ISSN 1729-3901

# Taurida Journal of Computer Science Theory and Mathematics

2013, No. 2

# TAURIDA JOURNAL OF COMPUTER SCIENCE THEORY AND MATHEMATICS

## 2013, No. 2

INTERNATIONAL THEORETICAL RESEARCH EDITION, PEER–REVIEWED JOURNAL
CRIMEAN SCIENTIFIC CENTER OF NATIONAL ACADEMY OF SCIENCES
MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
TAURIDA NATIONAL V. I. VERNADSKY UNIVERSITY

**FOUNDED IN 2002**
Ukrainian Registration Certificate
KB No7826 (04.09.2003)

**According to Ukrainian Supreme Certification Commission Decision No 1-05/4 (26.05.2010), Taurida Journal of Computer Science Theory and Mathematics is entered in the Ukrainian special edition specification for dissertation research publications on Computer Science and Mathematics**

CRIMEAN SCIENTIFIC CENTER OF NATIONAL ACADEMY OF SCIENCES
MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
TAURIDA NATIONAL V. I. VERNADSKY UNIVERSITY



OFFICE ADDRESS:
Taurida Journal of Computer Science Theory and Mathematics,
Crimean scientific center of National Academy of Sciences of Ukraine,
Vernadsky Avenue, 2, Simpferopol, Crimea, 95007, Ukraine

JOURNAL SITE: www.tvim.info

FOR CORRESPONDENCE:
Vernadsky Avenue, 4, Simpferopol, Crimea, 95007, Ukraine
**Tel**. +38 0652 637 542 – editor-in-chief
+38 0652 602 466 – office
**Email**: donskoy@tnu.crimea.edu – editor-in-chief
article@tvim.info – for correspondence

Taurida Journal of Computer Science Theory and Mathematics is a peer-reviewed journal, published by Crimea Scientific Center of National Academy of Sciences of Ukraine. The journal publishes research papers in the fields of computer science and mathematics.

## THEMATIC SECTIONS:

Algorithm Theory, Mathematical Logic, Discrete Optimization, Complexity Theory, Calculus Mathematics, Machine Learning, Pattern Recognition, Data Mining, Deductive Systems and Knowledge Bases, Decision Making Models;

Functional Analysis and Applications, Integral, Differential Equations, Dynamic Systems, Spectral and Evolutional Tasks, The mathematical problems of hydrodynamics.

## CONTENTS

# AN APPROACH TO RECONSTRUCT TARGET FUNCTION OF THE OPTIMIZATION PROBLEM WITH PRECEDENT INITIAL INFORMATION

© **Ayder Anafiyev, Alim Abdulkhairov**

Taurida National V. I. Vernadsky University
Mathematics and Computer Science Department
e-mail: *anafiyev@gmail.com*, *alim.abdulkhairov@gmail.com*

***Abstract.*** *The optimization problem with precedent (training sample) initial information is considered. Some approaches for reconstruction of the target function of such optimization problem are proposed. The open problems that must be solved to obtain better quality solutions of this problem are highlighted.*

## 1. Formulation of the problem

Let $X$, $Y$ and $W$ are the spaces of object (feature), target function value and admissible function value respectively, $f : X \to Y$ is a target function and $\Omega$ is an admissible set of objects. Consider the optimization problem

$$f(x) \to \max_{x \in \Omega \subseteq X} \tag{1}$$

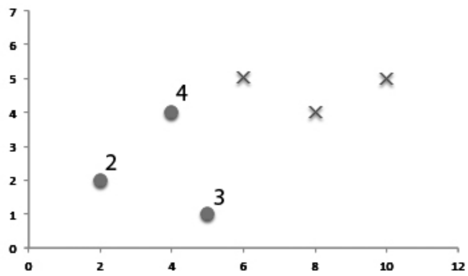
with initial information is represented by the set of triple $X^\ell = (x_i, y_i, w_i)_{i=1}^{\ell}$, where $x_i \in X$, $y_i \in Y$ and $w_i \in W$. The triple $(x_i, y_i, w_i)$ will be called *precedent* (training sample). If $W = \{0, 1\}$ then $w_i = 0$ means that the object $x_i \notin \Omega$, otherwise $(w_i = 1)$ means that the object $x_i \in \Omega$. If $W = [0, 1]$ then $w_i \in W$ could be interpreted as the probability that the object $x_i$ belongs to the set $\Omega$.

The problem (1) will be called the optimization problem with precedent initial information [1, 2, 3]. This is a problem with incomplete information. For solving this problem it's necessary to construct an algorithm which finds in $\Omega$ the optimal object(s) of the target function $f$ or reduces the problem to a certain optimization problem with a fully defined data and which allows to find an effective decision.

The optimization problem could be divided into two problems: the problem of reconstructing of the target function $f$ (regression problem) and the problem of reconstructing the admissible object set $\Omega$ (classification problem). There are many approaches for solving the regression and classification problems. *However it's still open the problem of synthesis of these two problems to get the better quality solution of the given optimization problem.*

**Example 1.** Lets consider the maximization problem with precedent initial information: $X = \mathbb{R}^2$, $Y = \mathbb{R}$, $W = \{0, 1\}$. The training sample $X^\ell$ is set as a training table:

| $x_1$ | $x_2$ | $y$ | $w$ |
|---|---|---|---|
| 2 | 2 | 2 | 1 |
| 5 | 1 | 3 | 1 |
| 4 | 4 | 4 | 1 |
| 6 | 5 | - | 0 |
| 8 | 4 | - | 0 |
| 10 | 5 | - | 0 |




Admissible objects are marked by circles, inadmissible — by crosses. In additional the objects with known values are labeled by the target function values.

As you can see from the figure when we come near to the imaginary border of the space $\Omega$ the value of the target function grows. Obviously this information would be useful for the decision making. The location of objects of different classes ("admissible" and "inadmissible" objects) may be very important during the target function reconstruction.

## 2. Reconstruction of the target function

2.1. **Linear regression.** Lets consider an input space $X = \mathbb{R}^n$ and output space $Y = \mathbb{R}$. The linear regression model $\phi(x, \alpha)$ is represented by

$$\phi(x, \alpha) = \sum_{i=1}^{n} \alpha_j x^j, \quad \alpha_j \in \mathbb{R}, \quad j = \overline{1, n}.$$

The optimal value of parameter $\alpha$ is selected from solution of the optimization problem

$$\alpha^* = \arg\min_{\alpha \in \mathbb{R}^n} \sum_{i=1}^{\ell} L(\alpha, x_i) = \arg\min_{\alpha \in \mathbb{R}^n} Q(a, X^\ell) \tag{2}$$

where $L(\alpha, x)$ is a loss function which is used to determine loss on the object $x$ and $Q(\alpha, X^\ell) = \sum_{i=1}^{\ell} L(\alpha, x_i)$ is an empirical risk.

When defining the loss function it must be considered the fact that we reconstruct the target function of the maximization problem (1). In this case we need more "detailed" study the objects on which the target function takes large values. Therefore the object importance depends on the value of the target function for this object. Thus the loss function will be considered as

$$L_\gamma(\alpha, x) = \gamma(x) L(\alpha, x)$$

where $\gamma(x)$ is the weight function which defines an importance of the object $x$ for the optimization problem.

Let us consider $L(\alpha, x) = (\phi(\alpha, x) - f(x))^2$ and $\gamma(x) = f(x)$.

Introduce the matrix notations $F = (x_i^j)_{\ell\times n}$, $y = (y_i)_{\ell\times 1}$, $\alpha = (\alpha_j)_{n\times 1}$ and $\gamma = \mathrm{diag}(\gamma(x_1), \ldots, \gamma(x_\ell)) = \mathrm{diag}(y_1, \ldots, y_\ell)$.

Let us write the optimization problem (2) in matrix form

$$Q(\alpha, X^\ell) = \gamma\|F\alpha - y\|^2 \to \min_\alpha .$$

The standard way to solve this optimization problem is to use a necessary condition of minimum

$$\frac{\partial Q}{\partial \alpha} = 2F^T\gamma(F\alpha - y) = 0.$$

Therefore

$$F^T\gamma F\alpha = F^T\gamma y.$$

If $F^T\gamma F$ is a nonsingular matrix then the solution of the system will

$$\alpha^* = (F^T\gamma F)^{-1}F^T\gamma y.$$

The result of the using the linear regression method for reconstruction of the target function $f$ of the optimization problem (1) is illustrated on the figure 1. As you can see a more detailed learning of the optimal objects could improve the solution[1].

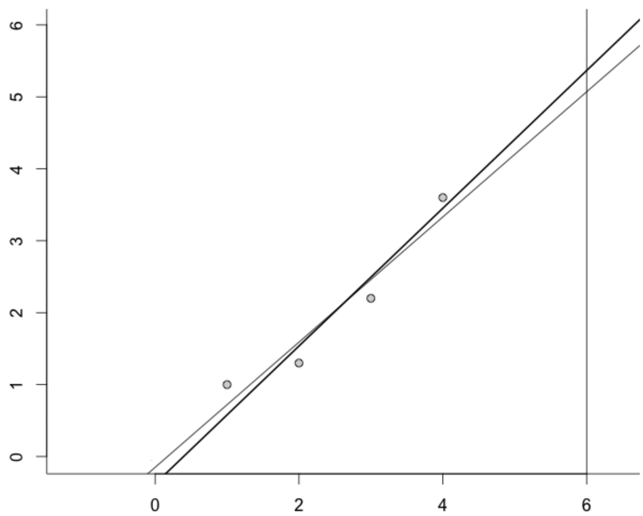



*Fig.* 1. Application of linear regression method for reconstruction of a target function of an optimization problem (bold line — with proposed loss function)

[1]Of course this approach requires more detailed research.

2.2. **Support Vector Regression.** The SVR [4] (Support Vector Regression) could be used to solve the problem of reconstructing the target function.The SVM (Support Vector Machines) method is used for the reconstruction of the admissible object set Ω. The support vector machines is one of the best classification algorithm nowadays. Learning of SVM leads to solving a quadratic (linear) programming problem. The position of the discriminant hyperplane depends only from a few support objects. In addition the use of kernel functions allows the efficient using this method for both: linearly separable and inseparable samples.

The figure 2 shows an example of the optimization problem with the precedent initial information with six objects: two are inadmissible and marked with a cross symbol and the other four belong to the space Ω and marked as a circle (the big radius circles have the larger target function value).

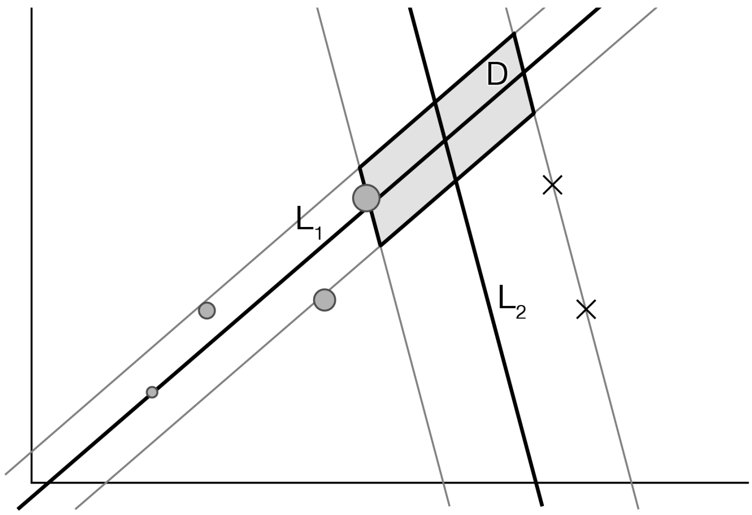



*Fig.* 2. SVR and SVM for solving the optimization problem with precedent initial information.

Line $L_1$ is the result of the solving regression problem and corresponds to the target function. Line $L_2$ is the result of solving classification problem on two classes: "admissible" and "inadmissible" objects. The set $D$ is very interesting from scientific point and needs to be researched.

Consider the applying of SVR method to the reconstruction of the target function of the optimization problem 1. The target function $f(x)$ is represented as

$$f(x) = \langle \alpha, x \rangle + \alpha_0,$$

where $\langle \cdot\,, \cdot \rangle$ is the scalar product.

The optimal set of parameters $(\alpha_1, \dots, \alpha_n, \alpha_0)$ is the solution of the optimization problem

$$\begin{cases} \frac{1}{2}\|\alpha\|^2 + C\sum_{i=1}^{\ell} L_\gamma(\alpha, x_i) \to \min\limits_{\alpha,\alpha_0}, \\ L_\gamma(\alpha, x_i) \leqslant \varepsilon + \xi_i, \\ \xi_i \geqslant 0, i = \overline{1,\ell}. \end{cases}$$

where $\xi_i$ is an error on the object $x_i$.

It's proposed to use $L_\gamma(\alpha, x)$ as a loss function

$$L_{\varepsilon,\gamma}(\alpha, x) = \begin{cases} 0, & \gamma(x)L(\alpha, x) \leqslant \varepsilon \\ \gamma(x)L(\alpha, x) - \varepsilon, & \gamma(x)L(\alpha, x) > \varepsilon, \end{cases}$$

where $\gamma(x)$ is the weight function which defines the importance of the object $x$ for optimization problem (1).

Let us $\xi_x = \gamma(x)L(\alpha, x) - \varepsilon$. Then

$$L_{\varepsilon,\gamma}(\alpha, x) = \begin{cases} 0, & \xi_x \leqslant 0, \\ \xi_x, & \xi_x > 0, \end{cases}$$

and

$$L_{\varepsilon,\gamma}(\alpha, x) = \frac{1}{2}\Big(|\xi_x| + \xi_x\Big).$$

Introduce additional variables $\xi_x^+$ and $\xi_x^-$:

$$\xi_x^+ = \frac{|\xi_x| + \xi_x}{2}, \quad \xi_x^- = \frac{|\xi_x| - \xi_x}{2}, \quad \xi_x^+ \geqslant 0, \quad \xi_x^- \geqslant 0.$$

Note that

$$\xi_x = \xi_x^+ - \xi_x^- \qquad \textit{and} \qquad |\xi_x| = \xi_x^+ + \xi_x^-.$$

As a result the optimization problem is got in the form below

$$\begin{cases} \frac{1}{2}\|\alpha\|^2 + C\sum_{i=1}^{\ell} \xi_{x_i}^+ \to \min\limits_{\alpha,\alpha_0}, \\ \gamma(x_i)L(\alpha, x_i) \leqslant \varepsilon + \xi_{x_i}^+ - \xi_{x_i}^-, \\ \xi_{x_i}^+ \geqslant 0, \quad \xi_{x_i}^- \geqslant 0, \quad i = \overline{1,\ell}, \end{cases}$$

which reduces to the problem of quadratic (linear) programming.

## Conclusion

The optimization problem with precedent (training sample) initial information is considered. The open problems that must be solved to obtain better quality solutions of this problem are highlighted. Proposed the weighted loss function which uses the importance of the object for the optimization problem. It's shown how to use such function for reconstruction of the target function using linear regression and SVR methods. It should be noted that the using of different weight loss functions (with weights depending on an optimization problem) provide more accuracy formalize the optimization problem and obtain better solutions.

## References

1. Anafiyev, A. and Blyschik, V. 2011. Optimization problem with precedent initial information. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 51–57.
2. Anafiyev, A. 2012. The approach to solving optimization problems with precedent initial information. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 7–12.
3. Taratynova, N. 2006. Building an optimization model for precedent initial information as a problem of nonlinear regression. *Iskusstvennyi intellekt*, 2, pp. 238–241.
4. Smola, A. 2004. Tutorial on Support Vector Regression. *Statistics and Computing*, 14, pp. 199–222.

# OPTIMIZATION PROBLEMS WITH PARTIAL DERIVATIVES AND ALGORITHMS FOR CONSTRUCTING GENERALIZED SOLUTIONS


**© Ivan Beyko, Olesya Shchyrba**

National Technical University of Ukraine "Kiev Politechnic Institute"

e-mail: *ivan.beyko@gmail.com*



***Abstract.*** *In the paper we define generalized solutions of the optimization problems for control systems with partial derivatives and develop two types of numerical algorithms for calculating the generalized solutions.*


We consider optimization problems of control systems that are described by the partial differential equations

$$\sum_{(k,i,j)\in K_n^1} a_{kij}^n(t,s)D^{ij}x_k(t,s) + \sum_{(k,i,j)\in K_n^2} b_{kij}^n(t,s)D^{ij}u_k(t,s) = f_n(t,s), n=\overline{1,m},$$

$$D^{ij}x_k(t,s) \triangleq \frac{\partial^{i+j_1+...+j_m}x_k(t,s)}{\partial t^i\partial s^{j_1}\partial s^{j_2}...\partial s^{j_m}}, D^{ij}u_k(t,s) \triangleq \frac{\partial^{i+j_1+...+j_m}u_k(t,s)}{\partial t^i\partial s^{j_1}\partial s^{j_2}...\partial s^{j_m}},$$

$$U \triangleq \{u(t,s)\in R^r | D^{ij}u_k(t,s)\in [u_{kij}^{\min}(t,s); u_{kij}^{\max}(t,s)],\ \ (t,s)\in D_k\subset D,\ \ (k,i,j)\in K_u\}.$$

The optimal control $u\in U$ is defined as the minimizer of the criteria functional

$$J(x,u)\triangleq \max_{l=\overline{1,L_1}} F(x,u,c^l,l)$$

under inequality constraints

$$F(x,u,d^l,l)\le g_l,\ l=\overline{L_1,L_2},$$

where

$$F(x,u,c^l,l)\triangleq \iint\limits_{D_l}\Big(\sum_{(k,i,j)\in K_l^{c1}} c_{1kij}^l(t,s)D^{ij}x_k(t,s) + \sum_{(k,i,j)\in K_l^{c2}} c_{2kij}^l(t,s)D^{ij}u_k(t,s)\Big)dtds+$$

$$+\int\limits_{T_l}\Big(\sum_{(k,i,j)\in K_l^{c3}} c_{3kij}^l(\phi_l(\tau),\psi_l(\tau))D^{ij}x_k(\phi_l(\tau),\psi_l(\tau))+$$

$$+\sum_{(k,i,j)\in K_l^{c4}} c_{4kij}^l(\phi_l(\tau),\psi_l(\tau))D^{ij}u_k(\phi_l(\tau),\psi_l(\tau))\Big)d\tau+$$

$$+\sum_{q\in Q_l}\left(\sum_{(k,i,j)\in K_l^{c5}} c_{5kij}^l(t_q^l,s_q^l)D^{ij}x_k(t_q^l,s_q^l) + \sum_{(k,i,j)\in K_l^{c6}} c_{6kij}^l(t_q^l,s_q^l)D^{ij}u_k(t_q^l,s_q^l)\right).$$

The continue differential functions $c^l_{pkij}, d^l_{pkij}$, $\phi_l(\tau)$, $\psi_l(\tau)$, $a^n_{kij}$, $b^n_{kij}$, $f_n$, and $u^{\min}_{kij}$, $u^{\max}_{kij}$ are defined on the given subsets $D \subset R^2$, $D_l \subset D$, $\{(\phi_l(\tau), \psi_l(\tau))|\tau \in T_l\} \subset D$, $\{(t^l_q, s^l_q)|q \in Q_l\} \subset D$, $K^{c_p}_l, K^{d_p}_l$, $g_l \in R$.

Numerical algorithms for optimal control approximations are based on the reductions of the primary optimal control problem to linear programming. The adequate reduction may be performed, in particular, by replacing partial derivatives $D^{ij}x_k(t,s)$ and $D^{ij}u_k(t,s)$ by correspondent difference approximations of adequate accuracy, and by implementation of appropriate numerical procedures for computing of integrals. The obtained linear programming problem is to be solved by interior point algorithms [1].

In general case of the nonlinear control systems

$$f_0(t,s,x(t,s),u(t,s),\frac{\partial x(t,s)}{\partial t},\frac{\partial u(t,s)}{\partial t},...,\frac{\partial^{\alpha^x} x(t,s)}{\partial t^{\alpha^x_t}\partial s^{\alpha^x_s}},\frac{\partial^{\alpha^u} u(t,s)}{\partial t^{\alpha^u_t}\partial s^{\alpha^u_s}})=0$$

and the nonlinear constraints

$$\iint\limits_{D_l} f_l(t,s,x(t,s),u(t,s),\frac{\partial x(t,s)}{\partial t},\frac{\partial u(t,s)}{\partial t},...,\frac{\partial^{\alpha^x} x(t,s)}{\partial t^{\alpha^x_t}\partial s^{\alpha^x_s}},\frac{\partial^{\alpha^u} u(t,s)}{\partial t^{\alpha^u_t}\partial s^{\alpha^u_s}})dsdt \le 0, l=\overline{1,n_1},$$

$$h_i(t,s,x(t,s),u(t,s),\frac{\partial x(t,s)}{\partial t},\frac{\partial u(t,s)}{\partial t},...,\frac{\partial^{\alpha^x} x(t,s)}{\partial t^{\alpha^x_t}\partial s^{\alpha^x_s}},\frac{\partial^{\alpha^u} u(t,s)}{\partial t^{\alpha^u_t}\partial s^{\alpha^u_s}}) \le 0, i=\overline{1,n_2}$$

iterative gradient methods of linearization and the modified interior point algorithms are used to built extreme controls [1, 2].

The practical example of such multidimensional optimization problem is the following inverse river pollution problem. In mathematical model of the river pollution transfer they denote by $x\,(t,z)$ the concentration of river water pollution at the distance coordinate $z$ (along the river) at the time moment $t$. The value of the concentration $x\,(t,z)$ depends on concentrations$x(t,0) = u_1(t,p)$ at the initial point $z = 0$, on concentrations $x(0,z) = u_2(z,p)$ at the initial time $t = 0$, on the pollution sources intensities $u_3\,(t,z,p)$ at points $z$ (industrial and agricultural production, sewage settlements, etc.), on the rate of flow $v\,(t,z,p)$ and on the coefficient of turbulent diffusion $a\,(t,z,p)$ at different points $z \in [0,b]$. These dependences are approximately described by differential equations with partial derivatives

$$\frac{\partial x(t,z)}{\partial t} = a(t,z,p)\frac{\partial^2 x(t,z)}{\partial z^2} + v(t,z,p)\frac{\partial x(t,z)}{\partial z} + u_3(t,z,p).$$

The solution of the inverse problem in search for pollution sources $u_3\,(t,z,p)$ is based on data measurements of concentrations $X\,(t_i,z_j)$ of river water contaminants at the observation points $z_j$, $j = 1, 2, \ldots, m$, in the time moments $t_i$ and may be calculated as minimizer of the maximum deviation $J\left(u\right)$,

$$J\left(u\right) = \max_i \max_j \left|x\left(t_i, z_j\right) - X\left(t_i, z_j\right)\right|,$$

on the given set $P$ of admissible parameters $p \in P$, that satisfy constraints

$$\left|u_1\left(t, p\right) - U_1\left(t\right)\right| \le C_1(t), \left|u_2\left(z, p\right) - U_2\left(z\right)\right| \le C_2(z),$$

$$\left|a\left(t, z, p\right) - A\left(t, z\right)\right| \le C_3(t, z), \left|v\left(t, z, p\right) - V\left(t, z\right)\right| \le C_4(t, z),$$

$$\left|f\left(t, z, p\right) - F\left(t, z\right)\right| \le C_5(t, z), \left|\frac{du_1(t, p)}{dt}\right| \le D_1(t), \left|\frac{du_2(z, p)}{dt}\right| \le D_2(t),$$

$$\left|\frac{\partial a(t, z, p)}{\partial t}\right| \le D_3(t, z), \left|\frac{\partial a(t, z, p)}{\partial z}\right| \le D_4(t, z), \left|\frac{\partial v(t, z, p)}{\partial t}\right| \le D_5(t, z),$$

$$\left|\frac{\partial v(t, z, p)}{\partial z}\right| \le D_6(t, z), \left|\frac{\partial f(t, z, p)}{\partial t}\right| \le D_7(t, z), \left|\frac{\partial f(t, z, p)}{\partial z}\right| \le D_8(t, z)$$

for the observed averaged values $U_1\left(t\right)$, $U_2\left(z\right)$, $A\left(t, z\right)$, $V\left(t, z\right)$, $F\left(t, z\right)$ of unknown $u_1(t, p)$, $u_2(z, p)$, $a(t, z, p)$, $v(t, z, p)$ and $f(t, z, p)$.

This inverse problem is a particular case of the general optimization problem in search for unknown functions (controls) $u : D \to R^r$ and $x : D \to R^n$, $(t, s) \in D \subset R \times R^{n_s}$, that satisfy integro-differential equations and inequalities

$$\bar{f}^k_{ij}(t, s, x, u) \triangleq f^k_{ij}(t, s, u(t, s), F^{f^k_{ij}}(x, t, s)) = 0, (t, s) \in D^i_j(x, u), k = \overline{1, k_{ij}},$$

$$\bar{g}^l_{ij}(t, s, x, u) \triangleq g^l_{ij}(t, s, u(t, s), F^{g^l_{ij}}(x, t, s)) \le 0, (t, s) \in D^i_j(x, u), l = \overline{1, l_{ij}},$$

where $f^k_{ij}$ and $g^k_{ij}$ are given functions on given subsets $D^i_j(x, u) \subset D$, $j = \overline{1, m+1}$, $D^i_0(x, u) \triangleq \{t^i_q(x, u), s^i_q(x, u)\}^{q_i}_{q=1} \subset D$, $i = \overline{1, i_j}$; $F^{f^k_{ij}}$ and $F^{g^k_{ij}}$ are given compositions of operators $F_1$, $F_2$ and $F_3$:

$$F_1(x, t, s, \alpha, \beta) \triangleq (x(t, s), \frac{\partial}{\partial t}x(t, s), \frac{\partial}{\partial s}x(t, s), ..., \frac{\partial^{\alpha+\beta}}{\partial t^\alpha \partial s^\beta}x(t, s)),$$

$F_2$ is defined by the set $\Omega(t, s) \triangleq \{t^i(t, s), s^i(t, s), \alpha^i, \beta^i\}^{n\Omega}_{i=1}$,

$$F_2(F_1, x, t, s, \Omega) \triangleq (F_1(x, t + t^1(x, t), s + s^1(x, t), \alpha^1, \beta^1),$$

$$F_1(x,t+t^2(x,t),s+s^2(x,t),\alpha^2,\beta^2),\dots,\ F_1(x,t+t^{n\Omega}(x,t),s+s^{n\Omega}(x,t),\alpha^{n\Omega},\beta^{n\Omega}))=$$

$$=(x(t+t^1(t,s),s+s^1(t,s)),...,\frac{\partial^{\alpha^1+\beta^1}}{\partial t^{\alpha^1}\partial s^{\beta^1}}x(t+t^1(t,s),s+s^1(t,s)),$$

$$x(t+t^2(t,s),s+s^2(t,s)),...,\frac{\partial^{\alpha^2+\beta^2}}{\partial t^{\alpha^2}\partial s^{\beta^2}}x(t+t^2(t,s),s+s^2(t,s)),...,$$

$$x(t+t^{n\Omega},s+s^{n\Omega}),\frac{\partial}{\partial t}x(t+t^{n\Omega},s+s^{n\Omega}),...,\frac{\partial^{\alpha^{n\Omega}+\beta^{n\Omega}}}{\partial t^{\alpha^{n\Omega}}\partial s^{\beta^{n\Omega}}}x(t+t^{n\Omega},s+s^{n\Omega})))$$

and $F_3$ is defined by the given operator $\phi$ on the given set $\tilde{\Omega}(t,s,x,u)\subset R\times R^{n_s}$,

$$F_3(x,u,t,s,\phi,\tilde{\Omega})\triangleq\iint\limits_{\tilde{\Omega}(t,s,x,u)}\phi(t,s,u(t,s),F_1(x,t+\tau,s+\sigma,\alpha,\beta))d\tau d\sigma.$$

In search for extremal solution of such generalized optimization problem we may implement subgradient methods. In case of convex functions one use generalized gradient algorithms to calculate approximated global optimal solutions. In this way the parameter set $\Omega(\alpha_r)$ of all the functions $(x,u)$, that satisfy the inequalities

$$\bar{f}^k_{ij}(t,s,x,u)\le\alpha_r,\ \ \bar{h}^k_{ij}(t,s,x,u)\le\alpha_r,(t,s)\in D^i_j(x,u),k=\overline{1,k_{ij}},$$
$$\bar{g}^l_{ij}(t,s,x,u)\le 0,(t,s)\in D^i_j(x,u),l=\overline{1,l_{ij}},$$
$$\bar{h}^k_{ij}\triangleq-\bar{f}^k_{ij},j=\overline{0,m+1},i=\overline{1,i_j}$$

is defined and the generalized solution is defined as a subsequence of the sequence $\{(x_r,u_r)\}_{r=1}^{\infty}\in\Omega(\alpha_r)$, that satisfy the inequalities $B(x_k,u_k)\le\inf\limits_{(x,u)\in\Omega(\alpha_r)}B(x,u)+\alpha_r$ at $\alpha_r\to 0$. The generalized solution is to be calculated by numerical methods [1,2] as a sequence of functions $(x_r(t,s),u_r(t,s))$, belonging to nested sets $X^{n_x(r)}\subset X^{n_x(r)+1}$, $U^{n_u(r)}\subset U^{n_u(r)+1}$ of parametric functions

$$(x_r(t,s),u_r(t,s))\triangleq(x_{n_x(r)}(p_r,t,s),u_{n_u(r)}(q_r,t,s))\in X^{n_x(r)}\times U^{n_u(r)}$$

that are defined by the parameters $p_r\in R^{n_x(r)}$, $q_r\in R^{n_u(r)}$, where for any value $\alpha>0$ there exists a number $r$ for which the parameters $p_r,q_r$ satisfy the inequalities

$$\max_{(t,s)\in D^i_j(x,u)}\bar{f}^k_{ij}(t,s,x_{n_x(r)}(p_r,\cdot,\cdot),u_{n_u(r)}(q_r,\cdot,\cdot))\le\alpha,k=\overline{1,k_{ij}}$$

$$\max_{(t,s)\in D^i_j(x,u)}\bar{h}^k_{ij}(t,s,x_{n_x(r)}(p_k,\cdot,\cdot),u_{n_u(r)}(q_k,\cdot,\cdot))\le\alpha,\ k=\overline{1,k_{ij}},\tag{1}$$

$$\max_{(t,s)\in D_j^i(x,u)} \bar{g}_{ij}^l(t,s,x_{n_x(r)}(p_k,\cdot,\cdot),u_{n_u(r)}(q_k,\cdot,\cdot)) \le 0, l=\overline{1,l_{ij}},$$

$$B(x_{n_x(r)}(p_k,\cdot,\cdot),u_{n_u(r)}(q_k,\cdot,\cdot)) \le \inf_{(x,u)\in\Omega(\alpha)} B(x,u)+\alpha.$$

Numerical algorithms for calculating generalized solutions are given by the following theorem.

**Theorem 1.** *If for each $\alpha > 0$ and for selected sequence of nested sets$X^r$, $U^r$, $r=\overline{1,\infty}$, convex on $(p,q)$ functionals*

$$B(x_r(p,\cdot,\cdot),u_r(q,\cdot,\cdot)), \bar{g}_{ij}^k(t,s,x_r(p,\cdot,\cdot),u_r(q,\cdot,\cdot))$$

*and for the linear functionals*

$$\bar{f}_{ij}^k(t,s,x_r(p,\cdot,\cdot),u_r(q,\cdot,\cdot)), k=\overline{1,k_{ij}}$$

*there exists a number $r$, for which the set of parameters $p\in R^r$ and $q \in R^r$, which satisfy the inequalities (1), has an open subset, then the generalized solution is contained in the sequence $\{x_{n_x(r)}(p_r,\cdot,\cdot),u_{n_u(r)}(q_r,\cdot,\cdot)\}_{r=2}^{\infty}$ and is calculated by the iterative algorithm:*

$$p_{r+1}=p_r-h_r v_r/||v_r||, q_{r+1}=q_r-h_r w_r/||w_r||,$$

$$(v_r,w_r)=\begin{array}{l} \bar{\nabla}_{(p,q)}\bar{f}_{ij}^k(t,s,x_{n_x(r)}(p_r,\cdot,\cdot),u_{n_u(r)}(q_r,\cdot,\cdot)),\ if \\ \qquad \bar{f}_{ij}^k(t,s,x_{n_x(r)}(p_r,\cdot,\cdot),u_{n_u(r)}(q_r,\cdot,\cdot))=z, \\ \bar{\nabla}_{(p,q)}\bar{h}_{ij}^k(t,s,x_{n_x(r)}(p_r,\cdot,\cdot),u_{n_u(r)}(q_r,\cdot,\cdot)),\ if \\ \qquad \bar{h}_{ij}^k(t,s,x_{n_x(r)}(p_r,\cdot,\cdot),u_{n_u(r)}(q_r,\cdot,\cdot))=z, \\ \bar{\nabla}_{(p,q)}\bar{g}_{ij}^k(t,s,x_{n_x(r)}(p_r,\cdot,\cdot),u_{n_u(r)}(q_r,\cdot,\cdot)),\ if \\ \qquad \bar{g}_{ij}^k(t,s,x_{n_x(r)}(p_r,\cdot,\cdot),u_{n_u(r)}(q_r,\cdot,\cdot))=z, \\ \bar{\nabla}_{(p,q)}f_0(x_{n_x(r)}(p_r,\cdot,\cdot),u_{n_u(r)}(q_r,\cdot,\cdot)),\ if\ z\le 0, \end{array}$$

$$z=\max\{\max_{j=\overline{0,m+1}}\max_{i=\overline{1,i_j}}\max_{k=\overline{1,k_{ij}}}\max_{(t,s)\in D_j^i(x,u)} \bar{f}_{ij}^k(t,s,x_{n_x(r)}(p_r,\cdot,\cdot),u_{n_u(r)}(q_r,\cdot,\cdot)),$$

$$\max_{j=\overline{0,m+1}}\max_{i=\overline{1,i_j}}\max_{k=\overline{1,k_{ij}}}\max_{(t,s)\in D_j^i(x,u)} \bar{h}_{ij}^k(t,s,x_{n_x(r)}(p_r,\cdot,\cdot),u_{n_u(r)}(q_r,\cdot,\cdot)),$$

$$\max_{j=\overline{0,m+1}}\max_{i=\overline{1,i_j}}\max_{k=\overline{1,k_{ij}}}\max_{(t,s)\in D_j^i(x,u)} \bar{g}_{ij}^k(t,s,x_{n_x(r)}(p_r,\cdot,\cdot),u_{n_u(r)}(q_r,\cdot,\cdot))\},$$

$$\lim_{r\to\infty} h_r=0, \lim_{r\to\infty} n_x(r)=\infty, \lim_{r\to\infty} n_u(r)=\infty, \sum_{r=1}^{\infty} h_r=\infty, h_r>0.$$

In case of linear optimization problems the optimal solution may be calculated by accelerated algorithms using interior point methods. In this way the original generalized optimization problem is approximated by the LP problem

$$\min c^T x | Ax = b, x \geq 0$$

that is solved simultaneously with the dual problem

$$\max b^T y \mid A^T y + z = c,\ z \geq 0.$$

By the Karush-Kuhn-Tucker theorem the solution of these LP is the solutions of the nonlinear system (and backwards)

$$Ax - b = 0,\ A^T y + z - c = 0,\ ZXe = 0, x \geq 0,\ z \geq 0,$$
$$e = (1,1,...,1),\ X = diag(x), Z = diag(z).$$

To calculate the solution $(x, y, z)$ of the last nonlinear system the Newton's iterative methods may be effectively implemented starting from any interior admissible point $(x^0, y^0, z^0), x^0 > 0,\ z^0 > 0$. At the $k$-th iteration the solution $(\delta x, \delta y, \delta z)$ of the linearized at the point $\left(x^k, y^k, z^k\right)$ Newton system

$$A\delta x = r_p, A^T \delta y + \delta z = r_d, Z^k \delta x + X^k \delta z = r_a,$$
$$r_p = b - Ax^k, r_d = c - z^k - A^T y^k, r_a = -X^k Z^k e$$

is calculated

$$! = \left(X^{-1}Z\right)^{-1},\ ACA^T \delta y = r_p + C(r_d - X^{-1} r_a), \delta x = CA^T \delta y - C(r_d - X^{-1} r_a),\ \delta z = r_d - A^T \delta y$$

To ensure the inequalities $x^{k+1} > 0,\ z^{k+1} > 0$ we calculate

$$\alpha_1 = \min_i \left(\frac{-x_i}{\delta x_i}\right) | \delta x_i < 0,\ \alpha_2 = \min_i \left(\frac{-z_i}{\delta z_i}\right) | \delta z_i < 0,$$

$$\tilde{\alpha} = \min\{\alpha_1, \alpha_2\}, \gamma^k = \left(x^k\right)^T z^k, \tilde{\gamma}^k = \left(x^k + \tilde{\alpha}^k \delta x^k\right)^T \left(z^k + \tilde{\alpha}^k \delta z^k\right),$$

$$\sigma^k = \left(\frac{\tilde{\gamma}^k}{\gamma^k}\right)^2, \mu^k = \sigma^k \left(\frac{\gamma^k}{n}\right),$$

$$r_a = \mu^k e - \Delta_a X^k \Delta_a Z^k e - X^k Z^k e,\ \Delta_a X^k = diag(\delta x),\ \Delta_a Z^k = diag(\delta z),$$

$$ACA^T \Delta y = r_p + C(r_d - X^{-1} r_a),$$
$$\Delta x = CA^T \Delta y - C(r_d - X^{-1} r_a), \Delta z = r_d - A^T \Delta y,$$
$$\left(x^{k+1}, y^{k+1}, z^{k+1}\right) = \left(x^k, y^k, z^k\right) + \alpha\left(\Delta x, \Delta y, \Delta z\right).$$

The approximate solution is obtained at the iteration satisfying the three inequalities $||\Delta x|| < e, ||\Delta y|| < e, ||\Delta z|| < e$. In general case of regular convex optimization problem the polynomial convergence of this algorithm was proved.

## Conclusions

Two types of numerical algorithms for calculating the generalized solutions of the generalized optimization control systems with partial derivatives is proposed: the gradient algorithm for calculating extremal solutions and the Newton type interior point algorithm for calculating the global optimal generalized solutions of linear control systems.

## References

1. Beyko, I. V., Zinko, P. M., and Nakonechny, O. H. 2012. *Problems, methods and algorithms of optimization.* Kyiv, Ukraine: Kyiv University Press.
2. Bejko, I. V. 1998. The unified methodology of solving operators as a new information technology in search for new knowledge and optimized decision making. *Proc. "The Information Technology Contribution to the Building of a Safe Regional Environment", AFCEA, Europe Seminar*, Kiev, 28–30.05.98, pp. 44–50.

# SOLVE-OPERATOR METHODS FOR OPTIMIZATION OF RISK CONTROLLED STOCHASTIC PROCESSES

© Ivan Beyko, Petr Zinko

National Technical University of Ukraine "Kiev Politechnic Institute"
Taras Shevchenko National University of Kyiv. The Faculty of Cybernetics
e-mail: *ivan.beyko@gmail.com, petro.zinko@gmail.com*

***Abstract.*** *In the paper we develop solve-operator methods for high order modelling, simulation and optimization of risk controlled stochastic processes described by general graph-operator control systems with incomplete data.*

The risk management includes increasing of the likelihood and impact of favorable events and reducing of the likelihood and impact of adverse processes. Development of new information technologies and computer based systems for solving risk minimization problems are based on optimization of adequate simulators of risk processes. A simulator is said to be adequate if it's practical implementation meets practical requirements to the allowable time $T(p)$ and error $E(p)$ of the calculations, where $p$ is a vector-parameter of the simulator.

To design the optimal simulator, that minimizes the criterion function $J(p) = KT(p) + E(p)$ one uses available sets of mathematical models (with different aggregation levels and different resolving power) and available sets of sources of useful information. The optimal information sources are evaluated by "functions of information evaluation" (FIE) and the simulators are optimized by their iterative decomposition into optimal subsystems to perform substantiated prediction of risk processes in limited time in uncertain environment [1, 2]. Using FIE the iterative optimization procedures detect (on each iterative step) those of the subsystems that ought to be decomposed and those to be aggregated.

Risk optimization problems belong to the most difficult problems of controlled stochastic processes optimization. Their solution requires either simulation-based stochastic quasi-gradient methods [3] dealing with a general distribution of the random parameters, or special decomposition methods [4, 5] dealing with the distribution approximated by finitely many scenarios. Most of the existing computational methods are applicable only to convex problems and converge to a local minimum of multi-extremal problems [6].

To solve global stochastic non-convex optimization problems one may use the stochastic branch and bound algorithm based on the idea of global deterministic branch and bound algorithms [8]. The branch and bound algorithms are designed to solve those global stochastic non-convex problems, for which one can calculate (within a reasonable time) a grate amount of alternative values of the objective function on allowable control sets.

To make it possible two optimization problems should be solved: the problem of mathematical models/simulators optimization and the problem of decision strategies optimization. In this way we implement solve-operator methods to design stochastic processes simulators and risk processes optimization under parametric uncertainties. In a rather general form the solution $u^*$ of a risk optimization problem may be defined as the minimizer

$$u^* = \arg\min_{u\in\Omega} \bar{F}(u) \tag{1}$$

of a risk function

$$\bar{F}(u) = E \max_{q\in Q} \bar{f}(u, q, \theta) \tag{2}$$

where $u$ is the control input, $q$ is an uncertainty parameter, $\theta$ is a random variable defined on a probability space $(\Theta, \Sigma, \mathrm{P})$, $\bar{f}(u, q, \theta)$ is a random performance function, $\bar{F}(u)$ is the expected performance indicator, $Q$ is a set of uncertainty, and $\Omega$ is a feasible control set.

We will consider time and space multidimensional interdependent risk processes where the random performance function

$$\bar{f}(u, q, \theta) \triangleq \tilde{f}(x(u, q, \theta), u, q, \theta)$$

depends on the stochastic process $x(u, q, \theta)$ simulated by the graph-operator system

$$A(x, u, q, \theta) \triangleq (A_1(x, u, q, \theta), \dots, A_{N_k}(x, u, q, \theta)) = 0, \tag{3}$$

$$A_k(x, u, q, \theta) \triangleq (A_{k1}(x_{k1}, z_{k1}, u_{k1}, q_{k1}, \theta_{k1}), \dots$$

$$\dots, A_{kN_{ks}}(x_{kN_{ks}}, z_{kN_{ks}}, u_{kN_{ks}}, q_{kN_{ks}}, \theta_{kN_{ks}})),$$

$$(x, u, q, \theta) \triangleq \{(x_k, u_k, q_k, \theta_k)\}_{k=1}^{N_k}, \ (x_k, u_k, q_k, \theta_k) \triangleq \{(x_{ks}, u_{ks}, q_{ks}, \theta_{ks})\}_{s=1}^{N_{ks}}.$$

where the $ks$-th subsystem

$$A_{ks}(x_{ks}, z_{ks}, u_{ks}, q_{ks}, \theta_{ks}) = 0 \tag{4}$$

of the graph's $k$-th knot describes interdependences between the $ks$-th subsystem states $x_{ks}$, subsystem controls $u_{ks}$, uncertainty parameters $q_{ks}$, random parameters $\theta_{ks}$, and influences $z_{ks}$ with the subsystem of environment,

$$z_{ks} = \varphi_{ks}(x, u, q, \theta), k = \overline{1, N_k}, s = \overline{1, N_{ks}}. \tag{5}$$

The designing of adequate computational procedures for calculating $x_{ks}(u_{ks}, q_{ks}, \theta_{ks}, z_{ks})$, $\tilde{f}(x(u, q, \theta), u, q, \theta)$, and for calculating optimal solutions

$$u^* = \arg\min_{u\in\Omega} E \max_{q\in Q} \tilde{f}(x(u, q, \theta), u, q, \theta)$$

depends on types and dimensions reducing of all the algebraic, differential, and algebraic-integral-differential equations, that are being implemented to describe the $ks$-th subsystem.

Main difficulties of the optimal solution calculation arise in cases of non-convex multi-extremal performance function $\bar{F}$. There are different numerical algorithms designed for non-convex stochastic optimization. In simple cases, where calculations of $\bar{F}(u)$ may be done for many different alternative $u$, the branch and bound algorithm for stochastic global optimization may be used, capable of solving within a reasonable time small problems with highly non-convex functions and with a large number of local minima. The idea of deterministic branch and bound algorithm is to subdivide the set $\Omega$ into smaller subsets and to estimate from above and from below the optimal value of the objective within these subsets and to delete non perspective subsets from the $\Omega$ partition by using current lower and upper bounds of the optimal value within the subsets. In the stochastic deletion rule they do not delete subsets at each iteration, but only after carrying out a sufficiently large number of iterations, and after deriving an independent estimate of the objective value at the current approximate solution.

To simplify calculation difficulties we may replace too complicated subsystems models (4), (5) by simplified subsystems for which while there is some loss of accuracy using the simplified models, the results actually match fairly closely with the full solution. In this way there ware many successful attempts in searching for adequate approximations of stochastic subsystems $A_{ks}(x_{ks}, z_{ks}, u_{ks}, q_{ks}, \theta_{ks}) = 0$ by some simplified stochastic differential equation subsystems (SDE), that allow simplification of computation procedures for calculating $x_{ks}(u_{ks}, q_{ks}, \theta_{ks}, z_{ks})$. For example, in many cases the adequate simplified approximation models may be described by simple SDE:

$$dx^1_{ks}(t) = a^1_{ks}(u_{ks}, q_{ks}, \theta_{ks}, z_{ks})dt + b^1_{ks}(u_{ks}, q_{ks}, \theta_{ks}, z_{ks})dw(t),$$

$$dx^2_{ks}(t) = x^2_{ks}(t)(a^2_{ks}(u_{ks}, q_{ks}, \theta_{ks}, z_{ks})dt + b^2_{ks}(u_{ks}, q_{ks}, \theta_{ks}, z_{ks})dw(t)),$$

$$dx^3_{ks}(t) = a^3_{ks}(u_{ks}, q_{ks}, \theta_{ks}, z_{ks})x^3_{ks}(t)dt + b^3_{ks}(u_{ks}, q_{ks}, \theta_{ks}, z_{ks})dw(t),$$

or by more general SDE linear systems

$$dx^4_{ks}(t) = (A(u(t), q(t), \theta(t), t)x^4_{ks}(t) + c(u(t), q(t), \theta(t), t))dt + +B(u(t), q(t), \theta(t), t)dw(t)$$

with Brownian movements $w(t) \triangleq (w_1(t), ..., w_m(t))$,

$$dw_i(t) \triangleq w_i(t + dt) - w_i(t), E(dw_i^2(t)) = \sigma_i^2 dt,$$

$E(dw_i(t)dw_j(t)) = 0$ for $i \neq j$.

The trajectories of these models are known to be:

$$x_{ks}^1(t) = x_{ks}^1(t_0) + a_{ks}^1(u_{ks}, q_{ks}, \theta_{ks}, z_{ks})(t - t_0) + b_{ks}^1(u_{ks}, q_{ks}, \theta_{ks}, z_{ks})w(t - t_0),$$

$$x_{ks}^2(t) = \exp((a_{ks}^2(u_{ks}, q_{ks}, \theta_{ks}, z_{ks}) - (b_{ks}^2)^2(u_{ks}, q_{ks}, \theta_{ks}, z_{ks})/2)(t - t_0)+$$
$$+b_{ks}^2(u_{ks}, q_{ks}, \theta_{ks}, z_{ks})w(t - t_0)),$$

$$x_{ks}^3(t) = x_{ks}^3(t_0)\exp(a_{ks}^3(u_{ks}, q_{ks}, \theta_{ks}, z_{ks})(t - t_0))+$$
$$+b_{ks}^3(u_{ks}, q_{ks}, \theta_{ks}, z_{ks}) \int_{t_0}^{t} \exp(a_{ks}^3(u_{ks}, q_{ks}, \theta_{ks}, z_{ks})(t - \tau))dw(\tau),$$

$$x_{ks}^4(t) = \Phi(t)x_{ks}^4(t_0) + \Phi(t) \int_{t_0}^{t} \Psi(\tau)(c(u(\tau), q(\tau), \theta(t), \tau) + +B(u(\tau), q(\tau), \theta(\tau), \tau)dw(\tau),$$

where $\Phi(\cdot)$ and $\Psi(\cdot)$ are the fundamental matrices of the associated homogeneous linear system and it's conjugate system.

In case of the nonlinear SDE

$$dx(t) = a(x(t), u, q, \theta)dt + b(x(t), u, q, \theta)dw(t),$$

$$x(t) \in R^n, a(\cdot) = \{a_i(\cdot), i = \overline{1, n}\}, b(\cdot) = \{b_{ij}(\cdot), i = \overline{1, n}, i = \overline{1, m}\},$$

the appropriate subsystem's risk increments of smooth random risk performance functions $\tilde{\tilde{f}}(t, x(t))$ satisfy the Ito formula

$$d\tilde{\tilde{f}}(t, x(t)) = [\partial_t \tilde{\tilde{f}}(t, x(t)) + a(x(t), u, q, \theta)\partial_x \tilde{\tilde{f}}(t, x(t))dt+$$
$$+0,5b^2(x(t), u, q, \theta)\partial_{xx}^2 \tilde{\tilde{f}}(t, x(t))]dt + b(x(t), u, q, \theta)\partial_x \tilde{\tilde{f}}(t, x(t))dw(t)$$

and the probability density $p \triangleq p(x, t|x_0, t_0, u, q, \theta)$ may be calculated as the solution of the Fokker-Planck equations

$$\frac{\partial p(x, t)}{\partial t} = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}[a_i(x(t), u, q, \theta)p(x, t)] + \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}[\bar{b}_{ij}(x(t), u, q, \theta)p(x, t)].$$

Using the calculated probability density $p$ we calculate $u^*$ as the solution of the significant simplified optimization problem

$$u^* = \arg\min_{u\in\Omega}\max_{q\in Q}\int \tilde{f}(x,u,q,\theta)dp(x,u,q,\theta).$$

In cases of convex optimization problems the global optimal solution $u^*$ may be estimated by stochastic quasi-gradient methods using numerical SDE simulators. For example, the iterative Euler-Maruyama simulator

$$x(t_{i+1}) = x(t_i) + a(x(t_i),u,q,\theta)(t_{i+1}-t_i) + b(x(t_i),u,q,\theta)(w(t_{i+1})-w(t_i))$$

or more accurate Milstein simulator

$$x(t_{i+1}) = x(t_i) + a(x(t_i),u,q,\theta)(t_{i+1}-t_i) + b(x(t_i),u,q,\theta)(w(t_{i+1})-w(t_i))+$$
$$+\frac{1}{2}b(x(t_i),u,q,\theta)b^T(x(t_i),u,q,\theta)((w(t_{i+1})-w(t_i))^2+t_i-t_{i+1}).$$

Using the Ito Formula and the stochastic Taylor expansions of functionals of SDEs many other convergent, consistent, and strictly or marginally stable simulators are developed and may be implemented.

We develop higher order solve operator methods to calculate trajectories $x_{ksi}(t,p,q,\theta)$ of stochastic control processes

$$t_0 = \bar{t}(u_{ks0},q_{ks0},\theta_{ks0},z_{ks0}) \in R, x_{ks0}(t_0) = \bar{x}_{ks0}(t_0,u_{ks0},q_{ks0},\theta_{ks0},z_{ks0}) \in R^{n_{ksx}},$$
$$dx_{ksi}(t) = a_{ksi}(x_{ksi}(t),u_{ksi},q_{ksi},\theta_{ksi},z_{ksi},t,\omega_{ksi}(x_{ksi}(t_i),t_i,q_{ksi},\theta_{ksi},z_{ksi},t))dt, \quad (6)$$
$$t\in[t_i,t_{i+1}],$$
$$t_{i+1} = \tau(x_{ks(i-1)}(t_i),t_i,u_{ksi},q_{ksi},\theta_{ksi},z_{ksi}) > t_i,$$
$$x_{ks(i+1)}(t_{i+1}) = \psi(x_{ksi}(t_{i+1}),t_{i+1},u_{ks(i+1)},\theta_{ks(i+1)},z_{ks(i+1)}),$$

were $\theta_{ksi}\in\mathbb{R}^{n_{ksi}}$ are random vectors defined by adequate evaluated distribution functions $\tilde{F}_{ksi}(\tilde{\theta}_{ksi}|(x_{ks(i-1)}(t_i),t_i,q_{ks(i-1)},\theta_{ks(i-1)},z_{ks(i-1)})$.

For the given $\bar{q}$, $\bar{u}$ and for the given realization $\bar{\theta}$ of $\theta$ the trajectory $x = x(\tau) \triangleq x(\tau,\bar{u},\bar{q},\bar{\theta})$ of the system (6) in the neighbourhood $O(t)\in\prod_i(t_i,t_{i+1})$ of $t\in\prod_i(t_i,t_{i+1})$ may be described by the system (7)

$$dx(\tau)/d\tau = f(x(\tau),\tau), \quad (7)$$

$$f(x(\tau),\tau) \triangleq a(x(\tau),\bar{u},\bar{q},\bar{\theta},\tau,\omega(x(t_i),t_i,\bar{u},\bar{q},\bar{\theta},\tau)) \quad (8)$$

The operator $F$ is said to be an asymptotic solve operator on the interval $\tau\in[t,t+H]\subset O(t)$ for the given function $v\left(x(t+H)\right)$ with respect to the continue function $Z\left(Q(\tau)\right)$ on the trajectory $x$ of the system (7) if for continue functions $p$ from the neighbourhood of $x$ holds the asymptotic neighbourhood

$$F(t,p,H,Z,Q) = v\left(x(t+H)\right) \;+\; \left(O(||Z(Q)||) + O(||p-x||)\right) H||p-x||.$$

And the operator $G(\tau)$ is said to be an s-asymptotic solve operator with respect to the parameter $h$ if for the function $v\left(x(t+h)\right)$ holds the asymptotic equality

$$G(h) = v\left(x(t+h)\right) + O(h^s).G(h) = v\left(x(t+h)\right) + O(h^s).$$

**Theorem 1.** [8] *If* $v\left(x(t+H)\right) \triangleq Q(t+H)x(t+H)$, $Z\left(Q(\tau)\right) \triangleq dQ(\tau)/d\tau + Q(\tau)A(\tau)$, *on the interval the functions* $Q(\tau)$, $A(\tau) \triangleq f'_x\left(p(\tau),\tau\right)$ *and* $Z\left(Q(\tau)\right)$ *are continuous and* $f'_x\left(p(\tau),\tau\right)$ *is a Lipschitz matrices with respect to* $p(\tau)$, *then the asymptotic solve operator* $F$ *is defined by the equality*

$$F(t,p,H,Z,Q) = Q(t+H)p(t+H) + \int\limits_t^{t+H} Q(\tau)\left(f(p(\tau),\tau) - dp(\tau)/d\tau\right)d\tau.$$

**Theorem 2.** *If in the conditions of the theorem 1 the functions* $Q(\tau)$, $A(\tau) = f'_x\left(p(\tau),\tau\right)$, $p(\tau)$ *and* $x(\tau)$ *satisfy on the interval* $\tau \in [t, t+h]$ *the asymptotic equality*

$$dQ\left(\tau\right)/d\tau = -Q\left(\tau\right)A\left(\tau\right) + \mathrm{O}(h^k),\; p\left(\tau\right) = x\left(\tau\right) + \mathrm{O}(h^l),\; p(t) = x(t),$$

*then s-asymptotic solve operator* $G(h)$, $s = k+l+1$, $l \le k$ *is defined by the equality*

$$G(h) = Q(t+h)p(t+h) + \int\limits_t^{t+H} Q(\tau)\left(f\left(p(\tau),\tau\right) - dp(\tau)/d\tau\right)d\tau. \tag{9}$$

The theorem statement follows from the given equalities

$$F(t,p,h,Z,Q) = Q(t+h)p(t+h) + \textstyle\int_t^{t+H} Q(\tau)\left(f\left(p(\tau),\tau\right) - dp(\tau)/d\tau\right)d\tau,$$
$$F(t,p,h,Z,Q) = v\left(x(t+h)\right) + \left(O\left(||Z(Q)||\right) + O\left(||p-x||\right)\right)||p-x||h.$$

Really, it follows

$$G(h) = F\left(t,p,h,Z,Q\right) = v\left(x(t+h)\right) + \left(O\left(||Z(Q)||\right) + O\left(||p-x||\right)\right)||p-x||h.$$

And taking into account

$$dQ(\tau)/d\tau = -Q(\tau)A(\tau) + O(h^k),\;\; p(\tau) = x(\tau) + O(h^l),$$

we obtain

$$\begin{aligned} G(h) &= v\left(x(t+h)\right) + h\left(O\left(||Z(Q)||\right) + O\left(||p-x||\right)\right)||p-x|| = \\ &= v\left(x(t+h)\right) + h\left(O(h^k) + O(h^l)\right)O(h^l), \end{aligned}$$

and for $l \leq k$ we obtain the required equality

$$G(h) = v\left(x(t+h)\right) + O(h^{k+l+1}) = v\left(x(t+h)\right) + O(h^s).$$

From the theorem 2 it follows that for any given $\bar{q}$, $\bar{u}$, and for given realization $\bar{\theta}$ of $\theta$, the $s$-order approximation $\bar{x}(t+h), \bar{x}(t+h) = x(t+h) + O(h^s), s = k+l+1$, the trajectory $x(\tau) \triangleq x(\tau, \bar{u}, \bar{q}, \bar{\theta})$ of the differential equation (7) may be calculated by the asymptotic solver-operator formula (10)

$$\bar{x}(t+h) = p(t+h) + \int\limits_t^{t+h} Q(\tau)\left(f\left(p(\tau),\tau\right) - dp(\tau)/d\tau\right) d\tau, \tag{10}$$

using $p(\cdot)$ and $Q(\cdot)$ that satisfy (11), (12)

$$p(\tau) = x(\tau) + O(h^l), \tag{11}$$

$$dQ(\tau)/d\tau = -Q(\tau)A(\tau) + O(h^k),\ Q(t+h) = I. \tag{12}$$

Using asymptotic solve-operators (10)–(12) we construct many of the following high-order simulators to calculate trajectories of stochastic processes realization (6) and (7). For example, using Lagrange polynomials

$$\begin{aligned} p_{n+1}(\tau) = \tfrac{1}{h^n} \Big[ & x(t) \tfrac{(\tau-t-h)\ldots(\tau-t-nh)}{(-1)\cdot(-2)\ldots(-n)} + \\ & + x(t+h) \tfrac{(\tau-t)(\tau-t-2h)\ldots(\tau-t-nh)}{1\cdot(-1)\cdot(-2)\ldots(-(n-1))} + \ldots \\ & + x(t+nh) \tfrac{(\tau-t)(\tau-t-h)\ldots(\tau-t-(n-1)h)}{n\cdot(n-1)\ldots 2\cdot 1} \Big]. \end{aligned}$$

for the given values $x(t+ih),\ \ i = \overline{0,n}$ we obtain the high-order simulators

$$\begin{aligned} x\left(t+(n+1)h\right) = p_{n+1}\left(t+(n+1)h\right) + \textstyle\int_t^{t+(n+1)h} \left[E - (\tau - t - (n+1)h) \times \right. \\ \left. \times f'_x\left(p_{n+1}\left(t+(n+1)h\right), t+(n+1)h\right)\right] \left[f\left(p_{n+1}(\tau),\tau\right) - \dot{p}_{n+1}(\tau)\right] d\tau, \end{aligned}$$

with simulators error $O(h^{n+3})$. And using the Newton-Cotes formula we obtain a number of numerical simulators

$$\begin{aligned} x\left(t+(n+1)h\right) = {} & p_{n+1}\left(t+(n+1)h\right) + \\ & + (n+1)h \textstyle\sum_{i=0}^{n+1} c_{i,n+1} \left[f\left(p_{n+1}(t+ih), t+ih\right) - \dot{p}(t+ih)\right] - \\ & - (n+1)h^2 f'_x\left(p_{n+1}\left(t+(n+1)h\right), t+(n+1)h\right) \textstyle\sum_{i=0}^{n+1} c_{i,n+1}(i-n-1) \times \\ & \times \left[f\left(p_{n+1}(t+ih), t+ih\right) - \dot{p}_{n+1}(t+ih)\right]. \end{aligned}$$

with estimated errors $O(h^{n+3})$. Using the Tylor's formula

$$\begin{aligned} x(t+mh) = \int\limits_t^{t+mh} & \left[E - (\tau - t - mh) f'_x\left(p(t+mh), t+mh\right)\right] \times \\ & \times \left[f\left(p(\tau),\tau\right) - \dot{p}(\tau)\right] d\tau + p(t+mh), \end{aligned}$$

we obtain numerical simulators with the error estimate $O(h^{s+2})$.

Similar high-order simulators are constructed to calculate the probability densities using Fokker-Planck equations. Numerical experiments proved the practical efficiency of the designed high-order simulators implementation to calculate $u$ minimizing the risk function

$$\mathrm{E} \max_{q \in Q} \tilde{f}(x(u,q,\theta),u,q,\theta)$$

using stochastic generalized gradient methods [9] and stochastic minimax algorihms [10].

## Conclusions

The developed high order solve-operator methods may be implemented to solve problems of the general graph-operator stochastic control systems modelling, simulation and optimization under incomplete data.

## References

1. Beyko, I. 1988. Extremal models of complex systems and decomposition methodology in numerical experiments. *Vestnik Kyiv University, Modeling and optimization of complex systems*, pp. 72–81.
2. Beyko, I. 1996. Functions for evaluation of information in the theory of optimal aggregate models. *Cybernetics and System Analysis*, 3, pp. 43–54.
3. Ermoliev, Yu. M. 1983. Stochastic quasi-gradient methods and their application to systems optimization, *Stochastics*, 4, pp. 1–37.
4. Mulvey, J. M., and Ruszczyn'ski, A. 1995. A new scenario decomposition method for large-scale stochastic optimization. *Operations Research*, 43, pp. 477–490.
5. Rockafellar, R. T., and Wets R. J.-B. 1991. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, 16, pp. 1–23.
6. Mikhalevich, V. S., Gupal, A. M. and Norkin, V. I. 1996. *Methods of Non-Convex Optimization.* Moscow: Nauka.
7. Norkin, V. I., Pug, G. Ch. and Ruszczyn'ski, A. 1996. A Branch and Bound Method for Stochastic Global Optimization. *IIASA WP-96-xxx, April 1996*, pp. 25.
8. Beyko, I. V. and Zinko, P. M. 2009. High-order methods for solving the Cauchy problems and the multidimensionalvalue problems by using an asymptotically-solving operators.*Mathematical and computer modeling* (Kamenets-Podolsky National University), 1, pp. 18–25.
9. Ermoliev, Y. M. and Norkin, V. I. 1997. Stochastic generalized gradient method with application to insurance risk management. IR-97-021, Int. Inst. for Appl. Syst. Anal., Laxenburg, Austria.
10. Beyko, I. V., Zinko, P. M., and Nakonechny, O. H. 2011. *Problems, methods and algorithms of optimization.* National University of Water Management and Nature Resources, Rivne.

# INCOMPLETENESS OF INITIAL INFORMATION AND THE PROBLEM OF PAYOFF FUNCTION RECONSTRUCTION

© Vladimir Blyshchik

Taurida National V. I. Vernadsky University
Mathematics and Computer Science Department
e-mail: *veb@land.ru*

***Abstract.*** *The paper introduces the classification of informational situations for a zero-sum game with incomplete information based on uncertainty level. For each case the possible ways to deal with uncertainty are considered.*

Zero-sum matrix game (antagonistic game) is one of the most popular game-theoretical models widely used in theory and practice [1, 2].

Game-theoretical modeling usually assumes the knowledge of all game components such as the set of all pure strategies and values for all entries of the game payoff matrix. However in practice not all values of payoff matrix elements for the antagonistic game which simulate a decision-making problem are possible to know. This prevents a wide use of game-theoretical models in decision support systems (DSS).

We will call a Partially Defined Antagonistic Game the following generalized form of two-person zero-sum game

1. Given is the set $X = \{1, 2, \ldots, m\}$ of all pure strategies of player I numbered with natural numbers $1, 2, \ldots, m$;
2. Given is the set $Y = \{1, 2, \ldots, n\}$ of all pure strategies of player II numbered with natural numbers $1, 2, \ldots, n$;
3. The payoff matrix $A = A_{m\times n} = (a_{ij})$ is given partially (values for some payoff matrix entries $a_{ij}$ are omitted).

Similar to the classical antagonistic game, $A_{ij}$ represents the winnings of player I when player I chooses pure strategy $i$ and player II chooses pure strategy $j$. The winnings of player I are equal to the loses of player II.

For a partially defined antagonistic game there are numbers $i_0 \in X$ and $j_0 \in Y$ such that the values of correspondent payoff matrix entries $A = A_{m\times n} = (a_{ij})$ are unknown. In the general case a payoff matrix may contain a lot of such elements.

Up to now the problem of getting solutions for the partially defined antagonistic games is described very scarcely in the modern scientific literature.

Different concepts of decision search for partially defined antagonistic games are possible under conditions of risk and uncertainty. The natural way of solution search for a partially defined antagonistic game lies in its correct reduction to some classical antagonistic game. The solution of such the antagonistic game with completely defined payoff matrix can be treated as an optimal solution of the initial partially defined antagonistic game.

Decision-making game model is given by a triplet $< X, Y, R >$, where $X = \{1, 2, \ldots, m\}$ is a set of pure strategies for player I, $Y = \{1, 2, \ldots, n\}$ is a set of pure strategies for player II, $A = A_{m\times n} = (a_{ij})$ is a partially defined payoff matrix of the antagonistic game. There is at least one or several elements $a_{ij}$ with unknown values. Our task is to find optimal strategies (possibly mixed) for players in a partially defined antagonistic game.

Classification of possible informational situations is given below. It is similar to the classification of informational situations given in [3, 4] where a comparison criteria is based on uncertainty level the Nature player encounters while choosing a possible state.

We shall call an Informational Situation (IS) the gradation level characterizing the uncertainty of elements $a_{ij}$ from a partially defined payoff matrix $A = A_{m\times n} = (a_{ij})$.

Informational situation classification can be represented by the following gradation

1. $I_1$ — the first IS: unknown elements of payoff matrix are all random values described by a known distribution;
2. $I_2$ — the second IS: all unknown elements of payoff matrix are represented by functions of one or several parameters;
3. $I_3$ — the third IS: all unknown elements of payoff matrix are restricted by a range of values;
4. $I_4$ — the forth IS: there is no any mathematical information about unknown elements of payoff matrix;
5. $I_5$ — the fifth IS: all unknown elements of payoff matrix takes the worst values for player I that is values preventing player I from reaching his/her aims;
6. $I_6$ — the sixth IS: all unknown elements of payoff matrix belong to a given fuzzy set [4], these elements are represented by fuzzy variables with known membership functions;
7. $I_7$ — the seventh IS: IS intermediate between $I_1$ and $I_6$.

Lets note the particular quality of $I_4$. The situation when all the elements $a_{ij}$ of payoff matrix $A = A_{m\times n} = (a_{ij})$ are unknown is forbidden only for informational situation $I_4$. Indeed, if a payoff matrix is completely unknown in situation $I_4$ then the formalization of completely undefined zero-sum game loses any mathematical meaning.

In virtually all cases of informational situations $I_l$ it is possible to evaluate unknown elements of the payoff matrix by interpolating (or extrapolating) corresponding functions or by using pattern recognition methods.

Consider some possible ways to deal with uncertainty.

In case $I_1$ all unknown elements of payoff matrix are all random values given by a distribution law. In this case it is reasonable to change all elements of the payoff matrix (which are the given random variables) with values of numerical characteristics of the corresponding probability distribution such as mathematical expectations, modal values, as well as variances, standard deviations, coefficients of variance, and other numerical characteristics of these random variables.

In case $I_2$ all unknown elements of payoff matrix are represented by given functions of one or several parameters. One approach to solving the partially defined antagonistic game for this case is based on investigating the effect of possible values of these parameters on the optimal solution of the corresponding game. For some cases this investigation will lead to consideration of analytic (functional) dependencies of the optimal solution. In other cases it will be based on the search over the finite set of the most typical (or most important) parameter values. Moreover, it is possible that the mathematical idea behind the partially defined antagonistic game under consideration requires either a single optimal solution, or a number of optimal solutions which are equivalent with respect to the chosen decision criterion. In this case the final choice of optimal solution may require other approaches (e.g. the operation research methods or the methods of expected utility theory).

In case $I_3$ all unknown entries of payoff matrix are restricted by a range of values. For example, the range of unknown elements can be defined by minimal and maximal values with the inequalities of the form $r_{i_0j_0}^{min} \le a_{i_0j_0} \le r_{i_0j_0}^{max}$. Here $a_{i_0j_0}$ is a payoff matrix element with unknown true value, $r_{i_0j_0}^{min}, r_{i_0j_0}^{max}$ are given numbers satisfying the strict inequality $r_{i_0j_0}^{min} < r_{i_0j_0}^{max}$. In such cases one can try an approach based on search among the most typical (and/or most important) values of the corresponding elements of the payoff matrix true values of which are unknown but should meet given restrictions. Though this approach entails a considerable increase in computational operations needed to solve a number of zero-sum games with completely defined payoff matrices.

In case $I_4$ we have only some elements of the payoff matrix. This enables us to say that we have an initial information (a learning data set) which can be used for restoration the unknown elements by the method of empirical generalization. In this case the initial information is treated as a training set containing all necessary information about the matrix. Assuming that there is a regularity (payoff function $H : X \times Y \to \mathbb{R}$) exhibited by the training set we can tackle the problem of function restoration which is incorrect in general case $H$.

In case $I_5$ all unknown elements of payoff matrix takes the values preventing player I (Decision Maker, DM) from reaching his/her aims. Here the economical or physical meaning of the payoff matrix elements plays a crucial role. In case $I_5$ uncertainty is considerably reduced especially when the players are enabled to use mixed strategies. In this case it is possible to create a payoff function of the zero-sum game. The unknown elements of the payoff function can be treated as some parameters which most typical values yield the lowest price for the game.

In case $I_6$ all unknown elements of payoff matrix are fuzzy variables with known membership functions. According to the definition of fuzzy set each element of the payoff function which value is unknown takes values from a definite set of numbers. These values are the elements of the corresponding fuzzy set of known reliability. The reliability function is defined on all elements of the set of numbers and maps it on numbers within the interval $[0, 1]$. In some cases, the values bringing the maximum of the reliability function can be uniquely detected. The unknown elements of the payoff matrix are to be substituted with these values. This replacement turns the partially defined antagonistic game into the classical zero-sum game with all known elements in natural way.

In case $I_7$ the solution of partially defined antagonistic game assumes an approach based on combination of above-mentioned methods of reducing partially defined antagonistic games to classical completely defined games. This combination severely depends on the unknown entries of the payoff matrix. For this case there are more then two unknown elements of the payoff matrix and these elements can be divided into several groups so that each group is represented by its own IS $I_l$, where $l = \overline{1, 6}$.

Review of possible informational situations allows to conclude the following:

1. A partially defined antagonistic game is a zero-sum matrix game with a payoff matrix containing a number of entries with unknown values.
2. One way of solving partially defined antagonistic game is based on its reduction to one or more completely defined zero-sum games. To evaluate the unknown values of the payoff matrix elements one can be use the algorithms of interpolation, extrapolation, as well as methods of pattern recognition.
3. The approach to solving partially defined antagonistic game depends on the informational situation in hand that characterizes the type and the level of uncertainty of the values of the the payoff matrix elements.
4. There are seven basic informational situations that characterize the level of uncertainty of the partially defined payoff matrix of the game.
5. The optimal solution search for the partially defined zero-sum game can contain the solutions of several completely defined zero-sum games.

## References

1. John von Neumann and Oskar Morgenstern. 2007. *Theory of Games and Economic Behavior* (60th Anniversary Commemorative Edition). Princeton: Princeton University Press.
2. Vorobyov, N. N. 1984. *The Basics Of Game Theory. Non-cooperative game*. Moscow: Nauka.
3. Vitlinskiy, V. V. and Nakonechniy, S. I. 1996. *Risk management.* Kiev: TOV Borisfen-M.
4. Trukhayev, R. I. 1981. *Models of decision making under uncertainties.* Moscow: Nauka.

# MATRIX "FEATURE VECTORS" AND GROUPING OPERATORS IN PATTERN RECOGNITION

© **Volodymyr Donchenko, Taras Zinko**

Taras Shevchenko National University of Kyiv

The Faculty of Cybernetics

e-mail: *voldon@bigmir.net, tzinko@ukr.net*

***Abstract***. *Problem of grouping information: recovering function, represented by its observations, and the of classification (problem) clusterization problem, — is of great importance for applied research. Choice of math object which represent the object under investigations largely determines the effectiveness: scalars, vectors or objects of other kinds. Such choice is determined by the richness of mathematical structures within which "representatives" are investigated. Euclidean spaces $R^n$ are common in this choosing. Euclidean spaces of $R^{m\times n}$ of all $m\times n$ matrices are natural as a math structure for "representatives", but the handling technique for such spaces is poorer in comparison with vector space. Just the development of the technique handling" for Euclidean space of $R^{m\times n}$, including SVD and Moore-Penrose inversion for the linear operators, constructive construction of orthogonal projectors and grouping operators for matrix spaces is the subject of the article. Important "grouping statements" about minimal ellipsoid, which covers elements of fixed sequence of matrices in $R^{m\times n}$ is represented. This statement generalize correspondent results for real valued vectors. "Grouping statements" is proposed to be the base for constructing correspondence distance in solving clusterization problem.*

## Introduction

The problem of grouping the information (grouping problem) is the fundamental problem of applied investigations. It appears in various forms and manifestations. All of them eventually are reduced to two forms. Namely, these are: the problem of recovering the function represented by their observations and the problem of clustering, classification and pattern recognition. State of art in the field is represented perfectly in [23, 25, 11, 10, 3].

It's opportune to mark what the information regarding the object or a collection of similar object is exposed to aggregating is. It is of principal importance that an object is considered as a set of its main components and fundamental for the object ties between them. Such consideration and only this one enable application of the math in object description, namely, for math modelling. It is due the fact that after Georg Cantor the objects of investigation in math (math structures) are the sets plus "ties" between its elements. There are only four (may be, five) fundamental mathematical means to describe these "ties". Namely, these are: relations, operations, functions and collections of subsets (or combinations of mentioned above). Thus, the mathematical description of the object (mathematical modelling) can not be anything other than representing the object structure by the means of mathematical structuring. It is applicable to the full extent to that objects which indicated by the term "complex system". A "complex system" should be understanding and, correspondingly, determined, as an objects with complex structure (complex "ties"). Namely, when reading attentively manuals by the theme (see, for example, [9, 26]) one could find correspondent allusions. It is reasonable understanding of "complex systems" instead of the its understanding as the "objects, consisting of numerous parts, functioning as an organic whole".

So, math modelling is designing in math "parts plus ties", which reproduce "part plus ties" in reality.

So it is principal question in math modelling which math objects represents "part" of the object and which the "ties" ones. The math object — representative should be chosen in such a way that variety of math structuring means were sufficient to convey the object structure.

It is commonly used approach for designing objects — representative to construct them as an finite ordered collection of characteristics: quantitative (numerical) or qualitative (non numerical). Such ordered collection of characteristics is determined by term cortege in math. Cortege is called vector when its components are numerical. In the function recovering problem objects — representatives are vectors and functions are used as a rule to design correspond mathematical "ties". In clustering and classification problem the collection may be both qualitative and quantitative. In last case correspond collection is called feature vector. It is reasonable to note that term "vector" means more, than simply ordered numerical collection. It means that curtain standard math "ties" are applicable to them. These "ties" are adjectives of the math structure called Euclidean space denoted be $R^n$. Namely these are: linear operations (addition and scalar multiplying), scalar product and correspond norm and distance.

It is noteworthy to say, that this variant of Euclidean space $R^n$ is not unique: the space $R^{m\times n}$ of all matrices of a fixed dimension $m \times n$ represents alternative example. The choice of the $R^n$ space as "environmental" math structure is determined by perfect technique developed for manipulation with vectors. These include classical matrix methods and classical linear algebra methods. SVD-technique and methods of Generalized or Pseudo Inverse according Moore-Penrose are comparatively new elements of linear matrix algebra technique [24] (see, also, [1, 2]). Outstanding impacts and achievements in this area are due to N.F Kirichenko (especially, [13, 18], see also [19]). Greville's formulas:forward and inverse -for pseudo inverse matrices, formulas of analytical representation for disturbances of pseudo inverse, - are among them. Additional results in the theme as to further development of the technique and correspondent applications one can find in [7, 19, 20, 21, 15, 6, 14, 22, 17].

As to technique designing for the Euclidean space $R^{m \times n}$ as "environmental" one see, for example [5]. Speech recognition with the spectrograms as the representative and the images in the problem of image recognition are the natural application area for the correspond technique.

As to the choice of the collection (design of cortege or vector) it is necessary to note, that good "feature" selection (components for feature vector or cortege or an arguments for correspond functions) determines largely the efficiency of the problem solution.

As noted above, the efficiency of problem solving group, the choice of representatives of right: space arguments or values of functions and suitable characteristics for features vectors. This phase in solving the grouping information problem must be a special step of the correspondent algorithm. Experience showed the effectiveness of recurrent procedures is largely determined just by successful selection of features vector. For correspond examples see,[12] with Ivachnenko's GMDH (Group Method Data Handling), [25] with Vapnik's Support Vector Machine. Further development of the recurrent technique one may find in [7, 20, 21, 15, 6, 14, 22]. The idea of nonlinear recursive regressive transformations (generalized neuron nets or neurofunctional transformations) due to Professor N. F. Kirichenko is represented in the works referred earlier in its development. Correspondent technique has been designed in this works separately for each of two its basic form f the grouping information problem. The united form of the grouping problem solution is represented here in further consideration. The fundamental basis of the recursive neurofunctional technique include the development of pseudo inverse theory in the publications mentioned earlier first of all due to Professor N.F. Kirichenko and his disciples.

The essence of the idea mentioned above is in the choice of the primary collection and changing it if necessary by standard recursive procedure. Each step of the procedure include detecting of insignificant components, excluding or purposeful its changing, control of efficiency of changes has been made. Correspondingly, the means for implementing the correspondent operations of the step must be designed. Methods of neurofunctional transformation (NfT) (generalized neural nets, nonlinear recursive regressive transformation: [7, 20, 21]).

## 1. Development of Pseudo Inverse Technique for matrices Euclidean spaces

The following are results that transfer basic features of describing the basic structures of Euclidean spaces [5] matrix Euclidean spaces. These are, first of all General Single Valued Decomposition (SVD) theorem and then determination of Pseudo Inverse (PdI) and designing the constructive methods for manipulating with basic structures within matrixes spaces on the base of the Pseudo Inverse. Such transfer make it necessary to introduce special objects and tools for handling them. Namely, these are matrix corteges and corteges operations.

First theorem below is the advanced form of SVD theorem for Euclidean spaces, which one can find in [5].

## 2. Matrices spaces and cortege operators

**Theorem 1.** *For an arbitrary linear operator between a pair of Euclidean spaces* $(E_i, (,)_i)$, $i = 1, 2$: $\wp_E : E_1 \to E_2$, *the collection of singularities* $(v_i, \lambda_l^2)$, $(u_i, \lambda_l^2)$, $i = \overline{1, r}$, $r = rank\wp_E$ *exists for the operators* $\wp_E^* \wp : E_1 \to E_1$, $\wp\wp_E^* : E_2 \to E_2$ *correspondingly, with a common for both operators* $\wp_E^*\wp, \wp\wp_E^*$ *set of Eigen values* $\lambda_l^2, i = \overline{1, r}$ : $\lambda_{i-1} \geq \lambda_i > 0$, $\overline{i = 2, r}$ *such that*

$$\wp_E x = \sum_{i=1}^{r} \lambda_i u_i (v_i, x)_1, \qquad \wp_E^* y = \sum_{i=1}^{r} \lambda_i v_i (u_i, y)_2.$$

*Besides, the following relations take place:*

$$u_i = \lambda_i^{-1} \wp v_i, \quad i = \overline{1, r},$$
$$v_i = \lambda_l^{-1} \wp_E^* u_i, \quad i = \overline{1, r}.$$

## 3. SVD — technique for matrices spaces

We denote by $R^{(m \times n), K}$ – Euclidean space of all matrices $K$-corteges from $m \times n$ matrices: $\alpha = (A_1 \vdots ... \vdots A_K) \in R^{(m \times n), K}$ with a "natural" component wise trace inner product:

$$(\alpha, \beta)_{cort} = \sum_{k=1}^{K} (A_k, B_k)_{tr} = \sum_{k=1}^{K} tr A_k^T B_k,$$
$$\alpha = (A_1 \vdots ... \vdots A_K), \beta = (B_1 \vdots ... \vdots B_K) \in R^{(m \times n), K}.$$

We also denote by $\wp_\alpha : R^K \to R^{m \times n}$a linear operator between the Euclidean space, determined by the relation:

$$\wp_\alpha y = \sum_{k=1}^{K} y_k A_k, \alpha = (A_1 \vdots \dots \vdots A_K) \in R^{(m \times n), K}, \tag{1}$$

$$y = \begin{pmatrix} y_1 \\ \cdots \\ y_K \end{pmatrix} \in R^K.$$

**Theorem 2.** *Range* $\Re(\wp_\alpha) = L_{\wp_\alpha}$*, which is linear subspace of* $R^{m\times n}$*, is the subspace spanned on the components of cortege* $\alpha = (A_1 \vdots \ldots \vdots A_K) \in R^{(m\times n),K}$*, that determines* $\wp_\alpha$*:*

$$\Re(\wp_\alpha) = L_{\wp_\alpha} = L(A_1, \ldots, A_K).$$

**Theorem 3.** *Conjugate for the operator, determined by* (1) *is a linear operator, which, obviously, acts in the opposite direction:* $\wp_\alpha^* : R^{m\times n} \to R^K$*, and defined as:*

$$\wp_\alpha^* X = \begin{pmatrix} trA_1^T X \\ \cdots \\ trA_K^T X \end{pmatrix} = \begin{pmatrix} trX^T A_1 \\ \cdots \\ trX^T A_K \end{pmatrix}.$$

**Theorem 4.** *A product of two operators* $\wp_\alpha^* \wp_\alpha : R^K \to R^K$ *is a linear operator, defined by the matrix from the next equation:*

$$\wp_\alpha^* \wp = \begin{pmatrix} trA_1^T A_1, ..., trA_1^T A_K \\ \cdots \\ trA_K^T A_1, ..., trA_K^T A_K \end{pmatrix}. \tag{2}$$

*Remark.* Matrix defined by (2) is the 'Gram' matrix for the elements of the cortege $\alpha = (A_1 \vdots \ldots \vdots A_K) \in R^{(m\times n),K}$, which determines the operator.

Singular value decomposition for a matrix (2) is obvious, as it is the classical matrix: symmetric and positive semi-definite, on vector Euclidean $R^K$. It is defined by a collection of singularities

$$||v_i|| = 1, v_i \perp v_j, i \neq j;\ i,j = \overline{1,r};\ \lambda_1 > \lambda_2 > ... > \lambda_r > 0,$$

$$\wp_\alpha^* \wp_\alpha v_i = \lambda_i^2 v_i,\ i = \overline{1,r}.$$

The operator $\wp_\alpha^* \wp_\alpha$ by itself and is determined by the relation

$$\wp_\alpha^* \wp_\alpha = \sum_{i=1}^{r} \lambda_i^2 v_i v_i^T = \sum_{i=1}^{r} \lambda_i^2 v_i (v_i,\ \cdot\ ).$$

Each of the row – vectors $v_i^T, i = \overline{1,r}$ will be written by their components:

$$v_i^T = (v_{i1}, ..., v_{iK}),\ i = \overline{1,r},$$

i.e. $v_{ik},\ i = \overline{1,r},\ k = \overline{1,K}$ is the component with the number $k$ of a vector $v$ with a number $I$.

**Theorem 5.** *Matrices* $U_i \in R^{m\times n} : U_i = \frac{1}{\lambda_i}\wp_\alpha v_i = \frac{1}{\lambda_i}\sum_{k=1}^{K} A_k v_{ik},\ i = \overline{1,r}$*, defined by the singularities* $(v_i, \lambda_i^2),\ i = \overline{1,r}$ *of the operator* $\wp_\alpha^* \wp_\alpha$ *are elements of a complete collection of singularities* $(U_i, \lambda_i^2),\ i = \overline{1,r}$ *of the operator* $\wp_\alpha^* : R^K \to R^{m\times n}$*.*

**Theorem 6.** *(Singular Value Decomposition (SVD) for cortege operator). Singularity of two operators* $\wp_\alpha^*\wp_\alpha, \wp_\alpha\wp_\alpha^*$*, obviously determine the SVD for* $\wp_\alpha,\ \wp_\alpha^*$*:*

$$\wp_\alpha y = \sum_{i=1}^{r} \lambda_i U_i v_i^T y, y \in R^K,$$

$$\wp_\alpha^* X = \sum_{i=1}^{r} \lambda_i v_i (U_i, X)_{tr}, X \in R^{m\times n}.$$

**Corollary 1.** *A variant is a SVD for the operator* $\wp_\alpha$ *is represented by the next relation:*

$$\wp_\alpha = \sum_{k=1}^{r} \lambda_k U_k v_k^T = \sum_{k=1}^{r} (\wp_\alpha v_k)\, v_k^T.$$

## 4. Pseudo Inverse Technique for matrices Euclidean spaces

Basic operators of Pseudo Inverse (PdI-operators) theory for a cortege operators are namely pseudo inverse by itself for linear operator, orthogonal projectors on fundamental subspaces of linear operators and grouping operators which also often called by "weighted projection" operators.

**Theorem 7.** *The pseudo inverse operators for* $\wp_\alpha,\ \wp_\alpha^*$ *are determined, correspondingly, by the relations*

$$\wp_\alpha^+ X = \sum_{k=1}^{r} \lambda^{-1} v_k\, (U_k, X)_{tr} = \sum_{k=1}^{r} \lambda^{-2} v_k\, (\wp_\alpha v_k, X)_{tr}\,,\ \forall X \in R^{m\times n},$$

$$(\wp_\alpha^*)^+\, y = \sum_{i=1}^{r} \lambda^{-1} U_i v_i^T y,\ \forall y \in R^K.$$

The basic orthogonal projectors PdI-theory are two pairs of orthogonal projectors. The first one is the pair of orthogonal projectors on the pair fundamental subspaces of $\wp_\alpha, \wp_\alpha^* : \Re(\wp_\alpha) = L_{\wp_\alpha}, \Re(\wp_\alpha^*) = L_{\wp_\alpha^*}$ – their ranges. These orthogonal projections will be designated in one of two equivalent ways:

$$P(\wp_\alpha^*) \equiv P_{L_{\wp_\alpha}} = P_{(A_1,\ldots,A_K)},\ L_{L_{\wp_\alpha}} \subseteq R^{m\times n},\ P(\wp_\alpha) \equiv P_{L_{\wp_\alpha^*}},\ L_{\wp_\alpha^*} \subseteq R^K.$$

The second pair is a pair of orthogonal projectors onto the orthogonal complement $L_{\wp_\alpha}^\perp \subseteq R^{m\times n},\ L_{\wp_\alpha^*}^\perp \subseteq R^K$ of the first pair of the subspaces. The complements, namely, are the Kernels of the correspondent operators. Each of these projectors will be denoted in one of two equivalent ways:

$$Z(\wp_\alpha) \equiv P_{L_{\wp_\alpha^*}^\perp},\ Z(\wp_\alpha^*) \equiv P_{L_{\wp_\alpha}^\perp},$$

obviously:

$$Z(\wp_\alpha) \equiv E_K - P(\wp_\alpha),\ Z(\wp_\alpha^*) \equiv E_{m\times n} - P(\wp_\alpha^*). \tag{3}$$

In accordance with the general properties of PdI, the next properties are valid:

$$P(\wp_\alpha) = \wp_\alpha^+ \cdot \wp_\alpha,\ P(\wp_\alpha^*) = (\wp_\alpha^*)^+ \cdot \wp_\alpha^* = \wp_\alpha \cdot \wp_\alpha^+.$$

Correspondingly:

$$Z(\wp_\alpha) \equiv E_K - \wp_\alpha^+ \cdot \wp_\alpha, \quad Z(\wp_\alpha^*) \equiv E_{m\times n} - \wp_\alpha \cdot \wp_\alpha^+.$$

Grouping operators, denoted below as $R(\wp_\alpha)$, $R(\wp_\alpha^*)$, are also "paired" operators, and are determined by the relations:

$$R(\wp_\alpha) = \wp_\alpha^+ \left(\wp_\alpha^+\right)^* = \wp_\alpha^+ \left(\wp_\alpha^*\right)^+, \quad R(\wp_\alpha^*) = (\wp_\alpha^*)^+ \left((\wp_\alpha^*)^+\right)^* = \left(\wp_\alpha^+\right)^* \wp_\alpha^+.$$

**Theorem 8.** *Grouping operators for the cortege operators $\wp_\alpha$, $\wp_\alpha^*$ can be represented by the next expression:*

$$R(\wp_\alpha^*)X = \sum_{k=1}^{r} \lambda_k^{-2} U_k (U_k, X)_{tr} = \sum_{k=1}^{r} \lambda_k^{-2} U_k tr U_k^T X = \sum_{k=1}^{r} \lambda_k^{-2} U_k tr X^T U_k,$$

*and the quadratic form $(X, R(\wp_\alpha^*)X)_{tr}$ is determined by the relation:*

$$(X, R(\wp_\alpha^*)X)_{tr} = \sum_{k=1}^{r} \lambda_k^{-2} (U_k, X)_{tr}^2,$$

*where*

$$\wp_\alpha^+ X = \sum_{k=1}^{r} \lambda^{-1} v_k \left(U_k, X\right)_{tr} = \sum_{k=1}^{r} \lambda^{-2} v_k \left(\wp_\alpha v_k, X\right)_{tr},$$

$$\left(\wp_\alpha^*\right)^+ y = \sum_{i=1}^{r} \lambda^{-1} U_i v_i^T y.$$

**Theorem 9.** *Quadratic form $(X, R(\wp_\alpha^*)X)_{tr}$ may be written as:*

$$(X, R(\wp_\alpha^*)X)_{tr} = \sum_{i=1}^{r} \lambda_i^{-4} v_i^T \begin{pmatrix} trA_1^T X trA_1^T X & trA_2^T X trA_2^T X & \cdots & trA_1^T X trA_K^T X \\ trA_2^T X trA_1^T X & trA_2^T X trA_2^T X & \cdots & trA_2^T X trA_K^T X \\ \cdots & \cdots & \cdots & \cdots \\ trA_K^T X trA_1^T X & trA_K^T X trA_1^T X & \cdots & trA_K^T X trA_1^T X \end{pmatrix},$$

$$v_i = \sum_{i=1}^{r} \lambda_i^{-4} \left\{ v_i^T \begin{pmatrix} trA_1^T X \\ \cdots \\ trA_K^T X \end{pmatrix} \right\}^2 = \sum_{i=1}^{r} \lambda_i^{-4} \left\{ v_i^T \wp_\alpha^* X \right\}^2.$$

Importance of grouping operators is determined by their properties, represented by the next two theorems.

**Theorem 10.** *For any $A_i,\ i=\overline{1,K}$ of $\alpha=(A_1\vdots\ldots\vdots A_K)\in R^{(m\times n),K}$ the next inequalities are fulfilled:*

$$(A_i, R(\wp_\alpha^*)A_i)_{tr} \le r,\ i=\overline{1,K},\ r=rank\wp_\alpha.$$

**Theorem 11.** *For any $A_i,\ i=\overline{1,K}$ of $\alpha=(A_1\vdots\ldots\vdots A_K)\in R^{(m\times n),K}$ the next inequalities are fulfilled:*

$$(A_i, R(\wp_\alpha^*)A_i)_{tr} \le r_{\min} \le r,\ i=\overline{1,K},\ r=rank\wp_\alpha,$$

$$r_{\min}=\min_{i=\overline{1,n}}(A_i, R(\wp_\alpha^*)A_i)_{tr} \le r_{\min} \le r,\ i=\overline{1,K},\ r=rank\wp_\alpha.$$

*Note.* Statement of theorem 11 is equivalent to that one ellipsoid

$$\frac{1}{r_{\min}}(X, R(\wp_\alpha^*))_{tr} \le 1 \tag{4}$$

is minimal to cover all matrices $A_i,\ i=\overline{1,K}$ of cortege $\alpha=(A_1\vdots\ldots\vdots A_K)\in R^{(m\times n),K}$.

**Definition 1.** Ellipsoid, defined by (4) we will call the minimum grouping ellipsoid for matrices collection $A_i,\ i=\overline{1,K}$.

## 5. Grouping operators and correspondence distances clasterization problems with feature matrix

The results, represented earlier one can apply to solve the grouping information problem in applied math with matrices 'representatives': matrices "feature vectors" or simply — "feature matrices". Indeed, in many important applied researches the objects under investigations are naturally represented by matrices. Spectrograms in speech recognition or digital images in image processing are appropriate examples of such situation. Important means for solving the clasterization problem is constructing and using of appropriate correspondence distance $\rho(X, Kl)$ from a cluster Kl, represented by learning sample of matrices $A_i,\ i=\overline{1,K}$. Such distance one can construct using characteristics of the minimal grouping ellipsoid from theorem 10, 11, built for cortege operator $\wp_\alpha$, generated by the $A_i,\ i=\overline{1,K}$ with $\alpha=(A_1\vdots\ldots\vdots A_n)$:

$$\rho^2(X, Kl)=\frac{1}{r_{\min}}(X, R(\wp_\alpha^*)X)_{tr}, r_{\min}=\min_{i=\overline{1,n}}(A_i, R(\wp_\alpha^*)A_i)_{tr}\le r.$$

## Conclusion

Development of the technique for manipulating with the basic structures of Euclidean spaces within matrices spaces is represented. This technique include General SVD theorem and Moore-Penrose pseudo inverse technique for matrices spaces. Designing the technique demanded introduction matrices corteges and of special cortege operators associated with them.

## References


1. Albert, A. 1972. *Regression and the Moore-Penrose pseudoinverse.* New York: Academic Press.
2. Ben-Israel, A. and Greville, T. N. E. 2003. *Generalized inverses.* New York: Springer.
3. Berry, M. W. 2004. *Survey of text mining.* New York: Springer.
4. Bublik, B. and Kirichenko, N. 1975. *Osnovyi teorii upravleniya.* Kyiv: Vischa shkola.
5. Donchenko, V. 2011. Evklidovyi prostranstva chislovyih vektorov i matrits: konstrutkivnyie metodyi opisaniya bazovyih struktur i ih ispolzovanie. *Information technologies & Knowledge*, 5 (3), pp. 203–216.
6. Donchenko, V., Kirichenko, M. and Krivonos, Y. 2007. Generalizing of neural nets: functional nets of special type. Institute of Information Theories and Applications FOI ITHEA.
7. Donchenko, V., Kirichenko, M. and Serbaev, D. 2004. "Recursive regression transformation and dynamical systems", paper presented at Computer Data analysis and Modeling: robustness and computer intensive methods, Minsk, September 6-10. Minsk: pp. 147–151.
8. Donchenko, V., Zinko, T. and Skotarenko, F. 2012. "Feature vectors in grouping information problem in applied mathematics: vectors and matrices", paper presented at Problems of Computer Intellectualization, Kyiv, Ukraine-Sofia Bulgaria: NASU, V. M. Glushkov Institute of Cybernetics, ITHEA, Kyiv: pp. 111–124.
9. Foster, J. and Helzl, W. 2004. *Applied evolutionary economics and complex systems.* Cheltenham: Elgar.
10. Friedman, M. and Kandel, A. 1999. *Introduction to pattern recognition.* London: Imperial College Press.
11. Haykin, S. S. 1999. *Neural networks.* Upper Saddle River, N.J.: Prentice Hall.
12. Ivahnenko, O. 1969. Samoobuchayuschiesya sistemyi raspoznavaniya i avtomaticheskogo upravleniya. Kyiv: Tehnika.
13. Kirichenko, M. 1997. Analytical Representation of Perturbation of Pseudoinverse Matrices. *Cybernetics and Systems Analysis*, 33 (2), pp. 98–107.
14. Kirichenko, M. and Donchenko, V. 2007. Pseudoinversion in clustering problems. *Cybernetics and Systems Analysis*, 4, pp. 73–92.
15. Kirichenko, M. and Donchenko, V. 2005. Zadacha terminal"noho sposterezhennya dynamichnoyi systemy: mnozhynnist rozv'yazkiv ta optymizaciya. *Journal of Computational & Applied Mathematics*, 5, pp. 63–78.
16. Kirichenko, M., Donchenko, V. and Serbaev, D. 2005. Nonlinear recursive nonlinear Transfomations: Dynamic systems and Optimizations. *Cybernetics and System Analysis*, 41 (3), pp. 364–373.

17. Kirichenko, M., Donchenko, V., Krivonos, Y., Krak, Y. and Kulyas, A. 2009. *Analiz ta syntez sytuacij v systemax pryjnyattya rishen.* Kyiv: Naukova dumka.
18. Kirichenko, N. 1997. Analiticheskoe predstavlenie vozmuscheniy psevdoobratnyih matrits. *Cybernetics and Systems Analysis*, 2.
19. Kirichenko, N. and Lepeha, N. 2002. Primenenie psevdoobratnyih i proektsionnyih matrits k issledovaniyu zadach upravleniya, nablyudeniya i identifikatsii. *Cybernetics and Systems Analysis*, 4, pp. 107–124.
20. Kirichenko, N. F., Krak, Y. V. and Polischuk, A. 2004. Psevdoobratnyie i proektsionnyie matritsyi v zadachah sinteza funktsionalnyih preobrazovateley. *Cybernetics and Systems Analysis*, 3, pp. 116–129.
21. Kirichenko, N., Donchenko, V. and Serbaev, D. 2005. Nelineynyie rekursivnyie regressionnyie preobrazovateli: dinamicheskie sistemyi i optimizatsiya. *Cybernetics and Systems Analysis*, 3, pp. 58–68.
22. Kirichenko, N., Krivonos, Y. and Lepeha, N. 2007. Sintez sistem neyrofunktsionalnyih preobrazovateley v reshenii zadach klassifikatsii. *Cybernetics and Systems Analysis*, 3, pp. 47–57.
23. Kohonen, T. 2001. Self-organizing maps. Berlin: Springer.
24. Nashed, M. Z. and Votruba, G. F. 1976. *A unified operator theory of generalized inverses.*
25. Vapnik, V. N. 1998. *Statistical learning theory.* New York: Wiley.
26. Wakefield, T. 2004. *Systems Analysis and Design.* Pearson Education UK.

# SOLVING OF THE PROBLEM OF DISCRETE FUZZY NUMBER CARRIER'S GROWING

© **Oleg Iemets, Oleksandra Yemets'**

Poltava University of Economics and Trade

Department of Mathematical Simulation and Social Informatics

e-mail: *yemetsli@mail.ru, yemets2008@ukr.net*

***Abstract.*** *The algorithm of reduction of the number of carrier elements of a discrete fuzzy number with the realization of an opportunity to save the information about values is proposed in the article. It is proposed that the information is given by the fuzzy number.*

## Introduction

During performing of a number of operations with fuzzy numbers [1–14], growing of the carrier of a discrete fuzzy number occurs. However several gradations are enough for describing qualitative phenomena. Therefore let set the problem of the reduction of the number of carrier elements with the realization of an opportunity to save the information about values. It is proposed that the information is given by the fuzzy number. One of methods is presented in the report.

## Formulation of a problem

Let we have a fuzzy number $A = \{(a_1|\mu_1), ..., (a_n|\mu_n)\}$ for describing a variable. It is necessary to describe this variable by a fuzzy number with a carrier which has $k < n$ carrier elements.

Let us consider an example.

Let $A = \{(1|0,5), (2|0,8), (3|0,9), (4|0,6), (5|0,4)\}$. Here $n = 5$.

Let $k = 3$. Get a corresponding fuzzy number.

Split the interval $[1,\ 3]$ into three intervals with the length of $\frac{4}{3}$:

$$\left[1,\ 2\frac{1}{3}\right);\qquad \left[2\frac{1}{3},\ 3\frac{2}{3}\right);\qquad \left[3\frac{2}{3};\ 5\right].$$

Find the middles of intervals:

$$b_1 = \frac{1 + 2\frac{1}{3}}{2} = 1\frac{2}{3};\qquad b_2 = \frac{2\frac{1}{3} + 3\frac{2}{3}}{2} = 3;\qquad b_3 = \frac{3\frac{2}{3} + 5}{2} = 4\frac{1}{3}.$$

These numbers make up the carrier of the fuzzy number.

Define membership functions in such way:

$$\mu_{b_1} = \frac{1 \cdot 0,5 + 2 \cdot 0,8}{1 + 2} = 0,7,\qquad \mu_{b_2} = \frac{3 \cdot 0,9}{3} = 0,9,$$

$$\mu_{b_3} = \frac{4 \cdot 0,6 + 5 \cdot 0,4}{9} = 0,4(8) \approx 0,49.$$

So, the result (see Fig. 1) is

$$B = \{(1\frac{2}{3}|0,7), (3|0,9), (4\frac{1}{3}|0,4(8))\}.$$

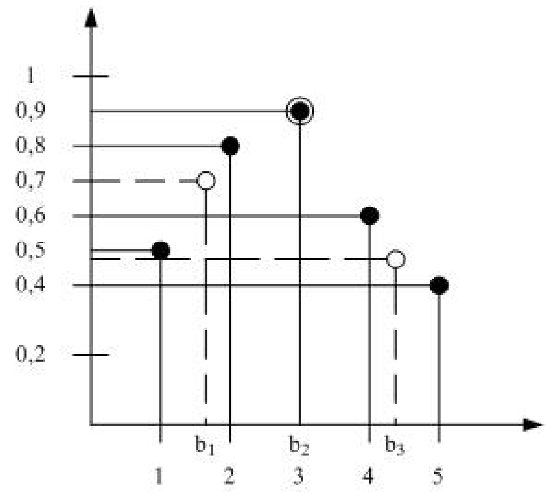



*Fig.* 1. Fuzzy number after the reduction

Describe this procedure by reduction $A$ to $B$ in a general view.

Let $A = \{(a_1|\mu_1), ..., (a_n|\mu_n)\}$, without restricting the generality, consider $a_1 < a_2 < ... < a_n$. It is necessary to get the reduction of the number $A$ down to the number $B$ (corresponding $A$) with $k$ elements.

Denote $B = \{(b_1|\eta_1), ..., (b_k|\eta_k)\}$.

Split the interval $[a_1, a_n]$ into $k$ intervals

$$\Delta_i = [a_1 + h(i-1),\ a_1 + hi), \qquad i = 1, 2, ..., k,$$

where

$$h = \frac{a_n - a_1}{k}.$$

For $i = k$ the right end of interval is included:

$$[a_1 + h(k-1),\ a_1 + hk].$$

Find their middles:

$$\mathrm{b_i} = a_1 + h(i - \frac{1}{2}) = a_1 + \frac{a_n - a_1}{k}(i - \frac{1}{2}), \qquad i = 1, 2, ..., k.$$

Find the value $\eta_i$ of membership function for each $b_i$:

$$\eta_i = \sum_{j:\, a_j \in \Delta_i} a_j \mu_j \; / \sum_{j:\, a_j \in \Delta_i} a_j, \qquad i = 1, 2, ..., k.$$

Thereby $B$ is defined.

## Conclusion

It is proposed to investigate the properties of the operation of reduction in further researches.

## References

1. Donets, G.A. and Yemets, A.O. 2009. The Statement and Solution of the Knapsack Problem with Fuzzy Data. *Journal of Automation and Information Science*, 41 (9), pp. 1–13.
2. Donets, G.A. and Yemets‘, A.O. 2011. The One Algorithm for the Combinatorial Problem of the Fuzzy Rectangles Packing. *Prykladna statystyka. Aktuarna ta feenansova matematyka*, 1–2, pp. 158–167.
3. Iemets, O. and Yemets‘, O. 2008. Operations and Relations on Fuzzy Numbers. *Naukovi Visti of the National Technical University of Ukraine “Kyiv Polytechnic Institute”*, 5, pp. 39–46.
4. Iemets, O. and Yemets‘, O. 2008. The Construction of the Mathematical Model of the One Combinatorial Problem of Rectangles Packing with Fuzzy Sizes. *Naukovi Visti of the National Technical University of Ukraine “Kyiv Polytechnic Institute”*, 6, pp. 25–33.
5. Iemets, O. and Yemets‘, O. 2012. Solving a Linear Problem of Euclidean Combinatorial Optimization on Arrangements with a Constant Sum of the Elements. *Kibernetika i sistemny analiz*, 4, pp. 83–94.
6. Iemets, O. and Yemets‘, O. 2011. *Solving of Combinatorial Optimization Problems on Fuzzy Sets: Monograph.* Poltava: PUET.
7. Iemets, O. and Yemets‘, O. 2012. “A Branch and Bound Method for Optimization Problems with Fuzzy Numbers”, paper presented at *Modeling and Simulation*: MS’2012, Minsk, 2–4 May. Minsk: Publ. Center of BSU, pp. 62–65.
8. Iemets, O.O. and Yemets‘, O.O. 2010. “One Problem of Rectangles Packing as Problem of Combinatorial Optimization with Fuzzy Parameters”, paper presented at *17th Zittau East-West Fuzzy Colloquium*, Zittau, 15–17 September. Zittau: Hochschule Zittau/Gurlitz: Univ. of Appl. Sciences, pp. 180–187.
9. Iemets, O., Yemets‘, A. and Parfonova, T. 2012. “Optimization on Fuzzy Sets by the Branch and Bound Method”, paper presented at *III Intern. Sc. Conf. “Mathematical stimulation, optimization and information technologies”*, Kishinev, 19-23 March. Kishinev, pp. 338–347..
10. Roskladka, A. A. and Yemets‘, A. O. 2007. Solving of the One Combinatorial Packing Problem under Conditions of Uncertain Data Which Are Described by Fuzzy Numbers. *Radioelektronika i informatika*, 2, pp. 132–141.

11. Yemets', A. 2011. "A Problem of Investments' Distribution in Fuzzy Formulation", paper presented at *Economics through the eyes of the youth: materials of the IV International economic forum of young scientists*, Vileyka, 3–5 June. Minsk: BGATU, pp. 326–329.

12. Yemets', O. O. 2004. One Combinatorial Optimization Problem on Permutations of Fuzzy Sets. *Volinskiy matematichny visnyk: Series Applied Mathematics*, 2 (11), pp. 101–106.

13. Yemets', O. O. 2007. The One Packing Problem as Combinatorial Optimization on Fuzzy Set of Partitions and Its Solving. *Radioelektronika i informatika*, 4, pp. 150–160.

14. Yemets', O. O. 2011. Two Properties of the Sum Operation and the Linear Order for Fuzzy Numbers with a Discrete Carrier. *Shtuchnyi Intelekt*, 4, pp. 285–290.

# MATHEMATICAL MODEL FOR CANCER PREVALENCE AND CANCER MORTALITY

© J. Kalas, J. Novotný, J. Michalek, O. Nakonechny

Institute of Mathematics. Faculty of Mechanical Engineering Brno UT
e-mail: *fusek.m@gmail.com*
Taras Shevchenko National University of Kyiv. Cybernetic Faculty
e-mail: *a.nakonechniy@gmail.com*

***Abstract***. *The first part of the paper designs a deterministic model to describe cancer prevalence and mortality in a population. Next the asymptotic properties of the model are investigated. In the second part, the model is applied to real-world data. For selected model data, a numerical solution is found to the differential equations describing the model, a long-term prediction is made with its results compared with those of predictions made by regression analysis, which are often used to model the prevalence and mortality in the present literature. It is shown that, although for short-term predictions (up to 10 years) both approaches are nearly equivalent, there is a major difference between them if a longer-term prediction is made and finding a reliable prediction for a period longer than 10 years based on short time series seems to be unlikely.*

## Introduction

Today, cancer is one of the major health risks of our civilisation. The statistics of the cancer prevalence and mortality related to the geographical distribution of such variables is a subject that has been receiving much attention in the present literature [3, 4, 6, 7]. The objective is to design good mathematical models that can be used to describe the changes in the prevalence numbers with respect to their prediction and to the prediction of mortality.

This paper is concerned with the design of a mathematical model based on differential equations for making reliable short-term predictions for a given population with the possibility of a long-term perspective. The model is then tested on real-world data and the resulting predictions are compared with the predictions obtained by regression analysis.

## 1. Model

We will use the following denotations in a population with cancer occurrence:

| | |
|---|---|
| $n_1(t)$ | number of people suffering from cancer (prevalence) at time t, |
| $n_2(t)$ | number of deaths from cancer (mortality) at time t. |

The time interval in which the prevalence $n_1(t)$ and mortality $n_2(t)$ is to be modelled is $\langle 0, T\rangle$ with $T$ being a time horizon and, denoting by $n(t)$ the population size at time $t$, $n(T)$ gives the size of the observed population at the time horizon $T$.

When constructing the model, we assume the prevalence change over a time interval $\Delta t$ to be proportional to the length of this interval next to the prevalence at $t$ and, finally, to the logarithm of $\frac{n(T)}{n_1(t)}$. Thus, as $t$ increases and $t$ is close to the time horizon $T$, the change in the growth rate $\frac{\mathrm{d}n_1(t)}{\mathrm{d}t}$ is slower and, when the time horizon $n(T)$ is reached, it almost vanishes. Similarly, we assume that the change in mortality over a time interval $\Delta t$ is proportional to the length of this interval and to the mortality $n_2(t)$, and, finally, to the logarithm of $\frac{n_1(t)}{n_2(t)}$. Thus, when describing the prevalence behaviour, we see that it does not change in the limit case if the mortality reaches the value of prevalence.

The given considerations lead to the following system of differential equations for prevalence $n_1$ and mortality $n_2$ :

$$\frac{\mathrm{d}n_1(t)}{\mathrm{d}t} = \alpha_1 n_1(t) \ln\left(\frac{n(T)}{n_1(t)}\right), \tag{1}$$

$$\frac{\mathrm{d}n_2(t)}{\mathrm{d}t} = \alpha_2 n_2(t) \ln\left(\frac{n_1(t)}{n_2(t)}\right). \tag{2}$$

These equations should be solved in terms of $n_1$ and $n_2$, subject to initial conditions $n_1(t_0) = n_{10}$ and $n_2(t_0) = n_{20}$. The model has two parameters, $\alpha_1$ and $\alpha_2$, which affect the shape of $n_1$ and $n_2$, respectively. When fitting the model to a particular population data, the initial conditions are given, while the parameters $\alpha_1$ and $\alpha_2$ are to be estimated. The constant $n(T)$ in equation (1), as mentioned above, denotes the size $n$ of the whole population (e.g. of a given country) at time $T$ - the horizon of the intended prognosis. This quantity should be estimated or based on an expert judgment.

## 2. The phase analysis of the model equations

It can be shown that the solutions of (1) have the form

$$n_1(t) = \exp\left\{\ln n(T) - c\exp\left(-\alpha_1 t\right)\right\}. \tag{3}$$

Inserting this into (2) yields an equation in $n_2$ and $t$ only, which is however nontrivial. Therefore we shall accomplish phase analysis of the autonomous two-dimensional system (1), (2) in the first quadrant of the phase space of (1), (2). It can be easily seen that, for the right-hand sides of (1) and (2), it holds that $\alpha_1 n_1 \ln\left(\frac{n(T)}{n_1}\right) > 0$ $(< 0)$ iff $0 < n_1 < n(T)$ $(n_1 > n(T))$ and $\alpha_2 n_2 \ln\left(\frac{n_1}{n_2}\right) > 0$ $(< 0)$ iff $n_1 > n_2 > 0$ $(0 < n_1 < n_2)$. Hence the direction field of (1), (2) looks as in Figure 1. The nulclines of (1), (2) are lines $n_2 = n_1$ and $n_1 = n(T)$. From the direction field we infer that any trajectory of (1), (2) starting in the interior $\mathring{\mathbb{R}}^2_+$ of the first quadrant $\mathbb{R}^2_+$ remains in $\mathring{\mathbb{R}}^2_+$ for $t \to \infty$ and any trajectory is bounded. Taking into account the practical meaning of $n_1$, $n_2$, it is obvious that only trajectories lying in the interior of the shaded triangle $T$ are admissible in our model.

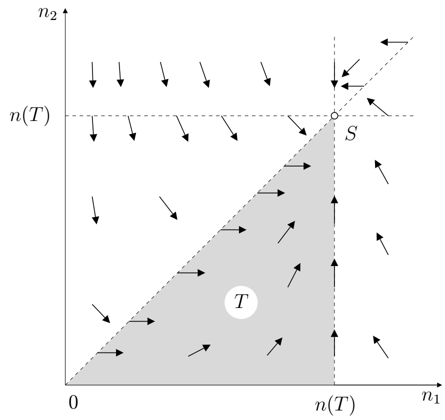



*Fig.* 1. Direction field of the system (1), (2).

**Theorem 1.** *The autonomous system* (1), (2) *has a unique stationary point* $S = (n(T), n(T))$ *in the interior of the first quadrant. The trajectory starting at a point* $(n(T), n_{20})$ *different from the stationary point* $S$ *is a part of a straight line* $n_1 = n(T)$*. Any trajectory starting in the interior* $\mathring{T}$ *of the triangle* $T$ *remains in* $T$ *for increasing* $t$ *and tends to the point* $S$ *as* $t \to \infty$ *(see Figure 2).*

*Proof:* Any stationary point of (1), (2) is an intersection of nulclines of (1), (2). Clearly, there is the unique intersection of the nulclines $n_2 = n_1$, $n_1 = n(T)$ in $\mathring{\mathbb{R}}^2_+$ at the point $S = (n(T), n(T))$. The solution with the initial point $(n(T), n_{20})$, where $n_{20} \in (0, n(T)) \cup (n(T), \infty)$, is of the form

$$(n_1(t), n_2(t)) = \left( n(T), n(T) \exp \left\{ \ln \frac{n_{20}}{n(T)} \exp[\alpha_2(t_0 - t)] \right\} \right).$$

The corresponding trajectory is a part of a straight line $n_1 = n(T)$. The Jacobi matrix of the mapping

$$(n_1, n_2) \mapsto \left( \alpha_1 n_1 \ln \left( \frac{n(T)}{n_1} \right), \alpha_2 n_2 \ln \left( \frac{n_1}{n_2} \right) \right)$$

is

$$J(n_1, n_2) = \begin{bmatrix} \alpha_1 \left( -1 + \ln \frac{n(T)}{n_1} \right) & 0 \\ \alpha_2 \frac{n_2}{n_1} & \alpha_2 \left( \ln \frac{n_1}{n_2} - 1 \right) \end{bmatrix}.$$

Thus

$$J(n(T), n(T)) = \begin{bmatrix} -\alpha_1 & 0 \\ \alpha_2 & -\alpha_2 \end{bmatrix}.$$

Since the eigenvalues of the matrix $J(n(T), n(T))$ are $\lambda_1 = -\alpha_1 < 0$, $\lambda_2 = -\alpha_2 < 0$, the stationary point $S = (n(T), n(T))$ is a stable node. With respect to the direction field of (1), (2), we observe that any trajectory starting in the interior $\mathring{T}$ of the triangle $T$ remains in $\mathring{T}$ for $t \to \infty$. In view of the Poincaré-Bendixson theory (see e. g. Hartman [2], Chapter VII), the $\omega$-limit set $\Omega(C^+)$ of any trajectory $C^+$ starting in $\mathring{T}$ is the set $\Omega(C^+) = \{(n(T), n(T))\}$. This implies $(n_1(t), n_2(t)) \to (n(T), n(T))$ as $t \to \infty$ for any solution $(n_1(t), n_2(t))$ of (1),(2) corresponding to the considered trajectory. □

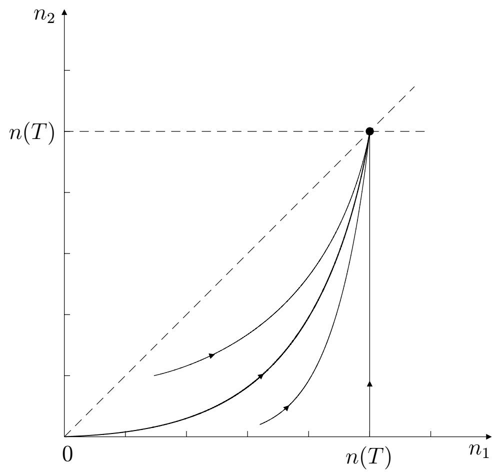



*Fig.* 2. Phase trajectories corresponding to the solution $(n_1(t), n_2(t))$.

**Theorem 2.** *If $\alpha_2 > \alpha_1$, then infinitely many trajectories of (1), (2) starting in $\mathring{T}$ approach the stationary point $S = (n(T), n(T))$ as $t \to \infty$ with the characteristic direction $(\alpha_2 - \alpha_1, \alpha_2)$ and there is at least one trajectory of (1), (2) starting at $T$ such that it approaches the point $S$ with the characteristic direction $(0, 1)$. Moreover,*

$$n_1(t) = n(T) + e^{-\alpha_1 t}\left[(\alpha_2 - \alpha_1)\varkappa + o(1)\right] \quad \text{as } t \to \infty,$$
$$n_2(t) = n(T) + e^{-\alpha_1 t}[\alpha_2 \varkappa + o(1)] \quad \text{as } t \to \infty$$

*for infinitely many solutions $(n_1(t), n_2(t))$ of (1), (2) starting in $\mathring{T}$, where $\varkappa$ is a nonzero real constant dependent on the solution $(n_1(t), n_2(t))$.*

*Proof:* Denote $' = \frac{\mathrm{d}}{\mathrm{d}t}$. The transformation $x_1 = n_1 - n(T)$, $x_2 = n_2 - n(T)$ converts the system (1), (2) into the system (written in a vector form)

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}' = \begin{bmatrix} -\alpha_1 & 0 \\ \alpha_2 & -\alpha_2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \alpha_1(x_1 + n(T)) \ln \frac{n(T)}{x_1+n(T)} + \alpha_1 x_1 \\ \alpha_2(x_2 + n(T)) \ln \frac{x_1+n(T)}{x_2+n(T)} - \alpha_2 x_1 + \alpha_2 x_2 \end{pmatrix}$$

with the singular point $(x_{10}, x_{20}) = (0, 0)$ corresponding to the singular point $S$ of (1), (2). The transformation

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{bmatrix} 0 & \alpha_2 - \alpha_1 \\ 1 & \alpha_2 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \tag{4}$$

yields

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}' = \begin{bmatrix} -\alpha_2 & 0 \\ 0 & -\alpha_1 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} + F(y_1, y_2), \tag{5}$$

where

$$F(y_1, y_2) = \begin{bmatrix} \frac{-\alpha_2}{\alpha_2-\alpha_1} & 1 \\ \frac{1}{\alpha_2-\alpha_1} & 0 \end{bmatrix} F_1((\alpha_2 - \alpha_1) y_2, y_1 + \alpha_2 y_2),$$

$F_1$ being defined by

$$F_1(x_1, x_2) = \begin{pmatrix} \alpha_1(x_1 + n(T)) \ln \frac{n(T)}{x_1+n(T)} + \alpha_1 x_1 \\ \alpha_2(x_2 + n(T)) \ln \frac{x_1+n(T)}{x_2+n(T)} - \alpha_2 x_1 + \alpha_2 x_2 \end{pmatrix}.$$

Notice that the inverse transformation to (4) is given by

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{bmatrix} \frac{-\alpha_2}{\alpha_2-\alpha_1} & 1 \\ \frac{1}{\alpha_2-\alpha_1} & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

It can be easily verified that $||F(y_1, y_2)|| \,/\, ||(y_1, y_2)||^{1+\varepsilon} \to 0$ as $(y_1, y_2) \to (0, 0)$ for some $\varepsilon > 0$, where $|| \cdot ||$ denotes the Euclidean norm in $\mathbb{R}^2$. The transformations used are regular affine, the triangle $T$ is converted to a new triangle $T'$ and $\mathring{T}'$ is an invariant set with respect to the system (5). Combining this with Theorem 3.1 from Chapter VIII of [2], we get that infinitely many solutions $(y_1(t), y_2(t))$ of (5) with $(y_1(t_0), y_2(t_0)) \in \mathring{T}'$ satisfy $(y_1(t), y_2(t)) \to (0, 0)$ and $(y_1(t), y_2(t))/||(y_1(t), y_2(t))|| \to (0, 1)$ as $t \to \infty$. Moreover, Theorem 3.5 from Chapter VIII of [2] provides the equations

$$y_1(t) = e^{-\alpha_1 t} o(1) \quad \text{as } t \to \infty,$$
$$y_2(t) = e^{-\alpha_1 t}(\varkappa + o(1)) \quad \text{as } t \to \infty$$

for these solutions, where $\varkappa$ is a nonzero real constant. Using the transformation (4), the characteristic direction $(0, 1)$ of (5) is converted to the characteristic direction $(\alpha_2 - \alpha_1, \alpha_2)$ and the relations $n_1 = n(T) + (\alpha_2 - \alpha_1)y_2$, $n_2 = n(T) + y_1 + \alpha_2 y_2$ yield the desired results. Note that the trajectory corresponding to the solution $(n_1(t), n_2(t)) = \left(n(T), n(T)\exp\left\{\ln \frac{n_{20}}{n(T)} \exp[\alpha_2(t_0 - t)]\right\}\right)$ tends to the singular point $S$ with the characteristic direction $(0, 1)$ as $t \to \infty$. □

The case $\alpha_2 < \alpha_1$ is analogous and from our data point of view is not important.

## 3. Parameter estimation

This section is concerned with parameter estimation $\alpha_1$ and $\alpha_2$. To estimate the parameters $\alpha_1$ and $\alpha_2$, we propose to minimize the L2 distance between the predictions and the real-world data. Consider real-world data for the years $t_0 \dots t_m$ denoting them by $n_{10}, \dots, n_{1m}$ and $n_{20}, \dots, n_{2m}$. Also denote the solution to (1), (2) by $n_1(\alpha_1, t)$, $n_2(\alpha_1, \alpha_2, t)$ where the dependence on the parameters $\alpha_1$ and $\alpha_2$ is stressed. The optimization problem can then be expressed as

$$\min_{\alpha_1,\alpha_2}[c_1 \sum_{i=0}^{m} (n_{1i} - n_1(\alpha_1, t_i))^2 + c_2 \sum_{i=0}^{m} (n_{2i} - n_2(\alpha_1, \alpha_2, t_i))^2], \tag{6}$$

$$\begin{aligned} \text{s.t.} \quad \alpha_1 &\geq 0, \\ \alpha_2 &\geq 0, \end{aligned}$$

where $c_1$ and $c_2$ are suitable weighting coefficients (in the basic setting $c_1 = 1$, $c_2 = 1$ ).

As mentioned in the previous section, the solutions to (1) have the form (3). Substituting (3) into (2) yields a non-trivial equation in $n_2$ and $t$ only. Thus it is better, using computer, to integrate the equations (1), (2) numerically and use a black-box type solver for the problem (6). In this case, the solver requires that the objective function of (6) is evaluated on a sequence of points $(\alpha_1, \alpha_2)$. For each such point, the equations (1), (2) are solved and subsequently the value of (6) is obtained.

By this approach, satisfactory results on the given data were achieved. We used Octave with the `lsode` ODE solver [5] to integrate the equations (1), (2), and the NOMAD [1] solver for the optimization.

## 4. Data

The model was tested for functionality using the data shown by Table 1. In processing prevalence the numbers of colon cancers were used (the cancer type being C18) in the Czech Republic's male population from 1989 to 2005, see [3]. The table is completed by further demographic data on the numbers of new born and deceased men as well as the total size of the Czech male population during the years in question.

*Table* 1. Men's population — C18 cancer type.

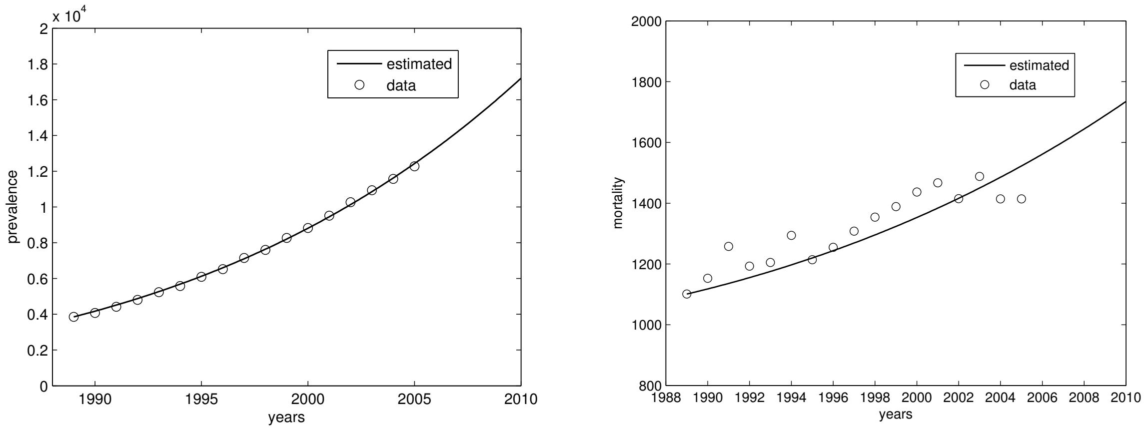
| | diseased | | | total | | |
|---|---|---|---|---|---|---|
| year | prevalence ($n_1$) | incidence | mortality ($n_2$) | births | deaths | population |
| 1989 | 3853 | 1505 | 1101 | n.a. | n.a. | n.a. |
| 1990 | 4075 | 1476 | 1153 | n.a. | n.a. | n.a. |
| 1991 | 4416 | 1730 | 1258 | 129354 | 63342 | 5006002 |
| 1992 | 4807 | 1710 | 1193 | 121705 | 61767 | 5013413 |
| 1993 | 5231 | 1756 | 1205 | 121025 | 59180 | 5019297 |
| 1994 | 5578 | 1835 | 1294 | 106579 | 58609 | 5020464 |
| 1995 | 6091 | 1886 | 1214 | 96097 | 58925 | 5016515 |
| 1996 | 6525 | 1951 | 1255 | 90446 | 56709 | 5012085 |
| 1997 | 7149 | 2234 | 1308 | 90657 | 56692 | 5008730 |
| 1998 | 7602 | 2163 | 1354 | 90535 | 55139 | 5005435 |
| 1999 | 8267 | 2325 | 1389 | 89471 | 54845 | 5001062 |
| 2000 | 8821 | 2323 | 1437 | 90910 | 54882 | 4996731 |
| 2001 | 9511 | 2459 | 1467 | 90715 | 53772 | 4967986 |
| 2002 | 10268 | 2603 | 1415 | 92786 | 54377 | 4966706 |
| 2003 | 10938 | 2559 | 1488 | 93685 | 55880 | 4974740 |
| 2004 | 11569 | 2460 | 1414 | 97664 | 54190 | 4980913 |
| 2005 | 12273 | 2622 | 1414 | 102211 | 54072 | 5002648 |

## 5. Results

Since the total population of the Czech Republic is steady, we estimate the value of $n(T)$ to be approximately 5 000 000. The estimated parameters of the model (1) and (2) are

$$\alpha_1 = 0.0111$$
$$\alpha_2 = 0.0119$$

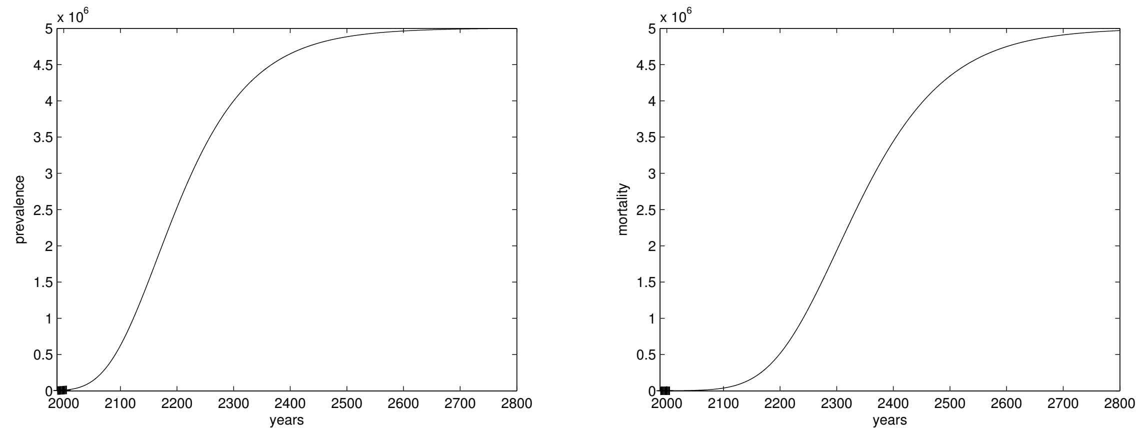
and the fitted time dependencies are shown in Figure 3. It can be seen that the short-time predictions obtained from this model are reasonable, especially for prevalence.




(A) Prevalence. (B) Mortality.

*Fig.* 3. Estimates.

In the event of a long-term prediction, the model achieves an equilibrium close to $n(T)$ — see Figure 4. It is obvious however, that the model does not give a satisfactory description of reality in the long term. It is clear from the pictures that, for a short time horizon (of up to ten years) the predictions obtained seem to be realistic. Predictions for a long time horizon, however, are rather debatable.




(A) Prevalence. (B) Mortality.

*Fig.* 4. Long term estimates.

## 6. Comparison with the regression model

In the medical community, linear regression models prevail nowadays. We present a regression model of both mortality and prevalence, based on the data of Table 1, which is to be compared with the model based on differential equations (DE model) developed in the previous section.

A linear dependence for mortality and a quadratic one for prevalence are the appropriate polynomial choices, as indicated by statistical tests of their coefficients differences from zero.

$$\begin{aligned}\text{Mortality:}\quad m &= \beta_0 + \beta_1(y-1989)\\ \text{Prevalence:}\quad p &= \beta_0 + \beta_1(y-1989) + \beta_2(y-1989)^2\end{aligned}$$

The regression coefficients are summarized in tables (3) and (2), and the fitted dependencies are depicted in Figure 5.

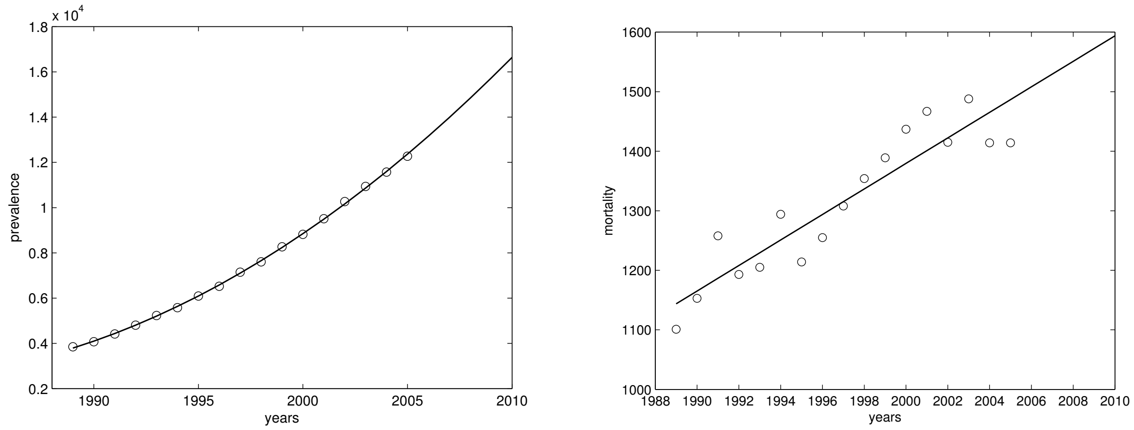
*Table* 2. Regression coefficients — mortality.

| parameter | estimate | conf. interval (95%) | |
|---|---|---|---|
| $\beta_0$ | 1144 | 1096 | 1190 |
| $\beta_1$ | 21.4 | 16.4 | 26.4 |

*Table* 3. Regression coefficients — prevalence.

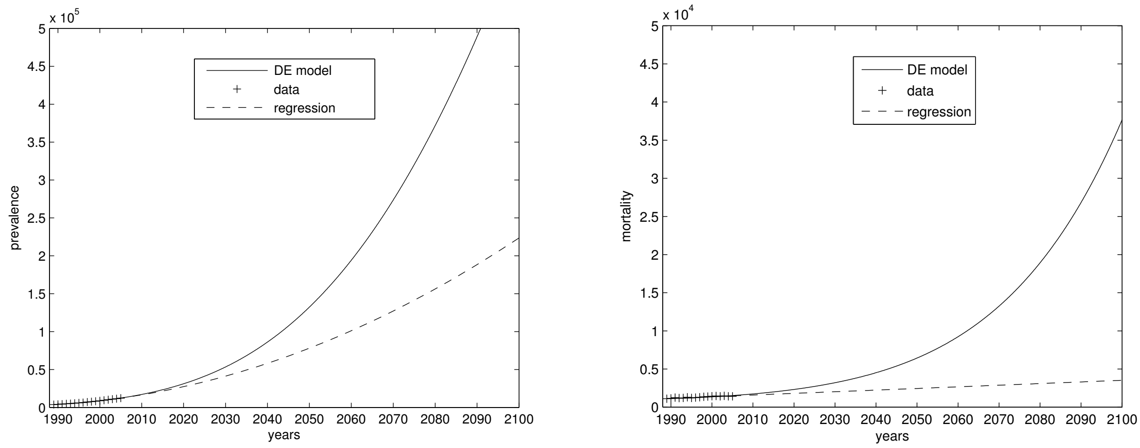
| parameter | estimate | conf. interval (95%) | |
|---|---|---|---|
| $\beta_0$ | 3792 | 3714 | 3870 |
| $\beta_1$ | 292 | 270 | 315 |
| $\beta_2$ | 15.2 | 13.8 | 16.6 |

Figure 6 shows a comparison of the regression and DE models. The models will differ by more than 50 percent by 2040 in the case of mortality, and by 2070 in the case of prevalence. This considerable difference may be accounted for by the regression model dependent variables growing at a polynomial while those of the DE model at an exponential rate. Because of this, the use of either of these models for long-term predictions is considerably limited. However, the graphics give an outline of the behaviour of the observed quantities. Based on the comparison of the models, it may be concluded that the regression predictions, used quite often nowadays, are applicable to short-term predictions (of up to ten years). The values predicted by the regression approach are similar to those obtained from the dynamic DE model. For long-term predictions extending beyond 10 years, however, the methods differ considerably thus making a reliable prediction for this period based on the short data series rather unrealistic.




(A) Prevalence. (B) Mortality.

*Fig.* 5. Regression model.




(A) Prevalence. (B) Mortality.

*Fig.* 6. Model results comparison.

*The paper was prepared with the support of the Czech Science Foundation – Projects MSM No. 0021622418 and CQR No. 1M06047 and of the grant 201/08/0469 of the Czech Science Foundation.*

## References

1. Abramson M. A., Audet C., Couture G., Dennis J. E., Jr., and Le Digabel S.: *The nomad project. Software available at www.gerad.ca/nomad.* August 2009.
2. Hartman, P.: *Ordinary differential equations,* Society for Industrial and applied mathematics (SIAM), Philadelphia, 2002,(Second edition).

3. Konečný M., Kubáček P., Štampach R., Kozel J. Strachoň Z., Dítě P., Kraus R., Koška P., Geryk E., Michálek J. and Odehnal J. *Cancer prevalence in the Czech Republic 1989– 2005–2015* PF MU Brno, 2008, pp. 1-69.
4. Levi F., Lucchini F., Negri E., Boyle P., La Vecchia C.:*Mortality from major cancer sites in European Union, 1955 - 1998.* Ann. Oncol. 2003, 14:490-495.
5. Radhakrishnan K. and Hindmarsh A. C., *Description and use of lsode, the livermore solver for ordinary differential equations, Tech. Rep., 1993.* [Online]. Available: http://gltrs.grc.nasa.gov/cgi-bin/GLTRS/browse.pl?2003/RP-1327.html. October 2009.
6. Stracci F., Canosa A., Minelli L., Petrinelli A. M., Cassetti T., Romagnoli C. and La Rosa F.: *Cancer mortality trends in the Umbria region of Italy 1978 - 2004: a jointpoint regression analysis* BMC Cancer 2007, 7;10 p. 1-9.
7. Wingo P. A., Cardinez C. J., Landis S. H., Greenlee R. T., Ries L. A., Anderson R. N., Thun M. J.: *Long-term trends in cancer mortality in the United States, 1930 - 1998.* Cancer 2003, 97 (Suppl 12]: 3133-3275. Erratum Cancer 2005, 103 Suppl 12:2658)

# AVERAGING IN THE OPTIMAL CONTROL PROBLEM FOR THE REACTION-DIFFUSION EQUATION WITH MULTIVALUED INTERACTION FUNCTION

© Olena Kapustian

Taras Shevchenko National University of Kyiv
Faculty of Cybernetics
e-mail: *olena.kap@gmail.com*

***Abstract****. In this paper consider the optimal control problem on infinite time interval with quadratic cost functional. State of this problem is defined by the evolutionary inclusion of reaction-diffusion type. We prove the solvability of such a problem. In the case of rapidly oscillating coefficients in coefficients of differential operator and multivalued interaction function we prove the convergence of* $\varepsilon$*-dependent optimal process to optimal process of the corresponding averaged problem.*

## Introduction

One of the main problems in the study of processes in micro-inhomogeneous media is the correctness of passing to the averaged problem [1]. Works [2] - [4] are devoted to the research on convergence in optimal control problems for distributed systems with perturbations in coefficients. In this paper we consider the optimal control problem on the solutions of reaction-diffusion type inclusion. Moreover, such an inclusion has perturbations in the differential operator coefficients and multivalued interaction function which has power growth. We investigate the issue of the solution dependence on the parameter for mentioned problem. However as opposed to [3, 4] the averaged problem is not degenerate into linear-quadratic one.

## 1. Problem setting

We consider the optimal control problem

$$\begin{cases} \frac{\partial y}{\partial t} \in div(a^{\varepsilon}(x)\nabla y) - F_{\varepsilon}(x,y) + h^{\varepsilon}(x)u(t), \ x \in \Omega, \ t > 0, \\ y(x,t) = 0, \ x \in \partial\Omega, \\ y(x,0) = y_0^{\varepsilon}, \end{cases} \tag{1}$$

$$u(t) \in U \subseteq L^2(0,+\infty), \tag{2}$$

$$J(y,u) = \int_0^{+\infty} \int_{\Omega} y^2(x,t)dxdt + \gamma \int_0^{+\infty} u^2(t)dt \to \inf, \tag{3}$$

where $\Omega \subset \mathbb{R}^n$ is bounded domain, $\varepsilon > 0$ is a small parameter, matrix $a^\varepsilon(x) = \{a^\varepsilon_{ij}(x)\}$ is measurable, symmetric and satisfies the condition of uniform ellipticity

$$\begin{gathered} \exists \lambda_1 > 0,\ \Lambda_1 > 0\ \forall \varepsilon > 0\ \forall \xi \in \mathbb{R}^n \\ \lambda_1 |\xi|^2 \le \sum_{i,j=1}^{n} a^\varepsilon_{ij}(x)\xi_i\xi_j \le \Lambda_1 |\xi|^2. \end{gathered} \tag{4}$$

Multivalued interaction function $F_\varepsilon(x, y)$ has a form

$$F_\varepsilon(x, y) = [b_\varepsilon(x) f_\varepsilon(y), d_\varepsilon(x) g_\varepsilon(y)].$$

Here $b_\varepsilon,\ d_\varepsilon$ are measurable, bounded functions in $L^\infty(\Omega)$, for which the following condition holds

$$\exists \beta > 0\ \forall x \in \Omega\ \forall\ \varepsilon > 0\ b_\varepsilon(x) \ge \beta,\ d_\varepsilon(x) \ge \beta. \tag{5}$$

Functions $f_\varepsilon,\ g_\varepsilon$ are bounded functions in $C(\mathbb{R})$, which satisfy the next conditions

$$\begin{gathered} \exists C_1 \ge 0,\ \exists \alpha > 0,\ \exists p \ge 0,\ \forall y \in \mathbb{R}\ \forall \varepsilon > 0 \\ |f_\varepsilon(y)| + |g_\varepsilon(y)| \le C_1(1 + |y|^{p-1}), \\ y f_\varepsilon(y) \ge \alpha |y|^p, y g_\varepsilon(y) \ge \alpha |y|^p. \end{gathered} \tag{6}$$

Functions $h^\varepsilon,\ y_0^\varepsilon$ are bounded in $L^2(\Omega)$, set of admissible controls $U$ is closed, convex and $0 \in U$.

**Definition.** For fixed $u \in U$ a function $y \in W = L^2_{loc}(0, +\infty; H^1_0(\Omega)) \bigcap L^p_{loc}(0, +\infty; L^p(\Omega))$ is called the solution of the problem (1) if this function is such that $y(0) = y_0^\varepsilon$, and for some function $l = l(t, x) \in L^q_{loc}(0, +\infty; L^q(\Omega))$, $\frac{1}{p} + \frac{1}{q} = 1$ it holds that $l(t, x) \in F_\varepsilon(x, y(t, x))$ almost everywhere (a. e.) and $\forall v \in H^1_0(\Omega) \bigcap L^p(\Omega)$, $\forall \eta \in C_0^\infty(0, T)$

$$\int_0^T (y, v)\eta_t dt - \int_0^T ((a^\varepsilon \nabla y, \nabla y) + (l, v) - u(t)(h^\varepsilon, v))\, \eta dt = 0. \tag{7}$$

Here and below $\|\cdot\|$ and $(\cdot, \cdot)$ indicate a norm and a scalar product in $L^2(\Omega)$.

By the conditions (5), (6) the global solvability of the problem (1) follows from [5] for $\forall u \in U$, $y_0^\varepsilon \in L^2(\Omega)$, if in the right-hand side we put a continuous selector of the mapping $F_\varepsilon$. However from the results of [6] it implies that the set of the solutions of (1) is not exhausted to the solutions of equations for continuous selectors of $F_\varepsilon$. It greatly increases the set of admissible processes in the problem (1) - (3).

The main aim of this paper is to prove convergence of optimal process of the problem (1) - (3) to optimal process of corresponding averaged problem.

## 2. Existence of solutions of the optimal control problem

From [3] - [6] it follows that any solution of the problem (1) belongs to class $C([0,+\infty); L^2(\Omega))$ and for almost all (a. a.) $t > 0$ next energy equality holds

$$\frac{1}{2}\frac{d}{dt}\|y(t)\|^2 + (a^\varepsilon \nabla y(t), \nabla y(t)) + (l, y(t)) = u(t)(h^\varepsilon, y(t)), \tag{8}$$

where $l(t,x) \in F_\varepsilon(x, y(t,x))$ a. e.

Moreover, by (4) - (6) $\forall t \geq s \geq 0$ we have

$$\frac{1}{2}\frac{d}{dt}\|y(t)\|^2 + \lambda_1 \|y(t)\|^2_{H_0^1} + \alpha\beta \|y(t)\|^p_{L^p} \leq |u(t)| \|h^\varepsilon\| \|y(t)\|. \tag{9}$$

From the Poincare inequality [5] one can obtain

$$\frac{d}{dt}\|y(t)\|^2 + \lambda_1 \|y(t)\|^2_{H_0^1} + 2\alpha\beta \|y(t)\|^p_{L^p} \leq C_2 \|h^\varepsilon\|^2 |u(t)|^2. \tag{10}$$

Applying the Gronwall inequality, we finally have $\forall t \geq s \geq 0$

$$\|y(t)\|^2 \leq \|y(s)\|^2 \exp^{-\lambda_1(t-s)} + C_3 |h^\varepsilon\|^2 \int_0^{+\infty} |u(t)|^2 dt. \tag{11}$$

Using the Poincare inequality again, by (10) $\forall t \geq s \geq 0$ we have

$$\int_s^t \|y(s)\|^2 ds \leq \frac{1}{\lambda_1}\big(\|y(t)\|^2 + \|y(s)\|^2 + C_2 \|h_\varepsilon\|^2 \int_s^t |u(s)|^2 ds\big). \tag{12}$$

Herefrom, in particular, this implies that $J(y,u) < \infty$.

The next lemma is needed for passing to the limit in the problem (1) and it follows from The Mazur Theorem [7].

**Lemma 1.** *Let $Q$ be a bounded set, $q \geq 1$ and functions $f_n$, $q_n$, $l_n \in L^q(Q)$ satisfy*

$$f_n(x) \leq l_n(x) \leq g_n(x) \text{ for a. a. } x \in Q,$$

$$f_n \to f,\ l_n \to l,\ g_n \to g \text{ weakly in } L^q(Q).$$

*Then*

$$f(x) \leq l(x) \leq g(x) \text{ for a. a. } x \in Q.$$

**Theorem 1.** *Under (4) - (6) for $\forall \varepsilon > 0$, $\forall y_0^\varepsilon \in L^2(\Omega)$ the optimal control problem (1) - (3) has at least one solution.*

*Доказательство.* Let $\widetilde{J}_\varepsilon$ be a value of the problem (1) - (3). We choose $\{u_n\} \subset U$ such that $\forall n \geq 1$

$$J(y_n, u_n) \leq \widetilde{J}_\varepsilon + \frac{1}{n}.$$

Then

$$\gamma \int\limits_0^{+\infty} |u_n(t)|^2 dt \le \widetilde{J}_\varepsilon + \frac{1}{n},$$

so $\{u_n\}$ is bounded in $L^2(0,+\infty)$ and for some $u \in U$ on subsequence

$$u_n \to u \text{ weakly in } L^2(0,+\infty).$$

From the estimates (10), (11) for $\forall T > 0$

$$\{y_n\} \text{ is bounded in } L^\infty(0,T;L^2(\Omega)) \bigcap L^2(0,T;H_0^1(\Omega)) \bigcap L^p(0,T;L^p(\Omega)). \tag{13}$$

By the condition (6) we have

$$\{f_\varepsilon(y_n)\}, \{g_\varepsilon(y_n)\} \text{ are bounded in } L^q(0,T;L^q(\Omega)).$$

For $l_n(t,x) \in F_\varepsilon(x, y_n(t,x))$ $\exists \lambda_n = \lambda_n(t,x) \in [0,1]$ such that for a. a. $(t,x)$

$$l_n(t,x) = \lambda_n b_\varepsilon(x) f_\varepsilon(y_n(t,x)) + (1-\lambda_n) d_\varepsilon(x) g_\varepsilon(y_n(t,x)).$$

And since $b_\varepsilon$, $d_\varepsilon$ are bounded in $L^\infty(\Omega)$, then

$$\{l_n\} \text{ is bounded in } L^q(0,T;L^q(\Omega)). \tag{14}$$

This implies that

$$\{\frac{\partial y_n}{\partial t}\} \text{ is bounded in } L^2(0,T;H^{-1}(\Omega)) + L^q(0,T;L^q(\Omega)). \tag{15}$$

From the Compactness Theorem [5] for some function $y \in W$ on subsequence

$$\begin{gathered} y_n \xrightarrow{w} y \text{ in } L^2(0,T;H_0^1(\Omega)), \\ y_n \to y \text{ in } L^2(0,T;L^2(\Omega)), \\ y_n(t) \xrightarrow{w} y(t) \text{ in } L^2(\Omega)\ \forall\, t \ge 0, \\ y_n(t) \to y(t) \text{ in } L^2(\Omega) \text{ for a.a. } t \ge 0, \\ y_n(t,x) \to y(t,x) \text{ a. e.}, \\ l_n \xrightarrow{w} l \text{ in } L^q(0,T;L^q(\Omega)). \end{gathered} \tag{16}$$

Passing to the limit in (7) at $n \to \infty$, we have that $\{y, u, l\}$ satisfies (7).

By Lions Lemma [5] $b_\varepsilon f_\varepsilon(y_n) \to b_\varepsilon f_\varepsilon(y)$, $d_\varepsilon g_\varepsilon(y_n) \to d_\varepsilon g_\varepsilon(y)$ at $n \to \infty$ weakly in $L^q((0,T)\times\Omega)$ and a.e. In this case for a. a. $(t,x)$

$$b_\varepsilon f_\varepsilon(y_n(t,x)) \le l_n(t,x) \le d_\varepsilon g_\varepsilon(y_n(t,x)).$$

Then from the Lemma 1 $l(t,x) \in F_\varepsilon(x, y(t,x))$ a. e.

Hence, $\{y, u\}$ is the admissible process in the problem (1) - (3), and inequality

$$J(y_n, u_n) \geq J_T(y_n, u_n) := \int_0^T \|y_n(t)\|^2 dt + \gamma \int_0^T |u_n(t)|^2 dt$$

implies that $\forall T > 0$

$$\widetilde{J}_\varepsilon \geq \varliminf_{n\to\infty} J(y_n, u_n) \geq \varliminf_{n\to\infty} J_T(y_n, u_n) \geq$$
$$\geq \varliminf_{n\to\infty} \int_0^T \|y_n(t)\|^2 dt + \gamma \varliminf_{n\to\infty} \int_0^T |u_n(t)|^2 dt \geq J_T(y, u).$$

It follows that $\widetilde{J}_\varepsilon = J(y, u)$, so $\{y, u\}$ is the optimal process of the problem (1) - (3).

$\square$

## 3. Convergence to optimal process of averaged problem

Let us consider now a limit averaged problem

$$\begin{cases} \frac{\partial y}{\partial t} = div(a^0(x)\nabla y) - F_0(x, y) + h_0(x)u(t), \ x \in \Omega, \ t > 0, \\ y(x, t) = 0, \ x \in \partial\Omega, \\ y(x, 0) = y_0, \end{cases} \tag{17}$$

$$u(t) \in U \subseteq L^2(0, +\infty), \tag{18}$$

$$J(y, u) = \int_0^{+\infty} \int_\Omega y^2(x, t) dx dt + \gamma \int_0^{+\infty} u^2(t) dt \to \inf, \tag{19}$$

where $F_0(x, y) = b(x) f(y)$ and for $\varepsilon \to 0$

$$\begin{gathered} a^\varepsilon \to a^0, \ h^\varepsilon \to h_0 \text{ in } L^2(\Omega), \\ y_0^\varepsilon \to y_0 \text{ weakly in } L^2(\Omega), \\ b_\varepsilon \to b, \ d_\varepsilon \to b \text{ *-weakly in } L^\infty(\Omega), \\ \forall R > 0 \sup_{|y|\leq R} (|f_\varepsilon(y) - f(y)| + |g_\varepsilon(y) - f(y)|) \to 0. \end{gathered} \tag{20}$$

By (20) this implies that the matrix $a(x)$ is symmetric and satisfies (4), $b(x)$ satisfies (5) and $f \in C(\mathbb{R})$ satisfies (6). Hence, by Theorem 1 the optimal control problem (17) - (19) has solutions and we can consider the problem (1) as the perturbed problem (17). Such a situation naturally arises when modeling of complex evolutionary processes in micro-inhomogeneous media.

The following condition is supposed to satisfy:

$$\forall u \in U \ \forall y_0 \in L^2(\Omega) \text{ the problem (17) has the unique solution.} \tag{21}$$

The following condition [5] is sufficient to carry out the condition (21):

$$f \in C^1(\mathbb{R}),\ f'(u) \geq -C_4\ \forall u \in \mathbb{R}.$$

**Theorem 2.** *Let the conditions (4) - (6), (20), (21) hold. Then*

$$\lim_{\varepsilon\to 0} |\widetilde{J}_\varepsilon - \widetilde{J}_0| = 0,$$

*where $\widetilde{J}_\varepsilon$ is the value of the problem (1) - (3), $\widetilde{J}_0$ is the value of the problem (17) - (19).*

*Доказательство.* Let $\{\widetilde{y}^\varepsilon, \widetilde{u}^\varepsilon\}$ be an optimal process of the problem (1) - (3). Note that for any admissible process $\{y, u\}$ in the problem (1) - (3) the estimates (10), (11) are valid. Therefore if $z^\varepsilon$ is the solution of (1) with control $u \equiv 0 \in U$, then by the optimality of $\widetilde{u}^\varepsilon$ we have

$$\int_0^{+\infty} |\widetilde{u}^\varepsilon(t)|^2 dt \leq \frac{1}{\gamma}\int_0^{+\infty} \|z^\varepsilon(t)\|^2 dt \leq \frac{1}{\gamma}\int_0^{+\infty} \|y_0^\varepsilon\|^2 e^{-\lambda_1 t} dt \leq \frac{\|y_0^\varepsilon\|^2}{\lambda_1\gamma}. \quad (22)$$

Hence $\{\widetilde{u}^\varepsilon\}$ is bounded in $L^2(0, +\infty)$ and for some $\widetilde{u} \in U$ on subsequence

$$\widetilde{u}^\varepsilon \to \widetilde{u} \text{ weakly in } L^2(0, +\infty).$$

Let $\widetilde{l}^\varepsilon$ corresponds to $\widetilde{y}^\varepsilon$, $\widetilde{l}^\varepsilon(t, x) \in F_\varepsilon(x, \widetilde{y}^\varepsilon(t, x))$ a. e. Then we can repeat thinking of the Theorem 1 and obtain the convergence (16) for some $\widetilde{y} \in W$, $\widetilde{l} \in L^q((0, T) \times \Omega)$.

Let's argue the passing to the limit in the equality (7). Since $a^\varepsilon \to a^0$ in $L^2(\Omega)$ then

$$\int_0^T (a^\varepsilon \nabla\widetilde{y}^\varepsilon, \nabla v)\eta dt \to \int_0^T (a\nabla\widetilde{y}, \nabla v)\eta dt\ \forall v \in H_0^1(\Omega), \forall\eta \in C_0^\infty(0, T).$$

Due to strong convergence $a^\varepsilon \to a^0$, $h^\varepsilon \to h_0$ in $L^2(\Omega)$, we can pass to the limit in the equality (7) and obtain that $\{\widetilde{y}, \widetilde{u}, \widetilde{l}\}$ satisfies (7) for $\forall T > 0$.

Prove that $\widetilde{l}(t, x) = b(x)f(\widetilde{y}(t, x))$ a. e. In fact, $f_\varepsilon(\widetilde{y}^\varepsilon) \to f(\widetilde{y})$ weakly in $L^q(0, T; L^q(\Omega))$ and a. e., $b_\varepsilon \to b$ *-weakly in $L^\infty(\Omega)$. Then

$$b_\varepsilon f_\varepsilon(\widetilde{y}^\varepsilon) - bf(\widetilde{y}) = b_\varepsilon(f_\varepsilon(\widetilde{y}^\varepsilon) - f(\widetilde{y})) + (b_\varepsilon - b)f(\widetilde{y}) = I_\varepsilon^{(1)}(t, x) + I_\varepsilon^{(2)}(t, x).$$

Since $b_\varepsilon$ is bounded in $L^\infty(\Omega)$, then $I_\varepsilon^{(1)}(t, x) \to 0$ a.e. and it is bounded in $L^q(0, T; L^q(\Omega))$. Hence, by Lions Lemma $I_\varepsilon^{(1)}(t, x) \to 0$ weakly in $L^q(0, T; L^q(\Omega))$.

On the other hand, $\forall\theta \in L^p(0, T; L^p(\Omega))$ $f(\widetilde{y}) \cdot \theta \in L^1((0, T) \times \Omega)$, therefore

$$\int_0^T \int_\Omega (b_\varepsilon(x) - b(x))f(\widetilde{y}(t, x))\theta(t, x) \to 0,$$

i. e. $I_\varepsilon^{(2)}(t, x) \to 0$ weakly in $L^q(0, T; L^q(\Omega))$.

Thus,

$$b_\varepsilon(x)f_\varepsilon(\widetilde{y}(t,x)) \le \widetilde{l}(t,x) \le d_\varepsilon(x)g_\varepsilon(\widetilde{y}(t,x)) \text{ a. e.,}$$

moreover,

$$b_\varepsilon \cdot f_\varepsilon(\widetilde{y}^\varepsilon) \to b \cdot f(\widetilde{y}),\ d_\varepsilon g_\varepsilon(\widetilde{y}) \to b \cdot f(\widetilde{y}) \text{ weakly in } L^q(0,T;L^q(\Omega)),$$

$$\widetilde{l}^\varepsilon \to \widetilde{l} \text{ weakly in } L^q(0,T;L^q(\Omega)).$$

Then by the Lemma 1 we have that $\widetilde{l}(t,x) = b(x)f(\widetilde{y}(t,x))$ a. e.

Moreover, $\widetilde{y}^\varepsilon \to \widetilde{y}$ in $C([\tau,T];L^2(\Omega))$ $\forall \tau > 0$. So $\forall T > 0$

$$\varliminf_{\varepsilon\to 0} \widetilde{J}_\varepsilon \ge \varliminf_{\varepsilon\to 0} J_T(\widetilde{y}^\varepsilon, \widetilde{u}^\varepsilon) \ge J_T(\widetilde{y},\widetilde{u}),$$

hence

$$\varliminf_{\varepsilon\to 0} \widetilde{J}_\varepsilon \ge J(\widetilde{y},\widetilde{u}). \tag{23}$$

Using Bellman optimality principle, we can argue [4] that $\{\widetilde{y},\widetilde{u}\}$ is an optimal process of the problem (17) - (19).

Let's prove that

$$\varlimsup_{\varepsilon\to 0} \widetilde{J}_\varepsilon \le J(\widetilde{y},\widetilde{u}). \tag{24}$$

From Bellman optimality principal we obtain that the process $\{\widetilde{y}^\varepsilon, \widetilde{u}^\varepsilon\}$ is optimal for the problem (1) – (3) on $[T,+\infty)$ with initial data $(T,\widetilde{y}^\varepsilon(T))$. Then for every $T > 0$ by (12) the following inequality holds

$$\int_T^{+\infty} \|\widetilde{y}^\varepsilon(t)\|^2 dt + \gamma \int_T^{+\infty} |\widetilde{u}^\varepsilon(t)|^2 dt \le \int_T^{+\infty} \|p^\varepsilon(t)\|^2 dt \le \frac{1}{\lambda_1}\|\widetilde{y}^\varepsilon(T)\|^2, \tag{25}$$

where $p^\varepsilon$ is the solution of the problem (1) with control $u = 0 \in U$ and initial data $(T,\widetilde{y}^\varepsilon(T))$.

Let $\omega^\varepsilon$ be a solution of the problem (1) with control $\widetilde{u}$. Then from (21) we have that $\omega^\varepsilon \to \widetilde{y}$ in the sense of (16). Moreover, we obtain the following estimates:

$$\begin{gathered} \int_0^T \|\widetilde{y}^\varepsilon(t)\|^2 dt + \gamma \int_0^{+\infty} |\widetilde{u}^\varepsilon(t)|^2 dt \le \\ \le \gamma \int_0^{+\infty} |\widetilde{u}(t)|^2 dt + \int_0^T \|\omega^\varepsilon(t)\|^2 dt + \int_T^{+\infty} \|\omega^\varepsilon(t)\|^2 dt \le \\ \le \frac{1}{\lambda_1}\|\omega^\varepsilon(T)\|^2 + \gamma \int_0^{+\infty} |\widetilde{u}(t)|^2 dt + \int_0^T \|\omega^\varepsilon(t)\|^2 dt + \frac{C_1}{\lambda_1}\int_T^{+\infty} |\widetilde{u}(t)|^2 dt. \end{gathered} \tag{26}$$

Then

$$\gamma \overline{\lim_{\varepsilon\to 0}} \int_0^{+\infty} |\widetilde{u}^{\varepsilon}(t)|^2 dt \leq \int_0^{+\infty} |\widetilde{u}(t)|^2 dt + \frac{2}{\lambda_1}\|\widetilde{y}(T)\|^2 + \frac{C_1}{\lambda_1}\int_T^{+\infty} |\widetilde{u}(t)|^2 dt$$

and for $T \to \infty$ we get

$$\overline{\lim_{\varepsilon\to 0}} \int_0^{+\infty} |\widetilde{u}^{\varepsilon}(t)|^2 dt \leq \int_0^{+\infty} |\widetilde{u}(t)|^2 dt, \tag{27}$$

which together with weak convergence guarantees strong convergence $\widetilde{u}^{\varepsilon} \to \widetilde{u}$ in $L^2(0, +\infty)$.

Further from inequalities (25), (26) we obtain the following inequality

$$\widetilde{J}_{\varepsilon} \leq J_T(\widetilde{u}^{\varepsilon}) + \frac{1}{\lambda_1}|\widetilde{y}^{\varepsilon}(T)|^2.$$

Then

$$\overline{\lim_{\varepsilon\to 0}}\, \widetilde{J}_{\varepsilon} \leq J_T(\widetilde{u}) + \frac{1}{\lambda_1}|\widetilde{y}(T)|^2$$

and for $T \to \infty$ we get (24), which means together with (23) that on some subsequence

$$\lim_{\varepsilon\to 0} \widetilde{J}_{\varepsilon} = J(\widetilde{y}, \widetilde{u}).$$

Assuming by contradiction that this convergence goes on not all sequence $\varepsilon \to 0$, we can repeat previous thinking and under uniqueness of optimal process $\{\widetilde{y}, \widetilde{u}\}$ we obtain the contradiction.

□

## Conclusion

In this paper the following results were obtained:

– we proved the solvability of optimal control problem (1)–(3),

– we proved convergence of the optimal process of the problem (1)–(3) to optimal process of corresponding averaged problem (17)–(19).

## References

1. Zhykov, V. V. and Kozlov, V. V., Oleynik, O. A. 1993. *Averaging of Differential Operators*. Moscow: Physmatlit.
2. Fursikov, A. V. 1999. *Optimal Control by Distributed Systems. Theory and Application*. Novosibirsk: Nauchnaya kniga.
3. Kapustyan, O. and Kapustian, O. and Sukretna, A. 2011. Approximated stabilization for nonlinear parabolic problem. *Ukrain. Math. J.*, 63 (5), pp. 759–767.

4. Kapustian, O. and Yasinski, V. 2012. Approximated solution of one infinite-dimensional optimal stabilization problem with non-autonomous perturbations in coefficients. *Naukovi visti KPI*, 4, pp. 111–115.
5. Temam, R. 1997. *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*. New York: Springer.
6. Shkliar, T. 2011. The global attractor of non-autonomous evolutionary inclusion of reaction-diffusion type. *Naukovi visti KPI*, 4(78), pp. 98–104.
7. Iosida, K. 1967. *Functional Analysis*. Moscow: Mir.

# THE MECHANISMS OF DECISION-MAKING INTELLECTUALIZATION BASED ON DISTRIBUTED COGNITIVE RESOURCES

**© Victor Krasnoproshin**

Belarusian State University
Department of Applied Mathematics and Informatics
e-mail: *krasnoproshin@bsu.by*

***Abstract.*** *In the article the theoretical and practical results of researches are generalized on the problem of intellectualization of decision-making in the conditions of global informatively-communication environment. Results are got within the framework of implementation of the proper scientific programs and realization of the different (including international) applied projects.*

*Under the decision-making the choice of the best (on some criterion) alternative is understood from the finite set of possible variants for permission of problem situation. The term "intellectualization" is examined as process of integration in the mechanisms of decision-making innovative knowledge, providing the competitive edges for the states, companies and having a special purpose groups of population. Innovative knowledge (further are knowledge) are the structured information giving the maximal competitive edges at the decision of concrete practical tasks. The specialized complexes of the programs, known as systems of support of decision-making, are the basic mean of intellectualization.*

*The problems of intellectualization are formulated in the directives documents of guidance of the largest countries of world. Interesting results which touch the economic and social aspects of problem mainly are got. The questions of design and a algorithmization of process of intellectualization are investigational, unfortunately, it is not enough, that hampers his realization in practice. In particular, major questions tied-up are opened: (1) with the choice of form of knowledge representation for their effective use; (2) with construction of stage providing access to knowledge of different categories of users (from the presidents of the states to the having a special purpose groups of population); (3) with integration of the knowledge got from local sources, in the global systems of accumulation; (4) with diminishing of speed of obsolescence of knowledge, i.e. by providing of their permanent competence; (5) with the choice of knowledge in most degree proper appropriate to preferences and level of provisioning of users.*

*The complex decision of tasks (1–5) is represented in the article based on the synthesis of elements of theory of decision-making, theory of organization and possibilities of modern communications and programmatic technologies. This decision of problem of intellectualization of decision-making is in-process considered on the basis of synthesis of elements of expert knowledge and cognitive resources the Internet. Conception of intellectualization is developed on the basis of the innovative knowledge represented in the form of subject collections. The models of subject collection and stage of realization of its life cycle are offered. The charts of algorithms of construction, choice, estimation and actualization of subject collections are represented. The variant of the system, realizing algorithms on the basis of multi agent approach in the form of portal, and experience of his application, is described for the decision of practical tasks.*

## Введение

В статье обобщены теоретические и практические результаты исследований по проблеме интеллектуализации принятия решений в условиях глобальной информационно-коммуникационной среды. Результаты получены в рамках выполнения соответствующих научных программ и реализации различных (в том числе и международных) прикладных проектов.

Под принятием решений понимается выбор лучшей (по некоторому критерию) альтернативы из конечного множества возможных вариантов для разрешения проблемной ситуации [1]. Термин "интеллектуализация" рассматривается как процесс интеграции в механизмы принятия решений инновационных знаний, обеспечивающих конкурентные преимущества для государств, компаний и целевых групп населения [2, 3, 4]. Инновационные знания (далее — знания) — это структурированная информация, дающая максимальные конкурентные преимущества при решении конкретных практических задач [3]. Основным средством интеллектуализации являются специализированные комплексы программ, известные как системы поддержки принятия решений (СППР) [1, 4].

Проблемы интеллектуализации сформулированы в директивных документах руководства крупнейших стран мира [5, 6] и исследуются в работах известных ученых Saaty [2], Schilling [3], Князева, Шрубенко [4] и др. Получены интересные результаты, которые касаются в основном экономических и социальных аспектов проблемы. Вопросы моделирования и алгоритмизации процесса интеллектуализации исследованы, к сожалению, недостаточно, что затрудняет его реализацию на практике. В частности, открытыми остаются важнейшие вопросы, связанные: (1) с выбором формы представления знаний для их эффективного использования; (2) с построением сцены, обеспечивающей доступ к знаниям различных категорий пользователей (от президентов государств до целевых групп населения); (3) с интеграцией знаний, полученных из локальных источников, в глобальные системы аккумуляции; (4) с уменьшением скорости устаревания знаний, т.е. обеспечением их постоянной компетентности; (5) с выбором знаний, в наибольшей степени соответствующих предпочтениям и уровню подготовки пользователей.

В статье представлено комплексное решение задач (1)–(5), основанное на синтезе элементов теории принятия решений, теории организации и возможностей современных коммуникаций и программных технологий.

## 1. Постановка задачи

В литературе описаны различные варианты постановки задачи интеллектуализации принятия решений (ЗИПР), которые в большей степени характерны для теоретических исследований [1–4]. Ниже предлагается вариант, который изначально ориентирован на технологическое применение и прошел апробацию в качестве основы для автоматизации ЗИПР в области медицины, криминалистики и программирования.

Задачу принятия решений (ЗПР) формально опишем в виде кортежа:

$$\mathrm{Z} = (\mathrm{S}, \mathrm{X}, \mathrm{G}, \mathrm{V}, \mathrm{L}, \mathrm{Kr}, \mathrm{fV}, \mathrm{fV}, \mathrm{fEs}, \mathrm{fCh}, \boldsymbol{V}), \tag{1}$$

где: S — проблемная ситуация; X — признаки, характеризующие эту ситуацию; G — цель разрешения ситуации; V — альтернативы (возможные вариантов решения); L — условия, которым должно удовлетворять решение; Kr — критерии выбора; fV, fEs и fEs — соответственно механизмы построения, оценивания альтернатив и выбора лучшей альтернативы; $\boldsymbol{V}$ — выбранная альтернатива.

В соответствии с (1), интеллектуализацию можно рассматривать как процесс разработки (поиска), использования и поддержки компетентности элементов fV, fEs, fCh, V. Под компетентностью понимается полнота, точность и актуальность всех элементов модели принятия решений.

Альтернативы V в классическом понимании, как правило, представлены однородными (гомогенными) строковыми переменными [1]. Однако в последние годы ситуация изменилась: альтернативы стали рассматривать как формализованные сложно структурированные разнородные (гетерогенные) знания, представляющие различные варианты полного (теоретического и практического) решения задачи. Для обеспечения конкурентоспособности альтернатива V должна носить инновационный характер. Исходя из этих соображений, предлагается постановка ЗИПР, которая подходит для многих типов организационных систем.

Пусть имеется организация C, которая периодически решает задачи Z. Конкурентоспособность решения определяют инновационные знания V, элементами которых $\mathrm{V}_1, \mathrm{V}_2, \ldots, \mathrm{V}_\mathrm{n}$ обладают распределенные эксперты $\mathrm{E}_1, \mathrm{E}_2, \ldots, \mathrm{E}_\mathrm{n}$. Знания V используются распределенными участниками организации (пользователями) U. Требуется разработать СППР, обеспечивающую интеллектуализацию решения Z.

Применяя принцип декомпозиции, выделим три основных подзадачи:

– построение моделей представления знаний (альтернатив) V и сцены решения задачи $Z$;

– разработка алгоритмов построения сцены и реализации жизненного цикла ЗИПР;

– разработка соответствующей архитектуры и системы интеллектуализации.

Основное требование к результату связано с использованием стандартного программного обеспечения (ПО), коммуникаций и средств доступа.

## 2. Модели

Первый вопрос, возникающий при реализации процесса интеллектуализации, связан с выбором формы представления знания, которая обеспечила бы их эффективное использование. В когнитивных системах компаний — IBM, CISCO, Microsoft — знания представлены в одном или двух форматах (например, в msdn — текст + видео), что значительно сужает сферу их применения. Кроме того, в них отсутствует удобные механизмы оценки полезности, актуальности и улучшения знаний. Для устранения этих недостатков предлагается вариант формы представления знаний, основанный на классической модели человеческой деятельности L. Mises [9], адаптированной к особенностям постиндустриальной эпохи:

$$\mathrm{mV_1} = \bigcup_{\mathrm{i=1}}^{\mathrm{n}} \big(\mathrm{nZ, Z, nZ_i, Z_i, Alg_i, Tech_i, Exp, ind}, \Delta\big), \tag{2}$$

где: nZ и Z — название и постановка задачи; n и i — количество и номер подзадач; Alg и Tech — алгоритм и средство решения; Exp — руководство пользователю; ind и $\Delta$ — независимые оценка (индекс) полезности контента и предложения по его улучшению.

Модель (2), в отличие от описанных в литературе вариантов представления знаний, обеспечивает многофункциональное использование контента (например, Z, Alg — для корпоративной библиотеки алгоритмов; Z, Tech — для локального тиражирования на CD; Z, ind — для рекламы и т.д.). Пользователи могут оценить полезность ind контента на основе собственного опыта его применения. Контент могут оценить члены независимых профессиональных сообществ в социальных сетях, их же можно рассматривать как источник материала $\Delta$ для усовершенствования Alg, Tech, Exp [7].

Для практического применения модели (2) необходима соответствующая сцена. В ней должны участвовать: центр (C), инициирующий решение задачи; эксперты (E), формирующие знания (контент альтернатив); пользователи контента (U); члены профессиональных сообществ (SocNet) в социальных сетях Twitter, Facebook (для независимой оценки контента) и др.; средства обеспечения диалога между участниками (dp) и регламентации доступа участников к контенту (Valid); глобальные коммуникации (com). Соответствующая сцена описывается кортежем:

$$\mathrm{Scene} = (\mathrm{C, E, U, SocNet, dp, Valid, com}). \tag{3}$$

Для построения универсальной модели участников сцены (акторов), предлагается использовать модель речевого акта J. Austin [10], адаптированную к условиям информационно-коммуникативной среды:

$$\mathrm{Actor} = (\mathrm{e\_adr}, \mathrm{Name}, \mathrm{p\_adr}, \mathrm{Status}, \mathrm{inf}, \mathrm{Role}, \mathrm{Dlg}(\mathrm{Q}, \mathrm{R})), \tag{4}$$

где: e_adr — адрес актора в сети; Name — имя; p_adr — административный адрес; Status — статус; inf — дополнительная информация; Role — роль; Dlg — диалог (Q — вопрос, R — ответ).

Согласно (4), одушевленный или искусственный актор рассматривается как участник сцены, который имеет определенный статус и реализует свою роль и диалог в инфраструктуре глобальной сети.

Для практического применения модели (2) в рамках сцены (3) ее необходимо дополнить атрибутами, обеспечивающими идентификацию контента в различных средах. Как минимум, это атрибуты глобальной (idG), предметной (idD) и корпоративной (idC) среды:

$$\mathrm{mV2} = (\mathrm{idG}, \mathrm{idD}, \mathrm{idC}, \mathrm{mV1}) \tag{5}$$

Для полноты картины кортеж (5) дополняется участниками сцены:

$$\mathrm{SC} = (\mathrm{Scene}, \mathrm{idG}, \mathrm{idD}, \mathrm{idC}, \mathrm{mV1}) \tag{6}$$

Представление инновационных знаний в форме (6) названа нами предметной коллекцией (ПрК). В отличие от существующих моделей она обеспечивает поддержку полного жизненного цикла инноваций (ЖЦ ПрК), который включает в себя стадии создания, поиска, применения, оценки и усовершенствования контента.

Для интеграции ПрК с глобальными системами (включая Semantic web systems) в структуру ПрК включена онтология SWO (Semantic Web Ontology), которая описывает контент на одном из специализированных языков (OWL, KIF, DAML, RDF):

$$\mathrm{SC} = (\mathrm{SWO}, \mathrm{Scene}, \mathrm{idG}, \mathrm{idD}, \mathrm{idC}, \mathrm{mV1}) \tag{7}$$

В результате у внешних систем появляется возможность быстрого поиска и включения контента ПрК в собственные базы знаний. Для интеграции ПрК с внешними локальными системами достаточно сохранить ПрК как XML-файл и использовать его как входную информацию для внешней системы. Общая схема реализации ЖЦ ПрК представлена на рис. 1.

Для реализации ЖЦ ПрК в рамках данной схемы необходим соответствующий комплекс алгоритмов.

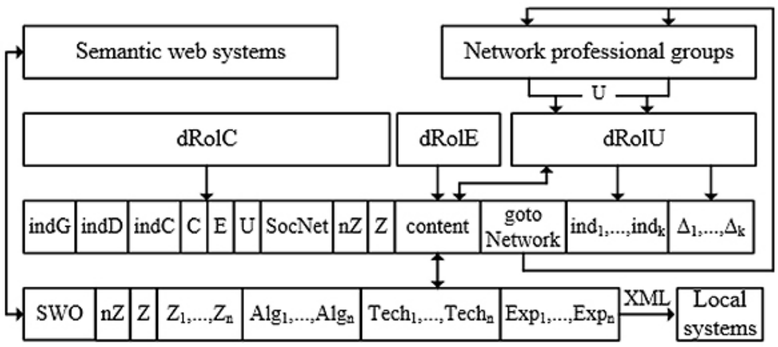



*Рис.* 1. Схема взаимодействия акторов сцены

## 3. Алгоритмы

Разработка алгоритмов для реализации ЖЦ ПрК является сложной задачей. Вход-выход таких алгоритмов представлен сотнями территориально распределенных узлов, соответствующих участникам сцены, которые выполняют различные роли и используют для этого разные типы диалогов. Фактически это новый класс алгоритмов, которым больше подходит термин "механизмы", т.к. они жестко связаны с "внешними" составляющими решения, включая аппаратную часть, коммуникации и системное глобальное программное обеспечение.

Согласно (6), необходимо разработать алгоритм построения сцены и алгоритмы формирования, выбора, оценивания и актуализации ПрК. Ниже приведены общие схемы таких алгоритмов.

**Алгоритм 1: построение сцены** (актор C).

Шаг 1. Анализ задачиn Z;

Шаг 2. Определение участников сцены решения: C, E, U, SocNet;

Шаг 3. Загрузка реквизитовC, E, U, SocNet в БД СППР;

Шаг 4. Загрузка в Valid паролей для C, E, U, SocNet;

В результате формируется БД акторов сцены, и создаются условия для реализации ЖЦ ПрК.

**Алгоритм 2: построение ПрК** (акторы C, E).

Шаг 1: Инициализация шаблона ПрК согласно (2);

Шаг 2: Отправка шаблона экспертам E;

Шаг 3: Формирование экспертами контента mV1;

Шаг 4: Отправка контента mV1 в центр C;

Шаг 5: Формирование SC.

В результате формируется база ПрК и появляется возможность их применения для решения практических задач различными категориями пользователей.

**Алгоритм 3: принятие решений на основе ПрК** (акторы

Шаг 1. Инициализация проблемы Z;

Шаг 2. Поиск SC, релевантных проблеме: SC1, SC2, . . . , SCn;

Шаг 3. Ранжирование SC по индексу полезности ind;

Шаг 4. Выбор:

– альтернативы SCi с максимальным значением индекса полезности ind;

– альтернативы SCj с меньшим значением уровня полезности, но более соответствующей возможностям и уровню подготовки пользователя;

Шаг 5: применение выбранной SC для решения практических задач.

Таким образом, обеспечивается выбор решения, соответствующего предпочтениям различных категорий пользователей.

**Алгоритм 4: оценка ПрК** (акторы U, SocNet).

Шаг 1. Поиск SC;

Шаг 2. Оценка пользователями U полезности ind контента на основе собственного опыта;

Шаг 3. Предложение $\Delta$ для улучшения контента.

При большом количестве оценок контента можно говорить об объективности значения индекса ind. Низкое значение индекса полезности говорит о необходимости актуализации контента.

**Алгоритм 5: актуализация ПрК** (акторы C, E).

Шаг 1. Поиск SC с минимальными значениями индекса полезности ind;

Шаг 2. Анализ индекса ind и предложений $\Delta$;

Шаг 3. Принятие или отклонение $\Delta$;

Шаг 4. Коррекция контента SC.

Наиболее эффективным считается вариант коррекции, который обеспечивает консенсус между мнениями экспертов и участниками профессиональных сообществ в социальных сетях.

Описанные выше модели и алгоритмы являются базой, на основе которой построена архитектура компьютерной системы интеллектуализации принятия решений.

## 4. Система

Сцена интеллектуализации (3) носит распределенный характер, поэтому для построения архитектуры целевой системы использовался многоагентный подход [11]. На основе классической архитектуры (sensor-effector-processor-memory) построены четыре агента. Агент Center реализует процессы, соответствующие алгоритмам 1, 2;

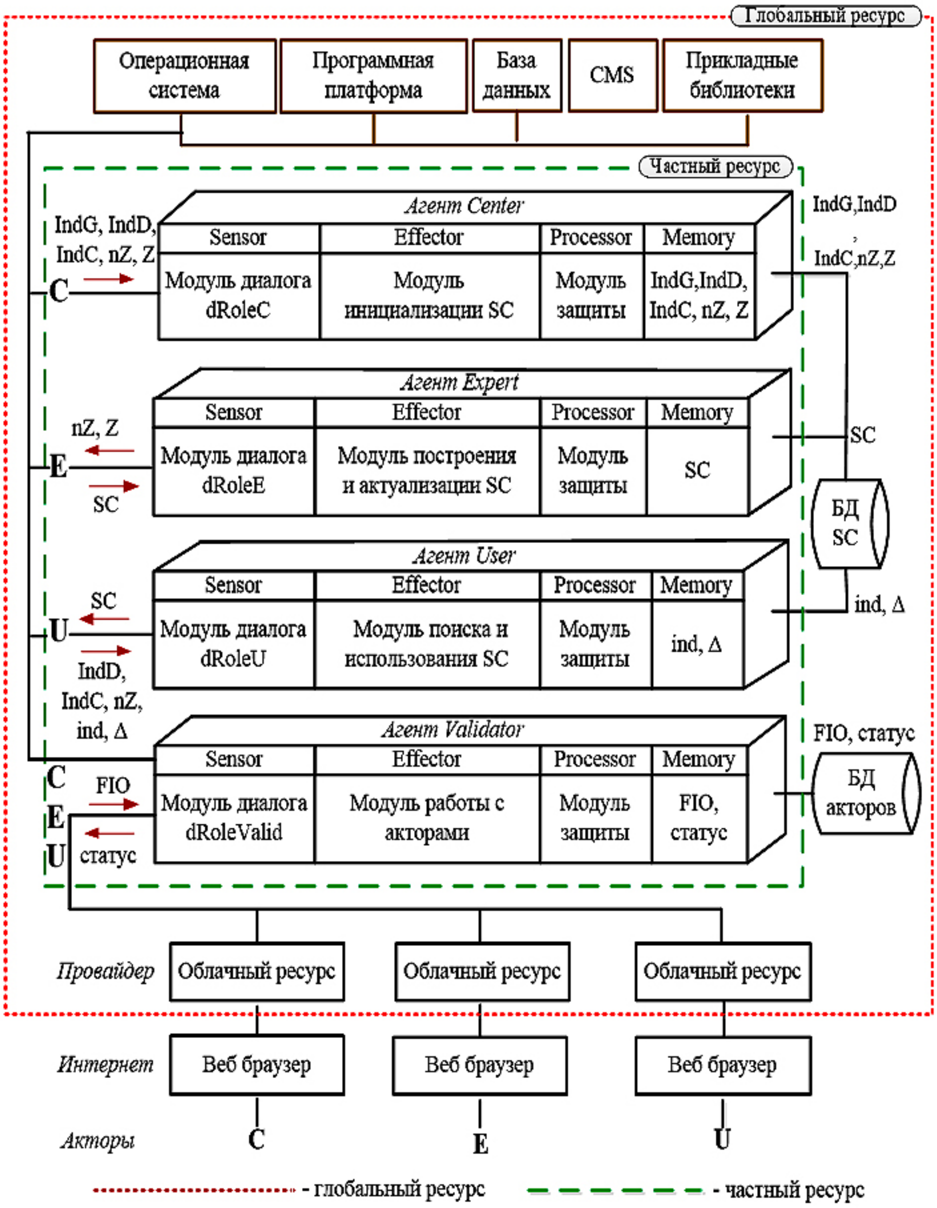



*Рис.* 2. Архитектура системы интеллектуализации

Expert — алгоритмам 2, 5; User — алгоритмам 3, 4; Validator осуществляет контроль доступа актров к ПрК. Для уменьшения расходов на поддержку жизненного цикла системы все ПО размещено как частный ресурс в облачном ресурсе компании Byelex. Доступ к прикладному ПО и ПрК осуществляется через стандартный браузер (MS Explorer, Chrome, Opera и др.). Соответствующая архитектура представлена на рис. 2.

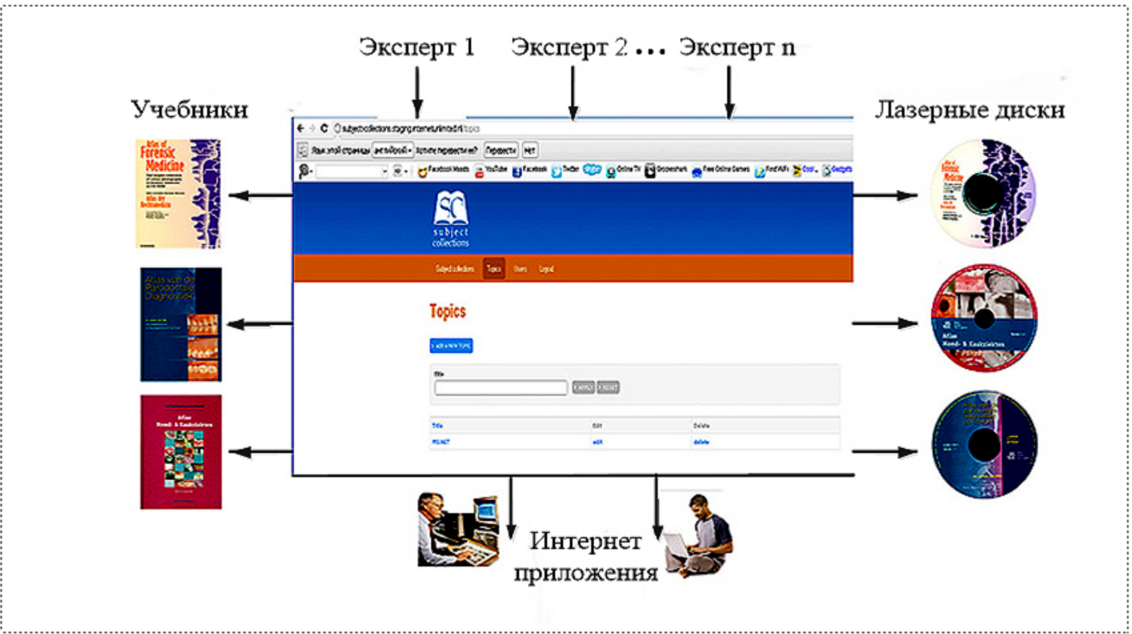



*Рис.* 3. Портал для реализации ЖЦ ПрК

Архитектура системы реализована в форме портала, которой прошел эволюцию от стандартного сайта для сбора информации по заданной тематике до хранилища ПрК для сферы ИТ (с широким спектром средств визуализации, публикации и актуализации). Для каждого нового проекта строится оригинальные ПрК, которые после завершения проекта интегрируются в частные ресурсы заказчика или правообладателя (рис. 3).

В период 2000–2013 гг. был накоплен значительный опыт эксплуатации портала. В частности, он использовался для реализации ряда международных проектов, включая “Orthopedishe Casuistiek”, “Atlasvande Parodontale Diagnostic”, “AtlasMond-&Kaakziekten” (для принятия решений в области ортопедии), “Atlasof Forensic Medicine” (для судебной медицины) и др. Построенные в рамках данных проектов ПрК были опубликованы на бумажных носителях (книги), лазерных дисках и размещены в сетевых ресурсах. До настоящего времени ПрК “Atlas of Forensic Medicine” является крупнейшей в мире коллекцией по данной тематике. В процессе реализации проектов структура предметной коллекции не претерпела существенных изменений, что говорит о ее соответствии современным требованиям. В целом, опыт применения портала подтвердил жизнеспособность и эффективность разработанного подхода к решению проблемы интеллектуализации принятия решений.

## Заключение

В работе рассмотрено комплексное решение проблемы интеллектуализации принятия решений на основе синтеза элементов экспертных знаний и когнитивных ресурсов Интернет. Разработана концепция интеллектуализации на основе инновационных знаний, представленных в форме предметных коллекций. Предложены модели предметной коллекции и сцены реализации ее жизненного цикла. Представлены схемы алгоритмов построения, выбора, оценки и актуализации предметных коллекций. Описан вариант системы, реализующий алгоритмы на основе многоагентного подхода в форме портала, и опыт его применения для решения практических задач.

## Список литературы


1. Петровский А. Б. Теория принятия решений / А. Б. Петровский. — Академия, 2009. — 400 с.
   Petrovsky, A. B. 2009. *Decision making theory.* Moskow: Academy.
2. Saaty, T. L. 2008. *Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World.* RWS Publications.
3. Schilling, M. 2009. *Strategic Management of Technological Innovation.* McGraw-Hill.
4. Князев С. Н. Интеллектуализация — стержневая основа развития экономики и управления / С. Н. Князев, А. Г. Шрубенко // Проблемы управления. — 2007. — № 2. — С. 16–25.
   Knyazev, S. N. and Shrubenko, A. G. 2007. Intellectualization — the backbone base of economic and control progress. *Control Problems*, 2, pp. 16–25.
5. Government ICT Strategy. 2011. Cabinet Office. Whitehall. — London: Crown.
6. The Seventh Framework Programme. of the European Community for Research, Technological Development and Demonstration Activities. 2007–2013. European Commission. — Brussels.
7. Krasnoproshin, V., Obraztsov, V and Vissia, H. 2010. Overall Approach to the Solution of Applied Problems. *"Lectures on Modeling and Simulation" Special Issue of International Association AMSE.* — Barcelona: Spain, pp. 166-174.
8. Виссия Х. Интеллектуализация принятия решений на основе предметных коллекций / Х. Виссия, В. В. Краснопрошин, А. Н. Вальвачев // Вестник БГУ. Сер. 1. — 2011. — №3. — С. 84–90.
   Vissia, H., Krasnoproshin, V. and Volvachev, A. 2011. Decision making intellectualization based on the subject collections. *Belarus State University Reporter*, 1 (3), pp. 84–90.
9. Mises, L. 2007. *Human action: a treatise on economics.* Liberty Fund.
10. Austin, J. 1962. *How to do things with words.* Oxford at the Clarendon press.
11. Uhrmacher, A. and Weyns, D. 2009. Multi-Agent Systems: Simulation and Applications. CRC press.

# ON SIMILARITY OF PETRI NETS LANGUAGES

**© Elena Lukyanova**

Taurida National V. I. Vernadsky University

e-mail: *lukyanovaea@mail.ru*

***Abstract***. *The concept of languages similarity of Petri nets is introduced. It is determined, that mapping of languages similarity of Petri nets is a surjective homomorphism. The similarity of languages of component Petri net and original detailed Petri model of the investigated parallel distributed system is considered. The work reveals that the language of the original detailed Petri net model can always be restored using the language of its component model.*

## Introduction

Currently, the development of research in the field of theoretical computer science is caused by the necessity of development of formal methods of modeling and analysis of parallel distributed systems having complex structural organization and operating in real time. The establishment of adequate systems of this type is not a trivial task. The solution to this problem depends on the nature of the problems under consideration, class of simulated systems, the level of their structure detailing and behavior, and requires complex fundamental research of various formal methods and tools. Petri nets are one of the most popular and convenient modern formalisms for modeling and analysis of parallel distributed systems. This formalism has several important advantages, such as visibility, availability of simple structures to describe concurrency structures (sequential composition, choice, parallel merging) and the solubility of many behavioral properties [1, 2]. Petri nets allow, with sufficient detailing level, to model the computational processes, management processes in parallel systems and communication protocols. The main advantage of Petri nets is the ability to display the interaction of multiple parallel sequential processes as a single structure. This formalism has several drawbacks. High ability of Petri nets modeling and complexity of the simulated systems can lead to larger nets [3], [4] and as a result — to the "state explosion" problem [4]. Petri nets do not describe explicitly the dynamics of states change (behavior), and in analyzing the behavior of Petri nets we have to simultaneously monitor the situation and several points to remember these situations. In the case of errors localization, the route (path) to error site is not indicated. These circumstances are essential for the analysis of Petri nets, errors identifying and eliminating in the real system. In this connection there is a need to find trails that lead to the suspicious or erroneous state in the net operation. Such an analysis is logical to perform by constructing the relevant languages. As for significant reduction of verification efficiency in this formalism due to the "state explosion" problem, preliminary reduction of Petri net is required, which models system. The way to solve the problems formulated is in studying languages of component Petri nets ($CN$-net) [5, 6] and the establishment of links between languages of the original detailed Petri net model and the reduced $CN$ model — component Petri net. The purpose of this work is to continue studies [7, 8, 9] of language connections of detailed and component Petri models of parallel distributed system and to establish the possibility of language recovery of detailed Petri net model of investigated parallel distributed system by its reduced model language — component Petri net.

## 1. Preliminary information

Component Petri net, introduced in works [5, 6], is a directed graph, described by the ordering quinary:

$$CN = (P, T, F, W, M_0),$$

where $P = P_1 \cup P_2$ is a finite set of places ($P_1$ is a finite set of component-places, $P_2$ — a finite set of places that are left after the separation of component-places); $T = T_1 \cup T_2$ — final set of transitions ($T_1$ is a finite set of components-transitions and $T_2$ — a finite set of transitions that are left after the separation of the component transitions); $F \subseteq P \times T \cup T \times P$ — the incidence relation between places and transitions; $W : F \to N \setminus \{0\}$ — the multiplicity function of arcs; $M_0$ — the initial marking of net.

It is stated in work [10] that allocation procedure in the initial detailed Petri net model of the system under consideration with concurrency of composite components (component places $C_p$ and component transitions $C_t$) is a structural transformation that can significantly reduce the number of nodes of the net $N$ while preserving its behavioral properties. This means that $CN$-net, built as a result of such transformations, is adequate, and hence preserving the description expressiveness of the original system under consideration. The proof of the correctness of such transformations is justified by defining component $\chi_1$ ratio at the set of nodes of reachable markings of detailed Petri net model [11], establishing homomorphism of graphs of reachable markings of investigated Petri $N$ and $CN$ models, and proof of bisimular equivalence of $N$ and $CN$nets.

In works [7, 8, 9] the following languages of component Petri net were introduced: language $L_t(CN)$ of component Petri net containing only components-transitions $C_t$, language $L_p(CN)$ of component Petri net containing only components-places $C_p$, language $L_{p,t}(CN)$ of component Petri net containing components-places $C_p$ and components-transitions $C_t$. In this case when determining the languages $L_p(CN)$ and $L_{p,t}(CN)$, operation of nets $N$ and $CN$ is described in terms of the set of net reachable markings, and in determining the language $L_t(CN)$ — in terms of sequences of transitions firing. It is connected with the structures of respective composite components [12] and with the fact that composite component information, accumulated in the nodes of the component net, should be reflected in the words of the language of the corresponding component net.

## 2. Similarity of Petri nets languages

An important concept of the theory of formal systems is the notion of equivalence of behaviors. Equivalence of this type provides an opportunity to compare the parallel and distributed systems, taking into account certain aspects of their functioning. One type of behavioral equivalences for parallel systems and programs is language equivalence [13], i. e. the equivalence of languages, generated by systems. Language equivalence allows us to compare the behavior of both serial and parallel systems. Analytical representation is convenient for Petri net models of these systems, using a formula in algebra nets [14, 15, 16]. Net formula is constructed from symbols that define some basic net from net operations. With the help of these operations, the net described is built from elementary nets. In this way it is possible to verify the equality or inclusion of generated languages [1]. And what if languages are 'similar'? What does it mean — "similar", by how much?

For languages of Petri nets we introduce the concept of similarity of languages.

**Definition 1.** Similarity of Petri nets languages is understood as such transformation of Petri nets languages, defined over the same alphabet, which allows recovering one Petri net language by means of language of the other.

**Statement 1.** *Languages $L_t(N)$ and $L_t(CN)$ are similar.*

**Argument.** Consider languages $L_t(N)$ and $L_t(CN)$ [7] of some Petri net $N$ and its component $CN$-net in which only components-transitions $C_t$ are allocated, respectively, over a finite alphabets $A$ and $B$ (let's recall that the functioning of nets $N$ and $CN$, when allocating only component-transitions, is described in terms of sequences of firing transitions). Then $A^*$ is a set of all words in the alphabet $A$, $B^* = (A \cup \{T_1^*, T_2^*, ...T_n^*\})^*$ — a set of all words in the alphabet $B = A \cup \{T_1^*, T_2^*, ...T_n^*\}$, where $T_k^*$ $(k = 1, 2, ...n)$ are the names of the various components-transitions $C_{t_k}$ $(k = 1, 2, ...n)$ in the $CN$-net.

Let some word $\tau \in A^*$ have a form $\tau = abt_1t_2cdt_3t_4h$, where the symbols $a, b, c, d, h$ denote the names of transitions of detailed model $N$, outside of any components-transitions $C_t$, and symbols $t_1, t_2, t_3, t_4$ are the names of transitions, which are elements of the components-transitions $C_t$. Making notations in the word $\tau$:

$$ab = \tau_1, t_1t_2 = \overline{\tau}_1, cd = \tau_2, t_3t_4 = \overline{\tau}_2, h = \tau_3,$$

we have a record of the original word $\tau$ as a concatenation of the words $\tau_1, \overline{\tau}_1, \tau_2, \overline{\tau}_2, \tau_3$, so that $\tau = \tau_1\overline{\tau}_1\tau_2\overline{\tau}_2\tau_3$ . In the transition from words of the language $L_t(N)$ to the words of the language $L_t(CN)$, the word $\tau$ is converted into word $\tau^{'}$ from B$^*$: $\tau^{'}$=$\tau_1 T_1^* \tau_2 T_2^* \tau_3$. At the same time, for the names $T_1^*$ and $T_2^*$ of the component net $CN$, their record is known as subwords of words of language $L_t(N)$. Taking into account [7], that the language, as described in terms of sequences of transition firing, of identical and single-type components is congruent, it is enough to remember the word(s) of one representative from identical and single-type components to substitute its record instead of the appropriate symbol for the component-transition in words of component $CN$ net language and get the words of original detailed Petri $N$ model. Languages $L_t(N)$ and $L_t(CN)$ are similar.

Let's consider free languages $L$ and $L^{'}$ of two Petri nets $N$ and $N^{'}$, over the same alphabet $W$. Let this alphabet represents the grouping of the alphabets $A$ and $B$, respectively, of the languages $L$ and $L^{'}$ under consideration. And also let there be mapping "onto" of one language onto another, for example, $L$ onto $L^{'}$. Let's mark this mapping by $S$. Then for each word $\tau^{'} \in L^{'}$ there should be a word $\tau \in L$ so that the equation $S(\tau) = \tau^{'}$ takes place. And because words of the language $L$ ($L^{'}$) are written as a sequence of characters of the corresponding alphabet $A$ ($B$), the mapping $S$ generates mapping $\phi$ that translates the characters of each word of the language $L$ into the characters of words of the language $L^{'}$. Given that the original mapping is "onto", mapping, then mapping $\phi$ is also "onto" mapping. Then when mapping $\phi$ the image of each character in the alphabet $A$ (letter of the alphabet $B$) has at least one prototype in the alphabet $A$. This means that some of the letters of the alphabet $B$ may be images of several letters of the alphabet $A$. Then, having the words of the language $L^{'} \subset B^*$ ($B^*$ is a set of all words in the alphabet $B$) and knowing the prototypes of the letters from $B$ being the letters of words of the language $L$, you can always restore the word from $A^*$ ($A^*$ is a set of all words in the alphabet $A$). And this means that the considered language $L$ of Petri net $N$. Theorem 1 holds:

**Theorem 1.** *Similarity of Petri nets languages is surjective mapping.*

Thus, the mapping $S$ of words in the language $L$ of Petri net $N$ on the set of words in language $L^{'}$ of Petri net $N^{'}$ is completely determined by the values on the letters of the alphabet $W$ so, that each character $a \in A$ is an image of at least one character $b \in B$, that is, at mapping $S$ for any $b \in B$ there is $a \in A$ so that $b = \phi(a)$. Then we can draw the following conclusions regarding mapping $S$:

1. $S(\tau\mu) = S(\tau)S(\mu)$ holds for all words $\tau$ and $\mu$ in concatenation of word $\tau\mu$ over $A \subset W$;
2. $S(e) = e$, where $e$ is an empty word;

3. $S(\tau) = \phi(a_1)\phi(a_2)...\phi(a_k)$ for words $\tau \in A^*$ of any length. Then for $L(N)$ of net $N$ language and $L^{'}(N^{'})$ of net $N^{'}$ language such equation is true: $L^{'}(N^{'}) = S(L(N)) = \{\tau^{'}/\tau^{'} = S(\tau),\ \ where\ \ \tau \in L(N)\}$.

Theorem 2 follows.

**Theorem 2.** *Similarity of Petri nets languages is a homomorphism.*

## 3. Similarity of languages of component Petri net, containing components-places among the allocated composite components

Consider the possibility of language recovery of detailed Petri net model by its component Petri net language, which contains components-places $C_p$ among its constituent components. This can be a component Petri net, containing only the components-places, or component Petri net, containing both types of components: components-transitions $C_t$ and components-places $C_p$. In this case, when among the constituent components of the net components-places are allocated, operation of the net has to be described in terms of the set of net reachable markings [8, 9].

**Statement 2.** *Languages $L_{p,t}(N)$ and $L_{p,t}(CN)$ are similar. Languages $L_p(N)$ and $L_p(CN)$ are similar.*

**Argument.** Consider only one variant of net. Let it be a net with both types of constituent components. Words of the language $L_{p,t}(N)$ of detailed Petri net model, with allocated constituent components $C_p$ and $C_t$, and the language $L_{p,t}(N)$ of its component Petri model, represent sequences obtained by writing out symbols of nodes along the paths in the graph of reachable markings of respectively nets $N$ and $CN$, starting at the initial marking and leading to each reachable net markup. Let $A$ be a finite alphabet for the detailed model $N$ language. It consists of a set of names, for example, $s$-dimensional vectors. Let $B$ be a finite alphabet of component net $CN$ with two types of composite components consisting of a set of names, for example, $r$-dimensional vectors. Then $r = s-k+l$, where $k$ is total number of places occurred in the allocated components, $l$ is a number of components-locations. Then $A^*$ is a set of all words in the alphabet $A$, $B^* = (\psi(A) \cup \{a_1^{'}, a_2^{'}, \ldots, a_n^{'}\})^*$ is a set of all words in the alphabet $B = \psi(A) \cup \{a_1^{'}, a_2^{'}, \ldots, a_n^{'}\}$, where $a_k^{'}(k = 1, 2, ..., n)$ are the names of the nodes of the graph of reachable markings of component Petri net $CN$, in which the nodes have moved or different parts of the graph of reachable markings of detailed Petri net $N$ encapsulated. Hereat $\psi$ — mapping that converts $s$-dimensional vectors of the graph of reachable markings of detailed Petri net in the $r$-dimensional vectors of the graph of reachable markings of component Petri net is surjective mapping [9].

Consider a word $\tau \in L_{p,t}(N)$. Let the word be of the form $\tau = a_1b_1b_2a_1a_3b_1'b_2'a_4$. Symbols $a_1, a_1, a_3, a_4$ mark the names of nodes of the graph of reachable markings of detailed net $N$, which are not nodes of any sections of the net, reflecting the operation of composite components. Symbols $b_1, b_2, b_1', b_2'$ are the names of nodes of the graph of reachable markings of detailed net $N$, which are nodes of such sections. At the transition from words of the language $L_{p,t}(N)$ to the words of the language $L_{p,t}(CN)$, the word $\tau$ is converted in the word $\tau'$ from $B^*$: $\tau' = \psi(a_1)a_1'\psi(a_2)\psi(a_3)a_2'\psi(a_4)$. In the word $\tau'$, symbol $\psi(a_i)$ $(i = 1,2,3,4)$ denotes the image of the corresponding node $a_i$of the graph of reachable markings of net $N$, which is not a node of any of the sections of the net, reflecting the operation of the composite component. This image is determined in a one-to-one manner. Symbol $a_j'$ $(j = 1,2,3)$ in the word $\tau'$ denotes the name of the node-encapsulent. Such a node is the image of all nodes from the sections of the graph of reachable markings of detailed Petri net $N$, which reflects the dynamics of the functioning of the composite components. For the names $a_j'$ of the language $L_{p,t}(CN)$ of the component net $CN$, their record as subwords of words of the language $L_{p,t}(N)$ is known by the construction of net component. Then, knowing all image prototypes of characters of any word from $L_{p,t}(N)$, the language $L_{p,t}(N)$ is easy to recover according to the words of language $L_{p,t}(CN)$ and get the language of original detailed Petri net model $N$, with allocated constituent components (components-places and components-transitions). Languages $L_{p,t}(N)$ and $L_{p,t}(CN)$ are similar. To establish the similarity of languages $L_p(CN)$ and $L_p(CN)$, the argument is similar.

## Conclusion

When modeling thoroughly functioning of parallel distributed systems, we have to deal with so-called problem of "state explosion", when the full system model becomes immensely large. This is the problem of building detailed models of real systems. Application of the component Petri nets for modeling of parallel distributed systems allows us to build smaller — reduced models. Study of languages of such networks allows us to investigate their behavioral properties. Proceeding with the problem of how "similar" languages of detailed model of the system under consideration and its component model are, we show that the language of the reduced model (component Petri nets) can restore the language of detailed Petri net model of the system in question. Languages of detailed and component models of parallel distributed systems are similar. The concept of language similarity of Petri nets, introduced in this work, allows to determine surjective homomorphism of the languages of such networks, and on this basis to carry out the qualitative analysis of the considered Petri nets languages.

## References

1. Kotov, V. 1984. *Petri nets*. Moscow: Nauka.
2. Peterson, J. L. 1977. Petri nets. *ACM Computing Surveys (CSUR)*, 9 (3), pp. 223–252.
3. Kryvy, S. and Matveeva, L. 2003. Formal Methods of Analysis of the Systems Properties. *Cybernetics and Systems Analysis*, 2, pp. 15–36.
4. Esparza, J. and Heljanko, K. 2008. *Unfoldings: A Partial-Order Approach to Model Checking. EATCS Monographs in Theoretical Computer Science*, ISBN: 978-3-540-77425-9, Springer-Verlag.
5. Lukyanova, E. 2012. On component modeling of systems with concurrency. *Naukovi Zapysky of NaUKMA. Computer Science*, t. 138, pp. 47–52.
6. Lukyanova, E. and Dereza, A. 2012. The study of the structural elements of the single-type *CN*-network during the component modeling and analysis of complex systems with concurrency. *Cybernetics and Systems Analysis*, 6, pp. 20–29.
7. Lukyanova, E. 2012. On the relationship between the language of CN-Model with component junctions, and the language of detailed Petri Model of parallel distributed system. *Visnyk of Kyiv Univ. im. Tarasa Shevchenka*, 4, pp. 145–150.
8. Lukyanova, E. 2013. Component modeling: on connections of detailed Petri model and component model of parallel distributed system. *ITHEA*, 2, pp. 15–22.
9. Lukyanova, E. 2013. On language of component Petri net with the components-places and components-transitions. *Bionics of intelligence*, 4.
10. Lukyanova, E. 2013. On Bisimulational Equivalence of Detailed Petri Net and Its CN-model of the Investigated Parallel Distributed System. *Visnyk of Kyiv Univ. im. Tarasa Shevchenka, special issue.*
11. Lukyanova, E. 2014. On homomorphism of component Petri net. *Cybernetics and Systems Analysis*, 1.
12. Lukyanova, E. 2012. The structural elements of a Petri net component. *Problemy programuvannya*, 2–3, pp. 25–32.
13. Jancar, P., Kucera, P. and Moller, F. 2000. USimulation and Bisimulation over One-Counter Processes. *Proc. of STACS'2000, 1770 of LNCS*, Springer, pp. 334–345.
14. Kotov, V. 1980. Algebra of Regular Networks. *Cybernetics*, 5, pp. 10–18.
15. Kotov, V. 1978. An algebra for parallelism based on Petri nets. *Computer Science*, 64, pp. 39–55.
16. Kapitonova, Yu. and Letichevsky, A. 1988. *Mathematical Theory of Computer Systems Designing.* Moscow: Nauka.

## ON REALIZING PRESCRIBED QUALITY OF A CONTROLLED SYSTEM'S PROCESS UNDER UNCERTAINTLY

© **Vyacheslav Maksimov**

Institute of Mathematics and Mechanics, Ural Federal University
Ural Branch of Russian Academy of Sciences
e-mail: *maksimov@imm.uran.ru*

***Abstract**. In this paper, we discuss a method of auxiliary controlled models and the application of this method to solving some problems of robust control for differential equations. As objects for the approbation of the method, a system of nonlinear differential equations describing some ecological and economic processes is used. A solving algorithm, which is stable with respect to informational noises and computational errors, is presented.*

### 1. Introduction. Statement of the problem

A dynamical model connecting main economic and climatic indices was suggested in [5]. This model is oriented to developing an economic strategy directed to deceleration of global warming. The main goal of the model analysis is to provide the means for tackling the following question: whether the reduction of emissions of greenhouse gases is justified from the economical viewpoint or not. The model takes into account global processes: it is assumed that the structure of economy is the same for all countries; the climate change is characterized by the average value of the temperature on Earth's surface and so on. This model contains three types of parameters.

1) Constant parameters (their list is presented in tables 2.3 and 2.4 on page 21, [5]).
2) Functions that are considered (for simplicity of the analysis) as exogenous with respect to the model and are a priori given.
3) Inner functions that are connected to one another and to exogenous parameters by means of some algebraic and differential equations. The list of these functions is presented in table 2.3. (see [5]), and the model equations are presented in table 2.2. Let us give the list of functions:
   - : $\mu(t)$ is a rate of emissions reduction with respect to uncontrollable emissions,
   - : $E(t)$ is an amount of emissions of greenhouse gases, below GHGs ($CO_2$ (carbonic acid gas) and chlorine-fluorine carbons only),
   - : $M_1(t) = (M(t) - 590)$ is an excess of the mass of GHGs in the atmosphere in comparison with the pre-industrial period,
   - : $T_0(t)$ is an average atmospheric temperature (on Earth's surface),

*Work was supported in part by the RFBR (project 12-01-00175a), by the Program of Basic Research of the Presidium of the Russian Academy of Sciences 38P (project 12-P-1-1038P), and by the Program for support of leading scientific schools of Russia (6512.2012.1).

- : $T_1(t)$ is an average deep-ocean temperature,
- : $I(t)$ is a gross investment,
- : $K(t)$ is a capital stock,
- : $F(t)$ is an atmospheric radiative forcing from GHGs,
- : $O(t)$ is a forcing of exogenous GHGs (i.e., of gases, which are considered as uncontrollable; there are all GHGs, besides $CO_2$ (carbonic acid gas) and chlorine-fluorine carbons),
- : $A(t)$ is a level of technology,
- : $\sigma(t)$ is the ratio of GHGs emissions to global output,
- : $L(t)$ is a population at time $t$, also equal to labor inputs,
- : $Q(t)$ is a gross world product.

Schematically, the connections between the inner functions can be pictured in the following way:

$$\begin{array}{ccccccc}
T_0^{**} & \longrightarrow & \Omega & & & & \\
\uparrow & & \uparrow & \searrow & & & \\
F & & \mu & & Q & & \\
\uparrow & & \downarrow & \swarrow & \uparrow & \nwarrow & \\
M_1^* & \longleftarrow & E & & K^* & \longleftarrow & I
\end{array}$$

Here the functions marked by asterisk are solutions of linear differential equations of the first order, the function $T_0(t)$ is a solution of a linear differential equation of the second order.

If we pass from the discrete model suggested by the authors to the "continuous" one, then the equations of the model $\Sigma$ take the form:

$$\begin{aligned}
\dot{T}_0(t) &= c_1T_0(t) + c_2T_1(t) + c_3F(t), \quad t \in [0, \vartheta] \\
\dot{T}_1(t) &= c_4(T_0(t) - T_1(t)) \\
\dot{M}_1(t) &= \beta E(t) - \delta_M M_1(t) \\
\dot{K}(t) &= -\delta_K K(t) + I(t),
\end{aligned} \tag{1}$$

where $t$ is time, $\vartheta$ is a terminal time moment,

$$F(t) = 4,1 \cdot \log_2 \left(1 + \frac{M_1(t)}{590}\right) + O(t),$$

$$E(t) = (1 - \mu(t))\sigma(t)Q(t),$$

$$Q(t) = (1 - b_1\mu(t)^{b_2})/(1 + \theta_1 T_0(t)^{\theta_2})A(t)K(t)^{\gamma}L(t)^{1-\gamma}.$$

An initial state of $\Sigma$, $x(0) = \{T_0(0), T_1(0), M_1(0), K(0))\}$, is assumed to be known and a priori given. It is natural to set $T_0(0) > 0$, $T_1(0) > 0$, and $K(0) > 0$. Functions $\mu(\cdot)$

and $I(\cdot)$ are considered as control parameters determining a strategy of global control of climate and economy. The numerical analysis of the model is performed in [5]. At that, the direct problem is solved, namely, possible strategies (rules of forming $\mu(\cdot)$ and $I(\cdot)$) are specified, and system's dynamics is computed. The comparative analysis of simulation results for different structures is performed. In addition, the analysis of sensitivity of results with respect to some model parameters is fulfilled.

In what follows, values $\mu$ and $I$, according to [5], are treated as controls and are denoted by the symbol $u$, i.e., $u = \{\mu, I\}$. We transform system (1) to the form (neglecting small values ($b_1 = 0,0686$, $\vartheta_1 = 0,00144$))

$$\begin{aligned} \dot{T}_0(t) &= c_1 T_0(t) + c_2 T_1(t) + 4,1 c_3 \cdot \log_2\left(1 + \frac{M_1(t)}{590}\right) + c_3 O(t), \quad t \in [0, \vartheta] \\ \dot{T}_1(t) &= c_4(T_0(t) - T_1(t)) \\ \dot{M}_1(t) &= E_1(t)(1 - \mu(t)) - \delta_M M_1(t) \\ \dot{K}(t) &= -\delta_K K(t) + I(t), \end{aligned} \tag{2}$$

where

$$E_1(t) = E_1(t, K) = \beta\sigma(t)A(t)K(t)^\gamma L(t)^{1-\gamma}.$$

Hereinafter, we consider the system $\Sigma$ of form (2). The symbol $x(\cdot) = x(\cdot; x(0), u(\cdot))$ stands for the solution of system (2) with an initial state $x(0)$ and a control $u(\cdot) = \{\mu(\cdot), I(\cdot)\}$.

Our aim differs from the aim of [5]. We consider an "inverse" problem; its essence consists in the following. Some system's dynamics, i.e., a function $x_*(\cdot) = \{T_{0*}(\cdot), T_{1*}(\cdot), K_*(\cdot), M_{1*}(\cdot)\}$ generated by some unknown controls $\mu = \mu_*(\cdot)$ and $I = I_*(\cdot)$ is given. These controls may be program or feedback controls; the latter is formed, for example, by the rule $\mu_*(t) = \mu(t, x_*(t))$, $I_*(t) = I(t, x_*(t))$. Thus, the functions $x_*(\cdot) = \{T_{0*}(\cdot), T_{1*}(\cdot), K_*(\cdot), M_{1*}(\cdot)\}$ satisfy the system of equations

$$\begin{aligned} \dot{T}_{0*}(t) &= c_1 T_{0*}(t) + c_2 T_{1*}(t) + 4,1 c_3 \cdot \log_2\left(1 + \frac{M_{1*}(t)}{590}\right) + c_3 O(t), \quad t \in [0, \vartheta] \\ \dot{T}_{1*}(t) &= c_4(T_{0*}(t) - T_{1*}(t)) \\ \dot{M}_{1*}(t) &= E_{1*}(t, K_*)(1 - \mu_*(t)) - \delta_M M_{1*}(t) \\ \dot{K}_*(t) &= -\delta_K K_*(t) + I_*(t), \end{aligned} \tag{3}$$

where, emphasize once again, the functions $\mu_*(\cdot)$ and $I_*(\cdot)$ are unknown. It is known only that they are subject to restrictions of the form

$$I_*(t) \in [I_-, I_+], \quad \mu_*(t) \in [f_-, f_+] \quad \text{при } t \in [0, \vartheta]. \tag{4}$$

Here

$$-\infty < f_- < f_+ < +\infty, \quad 0 \le I_- < I_+ < +\infty.$$

The initial state of system (3), $x_*(0) = \{T_{0*}(0), T_{1*}(0), M_{1*}(0), K_*(0)\}$, is assumed to be $x(0)$.

The problem under consideration may be formulated in the following way. At frequent enough time moments

$$\tau_i \in \Delta = \{\tau_i\}_{i=0}^m, \quad \tau_{i+1} = \tau_i + \delta, \quad \tau_0 = 0, \quad \tau_m = \vartheta,$$

values of $T_0(\tau_i)$, $T_1(\tau_i)$, and $K(\tau_i)$ are inaccurately measured. Results of measurements (vectors $\{\xi_{1i}^h, \xi_{2i}^h, \xi_{3i}^h\} \in R^3$) satisfy the inequalities

$$|T_0(\tau_i) - \xi_{1i}^h|^2 + |T_1(\tau_i) - \xi_{2i}^h|^2 + |K(\tau_i) - \xi_{3i}^h|^2 \le h^2, \tag{5}$$

where $h \in (0,1)$ is a level of informational noise. Here and below, the symbol $|\cdot|$ stands for the absolute value of a number, whereas the symbol $\|\cdot\|$, the Euclidean norm of a vector. Denote by $\Xi(x(\cdot), h)$ the set of admissible measurements, i.e., the set of all piece-wise constant functions $\xi^h(\cdot) \to R^3$, $\xi^h(t) = \xi_i^h$ for $t \in [\tau_i, \tau_{i+1})$, $\tau_i = \tau_{i,h}$, satisfying inequalities (5). Here

$$\xi_i^h = \{\xi_{1i}^h, \xi_{2i}^h, \xi_{3i}^h\}, \quad \xi^h(t) = \{\xi_{1i}^h, \xi_{2i}^h, \xi_{3i}^h\} \quad \text{for a.a.} \quad t \in [\tau_{i,h}, \tau_{i+1,h}).$$

The control problem under discussion in the paper is as follows. A number $\varepsilon > 0$ is given. It is required to construct an algorithm for forming a feedback control

$$u = u^h(t) = u(t; x^h(\cdot), x_*(\cdot), \xi^h(\cdot))$$

of system (2) providing fulfilment of the following condition. Whatever unknown possible Lebesgue measurable functions $\mu_*(\cdot)$ and $I_*(\cdot)$ with properties (4) may be, the distance between $x^h(t)$ and $x_*(t)$ at all moments $t \in [0, \vartheta]$ should not exceed the value of $\varepsilon$ provided the values of $h$ and $\delta$ are sufficiently small.

Here

$$x^h(\cdot) = x(\cdot; u^h(\cdot)) = \{T_0^h(\cdot), T_1^h(\cdot), M_1^h(\cdot), K^h(\cdot)\}$$

is the trajectory of $\Sigma$ generated by the control

$$u(t) = u^h(t; x^h, x_*, \xi^h) = \{\mu^h(t) = \mu(t; x^h(\cdot), x_*(\cdot), \xi^h(\cdot)),\ I^h(t) = I(t; x^h(\cdot), x_*(\cdot), \xi^h(\cdot))\} \in$$
$$\in U(t, x^h(\cdot), x_*(\cdot), \xi^h(\cdot)) \subset [I_-, I_+] \times [f_-, f_+],$$

which is formed according to the feedback principle. Thus, $x^h(\cdot)$ is the solution of system (2) with the feedback controls $\mu(\cdot) = \mu^h(\cdot)$ and $I(\cdot) = I^h(\cdot)$.

Hereinafter, the symbol $\mathscr{U}$ stands for the set of admissible controls, i.e., the set of Lebesgue measurable functions $u(\cdot) = \{\mu(\cdot), I(\cdot)\}$ such that $\mu(t) \in [f_-, f_+]$, $I(t) \in [I_-, I_+]$ for a.a. $t \in [0, \vartheta]$.

One of the approaches to solving the problems of guaranteed control (they are also called positional differential games) for dynamical systems described by ordinary differential equations was developed in [2, 6, 7]. In all the works cited above, the cases when the full phase state of a system is inaccurately measured at frequent enough time moments are considered. In the present work, from the position of the approach described in [2, 6, 7], the problems of guaranteed control under the measurement of a "part" of system's phase state (a "part" of coordinates) are investigated.

To form a control $u$ providing the solution of the problem, along with the information on the "part" of coordinates of the solution of the system $\Sigma$ (namely, on the values $\xi_i^h$ satisfying inequalities (5)), it is necessary to obtain some additional information on the coordinate $M_1(\cdot)$, which is missing. To get such a piece of information during the control process, it is reasonable, following the approach developed in [2, 6, 7], to introduce an auxiliary controlled system $M$. This system is described by a differential equation (the form is specified below). The equation has an output $w^h(t)$ and an input $v^h(t)$. The input $v^h(\cdot)$ is some new auxiliary control; it should be formed by the feedback principle in such a way that $v^h(\cdot)$ "approximates" the unknown coordinate $M_1(\cdot)$ in the mean uniform metric. Thus, along with the block of forming the control in the real system (it is called a controller), we need to incorporate into the control contour one more block (it is called a identifier) allowing to reconstruct the missing coordinate $M_1(\cdot)$ in the real time mode.

The scheme of algorithms for solving the problem is given in Figure 1.

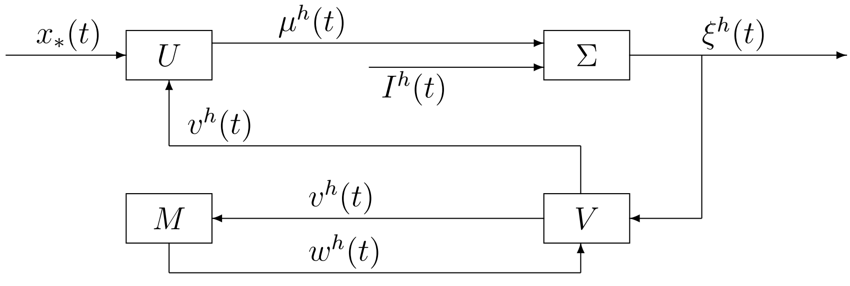



Figure 1.

In the beginning, an auxiliary dynamical system $M$ (a model) is introduced. This model functioning on the time interval $[0, \vartheta]$ has an input $v^h(t)$ and an output $w^h(t)$. The model $M$ with its control law $V$ forms the identifier. Before the algorithm starts, the value $h$ and the partition $\Delta$ with the step $\delta$, as well as the model $M$, are fixed. The process of synchronous feedback control of the systems $\Sigma$ and $M$ is organized on the interval $[0,\vartheta]$. This process is decomposed into $(m-1)$ identical steps. At the $i$th step carried out during the time interval $\delta_i = [\tau_i, \tau_{i+1})$, the following actions are fulfilled. First, at the time moment $\tau_i$ according to the chosen rules $U$ and $V$ the functions

$$v^h(t) = v_i^h \in V(\tau_i, \xi_i^h, w^h(\tau_i)), \quad t \in \delta_i, \tag{6}$$

$$u^h(t) = u_i^h \in U(\tau_i, v_i^h, \xi_i^h, x_*(\tau_i)), \tag{7}$$

are calculated by measurements $\xi_i^h$ and $w^h(\tau_i)$. Then (till the moment $\tau_{i+1}$) the control $u = u^h(t)$, $\tau_i \le t < \tau_{i+1}$, is fed onto the input of the system $\Sigma$ and the control $v = v^h(t)$, $\tau_i \le t < \tau_{i+1}$, onto the input of the model $M$. The values $\xi_{i+1}^h$ and $w^h(\tau_{i+1})$ are the results of the work of the algorithm at the $i$th step. The procedure stops at the moment $\vartheta$.

Thus, all complexity of solving these problems is reduced to an appropriate choice of a model $M$ and functions $U$ and $V$.

So, the problem may be formulated as follows. In the sequel, a family of partitions

$$\Delta_h = \{\tau_{i,h}\}_{h=0}^{m_h}, \quad \tau_{i+1,h} = \tau_{i,h} + \delta(h), \quad \tau_{0,h} = 0, \quad \tau_{m_h,h} = \vartheta$$

of the interval $[0,\vartheta]$ is assumed to be fixed.

*Problem of robust control.* It is required to specify differential equations of the model $M$

$$\dot{w}^h(t) = f_1(\xi_i^h, w^h(\tau_i), v_i^h), \quad t \in \delta_{h,i} = [\tau_{i,h}, \tau_{i+1,h}), \quad \tau_i = \tau_{i,h}, \tag{8}$$

$$w^h(0) = w_0^h, \quad w^h(t) \in R,$$

and the rule of choosing controls $v_i^h$ and $u_i^h$ at the moments $\tau_i$ being a mapping of form (6), (7) such that the inequality

$$\max_{t\in[0,\vartheta]} \|x^h(t) - x_*(t)\| \le \varepsilon \tag{9}$$

holds for $h \in (0, h_*(\varepsilon))$ and $\delta = \delta(h) \in (0, \delta(h_*(\varepsilon))$. Let the symbol $X(\cdot)$ denote the bundle of solutions of system (2), i.e.,

$$X(\cdot) = \{x(\cdot) : x(\cdot) = x(\cdot; x(0), u(\cdot)) = \{T_0(\cdot), T_1(\cdot), M_1(\cdot), K(\cdot)\}, u(\cdot) \in \mathscr{U}\}.$$

We assume that the following condition is fulfilled:

**Condition 1.**

$$d_* = \inf\Big\{\min_{t\in[0,\vartheta]} \Big(1 + \frac{M_1(t)}{590}\Big) : x(\cdot) = \{T_0(\cdot), T_1(\cdot), M_1(\cdot), K(\cdot)\} \in X(\cdot)\Big\} > 1.$$

In addition, the functions $\sigma(t)$, $A(t)$, $L(t)$, and $Q(t)$ are considered as known and continuous.

## 2. Algorithm for reconstructing $M_1(\cdot)$

First, we specify the algorithm for reconstructing $M_1(\cdot)$, which will be applied for solving the problem in question. Namely, we describe the identifier (see Fig. 1). To substantiate this algorithm, we use ideas from [6, 7, 1, 3].

Introduce the notation

$$T(t) = \{T_0(t), T_1(t)\}, \quad f(t, T(t)) = c_1T_0(t)+c_2T_1(t)+c_3Q(t), \quad \tilde{u}(t) = \log_2\left(1 + \frac{M_1(t)}{590}\right).$$

Here $x(\cdot) = \{T_0(\cdot), T_1(\cdot), M_1(\cdot), K(\cdot)\}$ is an arbitrary element of the set $X(\cdot)$. In this case, the first equation of system (2) is rewritten in the form

$$\dot{T}_0(t) = f(t, T(t)) + 4,1c_3\tilde{u}(t).$$

Note that one can specify a number $M_* > 0$ such that the following inequalities are valid:

$$\|\dot{T}(t)\| \le M_* \quad \text{for almost all} \quad t \in [0, \vartheta], \tag{10}$$

$$\|f(t, T(t)) - f(\tau_i, \xi_i^h)\| \le M_*(\delta + h + \omega(\delta)) \text{ for } t \in \delta_i = [\tau_i, \tau_{i+1}). \tag{11}$$

Here $\tau_i = \tau_{i,h}$, $\omega(\delta)$ is the continuity modulo of the function $t \to O(t)$, $t \in [0, \vartheta]$, i. e.,

$$\omega(\delta) = \sup\{|O(t) - O(t-\delta)| : t \in [\delta, \vartheta]\}.$$

Inequality (11) is a consequence of (5) and (10).

We fix a family $\Delta_h$ of partitions of the interval $[0, \vartheta]$ and some auxiliary function $\alpha(h) : (0, 1) \to (0, 1)$.

As the model $M$, we take a linear system described by a scalar differential equation of the form

$$\dot{w}^h(t) = f(\tau_i, \xi_i^h) + 4,1c_3v^h(t) \quad \text{for a.a.} \quad t \in \delta_i = [\tau_i, \tau_{i+1}), \tag{12}$$

$i \in [0 : m-1]$, $\quad \tau_i = \tau_{i,h}$, $\quad m = m_h$, with the initial condition

$$w^h(0) = T_0(0).$$

Let

$$v^h(t) = v_i^h \in V(\tau_i, \xi_i^h, w^h(\tau_i)) = -\frac{1}{\alpha}4,1c_3[w^h(\tau_i) - \xi_{1i}^h] \quad \text{for} \quad t \in \delta_i. \tag{13}$$

The control $v^h(t)$ in equation (12) is found from (13). Thus, the model control is specified by the feedback principle (see (6)). Consequently, equation (12) takes the form

$$\dot{w}^h(t) = f(\tau_i, \xi_i^h) - \frac{1}{\alpha}(4,1c_3)^2[w^h(\tau_i) - \xi_{1i}^h] \quad \text{for a.a.} \quad t \in \delta_i. \tag{14}$$

Let us describe the algorithm for reconstructing the unmeasured coordinate $M_1(\cdot)$ in the real time mode. Before the algorithm starts, we fix a value $h \in (0,1)$ and a partition $\Delta_h$. The work of the algorithm is decomposed into $m-1$ identical steps. At the $i$th step carried out during the time interval $\delta_i = [\tau_i, \tau_{i+1})$, $\tau_i = \tau_{i,h}$, the following actions are fulfilled. First, at the moment $\tau_i$, the control $v^h(t)$ is calculated by (13). This control is fed onto the input of model(12) on the interval $[\tau_i, \tau_{i+1})$. Under the action of this control, the model passes form the state $w^h(\tau_i)$ to the state $w^h(\tau_{i+1}) = w^h(\tau_{i+1}; \tau_i, w^h(\tau_i), v_i^h)$. The work of the algorithm stops at the moment $\vartheta$.

**Lemma 1.** *Let the conditions*

$$\alpha(h) \to 0,\ \delta(h) \to 0,\ \delta(h)\alpha^{-1}(h) \to 0,\ h\alpha^{-1}(h) \to 0 \text{ as } h \to 0 \tag{15}$$

*be fulfilled. Then, uniformly in all* $x(\cdot) \in X(\cdot)$, $h \in (0,1)$, $\xi^h(\cdot) \in \Xi(x(\cdot), h)$, $i \in [0 : m_h - 1]$, *the inequalities*

$$\int_{\tau_i}^{\tau_{i+1}} |\dot{w}^h(s)|\, ds \le C\delta, \tag{16}$$

*are valid. Here* $C = \text{const} > 0$, $\delta = \delta(h)$, *and* $\tau_i = \tau_{i,h}$.

**Lemma 2.** *Let conditions (15) be fulfilled. Let* $\delta^\gamma(h)\alpha^{-1}(h) \to +\infty$ *(for some* $\gamma \in (0,1)$*) as* $h \to 0$ *and*

$$u_e^h(t) = \begin{cases} \tilde{u}(0), & t \in [0, \delta^\gamma) \\ v^h(t), & t \in [\delta^\gamma, \vartheta]. \end{cases}$$

*Then, the inequality*

$$\sup_{t \in [0,\vartheta]} |u_e^h(t) - \tilde{u}(t)| \le$$

$$\le d_1^0\alpha(h) + d_2^0(h + \delta(h))\alpha^{-1}(h) + d_3^0\omega(\delta(h)) + d_4^0\alpha(h)\delta^{-\gamma}(h) + d_5^0\delta^\gamma(h)$$

*is valid. Here the constants* $d_j^0$, $j \in [1:5]$, *do not depend on* $h \in (0,1)$.

Introduce the notation

$$u_*^h(t) = 590(2^{u_e^h(t)} - 1).$$

The following Theorem is true.

**Theorem 1.** *Under the conditions of Lemma 2, the inequality*

$$\sup_{t \in [0,\vartheta]} |u_*^h(t) - M_1(t)| \le \nu(h, \delta(h), \alpha(h)) =$$

$$= d_1\alpha(h) + d_2(h + \delta(h))\alpha^{-1}(h) + d_3\omega(\delta(h)) + d_4\alpha(h)\delta^{-\gamma}(h) + d_5\delta^\gamma(h)$$

*holds. Here the constants* $d_j$, $j \in [1:5]$, *do not depend on* $h \in (0,1)$.

The Theorem follows from Lemma 2 and the inequality

$$|u_*^h(t) - M_1(t)| \leq 590|2^{u_e^h(t)} - 2^{\tilde{u}(t)}|.$$

## 3. Algorithm for solving control problem

Let us turn to the description of the algorithm for solving the control problem in question. From the above, it is necessary to specify model (8) and strategies $U$ and $V$ (6), (7) providing inequality (9).

We fix a family $\Delta_h$ of partitions of the interval $[0, \vartheta]$ and some function $\alpha(h) : (0, 1) \to (0, 1)$. Let the family $\Delta_h$ and function $\alpha(h)$ be such that the following condition holds:

**Condition 2.** *The convergences*

$$\alpha(h) \to 0,\ \delta(h) \to 0,\ \delta(h)\alpha^{-1}(h) \to 0,\ h\alpha^{-1}(h) \to 0, \quad \alpha^{-1}(h)\delta^{\gamma}(h) \to +\infty \quad \textit{as} \quad h \to 0$$

*take place for some* $\gamma \in (0, 1)$.

Let model (8) be of form (12), i.e.,

$$\dot{w}^h(t) = f(\tau_i, \xi_i^h) + 4,1c_3 v^h(t) \quad \text{for a.a.} \quad t \in \delta_i = [\tau_i, \tau_{i+1}), \tag{17}$$

$i \in [0 : m-1], \quad \tau_i = \tau_{i,h}, \quad m = m_h$, with the initial condition

$$w^h(0) = T_0(0).$$

Let rules $U$ (6) and $V$ (7) for forming the controls $u_i^h$ and $v_i^h$ be as follows:

$$v_i^h = V(\tau_i, \xi_i^h, w^h(\tau_i)) = -\frac{1}{\alpha}4,1c_3[w^h(\tau_i) - \xi_{1i}^h], \tag{18}$$

$$u_i^h = \{\mu^h(\tau_i), I^h(\tau_i)\} = U(\tau_i, v_i^h, \xi_i^h, x_*(\tau_i)) \quad \text{for} \quad t \in \delta_i.$$

Here

$$I^h(\tau_i) = \arg\min\{(\xi_{3i}^h - K_*(\tau_i))I : I \in [I_-, I_+]\}, \tag{19}$$

$$\mu^h(\tau_i) = \arg\min\{E_1(\tau_i, K_*)(u_*^h(\tau_i) - M_{1*}(\tau_i))\mu : \mu \in [f_-, f_+]\}, \tag{20}$$

$$u_*^h(\tau_i) = \begin{cases} \log_2\Big(1 + \frac{M_{*1}(t)}{590}\Big), & \text{if } \tau_i \leq \delta^{\gamma}(h) \\ v_i^h, & \text{otherwise.} \end{cases}$$

In what follows, we need

**Lemma 3.** [4] *Let the function $\varepsilon(t)$ be nonpositive for $t \in T$ and, for all $i \in [0 : m-1]$, satisfy the inequalities*

$$\varepsilon(\tau_{i+1}) \le \varepsilon(\tau_i)(1+\beta\delta) + \int\limits_{\tau_i}^{\tau_{i+1}} |\varphi(t)|\, dt,$$

*where $\tau_i \in \Delta$, $\beta = \mathrm{const} > 0$, and $\varphi(\cdot) \in L(T; R)$. Then,*

$$\varepsilon(\tau_i) \le \Big(\varepsilon(t_0) + \int\limits_{t_0}^{\tau_i} |\varphi(t)|\, dt\Big) \exp(\beta(\tau_i - t_0)).$$

Introduce

**Condition 3.** *The inequalities*

$$0 < C^{(1)} < K_*(t) < C^{(2)} < +\infty \quad \textit{for } t \in [0, \vartheta]$$

*are valid.*

**Theorem 2.** *For any $\varepsilon > 0$, one can specify $h_*(\varepsilon) \in (0,1)$ such that, for all $h \in (0, h_*(\varepsilon))$ and $\delta(h) \in (0, \delta(h_*(\varepsilon)))$, inequality (9) holds, if the model $M$ is given by equation (17), the strategies $V$ and $U$ are taken in form* (6), (7), (18)*–(20).*

The proof of the theorem is performed by the scheme of the proof of corresponding statements from [2] and is based on Theorem 3 and Lemma 4. In the process, the variation of the values

$$\varepsilon_1(t) = |K_*(t) - K^h(t)|^2, \quad t \in [0, \vartheta],$$
$$\varepsilon_2(t) = |M_{*1}(t) - M_1^h(t)|^2, \quad t \in [0, \vartheta]$$

is estimated and the inequalities

$$\varepsilon_1(t) \le C_*(\delta + h),$$

$$\varepsilon_2(t) \le C_{**}(h^{1/2} + \delta^{1/2} + \nu(h, \delta(h), \alpha(h))), \quad t \in [0, \vartheta]$$

are established. Here the function $\nu(h, \delta, \alpha)$ is defined in Theorem 3.

In the conclusion, we describe the algorithm of the problem under consideration. Thus, we have system (2) with the control $u = \{\mu, I\}$ and system (3) with the unknown control $u_* = \{\mu_*, I_*\}$. We choose a family $\Delta_h = \{\tau_{i,h}\}_{i=0}^{m_h}$ of partitions of the interval $[0, \vartheta]$ with a step $\delta(h) = \tau_{i+1,h} - \tau_{i,h}$ and a function $\alpha(h) : (0,1) \to (0,1)$ depending on the parameter $h$. The family $\Delta_h$ and function $\alpha(h)$ satisfy Condition 2. Before the algorithm starts, we fix some value of measurement accuracy $h$, the partition $\Delta = \Delta_h$ and number $\alpha = \alpha(h)$. The work of the algorithm is decomposed into $m-1$, $m = m_h$, identical steps.

At the $i$th step carried out during the time interval $\delta_i = [\tau_i, \tau_{i+1})$, $\tau_i = \tau_{i,h}$, the following actions are fulfilled. First, at the moment $\tau_i$, using the state $w^h(\tau_i)$ of model (17), the result $\xi_i^h$ (satisfying inequality (5)) of calculating the state of system (2), we calculate three numbers, namely, $v_i^h$ and $u_i^h = \{\mu^h(\tau_i), I^h(\tau_i)\}$, by formulas (18)–(20). Then, during the time interval $\delta_i$, the constant control $u^h(t) = u_i^h$ is fed onto the input of model (17). After these operations, at the moment $\tau_{i+1}$ the model state is recalculated (instead of the number $w^h(\tau_i)$, the number $w^h(\tau_{i+1}) = w^h(\tau_{i+1}; w^h(\tau_i), v_i^h)$ is found; in addition, the vector $\xi_{i+1}^h$ is determined). The analogous actions are performed till the moment $\tau_{m_h-1,h}$.

As follows from Theorem 1, if the fixed measurement accuracy $h$ is sufficiently small, then the described above algorithm for forming the control $u(\cdot)$ in system (2) provides "tracking" (in uniform metric) the solution $x_*(\cdot)$ of system (3) by the solution $x^h(\cdot)$ of system (2). Thus, the algorithm solves the problem of robust control.

## References

1. Blizorukova, M. S., Maksimov, V. I. and Pandolfi, L. 2002. Dynamic input reconstruction for a nonlinear time-delay system. *Automation and remote control.* 63 (2), pp. 171–180.
2. Krasovskii, N. N. and Subbotin, A. I. 1988. *Game-Theoretical Control Problems.* Springer Verlag, New York–Berlin.
3. Maksimov, V. I. 2013. An algorithm for reconstructing controls in a uniform metric. *Journal of applied mathematics and mechanics.* 77 (1), pp. 212–219.
4. Maksimov, V. I. 2011. The tracking of the trajectory of a dynamical system. *Journal of applied mathematics and mechanics.* 75, pp. 667–674.
5. Nordhaus, W. D. 1994. *Managing the global commons. The economics of climate change.* The MIT Press.
6. Osipov, Yu. S. and Kryazhimskii, A. V. 1995. *Inverse problems of ordinary differential equations: dynamical solutions.* Gordon and Breach.
7. Osipov, Yu. S. and Kryazhimskii, A. V. and Maksimov, V. I. 2011. *Methods of Dinamical Reconstruction of Inputs of Controlled Systems.* Ekaterinburg. ISBN 978-5-7691-2219-4.

# GUARANTEED ESTIMATES OF LINEAR FUNCTIONALS ON VELOCITY OF A VISCOUS INCOMPRESSIBLE FLUID UNDER UNCERTAINTIES

**© Olexander Nakonechny, Yuri Podlipenko**

Taras Shevchenko National University of Kyiv
The Faculty of Cybernetics
e-mail: *yourip@mail.ru*

***Abstract***. *The creation and justification of the methods for guaranteed estimation of linear functionals from solutions to the boundary value problems for linearized stationary Navier-Stokes equations in bounded open Lipschitzian domains are considered.*

## Introduction

Problems of optimal reconstruction of solutions of linearized stationary Navier-Stokes equations under incomplete data are investigated. These problems play an important role in mathematical physics. Depending on a character of an apriori information, stochastic or deterministic approach are possible. The choice is determined by nature of the parameters in the problem, which can be random or not. Moreover the optimality of estimations depends on a criterion with respect to which a given value is evaluated.

We assume that right-hand sides of linearized Navier-Stokes equations are unknown and belong to the given bounded subsets of the space of all square integrable functions in the considered domain and for solving the estimation problems we must have supplementary data (observations) depending on solutions of these equations. We suppose that observation errors (noises) are realizations of the stochastic fields, with unknown moment functions of the second order also belonging to certain given subsets.

Our approach is as follows. We are looking for linear with respect to observations optimal estimates of solutions of linearized Navier-Stokes equations from the condition of minimum of maximal mean square error of estimation taken over the above subsets.

We consider constructive methods for obtaining such estimates, which is expressed in terms of solutions of special variational equations.

Guaranteed estimation problems for some other types of ordinary and partial differential equations are investigated in [1]–[5].

## 1. Preliminaries and auxiliary results

If $X$ is a Hilbert space over $\mathbb{R}$ with inner product $(\cdot,\cdot)_X$ and norm $\|\cdot\|_X$, then by $J_X \in \mathscr{L}(X,X')$ we will denote an operator, called a canonical isomorphism from $X$ onto dual space $X'$, and defined by the equality $(v,u)_X = < v, J_X u >_{X\times X'}$ $\forall u,v \in X$, where $< x, f >_{X \times X'} := f(x)$ for $x \in X$, $f \in X'$, and $\mathscr{L}(X, Y)$ is the set of bounded linear operators mapping $X$ into a Hilbert space $Y$.

Further we use the following notations: $x = (x_1, \ldots, x_n)$ denotes a spatial variable that is varied in a bounded open Lipschitzian domain $D \subset \mathbb{R}^n$, with boundary $\Gamma$;

$dx = dx_1 \cdots dx_n$ is a Lebesgue measure in $\mathbb{R}^n$;

$\mathscr{D}(D)$ is the space of infinitely differentiable functions with compact support contained in $D$.

A continuous linear form on $\mathscr{D}(D)$ is called a distribution on $D$. We denote by $\mathscr{D}'(D)$ the set of distributions on $D$. If $T \in \mathscr{D}'(D)$ we denote by $< T, \phi >$ its value on the function $\phi \in \mathscr{D}(D)$.

If $T \in \mathscr{D}'(D)$ the derivative $D_i T = \frac{\partial T}{\partial x_i}$ which coincides with the usual differentiation of continuously differentiable functions, is defined by $< \frac{\partial T}{\partial x_i}, \phi >= - < T, \frac{\partial \phi}{\partial x_i} >$.

We denote by $L^2(D)$ the space of the real functions defined on $D$ with the second power absolutely integrable for the Lebesgue measure $dx$. This is a Hilbert space with the norm

$$\|u\|_{L^2(D)} = \left( \int_D |u(x)|^2 \, dx \right)^{1/2}.$$

and inner product

$$(u, v)_{L^2(D)} = \int_D u(x) v(x) \, dx.$$

The Sobolev space $H^1(D)$ is the space of functions in $L^2(D)$ with derivatives of order 1 also belonging to $L^2(D)$. This is a Hilbert space with the norm

$$\|u\|_{H^1(D)} = \left( \|u\|^2_{L^2(D)} + \sum_{j=1}^{n} \|D_j u\|^2_{L^2(D)} \right)^{1/2}$$

and inner product

$$(u, v)_{H^1(D)} = (u, v)_{L^2(D)} + \sum_{j=1}^{n} (D_j u, D_j v)_{L^2(D)}.$$

The closure of $\mathscr{D}(D)$ in $H^1(D)$ is denoted by $H^1_0(D)$.

We will also use the notation $\mathbf{L}^2(D) = \{L^2(D)\}^n$, $\mathbf{H}^1(D) = \{H^1(D)\}^n$, $\mathbf{H}^1_0(D) = \{H^1_0(D)\}^n$, $\boldsymbol{\mathscr{D}}(D) = \{\mathscr{D}(D)\}^n$, $\boldsymbol{\mathscr{D}'}(D) = \{\mathscr{D}'(D)\}^n$ for the product spaces consisting of vector functions $\mathbf{u} = (u_1, \ldots, u_n)$ whose componets belong to one of the spaces $L^2(D)$, $H^1(D)$, $H^1_0(D)$, $\mathscr{D}(D)$, $\mathscr{D}'(D)$ respectively, and we suppose that these product spaces are equipped with the usual product norm and inner product (except $\mathscr{D}(D)^n$ and $\mathscr{D}'(D)^n$ which are not normed spaces). For example, if

$\mathbf{u} = (u_1, \dots, u_n), \mathbf{v} = (v_1, \dots, v_n) \in \mathbf{L}^2(D)$ then

$$(\mathbf{u}, \mathbf{v})_{\mathbf{L}^2(D)} = \sum_{i=1}^{n} (u_i, v_i)_{L^2(D)}, \quad \|\mathbf{u}\|_{\mathbf{L}^2(D)} = (\mathbf{u}, \mathbf{u})^{1/2}_{\mathbf{L}^2(D)} = \Big\{ \sum_{i=1}^{n} \|u_i\|^2_{L^2(D)} \Big\}^{1/2}.$$

For every $v \in \mathscr{D}'(D)$ we put

$$\mathbf{grad}\, v := \left( \frac{\partial v}{\partial x_1}, \dots, \frac{\partial v}{\partial x_n} \right),$$

which defines the linear differential operator denoted by **grad** from $\mathscr{D}'(D)$ to $\boldsymbol{\mathscr{D}}'(D)$.

We define the linear differential operator denoted by div from $\boldsymbol{\mathscr{D}}'(D)$ to $\mathscr{D}'(D)$ by

$$\operatorname{div} \mathbf{v} := \sum_{i=1}^{n} \frac{\partial v_i}{\partial x_i} \quad \forall \mathbf{v} = (v_1, \dots, v_n) \in \boldsymbol{\mathscr{D}}'(D)$$

and the Laplace operator $\Delta$ from $\boldsymbol{\mathscr{D}}'(D) \to \boldsymbol{\mathscr{D}}'(D)$ by

$$\Delta \mathbf{v} = \left( \sum_{i=1}^{n} \frac{\partial^2 v_1}{\partial x_i^2}, \dots, \sum_{i=1}^{n} \frac{\partial^2 v_n}{\partial x_i^2} \right).$$

Let $\mathscr{V} = \{\mathbf{u} \in \boldsymbol{\mathscr{D}}(D),\ \operatorname{div} \mathbf{u} = 0\}$ and $V$ be the closure of $\mathscr{V}$ in $\mathbf{H}^1_0(D)$. In [8] it is shown that

$$V = \{\mathbf{u} \in \mathbf{H}^1_0(D),\ \operatorname{div} \mathbf{u} = 0\}.$$

The space $V$ is a Hilbert space with the inner product

$$(\mathbf{u}, \mathbf{v})_V := \sum_{i=1}^{n} (D_i \mathbf{u}, D_i \mathbf{v})_{\mathbf{L}^2(D)} = \sum_{i=1}^{n} (\mathbf{grad}\, u_i, \mathbf{grad}\, v_i)_{\mathbf{L}^2(D)}$$

and norm $\|\mathbf{u}\|_V = (\mathbf{u}, \mathbf{u})^{1/2}_V$, where $D_i \mathbf{u} = (D_i u_1, \dots, D_i u_n)$.

We will also apply the generalized Schwarz's inequality (see, for example, [10], page 186):

$$(x, y)^2_X \le (R^{-1}x, x)_X (Ry, y)_X \quad \forall x, y \in X, \tag{1}$$

where $R : X \to X$ is a linear bounded self-adjoint positive definite operator in Hilbert space $X$ over $\mathbb{R}$, and inequality (1) is transformed to an equality on the element $y = \lambda R^{-1} x,\ \ \forall \lambda \in \mathbb{R}$.

Let $H$ be a separable Hilbert space over $\mathbb{R}$. By $L^2(\Omega, H)$ we denote the Bochner space composed of random* variables $\xi = \xi(\omega)$ defined on a certain probability space $(\Omega, \mathscr{B}, P)$

*Random variable $\xi$ with values in Hilbert space $H$ is considered as a function $\xi : \Omega \to H$ mapping random events $E \in \mathscr{B}$ to Borel sets in $H$ (Borel $\sigma$-algebra in $H$ is generated by open sets in $H$).

with values in $H$ such that

$$\|\xi\|^2_{L^2(\Omega,H)} = \int_\Omega \|\xi(\omega)\|^2_H dP(\omega) < \infty. \tag{2}$$

In this case there exists the Bochner integral

$$\mathbb{E}\xi := \int_\Omega \xi(\omega)\, dP(\omega) \in H \tag{3}$$

called the expectation or the mean value of random variable $\xi(\omega)$ which satisfies the condition

$$(h, \mathbb{E}\xi)_H = \int_\Omega (h, \xi(\omega))_H\, dP(\omega) \quad \forall h \in H. \tag{4}$$

Being applied to random variable $\xi$ with values in $\mathbb{R}$ this expression leads to a usual definition of its expectation because the Bochner integral (3) reduces to a Lebesgue integral with probability measure $dP(\omega)$.

In $L^2(\Omega, H)$ one can introduce the inner product

$$(\xi, \eta)_{L^2(\Omega,H)} := \int_\Omega (\xi(\omega), \eta(\omega))_H\, dP(\omega) \quad \forall \xi, \eta \in L^2(\Omega, H). \tag{5}$$

Applying the sign of expectation, one can write relationships (2), (4), (5) as

$$\|\xi\|^2_{L^2(\Omega,H)} = \mathbb{E}\|\xi(\omega)\|^2_H, \tag{6}$$

$$(h, \mathbb{E}\xi)_H = \mathbb{E}(h, \xi(\omega))_H \quad \forall h \in H, \tag{7}$$

$$(\xi, \eta)_{L^2(\Omega,H)} := \mathbb{E}(\xi(\omega), \eta(\omega))_H \quad \forall \xi, \eta \in L^2(\Omega, H). \tag{8}$$

$L^2(\Omega, H)$ equipped with norm (6) and inner product (8) is a Hilbert space.

The Stokes problem consists of finding a vector function $\mathbf{v} = (v_1, \dots, v_n) : D \to \mathbb{R}^n$ and a scalar function $p : D \to \mathbb{R}$ from equations

In this paper we focus on the estimation problems for linearized Navier-Stokes equations

$$-\nu\mathbf{\Delta v} + \operatorname{grad} p = \mathbf{f} \quad \text{в} \quad D, \tag{9}$$

$$\operatorname{div} \mathbf{v} = 0 \quad \text{в} \quad D, \tag{10}$$

$$\mathbf{v} = 0 \quad \text{на} \quad \Gamma, \tag{11}$$

that simulate the motion of a viscous incompressible fluid in the domain $D$. Here vector-functions $\mathbf{v} = (v_1, \dots, v_n), \mathbf{f} = (f_1, \dots, f_n) : D \to \mathbb{R}^n$, and scalar function $p : D \to \mathbb{R}$ represent the velocity, body force, and the pressure fields, respectively, and the positive constant $\nu$ is the coefficient of kinematic viscosity.

It is known that in the case, when $\mathbf{f} \in \mathbf{L}^2(D)$, vector function $\mathbf{v}$ can be found from the following equations

$$\mathbf{v} \in V, \tag{12}$$

$$\nu \sum_{i=1}^{n} (D_i \mathbf{v}, D_i \mathbf{u})_{\mathbf{L}^2(D)} = (\mathbf{f}, \mathbf{u})_{\mathbf{L}^2(D)} \quad \forall \mathbf{u} \in V. \tag{13}$$

Problem (12)–(13), called the variational statement of the Stokes problem (9)–(11), is uniquely solvable [6]–[9].

Since in this paper, from observations of velocity $\mathbf{v}$ only the linear functionals of the form $l(\mathbf{v})$ will be evaluated, in future we will deal only with the variational statement (12)– (13) of the Stokes problem (9)– (11).

## 2. Setting of the estimation problem

The estimation problem consists in the following: from the observations

$$y = C\mathbf{v} + \xi, \tag{14}$$

find optimal in a certain sense estimate of the functional

$$l(\mathbf{v}) = (\mathbf{l}_0, \mathbf{v})_{\mathbf{L}^2(D)} = \int_D (\mathbf{l}_0(x), \mathbf{v}(x))_{\mathbb{R}^n}\, dx \tag{15}$$

in the class of estimates linear w.r.t. observations (14),

$$\widehat{l(\mathbf{v})} = (u, y)_H + c, \tag{16}$$

under the assumption that errors $\xi = \xi(\omega)$ in observations (14) are realizations of random variables defined on a certain probability space $(\Omega, \mathscr{B}, P)$ with values in a separable Hilbert space $H$ over $\mathbb{R}$, belong to the set $G_1$, and $\mathbf{f} \in G_0$, where

$$G_0 = \left\{ \tilde{\mathbf{f}} : \tilde{\mathbf{f}} \in \mathbf{L}^2(D), (Q\tilde{\mathbf{f}} - \mathbf{f}_0, \tilde{\mathbf{f}} - \mathbf{f}_0)_{\mathbf{L}^2(D)} \le \varepsilon_0 \right\}, \tag{17}$$

$$G_1 = \{\tilde{\xi} : \tilde{\xi} \in L^2(\Omega, H),\ \mathbb{E}\tilde{\xi} = 0,\ \mathbb{E}(Q_1\tilde{\xi}, \tilde{\xi})_H \le \varepsilon_1\}. \tag{18}$$

Here $\epsilon_k > 0$, $k = 0, 1$, are given constants; $u \in H$; $c \in \mathbb{R}$; $(\cdot, \cdot)_H$ is inner product in $H$; $\mathbf{l}_0, \mathbf{f}_0 \in \mathbf{L}^2(D)$ are given real-valued functions; $C \in \mathscr{L}(\mathbf{L}^2(D), H)$ is linear continuous operator; and $Q$, $Q_1$, are self-adjoint positive definite operators in $\mathbf{L}^2(D)$ and $H$, respectively, for which there exist bounded inverse operators $Q^{-1}$ and $Q_1^{-1}$. Further, without loss of generality we may set $\epsilon_k = 1$, $k = 0, 1$.

**Definition 1.** An estimate

$$\widehat{\widehat{l(\mathbf{v})}} = (\hat{u}, y)_H + \hat{c}$$

is called a minimax (or a guaranteed) estimate of $l(\mathbf{v})$ if element $\hat{u} \in H$ and a number $\hat{c} \in \mathbb{R}$ are determined from the condition

$$\inf_{u\in H,\, c\in\mathbb{R}} \sigma(u,c) = \sigma(\hat{u},\hat{c}),$$

where

$$\sigma(u,c) := \sup_{\tilde{\mathbf{f}}\in G_0,\, \tilde{\xi}\in G_1} \mathbb{E}[l(\tilde{\mathbf{v}}) - \widehat{l(\tilde{\mathbf{v}})}]^2,$$

$\tilde{\mathbf{v}}$ is a solution to problem (12) –(13) when $\mathbf{f} = \tilde{\mathbf{f}}$, $\widehat{l(\tilde{\mathbf{v}})} = (u,\tilde{y})_H + c$, $\tilde{y} = C\tilde{\mathbf{v}} + \tilde{\xi}$.

The quantity $\sigma := [\sigma(\hat{u},\hat{c})]^{1/2}$ is called the error of the minimax estimation of $l(\mathbf{v})$.

Thus, the minimax estimate is an estimate minimizing the maximal mean-square estimation error calculated for the “worst” implementation of perturbations.

## 3. Reducing of the estimation problem to the optimal control problem

To find representations for minimax estimates, we first reduce this problem to certain optimal control problem.

For every fixed $u \in H$ introduce a function $\mathbf{z}(x;u)$, as a solution to the following variational problem:

$$\mathbf{z}(\cdot;u) \in V, \tag{19}$$

$$\nu \sum_{i=1}^{n} (D_i\mathbf{z}(\cdot;u), D_i\mathbf{u})_{\mathbf{L}^2(D)} = (\mathbf{l}_0 - C^* J_H u, \mathbf{u})_{\mathbf{L}^2(D)} \quad \forall \mathbf{u} \in V, \tag{20}$$

where $C^* : H' \to \mathbf{L}^2(D)$ is an operator adjoint of $C$ defined by

$$(p, C^*g)_{\mathbf{L}^2(D)} =< Cp, g >_{H\times H'} \quad \forall p \in \mathbf{L}^2(D), g \in H'.$$

Then the following assertion is valid.

**Lemma 1.** *The problem of minimax estimation of $l(\mathbf{v})$ (i.e. the determination of $\hat{u}$ and $\hat{c}$) is equivalent to the problem of optimal control of the system described by equation (19),* (20) *with a cost function*

$$I(u) = (Q^{-1}\mathbf{z}(\cdot;u), \mathbf{z}(\cdot;u))_{\mathbf{L}^2(D)} + (Q_1^{-1}u, u)_H \to \inf_{u\in H}. \tag{21}$$

*Proof.* From relation (15) and (16) at $\mathbf{v} = \tilde{\mathbf{v}}$ , we have

$$l(\tilde{\mathbf{v}}) - \widehat{l(\tilde{\mathbf{v}})} = \int_D (\mathbf{l}_0(x), \tilde{\mathbf{v}}(x))_{\mathbb{R}^3}\, dx - \int_D (C^* J_H u(x), \tilde{\mathbf{v}}(x))_{\mathbb{R}^n}\, dx - (u,\tilde{\xi})_H - c$$

$$= \nu \sum_{i=1}^{n} (D_i\mathbf{z}(\cdot;u), D_i\tilde{\mathbf{v}})_{\mathbf{L}^2(D)} - (u,\tilde{\xi})_H - c. \tag{22}$$

Taking into account the fact, that $\tilde{\mathbf{v}}$ is a solution of problem (12) –(13) at $\mathbf{f} = \tilde{\mathbf{f}}$, and setting in (13) $\mathbf{u} = \mathbf{z}(\cdot; u)$, we come to the equality

$$\nu \sum_{i=1}^{n} (D_i\tilde{\mathbf{v}}, D_i\mathbf{z}(\cdot; u))_{\mathbf{L}^2(D)} = \int_D (\tilde{\mathbf{f}}(x), \mathbf{z}(x; u))_{\mathbb{R}^n}\, dx.$$

From (22) and the latter formula, we obtain

$$l(\tilde{\mathbf{v}}) - \widehat{l(\tilde{\mathbf{v}})} = \int_D (\tilde{\mathbf{f}}(x), \mathbf{z}(x; u))_{\mathbb{R}^n}\, dx - (u, \tilde{\xi})_H - c. \tag{23}$$

Applying to the right hand side of (23) the generalized Schwarz's inequality and (6)–(8), (17), (18), we find

$$\inf_{c\in\mathbb{R}^1} \sup_{\tilde{\mathbf{f}}\in G_0,\, \tilde{\xi}\in G_1} \mathbb{E}[l(\tilde{\mathbf{v}} - \widehat{l(\tilde{\mathbf{v}})}]^2 = \inf_{c\in\mathbb{R}^1} \sup_{\tilde{\mathbf{f}}\in G_0} \Big\{ \int_D (\tilde{\mathbf{f}}(x), \tilde{\mathbf{z}}(x; u))_{\mathbb{R}^n}\, dx - c \Big\}^2$$

$$+ \sup_{\tilde{\xi}\in G_1} \mathbb{E}\{(u, \tilde{\xi})_H\}^2 = \int_D (Q^{-1}\mathbf{z}(\cdot; u)(x), \mathbf{z}(x; u))_{\mathbb{R}^n}\, dx + (Q_1^{-1}u, u)_H.$$

with $c = \int_D (\mathbf{z}(x; u), \mathbf{f}_0(x))_{\mathbb{R}^n} dx$. The lemma is proved. □

## 4. Representation of guaranteed estimates of functionals from solutions of Stokes problem

Solving the optimal control problem (19) – (21) and applying arguments completely analogous to that used in the proof of Theorem 2 on page 62 from [2] , we prove the following.

**Theorem 1.** *The minimax estimate of $l(\mathbf{v})$ has the form*

$$\widehat{\widehat{l(\mathbf{v})}} = (\hat{u}, y)_H + \hat{c} = l(\hat{\mathbf{v}}) = \int_D (\mathbf{l}_0(x), \hat{\mathbf{v}})_{\mathbb{R}^n}(x)\, dx, \tag{24}$$

*where*

$$\hat{c} = \int_D (\hat{\mathbf{z}}(x), \mathbf{f}_0(x))_{\mathbb{R}^n}\, dx, \quad \hat{u} = Q_1 C\mathbf{p}, \tag{25}$$

*the functions $\mathbf{p}(x)$ and $\hat{\mathbf{z}}(x)$ are determined as a solution of the following problem:*

$$\hat{\mathbf{z}} \in V, \tag{26}$$

$$\nu \sum_{i=1}^{n} (D_i\hat{\mathbf{z}}, D_i\mathbf{u})_{\mathbf{L}^2(D)} = (\mathbf{l}_0 - C^* J_H Q_1 C\mathbf{p}, \mathbf{u})_{\mathbf{L}^2(D)} \quad \forall \mathbf{u} \in V, \tag{27}$$

$$\mathbf{p} \in V, \tag{28}$$

$$\nu\sum_{i=1}^{n}(D_i\mathbf{p},D_i\mathbf{u})_{\mathbf{L}^2(D)}=(Q^{-1}\hat{\mathbf{z}},\mathbf{u})_{\mathbf{L}^2(D)}\quad\forall\mathbf{u}\in V,\tag{29}$$

*and the function $\hat{\mathbf{v}}$ is determined from solution of the problem*

$$\hat{\mathbf{p}}\in V,\tag{30}$$

$$\nu\sum_{i=1}^{n}(D_i\hat{\mathbf{p}},D_i\mathbf{u})_{\mathbf{L}^2(D)}=(C^*J_HQ_1(y-C\hat{\mathbf{v}}),\mathbf{u})_{\mathbf{L}^2(D)}\quad\forall\mathbf{u}\in V,\tag{31}$$

$$\hat{\mathbf{v}}\in V,\tag{32}$$

$$\nu\sum_{i=1}^{n}(D_i\hat{\mathbf{v}},D_i\mathbf{u})_{\mathbf{L}^2(D)}=(Q^{-1}\hat{\mathbf{p}}+\mathbf{f}_0,\mathbf{u})_{\mathbf{L}^2(D)}\quad\forall\mathbf{u}\in V.\tag{33}$$

*Problems (26)–(29) and (30)–(33) are uniquely solvable.*

*The error of estimation $\sigma$ is given by an expression*

$$\sigma=[l(\mathbf{p})]^{1/2}=\left[\int_D(\mathbf{l}_0(x),\mathbf{p}(x))_{\mathbb{R}^n}\,dx\right]^{1/2}.\tag{34}$$

Note that the function $\hat{\mathbf{z}}(x)=\mathbf{z}(x;\hat{u})$, where $\mathbf{z}(x;u)$ is a solution of problem (19), (20), and $u=\hat{u}\in H$ is optimal control of the system governed by these equations with cost function (21) (see Lemma 1).

Also, as one can see from equations (30)–(33), the function $\hat{\mathbf{v}}$ entering into the representation $\widehat{\widehat{l(\mathbf{v})}}=l(\hat{\mathbf{v}})$ does not depend on the concrete functional $l$ and, hence, can be taken as a good estimate of an unknown solution $\mathbf{v}$ of the problem (12)–(13).

## 5. Approximate guaranteed estimates of linear functionals from solutions of Stokes problem. Theorems on convergence

Using the Galerkin method for solving problems (26)–(29) and (30)–(33), we obtain approximate guaranteed estimates via solutions of linear algebraic equations and show their convergence to the optimal estimates.

Introduce a sequence of finite-dimensional subspaces $V^h$ in the space $V$, defined by an infinite set of parameters $h_1,h_2,\dots$ with $\lim_{k\to 0}h_k=0$.

We say that sequence $\{V^h\}$ is complete in $V$, if for any $\mathbf{v}\in V$ and $\epsilon>0$ there exists an $\hat{h}=\hat{h}(\mathbf{v},\epsilon)>0$ such that $\inf_{\mathbf{w}\in V^h}\|\mathbf{v}-\mathbf{w}\|_V<\varepsilon$ for any $h<\hat{h}$. In other words, the completeness of sequence $\{V^h\}$ means that any element $\mathbf{v}\in V$ may be approximated with any degree of accuracy by elements of $\{V^h\}$.

Such finite-dimensional subspaces $V^h$ are constructed, for example, in [8], Ch 1, §4.

Take an approximate minimax estimate of $l(\mathbf{v})$ as

$$\widehat{l^h(\mathbf{v})} = (\hat{u}^h, y)_H + \hat{c}^h,$$

where

$$\hat{c}^h = \int_D (\hat{\mathbf{z}}^h(x), \mathbf{f}_0(x))_{\mathbb{R}^n}\, dx, \quad \hat{u}^h = Q_1 C \mathbf{p}^h, \tag{35}$$

and functions $\mathbf{p}^h(x)$, $\hat{\mathbf{z}}^h(x)$ are determined from the following system of variational equations:

$$\hat{\mathbf{z}}^h \in V^h, \tag{36}$$

$$\nu \sum_{i=1}^n (D_i \hat{\mathbf{z}}^h, D_i \mathbf{u}^h)_{\mathbf{L}^2(D)} = (\mathbf{l}_0 - C^* J_H Q_1 C \mathbf{p}^h, \mathbf{u}^h)_{\mathbf{L}^2(D)} \quad \forall \mathbf{u}^h \in V^h, \tag{37}$$

$$\mathbf{p}^h \in V, \tag{38}$$

$$\nu \sum_{i=1}^n (D_i \mathbf{p}^h, D_i \mathbf{u}^h)_{\mathbf{L}^2(D)} = (Q^{-1} \hat{\mathbf{z}}^h, \mathbf{u}^h)_{\mathbf{L}^2(D)} \quad \forall \mathbf{u}^h \in V^h. \tag{39}$$

The unique solvability of this system (and system (40)–(43)) follows from the proof of Theorem 1 in which $V$ is replaced by $V^h$.

**Theorem 2.** *Approximate minimax estimate* $\widehat{l^h(\mathbf{v})}$ *of* $l(\mathbf{v})$ *tends to a minimax estimate* $\widehat{\widehat{l(\mathbf{v})}}$ *of this expression as* $h \to 0$ *in the sense that*

$$\lim_{h\to 0} \sup_{\tilde{\mathbf{f}}\in G_0,\, \tilde{\xi}\in G_1} \mathbb{E}|\widehat{l^h(\tilde{\mathbf{v}})} - \widehat{\widehat{l(\tilde{\mathbf{v}})}}|^2 = 0$$

*and*

$$\lim_{h\to 0} \sup_{\tilde{\mathbf{f}}\in G_0,\, \tilde{\xi}\in G_1} \mathbb{E}|\widehat{l^h(\tilde{\mathbf{v}})} - l(\tilde{\mathbf{v}})|^2 = \sup_{\tilde{\mathbf{f}}\in G_0,\, \tilde{\xi}\in G_1} \mathbb{E}|\widehat{\widehat{l(\tilde{\mathbf{v}})}} - l(\tilde{\mathbf{v}}))|^2,$$

*where* $\tilde{\mathbf{v}}$ *is a solution of problem* (12) –(13) *at* $\mathbf{f} = \tilde{\mathbf{f}}$, $\widehat{l^h(\tilde{\mathbf{v}})} = (u^h, \tilde{y})_H + c^h$, $\tilde{y} = C\tilde{\mathbf{v}} + \tilde{\xi}$.

Now, we formulate an analogous result for the case when an estimate $\hat{\mathbf{v}}$ of $\mathbf{v}$ is directly determined from solution to the problem (30)–(33). Namely, the following result holds.

**Theorem 3.** *Let* $\hat{\mathbf{v}}^h \in V^h$ *be an approximate estimate of the vector-function* $\hat{\mathbf{v}} \in V$ *determined from the solution to the variational problem*

$$\hat{\mathbf{p}}^h \in V^h, \tag{40}$$

$$\nu \sum_{i=1}^n (D_i \hat{\mathbf{p}}^h, D_i \mathbf{u}^h)_{\mathbf{L}^2(D)} = (C^* J_H Q_1 (y - C\hat{\mathbf{v}}^h), \mathbf{u}^h)_{\mathbf{L}^2(D)} \quad \forall \mathbf{u}^h \in V^h, \tag{41}$$

$$\hat{\mathbf{v}}^h \in V^h, \tag{42}$$

$$\nu \sum_{i=1}^{n} (D_i \hat{\mathbf{v}}^h, D_i \mathbf{u}^h)_{\mathbf{L}^2(D)} = (Q^{-1}\hat{\mathbf{p}}^h + \mathbf{f}_0, \mathbf{u}^h)_{\mathbf{L}^2(D)} \quad \forall \mathbf{u}^h \in V^h. \tag{43}$$

*Then*

$$\|\hat{\mathbf{v}} - \hat{\mathbf{v}}^h\|_V \to 0 \quad \text{as} \quad h \to 0$$

*and the approximate minimax estimate* $\widehat{l^h(\mathbf{v})}$ *of* $l(\mathbf{v})$ *has the form*

$$\widehat{l^h(\mathbf{v})} = l(\hat{\mathbf{v}}^h) = \int_D (\mathbf{l}_0(x), \hat{\mathbf{v}}^h(x))_{\mathbb{R}^n}\, dx. \tag{44}$$

The proofs of Theorem 2 and Theorem 3 are similary to the proof of Proposition 3.2 on page 32 from [1].

Introducing the basis in the space $V^h$, problems (36)–(39) i (40)–(43) can be rewritten as a systems of liner algebraic equations. To do this, let us denote the elements of the basis by $\boldsymbol{\xi}_i$ $(i = 1, \ldots, N)$ where $N = \dim V^h$. The fact that $\hat{\mathbf{z}}^h, \mathbf{p}^h, \hat{\mathbf{p}}^h, \hat{\mathbf{v}}^h$ belong to the space $V^h$ means the existence of constants $\hat{z}_j$, $p_j$ and $\hat{p}_j$, $\hat{v}_j$ such that

$$\hat{\mathbf{z}}^h = \sum_{j=1}^{N} \hat{z}_j \boldsymbol{\xi}_j, \quad \mathbf{p}^h = \sum_{j=1}^{N} p_j \boldsymbol{\xi}_j \tag{45}$$

and

$$\hat{\mathbf{p}}^h = \sum_{j=1}^{N} \hat{p}_j \boldsymbol{\xi}_j, \quad \hat{\mathbf{v}}^h = \sum_{j=1}^{N} \hat{v}_j^{(2)} \boldsymbol{\xi}_j. \tag{46}$$

Setting in (37) and (39) and in (41) and (43) $\mathbf{u}^h = \boldsymbol{\xi}_i$ $(i = 1, \ldots, N)$, we obtain that finding $\hat{\mathbf{z}}^h$, $\mathbf{p}^h$ and $\hat{\mathbf{p}}^h$, $\hat{\mathbf{v}}^h$ is equivalent to solving the following systems of linear algebraic equations with respect to coefficients $\hat{z}_j$, $p_j$ and $\hat{p}_j$, $\hat{v}_j$ of expansions (45) and (46):

$$\sum_{j=1}^{N} a_{jl} \hat{z}_j + \sum_{j=1}^{N} a_{jl}^{(1)} p_j = b_l, \quad l = 1, \ldots, N, \tag{47}$$

$$\sum_{j=1}^{N} a_{il} p_j + \sum_{j=1}^{N} a_{jl}^{(2)} \hat{z}_j = 0, \quad l = 1, \ldots, N \tag{48}$$

and

$$\sum_{j=1}^{N} a_{jl} \hat{p}_j + \sum_{j=1}^{N} a_{jl}^{(1)} \hat{v}_j = b_l^{(1)}, \quad l = 1, \ldots, N, \tag{49}$$

$$\sum_{j=1}^{N} a_{il} \hat{v}_j + \sum_{j=1}^{N} a_{jl}^{(2)} \hat{p}_j = b_l^{(2)}, \quad l = 1, \ldots, N, \tag{50}$$

where

$$a_{jl} = \nu \sum_{i=1}^{n} (D_i \boldsymbol{\xi}_j, D_i \boldsymbol{\xi}_l)_{\mathbf{L}^2(D)}, \quad j, l = 1, \ldots, N, \tag{51}$$

$$a_{jl}^{(1)} = (C^* J_H Q_1 C \boldsymbol{\xi}_j, \boldsymbol{\xi}_l)_{\mathbf{L}^2(D)}, \quad j, l = 1, \ldots, N, \tag{52}$$

$$a_{jl}^{(2)} = -(Q^{-1} \boldsymbol{\xi}_j, \boldsymbol{\xi}_l)_{\mathbf{L}^2(D)}, \quad j, l = 1, \ldots, N, \tag{53}$$

$$b_l = (\mathbf{l}_0, \boldsymbol{\xi}_l)_{\mathbf{L}^2(D)}, \quad l = 1, \ldots, N, \tag{54}$$

$$b_l^{(1)} = (C^* J_H Q_1 y, \boldsymbol{\xi}_l)_{\mathbf{L}^2(D)}, \quad l = 1, \ldots, N, \tag{55}$$

$$b_l^{(2)} = (Q^{-1} \mathbf{f}_0, \boldsymbol{\xi}_l)_{\mathbf{L}^2(D)}, \quad l = 1, \ldots, N. \tag{56}$$

## References

1. Nakonechnyi, O. G. 1985. *Minimax Estimation of Functionals of Solutions to Variational Equations in Hilbert Spaces.* Kiev State University.
2. Nakonechny, A. G. 2004. *Optimal Control and Estimation in Partial Differential Equations.* Kyiv State University, Kyiv.
3. Nakonechny, A. G. and Podlipenko, Yu. K. 1997. Minimax Prediction of Solutions of Parabolic Equations under Incomplete Data. *Dopovidi Acad. Nauk Ukrainy*, 9, pp. 107–112.
4. Podlipenko, Yu. K., Ryabikova, A. V. 2005. Minimax Estimation Based on Incomplete Data of Solutions to Two-Point Boundary Value Problems for Systems of Linear Ordinary Differential Equations, *Zh. Vych. Matem. Matem. Fiz.*, 4, pp. 97–106.
5. Podlipenko, Y. and Shestopalov, Y. 2013. Guaranteed estimates of functionals from solutions and data of interior Maxwell problems under uncertainties, in *Springer Proceedings in Mathematics & Statistics*, Vol. 52. pp. 135–168.
6. Boyer, F. and Fabrie, P. 2013. *Mathematical Tools for the Study of the Incompressible Navier-Stokes Equations and Related Models.* New York, Dordrecht, Heidelberg, London: Springer.
7. Galdi, G. 2011. *An Introduction to the Mathematical Theory of the Navier-Stokes Equations. Steady-State Problems.* Second Edition. New York, Dordrecht, Heidelberg, London: Springer.
8. Temam, R. 1977. *Navier-Stokes Problem. Theory and Numerical Analysis.* Amsterdam, New York, Oxford: North-Holland Publishing Company.
9. Ladyzhenskaya, O. A. 1969. *Mathematical theory of viscous incompressible flow* (English translation, Second Edition). New York: Gordonand Breach.
10. Hutson, V., Pym, J. and Cloud, M. 2005. *Applications of Functional Analysis and Operator Theory.* Amsterdam: Elsevier.

# THE SCHEME OF PARTIAL AVERAGING FOR ONE CLASS OF HYBRID SYSTEMS

© **Olga Osadcha, Natalia Skripnik**

I. I. Mechnikov Odessa National University

e-mail: *olga.osadcha.ua@gmail.com*, *natalia.skripnik@gmail.com*

***Abstract***. *This paper contains the substantiation of the scheme of partial averaging for one class of hybrid systems where one equation is a differential equation with Hukuhara derivative and the other one is an ordinary differential equation.*

## Introduction

In practice there often appear the so-called hybrid systems — systems which contain equations of different nature: for example, one of the equations is a partial differential equation and the other one is an ordinary differential equation, or one of the equations is a discrete one and the other is a differential equation, etc. In this paper we consider the case of a hybrid system, when one of the equations is a differential equation with Hukuhara derivative and the other one is an ordinary differential equation. The interest in such systems follows from the fact, that some parameters of the model can be accurate, while the rest may contain the noise, errors and inaccuracies.

## 1. Main Definitions

Development of the theory of multivalued mappings led to the question what should be understood as a derivative of a multivalued mapping. The main cause of difficulties for the inducting of such definition was the nonlinearity of the space $comp(R^n)$, which led to the absence of the concept of difference. There are several approaches to define the difference of two sets, one of them is the Hukuhara difference.

**Definition 1.** [see [6]] Let $X, Y \in conv(R^n)$. The set $Z \in conv(R^n)$, where $X = Y + Z$, is called the Hukuhara difference of sets $X$ and $Y$ and is designated as $X \overset{h}{-} Y$.

Along with the inducted difference there appeared the concept of derivative.

**Definition 2.** [see [6]] A multuvalued mapping $X : I \to conv(R^n), I \subset R$, is called differentiable in the sense of Hukuhara at point $t \in I$ if there exists such $D_H X(t) \in conv(R^n)$ that the limits $\lim\limits_{\Delta t \to 0} \frac{1}{\Delta t} \left( X(t + \Delta t) \overset{h}{-} X(t) \right)$ and $\lim\limits_{\Delta t \to 0} \frac{1}{\Delta t} \left( X(t) \overset{h}{-} X(t - \Delta t) \right)$ exist and are equal to $D_H X(t)$. The set $D_H X(t)$ is called the Hukuhara derivative of the multivalued mapping $X : I \to conv(R^n)$ at point $t$.

In 1969 F.S. de Blasi and F. Iervolino first considered the differential equation with Hukuhara derivative [4, 2, 3, 1], which solution was a multivalued mapping. After that various existence, uniqueness theorems were proved, stability of solutions for this type of equations was investigated, integro-differential equations, impulse differential equations, differential equations with fractional derivatives, controlled differential equations with Hukuhara derivative were considered. The possibility of using some averaging schemes for such type of equations was studied in [5, 13, 11, 7, 12, 8, 9, 10].

Consider the hybrid system

$$\begin{cases} D_H X = F(t, X, y), \\ \dot{y} = g(t, X, y), \\ X(t_0) = X_0, \\ y(t_0) = y_0, \end{cases} \tag{1}$$

where $I = [t_0, T] \subset R$; $X : I \to conv(R^n)$ is a multivalued mapping; $y : I \to R^m$ is a vector function; $F : I \times conv(R^n) \times R^m \to conv(R^n)$ is a multivalued mapping; $g : I \times conv(R^n) \times R^m \to R^m$ is a vector function; $X_0 \in conv(R^n)$, $y_0 \in R^m$.

Consider a class $S$ of pairs $(X(\cdot), y(\cdot))$, where $X(\cdot)-$ is a continuously differentiable on $I$ in a sense of Hukuhara multivalued mapping, $y(\cdot)-$ is a continuously differential on $I$ vector-function.

**Definition 3.** A pair $(X(\cdot), y(\cdot)) \in S$ is called a solution of system (1), if it satisfies the system for all $t \in I$ (e.g for all $t \in I$ the following equalities fulfill $D_H X(t) = F(t, X(t), y(t))$, $\dot{y}(t) = g(t, X(t), y(t)))$ and $X(t_0) = X_0$, $y(t_0) = y_0$.

**Theorem 1.** *Let in the domain*

$$Q = \{(t, X, y) : t_0 \le t \le t_0 + a,\ h(X, X_0) \le b,\ \|y - y_0\| \le c\}$$

*the multivalued mapping $F(t, X, y)$ and the vector function $g(t, X, y)$ be continuous and satisfy the Lipschitz condition in variables $X$ and $y$, i.e. there exists such constant $\lambda > 0$ that*

$$h\left(F(t, X_1, y_1), F(t, X_2, y_2)\right) \le \lambda \left[h(X_1, X_2) + \|y_1 - y_2\|\right],$$
$$\|g(t, X_1, y_1) - g(t, X_2, y_2)\| \le \lambda \left[h(X_1, X_2) + \|y_1 - y_2\|\right].$$

*Then system (1) has the unique solution defined on the interval $[t_0, t_0 + d]$ where $d = \min\left(a, \frac{b}{M}, \frac{c}{M}\right)$, constant $M$ satisfies inequalities $|F(t, X, y)| \le M$, $\|g(t, X, y)\| \le M$ in the domain $Q$.*

## 2. Main Results

Consider the hybrid system with a small parameter

$$\begin{cases} D_H X = \varepsilon F(t, X, y), \\ \dot{y} = \varepsilon g(t, X, y), \\ X(0) = X_0, \\ y(0) = y_0, \end{cases} \tag{2}$$

where $t \geq 0$ is time, $X \in D_1 \subset conv(R^n)$, $y \in D_2 \subset R^m$, $\varepsilon > 0$ is a small parameter.

With system (2) the following partially averaged system is assigned:

$$\begin{cases} D_H \bar{X} = \varepsilon \bar{F}(t, \bar{X}, \bar{y}), \\ \dot{\bar{y}} = \varepsilon \bar{g}(t, \bar{X}, \bar{y}), \\ \bar{X}(0) = X_0, \\ \bar{y}(0) = y_0, \end{cases} \tag{3}$$

where

$$\lim_{T\to\infty} \frac{1}{T} h\left( \int_0^T F(t, X, y)\, dt, \int_0^T \bar{F}(t, X, y)\, dt \right) = 0, \tag{4}$$

$$\lim_{T\to\infty} \frac{1}{T} \left\| \int_0^T g(t, X, y)\, dt - \int_0^T \bar{g}(t, X, y)\, dt \right\| = 0. \tag{5}$$

**Theorem 2.** *Let in the domain $Q = \{(t, X, y):\ t \geq 0, X \in D_1, y \in D_2\}$ the following conditions hold:*

*1) the multivalued mappings $F(t, X, y), \bar{F}(t, X, y)$ and vector functions $g(t, X, y), \bar{g}(t, X, y)$ are continuous in $t$, uniformly bounded with constant $M$ and satisfy the Lipschitz condition in $X$ and $y$ with constant $\lambda$, i.e.*

$$|F(t, X, y)| \leq M, h\left(F(t, X_1, y_1), F(t, X_2, y_2)\right) \leq \lambda \left[h(X_1, X_2) + \|y_1 - y_2\|\right],$$

$$\left|\bar{F}(t, X, y)\right| \leq M, h\left(\bar{F}(t, X_1, y_1), \bar{F}(t, X_2, y_2)\right) \leq \lambda \left[h(X_1, X_2) + \|y_1 - y_2\|\right],$$

$$\|g(t, X, y)\| \leq M, \|g(t, X_1, y_1) - g(t, X_2, y_2)\| \leq \lambda \left[h(X_1, X_2) + \|y_1 - y_2\|\right],$$

$$\|\bar{g}(t, X, y)\| \leq M, \|\bar{g}(t, X_1, y_1) - \bar{g}(t, X_2, y_2)\| \leq \lambda \left[h(X_1, X_2) + \|y_1 - y_2\|\right];$$

*2) limits (4) and (5) exist uniformly with respect to $X \in D_1$ and $y \in D_2$;*

*3) the solution $\left(\bar{X}(t), \bar{y}(t)\right)$ of system (3) with the initial condition $\bar{X}(0) = X_0 \in D_1^{'} \subset D_1$, $\bar{y}(0) = y_0 \in D_2^{'} \subset D_2$ is defined for all $t \geq 0$, $\varepsilon \in (0, \sigma]$ and $\bar{X}(t)$ belongs with some $\rho$- neighborhood to the domain $D_1$, $\bar{y}(t)$ belongs with some $\xi$-neighborhood to the domain $D_2$.*

*Then for any $\eta > 0$ and $L > 0$ there exists such $\varepsilon_0(\eta, L) \in (0, \sigma]$ that for varepsilon $\in (0, \varepsilon_0]$ and $t \in [0, L\varepsilon^{-1}]$ the following inequalities fulfill:*

$$h\left(X(t), \bar{X}(t)\right) < \eta, \|y(t) - \bar{y}(t)\| < \eta,$$

*where $(X(\cdot), y(\cdot))$ and $\left(\bar{X}(\cdot), \bar{y}(\cdot)\right)$ are the solutions of systems (2) and (3) with the initial conditions $X(0) = \bar{X}(0) \in D_1'$, $y(0) = \bar{y}(0) \in D_2'$.*

*Proof.* From conditions 1) and 2) of the theorem it follows that systems (2) and (3) have unique solutions that are defined for $t \geq 0$ if $X(t)$ and $y(t)$ (accordingly $\bar{X}(t)$ and $\bar{y}(t)$) belong to the domains $D_1, D_2$. That is why for $D_1 = conv(R^n)$, $D_2 = R^m$ condition 3) follows from 1) and 2).

Replace systems (2) and (3) with the equivalent system of integral equations:

$$\begin{cases} X(t) = X_0 + \varepsilon \int\limits_0^t F(s, X(s), y(s))\, ds, \\ y(t) = y_0 + \varepsilon \int\limits_0^t g(s, X(s), y(s))\, ds, \end{cases} \tag{6}$$

$$\begin{cases} \bar{X}(t) = X_0 + \varepsilon \int\limits_0^t \bar{F}(s, \bar{X}(s), \bar{y}(s))\, ds, \\ \bar{y}(t) = y_0 + \varepsilon \int\limits_0^t \bar{g}(s, \bar{X}(s), \bar{y}(s))\, ds. \end{cases} \tag{7}$$

Then

$$h\left(X(t), \bar{X}(t)\right) =$$

$$= h\left(X_0 + \varepsilon \int\limits_0^t F(s, X(s), y(s))ds, X_0 + \varepsilon \int\limits_0^t \bar{F}(s, \bar{X}(s), \bar{y}(s))ds\right) =$$

$$= h\left(\varepsilon \int\limits_0^t F(s, X(s), y(s))ds, \varepsilon \int\limits_0^t \bar{F}(s, \bar{X}(s), \bar{y}(s))ds\right) \leq$$

$$\leq h\left(\varepsilon \int\limits_0^t F(s, X(s), y(s))ds, \varepsilon \int\limits_0^t F(s, \bar{X}(s), \bar{y}(s))ds\right) +$$

$$+h\left(\varepsilon \int\limits_0^t F(s, \bar{X}(s), \bar{y}(s))ds, \varepsilon \int\limits_0^t \bar{F}(s, \bar{X}(s), \bar{y}(s))ds\right) \leq$$

$$\leq \varepsilon \int\limits_0^t h\left(F(s, X(s), y(s)), F(s, \bar{X}(s), \bar{y}(s))\right) ds +$$

$$+\varepsilon h\left(\int\limits_0^t F(s,\bar X(s),\bar y(s))ds,\int\limits_0^t \bar F(s,\bar X(s),\bar y(s))ds\right)\le$$

$$\le\varepsilon\lambda\int\limits_0^t\left[h(X(s),\bar X(s))+\|y(s)-\bar y(s)\|\right]ds+$$

$$+\varepsilon h\left(\int\limits_0^t F(s,\bar X(s),\bar y(s))ds,\int\limits_0^t \bar F(s,\bar X(s),\bar y(s))ds\right). \qquad (8)$$

Similarly

$$\|y(t)-\bar y(t)\|\le$$

$$\le\varepsilon\lambda\int\limits_0^t\left[h(X(s),\bar X(s))+\|y(s)-\bar y(s)\|\right]ds+$$

$$+\varepsilon\left\|\int\limits_0^t\left[g(s,\bar X(s),\bar y(s))-\bar g(s,\bar X(s),\bar y(s))\right]ds\right\|. \qquad (9)$$

Divide the interval $[0,L\varepsilon^{-1}]$ in $p$ equal intervals by the points $t_i=\frac{iL}{\varepsilon p}, i=\overline{0,p}$. Define by $(\bar X_i,\bar y_i)=(\bar X(t_i),\bar y(t_i))$ the solution of system (2) in division points.

Let us estimate the expressions $\varepsilon h\left(\int\limits_0^t F(s,\bar X(s),\bar y(s))ds,\int\limits_0^t \bar F(s,\bar X(s),\bar y(s))ds\right)$ and $\varepsilon\left\|\int\limits_0^t\left[g(s,\bar X(s),\bar y(s))-\bar g(s,\bar X(s),\bar y(s))\right]ds\right\|$ in the interval $[t_k,t_{k+1}]$, where $0\le k\le p-1$.

$$\varepsilon h\left(\int\limits_0^t F(s,\bar X(s),\bar y(s))ds,\int\limits_0^t \bar F(s,\bar X(s),\bar y(s))ds\right)=$$

$$=\varepsilon h\left(\sum_{i=0}^{k-1}\int\limits_{t_i}^{t_{i+1}}F(s,\bar X(s),\bar y(s))ds+\int\limits_{t_k}^{t}F(s,\bar X(s),\bar y(s))ds,\right.$$

$$\left.\sum_{i=0}^{k-1}\int\limits_{t_i}^{t_{i+1}}\bar F(s,\bar X(s),\bar y(s))ds+\int\limits_{t_k}^{t}\bar F(s,\bar X(s),\bar y(s))ds\right)\le$$

$$\le\varepsilon\left[\sum_{i=0}^{k-1}h\left(\int\limits_{t_i}^{t_{i+1}}F(s,\bar X(s),\bar y(s))ds,\int\limits_{t_i}^{t_{i+1}}\bar F(s,\bar X(s),\bar y(s))ds\right)+\right.$$

$$+h\left(\int\limits_{t_k}^{t} F(s,\bar{X}(s),\bar{y}(s))ds,\int\limits_{t_k}^{t}\bar{F}(s,\bar{X}(s),\bar{y}(s))ds\right)\Bigg]\le$$

$$\le\varepsilon\left[\sum_{i=0}^{k-1}\left(h\left(\int\limits_{t_i}^{t_{i+1}} F(s,\bar{X}(s),\bar{y}(s))ds,\int\limits_{t_i}^{t_{i+1}} F(s,\bar{X}_i,\bar{y}_i)ds\right)+\right.\right.$$

$$+h\left(\int\limits_{t_i}^{t_{i+1}} F(s,\bar{X}_i,\bar{y}_i)ds,\int\limits_{t_i}^{t_{i+1}}\bar{F}(s,\bar{X}_i,\bar{y}_i)ds\right)+h\left(\int\limits_{t_i}^{t_{i+1}}\bar{F}(s,\bar{X}_i,\bar{y}_i)ds,\int\limits_{t_i}^{t_{i+1}}\bar{F}(s,\bar{X}(s),\bar{y}(s))ds\right)\Bigg)+$$

$$+h\left(\int\limits_{t_k}^{t} F(s,\bar{X}(s),\bar{y}(s))ds,\int\limits_{t_k}^{t} F(s,\bar{X}_k,\bar{y}_k)ds\right)+h\left(\int\limits_{t_k}^{t} F(s,\bar{X}_k,\bar{y}_k)ds,\int\limits_{t_k}^{t}\bar{F}(s,\bar{X}_k,\bar{y}_k)ds\right)+$$

$$\left.+\,h\left(\int\limits_{t_k}^{t}\bar{F}(s,\bar{X}_k,\bar{y}_k)ds,\int\limits_{t_k}^{t}\bar{F}(s,\bar{X}(s),\bar{y}(s))ds\right)\right]\le$$

$$\le\varepsilon\left[\sum_{i=0}^{k-1}\left(\int\limits_{t_i}^{t_{i+1}} h\left(F(s,\bar{X}(s),\bar{y}(s)),F(s,\bar{X}_i,\bar{y}_i)\right)\,ds+\right.\right.$$

$$+h\left(\int\limits_{t_i}^{t_{i+1}} F(s,\bar{X}_i,\bar{y}_i)ds,\int\limits_{t_i}^{t_{i+1}}\bar{F}(s,\bar{X}_i,\bar{y}_i)ds\right)+\int\limits_{t_i}^{t_{i+1}} h\left(\bar{F}(s,\bar{X}_i,\bar{y}_i),\bar{F}(s,\bar{X}(s),\bar{y}(s))\right)ds\Bigg)+$$

$$+\int\limits_{t_k}^{t} h\left(F(s,\bar{X}(s),\bar{y}(s)),F(s,\bar{X}_k,\bar{y}_k)\right)ds+$$

$$\left.+h\left(\int\limits_{t_k}^{t} F(s,\bar{X}_k,\bar{y}_k)ds,\int\limits_{t_k}^{t}\bar{F}(s,\bar{X}_k,\bar{y}_k)ds\right)+\int\limits_{t_k}^{t} h\left(\bar{F}(s,\bar{X}_k,\bar{y}_k),\bar{F}(s,\bar{X}(s),\bar{y}(s))\right)ds\right]\le$$

$$\le\varepsilon\left[\sum_{i=0}^{k}\int\limits_{t_i}^{t_{i+1}}\left(h\left(F(s,\bar{X}(s),\bar{y}(s)),F(s,\bar{X}_i,\bar{y}_i)\right)+h\left(\bar{F}(s,\bar{X}(s),\bar{y}(s)),\bar{F}(s,\bar{X}_i,\bar{y}_i)\right)\right)ds+\right.$$

$$+\sum_{i=0}^{k-1} h\left(\int\limits_{t_i}^{t_{i+1}} F(s,\bar{X}_i,\bar{y}_i)ds, \int\limits_{t_i}^{t_{i+1}} \bar{F}(s,\bar{X}_i,\bar{y}_i)ds\right) + h\left(\int\limits_{t_k}^{t} F(s,\bar{X}_k,\bar{y}_k)ds, \int\limits_{t_k}^{t} \bar{F}(s,\bar{X}_k,\bar{y}_k)ds\right)\Bigg].$$

Similarly

$$\varepsilon \left\| \int\limits_0^t \left[g(s,\bar{X}(s),\bar{y}(s)) - \bar{g}(s,\bar{X}(s),\bar{y}(s))\right] ds \right\| \leq$$

$$\leq \varepsilon \left[\sum_{i=0}^{k} \int\limits_{t_i}^{t_{i+1}} \left(\left\|g(s,\bar{X}(s),\bar{y}(s)) - g(s,\bar{X}_i,\bar{y}_i)\right\| + \left\|\bar{g}(s,\bar{X}(s),\bar{y}(s)) - \bar{g}(s,\bar{X}_i,\bar{y}_i)\right\|\right) ds + \right.$$

$$\left. + \sum_{i=0}^{k-1} \left\| \int\limits_{t_i}^{t_{i+1}} \left(g(s,\bar{X}_i,\bar{y}_i) - \bar{g}(s,\bar{X}_i,\bar{y}_i)\right) ds \right\| + \left\| \int\limits_{t_k}^{t} \left(g(s,\bar{X}_k,\bar{y}_k) - \bar{g}(s,\bar{X}_k,\bar{y}_k)\right) ds \right\| \right].$$

Notice that

$$h\left(\bar{X}(s),\bar{X}_i\right) = h\left(\bar{X}(s),\bar{X}(t_i)\right) \leq \varepsilon \int\limits_{t_i}^{s} h\left(\bar{F}(v,\bar{X}(v),\bar{y}(v)),\{0\}\right) dv \leq \varepsilon M(s-t_i),$$

$$\|\bar{y}(s)-\bar{y}_i\| = \|\bar{y}(s)-\bar{y}(t_i)\| \leq \varepsilon \int\limits_{t_i}^{s} \left\|\bar{g}(v,\bar{X}(v),\bar{y}(v))\right\| dv \leq \varepsilon M(s-t_i).$$

Then

$$\varepsilon \sum_{i=0}^{k} \int\limits_{t_i}^{t_{i+1}} h\left(F(s,\bar{X}(s),\bar{y}(s)), F(s,\bar{X}_i,\bar{y}_i)\right) ds \leq$$

$$\leq \varepsilon \sum_{i=0}^{k} \int\limits_{t_i}^{t_{i+1}} \lambda\left[h\left(\bar{X}(s),\bar{X}_i\right) + \|\bar{y}(s)-\bar{y}_i\|\right] ds \leq$$

$$\leq \varepsilon\lambda \cdot 2\varepsilon M \sum_{i=0}^{k} \int\limits_{t_i}^{t_{i+1}} (s-t_i)ds =$$

$$= 2\varepsilon^2\lambda M \sum_{i=0}^{k} \frac{(t_{i+1}-t_i)^2}{2} = \varepsilon^2\lambda M \cdot (k+1) \cdot \left(\frac{L}{\varepsilon m}\right)^2 \le \frac{\lambda M L^2}{m},$$

$$\varepsilon \sum_{i=0}^{k} \int_{t_i}^{t_{i+1}} h\left(\bar{F}(s,\bar{X}(s),\bar{y}(s)), \bar{F}(s,\bar{X}_i,\bar{y}_i)\right) \le \frac{\lambda M L^2}{m},$$

$$\varepsilon \sum_{i=0}^{k} \int_{t_i}^{t_{i+1}} \left\| g(s,\bar{X}(s),\bar{y}(s)) - g(s,\bar{X}_i,\bar{y}_i) \right\| ds \le$$

$$\le \varepsilon \sum_{i=0}^{k} \int_{t_i}^{t_{i+1}} \lambda \left[ h\left(\bar{X}(s),\bar{X}_i\right) + \|\bar{y}(s)-\bar{y}_i\| \right] ds \le$$

$$\le \varepsilon^2\lambda M \cdot (k+1) \cdot \left(\frac{L}{\varepsilon m}\right)^2 \le \frac{\lambda M L^2}{m},$$

$$\varepsilon \sum_{i=0}^{k} \int_{t_i}^{t_{i+1}} \left\| \bar{g}(s,\bar{X}(s),\bar{y}(s)) - \bar{g}(s,\bar{X}_i,\bar{y}_i) \right\| \le \frac{\lambda M L^2}{m}.$$

Using condition 2) of the theorem there exist such monotone decreasing functions $f_1(t)$ and $f_2(t)$ that tend to zero as $t \to \infty$ , that for all $(X,y) \in D_1 \times D_2$ we have:

$$h\left( \int_0^t F(s,\bar{X},\bar{y})ds, \int_0^t \bar{F}(s,\bar{X},\bar{y})ds \right) \le t \cdot f_1(t),$$

$$\left\| \int_0^t \left( g(s,\bar{X},\bar{y}) - \bar{g}(s,\bar{X},\bar{y}) \right) ds \right\| \le t \cdot f_2(t).$$

Then

$$\varepsilon h\left( \int_{t_i}^{t_{i+1}} F(s,\bar{X}_i,\bar{y}_i)ds, \int_{t_i}^{t_{i+1}} \bar{F}(s,\bar{X}_i,\bar{y}_i)ds \right) =$$

$$= \varepsilon h\left( \int_0^{t_{i+1}} F(s,\bar{X}_i,\bar{y}_i)ds \stackrel{h}{-} \int_0^{t_i} F(s,\bar{X}_i,\bar{y}_i)ds, \int_0^{t_{i+1}} \bar{F}(s,\bar{X}_i,\bar{y}_i)ds \stackrel{h}{-} \int_0^{t_i} \bar{F}(s,\bar{X}_i,\bar{y}_i)ds \right) \le$$

$$\le \varepsilon \left[ h\left( \int_0^{t_{i+1}} F(s,\bar{X}_i,\bar{y}_i)ds, \int_0^{t_{i+1}} \bar{F}(s,\bar{X}_i,\bar{y}_i)ds \right) + h\left( \int_0^{t_i} F(s,\bar{X}_i,\bar{y}_i)ds, \int_0^{t_i} \bar{F}(s,\bar{X}_i,\bar{y}_i)ds \right) \right] \le$$

$$\leq \varepsilon\left[t_{i+1}\cdot f_1(t_{i+1})+t_i\cdot f_1(t_i)\right]\leq 2\sup_{\tau\in[0,L]}\tau f_1\left(\frac{\tau}{\varepsilon}\right)=\gamma_1(\varepsilon),$$

$$\varepsilon\left\|\int\limits_{t_i}^{t_{i+1}}\big(g(s,\bar X_i,\bar y_i)-\bar g(s,\bar X_i,\bar y_i)\big)\,ds\right\|=$$

$$=\varepsilon\left\|\int\limits_{0}^{t_{i+1}}\big(g(s,\bar X_i,\bar y_i)-\bar g(s,\bar X_i,\bar y_i)\big)\,ds-\int\limits_{0}^{t_i}\big(g(s,\bar X_i,\bar y_i)-\bar g(s,\bar X_i,\bar y_i)\big)\,ds\right\|\leq$$

$$\leq\varepsilon\left[\left\|\int\limits_{0}^{t_{i+1}}\big(g(s,\bar X_i,\bar y_i)-\bar g(s,\bar X_i,\bar y_i)\big)\,ds\right\|+\left\|\int\limits_{0}^{t_i}\big(g(s,\bar X_i,\bar y_i)-\bar g(s,\bar X_i,\bar y_i)\big)\,ds\right\|\right]\leq$$

$$\leq \varepsilon\left[t_{i+1}\cdot f_2(t_{i+1})+t_i\cdot f_2(t_i)\right]\leq 2\sup_{\tau\in[0,L]}\tau f_2\left(\frac{\tau}{\varepsilon}\right)=\gamma_2(\varepsilon),$$

where $\tau=\varepsilon t$, a $\lim\limits_{\varepsilon\to 0}\gamma_1(\varepsilon)=0$, $\lim\limits_{\varepsilon\to 0}\gamma_2(\varepsilon)=0$. Similarly

$$h\left(\int\limits_{t_k}^{t}F(s,\bar X_k,\bar y_k)ds,\int\limits_{t_k}^{t}\bar F(s,\bar X_k,\bar y_k)ds\right)\leq$$

$$\leq \varepsilon\left[t\cdot f_1(t)+t_k\cdot f_1(t_k)\right]\leq 2\sup_{\tau\in[0,L]}\tau f_1\left(\frac{\tau}{\varepsilon}\right)=\gamma_1(\varepsilon),$$

$$\varepsilon\left\|\int\limits_{t_k}^{t}\big(g(s,\bar X_k,\bar y_k)-\bar g(s,\bar X_k,\bar y_k)\big)\,ds\right\|\leq$$

$$\leq \varepsilon\left[t\cdot f_2(t)+t_k\cdot f_2(t_k)\right]\leq 2\sup_{\tau\in[0,L]}\tau f_2\left(\frac{\tau}{\varepsilon}\right)=\gamma_2(\varepsilon).$$

So

$$\varepsilon h\left(\int\limits_{0}^{t}F(s,\bar X(s),\bar y(s))ds,\int\limits_{0}^{t}\bar F(s,\bar X(s),\bar y(s))ds\right)\leq\frac{2\lambda ML^2}{m}+(k+1)\gamma_1(\varepsilon)\leq$$

$$\leq\frac{2\lambda ML^2}{m}+m\gamma_1(\varepsilon)\equiv\phi_1(\varepsilon,m), \tag{10}$$

$$\varepsilon\left\|\int\limits_{0}^{t}\left[g(s,\bar X(s),\bar y(s))-\bar g(s,\bar X(s),\bar y(s))\right]ds\right\|\leq\frac{2\lambda ML^2}{m}+m\gamma_2(\varepsilon)\equiv\phi_2(\varepsilon,m). \tag{11}$$

If we substitute (10) in (8) and (11) in (9), we will get

$$h\left(X(t),\bar{X}(t)\right) \le \varepsilon\lambda \int\limits_0^t \left[h(X(s),\bar{X}(s)) + \|y(s)-\bar{y}(s)\|\right] ds + \varphi_1(\varepsilon,m),$$

$$\|y(t)-\bar{y}(t)\| \le \varepsilon\lambda \int\limits_0^t \left[h(X(s),\bar{X}(s)) + \|y(s)-\bar{y}(s)\|\right] ds + \varphi_2(\varepsilon,m).$$

Adding these two inequalities and applying Gronwall-Bellmann lemma we get

$$h\left(X(t),\bar{X}(t)\right) + \|y(t)-\bar{y}(t)\| \le e^{2\varepsilon\lambda\int\limits_0^t 1ds} \left(\phi_1(\varepsilon,m)+\phi_2(\varepsilon,m)\right) =$$
$$= e^{2\varepsilon\lambda t}\left(\frac{4\lambda ML^2}{m} + m\gamma_1(\varepsilon) + m\gamma_2(\varepsilon)\right) \le$$
$$\le e^{2\lambda L}\left(\frac{4\lambda ML^2}{m} + m\gamma_1(\varepsilon) + m\gamma_2(\varepsilon)\right).$$

Then for every summand the inequality holds:

$$h\left(X(t),\bar{X}(t)\right) \le e^{2\lambda L}\left(\frac{4\lambda ML^2}{m} + m\gamma_1(\varepsilon) + m\gamma_2(\varepsilon)\right),$$

$$\|y(t)-\bar{y}(t)\| \le e^{2\lambda L}\left(\frac{4\lambda ML^2}{m} + m\gamma_1(\varepsilon) + m\gamma_2(\varepsilon)\right).$$

Let $\eta_1 = \min\{\eta,\rho,\xi\}$. Choose $m$ to satisfy the inequality

$$e^{2\lambda L}\frac{\lambda ML^2}{m} < \frac{\eta_1}{12}.$$

Then fix $m$ and choose $\varepsilon_0 \in (0,\sigma]$ such that for $\varepsilon \in (0,\varepsilon_0]$ the inequalities hold

$$e^{2\lambda L}m\gamma_1(\varepsilon) \le \frac{\eta_1}{3}, e^{2\lambda L}m\gamma_2(\varepsilon) \le \frac{\eta_1}{3}.$$

Then $h\left(X(t),\bar{X}(t)\right) \le \eta_1$ and $\|y(t)-\bar{y}(t)\| \le \eta_1$ if the solution $(X(t),y(t))$ belongs to the domain $D_1 \times D_2$. And it follows from condition 3) of the theorem as $\eta_1 = \min\{\eta,\rho,\xi\}$.

So, we get that for any $\eta > 0$ and $L > 0$ there exists such $\varepsilon_0$, that for $\varepsilon \in (0,\varepsilon_0]$ and $t \in [0, L\varepsilon^{-1}]$ the following inequalities fulfill

$$h\left(X(t),\bar{X}(t)\right) \le \eta, \|y(t)-\bar{y}(t)\| \le \eta.$$

The theorem is proved. □

## 3. Conclusion

This paper contains the substantiation of the scheme of partial averaging for one class of hybrid systems where one equation is a differential equation with Hukuhara derivative and the other one is an ordinary differential equation. In case when the right-hand sides are periodic in time one can obtain a better estimate. Namely one can show that for any $L > 0$ there exist $C(L) > 0$ and $\varepsilon_0(L) > 0$ such that the conclusion of the theorem holds with $\eta = C\varepsilon$.

## References

1. Brandao Lopes Pinto, A., De Blasi, F. and Iervolino, F. 1970. Uniqueness and existence theorems for differential equations with compact convex valued solutions. *Boll. Unione Mat. Ital.*, 4, pp. 534–538.
2. De Blasi, F. and Iervolino, F. 1969. Equazioni differentiali con soluzioni a valore compatto convesso. *Boll. Unione Mat. Ital.*, 2 (4–5), pp. 491–501.
3. De Blasi, F. and Iervolino, F. 1971. Euler method for differential equations with set - valued solutions. *Boll. Unione Mat. Ital.*, 4 (4), pp. 941–949.
4. De Blasi, F. 1976. On the differentiability of multifunctions. *Pacific J. Math*, 66 (1), pp. 67–81.
5. Perestyuk, N., Plotnikov, V., Samoilenko, A. and Skripnik, N. 2011. Differential Equations with Impulse Effects. Berlin: De Gruyter.
6. Hukuhara, M. 1967. Integration des applications mesurables dont la valeur est un compact convexe. Func. Ekvacioj, 10, pp. 205–223.
7. Kisielewicz, M. 1976. Method of Averaging for Differential Equations with Compact Convex Valued Solutions. Rend. Math, 9 (3), pp. 397–408.
8. Plotnikov, V. and Rashkov, P. 2001. Averaging in differential equations with Hukuhara derivative and delay. *Funct. Differ. Equ.*, 8, pp. 371–381.
9. Plotnikov, V. and Kichmarenko, O. 2006. Averaging of controlled equations with the Hukuhara derivative. *Nonlinear Oscil.* (N. Y.), 9 (3), pp. 365–374.
10. Plotnikov, V. and Kichmarenko, O. 2007. Averaging of equations with Hukuhara derivative, multivalued control and delay. *Bulletin of the Odessa National University*, 12 (7), pp. 130–139.
11. Plotnikov, V., Plotnikov, A. and Vityuk, A. 1999. Differential equations with a multivalued right-hand side. Asymptotic methods. Odessa: AstroPrint.
12. Plotnikov, V. and Rashkov, P. 1997. "Existence, continuous dependence and averaging in differential equations with Hukuhara derivative and delay", paper presented at "Mathematics and education in mathematics": Proceedings of Twenty Sixth Spring Conference of the Union of Bulgarian Mathematicians, Plovdiv, Bulgaria, April 22–25. pp. 179–184.
13. Skripnik, N. 2008. Averaging of impulsive differential equations with Hukuhara derivative. Visn. Yuriy Fedkovich Chernivtsy National University, 374, pp. 109–115.

# A SIMULATION OF SUB-GAUSSIAN RANDOM FIELDS ON A SPHERE OF ORLICZ SPACES

© Anatoliy Pashko

Taras Shevchenko National University of Kyiv

e-mail: *pashkoua@mail.ru*

***Abstract***. *Estimates for the convergence speed models isotropic random fields on the sphere in the norms of Orlich space. The resulting estimates are used to construct models of random fields on the sphere. Models approximate the random field with given accuracy and reliability.*

## Introduction

This paper continues investigation of convergence rate of the random series [3]–[7]. We obtain estimates for sub-Gaussian trigonometric series in Orlicz spaces. Same estimations of Gaussian series were obtained at [3]–[5], and on the uniform metric [6]. The results are used to model homogeneous and isotropic random fields on the sphere. Methods for the random modeling fields can be found in [2].

## 1. Basic determinations

Let $(\Omega, A, P)$ — be a standard probability space.

**Definition 1.** A random variable $\xi$ is sub-Gaussian, if $E\xi = 0$ and $a \geq 0$ exists, such that for every $\lambda \in R^1$ following estimate occurs

$$E\exp\{\lambda\xi\} \leq \exp\left\{\frac{\lambda^2 a^2}{2}\right\}.$$

A space of sub-Gaussian variables $Sub(\Omega)$ is Banach relative to the following norm

$$\tau(\xi) = \sup_{\lambda \neq 0}\left[\frac{2\ln E\exp\{\lambda\xi\}}{\lambda^2}\right]^{\frac{1}{2}}.$$

**Definition 2.** A family of random variables $S_\Lambda \subset Sub(\Omega)$ called strictly sub-Gaussian, if every finite or countable set of random variables $\{\xi_i, i \in I\} \subset S_\Lambda$ for every $\lambda \in R^1$ performs

$$\tau^2\left(\sum_{i\in I}\lambda_i\xi_i\right) = E\left(\sum_{i\in I}\lambda_i\xi_i\right).$$

Let $(T, \sum, \mu), \mu(T) < \infty$ — be some measurable space, $L_U(T)$ — Orlicz space, that was generated from C-function $U = \{U(x), x \in R^1\}$.

**Definition 3.** Orlicz space, generated by $U(x)$, called a function family $\{f(t), t \in T\}$, and for each function $f(t)$ exists constant $r$, that $\int_T U\left(\frac{f(t)}{r}\right) d\mu(t) < \infty$.

Space $L_U(T)$ is Banach relative to norm $\|f\|_{L_U} = \inf\left\{r > 0 : \int_T U\left(\frac{f(t)}{r}\right) d\mu(t) \leq 1\right\}$. Norm $\|f\|_{L_U}$ called the Luxemburg norm.

**Definition 4.** Let $f = \{f_k(t), t \in T, k = 1, 2...\}$ — be a family of functions from the space $L_U(T)$. This family belongs to the class $D_U(c)$, if numeric sequence $c = \{c_k, k = 1, 2, ...\}, 0 \leq c_k \leq c_{k+1}$ exists, such that for every sequence $r = \{r_k, k = 1, 2, ...\}$ following inequality holds

$$\left\| \sum_{k=1}^{n} r_k f_k(t) \right\|_{L_U} \leq c_n \left\| \sum_{k=1}^{n} r_k f_k(t) \right\|_{L_2}.$$

**Definition 5.** Isotropic in the broad sense field will be called linear isotropic field, if the random variables $\xi_m^l$ are independent.

## 2. Simulation random fields on the sphere

Let $S_d$ sphere in $d$ — be a measurable space. A random continuous in mean-square homogeneous and isotropic field on the sphere $\xi(x)$ can be represented as [9]

$$\xi(x) = \sum_{m=0}^{\infty} \sum_{l=1}^{h(m,d)} \xi_m^l S_m^l(x),$$

where $\xi_m^l$ independent strictly sub-Gaussian random variables, $E\xi_m^l = 0$, $E\xi_m^l \xi_r^s = \sigma_m^2 \delta_m^r \delta_l^s$, $m = 0, 1, ...,$ $l = 1, ..., h(m, d)$, $S_m^l(x)$ — Spherical harmonic of $m$ degree, $h(m, d)$ — harmonic count and $\sum_{m=0}^{\infty} \sigma_m^2 h(m, d) < \infty$.

Field model construct as

$$\xi_M(x) = \sum_{m=0}^{M} \sum_{l=1}^{h(m,d)} \xi_m^l S_m^l(x),$$

Number of summand $M$ chosen in such way, where $\delta > 0$ and $0 < \alpha < 1$ and inequality holds $P\{\| \xi(x) - \xi_M(x) \| \geq \delta\} < 1 - \alpha$.

Next results were proved in papers [4, 5].

**Lemma 1.** *Let $\xi_1, \xi_2, ..., \xi_n$ — be an independent strictly sub-gaussian random variables, $E\xi_i^2 = \sigma_i^2, i = 1, 2, ..., n$. Then, for each $0 \leq u < 1$ and $N = 1, 2, ...$ following inequality* *holds*

$$E\exp\left\{\frac{u}{2Z_N}\sum_{l=1}^{n}\xi_l^2\right\}\le \exp\left\{\frac{1}{2}\sum_{l=1}^{n}\frac{1}{l}\left(\frac{uZ_l}{Z_N}\right)^l\right\},$$

*where* $Z_N = \left(\sum_{i=1}^{n}\sigma_i^{2N}\right)^{\frac{1}{N}}$.

**Lemma 2.** *Let* $\xi_1, \xi_2, ..., \xi_n, ...$ *— independent strictly sub-gaussian random variables. If* $\sum_{i=1}^{\infty}\sigma_i^2 < \infty$*, then for each* $0 \le u < 1$ *and* $N = 1, 2, ...$ *and following inequality holds*

$$E\exp\left\{\frac{u}{2Z_N}\sum_{i=1}^{\infty}\xi_i^2\right\}\le \exp\left\{\frac{1}{2}\sum_{l=l}^{\infty}\frac{1}{l}\left(\frac{uZ_l}{Z_N}\right)^l\right\},$$

*where* $Z_N = \left(\sum_{i=1}^{\infty}\sigma_i^{2N}\right)^{\frac{1}{N}}$.

**Lemma 3.** *Let* $\xi_1, \xi_2, ..., \xi_n, ...$ *— be an independent strictly sub-Gaussian random variables. If* $\sum_{i=1}^{\infty}\sigma_i^2 < \infty$*, then for such* $0 \le u < 1$ *and* $N = 1, 2, ...$ *following inequality holds*

$$E\exp\left\{\frac{u}{2Z_N}\sum_{i=1}^{\infty}\xi_i^2\right\}\le \exp\left\{\frac{1}{2}v_N(u) + w_N(u)\right\},$$

*where*

$$w_N(u) = \frac{1}{2}\sum_{l=N}^{\infty}\frac{u^l}{l},$$

$$v_1(u) = 0, v_N(u) = \sum_{l=1}^{N-1}\frac{(lZ_l)^l}{lZ_N^l}.$$

Have similar lemma

**Lemma 4.** *If* $\left(\sum_{i=m}^{\infty}h(i,d)\sigma_i^{2N}\right)^{\frac{1}{N}} < \infty$*, for* $N = 1, 2, ...$ *then for each* $0 \le u < 1$ *and* $m \ge 1$ *following inequality holds*

$$E\exp\left\{\frac{u}{2J(N,m)} \parallel \xi_m(x) \parallel_{L_2}^2 \right\}\le \exp\left\{\frac{1}{2}v_N(u) + w_N(u)\right\},$$

*where* $J(N,m) = \left(\sum_{i=m}^{\infty}h(i,d)\sigma_i^{2N}\right)^{\frac{1}{N}}$.

Using these results we obtain the following theorem.

**Theorem 1.** *If* $\left(\sum_{i=1}^{\infty} h(i,d)\sigma_i^{2N}\right)^{\frac{1}{N}} < \infty$, *for* $N = 1, 2, ...$ *then for each* $0 \le u < 1$ *and* $\varepsilon > 0$ *following inequality holds*

$$P\Big\{ \parallel \xi(x) - \xi_M(x) \parallel_{L_2} > \varepsilon \Big\} \le \exp\Big\{ -\frac{u\varepsilon^2}{2J(N, M+1)} \Big\} \exp\Big\{ \frac{1}{2} v_N(u) + w_N(u) \Big\},$$

*where* $w_N(u)$ *and* $v_N(u)$ *defined in Lemma 3,* $J(N,m)$ *— defined in Lemma 4.*

*Proof.* Compute $\parallel \xi(x) - \xi_M(x) \parallel_{L_2}^2 = \sum_{m=M+1}^{\infty} \sum_{l=1}^{h(m,d)} (\xi_m^l)^2$. According to Lemma 4 for $0 \le u < 1$ holds

$$E\exp\Big\{ \frac{u}{2J(N,M)} \parallel \xi(x) - \xi_M(x) \parallel_{L_2}^2 \Big\} \le \exp\Big\{ \frac{1}{2} v_N(u) + w_N(u) \Big\},$$

where $J(N,M) = \left(\sum_{i=M+1}^{\infty} h(i,d)\sigma_i^{2N}\right)^{\frac{1}{N}}$ and $N = 1, 2, ...$

Then, according to the Chebyshev inequality

$$P\Big\{ \parallel \xi(x) - \xi_M(x) \parallel_{L_2} > \varepsilon \Big\} = P\Big\{ \parallel \xi(x) - \xi_M(x) \parallel_{L_2}^2 > \varepsilon^2 \Big\} \le$$

$$\le E\exp\Big\{ \frac{u}{2J(N,M)} \parallel \xi(x) - \xi_M(x) \parallel_{L_2}^2 \Big\} \exp\Big\{ -\frac{u\varepsilon^2}{2J(N,M)} \Big\}.$$

Theorem proved. □

When $N = 1$ we have

$$P\Big\{ \parallel \xi(x) - \xi_M(x) \parallel_{L_2} > \varepsilon \Big\} \le \frac{\varepsilon}{(J(M))^{\frac{1}{2}}} \exp\Big\{ -\frac{\varepsilon^2}{2J(M)} \Big\} \exp\Big\{ \frac{1}{2} \Big\},$$

where $J(M) = \left(\sum_{i=M+1}^{\infty} h(i,d)\sigma_i^2\right)$. . When $N = 2$ we have

$$P\Big\{ \parallel \xi(x) - \xi_M(x) \parallel_{L_2} > \varepsilon \Big\} \le \left( \frac{\varepsilon^2 - J(M)}{(J(2,M))} + 1 \right)^{\frac{1}{2}} \exp\Big\{ -\frac{\varepsilon^2 - J(M)}{2J(2,M)} \Big\},$$

where $J(2,M) = \left(\sum_{i=M+1}^{\infty} h(i,d)\sigma_i^4\right)$.

Let

$$P_m(x) = \sum_{l=1}^{h(m,d)} \xi_m^l S_m^l(x),$$

$$Q_m^r(x) = \sum_{s=m}^{r} P_s(x),$$

$$R_m^r(x,b) = \sum_{s=m}^{r} b_s P_s(x),$$

where $\{b_s > 0\}$ — be a monotonically non-decreasing sequence. $R_m^r(x)$ — that trigonometric polynomial of $(d-1)$ — variable of order $m = (m, m, ..., m)$. that's why for $p > 2$ holds (Nikolskii inequality [8])

$$\| R_m^r(x,b) \|_{L_p} \leq 3^{d-1}(r)^{(d-1)\left(\frac{1}{2}-\frac{1}{p}\right)} \| R_m^r(x,b) \|_{L_2} .$$

**Theorem 2.** *Let a monotonically non-decreasing sequence exists* $\{b_k > 0\}$, $b_k \to \infty$, $k \to \infty$, *that following series convergent*

$$\sum_{s=1}^{\infty} c_s (J(s))^{\frac{1}{2}} < \infty$$

*where*

$$J(s) = \sum_{k=1}^{s} h(k,d) b_k^2 \sigma_k^2$$

*and*

$$c_s = 3^{d-1}(s)^{(d-1)\left(\frac{1}{2}-\frac{1}{p}\right)} \left(\frac{1}{b_s} - \frac{1}{b_{s+1}}\right),$$

*Then, for each*

$$\varepsilon > \sum_{s=M+1}^{\infty} c_s (J(s))^{\frac{1}{2}}$$

*estimate holds*

$$P\left\{ \| \xi(x) - \xi_M(x) \|_{L_p} > \varepsilon \right\} \leq \frac{\varepsilon}{(D(M))} \exp\left\{ -\frac{\varepsilon^2}{2D(M)^2} \right\} \exp\left\{ \frac{1}{2} \right\},$$

*where* $D(M) = \sum_{s=M+1}^{\infty} c_s (J(s))^{\frac{1}{2}}$.

*Proof.* Write Abel's transformation

$$Q_m^r(x) = \sum_{i=m}^{r-1} R_m^i(x,b) \left(\frac{1}{b_i} - \frac{1}{b_{i+1}}\right) + R_m^r(x,b) \frac{1}{b_{r+1}}.$$

Then

$$\| Q_m^r(x) \|_{L_p} = \sum_{i=m}^{r-1} \| R_m^i(x,b) \|_{L_p} \left(\frac{1}{b_i} - \frac{1}{b_{i+1}}\right) + \| R_m^r(x,b) \| +p \frac{1}{b_{r+1}} \leq$$

$$\sum_{i=m}^{r} c_i \| R_m^i(x,b) \|_{L_2},$$

where $c_i = 3^{d-1}(i)^{(d-1)\left(\frac{1}{2}-\frac{1}{p}\right)}\left(\frac{1}{b_i}-\frac{1}{b_{i+1}}\right)$, by $m \le i < r$, and $c_r = 3^{d-1}(r)^{(d-1)\left(\frac{1}{2}-\frac{1}{p}\right)}\left(\frac{1}{b_r}\right)$, by $i = r$.

Therefore, for some $y > 0$ holds

$$E\exp\left\{y^2 \parallel Q_m^r(x) \parallel_{L_p}^2\right\} \le E\exp\left\{\left(y\sum_{i=m}^{r} c_i \parallel R_m^i(x,b) \parallel_{L_2}\right)^2\right\}.$$

According to Jensen's inequality $\delta_i, i = m, .., r$ such that $\sum_{i=m}^r \delta_i = 1$, holds

$$E\exp\left\{\left(\sum_{i=m}^{r}\frac{y}{\delta_i}\delta_i c_i \parallel R_m^i(x,b) \parallel_{L_2}\right)^2\right\} \le E\exp\left\{\sum_{i=m}^{r}\delta_i\left(\frac{y}{\delta_i}c_i \parallel R_m^i(x,b) \parallel_{L_2}\right)^2\right\}.$$

According to the Holder inequality

$$E\exp\left\{y^2 \parallel Q_m^r(x) \parallel_{L_p}^2\right\} \le \prod_{i=m}^{r}\left(E\exp\left\{\left(\frac{y}{\delta_i}c_i \parallel R_m^i(x,b) \parallel_{L_2}\right)^2\right\}\right)^{\delta_i}.$$

Mark $u_i = 2y^2 c_i^2 \delta_i^{-2} J(N,m,i)$, where

$$J(N,m,i) = \left(\sum_{k=m}^{i} h(k,d)\sigma_k^{2N} b_k^{2N}\right)^{\frac{1}{N}}.$$

If $0 \le u_i = 2y^2 c_i^2 \delta_i^{-1} J(N,m,i) < 1$, then by Lemma 4

$$E\exp\left\{\left(\frac{y}{\delta_i}c_i \parallel R_m^i(x,b) \parallel_{L_2}\right)^2\right\} =$$

$$E\exp\left\{\frac{2y^2 c_i^2 \delta_i^{-2} J(N,m,i)}{2J(N,m,i)}\left(\parallel R_m^i(x,b) \parallel_{L_2}\right)^2\right\} \le$$

$$\exp\left\{\frac{1}{2}v_N(u_i) + w_N(u_i)\right\}.$$

If $N = 1$, then $E\exp\left\{\left(\frac{y}{\delta_i}c_i \parallel R_m^i(x,b) \parallel_{L_2}\right)^2\right\} = (1-u_i)^{\frac{1}{2}}$.

Then

$$E\exp\left\{y^2 \parallel Q_m^r(x) \parallel_{L_p}^2\right\} \le \prod_{i=m}^{r}\left((1-u_i)^{-\frac{1}{2}}\right)^{\delta_i} = \exp\left\{-\frac{1}{2}\sum_{i=m}^{r}\delta_i \ln(1-u_i)\right\}.$$

Set $\delta_i = \frac{\sqrt{2}yc_i J^{\frac{1}{2}}(m,i)}{V}$, where $V > 0$ and $\sum_{i=m}^r \delta_i = 1$.

Then

$$\sum_{i=m}^{r}\frac{\sqrt{2}yc_i J^{\frac{1}{2}}(m,i)}{V} = \frac{\sqrt{2}y}{V}\sum_{i=m}^{r} c_i J^{\frac{1}{2}}(m,i) = 1,$$

or

$$V=\sqrt{2}y\sum_{i=m}^{r}c_iJ^{\frac{1}{2}}(m,i).$$

And therefore,

$$u_i=2y^2c_i^2\delta_i^{-2}J(N,m,i)=2y^2\Bigg(\sum_{i=m}^{r}c_iJ^{\frac{1}{2}}(m,i)\Bigg)^2=V^2.$$

If $V<1$, then

$$E\exp\{y^2\parallel Q_m^r(x)\parallel_{L_p}^2\}\leq\exp\{-\frac{1}{2}\sum_{i=m}^{r}\delta_i\ln(1-u_i)\}=$$

$$\exp\Bigg\{-\frac{1}{2}\ln(1-V^2)\sum_{i=m}^{r}\frac{\sqrt{2}yc_iJ^{\frac{1}{2}}(m,i)}{V}\Bigg\}=(1-V^2)^{-\frac{1}{2}}.$$

Let set $y^2=\frac{V^2}{2}\Bigg(\sum_{i=m}^{r}c_iJ^{\frac{1}{2}}(m,i),\Bigg)^{-2}$, then

$$E\exp\Bigg\{\frac{V^2}{2\Bigg(\sum_{i=m}^{r}c_iJ^{\frac{1}{2}}(m,i),\Bigg)^2}\parallel Q_m^r(x)\parallel_{L_p}^2\Bigg\}\leq(1-V^2)^{-\frac{1}{2}}.$$

Consequently, according to the Chebyshev inequality,

$$P\{\parallel Q_m^r(x)\parallel_{L_p}^2>\varepsilon^2\}\leq\exp\Bigg\{-\frac{V^2\varepsilon^2}{2\Bigg(\sum_{i=m}^{r}c_iJ^{\frac{1}{2}}(m,i),\Bigg)^2}\Bigg\}(1-V^2)^{-\frac{1}{2}}.$$

If the series of converges $\sum_{i=1}^{\infty}c_iJ^{\frac{1}{2}}(1,i)$, then $\sum_{i=m}^{r}c_iJ^{\frac{1}{2}}(m,i)\to 0$ where $m\to\infty$, $r\to\infty$.

Consequently, $P\{\parallel Q_m^r(x)\parallel_{L_p}^2>\varepsilon^2\}\to 0$ where $m\to\infty, r\to\infty$. If we set $m=M+1$ and direct $r\to\infty$, then we will get following estimate

$$P\{\parallel Q_M^\infty(x)\parallel_{L_p}^2>\varepsilon^2\}\leq\exp\Bigg\{-\frac{V^2\varepsilon^2}{2\Bigg(\sum_{i=M+1}^{\infty}c_iJ^{\frac{1}{2}}(M+1,i)\Bigg)^2}\Bigg\}(1-V^2)^{-\frac{1}{2}}.$$

If we optimize right part by $V$, i.e., when

$$\varepsilon>\sum_{i=M+1}^{\infty}c_iJ^{\frac{1}{2}}(M+1,i)$$

set $V=1-\frac{1}{\varepsilon}\sum_{i=M+1}^{\infty}c_iJ^{\frac{1}{2}}(M+1,i)$, then we get estimate. Theorem proved. □

With a similar argument we can prove a next theorem.

**Theorem 3.** *If sequence convergence*

$$\sum_{i=1}^{\infty} C_i \frac{h(i,d)\sigma_i^2}{\left(\sum_{k=i}^{\infty} h(k,d)\sigma_k^2\right)^{\frac{1}{2}}} < \infty,$$

*where* $C_i = 3^{d-1}(i)^{(d-1)\left(\frac{1}{2}-\frac{1}{p}\right)}$, *then for each* $\varepsilon > G(M+1)$ *holds next estimate*

$$P\Big\{ \parallel \xi(x) - \xi_M(x) \parallel_{L_p} > \varepsilon \Big\} \leq$$

$$\frac{\varepsilon}{(G(M+1))} \exp\Big\{ -\frac{\varepsilon^2}{2G^2(M+1)} \Big\} \exp\Big\{ \frac{1}{2} \Big\},$$

*where*

$$G^2(M+1) = (1+\sqrt{2}) \sum_{i=M+1}^{\infty} C_i \frac{h(i,d)\sigma_i^2}{\left(\sum_{k=i}^{\infty} h(k,d)\sigma_k^2\right)^{\frac{1}{2}}}.$$

*Proof.* For chosen sequence $\{b_k\}$ get

$$J^{\frac{1}{2}}(M+1,i) = \left( \sum_{k=M+1}^{i} h(k,d)\sigma_k^2 b_k^2 \right)^{\frac{1}{2}} +$$

$$1+\sqrt{2}\left( \Big(\sum_{k=i}^{\infty} h(k,d)\sigma_k^2\Big)^{-1} - 1 \right)^{\frac{1}{2}}$$

and

$$\sum_{i=M+1}^{\infty} C_i J^{\frac{1}{2}}(M+1,i) \leq \sum_{i=M+1}^{\infty} C_i \Big(\frac{1}{b_i} - \frac{1}{b_{i+1}}\Big) \left( 1+\sqrt{2}\left( \Big(\sum_{k=i}^{\infty} h(k,d)\sigma_k^2\Big)^{-1} - 1 \right)^{\frac{1}{2}} \right) \leq$$

$$\left(1+\sqrt{2}\right) \sum_{i=M+1}^{\infty} C_i \frac{h(i,d)\sigma_i^2}{\left(\sum_{k=i}^{\infty} h(k,d)\sigma_k^2\right)^{\frac{1}{2}}}.$$

Theorem proved. □

In modeling of random fields ask the modeling accuracy $\varepsilon > 0$ and reliability $1-\alpha$, $0 < \alpha < 1$. For space $L_2$ number of summand $M$ in model (1) we found as minimum value, where inequality when $N = 1$

$$\frac{\varepsilon}{(J(M))^{\frac{1}{2}}} \exp\Big\{ -\frac{\varepsilon^2}{2J(M)} \Big\} \exp\Big\{ \frac{1}{2} \Big\} \leq 1-\alpha,$$

And when $N = 2$ inequality

$$\left( \frac{\varepsilon^2 - J(M)}{(J(2,M))} + 1 \right)^{\frac{1}{2}} \exp\Big\{ -\frac{\varepsilon^2 - J(M)}{2J(2,M)} \Big\} \leq 1-\alpha$$

For functional space $L_p, p > 2$ number of summand $M$ in model (1) we found from inequality

$$\frac{\varepsilon}{(D(M+1))}\exp\Big\{-\frac{\varepsilon^2}{2D^2(M+1)}\Big\}\exp\Big\{\frac{1}{2}\Big\}\leq 1-\alpha.$$

Left-side depends on the sequence $\{b_k\}$. As $\sum_{i=1}^{\infty} h(i,d)\sigma_i^2 < \infty$, then, without any loss of generality, we can assume that $\sum_{i=1}^{\infty} h(i,d)\sigma_i^2 = 1$ and choose $b_k = 1+\Bigg(\big(\sum_{i=k}^{\infty} h(i,d)\sigma_i^2\big)^{-1}-1\Bigg)$.

Consequently, number of summand in model $\xi_M(x)$ we can calculate from inequality

$$\frac{\varepsilon}{(G(M+1))}\exp\Big\{-\frac{\varepsilon^2}{2G^2(M+1)}\Big\}\exp\Big\{\frac{1}{2}\Big\}\leq 1-\alpha.$$

With a similar argument we can prove a next theorem

**Theorem 4.** *Let $U(x) = \{U(x), x \in \mathbb{R}$ be a C-Orlicz function, those function*

$$GU(x) = \exp\big\{(U^{(-1)}(x-1))^2\big\}, x \geq 1$$

*convex at $x \geq 1$, $U^{(-1)}(x)$ - inverse function to $U(x)$. Then for every $x$ such $x \geq \max(\mu(T),1)\tau\big(2+(U^{(-1)}(1))^{-2}\big)^{\frac{1}{2}}$, following inequality holds*

$$P\Big\{\parallel \xi(x)-\xi_M(x)\parallel_{L_U(x)} > \varepsilon\Big\}\leq$$

$$\frac{\varepsilon U^{(-1)}(1)}{\max(1,\mu(T))\tau}\exp\Big\{-\frac{\varepsilon^2(U^{(-1)}(1))^2}{2(\max(1,\mu(T))\tau)^2}\Big\}\exp\Big\{\frac{1}{2}\Big\},$$

**Theorem 5.** *Let $\xi(x)$ — be a strictly Orlicz field, $\xi_M(x)$ — those field model. If some $p > 2$ sequence convergence $\sum_{m=1}^{\infty} h(m,d)\sigma_m^2 m^{(d-1)\left(2-\frac{2}{p}\right)}$, then for any $\delta > 0$ following inequality holds*

$$P\Big\{\Big(\int\limits_{S_d}|\xi(x)-\xi_M(x)|^p dx\Big)^{\frac{1}{p}}\geq\delta\Big\}\leq\Bigg(U\Bigg(\delta^2\Big[C_{u_p}C(3^{d-1})^2\sum_{m=M+1}^{\infty}h(m,d)\sigma_m^2 m^{(d-1)\left(2-\frac{2}{p}\right)}\Big]^{-\frac{p}{2}}\Bigg)\Bigg)^{-1}.$$

*Proof.* Let use Nikolskii inequality. We have

$$P\Big\{\Big(\int\limits_{S_d}|\xi(x)-\xi_M(x)|^p dx\Big)^{\frac{1}{p}}\geq\delta\Big\}= P\Big\{\Big(\int\limits_{S_d}|\sum_{m=M+1}^{\infty}h(m,d)\sigma_m^2 m^{(d-1)\left(2-\frac{2}{p}\right)}|^p dx\Big)^{\frac{1}{p}}\geq\delta\Big\}$$

$$\leq\Bigg(U\Bigg(\delta^2\Big[C_{u_p}\sum_{m=M+1}^{\infty}\|\xi_m^l\|_{u_p}^2\big(\int\limits_{T}|S_m^l(x)|^p dx\big)^{\frac{2}{p}}\Big]^{-\frac{p}{2}}\Bigg)\Bigg)^{-1}.$$

As the $S_m^l(x)$ — a trigonometric polynomial of $(d-1)$ variables, then for $p > 2$ inequality holds $\|S_m^l\|_{L_p} \leq 3^{d-1}m^{(d-1)\left(2-\frac{2}{p}\right)}\|S_m^l\|_{L_2}$, a $\|S_m^l(x)\|_{L_2} = 1$. Therefore, when

$p > 2$ we have $\left(\int_T |S_m^l(x)|^p dx\right)^{\frac{1}{p}} \leq 3^{d-1} m^{(d-1)\left(2-\frac{2}{p}\right)}$ and $\|S_m^l\|_{u_p}^2 \leq C\sigma_m^2$. In that case following inequality holds

$$C_{u_p} \sum_{m=M+1}^{\infty} \sum_{l=1}^{h(m,d)} \|\xi_m^l\|_{u_p}^2 \Big( \int_T |S_m^l(x)|^p dx \Big)^{\frac{2}{p}} \leq C_{u_p} C (3^{d-1})^2 \sum_{m=M+1}^{\infty} h(m,d) \sigma_m^2 m^{(d-1)\left(2-\frac{2}{p}\right)}.$$

Theorem proved. □

When $p = 2$ holds following theorem

**Theorem 6.** *Let $\xi(x)$ - be a strictly Orlicz field, $\xi_M(x)$ - those field model. If such sequence convergence $\sum_{m=1}^{\infty} h(m,d)\sigma_m^2 m^{(d-1)}$, then for each $\delta > 0$ holds inequality*

$$P\left\{ \left( \int_{S_d} |\xi(x) - \xi_M(x)|^2 dx \right)^{\frac{1}{2}} \geq \delta \right\} \leq \left( U\left( \delta^2 \left[ C_{u_p} C \sum_{m=M+1}^{\infty} h(m,d)\sigma_m^2 m^{(d-1)} \right] \right) \right)^{-1}.$$

## Conclusion

The paper constructed a model of random fields on the sphere. The models of linear isotropic fields from Orlicz space were observed. The models approximate the field with given accuracy and reliability.

## References

1. Dzjamko, V., Moza, A. and Pashko, A. 2011. The accuracy in and reliability of linear isotropic fields simulation on the sphere. *Bulletin of Taras Shevchenko National University of Kyiv.Physics and Mathematics.*, 4, pp. 16–19.
2. Ermakov, S. and Mikhailov, G. 1982. Statistical simulation. Moscow: Nauka.
3. Kozachenko, Y. and Pashko, A. 2001. Accuracy estimation of simulation of random fields on the sphere in Lp. *Bulletin of Taras Shevchenko National University of Kyiv. Mathematics and Mechanics.*, 7, pp. 26–32.
4. Kozachenko, Y. and Pashko, A. 1998. Accuracy of simulation of stochastic processes in norms of Orlicz spaces.I. Theor. Probability and Math. Statist., 58, pp. 45–60.
5. Kozachenko, Y. and Pashko, A. 1998. Accuracy of simulation of stochastic processes in norms of Orlicz spaces.II. Theor. Probability and Math. Statist., 59, pp. 75–60.
6. Kozachenko, Y. and Pashko, A. 1999. Simulation of random processes. Kyiv: T. Shevchenko University.
7. Kozachenko, Y., Pashko, A. and Rozora, I. 2007. Simulation of random processes and fields. Kyiv: Zadruga.
8. Nikolskii, S. 1977. Approximation of functions of several variables and embedding theorems. Moscow: Nauka.
9. Yadrenko, M. 1980. Spectral theory of random fields. Kyiv: Vyshcha shkola.

# HARNESSING EMPIRICAL CHARACTERISTIC FUNCTION CONVERGENCE BEHAVIOUR

© Lino Sant

University of Malta

Department of Statistics and OR, Faculty of Science

e-mail: *lino.sant@um.edu.mt*

***Abstract.*** *Parameter estimation for Lévy processes has generated much research effort lately with a strong injection of interest coming from finance. Within this context the problem can be framed as estimation using increments from an infinitely divisible distribution, for which empirical characteristic functions (ecf) are convenient tools. However convergence of ecf's to Gaussian processes has not been exploited as fully as it might have been. In this paper we go back to strong convergence results derived from the Hungarian construction and use Brownian bridge approximations to construct new estimators. In particular we study one integrated square error estimator tailored to show deference to the variance structure of the corresponding Gaussian process. We prove some of its nice statistical properties and present simulation results obtained through its use.*

## 1. Introduction

The flexibility offered by Lévy processes for use in modeling has been acknowledged in various fields within the natural sciences, notably physics and chemistry, and in the applied science, with special mention in meteorology and geology. In more recent years applications in finance and insurance have given a big boost in the study and use of Levy processes. The possibility of including distributions with heavy tails as well as paths with jumps were two features which made these processes so attractive. Parameter estimation for Lévy processes progressed a lot with a large number of estimation techniques being proposed and developed over a number of papers. In this paper we are specifically interested in methods using the characteristic function. The Levy-Khinchine representation motivates the interest these methods have aroused. In particular the class of infinitely divisible distributions assume an important role seeing that the independent increments of Lévy processes belong this class. However, lately the interest runs deeper than that as researchers are trying to reconstruct Lévy measures through spectral methods applied to characteristic functions as in Belomestny (2010)[1].

Parzen's (1965)[16] idea of using the the empirical characteristic function for estimation was first used for stable distributions by Press (1972)[18]. Notable contributions to the area are those provided by Paulson, A. S., Holcomb and E. W., Leitch, (1975)[17], Heathcote (1977)[10], Koutrouvelis (1980)[14], Kogon and Williams (1998)[12], Feuerverger and McDunnough (1981a, 1981b)[8, 9].

## 2. Parameter Estimation of the Characteristic Function

2.1. **Uses of the Empirical Characteristic Function.** The search for good estimators of parameters within the Lévy context has been heavily influenced by earlier research on stable distributions. A characteristic function is defined by $\varphi(t) = \int e^{itx} dF(x) = \varphi^R(t) + i\varphi^I(t)$ and is associated uniquely with some distribution $F$. The class of characteristic functions for stable distributions happens to be parametrized by $\boldsymbol{\theta}$ a 4-dimensional vector as in $\varphi(t, \boldsymbol{\theta})$. In cases where an explicit formula for the distribution function is not known, characteristic functions are most useful. However the advantages of characteristic function methods in statistics, like robustness and smoothness of the functions involved, have been shown to be considerable in Paulson et al (1975)[17], Yu (2004)[19]. Their use has been quite extensive in model-based hypothesis testing and goodness-of-fit statistics.

In general readings from a Lévy process will give us increments which form a sequence of iid random variables $X_1, ..., X_n$ from an infinitely divisible distribution function $F$ . The empirical characteristic function (ecf) is defined by: $c_t^n = \frac{1}{n} \sum_{j=1}^{n} e^{itX_j}$.

Glivenko-Cantelli assures us that we have strong convergence of this sum of random variables to the characteristic function uniformly in $t$. Following the development of empirical process theory, a stochastic process $Y_t^n$ can be constructed out of the iid sample: $Y_t^n = \sqrt{n} \left( \frac{1}{n} \sum_{j=1}^{n} e^{itX_j} - \varphi(t) \right)$ which is called the normalized empirical characteristic function. The behaviour of this process was studied extensively from mid-1970's starting with Kent (1975)[11] onwards. The major result was that it converges weakly to a complex Gaussian process under certain conditions. These conditions were refined and related to a number of properties of the limit complex process which we denote by $Z_t = U_t + iV_t$, with $U$ and $V$ being both real processes. $Z_t$ has mean 0 and covariance function given by: $K(s, t) = \varphi(t - s) - \varphi(t)\varphi(-s)$.

One important property, which leading researchers were insisting on, was continuity of sample paths for $Z_t$ , or rather the existence of a version of the limit process which does have continuous paths. This condition guarantees that convergence occurs with reference to the measure generated by the paths of the stochastic process viewed as random elements in the space of continuous functions on some compact subset of $\mathbb{R}$, say $\mathscr{C}([-1, 1])$. The insistence that the limit measure has support on this Banach space had deep theoretical implications as discussed in Marcus (1981)[15]. However it is well known that there are Gaussian processes whose sample paths are not continuous in the sense above.

2.2. **Strong Approximations.** In practice one might well be happy working with an empirical characteristic function whose limiting Gaussian process might have paths in the space of right-continuous functions $\mathscr{D}(\mathbb{R})$. Path continuity might not be needed in some applications. There are a lot of interesting properties still around. This can be appreciated by the fact that by construction, $Y_t^n$ has $\varphi(t-s)-\varphi(t)\varphi(-s)$ as its covariance function. One particularly fruitful way of studying the asymptotic behaviour of $Y_t^n$ is provided by recourse to the Hungarian construction of the Brownian bridge and Kiefer process sequence approximations as first set up in Komlos, Major, Tusnady (1975) [13]. This technique was perfected, generalized and applied to many situations to obtain more manageable results by Csörgö (1981)[4].

The starting point is the empirical process $\sqrt{n}(F_n(t) - F(t))$ which can be approximated strongly by a sequence of Brownian bridges $B_t^n$ (to which we limit ourselves) at the following rates:

$$\mathbb{P}\Big[\omega : \sup_{0\le t\le T} |\sqrt{n}(F_n(t) - F(t)) - B^n_{F(t)}| = O\Big(\frac{\log n}{\sqrt{n}}\Big)\Big] = 1 \tag{1}$$

where we assume the sufficient condition given in Csörgö (1981)[5], namely:

**Condition 4.** *For some $\alpha > 0$, $x^\alpha F(-x) + x^\alpha(1 - F(x)) = O(1)$ when $x \to \infty$*

holds. These Brownian bridges live on the same probability space and thus can be used to approximate the empirical process on a set of probability 0.

Under this same condition, following Csörgö, we have a similar result for empirical characteristic functions. For an underlying probability space which is large enough to allow suitable constructions of the various processes involved, there exists a sequence of Brownian bridges $B_t^n$ defined on the same probability space for which we define the corresponding Fourier transform, written as a stochastic integral: $Z_t^n = \int_{-\infty}^{\infty} e^{itx} dB^n_{F(x)}$ such that:

$$\mathbb{P}\big[\omega : \sup_{T_1\le t\le T_2} |Y_t^n - Z_t^n| = O\big(\frac{(\log n)^{(\alpha+1)/\alpha+2}}{n^{\alpha/(2\alpha+4)}}\big)\big] = 1 \text{ where } -\infty < T_1 < T_2 < \infty. \tag{2}$$

2.3. **The Gaussian Limit Process.** It is not hard to see that the Csörgö perspective gives us another expression for the limit process $Z_t$, which is of course the same process introduced earlier on:

$$Z_t = \int_{-\infty}^{\infty} e^{itx} dB_{F(x)} = \int_0^1 e^{itF^{-1}(y)} dB_y = U_t + iV_t \tag{3}$$

Having an explicit form of the limit process, we can do a lot of computations with it for estimation purposes. We can experiment through simulation to get a good picture of the more probable paths of the process. The plots in Figure 1 shown give an idea of how the paths more likely to be generated by the normalized ecf look like for a process whose increments gave gamma distributed random variables.

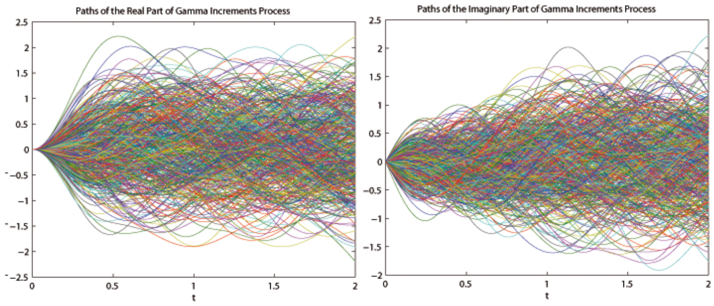



*Fig.* 1. Paths of a Gaussian Limit Process

We can treat the characteristic empirical function, $\mathbb{P}$ almost surely and hence distributionally , as $O\big(\frac{(\log n)^{(\alpha+1)/\alpha+2}}{n^{\alpha/(2\alpha+4)}}\big)$ close to the stochastic integral with respect to a Brownian bridge. If the distribution function $F$ or its inverse is not known, computation-wise we are still not defeated. We could approximate $F^{-1}$ by the empirical quantile process obtained from $F_n^{-1}$ whose approximation by Brownian bridges runs parallel to the one above and has been extensively studied by another Csörgö, Miklos (1983)[6]

If we only know the functional form of the characteristic function, as in the case of stable distributions, then we could apply the inverse Fourier transform on the characteristic function.

2.4. **Estimation using the Characteristic Function.** There are quite a few estimation techniques that have been developed to obtain estimates of parameters of the characteristic function proper using the ecf. We mention briefly two important ones and concentrate more on the technique which is closest in spirit to the ones we are proposing here.

The natural idea for using the ecf in estimation is to define some distance $d$ between the empirical characteristic function $c_t^n$ and any characteristic function $\varphi(\boldsymbol{\theta})$, call it $d(c_t^n, \varphi(\boldsymbol{\theta}))$ or some suitably defined functional of the difference between the two functions, and measure this distance cumulatively over some subset $O$ of the set over which $t$ varies. For instance if $O$ is finite, $O = \{t_1, t_2, ...t_K\}$ we could use $G(\theta) = \sum_{k=1}^{K} d(c_{t_k}^n, \varphi(t_k, \boldsymbol{\theta}))$ as the discrepancy measure between the ecf and a particular characteristic function over $O$. Then we compute the values of the parameter vector $\theta_0$ which minimizes this discrepancy and declare the corresponding vector to be our estimate: $\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{argmin} G(\boldsymbol{\theta})$.

This technique was developed quite a bit in Feurverger and Mureika (1977) [7] and was even analyzed through the GMM perspective by various authors Yu (2004) [19]. Within the GMM paradigm discrepancies are expressed vectorially as $\boldsymbol{d}\mathbf{B}\boldsymbol{d'}$for some suitably selected weighting matrix $\mathbf{B}$. Carrasco et al (2007)[2] take this idea further by proposing to take as $\mathbf{B}$ a suitably selected operator working on a Hilbert space and also to go from summation over time instants to integration over time into what they call a continuum of moment conditions CGMM. The intimate dependence of these methods on characteristic functions can best be appreciated in Carrasco and Kotchoni (2010) [3] but it is the integrating moment conditions over $t$ which is the more interesting to us. However, we shall not take this point further.

Of direct interest to us is the use of the related integrated square error function for parameter estimation technique successfully underpinned by a theory Heathcote (1977)[10]. A distance function between the ecf and a characteristic function is defined as the weighted integral of the square of the modulus of the difference. Its minimum gives the estimator: $\hat{\boldsymbol{\theta}} = \underset{\theta}{argmin} \int_{-\infty}^{\infty} |c_t^n(\boldsymbol{\theta}) - \varphi(\boldsymbol{\theta})|^2 dG(t)$. The development of the theory parallels that of the maximum likelihood method. But it has well-known problems of poor efficiency in comparison with this same method Yu (2004) [19] . Usually the weighting function is blamed on the choice of the weighting function. And this is where we strike. The choices of weighting functions were made to be dependent only on $t$ with absolutely no consideration of the characteristic functions itself.

2.5. **A New Type of Estimators.** We propose a class of estimators which are designed to exploit the strong Brownian bridge approximations. Such approximations can be useful for proving statistical properties of the estimators as well as for providing ways to compute associated asymptotic distributions through simulation. We shall define functionals of the type $J^n(\boldsymbol{\theta}) = \int_0^T \zeta(\boldsymbol{\theta}, \varphi, Y_t^n)dt$, which when suitably normalized, will converge strongly to a functional $J$ of the Gaussian limit process. Furthermore, passage to the asymptotic limit can be made to proceed through estimators of the type $\widehat{\boldsymbol{\theta}}$, given by $J^n(\widehat{\boldsymbol{\theta}}) = \underset{\boldsymbol{\theta}}{inf}\, J^n(\boldsymbol{\theta})$ and will lead to $J(\boldsymbol{\theta_0})$ the value at the true parameter vector $\boldsymbol{\theta_0}$. In contrast with the squared integrated error type estimators our estimator involves $\varphi$ more intricately in the integrand $\zeta$.

As examples we give:

$$J_1^n(\boldsymbol{\theta}) = \int_0^T \frac{(U_t^n(\boldsymbol{\theta}))^2}{\frac{1}{2}(1 + \varphi^R(2t, \boldsymbol{\theta}) - \varphi^R(t, \boldsymbol{\theta})^2} dt \text{ and } J_2^n(\boldsymbol{\theta}) = \int_0^T \frac{(V_t^n(\boldsymbol{\theta}))^2}{\frac{1}{2}(1 - \varphi^R(2t, \boldsymbol{\theta})) - \varphi^I(t, \boldsymbol{\theta})^2} dt.$$

And for our estimators we define: $\widehat{\boldsymbol{\theta}}_k = \underset{\theta}{\text{argmin}}\ J_k(\boldsymbol{\theta})$

The form of $\zeta$ has been purposely constructed so as to reflect the variance structure of the limit processes ( real and imaginary ). The integrand converges to a $\chi^2$ distributed random variable at the true value of the parameter. It can be suitably selected according to the type of distribution under investigation or to capture the features considered important. Besides appropriate statistical properties which may be needed to ensure the required asymptotic convergence, $\eta$ could be chosen so that simulation techniques can be applied on the corresponding stochastic integral of Brownian bridges to obtain numerical values for the required distributions. These estimators are more general than the integrated square error estimators in including directly the characteristic function in the "weighting"function for the integral. From now onwards we shall work with the minimizing function:

$$J^n(\boldsymbol{\theta}) = \int_0^T \frac{|Y_t^n(\boldsymbol{\theta})|^2}{1-|\varphi(t,\boldsymbol{\theta})|^2} dt \tag{4}$$

which has been designed to penalize mismatches between the variance of the normalized ecf and the variance given by the $\boldsymbol{\theta}$ choice. In some sense we are forcing on our choice of estimates a variance structure on the normalized empirical characteristic function which is close to that of the limit process. As an extension of this idea we propose another estimator, which enforces the covariance structure more rigorously as follows, while it seeks for the minimum of the functional: $J^n(\boldsymbol{\theta}) = \int_0^T \int_0^T \frac{|Y_t^n(\boldsymbol{\theta})\overline{Y_s^n(\boldsymbol{\theta})}|^2}{\varphi(t-s,\boldsymbol{\theta})-\varphi(t,\boldsymbol{\theta})\varphi(-s,\boldsymbol{\theta})} dsdt$ Working with this estimator may be a bit cumbersome, but from some simulation work we conducted, the results obtained were very encouraging. We shall revert to proving results for the simpler estimator 4. We prove a number of results about its statistical properties most of which should apply to similarly defined estimators along the lines indicated above.

2.6. **Basic results.** First a few definitions and elementary results:

Let $Y_t^n = \sqrt{n}(U_t^n + iV_t^n)$ so that $\sqrt{n}U_t^n = \mathfrak{Re}(Y_t^n)$ and $\sqrt{n}V_t^n = \mathfrak{Im}(Y_t^n)$.

The following equations hold:

$$\mathbb{E}[Y_t^n(\boldsymbol{\theta_0})] = 0 \text{ and } \mathbb{E}[Y_t^n(\boldsymbol{\theta_0})\overline{Y_s^n}(\boldsymbol{\theta_0})] = \varphi(t-s,\boldsymbol{\theta_0}) - \varphi(t,\boldsymbol{\theta_0})\varphi(-s,\boldsymbol{\theta_0})) \tag{5a}$$

$$\mathbb{E}[|Y_t^n|^2(\boldsymbol{\theta_0})] = 1 - |\varphi(t,\boldsymbol{\theta_0})|^2 \tag{5b}$$

$$\lim_{n\to\infty} c_t^n = \varphi(t,\boldsymbol{\theta_0}) \ \mathbb{P} \text{ a.s. uniformly in } t \tag{5c}$$

$$\lim_{n\to\infty} U_t^n(\boldsymbol{\theta_0}) = 0 = \lim_{n\to\infty} V_t^n(\boldsymbol{\theta_0}) \ \mathbb{P} \text{ a.s. uniformly in } t \tag{5d}$$

$$\mathbb{P}a.s. - \lim_{n\to\infty} Y_t^n(\boldsymbol{\theta_0}) = Z_t \tag{5e}$$

$$n\mathbb{V}ar[U_t^n(\boldsymbol{\theta_0})] = \frac{1}{2}(1+\varphi^R(2t,\boldsymbol{\theta_0})) - \varphi^R(t,\boldsymbol{\theta_0})^2 \tag{5f}$$

$$n\mathbb{V}ar[V_t^n(\boldsymbol{\theta_0})] = \frac{1}{2}(1 - \varphi^R(2t, \boldsymbol{\theta_0})) - \varphi^I(t, \boldsymbol{\theta_0})^2 \tag{5g}$$

$$n\mathbb{E}[U_t^n(\boldsymbol{\theta_0})V_t^n(\boldsymbol{\theta_0})] = \frac{1}{2}(\varphi^i(2t, \boldsymbol{\theta_0}) - 1) - \varphi^R(t, \boldsymbol{\theta_0})\varphi^I(t, \boldsymbol{\theta_0}) \tag{5h}$$

$$\mathbb{V}ar[Y_t^n(\boldsymbol{\theta_0})] = 1 - |\varphi(t, \boldsymbol{\theta_0})|^2 \tag{5i}$$

$$\frac{\partial U_t^n}{\partial \boldsymbol{\theta}} = \frac{\partial \varphi^R(t)}{\partial \boldsymbol{\theta}}, \frac{\partial V_t^n}{\partial \boldsymbol{\theta}} = \frac{\partial \varphi^I(t)}{\partial \boldsymbol{\theta}} \tag{5j}$$

2.7. **Consistency of BB Estimator.** We denote our estimator by $\widehat{\boldsymbol{\theta}} = \underset{\theta}{\text{argmin}}\ J^n(\boldsymbol{\theta})$, with $J$ as in 4 from now onwards, and we shall refer to it as the BB estimator. To simplify our proofs parametrization will involve only one variable. The generalization to vector $\boldsymbol{\theta}$ will be straightforward. We write: $\frac{1}{n}J^n(\theta) = \int_0^T \eta(t,\theta)dt = \int_0^T \frac{|c_t^n - \varphi(t,\theta)|^2}{1-|\varphi(t,\theta)|^2}dt = \int_0^T \frac{(U_t^n)^2+(V_t^n)^2}{1-|\varphi(t,\theta)|^2}dt$ Note that $\mathbb{E}[\eta(\theta_0)] = \frac{1}{n}$ and $\lim_{n\to\infty} \eta(t, \boldsymbol{\varphi_0}) = 0$ $\mathbb{P}$ a.s. uniformly in $t \in [0, T]$.
We shall need some other conditions which ensure that the integrals we use exist:

**Condition 5.** $\displaystyle\int_0^T \frac{|\frac{\partial\varphi^R}{\partial\boldsymbol{\theta}}|^2 + |\frac{\partial\varphi^I}{\partial\boldsymbol{\theta}}|^2}{1 - |\varphi(t)|^2}dt < \infty.$

**Condition 6.** $\displaystyle\int_0^T \frac{1}{(1 - |\varphi(t)|^2)^2}dt < \infty.$

**Condition 7.** *The usual regularity conditions, allowing the interchange of the integral and the differential operators, hold for integrands used.*

**Theorem 1.** *Under conditions 4, 5 , 6 and 7 the BB estimator in 4 is a strongly consistent estimator of* $\boldsymbol{\theta}$.

*Proof.* Firstly we observe that, assuming continuity of $\varphi$ with respect to $\theta$, $\displaystyle\lim_{n\to\infty}\frac{1}{n}\mathbb{E}[J^n(\theta)] = \int_0^T \frac{|\varphi(t,\theta_0) - \varphi(t,\theta)|^2}{1 - |\varphi(t,\theta)|^2}dt.$ $\mathbb{P}a.s.$ giving us $\displaystyle\lim_{n\to\infty}\frac{1}{n}J(\theta_0) = 0, \mathbb{P}a.s.$, which minimum value is achieved only at $\theta_0$ by the properties of characteristic functions. So that this minimum has to become isolated as $n$ increases. The nature of the functions, whose minima we are chasing, and the above allow us to conclude that the values of $\theta$ giving us the minimum are random variables which have to converge to the value of $\theta$ for which the ultimate limit 0 is achieved. In other words the estimator 4 converges strongly to $\theta_0$. □

2.8. **Asymptotic Distribution of the BB estimator.** We next set about proving the main result of this paper. We set the arguments in the sequel and present the theorem at the end of the section.

To make the notation a little less cumbersome we shall take our vector of parameters $\boldsymbol{\theta}$ as one-dimensional. Generalizing all our results to the multi-dimensional case is elementary.

Applying Taylor's theorem:
$\frac{\partial\eta}{\partial\theta}(\widehat{\theta}) = \frac{\partial\eta}{\partial\theta}(\theta_0) + (\widehat{\theta} - \theta_0)\frac{\partial^2\eta}{\partial^2\theta}(\theta_0 + \lambda(\widehat{\theta} - \theta_0))$ for some $|\lambda| < 1$
Also by the definition of the estimator: $\int_0^T \frac{\partial\eta}{\partial\theta}(\widehat{\theta})dt = 0$
**For the derivations which follow we are evaluating all functions at $\theta = \theta_0$.**

$$\int_0^T \frac{\partial\eta}{\partial\theta}(\theta_0)dt = 2\int_0^T \frac{U_t^n\frac{\partial\varphi^R}{\partial\theta} + V_t^n\frac{\partial\varphi^I}{\partial\theta}}{1-|\varphi(t)|^2}dt + 2\int_0^T \eta\frac{\frac{\partial\varphi^R}{\partial\theta}\varphi^R + \frac{\partial\varphi^I}{\partial\theta}\varphi^I}{1-|\varphi(t)|^2}dt$$

Both integrands in the RHS tend $\mathbb{P}$ almost surely to 0. The first term's asymptotic behaviour is given by:
$\sqrt{n}\int_0^T \frac{U_t^n\frac{\partial\varphi^R}{\partial\theta} + V_t^n\frac{\partial\varphi^I}{\partial\theta}}{1-|\varphi(t)|^2} \to \int_0^T \frac{U_t\frac{\partial\varphi^R}{\partial\theta} + V_t\frac{\partial\varphi^I}{\partial\theta}}{1-|\varphi(t)|^2}$ and it dominates the second term by an order of $n^{1/2}$. Also $\mathbb{E}\left[\frac{\partial\eta}{\partial\theta}\right] = \frac{\frac{\partial\varphi^R}{\partial\theta}\varphi^R + \frac{\partial\varphi^I}{\partial\theta}\varphi^I}{n(1-|\varphi(t)|^2)}$

Furthermore

$$\begin{aligned}
\int_0^T \frac{\partial^2\eta}{\partial\theta^2}(\theta_0)dt = {} & 2\int_0^T \frac{(\frac{\partial\varphi^R}{\partial\theta})^2 + (\frac{\partial\varphi^I}{\partial\theta})^2 + U_t^n\frac{\partial^2\varphi^R}{\partial\theta^2} + V_t^n\frac{\partial^2\varphi^I}{\partial\theta^2}}{1-|\varphi(t)|^2}dt \\
& + 4\int_0^T \Big(\frac{U_t^n\frac{\partial\varphi^R}{\partial\theta} + V_t^n\frac{\partial\varphi^I}{\partial\theta}}{1-|\varphi(t)|^2}\Big)\Big(\frac{\frac{\partial\varphi^R}{\partial\theta}\varphi^R + \frac{\partial\varphi^I}{\partial\theta}\varphi^I}{1-|\varphi(t)|^2}\Big)dt \\
& + 2\int_0^T \frac{\partial\eta}{\partial\theta}\frac{\frac{\partial\varphi^R}{\partial\theta}\varphi^R + \frac{\partial\varphi^I}{\partial\theta}\varphi^I}{1-|\varphi(t)|^2}dt \\
& + 2\int_0^T \eta\frac{\frac{\partial^2\varphi^R}{\partial\theta^2}\varphi^R + \frac{\partial^2\varphi^I}{\partial\theta^2}\varphi^I + (\frac{\partial\varphi^R}{\partial\theta})^2 + (\frac{\partial\varphi^I}{\partial\theta})^2}{1-|\varphi(t)|^2}dt \\
& + 4\int_0^T \eta\frac{\frac{\partial\varphi^R}{\partial\theta}\varphi^R + \frac{\partial\varphi^I}{\partial\theta}\varphi^I}{(1-|\varphi(t)|^2)}\Big(\frac{\frac{\partial\varphi^R}{\partial\theta}\varphi^R + \frac{\partial\varphi^I}{\partial\theta}\varphi^I}{1-|\varphi(t)|^2}\Big)dt
\end{aligned}$$

so that $\underset{n\to\infty}{lim}\mathbb{E}[\int\limits_0^T \frac{\partial^2\eta}{\partial\theta^2}(\theta_0)dt] = \int\limits_0^T \frac{|\varphi(t)|^2}{1-|\varphi(t)|^2}dt$
where all the terms on the right hand side are evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta_0}$.

So going back to the result derived from Taylor's theorem and using the results above and denoting the first term by $K_n$, we have:

$0 = K_n + (\widehat{\theta} - \theta_0)\frac{\partial^2\eta}{\partial^2\theta}(\theta_0 + \lambda(\widehat{\theta} - \theta_0))$ so that $\sqrt{n}(\widehat{\theta} - \theta_0) = \frac{-\sqrt{n}K_n}{\frac{\partial^2\eta}{\partial^2\theta}(\theta_0+\lambda(\widehat{\theta}-\theta_0))}$ Under the regularity assumptions, the denominator tends $\mathbb{P}a.s.$ to $\int_0^T \frac{|\varphi(t)|^2}{1-|\varphi(t)|^2}dt$ while the numerator is dominated by $W = \int_0^T \frac{U_t\frac{\partial\varphi^R}{\partial\theta}+V_t\frac{\partial\varphi^I}{\partial\theta}}{1-|\varphi(t)|^2}dt$ which is the sum of two centred normal random variables with covariance $C = \int_0^T \frac{\frac{1}{2}(\varphi^i(2t)-1)-\varphi^R(t)\varphi^I(t)}{(1-|\varphi(t)|^2)^2}\frac{\partial\varphi^R}{\partial\theta}\frac{\partial\varphi^I}{\partial\theta}dt$ But coming from $U_t^n$ and $V_t^n$ we can use the strong approximations using Brownian bridges we mentioned before.

Vectorizing our parameters, we have random vector $\boldsymbol{W}$ and matrix $\mathbb{C}$:

$$\boldsymbol{W} = \int\limits_0^T \frac{U_t\frac{\partial\varphi^R}{\partial\boldsymbol{\theta}} + V_t\frac{\partial\varphi^I}{\partial\boldsymbol{\theta}}}{1-|\varphi(t)|^2}dt \text{ and } \mathbb{C} = \int\limits_0^T \frac{\frac{1}{2}(\varphi^i(2t)-1)-\varphi^R(t)\varphi^I(t)}{(1-|\varphi(t)|^2)^2}\frac{\partial\varphi^R}{\partial\boldsymbol{\theta}}(\frac{\partial\varphi^I}{\partial\boldsymbol{\theta}})'dt$$

We comment again on the ability to work out numerically to excellent accuracy all the quantities we may require from random vector $\boldsymbol{W}$. The integrand can be simulated through the use of simulated paths from Brownian bridge. The parts needed from the characteristic function can be obtained as the corresponding quantities in $\varphi(t,\widehat{\boldsymbol{\theta}})$. Generating lots of proxy values for this random vector will allow us to approximate its variance, for instance, or obtain values for its distribution function. This same approximation $\varphi(t,\widehat{\boldsymbol{\theta}})$ can give us values for the entries of $\mathbb{C}$.

We state in generality the relevant theorem :

**Theorem 2.** *Given iid sequence $X_1, ...., X_n$, from a distribution with characteristic function $\varphi(t,\boldsymbol{\theta_0})$, and $T > 0$, under assumptions 1, 2, 3, 4 and :*

**Condition 8.** $\frac{\partial^2\varphi^R}{\partial\boldsymbol{\theta}^2}$ *and* $\frac{\partial^2\varphi^I}{\partial\boldsymbol{\theta}^2}$ *are dominated by a Lebesgue integrable functions over* $[0,T]$

*the estimator:* $\hat{\boldsymbol{\theta}} = \underset{\theta}{argmin}\int\limits_0^T \frac{|Y_t^n(\boldsymbol{\theta})|^2}{1-|\varphi(t,\boldsymbol{\theta})|^2}dt$ *is an asymptotically unbiased , consistent estimator of $\boldsymbol{\theta_0}$ for which the random vector $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta_0})$ converges $\mathbb{P}$ almost surely to a centred normally distributed random vector which has the same distribution as random vector $\boldsymbol{W}$ with covariance matrix $\mathbb{C}$.*

A comment at this stage should be made about the efficiency of the estimator under consideration. The rate of convergence given in the theorem above is clearly of the order of maximum likelihood, which asymptotically goes towards the optimal Cramér-Rao bound. We are technically in the same situation here.

## 3. Simulation Studies

Having obtained reassuring resultsabout our estimator, we next present results involving simulations using estimator 4. As a general guide, we tried to compare results from BB with those from maximum likelihood. MLE is the best there is in the business on a number of issues for a wide spectrum of distributions. So the comparison should be a stiff test for the viability of BB. Of primary importance, at this stage of preliminary testing, was the size of bias and of sampling variance. We should also mention the frequency of the data points, which naturally depend on the application, should also somehow come into the picture. Financial time series and climate statistics usually have data with very high frequency. But there are many other applications with more meagre datasets. Here we do just a preliminary exercise to check whether it is worthwhile to work further with BB. The choices of the parameters were not guided by some deep considerations and consequently they should be digested with caution.

We took samples with size varying in the medium range, 100 in steps of 100 to 500. Simulations with 5000 strong sample were also conducted to have a feel for how fast the convergence studied above moves in practice. Having started our discussion from a Lévy context , it makes only sense that we look at infinitely divisible distributions where MLE works well : normal and gamma. Tables 1 and 2 show clearly that as far as bias is concerned it is minimal for both estimators, in many cases the BB estimate being better. The situation with variance as expected is slightly in favour of MLE but not by much and furthermore as the sample dize increases the discrepancy in favour of MLE diminishes.

*Table* 1. Normally Distributed RV's

| True parameters are $\mu = -1.32$ , $\sigma^2 = 3.2$ and $T = 2$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Sample | MLE means of | | BB means of | | MLE variance of | | BB variance of | |
| Size | $\widehat{\mu}$ | $\widehat{\sigma^2}$ | $\widehat{\mu}$ | $\widehat{\sigma^2}$ | $\widehat{\mu}$ | $\widehat{\sigma^2}$ | $\widehat{\mu}$ | $\widehat{\sigma^2}$ |
| 100 | -1.2851 | 3.1813 | -1.2966 | 3.1985 | 0.1150 | 0.0467 | 0.1215 | 0.0724 |
| 200 | -1.3122 | 3.1770 | -1.3150 | 3.1658 | 0.0563 | 0.0315 | 0.0624 | 0.0544 |
| 300 | -1.2789 | 3.1813 | -1.2834 | 3.1949 | 0.0345 | 0.0169 | 0.0393 | 0.0248 |
| 400 | -1.3112 | 3.1722 | -1.3142 | 3.1884 | 0.0217 | 0.0122 | 0.0271 | 0.0168 |
| 500 | -1.3093 | 3.1791 | -1.3038 | 3.1811 | 0.0211 | 0.0129 | 0.0266 | 0.0181 |
| 5000 | -1.2982 | 3.1970 | -1.3016 | 3.2011 | 0.0018 | 0.0013 | 0.0020 | 0.0018 |

*Table* 2. Gamma Distributed RV's

| True parameters are $\alpha = 5.3$ , $\sigma^2 = 4.2$ and $T = 2$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Sample | MLE means of | | BB means of | | MLE variance of | | BB variance of | |
| Size | $\widehat{\alpha}$ | $\widehat{\beta}$ | $\widehat{\alpha}$ | $\widehat{\beta}$ | $\widehat{\alpha}$ | $\widehat{\beta}$ | $\widehat{\alpha}$ | $\widehat{\beta}$ |
| 100 | 5.3403 | 4.2770 | 5.3256 | 4.3063 | 0.6331 | 0.4170 | 0.8207 | 0.5412 |
| 200 | 5.3560 | 4.2033 | 5.3866 | 4.2137 | 0.3008 | 0.1794 | 0.5121 | 0.3207 |
| 300 | 5.3633 | 4.1724 | 5.3125 | 4.2146 | 0.2010 | 0.1326 | 0.2598 | 0.1836 |
| 400 | 5.3461 | 4.1600 | 5.3329 | 4.1793 | 0.1080 | 0.0691 | 0.1820 | 0.1133 |
| 500 | 5.3210 | 4.2144 | 5.3232 | 4.2213 | 0.1006 | 0.0634 | 0.1583 | 0.1001 |
| 5000 | 5.3057 | 4.2012 | 5.3157 | 4.1944 | 0.0129 | 0.0085 | 0.0195 | 0.0133 |

We also repeated the exercise with a stable distribution. The picture is very similar to the one we have just described for the other two distributions, though in this case the passage to the limit is more rough! Again the choice of parameters was casual as these results are preliminary in nature. The comparison here cannot be made with the MLE of course! So we used a method described in Koutrouvelis (1980)[14] to provide us with estimates from the same data for comparative purposes. Results can be seen in Table 3.

*Table* 3. Stable Distributed RV's

| True parameters are $\alpha = 1.3$ , $\beta = 0.2$ , $\gamma = 1.5$ , $\delta = 2.2$and $T = 2$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Sample | Koutrouvelis Method means of | | | | BB means of | | | |
| Size | $\widehat{\alpha}$ | $\widehat{\beta}$ | $\widehat{\gamma}$ | $\widehat{\delta}$ | $\widehat{\alpha}$ | $\widehat{\beta}$ | $\widehat{\gamma}$ | $\widehat{\delta}$ |
| 100 | 1.3134 | 0.1868 | 1.4677 | 2.1716 | 1.3017 | 0.2002 | 1.4800 | 2.2098 |
| 200 | 1.2997 | 0.2039 | 1.4873 | 2.2951 | 1.2746 | 0.2028 | 1.4839 | 5.9524 |
| 300 | 1.2814 | 0.2039 | 1.4913 | 2.3515 | 1.2784 | 0.1844 | 1.4974 | 2.3635 |
| 400 | 1.2932 | 0.2158 | 1.4802 | 2.2910 | 1.2920 | 0.2127 | 1.4877 | 2.3354 |
| 500 | 1.2927 | 0.2121 | 1.4885 | 2.2734 | 1.2846 | 0.2151 | 1.4905 | 2.2883 |
| | Koutrouvelis method variance of | | | | BB variance of | | | |
| | $\widehat{\alpha}$ | $\widehat{\beta}$ | $\widehat{\gamma}$ | $\widehat{\delta}$ | $\widehat{\alpha}$ | $\widehat{\beta}$ | $\widehat{\gamma}$ | $\widehat{\delta}$ |
| 100 | 0.0237 | 0.0719 | 0.0346 | 0.7752 | 0.0223 | 0.0791 | 0.0311 | 0.9740 |
| 200 | 0.0139 | 0.0461 | 0.0175 | 0.6410 | 0.0146 | 0.0437 | 0.0167 | 0.7981 |
| 300 | 0.0077 | 0.0260 | 0.0104 | 0.3785 | 0.0106 | 0.0236 | 0.0112 | 0.9468 |
| 400 | 0.0059 | 0.0209 | 0.0071 | 0.1878 | 0.0069 | 0.0202 | 0.0077 | 0.2971 |
| 500 | 0.0044 | 0.0119 | 0.0070 | 0.1416 | 0.0058 | 0.0173 | 0.0071 | 0.1956 |

## 4. Conclusion

Starting from a literature review of clever ecf uses in estimation problems for Lévy processes, one could well have a look at the integrated squared error method with two ideas in mind:

- The Brownian bridge approximation to the empirical characteristic functiom can be put to use more effectively.
- particular features of the type of characteristic function at hand could be incorporated suitably in the function whose mimnimum gives us the estimator

This strategy has worked well with our choice of estimator. The BB estimator has a variance-proxy term built out of the characteristic function embedded within the error function. Results obtained theoretically for this estimator give us an asymptotic behaviour close to that of the maximum likelihood. A few preliminary exercises using simulated data also gave promising results. More work needs to be done with the latter numerical efforts. Moreover, the ideas can be extended and particularized to specific distributions and Lévy process contexts so that more efficient and numerically stable methods can be devised.

## References

1. Belomestny, D. 2010. Spectral Estimation of the fractional order of a Lévy process. *Ann. Statist.*, 38, No. 1, pp. 317–351
2. Carrasco, M. and Chernov, Mikhail and Florens, Jean-Pierre and Ghysels, E. 2007. Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of econometrics*, 140 (2), pp. 529–573
3. Carrasco, M. and Kotchoni, R. 2010. Efficient Estimation using the Characteristic Function. *Working paper.* Université de Montréal.
4. Csörgö, S. 1981a. *Strong Approximations in Probability and Statistics. Academic Press.*
5. Csörgö, S. 1981b. Limit behavior of the empirical distribution function. *Annals of Probability*, Vol. 9, No. 1, pp. 130–144.
6. Csörgö, M. 1983. Quantile Processes with Statistical Applications. *SIAM, Philadelphia.*
7. Feuerverger, A., Mureika, R. A. 1977. The empirical characteristic function and its applications. *Ann. Statist.*, 5, pp. 88–97.
8. Feuerverger, A., McDunnough, P. 1981a. On efficient inference in symmetric stable laws and processes, in *Statistics and Related Topics. New York: North Holland*, pp. 109–122.
9. Feuerverger, A., McDunnough, P. 1981b. On some Fourier methods for inference. *J. Am. Statist. Assoc.*, 76, pp. 379–387.
10. Heathcote, C. R. 1977. The integrated square error estimation of parameters. *Biometrika.*, 64 (2), pp. 255–64.
11. Kent, J. 1975. A Weak Convergence theorem for the empirical characteristic function. *J. Appl. Probability*, 12, pp. 512–523.

12. Kogon, S. M. and Williams, D.B. 1998. Characteristic function based estimation of stable distribution parameters, *A Practical Guide to Heavy Tails*. Birkhauser, pp. 311–325.
13. Komlos, J., Major, P., and Tusnady, G. 1975. An approximation of partial sums of independent random variables and the sample distribution function. *Z. Wahrsch. Verw. Gebeite*, 32, pp. 111–131.
14. Koutrovelis, I.A. 1980. Regression-type estimation of the parameters of stable laws. *J. Am. Statist. Assoc.*, 75, pp. 918–928.
15. Marcus, M.M. 1977. Weak convergence of the characteristic function. *The Annals of Probability*, Vol. 9, No. 2, pp. 194–201.
16. Parzen, E. .1962. On estimation of a probability density function and mode. *Ann. Math. Statist.*, 33, pp. 1065–1076
17. Paulson, A.S., Holcomb, E.W., and Leitch, A.H. 1977. The integrated square error estimation of parameters. *Biometrika.*, 62, pp. 163–70.
18. Press, S.J. 1972. Estimation in univariate and multivariate stable distributions. *J. Am. Statist. Assoc.*, 67, pp. 842–846
19. Yu, J. 2004. Empirical Characteristic function estimation and its applications. *Econometric Reviews*, Vol. 23, No. 2, pp. 93–123.

# CONVERGENCE CONDITIONS OF THE TWO-PARAMETRIC SECANT TYPE METHOD FOR SOLVING NONLINEAR EQUATIONS TAKING INTO ACCOUNT ERRORS

© **Stepan Shakhno, Halyna Yarmola**

Ivan Franko National University of Lviv

e-mail: *shakhno@is.lviv.ua*, *halina_yarmola@ukr.net*

***Abstract.*** *In this paper we analyze the stability of the two-parametric secant type method to errors calculations for solving nonlinear equations and estimate the total error.*

## 1. Introduction

We consider the equation

$$F(x) = 0, \tag{1}$$

where $F$ is a nonlinear operator defined on a convex subset $D$ of a Banach space $X$ with values in a Banach space $Y$. Studying solving methods for the equation (1) does not always take into account all the errors that arise during equation solving with the help of numerical methods. These issues were researched by certain authors. The stability and the error perturbation of the Newton-Kantorovich method and its modification are investigated in [5]. The evaluation of the total error of the simple iteration method is obtained in the work [3]. The paper [4] studies conditions of convergence and evaluation of the total error of the two-step iterative-differential method. The stability analysis of the accelerated Newton method to calculation's errors is carried out in [8].

In this paper we investigate convergence conditions of the two-parametric secant type method with regard to the rounding errors. The two-parametric secant type method, proposed in [6], has the form

$$x_{k+1} = x_k - [F(u_k, v_k)]^{-1} F(x_k),\ k = 0,\ 1,\ 2, \ldots, \tag{2}$$

where $F(u_k, v_k)$ is divided difference of the first order of the operator $F$ at the points $u_k$ and $v_k$, $u_k = x_k + a_k(x_{k-1} - x_k)$, $v_k = x_k + b_k(x_{k-1} - x_k)$, $a_k \in [-1, 1]$, $b_k \in [0, 1]$. In the work [7] the semilocal convergence of the method (2) is examined.

**Definition 1.** Let $F$ be a nonlinear operator defined on a subset $D$ of a linear space $X$ with values in a linear space $Y$ and let $x,\ y$ be two points of $D$. A linear operator from $X$ into $Y$, denoted as $F(x, y)$, which satisfies the condition

$$F(x, y)(x - y) = F(x) - F(y).$$

is called a divided difference of $F$ at the points $x$ and $y$.

## 2. Convergence conditions of the perturbed method

Let us assume that the divided difference $F(x,y)$ is calculated with an error, the operator $F$ is calculated exactly. Let us consider the perturbed iterative process

$$x_{k+1} = x_k - [F(u_k, v_k) + \Gamma_k]^{-1} F(x_k),\ k = 0,\ 1,\ 2, \ldots. \tag{3}$$

Here $\{\Gamma_k\} \in L(X,Y)$ is a sequence of linear operators. For the iterative process (3) the following theorem is valid.

**Theorem 1.** *Let $x_{-1}, x_0 \in D$ be initial approximations, $S_0 = \{x \in D : \|x - x_0\| < R\}$. Assume that the following conditions hold*

*1) $\|x_{-1} - x_0\| = \alpha$;*

*2) there exist $A_0^{-1} = [F(u_0, v_0)]^{-1}$ and $\left\|A_0^{-1}\right\| \le \beta_0$;*

*3) $\|F(x_0)\| \le \zeta_0$, $\eta_0 = \beta_0 \zeta_0$;*

*4) $\|\Gamma_k\| \le \mu \eta_k$, $\beta_0 \mu \eta_0 < 1$, $k = 0,\ 1,\ 2,\ \ldots$, where $\{\eta_k\}$ is a numerical sequence;*

*5) divided differences of the first order of the operator $F$ satisfy Lipschitz condition*

$$\|F(x,y) - F(u,v)\| \le L(\|x-u\| + \|y-v\|), \quad x,\ y,\ u,\ v \in D, \quad L > 0$$

*Let us denote*

$$m = \beta_0 L \max\left\{\frac{\eta_0}{1-\beta_0\mu\eta_0} + (a+b)\alpha,\ (1+a+b)\frac{\eta_0}{1-\beta_0\mu\eta_0}\right\} + \beta_0\mu\eta_0,$$

*suppose that $|a_k| \le a$, $b_k \le b$ and the equation*

$$u\left(1 - \frac{m}{1-\beta_0 L((2+a+b)u + (a+b)\alpha) - \beta_0\mu u}\right) - \frac{\eta_0}{1-\beta_0\mu\eta_0} = 0 \tag{4}$$

*has at least one positive zero, let $R$ be the minimum positive one.*

*If $\beta_0 L((2+a+b)R + (a+b)\alpha) + \beta_0\mu R < 1$,*

$$M = \frac{m}{1-\beta_0 L((2+a+b)R + (a+b)\alpha) - \beta_0\mu R} < 1$$

*and $\bar{S}_0 \subset D$, then the sequence $\{x_k\}$, given by the iterative process (3) is well defined, remains in $\bar{S}_0$ and converges to a unique solution $x_* \in \bar{S}_0$ of the equation (1). Moreover, the following inequality holds*

$$\|x_k - x_*\| < \frac{1}{h(1-M)} M^{\Phi_k}, \tag{5}$$

*where* $h = \dfrac{\beta_0\Big[L(1+a+b)+\mu\Big]}{1-\beta_0 L((2+a+b)R + (a+b)\alpha) - \beta_0\mu R}$, $\Phi_{-1} = 0,\ \Phi_0 = 1,$
$\Phi_k = \Phi_{k-1} + \Phi_{k-2},\ k = 1,\ 2,\ \ldots$

*Proof.* Let us denote $A_k = F(u_k, v_k)$. By (3), we have

$$x_1 = x_0 - [A_0 + \Gamma_0]^{-1} F(x_0) = x_0 - [A_0(I + A_0^{-1}\Gamma_0)]^{-1} F(x_0) = \\ = x_0 - [I + A_0^{-1}\Gamma_0]^{-1} A_0^{-1} F(x_0).$$

Since $\|[I + A_0^{-1}\Gamma_0]^{-1}\| \leq \dfrac{1}{1 - \|A_0^{-1}\|\|\Gamma_0\|}$, then, taking into account the theorem's conditions, we get

$$\|x_1 - x_0\| = \left\|-[I + A_0^{-1}\Gamma_0]^{-1} A_0^{-1} F(x_0)\right\| \leq \frac{\left\|A_0^{-1}\right\| \|F(x_0)\|}{1 - \|A_0^{-1}\|\|\Gamma_0\|} \leq \\ \leq \frac{\eta_0}{1 - \beta_0\mu\eta_0} = \frac{1}{h}\Big(h\frac{\eta_0}{1 - \beta_0\mu\eta_0}\Big)^{\Phi_0} < R.$$

So, $x_1 \in S_0$.

Using the condition 5) of the theorem, we obtain

$$\left\|I - A_0^{-1}A_1\right\| \leq \left\|A_0^{-1}\right\| \|A_0 - A_1\| \leq \beta_0 L\left(\|u_0 - u_1\| + \|v_0 - v_1\|\right).$$

Since

$$\|u_0 - u_k\| = \|x_0 + a_0(x_{-1} - x_0) - x_k - a_k(x_{k-1} - x_k)\| \leq \\ \leq \|x_0 - x_k\| + |a_0|\,\|x_{-1} - x_0\| + |a_k|\,\|x_{k-1} - x_k\|,$$

$$\|v_0 - v_k\| = \|x_0 + b_0(x_{-1} - x_0) - x_k - b_k(x_{k-1} - x_k)\| \leq \\ \leq \|x_0 - x_k\| + b_0\,\|x_{-1} - x_0\| + b_k\,\|x_{k-1} - x_k\|$$

and $|a_k| \leq a$, $b_k \leq b$, then

$$\left\|I - A_0^{-1}A_1\right\| \leq \beta_0 L\left((2 + a + b)\|x_0 - x_1\| + (a + b)\|x_{-1} - x_0\|\right) \leq \\ \leq \beta_0 L\left[(2 + a + b)\frac{\eta_0}{1 - \beta_0\mu\eta_0} + (a + b)\alpha\right] < \beta_0 L\Big[(2 + a + b)R + (a + b)\alpha\Big] < 1.$$

By the Banach lemma, $A_1^{-1}$ exists and

$$\left\|A_1^{-1}\right\| < \frac{\beta_0}{1 - \beta_0 L\left((2 + a + b)R + (a + b)\alpha\right)}.$$

Let us denote:

$$M_{k-1} = \frac{\beta_0\Big[L\,\|x_k - x_{k-1}\| + L(a + b)\,\|x_{k-1} - x_{k-2}\| + \mu\eta_{k-1}\Big]}{1 - \beta_0 L\Big[(2 + a + b)R + (a + b)\alpha\Big]},\ k \geq 1,$$

$$C = \frac{1 - \beta_0 L\Big[(2 + a + b)R + (a + b)\alpha\Big]}{1 - \beta_0 L\Big[(2 + a + b)R + (a + b)\alpha\Big] - \beta_0\mu R},\ C_0 = \frac{1}{1 - \beta_0\mu\eta_0}.$$

It can be easily seen that $C_0M_0 \le M$ and $CM_k \le M$, $k \ge 1$.

From the definition of the first divided difference and (3) we can obtain

$$F(x_1) = F(x_0) - F(x_0, x_1)(x_0 - x_1) = (A_0 + \Gamma_0 - F(x_0, x_1))(x_0 - x_1).$$

Taking into account the theorem's condition 4) and Lipschitz condition 5), we get

$$\begin{aligned} \|F(x_1)\| &\le \Big[\|A_0 - F(x_0, x_1)\| + \|\Gamma_0\|\Big] \|x_1 - x_0\| \le \\ &\le \Big[L(\|u_0 - x_0\| + \|v_0 - x_1\|) + \mu\eta_0\Big] \|x_1 - x_0\| \le \\ &\le \Big[L((a+b)\|x_0 - x_{-1}\| + \|x_1 - x_0\|) + \mu\eta_0\Big] \|x_1 - x_0\|. \end{aligned}$$

Thus,

$$\left\|A_1^{-1}\right\| \|F(x_1)\| < \frac{\beta_0 (L(a+b)\|x_0 - x_{-1}\| + L\|x_1 - x_0\| + \mu\eta_0)}{1 - \beta_0 L((2+a+b)R + (a+b)\alpha)} \frac{\eta_0}{1 - \beta_0\mu\eta_0} =$$
$$= M_0 C_0 \eta_0 = \eta_1.$$

Let us show that $\eta_1 < \eta_0$. In fact,

$$\begin{aligned} \eta_1 &\le \frac{\beta_0 L\left((a+b)\alpha + \dfrac{\eta_0}{1 - \beta_0\mu\eta_0}\right) + \beta_0\mu\eta_0}{(1 - \beta_0 L((2+a+b)R + (a+b)\alpha))(1 - \beta_0\mu\eta_0)} \eta_0 \le \\ &\le \frac{m\eta_0}{1 - \beta_0 L((2+a+b)R + (a+b)\alpha) - \beta_0\mu\eta_0} \le \\ &\le \frac{m\eta_0}{1 - \beta_0 L((2+a+b)R + (a+b)\alpha) - \beta_0\mu R} = M\eta_0 < \eta_0. \end{aligned}$$

Above this we have

$$\eta_1 = \frac{CM_0}{C} \frac{\eta_0}{1 - \beta_0\mu\eta_0} < \frac{M}{C} \frac{\eta_0}{1 - \beta_0\mu\eta_0} < \frac{1}{Ch}\left(h \frac{\eta_0}{1 - \beta_0\mu\eta_0}\right)^{\Phi_1}.$$

Therefore $x_2$ is well defined and

$$\begin{aligned} \|x_2 - x_1\| &\le \left\|-[I + A_1^{-1}\Gamma_1]^{-1}\right\| \left\|A_1^{-1}\right\| \|F(x_1)\| \le \\ &\le \frac{\left\|A_1^{-1}\right\| \|F(x_1)\|}{1 - \|A_1^{-1}\|\|\Gamma_1\|} < C\eta_1 < \frac{1}{h}\left(h \frac{\eta_0}{1 - \beta_0\mu\eta_0}\right)^{\Phi_1}. \end{aligned}$$

In addition, $\|x_2 - x_1\| < M \dfrac{\eta_0}{1 - \beta_0\mu\eta_0}$. Since $R$ is a solution of the equation (4), then

$$\|x_2 - x_0\| \le \|x_2 - x_1\| + \|x_1 - x_0\| < (M+1)\frac{\eta_0}{1 - \beta_0\mu\eta_0} < R$$

and $x_2 \in S_0$.

Let us suppose that the following conditions are valid for $i = \overline{2, k-1}$:

- linear operators $A_i$ are invertible,
- $\left\|A_i^{-1}\right\| \|F(x_i)\| < \eta_i = M_{i-1} C \eta_{i-1} < \frac{1}{Ch}\Big(h\frac{\eta_0}{1-\beta_0\mu\eta_0}\Big)^{\Phi_i}$, $\eta_i < \eta_{i-1}$,
- $\|x_{i+1} - x_i\| < C\eta_i < \frac{1}{h}\Big(h\frac{\eta_0}{1-\beta_0\mu\eta_0}\Big)^{\Phi_i} \leq \frac{1}{h}M^{\Phi_i}$ and $x_{i+1} \in S_0$.

Then, for $i = k$ we also obtain

$$\left\|I - A_0^{-1}A_k\right\| \leq \left\|A_0^{-1}\right\| \|A_0 - A_k\| \leq \beta_0 L \left(\|u_0 - u_k\| + \|v_0 - v_k\|\right) \leq$$
$$\leq \beta_0 L\left[(2+a+b)\frac{\eta_0}{1-\beta_0\mu\eta_0} + (a+b)\,\alpha\right] < \beta_0 L\left[(2+a+b)\,R + (a+b)\,\alpha\right] < 1$$

and $\left\|A_k^{-1}\right\| < \dfrac{\beta_0}{1-\beta_0 L\left((2+a+b)\,R + (a+b)\,\alpha\right)}$.

From the definition of the first divided difference and (3) we can obtain

$$F(x_k) = F(x_{k-1}) - F(x_{k-1}, x_k)(x_{k-1} - x_k) =$$
$$= \left(A_{k-1} + \Gamma_{k-1} - F(x_{k-1}, x_k)\right)(x_{k-1} - x_k).$$

Taking into account the condition 5) of the theorem, we receive the following

$$\|F(x_k)\| = \|\left(A_{k-1} + \Gamma_{k-1} - F(x_{k-1}, x_k)\right)(x_{k-1} - x_k)\| \leq$$
$$\leq \Big[\|A_{k-1} - F(x_{k-1}, x_k)\| + \|\Gamma_{k-1}\|\Big]\|x_k - x_{k-1}\| \leq$$
$$\leq \Big[L\left(\|u_{k-1} - x_{k-1}\| + \|v_{k-1} - x_k\|\right) + \mu\eta_{k-1}\Big]\|x_k - x_{k-1}\| \leq$$
$$\leq \Big[L\|x_k - x_{k-1}\| + L(a+b)\|x_{k-1} - x_{k-2}\| + \mu\eta_{k-1}\Big]\|x_k - x_{k-1}\|.$$

Then

$$\left\|A_k^{-1}\right\| \|F(x_k)\| < \frac{\beta_0\Big[L\|x_k - x_{k-1}\| + L(a+b)\|x_{k-1} - x_{k-2}\| + \mu\eta_{k-1}\Big]C\eta_{k-1}}{1-\beta_0 L\left((2+a+b)\,R + (a+b)\,\alpha\right)} =$$
$$= M_{k-1}C\eta_{k-1} = \eta_k < \frac{\beta_0\Big[L\frac{1}{h}M^{\Phi_{k-1}} + L(a+b)\frac{1}{h}M^{\Phi_{k-2}} + \mu\frac{1}{Ch}M^{\Phi_{k-1}}\Big]\frac{1}{h}M^{\Phi_{k-1}}}{1-\beta_0 L\left((2+a+b)\,R + (a+b)\,\alpha\right)} <$$
$$< \frac{1}{Ch}M^{\Phi_{k-1}+\Phi_{k-2}} = \frac{1}{Ch}M^{\Phi_k}.$$

Since $M_{k-1}C \leq M < 1$, then $\eta_k < \eta_{k-1}$.

Thus,

$$\|x_{k+1} - x_k\| < C\eta_k < \frac{1}{h}M^{\Phi_k}.$$

Obviously that for $i = \overline{1,k}$ $\|x_{i+1} - x_i\| < M^i \dfrac{\eta_0}{1-\beta_0\mu\eta_0}$ is valid. Therefore, taking into account that $R$ is the solution of the equation (4), we get

$$\|x_{k+1} - x_0\| \le \|x_{k+1} - x_k\| + \|x_k - x_{k-1}\| + ... + \|x_1 - x_0\| <$$
$$< (M^k + M^{k-1} + ... + 1)\frac{\eta_0}{1-\beta_0\mu\eta_0} = \frac{1-M^{k+1}}{1-M}\frac{\eta_0}{1-\beta_0\mu\eta_0} < \frac{1}{1-M}\frac{\eta_0}{1-\beta_0\mu\eta_0} = R$$

and $x_{k+1} \in S_0$.

Let us show that $\{x_k\}$ is a Cauchy sequence. In fact,

$$\begin{aligned}
&\|x_{k+p} - x_k\| \le \|x_{k+p} - x_{k+p-1}\| + ... + \|x_{k+1} - x_k\| < \\
&< (M^{p-1} + M^{p-2} + ... + 1)\frac{1}{h}\left(h\frac{\eta_0}{1-\beta_0\mu\eta_0}\right)^{\Phi_k} = \\
&= \frac{1-M^p}{h(1-M)}\left(h\frac{\eta_0}{1-\beta_0\mu\eta_0}\right)^{\Phi_k} < \frac{1}{h(1-M)}\left(h\frac{\eta_0}{1-\beta_0\mu\eta_0}\right)^{\Phi_k}.
\end{aligned} \tag{6}$$

Therefore, $\{x_k\}$ is a Cauchy sequence and converges to $x_* \in \bar{S}_0$.

Now let us prove that $x_*$ is a unique solution of the equation (1). Since

$$\|F(x_k)\| \le \left[L(1+a+b)\frac{\eta_0}{1-\beta_0\mu\eta_0} + \mu\eta_0\right]\|x_k - x_{k-1}\|$$

and $\|x_k - x_{k-1}\| \to 0$ if $k \to \infty$, then $F(x_*) = 0$.

Suppose that there exists $x_{**} \in \bar{S}_0$, $x_{**} \ne x_*$ i $F(x_{**}) = 0$. Let us denote $F(x_{**}, x_*) = H$. From the definition of divided difference of the first order we get

$$H(x_{**} - x_*) = F(x_{**}) - F(x_*).$$

If the operator $H$ is invertible, then $x_{**} = x_*$. Indeed,

$$\left\|A_0^{-1}H - I\right\| = \left\|A_0^{-1}(H - A_0)\right\| \le \left\|A_0^{-1}\right\|\|H - A_0\| \le \beta_0 L\left[\|x_{**} - u_0\| + \|x_* - v_0\|\right] \le$$
$$\le \beta_0 L\left[\|x_{**} - x_0\| + \|x_* - x_0\| + (a+b)\|x_{-1} - x_0\|\right] < \beta_0 L(2R + (a+b)\alpha) < 1.$$

So, the operator $H^{-1}$ exists. From (6) we can obtain the following estimation

$$\|x_k - x_*\| < \frac{1}{h(1-M)}M^{\Phi_k}.$$

□

## 3. Evaluation of the total error of the method (2)

Let us assume that the operator $F$ is calculated approximately, i.e. we have perturbed equation

$$F_\varepsilon(x) = 0. \tag{7}$$

We assume that the operator $F_\varepsilon$ is "close" to operator $F$ in the sense that the following condition is valid

$$\|F_\varepsilon(x) - F(x)\| \le \delta(\varepsilon, x), \tag{8}$$

where $\delta(\varepsilon, x) \to 0$, if $\varepsilon \to 0$, $x \in D$.

Let us apply the method (2) for solving the equation (7)

$$x_{k+1}^\varepsilon = x_k^\varepsilon - [F_\varepsilon(u_k^\varepsilon, v_k^\varepsilon)]^{-1} F_\varepsilon(x_k^\varepsilon),\ k = 0,\ 1,\ 2, \ldots. \tag{9}$$

For the iterative process (9) the following theorem is true.

**Theorem 2.** *Let us suppose*

*1) conditions of Theorem 2 hold for the operator $F_\varepsilon$;*

*2) the equation (1) has at least one solution;*

*3) $\|[F(x,y)]^{-1}\| \le \beta$, for all $x, y \in D$;*

*4) the condition (8) holds.*

*If $k \to \infty$ and $\varepsilon \to 0$ then iterative process (9) converges to the solution $x_* \in \bar{S}_0$ of the equation (1) and the following inequality holds*

$$\|x_k - x_*\| < \frac{1}{h(1-M)} M^{\Phi_k} + \beta\delta(\varepsilon, x). \tag{10}$$

Let us suppose that the divided difference $F(x, y)$ and $F$ are calculated with errors. Then the following iterative process is studied

$$x_{k+1} = x_k - [F(u_k, v_k) + \Gamma_k]^{-1} [F(x_k) + \Psi_k],\ k = 0,\ 1,\ 2, \ldots, \tag{11}$$

where $\{\Gamma_k\} \in L(X, Y)$ is a sequence of linear operators, $\{\Psi_k\} : X \to Y$ is a sequence of operators.

**Theorem 3.** *Let $x_{-1}, x_0 \in D$ be initial approximations, $S_0 = \{x \in D : \|x - x_0\| < R\}$. We assume that the following conditions hold*

*1) $\|x_{-1} - x_0\| = \alpha$;*

*2) there exists $A_0^{-1} = [F(u_0, v_0)]^{-1}$ and $\left\|A_0^{-1}\right\| \le \beta_0$;*

*3) $\|F(x_0)\| \le \zeta_0$, $\eta_0 = \beta_0\zeta_0$;*

*4) $\|\Gamma_k\| \le \mu\eta_k$, $\beta_0\mu\eta_0 < 1$;*

*5) $\|\Psi_0\| \le \gamma\eta_0^2$, $\|\Psi_k\| \le \gamma\eta_k\eta_{k-1}$, $k \ge 1$;*

*6) divided differences of the first order of the operator $F$ satisfy Lipschitz condition*

$$\|F(x,y) - F(u,v)\| \le L(\|x-u\| + \|y-v\|),$$

*where $x,\ y,\ u,\ v \in D,\ L > 0$.*

*Let us denote*

$$m = \beta_0 L \max\left\{\frac{\eta_0(1+\beta_0\gamma\eta_0)}{1-\beta_0\mu\eta_0} + (a+b)\,\alpha, (1+a+b)\,\frac{\eta_0(1+\beta_0\gamma\eta_0)}{1-\beta_0\mu\eta_0}\right\} + \beta_0\mu\eta_0 + \beta_0\gamma\eta_0,$$

$$C^* = \frac{\beta_0\gamma\eta_0}{1-\beta_0 L\left((2+a+b)\,R + (a+b)\,\alpha\right)},$$

*suppose that $|a_k| \le a,\ b_k \le b$ and*

$$u\left(1 - \frac{m(1+C^*)}{1-\beta_0 L\left((2+a+b)\,u + (a+b)\,\alpha\right) - \beta_0\mu u}\right) - \frac{\eta_0(1+\beta_0\gamma\eta_0)}{1-\beta_0\mu\eta_0} = 0 \qquad (12)$$

*has at least one positive zero, let $R$ be the minimum positive one.*

*If $\beta_0 L\left((2+a+b)\,R + (a+b)\,\alpha\right) + \beta_0\mu R < 1$,*

$$M = \frac{m}{1-\beta_0 L\left((2+a+b)\,R + (a+b)\,\alpha\right) - \beta_0\mu R}(1+C^*) < 1$$

*and $\bar{S}_0 \subset D$, then the sequence $\{x_k\}$, given by iterative process (11), is well defined, remains in $\bar{S}_0$ and converges to a unique solution $x_* \in \bar{S}_0$ of the equation (1). Moreover, the following inequality holds*

$$\|x_k - x_*\| < \frac{1}{h(1-M)} M^{\Phi_k},$$

*where* $h = \dfrac{\beta_0\Big(L(1+a+b) + \mu + \gamma\Big)(1+C^*)}{1-\beta_0 L\left((2+a+b)\,R + (a+b)\,\alpha\right) - \beta_0\mu R}$, $\Phi_{-1} = 0,\ \Phi_0 = 1,$
$\Phi_k = \Phi_{k-1} + \Phi_{k-2},\ k = 1, 2, \ldots.$

## Conclusion

In this paper we analyze the stability of the two-parametric secant type method to errors calculations for solving nonlinear equations and estimate the total error. It indicates that the iterative process (2) is resistant to the rounding errors and does not change the convergence order if conditions of Theorems are true.

## References

1. Amat, S. and Busquier, S. 2003. On a higher order Secant method. *Appl. Math. Comp.*, 141, pp. 321–329.
2. Hernandez, M. A. and Rubio, M. J. 2002. The Secant method for nondifferentiable operators. *Appl. Math. Lett.*, 15, pp. 395–399.

3. Babich, M. D. and Ivanov, V. V. 1967. Total error estimate for the solution of a nonlinear operator equation by simple iterations (in Russian). *Zh. Vychisl. Mat. Mat. Fiz.*, 7(5), p. 988–1000.
4. Bartish, M. Ya. and Shcherbina, Yu. N. 1976. Investigation of convergence condition and complete error estimation of an iterative-difference method for the solution of non-linear operator equations (in Russian). *Comput. and Appl. Math.*, 28, p. 3–9.
5. Mikhlin, S. G. 1988. *Some questions of the theory of errors (in Russian).* Leningrad: Leningrad University.
6. Shakhno, S. and Grab, S. and Yarmola, H. 2009. Twoparametric secant type methods for solving nonlinear equations (in Ukrainian). *Visnyk Lviv Univ. Ser. Appl. Math. Comp. Sci.*, 15, pp. 117–127.
7. Shakhno, S. and Yarmola, H. 2011. Application of twoparametric difference method for solving nonlinear integral equations (in Ukrainian). *Visnyk Lviv Univ. Ser. Appl. Math. Comp. Sci.*, 17, pp. 37–46.
8. Shakhno, S. 1988. Construction and investigation of some Newton-Kantorovich type methods for solving nonlinear functional equations. The abstract of PhD thesis in Physics and Mathematics. (in Russian). Kyiv. 17 p.

**Anafiyev, A. and Abdulkhairov, A. 2013. An approach to reconstruct target function of the optimization problem with precedent initial information. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 4–9.**

*The optimization problem with precedent (training sample) initial information is considered. Some approaches for reconstruction of the target function of such optimization problem are proposed. The open problems that must be solved to obtain better quality solutions of this problem are highlighted.*

---

**Beyko, I. and Shchyrba, O. 2013. Optimization problems with partial derivatives and algorithms for constructing generalized solutions. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 10–16.**

*In the paper we define generalized solutions of the optimization problems for control systems with partial derivatives and develop two types of numerical algorithms for calculating the generalized solutions.*

---

**Beyko, I. and Zinko, P. 2013. Solve-operator methods for optimization of risk controlled stochastic processes. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 17–24.**

*In the paper we develop solve-operator methods for high order modelling, simulation and optimization of risk controlled stochastic processes described by general graph-operator control systems with incomplete data.*

---

**Blyshchik, V. 2013. Incompleteness of initial information and the problem of payoff function reconstruction. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 25–29.**

*The paper introduces the classification of informational situations for a zero-sum game with incomplete information based on uncertainty level. For each case the possible ways to deal with uncertainty are considered.*

**Donchenko, V. and Zinko, T. 2013. Solve-operator methods for optimization of risk controlled stochastic processes. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 30–39.**

*Problem of grouping information: recovering function, represented by its observations, and the of classification (problem) clusterization problem, — is of great importance for applied research. Choice of math object which represent the object under investigations largely determines the effectiveness: scalars, vectors or objects of other kinds. Such choice is determined by the richness of mathematical structures within which "representatives" are investigated. Euclidean spaces* $R^n$ *are common in this choosing. Euclidean spaces of* $R^{m \times n}$ *of all* $m \times n$ *matrices are natural as a math structure for "representatives", but the handling technique for such spaces is poorer in comparison with vector space. Just the development of the technique handling" for Euclidean space of* $R^{m \times n}$*, including SVD and Moore-Penrose inversion for the linear operators, constructive construction of orthogonal projectors and grouping operators for matrix spaces is the subject of the article. Important "grouping statements" about minimal ellipsoid, which covers elements of fixed sequence of matrices in* $R^{m \times n}$ *is represented. This statement generalize correspondent results for real valued vectors. "Grouping statements" is proposed to be the base for constructing correspondence distance in solving clusterization problem.*

---

**Kalas, J., Novotný, J, Michalek, J. and Nakonechny, O. 2013. Mathematical model for cancer prevalence and cancer mortality. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 44–54.**

*The first part of the paper designs a deterministic model to describe cancer prevalence and mortality in a population. Next the asymptotic properties of the model are investigated. In the second part, the model is applied to real-world data. For selected model data, a numerical solution is found to the differential equations describing the model, a long-term prediction is made with its results compared with those of predictions made by regression analysis, which are often used to model the prevalence and mortality in the present literature. It is shown that, although for short-term predictions (up to 10 years) both approaches are nearly equivalent, there is a major difference between them if a longer-term prediction is made and finding a reliable prediction for a period longer than 10 years based on short time series seems to be unlikely.*

**Kapustian, O. 2013. Averaging in the optimal control problem for the reaction-diffusion equation with multivalued interaction function.** ***Taurida Journal of Computer Science Theory and Mathematics*****, 2, pp. 55–63.**

*In this paper consider the optimal control problem on infinite time interval with quadratic cost functional. State of this problem is defined by the evolutionary inclusion of reaction-diffusion type. We prove the solvability of such a problem. In the case of rapidly oscillating coefficients in coefficients of differential operator and multivalued interaction function we prove the convergence of $\varepsilon$-dependent optimal process to optimal process of the corresponding averaged problem.*

---

**Krasnoproshin, V. 2013. The mechanisms of decision-making intellectualization based on distributed cognitive resources.** ***Taurida Journal of Computer Science Theory and Mathematics*****, 2, pp. 64–73.**

*This paper describes a comprehensive approach to the problem of intellectualization of decision-making based on a synthesis of expert knowledge and Internet cognitive resources. Models of a global scene and representation of innovative knowledge in the form of subject collections are offered. Algorithms of construction, application, assessment and updating of subject collections are developed. A multi-agent system architecture and implementation option intellectualization in the form of an Internet portal that is used to perform a number of international projects.*

---

**Lukyanova, E. 2013. On similarity of Petri nets languages.** ***Taurida Journal of Computer Science Theory and Mathematics*****, 2, pp. 74–80.**

*The concept of languages similarity of Petri nets is introduced. It is determined, that mapping of languages similarity of Petri nets is a surjective homomorphism. The similarity of languages of component Petri net and original detailed Petri model of the investigated parallel distributed system is considered. The work reveals that the language of the original detailed Petri net model can always be restored using the language of its component model.*

**Maksimov, V. 2013. On realizing prescribed quality of a controlled system's process under uncertainty. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 81–91.**

*In this paper, we discuss a method of auxiliary controlled models and the application of this method to solving some problems of robust control for differential equations. As objects for the approbation of the method, a system of nonlinear differential equations describing some ecological and economic processes is used. A solving algorithm, which is stable with respect to informational noises and computational errors, is presented.*

---

**Nakonechny, O. and Podlipenko, Yu. 2013. Guaranteed estimates of linear functionals on velocity of a viscous incompressible fluid under uncertainties. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 92–102.**

*The creation and justification of the methods for guaranteed estimation of linear functionals from solutions to the boundary value problems for linearized stationary Navier-Stokes equations in bounded open Lipschitzian domains are considered.*

---

**Osadcha, O. and Skripnik, N. 2013. The scheme of partial averaging for one class of hybrid systems. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 103–113.**

*This paper contains the substantiation of the scheme of partial averaging for one class of hybrid systems where one equation is a differential equation with Hukuhara derivative and the other one is an ordinary differential equation.*

---

**Pashko, A. 2013. A simulation of sub-Gaussian random fields on a sphere of orlicz spaces. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 114–123.**

*Estimates for the convergence speed models isotropic random fields on the sphere in the norms of Orlich space. The resulting estimates are used to construct models of random fields on the sphere. Models approximate the random field with given accuracy and reliability.*

**Sant, L. 2013. Harnessing empirical characteristic function convergence behaviour. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 124–136.**

*Parameter estimation for Lévy processes has generated much research effort lately with a strong injection of interest coming from finance. Within this context the problem can be framed as estimation using increments from an infinitely divisible distribution, for which empirical characteristic functions (ecf) are convenient tools. However convergence of ecf's to Gaussian processes has not been exploited as fully as it might have been. In this paper we go back to strong convergence results derived from the Hungarian construction and use Brownian bridge approximations to construct new estimators. In particular we study one integrated square error estimator tailored to show deference to the variance structure of the corresponding Gaussian process. We prove some of its nice statistical properties and present simulation results obtained through its use.*

---

**Shakhno, S. and Yarmola, H. 2013. Convergence conditions of the two-parametric secant type method for solving nonlinear equations taking into account errors. *Taurida Journal of Computer Science Theory and Mathematics*, 2, pp. 137–145.**

*In this paper we analyze the stability of the two-parametric secant type method to errors calculations for solving nonlinear equations and estimate the total error.*

# AUTHORS


***Alim Abdulkhairov*** — Postgraduate student, Department of Computer Science, Taurida National V. I. Vernadsky University
*e-mail: alim.abdulkhairov@gmail.com*

***Ayder Anafiyev*** — Candidate of Physico-Mathematical Sciences, Associate professor, Department of Computer Science, Faculty of Mathematics and Computer Science, Taurida National V. I. Vernadsky University
*e-mail: anafiyev@gmail.com*

***Ivan Beyko*** — Doctor of Engineering Sciences, Professor, Academician of the Academy of Sciences of Higher School of Ukraine
*e-mail: ivan.beyko@gmail.com*

***Vladimir Blyshchik*** — Candidate of Physico-Mathematical Sciences, Associate professor, Department of Computer Science, Faculty of Mathematics and Computer Science, Taurida National V. I. Vernadsky University
*e-mail: veb@land.ru*

***Volodymyr Donchenko*** — Doctor of Physico-Mathematical Sciences, Professor, Department of System Analysis and Decision Making Theory, Faculty of Cybernetics, Taras Shevchenko National University of Kyiv
*e-mail: petro.zinko@gmail.com*

***Oleg Iemets*** — Doctor of Physico-Mathematical Sciences, Professor, Head of the Department of Mathematical Modeling and Social Informatics, Poltava University of Economics and Trade, Poltava, Ukraine
*e-mail: yemetsli@mail.ru*

***Josef Kalas*** — RNDr., CSc. Associate Professor, Department of Mathematics and Statistics, Faculty of Science, Masaryk University, Czech Republic
*e-mail: kalas@math.muni.cz*

***Olena Kapustian*** Candidate of Physico-Mathematical Sciences, Head of the research sector "The Problems of Systems Analysis" of the Cybernetics Faculty, Taras Shevchenko National University of Kyiv
*e-mail: olena.kap@gmail.com*

***Victor Krasnoproshin*** Doctor of Engineering Sciences, Professor, Head of the Department of Management Information Systems, Faculty of Applied Mathematics and Computer Science, Belarusian State University, Minsk, Belarus
*e-mail: krasnoproshin@bsu.by*

***Elena Lukyanova*** Candidate of Physico-Mathematical Sciences, Associate professor, Department of Algebra and Functional Analysis, Taurida National V. I. Vernadsky University
*e-mail: lukyanovaea@mail.ru*

***Vyacheslav Maksimov*** Professor, Ural Federal University, Head of Department "Differential equations", Institute of Mathematics and Mechanics Ural Branch of Russian Academy of Sciences, Ekaterinburg, Russia
*e-mail: maksimov@imm.uran.ru*

***Jaroslav Michalek*** RNDr, CSc, Associate Professor, Institute of Mathematics, Faculty of Mechanical Engineering, Brno University of Technology, Brno, Czech Republic
*e-mail: michalek@fme.vutbr.cz*

***Olexander Nakonechny*** Doctor of Physico-Mathematical Sciences, Professor, Head of the Department of System Analysis and Decision Making Theory, Faculty of Cybernetics, Taras Shevchenko National University of Kyiv
*e-mail: a.nakonechniy@gmail.com*

***Jan Novotný*** PhD. student in Applied Mathematics-Optimization. Institute of Mathematics, Faculty of Mechanical Engineering, Brno University of Technology, Brno, Czech Republic
*e-mail: iannovot@gmail.com*

**Olga Osadcha** — Postgraduate student of the Department of Optimal Control and Economic Cybernetics, I. I. Mechnikov Odessa National University
*e-mail: olga.osadcha.ua@gmail.com*

**Anatoliy Pashko** — Candidate of Physico-Mathematical Sciences, Associate professor, Department of Information System, Faculty of Cybernetics, Taras Shevchenko National University of Kyiv
*e-mail: pashkoua@mail.ru*

**Yuri Podlipenko** — Doctor of Physico-Mathematical Sciences, Professor, Leading Researcher of Research Sector "Problems of System Analysis", Faculty of Cybernetics, Taras Shevchenko National University of Kyiv
*e-mail: yourip@mail.ru*

**Lino Sant** — Professor, Department of Statistics and OR, Faculty of Science, University of Malta
*e-mail: lino.sant@um.edu.mt*

**Stepan Shakhno** — Doctor of Physico-Mathematical Sciences, Associate professor, Ivan Franko National University of Lviv
*e-mail: shakhno@is.lviv.ua*

**Olesya Shchyrba** — Postgraduate student of the National Technical University of Ukraine "Kiev Politechnic Institute"

**Natalia Skripnik** — Candidate of Physico-Mathematical Sciences, Associate professor, Department of Optimal Control and Economic Cybernetics, I. I. Mechnikov Odessa National University
*e-mail: natalia.skripnik@gmail.com*

**Halyna Yarmola** — PhD student, Ivan Franko National University of Lviv
*e-mail: halina_ yarmola@ukr.net*

**Oleksandra Yemets'** — Candidate of Physico-Mathematical Sciences, Associate professor, Department of Mathematical Modeling and Social Informatics, Poltava University of Economics and Trade, Poltava, Ukraine
*e-mail: yemets2008@ukr.net*

**Petr Zinko** Candidate of Physico-Mathematical Sciences, Associate professor, Department of System Analysis and Decision Making Theory, Faculty of Cybernetics, Taras Shevchenko National University of Kyiv
*e-mail: petro.zinko@gmail.com*

**Taras Zinko** Junior research assistant of the research sector "The Problems of Systems Analysis" of the Cybernetics Faculty, Taras Shevchenko National University of Kyiv
*e-mail: tzinko@ukr.net*

*Taurida Journal of Computer Science Theory and Mathematics*, 2013, 2. ISSN 1729-3901
Printed by the Crimean Scientific Center Press