Institute of Information and Communication Technologies
Bulgarian Academy of Sciences

Advanced Computing for Innovation
FP7 Capacity Programme
Research Potential of Convergence Regions

ACOSA Advanced Control and Optimization: Step Ahead | May 8–10, 2014, Bankya, Bulgaria

INTERNATIONAL WORKSHOP ON

ADVANCED CONTROL AND OPTIMIZATION: STEP AHEAD '2014

PROCEEDINGS

Prof. Marin Drinov Academic Publishing House

ISSN 1314-4634

Institute of Information and Communication Technologies

Advanced Computing for Innovation, Grant Agreement: 316087, Funding: FP7 Capacity Programme, Research Potential of Convergence Regions

Advanced Control and Optimization: Step Ahead | May 8–10, 2014, Bankya, Bulgaria

INTERNATIONAL WORKSHOP ON

ADVANCED CONTROL AND OPTIMIZATION: STEP AHEAD' 2014

PROCEESINGS

Prof. Marin Drinov Academic Publishing House

Institute of Information and Communication Technologies
Bulgarian Academy of Sciences

International Workshop on

Advanced Control and Optimisation: Step Ahead

ACOSA

8-10 May 2014

Hotel "Bankya Palace", Bankya, Bulgaria

Supported by FP7 Grant "AcomIn" No. 316087

Programme Committee of ACOSA'2014

Chairs

Acad. Vassil Sgurev (BG)
Acad. Janusz Kacprzyk (PL)

Program Committee

Acad. Ivan Popchev (BG)
Corr. Member Mincho Hadjiski (BG)
Prof. Anthony Shannon (AUS)
Prof. Eulalia Szmidt (PL)
Prof. František Čapkovič (SK)
Prof. Jacques Richalet (FR)
Prof. Panagiotis Chountas (UK)
Prof. Taekyun Kim (KOR)
Prof. Maciek Krawczak (PL)
Prof. Franziska Klügl (SE)
Prof. Dimitar Karastoyanov (BG)

Organising Committee

Assoc. Prof. Lyubka Doukovska (BG) – Chair
Dr. Vassia Atanassova (BG) – Co-Chair

Members

Assoc. Prof. Vladimir Monov (BG)
Assoc. Prof. Vera Angelova (BG)
Assist. Prof. Irina Radeva (BG)
Assist. Prof. Yancho Todorov (BG)
Assist. Prof. Stefan Koynov (BG)

Proceedings of ACOSA'2014

I. Contemporary methods for assessment and control of complex systems

II. Application aspects of the analysis of the contemporary methods for assessment, control and optimisation of complex processes

III. Applications of the contemporary optimization procedures in the financial and industrial structures

IMPROVED RESIDUAL BOUND OF THE MATRIX EQUATION $X + A_2^{\mathrm{H}} X^{-1} A_2 = A_1$

Ivan P. Popchev, Vera A. Angelova
Institute of Information and Communication Technologies - BAS,
Akad. G. Bonchev Str., Bl. 2, Sofia 1113, Bulgaria
ipochev@iit.bas.bg, vangelova@iit.bas.bg

Keywords: Perturbation analysis, Residual bound, Non-linear matrix equation

Abstract: Consider the nonlinear matrix equation $X + A_2^{\mathrm{H}} X^{-1} A_2 = A_1$, with $n \times n$ complex matrices A_1 and A_2. The matrix A_1 is Hermitian positive definite. A_2^{H} denotes the conjugate transpose of A_2. Improved residual bound for the error in the computed solution X is proposed. The bound is simple and of practical use to assess the accuracy of the approximate solution obtained by an iterative algorithm. The effectiveness of the bound proposed is compared to other known in the literature bound.

1 INTRODUCTION

We consider the nonlinear complex matrix equation

$$X + A_2^{\mathrm{H}} X^{-1} A_2 = A_1, \tag{1}$$

where A_1 is a Hermitian positive definite $n \times n$ complex matrix, A_2 is a complex matrix of order n and A_2^{H} is its complex conjugate transpose and X is the Hermitian positive definite solution of equation (1). Equation (1) finds wide applications in various areas as control theory, dynamic programming, statistics, ladder network, stochastic filtering. This determines the interest of many authors to equation (1). Necessary and sufficient conditions, as well as iterative methods to obtain its the Hermitian positive definite solution X are considered in (Engwerda et al., 1993; Meini, 2001). Perturbation analysis of the solution is made in (Xu, 2001; Konstantinov et al., 2002; Sun and Xu, 2003; Hasanov, 2010). Comparison of some perturbation bounds of the solution of equation (1) is made in (Popchev and Angelova, 2010). Complete perturbation analysis of the more general case $A_0 + \sum_{i=1}^{k} \sigma_i A_i^{\mathrm{H}} X^{p_i} A_i = 0$, $\sigma_i = \pm 1$, $p_i \in \mathbb{R}$ can be find in (Konstantinov et al., 2011; Popchev et al., 2014).

Assume that the necessary and sufficient conditions (Engwerda et al., 1993; Ferrante and Levy, 1996) for the existence of a positive definite solution of equation (1) are fulfilled. In this paper we derive a simple, effective and easy computable norm-wise non-local residual bound for the computed by an iterative algorithm approximate solution to (1). The residual bound can be used to assess the accuracy of the approximate solution in an iterative procedure.

Throughout the paper, $\mathbb{C}^{n\times n}$ is the set of $n \times n$ complex matrices; $\mathbb{R}$ is the set of real numbers; $A^{\top}$ is the transpose of A; $A \otimes B = (a_{ij}B)$ is the Kronecker product of A and B; $\mathrm{vec}(A) = [a_1^{\top}, a_2^{\top}, \ldots, a_n^{\top}]^{\top}$ is the vector representation of the matrix A, where $A = [a_{ij}]$ and $a_1, a_2, \ldots, a_n \in \mathbb{C}^n$ are the columns of A; $\|\cdot\|_2$ and $\|\cdot\|_{\mathrm{F}}$ are the spectral and the Frobenius matrix norms, respectively, $\|\cdot\|$ is a unitary invariant norm such as the spectral norm $\|\cdot\|_2$ or the Frobenius norm $\|\cdot\|_{\mathrm{F}}$. The notation ':=' stands for 'equal by definition'.

The paper is organized as follows. The problem is stated in Section 2. In Section 3 a computable residual bound expressed in terms of the computed approximate solution to equation (1) is obtained using the techniques of Lyapunov majorants and fixed point principles. In Section 4 the effectiveness of the bound proposed is compared to the bound from (Xu, 2001) on the base of the numerical Example 4.3 from (Xu, 2001).

2 STATEMENT OF THE PROBLEM

Let $\hat{X} = X + \delta X$ be the approximate Hermitian positive definite solution X of (1) obtained by some iterative algorithm, and let

$$R(\hat{X}) := \hat{X} + A_2^{\mathrm{H}} \hat{X}^{-1} A_2 - A_1 \tag{2}$$

be the residual of (1) with respect to $\hat{X}$. The term δX, $\|\delta X\|_{\rm F} \le \varepsilon \|X\|_2$, reflects the presence of round-off errors and errors of approximation in the solution $\hat{X}$ computed in environment with machine precision ε.

We apply the matrix inversion lemma on (2)

$$\begin{aligned}(X+\delta X)^{-1} &= X^{-1} - X^{-1}\delta X \hat{X}^{-1} \\ &= X^{-1} - (\hat{X}-\delta X)^{-1}\delta X \hat{X}^{-1},\end{aligned}$$

and we obtain an expression for the error δX in the calculated approximate solution $\hat{X}$

$$\delta X = R(\hat{X}) + A_2^{\rm H}(\hat{X}-\delta X)^{-1}\delta X \hat{X}^{-1} A_2. \quad (3)$$

Estimating norm-wisely the error δX, we will obtain norm-wise non-local residual bound for the approximate solution $\hat{X}$ to equation (1).

3 RESIDUAL BOUND

Written in a vector form equation (3) remains

$$\begin{aligned}\mathrm{vec}(\delta X) &= \mathrm{vec}(R(\hat{X})) \quad (4)\\ &+ (A_2^{\top}\hat{X}^{-1} \otimes A_2^{\rm H})\mathrm{vec}((\hat{X}-\delta X)^{-1}\delta X).\end{aligned}$$

Taking the spectral norm of both sides of (4), we obtain

$$\begin{aligned}\|\delta X\|_{\rm F} &\le \|R(\hat{X})\|_{\rm F} \quad (5)\\ &+ \|A_2^{\top}\hat{X}^{-1} \otimes A_2^{\rm H}\|_2 \|(\hat{X}-\delta X)^{-1}\|_2 \|\delta X\|_{\rm F}.\end{aligned}$$

Suppose that $\|\delta X\|_{\rm F} \le \frac{1}{\|\hat{X}^{-1}\|_2}$, which is obvious, because of the nature of δX. For $\|(\hat{X}-\delta X)^{-1}\|_2$ it follows that

$$\|(\hat{X}-\delta X)^{-1}\|_2 \le \frac{\|\hat{X}^{-1}\|_2}{1-\|\hat{X}^{-1}\|_2\|\delta X\|_{\rm F}} \quad (6)$$

Replacing (6) in (5) we obtain

$$\begin{aligned}\|\delta X\|_{\rm F} &\le \|R(\hat{X})\|_{\rm F} \quad (7)\\ &+ \frac{\|A_2^{\top}\hat{X}^{-1} \otimes A_2^{\rm H}\|_2 \|\hat{X}^{-1}\|_2}{1-\|\hat{X}^{-1}\|_2\|\delta X\|_{\rm F}}\|\delta X\|_{\rm F}.\end{aligned}$$

Denote by $a := \|A_2^{\top}\hat{X}^{-1} \otimes A_2^{\rm H}\|_2\|\hat{X}^{-1}\|_2$, $\chi := \|\hat{X}^{-1}\|_2$, $r := \|R(\hat{X})\|_{\rm F}$. For equation (7) we obtain

$$\begin{aligned}\|\delta X\|_{\rm F} &\le r + \frac{a}{1-\chi\|\delta X\|_{\rm F}}\|\delta X\|_{\rm F} \\ &\le \chi\|\delta X\|_{\rm F}^2 - (r\chi - a)\|\delta X\|_{\rm F} + r \\ &\le \chi\|\delta X\|_{\rm F}^2 + k\|\delta X\|_{\rm F} + r, \quad (8)\end{aligned}$$

with $k := a - r\chi$.

Let $W, V, S \in \mathbb{C}^{n\times n}$ be given arbitrary, and define the linear operator $\Phi : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ by

$$\Phi(H) = S + W^{\rm H}(V-H)^{-1}HV^{-1}W.$$

Written in operator form equation (3) becomes

$$\delta X = \Phi(\delta X). \quad (9)$$

For $\rho > 0$ denote by $\mathcal{B}(\rho) \subset \mathbb{C}^{n\times n}$ the set of all matrices $M \in \mathbb{C}^{n\times n}$ satisfying $\|M\|_{\rm F} \le \rho$. For $H \in \mathcal{B}(\rho)$, having in mind (8), we have

$$\|\Phi(H)\|_{\rm F} \le r + k\rho + \chi\rho^2.$$

The function $h(\rho) = r + k\rho + \chi\rho^2$ is a Lyapunov majorant for equation (8) and the majorant equation for determining a non-local bound ρ for $\|\delta X\|_{\rm F}$ is

$$\chi\rho^2 - (1-k)\rho + r = 0. \quad (10)$$

Suppose that $r \in \Omega$, where

$$\Omega = \left\{0; \frac{(1-\sqrt{a})^2}{\chi} \cup \frac{(1+\sqrt{a})^2}{\chi}; +\infty\right\}. \quad (11)$$

Then equation (10) has nonnegative roots $\rho_1 \le \rho_2$ with

$$\begin{aligned}\rho_1 &= \phi(r), \quad (12)\\ \phi &:= \frac{2r}{1-a+r\chi+\sqrt{(1+a-r\chi)^2-4a}}.\end{aligned}$$

The operator Φ maps the closed central ball

$$\mathcal{B}_\psi = \left\{\mathrm{vec}(\delta X) \in \mathbb{C}^{n^2} : \|\mathrm{vec}(\delta X)\|_{\rm F} \le \phi(r)\right\}$$

of radius $\phi(r)$ into itself, where $\phi(r)$ is continuous and $\phi(0) = 0$. Then, according to the Schauder fixed point principle, there exists a solution $\delta X \in \mathcal{B}_{\phi(r)}$ of equation (9). In what follows, we deduced the following statement.

Theorem 1. Consider equation (1) for which the solution X is approximated by $\hat{X}$, obtained by some iterative algorithm with residual $R(\hat{X})$ (2).

Let $r := \|R(\hat{X})\|_{\rm F}$, $a := \|A_2^{\top}\hat{X}^{-1} \otimes A_2^{\rm H}\|_2\|\hat{X}^{-1}\|_2$ and $\chi := \|\hat{X}^{-1}\|_2$.

For $r \in \Omega$, given in (11) the following bounds are valid:

- non-local residual bound

$$\begin{aligned}\|\delta X\|_{\rm F} &\le \phi(r), \quad (13)\\ \phi(r) &:= \frac{2r}{1-a+r\chi+\sqrt{(1+a-r\chi)^2-4a}};\end{aligned}$$

- relative error bound in terms of the unperturbed solution X

$$\frac{\|\delta X\|_{\rm F}}{\|X\|_2} \le \frac{\phi(r)}{\|X\|_2}. \quad (14)$$

- relative error bound in terms of the computed approximate solution $\hat{X}$

$$\frac{\|\delta X\|_{\rm F}}{\|X\|_2} \le \frac{\phi(r)/\|\hat{X}\|_2}{1-\phi(r)/\|\hat{X}\|_2}. \quad (15)$$

4 NUMERICAL EXAMPLE

We consider Example 4.3. from (Xu, 2001) in order to compare the accuracy of est (15), proposed in Section 3 to this of the residual bound proposed in (Xu, 2001).

Consider equation $X + A_2^{\mathrm{H}} X^{-1} A_2 = A_1$ with coefficient matrices

$$\begin{aligned} A_2 &= \frac{1}{10} \begin{pmatrix} -1 & 0 & 0 & 0 & 1 \\ -1 & 1 & 0 & 0 & 1 \\ -1 & -1 & 1 & 0 & 1 \\ -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 \end{pmatrix}, \\ A_1 &= X + A_2^{\mathrm{H}} X^{-1} A_2 \end{aligned}$$

and solution $X = \mathrm{diag}(1,\ 2,\ 3,\ 2,\ 1)$. The approximate solution $\hat{X}$ of X is chosen as

$$\hat{X} = X + 10^{-2j} X_0; \quad X_0 = \frac{1}{\|C^\top + C\|}(C^\top + C),$$

where C is a random matrix, generated by MatLab function **rand**. The error bound (15) for $\hat{X}$, defined in Theorem 1 is compared to the proposed in (Xu, 2001) error bound

$$\frac{\|\delta X\|}{\|X\|} \leq \frac{1}{\frac{1}{2} - \|A_2\| \|A_1^{-1}\|} \frac{\|R(\hat{X})\|}{\|A_1\|}.$$

The results for $j = 1,2,3,4,5$ are listed in Table 1.

Table 1: Numerical results.

j	$\frac{\|\delta X\|_{\mathrm{F}}}{\|X\|_2}$	est (Xu, 2001)	est (15)
1	3.33e-003	1.73e-002	1.39e-002
2	3.33e-005	1.73e-004	1.36e-004
3	3.33e-007	1.73e-006	1.36e-006
4	3.33e-009	1.73e-008	1.36e-008
5	3.33e-011	1.73e-010	1.36e-010

The results illustrate the sharpness of the bound (15) proposed in Theorem 1.

ACKNOWLEDGMENTS

The research work presented in this paper is partially supported by the FP7 grant AComIn No 316087, funded by the European Commission in Capacity Programme in 2012-2016.

REFERENCES

Engwerda, J., Ran, A., and Rijkeboer, A. (1993). Necessary and sufficient conditions for the existence of a positive definite solution of the matrix equation $X + A^* X^{-1} A = Q$. *Linear Algebra Appl.*, 186:255–275.

Ferrante, A. and Levy, B. (1996). Hermitian solutions of the equation $X = Q + NX^{-1}N^*$. *Linear Algebra Appl.*, 247:359–373.

Hasanov, V. (2010). Notes on two perturbation estimates of the extreme solutions to the equations $X \pm A^* X^{-1} A = Q$. *Appl. Math. Comput.*, 216:1355–1362.

Konstantinov, M., Petkov, P., Angelova, V., and Popchev, I. (2002). Sensitivity of a complex fractional–affine matrix equation. In *Proc. Jub. Sci. Conf. Univ. Arch. Civil Eng. Geod.*, 8, pages 495–504, Sofia.

Konstantinov, M.M., Petkov, P.H., Popchev, P.I., and Angelova, V.A. (2011). Sensitivity of the matrix equation $A_0 + \sum_{i=1}^{k} \sigma_i A_i^{\mathrm{H}} X^{p_i} A_i = 0$, $\sigma_i = \pm 1$, In *Int. J. Appl. Comput. Math.*, 10(3):409–427.

Meini, B. (2001). Efficient computation of the extreme solutions of $X + A^* X^{-1} A = Q$ and $X - A^* X^{-1} A = Q$. *Math. Comp.*, 71:1189–1204.

Popchev, I. and Angelova, V. (2010). On the sensitivity of the matrix equations $X \pm A^* X^{-1} A = Q$. *Cybernetics and Information Technologies*, 10(4):36–61.

Popchev, P.I., Konstantinov, M.M., Petkov, P.H., and Angelova, V.A. (2014). Norm-wise, mixed and component-wise condition numbers of matrix equation $A_0 + \sum_{i=1}^{k} \sigma_i A_i^{\mathrm{H}} X^{p_i} A_i = 0$, $\sigma_i = \pm 1$. *Int. J. Appl. Comput. Math.*, 13(1):18–30.

Sun, J.-G. and Xu, S. (2003). Perturbation analysis of the maximal solution of the matrix equation $X + A^* X^{-1} A = P$. II. *Linear Algebra Appl.*, 362:211–228.

Xu, S. (2001). Perturbation analysis of the maximal solution of the matrix equation $X + A^* X^{-1} A = P$. *Linear Algebra Appl.*, 336:61–70.

SYNERGY EVALUATION IN ECONOMIC CLUSTERS: APPROCHES AND SOLUTIONS

Irina Radeva
Institute of Information and Communicaiton Technologies, akad. G. Bonchev blvd., bl. 2, Sofia, Bulgaria
iradeva@iit.bas.bg

Keywords: Economic clusters, Synergy, Synergy effects

Abstract: The following work presents a study of some theoretical approaches and solutions to the assessment of synergy and synergy effects in regard to the application of identification and management of synergy in economic clusters. Presented are five different approaches to problem formulation and sources of synergy effects. They solve specific problems and use indicators such as economic growth, cash flow, efficiency, entropy level, and evaluation of sustainable development. Approaches are analysed according to the general formulation, types of synergy effects evaluation, necessary data, and results interpretation. The analysis shows that the problems of synergy and its effects assessment have a partial solution and need systematization, extension, and additions to the conceptual and methodological direction, so that they can be applicable to economic clusters.

1 INTRODUCTION

Viewing an economic cluster as a system allows for the assumption that the interaction between its participants leads to the manifestation of synergy. This statement is intuitively supported by the various definitions of the concept of economic clusters, which generally define it as a group or voluntary union of enterprises. The most popular definition, suggested by M. Porter, is that a cluster is a geographic group of companies, affiliated in a particular field and united by similarities and mutual complementation (Porter, 1998). Over years of researching this economic phenomenon the definition has been added to, expanded and clarified, always towards the affirmation of the positive effects of mutual and equal cooperation while preserving and developing the competitive positions of the enterprises involved. Synergy (Greek – collaboration) is a combined action of several components, where the overall effect is greater than the effect of each individual component. (Ansoff, 1988) defines synergy as a measure of collaborative effects, when the income from the joint utilization of resources exceeds the sum of incomes from utilizing the same resources separately. This is a “2 + 2 = 5” effect (Bradley, M., A. Desai, E. Kim, 1983). (Copland, T., Koller, T., Murrin, J., 2002) names synergy as one of the most important factors influencing the business state of the individual economic entities which are united in clusters.

Synergy can manifest in any group of enterprises. It is a potential, which can be realised under proper management and allocation of resources in the group (Kaplan R., D. Norton, 2006). Its expressions are diverse and are known as “synergy effects”. However these effects cannot manifest themselves, no matter how potentially big they are. They have to be planned and derived.

Regardless of the widespread use of the concepts of synergy and synergy effects, the methods for their quantitative and qualitative evaluation as well as those for their integration in decision-making systems, have not been sufficiently systematised. A large number of proposed models apply only to isolated cases, for instance: evaluating the synergy in the merger or takeover of companies; choosing for merger and takeover; diversification of business activities; evaluating the synergy due to increasing the efficiency of business activities; evaluating the synergy of integrating corporate structures or economic systems; or as an indirect condition of positive synergy between economic agents in the evaluation of the investment preference of cluster structures.

Analysis of different research sources leads to the conclusion that topics such as the identification of clusters (Velev, 2007), (Radeva, I., T. Naneva, 2007), the evaluation of cluster effect on the economy and predicting the efficiency of their functioning (Bergman, E.M. and Feser, E.J., 1999) have been studied thoroughly enough. There remain, however, the problems of argumentation of the mechanisms of the relationships between enterprises in the cluster, as well as the choice of participants to be included in clusters (Radeva I. , 2013), (Vom Hofe R., K. Chen., 2006). There is practically no scientific applied research, which can utilise a systematic approach in evaluating the functional efficiency of clusters.

The goal of the following work is researching some theoretical approaches and solutions to the evaluation of synergy, synergy effects and their utilization, which can be applied to manifest and manage synergy in economic clusters. To that end five approaches are presented, which evaluate or make use of synergy effects. They differ significantly in formulating issues and the sources of synergy effects they address, but they could find their interpretation in the context of managing cluster activities. The approaches are analysed according to their general formulation, the synergy effect type they assess or utilize, the data required and the extent of result interpretation.

2 SYNERGY EVALUATION APPROACHES

2.1 Synergetic Modelling of Economic Growth in Accounting the Merger and Takeover of Companies

The approach assumes that as a result of merger or takeover, there is a complete or partial transition of funds from one company to another. It is applicable to enterprises or groups of enterprises, clusters, holdings etc. (Serkov, 2009). The approach is built on the following assumptions:

- two groups of enterprises: enterprise-centres which can annex and enterprise-peripheries, which can be annexed;
- when enterprise-centres are not present, the speed of capital accumulation in enterprise-peripheries is described through a logistic function;
- the enterprise-centres can exist in one of two states: a state of development (growth) or a state of accumulation (when they are annexing);
- the capital accumulation of enterprise-centres occurs according to the single-sector model of Solow with the accounting of synergy effects stemming from the use of labour resources.

The starting system of equations is as follows:

$$x' = A + \lambda . x . \left(\frac{1-x}{N}\right) - \theta . x . y; \quad (1)$$

$$y' = -\theta . x . y + \frac{s}{\tau_R} \times \frac{F(K, z.L)}{L} - \mu . y; \quad (2)$$

$$k' = s . f(k, z) - (\delta + n) . k, \quad (3)$$

where x, y and k are endogenous specific indicators of capital accumulation of the corresponding enterprises: $x = \frac{X}{L_1}, y = \frac{Y}{L_2}, k = \frac{K}{L}$; X, Y, K are the absolute values of the capital of corresponding enterprises; $L_1, L_2, L = L_1 + L_2$ is the number of employees in these enterprises; $L = L_0 \exp(\mathrm{n} + 1)$, n is the annual growth rate of the number of employees.

The derivatives x', y', k' characterise the accumulation of capital respectively by x (enterprise-peripheries, which are annexed first), y (enterprise-centres) and k (enterprises formed by the annexing of the former by the latter). The k enterprises can be interpreted as a cluster because they are formed as a result of the inclusion of the *x* enterprises to the *y* enterprises.

The first equation describes the evolution of the object under the conditions of limited resources and competition, which limits growth. The A parameter is the constant speed of fund accumulation in enterprise-peripheries, related to the market entrance of new enterprises, λ is the absolute speed of the rate of capital gain of these enterprises, N is the stock market potential. The last article $\theta . x . y$ describes the process of uniting both types of enterprises; θ is the speed limit of enterprise union.

In the second equation s is the rate of capital accumulation in the merged enterprises $(0 < s < 1)$, τ_R is the life-cycle of these enterprises between two consecutive mergers, $F(K, z.L) = K^{\alpha}(z.L)^{1-\alpha}$ is a production function of the common company which describes the production volume in terms of value. According to the authors' comment, *the synergy effect of merger or takeover consists of the rational utilization of labour resources and human capital.*

The parameter z is a synergetic parameter of utilizing labour resources:

$$\frac{F(K, z.L)}{L} = f(k, z) = k^{\alpha} . z^{1-\alpha},$$

where α is a regression parameter ($0 < \alpha < 1$), $k = \frac{K}{L}$ and $z > 1$ *when there is a positive synergetic effect from the inclusion of companies and* $0 < z < 1$ *when there is a negative synergetic effect*.

The last part $\mu.y$ describes a reduction of capital in enterprise-centres related to the lack of resources and competition; μ is the absolute speed of this process.

The third equation describes the accumulation of capital in the joint enterprises (complexes) in accordance to Solow's single-sector model. It uses the exogenous indicators: s - rate of accumulating capital, δ – share of the reduction of capital, ($0 < \delta < 1$) and n - annual growth rate of the number of employees ($-1 < n < 1$).

The model described in the approach uses parameters such as capital, number of employees, and their corresponding rate of growth under the conditions of the production function of Cobb-Douglas. The evaluated synergy effect can be partially related to the operational effect, including the rational utilisation of labour resources and human capital in accordance to the Ansoff classification (Ansoff, 1988). The positive or negative manifestation is accounted. Limitations of utilization that can be highlighted are the difficult gathering and processing of the required data, as well as the need for adaptation to using for clusters due to the voluntary and equitable organisation of the relationships between the enterprises involved. An advantage is the complex approach, the formal mathematical description and the specific synergetic effect it evaluates.

2.2 Assessment of Synergy Effects Based on the Discounted Cash Flows Method

The model described by (Khasanova G.F., I.V. Burenina, 2011) is following:

$$\Delta V = V_{synergy} = \frac{\sum_{i=1}^{n} \Delta FCFE_i}{(1+re)^i}, \quad (4)$$

where:

- $\Delta FCFE_i = \Delta S - \Delta C - \Delta T - \Delta NWC - \Delta Capex + \Delta Debt$;
- ΔS (sales) – income growth;
- ΔC (costs) – cost savings;
- ΔT (tax) – profiteer tax savings;
- ΔNWC (net working capital) – working capital investment savings;
- $\Delta Capex$ (capital expenditure) – investment savings;
- $\Delta Debt$ - clear debt alterations;
- re - share capital profitability;
- i – reference period of time.

The advantage of this model is its ease of use, due to the standards included in it, as well as easily interpreted parameters. In regards to a single enterprise this type of evaluation is easy to apply and interpret. In case of evaluating clusters preliminary data processing is necessary so as to eliminate duplication and to differentiate and affirm the causes and sources of the observed results. The authors do not state the type of evaluated synergetic effect but it can be referred to Development in the Ansoff classification. There remains, however, the question of what part of the thus evaluated discounted cash flows are a result of synergetic effects realised in the cluster, i.e. whether they stem from interaction or from other factors.

2.3. Synergy Through Evaluation of the Activity's Efficiency

In this model (Ekimova, K., E. Fedina, 2010), the assessment of functional efficiency of a cluster is based on two elements: evaluating the efficiency of a single enterprise and evaluating the efficiency of the group of enterprises as a whole. The efficiency on the level of a single enterprise includes the share of synergetic effect granted to said enterprise:

$$E_i = W_i + S_i, \quad (5)$$

where:

- E_i is efficiency of i-th enterprise;
- W_i is efficiency of the activity of i-th enterprise;
- S_i is share of synergetic effect granted to i-th enterprise.

The efficiency of the enterprise participating in the cluster is given as a ratio between the acquired result and the total expenditure it required. The value of efficiency in (5) should be maximised. It is defined as a ratio between the effect of creating a cluster and the total expenditure necessary for it. In this case it is reasonable to use the formula for evaluating investments as far as a cluster can be considered an investment project. The effect of an enterprise functioning in a cluster is represented as the difference between inflow and outflow over a given period of time. The expenditures necessary to achieve the effect are considered as investments made in order to join the cluster. When accounting for inflation and discount rate the formula for calculating the efficiency of the cluster is following:

$$E_c = \frac{\sum_{i=0}^{T}\frac{ICF_i^t}{[(1+d)(1+p)]^t} - \sum_{i=0}^{T}\frac{OCF_i^t}{[(1+d)(1+p)]^t}}{\sum_{i=0}^{T}\frac{K_i^t}{[(1+d)(1+p)]^t}}, \quad (6)$$

where:

- E_c is the efficiency of their activity of the cluster;
- ICF_i^t is the inflow of i-th enterprise over a given period of time t;
- OCF_i^t is the outflow of i-th enterprise over a given period of time t;
- K_i^t is the investment of i-th enterprise over a given period of time t;
- d is the annual interest rate on bank deposits;
- p is the annual inflation rate.

An advantage of the model is the ease of interpretation of the results and the accessibility of the data involved in the calculations. What is assessed, though, is efficiency, while assuming that the share of synergetic effect of a given enterprise is the difference between the overall efficiency and the efficiency of the activities of the enterprise. The overall efficiency of the cluster is more difficult to evaluate because remaining on the level of a single enterprise leads to accounting for the shares of mutual shared resources utilised by the enterprise as part of the cluster. This demands adaptations in the system of accounting of the enterprises and the management unit of the cluster.

2.4 Entropy as an Assessment of the Synergetic Effect in the Activities of Economic Systems

A systematic approach is utilised here, which evaluates the synergy in an economic system. The synergy is assessed through entropy. This approach can be adapted to clusters considering they are economic systems. Entropy is a quantitative measure of uncertainty in an economic system.

In the approach, *the economic system is described with two characteristics: with a function of its state and with a degree of its organization.*

The activity is a combination of functional and structural processes. The functional processes derive directly form the content of the cluster. The structural processes support and develop the cluster's organisation and activity. The realisation of the processes defines the expenses, which are presented as functional and structural (practically equated to the transactional expenses). Structural expenses are not utilized directly in the manufacture of production. In order to describe the state of the system (Krasnov, G.A., V.V. Vinogradov, A.A.Krasnov, 2009) a *function of state* is defined:

$$Q = P + F + C, \quad (7)$$

where:

- Q is the amount of resources obtained as a result of economic activity;
- P is the amount of resources which provide profits for the economic system;
- F is the amount of resources spent on functional expenses and
- C is the amount of resources spent on structural expenses.

A cluster is an economic system formed as a result of the interaction of the elements incorporated in it. Following this analogy, a cluster's function of state is expressed as a unification of structural and functional resources, as well as the resources providing the overall gain:

$$Q_c = P_c + F_c + C_c \quad (8)$$

When there is no synergetic effect, assuming that the system is comprised of an n number of enterprises which utilize technological processes sufficiently, (8) is presented thusly:

$$Q_c = \sum_{i=1}^{n} Q_i = \sum_{i=1}^{n} P_i + \sum_{i=1}^{n} F_i + \sum_{i=1}^{n} C_i, \quad (9)$$

A positive synergetic effect at the expense of superadditivity is accomplished when:

$$P_c > (P_1 + P_2 + \cdots . + P_n),$$

where:

- P is the amount of resources, which provide profits for the enterprises;
- n is the number of enterprises;
- P_c is the total resources.

Defining the structural expenses C is not easy, due to the high degree of uncertainty of the changes in environment and the uncertainty of the changes in the economic interaction between the enterprises in the system.

As an economic system, a cluster can also be characterised by the *degree of its organisation.* (Alexandrov E.A., V.P. Bogolepov, 1968) utilised for for assessment the so-called "Shannon surplus" R:

$$R = 1 - S/S_{max}, \quad (10)$$

where:
- R is surplus;
- S is the rate of entropy in the system;
- S_{max} is the maximum rate of entropy in the system

(Krasnov, G.A., V.V. Vinogradov, A.A.Krasnov, 2009) suggests using statistical entropy, where $S_{max} = 1$, which transforms (10) into:

$$R = 1 - S, \quad (11)$$

The condition of positive synergetic effect when integrating a n number of economic systems is related to the entropy of the integrated system as following:

$$S < F \times (1 - S_n) \times \frac{\left[\frac{Q_{n-1} \times F_n \times (1 - S_{n-1})}{Q_n \times F_{n-1} \times (1 - S_n)} - 1\right]}{\left[F_n \frac{Q_{n-1}}{Q_n} - 1\right]}, \quad (12)$$

where:
- S is the entropy in the integrated system;
- $F = F_1 + F_2 + \cdots + F_n$ are the total functional expenses of the unifying economic systems;
- F_n are the functional expenses of the n-th economic system;
- S_n is the entropy in the n-th economic system;
- S_{n-1} is the rate of entropy resulting from the unification of $n - 1$ economic systems;
- Q_n is the amount of resources accumulated as a result of the manufacturing activity of the n-th economic system;
- Q_{n-1} is the total amount of resources accumulated as a result of the manufacturing activity of $n - 1$ economic systems.

According to (Prangishvili, 2003), the sustainable evolutionary development of a complex economic system requires the ratio between order and disorder in the system to satisfy the conditions of the golden section:

$$R/S = (\sqrt{5} - 1)/2.$$

where:
- R is a measure of order in the system and
- S is a measure of disorder (a measure of chaos).

Therefore in order to have a sustainable evolutionary development $S > R = 0.62$ or the development of the system implies a relative growth of entropy.

The presented approach is principally different than the ones showcased above. The limitation that can be highlighted is that the value of the final assessment is difficult to interpret in the context of standard economic indicators and cannot be directly linked with them. There are, however, multiple advantages: a simply formulated function of state, which includes both income and expense; an evaluation of the state of the whole system, including the process of expansion of said system. One of its greatest strengths is the formulation of a condition of positive synergetic effect which has two consequences: firstly it is an element which is easily included as a criterion in the process of decision making; secondly it assesses the ability to express a negative synergetic effect. In the Ansoff classification, the synergy of management is the one, which has potential to exhibit negative synergy.

2.5. Utilizing a Condition of Positive Synergy in Cluster's Design

(Radeva I. , 2013), (Radeva I. , 2012) develops a decision making system for cluster's design. It combines a number of models of multi-criteria choice. In the process of formulating criteria it utilises the assumption that if enterprises included have similar rates of *sustainable development* this will create the conditions necessary for attaining positive synergy. In other words, if the cluster selects participants that are comparable in regard to pre-established indicators, which evaluate the level of technological advancement, productivity, efficiency, market share etc., they will be able to form mutually beneficial relationships in the conditions of equality inside the cluster structure. If the enterprises differ widely in the assessment of these criteria groups, there is a risk of relationships not developing under the conditions of equality and this may lead to acquisition or the taking over of functions by the stronger enterprises at the expense of the weaker enterprises.

The system includes two model groups: models for selection participant enterprises for the cluster and models for a cluster structure selection.

The first group allows selection the enterprises according to the *evaluation of their state of development* (ESD) in two stages. ESD is a complex assessment, acquired through a system of indicators. The first stage is to make the choice following two ranking of enterprises according to their ESD, respectively according to nodes in a given technological network and in the whole network. The resulting ranking is compared while choosing a set of enterprises whose assessment value is above a specifically set threshold. The second stage is to assess the chosen set of enterprises according to the criteria of *economic creativity* and the *growth through competitiveness,* which are also complex evaluations under a given system of indicators. The assessment of both criteria forms a pair of values that allow the positioning of enterprises in a two-dimensional space, called *polygon of sustainable development.* The enterprises analysed may or may not form a cluster on the polygon. The proximity in the cluster is considered a prerequisite for the manifestation of positive synergy. If such exists, the relevant enterprises transition to the next stage where versions of the potential cluster structure are designed. In case of insufficient data the process is returned to a previous stage for adjustment. Obtaining the arrangements and complex evaluations is done through utilizing the model PROMETHEE II (Brans, 1994), realised through DSS MKA – 2 (Vassilev, V., S. Konstantinova, 2005).

The second group of models facilitates the selection of the cluster by criteria of investment preference (IP) evaluation. With risk-free investments the IP assessment includes a developed Balanced Scorecard system (BS_C). The evaluation is based on a *consolidated budget,* calculated according to a cluster's activity programme. The assessment under risk conditions is done by additionally calculating NPV, IRR and the standard deviation of cash flows, based on developed pessimistic, realistic and optimistic scenarios of consolidated budgets. The procedure of decision-making is done vie the algorithms LINCOM and MAXIMIN (Popchev, 2009).

The last stage of decision-making is an assessment of the state of enterprises through prognosis evaluation of the indicators of efficiency over a period of time after clusterisation. The selection of cluster structure is finalised if all participants exhibit improvement or at least the sustaining of the initial values of the indicators. In this case it is accepted that the wanted synergy resulting from the unification of enterprises has been attained.

An advantage of this system is the provision of a thorough interactive approach towards the design and selection a cluster structure. The condition of positive synergy, expressed as a requirement of comparability and relation according to given criteria, is integrated into the process of choosing participants. The approach is applicable both in the design and optimisation of cluster structures. The required data is accessible and easily interpreted. The disadvantage that can be observed is not specifying and not formalising the sources of synergetic effects that makes their subsequent management difficult.

4 CONCLUSIONS

The presented approaches and solutions for assessing synergy have been developed for specific tasks and use different indicators – economic growth, cash flows, efficiency, and rate of entropy, sustainable development. The first method uses a strictly formalised description of the economic growth of enterprises, participating in the processes of merger and takeover and assesses the synergy of utilising labour resources. The second utilises discounted cash flows calculated based on savings or the growth of relevant components constituting the cash flow. The third one evaluates synergy effects through efficiency. The fourth one applies a systematic approach and assesses synergy through entropy, developing a condition of positive synergetic effect. The fifth one utilises a condition of positive synergy that can be fulfilled through the cluster design in a polygon of sustainable development, with the goal of selecting comparable enterprises and preventing consequences such as takeover and the seizure of activities and markets.

Conceptually, the last two methods come closest to the expression of a general effect from the interaction of enterprises, although they lack specific and easily interpreted values of synergy assessment. The analysis made shows that the problems of synergy and its effects have partial solutions and need systematisation, additions and extensions both conceptually and methodologically so that they can be applied towards the expression and management of synergy in economic clusters.

In regard to future directions for research on synergy and its effects it is appropriate to search for additional analogies with practical experience accumulated in the field of risk identification and risk management. Synergy and risk are phenomena that exist in any system. Uncovering the correlation between them could provide an effective instrument for influencing the formation, activities and optimisation of cluster management in economics. Questions such as: what is the desired synergetic effect, does it manage or assess, at what stage of a cluster's life cycle, where in the cluster is synergy assessed, how to account interactions, "together-ness", "resonance" in the functioning of enterprises should be at the core of formulating the problems in this field of application.

ACKNOWLEDGEMENTS

This work is supported by project FP7 - REGPOT–2012–2013–1, grant agreement 316087 AComIn.

REFERENCES

Alexandrov E.A., V.P. Bogolepov. (1968). On Some Organizational System Performance Criteria. In *Organization and Management* (p. 275). Moscow: Mir.

Ansoff, H. I. (1988). *The New Corporate Strategy* (Revised Ediiton ed.). Wiley.

Bergman, E.M. and Feser, E.J. (1999). *Industrial and Regional Clusters: Concepts and Comparative Applications.* WVU Regional Research Institute.

Bradley, M., A. Desai, E. Kim. (1983). The Rational Behind Interfirm Tender Offers: Information ir Synergy. *Journal of Financial Economics , 2*, 195.

Brans, J. P. (1994). The PROMCALC & GAIA Decision Support System for Multicriteria Decision Aid. *Decision Support System* (12), 297-310.

Copland, T., Koller, T., Murrin, J. (2002). *Valuation: Measure and Managing the Value of Companies.* New York: John Wiley.

Ekimova, K., E. Fedina. (2010). Organizational-economic mechanism of interaction between participants of regional economic cluster of artificial. *Herald Chelyabenskogo State University, 26*, 179-183.

Kaplan R., D. Norton. (2006). *Alignment: Using the Balanced Scorecard to Create Corporate Synergies.*

Khasanova G.F., I.V. Burenina. (2011). Synergy as Method for Increasing Company's Effectivness. *Electronic Journal Oil and Gas Business* , 188-196.

Krasnov, G.A., V.V. Vinogradov, A.A.Krasnov. (2009). Conditions for Manifestation of Synerghetic Effect in the Process if Integration of Economic Systems. (4), 219-222.

Popchev, I. (2009). Multi-Criteria Selection of Design Decisions: some practical algorithms. *Financial Decisions: Research and Practice*, 23-51.

Porter, M. (1998). *On Competition, Clusters and Competition: New Agendas for Companies, Governments, and Institutions.* . Boston: Harvard Business School Press.

Prangishvili, I. (2003). *Entropy and Other Systemic Laws: Issues of Complex Systems.* Moscow: Institute of Control Sciences. VA Trapeznikova, Nauka.

Radeva, I. (2012). *Decision Making Models for Cluster Design - Dissertation.* Sofia: IICT - BAS.

Radeva, I. (2013). Multi-Criteria Models for Cluster Design. *Cybernetics and Information Technologies* , 18-33.

Radeva, I., T. Naneva. (2007). Identification of Economic Clusters. Concepts and Methods. *Automatics and Informatics , 4*, 41-44.

Serkov, L. (2009). Synergetic Modelling of Economic Growth in Accounting the Merger and Takeover of Companies. *Herald Chelyabenskogo State University , 9 (147)* (20), 113-117.

Vassilev, V., S. Konstantinova. (2005). Multicriteria Macroeconomic Ranking of Two Groups of European Countries by Decision Support System MKA-2. *International Workshop Proceedings "Distributed Computer and Communication Networks"*, (pp. 153-164). Bulgaria, Sofia.

Velev, M. (2007). *Cluster Approach to Increase of Competitivness.* Sofia: Softrade.

Vom Hofe R., K. Chen. (2006). Whither or Not Industrial Clysters: Conclusions or Confusions. *The Industrial Geographer , 4* (1), 2-28.

SOLVING SCALAR NONLINEAR EQUATIONS

High Precision Computation with .NET Framework C# and X-MPIR

Velichko Dzhambov
Institute of Information and Communcaition Technologies
Acad. G. Bontchev Street, Sofia, Bulgaria
vili_jambov@abv.bg

Keywords: High precision computation, Solving nonlinear algebraic equation, Computational mathematics, .NET Framework, X-MPIR

Abstract: This paper concerns root finding of nonlinear algebraic equation with high precision in specific environment, namely .NET Framework. The work is part of a series tracing the progress of creating tools for high precision computations in this environment and may be considered as continuation, in this direction, of the beginning, in (Dzhambov, 2011), that includes special function calculations with arbitrary precision. The results are illustrated with the help of applications, using the current state-of-art library being created for implementation of functions and numerical methods of arbitrary precision in a given environment.

1 INTRODUCTION

Arbitrary precision computations are not a self-purpose. They are related to receiving precise values when solving mathematical models in different area, including often for instance as (sub)task, solving nonlinear algebraic equations. In (Dzhambov, 2011) an announcement is made for creating of high precision computation library in the environment of .NET Framework. The application SFCALC was represented there, illustrating a part of the potentialities of the library being created. Some extensions of this library are added further concerning computing of definite integrals (Dzhambov, 2014), solving of systems of nonlinear ordinary differential equations, and root finding of nonlinear algebraic .equation (real roots), described bellow.

This task is subdivided naturally in different tasks, according to the information available a priori: roots of general nonlinear algebraic equation (some general assumptions concerning differentiability or root's multiplicity are also needed even in this case, before we choose an appropriate method), and roots of some special functions. In this last case (for instance all zeros of the classical orthogonal polynomials are real and distinct, Bessel's functions are solutions of specifically parametrized second order homogenous linear ordinary differential equation) specific, more elaborate methods may be used. Another natural classification is based on properties of the method: global or local, iterative or analytic, derivatives used (no, first, second...), order of convergence, one-point or multi-point, with memory or not, computational efficiency and so on. Root finding is one of the oldest numerical analysis areas and the number of references, related to the topic is enormous. (Traub, 1964), (Ostrowski, 1966), (McNamee, 2007), (Sharma, 2011), (Gil, 2003), (Allgower, 1990) are some of the sources, we used.

One of the purposes is to demonstrate that useful and mutually complementing computational tools for solving non-trivial problems with high precision computations may be implemented in a concrete environment which is wide spread for personal computers but it seems to be underestimated by the software developers for scientific applications. There are reasons of course for this, but in our opinion, there are advantages also: comparatively easy integration in various functionalities such as visualization and interactivity in one and the same application if, of course, the necessary methods for solving the concrete problems are available which is the purpose of the library being created. We of course keep in mind that the creation of a complete system covering a great range of tools for solving of various mathematical models is a task too ambitious due at least to the huge amount of the work required. But the implementation of some basic tools makes it possible for the interested investigator to create own applications in the given environment, exactly matching to his concrete interests. To viz. create and not be a user of a ready system. Another purpose which hope to achieve is to provide such a possibility.

2 GENERAL STRUCTURE AND EXTENSIONS OF THE LIBRARY, CONCERNING ROOT FINDING OF NONLINEAR ALGEBRAIC EQUATION

The arbitrary precision of the calculations (with real numbers) is provided by the library X-MPIR. X-MPIR ensure the link from C# to MPIR (for details search the Web, there is no independent site, devoted to X-MPIR at this moment). On the base of X-MPIR, which provides the basic arithmetic operations only, a library of elementary and special functions is implemented, as well some numerical methods.

The library is grouped in several files that we shall briefly describe. Only the components, referring to the problem solved, will be mentioned.

funcs.cs: Basic implementation of the functions and constants, described in (Dzhambov, 2011), combined in a namespace MPMath_R, most in the public static class MPMath, and several general numeric methods, plus a public class TestDer for transforming mathematical expression (text representation) in postfix notation (ExpressionParser in mpir_pars_fx.cs, described below, is used), finding and then simplifying derivative of given mathematical expression (in namespace Zero_R). The method used is non-standard. The derivative is produced "in place" as postfix record, directly from the postfix record of the function, without building expression tree. This approach is very similar of that, described in (Krtolica, 2001). The class TestDer contains also method "internal mpir.mpf_t Compute(int derivative)", which is simply an interpreter. Functions for zeros of Airy functions and their derivatives and zeros of Bessel functions are also provided. We will discuss some of them later in section 3.

funcs_c.cs: It accomplishes complex arithmetic of arbitrary precision with the help of an internal class MPComplex, and the functions, existing in funcs.cs are implemented in the namespace MPMath_C, public static class CMPMath. An integral logarithm is added as well, an integral exponent and the elliptic Jacobi functions. This file is mentioned because some lastly added functions exist only there. Real variants simply pass complex number with zero imaginary part (in the valid range) and receive the real part of the complex result.

mpir_parse_fx.cs: Parser for mathematical expressions (public class ExpressionParser and some auxiliary classes implementing the concept of token in namespace mpir_parse_fx), recognizing arithmetic operations and the functions and constants, implemented in the previous two files. If no errors are available, the parser creates the converted in a postfix form list (List<Token>) from the elements of the expression, each one possessing specific information about the list element. In Token class there are fields foreseen for this information: a mathematic or literal constant, variable, arithmetic or functional operation, text presentation, priority, number of arguments, type of associativity, eventual value. This list is used by a concrete application, invoking the parser, in order to form a list of simplified elements (struct SToken {public int opcode; internal mpir.mpf_t val} or probably internal MPComplex val), including the operation code only, which may be updated by the given application in the process of the problem solution. This list is already prepared for interpretation with the help of a well known algorithm. The parser is designed to accept more than one variable and it denotes as parameters all variables that are not names of mathematic constants or name of the basic variable of the model being solved. This gives additional flexibility. If an error occurs, the parser acknowledges the type and position of the error in the expression.

3 APPLICATIONS

Even using only a part of the tools in the first two files described in section 2, one can create console applications for certain problems or useful auxiliary applications with graphical user interface, such as SFCALC, presented in (Dzhambov, 2011). The third enable the design of applications with considerably more friendly user interface, allowing the solution of a given class of problems without changing the program code, additional compilation, etc. An example of application of this type with graphical interface is NQTS, presented in (Dzhambov, 2014). This time some console applications are described. No point in doing graphical user interface for this type of problems. These applications use supplementary files. We begin with the case of general algebraic nonlinear equation. The only assumption is that the function is derivable. In subsection 3.2 applications for root finding of some special function are described.

3.1 Applications for general nonlinear algebraic equation

3.1.1 Local methods

MPRootFindTestLocal is an application for root finding of general nonlinear algebraic equation

$$f(x) = 0 \quad (1)$$

allowing direct input (text form) and choice of method. Actually this application was created originally for testing and comparing different methods, including these, described in (Sharma, 2011). Input parameters are: mathematical expression of the function, initial point, desired accuracy (number of decimal digits) and method. For example

```
MPRootFindTestLocal "x*exp(x*x)-
PowInt(sin(x),2)+3*cos(x)+5" -1 200
M8_1
```

(function 7, Table 2.1 in (Sharma, 2011)). Quotation marks are optional if there are no spaces in the functional expression. Here M8_1 is the name of the method (4.5.15), section 4.5 "Optimal eight order methods" in (Sharma, 2011) with $\alpha = 1$. Besides the result -1.2076... with error 2.4043e-212, we get also some additional information: 4 iterations, 12 function evaluations and 4 evaluations of the derivative. The same equation solved with the classical Ostrowski method give us 5 iterations, 9 function evaluations, and 4 evaluations of the derivative. For the Newton method these values are respectively 10, 10, 9.

When local methods are used, the closeness of the initial guess point to the solution is crucial. That is why an auxiliary tool assisting this choice is desirable. Fortunately we have one. It is a part of the functionality of the application VODE2 for solving second order ordinary differential equation (concerning another part of the library and not described here in details; uses internally gnuplot by automatic redirecting the output). This application has so called "expression mode" allowing visualization of virtually arbitrary function. Here is an example with the above function (interval [–1.5; 1]).

MPRootFindTestLocal uses separate file including currently implemented methods and is therefore open for adding new ones.

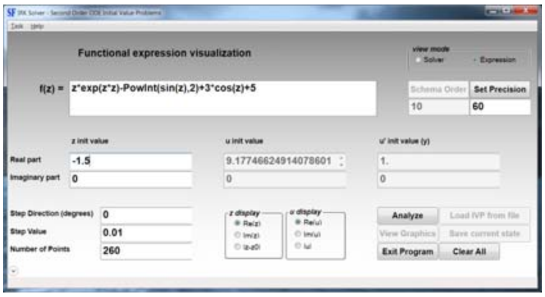


Figure 1: VODE2 in "expression mode".

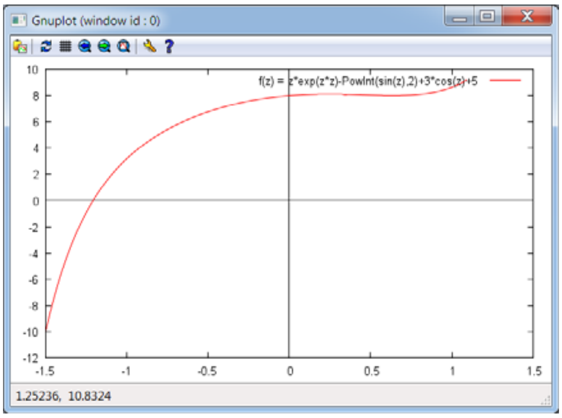


Figure 2: Visualization from "expression mode".

3.1.2 An attempt to solve the local convergence problem using homotopy continuation.

Of course, the local convergence problem can be solved in this case (one dimensional) using bisection. However, the homotopy continuation method has some advantages (besides the undoubted elegance): can be used in multidimensional cases and can be used for tracking multiple solutions (see (Rahimian, 2011) for one possible approach and also the references there for numerical methods that seek to find multiple solutions). Our approach is elementary: we use Newton homotopy

$$H(t,x) = f(x) - (1-t) f(x_0) \quad (2)$$

or fixed point homotopy

$$H(t,x) = (1-t)(x - x_0) + t f(x) \quad (3)$$

for some x_0, following the parameterized curve of the solution obtained by derivation of

$$H\big(t(s), x(s)\big) = 0 \quad (4)$$

from t=0 to t=1, or equivalently, we have to solve the initial value problem

$$\begin{aligned}
\frac{\mathrm{d}x}{\mathrm{d}s} &= -\frac{\partial H}{\partial t}, \\
\frac{\mathrm{d}t}{\mathrm{d}s} &= \frac{\partial H}{\partial x}, \qquad (5) \\
(x,t)\big|_{s=0} &= (x_0, 0)
\end{aligned}$$

until we reach $t(s)$=1. (For all possible scenarios in the general case, turning and branch (bifurcation) points and so on, we refer the reader to the very clear exposition in (Judd, 1998)).) The value $x(1)$ obtained this way is then used as initial guess point of some rapidly convergent local method. We have an appropriate tool in the library, namely public class IRK_Solver, designed for solving systems of ordinary differential equations using implicit Runge-Kutta schema of arbitrary order (accuracy is parameter too). One console application that combines the ideas described above is MPRootFindTestHN (besides functional expression and initial point, desired accuracy and schema order are required parameters). Let the function be (example took from (Judd, 1998))

$$f(x) = 2x - 4 + \sin(2\pi x). \qquad (6)$$

Although simple, its graphic worth to be displayed, because it gives us an idea why it is often difficult for the local methods to find the root (x=2, unique, multiplicity one) starting from relatively distant point.

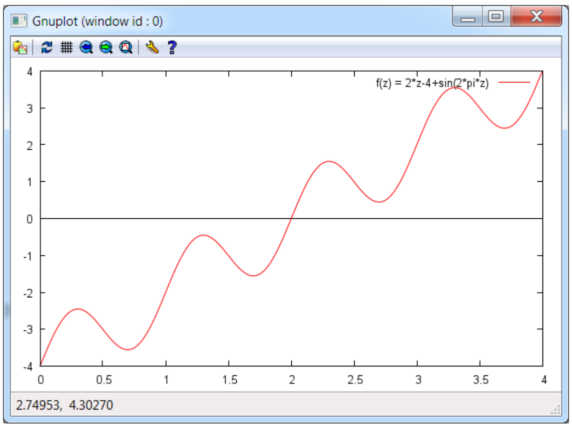


Figure 3: Graph of the function (6).

For example all of the three methods, mentioned in section 3.1.1 diverge for initial point $x = -8$ (and say accuracy 100 and schema order 8), but MPRootFindTestHN has no problems and (in this case, that is not general rule) even calculates the solution with the desired accuracy without to have passing the value of $x(1)$ for refining. Actually, when t=1 is reached (more precisely, surpassed), the program repeat several times the last step (dividing in half the step for IRK_Solver, increasing this way the accuracy).

3.2 Applications for finding zeros of some special functions

Besides their role in mathematical physics, the zeros of some special function are often used in different numerical analysis areas: in coefficients of asymptotical expansions (zeros of Airy functions and their derivative and zeros of Bessel functions of first and second kind) or quadrature formulae of gaussian type (zeros of classical orthogonal polynomials), to cite some applications. Such kind of root finding is used internally in some places of the library. As mentioned in section 3.1 some particular methods are available in this case and we will briefly describe some of them bellow. They are always alternatives. We don't discuss all possible options. In practice such choices compromise different criteria (effectiveness, simplicity, that is to say easy implementation, universality, namely ease of embedding as building blocks, and so on).

3.2.1 Roots of classical orthogonal polynomials

ADK (Aberth, Durand, Kerner) method is particularly convenient, because in this case all roots are real, distinct and single (multiplicity one). The ADK method is a modification of the Newton method with implicit deflating. If $x^{(1)}$, $x^{(2)}$, ... , $x^{(k)}$ are previously calculated roots of the polynomial $p(x)$, then the iterative formula

$$x_{n+1} = x_n - \frac{p(x_n)}{p'(x_n) - p(x_n)\sum_{j=1}^{r} \frac{1}{x_n - x^{(j)}}} \qquad (7)$$

is convergent to another root of the polynomial. The strategy consists in choosing an initial guess point $x=R$ greater than the largest root of the polynomial, and successive finding of all roots applying formula (7). R can be taken respectively as:

Polynomial	R
Jacobi	1
Legendre	1
Laguerre	$2n+\alpha-2+\sqrt{1+4(n-1)(n+\alpha-1)}$
Hermite	$\sqrt{2n+1}$

Table 1: Starting point for different orthogonal polynomials tu use with the ADK method.

Remark: Important special class of the Jacobi polynomials $P_n^{(\alpha,\beta)}(x)$ are the ultraspherical polynomials (called also Gegenbauer polynomials)

$$C_n^{(\lambda)}(x) = P_n^{\left(\lambda-\frac{1}{2},\lambda-\frac{1}{2}\right)}(x)\,.$$

We use currently separate implementation for the special case of the Legendre polynomials $P_n(x) = C_n^{\left(\frac{1}{2},\frac{1}{2}\right)}(x) = P_n^{(0,0)}(x)$ and not for the Chebyshev polynomials of first and second kind

$$(T_n(x) = C_n^{(0)}(x) = P_n^{\left(-\frac{1}{2},-\frac{1}{2}\right)}(x)\,,$$

$$U_n(x) = C_n^{(1)}(x) = P_n^{\left(\frac{1}{2},\frac{1}{2}\right)}(x)\;).$$

They have to be calculated directly (as Jacoby polynomials special cases: $\alpha=\beta=-1/2$ and $\alpha=\beta=1/2$).

If the polynomial is of high degree and high precision is required, it is convenient to use ADK method with smaller precision (half of the required for instance) as first stage and then pass the roots as initial points of rapidly convergent local method (the so called refiner, we use currently the method of Laguerre). As a test, the roots of all classical orthogonal polynomial of degree 25,000 with accuracy of 200 decimal digits was calculated along with the weights of their corresponding quardature formulae (7 hours on the author's laptop for the hardest problem of general Jacobi polynomials). It is understood that in this case the derivatives (first and second for the Laguerre method) are calculated using separate procedures and not interpreted after symbolical derivation as for the general purpose applications mentioned in section 3.1. First derivatives are calculated using directly the recurrent relations

$(2n+\alpha+\beta)(1-x^2)\frac{\mathrm{d}}{\mathrm{d}x}P_n^{(\alpha,\beta)}(x) =$ $n(\alpha-\beta-(2n+\alpha+\beta)x)P_n^{(\alpha,\beta)}(x) +$ $2(n+\alpha)(n+\beta)P_{n-1}^{(\alpha,\beta)}(x)$
$(1-x^2)\frac{\mathrm{d}}{\mathrm{d}x}P_n(x) = -2xP_n(x)+2P_{n-1}(x)$
$\frac{\mathrm{d}}{\mathrm{d}x}L_n^{(\alpha)}(x) = -L_{n-1}^{(\alpha+1)}(x)$
$\frac{\mathrm{d}}{\mathrm{d}x}H_n(x) = 2nH_{n-1}(x)$

Table 2: Recurrent relations for the first derivatives.

For the second derivatives the differential equation

$$A(x)f''(x)+B(x)f'(x)+\lambda_n f(x)=0 \qquad (8)$$

of the polynomial is used with the corresponding coefficients

	$A(x)$	$B(x)$	λ_n
$P_n^{(\alpha,\beta)}(x)$	$1-x^2$	$\beta-\alpha-(\alpha+\beta+2)x$	$n(n+\alpha+\beta+1)$
$P_n(x)$	$1-x^2$	$-2x$	$n(n+1)$
$L_n^{(\alpha)}(x)$	x	$\alpha+1-x$	n
$H_n(x)$	1	$-2x$	$2n$

Table 3: Coefficients in equation (8) for corresponding polynomials.

3.2.2 Zeros of Bessel functions of first and second kind

Actually the schema we use is applicable to considerably larger class of functions. It is very simple, but is based on deep results due to (Gil, 2003) and (Segura, 2002). We firstly give the result in the special case we are interested and then attempt to sketch the idea behind, allowing to understand why this schema works. For detailed and profound explanation and much more results we refer the reader to the papers cited above.

Suppose we want to calculate the first *N* zeros of Bessel function of first kind with index *p*, denoted as usually by $J_p(x)$. For each successive zero we use fixed point iteration

$$x_{k+1} = x_k + \arctan\left(\frac{J_p(x_k)}{J_{p+1}(x_k)}\right) \qquad (9)$$

For the initial point of the iteration we get

$x_0^{(n)} = x_*^{(n-1)} + \Delta^{(n)}$ where $x_*^{(n)}$ denotes the n-th zero and for $\Delta^{(n)}$ we distinguish two cases. If the number of currently founded zeros is less than two, we pose $\Delta^{(n)} = \frac{\pi}{2}$, otherwise we pose $\Delta^{(n)} = x_*^{(n-1)} - x_*^{(n-2)}$ (which gives us a better approximation). The fixed point iteration (9) is globally convergent and its order of convergence is 2. In (9) we can replace $J_p(x)$ with an arbitrary Bessel function of first kind, second kind or a linear combination of them (provided they have same index).

This schema is applicable to function, which is solution of linear homogeneous ordinary differential equation of second order

$$y'' + B_y(x)y' + A_y(x)y = 0 \qquad (10)$$

with continuous $A_y(x)$ and $B_y(x)$ in some interval I and also satisfies the following supplementary conditions: There exists a so called contrast function, which is solution of an equation

$$w + B_w(x)w + A_w(x)w = 0 \qquad (11)$$

with continuous $A_w(x)$ and $B_w(x)$ in I in such a way, that for every two pairs of fundamental solutions $\{y^{(1)}, y^{(2)}\}$ and $\{w^{(1)}, w^{(2)}\}$ of (10) and (11), these solutions are related by a first order system

$$\begin{cases} y' = \alpha(x)y + \delta(x)w \\ w' = \beta(x)w + \gamma(x)y \end{cases} \qquad (12)$$

with continuous coefficients $\alpha(x)$, $\beta(x)$, $\gamma(x)$, $\delta(x)$ in I.

Our first goal is to transform this system to the form of equation (19), where H depends implicitly on y and w, and for which a globally fixed point iteration is applicable. We achieve this goal in several steps.

The first transformation changes the dependent variables

$$\begin{cases} y(x) = \lambda_y(x)\overline{y}(x) \\ w(x) = \lambda_w(x)\overline{w}(x) \end{cases} \qquad (13)$$

with continuous, different from zero $\lambda_y(x)$ and $\lambda_w(x)$ in I, so that $\overline{y}(x)$ and $\overline{w}(x)$ satisfy the system

$$\begin{cases} \overline{y}' = \overline{\alpha}\,\overline{y} + \overline{\delta}\,\overline{w} \\ \overline{w}' = \overline{\beta}'\overline{w} + \overline{\gamma}\,\overline{y} \end{cases} \qquad (14)$$

with $\overline{\delta} > 0$ and $\overline{\delta} = -\overline{\gamma}$. This can be done posing

$$\lambda_y = \text{sign}(\delta)\lambda_w\sqrt{-\frac{\delta}{\gamma}} \qquad (15)$$

where the sign of the square root is positive if y and/or w has at least two roots in I. The new functions $\overline{y}$ and $\overline{w}$ have the same roots as y and w.

After the change of variable

$$z(x) = \int^x \overline{\delta}(s)\mathrm{d}s \qquad (17)$$

the system becomes

$$\begin{cases} \dot{\overline{y}} = \overline{a}\,\overline{y} + \overline{w} \\ \dot{\overline{w}} = \overline{b}\,\overline{w} - \overline{y}, \end{cases} \qquad (18)$$

where $\overline{a} = \overline{\alpha}/\overline{\delta}$, $\overline{b} = \overline{\beta}/\overline{\delta}$ and the dots denote derivations with respect to z. From this system we can see that the ratio $H(z) = \overline{y}/\overline{w}$ satisfies the nonlinear ODE of first order

$$\dot{H} = 1 + H^2 - 2\eta H \qquad (19)$$

where $\eta = (\overline{b} - \overline{a})/2$. For small η H has a behaviour, that is similar to the tangent function with zeros and singularities interlaced. In (Segura, 2002) it is shown that for ratio H of this type, satisfying (19) the fixed point iteration

$$T(z) = z - \arctan(H(z)) \qquad (20)$$

is globally convergent to the zeros of the function $y(x(z))$ in intervals where η does not change sign. Here $x(z)$ is the inverse of $z(x)$ in (17).

Now we are much closer to the example in the beginning of this section. Let an equation

$$y_k''(x) + B_k(x)y_k'(x) + A_k(x)y_k(x) = 0 \qquad (21)$$

is given, with a continuous dependence of k and two families of independent solutions $\{y_k^{(1)}\}$ and $\{y_k^{(2)}\}$ satisfying the following difference-differential equations

$$\begin{aligned} y_k'(x) &= a_k(x)y_k(x) + d_k(x)y_{k-1}(x), \\ y_{k-1}'(x) &= b_k(x)y_{k-1}(x) + e_k(x)y_k(x). \end{aligned} \qquad (22)$$

These equations give us the possibility to build contrast function for y_n (if permitted be the range of k) in two ways, taking $k=n$ and $k=n+1$. Following the steps described above we can build two fixed point iterations

$$T_i(z) = z - \arctan\left(H_i(z)\right), \quad i = \pm 1 \qquad (22)$$

where

$$H_i(z) = -i\text{sign}(d_{n_i})K_i\frac{y_n(x(z))}{y_{n+i}(x(z))}, \qquad \begin{cases} n, & i = -1 \\ n+1, & i = +1 \end{cases} \qquad (23)$$

and

$$z(x)=\int^{x}\sqrt{-d_{n_i}(s)e_{n_i}(s)}\,\mathrm{d}s\,,\quad K_i=\left(-\frac{d_{n_i}}{e_{n_i}}\right)^{\frac{i}{2}}. \qquad (24)$$

And now, to be quite specific, we rewrite (22) for $y_n(x)=J_n(x)$ (this is very special case of more general relations for linear combinations of Bessel functions, see for example 10.6.6 in (Olver, 2010))

$$\begin{aligned} J'_k(x) &= -\frac{k}{x}J_k(x)+J_{k-1}(x),\\ J'_{k-1}(x) &= \frac{k-1}{x}J_{k-1}(x)-J_k(x). \end{aligned} \qquad (25)$$

Taking $i = 1$, in this case we have $d_k(x) = 1$, $e_k(x) = -1$, $z(x) = x$, $K_1 = 1$ and (skipping the index i)

$$\begin{aligned} H(z) &= -\frac{J_n(x)}{J_{n+1}(x(z))},\\ T(x) &= x+\arctan\left(\frac{J_n(x)}{J_{n+1}(x(z))}\right). \end{aligned} \qquad (26)$$

which is exactly the starting example.

4 CONCLUSIONS

The paper presents a part of the currently done implementation of a library and tools for computing with arbitrary accuracy in environment, not typical for this purpose, namely .NET Framework. The combination of well selected methods and the excellent possibilities for integration of different functionalities in this environment are worth the efforts, and enable the achievement of non-trivial results even on a home computer. The applications, herein presented, though useful, are mainly illustrations of the library possibilities and could be further improved in different aspects, including the use of the multi-core architecture of the modern processors, executing parallel calculations at some stages of the algorithms used.

ACKNOWLEDGEMENTS

The research work reported in the paper is partially supported by the project AComIn “Advanced Computing for Innovation”, Grant 316087, funded by FP7 Capacity Programme (Research Potential of Convergence Regions).

REFERENCES

Olver, J., Lozier, D., Boisvert, R., Clark. C., 2010. *NIST Handbook of Mathematical Functions*, Cambridge University Press.

Abramovitz, M., Stegun, I., 1964. *Handbook of Mathematical Functions*, Dover Publications.

Traub, J., 1964. *Iterative Methods for the solution of Equations*, Prentice-Hall. London.

Ostrowski, A., 1966. *Solution of Equations and Systems of Equations*, Academic Press. New York and London, 2nd edition.

McNamee, J., 2007. *Numerical Methods for Roots of Polynomials - Part I*, Elsevier.

Scharma, R., 2011. *Iterative Methods for the Solution of Nonlinear Equations: A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Mathematics*, Department of Mathematics Sant Longowal Institute of Engineering and Technology Longowal- 148 106, District Sangrur, Punjab (India).

Gil, A., Segura, J., 2003. *Computing the Zeros and Turning Points of Solutions of Second Order Homogeneous Linear ODES*, Siam J. Numer. Anal., Vol. 41, No. 3, pp. 827-855.

Segura, J., 2002. *The Zeros of Special Functions From a fixed Point Method*, SIAM J. Numer. Anal., 40 (2002), pp. 114-133.

Allgower, E., Georg, K., 1990. *Numerical Continuation Methods*, Springer-Verlag.

Volpi, L., 2006. *Non Linear Equations: Practical Methods for Root Finding, Part II - Zeros of Polynomials*, Foxes Team.

Szegö, G., 1939. *Orthogonal Polynomials*, American Mathematical Society.

Judd, K., 1998. *Numerical Methods in Economics*, The MIT Press, Cambridge, Massachusetts, London England.

Rahimian, S., Jalali, F., White, R., 2011. *A New Homotopy for Seeking All Real Roots of a Nonlinear Equation*, Computers and Chemical Engineering 35 (2011) 303-411.

Krtolica, P., Stanimirovi'c, P., 2001. *Symbolic Derivation without Using Expression Tree*, Yugoslav Journal of Operations Research 11 (2001) Number 1. pp. 61-75.

Dzhambov, V., Drangajov, S., 2011. *Computing of Special Function with Arbitrary Precision in the Environment of .NET Framework*, Cybernetics and Information Technologies, Vol. 11, No. 2, 2011, pp. 32-45.

Dzhambov, V., 2014. *High Precision Computing of Definite Integrals with .NET Framework C# and X-MPIR*, Cybernetics and Information Technologies, Vol. 14, No. 1, 2014, pp. 172-182.

SEMI FUZZY NEURAL NETWORK. PART 1: NONLINEAR SYSTEM IDENTIFICATION

Margarita Terziyska, Lyubka Doukovska
Institutes of Information and Communication Technologies, BAS,
Acad. G. Bonchev st., block 2, floor 5, 1113 Sofia, Bulgaria
terziyska@dir.bg, l.doukovska@mail.bg

Keywords: Semi fuzzy neural network, Fuzzy-neural models, Nonlinear identification, Takagi-Sugeno fuzzy inference, NARX model, Chaotic time series

Abstract: Semi Fuzzy Neural Network (SFNN) model with a second order Takagi-Sugeno inference mechanism is presented in this paper. This kind of model is similar in structure to a linear model, therefore, can also be called semi- or quasi-linear. The proposed approach lies on the simple idea to reduce the number of the fuzzy rules and the computational burden, when they are modeling nonlinear systems. As a learning procedure for the designed structure it is used two step gradient descent algorithm. To demonstrate the potentials of the SFNN model test experiments with two benchmark chaotic systems - Mackey-Glass and Rossler chaotic time series are studied. The proposed simplified model structure shows an accurate performance and may be a promising solution to model complex nonlinear processes in a model predictive control schemes.

1 INTRODUCTION

Neural networks and fuzzy logic are used for identification of complex nonlinear industrial processes for many years. They are well known as universal approximators. The fusion of the fuzzy logic with the neural networks allows combining the learning and computational ability of neural networks with the human like IF-THEN thinking and reasoning of fuzzy system. Combining neural networks and fuzzy systems in one unified framework has become popular in the last few years.

A lot of architectures have been proposed in the literature that combines fuzzy logic and neural networks. Some of the most popular are ANFIS (Jang, 1993), DENFIS (Kasabov, 2002). They are all composed of a set of if-then rules. In principle, the number of fuzzy rules dependent exponentially on the number of inputs and membership functions. If **n** is number of inputs and **m** is number of membership functions, then the required number of fuzzy rules is $\mathbf{m}^{\mathbf{n}}$. The huge number of rules leads to determination of a large number of parameters during the learning procedure, which makes the neural fuzzy system difficult to implement.

In order to reduce the number of fuzzy rules without accuracy losses different fuzzy clustering approaches can be used, such as fuzzy C-means clustering (Žalik, 2006), K-means clustering (Gallucc et al, 2012). Evolving fuzzy systems (Kasabov, 2002), (Kasabov and Filev, 2006), (Angelov, 2013) are used to evolve clustering and formed dynamically bases on fuzzy rules that had been created during the past learning process.

Another possibility to reduce the number of fuzzy rules gives self-constructing and self-organizing fuzzy neural network structures (Allende-Cid et al, 2008), (Ferreyra, 2006). In this type of structures, during the training procedure, inactive rules are dropped, which consequently leads to a reduction in the number of trained parameters.

In order to deal with the rule-explosion problem, hierarchical fuzzy neural networks could be used. But the normal learning method in these structures is very complex (Wen Yu et al, 2007). A method that compresses a fuzzy system with an arbitrarily large number of rules into a smaller fuzzy system by removing the redundancy in the fuzzy rule base is presented in (Gegov, 2007). Review of the most of existing rule base reduction methods for fuzzy systems, summary of their attributes and introduction of advanced techniques for formal presentation of fuzzy systems based on Boolean matrices and binary relations, which facilitate the overall management of complexity, is made in (Gegov et al, 2007).

In this paper it is presented a SFNN model that works with a reduced number of fuzzy rules. Firstly, to demonstrate the potentials of the SFNN model test experiments with two benchmark chaotic systems - Mackey-Glass and Rossler chaotic time series are studied. After that, the SFNN model is incorporated in GPC and its efficiency is tested by simulation experiments in MATLAB environment to control a CSTR.

2 CLASSICAL FUZZY NEURAL NETWORK

In this section it is described so-called Classical Fuzzy Neural Network (CFNN).It is so named to distinguish it from the proposed in next section SFNN. The structure of CFNN is shown on figure 1.
Layer 1: This layer accepts the input variables and then nodes in this layer only transmit the input values to the next layer directly.
Layer 2: Each node in this layer does the fuzzyfication via Gaussian membership function:

$$\mu_{Xp,m}^{(n)} = \exp\frac{-(x_p - c_{Xp,m})^2}{2\sigma_{Xp,m}^2} \quad (1)$$

where Xp are the input values, cXp,m and σXp,m are the center and the standard deviation of the Gaussian membership function.
Layer 3: This layer is a kind of rules generator as it formed the fuzzy logic rules. Their number depends on the number of inputs p and the number of their fuzzy sets m, and is calculated according to the expression N=mp. In this layer, each node represents a fuzzy in following form:

$$R^{(i)} : if\ x_1\ is\ \tilde{A}_1^{(i)}\ and\ x_p\ is\ \tilde{A}_p^{(i)}\ then\ f_y^{(i)}(k) \quad (2)$$

$$f_y^{(i)}(k) = a_1^{(i)} y(k-1) + a_2^{(i)} y(k-2) + \cdots + a_{ny}^{(i)} y(k-n_y) + b_1^{(i)} u(k) + b_2^{(i)} u(k-1) + \cdots + b_{nu}^{(i)} u(k-n_u) + c_0^{(i)} \quad (3)$$

Layer 4: In the fourth layer is realized implication operation:

$$\mu_{yq}^{(n)}(k+j) = \mu_{x_1,m}^{(n)}(k+j) * \mu_{x_2,m}^{(n)}(k+j) * \ldots * \mu_{x_p,m}^{(n)}(k+j) \quad (4)$$

Layer 5: In the fifth last layer takes the removal decision which consists in determining the value of the model output by the expression:

$$\hat{y}(k+j) = \frac{\sum_{i=1}^{q} f_y^{(i)}(k+j)\mu_y^{(i)}(k+j)}{\sum_{i=1}^{q} \mu_y^{(i)}(k+j)} \quad (5)$$

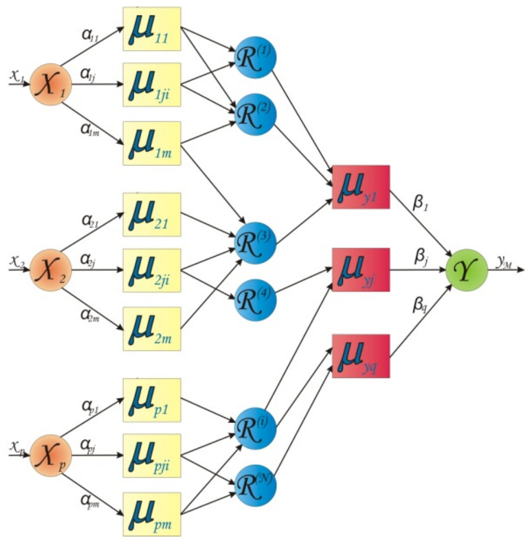


Figure 1: Structure of Classical Fuzzy Neural Network.

The learning algorithm for the CFNN model is very simple. It is based on minimization of an instant error measurement function between the real plant output and the process output, calculated by the CFNN:

$$E(k) = \frac{(y(k) - \hat{y}(k))^2}{2} \quad (6)$$

where y(k) denotes the measured real plant output and ŷ(k) is the CFNN model output.
During the gradient learning procedure it is needed to adjusted two groups of parameters in CFNN and they are: premise and consequent parameters on the fuzzy rules. The consequent parameters are the coefficients a1, a2...any, b1, b2...bnu in the Sugeno function fy and they are calculated at first step by the following equations:

$$\beta_{ij}(k+1) = \beta_{ij}(k) + \eta(y(k) - y_M(k))\bar{\mu}_y^{(j)}(k)x_i(k) \quad (7)$$

$$\beta_{0j}(k+1) = \beta_{0j}(k) + \eta(y(k) - y_M(k))\bar{\mu}_y^{(j)}(k) \quad (8)$$

where η is the learning rate and β_{ij} is an adjustable i-th coefficient (a_i or b_i) in the Sugeno function of the j-th activated rule.

The premise parameters are the centre c_{ij} and the deviation σ_{ij} of a Gaussian fuzzy set. They can be calculated at second step using the following equations:

$$c_{ij}(k+1) = c_{ij}(k) + \eta(y - y_M)\bar{\mu}_y^{(j)}(k)[f_y^{(i)} - \hat{y}(k)]\frac{[x_i(k) - c_{ij}(k)]}{c_{ij}^2(k)} \quad (9)$$

$$\sigma_{ij}(k+1) = \sigma_{ij}(k) + \eta(y - y_M)\bar{\mu}_y^{(j)}(k)[f_y^{(i)} - \hat{y}(k)]\frac{[x_i(k) - \sigma_{ij}(k)]^3}{\sigma_{ij}^2(k)} \quad (10)$$

3 SEMI FUZZY NEURAL NETWORK

The structure of the proposed SFNN model is shown on Fig. 2. SFNN model is a modification of the CFNN model described in previous section. Actually, the SFNN model is also five-layer architecture with the Takagi-Sugeno inference mechanism. However, in SFNN model a part of input signals are not fuzzyfied, but they come with their real values, weighted by the appropriate coefficient, into the third layer (fuzzy rules layer), i.e. directly into the THEN part of the functions of Sugeno. Thus, on the one hand it is reduced the number of fuzzy rules with which the model is working and on the other hand – it is reduced the number of the parameters that must be determined during learning procedure. For example, for implementation of NARX model with 4 inputs, each of which is fuzzyfied with 3 fuzzy sets, then for CFNN model is needed 81 fuzzy rules while SFNN works with only 9 fuzzy rules. Furthermore, in CFNN model during the learning procedure it is needed to obtain 405 linear and 648 nonlinear parameters or total 1053 parameters! The number of parameters that must be determined during the training procedure is only 81 (45 linear and 36 nonlinear parameters).

SFNN model is similar in structure to a linear model, therefore, it can also be called semi- or quasi-linear. This ensure reduced computational burden during the optimization procedure for the calculation of the optimal value of the control action, when the SFNN model is used as a part of a NMPC algorithm.

In this paper on the basis on SFNN model it is realized a NARX model. Two types of SFNN models are used during the simulation experiments. The first one is a case in which the values of vector-regressor **y**(y(k-1), y(k-2)) are fuzzyfied inputs and the values of vector-regressor **u** (u(k), u(k-1)) are nonfuzzyfied inputs – SFNN Type I. The second one is an opposite – the values of vector-regressor **y**(y(k-1), y(k-2)) are nonfuzzyfied inputs and the values of vector-regressor **u** (u(k), u(k-1))are fuzzyfied inputs – SFNN Type II .

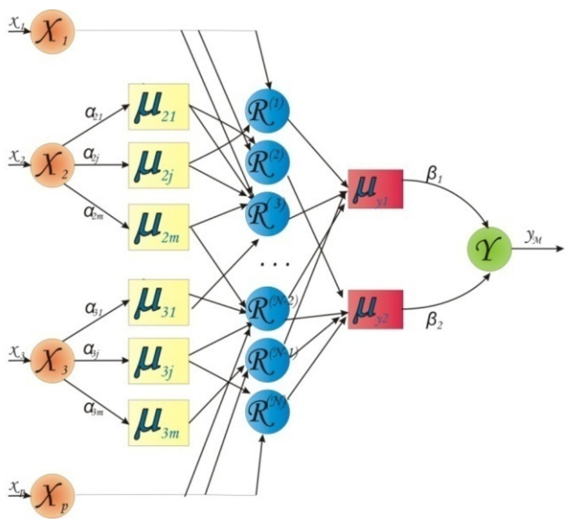


Figure 2: Structure of the proposed Semi Fuzzy Neural Network.

4 SIMULATION RESULTS

To investigate the 20odelling potentials of the proposed SFNN model, Mackey-Glass (MG) and Rossler chaotic system series benchmark models, have been used. The used time series will not converge or diverge, and the trajectory is highly sensitive to initial conditions. The MG time series is described by time-delay differential equation:

$$x(i+1) = \frac{x(i) + ax(i-s)}{(1 + x^C(i-s) - bx(i))} \quad (11)$$

where a=0.2; b=0.1; C=10; and x(0)=0.1 and s= 17s. The results are given on Fig. 3 and Fig. 4.

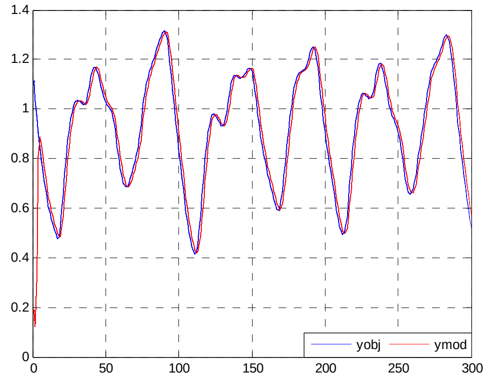


Figure 3: Model validation of the proposed SFNN Type I model by using Mackey-Glass chaotic time series.

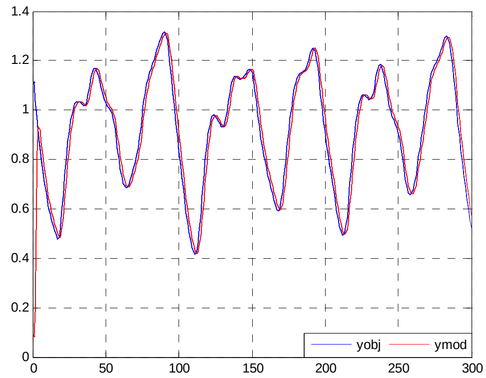


Figure 5: Model validation of the proposed SFNN Type II model by using Mackey-Glass chaotic time series.

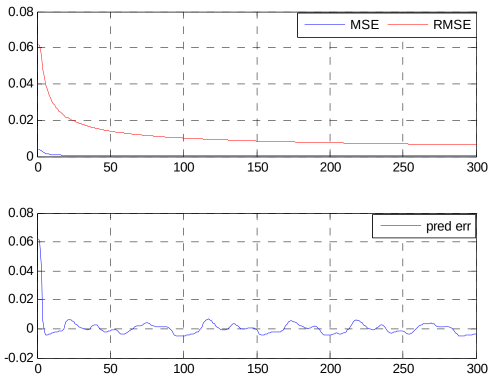


Figure 4: Mean Square Error (MSE), Root Mean Square Error (RMSE) and predicted error (SFNN Type I).

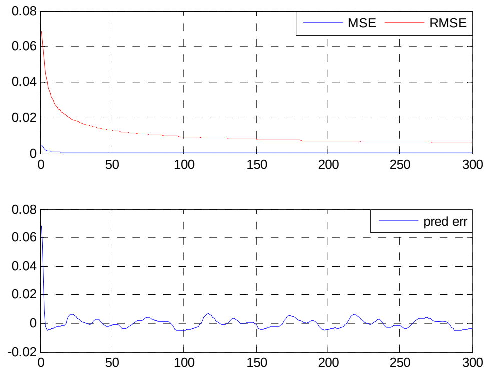


Figure 6: Mean Square Error (MSE), Root Mean Square Error (RMSE) and predicted error (SFNN Type II).

On Fig. 3 and Fig. 4 are shown results on SFNN Type I validation by using Mackey-Glass chaotic time series. As it can be seen, the proposed model structure predicts accurately the generated time series, with minimum prediction error and fast transient response of the RMSE, reaching value closer to zero. The values of the RMSE and MSE in the 50-th time step are respectively 0,014 and 0.00019.

Similarly, on Fig. 5 and Fig. 6 are shown results on SFNN Type II validation by using Mackey-Glass chaotic time series.

Another test of the proposed SFNN model is made with Rossler chaotic time series. These series is described by three coupled first-order differential equations:

$$\frac{dx}{dt} = -y - z$$
$$\frac{dy}{dt} = x + ay \tag{12}$$
$$\frac{dz}{dt} = b + z(x - c)$$

where a=0.2; b=0.4; c=5.7; initial conditions x_0=0.1; y_0=0.1; z_0=0.1. The validation of SFNN Type I and SFNN Type II models with Rossler chaotic time series are given respectively on Fig.7, Fig.8 and Fig 9, Fig. 10.

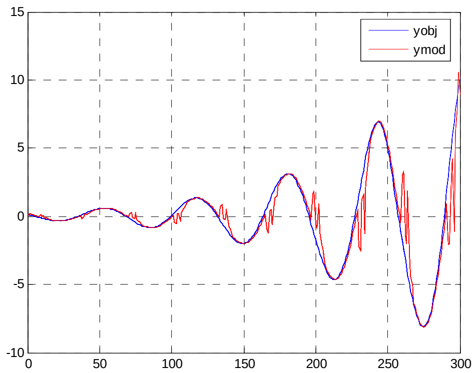


Figure 7: Model validation of the proposed SFNN Type I model by using Rossler chaotic time series.

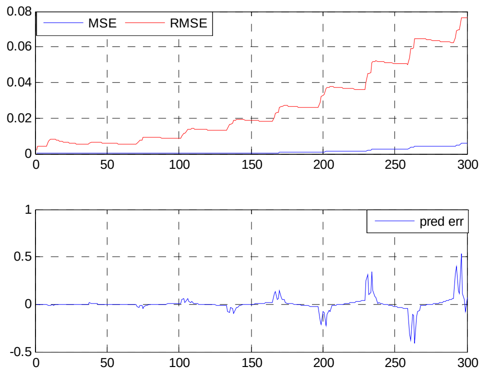


Figure 8: Mean Square Error (MSE), Root Mean Square Error (RMSE) and predicted error (SFNN Type I).

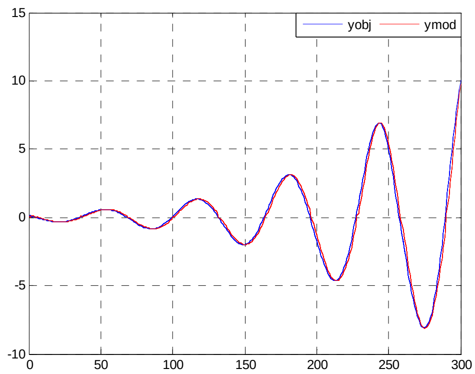


Figure 9: Model validation of the proposed SFNN Type II model by using Rossler chaotic time series.

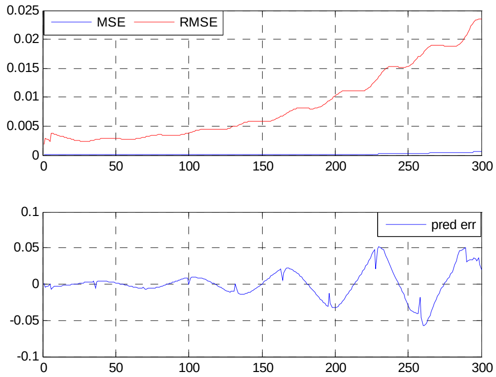


Figure 10: Mean Square Error (MSE), Root Mean Square Error (RMSE) and predicted error (SFNN Type II).

The values of MSE and RMSE in interval of 50 steps are summarized in Table 1. From these data it can be concluded that the SFNN Type II model is more accurate than the SFNN Type I model, ie the case that u(k), u(k-1) are fuzzyfied inputs.

4 CONCLUSIONS

In this paper it is presented Semi Fuzzy Neural Network (SFNN) model. To demonstrate the potentials of the SFNN model test experiments with two benchmark chaotic systems - Mackey-Glass and Rossler chaotic time series are studied. The proposed SFNN model predicts accurately the generated time series, with minimum prediction error and fast transient response of the RMSE, reaching values closer to zero. The main advantage of SFNN is that operated by a small number of rules and respectively has a smaller number of parameters for learning. Thus, it carries out the modelling of non-linear systems with considerably less calculation in comparison with the classical FNN. Furthermore, SFNN has other advantages - it is not require a priori data and is not bound by additional procedures, such as clustering. This makes it suitable for real-time applications such as predictive controllers.

Table 1: SFNN Type I and SFNN Type II Comparison

Mackey-Glass chaotic time series					Rossler chaotic time series				
Steps	SFNN Type I		SFNN Type II		Steps	SFNN Type I		SFNN Type II	
	MSE	RMSE	MSE	RMSE		MSE	RMSE	MSE	RMSE
50	0,00019	0,014	0,000168	0,013	50	$8{,}06.10^{-5}$	0,00898	$8{,}3.10^{-6}$	0,0029
100	$9{,}8.10^{-5}$	0,0099	$8{,}85.10^{-5}$	0,0093	100	0,000109	0,0104	$1{,}4.10^{-5}$	0,0037
150	$6{,}81.10^{-5}$	0,0083	$6{,}16.\ 10^{-5}$	0,0079	150	0,00038	0,0195	$3{,}45.10^{-5}$	0,0058
200	$5{,}33.10^{-5}$	0,0073	$4{,}85.\ 10^{-5}$	0,007	200	0,00109	0,0331	0,000104	0,0102
250	$4{,}43.10^{-5}$	0,0067	$4{,}05.\ 10^{-5}$	0,006	250	0,00263	0,0531	0,00023	0,0150
300	$3.85.10^{-5}$	0,0062	$3{,}5.\ 10^{-5}$	0,0051	300	0,00595	0,0772	0,00055	0,0234

ACKNOWLEDGEMENTS

The research work reported in the paper is partly supported by the project ***AComIn*** "Advanced Computing for Innovation", grant 316087, funded by the FP7 Capacity Programme (Research Potential of Convergence Regions).

REFERENCES

Moore, R., Lopes, J., 1999. Paper templates. In *TEMPLATE'06, 1st International Conference on Template Production*. SciTePress.

Smith, J., 1998. *The book*, The publishing company. London, 2nd edition.

Jang, J.-S. Roger, "ANFIS: Adaptive-network-based fuzzy inference system", *IEEE Transactions on System, Man, and Cybernetics*, No. 5, Vol. 23, PP. 665-685, May 1993.

Kasabov, N. K., Qun Song, DENFIS: Dynamic Evolving Neural-Fuzzy Inference System and Its Application for Time-Series Prediction, *IEEE Transactions on Fuzzy Systems*, 10(2), April 2002.

Žalik, K.R., Fuzzy C-Means Clustering and Facility Location Problems, *Artificial Intelligence and Soft Computing*, 2006.

Gallucc, Laurent et al, Graph based k-means clustering, *Signal Processing*, Vol. 92, Issue 9, September 2012, pp. 1970-1984.

Kasabov, N. K. D. Filev, Evolving Intelligent Systems: Methods, Learning & Applications, International Symposium of Evolving Fuzzy Systems, September 2006.

Angelov, P., *Autonomous Learning Systems*, Wiley, 2013.

Allende-Cid, Hector et al, Self-Organizing Neuro-Fuzzy Inference System, *Iberoamerican Congress on Pattern Recognition CIARP*, pp. 429-436, 2008.

Ferreyra, A., Jose de Jesus Rubio, A new on-line self-constructing neural fuzzy network, *Proceedings of the 45th IEEE Conference on Decision & Control*, San Diego, CA, USA, December 13-15, 2006.

Yu, Wen et al, System Identification using hierarchical fuzzy neural networks with stable learning algorithm, *Journal of Intelligent & Fuzzy Systems*, 18 (2007) 171-183.

Gegov, A., Complexity Management in Fuzzy Systems, *Studies in Fuzziness and Soft Computing*, Volume 211, 2007.

Gegov, A., N. Gobalakrishnan, Advanced Inference in Fuzzy Systems by Rule Base Compression, *Mathware & Soft Computing* 14 (2007), 201-216.

SEMI FUZZY NEURAL NETWORK. PART 2: PREDICTIVE CONTROL

Margarita Terziyska, Lyubka Doukovska
Institutes of Information and Communication Technologies, BAS,
Acad. G. Bonchev st., block 2, floor 5, 1113 Sofia, Bulgaria
terziyska@dir.bg, l.doukovska@mail.bg

Keywords: Semi fuzzy neural network, Nonlinear predictive control, Generalized predictive controller, Levenberg-Marguard optimization algorithm, CSTR

Abstract: This paper presents a Nonlinear Model Predictive Controller based on Semi Fuzzy Neural Network (SFNN) model. SFNN works with small fuzzy rules base; not requires a priori data and is not bounds by any additional procedures, such as clustering, which make it suitable for real-time applications like predictive controllers. Levenberg-Marguard algorithm is used to compute the optimal control value in the designed Generalized Predictive Controller (GPC). The efficiency of the proposed controller is demonstrated by simulation experiments in MATLAB environment to control a Continuous Stirred Tank Reactor.

1 INTRODUCTION

Model Predictive Control (MPC) is advanced control methodology that originates in the late seventies. MPC is optimal control that relies on dynamic model of the process. The models in MPC are used to predict the future response of a plant. Then, MPC algorithm computes optimal control by minimizing a prescribed cost function. One of the key advantages of MPC is its ability to deal with input and output constraints while it can be applied to multivariable process control. That's why MPC very quickly became popular and nowadays MPC is a well-known, classical control method.

MPC algorithms based on linear models have been successfully used for years in numerous industrial applications (Qin and Badgwell, 2003). It is provided a review (Holkar et al, 2010) of the most commonly used methods that have been embedded in industrial model predictive control. Since, the nature of processes is inherently nonlinear and this implies the use of nonlinear models and respectively Nonlinear Model Predictive Control (NMPC) algorithms.

NMPC is a variant of MPC that is characterized by the use of nonlinear system models in the prediction. As in linear MPC, NMPC requires the iterative solution of optimal control problems on a finite prediction horizon. While these problems are convex in linear MPC, in nonlinear MPC they are not convex anymore. The nonlinear optimization task must be solved online. Therefore, the success of NMPC algorithm depends critically on mathematical model (Rawlings and Mayne, 2013). It is very important to find a predictive model that effectively describes the nonlinear behavior of the system and can easily be incorporated into NMPC algorithm. One possibility is to use first principle models such as nonlinear ordinary differential equations, partial differential equations, integro-differential equations and delay equations models. Such models can be accurate over a wide range of operating conditions, but they are difficult to develop for many industrial cases and may lead to numerical problems (e.g. stiffness, ill-conditioning). The other possibility is to use empirical or black–box models (e.g. neural networks, fuzzy models, polynomial models, Volterra series models). How to select nonlinear model for NMPC is detailed described in (Pearson, 2003).

Fuzzy neural network models are used widely in NMPC control schemes. Many methods have been proposed in the literature that combine fuzzy neural network and model predictive control algorithm. Different multilevel neuro-fuzzy predictive control algorithms are described in (Grosso et al, 2012).

A hybrid fuzzy-neuro model based predictive control (HFNMBPC) is proposed in (Paulusová et al, 2009). Predictive controller based on block-oriented nonlinear model is presented in (Terziyska et al, 2006). GPC using recurrent fuzzy neural networks as a model is described in (Lu and Tsai, 2007). Another GPC algorithm with feedforward neuro-fuzzy network is presented in (Liu and Chan, 2006).

In this paper it is presented a Nonlinear Model Predictive Controller based on Semi Fuzzy Neural Network (SFNN) model. SFNN works with small fuzzy rules base; not requires a priori data and is not bounds by any additional procedures, such as clustering, which make it suitable for real-time applications like predictive controllers. Levenberg-Marguard algorithm is used to compute the optimal control value in the designed Generalized Predictive Controller (GPC). The efficiency of the proposed controller is demonstrated by simulation experiments in MATLAB environment to control a Continuous Stirred Tank Reactor.

2 SEMI FUZZY NEURAL NETWORK MODEL

Semi Fuzzy Neural Network model is five-layer architecture with the Takagi-Sugeno inference mechanism and reduced fuzzy rules base. In this paper the SFNN model it is used to realized NARX model. More details about SFNN model can be found in first part of this paper (Terziyska, 2014). Here, it is examined two types of SFNN models:

- *SFNN Type I*

This is a case in which the values of vector-regressor **y**(y(k-1), y(k-2)) are fuzzyfied inputs and the values of vector-regressor **u** (u(k), u(k-1)) are nonfuzzyfied inputs. The SFNN Type I model output can be calculated by next expression:

$$\hat{y}_{MI}(k+j)=\frac{\sum_{i=1}^{q} f_y^{(i)}(k+j)\mu_y^{(i)}(k+j)}{\sum_{i=1}^{q}\mu_y^{(i)}(k+j)} \quad (1)$$

where f_y and μ_y are respectively the Sugeno output function and the corresponded membership degrees of the quantization levels and can be obtained by:

$$f_y^{(i)} = b_1^{(i)}u(k) + b_2^{(i)}(k-1) + ...+ b_{n_u}^{(i)}u(k-n_u) + b_o^{(i)} \quad (2)$$

$$\mu_y^{(i)} = \mu_{y1}^{(i)} * \mu_{y2}^{(i)} * ......\mu_{yp}^{(i)} \quad (3)$$

where **p** is the number of fuzzyfied inputs (in our case p=2).

- *SFNN Type II*

This is the case in which the values of vector-regressor **y**(y(k-1), y(k-2)) are nonfuzzyfied inputs and the values of vector-regressor **u** (u(k), u(k-1))are fuzzyfied inputs. The SFNN Type II model output can be calculated by next expression:

$$\hat{y}_{MII}(k+j)=\frac{\sum_{i=1}^{q} f_u^{(i)}(k+j)\mu_u^{(i)}(k+j)}{\sum_{i=1}^{q}\mu_u^{(i)}(k+j)} \quad (4)$$

where f_u and μ_u are respectively the Sugeno output function and the corresponded membership degrees of the quantization levels and can be obtained by:

$$f_u^{(i)} = b_1^{(i)}u(k) + b_2^{(i)}(k-1) + ...+ b_{n_u}^{(i)}u(k-n_u) + b_o^{(i)} \quad (5)$$

$$\mu_u^{(i)} = \mu_{u1}^{(i)} * \mu_{u2}^{(i)} * ......\mu_{up}^{(i)} \quad (6)$$

As a learning algorithm for both SFNN Type I and SFNN Type II it is used two steps gradient procedure (Todorov, 2012).

3 LEVENBERG-MARGUARD OPTIMIZATION ALGORITHM

Using the proposed SFNN model in its two forms (Type I and Type II), the *Optimization Algorithm* computes the future control actions at each sampling period, by minimizing the proposed by (Clarke et al, 1987) cost function which is typical for Generalized Predictive Control (GPC):

$$J(k,u(k)) = \sum_{i=N_1}^{N_2}(r(k+i) - \hat{y}(k+i))^2 + \rho\sum_{i=1}^{N_u}\Delta u(k+i-1)^2 \quad (7)$$

where $\hat{y}$ is the predicted model output ($\hat{y}_{MI}$ or $\hat{y}_{MII}$), r is the system reference, u is the control action, N_1 is the minimum prediction horizon, N_2 is the maximum prediction horizon, N_u is the control horizon and ρ is the weighting factor penalizing changes in the control actions.

In this paper it is used Levenberg-Marquardt optimization algorithm, which uses the approximated Hessian and the information in the gradient, taking into account some regularization factors. Important principal in Levenberg-Marguard method is that the cost function must be a quadratic one and the Hess matrix must be positive definite. The algorithm iterates using the following general equation:

$$x^{(k+1)} = x^{(k)} - [H^{-1}(x^{(k)}) + \lambda E]\nabla Q(x^{(k)}) \quad (8)$$

where H is the Hessian matrix with second partial derivatives as elements, E is the identity matrix and λ is the Levenberg-Marquardt parameter, which adjust the direction of movement to extremes, from gradient method at a great value ($\lambda > 10^3$) to the Newton optimization method when $\lambda = 0$.

When the criterion function is a quadratic one and there are no constraints on the control action, the cost function can be minimized analytically. If the criterion **J** is minimized with respect to the future control actions **u**, then their optimal values can be calculated by applying the condition:

$$\nabla J[k,U(k)] = \left[\frac{\partial J[k,U(k)]}{\partial u(k)}, \frac{\partial J[k,U(k)]}{\partial u(k+1)}, \ldots, \frac{\partial J[k,U(k)]}{\partial u(k+N_u-1)}\right] = 0 \quad (9)$$

Each element from this gradient vector can be calculated using the following equation:

$$\frac{\partial J[k,U(k)]}{\partial U(k)} = \left[-2\left[R(k)-\hat{Y}(k)\right]^T \frac{\partial \hat{Y}(k)}{\partial U(k)} + 2\rho\, \hat{U}(k)^T \frac{\partial \hat{U}(k)}{\partial U(k)}\right] \quad (10)$$

From the above expression can be seen that it is needed to obtain two groups of partial derivatives. The first one is $\left[\frac{\partial \hat{Y}(k)}{\partial U(k)}\right]$, and the second one is $\left[\frac{\partial \hat{U}(k)}{\partial U(k)}\right]$. Each element from the first group of partial derivatives is calculated with the following equations:

$$\frac{\partial \hat{y}(k)}{\partial u(k)} = \sum_{i=1}^{N} b_1^{(i)} \overline{\mu}_y^{(i)}(k) \quad (11)$$

$$\frac{\partial \hat{y}(k+1)}{\partial u(k)} = \sum_{i=1}^{N}\left[a_1^{(i)} \frac{\partial \hat{y}(k)}{\partial u(k)} + b_2^{(i)}\right] \overline{\mu}_y^{(i)}(k+1) \quad (12)$$

$$\frac{\partial \hat{y}(k+N_2)}{\partial u(k)} = \sum_{i=1}^{N}\left[\begin{array}{l} a_1^{(i)} \frac{\partial \hat{y}(k+N_2-1)}{\partial u(k)} + \ldots \\ + a_2^{(i)} \frac{\partial \hat{y}(k+N_2-2)}{\partial u(k)} \end{array}\right] \overline{\mu}_y^{(i)}(k+N_2) \quad (13)$$

The second group partial derivatives have the following form:

$$\frac{\partial \hat{U}(k)}{\partial U(k)} = \begin{bmatrix} \frac{\partial \Delta u(k)}{\partial u(k)} & \ldots & \frac{\partial \Delta u(k)}{\partial \Delta u(k+N_u-1)} \\ \ldots & \ldots & \ldots \\ \frac{\partial \Delta u(k+N_u-1)}{\partial u(k)} & \ldots & \frac{\partial \Delta u(k+N_u-1)}{\partial \Delta u(k+N_u-1)} \end{bmatrix} \quad (14)$$

Since $\Delta u(k) = u(k) - u(k-1)$, this is:

$$\frac{\partial \hat{U}(k)}{\partial U(k)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \quad (15)$$

So, this is the way to obtain the first derivative that is needed for the first order gradient optimization algorithms. But, the Newton algorithm belongs to the second order gradient optimization algorithms. So, it is needed to calculate the second partial derivatives of the cost function. In this case (13) can be rewritten as follows:

$$H[k,U(k)] = \left[\frac{\partial^2 J[k,U(k)]}{\partial u^2(k)}, \frac{\partial^2 J[k,U(k)]}{\partial u^2(k+1)}, \ldots, \frac{\partial^2 J[k,U(k)]}{\partial u^2(k+N_u-1)}\right] \quad (16)$$

Each element from this vector is calculated as follows:

$$H[k,U(k)] = \left[-2\left[R(k)-\hat{Y}(k)\right]^T \frac{\partial \hat{Y}(k)}{\partial U(k)} + 2\rho\, \hat{U}(k)^T \frac{\partial \hat{U}(k)}{\partial U(k)}\right]' =$$

$$= \left[\begin{array}{l} -2\left(\left[R(k)-\hat{Y}(k)\right]^T\right)' \frac{\partial \hat{Y}(k)}{\partial U(k)} - 2\left[R(k)-\hat{Y}(k)\right]^T \left(\frac{\partial \hat{Y}(k)}{\partial U(k)}\right)' + \ldots \\ + 2\rho\left(\hat{U}(k)^T\right)' \frac{\partial \hat{U}(k)}{\partial U(k)} + 2\rho\, \hat{U}(k)^T \left(\frac{\partial \hat{U}(k)}{\partial U(k)}\right)' \end{array}\right] \quad (17)$$

Since $\left(\frac{\partial \hat{Y}(k)}{\partial U(k)}\right)' = 0$ and $\left(\frac{\partial \hat{U}(k)}{\partial U(k)}\right)' = 0$, finally:

$$H[k,U(k)] = -2\left(\frac{\partial \hat{Y}}{\partial U(k)}\right)^2 + 2\rho\left(\frac{\partial \hat{U}(k)}{\partial U(k)}\right)^2 \quad (18)$$

After calculating the Hessian matrix it is need to obtain its inverse matrix, which is the main difficult in the Levenberg-Marguard algorithm.

Once calculated individual gradient elements, they are made equal to zero. As a result, produce a system of equations that can be solved to control effects $u(k)$, $u(k+1)$, …, $u(k+N_u+1)$:

$$\frac{\partial J[k,U(k)]}{\partial u(k)} = -2\hat{e}(k+N_1)\frac{\partial \hat{y}(k+N_1)}{\partial u(k)} - ... \\ -2\hat{e}(k+N_2)\frac{\partial \hat{y}(k+N_2)}{\partial u(k)} \\ +2\lambda\Delta u(k) - 2\lambda\Delta u(k+1) = 0 \quad (19)$$

$$\frac{\partial J[k,U(k)]}{\partial u(k+1)} = -2\hat{e}(k+N_1)\frac{\partial \hat{y}(k+N_1)}{\partial u(k+1)} - ... \\ -2\hat{e}(k+N_2)\frac{\partial \hat{y}(k+N_2)}{\partial u(k+1)} \\ +2\lambda\Delta u(k+1) - 2\lambda\Delta u(k+2) = 0 \quad (20)$$

$$\frac{\partial J[k,U(k)]}{\partial u(k+N_u-1)} = -2\hat{e}(k+N_1)\frac{\partial \hat{y}(k+N_1)}{\partial u(k+N_u-1)} - ... - \\ -2\hat{e}(k+N_2)\frac{\partial \hat{y}(k+N_2)}{\partial u(k+N_u-1)} + 2\lambda\Delta u(k+N_u-1) = 0 \quad (21)$$

This system of equations can be solved very simply by starting from the equation (21), which is calculated by the control action Δu (k + Nu-1). Resulting value is substituted in the previous equation and it is determined Δu (k + Nu-2). Similarly, the entire sequence is calculated by the control effects in the control horizon.

Implementation of the presented Levenberg-Marguard algorithm is simple and according to (8) is based on the following equation:

$$\Delta u(k) = \Delta u(k+1) + \rho^{-1}[H^{-1}(x^{(k)}) + \lambda E]^* \begin{bmatrix} e(k+N_1)\frac{\partial \hat{y}(k+N_1)}{\partial u(k)} + ... \\ +e(k+N_2)\frac{\partial \hat{y}(k+N_2)}{\partial u(k)} \end{bmatrix} \quad (22)$$

4 SIMULATION EXPERTIMENTS

It is used a nonlinear model of a continuous stirred tank reactor (CSTR) for the purpose of experimental investigations of the control system with the proposed SFNN based predictive controller. The dynamic equations of the nonlinear CSTR are given by (Ray, 1981) as follow:

$$\dot{x}_1 = -x_1 + D_a(1-x_1)\exp\left(\frac{x_2}{1+x_2/\varphi}\right) \quad (23)$$

$$\dot{x}_2 = -(1+\delta)x_2 + BD_a(1-x_1)\exp\left(\frac{x_2}{1+x_2/\varphi}\right) + \delta u \quad (24)$$

where x_1 and x_2 represent the dimensionless reactant concentration and the reactor temperature, respectively. The control action u is dimensionless cooling jacket temperature. The physical parameters in the CSTR model equations are Da, φ, B and δ which correspond to the Damkhler number, the activated energy, the heat of reaction and the heat transfer coefficient, respectively. Based on the nominal values of the system parameters, Da=0.072, φ=20, B=8 and δ=0.69, the open-loop CSTR exhibits three steady states (x1,x2)A=(0.144, 0.886), (x1,x2)B=(0.445, 2.75) and (x1,x2)C=(0.765, 4.705), where the upper and the lower steady states are stable, whereas the middle one is unstable. The control objective here is to bring the nonlinear CSTR from the stable equilibrium point (x1,x2)A to the unstable one (x1,x2)B. All of the results presented are based on the reactor temperature x2, that is y(t)=x2(t).

The number of the Gaussian fuzzy membership functions for each input variable is chosen to be equal to three. Their initial parameters are normalized and uniformly distributed into the universe of discourse ϵ [0, 1]. The simulations were performed in various cases with different plant parameters. The SFNN based predictive controller has the following parameters: N1=1; Np=5; Nu=3; ρ=0.15. The results are shown on Fig. 1 and Fig. 2.

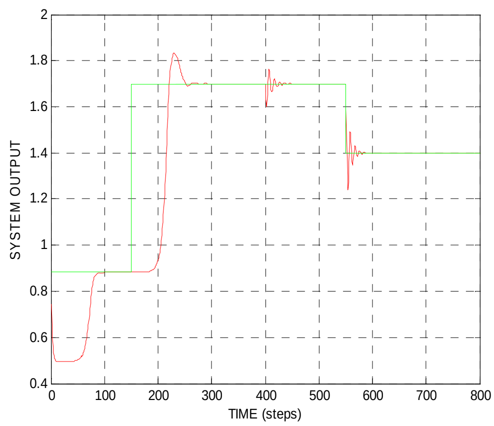


Figure 1: Transient process response in case of variable system reference and plant parameter changes in a case of SFNN Type I.

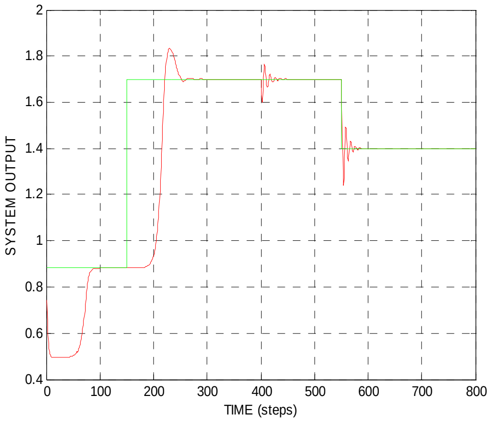


Figure 2: Transient process response in case of variable system reference and plant parameter changes in a case of SFNN Type II.

5 CONCLUSIONS

This paper presents a Nonlinear Model Predictive Controller based on two types Semi Fuzzy Neural Network (SFNN) model – SFNN Type I and SFNN Type II. As an optimization algorithm it is used Levenberg-Marguard algorithm. The efficiency of the proposed controller is demonstrated by simulation experiments in MATLAB environment to control a Continuous Stirred Tank Reactor. It was shown for both cases (SFNN Type I and SFNN Type II) that the NMPC with LM optimization approach ensures a reliable system performance in case of variable system reference and plant parameter changes. The proposed LM algorithm is rapidly converging when an analytical minimization approach is considered. Its main disadvantage is the need to obtain the invert Hesse matrix during the algorithm iteration. It is important to note, that the considered two types of SFNN models have insignificant differences in the accuracy of prediction of chaotic system series, but generally provide identical control of CSTR.

ACKNOWLEDGEMENTS

The research work reported in the paper is partly supported by the project ***AComIn*** "Advanced Computing for Innovation", grant 316087, funded by the FP7 Capacity Programme (Research Potential of Convergence Regions).

REFERENCES

Qin,S. J., T. A. Badgwell, A survey of industrial model predictive control technology, *Control Eng. Pract.* 11(2003)733–764.

Holkar, K. S., L. M. Waghmare, An Overview of Model Predictive Control, *International Journal of Control and Automation* Vol. 3 No. 4, December, 2010

Pearson, R. K., Selecting nonlinear model structures for computer control, *Journal of Process Control* 13, 2003.

Grosso, J.M., C. Ocampo-Martınez, V. Puig, Adaptive Multilevel Neuro-Fuzzy Model Predictive Control for Drinking Water Networks, 2012 *20th Mediterranean Conference on Control & Automation (MED)*, Barcelona, Spain, July 3-6, 2012.

Paulusová, J. ,M. Dúbravská, Hybrid Predictive Controller Based on Fuzzy-Neuro Model, *17th International Conference on Process Control*, Slovak Republic, June 9 – 12, 2009.

Terziyska, M., Y. Todorov, A. Mitev, M. Pertov, Nonlinear Model Predictive Controller using a Fuzzy-Neural Hammerstein model, *CEEPUS International Summer School Kosice, Slovak Republic*, 2006.

Lu, Chi-Huang, Ching-Chih Tsai, Generalized predictive control using recurrent fuzzy neural networks for industrial processes, *Journal of Process Control* 17 (2007) 83–92.

Liu, X. J., C. W. Chan, Neuro-Fuzzy Generalized Predictive Control of Boiler Steam Temperature, *IEEE Transactions on Energy Conversion*, Vol. 21, No. 4, December 2006.

Clarke, D. W., C. Mohtai, P. S. Tuffs, "Generalized predictive control – part I. The basic algorithm", *Automatica*, vol. 23, no. 2, 1987, pp. 137-148.

Ray, W. H., *Advanced Process Control*, McGraw-Hill, New York, 1981.

Terziyska, M., L. Doukovska, Semi Fuzzy Neural Network. Part 1: Nonlinear System Identification, International Workshop "Advanced Control and Optimisation: Step Ahead" ACOSA, 8-10 May, 2014, Bankya.

Todorov, Y., M. Terziyska, S. Ahmed, M. Petrov, Fuzzy-Neural Predictive Control using Levenberg-Marquardt optimization approach, In Proceedings of International IEEE conference on Innovations in Intelligent Systems and Applications, INISTA'2013, Albena, Bulgaria, pp. 1-5, 2013.

SOLUTION ESTIMATES FOR THE DISCRETE-TIME PARAMETER DEPENDENT LYAPUNOV EQUATION

Svetoslav Savov, Ivan Popchev
Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Acad. G. Bonchev str. Bl.2, Sofia, Bulgaria,
savovsg@mail.bg

Keywords : Discrete-time systems, Stability, Lyapunov equation.

Abstract: This paper studies the solution bounds derivation problem for one class of parameter dependent discrete-time algebraic Lyapunov equations (DALEs). The coefficient matrix is assumed to be a convex combination of *N* stable matrices. Several upper parameter independent bounds for the solution, its maximal eigenvalue and trace are proposed. The main contribution consists in the fact that fixed bounds are obtained for a parameter dependent solution, which can't be directly computed.

1 INTRODUCTION

The Lyapunov equation arises in many different perspectives such as control and system theory, identification, optimization, boundary value problems, power systems, signal processing, etc. Due to broad applications (Kouikoglu and Phillis, 1993), (Lee and Lee, 2001), (Lee, Li and Kung, 1993), (Wang and Lin, 1987), (Yedavalli, 1993), etc., this equation has been a subject of active research during the past 60 years. Digital technology also spelled the need for the solution of the Lyapunov equation for discrete-time systems, called the DALE.

The problem of valid bounds computation for the solution of the algebraic Lyapunov equation has a long history. A survey on bounds for the solution of algebraic equations in control theory may be found in (Savov, 2014). In some cases, the direct solution of this equation is impossible, due to its high order and in other ones, it is sufficient to have at disposal only some estimates for it.

This paper is an attempt to study the solution bounds derivation problem for one class of parameter dependent DALEs. Since it is not realistic to expect, that the exact model of any system is available, it is assumed that the coefficient matrix is not fixed, but it is subjected to a structured real parametric uncertainty, belonging to a compact vector set, namely, the unit simplex. The problems in getting solution bounds, especially upper ones, are well known (Savov, 2014). Additional difficulties arise here from the simple fact, that the coefficient matrix is not exactly known any more, in this case.

Several upper parameter independent bounds for the solution, its maximal eigenvalue and trace are proposed. The main contribution consists in the fact that fixed bounds are obtained for a parameter dependent solution, which can't be directly computed. By making use of them, a quadratic performance index associated with a discrete-time dynamic system can be easily estimated.

2 PRELIMINARIES

The following standard notations will be used in the sequel: $M > N, M \geq N$ means that $M - N$ is a positive definite, positive semi-definite matrix, respectively; the maximal eigenvalue of a symmetric matrix M is $\lambda_1(M)$; a real N x 1vector with entries a_i is denoted $a = (a_i)^T \in \mathbf{R}^N$, $|a|$ is the sum of its entries and $a \geq 0$ means that its entries are nonnegative; I denotes the identity matrix; all matrices are n x n. Define also the vector sets:

$$\Omega_N \equiv \{\omega = (\omega_i)^T \in \mathbf{R}^N : \omega \geq 0, |\omega| = 1\}$$

$$\overline{\Omega}_N \equiv \{\overline{\omega}_i = (\omega_i)^T \in \mathbf{R}^N, i = 1,2,...,N : \omega_i = 1\}$$

Consider the parameter dependent DALE

$$A^T(\alpha)P(\alpha,\beta)A(\alpha) - P(\alpha,\beta) = -B(\beta) \qquad (1)$$

The perturbed coefficient and right-hand side matrices are given by:

$$A(\alpha) = \alpha_1 A_1 + \alpha_2 A_2 + ... + \alpha_a A_a, \ \alpha \in \Omega_a$$

$$B(\beta) = \beta_1 B_1 + \beta_2 B_2 + ... + \beta_b B_b, \ \beta \in \Omega_b$$

All matrices $A_i, B_j = B_j^T$ are fixed, while the vectors α, β are constant, but not fixed. Note also, that the uncertain vectors reflect the action of independent perturbation factors.

Definition 1. If $\alpha \in \Omega_a, \beta \in \Omega_b$, it is said that the vector couple (α, β) belongs to the set $\Omega_{a,b}$. If $\overline{\alpha}_i \in \overline{\Omega}_a, i = 1,2,..,a, \overline{\beta}_j \in \overline{\Omega}_b, \ j = 1,2,..,b$, it is said that the vector couples ($\overline{\alpha}_i, \overline{\beta}_j$) belong to the set $\overline{\Omega}_{a,b}$.

It is well known that the solution $P(\alpha,\beta)$ in (1) is a positive (semi-) definite matrix for all $(\alpha,\beta) \in \Omega_{a,b}$ for any given positive (semi-) definite matrix $B(\beta)$, if and only if $A(\alpha)$ is a stable in the discrete-time domain matrix, i.e. its eigenvalues lie inside the unit circle for all α . The main purpose of this research is to derive parameter independent upper bounds for the solution, its maximal eigenvalue and trace. Before that, the following important result will be presented.

Lemma 1. If M denotes some symmetric matrix, $M(\alpha), \alpha \in \Omega_a$ denotes a polytope with vertices $M_i, i = 1,2,..,a$,

$$\tilde{M}(\alpha) = M - M^T(\alpha)M(\alpha)$$

$$M_{ii} = M - M_i^T M_i \ ,$$

then the following statements hold for all vectors α, β and all integers i, j and s:

$$\tilde{M}(\alpha) > 0, \Leftrightarrow M_{ii} > 0, \qquad (2)$$

$$M - B(\beta) \geq 0, \Leftrightarrow M - B_j \geq 0 \qquad (3)$$

$$\tilde{M}(\alpha) - B(\beta) \geq 0 \Leftrightarrow M - M_i^T M_i - B_s \geq 0 \qquad (4)$$

Proof. Suppose that $\tilde{M}(\alpha) > 0, \ \forall \alpha$.

Since $\overline{\Omega}_a \subseteq \Omega_a$, then $M_{ii} > 0, \ \forall \overline{\alpha}_i \in \overline{\Omega}_a$, by necessity, which proves the necessity part of (2). If all matrices M_{ii} are positive definite, then for all α one has:

$$\overline{M}(\alpha) = \sum_{i=1}^{a} \alpha_i^2 M_{ii} + \sum_{i,j=1, i<j}^{a} \alpha_i \alpha_j (M_{ii} + M_{jj}) > 0 \quad (5)$$

Since

$$M_i^T M_i + M_j^T M_j \geq M_i^T M_j + M_j^T M_i,$$

it follows that

$$0 < \overline{M}(\alpha) \leq \tilde{M}(\alpha), \forall \alpha \qquad (6)$$

which proves the sufficiency part of the Lemma.

The proof of statement (3) is obvious and is omitted.

Suppose that matrix $\tilde{M}(\alpha) - B(\beta)$ is positive semi-definite on $\Omega_{a,b}$. This means that

$$M_{ii} - B_s \geq 0, \ \forall (\overline{\alpha}_i, \overline{\beta}_s) \in \overline{\Omega}_{a,b}$$

by necessity, which proves the necessity part of statement (4). Let all matrices $M_{ii} - B_s$ are positive semi-definite. Having in mind (5) and (6), for any given $s = 1,2,..,b$, one gets:

$$0 \leq \tilde{M}(\alpha) - B_s = M - M^T(\alpha)M(\alpha) - B_s$$

By making use of (3) the proof of the sufficiency part of statement (4) follows easily.

3 UPPER MATRIX BOUNDS

Denote

$$S(\alpha) = X - A^T(\alpha)XA(\alpha), \ \alpha \in \Omega_a \qquad (7)$$

$$S_i = X - A_i^T X A_i, \ i = 1,2,..., a \qquad (8)$$

Theorem 1. Let all matrices B_j in (1) are positive (semi-) definite. If there exists some positive definite matrix X, such that all matrices in (8) are positive definite, then the positive (semi-) definite solution of the DALE (1) has the following upper parameter dependent bounds:

$$P_{U1}(\alpha,\beta) = \overline{\lambda}_1(\alpha,\beta)X \qquad (9)$$

$$P_{U2}(\alpha,\beta) = A^T(\alpha)P_{U1}(\alpha,\beta)A(\alpha) + B(\beta) \quad (10)$$

$$\overline{\lambda}_1(\alpha,\beta) = max\, \lambda_1[B(\beta)S^{-1}(\alpha)] \qquad (11)$$

Proof. Obviously, $B(\beta)$ in (1) is a positive (semi-) definite on Ω_b matrix if and only if matrices B_j are positive (semi-) definite. If all matrices in (8) are positive definite, then $S(\alpha)$ in (7) is a positive definite on Ω_a matrix, in accordance with statement (2) applied for

$$M = X, M_i = X^{1/2} A_i, i = 1,2,..,a.$$

These facts guarantee that $A(\alpha)$ is a stable matrix for all vectors in Ω_a and $P(\alpha,\beta)$ in (1) is a positive (semi-) definite matrix for all vector couples $(\alpha,\beta)\in\Omega_{a,b}$, in accordance with Lyapunov's stability theory for discrete-time systems. From (11) it follows that

$$\overline{\lambda}_1(\alpha,\beta)S(\alpha) \ge B(\beta), (\alpha,\beta)\in\Omega_{a,b}$$

Denote

$$\Delta(\alpha,\beta) = P_{U1}(\alpha,\beta) - P(\alpha,\beta)$$

Having in mind (7), (9) and (11) one gets:

$$-B(\beta) \ge A^T(\alpha)P(\alpha,\beta)A(\alpha) - P(\alpha,\beta)$$

Using (1), it follows that

$$A^T(\alpha)\Delta(\alpha,\beta)A(\alpha) - \Delta(\alpha,\beta) \le 0$$

This is possible if and only if $\Delta(\alpha,\beta)$ is a positive (semi-) definite matrix, i.e.

$$P(\alpha,\beta) \le P_{U1}(\alpha,\beta), \quad \forall(\alpha,\beta)\in\Omega_{a,b}$$

This proves the bound (9). Taking into account (1) and (9) the bound (10) follows.

Since the purpose is to obtain parameter independent solution bounds, matrices (9) and (10) need to be additionally bounded.

Definition 2. A scalar $\overline{\lambda}_1$ is said to be the maximal eigenvalue of a parameter dependent matrix $N(\alpha,\beta)$ if $\overline{\lambda}_1$ is an eigenvalue of $N(\alpha,\beta)$ for some $(\alpha,\beta)\in\Omega_{a,b}$ and $\overline{\lambda}_1 \ge \lambda_1[N(\alpha,\beta)]$ for all $(\alpha,\beta)\in\Omega_{a,b}$.

Lemma 2. Under the assumptions made in Theorem 1, the maximal eigenvalue of the matrix product $N(\alpha,\beta) = B(\beta)S^{-1}(\alpha)$ in (11) is

$$\overline{\lambda}_1 = \lambda_1(B_j S_i^{-1}), i = 1,..,a, \ j = 1,..,b \qquad (12)$$

Proof. If the assumptions in Theorem 1 hold, for all couples $(\alpha,\beta)\in\Omega_{a,b}$, one has: $B(\beta)\ge 0$, $S(\alpha) > 0$, matrix $N(\alpha,\beta)$ exists, and its eigenvalues are nonnegative. Since all vector couples $(\overline{\alpha}_i,\overline{\beta}_j)\in\overline{\Omega}_{a,b}\subseteq\Omega_{a,b}$, the defined in (12) scalar is an eigenvalue of $N(\alpha,\beta)$. It remains to show that $\overline{\lambda}_1 \ge \lambda_1[N(\alpha,\beta)]$ for all (α,β). From (12) one has:

$$0 \le \overline{\lambda}_1 S_i - B_j$$

or, equivalently, for all $i = 1,..,a$, and $j = 1,..,b$

$$\overline{\lambda}_1 X - A_i^T(\overline{\lambda}_1 X)A_i - B_j \ge 0$$

Denote

$$M = \overline{\lambda}_1 X, M_i = \sqrt{\overline{\lambda}_1} X^{1/2} A_i, i = 1,2,..,a$$

From statement (4) follows that the above set of inequalities is valid if and only if

$$\overline{\lambda}_1 X - A^T(\alpha)(\overline{\lambda}_1 X)A(\alpha) - B(\beta) \ge 0$$

In other words $\overline{\lambda}_1 S(\alpha) - B(\beta)$ is a positive (semi-) definite matrix for all $(\alpha,\beta)\in\Omega_{a,b}$, i.e.

$$\overline{\lambda}_1 \ge \overline{\lambda}_1(\alpha,\beta), (\alpha,\beta)\in\Omega_{a,b}$$

This proves the Lemma.

Corollary 1. The solution in (1) has the following bounds for all $(\alpha,\beta)\in\Omega_{a,b}$:

$$P_{U1}(\alpha,\beta) = P_{U1} = \overline{\lambda}_1 X \qquad (13)$$

$$P_{U2}(\alpha,\beta) = A^T(\alpha)P_{U1}(\alpha,\beta)A(\alpha) + B(\beta) \qquad (14)$$

where $\overline{\lambda}_1$ is given in (12).

Proof. The parameter independent bound (13) and the parameter dependent one (14) follow easily having in mind (9), (10) and (12).

The next result shows how the bound (14) can be used to improve (13) in sense of tightness.

Lemma 2. Consider (12) and suppose that for some integers $x = 1, 2,.., a$ and $y = 1, 2,.., b$, one has:

$$\overline{\lambda}_1 = \lambda_1(B_y S_x^{-1}) > \lambda_1(B_j S_i^{-1}), \forall i \ne x, \forall j \ne y$$

Denote $\Delta_{xy} = \overline{\lambda}_1 S_x - B_y$. The solution of the DALE has the following parameter independent upper matrix bound

$$P_{U2} = P_{U1} - \mu\Delta_{xy}, \forall(\alpha,\beta)\in\Omega_{a,b} \qquad (15)$$

$$\mu = \min\{1, \min\frac{1}{\lambda_1[\Delta_{xy}(\overline{\lambda}_1 S_i - B_j)^{-1}]}\} \qquad (16)$$

Proof. If the above supposition holds, then

$$\Delta_{xy} \ge 0, \quad \overline{\lambda}_1 S_i - B_j > 0, \ \forall i \ne x, \forall j \ne y$$

The scalar in (16) guarantees the inequalities

$$\overline{\lambda}_1 S_i - B_j = P_{U1} - A_i^T P_{U1} A_i - B_j \ge \mu\Delta_{xy},$$

for all $i = 1,..,a$ and $j = 1,..,b$. With

$$M = P_{U1} - \mu\Delta_{xy}, M_i = P_{U1}^{1/2} A_i, i = 1,2,..,a,$$

the above set of inequalities takes the form (4), i.e., it is equivalent to:

$$0 \le P_{U1} - A^T(\alpha)P_{U1}A(\alpha) - B(\beta) - \mu\Delta_{xy}$$

It follows that

$$A^T(\alpha)P_{U1}A(\alpha) + B(\beta) \le P_{U1} - \mu\Delta_{xy} = P_{U2}$$

Having in mind (14), the bound (15) is proved.

4 UPPER SCALAR BOUNDS

The obtained matrix bounds can be used to derive based on them upper estimates for the maximal eigenvalue and the trace of solution in (1), having in mind that if $X \le Y$, then $\lambda_1(X) \le \lambda_1(Y)$ and $tr(X) \le tr(Y)$ for arbitrary symmetric matrices *X* and *Y*.

Lemma 3. Consider the bounds P_{U1} in (13), $P_{U2}(\alpha,\beta)$ in (14) and P_{U2} in (15). The following scalar bounds for the DALE solution hold for all $(\alpha,\beta) \in \Omega_{a,b}$:

$$\lambda_1[P(\alpha,\beta)] \le l = \min\{l_1, l_2\} \le l_3 \qquad (17)$$

$$l_1 = \max \lambda_1(A_i^T P_{U1} A_i + B_j), i = 1,..,a,\ j = 1,..,b \quad (18)$$

$$l_2 = \lambda_1(P_{U2}),\ l_3 = \lambda_1(P_{U1}) \qquad (19)$$

$$tr[P(\alpha,\beta)] \le t = \min\{t_1, t_2\} \le t_3 \qquad (20)$$

$$t_1 = \max tr(A_i^T P_{U1} A_i + B_j) \qquad (21)$$

$$t_2 = tr(P_{U2}),\ t_3 = tr(P_{U1}) \qquad (22)$$

Proof. The bounds (19) and (22) are obvious. Consider the upper parameter dependent upper matrix bound in (14). It follows that

$$\lambda_1[P(\alpha,\beta)] \le \lambda_1[P_{U2}(\alpha,\beta)A(\alpha)],$$

for all vector pairs $(\alpha,\beta) \in \Omega_{a,b}$. From (18) one gets:

$$l_1 I - A_i^T P_{U1} A_i - B_j \ge 0,\ i = 1,..,a,\ j = 1,..,b$$

When statement (4) is applied for

$$M = l_1 I,\ M_i = P_{U1}^{1/2} A_i,\ i = 1,2,..,a\ ,$$

one gets:

$$l_1 I \ge A^T(\alpha)P_{U1}(\alpha,\beta)A(\alpha) + B(\beta)],$$

for all $(\alpha,\beta) \in \Omega_{a,b}$, which proves the bound in (18). In fact, l_1 is the exact maximal eigenvalue of the parameter dependent matrix $P_{U2}(\alpha,\beta)$ in (14), in accordance with Definition 2.

Consider the trace bound in (21) and the based on it inequalities:

$$tr(A_i^T P_{U1} A_i) + tr(B_s) = a_i + b_s \le t,\ \forall i, s$$

Denote $a_{is} = t - a_i - b_s$. For any given $\alpha \in \Omega_a$ and *s* = 1,2,..,*b*, one gets:

$$0 \le t(\alpha) = \sum_{i=1}^{a} \alpha_i^2 a_{is} + \sum_{i,j=1,\, i<j}^{a} \alpha_i \alpha_j (a_{is} + a_{js})$$

Since

$$t(\alpha) = t_1 - \sum_{i=1}^{a} \alpha_i^2 a_i - \sum_{i,j=1,\, i<j}^{a} \alpha_i \alpha_j (a_i + a_j) - b_s$$

$$a_i + a_j = tr(A_i^T P_{U1} A_i + A_j^T P_{U1} A_j) \ge a_{ij}\ .$$

where

$$a_{ij} = tr(A_i^T P_{U1} A^j + A_j^T P_{U1} A^i)$$

it follows that

$$t(\alpha) \le \tilde{t}(\alpha) = t_1 - \sum_{i=1}^{a} \alpha_i^2 a_i + \sum_{i,j=1,\, i<j}^{a} \alpha_i \alpha_j a_{ij} - b_s$$

Having in mind that

$$\tilde{t}(\alpha) = t_1 - tr[A^T(\alpha)P_{U1}A(\alpha)] - tr(B_s) \ge 0$$

one easily gets

$$\tilde{t}(\alpha) = t_1 - tr[A^T(\alpha)P_{U1}A(\alpha)] - \sum_{s=1}^{b} \beta_s tr(B_s)$$

Therefore,

$$t_1 - tr[A^T(\alpha)P_{U1}A(\alpha) - B(\beta)] \ge 0$$

for all $(\alpha,\beta) \in \Omega_{a,b}$, which proves the trace bound in (21).

From (12) and (13) one gets:

$$0 \le \overline{\lambda}_1 S_i - B_j = P_{U1} - A_i^T P_{U1} A_i - B_j$$

for all *i* = 1,.., *a*, and *j* = 1,.., *b*. Therefore,

$$A_i^T P_{U1} A_i + B_s \le P_{U1}$$

and the inequalities $l_1 \le l_3,\ t_1 \le t_3$, follow.

Therefore, bounds *l* and *t* are always tighter than l_3 and t_3, respectively.

ACKNOWLEDGEMENTS

This research is supported by the FP7 grant AComIn No316087, funded by the European Commission in Capacity Programme in 2012-2016

REFERENCES

Kouikoglu, V., Phillis, Y., 1993. Trace bounds on the covariances of continuous-time systems with multiplicative noise, *IEEE Trans. on Autom. Control*. 38, 138-142.

Lee, C., Lee, T., 2001. On the estimation of solution bounds of the generalized Lyapunov equation and the robust root clustering for the linear perturbed systems, *Intern. Journ. Contr*., v.74, 996-1008.

Lee, C., Li, T., Kung, F., 1993. A new approach for the robust stability of perturbed systems with a class of non-commensurate time delays, IEEE Trans. on Circuits and Systems, v. 40, 605-608.

Savov, S., 2014. Solution bounds for algebraic equations in control theory, Prof. Marin Drinov Academic Publishing House, Sofia.

Wang, S., Lin, T., 1987. Robust stability of uncertain time delay systems, Intern. Journ. Contr., v.46, 963-976.

Yedavalli, R., 1993. Robust root clustering for linear uncertain systems using generalized Lyapunov theory, Automatica, v. 29, 237-240.

A METHOD FOR FINDING OPTIMAL NETWORK CYCLES UNDER CONSTRAINTS AT A TOUR OF POINTS UNDER SURVEILLANCE

Vasil Sgurev, Stanislav Drangajov
Institute of Information and Communication Technologies – BAS
G.Bonchev, str, 113 Sofia, Bilgaria
e-mail: *vsgurev@gmail.com; sdrangajov@gmail.com*

Keywords: Network flow optimization; Optimal constrained flow cycles; Iterative method of network flow optimization.

Abstract: In this paper we consider an abstract network whose certain nodes must be visited and inspected at a given by the decision maker time intervals. The nodes under surveillance are previously known as well as the time prescribed between two successive inspections. Our purpose is to determine the optimal tours around the nodes, minimizing the cost of the tour, the number of agents (transportation units) performing the tour(s) keeping the time requirements. This task may be performed by the approach of the network flow programming with some new methods for determining the most cost effective cycles and the number of agents. Two exhaustive numerical examples are given which illustrate the method in details.

1 PRELIMINARY COMMENTS

The problem of cyclic tour of objects (points) on an arbitrary network by mobile transport units, including mobile robots, for executing operations there – goods transportation, points control, information gathering, etc. is a present day one of the daily practice in many business activities. When cyclic tour is a must the necessity of circulation of the transport units (agents), the cost minimization and thereafter –the route optimization is evident. This includes minimization of agents' number (transportation units), of time (distance), passing cost etc.

In many cases along with the pointed out requirements additional constraints are imposed according to which the time of one cycle must not exceed the minimum time necessary between two successive visits of each object under surveillance. This complicates considerably the optimization problem as it cannot be carried out with the classic optimization procedures for the circulation on a network.

A procedure, similar at first glance is described in (Sgurev, V, S. Drangajov, Optimal Circulation of Mobile Robots) but without the additional constraints for limitation of the time between two successive inspections. In the present work we try to do this, viz. – to propose an iterative procedure for solving of the class of problems described above together with the additional constraints.

2 FORMALIZATION

An oriented graph *G(X,U)* is introduced (Sgurev, V Network Flows with General Constraints,1991,) with a set of nodes (vertices, points) $X=\{x_1, x_2, ..,x_n\}$ and a set of edges (arcs) $U=\{u_1, u_2, ..,u_m\}$.

The following denotations will be used:

$u_k = (x_i, x_j) \in U$ — a single arc;

$I(U) = \{(i, j)\} / (x_i, x_j) \in U\}$ — the set of indices' pairs of existing arcs. At that

$X_0 \subset X$ — is the set of all points subject to compulsory monitoring;

$X' = X \setminus X_0$ — a set of intermediate points;

Γ_x^{-1} — direct mapping (all nodes connected by outgoing arcs) of node $x \in X$;

Γ_x^{-1} — reverse mapping (nodes connected by incoming arcs) of node $x \in X$;

$g_{ij} = g(x_i, x_j)$ — time (value) for passing of a transportation unit (agent) along the arc (x_i, x_j);

$R_i = R(x_i) \in X_0$ — the maximum admissible time interval between two inspections of the node (object) $x_i \in X_0$;

$f_{ij} = f(x_i, x_j); (i,j) \in I(U)$ — arc flow function (Sgurev, V, Network Flows with General Constraints, 1991)

At the assumptions made the motion of the transportation unit (agent) may be described in general as network flow circulation (Sgurev, V, Network Flows with General Constraints, 1991) of a discrete network flow. The tour of the points of X_0 may be carried out by one or several cycles. Near to this problem is the Chinese postman problem but with specific additional constraints.

The flow conservation equations are as follows for each node from X:

$$\sum_{j \in \Gamma_i^1} f_{ij} - \sum_{j \in \Gamma_i^{-1}} f_{ji} = 0; x_i \in X.$$

When defining the agents' circulation all nodes from X_0 must enter in at least one cycle of some transportation unit. For this purpose for each $x_i \in X_0$ we put

$$\sum_{j \in \Gamma_i^1} f_{ij} = 1.$$

The minimization of the functional

$$\sum_{(i,j) \in I(U)} g_{ij} f_{ij} \to \min$$

guarantees total minimal run of the transportation units being used.

The optimal control of the circulating transportation units (agents) will be determined through the following integer programming problem:

$$L = \sum_{(i,j) \in I(U)} g_{ij} f_{ij} \to \min; \quad (1)$$

under constraints: for each $x \in X$

$$\sum_{j \in \Gamma_i^1} f_{ij} - \sum_{j \in \Gamma_i^{-1}} f_{ji} = 0; \quad (2)$$

$$\sum_{j \in \Gamma_i^1} f_{ij} = 1 \text{ for each } x \in X_0; \quad (3)$$

$f_{ij} \geq 0$ for each $(i,j) \in I(U)$; (4)

for integer values of $\{f_{ij}\}$.

Solving the problem (1) to (4) allows defining the minimal number of vehicles (agents) necessary for a tour around all objects under surveillance as well as the minimal total run of these vehicles. Obtaining a single cycle means that it could be carried out by a single transportation unit, but this is just a dreamed particular case, which may not be achieved.

Let by solving the problem (1) – (4) k in number cycles be received for which

$C = \{C_1, C_2, \ldots, C_k\} = \bigcup_{r \in I_k} C_r$ is a set of cycles of indices from $I_k = \{r/C_r\}$;

X_r is the set of nodes in the cycle C_r;

U_r — the set of arcs in the cycle C_r;

$I(U_r) = \{(i,j)/(x_i, x_j) \in U_r\}$ — a set of all pairs of indices of arcs from the cycle of index r;

$G_r = \sum_{(i,j) \in I(U_r)} g_{i,j}$ — total time needed for a tour along all arcs of the cycle C_r.

It follows from equation (3) of the optimization problem, that for each cycle of index $r \in I_k$ it is true, that $|X_r \cap X_0| \geq 1$, and the condition for two successive inspections requires: for each $r \in I_k$ and $x_i \in |X_r \cap X_0| = X_{r,0} \neq \varnothing$

$R_i \geq G_r$; (5)

where $\varnothing$ stands for the empty set.

For each cycle $r \in I_k$ the set $X_{r,0}$ may be split into two subsets $X'_{r,0}$ and $X''_{r,0}$ depending on the observation of the requirements (5) and namely:

$X'_{r,0} = \{x_i / x_i \in X_{r,0}; R_i \geq G_r\}$; (6)

$X''_{r,0} = \{x_i / x_i \in X_{r,0}; R_i < G_r\}$; (7)

$X'_{r,0} \cup X''_{r,0} = X_{r,0}; X'_{r,0} \cap X''_{r,0} = \varnothing$.

For the nodes under surveillance from $X'_{r,0}$ of the optimal cycle $r \in I_k$ relations (5) are observed and therefor they should be serviced by a single transportation unit. The other subset $X''_{r,0}$ does not observe (5) and for it another cycle is being sought serviced by another transportation unit. Two new procedures – A and B, are proposed for solving this problem, resulting in optimal cycles, keeping the requirements (5). The following changes are made in the relations (1) to (4):

A. Instead of equations (3) the relations below are used:

$$\sum_{j \in \Gamma_i^1} f_{ij} = 1, \text{ for each } x_i \in X''_{r,0}; \quad (8)$$

$$\sum_{(i,j) \in I(U)} g_{ij} f_{ij} \leq R''_{r,o}; \quad (9)$$

where $R''_{r,0} = \min_{x_i \in X''_{r,0}} R(x_i)$ is the minimum admissible time between two successive inspections of the nodes from $X''_{r,0}$.

Solving the problem (1), (2), (4), (8) and (9) results in defining of one or more new optimal cycles in which for the nodes of $X''_{r,0}$ the requirements of (5) will be observed..

B. Analogically the new optimal cycle for the nodes of $X'_{r,0}$ may be defined whose duration may turn out to be less than the cycle *r*. For this purpose instead of (3) the following relations are used:

$$\sum_{j \in \Gamma_i^1} f_{ij} = 1 \text{ , за всяко } x_i \in X'_{r,0}; \qquad (10)$$

$$\sum_{(i,j) \in I(U)} g_{ij} f_{ij} \leq R'_{r,o} \; ; \qquad (11)$$

where $R'_{r,0} = \min_{x_i \in X'_{r,0}} R(x_i)$ is the minimum admissible time between two successive inspections of the nodes from $X'_{r,0}$.

Through problem (1), (2), (4), (10) and (11) a new optimal cycle is defined with a less total time than the previous and observing the requirements (5).

Two possibilities exist after the improvement of the cycle $r \in I_k$:

a) If it turns out that for the new cycles nodes under surveillance from $X_{r,0}$ appear for which $X''_{r,0} \neq \varnothing$, then procedure A is executed again until reaching $X''_{r,0} = \varnothing$.

b) If $X'_{r,0} = X_{r,0}$, then for each of the cases under surveillance the requirements (5) are observed and the new cycles obtained are locally optimal. A transfer to the next cycle of index *r*+1 is made with a new execution of procedures A and B.

If after repeated application of procedure A nodes from $X_{r,0}$ remained for which the requirement (5) is not observed, then no solution exists for these nodes. The procedures described result in locally optimal solutions for the necessary number of transportation units (cycles) and the tour planning.

Two numerical examples are given below which demonstrate the abilities and the efficiency of the methods proposed.

3 NUMERICAL EXAMPLES

EXAMPLE # 1: An oriented graph of 11 nodes ($|X| = 11$) and 15 arcs ($|U| = 15$) is shown in Fig. 1. Objects under surveillance are:

$X_0 = \{x_3, x_4, x_9, x_{10}\}$; (12)

and they are marked by thick lines and the remaining by dotted lines. The maximum admissible time between two successive visits is shown beside each node under surveillance of Fig. 1.

$R_2 = R(x_2) = 15; R_4 = 8; R_9 = 9; R_{10} = 22.$

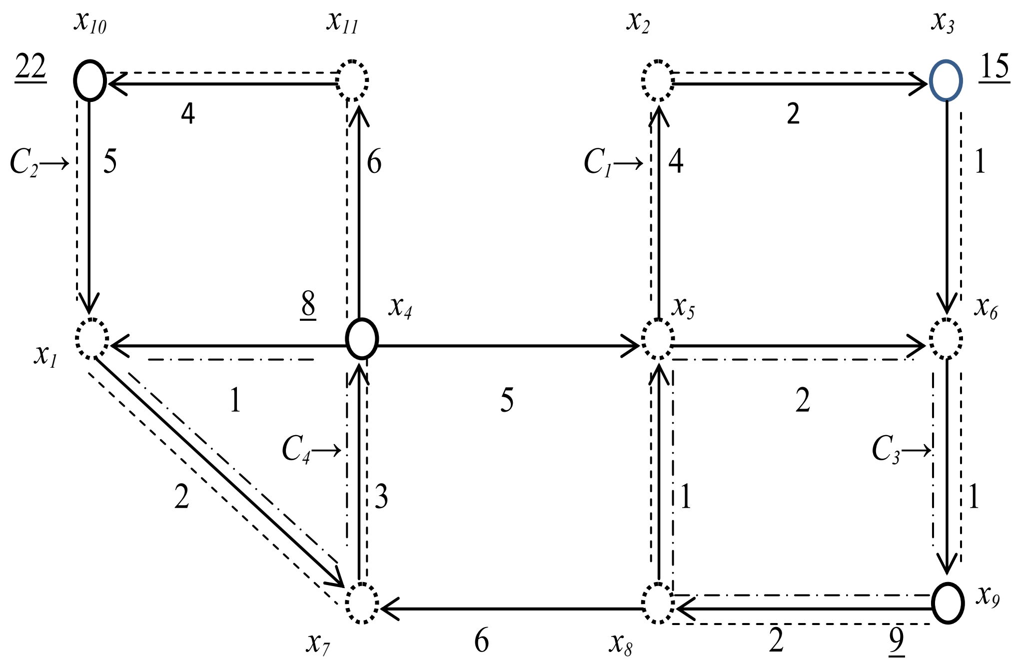


Fig. 1

In the same Fig.1 the nodes' indices $\{x_i\}$ are marked, and next to each arc is the given the numerical value of the arc estimation $\{g_{ij}\}$. These data are summarized in Table #1.

Table #1

#	#Arc index $\{x_{ij}\}$	$x_{1,7}$	$x_{2,3}$	$x_{3,6}$	$x_{4,1}$	$x_{4,5}$	$x_{5,2}$	$x_{5,6}$	$x_{6,9}$	$x_{7,4}$	$x_{8,5}$	$x_{8,7}$	$x_{9,8}$	$x_{4,11}$	$x_{11,10}$	$x_{10,1}$
1	Arc estimation $\{g_{ij}\}$	2	2	1	1	5	4	2	1	3	1	6	2	6	4	5
2	Arc flows $\{f_{ij}\}$ for cycles C_1 and C_2	1	1	1	0	0	1	0	1	1	1	0	1	1	1	1
3	Arc flows $\{f_{ij}\}$ for cycle C_3	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0
4	Arc flows $\{f_{ij}\}$ for cycle C_4	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0
5	Arc flows $\{f_{ij}\}$ for cycle C_5	1	1	1	0	1	1	0	1	2	0	1	1	1	1	1

The objective function L and equations (1) to (4) are as follows:

$L = 2f_{1,7}+ 2f_{2,3}+ f_{3,6}+ f_{4,1}+ 5f_{4,5}+ 4f_{5,2}+ 2f_{5,6}+ f_{6,9}+3$

$f_{7,4}+ f_{8,5}+ 6f_{8,7}+ 2f_{9,8}+ 6f_{4,11}+ 7f_{11,10}+ 5f_{10,1} \rightarrow min;$ (13)

observing constraints:

1. $f_{1,7} - f_{4,1} - f_{10,1} = 0;$
2. $f_{2,3} - f_{5,2} = 0;$
3. $f_{3,6} - f_{2,3} = 0;$
4. $f_{4,1} + f_{4,5} + f_{4,11} - f_{7,4} = 0;$
5. $f_{5,2} + f_{5,6} - f_{4,5} - f_{8,5} = 0;$
6. $f_{5,6} - f_{4,5} - f_{8,5} = 0;$
7. $f_{7,4} - f_{1,7} - f_{8,7} = 0;$
8. $f_{8,5} + f_{8,7} - f_{9,8} = 0;$
9. $f_{9,8} - f_{6,9} = 0;$
10. $f_{10,1} - f_{11,10} = 0;$
11. $f_{11,10} - f_{4,11} = 0;$
12. $f_{3,6} = 0;$
13. $f_{4,5} + f_{4,1} + f_{4,11} = 1;$
14. $f_{9,8} = 1;$
15. $f_{9,8} = 1;$
16. $f_{i,j} \geq 0$ for each (i,j);

at integer values of the variables $\{f_{i,j}\}$.

For solving the problem (1) to (4) the specialized software Weboptim (Genova K., et. al., 2011) was used, designed and developed in IICT-BAS. The optimal solution for $\{f_{i,j}\}$ without observing the requirement (5) is shown in row 2 of Table 1. The maintenance of the four points under surveillance $\{x_3, x_4, x_9, x_{10}\}$ is carried out by two transportation units circulating in two cycles with parameters $C = \{C_1, C_2\}$; $I_k = \{1,2\}$.

1. First cycle - C_1, for which

$X_1 = \{x_2, x_3, x_6, x_9, x_8, x_5, x_2,\}$;

$U_1 = \{x_{2,3}, x_{3,6}, x_{6,9}, x_{9,8}, x_{8,5}, x_{5,2}\}$;

$I(U_1)$ ={(2,3), (3,6), (6,9), (9,8), (8,5), (5,2)};

$G_1 = 11$; $|X_1 \cap X_0| = 2$;

$X_{1,0} = \{x_3, x_9\}$, at which for the node under surveillance x_3 condition (5) is observed. $R_3 = 15 \geq G_1 = 11$. Therefore $X'_{1,0} = \{x_3\}$ and the node should be maintained by the transportation unit circulating along the first cycle C_1.

The requirement (5) is not observed for node x_9, as $R_9 = 9 < G_1 = 11$ and as so $X''_{1,0} = \{x_9\}$. Another appropriate cycle is sought for this node.

2. Second cycle - C_2, for which

$X_2 = \{x_1, x_7, x_4, x_{11}, x_{10}, x_1\}$;

$U_2 = \{x_{1,2}, x_{7,4}, x_{4,11}, x_{11,10}, x_{10,1}\}$;

$I(U_2)$ ={(1,2), (7,4), (4,11), (11,10), (10,1)};

$G_2 = 20$; $|X_2 \cap X_0| = 2$;

$X_{2,0} = \{x_{10}, x_4\}$, at which for the node under surveillance x_{10} condition (5) is observed, as $R_{10} = 22 \geq G_2 = 20$. Therefore $X'_{2,0} = \{x_{10}\}$ and node x_{10} should be maintained by the transportation unit circulating on the second cycle C_2.

The requirement (5) is not observed for node x_4, as $R_4 = 8 < G_2 = 20$ and as so $X''_{2,0} = \{x_4\}$. Another appropriate cycle should be sought and a new transportation unit for this node. The following procedures A result in defining of new cycles for $\{x_9, x_4\}$ for which conditions (5) are observed.

3. Third cycle - C_3, for which $X'_{1,0} = \{x_9\}$ and instead of (3), (8) and (9) are used, viz.

$f_{9,8} = 1;$ (14)

$$\sum_{(i,j)\in I(U)} g_{ij} f_{ij} \leq 9; \quad (15)$$

The optimal solution of the problem (1), (2), (14), (15), (4) is shown in row 3 of Table 1.

$X_3 = \{x_5, x_6, x_9, x_8, x_5\}$;

$U_3 = \{x_{5,6}, x_{6,9}, x_{9,8}, x_{8,5}\}$;

$I(U_3) = \{(5,6), (6,9), (9,8), (8,5)\}$.

The node under surveillance x_9 will be maintained by a new transportation unit circulating along C_3.

4. Fourth cycle - C_4 is related to finding of optimal solution, observing requirements (5) for $X''_{2,0} = \{x_4\}$. In the problem (1) to (4) instead of (3) the relations

$$f_{4,5} + f_{4,1}\ f_{4,11} = 1; \quad (16)$$

$$\sum_{(i,j)\in I(U)} g_{ij} f_{ij} \le 8; \quad (17)$$

are used.

The case is solved by the same software product and the solution is shown in row 4 of Table 1. It shows the following parameters of the cycle C_4:

$X_4 = \{x_4, x_1, x_7, x_4\}$;

$U_4 = \{x_{4,1}, x_{1,7}, x_{7,4}\}$;

$I(U_4) = \{(4,1), (1,7), (7,4)\}$.

The node under surveillance x_4 will be maintained by a new transportation unit circulating along C_4. At that for each of them the time of two successive visits is less than the admissible i.e. keeping the requirements (5).

In Example #1 considered locally optimal solutions are obtained by the method and the procedures proposed for periodical maintenance of the four points (objects) under surveillance $\{x_3, x_4, x_9, x_{10}\}$ by the respective transportation units moving along cycles $\{C_1, C_2, C_3, C_4\}$.

EXAMPLE # 2: Let under the same conditions of Fig. 1 and Table 1 a problem be set for cyclic tour of the four points under surveillance by a single cycle and a single transportation unit. For this purpose the following changes should be made in the problem (1) to (4):

a) instead of (3) we put

$f_{3,6}=1; f_{8,7}=1; f_{10,1}=1$;

$f_{4,5} + f_{4,1}\ f_{4,11} \ge 1$.

b) instead of (4) the following is required:

$f_{i,j} \ge 0$ for each $(i,j) \in I(U)$.

Under these conditions the optimal solution for $\{f_{i,j}\}$ is shown in row 5 of Table 1.

5. Fifth cycle - C_5 for which

$X_5 = \{x_5, x_2, x_3, x_6, x_9, x_8, x_7, x_4, x_{11}, x_{11}, x_{10}, x_1, x_7, x_4, x_5,\}$;

$U_5 = \{x_{5,2}, x_{2,3}, x_{3,6}, x_{6,9}, x_{9,8}, x_{8,7}, x_{7,4}, x_{4,11}, x_{11,10}, x_{10,1}, x_{1,7}, x_{7,4}, x_{4,5}\}$;

$I(U_5) = \{(5,2), (2,3), (3,6)\ (6,9), (9,8), (8,7), (7,4), (4,11), (11,10), (10,1), (1,7), (7,4), (4,5\}$.

$G_5 = 44$; $|X_5 \cap X_0| = 4$;

$X_{5,0} = \{x_3, x_4, x_9, x_{10}\}$, at which for all four nodes under surveillance requirement (5) is not observed as $R_3 = 15 < G_5 = 44$; $R_4 = 8 < G_5 = 44$; $R_9 = 9 < G_5 = 44$; $R_{10} = 22 < G_5 = 44$. So, one cycle with a single transportation unit is not enough to satisfy the requirement for maximum admissible time intervals for visiting the nodes under surveillance.

The cycle is shown in the next Fig. 2.

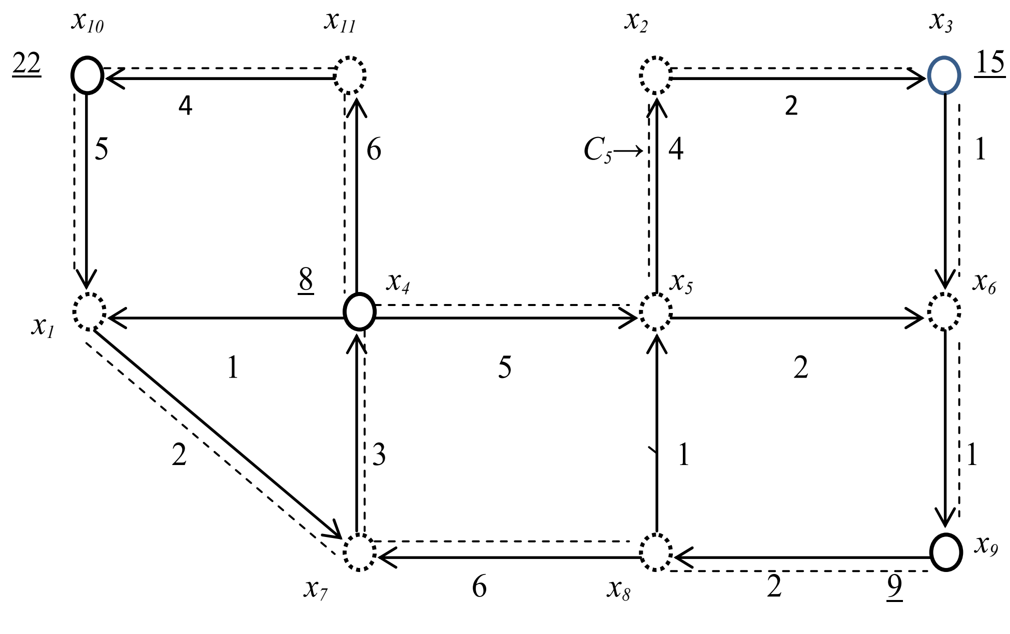


Fig. 2

4 SUMMING UP

Comparing examples #1 and #2 the following conclusions may be drawn up:

1. When using a single optimal cycle C_5, one optimal unit respectively neither of the requirements (5) for the points under surveillance are satisfied. This is due to the significant time – 44 units for the tour of these nodes.

2. With two optimal cycles - C_1 and C_2 two agents are necessary. For a part of the nodes $\{x_3, x_{10}\}$ the requirements (5) are observed and for the two other nodes - $\{x_4, x_9\}$ this is not possible. The tour duration for the two cycles are $G_1 = 11$; $G_2 = 20$ units respectively.

3. With four locally optimal cycles - C_1, C_2, C_3, C_4 all technological requirements (5) concerning the tour duration are observed. The durations are 11, 20, 6, and 6 units respectively.

The two numerical examples described demonstrate the abilities and the efficiency of the iterative method for discrete optimization of network flow cycles with additional constraints proposed in the present work.

ACKNOWLEDGEMENTS

The research work presented in this paper is partially supported by the FP7 grant AComIn No 316087, funded by the European Commission in Capacity Programme in 2012-2016.

REFERENCES

Sgurev, V, S. Drangajov, Optimal Circulation of Mobile Robots, In: Proceedings of XXI Int. Symposium on “Control of Heat Power Objects and Systems”, Sofia, Nov. 7-8 2013, pp. 77-80, UAI, ISSN 1113-2237.

Pennsylvania bracing for possible fight to protect military bases, State Sen. Richard Alloway and U.S. Rep. Bill Shuster prepare to defend Letterkenny Army Depot, http://articles.herald-mail.com/2012-02-08/news/31040178_1_military-bases-brac-process-closure-process

Tooele Army Depot, http://en.wikipedia.org/wiki/Tooele_Army_Depot

Risk Management in Warehouses, Subba Rao, http://www.cholarisk.com/uploads/risk%20management%20in%20warehouse.pdf

Don Philips and Alberto Garcia-Diaz, Fundamental of Network Analysis, Prentice-Hall, Englewood Cliffs, N.J., 1981

Sgurev, V, Network Flows with General Constraints, Publishing House of the Bulgarian Academy of Sciences, Sofia, 1991

Genova K., L. Kirilov, V.Guliashki, B. Staykov, D. Vatov (2011) "A Prototype of a Web-based Decision Support,System for Building Models and Solving Optimization and Decision Making Problems", In: Proc. of XII Int. Conf. on Computer Systems and Technologies - (Eds. B. Rachev, A. Smrikarov), Wien, Austria, 16-17 June, 2011, vol. 578, 167-172.

INTEGRATED INFORMATION SYSTEM FOR ENTERPRISE MANAGEMENT

Tatiana Atanasova
Institute of Information and Communication Technologies - BAS, Acad. G. Bonchev str., Sofia, Bulgaria
atanasova@iit.bas.bg

Janko Atanasov
Votan Lt., Sofia, Bulgaria
janko.a@abv.bg

Keywords: Information system, Integration, Barcodes.

Abstract: In the paper an information system is described that is implemented as an integrated software product tailored to the business and organizational structure of particular company. The system allows the management staff of the company to effectively control and optimize the manufacturing and realization of its products. The purpose of the system is to achieve integration of various processes related to the main activity of the company - production, storage and marketing in its own retail network - into a single information system.

1 INTRODUCTION

Modern business environment is very changeable. Businesses face the challenges of growing competition, increased consumer expectations and expanding markets. This requires looking at ways to reduce costs throughout the supply chain, to accelerate the production process, to reduce the warehouse, to expand the product range, ensuring timely deliveries to customers and better overall service with effective coordination of global demand, supply and production.

Effective use of resources together with integration (Daskalova, Kolchakov, 2000) of different departments during planning and implementation of production goals attracts more and more strategic attention in business management. Every year organizations continue to increase spending on information technology (IT) and their budgets continue to rise, even in the face of potential economic downturns (Petter et al, 2008).

The IT tools for modeling (Tashev, Hristov, 2001), data management and deployment of information applications evolve rapidly. For example, an integration of system dynamics and multi-objective optimisation is introduces in (Aslam et al, 2014). The research in (Lu et al, 2014) examines the possible adoption of distributed intelligence approaches in warehouse management systems. In order to benefit more from the system integration and information technology implementation, the scope of enterprise integration and interoperation is extending from intra-enterprise to inter-enterprise (Li et al, 2010). Thus the software tools for modelling, problem solving, data management and deployment of optimization applications have coevolved with generations of software systems (Valente, Mitra, 2007).

The research of the impact of information technologies to organization performance leads to recognition that it is a need of connecting the production and information flows (Ilie-Zudor et al, 2011). The effective analysis of information flows in an enterprise should include some properties of the enterprise as well (Petrauskas, 2006).

There are some realizations of software component architecture for supply chain management across dynamic organizational networks that consist of existing systems supplemented with new software components (Verwijmeren, 2004). The progress of market economy and information technology puts forward design ideas based on the traceability requirements of common model of the supply chain, and the whole process of dynamic management (Xu and Zhao, 2014).

To remain competitive, business organizations have to change their own practice and procedures. Many of them consider implementation of Enterprise resource planning (ERP) systems. But the implementation of ERP systems is difficult and places tremendous demands on corporate time and resources. Adaptation of the specific business model and integrated processes that are strictly individual for the company to packaged software is very complex and sometimes even impossible. The theory and practice of ERP do not provide a uniform approach for determining the sequence of implementation of the various possible components (particular modules) of the system. Many ERP implementations have been classified as failures because they did not achieve predetermined corporate goals (Umble et al, 2003). It is still a need of development of new tools for business that provides rapid and universal integration of various software applications based on business processes and prospects for the enterprise growth and progress. The above consideration allows us to define several requirements for these tools as:

- connection of the production and information flows;
- use modern tools for modeling and data management;
- be a basis for generations of software systems for IT applications;
- creation conditions for sustainable competitive advantages in a dynamic environment.

The proposed in this paper system tries to implement these requirements. It aims to achieve integration of various processes related to the main manufacture, storage, sales in shops of the company in a unified information structure. Thus the overall production process and the management of the company can increase their effectiveness. The purpose of the proposed integrated system is to design a framework tailored to operations and organizational structure of one company.

2 INTEGRATION IN TAILORED FRAMEWORK

Associating information, processing power and communication capabilities with products themselves and their environment may be an effective way to help people and corporate management systems.

The aim is to create sustainable competitive advantages in dynamic environments according to appropriate strategic moves.

The company realized that implementation of ERP systems is difficult because their underlying business processes do not coincide with those of the enterprise. The company has specific strategic goals, namely, entering and expansion on foreign markets with different trade policies. It has to be taken into account the geographical aspects of the organization of the company. Its employees, departments, assets, retail network, foreign contractors and customer base are distributed. Dynamically changing market environment and regulations is another complex factor. It has been considered that purchasing and adaption of ERP system require high cost and time resources to reflect specific characteristics of activity of the enterprise, the characteristics of its production, its organizational structure and its strategic and long-term development goals.

The integrated information system associates various processes incorporated with the main production, storage, and store sales of the company in a single data structure. The information system is fully compatible with

- Specifics of the company;
- Its characteristics;
- Type of production;
- Organizational structure;
- Strategic development goals.

The integrated software system is built on a modular basis, comprising all of the typical company processes and activities in order to improve and support their effective management. Thus the modules of the integrated system are associated with:

1. planning and production management;
2. control of basic materials for manufacturing;
3. control and monitoring of the production process;
4. effective management of human resources;
5. control and effective management in the realization of finished products;
6. efficient processing of data and feedback from customers.

2.1.1 Planning

The “planning” is a very essential module for the integrated system. It includes the following information:

- Description of the product according to the stages of its production;
- Overall planning of interrelated contracts for production;
- Assessment of physical security and utilization of various types of production resources;
- Consideration of quantitative and qualitative results of the production.

The integrated software is based on a custom system, i.e. the production department receives orders for production. The orders are individual or serial.

Mandatory elements of each order are:

- Type of the product;
- Season and year;
- Basic materials;
- Model quantities in size;
- Sales prices in the respective countries.

In production planning it is necessary to consider all materials that are required for the respective produced article, its type and the ordered quantity (by size). For this purpose, for each article a unique (nonrecurring) barcode that includes the order number, year and season, model, basic material and size is formed. Through this barcode the production stage of each product can be monitored. It also facilitates the storing (warehouse data) and its subsequent realization in shops in the country or abroad. Serial orders require the formation of so called “process maps” (Technology Map) that are used as guides during the production process. The Technology Map contains also the specifications of the auxiliary materials used in the production of the respective product. For individual orders it is necessary to keep records of the individuals that have ordered the particular article.

2.1.2 Control of basic materials

The module “control of basic materials for manufacturing” contains detailed information about material procurement with the aim to optimize the stock of basic and auxiliary materials, ensuring the availability of supplies in production planning and taking serial and individual orders.

The basic supplies for the production are different materials. Accountability of the materials includes information about the suppliers and their references. For the statistics purposes of marketed production it is necessary to account a number of specific characteristics of the material – its composition, type, origin, colour and other characteristics fully describing the materials. These are elements that are automatically printed on the label of the finished product.

Upon delivery of material, it receives a serial number according to the provider and it is assigned a barcode taking into account the season and year of delivery. Basic material’ catalogues are distributed with the associated barcodes and thus facilitate the customer’s order system.

2.1.3 Production process

The module “control and monitoring of the production process” includes a wide range of information and combines and systematizes the various indicators of overall production process - from the time of receipt of an order to completion of the product. These indicators are related to specific technology production period, quality indicators, etc. The main item in this module is the customer’s order system that forms a unique barcode for each item. Besides the elements of barcode described in the above modules, a serial number of each order is added here. This “extended” barcode is affixed on the individual order or on the technology map for the serial orders.

The automated system of autonomous barcode readers is built in order to gather information from different control points in the enterprise, thus providing means to control and monitor the production process. Control points are determined by the specialists from the technical control department (quality inspectors). When a product is identified by its barcode on the control point, the readers fix it in an event that reflects the successful passing of the product. This method of tracking the products during the production process provides real time information about the certain stage of the process and the overall status. The control points can be equipped with a display that shows the last item and the total number of items that have already passed through it. The item is accounted in the system as “ready”, once it passes through the last control point and is recognized by the system as “ready production”. This method of data and indicators collection provides comprehensive information about events, the stage of the production process of the serial orders and the possible disruption in the production process, etc.

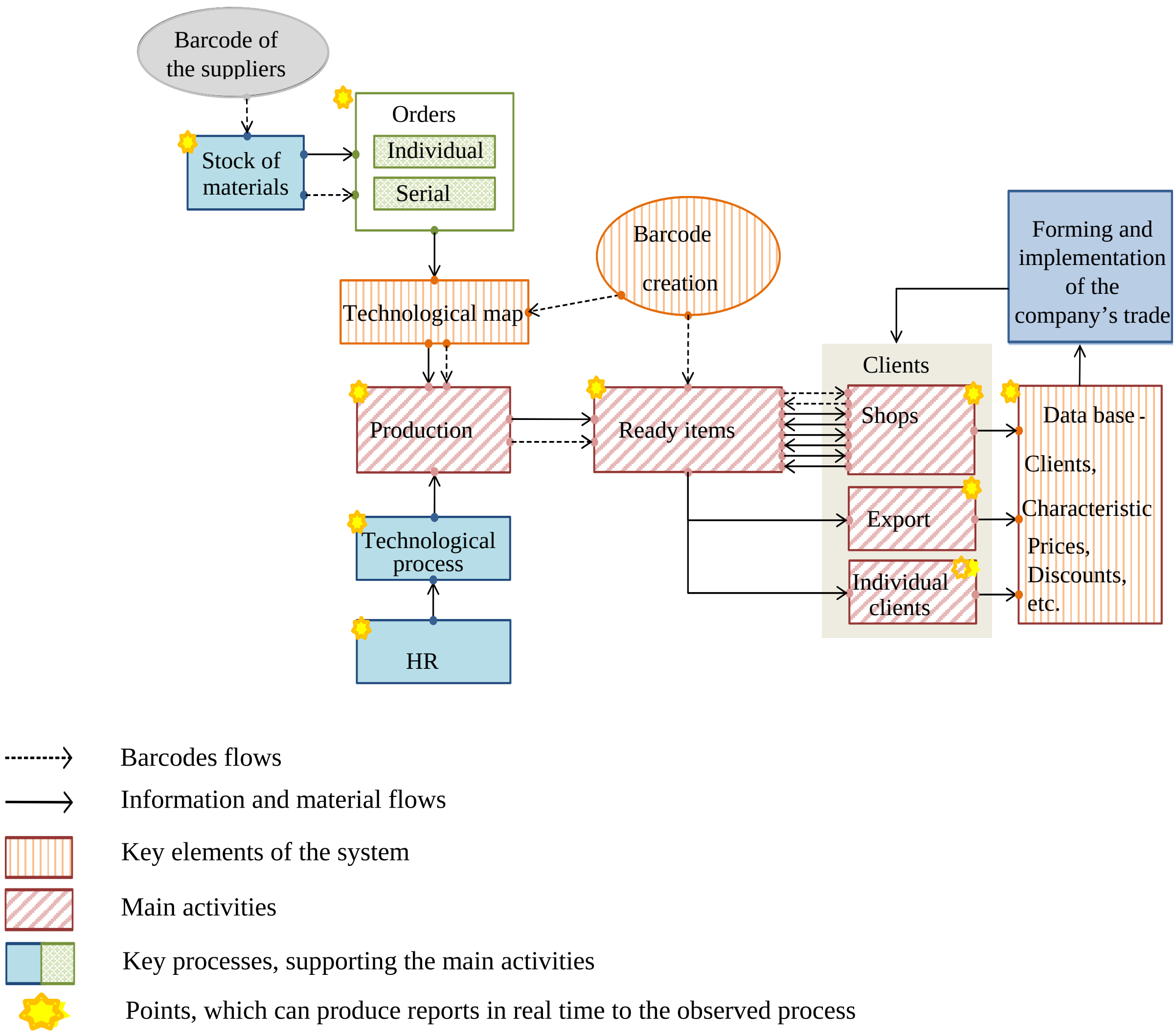


Figure 1.

2.1.4 Management of human resources

The module "management of human resources (HR)" includes information from the various existing and integrated enterprise systems that measure certain indicators regarding staffing, and technological operations performed by each single worker or team. This allows different indicators to be aggregated, classified and associated to the manufacturing process in order to give more complete information about the company's activity. Based on the company's long lasting traditions and experience in its production it has defined a complete set of basic technological operations for the preparation a certain type of products. There is a complete listing of various operations, containing the required time for their implementation that was experimentally defined. The human resource management system takes into consideration the experience of each employee gained so far, the activities each one performs, the employee's daily operations; and prepares daily, weekly and monthly reports on labour participation rate. It also provides information how many hours/days a worker was not working due to ineffective organization, workload planning, etc. The data from these subsystems is used in the other modules in the integrated system for determining the remuneration of workers. These data can also be used for the assessment of the skills of each worker and assigning the most appropriate process in which each one can participate and be most effective.

Thus the software collects, correlates and complements indicators related to the company's human resources with data according to the overall activity of the company and allows the analytical treatment of the resulting information.

2.1.5 Control and management of finished products

The module "control and effective management in the realization of finished products" includes information on finished products, their location (warehouse, retail network, in transit in the country, abroad etc.), specifications, stocks, etc.

The information system "Retail network (Stores)" is presented virtually as a whole, but geographically it is distributed in its individual parts (as separate stores or shopping stocks) and it is reflected in additional indicators.

For the purposes of the effective stores management, the barcodes that were developed during the production of each unique item are used here. All activities of the warehouse (e.g. "entering the store ", "discharge of store and exit to store"", "revision") are automated and facilitated through registering only by barcode reader. The barcode of each unit received in the warehouse contains information on:

The order, from which it was produced, including date and season;

- Material, with its characteristics – supplier, composition, colour and type;

- Cost and price, which is available in the retail network (in the country and abroad);

- Current location (warehouse or store);

- Once an item is sold – the information about its selling price, date and location on which it was sold.

All this information is stored in the common database.

An important part of the integrated system is the automatic preparation of a label, which meets the requirements of the current legislation in the country where the product will be sold, including all the relevant documentation. The integrated system allows connection via the Internet between the central warehouse of the company and the cash registers in its stores.

This allows safe and reliable supply, automatic setting of prices, revisions, transfers from one shop to another, etc. It is possible to draw a quick inquiry for units by different characteristics (e.g. barcode of similar product, providing a list sorted by key indicators such as order, material, pattern, colour, etc.). Sales are reported automatically via the Internet, each event is reflected in the common database.

The system offers the opportunity to use a mobile computer working on a wireless network that scans bar codes on products. The mobile application for warehouse management is developed, which helps to integrate the material and information flows in the warehouse at:

- An inventory of stocks;
- Acceptance of goods in a remote object.

2.1.6 Feedback from customers

This module contains the data about the customers of the company; their main characteristics used to prepare and specify a certain trade policy: pricing (for individuals, corporate, or general customers), discounts, terms, payment schemes, etc.

An important element in the overall system is a module that contains information about the regular customers of the company. This statistical information regarding the main customers, allows the company to provide better and faster services and to formulate specific trade policy by:

- Setting prices - individual, corporate, general;
- Determination of discounts for a specific client;
- Order prioritization, time for processing orders, production time and terms;
- Offer individual services (measurements, samples, product delivery and more);
- Individual payment schemes, etc.

2.2 Analytical data processing

The collected information allows making informed decisions on different levels:

- On the operational level: to monitor the actual recruitment of the each employee‘s contribution in the production process, salary formation, etc.;
- On the strategic level: the information is used in processes of strategic planning of staffing needs, planning costs for salaries, creating a system for benefits and rewards;
- At a tactical level: the information is used to manage the recruitment, distribution and motivation of the employees.

3 CONCLUSIONS

The main advantage of the proposed integrated system is the improved coordination management of all resources in the company.

The information from the receiving of the order for a certain product until its realization on the market is transferred between the various subsystems automatically and in real time. Improved efficiency with less working resource is achieved.

Furthermore, the support of the information system for all integrated processes allows the management team widely applicable tool for the analysis of the main activity. It allows real-time receipt of complete information about the observed processes and, therefore, the discovery of important trends in the business.

ACKNOWLEDGEMENTS

The research work reported in the paper is supported by the project AComIn "Advanced Computing for Innovation", grant 316087, funded by the FP7 Capacity Programme (Research Potential of Convergence Regions).

REFERENCES

Petter S., DeLone, W. and E. McLean, 2008. Measuring information systems success: models, dimensions, measures, and interrelationships, *European Journal of Information Systems*, 17, 236–263.

Petrauskas V., 2006. The use of information flow analysis for building an effective organization. Information Technology and Control, 2006, Vol.35, No.4, pp. 441

Ilie-Zudor E., Kemeny Z., F. Blommestein, L. Monostori, A. Meulen, 2011. A survey of applications and requirements of unique identification systems and RFID techniques. *Computers in Industry*, Vol. 62, Issue 3, pp. 227-252.

Li Q. et al. 2010. Business processes oriented heterogeneous systems integration platform for networked enterprises, *Computers in Industry*, Vol. 61, pp. 127–144.

Tashev T.D., Hristov H., 2001. Modelling and Synthesis of Information Interactions, *Problems of Technical Cybernetics and Robotics*, No 52,. pp. 75-80.

Daskalova H., Kolchakov K., 2000. Some investigations for optimizations of information interchange. Int. Conf. *"Communication, Electronic and Computer Systems"*, Vol.1, pp.206-210.

Xu, W., Zhao, X., 2014. A two-factor product anti-counterfeiting and secure tracing management system based on RFID and two-dimensional bar code, *Applied Mechanics and Materials*, Vol. 469, 2014, pp. 490-493.

Aslam, T., Ng, A.H., Karlsson, I., 2014. Integrating system dynamics and multi-objective optimisation for manufacturing supply chain analysis, International Journal of Manufacturing Research, Vol. 9 (1), 2014, pp. 27-57.

Lu, W., Giannikas, V., McFarlane, D., Hyde, J, 2014. The role of distributed intelligence in warehouse management systems, *Studies in Computational Intelligence*, Vol. 544, pp. 63-77.

Umble E., R. Haft, M. Umble, 2003. Enterprise resource planning: Implementation procedures and critical success factors, *European Journal of Operational Research*, 146, pp. 241–257.

Verwijmeren M., 2004. Software component architecture in supply chain management, *Computers in Industry*, 53, pp. 165–178.

Valente P., G. Mitra, 2007. The evolution of web-based optimisation: From ASP to e-Services, *Decision Support Systems*, 43, pp. 1096–1116.

EXAMINATION OF AN ALGORITHM FOR NON-CONFLICT SCHEDULE WITH DIAGONAL ACTIVATION OF JOINT SUB MATRICES IN A LARGE SCALE SWTCHING MATRIX

Kiril Kolchakov, Vladimir Monov

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences,
Acad. G. Bonchev Str., Sofia, Bulgaria,
kkolchakov@abv.bg , vmonov@iit.bas.bg.

Keywords: Network nodes, Message switching Node traffic, Crossbar switch, Conflict elimination, Packet messages.

Abstract: The aim of this study is to determine the influence of the size of submatrices in a lagre scale connections matrix on the performance of an algorithm for non conflict scheduling with diagonal activation of joint submatrices (ADAJS). We have used a software model (SMADAJS) of the algorithm developed in MATLAB programming environment. A comparison is made with similar algorithms with diagonal activation of the connections matrix with respect to performance time and memory required.

1 INTRODUCTION

The traffic via Crossbar switching nodes is casual and depends on the users. The formulation of a conflict issue during operation of the switching nodes is as follows. The dimensions of switches in the switching nodes are N x N, where N sources of packet messages are connected to N receivers via the switch of the switching node. The traffic is random by nature and conflicts are available in the following two cases:

- When one source of message requests communication to two or more message receivers
- When one message receiver receives communication requests from two or more message sources.

The evasion of conflicts is directly related to the switching node performance.

The status of the switch of the switching node is represented with the so called connection matrix. For N x N dimensional switch the dimension of the connection matrix T is N x N also, where every element $T_{ij} = 1$ if the connection request from i- source to j- receiver exists. In the opposite case $T_{ij} = 0$.

A conflict situation arises if any row of the connection matrix has more than a single 1, which corresponds to the case when one source requests a connection with more than one receiver. The presence of more than a single 1 in any column of the matrix T also indicates a conflict situation, it means that two or more sources have requested a connection with the same receiver (Kolchakov K, Monov V, 2013).

The problem of conflict situations is solved by compiling a non-conflict schedule. Algorithms using different approaches to obtain non-conflict schedules are described in (Tashev T. , 2010). Algorithms with diagonal activation of connections matrix to obtain non-conflict schedule are described in (Kolchakov K., Monov V., 2013, Kolchakov K., 2013, Kolchakov K., 2012). Software models to obtain a non-conflict schedule through the sparse matrix-masks are described in (Kolchakov K., 2009). In tis paper, we study the performance of an algorithm for non conflict scheduling with diagonal activation of joint submatrices (ADAJS) by using a software model (SMADAJS) of the algorithm developed in MATLAB programming environment.

2 DESCRIPTION OF THE ALGORITHM

The connections matrix T with N x N size, where N is being the degree of two, is divided into sub matrices (S) with dimension n x n, (n also is a degree of two), i.e:

The sets of sub matrices located along the main diagonal are processed simultaneously in each of the diagonals. For submatritces in diagonals parallel to the main one, the principle of reconciliation is used (Kolchakov K, Monov V, 2013).

The idea of synthesis of the algorithm ADAJS (Algorithm with diagonal activations of joint sub-switching matrices) is based on the knowledge that the diagonal sub matrices with requests for service in the matrix T are non-conflict in the diagonal where they are located. There are diagonals with sub matrices of requests that are non-conflict to one another. Figure 1 shows joint couples of non-conflict diagonals with sub matrices of requests for service and the main diagonal of sub matrices that can not be jointed with anyone else (Kolchakov K, Monov V, 2013).

T = [S i j] , i = 1- n, j = 1- n

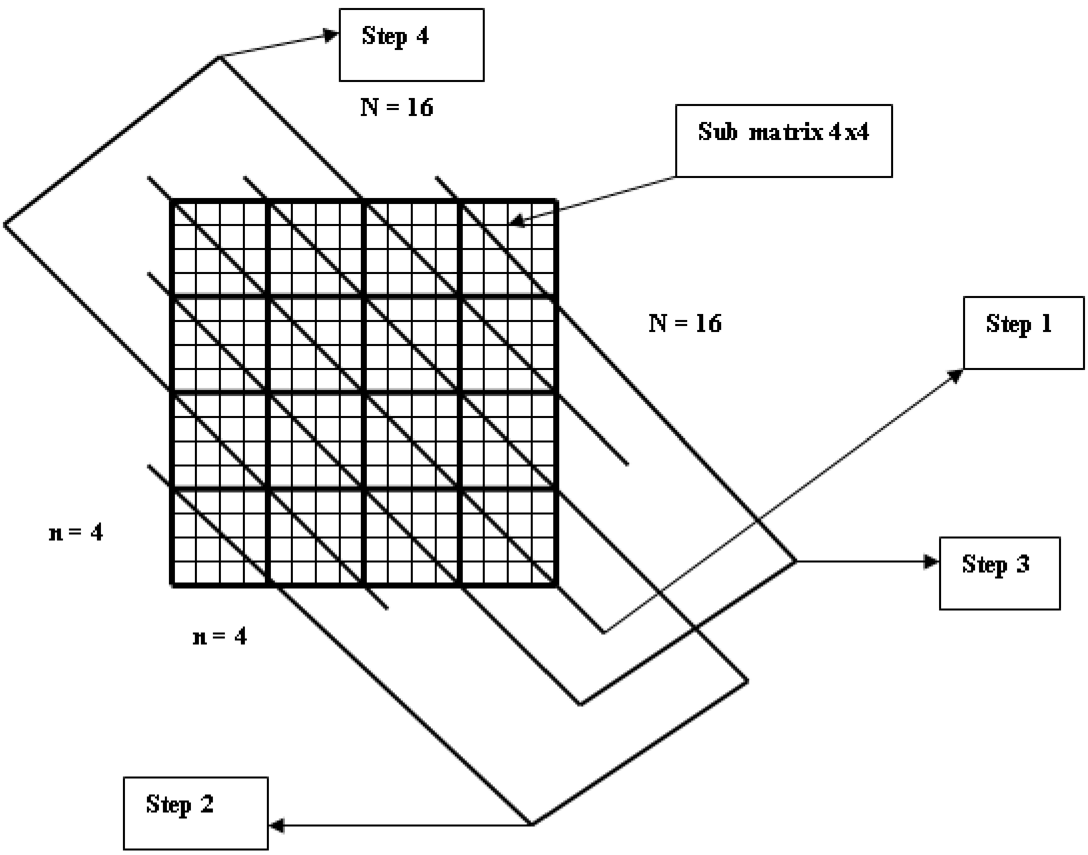


Figure 1: Diagonal activation of joint sub matrices.

The whole process of the implementation of ADAJS algorithm for obtaining a non-conflict schedule is divided into steps. The first step refers to the main diagonal sub matrices processed simultaneously and without conflict. The next steps are related to the reconciliation of the diagonals parallel to the main diagonal by pairs (Figure 1).

The analytical description of the steps shown in Figure 1 is as follows:

Step1 : S_{11} , S_{22} , S_{33} , S_{44} Step3 : S_{21}, S_{32} , S_{43} , S_{14}

Step2 : S_{41} , S_{12} , S_{23} , S_{34} Step4 : S_{31}, S_{42} , S_{13}, S_{24}

$T = [\, S_{ij} \,]$, $i = 1 - 4$, $j = 1 - 4$

The size (n) of the sub matrix determines the number of steps (I) as follows

$$I=N/n \qquad (1)$$

For **N** = const., I = f (n), where $1 < n \leq N / 2$.

3 TEST RESULTS WITH ADAJS FOR LARGE VALUES OF N

The software model SMADAJS, describing the algorithm ADAJS is written in MATLAB programming language. Our study of SMADAJS software model is performed on Dell OPTIPLEX 745 (Core 2 Duo E6400 2,13GHz, RAM 2048).

Figure 2 shows in a grphic form the performance time of the algorithm for values of N from N = 1024 to N=32768. Figure 3 presents data on the speed, when N is ranging from 65536 to 1048576.

The results from figures 2 and 3 enable us to determine the optimal values of the size size n of submatrices with respect to the performance time of the algorithm for different sizes N of the connections matrix. These values are shown in Table 1.

Our study with SMADAJS also shows that the required memory M [KB] of the algorithm depends only on the size of the submatrix n, at
N = const. This is illustrated in figure 4.

Table 1: Optimal values of the size n.

N	1024	2048	4096	16384
$n_{opt.}$	8	8	4	8
N	32768	65536	131072	262144
$n_{opt.}$	8	8	4	4
N	524288	1048576	-	-
$n_{opt.}$	4	8	-	-

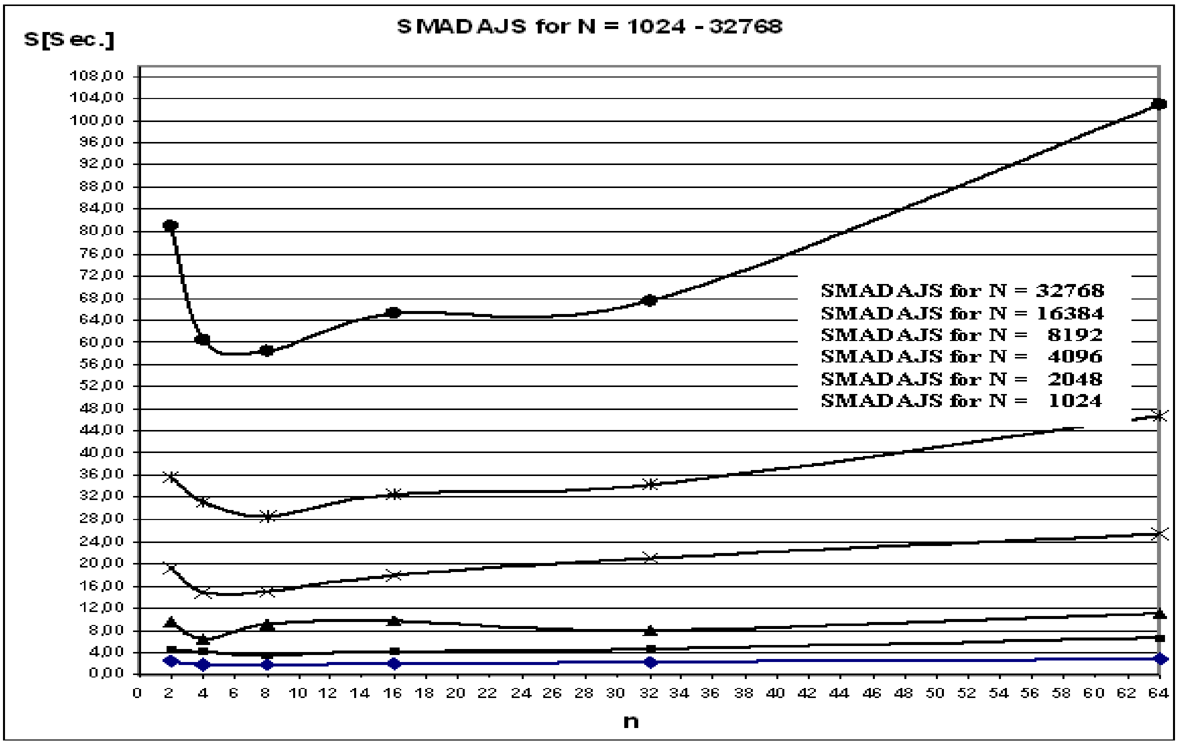


Figure 2: The performance time of the algorithm for values of N from N = 1024 to N=32768.

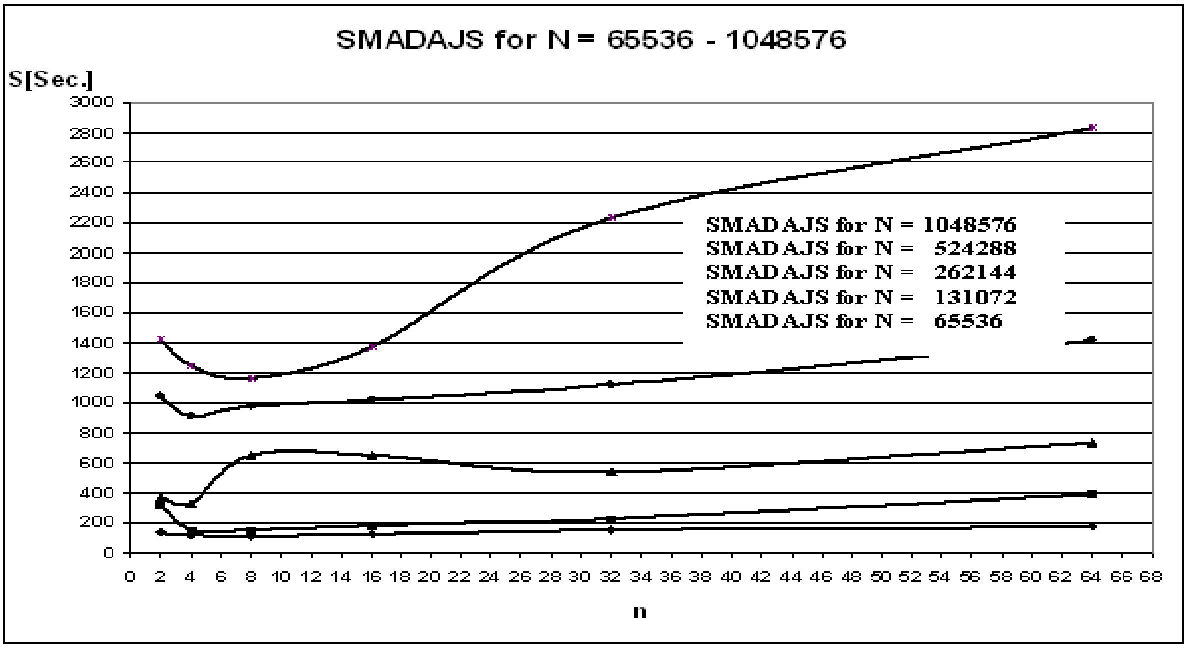


Figure 3: The speed, when N is ranging from 65536 to 1048576.

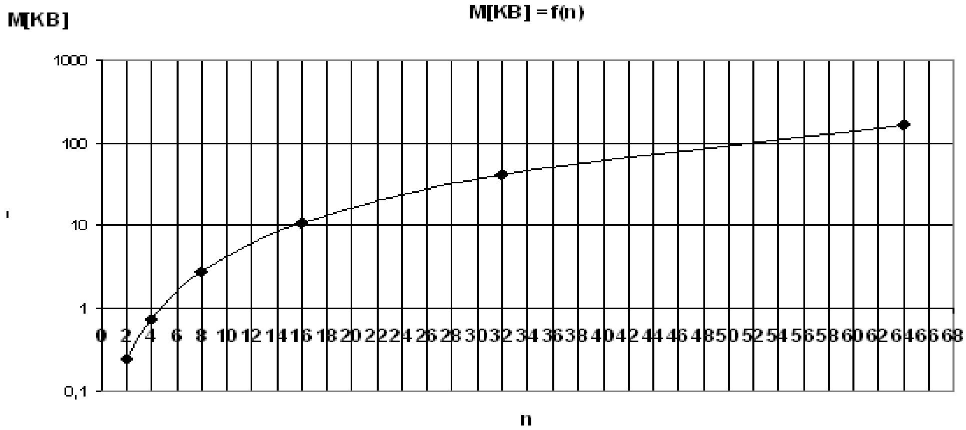


Figure 4: Required memory.

4 A COMPARISON OF THREE SOFTWARE MODELS PERFORMANCE

A software model's performance (P) is defined as a ratio of the non- nil resolutions to the total number of the solutions. R(v) is the set of the nil solutions, R(w) is the set of the non-nil solutions, and R is a set of the all solutions (Kolchakov K., 2013).

$$R=R(v)+R(w) \quad (2)$$

$$P=(R(w)/R).100[\%] \quad (3)$$

From formula (3) it is seen that when the number of nil solutions R(v) vanishes to nil, then the performance P tends to 100% (Kolchakov K., 2013).

To facilitate the performance examination, five kinds of matrices for simulation of the input connectivity matrix T are chosen. The special input matrices 2A, 2B, 2C, 2D and 2E (Kolchakov K., 2013) are represented on Figure 5.

1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1

2A

1	1	0	0	0
1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	0	1	1

2B

0	0	1	1	1
0	0	0	1	1
1	0	0	0	1
1	1	0	0	0
1	1	1	0	0

2C

0	0	0	1	1
1	0	1	0	1
0	1	1	0	0
1	1	1	1	0
1	0	1	1	1

2D

1	1	1	1	1
0	1	1	1	1
0	0	1	1	1
0	0	0	1	1
0	0	0	0	1

2E

Figure 5: The special input matrices.

In Table 2, we have represented the investigation results related to the performance P of the software models SMADA, SMAJDA and SMADAJS for each of the above input matrices.

P[%]	2A	2B	2C	2D	2E
SMADA	6,66	20	80	80	53,3
SMAJDA	6,66	20	80	80	100
SMADAJS	12,5	37,5	87,5	100	100

Table 2: Comparison in terms of performance

A comparison of the results in Table 2 shows that the performance of SMADAJS is superior than the performance of SMADA and SMAJDA.

5 CONCLUSIONS

The main conclusions of the paper are as follows. The optimal size of the submatrices n_opt. with respect to the speed is characterized by two values n_opt. = 4 and n_opt. = 8 for the different sizes of the connections matrix. The memory required is determined by the size of submatrices n and it does not exceed 200 MB. The performance of SMADAJS is faster than that of SMADA and SMAJDA.

ACKNOWLEDGEMENTS

The research work reported in the paper is supported by the project AComIn "Advanced Computing for Innovation", grant 316087, funded by the FP7 Capacity Programme (Research Potential of Convergence Regions).

REFERENCES

Kolchakov K., Monov V., 2013. An Algorithm for Non – Conflict Schedule with Diagonal Activation of Joint Sub Matrices, *17-th International Conference on "Distributed Computer And Communication Networks (DCCN-2013):"Control, Computation, Communications, Moscow, October 07-10, 2013, pp. 180-187, ISBN 978-5-94836-366-0.*

Kolchakov K., 2013. Comparative Analysis of Class Algorithms for Non-Conflict Schedule in Switching Nodes, *17-th International Conference on "Distributed Computer And Communication Networks (DCCN-2013): "Control, Computation, Communications, Moscow, October 07-10, 2013, pp. 172-179, ISBN 978-5-94836-366-0.*

Kolchakov K., 2012. Examination on Algorithms for Non Conflict Schedule with Diagonal Activation in Case of a Large Size Switching Matrix. *International Conference Automatics and Informatics `12 03-05.10.2012 Sofia, Bulgaria. Proceedings CD: ISSN 1313 – 1869. pp.341-344.*

Kolchakov K., 2011. "An Algorithm Synthesis of Non-Conflict Schedule by Diagonal Connectivity Matrix Activation" *Proceedings of the International Conference AUTOMATICS AND INFORMATICS`11, John Atanasoff Society of Automatics and Informatics, Bulgaria, Sofia 03.10-07.10.2011., pp. B-247 – B251, Proceedings ISSN 1313-1850, CD ISSN 1313-1869.*

Tashev T. , 2010. Computering simulation of schedule algorithm for high performance packet switch node modelled by the apparatus of generalized nets. *Proceedings of the 11th International Conference CompSysTech'2010, 17-18 June 2010, Sofia, Bulgaria. ACM ICPS, Vol.471, pp.240-245.*

Kolchakov K., 2009. Software Models with Sparse Mask Matrixes for Non-Conflict Schedule obtaining in a Switching Node, In *"Distributed Computer and communication networks DCCN 2009", Moscow: R&D Company "Information and Networking technologies" ISBN 978-5-9901871-1-5.*

Wanjari P., Chonhari A., 2011. "Implementation of 4x4 crossbar switch for Network Processor", *International Journal of Emerging Technology and Advanced Engineering,* Website: www.ijetae.com, Volume 1, Issue 2, December 2011).

Kim D., Lee K. and Yoo H., 2005. A Reconfigurable Crossbar Switch with Adaptive Bandwidth Control for Networks-on-Chip, *IEEE Int. Symp.on Circuits and Systems, 2005.*

LOAD OPTIMIZATION IN A GRID STRUCTURE FOR PARALLEL COMPUTER SIMULATIONS OF THE THROUGHPUT OF A CROSSBAR SWITCH NODE

Tasho Tashev, Vladimir Monov
Institute of Information and Communication Technologies, Bulgarian Academy of Sciences,Acad. G.Bonchev Str Bl.2, Sofia, Bulgaria
ttashev@iit.bas.bg, vmonov@iit.bas.bg

Radostina Tasheva
Department of Applied Physics, Technical University,St.Kl.Ohridski Bul 8, Sofia, Bulgaria
rpt@tu-sofia.bg

Keywords: Computer Technologies, Simulations, Optimization, Grid structures, Crossbar switch.

Abstract: The In the present paper we employ the grid structure of IICT-BAS for parallel computer simulations of the throughput of a crossbar switch node. In our simulations we use PIM-algorithm for non-conflict scheduling in a crossbar node with hotspot load traffic. The obtained simulation results enable us to propose a procedure for optimizing the load of a grid structure in order to minimize the overall time of performance.

1 INTRODUCTION

Modern digital information systems are built according to the principle of exchange of discrete portions of information called packets. Communication nodes in these systems are called router and switch. A crossbar switch node routes traffic from the input to output where a message packet is transmitted from the source to the destination. The randomly incoming traffic must be controlled and scheduled to eliminate conflict at the crossbar. The goal of the traffic-scheduling for the crossbar switches is to maximize the throughput of packet through a switch and to minimize packet blocking probability and packet waiting time (Kang, et all., 2013). This is assured by the algorithm for calculation of non-conflict schedule which is running in the control unit of the switch (Scheduler, Fig.1) (Csaszar, et all, 2007).

The problem of calculating of a non-conflict schedule is NP-complete (Chen, et all, 1990). Increasing data volumes (Атанасова, 2010) and increasing the speed of transmission lines of communication require new, more efficient algorithms for the calculation of the conflict-free schedule. The efficiency of these algorithms can be verified with a formal or simulation tools.

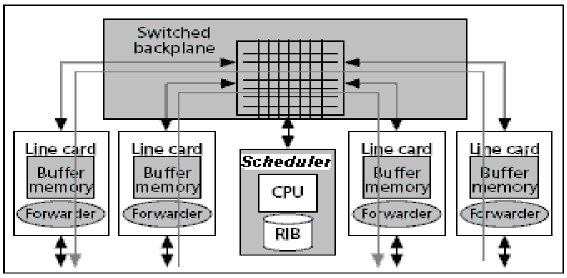


Figure 1: Third generation crossbar switch structure.

The efficiency of the algorithms for switches in the first place can be evaluated by using bandwidth output channels (throughput) (Kolchakov, 2010, Kang, et all., 2013). The incoming traffic may be uniform or non-uniform. The study of an algorithm's throughput begins with modeling of the switch throughput under uniform load traffic (uniform i.i.d. Bernoulli traffic). The next step is to investigate the algorithm's throughput under non-uniform traffic (.Chang, 2012).

In the previous paper (Ташев, et all, 2013) we proposed a numerical procedure for calculating the upper boundary of the throughput for crossbar switch node using the results of a simulation throughput. If throughput of a crossbar node increases to a certain limit (monotonically), the procedure provides one unique solution. We performed simulations for a specific algorithm for non-conflict schedule, a model for incoming traffic and a load intensity. Our modeling of the throughput utilizes PIM-algorithm (Anderson, et all, 1993), Chao-model for hotspot load traffic (Chao-Lin, Yu, et all , 2007) and ρ=100% load intensity of each input (i.i.d. Bernoulli) (Tashev, 2011). In this case the throughput of the crossbar node increases monotonically to a certain limit. In case of a size n of the switch field: $n \in [3,70]$ and the scale i of input buffers loading for Chao-model : $Chao_i$, $i \in [1,10]$ the results are shown in Fig.2 (Tashev, 2011).

To increase the accuracy of the numerical procedure the simulations should be performed for a large size of the switch field and in a large scale of input buffers loading for Chao- model. Such simulations and the necessary computations are commonly carried out by using grid-computer structures. In this case we should solve the problem for optimal loading resources of the grid structure for parallel computer simulations.

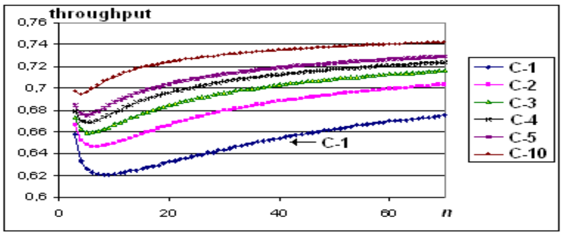


Figure 2: Throughput of PIM-algorithm with Chao (C-1, 2, 3, 4, 5, 10) traffic (Tashev, 2011).

In this paper we will research in what conditions for parallel simulations the values of the time load of processors of grid structure will be equal. In particular, for scale i=const, how to divide the interval [$n1$, $n2$] modeling the size of switch field n by subintervals for this aim.

2 CONDITIONS FOR THE SIMULATION

Our simulation of the throughput utilizes a PIM-algorithm specified by the apparatus of Generalized nets (Atanassov, 1997). The utilized GN-model for the PIM-algorithm is specified in (Tashev, Monov, 2012). The transition from a GN-model to executive program is performed as in (Tashev, Vorobiov, 2007) using the program package VFort of the Institute of mathematical modeling of Russian Academy of Sciences (Vabushkevich, 2009). The source code has been executed by means of Windows XP SP2 and IBM PC compatible computer with Intel Pentium IV 3 GHz and 2 GB RAM.

We utilize a family of patterns of a traffic matrix ***T*** for a non-uniform traffic simulation based on the hotspot (Chao) model (Tashev, 2011). This model is given by: $\lambda_{ii} = 0,5\rho$ for $i = j$ and $\lambda_{ij} = 0,5\rho/(n\text{-}1)$ otherwise, $i, j \in 1, \ldots , n$, where ρ is the load intensity (i.i.d. Bermoulli) of each input (Chao-Lin, Yu, et all , 2007).

The times of simulations for $Chao_i$, $i\in[1, 10]$ and $n\in[3, 70]$ are shown in Fig.3 (the average values of time by 10 000 simulations for each n). In the figures below, $Chao_i$ is denoted as C-i for i = 1, 2,... . The time of simulation depends linearly by scale i of Chao-model and depends approximately on the third power of the size n of commutation field.

Let the source code has been tested on Vfort and then compiled by means of the grid-structure BG01-IPP of the Institute of information and communication technologies - Bulgarian Academy of Sciences (http:// www.grid.bas.bg).

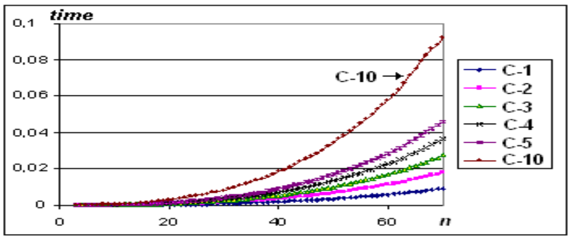


Figure 3: Simulation time of PIM-algorithm with Chao (C-1, 2, 3, 4, 5, 10) traffic (Tashev, 2011).

The resulting executive code is performed on the grid-structure. One main restriction is the time for execution. The results for throughput and time executions of PIM-algorithm with $Chao_1$ are shown in Fig.4 (10 000 simulations for each n, executed sequentially).

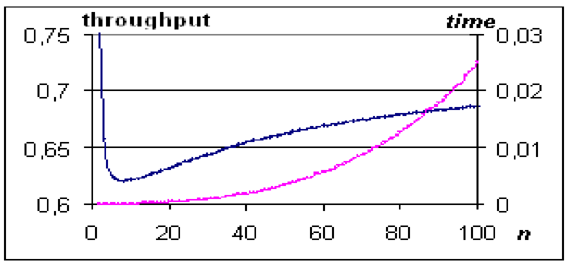


Figure 4: Throughput and time of PIM-algorithm with $Chao_1$ traffic by grid structure.

The results for throughput are identical with IBM PC. The results for time execution are similar - approximately on the third power. We can proceed to large-scale simulations using a grid structure. A main restriction is the time for execution. How to divide the interval [*n1*, *n2*] by subintervals for parallel execution each, for our aim?

3 CALCULATING OF SUBINTERVALS

Let us assume that the variable *n* is continuous and designate it with *x*. Then the time execution is of the type $y=a.x^3$, where *a* is a constant depending on the hardware used. In this case the area S under the curve $y=a.x^3$ is equal to the general time of simulation. Therefore, if we divide this area into two equal parts, we will receive the required subintervals (along *x*) for the case of two parallel executed tasks. If we divide the area into three, we will receive the subintervals for 3 parallel executed tasks etc. Formally this leads to the following calculations (showed in Fig.5):

$$S(x)=\int_A^B ax^3dx \rightarrow S(x)=\frac{a}{4}x^4|_A^B$$

For our case A=0, B>A.

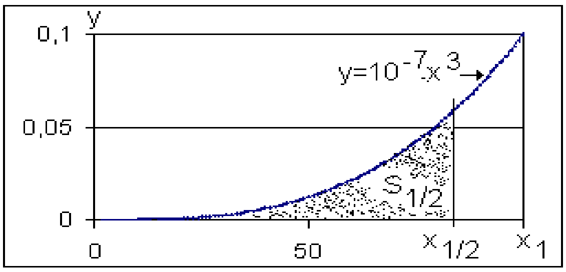


Figure 5: Analog approximation of half-time execution.

We will designate the time for simulation from dimension A=0=n1 to dimension B (x_1=B=n2 is given) as:

$$S_1(x_1) = (1/4).a.x_1^4 \qquad (1)$$

The time for simulation equal to the half of the latter will be:

$$S_{1/2}(x_{1/2}) = (1/4).a.x_{1/2}^4 \qquad (2)$$

If $S_{1/2}$ is calculated as half of S_1:

$$S_{1/2}(x_{1/2}) = (1/2).(1/4).a.x_1^4 \qquad (3)$$

We can write:

$$(1/4).a.x_{1/2}^4 = (1/2).(1/4).a.x_1^4 \qquad (4)$$

And then:

$$x_{1/2}^4 = (1/2).x_1^4, \qquad (5)$$

therefore

$$x_{1/2}=(1/2)^{1/4}.x_1 \qquad (6)$$

Analogically when we divide into 3 equal parts:

$$x_{1/3}=(1/3)^{1/4}.x_1 \text{ , } x_{2/3}=(2/3)^{1/4}.x_1 \qquad (7)$$

Analogically 4 equal parts, etc.:

$$x_{1/4}=(1/4)^{1/4}.x_1$$
$$x_{2/4}=(2/4)^{1/4}.x_1 \qquad (8)$$
$$x_{3/4}=(3/4)^{1/4}.x_1$$

Of course, the boundary of last interval is:

$$x_{4/4}=(4/4)^{1/4}.x_1= x_1 \qquad (9)$$

The results $x_{i/k}$ up to 8 equal parts are given in Table 1.

Table 1: Numbers of boundary of subintervals.

Tasks *k*	2	4	8
subintervals *i* 1/8			0,594604
2/8		0,707107	0,707107
3/8			0,782542
4/8	0,840896	0,840896	0,840896
5/8			0,889139
6/8		0,930605	0,930605
7/8			0,967168
8/8	1	1	1

Using the formula $x_{i/k} = (i/k)^{1/4}.x_1$ we can calculate the boundaries for each *i*-subinterval in a given number *k* of subintervals.

4 ACCURACY OF THE LOAD OPTIMIZATION WITH SUBINTERVALS

Let us assume that the variable *x* is analog (Fig. 5). We calculate the differences between different *x* and different $S_{1/16}$ (*x*) ($y=x^3$) time executions - for 16 subintervals. We build a line for differences as shown in Figure 6.

We have 2 lines: Δ_1 = x (accuracy of 4 significant digits -Table 1) – x (exact value) (down), Δ_2 = y (accuracy of 4 significant digits for x -Table 1) – y (exact value of y) (up). Differences between different x are smaller than 0.005%.

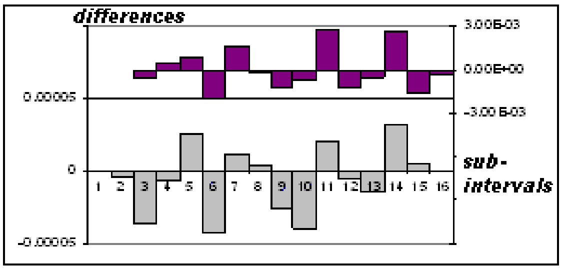


Figure 6: Differences between different time for S (top - analog 4 digits) and between different x (down) for 16 tasks.

Time execution, normalized for $S_{1/16}$ is shown in Fig.7. Over a continuous x the time differences are smaller than 0.3% relative to the theoretical time of $S_{1/16}$. The results for subintervals (x is analog) is very good.

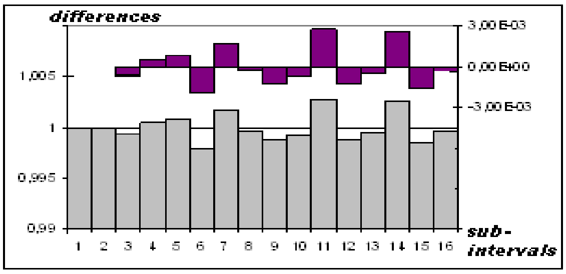


Figure 7: Differences – normalized by average for $S_{1/16}$ (down).

Let us assume that the variable x is discrete (Fig.8). For $n2$=100, a=1.10^{-7}, 2 subintervals, the error for continuous x is 2% (good). We choose the boundary of subintervals to be the nearest integer.

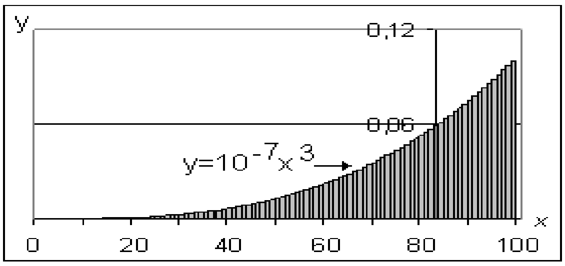


Figure 8: Discrete approximation of half-time execution.

For $n2$=100, a=1.10^{-7}, 16 subintervals, the differences between different S for discrete x reaches 40% (bad – Fig. 9). The differences between different discrete x are smaller than 0.5%. We choose the boundary of subintervals to be the nearest integer. In Fig. 10 are shown the times for S normalized by average for $S_{1/16}$.

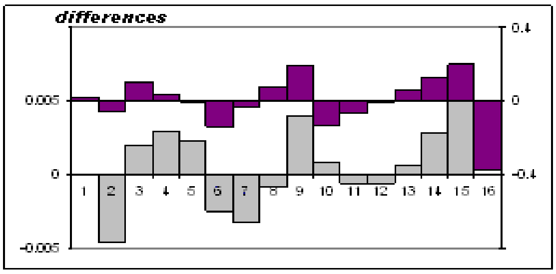


Figure 9: Differences between the times for $S_{1/16}$ (top) and x (down - discrete 2 digit).

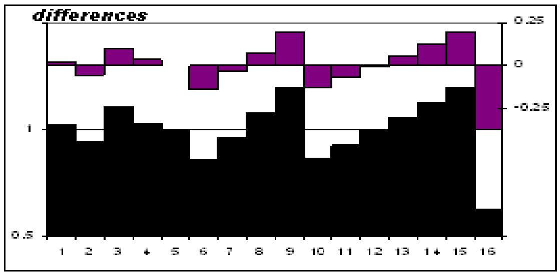


Figure 10: Differences between different S (down-normalized)

Can we reduce the difference of 40%? We may choose for boundaries of subintervals the nearest integer x “plus” shift to the left by 1 (minus 1). The results of this choice are shown on Fig.11 ($n2$=100, a=1.10^{-7}, 16 subintervals).

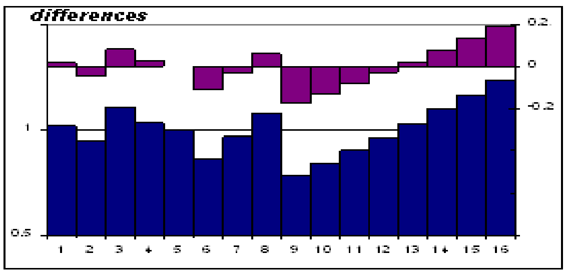


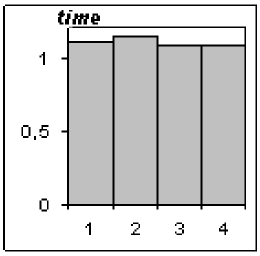
Figure 11: Differences between diff. S for x shift left by 1.

The difference between different S reaches 25%. That's the best we can get, as the following shifts do not improve the situation. We expect that this is possible when you select the interval type [1/2.$n2$, $n2$], $n2 \geq 200$. This requires new research. Now we have to check the real situation: the execution time of simulations on the grid structure.

5 GRID SIMULATION USING THE SUBINTERVALS

The results of the computational experiments for a given patterns for hotspot model of incoming traffic and model of PIM-algorithm, by means of the grid-structure BG01-IPP (http://www.grid.bas.bg).

For $Chao_{10}$, $n2$=100 ($y \approx 3.10^{-7}x^{2.9}$), 4 subintervals, the differences between different S for discrete x reaches 3,5% (Fig.12 up). For $Chao_{10}$, $n2$=200 ($y \approx 5.10^{-7}x^{2,8}$), 4 subintervals, the differences between different S for discrete x reaches 2,1% (Fig. 12 down).

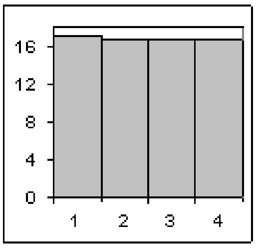


	time
70	1,104078
84	1,143542
93	1,086461
100	1,083513

Average $S_{1,4}$= 1,104

[time] x10^4sec

difference max 3,54%



	time
141	17,16664
168	16,80988
186	16,57685
200	16,69185

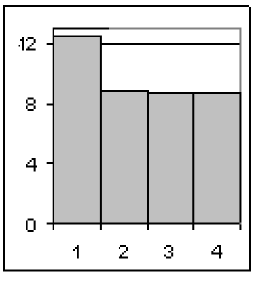
Average $S_{1,4}$=16,8113

[time] x10^4sec

difference max 2,11%

Figure 12: Time S for $Chao_{10}$, $n2$=100 and 200 with 4 tasks.

For $Chao_5$, $n2$=200 ($y \approx 4.10^{-8}x^{3,2}$), 4 subintervals, the differences between different S for discrete x reaches 43% (Fig.13).



	time
141	12,5088
168	8,780926
186	8,683511
200	8,747368

Average $S_{1,4}$= 9,6801?

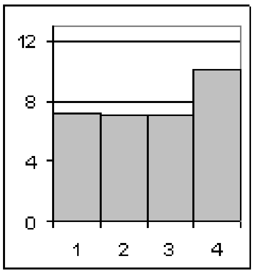
[time] x10^4sec

difference max 43%

Figure 13: Time S for $Chao_5$, $n2$=200 with 4 tasks

For $Chao_4$, $n2$=200 ($y \approx 3.10^{-8}x^{3,1}$), the differences between different S for discrete x reaches 43% (Fig.14).

We observed large as well as small differences in the execution time of simulations, when using patterns $Chao_3$ and $Chao_8$ (4 tasks). Above we have shown the extreme cases.



	time
141	7,23771
168	7,061804
186	7,015488
200	10.10794

Average $S_{1,4}$= 7,8557?

[time] x10^4sec

difference max 43%

Figure 14: Time S for $Chao_4$, $n2$=200 with 4 tasks

Why is that so? Such differences are expected in 16 tasks. Why in the last two cases we have such differences? During the simulation for $Chao_{10}$, the grid structure has performed 6 tasks (4 ours and two others). During the simulation for $Chao_5$ and $Chao_4$, the grid structure has performed more than 20 tasks (our tasks are 4). From this we can conclude that the constant a has different values depending on the grid load. We must keep this in mind. The building of Table 1 is just the beginning of the calculation optimization. For this purpose, we have to optimize our simulations using a practical procedure.

6 THE PROCEDURE FOR CALCULATING THE SUBINTERVALS

We give an informal description of the procedure for determining the subintervals for simulations the throughput of the switch.

Step 1. We have a program for simulations that is already tested. We determine the interval [1,n2] for 1 hour simulation with 1 task. We run serial simulation.

Step 2. We check the coefficients in $y=a.x^3$. We run 5–hour serial simulation (~interval [2, 1.5 $n2$]) by 2 tasks.

Step 3. We check the coefficients in $y=a.x^3$. We run 16–hour serial simulation (~interval [2, 2 $n2$]) by 4 tasks.

Step 4. We determine the differences of time for the tasks.

Step 5. We choose the interval simulation according to our authorized resources (CPU and time), starting by 5 tasks, ½ autorized time by using expansion Table 1.

Step 6. We run large-scale simulations with *k* subintervals.

Step 7. We obtain the results. If the results satisfy us, we stop. If the results do not satisfy us, we repeat from **Step 4**.

Approbation of the procedure is done with the patterns for hotspot model of incoming traffic and model of PIM-algorithm.

The procedure is applicable to simulations of throughput of other algorithms and different patterns of incoming traffic (with the complexity of calculation of $O(n^2)$ to $O(n^5)$).

7 CONCLUSIONS

In the present paper the possibility is investigated for "a priory" distribution of the resources of a grid-structure at parallel simulation of a task with calculation complexity $O(n^3)$.

An equation is derived for dividing the interval [1,x] of the argument *n* into finite number of subintervals which determine the processor tasks of the grid-structure. The aim is for the separated tasks to finish simultaneously – i.e. the loading to be uniform. Theoretically the suggested formula offers less than 2 % time lag difference in the task execution. During the fulfilled simulations (simulation of throughput of PIM-algorithm with hotspot incoming traffic) are observed bigger differences in the time for parallel execution of 4 task.

A procedure is suggested for decreasing of the differences. This procedure is useful for the initial period of a series of simulations. When large scale simulations are implemented, though, investigations show that it is better a directive for dynamic optimization to be used, for example those of OpenMP.

ACKNOWLEDGEMENTS

The research work reported in the paper is supported by the project AComIn "Advanced Computing for Innovation", grant 316087, funded by the FP7 Capacity Programme (Research Potential of Convergence Regions).

REFERENCES

Kang, K., Park, K., Sha L., Wang Q., 2013. Design of a crossbar VOQ real-time switch with clock-driven scheduling for a guaranteed delay bound. Real-Time Systems, January 2013, Vol. 49, Issue 1, pp. 117-135..

Csaszar, A., Enyedi, G., Retvari, G., Hidell, M., Sjodin, P., 2007. Converging the Evolution of Router Architecture and IP Networks. IEEE Network, Vol.4, 8-14.

Chen, T., Mavor, J., Denyer, Ph., Renshaw, D., 1990. Traffic routing algorithm for serial superchip system customisation. IEE Proc.- E.part, Jan 1990, Vol. 137, No.1, pp. 65-73.

Атанасова, Т., 2010. E-Home - агрегиране на данни за повишаване на енергийна ефективност. Съвместен научен семинар „Моделиране и управление на информационните процеси". Сборник трудове, КТП, София, стр. 69-75.

Kolchakov, K., 2010. An Approach for Performance Improvement of Class of Algorithms for Synthesis of Non-conflict Schedule in the Switch Nodes. Proc. of the 11th Int. Conf. CompSysTech'10, 17-18 June 2010, Sofia, Bulgaria. ACM Press, ICPS Vol. 471, pp. 235-239.

Chang, H., Qu, G., Zheng, S., 2012. Performance of CTC(N) Switch under Various Traffic Models. Springer, LNEE, Vol. 126, pp. 785-793.

Ташев, Т., Баканова, Н., Ташева, Р., 2013. Исследование верхней границы пропускной способности коммутационного узла при входящем трафике типа "горячей точки". International Journal "Information Technologies & Knowledge", Vol.7, No. 2, pp. 182-189.

Anderson T., Owicki S., Saxe J., and Thacker C., 1993. High speed switch scheduling for local area networks. ACM Trans. Comput. Syst., Vol. 11, No.4, Nov. 1993, pp. 319-352.

Chao-Lin, Yu, Chang C-S., Lee. D-S., 2007. CR Switch: A Load-Balanced Switch with Contention and Reservation. IEEE/ACM Transactions on Networking, Vol. 17 No.5, October 2007. pp.1659–1671.

Tashev, T., 2011. Modelling throughput crossbar switch node with nonuniform load traffic Proc. of the International Conference "DCCN-2011", October 26-28 2011, Moscow, Russia. Moscow, R&D Company "Information and Networking Technologies". pp. 96-102. (in Russian)

Atanassov, K., 1997. Generalized Nets and System Theory. Acad.Press "Prof. M. Drinov", Sofia, Bulgaria.

Tashev, T., Monov, V., 2012. Modeling of the Hotspot Load Traffic for Crossbar Switch Node by Means of Generalized Nets. Proc. of Intelligent Systems (IS), 6th IEEE Int. Conference, 6-8 Sept. 2012, Sofia, Bulgaria. pp. 187-191

Tashev, T., Vorobiov, V., 2007. Generalized Net Model for Non-Conflict Switch in Communication Node. Proc. of Workshop "DCCN'2007", Sept 10-12 2007, Moscow, pp. 158-163.

Vabushkevich, P., 2009. http://www.imamod.ru/~vab/vfort/download.html

OPTIMISATION PROCEDURES IN SMEs FINANCIAL MECHANISM

George L. Shahpazov, Lyubka A. Doukovska

Institute of Information and Communication Technologies
Bulgarian Academy of Sciences
Acad. G. Bonchev str., bl. 2, 1113 Sofia, Bulgaria
atlhemus@abv.bg

Keywords: SMEs financial mechanism, Decision making, Intelligent systems

Abstract: In the present paper is discussed the mechanism of bank support of small and medium-sized enterprises (SMEs). Analysis is made of the effectiveness of the bank's internal financial structural unit and hierarchy, and it is shown how the concept of Multi-Criteria Decision Analysis (MCDA) can be applied to the process of evaluating creditworthiness of the SMEs applications for bank loans, from the bank's perspective. The presented approach aims to yield estimations of the effectiveness of the process, taking consideration of the aspects of uncertainty, which is an inherent part of the processes of evaluation of applications for bank support and evaluation of the process itself.

1 INTRODUCTION

Multi-Criteria Decision Analysis (MCDA) is a valuable tool that we can apply to many complex decisions. It is most applicable to solving problems that are characterized as a choice among alternatives. It has all the characteristics of a useful decision support tool. It helps us focus on what is important, is logical and consistent, and is easy to use. The MCDA is useful for:

- Dividing the decision into smaller, more understandable parts;
- Analyzing each part;
- Integrating the parts to produce a meaningful solution.

When used for group decision making, MCDA helps groups talk about their decision opportunity (the problem to be solved) in a way that allows them to consider the values that each views as important.

It also provides a unique ability for people to consider and talk about complex trade-offs among alternatives. In effect, it helps people think, rethink, query, adjust, decide, rethink some more, test, adjust, and finally decide.

MCDA problems are comprised of five components:

1. Goal;
2. Decision maker or group of decision makers with opinions (preferences);
3. Decision alternatives;
4. Evaluation criteria (interests);
5. Outcomes or consequences associated with alternative or interest combination.

In general, there exist two distinctive types of MCDM problems due to the different problems settings: one type having a finite number of alternative solutions and the other an infinite number of solutions. Normally, in problems associated with selection and assessment, the number of alternative solutions is limited. In problems related to design, an attribute may take any value in a range. Therefore the potential alternative solutions could be infinite. If this is the case, the problem is referred to as multiple objective optimisation problems instead of multiple attribute decision problems.

A MCDM problem may be described using a decision matrix. Suppose there are m alternatives to be assessed based on n attributes, a decision matrix is a $m \times n$ matrix with each element Y_{ij} being the j-th attribute value of the i-th alternative.

All criteria in a MCDM problem can be classified into two categories. Criteria that are to be maximised are in the profit criteria category, although they may not necessarily be profit criteria. Similarly criteria that are to be minimised are in the cost criteria category. An '*ideal solution'* to a MCDM problem would maximise all profit criteria and minimise all cost criteria. Normally this solution is not obtainable. The question is what would be a best solution for the decision maker and how to obtain such a solution?

In paper (Xu, Yang, 2003) is shown that many small and medium enterprises (SMEs) in the UK use the beta (Business Excellence Through Action) approach to the EFQM Excellence Model to conduct business excellence self-assessment, which is in essence an MCDA problem. This paper introduces a decision support software package called Intelligent Decision System (IDS) to implement the beta approach. It is demonstrated in the paper that the IDS-beta package can provide not only average scores but also the following numerical results and graphical displays on:

- Distributed assessment results to demonstrate the diversity of company performances.
- The performance range to cater for incomplete assessment information.
- Comparisons:
  - between current performances and past performances,
  - among different companies,
  - among different action plans.
- Strengths and weaknesses.

The IDS-beta package also provides a structured knowledge base to help assessors to make judgments more objectively. The knowledge base contains guidelines provided by the developers of the beta approach, best practices gathered from research on award winning organizations, evidence collected from companies being assessed and comments provided by assessors to record the reasons why a specific criterion is assessed to a certain grade for a company. Four small UK companies, the industry partners of the research project, have carried out the preliminary self-assessment using the package. The results and experience of the application are discussed at the end of the paper (Xu, Yang, 2003).

2 COOPERATION AND COMPETITION WITH THE RELEVANT PARTNERS IN THE MARKET

Competition for investments from other PE equity funds is expected to be low due to small number of locally active funds, the fact that most are fully invested, their investment focus, and as well negative track record of some international players. Based on geography mandated and on-the-ground presence competing funds could be broadly classified in two categories.

- Funds with entirely local focus and presence like Advance Equity Holding (generalist, fully invested and preparing for exit), NEVEQ (IT focus, venture capital, fully invested)
- Regional or global with mandates for transactions in the country like Global Finance (PE firm with funds targeting SE Europe), GED Eastern Europe Fund (generalist; targeting Bulgaria and Romania), Bancroft (generalist, targeting Eastern Europe), 3TS Capital Partners (generalist, targeting CE Europe), NBG Private Equity (generalist, SE Europe), Alpha Ventures which usually have limited local presence through a representative office with 1-2 professionals.

The second group of funds usually target investments above EUR 5 million with preferred size of transactions in range of EUR 10-15 million. Their approach to investments is usually more opportunistic and targets the higher end of the mid-size company sector and companies in later stages of development.

90% of the SMEs remain unevaluated from the existing local PE sector. Traditional bank financing remains currently as well hardly accessible for SMEs. The result is a general inaccessibility to growth funding for the most of SMEs in Bulgaria.

The investment appetite of foreign PE is additionally reduced by the fact that several PE funds experienced a complete loss of single investments due to fraud and weak legal execution.

The fund management maintains excellent professional relationship with representatives of local and regional private equity funds and institutional investors. It is highly esteemed in the professional community and possesses a good network of contacts due to multitude and variety of roles in the industry and this is expected to be of vital importance in co-operating with outside investors. The fund will seek in single cases cooperation with external capital providers where sizeable follow-up rounds of financing is needed (e.g. acquisitions, elevated CAPEX needs) or in cases where the initial investment is too high or risky. Expectation at this kind of cooperation is to maintain possibility to invest in high return opportunities by sharing risk. The cooperation with other PE funds will be important for the execution of the exit strategy e.g. seeking of pre-IPO financing, secondary sale etc.

3 INVESTMENT STRATEGY

The individual investments in each fund's portfolio should be selected based on the combination between the mandatory and at least one of the optional criteria:

Mandatory Criteria:

- Management team and human resources' potential;
- Profound market and industry knowledge;
- Business model scalability;
- Distinctive competitive advantages;
- Double digit growth potential of the companies revenues;
- Clear Exit Route.

Optional Criteria:

- Value-adding opportunities through process optimization, strategy fine-tuning;
- Market scalability of the products (export);
- Potential for horizontal or vertical integration.

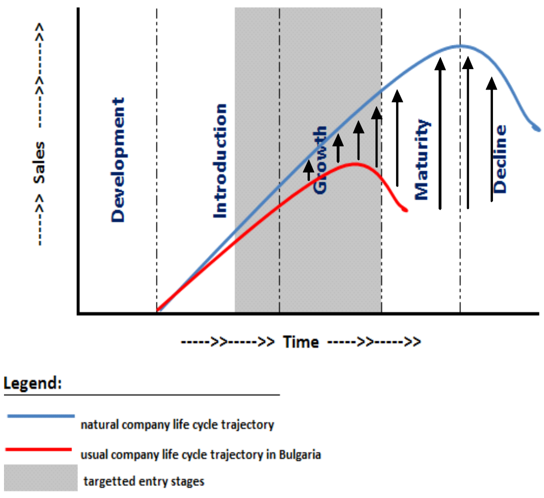


Figure 1: Expansion and optimization

The majority of SME companies in Bulgaria experience difficulties in maintaining a normal life cycle and tend to suffer from early maturity and decline without being able to materialize its full potential. There are many reasons for this, with the most common being – poor management and lack of financing. The Fund will aim in this cases at eliminating these factors with different optimization strategies, so the company converges to its natural development path and then seek expansion opportunities. Companies that have already accomplished this stage of their life cycle will be prepared for the next level.

By providing equity financing, business expansion and optimization can be achieved primarily through the implementation of various strategies: production capacity expansion; new product or a new line of products launch; commercial network development, process improvement and efficiency

More than 80% of the companies are managed with outdated structures, based on personal skills and single person's authority. We believe that implementation of modern business processes and process management would increase significantly profitability.

Optimization of the marketing strategy and establishment of adequate financial management will be in most of the investment cases the other substantial driver for successful expansion.

The step to the next lifecycle stage of the company will be achieved by providing equity capital and financial structuring of the implementation of one or several of the following strategies:

- Organic growth for companies with interesting and multipliable business models;
- Non-organic growth, horizontal integration;
- Non-organic growth, vertical integration across the value chain;
- Creating regional leaders and consolidation plays.

<u>Investment Case 1</u> – initial expansion

A diary company developing was established 2006. Original funding came from mortgaging owner's land and real estate property. These sources were sufficient to enable the company to become operational and to establish an organic production cycle. The sales experienced a steady double digit growth, but production is distributed under different conventional brands by the different contracted resellers. The owner wants to create an own brand, to be certified as an organic/bio producer and establish own/ new sales channels. The EBITDA increase after implementation is projected at 50%.

<u>Investment Case 2</u> – expansion and optimization

Food producer established in 2001 produces a high class product, higher price segment. Clear quality leader in the segment, but brand is not well established on the market. Company needs to develop and execute winning marketing strategy. Need for expansion of the production cycle to achieve economy of scale. Need for optimization of process and overall management. Due to liquidity problems and poor financial management the company is currently valuated from the owners at 2.5 x EBITDA. After expansion and optimization EBITDA is expected to triple in 4 years and exit at 5 to 6 x EBITDA would be achievable.

***Investment Case 3* – take to the next level, organic growth with a scalable business model in a highly fragmented segment**

Auto Repair Shop, all in one concept with 13 different activities brought to one place, incl. car sales, leasing and insurance consulting, guarantee repairs, general repairs, sale of parts & tires, paint, carwash, accessories, restaurant and coffee and education centre for car mechanics.

Company has 12 years experience in the different activities as single shops. The company has developed and optimized over the years an own business process management system for combining all 13 activities including the education of the employees, fully automated and centralized. This management system results in a 3 times higher productivity of the single employee compared to the averages in the industry. The resulting excess work free time is invested in education and bonus holiday as incentive.

The company plans to multiply the concept in several large Bulgarian cities.

***Investment Case 4* – creating regional player through vertical and horizontal integration**

A mid size raw material processing company with good marketing strategy and well established product basket wants parallel to widen the production cycle and to increase the output and sales. Vertical integration through acquisition is targeted to complete the production cycle, the synergy effects of the integration would increase the joint EBITDA with 25%. At a second step an acquisition of a competing company would provide additional production capacity, a further increase of the synergy effects of the integration and additional market share. The company plans to fund their growth strategy with a 50% capital increase and debt. The investor will have an exit option after 5 years with a projected IRR of 18%.

4 STRATEGY FOR CREDIT RISK DIVERSIFICATION OF FUND'S CAPITAL

Investing in growth capital in the SME sector involves substantial risk in general and particularly in emerging markets like Bulgaria.

A significant portion of this risks results from the lack of business ethics in the market and a legislation, which doesn't support in particular this kind of investments. Several cases from the experience of international PE players in Bulgaria have shown that even a complete loss of the investments is possible due to fraud and weak legal execution. We believe that the unique combination of experience in our team combining more than 120 successful financial deals in the local business environment and our understanding of the peculiarities of the execution of financial deals in Bulgaria will be crucial for mitigating the legislative and fraud risk.

In order to mitigate the business and industry risks, it is necessary to achieve a relative diversification in stages/ types of investment, industries, size and number of portfolio companies. We believe that fund needs to be able to invest in no less than eight companies in its total lifetime and not more than twelve at any moment of it.

The main purpose of the fund per definition is to support SME growth and not to takeover companies. Therefore our general intention is to hold not more than 50% of the company's equity. Although, as the mentioned negative experience of other PE investors in the country shows, even as minority shareholder it is appropriate to implement irrevocable control mechanisms over the decisions process of the company's management as guarantee that the invested capital is used for its original goals. Attendance in the management board of the company will be just one of these mechanisms.

Generally the management processes of the companies will be reviewed and if needed adjusted. We would prefer to invest in companies that have already existing or are willing to implement modern business and management processes, which are detached and independent from individual talent skills and single persons authority. The latter is unfortunately still the business standard for the majority of SMEs in Bulgaria, and bares a high potential business risk in the cases of disloyalty of this key people.

As the fund will be investing in growth, the equity investments as a general rule will be done as a capital increase and not as a partial or full shareholders exit. Exceptions to this rule can be evaluated if one or some of the shareholders hinder the development of the company.

Given our experience and the targeted industries the ideal investment sizes are between EUR 1.5 million (smallest investment, as described in the example cases) and EUR 8 million (large). This numbers show the initial investment size. For follow-up capital increases the fund will keep special reserves of 10% to 15% of the total fund capital. Ideally, capital injections will be scheduled in tranches tied to performance and/or investment cornerstones.

The general holding period of an investment will be around 5 years, depending on the industry, life cycle of the company and the general economic cycle. Overall fund target will be an IRR of 18%. Some of the companies will need to be prepared for acquisition by international buyers due to the natural limitation of the local market. Such companies need to have grown to a size and stage that will make such acquisitions possible.

Following additional investment rules will apply to cover the principles described above:

- A single investment should not exceed EUR 10 million, and if it does, then a decision of the supervisory board will be needed. Single investments below EUR 1.5 million will be not evaluated.
- To assure diversification of companies, Top 4 investments should not exceed EUR 30 million
- To assure diversification in the targeted industries, the limit per single industry will be 30% of the funds capital.
- A balance (50/50) between the two types of investment will be targeted

In the initial year the aim is to complete 3 deals from different industries and different investment types. The investment cases show a generalized summary of some of the existing projects/ deals in our pipeline. In the following years 3-4 deals per year will be the target.

The fund will utilize a two-pronged approach for deal flow generation. The origination will benefit from the commitment of the Bank, a commercial bank specialized in SME financing, and the partners' extensive network and experience in Bulgaria's financial market. The partners' combined network covers more than 50% of the Bulgarian SMEs in the targeted size and industries.

The fund will strive to maximize value creation by following policies providing in-direct and direct support to capital beneficiaries. Taking into consideration the business environment in the country and the usual practices of SMEs, the fund management is convinced that the potential for value-creation is immense and should be addressed adequately. As already several times described in this document even well-run Bulgarian companies often suffer from inefficient management, un-timely access to performance metrics, no monitoring procedures, and lack of access (or unwillingness to rely) to third-party advisory/consulting services. Post-investment strategies for value creation will also be designed as to address the mentioned risks of fraudulent behaviour, and to handle resistance for cooperation.

5 CONCLUSION

In this paper, it has been proven that given the development stage and nature of the SMEs in Bulgaria the most suitable instruments created by Funds management have to be as plain and simple as possible. Sophisticated financial products generally create mistrust on the local market. Thus each Fund must intend to use for its investment needs primarily direct participation in the companies via investing in common stock and in certain cases trough a combination with investments in preferred stock of the company.

Structured Funds under JEREMIE most likely will aim at purchasing a significant portion of a particular company in order to be able to have a larger influence in its governing and to speed up its growth via the experience and know-how of its investment team. Typically Funds will seek to participate via a capital increase aiming at further strengthening the shareholder's equity, and support the continued growth through acquisitions as well as organic growth.

In order to protect its investment each Fund might seek also participation trough preferred stock as it has many advantages including a greater claim of the assets than common stock thus limiting the downside of the investment. Buying preferred stock could include the option of converting them into common stock at any point of time, in which case the owners will lose the right of a dividend, but will gain the ability to participate in the decision making process. Preferred stocks could be flexible in terms of the dividend rates that they hold, which could be adjusted along the way so that it does not interfere with the company's sustainable growth.

In limited number of cases each Fund have to aim at lending different types of hybrid loan products, suited to best fit the business needs of each company. A common type of debt product that Funds will be looking at will be the convertible debt, where the loan is secured via the right to convert it to common stocks at certain predetermined conditions. This will reduce both the risk to each Fund and the requirement to the company to provide collateral, which as we have mentioned before proves to be a major obstacle for the SMEs on their way to receiving a proper financing.

ACKNOWLEDGEMENTS

The research work reported in the paper is partly supported by the project AComIn "Advanced Computing for Innovation", grant 316087, funded by the FP7 Capacity Programme (Research Potential of Convergence Regions), and partly supported by the European Social Fund and Republic of Bulgaria, Operational Programme "Development of Human Resources" 2007-2013, Grant № BG051PO001-3.3.06-0048.

REFERENCES

Bulgarian National Bank (BNB) - www.bnb.bg

Bulgarian Small and Medium Enterprise Promotion Agency – www.sme.government.bg

http://www.Investor.bg

http://www.ncsu.edu/nrli/decision-making/MCDA.php

http://www.oxfordeconomics.com

Xu D. L., J. B. Yang, 2003. Intelligent decision system for self-assessment, *Journal of Multiple Criteria Decision Analysis*, Vol.12, pp. 43-60, 2003.

SIGNIFICANCE OF THE ADVANCED CONTROL AND OPTIMISATION FOR SMEs

Vassil G. Nikov, Lyubka A. Doukovska

Institute of Information and Communication Technologies
Bulgarian Academy of Sciences
Acad. G. Bonchev str., bl. 2, 1113 Sofia, Bulgaria
vasilnikov@abv.bg

Keywords: Advanced process control (APC), Advanced Integrated Maintenance Management System (AIMMS), Small and medium-sized enterprises (SMEs).

Abstract: The Advanced Process Control (APC) refers to a broad range of techniques and technologies implemented within industrial process control systems. Advanced process controls are usually deployed optionally and in addition to basic process controls. Basic process controls are designed and built with the process itself, to facilitate basic operation, control and automation requirements. In this paper is discussed the efficiency of the APC implementation within SMEs. Based upon our analysis a methodology for an integrated management system for predictive maintenance or the Advanced Integrated Maintenance Management System (AIMMS) is presented. The results presented in this paper show that AIMMS supports strategic decisions for predictive maintenance and it helps increase the equipment effectiveness by prioritizing the criticality of the equipment and increasing profits based on the advanced process control.

1 INTRODUCTION

Advanced process control (APC) is the application of a broad range of disciplines and technologies to either refine process control or address a specific issue associated with a process. Interdisciplinary in nature, APC draws knowledge in its application from statistics, decision theory, engineering, signal processing and artificial intelligence. Most often it is leveraged to solve process control issues that have multiple variables or problems that are detached from the control process. To accomplish its intended objectives, an APC technology is usually linked with the distributed control system (DCS) that controls the process. Due to its inherent tendencies to solve issues that have multiple variables, advanced process control is used in a variety of industries.

In control theory Advanced process control (APC) is a broad term composed of different kinds of process control tools and methodologies, often used for solving multivariable control problems or discrete control problem. Advanced control describes a practice which draws elements from many disciplines ranging from control engineering, signal processing, statistics, decision theory and artificial intelligence.

Traditionally, the key benefits from the use of advanced process control are improved bottom line returns:

- Efficiency gains (yield increase, throughput increase, reduction in energy and raw material per unit of product);
- Quality gains (consistent product quality with minimum production costs)
- Agility gains (manufacturing flexibility, customer responsiveness, need to reduce working capital by processing to order not to stock).
- Another source of benefits is achieved during the implementation of APC application:
- A systematic check of instrumentation and analyzers and resolution of bad actors.
- During testing phase, key process parameters are moved from nominal values that might lead to more beneficial operating conditions.
- Synchronize work of shifts.

These benefits might be several times bigger than those obtained through reduction of controlled variables variability. The effective maintenance strategy includes preventive maintenance, root-cause failure analysis, optimized inspection frequencies, more disciplined planning, better spare part management etc.

In world practice is increasingly accepted that predictive maintenance can play a key role in the long-term profitability of a company in the manufacturing sector with a major impact on timely delivery, product quality and its ultimate cost. The importance of maintenance increases in terms of increasing both the productivity and also the quality requirements which can only be achieved with a well-developed and organized maintenance strategy.

In this sense, according to recent advances in technology there have been developed and tested many methodologies, tools, techniques and strategies. This paper focuses on identifying difficulties for the implementation of APC within SMEs.

In recent years there has been increasing interest in the operation and management of industrial maintenance in a number of organizations. This is due to the increasing pressure on manufacturing organizations to meet customer and corporate requirements; the available equipment and productivity are central to achieving this goal.

According to the paper (Chan, 2005), "Recent trends show that on the whole many production systems do not function as expected when it comes to cost effectiveness in terms of their operation and maintenance. Lots of companies often operate with reduced capacities and reduced productivity, while the prices of their products are high." The number of modern maintenance practices for technical support is designed to allow organizations to target strategic resources to achieve the maintenance tasks that are considered crucial for the effective and efficient operation of the equipment. A number of organizations announce improvements in existing equipment, reliability and reduction in the maintenance costs after the implementation of APC, the part of presented results are published in (Blanchard, 1997; Cooke, 2000). It is assumed that APC is used primarily because it integrates the functions of production and maintenance, but more importantly is that it redefines the role of operators and support engineers.

The goal is to increase the efficiency of the equipment as well as to maximize the volume of production from this equipment. This is the result of an effort to achieve and maintain optimal conditions for the equipment in order to prevent unexpected failures, the speed decrease and the qualitative defects during the manufacturing process according to (Bamber, 1999).

2 TYPES OF ADVANCED PROCESS CONTROL SYSTEMS

In this paper following is presented a list of the best known types of advanced process control:

- Advanced regulatory control (ARC) refers to several proven advanced control techniques, such as feed forward, override or adaptive gain. ARC is also a catch-all term used to refer to any customized or non-simple technique that does not fall into any other category. ARCs are typically implemented using function blocks or custom programming capabilities at the DCS level. In some cases, ARCs reside at the supervisory control computer level.
- Multivariable Model predictive control (MPC) is a popular technology, usually deployed on a supervisory control computer that identifies important independent and dependent process variables and the dynamic relationships (models) between them, and uses matrix-math based control and optimization algorithms, to control multiple variables simultaneously. MPC has been a prominent part of APC ever since supervisory computers first brought the necessary computational capabilities to control systems in the 1980s.
- Inferential control: The concept behind inferential is to calculate a stream property from readily available process measurements, such as temperature and pressure, that otherwise would require either an expensive and complicated online analyzer or periodic laboratory analysis. Inferential can be utilized in place of actual online analyzers, whether for operator information, cascaded to base-layer process controllers, or multivariable controller CVs.
- Sequential control refers to discontinuous time and event based automation sequences that occur within continuous processes. These may be implemented as a collection of time and logic function blocks, a custom algorithm, or using a formalized Sequential function chart methodology.
- Compressor control typically includes compressor anti-surge and performance control.

3 APC BUSINESS APPLICATIONS

Those responsible for the design, implementation and maintenance of APC applications are often referred to as APC Engineers or Control Application Engineers. Usually their education is dependent upon the field of specialization. For example, in the process industries many APC Engineers have a chemical engineering background, combining process control and chemical processing expertise.

Most large operating facilities, such as oil refineries, employ a number of control system specialists and professionals, ranging from field instrumentation, regulatory control system (DCS and PLC), advanced process control, and control system network and security. Depending on facility size and circumstances, these personnel may have responsibilities across multiple areas, or be dedicated to each area. There are also many process control service companies that can be hired for support and services in each area.

Since 1996, the management of industrial sites is aware of the need of implementing systems to support all business processes (Dochain, 2008).

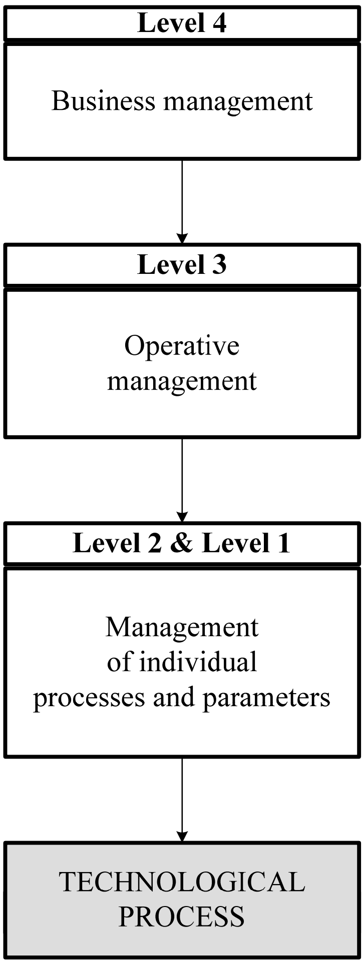


Figure 1: Production processes

In this paper is presented an international standard for the integration of business systems and control systems for the production processes shown in Figure 1.

This structure is operative and it includes next four hierarchical levels.

Level 1 and Level 2 are functionally grouped and comprise the management of individual devices and parameters.

Level 3 comprises all tasks of operative management. The timeline of Level 3 are days, hours and minutes.

Level 4 is the level of business planning and logistics.

4 THE BEST PRACTICES

In every industrial enterprise there are potential opportunities for improvement, which are in the range 1–10% by economy of energy resources and raw materials of 3–5%, 2–7% for increased production (http://www.automation.siemens.com, http://www.honeywellprocess.com/en-us). Doing good business and in particular industrial management becomes more and more expensive.

The best approach is the old control systems to continue to work, by adding new technologies of control (Figure 2).

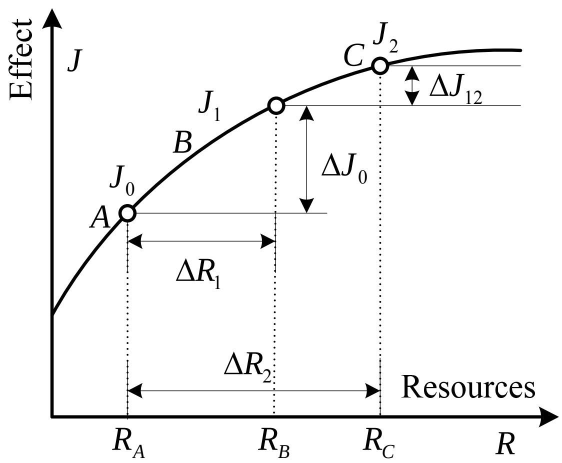


Figure 2: Process control effectiveness

Only a holistic approach can be effective. Information integration basically influences its realization.

5 CONCLUSIONS

Process control and management of industrial complexes is gaining acceptance as a functional problem with multiple criteria the decision of which is possible only via a holistic approach.

The methods of artificial intelligence (neural networks, fuzzy logic and genetic algorithms) and those based on knowledge (expert systems, autonomous agents, CBR) and especially the hybrid systems are still used only occasionally, but they possess a significant potential.

Generally modern process control systems are developed successfully to increase the competitiveness of industrial complexes in the globalized world.

The successful implementation of an approach to predictive maintenance taking into account the nature of activities in SMEs, it must be based on three main objectives. First, to allow the user to formulate a strategy of maintenance. Second, to convert the limited available data into knowledge to develop a strategy for maintenance. Third, to allow the user to record and measure the effect of the new strategy for support, to ensure that future decisions are based on facts and accurate data. The successful implementation of the formed in this way model would lead to the possibility the structure and the elements of the model to create efficient and effective strategy to support SMEs.

Progress is possible if it is use an advanced control system and a simple system for recording and analysis that can be easily accessed and updated. The present research shows that technology is able to solve emerging needs but it requires innovative adaptations to solve the existing problems.

ACKNOWLEDGEMENTS

The research work reported in the paper is partly supported by the project AComIn "Advanced Computing for Innovation", grant 316087, funded by the FP7 Capacity Programme (Research Potential of Convergence Regions), partially supported by the European Social Fund and Republic of Bulgaria, Operational Programme "Development of Human Resources" 2007-2013, Grant № BG051PO001-3.3.06-0048.

REFERENCES

http://www.wib.nl

http://www.automation.siemens.com

http://www.honeywellprocess.com/en-us

Bamber C. J., J. M. Sharp., M. T. Hides, 1999. Factors affecting successful implementation of TPM: A case study perspective. In *Journal of Quality in Maintenance Engineering* 5 (3), 162–181.

Blanchard B., 1997. An enhanced approach for implementing TPM in the manufacturing environment, In *Journal of Quality in Maintenance Engineering* 3 (2), 69-80.

Chan F., F. Lau, S. Chan, S. Kong, 2005. Implementation of Total Productive Maintenance: A case study. In *International Journal of Production Economics* 95(1), 71–94.

Cooke F. L., 2000. Implementing TPM in plant maintenance: Some organizational barriers. In *International Journal of Quality & Reliability Management* 17 (9), 1003–1016.

Dochain D., W. Marquardt, S. Won, O. Malik, M. Kinnaert, J. Lunze, 2008. Monitoring and Control of Process and Power Systems: Adapting to Environmental Challenges, Increasing Competitivity and Changing Customer and Consumer Demands, In *IFAC World Congress*, Seoul, Korea, 7160–7171.

FORECASTING FINANCIAL MARKETS WITH ARTIFICIAL INTELLIGENCE

Veselin L. Shahpazov, Lyubka A. Doukovska

Institute of Information and Communication Technologies
Bulgarian Academy of Sciences
Acad. G. Bonchev str., bl. 2, 1113 Sofia, Bulgaria
veselin_georgiev@abv.bg

Keywords: Artificial Intelligence, Neural Networks, Financial Market, Stock Prices, Credit risk, Intelligent Systems

Abstract: Interest in using artificial intelligence techniques for forecasting has led to a tremendous surge in research activities over time. In this paper is assumed that Artificial Neural Networks (ANNs) are flexible computing frameworks and universal approximations that can be applied to a wide range of time series forecasting problems with a high degree of accuracy. They are an artificial intelligence method for modelling complex target functions. Forecasting problems arise in so many different disciplines and the literature on forecasting using artificial intelligence procedures is scattered in so many diverse fields that it is hard for a researcher to be aware of all the work done to date in the area.

1 INTRODUCTION

One characteristic that all stock markets have in common is the uncertainty, which is related with their short and long-term future state. This feature is undesirable for the investor but it is also unavoidable whenever the stock market is selected as the investment tool. The best that one can do is to try to reduce this uncertainty. Stock Market Prediction (or Forecasting) is one of the instruments in this process.

The stock market prediction task divides researchers and academics into two groups those who believe that we can devise mechanisms to predict the market and those who believe that the market is efficient and whenever new information comes up the market absorbs it by correcting itself, thus there is no space for prediction (EMH). Furthermore they believe that the Stock Market follows a Random Walk, which implies that the best prediction you can have about tomorrow's value is today's value.

In literature a number of different methods have been applied in order to predict Stock Market returns. These methods can be grouped in four major categories: 1) Technical Analysis Methods, 2) Fundamental Analysis Methods, 3) Traditional Time Series Forecasting and 4) Machine Learning Methods. Technical analysts, known as chartists, attempt to predict the market by tracing patterns that come from the study of charts which describe historic data of the market. Fundamental analysts study the intrinsic value of a stock and they invest on it if they estimate that its current value is lower that its intrinsic value. In Traditional Time Series forecasting an attempt to create linear prediction models to trace patterns in historic data takes place. These linear models are divided in two categories: the univariate and the multivariate regression models, depending on whether they use one of more variables to approximate the stock market time series. Finally a number of methods have been developed under the common label Machine Learning these methods use a set of samples and try to trace patterns in it (linear or non-linear) in order to approximate the underlying function that generated the data. The aim is to draw conclusions from these samples in such way that when unseen data are presented to a model it is possible to infer the to-be explained variable from these data. These methods have been applied to market prediction; particularly for Neural Networks there is a rich literature related to the forecast of the market on daily basis.

Forecasting problems arise in so many different disciplines and the literature on forecasting using ANNs is scattered in so many diverse fields that it is hard for a researcher to be aware of all the work done to date in the area. One of the first successful applications of ANNs in forecasting is reported by Lapedes and Farber. Using two deterministic chaotic time series generated by the logistic map and the Glass Mackey equation, they designed the feed-forward neural networks that can accurately mimic and predict such dynamic nonlinear systems. Their results show that ANNs can be used for modelling and forecasting nonlinear time series with very high accuracy (Deppisch *et al.*, 1991). After that a number of papers were devoted to using ANNs to analyse and predict deterministic chaotic time series with and/or without noise. Chaotic time series occur mostly in engineering and physical science since most physical phenomena are generated by nonlinear chaotic systems. As a result, many authors in the chaotic time series modelling and forecasting are from the field of physics. Another major application of neural network forecasting is in electric load consumption study.

Many other forecasting problems have been solved by ANNs. A short list includes airborne pollen commodity prices, environmental temperature; helicopter component loads international airline passenger traffic macroeconomic indices, ozone level, personnel inventory, rainfall, river flow, student grade point averages, tool life, total industrial production, transportation, and wind pressure and wind pressure and many more.

There is an extensive literature in financial applications of ANNs. They have been used for forecasting bankruptcy and business failure, foreign exchange rate, stock prices off course.

2 ARTIFICIAL NEURAL NETWORKS FOR STOCK MARKET FORECASTING

The idea of using neural networks for predicting problems was first expressed by (Hu, 1964) who used it for weather forecasting, the conclusion made in his paper was: "...adaptive systems have the capability of making useful prediction or specifications of weather without complete understanding of the dynamics or complete measurement of the physical parameters involved." The absence of any learning method for multi-layer networks made it impossible to apply these networks to complex prediction problems. But in 1980's the back-propagation algorithm was introduced for training an MLP neural network. Werbos used this technique to train a neural network (Werbos, 1988) and claimed that neural networks are better than regression methods and Box-Jenkins model in prediction problems.

In the years that followed, many researchers have been done on neural networks to predict the financial markets and the stock market changes in particular. Probably the first paper in the vast field, at least from today's perspective, of stock market prediction with ANN's is (White, 1988), in that paper the main focus is to try and prove wrong the then prevailing Efficient Market Hypothesis. In its simplest form, this hypothesis asserts that asset prices follow a random walk that is, the movement of an asset's price is completely unpredictable from publicly available information such as the price and volume history for the asset itself or that of any other asset. The purpose of the paper was to illustrate how the search for such regularities using neural network methods might proceed, using the case of IBM daily common stock returns as an example. The necessity of dealing with the salient features of economic time series highlights the role to be played by methods of statistical inference and also requires modifications of neural network learning methods which may prove useful in general contexts. The target variable of interest in the study is the one day rate of return to holding IBM common stock. Of the available 5000 days of returns data, a sample was selected consisting of 1000 days for training purposes, together with samples of 500 days before and after the training period which are used for evaluating whatever knowledge the networks have acquired. The training sample covers trading days during the period 1974 through 1978. The evaluation periods cover 1972 through 1974 and 1978 through 1980.

The network architecture used is the standard single hidden layer architecture, with inputs, passed to a hidden layer (with full interconnections) and then with hidden layer activations passed to the output unit. The author trained a three layer feed-forward network with the same five inputs and five hidden units over the same training period. The choice of five hidden units, it represents a compromise between the necessities to include enough hidden units so that at least simple nonlinear regularities can be detected by the network. Training was done with the back-propagation (one of the most popular algorithms to date). In order to claim that the simple Efficient Markets Hypothesis has been statistically refuted, White conducted an out-of-sample forecasting experiment. This exercise was carried out for a post sample period of 500 days, and a pre-sample period of 500 days. For the post-sample period the correlation observed was of – 0.699 for the pre-sample period, it was 0.0751.

The conclusion made was: "Such results do not constitute convincing statistical evidence against the efficient markets hypothesis. The in-sample (training period) results are now seen to be over-optimistic, being either the result of over-fitting (random fluctuations recognized incorrectly as nonlinearities) or of learning evanescent features (features which are indeed present during the training period, but which subsequently disappear). In either case the implication is the same: the present neural network is not a money machine." The objective for future work stated, was to try and elaborate the network to allow additional inputs such as traded volume, other stock prices and volume, leading indicators, macroeconomic data, etc. and by permitting recurrent connections. Another important conclusion made in the paper was the fact that", an important limitation of the present exercise is that the optimization methods used here are essentially local. Although the final weight values were determined as giving the best performance over a range of different starting values for our iterations, there is no guarantee that a global minimum was found. A global optimization method such as simulated annealing or the genetic algorithm would be preferable." The study did not find evidence against the EMH because the network used was a single-layer feed-forward network, which did not have a lot of generalization power.

One of the first efforts to predict a broad market index was by Kimoto and his colleagues in which they used neural networks to predict the index of Tokyo stock market (Kimoto *et al.*, 1990). They used several neural networks trained to learn the relationships between past values of various technical and economic indices for obtaining the expected returns of the TOPIX. The TOPIX is a weighted average of all stocks listed on the Tokyo Stock Exchange. The used technical and economic indices were: the vector curve (an indicator of market momentum), turnover, interest rate, foreign exchange rate and the value of the DJIA (Dow Jones Industrial Average). The desired output of the networks is a weighted sum, few weeks, of the logarithm of the ratio of the TOPIX at the end of week t to the TOPIX value at the end of week $(t - 1)$ and the desired output is a weighted sum of r_t for some weeks. The future extraction is not explained in this paper, except for the fact that some irregularity is removed and logarithm function is used before normalization. The authors claim that the use of weighted sum of the outputs of many neural networks reduces the error, especially since the returns are predicted for a few weeks. The buy/sell system is setup based on the predicted returns and this system is shown to perform much better than s buy and hold strategy. However the training data used future returns, so that this method cannot be used for actual stock trading, fact that is outlined by the authors.

In the paper by (Kamijo and Tanigawa, 1990) the use of "Elman recurrent net" is proposed (recurrent neural network is class of neural network, where connections between units form a directed cycle. This creates an internal state of the network which allows it to exhibit dynamic temporal behaviour) for predicting the future stock prices using extracted features from past daily high, low and closing stock prices. Unlike feed-forward neural networks, RNNs can use their internal memory to process arbitrary sequences of inputs. The method used tries to extract triangle patterns in stock prices which are seen on the daily high, low and closing graph. A triangle is usually seen as a beginning of a sudden stock price rise after that the high and low prices appear and the price oscillates for a period of time before the lines converge. The ANN is trained to recognize this pattern in the stock prices.

Matsuba uses a feed-forward NN with the last n stock index values as inputs and the next $N - n$ values as the outputs (Matsuba, 1991). This is an $N - n$ step ahead prediction. Thus, if the index for the $n^{\text{-th}}$ day is denoted by X_n then, the inputs are X_1, X_2, …, X_n and the outputs are X_{n+1}, X_{n+2}, …, X_N. If such a network is trained, any correlation between the index values for $n + 1$ through $N^{\text{-th}}$ day will be neglected. To ensure that this does not happen, the network is trained with errors between the desired and actual outputs in addition to the n inputs. These errors will then be $X_{n-1} - Y_{n+1}$, …, where Y is the output of the network. As the training proceeds this error will tend to zero and these additional inputs are not required in the testing phase. This work also uses two neural networks, one to learn the global features and another to learn the local features or small fluctuations.

In his work Freisleben used a simple feed-forward neural network trained with past and present data to predict the value of the FAZ Index (Freisleben, 1992). Input data included the moving average of past 5 and 10 weeks of the FAZ Index a first order difference of the FAZ Index and its moving average, the present bond market index and its first order difference and the Dollar-Mark exchange rate along with its first order difference. The value of the FAZ Index was predicted for the next week based on this data. The neural network was trained for the past M weeks and is then tested based on data for the next L weeks, where M is the training window and L is the testing window. For every successive prediction, the windows were shifted ahead and the network is retrained. Three different networks are compared each having a different set of inputs, one of which had only the last 10 values of the FAZ Index as an input. The observations point that the network fed with data from the technical indicators performed better than the one trained with only the past values of the index. The normalization of the training data was done in order to keep the data within the 0.1 – 0.9 range, however no specification on the normalization method is given.

Azoff in (Azoff, 1994) outlined that networks are computer programs that can recognize patterns in data, learn from this and make forecasts of future patterns. At the time, there were just over 20 commercially available neural network programs designed for use on financial markets and there have been some notable reports of their successful application. However, like any other computer program, neural networks are only as good as the data they are given and the questions that are asked of them. Proper use of a neural network involves spending time understanding and cleaning the data: removing errors, pre-processing and post-processing. His book provides the knowledge that is required for the proper design and use of ANN's in financial markets forecasting – with an emphasis on futures trading.

In their study (Kaastra, Boyd, 1995) provided a practical, non-technical introduction to designing a neural network forecasting model using economic time series data. The procedure of designing a model is divided into eight steps: 1) variable selection; 2) data collection; 3) data pre-processing; 4) training, testing and validation sets; 5) neural network paradigms; 6) evaluation criteria; 7) neural network training; and 8) implementation. Three major conclusions are made, the first being that researchers must have the time resources and patience to experiment mainly because of the nature financial markets. The second is that NN software must allow automated routines such as walk-forward testing, optimization of hidden neurons and testing of input variable combinations, either through direct programming or the use of batch or script files. And third: the researcher must maintain a good set of records that list all parameters for each network tested since any parameter may turn out to cause a significant change in neural network performance.

In paper (Leung *et al.*, 2000) conducted a research which focuses on estimating the level of return on different stock market index. Given the notion that a prediction with little forecast error does not necessarily translate into capital gain, they evaluate the efficacy of several multivariate classification techniques relative to a group of level estimation approaches. Among the level estimation counterparts, which forecast the level, are exponenttial smoothing are the multilayered feed-forward neural network and a probabilistic neural network. The data used covers the period between 1967 and 1995 and includes the daily prices of S&P 500 FTSE 100 and Nikkei 225 as well as some macroeconomic indicators like CPI and industrial Production. The comparison of the predictive strength of classification and level estimation models shows that the best performance in correctly predicting the direction of the index return over the 60 out-of-sample forecasting periods from January 1991 through December 1995 is that of the Multi-layered feed-forward neural network and Probabilistic neural network. To provide a comparison with the PNN classifier, they test the performance of the multi-layered feed-forward neural network. Unlike the PNN which suggests a group classification for a given set of inputs, MLFN provides a point estimate or forecast as the output. They also refer to a description and comparison of the architectures and mathematical foundation of PNN and MLFN by Wasserman in 1993 paper. Based on the results from their study, they select the network architecture which leads to consistent and reasonable performance in the validation sample.

With the introduction of electronic communication networks (ECN) as electronic trading systems facilitating trading of stocks and other financial products in the world's leading stock exchanges at first and later on other non-mainstream stock markets, and the constantly growing interest by both retail and institutional investors all around the world in stock's investing, the research in this field exploded. The advancement in computational and communicational power allowed researchers to develop models using artificial neural networks that are fed with real time data and capable to produce real time buy and sell signals.

In (Pan *et al.*, 2005) a computational approach is presented for predicting the Australian stock market index – AORD using multi-layer feed-forward neural networks from the time series data of AORD and various interrelated markets. This effort aims to discover an effective neural network or a set of adaptive neural networks for this prediction purpose, which can exploit or model various dynamical swings and inter-market influences discovered from professional technical analysis and quantitative analysis. Three aspects of effectiveness on data selection are considered: effective inputs from the target market (AORD) itself, a sufficient set of interrelated markets, and effective inputs from the interrelated markets. Two traditional dimensions of the neural network architecture are also considered: the optimal number of hidden layers, and the optimal number of hidden neurons for each hidden layer. Three important results were obtained: A 6-day cycle was discovered in the Australian stock market during the studied period; the time signature used as additional inputs provides useful information; and a basic neural network using six daily returns of AORD and one daily returns of S&P500 plus the day of the week as inputs exhibits up to 80 % directional prediction correctness.

Kalyvas attempts to predict the daily excess returns of FTSE 500 and S&P 500 indices over the respective Treasury bill rate returns (Kalyvas, 2001). Then the author applies two different types of prediction models: autoregressive (AR) and feed-forward neural networks to predict the excess returns time series using lagged values. For the NN models a Genetic Algorithm (GA) is constructed in order to choose the optimum topology. Data consists of 3275 daily observations of FTSE-100 index, UK T-Bill Rates and 3277 observations of S&P-500 index and US T-bill rates from 4 Jan 1988 until 12 Dec 2000. Finally he evaluates the prediction models on four different metrics and concludes that they do not manage to outperform significantly the prediction abilities of naive predictors.

In their study (Chen *et al.*, 2003) attempt to model and predict the direction of market index of the Taiwan Stock Exchange, one of the fastest growing financial exchanges in the developing Asian countries (considered an emerging market). The probabilistic neural network (PNN) is used to forecast the direction of index return after it is trained by historical data. Statistical performance of the PNN forecasts are measured and compared with that of the generalized methods of moments (GMM) with Kalman filter. Moreover, the forecasts are applied to various index trading strategies, of which the performances are compared with those generated by the buy-and-hold strategy as well as the investment strategies guided by forecasts estimated by the random walk model and the parametric GMM models. They conclude that empirical results show that the PNN-based investment strategies obtain higher returns than other investment strategies examined in this study.

In (Versace *et al.*, 2004) the performance of a heterogeneous mixture of neural network algorithms for predicting the exchange-traded fund DJIA is evaluated. A genetic algorithm is utilized to find the best mixture of neural networks, the topology of individual networks in the ensemble, and to determine the features set. The genetic algorithm also determines the window size of the input time-series supplied to the individual classifiers in the mixture of experts. The mixtures of neural network experts consist of recurrent back-propagation networks, and radial basis function networks. A list of the variables employed in this study is comprised of data from several Dow sub-indexes and other well-known indexes, foreign exchange rates commodities and benchmark bonds from 11th November 2001 through 12th February 2003 or 320 total trading days. The conclusion made is that artificial neural networks and genetic algorithms performed relatively well and promise as a good technique for forecasting stochastic time series like those seen in stock market data.

In paper (Kim, Lee, 2004) compared a feature transformation method using genetic algorithm with two conventional methods for artificial neural networks. The genetic algorithm is incorporated to improve the learning and generalization abilities of ANN's for stock market prediction. Daily predictions are conducted and their accuracy is measured. The authors use the proposed model to predict South Korea composite stock price index (KOSPI). The comparison of the results achieved by a feature transformation method using a genetic algorithm to other feature transformation methods shows that the proposed model performs better. Experimental results show that the proposed model reduces the dimensionality of the feature space and decreases irrelevant factors for stock market predictions.

In work (Kim, 2006) a genetic algorithm approach to instance selection in artificial neural networks for financial data mining is proposed. He notes that artificial neural networks have preeminent learning ability, but often exhibit inconsistent and unpredictable performance for noisy data. In addition, it may not be possible to train ANN's or the training task cannot be effectively carried out without data reduction when the amount of data is so large. The proposed model uses a genetic algorithm to optimize simultaneously the connection weights between layers and a selection task for relevant instances. The globally evolved weights mitigate the well-known limitations of gradient descent algorithm. In addition, genetically selected instances shorten the learning time and enhance prediction performance. Experimental results show that the genetic algorithm approach is a promising method for instance selection in ANN for models for stock market analysis.

In paper (Madden, O'Connor, 2006) evaluated the effectiveness of using external indicators, such as commodity prices and currency exchange rates, in predicting movements in the Dow Jones Industrial Average index. The performance of each technique is evaluated using different domain-specific metrics. A comprehensive evaluation procedure is described, involving the use of trading simulations to assess the practical value of predictive models, and comparison with simple benchmarks that respond to underlying market growth. In the experiments presented, basing trading decisions on a neural network trained on a range of external indicators resulted in a return on investment of 23.5% per annum, during a period when the DJIA index grew by just 13.03% per annum. A substantial dataset has been compiled and is available to other researchers interested in analysing financial time series.

In paper (Tilakaratne *et al.*, 2007) is presented to explore the inter-market connections and relations. The study forecasts trading signals of the Australian All Ordinary Index (AORD), one day ahead. These forecasts were based on the current day's relative return of the Close price of the US S&P 500 Index, the UK FTSE 100 Index, French CAC 40 Index and German DAX Index as well as the AORD. The forecasting techniques examined were feed-forward and probabilistic neural networks. Performance of the networks was evaluated by using classification/ misclassification rate and trading simulations. For both evaluation criteria, feed-forward neural networks performed better. Trading simulations suggested that the predicted trading signals are useful for short term traders.

In (Maciel, Ballini, 2009) is analysed the neural networks for financial time series forecasting. Specifically the ability to predict future trends of North American, European and Brazilian Stock Markets. Accuracy is compared against a traditional forecasting method, generalized autoregressive conditional heteroscedasticity (GARCH). They also make a very extensive review of the work done in the field, while only briefly discussing neural network theory this research determined the feasibility and practicality of using neural networks as a forecasting tool for the individual investor. It was concluded that neural networks do have a powerful capacity to forecast all stock market indexes studied and, if properly trained, the individual investor could benefit from the use of this forecasting tool against current techniques.

In paper (Hadavandi *et al.*, 2010) claim that the main idea to successful stock market prediction is achieving best results using minimum required input data and the least complex stock market model. To achieve these purposes their article presents an integrated approach based on genetic fuzzy systems and artificial neural networks for constructing a stock price forecasting expert system. They first use stepwise regression analysis to determine factors which have most influence on stock prices. At the next stage they divide the raw data into k clusters by means of self-organizing map neural networks. Finally, all clusters are fed into independent genetic fuzzy system models with the ability of rule base extraction and data base tuning. Evaluation of the capability of the proposed approach I done by applying it on stock price data gathered from IT and Airlines sectors, and compare the outcomes with previous stock price forecasting methods using mean absolute percentage error. Results show that the proposed approach outperforms all previous methods, so it can be considered as a suitable tool for stock price forecasting problems.

In (Gosh, 2012), a hybrid neural-evolutionary methodology for forecast time-series data is presented and prediction of the NASDAQ stock price in particular. The methodology is hybrid because an evolutionary computation-based optimization process is used to produce a complete design of a neural network. The produced neural network, as a model, is then used to forecast the time-series (Lowe, Webb, 1990). The model identification process involves data manipulation and a highly experienced statistician to do the work. Compared to previous work, this paper approach is purely evolutionary, while others use mixed, mainly combined with back-propagation, which is known to get stuck in local optima. On the direction of model production, the evolutionary process automates the identification of input variables, allowing the user to avoid data pre-treatment and statistical analysis. The study proves the nimbleness of ANN as a predictive tool for Financial Time series Prediction. Furthermore, Conjugate Gradient Descent is proved to be an efficient Back-propagation algorithm that can be adopted to predict the average stock price of NASDAQ.

In (Chen, Du, 2013) the interactions between social media and financial markets are studied. The authors use a popular online Chinese stock forum Guba.com.cn as well as traditional sentimental analysis methods, for each stock, they build a Social Behaviour Graph based on human's online behaviour, calculate key characteristics of the graph, and find out the correlations between trading volume/price and those characteristics. They make use of a back-propagation neural network to predict the trading volume and price of stocks from the Shanghai/Shenzhen Stock Exchange in China. Their method has achieved better outcome compared to the traditional trading price based time series models.

In paper (Shahpazov, Velev, Doukovska, 2013), the problem of predicting the price of Bulgarian Stock Exchange's Sofix index using neural networks is considered. The analyzed period is of two years and two months or from 04.01.2011 until 08.03.2013. Data used for the case consists of the daily values of last price, open, high, low and volume traded 30 day moving average, 60 day moving average, 200 day moving average, the 14 day relative strength index and the 30 day relative strength index. The input data was preprocessed and transformed from values into daily changes. Initial readings showed that better results would be achieved if the input is one compared to using all or fragments of the initial data set. Smoothing ranging from 3 to 9 days was performed in order to eliminate the effects of the low liquidity and higher volatility in the market.

Results achieved in paper (Shahpazov, Velev, Doukovska, 2013) showed that this data manipulation managed to bring down the test error considerably. The produced neural network was structured by 1 input (20 time lagged steps) 13 nodes in the hidden layer and 1 output, training consisted of 100 epochs using back-propagation and 34 epochs using the conjugate gradient descent algorithm, the test error amounted to 0.057903. In comparison the best performing network using all or partial input data managed an error of 0.068935.

3 CONCLUSION

In the presented paper surveyed the application of neural networks to financial systems. It demonstrated how neural networks have been used to test the Efficient Market Hypothesis and how they outperform statistical and regression techniques in forecasting share prices. Although neural networks are not perfect in their prediction, they outperform all other methods and provide hope that one day we can more fully understand dynamic, chaotic systems such as the stock market.

Artificial Neural Networks are flexible computing frameworks and universal approximators that can be applied to a wide range of time series forecasting problems with a high degree of accuracy.

Predicting stock prices with traditional time series analysis has proven to be difficult. An artificial neural network may be more suitable for the task. Primarily because no assumption about a suitable mathematical model has to be made prior to forecasting. Furthermore, a neural network has the ability to extract useful information from large sets of data, which often is required for a satisfying description of a financial time series.

ACKNOWLEDGEMENTS

The research work reported in the paper is partly supported by the project AComIn "Advanced Computing for Innovation", grant 316087, funded by the FP7 Capacity Programme (Research Potential of Convergence Regions), partially supported by the European Social Fund and Republic of Bulgaria, Operational Programme "Development of Human Resources" 2007-2013, Grant № BG051PO001-3.3.06-0048.

REFERENCES

Azoff E. M., 1994. *Neural Network Time Series Forecasting of Financial Markets*. John Wiley&Sons.

Chen A. S., M. T. Leung, H. Daouk, 2003. Application of neural networks to an emerging financial market: Forecasting and trading the Taiwan Stock Index. *Computers & Operations Research*, Vol. 30, Issue 6, pp. 901–923.

Chen Z., X. Du, 2013. Study of Stock Prediction Based on Social Network, In: *Proceedings of 2013 International Conference on Social Computing*, September 8–14 Alexandria, VA, USA, pp. 913–916.

Deppisch J., Bauer H. U., Geisel T., 1991. Hierarchical training of neural networks and prediction of chaotic time series. *Physics Letters* 158, 57–62.

Freisleben B., 1992. Stock market prediction with back-propagation networks. In: *Proceedings of 5th Intl. Conf. on the industrial and engineering application of artificial intelligence and expert systems*, Germany, June 1992, pp. 451–460.

Tilakaratne C. D., S. A. Morris, M. A. Mammadov, C. P. Hurst, 2007. Predicting Stock Market Index Trading Signals Using Neural Networks, Centre for Informatics and Applied Optimization School of Information Technology and Mathematical Sciences University of Ballarat Victoria, Australia.

Gosh A., 2012. Comparative study of Financial Time Series Prediction By Artificial Neural Network with Gradient Descent Learning. *Brain*, Vol. 3 Issue 1, pp. 41–49.

Hadavandi E., A. Ghanbari, H. Shavandi, 2010. Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting Knowledge-Based Systems 23 (2010), pp. 800–808.

Hu M. J. C., 1964. *Application of the Adaline System to Weather Forecasting,* Master Thesis, Technical Report 6775-1, Stanford Electronic Laboratories, Stanford, CA.

Kaastra I., M. Boyd, 1995. Forecasting futures trading volume using neural networks *Journal of Futures Markets*, Vol. 15, Issue 8, pp. 953–970.

Kalyvas E., 2001. *Using Neural Networks And Genetic Algorithms to Predict Stock Market Returns*. Department of Computer Science, The University of Manchester, UK.

Kamijo K., T. Tanigawa, 1990. Stock market pattern recognition – A recurrent neural network approach *Proc. IJCNN*, San Diego, Vol. 1, pp. 215–221.

Kim K., 2006. Artificial neural networks with evolutionary instance selection for financial forecasting. *Expert Systems with Applications*, Vol. 30, Issue 3, pp. 519–526.

Kim W., B. Lee, 2004. Stock Market Prediction using artificial neural networks with optimal feature transformation. *Neural Comput & Applic*, Vol. 13, pp. 255–260.

Kimoto T., K. Asakawa, M. Yoda, M. Takeoka, 1990. Stock market prediction system with modular neural network. *Proceedings of the International Joint Conference on Neural Networks*, pp. 1–6.

Leung M. T., H. Daouk, A.-S. Chen, 2000. Forecasting stock indices: A comparison of classification and level estimation models. *International Journal of Forecasting*, Vol. 16, pp. 173–190.

Lowe D., Webb A. R., 1990. Time series prediction by adaptive networks: A dynamical systems perspective *IEEE proceedings-F138* (1), pp. 17–24.

Maciel L. S., R. Ballini, 2008. Design a neural network for time series financial forecasting: Accuracy and robustness analysis, Instituto de Economia, Universidade Estadual de Campinas, Sao Paulo-Brasil, 2008.

Madden M. G., N. O'Connor, 2006. Neural Network Approach to Predicting Stock Exchange Movements using External Factors. In: *Applications and Innovations in Intelligent Systems* XIII, pp. 64–77.

Matsuba I., 1991. Neural sequential associator and its application to stock price prediction *Proc. IECON 91*, Japan, Vol. 2, pp. 1476–1479.

Pan H., I. Tilakaratne, J. Yearwood, 2005. Predicting Australian Stock Market Index Using Neural Networks Exploiting Dynamical Swings and Intermarket Influences, *Journal of Research and Practice in Information Technology*, Vol. 37, No. 1, pp. 43–55.

Shahpazov V., V. Velev, L. Doukovska, 2013. Forecasting Price Movement of Sofix Index on the Bulgarian Stock Exchange – Sofia Using an Artificial Neural Network Model, *Proc. of the International Symposium on Business Modelling and Software Design – BMSD'13*, Noordwijkerhout, The Netherlands, ISBN 978-989-8565-56-3, 2013, pp. 298–303.

Versace M., R. Bhatt, O. Hinds, M. Shiffer, 2004. Predicting the exchange traded fund DIA with a combination of genetic algorithms and neural networks, *Expert Systems with Applications*, Vol. 27, pp. 417–425.

Werbos P. J. 1988. Generalization of back-propagation with application to a recurrent gas market model, *Neural Networks*, Vol. 1, 1988, pp. 339–356.

White H. 1988. Economic prediction using Neural Networks: The case of IBM daily stock returns. In: *Proc. of the IEEE International Conference on Neural Networks*, pp. 451–458.

MILL-FAN LOAD AND THE RESIDUAL-USEFUL-LIFE PREDICTION

Mincho Hadjiski, Kosta Boshnakov
University of Chemical Technology and Metallurgy - Sofia, 8 Kl. Ohridsky blvd., 1756 Sofia, Bulgaria
hadjiski@uctm.edu, kb@uctm.edu

Stefan Koynov
Dptmt. of Intelligent Systems, IICT-BAS, 1113 Sofia, Bulgaria
slk@iinf.bas.bg

Keywords: Condition Based Maintenance (CBM), Mill Fan, Residual Useful Life (RUL), trend analysis, vibration.

Abstract: A combination of multivariable vibration trend prediction with the current operational requirements for a Mill Fan (MF) in pulverizing system of power steam generator is considered. Using condition monitoring data provided from DCS and softsensing approach an optimal coal flow rate is determined in order to reach the predescribed RUL (Residual Useful Life) on the MF. Based on real experimental data a set of prediction models is derived usable in this stochastical case of machinery deterioration.

1 INTRODUCTION

Regardless of the problems that carbon energy creates worsening conditions for sustainable development of the environment, for many of world's leading economies (U.S., China, Russia, India) it retains a significant share for the next 50 years according to the national forecasts. Similarly for Bulgaria over 35% of electricity production is based on coal heating, mainly from local sources. Difficulties with burning lignite Bulgarian inferior lignite calorific 1300-1600 kcal / kg, up to 35% ash content and moisture – up to 50% are well known, but Bulgarian heating engineers for 40 years gained considerable experience in their combustion (Bata, 1995, Bonev, 1998). The dust-generating system determines the high efficiency of the power steam generators. Mill fans (MF) are most widely applied for grinding low-quality lignite according to national (Bata, 1995, Bonev, 1998) and worldwide (Klepikov, 2008, Tobaszewski, 2006) practice. Unlike dust-generating systems with quality coal, which use ball mills and rotor mills with sufficiently well developed mathematical models (Rees, 2003, Wei, 2009), there are no established models for mill fans, despite some successful private solutions (Hadjiski, 1995, Hadjiski, 2000, Tabaszewski, 2006).

In the current economic conditions of a complex balance between the components of the energy mix, the problem with the price of the generated electricity is particularly important. During the last two decades a considerable experience was accumulated using the possibilities not only for optimization of energy-transformation processes (Bata, 1995, Hadjiski, 1995, Rees, 2003), but also for maintenance of dust-generating facilities with a focus on the coal mills. The inevitable wear of the grinding surfaces set as particularly relevant the issues for monitoring the technical condition of the mill fan and also the maintenance decisions.

The paper treats the problem of rational load of MF based on the prediction of its technical condition according to the trend of the overall vibration of the bearing block as an indirect assessment of degradation and imbalance. Ideas on maintenance based on an assessment of the remaining lifetime resources developed in the last 15 years (Goode, 2000, Van, 2012, Zhang, 1997) are further developed in the direction of the combined assessment of Residual Useful Life (RUL) based on multivariate statistical analysis of the trend of overall vibration and also on the parametric influence over RUL via managing the MF workload in raw coal. Dynamic forecasting mathematical models are at the basis of the research.

Unlike a number of studies on the vibration analysis of rotating items (gears, bearings), wherein the primary focus is the isolation and identification of the failure mode (Dong, 2004, Yang, 2001, Zhang, 1997), in the present study the vibration state is predicted to eliminate the degree of degradation of the impeller (Hadjiski, 1995, Hadjiski, 2000, Khatab, 2012, Van, 2012) and also the operative actions to be taken in view of the current requirements for the steam generator.

2 FEATURES OF THE MILL FAN AS AN OBJECT OF PERFORMANCE DEGRADATION

Mill fans have exploitation resources at nominal load in the range of 2200-2500 operative hours.

Fig. 1 shows the graph of a mill fan of a power steam generator.

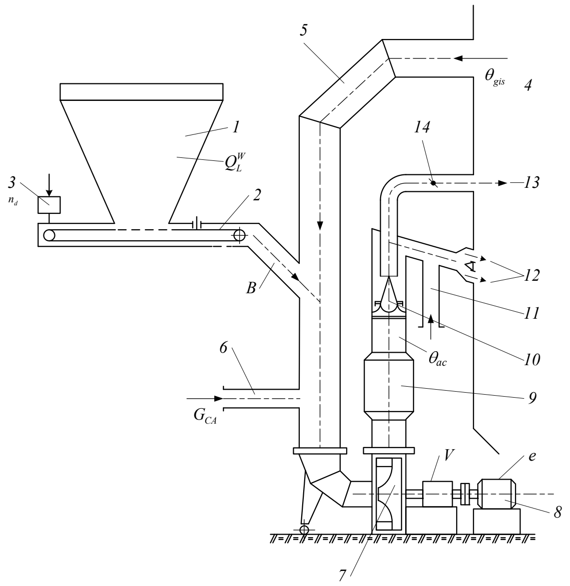


Figure 1: The graph of a mill fan for a steam generator

Raw coal from the hopper 1 is fed on conveyor 2, dosing is done by the speed feeder 3 or by the position of the slider 4. MF dries the coal sucking exhaust flue gases with temperature θ_{GIS}, which along the duct 5 are mixed with coal *B* and dry them. In certain situations, additional air 6 is supplied. The MF impeller 7, which is the main object of our consideration, is driven by the asynchronous motor 8 with unregulated speed. The coarse particles return back to the MF in the separator 9 for regrinding and small fractions are exported to burners 11 and 12. The received air mixture contains coal particles, air, combustion products and water vapor with temperature θ_{am} before the dust concentrator 10.

The impeller 7 of MF wears due to the abrasive effect of the ash in the coal and also due to the percussion of falling into it initially crushed coal. The linear wear of the impeller blades depends on the amount of the fed coal *B* t/h, the ash content of the coal A^c, the crystal structure and the chemical content of the mineral part *G* and also on the revs *n*:

$$\Delta L = f(B, A^c, G, n) \quad (1)$$

The relationship (1) is difficult to obtain analytically or using data, as with the exception of the revs *n*, the other quantities are difficult or impossible to measure. Wear of the blades can be judged by some indirect indicators, as shown in Fig. 2.

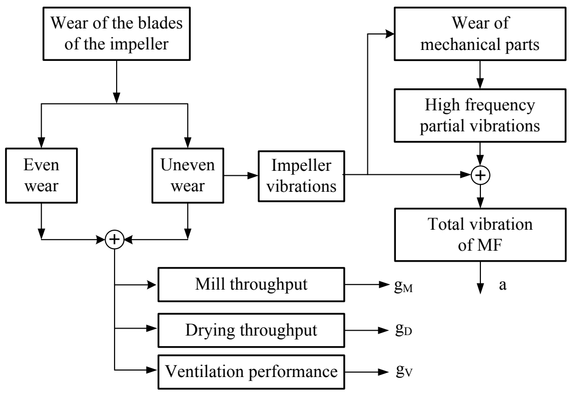


Figure 2: Indirect indicators for the wear of the blades

The mill throughput, the drying throughput and also the ventilation performance of the MF are reduced in all cases as a result of wear – respectively g_M, g_D and g_V. This paper treats another indicator of deterioration in the MF performance – the total vibration of the impeller, which occurs in uneven wear of the blades due to unbalance. The overall vibration *a* is a weak diagnostic signal, but it allows the degradation degree assessment. The total vibration is measured by DCS full-time sensors with discretization rate of 1s. These measurements cannot diagnose mechanical damage of the bearing block (misalignment, breakage of components, poor lubrication, unstable strengthening), but they allow the estimation of acceptable conditions for safe operation of the MF as a whole unit. On the MF, where the experiments were conducted, the amount of fuel is amended by the regulator, but the position of the effector u (Fig. 1) does not determine it uniquely. In the simplified functional diagram of Fig.3, the following data are measurable: the position of the effector and the inlet temperature of the gas intake shaft θ_{gis}, the air mixture temperature θ_{am}, the power of the electric motor P and the amplitude of the total vibration a.

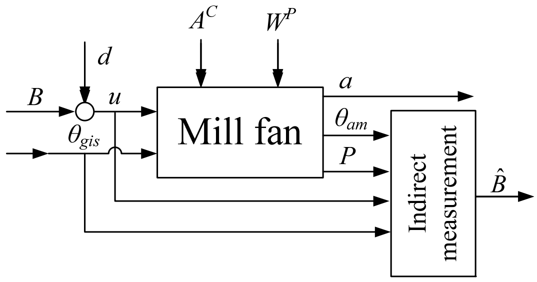


Figure 3: Simplified functional graph of a MF

The indirect measurement of the actual fuel consumption $\hat{B}$ may be replaced in the relationship (1)

$$\Delta L = f_1(u, \theta_{gis}, \theta_{am}, A^c, G, W^P, n) \quad (2)$$

Since the unknown unbalance of the impeller is connected to irregular wear

$$\Delta L_H = f_2(\Delta L) \quad (3)$$

and the vibration of the bearing block caused by it

$$a = f_3(\Delta L_H) \quad (4)$$

It is possible to write the qualitative dependency of the total vibration of the basic exploitation factors

$$a = f_4(u, \theta_{gis}, \theta_{am}, P, A^c, G, W^P, n) \quad (5)$$

The immeasurable factors are associated with the features of the fuel $K = (A^c, G, W^P)$ and they may be considered as a stochastic random process with mean $\overline{K}$ and deviation ΔK

$$K = \overline{K} + \Delta K; \quad (6)$$

the deviations ΔK will be added additively and possibly multiplicatively to the inaccuracies in the measurements of u, θ_{gis}, θ_{am} and P.

Thus, the total vibration can be considered as a multifactorial statistical dependence of u, θ_{gis}, θ_{am} and P, as shown in Fig.4.

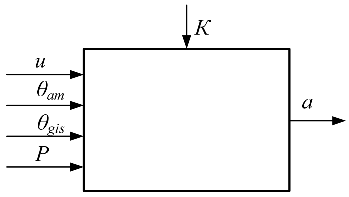


Figure 4: Block diagram of the functional dependence of the total vibration

3 FORMULATION OF THE PROBLEM

In this research, the main focus is on the gradual degradation of MFs due to the progressive wear of the impeller blades. The following tasks are determined:

3.1 Regions with increased total vibration

With a data source DCS on u, θ_{gis}, θ_{ac} and P to identify regions with increasing total vibration due to the impeller imbalance, due to an uneven wear of the blades (Fig.2).

3.2 Types of predictive models

Develop two types of predictive models of a sufficiently general form for areas with increasing amplitude of vibration:

- Autoregressive models
- Predictive dynamic models with controlling signals and disturbances

3.3 Predictive concrete models

Derive predictive concrete models for the MF degradation as multifactor trending analysis from the obtained experimental data.

3.4 Precision of the models

Estimate the precision of the obtained mathematical models.

3.5 Integrating results from models

Develop a method to integrate the results of the prediction by the two types of models so that

- Improve the predictive precision
- Preserve the ability to assess the impact of actions taken to actively change the load on the MF

3.6 The inverse problem

Develop a method to solve the inverse problem based on the unified model

$$\tilde{a} = F\left(u, \theta_{am}, \theta_{gis}, P, T_f\right) \tag{7}$$

so that at a predetermined magnitude of the boundary vibration $\tilde{a} = a_f$ and the time T_f to reach it, to determine the required load u_f satisfying these conditions:

$$u_f = F^{-1}\left(a_f, \overline{\theta}_{am}, \overline{\theta}_{gis}, \overline{P}, T_f\right) \tag{8}$$

where $\overline{\theta}_{am}, \overline{\theta}_{gis}, \overline{P}$ are the expected average values of the denoted measurable parameters.

4 EXPERIMENTAL DATA

Three types of time series are necessary for the purposes of vibration analysis of MFs in this research:

4.1 Discretization rate 1 hour

Time series with discretization rate $T_{01} = 1h$. These data allow to analyze the temporal trend of the total MF vibration in the range of several weeks. Fig. 5 shows a fragment of such series.

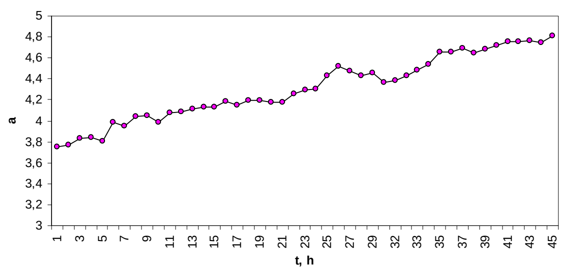


Figure 5: Total-vibration time-series fragment

4.2 Discretization rate 1 second

Series with the usual for DCS discretization rate of 1s for vibration phenomena with frequency up to 1 Hz.

4.3 High frequency diagnostics

High-frequency vibration analysis is performed with specialized firm equipment and well-known algorithms for isolated turbo (SKF Vibration Diagnosis Guide, www.skfreliability.com).

In our research efforts are focused on the prediction and analysis of vibrotrends as they allow to combine actions on MF technical maintenance with operational activities subject to the overall behavior of the power unit.

5 MATHEMATICAL MODELS

For the mathematical description of the wear of the MF blades based on the total vibration *a* and also for prognosticating the remaining lifetime resource evaluated for $\hat{a}$ there are developed two types of the following below mathematical models:

- Autoregressive model (MM_1), which is established on the basis of temporal series of the total vibration. The general form of the mathematical model is as follows:

$$\hat{a}(k+1) = \alpha_1 a(k) + \alpha_2 a(k-1) + \alpha_3 a(k-2) \tag{9}$$

For the interval in which the general vibrations approach the threshold value a_f, the modification of which is shown in Fig. 5, the coefficients in the autoregressive model have the following values: $\alpha_1 = 0.723$, $\alpha_2 = 0.18$ and $\alpha_3 = 0.1$.

- Predictive dynamic model (MM_2), including the control action *u* and the main disturbances: the temperature of the gas intake shaft θ_{gis}, the air mixture temperature θ_{am} and the power consumption of the electric motor *P*.

$$\hat{a}(k+1) = \alpha_1 a(k) + \alpha_2 a(k-1) + \alpha_2 a(k-2) + \\ + \beta_1 u(k) + \beta_2 \theta_{am}(k) + \beta_3 \theta_{gis} + \beta_4 P(k) \tag{10}$$

This mathematical model allows the load control CF to achieve the desired operational resources.

6 TECHNICAL-CONDITION BASED CONTROL OF MF

Based on the developed predictive models for multifactor analysis of the vibration trend of the total amplitude of the MF bearing block, a new type of procedure for combined operational/technological supporting management is introduced, the main points of which are summarized in the scheme of Fig. 6 and includes the following steps.

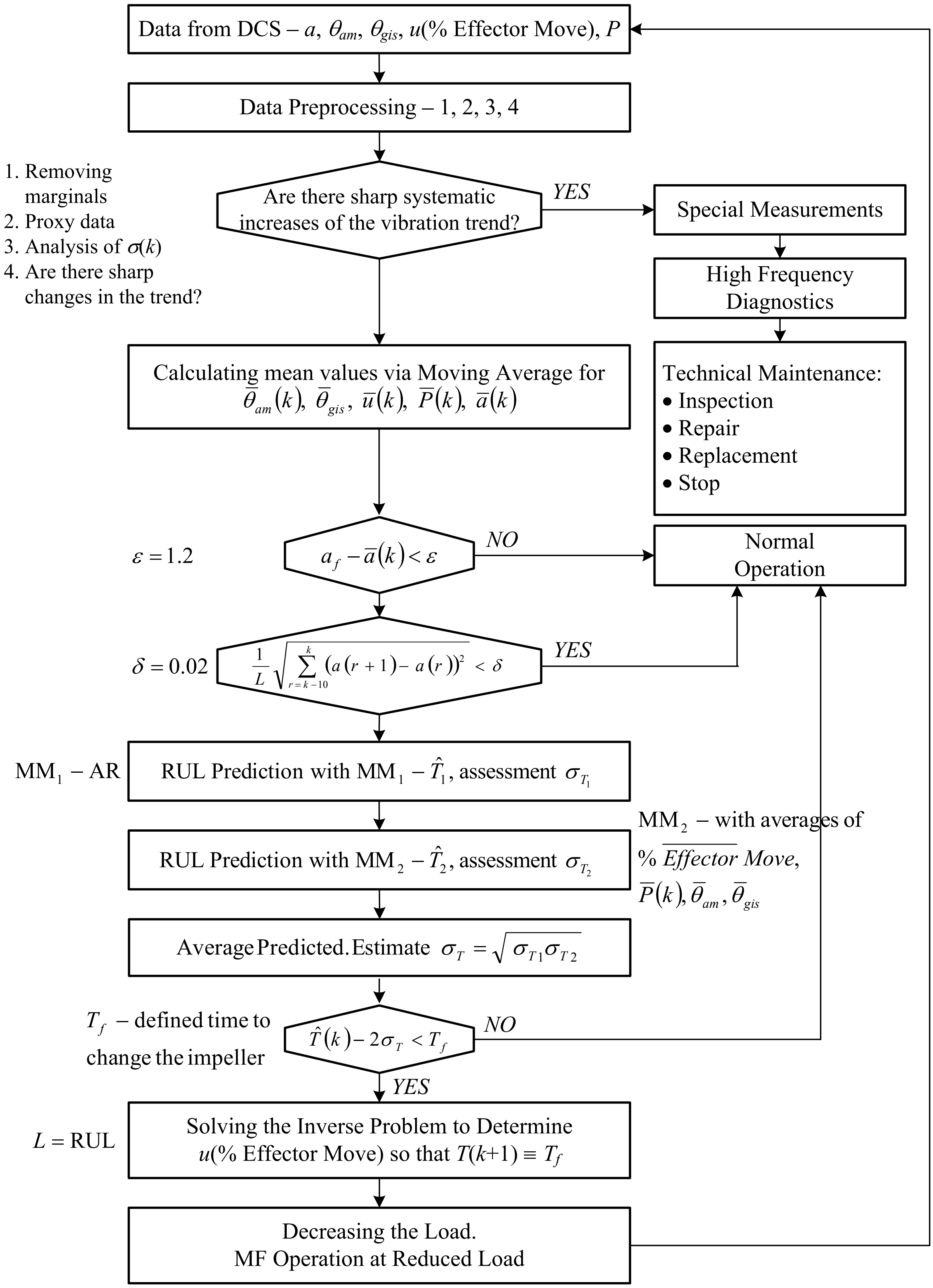


Figure 6: Procedure for combined operative/technological supporting management

6.1 Data from DCS

Systematization and synchronization of data from DCS for the basic in the treated problem measurable process parameters $a(k)$, $u(k)$, $\theta_{am}(k)$, $\theta_{gis}(k)$, $P(k)$.

6.2 Data preprocessing

Data preprocessing includes the following items:

- Determining the areas with growing trends of total vibration
- Removal of erroneous data (and sections) and replacing them with plausible values (linear interpolation)
- Evaluation for the presence of significant changes in the trend of total vibration to isolate cases of significant damages in the rest of the MF mechanical items (bearings, lubrication systems, fastening)
- Clarify the established vibration bounds and the defined by the planning level Residual Useful Life (RUL) for the MF (planned and forecast based on indirect measurements in the combustion chamber)

6.3 Final assessment

Are there sections in the treated trends with sharp and significant systematic increases of the vibration amplitude? If "yes" then is activated the branch with specialized apparatus for vibration diagnosis. Depending on the diagnostic result a decision is made for more frequent inspections, repair (temperature and level of lubricating fluid, tighten the fastening bolts), replacing the faulty item with a short stop, MF full stop and referral for repair).

6.4 Monotonous increasing trend

In the case of monotonically increasing trend of total vibration calculate the values of all measured variables $\overline{a}(k)$, $\overline{u}(k)$, $\overline{\theta}_{am}(k)$, $\overline{\theta}_{gis}(k)$, $\overline{P}(k)$ with Moving Average

$$\overline{z}(k) = \frac{1}{n}\sum_{i=k-n+1}^{k} z(i) \tag{11}$$

where $z(k)$ is any of the problematic variables.

6.5 Closeness to the threshold value

Check whether at time $t = kT_0$ vibration $\overline{a}(k)$ is far enough away from the set limit a_f (ε is an expert evaluation of proximity).

$$a_f - \overline{a}(k) < \varepsilon \tag{12}$$

6.6 In the zone of caution

Provided that the MF technical condition assessed by the total vibration $\overline{a}(k)$ is in the zone of caution ($a_f - \varepsilon$) check whether the current trend gradient $\| a(r+1) - a(r) \|$ exceeds the set limit δ using the expression

$$\frac{1}{L}\sqrt{\sum_{r=k-L+1}(a(r+1) - a(r))^2} < \delta \tag{13}$$

6.7 Autoregressive model

If the trend is significant use autoregressive model (AR) (9) to predict the Residual Useful Life (RUL) $\hat{T}_1 = RUL_1$ and the corresponding standard deviation prediction σ_{T1}.

6.8 Dynamic regression model

Similarly, by a regression dynamic model MM_2 (10) is defined the predicted RUL $\hat{T}_2 = RUL_2$ and the corresponding standard deviation of the prediction σ_{T2}.

6.9 Weighted estimate of the predicted time

A weighted estimate of the predicted time $\hat{T}$ is made:

$$\hat{T} = \alpha\hat{T}_1 + (1-\alpha)T_2, \ \alpha \le 1 \tag{14}$$

and the corresponding standard deviation of the evaluation $\sigma_{\hat{T}}$

$$\sigma_{\hat{T}} = \sqrt{\sigma_{T1}\sigma_{T2}} \tag{15}$$

6.10 $\hat{T}$ and T_f

Evaluate whether the predicted RUL time $\hat{T}$ coincides with the set for operational reasons time T_f.

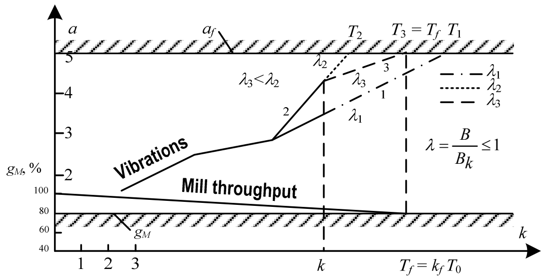


Figure 7: Combined MF control considering vibration

Fig. 7 presents graphically the proposed method of combined MF control with simultaneous consideration of the technological requirements (time limit reduction in the mill throughput $g_M \to T_f$) and conditions to avoid larger than a_f amplitudes of vibration. In a relatively small load λ_1, for time of reaching $a_f \to T_1 > T_f$, the impeller goes into repair earlier, having residual untapped resource $T_1 - T_f$. If the load λ_2 is very high $T_2 < T_f$ then the MF must be stopped because of the increased vibration, though that the machine has still satisfactory mill performance. If at point A the MF load is reduced according to equation (8), then $\lambda_1 < \lambda_3 < \lambda_2$ and the time of the remaining operational resources T_3 becomes equal to the defined time T_f. Thus, there is a simultaneous exhaustion of the MF resources related both to the mill throughput and also to the vibrations. The defined problem is solved. Because of the statistical nature of the predicted time $\hat{T}$ there are possible iterative corrections. Usually they are not necessary, since the very prescribed value of $RUL \to T_f$ is an estimated result of another model prediction, which is not subject to this research.

7 CONCLUSIONS

The proposed method allows the coordination of activities for support of the technological mode defined at a supervisory level and also maintaining the safety of MF operation within a dust-generating system of high-power steam generators. The resulting combined control allows technological requirements be considered as constraints in the task of finding the MF loading mode, so that the RUL resource be expired simultaneously in terms of various exploitation performance indices. Thus it is possible to realize economies of premature repair or of costs from unforeseen emergencies.

ACKNOWLEDGEMENTS

The research in this paper is partly funded by project FP7 – AcomIn № 316 087 and the Bulgarian NSF-Ministry of Education under contract № DVU-10-0267/10.

REFERENCES

Bata, S., Bonev, B., Totev, T., 1995 Problems and Decisions of the use of Bulgarian Low-Rank Lignite Coal for Electric Power Generation. In *Proceedings of STC*. Ohrid.

Bonev, B., Totev, T., Artakov, J., Nikolov, M., 1998. Diagnostic of Coal Dust Preparation Systems with Milling Fans. In *Proceedings of the 3rd International Conference "New Trends in Automation of Energetic Processes'98"*. Zlin, Czeh Republic.

Dong, Y.L., Gu, Y.J., Yang, K., Zhang, W.K., 2004. A combining Condition Prediction Model and its Application in Power Plant. In *Proceedings of International Conference on Machine Learning and Cybernetics, Vol.6*.

Goode, K.B., Moore, J., Roylance, B.J., 2000. Plant Machinery Working Life Prediction Method Utilizing Reliability and Condition Monitoring Data. In *Proceedings of the Institute of Mechanical Engineering, Part D*.

Hadjiski, M., Dukowski, S., 1995. Adaptive Control of 210 MW Boiler Milling System – Comparative Analysis of Several Approaches. In *International Symposium on Adaptive Control and Signal Processing*. Budapest, Hungary.

Hadjiski, M., Totev, V., 2000. Hybrid Modeling of Mill Fan of Steam Generators in TPP. In *Automatics and Informatics, № 4*. Sofia, Bulgaria (in Bulgarian).

Khatab, A.D., Ait Kadi, Rezg, N., 2012. A Condition-Based Maintenance Model for Stochastic Degradation Systems. In *9th International Conference on Modeling, Optimization and Simulation*. Bordeau, France.

Klepikov, N.Ts., Abudennikov, V.V., Volkov, A.B., et al., 2008. Dust-generating systems of mill fans for power boilers. In *Teploenergetika, № 9* (in Russian).

Rees, N.W., Fan, G.Q., 2003. Modeling and Control of Pulverized Fuel Coal Mills. In *IEE Power & Energy Series, 43*.

SKF Vibration Diagnosis Guide. `www.skfreliability.com`.

Tabaszewski, M., 2006. Forecasting of Residual Time of Milling Fans by Means of Neural Networks. In *Diagnostika'3* (*39*).

Van, P.D., Levart, E., Voisin, A., Inng, B., 2012. Remaining Useful Life (RUL) based Maintenance Decision Making for Deterioration Systems. In *2nd IFAC Workshop on Advanced Maintenance Engineering, Service and Technology*. Seville, Espagne.

Wei, J., Wang, J., Guo, S., 2009. Mathematical Modeling andCondition Monitoring of Power Plant Tube-Ball Mill Systems. In *Procedures of ACC*. St.Louis, USA.

Yang, W., 2001. Towards Dynamic Model Based Prognostics for Transmission Gear. In *Proceedings of SPIE Conference*.

Zhang, S., Ganesan, R., 1997. Multivariable Trend Analysis Using Neural Networks for Intelligent Diagnosis of Rotary Machinery. In *Transactions of ASME "Journal of Engineering for Gas Turbines and Power", Vol. 119*.

MAIN TYPES, COMPARISONS AND WORKING OF WIND GENERATORS

Chavdar Korsemov, Hristo Toshev

Institute of Information and Communication Technologies
Bulgarian Academy of Sciences
Acad G. Bonchev str., bl.2, Sofia, Bulgaria
chkorsemov@iinf.bas.bg, hr_toshev@mail.bg

Keywords: Wind turbine generators (WTG), exploitation of the WTG, electrical network

Abstract: The paper presents main types, comparison and working of the Wind Turbine Generators (WTG). The advantages and the shortcomings of the different groups of WTG are discussed. The specific features of the exploitation of WTG, located in wind parks, are considered with the purpose to obtain significant power for electric system. The problems, connected with WTG exploitation in the electrical network, are described. The stochastic feature, introduced in the network by the wind farms in case of relatively large participation in the general power, is a potential source for instability.

1 INTRODUCTION

The possibilities for rational use of the wind energy are indeed remarkable. About 1÷2% of the sunny energy, falling on earth is converted into wind. For comparison, the energy, absorbed by plants in photosynthesis, and its conversion into bio mass is only 0.02÷0.03%, or about hundred times smaller. Modern civilization, often regarded as lavish, consumes totally about 0.005÷0.006% of this energy (Danish Wind Energy Association).

According to the estimates in (Archer, 2005), the actual energy consumption can be satisfied by obtaining wind energy at a height of 80 m over 20% of the regions with average annual wind speed 6.9m/sec.

The world stores of easy obtained wind energy are estimated at 1 500 GW, with annual production of $3{\cdot}10^{12}$ kWh, that makes 500 kWh per every inhabitant of the world population of 6 milliards (Wind Force 12, 2005).

A specific feature of the wind generators, caused by the random character of the wind speed and the big dispersion, is the multiple difference between the average power generated and the installed nominal power. The relation between them, defined as utilization, is often determined as number of operation hours at nominal power, during which the same quantity of energy is produced, as the really generated for all 8760 hours of the year. Utilization of 23.5 % is accepted as very good utilization, which corresponds to 2000 hours work at nominal power.

The achievements in generators construction, computer technologies and power electronics enabled the creation of new generations of generators with full conversion and variable speed, which allow maximal deriving of energy at any wind speed. They comprise mainly the synchronous generators with constant magnets made from rare metals, as well as the asynchronous doubly fed generators. The generators with full conversion enable operation into a three-quarter mode − active power, reactive inductive and reactive capacitive with arbitrary ratios in-between. Their great advantage is that even under conditions of weak winds they can generate reactive power equal to the nominal fictitious power.

2 MAIN TYPES AND COMPARISONS OF WIND GENERATORS

The main type of generators, used until 1990 for power up to 1.5 MW, are the asynchronous ones, with a cage rotor, directly connected to the power line, from which they obtain the excitation – a Danish concept for WTG. They differ by their simplicity, reliability and a century developed technology of mass production. They operate at constant revolutions which does not allow maximal extraction of the wind energy. It is obtained for one single wind speed. The connection to the network is done by controlled valves restricting the big current overloading. After the transition process of switching $(0,2 \div 2s)$, the valves are shunted by contactors. In order to compensate for the magnetizing (excitation) current, some adjusted capacitor batteries are joined. The excitation current depends on the degree of loading and for its constant compensation, the battery must be adjustable.

A serious disadvantage of this type of generators is its firm characteristic resistance moment/ revolutions. The typical for the wind speed changeability of frequent stormy nature causes stress loading on the propeller shaft and especially on the speed box.

The introduction of a winded rotor is an improved modification of the asynchronous motor. The slippage is controlled by the degree of loading on the rotor windings, and thus on the revolutions, and the characteristic moment/revolutions becomes soft. The wind blow and the impulse loading of the generator are smoothed. The operation revolutions can be co-coordinated with the wind speed for maximal energy extraction. They are available in brush and non-brush variant. The heat mode of the rotor is lightened in the brush variant. Usually the revolutions differ from the synchronous ones by 1-10% (Gertmar et al., 2007).

The most widely used type of a generator in WTG is the Doubly Fed Induction asynchronous Generator (Doubly-Fed Induction Generator – DFIG). Regulating the energy, obtained by the rotor windings, the speed of propeller's rotation is controlled.

The double inversion enables efficient control of the slippage power (which reaches up to 30% of the nominal power), as well as its transfer together with this of the stator to the network. The new valves of IGBT type allow inversion with a high yield.

In order to design generators with acceptable dimensions, and hence – weights, their revolutions are most often chosen within the limits 1000-2000 rev./min and the poles number - 4 or 6. To achieve these revolutions, three-degree speed boxes are used. It is considered, that the losses in every transmission unit, if not higher than 6 is about 1% of the transmitted power in case of good support.

The gear box is an important element of a WTG. It is considerably affected by wind stress loading. The nacelle, or the WTG itself, containing a propeller, a gear box, a generator, attaching and protecting units, is a large construction with big weight. Due to the restrictions for few gear degrees, the generator itself is with big weight too. One of the known types of WTGs, that of Vestas, with power of 3 GW, weights about 70 tons (Gertmar et al., 2007). Enercon produce WTGs with direct transmission without a gear box and power of 4.5 MW, but the nacelle weight rises up to 450 tons.

Table 1:

Wind turbine characteristics	
Rated grid power (MW)	3
Rotor diameter (m)	90
Rated wind speed (m/s)	12
Rated speed (rpm)	15
Optimum tip speed ratio (blade tip speed divided by wind speed)	8
Maximum aerodynamic rotor efficiency (%)	48
Mass density of air (kg/m^3)	1.225
Generator material characteristics	
Slot filling factor k_{sfil}	0.6
Remanent flux density of the magnets B_{rm} (T)	1.2
Recoil permeability of the magnets μ_{rm}	1.06
Resistivity of copper at 120° ρ_{Cu} (μΩm)	0.025
Eddy-current losses in laminations at 1.5 T and 50 Hz P_{Fe0e} (W/kg)	0.5
Hysteresis losses in laminations at 1.5 T and 50 Hz P_{Fe0h} (W/kg)	2
Loss modeling	
Maximum losses in a single-stage gearbox P_{gearm} (kW)	45
Maximum losses in a three-stage gearbox P_{gearm} (kW)	90
Maximum losses in a 3 MW VSI P_{convm} (kW)	90
Maximum losses in a 1 MW VSI P_{convm} (kW)	30
Cost modeling	
Single-stage gearbox (ratio 6) cost (kEuro)	120
Three-stage gearbox (ratio 80) cost (kEuro)	220
Power electronics cost (Euro/kW)	40
Laminations cost (Euro/kg)	3
Copper cost (Euro/kg)	15
Permanent magnet cost (Euro/kg)	25
Rest of wind turbine cost (kEuro)	1300
Margin for company profit (kEuro)	250

Synchronous and asynchronous generators with variable rotational speed and with double conversion of the whole power – both the main one and that of slippage, are used. The entire disconnection between the propeller rotational speed and the network frequency enable the operation in four quadrants – in a mode of a generator, of a motor, of a consumer, and a generator or reactive power (capacitive, induction). This is an important feature, the use of which can improve the reliability of the network, to which it is attached. For small active power due to the wind low speed, such a generator can provide reactive power to the network up to its nominal apparent power.

The synchronous generators can be designed with electric excitation and constant magnets as well. The last variant leads to weight decrease. However, the price of the magnets made from rare metals, remains still quite high, which limits their use. Some realizations of synchronous generators with constant magnets and direct transmission without a gear box and direct generation of high voltage are known, which avoid the necessity of a step-up transformer for connection to the network. This approach potentially assumes high efficient use of the wind energy. But there are not any data available about their including in a real power line and that is why they are not discussed in detail.

Table 2:

	DFIG 3G	DD SG	DD PMG	PMG 1G	DFIG 1G
Generator dimensions					
Stator radius r_s (m)	0.42	2.5	2.5	1.8	1.8
Stack length l_s (m)	0.75	1.2	1.2	0.4	0.6
Number of pole pairs p	3	40	80	56	40
Number of slots per pole per phase q	6	2	1	1	2
Air gap g (mm)	1	5	5	3.6	2
Stator slot width b_{ss} (mm)	12.9	15	15	15	11
Stator tooth width b_{st} (mm)	11.5	18	18	19	12.5
Stator slot height h_{ss} (mm)	60	80	80	80	60
Stator yoke height h_{sy} (mm)	100	60	40	40	50
Rotor slot width b_{rs} (mm)	10	-	-	-	10
Rotor tooth width b_{rt} (mm)	11.5	-	-	-	12.5
Rotor slot height h_{rs} (mm)	60	-	-	-	60
Rotor yoke height h_{ry} (mm)	100	60	40	40	50
Pole/magnet height h_p / l_m (mm)	-	140	15	15	-
Rotor pole width b_p (mm)	-	137	79	82	-
Generator parameters					
Main inductance L_m (mH)	99	46	4	0.88	17
Stator leakage induct. $L_{s\sigma}$ (mH)	0.99	8.2	7.3	1.4	1.7
Stator resistance R_s (mΩ)	26	119	88	22	44
Rotor leakage induct. $L_{r\sigma}$ (mH)	1.2	-	-	-	1.8
Rotor resistance R_r (mΩ)	35	-	-	-	48
Generator material weight (ton)					
Iron (ton)	4.03	32.5	18.1	4.37	8.65
Copper (ton)	1.21	12.6	4.3	1.33	2.72
PM (ton)	-	-	1.7	0.41	-
Total (ton)	5.25	45.1	24.1	6.11	11.37
Cost (kEuro)					
Generator active material	30	287	162	43	67
Generator construction	30	160	150	50	60
Gearbox	220	-	-	120	120
Converter	40	120	120	120	40
Generator system cost	320	567	432	333	287
Other wind turbine parts appr.	1300	1300	1300	1300	1300
Margin for company costs	250	250	250	250	250
Total cost	1870	2117	1982	1883	1837
Annual energy					
Copper losses (MWh)	82	456	189	48	232
Iron losses (MWh)	70	40	89	123	123
Converter losses (MWh)	78	243	235	232	79
Gearbox losses (MWh)	533	-	-	273	267
Total losses (MWh)	763	739	513	674	701
Energy yield (GWh)	7.73	7.88	8.04	7.84	7.80
Annual energy yield / total cost					
(kWh/Euro)	4.13	3.72	4.05	4.16	4.25

The achievements in generators construction, computer technologies and power electronics have enabled the design of new generations of generators with complete conversion and variable speed that allows maximal extraction of the energy for any wind speed. These are mainly the synchronous generators with constant magnets made of rare metals and the asynchronous doubly fed generators. The generators with complete conversion enable operation in a three-quadrant mode – active power, reactive induction and reactive capacitive in arbitrary portions. Their main advantage is that even at weak wind they can generate reactive power equal to the nominal apparent one.

Interesting comparison is presented in (Polinder et al., 2006) between 5 different types of WTG with nominal power of 3 MW – Table1 and Table2:

- doubly fed asynchronous generators with complete conversion and a three-degree transmission box (transmission relation 1:80) – DFIG3G;
- directly connected to the propeller (without a speed box) 80-poles synchronous generators with electric excitation and complete conversion – DDSG;
- directly connected to the propeller 160-poles synchronous generators with constant magnets and complete conversion – DDPMG;
- synchronous 112-poles generators with one-degree transmission box with constant magnets and complete conversion – PMG1G;
- doubly fed asynchronous generators 80-poles generators with one-degree transmission boxes with complete conversion – DFIG1G.

3 EXPLOITATION OF WIND GENERATORS

The singular power of a WTG, compared to the conventional thermal or nuclear generators is small — within the limits of 1-5 MW. No considerable increase of this power can be expected. It is connected with increase of the surface, covered by the propeller and increases linearly with the second degree of its diameter, that leads to many constraints of another nature − weight, dimensions, possibility for transporting, mounting, available strong materials, lightning resistance, etc. For example, a WTG with power of 2.2 MW, has a propeller with a diameter of 112 m, weight of 20 tons and it obtains the energy from a section of 9800 m^2. (Thresher et al., 2007).

The connection of a separate generator to the power line is an isolated case, connected with big expenses for the joining itself. Its influence on the network can be regarded as an equivalent consumer.

The connection of WTG farms to a power line is related with the solution of many problems. The generated power has random nature with a large range of alteration – from zero up to the nominal value and limited predictability. The null power is often with a probability above 70%. The power control is possible only in direction of random maximum decrease. With this feature the network gets one more stochastic process together with the loads. The two processes are not correlated, compensation is possible in certain periods, in others – depositing, but in general, the dispersion increases. Due to the high probability for farm inactiveness, the nominal power of every farm must be reserved by conventional sources power. This reservation is especially heavy at relatively big share of the wind energy in the network. The reservation could be decreased with the help of the construction of several farms, located in sites with non-correlated wind energy. This requires good knowledge of the wind picture of the network region and appropriate planning of wind energetic development. Meteorological services data can be used for initial evaluation of a more continuous period of one up to three years.

The selection of farms sites is a complex optimization task. Besides the wind power, "statistic" independence, the location of the consumers, the distance to the electric network, the allocation of the conventional generators, their dynamics for compensating the changes of the wind energy must also be taken into consideration. The connection to the farm end (the generators remote points) may lead to decrease in network losses.

The compensating of short time alterations of the wind power is a separate task. It is done by rotating reserve powers and connected with additional dynamic losses. According to UCTE requirements, each change of the power consumed (for the wind power – positive or negative) must be compensated in an interval up to 15 min. The WTG with a cage rotor and especially those with complete conversion, have high dynamics and can participate in the compensation process, caused by load changes.

The predicting of the generated power has great significance in restricting the reserve and particularly the reserve rotating powers for every farm. It is usually assumed that the wind farms introduce strong disturbances in the energy network operation. The use of short-time (15 min) and daily (up to 38 h) prediction enables considerable decrease of the necessary rotating reserve powers. There are some well known methods, which, using the prognostic information, allow the short-time planning of the power, supplied by the conventional and wind generators for optimizing the value of the generated energy and the network reliability (Thresher et al., 2007), (Hansen et al., 2007), (Ummels et al., 2007).

4 CONCLUSIONS

The stochastic feature, introduced in the network by the wind farms in case of relatively large participation in the general power, is a potential source for instability. The operative staff must often distribute the generators loading, as well as the rotating reserve powers, including the wind, for minimizing the cost of the energy produced, accounting the constraints for:

- consumed electric energy,
- consumed thermal power by thermal and nuclear stations,
- possibilities of the generators for impulse (short-time) alteration of the generated power,
- actual prognosis for possible wind energy,
- special operation mode at small consumption – night hours, off days, holidays,
- water consumption in hydro electrical stations,
- electrical energy accumulation at PAWEC.

The rules for electric network operation, as well as the including of new energy sources is regulated by grid codes, specific for different countries and regions, but not differing much. Due to the principal difference of the wind energy generation in comparison with the conventional sources, there are introduced specific requirements in the countries with advanced wind energetic.

ACKNOWLEDGEMENTS

The research work reported in the paper is partly supported by the project AComIn "Advanced Computing for Innovation", grant 316087, funded by the FP7 Capacity Programme (Research Potential of Convergence Regions).

REFERENCES

Archer C. L., M.Z. Jacobson, 2005. "Evaluation of global wind power", *J. Geophis. Res.* Vol. 110, p.D 12110, 2005.

Danish Wind Energy Association, http://www.windpower.org/en/core.htm/fr/tour/wres/index

Gertmar L., L.Liljestrand, H.Lendenmann, 2007. Wind Energy Powers-That-Be Successor Generation in Globalization, *IEEE Trans. Energy Convers.*, vol.22, No 1, March 2007, pp.13-28.

Hansen A., P.Sorensen, F.Iov, F.Blaabjerb. 2007, Power Control of a Wind Farm with Active Stall Wind Turbines and AC Grid Connection Riso National Laboratory Wind Energy Department, P.O.Box 49, DK-4000, Roskilde, Dennmark.

Polinder H., F.Pijil, G.Vilder, P.Tavner, 2006. Comparison of Direct-Drive and Geared Generator Concepts for Wind Turbines, *IEEE Trans. Energy Convers.*, vol.21, No 3, Sept.2006, pp.725-733.

Thresher R., M.Robinson, P.Veers, 2007. To Capture the Wind, *IEEE Power & Energy Mag.*, No 6, Nov./Dec. 2007, pp.34-46.

Ummels B., M.Gibescu, E.Pelgrum, W.Kling, 2007. A. Brand. Impact of Wind Power on Thermal Generation Unit Commitement and Dispatch, *IEEE, Trans. E. C.* No 1, March 2007, pp.44-51.

Wind Force 12, 2005, http://ewea.org/fileadmin/ewea_documents/documents/publications/WF12/wf12-2005.pdf