte, P.L. Lions, J.M. Morel, T. Coll, *Image selective smoothing and edge detection by nonlinear diffusion*, SIAM Journal of Numerical Analysis, 29(1):182-193, 1992.

[4] T. Cazenave, A. Haraux, *Introduction aux Problemes d'Evolution Semi-Linéares*, (Introduction to Semilinear Evolution Problems), Mathématiques & Apllications, Ellipses, 1990.

[5] A.J.Chorin, J.Marsen, *A Mathamatical Intoduction to Fluid Mechanics*, Springer-Verlag, (1993).

[6] R.Deutray, J-L.Lions, *Mathematical Analysis and Numerical Methods for Science and Tecnology*, **vol.4,6**, Springer-Verlag, 2000.

[7] M. Gondran, *Analyse MINPLUS*, Analyse fonctionnelle/Functional Analysis, C. R. Acad. Sci. Paris, t. 323, Série I, p. 371-375, 1996.

[8] M. Gondran, M. Minoux, *Graphes, dioides et semi-anneaux*, Editions TEC & DOC, Londres- Paris- New York, 2001.

[9] E. P. Klement, R. Mesiar, E. Pap, *Triangular Norms.* Kluwer Academic Publishers, Dordrecht, 2000.

[10] V. N. Kolokoltsov, V. P. Maslov, *Idempotent calculus as the apparatus of optimization theory. I*, Functional. Anal. i Prilozhen 23, no. 1, (1989), 1-14. Kolokoltsov, V.N.

[11] V. N. Kolokoltsov, *Nonexpansive maps and option pricing theory. Kibernetika* 34(6) (1998), 713-724.

[12] V. N. Kolokoltsov, V. P. Maslov, *Idempotent Analysis and Its Applications,* Kluwer Academic Publishers, Dordrecht, Boston, London, 1997.

[13] W. Kuich: *Semirings, Automata, Languages*, Berlin, Springer-Verlag, 1986.

[14] P. L. Lions, *Generalized solutions of Hamilton-Jacobi equations.* London, Pitman, 1982.

[15] G. L. Litvinov, *The Maslov Dequantization, Idempotent and Tropical Mathematics: a very Brief Introduction*, Cont. Mathematics 377, AMS, (2005), 1-17.

[16] V. P. Maslov, S.N. Samborskij (eds.), *Idempotent Analysis*, Advances in Soviet Mathematics 13, Providence, Rhode Island, Amer. Math. Soc.., 1992.

[17] E. Pap, *An integral generated by decomposable measure*, Univ. u Novom Sadu Zb. Rad. Prirod.-Mat. Fak. Ser. Mat. 20 (1) (1990), 135-144.

[18] E. Pap, *g-calculus*, Univ. u Novom Sadu Zb. Rad. Prirod.-Mat. Fak. Ser. Mat. 23 (1) (1993), 145-156.

[19] E. Pap, *Applications of decomposable measures*, in Handbook Mathematics of Fuzzy Sets-Logic, Topology and Measure Theory (Ed. U. Höhle, R.S. Rodabaugh), Kluwer Academic Publishers, 1999, 675-700.

[20] E. Pap, *Null-Additive Set Functions*, Kluwer Academic Publishers, Dordrecht- Boston-London, 1995.

[21] E. Pap, *Decomposable measures and nonlinear equations*, Fuzzy Sets and Systems 92 (1997) 205-222.

[22] E. Pap, *Pseudo-Additive Measures and Their Aplications*, Handbook of Measure Theory (Ed. E. Pap), Elsevier, Amsterdam, 2002, 1403-1465,

[23] E. Pap, *Applications of the generated pseudo-analysis on nonlinear partial differential equations*, Proceedings of the Conference on Idempotent Mathematics and Mathematical Physics (Eds G.L. Litvinov, V.P. Maslov), *Contemporary Mathematics* 377, American Mathematical Society, 2005, 239-259.

[24] E. Pap, N. Ralević, *Pseudo-Laplace transform*, Nonlinear Analysis 33 (1998) 553-560.

[25] E. Pap, M. Štrboja, *Image processing based on a partial differential equation satisfying the superposition principle*, Idempotent and Tropical Mathematics and Problems of Mathematical Physics, Volume II, Moscow, August 25-30, 2007, 38-42.

[26] E. Pap, D. Vivona, *Non-commutative and associative pseudo-analysis and its applications on nonlinear partial differential equations*, J. Math. Anal. Appl. 246 (2) (2000) 390-408.

[27] E. Pap, D. Vivona, *Non-linear superposition principle in fluid mechanics: Euler equations, Prandtl equations, Navier-Stokes equations* (preprint), 2004.

[28] P. Peron, J. Malik, *Scale-space and edge detection using anisotropic diffusion*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(7): 629-639, 1990.

A recursive solution concept for multichoice games

Fabien Lange, Michel Grabisch

Université Paris I - Panthéon-Sorbonne
CERMSEM, 106-112 bd de l'Hôpital
75013 Paris, France
E-mail: Fabien.Lange@univ-paris1.fr, Michel.Grabisch@lip6.fr

Abstract: We propose a new axiomatization of the Shapley value for cooperative games, where symmetry and efficiency can be discarded and replaced with new natural axioms. From any game, an excluded-player game is built by discarding all coalitions that contain a fixed player. Then it is shown that the Shapley value is the unique value satisfying the linearity axiom, the nullity axiom, the excluded-null-player axiom, and the equity axiom. In the second part, by generalizing the above material, the Shapley value for multichoice games is worked out.

Keywords: Shapley value, multichoice games, equity, generalized nullity axiom

1 Introduction

The value or solution concept of a game is a key concept in cooperative game theory, since it defines a rational imputation given to the players if they join the grand coalition. In this respect, the Shapley value remains the best known solution concept [11], and it has been axiomatized by many authors in various ways (see especially Weber [13], or the survey by Monderer and Samet [8]).

If the definition and axiomatization of the Shapley value is well established for classical cooperative TU-games, the situation is less clear when considering variants of classical TU-games, like multichoice games [7], games in partition function form [12], etc. In this paper, we focus on multichoice games, where players are allowed to have several (and totally ordered) levels of participation. Hence, a solution for multichoice games assigns a numerical value to each possible participation level and to each player. The original proposal of Hsiao and Raghavan [7] for the Shapley value has been, up to our knowledge, scarcely used due to its complexity. Another proposal is due to Faigle and Kern [5], and compared to the former one by Branzei et al. [3], and also by the authors [6]. The value proposed by Faigle and Kern, although elegant but still with a very high computational complexity, is more rooted in combinatorics than in game theory, and takes as a basis the expression of the Shapley value using maximal chains in the lattice of coalitions. In [6], we proposed an alternative

view having the same complexity than the usual Shapley value for classical TU-games. It turned out that our value is identical to the egalitarian value proposed by Peters and Zank [10], although they use different axioms and impose some restrictions (namely, all players should have the same set of participation levels).

Although close to the axiomatization proposed by Weber for classical TU-games, our axiomatization in [6] suffered from a complex symmetry axiom, hard to interpret, the fundamental problem there being that the classical notion of symmetry among players cannot hold since two different players may have a different set of participation levels (note that this difficulty was avoided by Peters and Zank, since they considered multichoice games with all players having the same set of participation levels).

In this paper, we propose a new axiomatization for the so-called egalitarian value, which is based essentially on carriers and on a recursive scheme, and which does not make use of a symmetry axiom. In Section 3, we present the main ideas applied on classical TU-games, and we come up with a very simple and natural axiomatization using linearity, a nullity axiom which uses also carriers, and an equity axiom stating that the sharing should be uniform and efficient for the unanimity game based on the grand coalition (this is in fact a very weak version of the efficiency axiom). In Section 4, the same process is applied to multichoice games. An additional axiom (called decreased level axiom) is used, to take into account the case where a player does not participate at the highest level.

In the sequel, $\mathbb{N}$ refers to the set of positive integers. In order to avoid a heavy notation, we will often omit braces for subsets, by writing i instead of $\{i\}$ or 123 for $\{1, 2, 3\}$. Furthermore, cardinalities of subsets $S, T, \ldots$ will be denoted by the corresponding lower case letters $s, t, \ldots$

2 Mathematical background

We begin by recalling necessary material on lattices (a good introduction on lattices can be found in [4]), in a finite setting. A *lattice* is a set L endowed with a partial order $\leq$ such that for any $x, y \in L$ their least upper bound $x \vee y$ and greatest lower bound $x \wedge y$ always exist. For finite lattices, the greatest element of L (denoted $\top$) and least element $\bot$ always exist. x *covers* y (denoted $x \succ y$) if $x > y$ and there is no z such that $x > z > y$. A *ranked lattice* is a pair (L, r), where L is a lattice and the rank function $r : L \to \mathbb{N}$ satisfies the property that $r(y) = r(x)+1$ whenever y covers x in L. The lattice is *distributive* if $\vee, \wedge$ obey distributivity. An element $j \in L$ is *join-irreducible* if it cannot be expressed as a supremum of other elements. Equivalently j is join-irreducible if it covers only one element. The set of all join-irreducible elements of L is denoted $\mathcal{J}(L)$.

An important property is that in a distributive lattice, any element x can be written as an irredundant supremum of join-irreducible elements in a unique way (this is called the *minimal decomposition* of x). We denote by $\eta^*(x)$ the set of join-irreducible elements in the minimal decomposition of x, and we denote by $\eta(x)$ the *normal decomposition* of x, defined as the set of join-irreducible elements smaller or equal to x, i.e., $\eta(x) := \{j \in \mathcal{J}(L) \mid j \leq x\}$. Let us rephrase differently the above result. We say that $Q \subseteq L$ is a

downset of L if $x \in Q$ and $y \leq x$ imply $y \in Q$. For any subset P of L, we denote by $\mathcal{O}(P)$ the set of all downsets of P. Then, by Birkhoff's theorem [2], the mapping η is an isomorphism of L onto $\mathcal{O}(\mathcal{J}(L))$.

Given lattices $(L_1, \leq_1), \ldots, (L_n, \leq_n)$, the product lattice $L = L_1 \times \cdots \times L_n$ is endowed with the product order $\leq$ of $\leq_1, \ldots, \leq_n$ in the usual sense. Elements of L can be written in their vector form $(x_1, \ldots, x_n)$. The set L_{-i} denotes $\prod_{k \neq i} L_k$ if $n > 1$, and the singleton set $\{()\}$ otherwise. By this way, for any vector x, $\big((), x\big)$ simply denotes x. All join-irreducible elements of L are of the form $(\bot_1, \ldots, \bot_{i-1}, j_i, \bot_{i+1}, \ldots, \bot_n)$, for some i and some join-irreducible element j_i of L_i. A *vertex* of L is any element whose components are either top or bottom. We denote $\Gamma(L)$ the set of vertices of L.

3 A new axiomatization of the Shapley value for classical cooperative games

In the whole paper, we consider an infinite denumerable set Ω, the universe of players. As usual, a *game* on Ω is a set function $v : 2^\Omega \to \mathbb{R}$ such that $v(\emptyset) = 0$, which assigns to each *coalition* $S \subseteq \Omega$ its *worth* $v(S)$. We denote by 2^Ω (power set of Ω) the set of coalitions. In this section, we focus on the particular case of *classical cooperative games*, that is to say, each player has the only choice to cooperate or not.

A set $N \subseteq \Omega$ is said to be a *carrier* of a game v when for all $S \subseteq \Omega$, $v(S) = v(N \cap S)$. Thus a game v with carrier $N \subseteq \Omega$ is completely defined by the knowledge of the coefficients $\{v(S)\}_{S \subseteq N}$ and the players outside N have no influence on the game since they do not contribute to any coalition. In this paper, we restrict our attention to finite games, that is to say, games that posses a finite carrier N with n elements. We denote by $\mathcal{G}(N)$ the set of games with the finite carrier N. For the sake of clarity, and to avoid any ambiguity, the domain of $v \in \mathcal{G}(N)$ will be restricted to the elements of 2^N. $\mathcal{G}$ denotes the set of all finite games:

$$\mathcal{G} := \{\mathcal{G}(N) \mid N \subseteq \Omega, n \in \mathbb{N}\}.$$

Identity games of $\mathcal{G}(N)$ are particular games defined by

$$\forall S \subseteq N \setminus \{\emptyset\}, \quad \delta_S(T) := \begin{cases} 1 \text{ if } T = S, \\ 0 \text{ otherwise.} \end{cases}$$

A *value* on $\mathcal{G}(N)$ is a function $\Phi : \mathcal{G}(N) \times N \to \mathbb{R}$ that assigns to every player i in a game $v \in \mathcal{G}(N)$ his prospect $\Phi(v, i)$ for playing the game. For instance, the Shapley value [11] for cooperative games Φ_{Sh} is defined by

$$\forall v \in \mathcal{G}(N), \forall i \in N,$$
$$\Phi_{Sh}(v, i) := \sum_{S \subseteq N \setminus i} \frac{s!\,(n-s-1)!}{n!} \left(v(S \cup i) - v(S)\right). \tag{1}$$

The axiomatization is well-known. Φ_{Sh} is the sole value given on $\mathcal{G}(N)$ satisfying (see also Weber [13]):

Linearity (L): for any $i \in N$, $\Phi(v, i)$ is linear w.r.t the variable v.

Player $i \in N$ is said to be *null* for v if $\forall S \subseteq N \setminus i$, $v(S \cup i) = v(S)$.

Nullity (N): for any game $v \in \mathcal{G}(N)$ and any $i \in N$ null for v, $\Phi(v, i) = 0$.

For any permutation σ on N, we denote $v \circ \sigma$ the game defined by $v \circ \sigma(S) := v(\sigma(S)),\ \forall S \in 2^N$.

Symmetry (S): for any permutation σ on N, any game $v \in \mathcal{G}(N)$ and any $i \in N$, $\Phi(v, \sigma(i)) = \Phi(v \circ \sigma, i)$.

This means that Φ must not depend on the labelling of the players.

Efficiency (E): for any game $v \in \mathcal{G}(N)$, $\sum_{i \in N} \Phi(v, i) = v(N)$.

That is to say, the values of the players must be shared in proportion of the overall resources $v(N)$.

We now introduce a new axiomatization of the Shapley value for classical cooperative games. For any game $v \in \mathcal{G}(N)$ and any coalition $S \in 2^N$, we denote by $v^S \in \mathcal{G}(S)$ the *restricted game v to the power set of S*. For $i \in N$, v^{-i} denotes the restricted game $v^{N \setminus i}$. Let us consider the following axioms for values on $\mathcal{G}$.

Excluded-null-player (ENP): for any finite set $N \subseteq \Omega$ and any game $v \in \mathcal{G}(N)$, if $i \in N$ is null for v,

$$\forall j \in N \setminus i, \quad \Phi(v, j) = \Phi(v^{-i}, j).$$

This simply means that if a null player leaves the game, then other players should keep the same value in the associated restricted game. Note that this axiom completes in a certain sense the above axiom **(N)** since the former deals with null players whereas the latter addresses the others. Therefore, one can merge **(N)** and **(ENP)**:

Generalized nullity (GN): for any finite set $N \subseteq \Omega$ and any game $v \in \mathcal{G}(N)$, if $i \in N$ is null for v,

$$\begin{cases} \Phi(v, i) = 0, \\ \Phi(v, j) = \Phi(v^{-i}, j), \text{ for any player } j \in N \setminus i. \end{cases}$$

We define the particular *unanimity game* of $\mathcal{G}(N)$ by $u_N(S) := \begin{cases} 1, \text{ if } S = N, \\ 0, \text{ otherwise.} \end{cases}$

Equity (Eq): for any finite set $N \subseteq \Omega$, for any player $i \in N$,

$$\Phi(u_N, i) = \frac{1}{n}.$$

This natural axiom simply states that in the particular game where the grand coalition is the unique to produce a unitary worth (all others giving nothing), all players should share the same fraction of this unit.

Theorem 3.1 *Φ_{Sh} is the sole value on $\mathcal{G}$ satisfying axioms **(L)**, **(GN)** and **(Eq)**.*

Note that since the result is given over $\mathcal{G}$, axioms **(L)** and **(N)** should be adjusted in accordance with the arbitrariness of the choice of N. Actually, it is sufficient to specify for these axioms "for any finite set $N \subseteq \Omega$, for any game $v \in \mathcal{G}(N)$".

An important remark is that this new axiomatization has the advantage of characterizing Φ_{Sh} for all games of $\mathcal{G}$, and not only for the games of $\mathcal{G}(N)$, where N is a fixed finite set. This is due to the recursive nature of the axiom **(ENP)**.

We present now another axiomatization of Φ_{Sh}, where the generalized nullity axiom is outlined in another way.

Definition 3.2 *Let $v \in \mathcal{G}(N)$ be any finite game. We call* support *of v, denoted by $\mathfrak{S}(v)$, the minimal carrier of v, that is,*

$$\mathfrak{S}(v) := \bigcap_{C \text{ is a carrier of } v} \{C \in 2^N\}.$$

Actually, a *carrier axiom* has been introduced for the first time by Myerson [9], saying that, if C is a carrier for the game v, then the worth $v(C)$ should be shared only among the members of the carrier. It is shown that this axiom is equivalent to the conjunction of the above axioms **(N)** and **(E)**. With regard to our work, we focus our attention on the support of the game and give an axiom for players in accordance with their membership of the support of the game. If there is no ambiguity, we denote by $v^{\mathfrak{S}}$ the restricted game $v^{\mathfrak{S}(v)}$.

> **Restricted-support games (RS)**: for any finite set $N \subseteq \Omega$, any game $v \in \mathcal{G}(N)$, and any player $i \in N$,
>
> $$\Phi(v,i) = \begin{cases} \Phi(v^{\mathfrak{S}}, i) \text{ if } i \in \mathfrak{S}(v), \\ 0 \text{ otherwise.} \end{cases}$$

Corollary 3.3 *Φ_{Sh} is the sole value on $\mathcal{G}$ satisfying axioms **(L)**, **(RS)** and **(Eq)**.*

To show this result, we propose an alternative characterization of the support of a game:

Lemma 3.4 *Let $v \in \mathcal{G}(N)$ be any game. Then $\mathfrak{S}(v)$ is the set of players which are not null for v.*

4 The Shapley value of multichoice games

In previous section, the lattice representing actions of players was $L := \{0,1\}^{\Omega}$, 0 (resp. 1) denoting absence (resp. presence) of a player. Now, for every player i belonging to a finite carrier of players N, it is assumed that she may act at a level of participation $k \in L_i$ to the game. Actually, $L_i := \{0, 1, 2, \ldots, \top_i\}$ is a linear lattice, where 0 means absence of participation and $\top_i$ represents the maximal participation to the game. Thus $L = L_1 \times \cdots \times L_n$ is the set of all possible joint actions of players of N. We denote by $\mathcal{L}(N)$ the set of all cartesian products of finite linear lattices over N, and by $\mathcal{L}$, the union of all these ones for every finite set N:

$$\mathcal{L}(N) := \{\prod_{i=1}^{n} L_i \mid \top_1, \ldots, \top_n \in \mathbb{N}\},$$
$$\mathcal{L} := \{\mathcal{L}(N) \mid N \subseteq \Omega, n \in \mathbb{N}\}.$$

Note that it shall be useful for the sequel to introduce the following binary relation over $\mathcal{L}$ defined for all $L \in \mathcal{L}(N), L' \in \mathcal{L}(N')$, by

$$L \mathcal{R} L' \text{ iff } \begin{cases} n = n', \\ (\top'_1, \ldots, \top'_n) \text{ is a permutation of } (\top_1, \ldots, \top_n). \end{cases}$$

This relation is obviously an equivalence relation. We denote by $\overline{\mathcal{L}}$ the quotient set $\mathcal{L}/_{\mathcal{R}}$.

Thus, it turns out that $\overline{\mathcal{L}}$ is isomorphic to the set of the partitions of positive integers, where a *partition* of a positive integer m is a finite nonincreasing sequence[1] of positive integers $(\lambda_1, \ldots, \lambda_n)$ such that $\sum_{i=1}^{n} \lambda_i = m$ (see [1]). The λ_i's, corresponding to the maximal levels of participation of players, are called the *parts* of the associated partition. With a slight abuse of notation, we may assimilate $\overline{\mathcal{L}}$ to the set of partitions of positive integers. For any $\lambda := (\lambda_1, \ldots, \lambda_n) \in \overline{\mathcal{L}}$, $|\lambda|$ is the sum of the λ_i's, i.e., the unique integer whose partition is given by λ. Also, let us endow $\overline{\mathcal{L}}$ with the following order. For all $\lambda := (\lambda_1, \ldots, \lambda_n) \in \overline{\mathcal{L}}, \lambda' := (\lambda'_1, \ldots, \lambda'_{n'}) \in \overline{\mathcal{L}}$,

$$\lambda' \leq \lambda \text{ iff } \begin{cases} n' \leq n, \\ \forall i \in \{1, \ldots, n'\}, \lambda'_i \leq \lambda_i \end{cases}.$$

For instance, we have $(2,1,1) \leq (4,3,2,1)$. Note that $\lambda := (1)$ is the bottom of $(\overline{\mathcal{L}}, \leq)$.

Proposition 4.1 *$(\overline{\mathcal{L}}, \leq)$ is a ranked lattice, whose rank function is given by $r(\lambda) = |\lambda|$, $\forall \lambda \in \overline{\mathcal{L}}$.*

For $L \in \mathcal{L}$, $\mathcal{G}(L)$ denotes the set of functions defined on L which vanish at $\perp := (0, \ldots, 0)$: this corresponds to *multichoice games* as introduced by Hsiao and Raghavan [7], where each player has a set of possible ordered actions. For the sake of commodity, we will assimilate any element L of $\mathcal{L}$ with its

[1] In the sequel, elements of $\overline{\mathcal{L}}$ are assumed to be given under this form.

representative element in $\overline{\mathcal{L}}$. In this way, for any $\lambda := (\lambda_1, \ldots, \lambda_n) \in \overline{\mathcal{L}}$, $v \in \mathcal{G}(\lambda)$ means that v is any game with n players such that their maximal participation levels are given *up to the order of players* by $\lambda_1, \ldots, \lambda_n$. We denote by $\mathcal{G}^{\mathcal{M}}$ the set of all multichoice games, that is to say,

$$\mathcal{G}^{\mathcal{M}} := \{\mathcal{G}(L) \mid L \in \mathcal{L}\}.$$

The set $\mathcal{J}(L)$ of join-irreducible elements of L is $\{(0_{-i}, k_i) \mid i \in N, k \in L_i \setminus \{0\}\}$, using our notation for compound vectors (see Section 2); hence each join-irreducible element $(0_{-i}, k_i)$, which we will often denote by k_i if no ambiguity occurs, corresponds to a single player playing at a given level. Thus a value on $\mathcal{G}(L)$ is a function $\Phi : \mathcal{G}(L) \times \mathcal{J}(L) \to \mathbb{R}$ that assigns to every player i playing at the level k in a game $v \in \mathcal{G}(L)$ his prospect $\Phi(v, k_i)$. Our aim is to define the Shapley value $\Phi(v, k_i)$ for each join-irreducible element k_i.

Our approach will take here a similar way, such as the axiomatization given for classical cooperative games. Note that an axiomatization of the Shapley value for multichoice games has already been done in [6] and [10]. The computed formula is the same. However, the former uses a symmetry axiom which is not really natural, whereas the latter is less intuitive and requires more material. Another important difference in [10] is that the extended Shapley value is only given for multichoice games where the number of possible actions is the same for all players. Moreover, none are given in a simple recursive way on the whole set $\mathcal{G}^{\mathcal{M}}$.

Let us first give the following axioms generalizing the ones given for classical games.

> **Linearity ($\mathbf{L}^{\mathcal{M}}$)**: for any $L \in \mathcal{L}$, for all join-irreducible $k_i \in \mathcal{J}(L)$, $\Phi(v, k_i)$ is linear on the set of games $\mathcal{G}(L)$, which directly implies
>
> $$\Phi(v, k_i) = \sum_{x \in L} p_x^{k_i} \, v(x), \quad \text{with } p_x^{k_i} \in \mathbb{R}.$$

For some $k \in L_i, k \neq 0$, player i is said to be *k-null* (or simply k_i is *null*) for $v \in \mathcal{G}(L)$ if $v(x, k_i) = v(x, (k-1)_i)$, $\forall x \in L_{-i}$. If $\top_i$ is null for v and $\top_i = 1$, player i is simply said to be null for v.

> **Nullity ($\mathbf{N}^{\mathcal{M}}$)**: for any $L \in \mathcal{L}$, for any game $v \in \mathcal{G}(L)$, for any player i who is k-null for v,
>
> $$\Phi(v, k_i) = 0.$$

For some $i \in N$, and $v \in \mathcal{G}(L)$, if $\top_i \neq 1$, we define by $v^{-\top_i}$ the restriction of v to the product $L_{-i} \times (L_i \setminus \top_i)$. Moreover, v^{-i} denotes the mapping defined over $L_{-i} : x \mapsto v(x, 0_i)$.

> **Excluded-null-player ($\mathbf{ENP}^{\mathcal{M}}$)**: for any $L \in \mathcal{L}$, for any game $v \in \mathcal{G}(L)$, for any player $i \in N$ such that $\top_i = 1$, if i is null for v,
>
> $$\forall j \in N \setminus i, \quad \Phi(v, \top_j) = \Phi(v^{-i}, \top_j).$$

> **Decreased-level ($\mathbf{DL}^{\mathcal{M}}$)**: for any $L \in \mathcal{L}$, for any game $v \in \mathcal{G}(L)$, for any player $i \in N$ such that $\top_i \neq 1$, if $\top_i$ is null for v,

(i) $\forall k \in L_i \setminus \{0, \top_i\}, \quad \Phi(v, k_i) = \Phi(v^{-\top_i}, k_i)$.

(ii) $\forall j \in N \setminus i, \quad \Phi(v, \top_j) = \Phi(v^{-\top_i}, \top_j)$.

Likewise the previous section, $(\mathbf{N}^{\mathcal{M}})$, $(\mathbf{ENP}^{\mathcal{M}})$ and $(\mathbf{DL}^{\mathcal{M}})$ may be merged in the following axiom:

Generalized nullity $(\mathbf{GN}^{\mathcal{M}})$: for any $L \in \mathcal{L}$, for any game $v \in \mathcal{G}(L)$, for any player i which is k-null for v, any player $j \in N$ and any level $l \in \{1, \ldots, \top_j\}$,

$$\Phi(v, l_j) = \begin{cases} 0 \text{ if } j = i \text{ and } l = k, \\ \Phi(v^{-i}, l_j) \text{ if } j \neq i \text{ and } k = \top_i = 1, \\ \Phi(v^{-\top_i}, l_j) \text{ if } j \neq i \text{ and } k = \top_i \neq 1. \end{cases}$$

Note that this axiom is stronger than the simple concatenation of $(\mathbf{N}^{\mathcal{M}})$, $(\mathbf{ENP}^{\mathcal{M}})$ and $(\mathbf{DL}^{\mathcal{M}})$. Thus its validity is easily verifiable by checking the formulae are true.

For any $L \in \mathcal{L}$, we define the particular *unanimity game* of $\mathcal{G}(L)$ by

$$u_\top(x) := \begin{cases} 1, \text{ if } x = \top, \\ 0, \text{ otherwise.} \end{cases}$$

Equity $(\mathbf{Eq}^{\mathcal{M}})$: for any $L \in \mathcal{L}$, for any player $i \in N$,

$$\Phi(u_\top, \top_i) = \frac{1}{n}.$$

Theorem 4.2 *Under axioms $(\boldsymbol{L}^{\mathcal{M}})$, $(\boldsymbol{GN}^{\mathcal{M}})$, and $(\boldsymbol{Eq}^{\mathcal{M}})$, Φ is given on $\mathcal{G}^{\mathcal{M}}$ by:*

$$\Phi(v, k_i) = \sum_{x \in \Gamma(L_{-i})} \frac{h(x)!\,(n - h(x) - 1)!}{n!} \times [v(x, k_i) - v(x, (k-1)_i)], \tag{2}$$

for any finite set $N \subseteq \Omega, \forall L \in \mathcal{L}(N), \forall v \in \mathcal{G}(L)$, $\forall k_i \in \mathcal{J}(L)$, and where $h(x) := |\{j \in N \setminus i \mid x_j = \top_j\}|$.

Sketch of the proof

It is quite easy to show that the formula satisfies the axioms.

Conversely, we have to show that the formula is uniquely determined by the axioms. First, under $(\mathbf{L}^{\mathcal{M}})$ and $(\mathbf{N}^{\mathcal{M}})$, Φ is given by

$$\Phi(v, k_i) = \sum_{x \in L_{-i}} p_x^{k_i}(L)\,[v(x, k_i) - v(x, (k-1)_i)], \tag{3}$$

for any finite set $N \subseteq \Omega, \forall L \in \mathcal{L}(N), \forall v \in \mathcal{G}(L), \forall k_i \in \mathcal{J}(L)$,

with $p_x^{k_i}(L) \in \mathbb{R}$.

Now, the coefficients $p_x^{k_i}(L)$'s of (3) are computed by a basic *transfinite induction*, which is an extension of mathematical induction on sets endowed with a wellfounded relation. A binary relation R is *wellfounded* on a set E if every nonempty subset of E has an R-minimal element; that is, for every nonempty subset X of E, there is an element m of X such that for every element x of X, the pair (x, m) is not in R. Considering the strict order $<$ associated to $\leq$, it is easy to see that $<$ is wellfounded on $\overline{\mathcal{L}}$. Thus, the inductive step rests on showing the formula over $\mathcal{G}(\lambda)$ if it is true for games defined over all predecessors of λ in $(\overline{\mathcal{L}}, \leq)$. Consequently, if the formula is also satisfied on $\mathcal{G}\big((1)\big)$, then the induction hypothesis applies and the result is satisfied for any game of $\mathcal{G}^{\mathcal{M}}$.

The case $\lambda = (1)$ corresponds to classical cooperative games with one player, which correponds to Theorem 3.1. Indeed, in this case, $\mathcal{J}(L)$ has only one element one can denote by 1_1 (which is also one of the only two elements of L), for which (3) under $\mathbf{(Eq^{\mathcal{M}})}$ writes $\Phi(v, 1_1) = v(1_1)$.

For any $\lambda := (\lambda_1, \ldots, \lambda_n) \in \overline{\mathcal{L}} \setminus \{(1)\}$, let us assume that (2) holds for all games of $\mathcal{G}(\lambda')$ such that $\lambda' \prec \lambda$. We now show that under $\mathbf{(ENP^{\mathcal{M}})}$, $\mathbf{(DL^{\mathcal{M}})}$ and $\mathbf{(Eq^{\mathcal{M}})}$, the unicity of all coefficients in (3) is given for any game $v \in \mathcal{G}(\lambda)$. This being done, as it has been checked that (2) satisfies the axioms, the result will be proved. Let N be any set of players of cardinality n, and L be any linear lattice such that maximum levels $\top_1, \ldots, \top_n$, in any order, are given by λ.

- We first show the unicity of the $\Phi(v, k_i)$'s, for any player $i \in N$ such that $\top_i \neq 1$, and any level $k < \top_i$. Assuming that $\top_i$ is null for a particular game of $\mathcal{G}(L)$, and denoting by L' the lattice $L_{-i} \times (L_i \setminus \top_i)$, axiom $\mathbf{(DL^{\mathcal{M}})}$**-(i)** is used in order to identify the coefficients $p_x^{k_i}(L)$'s with the $p_x^{k_i}(L')$'s. However, since associated partition of L' is one of the predecessors of λ, thus all $p_x^{k_i}(L')$ are known by assumption. Thus the $p_x^{k_i}(L)$'s and then $\Phi(v, k_i)$'s in this situation are given.

- Then, let $i \in N$ be any player and $j \in N \setminus i$ such that $\top_j \neq 1$. Then for another particular game for which $\top_j$ is null, $\mathbf{(DL^{\mathcal{M}})}$**-(ii)** is used to identify the $p_x^{\top^i}(L)$'s with the $p_x^{\top^i}(L')$'s, where $L' := L_{-j} \times (L_j \setminus \top_j)$. Consequently, we have proved the unicity of coefficients $p_x^{\top^i}(L)$ for all $i \in N$, and for all $x \in L_{-i}$ such that $\exists j \in N \setminus i, x_j \neq \top_j, \top_j - 1$.

- Lastly, it remains to show the unicity of the $p_x^{\top^i}(L)$'s, where $i \in N$ and $x \in L_{-i}$ such that $\forall j \in N \setminus i$, $x_j \in \{\top_j, \top_j - 1\}$. This in view, we consider the partition $\big\{C_{i,m}\big\}_{i \in N; 0 \leq m \leq n-1}$ of these indices, where $C_{i,m}$ denotes the set of elements of L_{-i} whose m coordinates x_j are $\top_j - 1$ and the others are $\top_j$. For any $i \in N$, we show the unicity of the $p_x^{\top^i}(L)$'s by induction on m. For $x \in C_{i,0}$, that is to say, $x = \top_{-i} := (\top_1, \ldots, \top_{i-1}, \top_{i+1}, \ldots, \top_n)$, $p_x^{\top^i}(L)$ is given by $\mathbf{(Eq^{\mathcal{M}})}$:

$$\Phi(u_\top, \top_i) = p_{\top_{-i}}^{\top_i}(L) = \frac{1}{n}.$$

  Now, the unicity of the $p_x^{\top^i}(L)$'s for $x \in C_{i,m}$, $1 \leq m \leq n-1$ is shown by induction on m: assuming that all $p_x^{\top^i}(L)$'s are given for all elements of $C_{i,m}$ (m being fixed in $\{0, \ldots, n-2\}$), every $x \in C_{i,m+1}$ is considered

and associated to any $j_0 \in N \setminus i$ such that $x_{j_0} = \top_{j_0} - 1$. Now, two situations may arise: either $\top_{j_0} \neq 1$ or $\top_{j_0} = 1$. In the first case, the approach is the same as in the previous item, where $j := j_0$: by identification of coefficients in (3) with coefficients given by **(DL$^{\mathcal{M}}$)-(ii)**, we show that $p_x^{\top_i}(L) = p_x^{\top_i}(L') - p_{x'}^{\top_i}(L)$, where $x' \in C_{i,m}$, and is defined $x'_j := \begin{cases} \top_{j_0} \text{ if } j = j_0, \\ x_j \text{ otherwise} \end{cases}$ ($p_{x'}^{\top_i}(L)$ is given by hypothesis in the current induction, and $p_x^{\top_i}(L')$ is given by hypothesis in the backward transfinite induction). Finally, if $\top_{j_0} = 1$, for any game $v \in \mathcal{G}(L)$ for which j_0 is null, **(ENP$^{\mathcal{M}}$)** is used to compute $p_x^{\top_i}(L)$ in terms of $p_x^{\top_i}(L')$ and $p_{x'}^{\top_i}(L)$ (the formula is the same as above), where this time $L' := L_{-j_0}$. Note that even if $m+1$ choices of j_0 are possible, one cannot guarantee the existence of such an index such that $\top_{j_0} = 1$ for all $i \in N$, or such that $\top_{j_0} \neq 1$ for all $i \in N$. As a consequence, axioms **(DL$^{\mathcal{M}}$)-(ii)** and **(ENP$^{\mathcal{M}}$)** are both necessary.

This ends the proof of the current inductive step: $\forall i \in N$, all $p_x^{\top_i}(L)$'s are given for any $x \in L_{-i}$ such that $\forall j \in N \setminus i$, $x_j \in \{\top_j, \top_j - 1\}$.

Consequently, for all linear lattice L associated to λ, $\forall k_i \in \mathcal{J}(L)$, $\forall x \in L_{-i}$, all $p_x^{k_i}(L)$'s are given, which also completes the inductive step of the transfinite induction. ■

Acknowledgment

The paper was supported by the French-Serbian project "Pavle Savic" under the name "Aggregation Functions for Decision Making". Fabien Lange thanks Pr. Endre Pap for his valuable advice.

References

[1] G.E. Andrews, *The theory of partitions*, Addison-Wesley, 1976.

[2] G. Birkhoff, *Lattice theory*, 3d ed., American Mathematical Society, 1967.

[3] R. Branzei, D. Dimitrov, and S. Tijs, *Models in cooperative game theory: crisp, fuzzy and multichoice games*, Springer Verlag, to appear.

[4] B.A. Davey and H.A. Priestley, *Introduction to lattices and orders*, Cambridge University Press, 1990.

[5] U. Faigle and W. Kern, *The Shapley value for cooperative games under precedence constraints*, International Journal of Game Theory **21** (1992), 249–266.

[6] M. Grabisch and F. Lange, *Games on lattices, multichoice games and the Shapley value: a new approach*, Mathematical Methods of Operations Research **65** (2007), 153–167.

[7] C.R. Hsiao and T.E.S. Raghavan, *Shapley value for multichoice cooperative games, I*, Games and Economic Behavior **5** (1993), 240–256.

[8] D. Monderer and D. Samet, *Variations on the shapley value*, Handbook of Game Theory No III (Ed. R.J. Aumann and S. Hart, eds.), Elsevier Science, 2002.

[9] R.B. Myerson, *Values of games in partition function form*, Int. J. of Game Theory **6** (1977), 23–31.

[10] H. Peters and H. Zank, *The egalitarian solution for multi-choice games*, Annals of Operations Research **137** (2005), 399–409.

[11] L.S. Shapley, *A value for n-person games*, Contributions to the Theory of Games, Vol. II (H.W. Kuhn and A.W. Tucker, eds.), Annals of Mathematics Studies, no. 28, Princeton University Press, 1953, pp. 307–317.

[12] R.M. Thrall and W.F. Lucas, *N-person games in partition function form*, Naval Research Logistics Quarterly **10** (1963), 281–293.

[13] R.J. Weber, *Probabilistic values for games*, The Shapley Value. Essays in Honor of Lloyd S. Shapley (A.E. Roth, ed.), Cambridge University Press, 1988, pp. 101–119.

Priority, Weight and Threshold in Fuzzy SQL Systems

Aleksandar Takači

Faculty of Technology, University of Novi Sad
Bulevar Cara Lazara 1
Serbia
stakaci@tehnol.ns.ac.yu

Srdjan Skrbic

Faculty of Sciences, University of Novi Sad
Trg Dositeja Obradovića 3
Serbia
shkrba@uns.ns.ac.yu

Abstract: PFSQL is the query language used for querying fuzzy relational databases. One of the most distinguished features of PFSQL is the possibility to prioritize conditions. Priorities are most often confused with weights. In this paper we compare queries with prioritized conditions with queries with weighed conditions. Since PFCSP systems are the theoretical background for PFSQL and similarly WFCSP are the theoretical background for weighted queries we elaborate these two systems. Queries with thresholds are another feature of PFSQL. When a threshold is attached to a condition only the tuples that satisfy the condition with a higher value then the threshold are displayed in the result. Through examples we compare these features of PFSQL.

Keywords: PFSQL, PFCSP, priority, threshold, weight

1 Introduction

The representation of imprecise, uncertain or inconsistent information is not possible in relational databases, thus they require add-onns to handle these types of information. One possible add-on is to allow the attributes to have values that are fuzzy sets on the attribute domain, which results in fuzzy relational databases (FRDB) [2].

SQL (Structured Query Language) is the most influential commercially marketed database query language. It uses a combination of relational algebra and relational calculus constructs to retrieve desired data from a database. FSQL (Fuzzy Structured Query Language) is SQL that can handle fuzzy attribute values [1]. The main difference between SQL and FSQL is that SQL returns a subset of the database that matches search criteria as the query result. On the other hand, FSQL returns a subset of the database together with value in the unit interval for each data row that marks how much does that particular data row matches search criteria.

When attributes with fuzzy values appear in the query it is transformed into a query that can be handled by SQL. Finally, results obtained from the SQL query are post processed in order to obtain the desired information as explained above.

Priority is implemented within Prioritized fuzzy constraint satisfaction problem (PFCSP). PFCSP is actually a fuzzy constraint satisfaction problem (FCSP) in which the notion of priority is introduced [4]. Perhaps, the key factors in that implementation are priority t-norms. They are introduced in such a way that the smallest value, usually the value with the biggest priority, has the largest impact on the result given by a priority t-norm. It is introduced by an axiomatic framework. More details about PFCSP are given later in the paper. PFCSP is the theoretical background for incorporating priority into FSQL.

PFSQL allows conditions in the WHERE clause of the FSQL query to have a certain priority i.e. importance degree [5]. Priorities are most often confused with weights. We compare weighed FSQL queries with PFSQL queries. Also each condition in the WHERE clause can have a threshold. If the threshold is not satisfied the data row is dropped from query result.

2 FCSP, PFCSP and Threshold

I will first define FCSP as the background of PFCSP and WFCSP.

Definition 1 A fuzzy constraint satisfaction problem (FCSP) is defined as a 3-tuple (X, D, C^f) where:

1 $X = \{x_i \mid i = 1, 2, \cdots, n\}$ is a set of variables.

2 $D = \{d_i \mid i = 1, 2, \cdots, n\}$ is a finite set of domains. Each domain d_i is a finite set containing the possible values for the corresponding variable x_i in X.

3 C^f is a set of fuzzy constraints. That is,

$$C^f = \{R^f \mid \mu_{R_i^f} : (\prod_{x_j \in var(R^f)} d_j) \to [0,1] \tag{1}$$

where $i = \{1, 2, \cdots n\}$.

PFCSP systems are an extension of FCSP and they are introduced axiomatically. Instead of giving the formal definition of axioms we will briefly explain each of them.

The first axiom states that a zero value of the local satisfaction degree of the constraint with the maximum priority implies a zero value of the local satisfaction degree. The second axiom states that, in the case of equal priorities, the PFCSP becomes a FCSP. The third axiom captures the notion of the priority. If one constraint has a larger priority then, the increase of the value on that constraint should result in a bigger increase of the global satisfaction degree than when the value with the smaller priority has the same increase. It captures the concept of priority in a linear sense. For example take two investments where one of them results in a bigger profit (larger priority) then it is expected that it is better to invest in a more profitable investment then in a less profitable one, if the profit increase is linear to the investment sum.

The fourth axiom is the monotonicity property, and finally the fifth is the upper boundary condition.

When the t-norm T_L is used together with the s-norm S_P we obtain the system that satisfies the given axiomatic framework. The global satisfaction degree in this system is calculated using the following formula:

$$\alpha_\rho(v_X) = \bigoplus\{\frac{\rho(R^f)}{\rho_{max}} \Diamond \mu_{R_i^f}(v_{var(R^f)}) \mid R^f \in C^f\}, \tag{2}$$

where $\bigoplus(x, y) = T_L(x, y)$, and $\Diamond(x, y) = S_P(1 - x, y)$. We will call this system $T_L - S_P$.

Similarly, we obtain the $\min - \max$ system if we take $\bigoplus(x, y) = T_M(x, y)$, and $\Diamond(x, y) = S_M(1 - x, y)$.

Now, describe how a PFCSP works. Priority of every constraint R_f is evaluated by function $\rho : R_f \to [0, \infty)$. The larger the value of ρ is the larger the priority. After the normalization of the priority values which is done by dividing each priority by $\rho_{\max} = \max\{\rho(R_f), R_f \in C_f\}$ every priority obtains a value in the unit interval. Standard implication aggregates priority of each constraint with its

value. This is done in a way that the larger the priority, the more chance it has for the resulting value to stay the same as it was before aggregation. If the priority of constraint is small, then the aggregated value is closer to 1. This leads to greater values for constraints with the smaller priority. It makes sense, since when these aggregated values are again aggregated with a Scour-concave t-norm T, the smaller values have more impact on the result due to properties of Scour-concave t-norms [3, 6]. We have given two concrete PFCSP systems, min-max and $T_L - S_P$ that satisfy the previously given axioms. Now, we will describe how the global satisfaction degree of this system is calculated.

The function ρ represents the priority of each constraint. Operator $\Diamond$ aggregates priority of each constraint with the value of that constraint. These are then aggregated by the operator $\bigoplus$, which results in the satisfaction degree of an evaluation.

Priorities in PFCSP are most confused with the concept of weights. We can define a WFCSP – weighted fuzzy constraint satisfaction problem, where for each constraint C_i we have an assigned weight w_i. The global satisfaction degree for a valuation v_X, $\alpha_W(v_X)$ in WFCSP is calculated by a known formula:

$$\alpha_W(v_X) = T(c_1 * w_1, c_2 * w_2, \ldots, c_n * w_n), \tag{3}$$

where $c_i = \mu_{C_i}(v_X)$ is the local satisfaction degree of a constraint C_i and T is a t-norm. In order to have an adequate comparison between WFCSP and PFCSP we take $T = T_M$ and $T = T_L$. When T_M is used we get the global satisfaction degree $\alpha_W^{T_M}(v_X)$, and analogously when T_L is used we get $\alpha_W^{T_L}(v_X)$.

In FSQL we can assign a threshold (THOLD) to each constraint. We will now point out the difference between threshold and priority in order to avoid any confusion. If there is a THOLD quantifier attached to a condition, FSQL automatically discards the data row which does not satisfy the condition with a given threshold. On the other hand, if the value of the PRIORITY exists, PFSQL calculates the satisfaction degree for each data row regardless of its satisfaction degree as it will be shown in an example in the following section.

3 FRDB and PFSQL

If we allow the attributes in classical RDB to have values that are fuzzy subsets of the attribute domain, the result will be fuzzy relational databases (FRDB).

Our idea is to have the most common fuzzy set types implemented and that the attribute values in FRDB are most often standard fuzzy sets, and only a small percentage of attribute values are generalized fuzzy sets specified by the user, though our model works with general fuzzy sets in every aspect of FRDB - storing, querying, etc [8]. We introduce one more extension of the attribute value, the linguistic label. Linguistic labels are used to represent most common and widely used expressions of a natural language (such as 'tall people', 'small salary' or 'mediocre result'). Linguistic labels are in fact named fuzzy values from the domain. In order to use a linguistic label on some domain, first we must define this label. For instance, we can define the linguistic label 'tall man' as a fuzzy quantity that has an increasing linear membership function from the point (185,0) to the point (200,1). Considering these extensions, we can define a domain of a fuzzy attribute as:

$$D = D_C \cup F_D \cup L_L, \tag{4}$$

where D_C is a classical attribute domain, F_D is a set of all fuzzy subsets of the domain, and L_L is the set of linguistic labels. In our model we allow triangular fuzzy numbers and fuzzy quantities for F_D.

The basic difference between SQL and PFSQL is in the way the database processes records. In a classical relational database, queries are executed so that a tuple is either accepted in the result set, if it fulfills conditions given in a query, or removed from the result set if it does not fulfill the conditions. In other words, every tuple is given a value true (1) or false (0). On the other hand, as the result set PFSQL returns a fuzzy relation on the database. Every tuple considered in the query is given a value from the unit interval. This value is calculated using operators of fuzzy logic. The question is what elements of the classical SQL should be extended. Because variables can have both crisp and fuzzy values, it is necessary to allow comparison between different types of fuzzy values as well as between fuzzy and crisp values. In other words, PFSQL has to be able to calculate expressions like $height = triangle(180, 11, 8, lin)$, regardless of what value of height is in the database – fuzzy or crisp.

In classical SQL it is clear how to assign truth value to every elementary condition. With fuzzy attributes, situation is more complex. At first, we assign truth value from the unit interval to every elementary condition. Only way to do this is to give algorithm that calculates truth value for every possible combination of values in query and values in the database. For instance, if a query contains

condition that compares a fuzzy quantity value with a triangular fuzzy number in the database, we must have algorithm to calculate compatibility of the two fuzzy sets using a similarity relation. After the truth values from unit interval are assigned, they are aggregated using fuzzy logic. We use t-norm in case of operator AND, and its dual t-conorm in case of operator OR. For negation we use strict negation: $N(x) = 1 - x$.

In case of priority statements, mechanisms deduced from PFCSP systems are used to calculate the result [7]. With normalization of the priority values, every priority obtains a value in the unit interval and also one of the priorities has the value 1. Moreover, with standard implication ($S(1\text{-}p, v)$, S is a s-norm) we aggregate priority of each constraint with its value. This is done in a way that the larger the priority, the more chance it has for the resulting value to stay the same as it was before aggregation. If the priority of constraint is small, then the aggregated value is closer to 1. This leads to greater values for constraints with the smaller priority. It makes sense, since when these aggregated values are again aggregated with either T_M or T_P the smaller values have more impact on the global satisfaction degree.

We now describe processes that allow PFSQL queries to be executed. The basic idea is to first transform PFSQL query in something that classical SQL interpreter understands. Namely, conditions with fuzzy attributes are removed from WHERE clause and those fuzzy attributes are moved up in the SELECT clause. In this way, conditions containing fuzzy constructs are eliminated, so that the database will return all the tuples – ones that fulfill fuzzy conditions as well as the ones that do not. As a result of this transformation, we get a classical SQL query. Then, when this query is executed against the database, results are interpreted using fuzzy mechanisms. These mechanisms assign a value from unit interval to every tuple in the result set.

More precisely, processing the PFSQL query comprises of four phases: query syntax checking, loading the query into memory structure, transformation of the query, and fuzzy interpretation of the results returned by the database. First, the given query is checked for correct syntax using scanner and parser constructed for this task. If the query is correct, the result of syntax analysis done by the parser is a memory structure that represents this query. Next step is the transformation of this structure in already described fashion. We need to check whether an attribute is fuzzy or not. When the fuzzy attributes are identified, conditions in the WHERE clause that they appear in are removed, and the attributes are added to the SELECT clause. Removed conditions are put in another memory structure, because they will be used to interpret the result set. Result is a classical SQL query which can be directly executed against the database. After the query is executed, returned results are further processed. A measure of condition fulfillment is assigned to every tuple in the result set. This measure is a value from the unit interval which is defined by a similarity relation described in the

following section. In this phase we use a memory structure with fuzzy conditions removed from WHERE clause to calculate measures using priority fuzzy logic.

4 Example

Now we will describe the scenario. Suppose we have to pick a soccer player and a basketball player. We evaluate candidates based on their *Height*, *Speed* and *Stamina*. Depending on the sort, each attribute will have a certain priority. We will suppose that for the soccer player *Speed* is the most important, *Stamina* is mildly important and *Height* is the least important. For the basketball player *Height* is the most important, *Stamina* is mildly important and *Speed* is the least important. The min-max and T_L - S_P systems will be used for evaluation.

The results for five athletes are given in the following tables. Table 1 represents evaluations for the soccer player. We assume that the priority of *Speed* constraint is 1, *Stamina* has priority 0.7 and finally *Height* has priority 0.2. Similarly, Table 2 represents evaluations for the basketball player where priority *Speed* constraint is 0.4, *Stamina* has priority 0.6 and finally *Height* has priority 1.

The example for the soccer player can be interpreted as the following PFSQL query.

SELECT *
FROM Athletes
WHERE (*Height*='tall') PR 0.2
AND (*Stamina*='Excellent') PR 0.6
AND (*Speed*='fast') PR 1

The satisfaction degrees for the query are given in Table 1.

no.	*Spd*	*Sta*	*Hei*	T_M	T_L
1	1	0.6	0.2	0.6	0.53
2	0.55	0.65	0.7	0.55	0.31
3	0.1	0.6	1	0.1	0
4	0.8	0.7	0.7	0.7	0.59
5	0.9	0.6	0.3	0.6	0.59

Table 1
Satisfaction degree for the Soccer player

Similarly, the example for the basketball player can be interpreted as the following PFSQL query.

SELECT *
FROM Athletes
WHERE (*Height*='tall') PR 1
AND (*Stamina*='Excellent') PR 0.6
AND (*Speed*='fast') PR 0.4
The satisfaction degrees for the query are given in Table 2.

no.	*Spd*	*Sta*	*Hei*	T_M	T_L
1	1	0.6	0.2	0.7	0.575
2	0.55	0.65	0.7	0.25	0.145
3	0.1	0.6	1	0.1	0.07
4	0.8	0.7	0.7	0.7	0.5
5	0.9	0.6	0.3	0.6	0.51

Table 2
Satisfaction degree for the Basketball player

If we want to use a WFCSP - T_M query for choosing the athletes, the query for the soccer player would be the following.

SELECT (MIN(Height*0.2,Stamina*0.6,Speed*1))
FROM Athletes
WHERE (*Height*='tall')
AND (*Stamina*='Excellent')
AND (*Speed*='fast')

Similarly, for WPFCSP - T_L the query should have the following form:

SELECT (MAX(Height*0.2+Stamina*0.6+Speed*1,0))
FROM Athletes
WHERE (*Height*='tall')
AND (*Stamina*='Excellent')
AND (*Speed*='fast')

The satisfaction degrees for the queries are given in Table 3.

no.	*Spd*	*Sta*	*Hei*	T_M	T_L
1	1	0.6	0.2	0.2	0
2	0.55	0.65	0.7	0.11	0
3	0.1	0.6	1	0.002	0
4	0.8	0.7	0.7	0.16	0
5	0.9	0.6	0.3	0.18	0

Table 3
Weighted satisfaction degrees for the Soccer player

We see that the results in Table 3 differ completely from Table 1. Moreover, if we would run the weighed query for the basketball player we would obtain similar results. This leads to a conclusion that weighted queries are completely different from priority queries.

Finally we can add a threshold to each of the constraints. Assume that we insist that a basketball player must be tall, has good stamina with the degree of 0.6 and fast with the degree of 0.4. This would translate into the following query.

SELECT *
FROM Athletes
WHERE (*Height*='tall') THRESHOLD 1
AND (*Stamina*='Excellent') THRESHOLD 0.6
AND (*Speed*='fast') THRESHOLD 0.4

It is obvious that none of the athletes would fulfill these requirements. This situation is completely different than when we used priority or weighted queries.

Conclusions

In this paper we have presented the PFSQL language and PFCSP systems as the theoretical background for PFSQL, as well as some directions about PFSQL implementation. Similarly we have given a brief description of WFCSP as the theoretical background for weighted queries. In addition, a description of queries with thresholds is also given as another feature of PFSQL.

We have discussed differences and given a comparison of priority, weighted and threshold queries using an example of choosing athletes. We have shown that these three types of queries make different evaluations for the same data i.e. they are essentially different. The threshold is the most strict decision mechanism. Weighted and priority queries are based on different assumptions and must not be confused.

Acknowledgment

The authors would like to acknowledge the support of the Serbian Ministry of Science and Environmental Protection, project 'Mathematical models of non-linearity, uncertainty and decision making', No. 144012 and project 'Abstract Methods and Applications in Computer Science' No. 144017A, also the support of the Ministry of Science, Technology and Environmental Protection of Vojvodina.

References

[1] Galindo, J., Urrutia, A., Piattini, M.: Fuzzy Databases: Modeling Design and Implementation. Hershey, USA: IDEA Group, 2006

[2] Kerre, E. E., Chen, G. Q.: Fuzzy Data Modeling at a Conceptual Level: Extending ER/EER Concepts, In Knowledge Management in Fuzzy Databases, 2000, pp. 3-11

[3] Klement, E., Mesiar, R., Pap, E.: Triangular Norms, Series: Trends in Logic (8), Dordrecht: Kluwer Academic Publishers, 2000

[4] Leung, H., Jennings, N. R.: Prioritized Fuzzy Constraint Satisfaction Problems: Axioms, Instantiation and Validation, Fuzzy Sets and Systems 136(10), 2003, pp. 151-188

[5] Takači, A., Škrbić, S.: How to Implement FSQL and Priority Queries. Proceedings of 3rd Serbian-Hungarian Joint Symposium on Intelligent Systems, Subotica, Serbia: Budapest Tech Polytechnical Institution, 2005, pp. 98-104

[6] Takači, A.: Schur-Concave Triangular Norms: Characterization and Application in PFCSP, Fuzzy Sets and Systems, 155(1), 2005, pp. 50-64

[7] Takači, A., Škrbić, S.: Data Model of FRDB with Different Data Types and PFSQL, Handbook of Research on Fuzzy Information Processing in Databases, Hershey, PA, USA: Information Science Reference, in print, 2008

[8] Zvieli, A., Chen, P.: ER Modeling and Fuzzy Databases, Proceedings of the Second International Conference on Data Engineering, Los Angeles: IEEE Computer Society, 1986, pp. 320-327

Autonomous Hexapod Walker Robot "Szabad(ka)"

Ervin Burkus, Peter Odry

Polytechnical Engineering College "VTS" Subotica
Marka Oreskovica 16, 24000 Subotica, Serbia
ervinbur@freemail.hu, odry@vts.su.ac.yu

Abstract: "Szabad(ka)" is a hexapod walker constructed at the Polytechnical Engineering College "VTS" Subotica to test and help implementing algorithms, designed by the Hungarian science institute called "KFKI", and our college. These algorithms are connected to combined force and position control, and their primary goal is to achieve robust, adaptable walking in rough and unknown environment, and to calculate the prospective and best route. The article describes the problems which appeared during the process of realization. Starting from the imperfections of the designed construction we describe the justification of the implementation of ANFIS in the further development of the robot.

1 Introduction

Walking machines are desirable because they can navigate terrain features that are similar in size to the size of the robot, whereas wheeled and tracked vehicles are only suitable for obstacles smaller than half the diameter of the wheel. Furthermore, if given an ability to find locally horizontal footholds in regionally steep terrain, they can climb extreme angles. Applications potentially include reaching territories which are unreachable or dangerous for humans, exploration, mining, military, rescue, and industrial environments, on earth and beyond.

For walking machines, mostly two legged (biped), four legged (quadruped), and six legged (hexapod) constructions are used. The hexapod is the most stable of the named machines. This is why an autonomous hexapod robot was built with 18 DOFs (degree of freedom) [1].

Before having started building with Szabad(ka) robot, a simplier hexapod robot was built. This robot was the starting point and the source of experience for the designing of Szabad(ka). This hexapod was assembled with -from vitroplast sheets and simple screws. For driving two RC servo motors per feet (2 DOF) were used. The dimensions of the final robot are as follows: 300 mm x 180 mm x 120 mm.

The robot's weight is 1.5 kg. On the body of the robot there was a single microcontroller and was in contact with the base computer through a radio module as used in modeling. With this robot we controlled the movement of the feet, it was not integrated into the controlling cycle in the feet's movement algorithm.

Szabad(ka) was designed using the previously gained experience. Figure 1 shows Szabad(ka) in the phase of design, testing its walk.

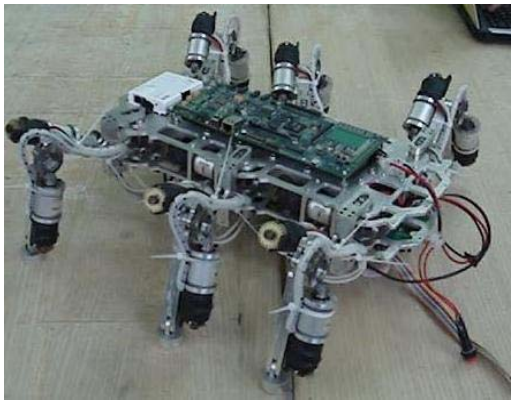


Figure 1
Testing the hexapod Szabad(ka)

Since the robot was created on a tight budget, yet had to meet great expectations, the necessary electronics and mechanical elements, as well as the animation tools for development were also created by the team itself. In order to realize the robot and its development we were forced to set off into various side-projects. These include among other the following:

- 3D robot animation, which helped the robot's mechanical design. This animation helped us control the movement of the designed mechanical structure depending on the developed movement algorithms. We were able to spot the possible clashes of parts, either by redesigning the robot's parts or by changing the movement algorithm.
- The robot's stereo camera was also designed by ourselves. The starting point in the camera development was a camera chip.
- We also designed a CNC milling tool.

Autonomous hexapod walker robot "Szabad(ka)" is developed for testing and developing algorithms connected to motion, robot vision, decision making and robot networking. This hexapod robot was given the name "Szabad(ka)" because it incorporates the name of the city where it was designed as well as hinting at its main feature, namely that it can be openly ('szabad') developing platform for user specific needs.

Szabad(ka) was created to help with the navigational processes of brain research for the Hungarian Academy of Sciences.

2 Construction of Robot

The robot is a complex system both considering its mechanical structure as well as its electronic developments and processor structure.

The robot has MATLAB development platform. It contains one TMS320C6455 DSP (1 GHz frequency) and 10 MSP430F6412 processors, an access point and fully-built it contains 2 cameras, 2 ultra sound radars, several accelerometers, gyroscopes and other sensors for navigation and motion control.

The robot can be used for testing and developing algorithms connected to motion control, visioning, decision making, and networking. In the robot's 10 microcontrollers, there are various algorithms for sensor processing and basic motion control. Higher level software can be written on personal computers, (in C++ or in MATLAB) and following that it can be implemented into the robots DSP processor which has really high processing abilities. Its software platform (with the already written algorithms) is designed in a modular way, which makes it possible for the developers and researchers to develop codes in their own fields of interest without having to be familiar with other software parts of the robot.

The DSP and the 10 MSP processors are connected into network through SPI and I2C protocols. The DSP processor is on a DSP Starter Kit (DSK) made by a third party company, and the 5 PCB-s (2 MSP controllers on each) are designed at the college. From these 5 boards, 3 are for controlling the legs, and the other two are for processing signals from 2 gyroscopes, 2 accelerometers, 2 radars, several IR collusion detectors. Also every leg has an accelerometer and a force sensor in its foot. The communication with the computer is realized wirelessly with an integrated high speed access point. A Video Interface is implemented for connecting 2 digital, high resolution cameras. This is used for stereo visioning.

The robot's body is made from more than 150 aluminium and steel parts. All of them were designed in AutoCAD. The parts were manufactured with CNC milling machines.

2.1 Mechanical

The robot weighs about 10 kg and it is 300 mm high if it stands. All the parts of the legs and the body are made mostly from aluminum. This material is strong and light enough for our needs.

Figure 2 shows the body and the locations of the six identical legs with the approximate movements available about the α axis of each leg before they hit the bumpers on the chassis.

The leg attachments all lie in the same plane, with all the α axes parallel. Besides holding the legs, the function of the chassis is to hold the electronics and the accumulator, too.

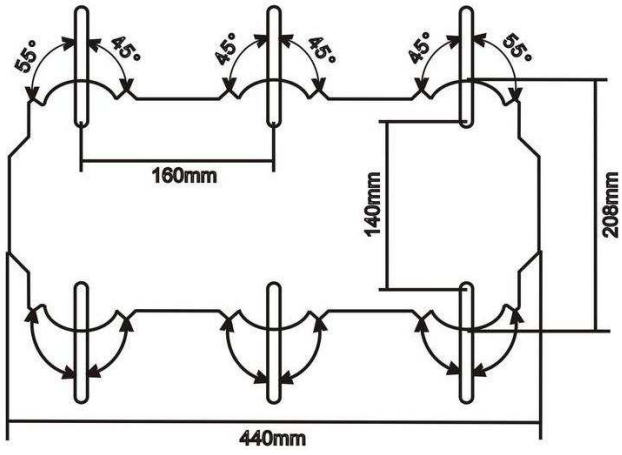


Figure 2
The body

The legs are all identical and have three revolute joints each. The first two are orthogonal to each other and the third is parallel with the second.

All the joints use identical 10W DC motors running through 1:100 planetary reduction gearboxes. After the gearbox, there is a metal bevel gear pair (with 12-36 teeth), providing 1:3 additional reduction, for smoother moving, and more torque.

For angle measurements on the servos home made optical quadrature encoders were used. Calibration of their offset can be done with software by moving the joints until they hit the bumpers, which generates known reference angles for each joint.

In every jointed foot there is a force measuring stamp and a 3D accelerometer providing additional data.

2.2 Electronics

The electronics consists of the DSP Starter Kit (DSK), the MSP boards, and the Stereo Video Interface.

The DSK has a high-end DSP TMS320C6455 processor, running on 1 GHz, having 128 MB of RAM memory, 4 MB of ROM memory, an Ethernet connector and an audio in/out port. Because this DSP is Texas Instruments' most advanced processor, it is a good choice for the current needs. The DSP and one MSP board is connected through SPI protocol and communication with the computer is established wirelessly through an access point and the Ethernet connector.

From the 5 MSP boards, there are:

- 1 Sensors board
- 1 Communications – Motion Algorithm board, and
- 3 Inverse Kinematics boards.

The Communications – Motion Algorithm board, contains 2 MSP processors. One of them is the one, which is connected to the DSP processor. Its main tasks are to transceivers the commands between the DSP and the other MSP-s, and to generate the walking coordinates in dependence of time. The other MSP's job is to process the rear gyroscopes, the rear accelerometers, and some infra red collision signals.

The task of the 3 Inverse Kinematics boards is to receive the coordinates from the Communications – Motion Algorithm board and to generate the desired angles of 3 joints for a leg. Logically, every IK board has 2 MSP-s, one for each leg. This board's other task is to process the data received from the force sensors and the accelerometers placed in the feet.

The **Sensors** board's first MSP controls the 2 ultra sound radars (it moves RC servo motors, and processes data), and the second MSP's job is to process the front gyroscopes, the front accelerometers, and some infra red collision signals.

The **Stereo Video Interface** connects 2 digital, high resolution cameras with the DSP-s EMIF (External Memory Interface). This board contains a FIFO memory. This is needed because the camera is sending its video data continuously and slowly, but the DSP can read data only rapidly, and in smaller parts. Because of the current special needs, the Digital Camera boards are also home made, using OV7640 video IC-s.

2.3 Software

The robot's software is physically divided into 3 parts. These are: the software running on the PC, the DSP software, and the MSP software.

For the PC, currently there is some controlling software written in JAVA, and a MATLAB platform. Both of them are capable for moving the robot in various directions, with various speeds, and with other options. The JAVA software is also prepared for receiving video data and other information from the robot. Further, it can set some behaviour.

About the DSP software: currently its only task is to transmit data between the PC and the Communication MSP but it will be used for image processing and decision making as soon the Stereo Video Interface will be finished. Some contour recognition and other algorithms are already written for it, and ready to use. The DSP software is written in C++ with Code Composer Studio.

The tasks of the software on the 5 MSP boards (on the 10 MSP controllers) were already described under the electronics section, thus 6 MSP-s are for Inverse

Kinematics, 2 MSP-s are processing one gyroscope, one accelerometer, and some IR sensors (per each controller), 1 MSP is for controlling, and processing 2 US radars, and 1 MSP is for communications, and for motion algorithms.

2.4 Animation

In order for our calculation results to be visualized more easily, without a implementing or using the robot, we created an animation ‘subsystem’.

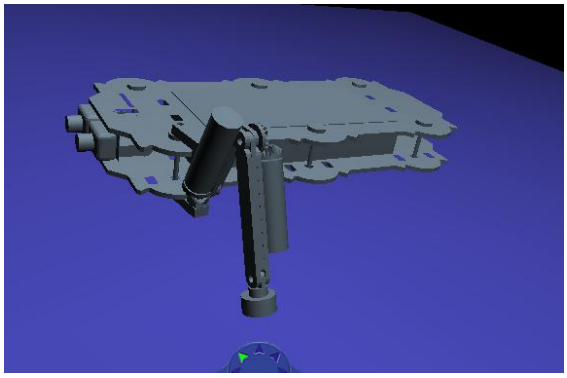


Figure 3
MATLAB simulation for one leg

A 3D Studio Max robot model, which is able to create an animation, based on data from imported using Matlab generated databases containing movement data. This procedure produced good results, but its drawback is that it needs an outside program (outside of Matlab), therefore the animation is not real-time.

So as to fix this problem we started to develop an ‘animation system’ within Simulink environment. With this we are currently able to visualize an animation within a Matlab environment, and are currently working on making it real-time. Figure 3 shows a detail of the robot’s MATLAB Simulink animation.

3 Robot Geometry Calculations

The following section will present several explanatory drawings which illustrate the geometrical calculations.

The first mechanical drawing (Figure 4) shows how many degrees the joints can turn compared to their starting point. Figures 5 and 6 will help to define the maximum tilt angles of the joints. Fig. 6 shows the basic state assumed at the DK (direct kinematics) and IK (inverse kinematics) calculations. The ‘difference angles’ used in calculations is shown in Figure 7.

In order to carry out the calculations we first have to define the dimensions of the leg structure's elements, as well as the maximum tilt angles of the joints.

Figure 4 presents the drawing of the mechanical structure of the robot leg. The thick red line connecting the joints shows the structure used with geometrical calculations. It was a great challege and a very hard job to creat a good and optimal robot-leg construction, so we can say, that this was one of the main problems [2].

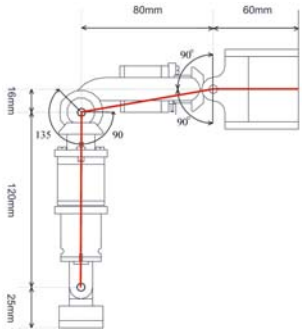


Figure 4
Mechanical construction a foot of robot

Let the horizontal planar angle between the robot's body and the root of the leg be marked *alpha.* Let the perpendicular planar angle between and the root of the leg and the upper leg be marked *theta1* and the angle between the upper leg and the lower leg be marked *theta2.* The length of the root of the leg will be marked with *L0*, the length of the upper leg *L1*, and the lower leg *L2* (Figure 7).

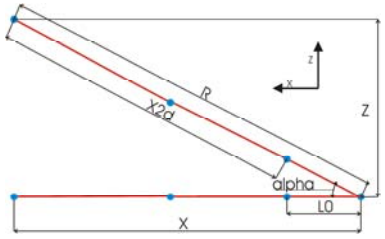


Figure 5
Illustration of the foot geometry`s modification when the robot moves

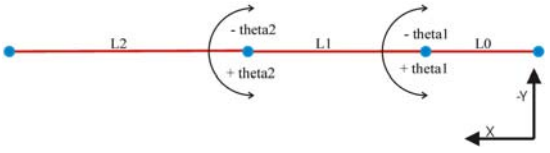


Figure 6
Presentating the basic positions to the IK and DK calculations

The end positions of the alpha angle joint remain the same, while starting from the original *theta1* the *theta1diff* will be subtracted, whereas the *theta2* will be added *theta2diff*. The calculated angles of difference are the following:

$$theta1diff = \arcsin\frac{16}{81.5843} = 11.31°$$

$$theta2diff = 90° - \arcsin\frac{16}{81.5843} = 78.69°$$

Thus the maximum angle rotations are:

alpha = -45° / +45°,

theta1 = -101.31° / +78.69°,

theta2 = -56.31° / +168.69°,

The role of direct kinematics is in our case that it defines the *x, y, z* positions of the robot leg's end point (the foot), if the angles of the joints connecting the parts (*alpha, theta1* and *theta2*) are known. Direct kinematics will only be used later, for verifying purposes.

Based on the Figures 5, 6 and 7 we can deduct the expressions necessary for the DK calculations:

$$\begin{aligned} x_{2d} &= l_1 * \cos\theta_1 + l_2 * \cos(\theta_1 + \theta_2) \\ y &= l_1 * \sin\theta_1 + l_2 * \sin(\theta_1 + \theta_2) \\ r &= x_{2d} + l_0 / \cos\alpha \\ x &= r * \cos\alpha \\ z &= \sqrt{r^2 - x^2} \end{aligned} \quad (1)$$

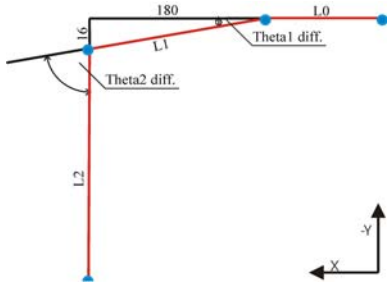


Figure 7
Presentation of the differential angles

4 Calculation of the Course of the Robot's Foot

We were looking for the numerically simplest solutions during the initial phase of development. The aim of this minimization was for the robot to be able to walk on a given course. The aim in this phase was not to include complex controlling conditions, which implement the signals of the sensors of the several accelerometers, gyroscopes, etc.

The calculations for the movement of the flat base of the foot can be put into three groups:

- first the course of the base of the foot in general is described using spline. A certain value will be given to the variable in this expression for the course, depending on how many points the wished course is to be converged to.
- the second calculation the coordinates thus received will help us calculate the necessary angle shift, i.e. we carry out the IK calculation.
- the third calculation refers to the operation of the motors helping the foot's movement using PID controller.

4.1 Calculation of the Leg's Trajectory

In a general case for the description of the current position of the robot foot the Hermite cubic spline calculation method is used for given points *P1*, *P4* and tangent vectors *R1*, *R4*. Look at *x* component:

$$x(t) = a_x t^3 + b_x t^2 + c_x t + d_x \tag{2}$$

So want to solve for a_x, b_x, c_x and d_x using the four continuity conditions, i.e. for two curve segments there is C_0 and C_1 continuity. So we have: $x(0) = P_{1x}$, $x(1) = P_{4x}$, $x'(0) = R_{1x}$, $x'(1) = R_{4x}$, end form of eqution is for *x:*

$$x(t) = P_{1x}(2t^3 - 3t^2 + 1) + P_{4x}(-2t^3 + 3t^2) + R_{1x}(t^3 - 2t^2 + t) + R_{4x}(t^3 - t^2) \tag{3}$$

with similar expressions for *y, z* coordinats.

The use of spline calculation method for the description of foot movement is a great advantage if the ground is not even or there are obstacles on the ground. But in the realization of initial movement this is not the goal, but only to realize the simplest walk algorithm.

The number of operations is shown in Table 1 broken down according to the type of operations.

In the initial phase of robot development the aim was to use the simplest possible algorithms in the realization. In this simple case the 3D foot movement can be broken down to two functional parts:

- the first, along the *z* axis (movement along the axis on the body), as shown the movement by Fig. 8a.
- the second, the movement of the two arms of the foot, upward and downward movement along the axis *z*, seen in Fig. 8b.

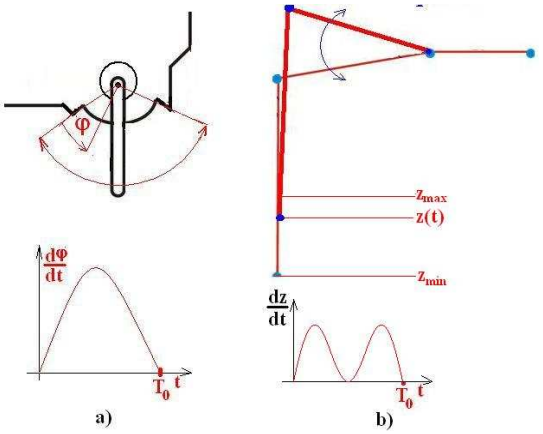


Figure 8

Simple algorithm for robot foot moving

The realization of these two functions assumes simultaneity, and a slower and faster section. The 3D spline description would be simpler in the sense of realization, but the numerical need towards processor capacity is great (Table 1).

4.2 Inverse Kinematics – Calculations of Angles

The role of inverse kinematics – just like in any other system – is that it creates the necessary angles of joints (in our case the *alpha, theta1* and *theta2*) if the end point (base of the foot) coordinates are given – *x, y* and *z*.

Figures 5, 6 and 7 are the basis for carrying out the following calculations:

$$
\begin{aligned}
r &= \sqrt{x^2 + z^2} \\
\alpha &= \arccos(x / r) \\
x_{2d} &= r - 6/\cos\alpha \\
c_2 &= \frac{{x_{2d}}^2 + y^2 - {l_1}^2 - {l_2}^2}{2 * l_1 * l_2} \\
s_2 &= \sqrt{1 - {c_2}^2} \\
\theta_2 &= \arctan(s_2 / c_2) \\
k_1 &= l_1 + l_2 * c_2 \\
k_2 &= l_2 * s_2 \\
\theta_1 &= \arctan(y / x_{2d}) - \arctan(k_2 / k_1)
\end{aligned}
\qquad (4)
$$

4.3 Numerical Calculations with Microcontrollers

The 6 inverse kinematics calculations necessary for the movement of the feet (one for each foot) are carried out by a microcontroller.

The section will discuss the numerical needs of the kinematics calculations for moving the 3 DOF robot legs using a 8 MHz clock signal 16-bit inner architecture microcontroller from the MSP430F16xx family. This microcontroller family is of fix point arithmetic structure and has a 16-bit parallel multiplier.

The MSP1612 was used because it has a 16-bit parallel multiplier and a low consumption, as well as the highest number of operations (performance) index among microcontrollers.

As you can see from the expressions related to inverse kinematics and calculation of the course, several types of calculations have to be carried out: addition, multiplication, division, cos, arcos, raising to the power, square root, arcustangens, as seen in Table 1.

Table 1

The number of operations with one spline and one IK calculation

op	$\sqrt[2]{\ }$	+/-	/	acos	cos	*	atan
trajectory	-	10	-	-	-	29	-
IK	2	8	6	1	1	11	3

The statistics in Table 2 shows several of the calculation needs in MSP430 microprocessors.

Table 2

Required number of clock for calculation several instruction in MSP430 microprocessors

instuction	**float**	**duble**	**optimised float**
add	144	144	142
multiplied	331	331	329
/	369	369	369
$\sqrt[2]{\ }$	1990	1990	1990
cos( )	3646	3646	3646
acos( )	7025	7025	7025
atan()	3070	3070	3068

Using data from Tables 1 and 2, the calculation in MSP430 processors, of spline for a trajectory point and the IK calculation needs *41907* clock signal times. Based on this we assume that the robot foot makes one step in a second and the foot course is converged to in 15 points, then the processor's expected performance is *3771630* clock signals per second (in this calculation is not therein the calculation of PID control, communication between processors, observing sensors and

calculation for the preprocessing sensor signals and other operations and calculation in process).

As can be seen, the processors using the floating point calculation methods are demanding, thus in a given time unit they can carry out fairly few calculations. The calculations cannot be solved by floating point mathematics.

Because of the above-described we searched for different possibilities for the solution. We applied fix point arithmetic and for the most demanding operations, the table-method. This process was helped by the small-resolution encoder and the reducer pair on the motors as well as the realized robot's walking mechanical faults.

First we have to define what the minimal angle fault can be allowed during the calculations. The division of the reducer is originally *1:100*, then with the help of an auxiliary cogwheel the division was multiplied by three. This was necessary for two reasons: the applied motor static momentum was estimated in the form, thus it was unable to hold the robot; on the other hand, the encoders` resolution that we designed was not enough for the driving quality. The initial encoder resolution was of an 8 division, after further development it had a division of 24. With the initial tool the fault with the encoder step was $360^o/(8*100)= 0.45^o$ (this is *800* values in the table, but if we use the trigonometric symmetries, then it is reduced to *200*). For the further developed item it was reduced to $360^o/(3*24*100)= 0.05^o = 3'$ (i.e. in the table, to *1800* elements). Naturally, it greater resolution would be called for but the budget did not allow for this.

In this way we can calculate that one if one encoder sheet moves, how many hour cycle we have for carrying out the calculations. This value was acceptable for the realization of the walking system, there was enough capacity for the microcontrollers. In order to increase the calculation resolution significantly, there are still numerous changes to be done. The alterations have to be carried out in the fields of mechanics, electronics, processors as well as on a theoretic level, too.

This is why we are still searching for alternative methods which can even be implemented on this robot.

The adaptive neuro-fuzzy system seems to be the most suitable for further development (commonly called ANFIS). Based on previous experience, considering the fuzzy [3] [4] and the neuro [6] networks seem suitable for using the MSP430 processor for the solution of the given problem. With this new system, with the increased speed, it will be possible to introduce further control branches.

5 ANFIS

ANFIS method is a hybrid neuro-fuzzy technique that brings learning capabilities of neural networks to fuzzy inference systems [5]. The learning algorithm in adaptive neural technique tunes the membership functions of a Sugeno-type Fuzzy Inference System using the training input-output data. In our case, the input data is coordinate dataset and output data angles dataset. The learning-training algorithm ANFIS map the co-ordinates to the angles.

With the help of Fuzzy logic we create a Fuzzy interference system which is able to conclude the inverse kinematics; if the task's forward kinematics is known.

Since the forward kinematics of the three-degrees-freedom robot foot is known, the base foot's *x, y* and *z* coordinates can be defined for the entire domain of the three joints' angles. The coordinates and their respective angles are stored for the training of our ANFIS network.

During the process of training the ANFIS systems learns how to pair up the coordinates (*x,y,z*) with the angles (*alpha, theta1, theta2*). The trained ANFIS network is thus capable of interference the joint angles based on the required position of the base foot.

The next section will turn to the presentation of MATLAB implementation. In the future course of research the robot will be implemented in a multiprocessor environment.

5.1 Creation of Input Data

The *x, y* and *z* coordinates are defined for *alpha, theta1* and *theta2* in all structurally possible combinations with the given resolution, using forward kinematics (1). The results are stored in three columns:

data1 (x, y, z, alpha),

data2 (x, y, z, theta1),

data3 (x, y, z, theta2).

5.2 Building the ANFIS Network

For the ANFIS system solution we have to create 3 ANFIS networks.

For the ANFIS network to be able to predict the necessary angles, first it has to be trained with the suitable input and output data. The first ANFIS network will be trained for *x, y, z* input data and alpha output data. The previously created *data1* matrix contains the necessary *x,y,z – alpha* data set.

Similarly, the second ANFIS network will be trained for the *x,y,z-theta1* data set stored in the *data2* matrix, whereas the third ANFIS network will be trained for the *x,y,z-theta2* stored in the *data3* matrix. In the current situation we can train the ANFIS network with MATLAB ***anfis*** function. If the function is requested with a simple syntax, it automatically creates a Sugeno-type FIS, and trains it with the given data.

```
anfis1 = anfis(data1, 6, 150, [0,0,0,0]); % train first ANFIS network
anfis2 = anfis(data2, 6, 150, [0,0,0,0]); % train second ANFIS network
anfis3 = anfis(data3, 6, 150, [0,0,0,0]); % train third ANFIS network
```

The first parameter of ANFIS is the training data, the second the number of membership functions, the third is the number of epochs. In order to reach the right exactness, the number of membership functions and the epochs can be defined after several attempts.

After the training anfis1, anfis2, and anfis3 represent the three trained ANFIS networks. After the training is finished, the 3 ANFIS networks have to be able to approximately define the output angles based on the *x, y* and *z* input functions. One of the advantages of the fuzzy approach is tat the ANFIS network is able to define even such angles that do not match with, only bear similarity with the angles used in training. With the help of the network we can define the angles of those points which are found between two points, found in the practice table.

5.3 Control of the ANFIS Network

After the training of the network another important task is to verify our network, to find out how precisely our ANFIS network is able to converge to the results of inverse kinematics. Since in our case we know the inverse kinematics deductions, there is no obstacle to doing this.

We have to generate a 3D matrix which contains the coordinates *x, y* and *z* with the given resolution for points of the chosen work space. After this we define the angles *alpha, theta1* and *theta2* for the given points using inverse kinematics (4):

THREE blocks: data1, data2, data3

In the following step with the help of the trained network we define their angles using the same coordinates. In the end the gained results are compared and we can see how close the FIS outputs are to the results of our inverse kinematics.

Figure 9 shows the error in the calculation of angle α with the application of the proposed method for 3 fuzzy membership functions, the accumulated error already with 3 membership functions within the error limit of the robot mechanics. By the increase of membership functions we could receive better results, but since the key factor in our system is the decrease of calculation time, and the preciseness is

suitable, the number of membership functions was not raised. With other systems, naturally, much better results can be reached.

In the first realization of the ANFIS system, we use the model proposed in Figure 10 to be used without a feedback.

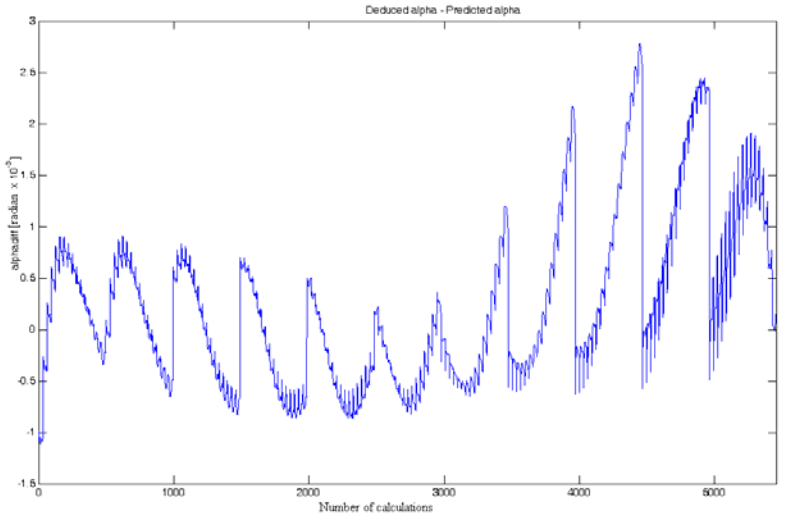


Figure 9

Error in calculation of angle *alpha* with the application of the proposed method

For this test, our robot design is quite suitable, because the MSP microcontroller is appropriate for the FIS and the DSP processor is appropriate for the adaptive calculations. Tests are conducted on a flat ground without any obstacles. In the second phase of introduction of ANFIS technology we introduce feedback. The feedback is obtained from the accelerometer and the gyroscope from the robot's body. The first goal to achieve with feedback is to obtain the horizontal position of the robot's body on an uneven terrain while moving.

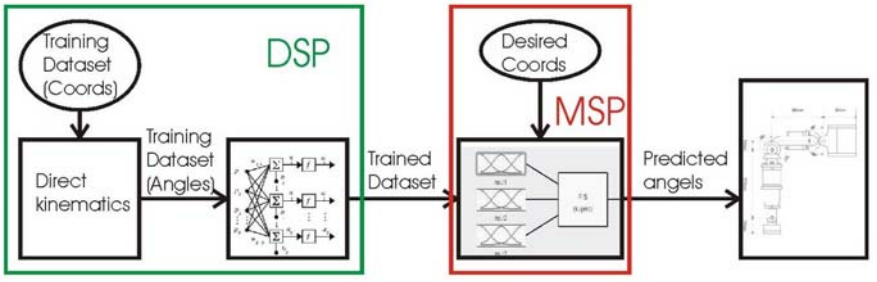


Figure 10

ANFIS model of first real test in our robot application

Conclusions

The robot has been developed for the control of brain research calculations. As it is equipped with high-speed (1 Gbit/s) internet (WiFi) connection, as well as numerous sensors, and currently the highest performance DSP processor, it has large calculation capacity. This device is completely autonomous besides being equipped with the appropriate algorithm structure, so it is capable of carrying out more complicated tasks, as well.

The mechanical development of the robot has been supported with animation while analyzing the crash test of the part of the device. The animation program device has been designed as follows:

1 With the help of MATLAB and the inverse kinematics and spline curve calculations we have calculated the joint coordinates for a given memory position for the point-to-point position of all the joints.

2 All the elements have been designed in mechanical design, with the help of the AUTOCAD program.

3 With then help of the STUDIO 3 MAX the results of points 1 and 2 have been compiled and the animation carried out. Using the animation the robot geometry was set, and also we could analyze the robot's step algorithm. The animation was able to help improve the crashes of the robot's legs. All these have greatly contributed to the development of the robot.

All these would not have been necessary had we disposed of the Solid Works program or a similar 3D animation program. However, in lack of these the tools had first had to be built so the dynamics of the complex geometrical and mechanical structure could be monitored.

In the course of the electronic development the most effective solution was chosen. Each pair of legs has been allocated a processor. Such a processor calculates the spline of the leg route to a pair of legs along with the inverse kinematics and the necessary soft computing calculations at the motor drive taking into consideration the force sensors on the legs and the route parameters.

During our experimentation with the training of ANFIS networks we tried using 3 - 7 membership functions. It was assumed that by increasing the number of membership functions, the degree of error will decrease. This did not happen in that degree, we expected. Based on these, we concluded that if we will not have much processor resources, we can freely reduce the number of membership functions to 3.

Acknowledgement

Special thanks go to the Hungarian Academy of Sciences research group dealing with brain research at the KFKI-RMKI Institute, because they took it on

themselves to finance this robot and continually monitor our activities around the development of the robot.

We are also grateful to the College of Dunaújváros for having agreed to carry out the robot's etch work using CNC. The development of the robot has furthermore been supported by Zsolt Takács from Subotica and the Gordos family from Ada with their work and advice.

Thanks for Tibor Szakáll helps in calculating the necessary processor aqusition times for trigonometric and algebric calculations.

We have to say thank you for Péter Dukán, Mihály Klosák and Árpád Miklós who have helped us in the final operation of the robot project.

References

[1] M. R. Fielding, R. Dunlop, C. J. Damaren: Hamlet: Force/Position controlled Hexapod Walker – Design and Systems, Proc. IEEE Int. Conf. of Controll Appl., Mexico, 2001, pp. 984-989

[2] P. Gonzales de Santos, E. Garcia, J. Estremera: Improving Walking-Robot Performances by Optimizing Leg Distribution, Auton. Robot, Vol. 23, 2007, pp. 247-258

[3] Odry Péter et. al.: 'Fuzzy Logic Motor Control with MSP430x14x', Application Report, Texas Instruments, (SLAA 235), February 2005

[4] Péter Odry, Szabolcs Divéki, Nándor Búrány, László Gyantár: *'Fuzzy Control of Brush Motor - Problems of Computing'*, Plenary section, invited paper, SISY 2004, Subotica, pp. 37-46, ISBN 963 7154 32 9

[5] Jyh-Shing Roger Jang: ANFIS: Adaptive-Network-based Fuzzy Inference System, IEEE Trans. on Systems, Man and Cibernetics, Vol. 23, No. 3, May/June 1993, pp. 665-685

[6] Péter Odry, Gábor Kávai: 'Neural Motor Control', SISY 2005, Subotica, pp. 15-21, ISBN 963 7154 418

Orders in Semirings of Transition Bistochastic Matrices Induced by Mobility Measure of Social Sciences

Branka Nikolic[1], Endre Pap[2]

[1] Nursery School Teacher Training College
University of Novi Sad, 21000 Novi Sad, Serbia
E-mail: bmnikolic@ptt.yu

[2] Department of Mathematics and Informatics
University of Novi Sad, 21000 Novi Sad, Serbia
E-mail: pape@eunet.yu

Abstract: The order of transition matrices induced by mobility measure is presented. A semiring is formed over the set of all bistochastic matrices in which the order is induced by mobility measure which satisfies relaxed Shorrocks monotonicity condition and all other Shorrocks axioms.

Keywords: mobility measure, transition matrix, semiring

1 Introduction

In 1978, Shorrocks [11] defined mobility index in the social sciences, as a continuous function over the set of transition matrices, and he was first to provide axiomatic approach to mobility indices, see [1], [4]. However, Shorrocks himself showed that the axioms he proposed are not consistent for all mobility indices, i.e., there is no mobility index which satisfies all axioms. Different indices detect various mobility aspects. Mobility is defined as movement of dynamic system from one state to the other in time. In social sciences, there are important different types of mobility, as social classes mobility, intergenerational mobility, intragenerational mobility, etc [4]. The selection of the mobility index is very important. Namely, it should satisfy different motivations for measuring mobility. It is common practice, when selecting mobility indices, to initially define the desired features which these indices should satisfy, and that these features have to be consistent. The applied mobility indices have a great influence on transition matrices ranking.

In this paper, the motivation is to measure mobility as movement by using mobility indices which induce partial order on the set of transition matrices. Then every two transition matrices can be compared by using such mobility indices. Transition matrices, which have more movement in them, must have higher mobility index. In this paper the set of transition matrices is reduced to the set of transition bistochastic matrices, as well as the properties of mobility indices are given. Mobility indices with selected properties induces partial order on the set of transition bistochastic matrices and semirings are formed.

Since mobility index is a bounded function, the mobility index minimal value is zero, and the maximal is one. There are also controversies over the selection of transition matrices with minimal and maximal mobility. Transition matrices, as non-negative matrices, are closely related to the class of stochastic processes which are Markov chains. Markov chains are used as theoretical models for description of a system which can be found in different states. In Section 2, an overview of definitions related to Markov Chains and transition matrices are given, and specially important homogenous regular Markov Chains. In Section 3, Shorrocks axiomatic approach to mobility measures is described with stress laid on the inconsistency of these axioms and possibilities of overcoming of this inconsistency. A brief overview of the influences of mobility measure on the order of transition matrices and the problem area of selecting transition matrices which have the values of mobility indices 0 or 1 are presented. In Section 4 a semiring is formed over the set of all the bistochastic transition matrices in which partial ordering is induced by mobility measure, which satisfies some of the Shorrocks axioms. In other words, mobility measure which induces the order in the formed semiring satisfies all the Shorrocks axioms except the monotonicity axiom. The way out is that the mobility measure satisfies the relaxed monotonicity condition thereby achieving the consistency of the axioms.

2 Transition Matrix of Markov Chain

Markov chains (MCs) are used to describe a system which can be found in different states, see [11]. The system passes from one state to the other in time, and this transition is described by the set of transition probabilities $p_{ij}(k)$. If the behaviour of the system is known at the initial time (time 0), the set of transition probabilities determines the behaviour of the system. According to [10] we have the following definitions.

Definition 2.1: For a given a countable state space $S=\{s_0, s_1, s_2,..\}$ a sequence of random variables $(X_k)_{k\in N'}$, where $N' = N \cup \{0\}$, taking values in S, is called *Markov Chain* if it has the following property: if $x_0, x_1, ..., x_{k+1}$ are elements of S, then

$$P(X_{k+1} = x_{k+1}/X_k = x_k,...,X_0 = x_0) = P(X_{k+1} = x_{k+1}/X_k = x_k)$$

if $P(X_k = x_k,...,X_0 = x_0) > 0$. We call the probability $P(X_{k+1} = s_j / x_k = s_i)$ the *transition probability* from state s_i to state s_j and write it as $p_{ij}(k+1)$, $s_i, s_j \in S$, $k \in \mathrm{N}$.

We denote the row vector of the initial distribution by $\prod_0^{'}$. We have by [3], [10].

Definition 2.2: (i) For fixed k in N the matrix $\mathbf{P_k}$= [p_{ij}(k)], $s_i, s_j \in S$, is called *the transition matrix* with non-negative elements.

(ii) $\Omega = \{ \mathbf{P} / p_{ij} \geq 0 \ \forall p_{ij}, \ \sum_{j=1}^{n} p_{ij} = 1 \}$ is called the class of *stochastic matrices*.

(iii) $\Gamma = \{ \mathbf{P} / p_{ij} \geq 0 \ \forall p_{ij}, \ \sum_{i=1}^{n} p_{ij} = 1 , \ \sum_{j=1}^{n} p_{ij} = 1 \}$ is called the class of *bistochastic matrices.*

(iv) If $\mathbf{P_1 = P_2 =...= P_k}$... the Markov chain is said to have *stationary transition probabilities* or is said to be *homogeneous*. Otherwise it is non-homogeneous.

Let $\mathbf{P_k}$ be a transition matrix. Denote by $\prod_k^{'}$ the row vector of the probability distribution of X_k. We shall use notation

$T_{p,r} = \mathbf{P_{p+1}P_{p+2}...P_{p+r}}$,

and write $\prod_k^{'} = \prod_0^{'} T_{0,k}$. For $k > p$ it is

$\prod_k^{'} = \prod_p^{'} T_{p,k-p}$.

For homogeneous Markov chain it holds: $T_{p,k} = \mathbf{P}^k$.

Definition 2.3: (i) A square non-negative matrix $\mathbf{P}$ is said to be *primitive* if there exists a positive integer k such that $\mathbf{P}^k > 0$.

(ii) Any initial probability distribution $\prod_0$ is said to be *stationary*, if $\prod_0^{'} = \prod_k^{'}$, and a Markov chain with such an initial distibution is itself said to be stationary.

Let us denote the *stationary distributon* by π.

Definition 2.4: (i) An n×n non-negative matrix $\mathbf{P}$ is *irreducible* if for every pair *i, j* of its index set, there exist a positive integer $m \equiv m(i,j)$ such that $p_{ij}^{(m)} > 0$.

(ii) MCs is *irreducible* when its transition matrix is irreducible.

Irreducible matrix cannot have a zero row or column, see [10].

Theorem 2.5: An irreducible MCs has a unique stationary distribution π', given as a solution of the equations $\pi'\mathbf{P} = \pi'$ and $\pi'\mathbf{1} = 1$, where $\mathbf{1}$ is vector column with unity in each position, and $\mathbf{1}\pi'$ is a transition matrix whose rows are all equal to π'.

We have by [10].

Theorem 2.6: (Ergodic Theorem for primitive MCs) For a primitive MCs we have:

$\lim_{k\to\infty} \mathbf{P}^{\mathbf{k}} = \mathbf{1}\pi'$ elementwise, where π is the unique stationary distribution of the MCs, and the rate of approach to the limit is geometric.

Following the literature on mobility indices, we assume that the transition matrix $\mathbf{P}$ is homogeneous, irreducible and primitive. Then there exist a unique stationary distribution π (vector column of probability distribution). Morever, $\lim_{k\to\infty} \mathbf{P}^{\mathbf{k}} = \mathbf{1}\pi'$.

3 Mobility Measure on Transition Matrices

In 1978, Shorrocks [11] defined mobility measure as a continuous real function M over the set $\mathtt{P}_{\mathbf{P}}$ of all transition matrices.

Definition 3.1: Mobility index in the Schoorrocs sense is a function M: $\mathtt{P}_{\mathbf{P}} \to \mathrm{R}$, which satisfies the following axioms:

(N) Normalization: $0 \le M(\mathbf{P}) \le 1$, for all $\mathbf{P} \in \mathtt{P}_{\mathbf{P}}$.

(M) Monotonicity: Mobility index reflects the change of increase in the matrix off-diagonal elements at the expense of diagonal elements. Thus, we write $\mathbf{P} \succ \mathbf{P'}$ when $p_{ij} \ge p'_{ij}$ for all the $i \ne j$ and $p_{ij} > p'_{ij}$ for a $i \ne j$. We have that $\mathbf{P} \succ \mathbf{P'}$ implies $M(\mathbf{P}) > M(\mathbf{P'})$.

(I) Immobility: $M(\mathbf{I}) = 0$, where $\mathbf{I}$ is identity matrix.

(PM) Perfect Mobility: Matrices with identical rows have maximal mobility 1.

(SI) Strong immobility: $M(\mathbf{P}) = 0$ if and only if $\mathbf{P} = \mathbf{I}$.

(SPI) Strong perfect mobility: $M(\mathbf{P}) = 1$ if and only if $\mathbf{P}$ has identical rows.

Example 3.2: Shorrocks gives in [11] an example which show that (M) and (PM) are into conflict. Consider the following matrices:

$$P_1 = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \qquad P_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Axioms (M) and (PM) imply $M(\mathbf{P_2}) > M(\mathbf{P_1})=1$, which violets (N).

Shorrocks assumes that a perfectly mobile structure is given by the maximum value of mobilty index, and that the exact index ranking is unimportant, so that the basic conflict is between the axioms (PM) and (M). As one of the ways for solving this conflict Shorrocks proposes adjusting the monotonicity condition by substituting the $M(\mathbf{P}) > M(\mathbf{P'})$ by a weaker condition.

Relaxed monotonicity: Mobility index reflects the change of increase in the matrix off-diagonal elements at the expense of diagonal elements. Thus, we write $\mathbf{P} \succ \mathbf{P'}$ when $p_{ij} \geq p'_{ij}$ for all the $i \neq j$ and $p_{ij} > p'_{ij}$ for a $i \neq j$. Then $\mathbf{P} \succ \mathbf{P'}$ implies $M(\mathbf{P}) \geq M(\mathbf{P'})$.

In this way, consistence is restored, since maximum mobility is assigned to all the transition matrices whose off-diagonal elements are not smaller than some perfectly mobile structure.

Different mobility measures can give different transition matrices ranking. Dardanoni (1993), gives an illustration of ranking of these matrices on the example of three transition matrices and five mobility measures [2]. Dardanoni examines ordering of transition matrices by applying the following mobility measures:

1 *Eigenvalue:* The second highest characteristic square according to the module $|\lambda_2|$.

2 *Trace: trace* = $\frac{trace(P)-1}{n-1}$. This mobility index ignores the extradiagonal transition probabilities.

3 *Determinant:*

Determinant = $|P|^{\frac{1}{n-1}}$.

This mobility index gives the minimum mobility value to the transition matrices which have any two rows or columns equal.

4 *Mean first passage*:

Mean first passage = $\pi' M^P \pi$.

5 *Bartholomew*

Bartholomew = $\frac{1}{n-1} \sum_i \sum_j \pi_i p_{ij} |i-j|$

where π_i is the i-th coordinate of π.

Mean first passage and Bartholomew mobility indices are called *equilibrium indices*. This indices measure mobility where the probability distribution remains unchanged over time, i.e. remains equal to unique stationary distribution π.

Example 3.3: Consider the following three matrices:

$$P_1 = \begin{bmatrix} 0.6 & 0.35 & 0.05 \\ 0.35 & 0.4 & 0.25 \\ 0.05 & 0.25 & 0.7 \end{bmatrix}; P_2 = \begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.1 & 0.2 & 0.7 \end{bmatrix}; P_3 = \begin{bmatrix} 0.6 & 0.4 & 0 \\ 0.3 & 0.4 & 0.3 \\ 0.1 & 0.2 & 0.7 \end{bmatrix}.$$

Each of these three matrices can be most mobile, depending on the selected mobility measure. The ordering of the transition matrices $\mathbf{P_1}$, $\mathbf{P_2}$, $\mathbf{P_3}$ induced by the selected mobility measures as most mobile are the following:

1	Eigenvalue	$\mathbf{P_2}$,
2	Trace	$\mathbf{P_2}$, $\mathbf{P_3}$,
3	Determinant	$\mathbf{P_1}$,
4	Mean First Passage	$\mathbf{P_3}$,
5	Bartholomew	$\mathbf{P_1}$, $\mathbf{P_2}$, $\mathbf{P_3}$.

4 Semirings

Let $\mathtt{S}$ be non-empty set endowed with a partial order $\leq$. The operation $\oplus$ (pseudo-addition) is function $\oplus : \mathtt{S} \times \mathtt{S} \to \mathtt{S}$ which is commutative, non-decreasing, associative and has a zero element, denoted by $\mathbf{0}$. Let $\mathtt{S}_+ = \{x: x \in \mathtt{S}, x \geq \mathbf{0}\}$. The operation $\otimes$ (pseudo-multiplication) is a function $\otimes: \mathtt{S} \times \mathtt{S} \to \mathtt{S}$ which is positively non-decreasing, i.e., $x \preceq y$ implies $x \otimes z \preceq y \otimes z$, $z \otimes x \preceq z \otimes y$, $z \in \mathtt{S}_+$, associative and for which there exist a unit element $\mathbf{1} \in \mathtt{S}$, i.e., for each $x \in \mathtt{S}$, $\mathbf{1} \otimes x = x$. We suppose $\mathbf{0} \otimes x = \mathbf{0}$ and that $\otimes$ is a distributuve pseudo-multiplication with respect to $\oplus$, i.e., $x \otimes (y \oplus z) = (x \otimes y) \oplus (x \otimes z)$. The structure $(\mathtt{S}, \oplus, \otimes)$ is a semiring (see [5], [6], [7], [8], [9]).

Example 4.1: Two special important real cases are $([0, \infty), min, +)$ and *g-calculus*, i.e., when there exist a bijection $g:[a, b] \to [0, \infty]$ such that $x \oplus y = g^{-1}(g(x)+g(y))$ and $x \otimes y = g^{-1}(g(x)g(y))$, where $[a, b] \subset [-\infty, \infty]$.

We denote by $\mathtt{P}_\mathtt{P}$ the set of transition matrices $\mathbf{I}$, $\mathbf{P^1}$, $\mathbf{P^2}$, $\mathbf{P^3}$,..,$\mathbf{P^k}$,.. , where $\mathbf{P}$ is a primitive homogenous transition matrix. According to Theorem 2.6, the sequence $(\mathbf{P^k})_{k \in N}$ converges with exponential growth to stationary regime which has all rows equal.

Theorem 4.2: ($\mathtt{P}_{\mathbf{P}}$, *min,* *) is a *semiring*, where *min:* $\mathtt{P}_{\mathbf{P}}^2 \rightarrow \mathtt{P}_{\mathbf{P}}$ is an idempotent operation, which induces the order on $\mathtt{P}_{\mathbf{P}}$, and it is defined for every two matrices $\mathbf{P}^i$, $\mathbf{P}^j$ from $\mathtt{P}_{\mathbf{P}}$ in the following way

$$min(\mathbf{P}^i, \mathbf{P}^j) = \mathbf{P}^i \text{ if } M(\mathbf{P}^i) \leq M(\mathbf{P}^j), \quad (1)$$

where M is the *mobility measure* which satisfies all Shorrocks axioms, and * is matrix multiplication.

Proof. Let us observe, without loss of generality, transition matrices $\mathbf{P}^1$, $\mathbf{P}^2$ and $\mathbf{P}^3$, and mobilty index M which satisfies all Shorrocks axioms. By matrix multiplication transition probabilities increase at the expense of diagonal elements, and thus

$M(\mathbf{P}^1) < M(\mathbf{P}^2) < M(\mathbf{P}^3)$.

Operation min given by (1) is closed in the set $\mathtt{P}_{\mathbf{P}}$. It is associative:

$$\begin{aligned} min(\mathbf{P}^1, min(\mathbf{P}^2,\mathbf{P}^3)) &= min(\mathbf{P}^1, \mathbf{P}^2) \\ &= \mathbf{P}^1 \\ &= min(\mathbf{P}^1, \mathbf{P}^3) \\ &= min(min(\mathbf{P}^1, \mathbf{P}^2), \mathbf{P}^3). \end{aligned}$$

It is commutative: $min(\mathbf{P}^1, \mathbf{P}^2) = min(\mathbf{P}^2, \mathbf{P}^1)$, and the neutral element $\mathbf{0}$ is the matrix which has all rows equal, and it is stationary distribution matrix $\mathbf{1}\pi'$. Mobility index of this matrix is 1. For every $\mathbf{P} \in \mathtt{P}_{\mathbf{P}}$ we have $min(\mathbf{P}, \mathbf{1}\pi') = min(\mathbf{1}\pi', \mathbf{P}) = \mathbf{P}$.

Operation * of the multiplication operation of transition matrices is closed in set the $\mathtt{P}_{\mathbf{P}}$, and associative, since the matrix multiplication, in general is associative. The neutral element $\mathbf{1}$ is the unit matrix $\mathbf{I}$, and its mobilty index is zero, i.e., for every $\mathbf{P} \in \mathtt{P}_{\mathbf{P}}$ we have $\mathbf{I} * \mathbf{P} = \mathbf{P} * \mathbf{I} = \mathbf{P}$.

Distributivity of the matrix multiplication according to min follows in the following way

$$\mathbf{P}^1 * min(\mathbf{P}^2, \mathbf{P}^3) = \mathbf{P}^1 * \mathbf{P}^2 = min(\mathbf{P}^1 * \mathbf{P}^2, \mathbf{P}^2 * \mathbf{P}^3).$$

For every $\mathbf{P} \in \mathtt{P}_{\mathbf{P}}$ we have $\mathbf{P} * \mathbf{0} = \mathbf{0} * \mathbf{P} = \mathbf{0}$. Matrix $\mathbf{1}\pi'$ is the left zero for matrix multiplication: $\mathbf{1}\pi' * \mathbf{P} = \mathbf{1}\pi'$. If we multiply the equation from the right side by matrix $\mathbf{P}$, we get: $\mathbf{1}\pi' * \mathbf{P}^2 = \mathbf{1}\pi' * \mathbf{P} = \mathbf{1}\pi'$. Continuing this procedure, it follows that

$$\mathbf{1}\pi' * \mathbf{P}^k = \mathbf{1}\pi' * \mathbf{P}^{k-1} = \ldots = \mathbf{1}\pi'.$$

Matrix $\mathbf{1}\pi'$ is the right zero for matrix multiplication: $\mathbf{P} * \mathbf{1}\pi' = \mathbf{1}\pi'$. Let us show, witout loss of generality, that this equation is fulfilled for n = 3,

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \\ b_1 & b_2 & b_3 \end{bmatrix},$$

and this is valid because $\sum_{j=1}^{n} p_{ij} = 1$. If we multiply the last equation from the left side by **P** matrix, we get that for every $\mathbf{P}^{\mathbf{i}}$ stationary probabilty matrix $\mathbf{1}\pi'$ is the right zero for multiplication of homogenous transition matrices. □

Remark 4.3: In Theorem 4.2 the semiring is formed on the set $\mathtt{P}_{\mathbf{P}}$ of transition matrices **I**, $\mathbf{P}^{\mathbf{1}}$, $\mathbf{P}^{\mathbf{2}}$, $\mathbf{P}^{\mathbf{3}}$,..,$\mathbf{P}^{\mathbf{k}}$,.. , where **P** is a primitive homogenous transition matrix. By matrix multiplication, transition probabilities increase at the expense of diagonal elements, and mobility index reflects this change. According to theorem 2.6, for some *k,* the matrix $\mathbf{P}^{\mathbf{k}}$ is stationary distribution matrix with all rows equal. For $m > k$, $\mathbf{P}^{\mathbf{m}}$ is equal to stationary distribution matrix. Counterexample from Example 3.2 is out of the present situation, and the axiom of monotonicity is satisfied.

Theorem 4.4: (P, min, *) is a semiring where P is the set of all primitive homogenous transition bistochastic matrices and unit matrix **I**, *min*: $\mathtt{P}^2 \to \mathtt{P}$ is an idempotent operation, which induces the order on P, and it is defined for every two matrices $\mathbf{P_i}$, $\mathbf{P_j}$ from P in the following way

$$min(\mathbf{P_i}, \mathbf{P_j}) = \mathbf{P}_i \text{ if } M(\mathbf{P}_i) \leq M(\mathbf{P}_j), \tag{2}$$

where *M* is the mobility measure, which fulfils the condition of relaxed monotonicity and all Shorocks' axioms (except monotonicity), and * is the matrix multiplication.

Proof. Let us observe, by not taking away from generality, homogenous transition bistochastic matrices $\mathbf{P_1}$, $\mathbf{P_2}$ i $\mathbf{P_3}$ and mobility index *M* which fulfils the conditions of theorem. Without loss of generality we suppose that

$M(\mathbf{P}_1) \leq M(\mathbf{P}_2) \leq M(\mathbf{P}_3)$.

Operation min is given by (2) is closed in the set P and it is associative:

$$\begin{aligned} min(\mathbf{P}_1, min(\mathbf{P}_2, \mathbf{P}_3)) &= min(\mathbf{P}_1, \mathbf{P}_2) \\ &= \mathbf{P}_1 \\ &= min(\mathbf{P}_1, \mathbf{P}_3) \\ &= min(min(\mathbf{P}_1, \mathbf{P}_2), \mathbf{P}_3). \end{aligned}$$

It is commutative: $min(\mathbf{P_1}, \mathbf{P_2}) = min(\mathbf{P_2}, \mathbf{P_1})$, and the neutral element **0** is the matrix which has all rows equal. On the set of all homogenous bistochastic matrices, every stationary distribution matrix of the $\mathbf{1}\pi'$ form for a sequence $\mathbf{P_i^k}$ has the mobility index 1. From the set of matrices with the mobility index 1, only

the form matrix $\left[\frac{1}{n}\right]_{nxn}$ is at the same time zero element for multiplying, so let us take this matrix as a neutral element for the min operation. For every $\mathbf{P} \in \mathtt{P}$ we have $\min(\mathbf{P}, \left[\frac{1}{n}\right]_{nxn}) = \min(\left[\frac{1}{n}\right]_{nxn}, \mathbf{P}) = \mathbf{P}$.

Operation * is multiplication operation of transition matrices, and it is closed in set P, and associative, since matrix multiplication, in general, is associative. The neutral element **1** is the unit matrix **I**, and its mobility index is zero, i.e., for every $\mathbf{P} \in \mathtt{P}$ we have $\mathbf{I} * \mathbf{P} = \mathbf{P} * \mathbf{I} = \mathbf{P}$.

Distributivity of the matrix multiplication according to min follows in the following way

$\mathbf{P_1} * min(\mathbf{P_2},\mathbf{P_3}) = \mathbf{P_1} * \mathbf{P_2} = min(\mathbf{P_1} * \mathbf{P_2}, \mathbf{P_2} * \mathbf{P_3})$.

For every $\mathbf{P} \in \mathtt{P}$ we have $\mathbf{P} * \mathbf{0} = \mathbf{0} * \mathbf{P} = \mathbf{0}$. Without loss of generality, let n = 3. The following then applies:

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \end{bmatrix} = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \end{bmatrix},$$

this is valid because $\sum_{j=1}^{3} p_{ij} = 1$.

$$\begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} & \frac{1}{n} \end{bmatrix}$$

this is valid because $\sum_{i=1}^{3} p_{ij} = 1$. □

Conclusions

In sociological researches often occurs the problem of determining the minimal and maximal mobile transition matrices, as well as the problem of ordering transition matrices whose mobility measures are between 0 and 1. Many authors keep considering this problem by introducing partial order on the set of transition matrices, induced by mobility measure. So far, this problem has not been considered by forming a semiring on the set of transition bistochastic matrices. In

this paper, a semiring on the set of transition bistochastic matrices has been formed, which is induced by mobility measures, which represents a consistent set of axioms. The research will be continued in the direction of determining mobility measures which satisfy the mentioned set of axioms, as well as comparing orders which induce these measures on the set of transition bistochastic matrices.

Acknowledgement

The second author would like to thank for the support in part by the project MNZŽSS 144012, grant of MTA of HTMT, French-Serbian project 'Pavle Savić', and the project 'Mathematical Models for Decision Making under Uncertain Conditions and Their Applications' of Academy of Sciences and Arts and Technological Development of Vojvodina.

References

[1] Arrow, K. J., Sen, A. K., Suzamura, K.: Handbook of Social Choice and Welfare, Elsevier, Amsterdam, (2002)

[2] Checchi, D., Dardanoni, V.: Mobility Comparisons: Does Using Different Measures Matter?, EconPapers, (2002)

[3] Gear, D., Schokkeart, M., Martinez, M.: Three Meanings of Intergenerational Mobility, Economica 68 (2001) 519-537

[4] Gear, D., Schokkeart, M., Martinez, M.: Measuring Intergenerational Mobility and Equality of Opportunity, Public Economics, Center for Economic Studies, Discussion Paper Series DPS 98.10 (1998)

[5] Gondran, M., Minoux, M.: Graphes, dioides et semi-anneaux, Editions TEC & DOC, Londres-Paris-New York, 2001

[6] Klement, E. P., Mesier, R., Pap, E.: Triangular Norms. Kluwer Academic Publishers, Dordrecht, 2000

[7] Kuich, W.: Semirings, Automata, Languages, Berlin, Springer-Verlag, 1986

[8] Pap, E., Null-Additive Set Functions, Kluwer Academic Publishers, Dordrecht-Boston-London, 1995

[9] Pap, E., Pseudo-Additive Measures and Their Aplications, Handbook of Measure Theory (Ed. E. Pap), Elsevier, Amsterdam, 2002, 1403-1465

[10] Seneta, E.: Non-negative Matrices and Markov Chains, Second Edition, Springer - Verlag, New York, Heidelberg, Berlin, (1981)

[11] Shorrocks, A. F.: The Measurement of Mobility, Econometrica 46, (1978) 1013-1024

An Application of the Interpretation Method in the Axiomatization of the Lukasiewicz Logic and the Product Logic

Aleksandar Perović, Maja Jovanović, Aleksandar Jovanović

Group for Intelligent Systems, Faculty of Mathematics
Vojvode Stepe 305, 11000 Belgrade, Serbia
pera@sf.bg.ac.yu

Abstract. During the last two decades, Group for intelligent systems at Mathematical faculty in Belgrade has developed several theorem provers for different kind of formal systems. Lately, we have turned our attention to fuzzy logic and development of the corresponding theorem prover. The first step is to find the suitable axiomatization, i.e., the formalization of fuzzy logic that is sound, complete and decidable. It is well known that there are fuzzy logics (such as Product logic) that require infinitary axiomatization in order to tame the non-compactness phenomena. Though such logics are strongly complete (every consistent set of formulas is satisfiable), the only possible decidability result is the satisfiability of a formula. Therefore, we have adapted the method of Fagin, Halpern and Megiddo for polynomial weight formulas in order to interpret the Lukasiewicz and the Product logic into the first order theory of the reals.

1 Introduction

The fuzzy logic emerged in mid sixties of the 20th century in order to mathematically capture the notion of uncertainty and enable mathematical tools for reasoning about notions with inherited fuzziness, such as being tall, young, fat, bald etc. The new semantics involve t-norms as 'and' operators, and s-norms as 'or' operators. A t-norm is any function $T:[0,1]^2 \to [0,1]$ such that:

- $T(x,y) = T(y,x)$
- $T(x,T(y,z)) = T(T(x,y),z)$
- $T(x,1) = x$
- $T(x,z) \le T(y,z)$ whenever $y \le z$.

The corresponding s-norm S is defined by $S(x, y) = 1 - T(1 - x, 1 - y)$. For instance, the truth evaluation of the Product conjunction and negation is defined by the following two clauses:

- $e(\alpha \wedge \beta) = e(\alpha)e(\beta)$
- $e(\neg\alpha) = 1$ if $e(\alpha) = 0$, otherwise $e(\alpha) = 0$.

The underlying t-norm is the product norm $T(x, y) = xy$.

Development of formal systems for fuzzy logic is a well worked area (see [1] and [3]). The main goal of any axiomatization is to achieve some variant of completeness. For our purposes, two of those are of interest:

- *Simple completeness*: a formula α is a theorem iff it is valid (satisfied in each model).
- *Strong completeness*: every consistent set of formulas has a model.

For more on the completeness and other basic model theoretical notions we refer the reader to [3] and [4]. Some fuzzy logics, such as $L\Pi\frac{1}{2}$ logic, one axiomatization of the Lukasiewicz logic and the Product logic, are only simply complete. Detailed treatment and the axioms of $L\Pi\frac{1}{2}$ can be found in [1]. In the case of $L\Pi\frac{1}{2}$ we can define a consistent theory that resembles the type of a proper infinitesimal ($\varepsilon > 0$ is a proper infinitesimal if $n\varepsilon < 1$ for all $n = 1,2,3,\ldots$):

$$\Sigma = \{\neg_{\Pi}(p \rightarrow_{\Pi} 0)\} \cup \left\{p \rightarrow_{\Pi} \frac{1}{n} : n = 1,2,3,\ldots\right\}.$$

Namely, theory Σ says that the truth value of p is greater than 0 and lesser than each $\frac{1}{n}$, so it must be a proper infinitesimal. Hence, Σ is unsatisfiable, i.e., there is no Archimedean truth evaluation $e : For \rightarrow [0,1]$ such that $e(\alpha) = 1$ for all $\alpha \in \Sigma$. However Σ is a consistent set of formulas in $L\Pi\frac{1}{2}$. Consequently,

$L\Pi\frac{1}{2}$ is not strongly complete logic. A strongly complete axiomatization of the Lukasiewicz logic and the Product logic can be found in [1] and [3].

The main result of this paper is the application of the interpretation method in the axiomatization of the Lukasiewicz logic and the Product logic. Namely, we have interpreted those two fuzzy logics in the first-order theory of real closed fields (RCF). Our methodology is similar to the one described in [2]. Introduced interpretation allows development of a theorem prover for the Lukasiewic logic and the Product logic that is within PSPACE (polynomial complexity).

The rest of the paper is organized as follows: real-valued propositional logics are discussed in Section 2; Section 3 introduces the interpretation of the Lukasiewicz logic and the Product logic in the theory of real-closed fields; concluding remarks are in Section 4.

2 Real-valued Propositional Logics

First we will build a formal propositional language, then semantically define the notion of the real-valued propositional logic, and, finally, say something about axiomatization of such logics. Like in any formal language, we will start with some basic symbols, and then define the word-formation rules which will be applied in the recursive construction of formulas.

Our basic symbols are propositional letters, truth constants and unary and binary connectives. The set of all propositional letters will be denoted by P, while its elements will be denoted by p, q and r, indexed if necessary. The truth constants will be denoted by c_s, where s is any rational number from the real unit interval $[0,1]$. The unary connectives will be denoted by U, indexed or primed, while the binary connectives will be denoted by B, indexed if necessary. The set For of propositional formulas is recursively defined as follows:

- Propositional letters and truth constants are propositional formulas.
- If α is a propositional formula and U is a unary connective, then $U\alpha$ is a propositional formula.
- If α, β are propositional formulae and B is a binary connective, then $(\alpha B \beta)$ is a propositional formula.
- Propositional formulae can be obtained only by the finite application of the above steps.

A real-valued propositional logic (RVPL) is any function $\Lambda : [0,1]^P \to [0,1]^{For}$ with the following properties:

1 $\Lambda f(p) = f(p)$ for all $f \in [0,1]^P$ and all $p \in P$.

2 $\Lambda f(c_s) = s$ for all $f \in [0,1]^P$ and all c_s.

A propositional formula α is Λ-valid if $\Lambda f(\alpha) = 1$ for all $f \in [0,1]^P$. A RVPL Λ is a truth-functional with respect to the unary connective U if there is a function $F_U : [0,1] \to [0,1]$ such that

$$\Lambda f(U\alpha) = F_U(\Lambda f(\alpha))$$

for all $\alpha \in For$ and all $f \in [0,1]^P$. Similarly, Λ is truth-functional with respect to the binary connective B if there is a function $F_B : [0,1]^2 \to [0,1]$ such that

$$\Lambda f(\alpha B \beta) = F_B(\Lambda f(\alpha), \Lambda f(\beta))$$

for all $\alpha, \beta \in For$ and all $f \in [0,1]^P$. Every fuzzy logic is a RVPL that is truth-functional with respect to some finite set of connectives.

The Lukasiewicz-Product logic is a RVPL $\Lambda_{L\Pi}$ that is truth functional with respect to the binary connectives $\wedge_\Pi$ (Product conjunction), $\to_\Pi$ (Product implication) and $\to_L$ (Lukasiewicz implication), where:

- $F_{\wedge_\Pi}(x, y) = xy$.
- $F_{\to_\Pi}(x, y) = 1$ if $x \le y$, otherwise $F_{\to_\Pi}(x, y) = \frac{y}{x}$.
- $F_{\to_L}(x, y) = \min(1 - x + y, 1)$.

For the sake of simplicity, we may assume that the above three connectives are the only connectives.

Next we will turn to the syntactical propositional logics (SPL's). A SPL Γ is a pair $\langle A_\Gamma, R_\Gamma \rangle$, where A_Γ (the set of axioms) is a subset of For, while R_Γ (the set of derivation rules) is a subset of the set of all partial functions from the power set of For to For. A proof in Γ is any sequence S of propositional formulae with the following properties:

- The order type of S is a successor ordinal.

- For each S_ξ (S_ξ is the ξ-th member of S), $S_\xi \in A_\Gamma$ or $S_\xi = F(X)$, where $F \in R_\Gamma$ and $X \subseteq \{S_\eta : \eta < \xi\}$.

A formula α is a theorem of Γ if it is the last member of some proof in Γ. A SPL Γ is an axiomatization of a RVPL Λ if, for all $\alpha \in For$, α is a theorem of Γ iff α is Λ-valid. For instance, $L\Pi\frac{1}{2}$ is an axiomatization of $\Lambda_{L\Pi}$.

3 Interpretation of $L\Pi\frac{1}{2}$ in RCF

We will assume that the only connectives appearing in propositional formulae are Product conjunction, Product implication and Lukasiewicz implication. Let $L_{OF} = \{+,\cdot,\leq,0,1\}$ (i.e. L_{OF} is a first order language of the theory of ordered fields). As it is usual, by RCF we will denote the L_{OF}-theory of the real closed fields. The axioms of RCF can be found in [4]. Here we will just say that every model of RCF (in the sense of the first order predicate logic, see [4, 6]) is an ordered field in which every polynomial of the odd degree has a root.

By $For_{L\Pi}$ we will denote the set of all propositional formulae built over the countable set of propositional letters and the set of the truth constants $\{c_s : s \in [0,1] \cap Q\}$ by means of $\wedge_\Pi$ (Product conjunction), $\rightarrow_\Pi$ (Product implication) and $\rightarrow_L$ (Lukasiewicz implication). In other words, $For_{L\Pi}$ is the set of all formulas of the Lukasiewicz – Product fuzzy logic. Our aim is to interpret this logic in RCF (for the necessary background on the interpretation method we refer the reader to [6]).

First of all, we will extend the L_{OF} with the countably many new constant symbols C_α, where $\alpha \in For_{L\Pi}$. The intended meaning of C_α is to represent the truth value of α. To provide this, we will extend the theory RCF with the following axioms (in the first order predicate calculus):

- $C_{\alpha \rightarrow_L \beta} = \min(1 - C_\alpha + C_\beta, 1)$.
- $C_{\alpha \wedge_\Pi \beta} = C_\alpha \cdot C_\beta$.
- $C_\alpha \leq C_\beta \rightarrow C_{\alpha \rightarrow_\Pi \beta} = 1$.

- $C_\beta < C_\alpha \to C_\alpha = C_\beta \cdot C_{\alpha \to_\Pi \beta}$.
- $0 \le C_p \le 1$.
- $C_{c_s} = s$.

The above axioms actually follow the standard definition of the truth evaluation in the case of the Lukasiewicz implication, Product conjunction and Product implication. The last axiom provides the usual behavior of the truth constants (i.e. they are, up to equivalence, rational numbers between 0 and 1. Obtained first order theory will be denoted by $RCF_{L\Pi}$.

Using the compactness theorem for the first order predicate logic, one can easily show that $RCF_{L\Pi}$ is a consistent first order theory. It is well known that α is a theorem of $L\Pi\frac{1}{2}$ (see [1, 3]) if and only if α is $\Lambda_{L\Pi}$-valid. An immediate consequence of the definition of $RCF_{L\Pi}$ is the fact that α is $\Lambda_{L\Pi}$-valid if and only if $C_\alpha = 1$ is a theorem of $RCF_{L\Pi}$. Thus, we have interpreted the logic $L\Pi\frac{1}{2}$ in the theory $RCF_{L\Pi}$. It remains to interpret $RCF_{L\Pi}$ in RCF.

Here we will give only the sketch of the proof. The detailed proof of this fact would be given elsewhere. The axioms for new constants provide the following fact: for each $\alpha \in For_{L\Pi}$, there is an RCF-definable function symbol F such that the formula

$$C_\alpha = F\left(C_{p_1}, \ldots, C_{p_n}\right),$$

where $p_1, \ldots, p_n$ are all propositional letters appearing in α. Thus, each sentence of the extended language is $RCF_{L\Pi}$ equivalent to some sentence of the form $\phi\left(C_{p_1}, \ldots, C_{p_n}\right)$. Finally, such a sentence $\phi\left(C_{p_1}, \ldots, C_{p_n}\right)$ is a theorem of $RCF_{L\Pi}$ if and only if the sentence

$$\exists x_1 \ldots \exists x_n \left(0 \le x_1 \le 1 \wedge \cdots \wedge 0 \le x_n \le 1 \wedge \phi\left(x_1, \ldots, x_n\right)\right)$$

is a theorem of RCF. Thus, we have interpreted the Lukasiewicz – Product logic into the first order theory of the reals.

Conclusion

It is a well known fact that the first order theory of the reals is decidable. Though the general decision procedure for RCF is in EXPSPACE, the satisfiability of $\alpha \in For_{L\Pi}$ can be decided by a PSPACE procedure. Namely, the satisfiability of $\alpha \in For_{L\Pi}$ can be equivalently reduced to the existence of the solution of a system of polynomial inequalities. The later problem can be expressed as a purely existential sentence - the universal quantifier does not appear in it in any form (implicit or explicit). As it is shown in [0], this can be decided by the procedure that is in PSPACE.

Our future work will include the implementation of a PSPACE procedure for the existential theory of the reals as well as the construction of the theorem prover for the Lukasiewicz – Product logic based on it.

References

[1] J. Canny. Some Algebraic and Geometric Computations in PSPACE. Proc. XX ACM Symposium on Theory of Computing, pp. 460-467, 1988

[2] F. Esteva, L. Godo, P. Hajek, M. Navara. Residuated Fuzzy Logics with an Involutive Negation. Arch. Math. Logic (2000) 39: 103-124

[3] R. Fagin, J. Y. Halpern, N. Meggido. A Logic for Reasoning about Probabilities. Information and Computation 87 (1/2), 789-128, 1990

[4] P. Hajek. Metamathematics of Fuzzy Logic. Kluwer Academic Publishers, 1998

[5] Ž. Mijajlović. An Introduction to Model Theory. University of Novi Sad, 1987

[6] Z. Ognjanović, M. Rašković. Some Probability Logics with New Types of Probability Operators. J. Logic Computat. Vol. 9, No. 2, 181-195, 1999

[7] J. Shoenfield. Mathematical Logic. Addison-Wesley 1967

An Axiomatizatin of Qualitative Probability

Zoran Ognjanović, Aleksandar Perović, Miodrag Rašković

Mathematical Institute SANU, Kneza Mihaila 36, 11000 Belgrade, Serbia
E-mail: zorano@mi.sanu.ac.yu, pera@sf.bg.ac.yu, miodragr@mi.sanu.ac.yu

Abstract: Qualitative reasoning attracts special attention over the last two decades due to its wide applicability in every-day tasks such as diagnostics, tutoring, real-time monitoring, hazard identification, etc. Reasoning about qualitative probabilities is one of the most common cases of qualitative reasoning. Here we will present a part of our work on the problem of sound, strongly complete and decidable axiomatization of the notion of qualitative probability.

1 Introduction

Reasoning about qualitative probabilities is one of the most prominent cases of the qualitative reasoning. Some varieties of qualitative probability are discussed in [18]. Here we will present a part of our work on the axiomatization of the notion of qualitative probability within the framework of probabilistic logic. Though they are infinitary, our logics are sound, strongly complete and decidable.

The standard approach to probabilistic logic [3], [9] (see also the database [10]) involves extension of the classical propositional or predicate calculus with the modal-like operators, in our notation $P(\alpha) \geq s$, with the attended meaning 'the probability of α is at least s', where s ranges over the predefined index set S. The corresponding semantics is defined as special kind of Kripke models with probability measures on worlds.

As it is well known (see [9], [17]), the key issue for this kind of logics is the non-compactness phenomena: there are examples of finitely satisfiable inconsistent sets of formulas. There are several ways to overcome this problem. In [3] a finitary axiomatization which involves the reasoning about linear inequalities was provided. However, only simple completeness (every consistent formula is satisfiable, in contrast to the strong completeness: every consistent set of formulas is satisfiable) can be proved for that logic. As a consequence, there are examples of consistent unsatisfiable sets of formulas.

In [8], [9], [11], [12], [14], [15] some infinitary probabilistic logics were presented and the corresponding strong completeness theorems were proved. In those logics

we keep formulas finite and allow countably infinite inference rules (conclusions might have countably many premises). In [9], [13] and [17], some probabilistic logics with a fixed finite range for probability measures were given.

Lately, probabilistic logics with the non-Archimedean measures were introduced. For instance, in [15] was introduced a non-Archimedean probabilistic formalism that can be used for the modeling of default reasoning.

In the presence of the probabilistic operators $P(\alpha) \geq s$ one can semantically express the notion of the qualitative probability in the following way: a formula β is at least probable as a formula α iff $P(\beta) \geq s$ implies $P(\alpha) \geq s$ for all $s \in S$. Building on our previous work, we have extended the probability language with an additional binary operator $\alpha \prec \beta$ with the intended meaning 'β is at least probable as α'. Depending on the choice of the index set S and the range of the models, we have developed several formal systems (see [12]). Here we will only present the basic ideas.

2 Syntax and Semantics

By Var we will denote the set of propositional variables. We assume that there are countably many propositional variables. The corresponding set of propositional formulas will be denoted by For_C. Propositional formulas will be denoted by α, β and γ, indexed if necessary. The index set S is defined as the set of all rational numbers in the real unit interval $[0,1]$. The elements of S will be denoted by r and s, indexed if necessary.

A basic probabilistic formula is any formula of the following two forms:

- $P(\alpha) \geq s$;
- $\alpha \prec \beta$.

Abbreviations such as $P(\alpha) > s$, $P(\alpha) \leq s$, etc. are defined as usual (see [12]). The set For_P of all probabilistic formulas is the Boolean closure of the basic probabilistic formulas. Probabilistic formulas will be denoted by ϕ, ψ and θ, indexed if necessary.

A model is any structure $M = \langle W, H, \mu, v \rangle$ with the following properties:

- W is a nonempty set.

- H is an algebra of sets on W.
- $\mu : H \to [0,1]$ is a finitely additive probability measure.
- $v : For_C \times W \to \{0,1\}$ is a truth assignment.

For $\alpha \in For_C$, by $[\alpha]$ we will denote the set of all $w \in W$ such that $v(\alpha, w) = 1$. A model M is measurable if $[\alpha] \in H$ for all $\alpha \in For_C$.

Let $M = \langle W, H, \mu, v \rangle$ be a measurable model. The satisfiability relation is defined recursively as follows:

- M satisfies α if $[\alpha] = W$.
- M satisfies $P(\alpha) \geq s$ if $\mu([\alpha]) \geq s$.
- M satisfies $\alpha \prec \beta$ if $\mu([\alpha]) \leq \mu([\beta])$.
- M satisfies $\phi \wedge \psi$ if M satisfies ϕ and M satisfies ψ.
- M satisfies $\neg\phi$ if M doesn't satisfy ψ.

A formula is satisfiable if there is a measurable model that satisfies it. A formula is valid if it is satisfied in every measurable model. A set of formulas is satisfiable if there is a measurable model that satisfies every formula from the set.

3 Axiomatization

In [12] we have shown that the following axioms and inference rules give a strongly complete characterization of valid probabilistic formulas.

Axioms

1 Substitutional instances of tautologies.

2 $P(\alpha) \geq 0$.

3 $P(\alpha) \geq r \to P(\alpha) > s$, whenever $r > s$.

4 $P(\alpha) > s \to P(\alpha) \geq s$.

5 $P(\alpha) \geq s \leftrightarrow P(\beta) \geq s$, whenever $\alpha \leftrightarrow \beta$ is a tautology.

6 $(P(\alpha) \geq r \wedge P(\beta) \geq s \wedge P(\alpha \wedge \beta) = 0) \to P(\alpha \vee \beta) \geq \min(1, r + s)$.

7 $\left(P(\alpha) \leq s \wedge P(\beta) \geq s\right) \rightarrow \alpha \prec \beta$.

8 $\left(\alpha \leq \beta \wedge P(\alpha) \geq s\right) \rightarrow P(\beta) \geq s$.

Inference Rules

1 Modus ponens for propositional formulas and modus ponens for probabilistic formulas.

2 Necessitation: from α derive $P(\alpha) = 1$.

3 Archimedean rule: from the set of premises $\left\{\phi \rightarrow P(\alpha) \geq s - n^{-1} : n > s^{-1}\right\}$ infer $\phi \rightarrow P(\alpha) \geq s$.

4 $\prec$ rule: from the set of premises $\left\{\phi \rightarrow \left(P(\alpha) \geq s \rightarrow P(\beta) \geq s\right) : s \in S\right\}$ infer $\phi \rightarrow \alpha \prec \beta$.

Let us briefly comment axioms and inference rules. The first axiom is necessary since all tautology instances are valid formulas (see the last two items in the definition of the satisfiability). The 2nd axiom and the necessitation rule provide that the P - value of each propositional formula is between 0 and 1. The 3rd and 4th axiom provide the usual properties of $\geq$. The 5th axiom provides that the equivalent formulas have the same P - values. The 6th axiom provides the finite additivity. The last two axioms and the $\prec$ rule axiomatize qualitative probability. Finally, the Archimedean rule provides the strong completeness of our system.

4 Decidability

An immediate consequence of the proposed axiomatization is the fact that each probabilistic formula has a disjunctive normal form, i.e., it is equivalent to a finite disjunction of literals, where a literal is either a basic probabilistic formula or its negation.

Thus, the question of satisfiability of probabilistic formulas is reduced to the question of satisfiability of finite conjunctions of literals. Using the standard technique (see [8]), we can equivalently reduce satisfiability of probabilistic formula to the existence of a solution of the adjoined system of linear inequalities. It is well known that the later problem is decidable.

Conclusion

The paper presents a probabilistic logic in which the notion of the qualitative probabiliry is completely axiomatized. Our logic involve infinitary rules in order to achieve the strong completeness, which is impossible in the finitary setting

(assuming the real valued semantics and the infinite number of propositional variables). Detailed exposition with all proofs can be found in [12].

We are aware of only a few papers which present a syntactical approach to qualitative probability. An early result on the first order axiomatization of qualitative probability is due to Scott [16]. Some variants of the first order approach in the infinite setting were discussed in [6]. Qualitative probabilities are expressible in the systems introduced in [3]. However, those logics are only simply complete (finitary axiomatization and the real valued semantics). The paper [5] also provides a simply complete axiomatization of qualitative probability.

In [2], [7], [9] and [11], nesting of probabilistic operators is allowed and higher order probabilities are expressible. Our methodology can be easily extended to those cases as well.

References

[1] D. Dubois, H. Prade. Qualitative Possibility Functions and Integrals. In: E. Pap (Ed.), Handbook of Measure Theory, North-Holland, 1499-1522, 2002

[2] R. Fagin, J. Halpern. Reasoning about Knowledge and Probability. Journal of the ACM 41 (2), 340-367, 1994

[3] R. Fagin, J. Y. Halpern, N. Meggido. A Logic for Reasoning about Probabilities. Information and Computation 87 (1/2), 78-128, 1990

[4] D. Lehmann. Generalized Qualitative Probability: Savage Revisited. Proc. of 12th Conference on Uncertainty in Artificial Intelligence (UAI-96), E. Horvitz and F. Jensen (Eds.), 381-388, 1996

[5] E. Marchioni, L. Godo. A Logic for Reasoning about Coherent Conditional Probability: a Modal Fuzzy Logic Approach. In J. Leite and J. Alferes (Eds.), 9th European Conference Jelia'04, lecture notes in artificial intelligence (LNCS/LNAI), 3229, 213-225, 2004

[6] L. Narens. On Qualitative Axiomatizations for Probability Theory. Journal of Philosophical Logic, Vol. 9, No. 2, 143-151, Springer 1980

[7] Z. Ognjanović, M. Rašković. A Logic with Higher Order Probabilities. Publications de l'institute mathematique, nouvelle serie, tome 60 (74), 1-4, 1996

[8] Z. Ognjanović, M. Rašković. Some Probability Logics with New Types of Probability Operators. J. Logic Computat. Vol. 9, No. 2, 181-195, 1999

[9] Z. Ognjanović, M. Rašković. Some First-Order Probability Logics. Theoretical Computer Science 247 (1-2), 191-212, 2000

[10] Z. Ognjanović, T. Timotijević, A. Stanojević. Database of Papers about Probability Logics. Mathematical Institute Belgrade. http//problog.mi.sanu.ac.yu/. 2005

[11] Z. Ognjanović, Z. Marković, M. Rašković. Completeness Theorem for a Logic with Imprecise and Conditional Probabilities. Publications de l'institute mathematique, nouvelle serie, tome 78 (92), 35-49, 2005

[12] Z. Ognjanović, A. Perović, M. Rašković. Logics with the Qualitative Probability Operator. Logic Journal of the IGPL. doi:10.1093/jigpal/jzm031, 2007

[13] M. Rašković. Classical Logic with some Probability Operators. Publications de l'institute mathematique, nouvelle serie, tome 53 (67), 1-3, 1993

[14] M. Rašković, Z. Ognjanović. A First Order Probability Logic LP_Q. Publications de l'institute mathematique, nouvelle serie, tome 65 (79), 1-7, 1999

[15] M. Rašković, Z. Ognjanović, Z. Marković. A Logic with Conditional Probabilities. In J. Leite and J. Alferes (Eds.), 9th European Conference Jelia'04, lecture notes in artificial intelligence (LNCS/LNAI), 3229, 226-238, Springer 2004

[16] D. Scott. Measurement Models and Linear Inequalities. Journal of Mathematical Psychology, 1, 233-247, 1964

[17] W. van der Hoek. Some Considerations on the Logic $P_F D$: a Logic Combining Modality and Probability. Journal of Applied Non-Classical Logics, 7 (3), 287-307, 1997

[18] M. P. Wellman. Some Varaities of Qualitative Probability. Proc. of 5th International Conference on Information Processing and the Management of Uncertainty, Paris 1994

Type-2 Fuzzy Sets and SSAD as a Possible Application

Károly Nagy

Polytechnical Engineering College in Subotica
M. Oreskovica 16, 24000 Subotica, Serbia
nkaroly@vts.su.ac.yu

Márta Takács

Budapest Tech
Bécsi út 96/B, H-1034 Budapest, Hungary
takacs.marta@nik.bmf.hu

Abstract: In the paper a short review of basic type-2 terms is given. One of the possible type-2 Fuzzy Logic System applications is represented for signal processing problems, because type-2 FLSs can handle second level of uncertainties of the stochastic error measurements using Stochastic adding A/D conversion.

Keywords: type-2 fuzzy sets, distance based operators

1 Introduction

Since the beginning of the fuzzy world, there are applications of type-1 fuzzy systems (FS) in which fuzzy system are used to approximate random data or to model an environment that is changing in an unknown way with time [1]. L. A. Zadeh introduced type-2 and higher-types FS in 1975 [6], to eliminate the paradox of type-1 fuzzy systems which can be formulated as the problem that the membership grades are themselves precise real numbers. It is not a serious problem for many applications, but in the cases when the data generating system is known to be time-varying but the mathematical description of the time-variability is unknown, or when the measurement noise is non-stationary and the mathematical description of the non-stationarity is unknown, furthermore when the features in a pattern recognition application have statistical attributes that are non-stationary and the mathematical description of the non-stationarity is

unknown, or knowledge is gathered from a group of experts using questionnaires that involve uncertain words linguistic terms are used and have a non-measurable domain, the results of type-1 fuzziness are imprecise boundaries of FS-s [5].

The solution for this problem can be type-2 fuzziness, where fuzzy sets have grades of membership that are themselves fuzzy [1], [2]. At each value of the primary variable x on the universe X, the membership is a function, and not just a point value (characteristic value). It is the second level, or secondary membership function, whose domain is the primary membership value set. The secondary membership function is a function $MF2:[0,1]\rightarrow[0,1]$. It can be concluded that $MF2$ gives a type-2 fuzzy set which is three-dimensional, and the third dimension offers a certain degree of freedom for handling uncertainties.

In [1] Mendel defines and differentiates two types of uncertainties, random and linguistic. The first one is characteristic for example in statistical signal processing, and the linguistic uncertainties characteristic have in word-information based imprecision systems.

Operations on type-2 fuzzy sets are extended based on type-1 union, intersection, and complementation and usually apply t-norms and conorms.

In the process of T2 fuzzy logic inference the uncertainties of fuzzy membership are included in the calculation. Using the maximum and minimum operators for interval type-2 fuzzy systems (IT2 FS) the description of the Mamdani type T2 inference and defuzzification is given in [2].

In the paper a short review of basic type-2 terms is given. A possible application of this method is related to signal processing [7]. The developed measure method SAADK in [3], stochastic adding A/D conversion has several advantages compared to standard digital instrumentation. Its main advantages are extremely simple hardware and, consequently, simple implementation of parallel measurements, and on the other hand, possibility to trade speeds for accuracy. This instrument can be fast and less accurate, or slow and very accurate, depending on the choice of frequency of reading its output, therefore the membership on the universe X is time-dependant and with random uncertainties, it means suitable for representing it as T2 FS. The hardware realization of the problem is based only on the min and max operators, therefore distance based operators, which are also defined using min and max, making them applicable in the process of realization. Furthermore, it is an opportunity to use simply hardware realization of the operators and approximate reasoning process in the control problems based on the SAADK method and distance based operators [4].

2 Type-2 Fuzzy Sets

A type-1 fuzzy set (T1 FS) has a grade of membership that is crisp, whereas a type-2 fuzzy set (T2 FS) has a grade of membership that is fuzzy, so T2 FS are 'fuzzy-fuzzy' sets.

One way of representing the fuzzy membership of fuzzy sets is to use the footprint of uncertainty (FOU), which is a 2-D representation, with the uncertainty about the left end point of the left side of the membership function, and with the uncertainty about the right end point of the right side of the membership function. Let $x \in X$ from the universe of basic variable of interest. Let $MF(x)$ be the T1 fuzzy membership function of the fuzzy linguistic variable or other fuzzy proposition. The functions $UMF(x)$ and $LMF(x)$ are functions of the left-end and right-end point uncertainty (Fig. 1). In a fix point x' of the universe X it is possible to define so-called vertical slices of the uncertainty, describing it for different possibilities of the $MF_i(x)$ functions (i=1,2,..N), included in the shading of FOU.

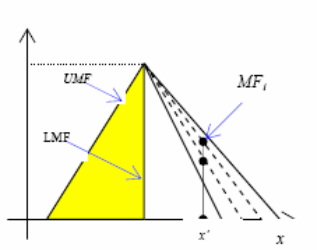


Figure 1
FOU type T2 FS

In this case, for example if we have a Gaussian primary membership function (MF), very often the uniform shading over the entire FOU means uniform weighting or uniform possibilities. T2 FS with FOU representation and uniform possibilities on FOU is called the interval type-2 FS (IT2 FS).

The second way is to use 3D representation, where in the domain xOy the F1 FS $A(x)$ is represented, and in the third dimension for every crisp membership value $A(x)$ of the basic variable x a value of possibility (or uncertainty) is given as the function $MF(x,A(x))= \mu(x, A(x))$. It is the embedded 3D T2 FS (Fig. 2). In Fig. 2 the value $\mu(x, A(x))$ is a random value from the interval [0,1].

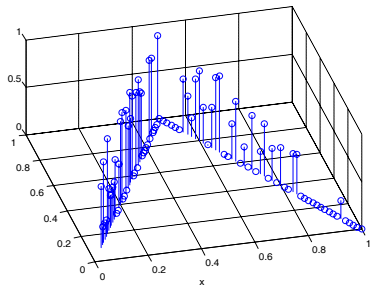


Figure 2
FOU type T2 FS

3 Stochastic Additive Analog to Digital Converter

In classical measurement in the whole measurement range absolute error is, more or less, is the same. This is characteristic of uniform quantizer, i.e. analog to digital converter. On the other hand, in case of control we know that all measurement subintervals are not of equal signification. This means that equal precision is not necessary for control applications. In less significant subintervals it is sufficient to confirm that the system is in it. No more information is needed. Fuzzy systems serve as good mathematical models for this simple situation.

An interesting question arises: does the fuzzyfication belong to domain of measurements and metrology? The answer can be found in several definitions.

Metrology is the science of measurement. Metrology includes all aspects both theoretical and practical with reference to measurements, whatever their uncertainty, and in whatever fields of science or technology they occur. Measurement is set of operations having the object of determining a value of a quantity. Measurable quantity is an attribute of a phenomenon, body or substance that may be distinguished qualitatively and determined quantitatively. Analog to digital conversion is the conversion of an analog quantity in its digital counterpart.

The possible conclusion, from the above definitions, is that fuzzyfication can be accepted as a kind of measurement, since performing fuzzyfication we confirm membership functions, i.e. that fuzzyfied quantity has characteristics of defined fuzzy sets, depending on the moment- the time variable.

Human beings are not capable to process to many information, so they process one information at a moment, possible, most significant in that moment. The process of choosing this most significant information is a kind of control process. Human have capability to do this.

Adaptive measurement system put this human characteristic in the area of automatic control systems. Let us imagine that we have system with many inputs, one multipexer and one system for processing inputs. On input we have information with low resolution, defined with some synchronizing pulse, but using longer processing time we get information with higher resolution. If processing system works longer on specific input the information is more detailed and more reliable, in the opposite case information is rough and less reliable.

Processing cycle for all inputs lasts the same time, in every case, but work on specific inputs can be variable, depend on system state. So, there is a need for instrument capable to give rough and fast measurement information in the shorter time or precise and reliable measurement information in longer time interval. One of such instruments is stochastic additive A/D converter (SSAD).

The developed measure method, stochastic adding A/D conversion has several advantages compared to standard digital instrumentation [3]. Its main advantages are:

- extremely simple hardware and, consequently, simple implementation of parallel measurements, and
- possibility to trade speed for accuracy.

This instrument can be either fast and less accurate, or slow and very accurate. The choice can be made specifying the frequency of reading its output. This is a kind of adaptation which master processor performs. In the case of multi-channel measurements, adaptation can be performed on each channel independently.

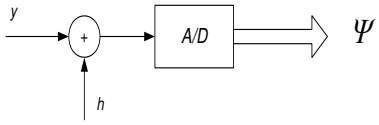


Figure 3
The most simple outline of the instrument

In the Fig. 3 a schematic of the instrument is shown. Dithering signal h is random, uniform and satisfies Widrow's conditions

$$0 \le |h| \le \frac{a}{2} \tag{1}$$

$$p(h) = \frac{1}{a}, \tag{2}$$

where a is a quantum of the uniform quantiser, and $p(h)$ is the corresponding probability density function of h.

3.1 Theory of Operation - DC Inputs

Let us observe the output of *AD* converter Ψ. Let $y = const = na + |\Delta a|$ be the corresponding input voltage located between quantum level na and $(n+1)a$, at the distance $|\Delta a| \leq a/2$ from the closest quantum level na shown in Figure 4.

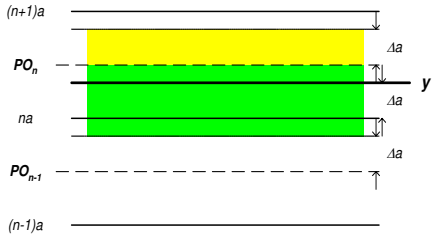


Figure 4
The situation for y=const

For the situation depicted in Fig. 4 the quantized level of $y+h$ is $\Psi \in \{na, (n+1)a\}$. The expectation $\overline{\Psi}$ is given by

$$\overline{\Psi} = \Psi_1 \cdot p_1 + \Psi_2 \cdot p_2 = (n+1) \cdot a \cdot \frac{|\Delta a|}{a} + na \frac{(a - |\Delta a|)}{a} =$$
$$= n \cdot |\Delta a| + |\Delta a| + na - n \cdot |\Delta a| = na + |\Delta a| = y$$

$$\overline{\Psi} = y \tag{3}$$

The corresponding variance is:

$$\overline{e^2} = \sigma_\Psi^2 = (\Psi_1 - \overline{\Psi})^2 \cdot p_1 + (\Psi_2 - \overline{\Psi})^2 \cdot p_2 =$$
$$(a - |\Delta a|)^2 \cdot \frac{|\Delta a|}{a} + |\Delta a|^2 \cdot \frac{(a - |\Delta a|)}{a} =$$
$$\sigma_\Psi^2 = (a - |\Delta a|) \cdot \left((a - |\Delta a|) \cdot \frac{|\Delta a|}{a} + \frac{|\Delta a|^2}{a} \right) =$$
$$(a - |\Delta a|) \cdot (|\Delta a|)$$

$$\sigma_\Psi^2 = (a - |\Delta a|) \cdot (|\Delta a|) \tag{4}$$

An interesting question is: what is the error if we have finite number N of dithered samples? The answer gives theory of samples and central limit theorem. Suppose that we have next set of samples: $\Psi_1, \Psi_2, ..., \Psi_n$.

Then measurement quantity is

$$\overline{\Psi} = \frac{1}{N}\sum_{i=1}^{N}\Psi_i \approx const \tag{5}$$

Central limit theorem gives next result

$$\sigma^2_{\overline{\Psi}} = \frac{\sigma^2_{\Psi}}{N}. \tag{6}$$

The equations (4), (5) and (6) completely define the situation if we have a finite number N of dithered samples.

The shape of σ^2_{Ψ} as function of Δa, for the case of $y=const$, is given in Fig. 4.

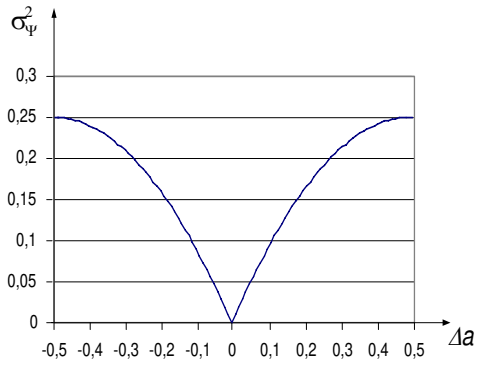


Figure 5
The diagram of σ^2_{Ψ} (Δa) for y=const and a=1

From the Fig. 5 is obvious that the maximum value of error is in the case of $y = c = \frac{a}{2} + na$, i.e. when measurement voltage has value of threshold voltage, but the minimum value of error is when $y=c=na$, i.e. when measurement voltage has value of quantum level.

3.2 Function-based Measurements and the Numeric Processing of the Measurements

Stochastic additive A/D converter basically works with equidistant comparators, but this is not obligatory. If we make A/D converter with non-equidistant threshold levels, then we have stochastic additive analog to fuzzy converter (SAAFC).

Defining the membership functions on the output of the stochastic additive analog to fuzzy converter the measurement range is divided in to *n* intervals. There is one deciding threshold in every interval. (*n* +1) sets are defined by *n* deciding thresholds. The first set is approximately the minimum value of the measurement range, $(n+1)^{st}$ set is approximately the maximum value of the measurement range. Between these two sets there are (*n*-1) sets 'approximately A_j' where x_{A_j} represents the middle of the two deciding thresholds: PO_j and PO_{j-1} where *j*=2,...,*n*. The deciding thresholds are marked from PO_1 to PO_n. There exists the following limitation: $PO_1 > PO_{j-1}$ where $j = 2,...,n$.

In general case, the membership of the defined sets can be described in the following way:

$$A_j(x) = \left\{ x_i \middle| x_i \geq PO_{j-1} \wedge x_i < PO_j \right\} \qquad \text{where } j = 2,...,n \tag{7}$$

For the first and the last set:

$$A_1(x) = \left\{ x_i \middle| x_i < PO_1 \right\} \tag{8}$$

$$A_{n+1}(x) = \left\{ x_i \middle| x_i \geq PO_n \right\} \tag{9}$$

The fuzzy intervals or fuzzy numbers are defined on these sets.

SAAFC makes the quantification on this sum:

$$x_i = x(t) + h(t) \tag{10}$$

where x_i is the *i*th value of input sum into the flash A/D converter, *i* is the serial number of quantification in the measurement cycle, $x(t)$ is the value of the variable which is fuzzyficated in the moment of sampling, $h(t)$ is the value of random variable of uniform distribution in the moment of sampling.

Membership function of fuzzy sets (A_j) is defined depending on the relative frequency of appearing of the measuring result in (A_j) during the measuring cycle.

$$\mu_{A_j}(x) = \frac{a_j}{N} \tag{11}$$

where: *j* is the serial number of the fuzzy set, ($j = 1, 2,...,(n+ 1)$), a_j is the number of appearing value on the output of the fuzzy set A_j,*N* is the total quantification number during the measuring cycle.

The following limitation is in force for *h*:

$$h \leq \min(PO_j - PO_{j-1}) \text{ where } j = 2,...,n. \tag{12}$$

where *h* must be smaller or equal to the minimum difference between the two neighboring deciding thresholds.

The elements of continuous set are ordered to the membership function values as a secondary fuzzy set, constructed a type-2 fuzzy environment.

If we deal with system with great number of inputs and with limited capability, not only for measurements, but for processing as well, we can use this system more efficiently if we apply above mentioned idea: if processing system works longer on specific input the information is more detailed and more reliable, in the opposite case information is rough and less reliable. Similarly to human reasoning, system can concentrate on most significant input and most significant information is processed. For other inputs, system only confirms that they are under control.

The appearance of the signal in the first instant of quantization indicates rough estimation of specific input and it can define processing time interval. If estimation tells that input is under control, we can process immediately the next input. But if estimation tells that input is not under control, system pay more attention (longer processing time) to go back under control.

Conclusions

The T2 FS are of primary importance in all preliminary works in order to later construct inference mechanism based on this family of fuzzy operators and T2 fuzzy logic.

A possible application of type-2 fuzzy sets, related to signal processing and measure method SAADK was represented. The main advantages of the stochastic adding A/D conversion compared to standard digital instrumentation are its extremely simple hardware and simple implementation of parallel measurements, towards the possibility to trade speeds for accuracy. This instrument can be fast and less accurate, or slow and very accurate, depending on the choice of frequency of reading its output, therefore the membership on the universe X is time-dependant and with random uncertainties, it means suitable for representing it as T2 FS. The hardware realization of the problem is based only on the min and max operators, therefore distance based operators, which are also defined using min and max, making them applicable in the process of realization. It is an opportunity to use simply hardware realization of the operators and approximate reasoning process in the control problems based on the SAADK method.

Acknowledgement

The research was partially supported by the Hungarian Scientific Research Project OTKA T048756 and by the project ‘Mathematical Models for Decision Making under Uncertain Conditions and Their Applications’ financed by the Vojvodian Provincial Secretariat for Science and Technological Development.

References

[1] J. M. Mendel: Type-2 Fuzzy Sets: Some Questions and Answers, IEEE Neural Network Society, Aug. 2003

[2] J. M. Mendel, R. I. B. John: Type-2 Fuzzy Sets Made Simple, IEEE Transactions on Fuzzy Systems, 10/2, pp. 117-127, 2002

[3] K. Nagy, V. Vujicic, Z. Mitrovic, M. Takács: Fuzzyfication and Measurement Using Stochastic Approach, SISY 2007 Conference

[4] Rudas, I. J.: Evolutionary Operators: New Parametric Type Operator Families, Fuzzy Sets and Systems 23, pp. 149-166, 1999

[5] Tuhran Ozen, J. M. Garibaldi: Investigating Adaptation in Type-2 Fuzzy Logic System, Applied to Umbilicial Acid-Base assessment

[6] Zadeh, L. A.: The Concept of a Linguistic Variable and its Application to Approximate Reasoning – 1, Information Sciences, Vol. 8. pp. 199-249, 1975

[7] Mendel, J. M.: Uncertainty, Fuzzy Logic, and Signal Processing, Signal Processing, Vol. 80, Issue 6, 2000

Entropy and Gaussianity - Measures of Deterministic Dynamics of Heart Rate and Blood Pressure Signals of Rats

Tatjana Loncar-Turukalo[1], Sonja Milosavljevic[2], Olivera Sarenac[2], Nina Japundzic-Zigon[2], Dragana Bajic[1]

[1] Department of Communications and Signal Processing
Faculty of Technical Sciences
University of Novi Sad
Trg Dositeja Obradovia 6
21000 Novi Sad, Serbia

[2] Department of Pharmacology, Clinical Pharmacology and Toxicology
Faculty of Medicine
University of Belgrade
dr Jovana Subotia 1
11000 Belgrade, Serbia

Abstract: Heart rate and blood pressure short-term variability analysis represent promising quantitative measures of the cardiovascular autonomic controls. The analysis include traditional statistical analytical tools and a number of methods based on nonlinear system theory, recently developed to give better insight into complex HR and BP time series. These methods might reveal abnormalities that may not be uncovered by traditional measures. This paper investigates the measure of entropy of HR and BP time series as a consequence of the involvement of the autonomic nervous system, assessed in conscious telemetred rats under blockade of β-adrenergic, α-adrenergic and M-cholinergic receptors.

1 Introduction

Physiological processes are complex phenomena, outcomes of multiple inputs including autonomic nervous system and humoral controls. Measure of entropy quantifies the unpredictability of fluctuations in a time series, reflecting the likelihood that 'similar' patterns of observations will not be followed by additional 'similar' observations. Therefore, a time series containing many repetitive patterns has relatively small entropy; a less predictable process (with more disorder) has higher entropy.

Entropy evaluation of a time series is an easy task if a process is well described; it may be still reliable if there is an inherent knowledge of time series statistics (Shannon's experiments, e.g. [13]); yet, it can lead to incorrect conclusions and unfounded extrapolations if a time series is controlled by multiple not completely known factors.

The aim of this paper is to try to quantify the contribution of the autonomic nervous system: the adrenergic and the holinergic part to the heart rate and the blood pressure disorder. Pulse pressure signals were recorded in freely moving radiotelemetred rats. Entropies were estimated on systolic arterial pressure (SAP) and heart rate (HR) time series.

2 Materials and Methods

2.1 Animals

Experiments were done in conscious male Wistar outbred rats (320-350 g) during daytime (10-14 h), housed separately in plexiglas cages under standard laboratory conditions with water and food ad libitum.

2.2 Surgery

Rats were submitted to surgical procedure under combined 2% xylazine and 10% ketamine anesthesia during which implants TA11 PA-C40 (Transoma Medical, DSI Inc., USA) were inserted in aorta. After full recovery period (10 days), rats were operated again under halothane anesthesia (4% concentration in the chamber for induction of anesthesia and 1.7% for maintenance under the mask) for quick insertion of catheter in jugular vein for drug injections. Two days later rats were submitted to four different protocols.

2.3 Protocols

Protocol 1 was designed to investigate the role of the sympathetic nervous system directed to the blood vessels in the genesis of the disorder in the cardiovascular signals, by selective blockade of α_1 adrenergic receptors by prazosine (1 mg/kg i.v. bolus continued by 0.5 mg/kg/h infusion) in n=6 conscious rats. *Protocol* 2 was designed to investigate the role of the sympathetic nervous system directed to the heart in the genesis of the disorder in the cardiovascular signals, by selective blockade of β_1 adrenergic receptors by metoprolol (2 mg/kg i.v. bolus continued by 1 mg/kg/h infusion) in n=6 conscious rats. In *protocol* 3 we investigated the

contribution of the parasympathetic part of the autonomic nervous system to the creation of the disorder of cardiovascular signals by blocking the muscarinic receptors by atropine methyl bromide (1 mg/kg i.v. bolus, followed by 0.5 mg/kg/h i.v. infusion). *Protocol* 4 was designed as a control group in which saline (0.9% NaCl) was injected to n=9 rats (1 ml/kg i.v. followed 0.5 ml/kg/h i.v. infusion). In addition, an artificial control signal was simulated by generator of random numbers with normalized and centralized Gaussian distribution.

2.4 Drugs

Prazosin chloride, atropin methyl bromide and metoprolol chloride were purchased from Sigma-Aldrich (Uni-Chem, Belgrade). Ketamine and xylazine were purchase from Richter Pharma (Germany) and Ceva Sante Animal (Hungary), respectively. Halothane was donated by Jugoremedia (Belgrade). All drugs were dissolved in saline (0.9% NaCl).

2.5 Signal Processing

Recording of blood pressure pulse wave was done using DSI radiotelemetry system. Pulse pressure was sampled at 1000 Hz. Systolic blood pressure was obtained by identification of the maxima in the pulse wave signal. Heart rate [in bpm] was calculated from the heart period measured in seconds between two maxima (60/heart period [in s]). The slow components of created time series of SAP and HR were removed using the filter for detrending the biomedical data (λ=10) [10].

$$\mathbf{y}_{\mathrm{DTR}} = (\mathbf{I} - (\mathbf{I} + \lambda^2 \cdot \mathbf{D}^{\mathrm{T}} \cdot \mathbf{D})^{-1}) \cdot \mathbf{y}\,, \quad \mathbf{D} = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -2 & 1 \end{bmatrix} \tag{1}$$

Both original and detrended signals were used to estimate the uncertainty and disorder of time series. This evaluation is based upon the series of probability estimations. It starts from a time series $\mathbf{y} = [y(j)]$, j=1,…,N. A sequence of length m is defined as a short vector $\mathbf{x}_m(i) = [y(i+k\text{-}1)]$, k = 1,…,m, i=1,…,N-m+1 that is a part of a long time series. For each pair of sequences a distance $\mathrm{d}(\mathbf{x}_n(i), \mathbf{x}_m(j))$, i,j=1,…,N-m+1 is defined. It can be maximal absolute distance, mean square distance or any other distance suitable for the current investigation. The number of m-tuples for which the distance is bounded by some value r is B_i, i = 2,…,N-m+1. The function

$$C_i^m(r) = \frac{B_i}{N-m+1} \tag{2}$$

estimates the probability that any vector $\mathbf{x}_m$(j) is within the distance r from the sequence $\mathbf{x}_m$(i). Another function

$$\Phi^m(r) = \frac{1}{N-m+1} \cdot \sum_{i=1}^{N-m+1} \ln\left[C_i^m(r)\right] \tag{3}$$

is average of the natural logarithms of the previous function. It is proven [12, 1] that the entropy of the underlying process can be approximated for the finite data sets:

$$\lim_{r\to 0} \lim_{m\to\infty} \lim_{N\to\infty} \left[\Phi^m(r) - \Phi^{m+1}(r)\right] \approx ApEn(m,r,N) = \Phi^m(r) - \Phi^{m+1}(r) \tag{4}$$

Dependency on the record length and lack of relative consistency of this approach were shown in [7] where a *heuristic* improvement SampEn was introduced. If $B^m(r)$ ($A^m(r)$) is an estimate of the probability that two sequences will match for m (m+1) points (with A_i being a m+1 counterpart of B_i):

$$B^m(r) = \frac{1}{N-m} \cdot \sum_{i=1}^{N-m} \frac{B_i - 1}{N-m-1}, \quad A^m(r) = \frac{1}{N-m} \cdot \sum_{i=1}^{N-m} \frac{A_i - 1}{N-m-1} \tag{5}$$

SampEn is then estimated as [7]:

$$\mathrm{SampEn}(m,r,N) = -\ln\left(\frac{A^m(r)}{B^m(r)}\right) \tag{6}$$

For each pair of HR-BP series a cross-entropy was evaluated. Cross-entropy is an asymmetric measure, evaluated in the same way as ApEn, except that the compared data sets are of different origin (HR and SAP in this case) [6].

The estimates require normalized and (in a case of cross-entropy) centralized time series for which the first and the second moments are required, the features that could be estimated *from stationary data only* [2]! For this reason, a stationarity test frequently employed for biological time series [3,11] has been performed. The data collected from 6 rats (the first three protocols) or 9 rats (the fourth protocol) were segmented into series of 1024 samples each; the ones (either original or detrended) that had not passed the tests considering both the mean value and standard deviation of samples were rejected. Therefore, the study sample included 84 time series for protocol 1; 28 time series for protocol 2; 21 time series for protocol 3 and 12 time series for protocol 4.

While ApEn (and SampEn) are estimates devised for time-series, other entropy measures can also be performed. One of them is devised within the joint symbol dynamics analysis, where two windows of size m slide simultaneously along differentially coded SAP and HR (or, alternatively, from BBI – bit to bit interval) data [8]. Binary content of a window defines a symbol, therefore number of different symbols is 2^m; a relative freqquence of pairs of symbols (observed in two simultaneous windows) is evaluated from the SAP and BBI as:

$$p_{i,j} = \frac{\text{number of occurences of SAP symbol } i \text{ and BBI symbol } j}{(N-m+1)^2}, \quad i,j = 1,\cdots,2^m \tag{7}$$

JSD Shannon is then estimated as:

$$\text{JSDSh} = \sum_{i=1}^{2^m}\sum_{j=1}^{2^m} p_{i,j} \cdot ld\left(\frac{1}{p_{i,j}}\right) \quad [\text{Sh}]. \tag{8}$$

2.6 Statistics

Results in the figures are plotted as mean, while the variance is presented, for particular values, in Table 1. Differences between experimental groups were assessed using Student's T test for unpaired observations in GraphPad Prism 4 software. Statistical significance was considered at $p < 0.05$.

	Protocol 1	Protocol 2	Protocol 3	Protocol 4	Gauss
HR-ApEn	**1.049±0.18**	**1.075±0.09**	**1.018±0.10**	**1.292±0.35**	**1.651±0.12**
SAP-ApEn	**1.347±0.30**	**1.536±0.45**	**1.474±0.30**	**1.093±0.38**	**1.651±0.12**
HRD-ApEn detrended	**1.291±0.25**	1.366±0.21	1.260±0.70	**1.530±0.18**	1.651±0.12
SAPD-ApEn detrended	1.514±0.34	1.481±0.45	1.506±0.69	**1.523±0.38**	1.651±0.12
HR-SampEn	1.028±0.28	1.229±0.10	0.976±0.07	**1.078±0.35**	**1.782±0.16**
SAP-SampEn	1.432±0.24	1.789±0.45	1.765±0.67	**1.298±0.38**	**1.782±0.16**
HRD-SampEn detrended	1.304±0.28	1.533±0.09	1.173±0.10	**1.413±0.35**	1.782±0.16
SAPD-SampEn detrended	1.567±0.30	1.495±0.45	1.568±0.70	**1.592±0.38**	1.782±0.16
JSDSh	2.66±0.24	2.65±0.22	2.68±0.33	**2.69±0.34**	2.73±0.29

Table 1
Mean ± SE (standard error) of entropy estimates for different protocols; gray results differs significantly from control group (Protocol 4)

3 Results

The entropy estimate is done observing the usual measure (m=2) corresponding to pairs of samples vs. triplets of samples, but instead of fixed value for comparison r/σ, a range of values is plotted. ApEn and its detrended counterpart for heart rate and systolic blood pressure are shown in Figs. 1 and 2, while SampEn is shown in

Figs 3 and 4. A sample of cross-entropy is shown in Fig. 5. JSDSh is observed considering the windows of size m=3 (Table 1).

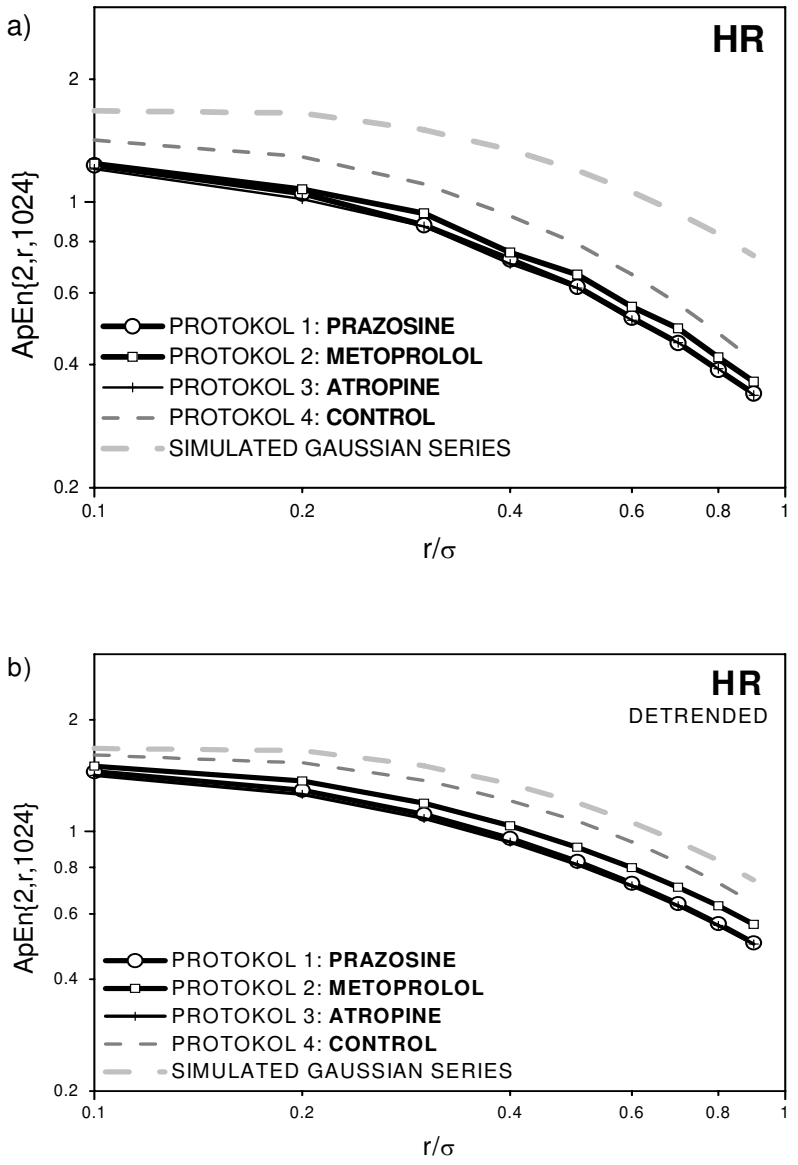


Figure1
ApEn of HR – original (a) and detrended (b) signals

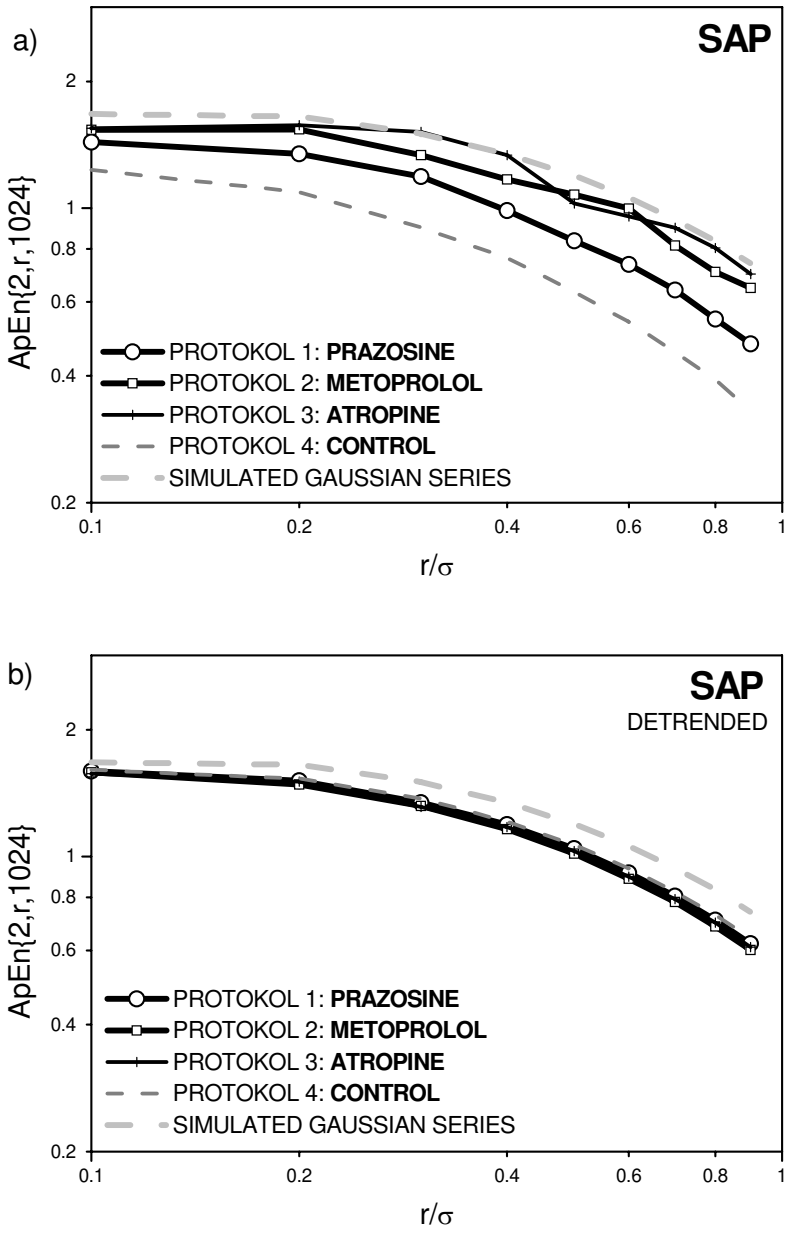


Figure2
ApEn of SAP – original (a) and detrended (b) signals

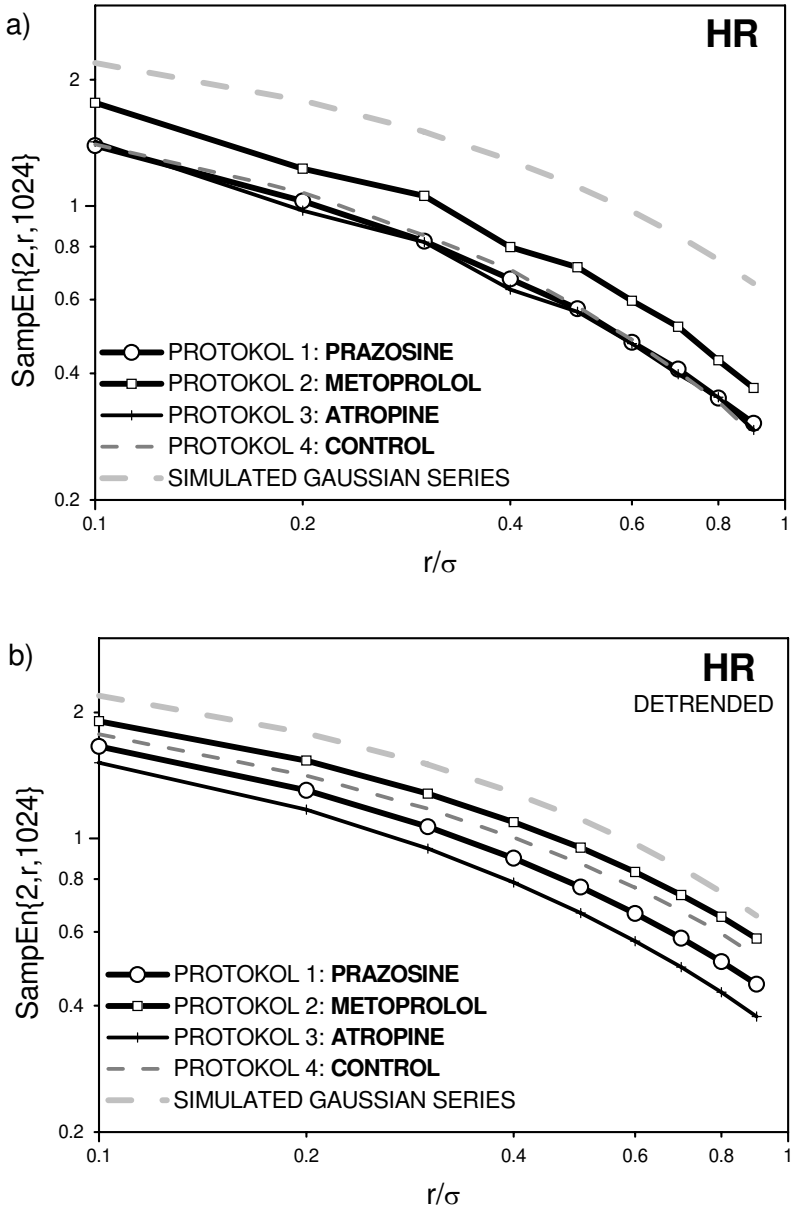


Figure 3
SampEn of HR – original (a) and detrended (b) signals

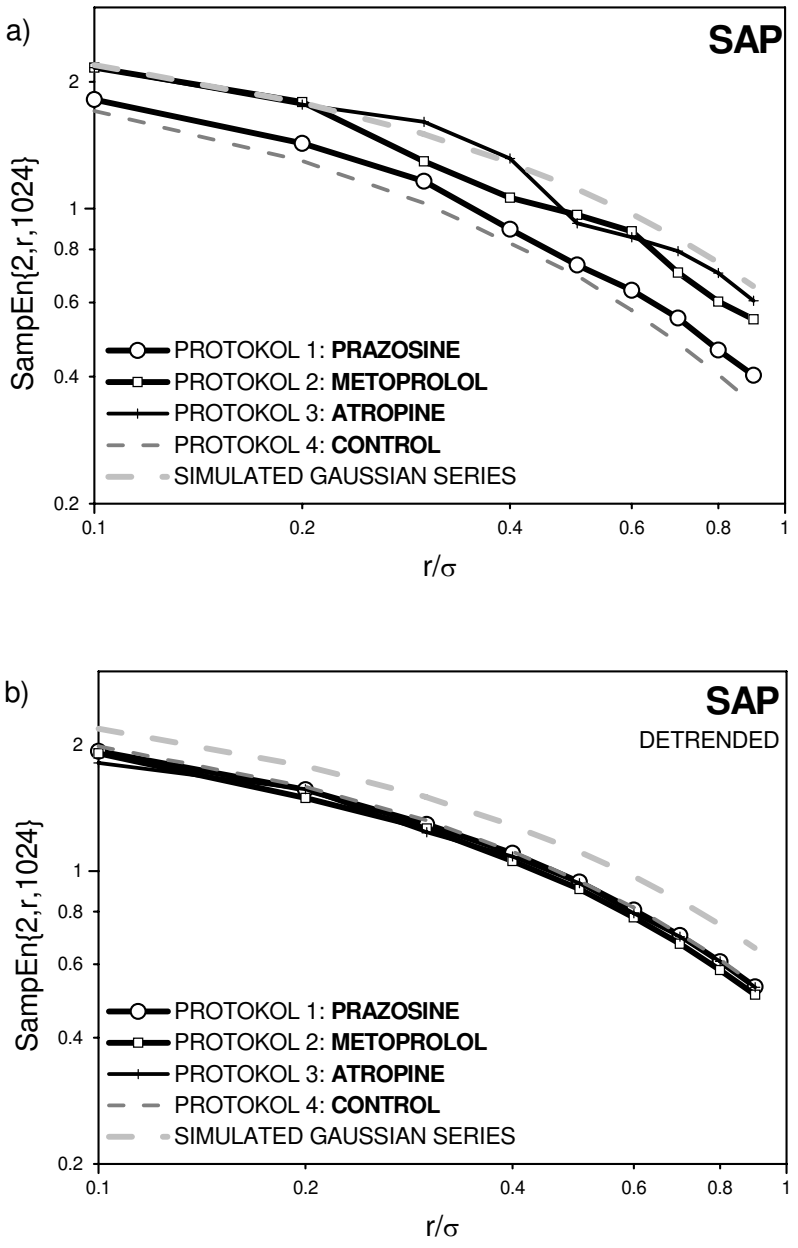


Figure 4
SampEn of SAP – original (a) and detrended (b) signals

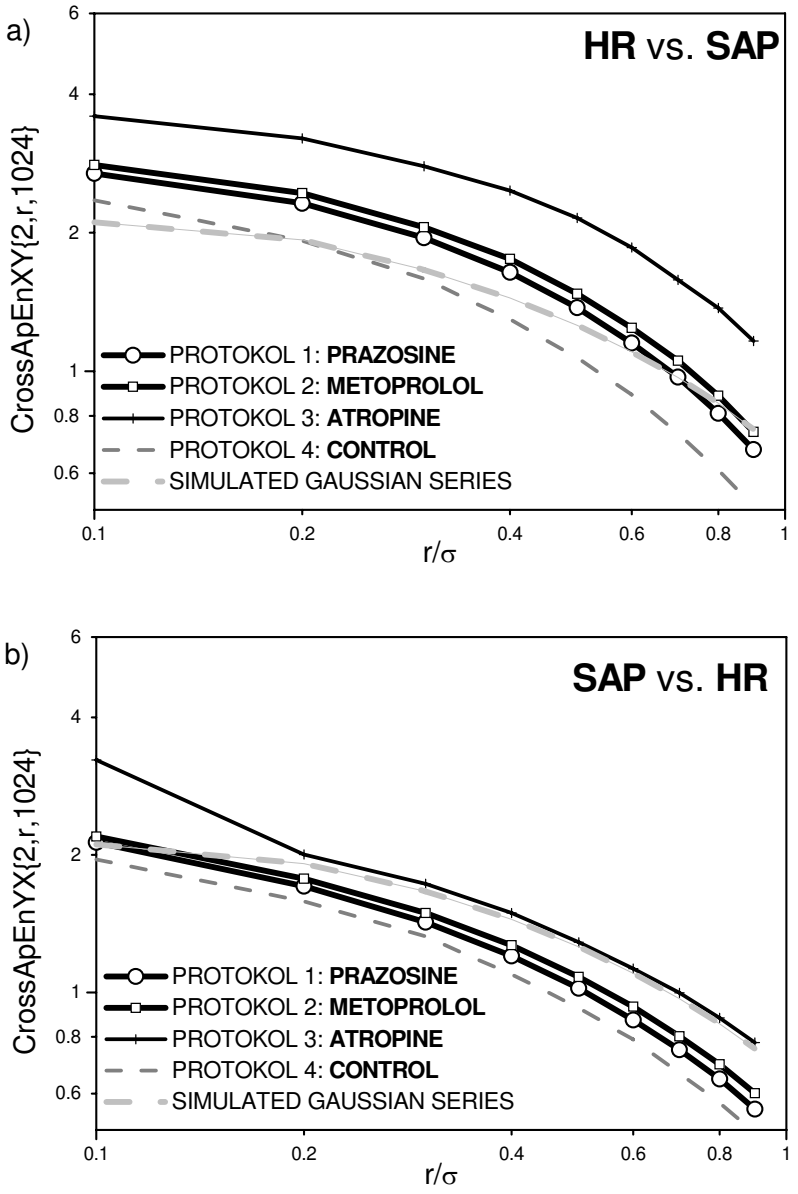


Figure 5
Cross-ApEn HR vs. SAP (a) and SAP vs. HR (b)

4 Discussion

HR and SAP time series show chaotic properties due to functioning multiple regulation mechanisms => high entropy characterizes a healthy cardiovascular system. The autonomic nervous system (functionally divided into adrenergic and cholinergic part) is the fastest acting regulation mechanism that coordinates the changes of blood pressure and heart rate.

As it could be noticed in Fig. 1, ApEn shows the greatest disorder for simulated Gauss signals and the smallest for treated animals. Detrended signal show increase of entropy (some of the predictive components has diminished), but the relative positions of curves for different protocols remain the same as for the original data. Table I shows value samples of sample mean and variance for a characteristic value of $r = 0.2$.

The results are different for blood pressure (Fig. 2) – the smallest entropy is found in control rats, suggesting the possible involvement of autonomic cardiovascular reflexes i.e. the baroreflex. For detrended SAP signals entropy values overlap, quite as expected – the significant amount of SAP signal energy is concentrated within the lower part of spectrum, removed by detrending.

Both the HR and SAP estimates using approach of original data does recognize autonomic contribution to SAP and HR, but does not discern adrenergic from holinergic contribution. Detrended (i.e. filtered) data lose this property for SAP time series. Besides, the detrended data do not show the significant difference regarding the control group, according to t-test. The suggestion is that data should not be detrended, but analyzed in its original form. Also, instead of absolute distance, a normal distance should be measured, to make more refined distinction between the components (drugs) involved.

SampEn approach, on the other hand, has a serious drawback in being a heuristic and not mathematically derived estimate. It is not the same measure, as could be seen from the corresponding equations: roughly speaking, ApEn is a sum of logarithms, while SampEn is a logarithm of sums. Therefore, SampEn *is* a measure, a probabilistic measure, but it is a measure only *similar* to entropy approximation and not an equal one. However, the results obtained by this method did not pass the significance test.

The results of JSDSh (Table I) show extremely high entropy values, compared to previous estimates. It could have been expected, since Eq. (8) is valid for *statistically independent symbols* only [5], while the contents of sliding window certanly form a Markov source: after the symbol (m-tuple, window content) 011, the only possible followers are 101 and 001. Therefore, conditional probabilities $\Pr\{p_{k,l}/p_{i,j}\}$ – relative frequencies of events that pair k,l would follow pair i,j, for $i,j,k,l, = 1,\ldots,2^m$ – must be estimated from differentially coded SAP and BBI time

series. Then a correct formula for statistically dependent symbols could be applied [5]:

$$\text{JSDSh with memory} = \sum_{i=1}^{2^m}\sum_{j=1}^{2^m}\sum_{k=1}^{2^m}\sum_{l=1}^{2^m} p_{i,j} \cdot \Pr\{p_{k,l} / p_{i,j}\} \cdot ld\left(\frac{1}{\Pr\{p_{k,l} / p_{i,j}\}}\right) \quad \text{[Sh]}. \qquad (9)$$

Unfortunatelly, there are $2^{4 \cdot m}$ different transition probabilities $\Pr\{p_{k,l}/p_{i,j}\}$, (4096 for the case m=3), making the estimates quite unrelliably [9], except for very long time series.

Conclusion

ApEn has mathematical foundation and if it is not detrended it does recognize autonomic contribution to BP and HR, but does not discern adrenergic from holinergic contribution. SampEn is a heuristic improvement of ApEn, but it does not recognize autonomic contribution to BP and HR. JSDSh ought to be modified to include the memory of sliding window estimate, therefore it could be implemented with long data series.

Further research should be focused on entropy estimates of the increases and decreases of the signal difference separately. Other analytical methods, both for series of samples and series in time [4] would be included to find out which one show the greatest distinction between the components included; knowing the mechanism of the each one of the methods, more of the complexity of the signal would be revealed.

Acknowledgement

This paper was supported in part by Fundamental research grant no. 145062, Ministry of Science, Serbia.

References

[1] A. L. Goldberger, S. M. Pincus: Physiological Time-Series Analysis: What does Regularity Quantify? Am J Physiol, Vol. 266 (Heart Circ. Physiol.), pp. H1643-H1656, 1994

[2] A. Papoulis: Probability, Random Variables and Stochastic Processes, McGraw-Holl International Edition, 1984

[3] Bendat and Piersol: Random Data Analysis and Measurement Procedures, New York – Wiley Interscience, 1986

[4] D. Hoyer, B. Pompe, H. Friedrich, U. Zwiener, R. Baranowski, U. Müller-Werdan, H. Schmidt: Autonomic Information Flow during Awakeness, Sleep, and Multiple Organ Dysfunction Syndrome assessed by Mutual Information Function of Heart Rate Fluctuations, In Proceedings of IEEE EMBS, San Francisco, CA, USA September, 2004

[5] D. L. Isaacson, R. W. Madsen: Markov Chains Theory and Applications, Wiley, NY, 1976

[6] J. S. Richman, J. R. Moorman: Physiological Time-Series Analysis Using Approximate Entropy and Sample Entropy Am. J. Physiol. Heart Circ. Physiol. Vol. 278(6), pp. H2039-H2049, 2000

[7] L. A. Lipsitz: Dynamics of Stability, J. of Gerontology, Vol. 57A, No. 3, pp. B115-B125, 2002

[8] M. Baumert, V. Baier, S. Truebner et al: Short-and Long-Term Joint Symbolic Dynamics of Heart Rate and Blood Pressure in Dilated Cardiomyopathy, IEEE Trans. on Biomed. Eng, Vol. 52; pp. 2112-2115, 2005

[9] M. Jeruchim: Techniques for Estimating the Bit Error Rate in the Simulation of Digital Communication Systems, IEEE JSAC, Vol. 2, No. 1, pp. 153-170, Jan. 1984

[10] M. P. Tarvainen, P. O. Ranta-aho, P. A. Karjalainen: An Advanced Detrending Method with Application to HRV Analysis, IEEE Trans. on Biomed. Engineering, Vol. 49, pp. 172-175, February 2001

[11] R. Karvajal et al.: Dimensional Analysis of HRV in Hypertrophic Cardiomyopathy Patientes, IEEE Engineering in Medicine and Biology, 21(4), pp. 71-78, July 2002

[12] S. M. Pincus: Approximate Entropy as a Measure of System Complexity, in Proceedings of Nat. Acad. Sci. USA;Vol. 88, pp. 2297-2301, 1991

[13] T. Bell, J. Cleary, I. Witten: Text Compression, Prentice Hall, New Jersey, 1990

Heart Rate Analysis and Telemedicine: New Concepts & Maths

Sándor Khoór[1], István Kecskés[2], Ilona Kovács[3], Dániel Verner[4], Arnold Remete[5], Péter Jankovich[6], Rudolf Bartus[7], Nándor Stanko[8], Norbert Schramm[9], Michael Domijan[10], Erika Domijan[11]

[1, 3] Szent István Hospital, Budapest, Hungary
[3] BION Ltd2, Budapest, Hungary
[2, 4, 5, 6, 7, 8, 9] UVA, Subotica, Serbia
[10, 11] UVA, Toronto, Canada

sandor.khoor@uva.ca, istvan.kecskes@uva.ca, ilona.kovacs@uva.ca, vernerd@gmail.com, rarnold83@gmail.com, fanatick@gmail.com, bartusr@gmail.com, nandor.stanko@gmail.com, norbert.schramm@gmail.com, michael.domijan@uva.ca, erika.domijan@uva.ca

Abstract: Our paper deals with some new aspects of ambulatory (Holter) ECG monitoring extending its indications and using for risk management purpose. Remote sensing consists of the transmittal of patient information, such as ECG, x-rays, or patient records, from a remote site to a collaborator in a distant site. Our earlier developed internet based ECG system was unique for on/off-line analysis of long-term ECG registrations. After the 5-year experience in a smaller region of Budapest, Hungary involving a municipal hospital and the surrounding outpatient cardiology departments and general practitioners, we decided to integrate into our new ECG equipment, the CardioClient the results. In the first clinical study of the four was a wavelet, non-linear heart rate analysis in sudden cardiac death patients using the Internet and the GPRS mobile communication. After the wavelet transformation by the Haar wavelet and the Daubechies 10-tap wavelet, the phase-space of the wavelet-coefficient standard deviation and the scale parameters showed an excellent separation in the scale-range of 3-6 between the two groups: in that region, the average scaling exponents was 0.14+-0.04 for Group-A, and 1.22+-0.27 for Group-B ($p<0.001$). In the next study, we used the Internet database of long-term ambulatory, mobile, GPRS electrocardiograms for the for risk stratification of patients through the cardiovascular continuum. From our ambulatory mobile GPRS ECG database the following a priori groups were defined after a 24 months follow-up: G1: N=227 patients (without manifest cardiovascular disease, clusterized 'boxes' based on the age, sex, cholesterol level, diabetes, hypertension); G2: N=89 patients (postinfarction group); G3: N=66 (patients with chronic heart failure) with (+) or without (-): all-cause death (acD), myocardial infarction (MI), malignant ventricular arrhythmia (MVA), sudden cardiac death (SCD). The actual vs. predicted values were analyzed with chi-square test. The best significance levels ($p<0.001$) were found with method in G1/MI+, G2/SCD+, G3/acD+, G3/SCD+

groups. In the third study a wavelet analysis of late potentials based on long-term, high-resolution, mobile, GPRS ECG data was performed. These pathological changes were also detected by the Haar and Daubechies_4 wavelets, but in a narrower space (110-128 ms and 180-240) and with lesser significance ($p<0.01$). Late potentials were found in Group-A (N=21) in 18 cases with Morlet, 16 with Haar, 19 with Daub-4 analysis, and in 15 cases using all the 3 waves; for Group-B the data were 5, 9, 8, 5, respectively. In the fourth clinical study the prognostic value of the nonlinear dynamicity measurement of atrial fibrillation waves detected by GPRS internet long-term ECG monitoring were analyzed. The multivariate discriminant model selects the best parameters stepwise, the entry or removal based on the minimalization of the Wilks' lambda. Three variables remained finally: x1 = CI mean-value at log r=-1.0 (m9-14), x2 = CI mean-value at log r=-0.5 (m12-17), and x3 = CD_cg. The Wilks' lambda was 0.011, chi-square 299.68, significancy: $p<0,001$.

1 Introduction

Telemedicine can be divided into three areas: aids to **decision-making, remote sensing, and collaborative arrangements for the real-time management of patients at a distance**. As an aid to decision-making, telemedicine includes areas such as remote expert systems that contribute to patient diagnosis or the use of online databases in the actual practice of medicine. Collaborative arrangements consist of using technology to actually allow one practitioner to observe and discuss symptoms with another practitioner whose patients are far away.

2 Technical Aspects

In the older telemedicine wireless system (HeartSpy, Heart Observer), the mobile ECG equipment transfers the continuously stored signals to the WEB server by a GPRS (General Packet Radio System protocol) route. The high resolution (24 bit, 0.5 ms sampling rate) full digital ECG recorder sends the compressed data via wireless network. The size of each ECG packet is 120 byte, the averaged communication bandwidth between 56 and 118kbit/s. The WEB server contains the ECG Knowledge- and Data-base, and broadcasts the ECGs to the medical users. The medical staff could real-time, continuously monitor the patients with PC, or mobile phone [1]. In the CardioClient, the 12-lead ECG is registered and two leads are served for the heart rate analysis. ***Figure 1*** represents the architecture of the internet, wireless ECG system.

Our online monitoring has two meanings. The first is the conventional, i.e. during the registration the cardiologist could observe continuously the ECG. Secondly, some measured non-linear parameters are calculated in every four hours, and a few

hours time-delay a message generate. In the case of significant changes of these non-invasive parameters, the pts get a message to attend the cardiologist. The less significant changes of the parameters elongate the monitoring for days (it means the 'continuous'). During the on-line monitoring, the conventional pathologic ECG signs (arrhythmia, definitive morphological changes) send a warning massage to the patient, the GP and the cardiologist.

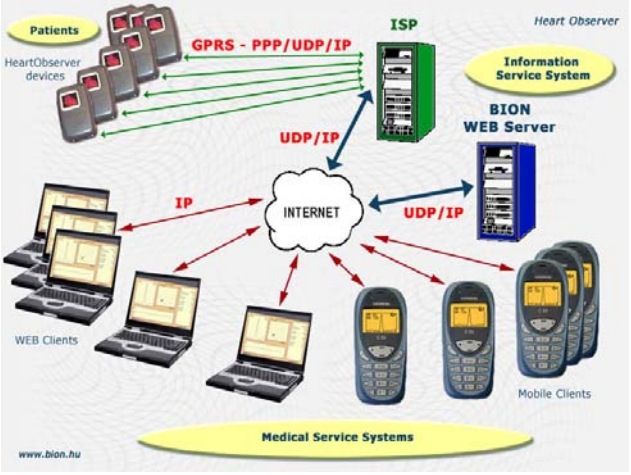


Figure 1
The architecture of the mobile telemedicine system

3 Clinical Concerns

Up to now, the prediction of atrial fibrillation (AF) recurrence and of sudden cardiac death (SCD) is unsolved [2-5]. Sudden cardiac deaths occur >= 75% at home and the patients` chance of arriving the hospital alive is in the range of 3% to 5%. How to do the prevention more effectively and economically - that is the important question. The arrhythmia risk stratification in CHF or other cardiac disease for the improvement of the prognosis and the prevention of SCD is essential. Using our mobile, internet, long-term ECG, the indications of the ambulatory (Holter) ECG could be extended to a larger population. Some works in these areas are presented in this paper, where more sophisticated math analysis was used on the long-term ECG recordings.

3.1 Wavelet, Non-Linear Heart Rate Analysis in Sudden Cardiac Death Patients Using the Internet and the GPRS Mobile Communication

The first aim of our study was to develop an effective ECG surveillance system for the prevention of sudden cardiac death (SCD) and determine the most powerful beat-to-beat heart rate dynamic values as indicators for this monitoring. The second aim was to determine the role of non-linear heart rate variability indexes in a sudden cardiac death (SCD) population [6-8]. The analysis based on the data of 27 chronic heart failure patients with SCD (Group-A) and 27 without (Group-B) it, which was selected from a 168 chronic heart failure patients population monitored for 24 hours two weekly. The inclusion criteria were Holter recordings at least 2 weeks before the SCD (the patients were monitored weekly), absence of acute myocardial infarction in the previous 1 year. SCD was defined as death occurring within 15 minutes of a change in symptoms or during sleep. Clinical features of the two groups: male/female (Group-A: 14/13, Group-B 14/13; age: 62.4±7, 59.7±6; CAD: 22, 21; other cause: 5, 6; NYHA II. class: 18, 17; NYHA III. Class 9, 10; EF: 36.2±6, 36.9±5, respectively).

We used multiresolution wavelet analysis for the 24 hour R-R interbeat time-series. After the wavelet transformation by the *Haar* wavelet and the *Daubechies* 10-tap wavelet, the phase-space of the wavelet-coefficient standard deviation and the scale parameters showed an excellent separation in the scale-range of 3-6 between the two groups: in that region, the average scaling exponents was 0.14±0.04 for Group-A, and 1.22±0.27 for Group-B ($p<0.001$) (***Figure 2***).

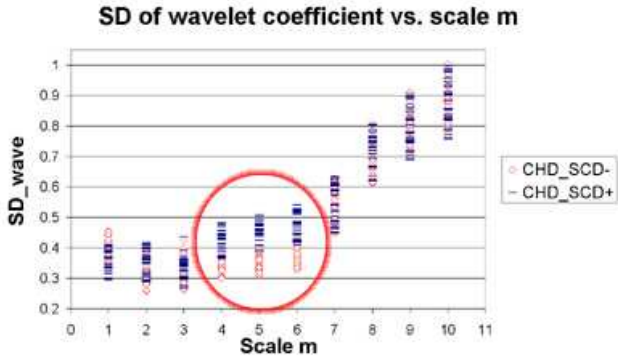


Figure 2
Discrimination value of wavelet coefficient between SCD+ and SCD- patients groups

3.2 Internet Database of Long-Term Ambulatory, Mobile, GPRS Electrocardiograms for the for Risk Stratification of Patients through the Cardiovascular Continuum

In this work we present a telemedicine application using an internet ECG database for risk stratification of patients with various cardiovascular disease state. The data mining [9,10] serves for indexing (finding the most similar time series in the database given a query time series Q, and some similarity/dissimilarity measure D(Q,C). The PAA method was used in this study.

From our ambulatory mobile GPRS ECG database the following a priori groups were defined after a 24 months follow-up: G1: N=227 patients (without manifest cardiovascular disease, clusterized 'boxes' based on the age, sex, cholesterol level, diabetes, hypertension); G2: N=89 patients (postinfarction group); G3: N=66 (patients with chronic heart failure) with (+) or without (-): all-cause death (acD), myocardial infarction (MI), malignant ventricular arrhythmia (MVA), sudden cardiac death (SCD).

The dimensionality reduction via PAA was used, where a time series C of length n can be represented in a w-dimensional space by a vector $\bar{C}=\bar{c1},\ldots \bar{cw}$. The ith element of $\bar{C}$ is calculated:

$$\bar{Ci} = w/n \sum_{j=n/w(i-1)+1}^{(n/w)\,I} C_j$$

The dimension of time series is reduced from n to w, the data is divided into w equal sized frames. The mean value of the data falling within a frame is calculated and a vector of these values becomes the data-reduced representation. The Haar wavelet approximation was used in our study – the principle was the same as in PAA. Each time series was normalized (a mean of zero and a standard deviation of one). A 128 length data segment was used for discretization, the lookup table contains the breakpoints calculated from the Gaussian distribution. The Euclidean distance of two time series Q and C of the same length n is:

$$D(Q,C) = \sqrt{}\,\bar{} \sum_{i=1}^{n} (qi - ci)^2$$

During the 18 months testing period an other 136 patients (the patient groups were identical: G1, G2, and G3) examined and the results (predicted four end-point: acD, MI, MVA, and SCD) formed the a posteriori groups. The actual vs. predicted values were analyzed with chi-square test. The best significance levels ($p<0.001$) were found with method in G1/MI+, G2/SCD+, G3/acD+, G3/SCD+ groups.

3.3 Wavelet Analysis of Late Potentials based on Long-Term, High-Resolution, Mobile, GPRS ECG Data

The time-frequency analysis of left ventricular late potentials is a more sophisticated method than the commercial ones. Some authors used the wavelet method which would be the best solution for the analysis of this kind of data [11-14].

The ECGs was recorded with 32-bit A/D converter and a sampling frequency of 1 kHz, and a modified V1-V3 bipolar leads were used for the registrations. The analysis based on 10,000 (approximately 3 hours) QRS windows starting 100 ms before and 250 ms after the QRS onset. The Haar, the Daubechies_4 and the Morlet (frequency parameter (c) = $2*\pi*5.33$) wavelets were used in the calculations. The Morlet wavelet HR-ECG analysis used a combined wavelet stratification method of Couderc and Selmaoui [10, 11]. In this case, the wavelet transform was applied on 512 ($=2^9$) points, from 128 ms before the beginning of the QRS to 384 ms after QRS onset. Ten different scales were calculated from the modified V1-V3 leads of our GPRS ECG. This single lead was used as the calculated magnitude vector (MV) of standard are differ from the XYZ leads analysis. Using the discrete wavelet transform for the Morlet wavelet, first it computed in the frequency domain and then in the time domain. Using a set of discrete scaling parameters each signal is decomposed into a set of 10 bandpass beat signal logarithmically equally spaced with a centre frequencies from 40 Hz to 250 Hz. The scaling exponent (m) varying linearly between 1.96 and 4.20 by steps of 0.25. The energy was the third dimension of the time-scale plane (means of the wavelet coefficients values at scale s and sample n) which was plotted with a color scale.

The discrete, dyadic wavelet calculations were used for the the Haar, and the Daubechies_4 wavelets. The 4 coefficients of Daub-4 waves: $C_0 = 0.6830127$, $C_1 = 1.1830127$, $C_2 = 0.3169873$, $C_3 = -0.1830127$. The DWT of Haar and Daub-4 wavelet transform was applied on the same 512 ($=2^9$) points as in the Morlet transform.

The study population consists of two postinfarction groups: Group-A: malignant ventricular arrhythmia or sudden cardiac death (N= 21; age: 60.3±11; male:13), Group-B: without them (N= 96; age:61.7±12; male 52). All patients were followed for 24 months, the GPRS 24 hour mobile ECG was repeated monthly.

For the analysis of the two groups the wavelet energy was changed for the mean square of wavelet coefficients, and the statistical p-value of the ANOVA test between the two groups was used. Abnormal time-frequency components were found between 90 and 130 ms after QRS onset in the 55-106 and in the 155-250 Hz frequency range, the p-values were < 0.001 with the Morlet waves (***Figure 3***).

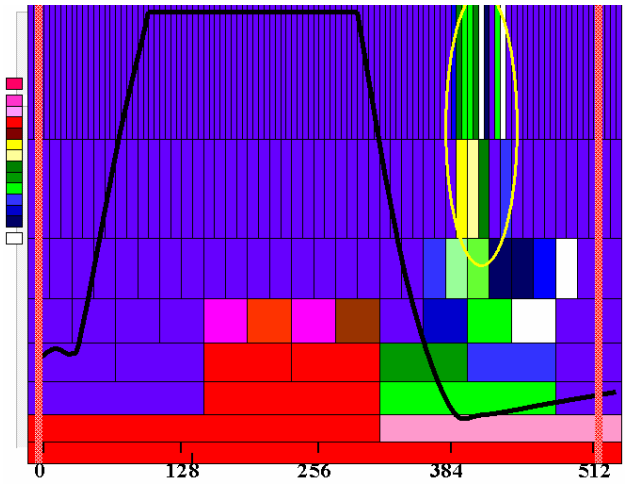


Figure 3
The ellipse shows the Region-of-Interest (ROI) of late potentials in the scalogram. The colors represents the SD values in the time-scale map. Black line: part of the ECG, the time-window: 512 ms.

These pathological changes were also detected by the Haar and Daubechies_4 wavelets, but in a narrower space (110-128 ms and 180-240) and with lesser significance ($p<0.01$).

Late potentials were found in Group-A (N=21) in 18 cases with Morlet, 16 with Haar, 19 with Daub-4 analysis, and in 15 cases using all the 3 waves; for Group-B the data were 5, 9, 8, 5, respectively.

Our work might be the first in the comparative wavelet analysis of long-term mobile, GPRS electrocardiography with high resolution ECG hardware. The results showed that using a very long analyzing segment (10,000 QRS-complex) the signal-to-noise ratio could be beneficially changed for analyzing late potentials of ambulatory wireless ECG data.

3.4 Prognostic Value of the Nonlinear Dynamicity Measurement of Atrial Fibrillation Waves detected by GPRS Internet Long-Term ECG Monitoring

In this study, a five-minute ECG was recorded with our mobile-internet equipment in 68 patients with paroxysmal atrial fibrillation (t<24 hour). Immediately after the arrhythmic episode, a 28-day continuous mobile, internet ECG was recorded for monitoring the atrial fibrillation recurrence. The nonlinear dynamicity of the f-waves was determined with advanced math method. Multivariate discriminant analysis was used analyzing the difference between the two groups (recurrent PAF [Group-2, N=29] or not [Group-1, N=39). The ECG pre-processing consists of the

R-wave detection (smooth – first derivative – largest deflection), the signal averaging in all time windows around the detected R-waves, the determination of template QRS by averaging the deflections in the corresponding time. For the measurement of complexity [15-17], the Grassberger-Procaccia Algorithm (GPA) was used, its main principle is to determine the correlation dimension using the correlation integral.

The CI of a (chaotic) deterministic system is given by

$C_m(r) = Ar^D e^{-mld\Delta tK}$,

A is a constant, D the correlation dimension, K is the correlation entropy, m the embedding dimension, l the embedding delay and Δt the sample interval.

$Cm(N,r)= 1/N_{dis} \Sigma_{i=\{i\_ref\}} \Sigma_{j=/i/>=W} \Theta_{(r-|x(i)-x(j)|)}$

where $N = L-(m-1)l$ is the number of delay vectors resulting from the time series of length L in reconstructed phase space of embedding dimension m. The Heaviside Θ (theta) is 1 for positive arguments and 0 otherwise. The inner sum counts the number of delay vectors within a distance r from a reference vector. The outer sum adds the results over a set $\{i_{ref}\}$ of reference vectors and the normalization factor N_{dis} is the total number of distance involved int he summations. We used 10000 randomly chosen reference vectors, which is equal to one third of the number of samples int he time series (sampled down to 30000 samples) as suggested Theiler.

The steps of the GPA were:

- The Correlation Integral ($C_m(r)$) dimension for different embedding (delayed) dimension (m) is calculated.
- If ($C_m(r)$) shows scaling (=linear part on double logarithmic scale) the Correlation Dimension (D) and Correlation entropy (K) are estimated with coarse-grained D_{cg} and $K_{cg.}$
- If (Cm(r)) shows no scaling a distance r and an embedding dimension m are chosen at which the coarse-grained Dcg and Kcg are estimated. ***Figure 4*** shows the Correlation Integral ($C_m(r)$) dimension for different embedding (delayed) dimension (m); if ($C_m(r)$) shows scaling (=linear part on double logarithmic scale) the Correlation Dimension (D) and Correlation entropy (K) are estimated with coarse-grained D_{cg} and K_{cg}.

The amplitude values of CI, CD, CE at various m were determined with their coarse-grained values. The DSC model selects the best parameters stepwise, the entry or removal based on the minimalization of the Wilks' lambda. Three variables remained finally: x1=CI mean-value at log(r)=-1.0 (m9-14), x2=CI mean-value at log(r)=-0.5 (m12-17), and x3=CD_cg. The Wilks' lambda was 0.011, chi-square 299.68, significancy: $p<0,001$.

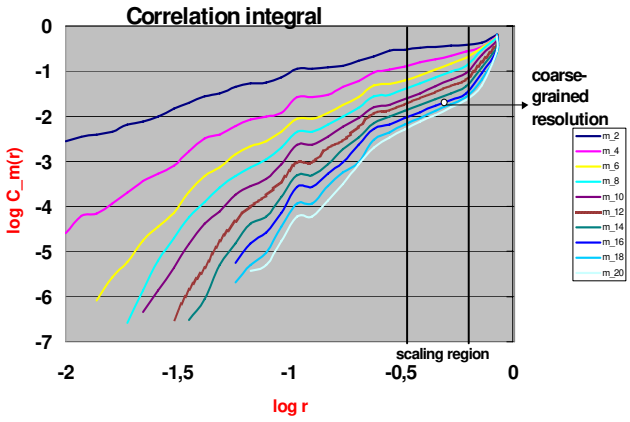


Figure 4
The relation of Correlation Integral ($C_m(r)$) to r value with the scaling region

Acknowledgment

The internet based ECG system combines the advantages of online and offline monitoring. Using various risk scores as predictor values, the telemedicine ECG management would be designed. The worsening indicator parameters indicate immediate change of patient management (re-checking the clinical signs and symptoms, change of therapy, hospital admission). In the case of borderline decision situation (mild change of the indicator values) the ambulatory registration will extend for longer time or repeat more frequently. The accessibility to the standard medical care of the moderate to high risk cardiovascular patients is markedly increased. The frequent and repeated ECG monitoring put the patient to the 'good place in good time' preventing from serious or lethal complications. The opinion of The Task Force Committee on Heart Rate Variability (in 1996!) was: 'At present, the nonlinear methods represent potential tool for HRV assessment. Advances in technology and the interpretation of the results of nonlinear methods are needed before these methods are ready for physiological and clinical studies.' Our studies show that more sophisticated math analysis of heart-rate variability, beat-to-beat analysis in atrial fibrillation would be real tools in cardiology. As repeating our question (Congress of ESC, 2002): 'Internet-based continuous Holter monitoring for the prevention of sudden cardiac death: Is this the Rosetta stone and Who will be Mr. Champollion?'.

References

[1] Khoór S, Nieberl J, Fügedi K, Kail E: Internet-based, GPRS, Long-Term ECG Monitoring and Non-Linear Heart-Rate Analysis for Cardiovascular Telemedicine Management. Computers in Cardiology, 2003; 28:209-212

[2] Malik M et al: Heart Rate Variability. Standards of Measurement, Physiological Interpretation and Clinical Use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Circulation, 93:1043-1065, 1996

[3] Priori SG, Aliot E, Blomstrom-Lundqvist C, et al. Task Force on Sudden Cardiac Death of the European Society of Cardiology. Eur Heart J, 2001; 22: 1374-450

[4] Fuster V, Ryden LE, Asinger RW, et al. ACC/AHA/ESC Guidelines for the Management of Patients with Atrial Fibrillation. Circulation, 2001; 104:2118-50

[5] Grimm W, Glaveris C, Hoffmann J, et al: Arrhythmia Risk Stratification in Idiopathic Dilated Cardiomyopathy Based on Echocardiography and 12-Lead, Signal Averaged, and 24-Hour Holter Electrocardiography. Am H J, 140:43-51, 2000

[6] Huikuri HV, Seppänen T, Koistinen MJ, et al.: Abnormalities in Beat-to-Beat Dynamics of Heart Rate before the Spontaneous Onset of Life Threatening Ventricular Tachyarrhythmias in Patients with Prior Myocardial Infarction. Circulation, 1996; 93:1836-44

[7] Thurner S, Feurstein MC, Teich MC. Multiresolution Wavelet Analysis of Heartbeat Intervals Discriminates Healthy Patients from those with Cardiac Pathology. Phys Rev Lett, 1998; 80:1544-47

[8] Ashkenazy Y, Lewkowicz M, Levitan J, et al.: Scale-Specific and Scale-Independent Measures of Heart Rate Variability as Risk Indicators. Europhys Lett, 2001; 53:709-15

[9] Berry MJA, Linoff G: Data Mining Techniques. J Wiley and Sons, 1997

[10] E Keogh, S Kasetty: On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration. Data Mining and Knowledge Discovery, 2003; (7):349-371

[11] Kennedy HL, Bavishi NS, Buckingham: Ambulatory (Holter) Electrocardiography Signal-Averaging: a Current Perspective. Am Heart J 124:1339-1346, 1992

[12] Morlet D, Couderc JP, Touboul P, Rubel P: Wavelet Analysis of the High-Resolution ECGs in Post-Infarction Patients: Role of the Basic Wavelet and of the Analyzed Lead. Int J Biomed Comp, 39:311, 1995

[13] Couderc JP, Fareh S, Chevalier P, et al: Stratification of Time-Frequency Abnoralities in the Signal-Averaged High-Resolution ECG in Postinfarction Patients with and without Ventricular Tachycardia and Congenital Long QT Syndrome. J Electrocardiol, 29(suppl):180-188, 1996

[14] Selmaoui N, Rubel P, Chevalier P, Frangin GA: Assessment of the Value of Wavelet Analysis of Holter Recordings for the Prediction of Sudden Cardiac Death. Computers in Cardiology, 28:81-84, 2001

[15] Grassberger P, Procaccia I: Measuring the Strangeness of Strange Attractors. Physica D 9:189-208, 1983

[16] F Takens: Detecting Strange Attractors in Turbulence, in Dynamical Systems and Turbulence (Warwick 1980), Lecture Notes in Mathematics, edited by DA Rand and LS Young (Springer, Berlin,1981), Vol. 898, pp. 366-381

[17] Theiler J: Statistical Precision of Dimension Estimators. Phys Rev 41:3038-3051, 1990