CONTENTS

ANALYSIS OF FLOW AND THERMAL PHENOMENA IN EVACUATED TUBE COLLECTORS

Anna DEMIANIUK*, Sławomir Adam SORKO*

*Faculty of Environmental Engineering, Department of Heat Engineering, Bialystok University of Technology, ul. Wiejska 45 E, 15-351 Białystok, Poland

a.b.demianiuk@10g.pl, s.sorko@pb.edu.pl

Abstract: The subject of this case study is an issue of optimisation of flat tube solar collectors. Basic elements of energy analysis of performance parameters described by Hottel-Whillier equation are presented in the article. It is considered to be crucial to precisely analyse fluid flow through flow elements in evacuated tube collectors. It is especially important in the case of systems with channels of cross-sections shapes different from circular and for the use of detailed mathematical description of complex film conduction phenomena. It is presented that the advanced analysis of the flow and thermal phenomena in complex heat transfer systems, represented by evacuated tube collectors, enables engineering rationalisation of technical solutions for these devices.

Keywords: Flat Plate Solar Collector, Heat Transfer, Fluid Flow in Pipes, Optimum Mass Flow Rate, Boundary Equation Method

1. INTRODUCTION

The primary objective of every active and passive solar system is to gain solar radiation energy and to transfer it to the recipient in a planned, relatively simple way with highest possible efficiency of conversion. In many active and passive solar systems the basic element that absorbs solar energy is a flat plate collector. Its basic element is a plate or tube absorber. In the case of the plate absorber the medium which receives the heat flows directly under the absorber surface. The heat absorbing medium can be a liquid as well as a gas. The majority of collectors' solutions employ tube absorbers, where tubular elements are attached to the absorber plate. Pipes are arranged parallel to each other forming series of channels. In majority of installation solutions pipes or channels are filled up with a liquid, whereas in passive systems, typically, the air is the medium. Another form of tube collectors are evacuated tube collectors. In this particular type of solution pipes transporting heat are placed inside glass vacuum pipes in order to reduce heat loss from the medium to the ambient (Chwieduk, 2011 and Kalogirou, 2004). In the advanced constructions, designed for conversion of solar radiation energy into useful energy, panels of photovoltaic cells are placed on the absorber surface. In these cells types (PV/T) solar radiation energy is converted into heat and electricity in the rate dependent on the construction of the device (Ibrahim et al. 2011)

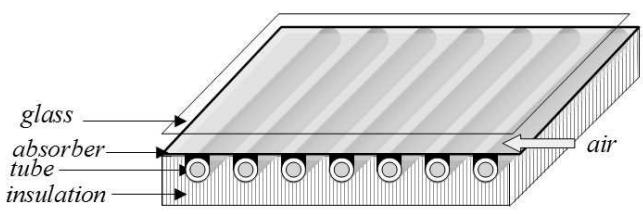


Fig. 1. Flat plate liquid collector with circular cross-section tubes

Variations design versions and types of solar collectors depending on kind of application (passive and active systems), type of medium used (air and liquid), connections for photovoltaic modules, and applied materials are described in detail in the literature (Zondag, 2008; Zhang et al., 2012; Charalambous et al., 2007).

The cross-section of flat plate liquid collector is presented in Fig. 1.

2. COLLECTOR HEAT BALANCE EQUATION

Deliberation on the thermal phenomena occurring in the collector is based on its thermal balance calculation. Balance equations of the configuration absorber-pipe system in the simplest form are obtained having analysed heat fluxes in the plate of the absorber and heat flux transferred to the medium inside pipe system of the collector. Fig. 2 presents a cross-section of flat plate collector with circular pipes system.

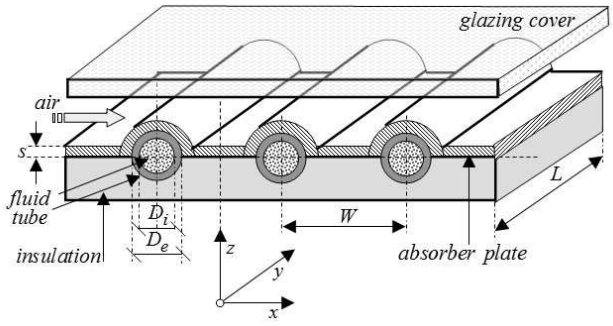


Fig. 2. The layout of a flat plate liquid collector of various dimensions

As investigated by Luminosu and Fara (2005) as well as Farahat et al., (2009), one of the most effective ways to obtain

the highest possible efficiency of the collector operation is to reduce heat loss to a minimum. Therefore it is crucial that thermal resistance of elements separating the main body of the solar radiation receiver, i.e. the absorbing surface (absorber), from the ambient was possibly the highest (Vestlund et al., 2009). Additionally, when absorbing surface is taken into consideration it is necessary to intensify processes of heat transfer in the absorber. It is important that the thermal resistance is possibly the lowest across absorber plate (between pipes or channels) and between the plate and the pipe or channel. A heat transfer phenomenon that occurs when the solar radiation energy is absorbed on the panel of the absorber and during its transfer to the medium flowing in pipe elements is analysed below.

Application of Hottel-Whillier model (Zondag, 2008, Alvarez et al., 2010 and Herrero Martín et al., 2011) in order to analyse thermal processes occurring in the collector facilitates describing temperature distribution $T_p \equiv T_p(x)$ in absorber plate and liquid temperature $T_f \equiv T_f(y)$ inside pipe system of the collector with equations:

$$\frac{\partial^2 T_p(x)}{\partial x^2} - C_p T_p(x) = -C_p\left(T_a + \frac{G_{sa}}{U_L}\right), \tag{1a}$$

where $c_p = \frac{U_L}{\lambda_p \delta}$ with boundary conditions:

$$T_p(x)\Big|_{X=0} = 0 \quad ; \quad T_p(x)\Big|_{X=h=\frac{(W-D_e)}{2}} = T_b \tag{1a*}$$

$$\frac{\partial T_f(y)}{\partial y} - C_f T_f(y) = C_f\left(T_a + \frac{G_{sa}}{U_L}\right) \tag{1b}$$

where $C_f = \frac{nWU_o}{\dot{m}c_p}$

with the boundary condition:

$$C_f = \frac{nWU_o}{\dot{m}c_p} \quad T_f(y)\Big|_{y=0} = T_i \tag{1b*}$$

where: U_L – overall collector heat loss coefficient [W/(m2K)], U_0 – coefficient of heat loss between ambient air and the medium inside pipes of the collector, G_{sa} – solar irradiation [W/m2], T_a – temperature of the ambitne [K], $\dot{m}$ – fluid mass flow rate in pipe collector [kg/s], c_p – heat capacity of fluid [J/(kg K)], λ_p – transmittance of the material [W/(m K)], δ – fin thickness [m], W – tube spacing [m], n – number of pipes in flat plate collector [-], T_b – temperature in the contact area for plate of the absorber and tube element [K], T_i – liquid temperature in the inlet to collector pipe [K].

Solution of the above equations provide to relations which describe temperature distribution in the absorber plate and in the medium flowing through the pipe elements:

$$T_p(x) = \frac{\cosh(C_p x)\left(T_b - T_a + \frac{G_{sa}}{U_L}\right)}{\cosh(C_p h)} + T_a + \frac{G_{sa}}{U_L} \tag{2}$$

Introducing F_R heat-removal factor for the collector:

$$F_R = \frac{\dot{m}c_p(T_o - T_i)}{S_c\left(G_{sa} - U_L(T_i - T_a)\right)} \tag{3}$$

where: S_c – area of the collector [m^2], $T_0 = T_f(y = L)$ – temperature of the medium in the outlet of the collector pipe obtained by calculation from equation (2b).

Density of energy gained by the collector and its useful thermal power can be described by following equations:

$$q_u = F_R\left(G_{sa} - U_L(T_i - T_a)\right) \; q_u = F_R\left(G_{sa} - U_L(T_i - T_a)\right) \tag{4a}$$

$$Q_u = S_c \cdot q_u \tag{4b}$$

Relating the density of the energy converted by the collector defined by Eq. 4a to the solar irradiation the expression determining efficiency of energy generation by the solar collector is as follows:

$$\eta_{c\,th} = \frac{q_u}{G_{sa}} \tag{5}$$

Fluid mass flow rate through system of pipes of the collector is a significant factor. This factor determines solar collector energetic parameters along with the thermal properties of the materials that the absorber is made of, pipe system, medium and with temperature difference between the medium and the ambient air.

Determination of the fluid mass flow of the medium inside the pipe system of the collector is easy in the case when the cross-section of pipes is circular.

In Fig. 3 the tube collectors are presented. There are three options of the cross-section shapes – elliptic, triangular and rectangular. Canals depending on the various construction solutions can be of various arbitrary cross-section shapes which can be obtained by profiling flow systems from steel sheets.

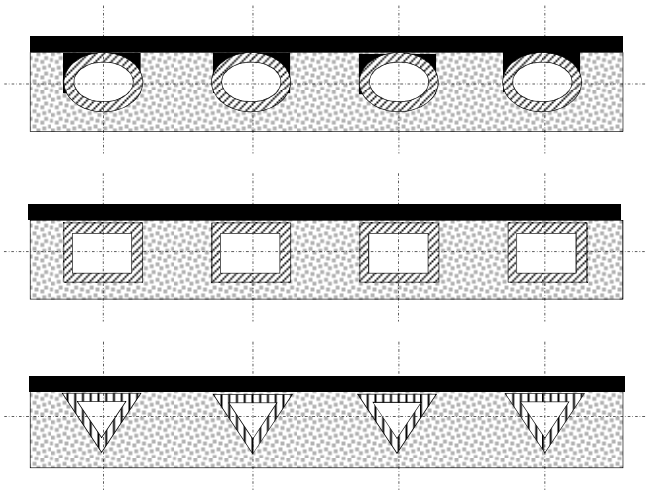


Fig. 3. Flat plate liquid collector with non-circular shape tubes

In the case when the collector pipes cross-sections are different from circular it becomes necessary to determine the flow velocity field and to calculate mass and volume flow rate of the medium by integration of velocity distribution inside the area. The area of calculations is described by canal profile contour. These calculations are essential in order to optimise construction parameters of the collector.

3. FLOW THROUGH NON-CIRCULAR CROSS-SECTION SHAPE PIPES COLLECTOR DETERMINATION

The sufficient model for fluid flow through pipes or hydraulic ducts of the collector flow system is the model of steady, uniform unidirectional laminar flow of incompressible viscous fluid (Batchelor, 1967).

3.1. Flow problem formulation

Unidirectional, slow flow of viscous fluid through a straight-axis pipe in the cross-section (Λ) of arbitrary shape inside the borderline (L), can be described by Stokes' equation with boundary condition which is the assumption of zero velocity value at inflexible and impermeable material border (at the duct partition):

$$\left(\frac{\partial^2 c_z}{\partial x^2}+\frac{\partial^2 c_z}{\partial y^2}\right)=-\Delta P\,, \tag{6}$$

where: $\dfrac{1}{\mu}\dfrac{\partial p}{\partial z}=-\Delta P$

with boundary condition:

$$c_z(x,y)=0 \;\;;\;\; \forall (x,y)\in L \tag{6a}$$

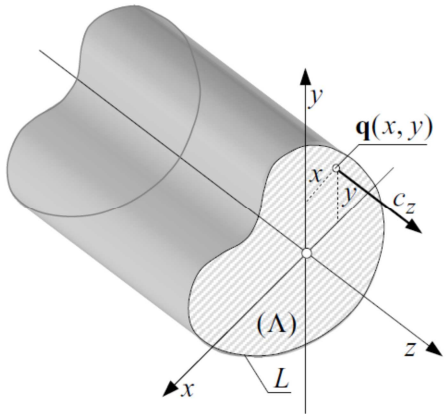


Fig. 4. Unidirectional flow through a straight duct

Among methods of solving the above-formulated boundary problem for Poisson equation (6) there is the decomposition of function ($c_z(x,y)$) into the homogeneous part ($\bar{c}_z(x,y)$), which satisfies Laplace equation, and the non-homogeneous part ($\tilde{c}_z(x,y)$), which satisfies Poisson equation.

$$c_z(x,y)=\bar{c}_z(x,y)+\tilde{c}_z(x,y)\,, \tag{7}$$

where:

$$\nabla^2\bar{c}_z(x,y)=0\,,\qquad \forall (x,y)\in\Lambda\cup L\,, \tag{7a}$$

$$\nabla^2\tilde{c}_z(x,y)=-\Delta P\,,\;\; \forall (x,y)\in\Lambda\cup L\,, \tag{7b}$$

where its non-homogeneous part can be written:

$$\tilde{c}_z(x,y)=-\Delta P\,\upsilon(x,y)\,, \tag{8}$$

where $\upsilon_z(x,y)$ is arbitrary chosen function which satisfies Poisson equation.

$$\nabla^2\upsilon(x,y)=1\;,\;\; \forall (x,y)\in\Lambda\cup L\;. \tag{8a}$$

One of the possible forms of this function is the square of the radius vector of any point $\mathbf{q}\equiv\mathbf{q}(x_\mathbf{q},y_\mathbf{q})\in\Lambda\cup L$,(Batchelor, 1967) i.e.:

$$\upsilon(\mathbf{q})=\frac{1}{4}r^2(\mathbf{q}) \tag{8a*}$$

Given the decomposition (7) of function $c_z(x,y)$ boundary condition $c_z(x,y)=0\;;\;\forall(x,y)\in L$ takes the form:

$$\bar{c}_z(x,y)=-\tilde{c}_z(x,y)=\Delta P\,\upsilon(x,y)\;\;\forall(x,y)\in L\;. \tag{9}$$

3.2. Integral formulation of the problem for unidirectional flow

Using Green's second identity, homogeneous component of flow velocity (function ($\bar{c}_z(\mathbf{p})$) satisfying Laplace equation in area (Λ) limited borderline (L) can be described by the following equation (Brebbia et al 1984):

$$\bar{c}_z(\mathbf{p})=-\int_{(L)}\frac{\partial\bar{c}_z(\mathbf{q})}{\partial n_\mathbf{q}}K(\mathbf{p},\mathbf{q})\,dL_\mathbf{q}+\int_{(L)}\bar{c}_z(\mathbf{q})E(\mathbf{p},\mathbf{q})\,dL_\mathbf{q} \tag{10}$$
$$(\mathbf{p})\in\Lambda\;;(\mathbf{q})\in L$$

where $\mathbf{p}(x_\mathbf{p},y_\mathbf{p})$ and $\mathbf{q}(x_\mathbf{q},y_\mathbf{q})$ are respectively the set point and the current integration point, and function $K(\mathbf{p},\mathbf{q})$ is a fundamental solution of Laplace equation:

$$K(\mathbf{p},\mathbf{q})=\frac{1}{2\pi}\ln\left(\frac{1}{r_\mathbf{pq}}\right)\,, \tag{10a}$$

$$E(\mathbf{p},\mathbf{q})=\frac{1}{2\pi}\frac{(x_\mathbf{p}-x_\mathbf{q})\,n_\mathbf{q}^y-(y_\mathbf{p}-y_\mathbf{q})\,n_\mathbf{q}^x}{r_\mathbf{pq}^2}\,, \tag{10b}$$

where: $\mathbf{n}_\mathbf{q}=\left[n_\mathbf{q}^x,n_\mathbf{q}^y\right]=\left[\dfrac{\delta y_\mathbf{q}}{\delta L_\mathbf{q}},\dfrac{\delta x_\mathbf{q}}{\delta L_\mathbf{q}}\right]$, is a versor normal to the boundary line (L) at point $\mathbf{q}(x_\mathbf{q},y_\mathbf{q})$.

After inserting dependence (9) into integral equation (10) and separating principal value from second integral, on right-hand side due to characteristic of kernel function $E(\mathbf{p},\mathbf{q})$ on boundary line (L) when $\mathbf{p}(x_\mathbf{p},y_\mathbf{p})\equiv\mathbf{q}(x_\mathbf{q},y_\mathbf{q})$ with an assumption that the border of the area is smooth, boundary integral equation is obtained:

$$\int_{(L)}g(\mathbf{q})K(\mathbf{p},\mathbf{q})\,dL_\mathbf{q}=\Delta P\left[-\frac{1}{2}\upsilon(\mathbf{p})+\int_{(L)}\upsilon(\mathbf{q})E(\mathbf{p},\mathbf{q})\,dL_\mathbf{q}\right] \tag{11}$$

where:

$$g(\mathbf{q})=\frac{\partial\bar{c}_z(\mathbf{q})}{\partial n_\mathbf{p}}\;\text{ and }\;\upsilon(\mathbf{q})=\frac{r(\mathbf{q})^2}{4} \tag{11a}$$

Equation (9) is a Fredholm integral equation of the first kind regarding density $g(\mathbf{q})$ of the function $\bar{c}_z(\mathbf{q})$ on the boundary of the area. The integral on right-hand side with the integrand $E(\mathbf{p},\mathbf{q})$ described by dependence (8b) is understood in the meaning of Cauchy principal value.

Having solved the integral equation (11) values of function $\bar{c}_z(\mathbf{p})$ at points $\mathbf{p}(x_p, y_p)$ inside the area (Λ) are determined according to the following integral relation:

$$\bar{c}_z(\mathbf{p}) = -\int_{(L)} g(\mathbf{q})K(\mathbf{p},\mathbf{q})\,dL_{\mathbf{q}} + \Delta P \int_{(L)} \upsilon(\mathbf{q})E(\mathbf{p},\mathbf{q})\,dL_{\mathbf{q}} \quad (12)$$
$$(\mathbf{p}) \in \Lambda\,;\,(\mathbf{q}) \in L$$

The volumetric flow rate of unidirectional flow through a duct of the cross-sectional area (Λ) is equal to:

$$Q = \iint_{(\Lambda)} c_z(\mathbf{q})\,d\Lambda_{\mathbf{q}} \quad (13)$$

Substituting expressions (7) and (8) as the integrand of the equation (13) and application of Green's second identity one can obtain the expression describing the flow rate of the creeping flow in an arbitrary shape cross-section straight pipe:

$$Q = \int_{(L)} \bar{c}_z(\mathbf{q})\frac{\partial \upsilon(\mathbf{q})}{\partial n_{\mathbf{q}}}\,dL_{\mathbf{q}} - \Delta P \int_{(L)} \upsilon(\mathbf{q})\frac{\partial \bar{c}_z(\mathbf{q})}{\partial n_{\mathbf{q}}}\,dL_{\mathbf{q}}\ . \quad (14)$$

Mass flow rate of the flow is the product of the volume flow rate described by the above expression and the fluid specific weight.

3.3. Numerical solution

The simplest way of discretisation of the integral equation is to approximate the curved closed boundary line (L) with J-element system of straight line segments with central collocation points of constant function density at each element.

When the boundary line (L) of the considered area (Λ) is approximated with the boundary linear elements of the constant density of function distribution $g(\mathbf{q}_j)$ and $\upsilon(\mathbf{q}_j)$ on each element ΔL_j, the integral equation (11) is reduced to the system of (J) linear algebraic equations with unknown function $g(\mathbf{q}_j)$ at points $\mathbf{q}_j\,(j = 1,\ldots, J)$ on the edge of the area:

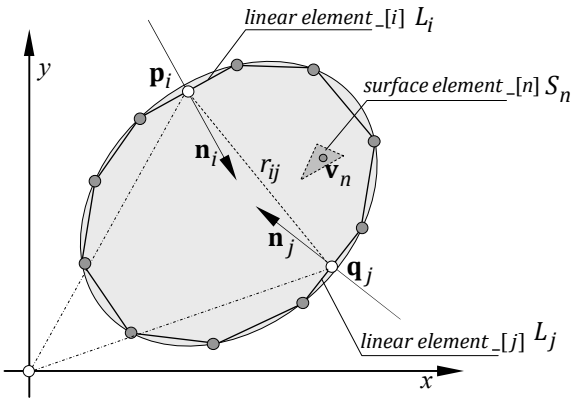


Fig. 5. Discretization of the boundary line with constant elements

$$\sum_{j=1}^{J} g(\mathbf{q}_j) \int_{(L_j)} K(\mathbf{p}_i,\mathbf{q}_j)\,dL_j =$$

$$= \Delta P\left[-\frac{1}{2}\upsilon(\mathbf{q}_i) + \sum_{\substack{j=1\\ j\neq i}}^{J} \upsilon(\mathbf{q}_j) \int_{(L_j)} E(\mathbf{p}_i,\mathbf{q}_j)\,dL_j\right] \quad (15)$$

where:

$$K(\mathbf{p}_i,\mathbf{q}_j) = \frac{1}{2\pi}\ln\left(\frac{1}{r_{ij}}\right) \quad (15a)$$

$$E(\mathbf{p}_i,\mathbf{q}_j) = \frac{1}{2\pi}\frac{(x_i - x_j)\,n_j^y - (y_i - y_j)\,n_j^x}{r_{ij}^2} \quad (15b)$$

and:

$$r_{ij} = \left[(x_i - x_j)^2 + (y_i - y_j)^2\right]^{\frac{1}{2}} \quad (15c)$$

$$\upsilon(\mathbf{q}_j) = \frac{r_j^2}{4} = \frac{1}{4}\left[(x_j - x_o)^2 + (y_j - y_o)^2\right] \quad (15d)$$

After determining the density of function distribution $g(\mathbf{q}_j)$ at the edge of the area, values of the function $\bar{c}_z(\mathbf{p}_n)(n = 1,\ldots,N)$ at points $(\mathbf{p}_n)(n = 1,\ldots,N)$ inside the area (Λ) are calculated using the following relation:

$$\bar{c}_z(\mathbf{p}_n) = \sum_{j=1}^{J} g(\mathbf{q}_j) \int_{(L_j)} K(\mathbf{p}_n,\mathbf{q}_j)\,dL_j +$$

$$+\sum_{\substack{j=1\\ j\neq i}}^{J} \upsilon(\mathbf{q}_j) \int_{(L_j)} E(\mathbf{p}_n,\mathbf{q}_j)dL_j \quad (16)$$

Finally, the velocity $c_z(\mathbf{p}_n)(n = 1,\ldots,N)$ at the points $(\mathbf{p}_n)(n = 1,\ldots,N)$, according to the equation (7), is described by the following sum:

$$c_z(\mathbf{p}_n) = \bar{c}_z(\mathbf{p}_n) + \Delta P\,\upsilon(\mathbf{p}_n) \quad (n = 1,\ldots,N) \quad (17)$$

3.4. Velocity field in elliptical, rectangular and triangular cross-section straight tubes – calculations principles

The flow velocity field is determined for the flow through straight pipe or closed duct thoroughly filled with liquid of density ρ=1000.00 kg/m^3, and viscosity μ=1.00*10^{-3} Pa·s (ethylene glycol 20% H_2O) at temperature 80°C) and pressure difference ΔP=10.00 s^{-1}m^{-1}.

3.4.1 Flow through an elliptic cross-section duct

The flow velocity field is determined for elliptic dust of the following dimensions: a=0.050 m, b=0.025 m. The results of the calculations are presented graphically in Fig. 6.a – c.

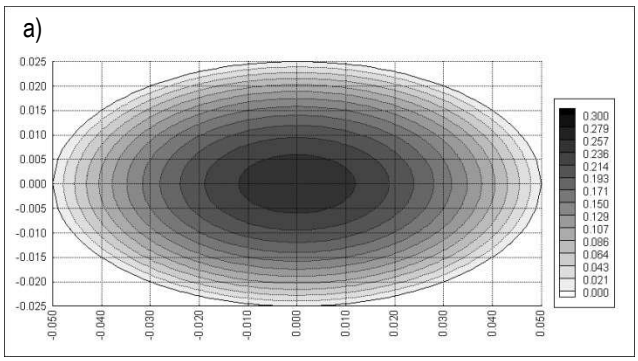


Fig. 6. Velocity distribution in a straight duct of elliptical cross-section:
a) Contour line chart of the velocity field in the duct
b) Velocity chart in the cross-sections: x=const
c) Velocity profile

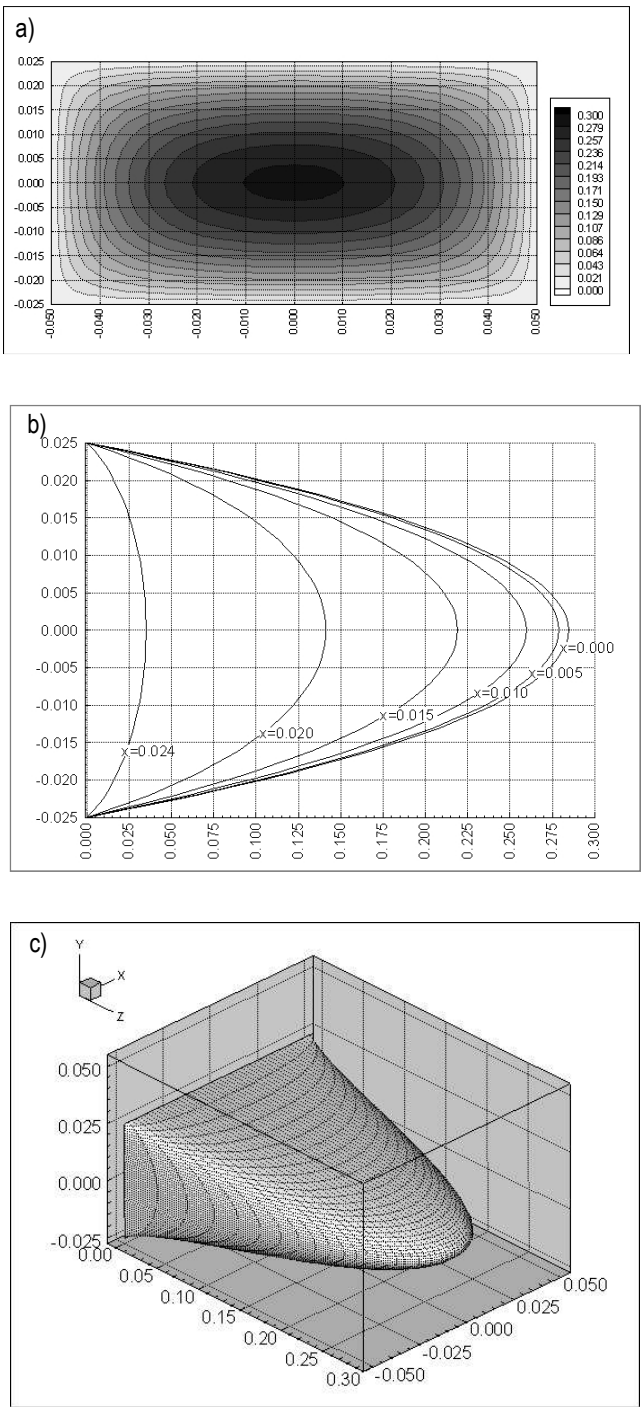


Fig. 7. Velocity distribution in a rectangular cross-section duct:
a) Contour line chart of the velocity field in the duct
b) Velocity chart in the cross-sections: x=const
c) Velocity profile

3.4.2. Flow through a rectangular cross-section duct

The flow velocity field is determined for rectangular duct of edges dimensions: a=0.050 m, b=0.025 m. The results of the calculations are presented graphically in Fig. 7.a – c.

3.4.3. Flow through a triangular cross-section duct

Velocity field is determined in the duct of a triangular cross-section (equilateral) of side size a=0.050 m.

The results of the calculations are presented graphically in Fig. 8.a – c.

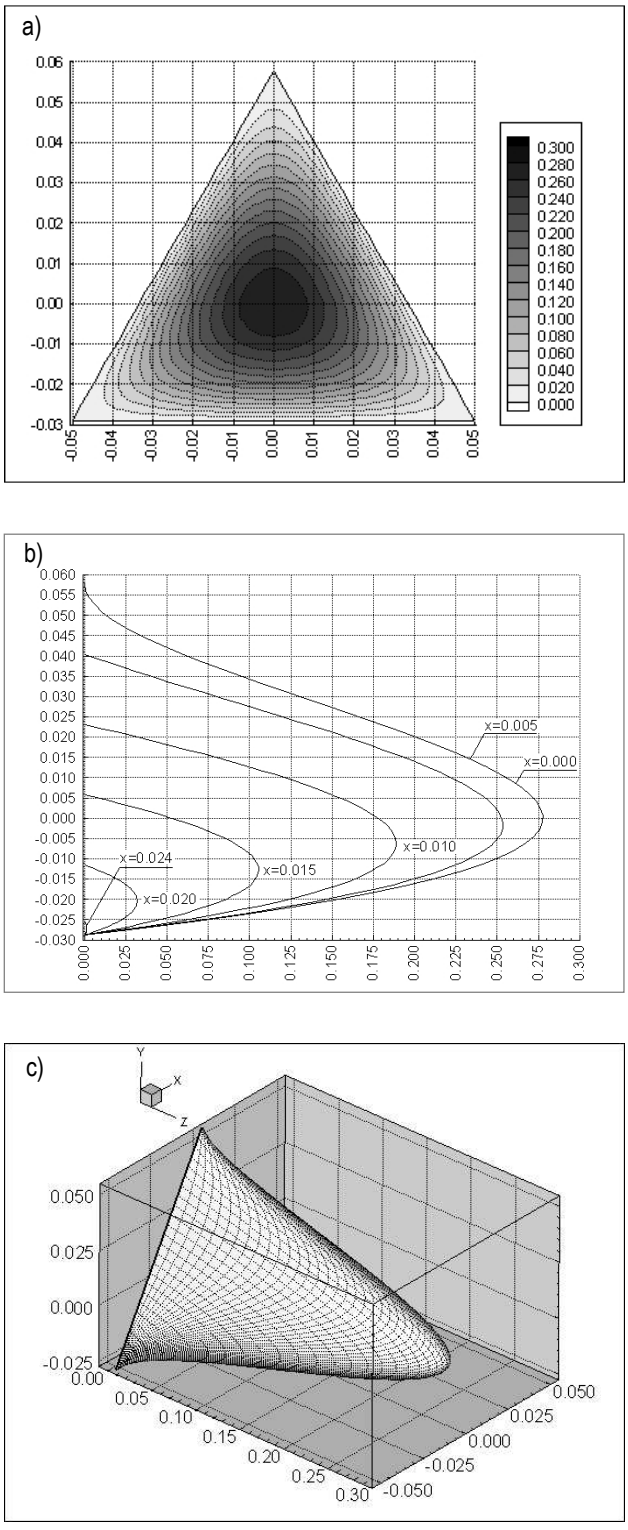


Fig. 8. Velocity distribution in a triangular cross-section duct.
a) Contour line chart of the velocity field in the duct
b) Velocity chart in the cross-sections: x=const
c) Velocity profile

4. CONCLUSIONS

The article presents classic model of calculations of temperature distribution in the absorber plate and in the pipe system in a flat plate solar collector. Further in the paper it is shown, how an energy efficiency of the device is determined. The model describes phenomena of flow and convective heat transfer in the systems of pipes of circular cross-sections. Determining velocity fields and temperature fields in the flow systems of liquid collectors is the principal element of flow and thermal optimisation of solar collectors.

The presented method uses Boundary Element Method that enables modelling and determining of flows in solar collectors pipe systems where cross-section shapes are different from circular. The method offers the possibility of determining velocity field and flow rate in the laminar flow of viscous fluid through straight-axis pipes and closed channels of arbitrary cross-section shape.

REFERENCES

1. **Alvarez A., Cabeza O., M.C. Muñiz M.C., Varela L.M.** (2010), Experimental and numerical investigation of a flat-plate solar collector, *Energy,* 35, 3707-3716.
2. **Batchelor G.K.,** (1967), *An Introduction to Fluid Dynamics*, Cambridge Univ. Press.
3. **Brebbia K., Telles, J.C.F., Wrobel, L.C.** (1984), Boundary Element Techniques. Theory and Applications in Engineering, *Springer-Verlag*.
4. **Charalambous P.G., Maidment G.G., Kalogirou S.A., Yiakoumetti K.** (2007), Photovoltaic thermal (PV/T) collectors: A review, Applied *Thermal Engineering,* 27, 275-287.
5. **Chwieduk D.** (2011), *Solar Energy in Buildings (in Polish),* Arkady.
6. **Farahat S., Sarhaddi F., Ajam H.** (2009), Exergetic optimization of flat plate solar collectors, *Renewable Energy*, 34, 1169–1174.
7. **Herrero Martín R., Pérez-García J., García A., García-Soto F.J., López-Galiana E.** (2011), *Simulation of an enhanced flat-plate solar liquid collector with wire-coil insert devices*, Solar Energy, 85, 455–469.
8. **Ibrahim A., Othman M.Y., Ruslan M.H., Mat S., Sopian K.** (2011), Recent advances in flat plate photovoltaic/thermal (PV/T) solar collectors, *Renewable and Sustainable Energy Reviews*, 15, 352-365.
9. **Kalogirou S.A.** (2004), Solar thermal collectors and applications, *Progress in Energy and Combustion Science*, 30, 231–295.
10. **Luminosu I., Fara L.** (2005), Determination of the optimal operation mode of a flat solar collector by exergetic analysis and numerical simulation, *Energy*, 30, 731–747.
11. **Vestlund J., Rönnelid M., Dalenbäck J.O.** (2009), Thermal performance of gas-filled flat plate solar collectors, *Solar Energy*, 83, 896–904.
12. **Zhang X., Xudong Zhao X., Smith S., Xu J., Yuc X. (**2012), Review of R&D progress and practical application of the solar photovoltaic/thermal (PV/T) technologies, *Renewable and Sustainable Energy Reviews,* Vol. 16, 594-617.
13. **Zondag H. A.** (2008)**,** Flat-plate PV-Thermal collectors and systems: A review, *Renewable and Sustainable Energy Reviews,* 12, 891–959.

Acknowledgement: The work described in this article was supported by Bialystok University of Technology Research Project S/WBiIŚ,5/11.

ASSESSMENT OF MUSCLE FORCES AND JOINT REACTIONS IN LOWER LIMBS DURING THE TAKE-OFF FROM THE SPRINGBOARD

Krzysztof DZIEWIECKI[*], Zenon MAZUR[*], Wojciech BLAJER[*]

[*]Institute of Applied Mechanics and Power Engineering, Faculty of Mechanical Engineering, Technical University of Radom
ul. Krasickiego 54, 26-600 Radom, Poland

krzysztof.dziewiecki@pr.radom.pl, z.mazur@pr.radom.pl, w.blajer@pr.radom.pl

Abstract: Computer simulation methods, based on the biomechanical models of human body and its motion apparatus, are commonly used for the assessment of muscle forces, joint reactions, and some external loads on the human body during its various activities. In this paper a planar musculoskeletal model of human body is presented, followed by its application to the inverse simulation study of a gymnast movement during the take-off from the springboard when performing the handspring somersault vault on the table. Using the kinematic data of the movement, captured from optoelectronic photogrammetry, both the internal loads (muscle forces and joint reactions) in the gymnast's lower limbs and the external reactions from the springboard were evaluated. The calculated vertical reactions from the springboard were then compared to the values assessed using the captured board displacements and its measured elastic behaviors.

Key words: Musculoskeletal Human Models, Inverse Dynamics Simulation, Muscle Forces, Joint Reactions

1. INTRODUCTION

In studying human movements, the non-invasive experiments are usually limited to photogrammetry from which the positions and orientations of the body segments are captured, electromyography (EMG) used to record the sequence and timing of muscle activity, and measurements of the ground reaction forces. Direct recording of muscle forces and joint reactions in vivo are currently infeasible. In this situation, inverse dynamics simulation, based on human body modeling and the non-invasive measurements, is still the prevailing method for the assessment of the internal loads during various human activities.

The inverse dynamics methodology, aimed at the determination of muscle forces and joint reactions in the motion apparatus, can be divided into four main stages:

1. design of the physical (musculoskeletal) model of the human body and its motion apparatus;
2. formulation of the mathematical model;
3. capturing the movement kinematic data and (possibly) other experimental data;
4. calculations using appropriate numerical codes.

2. PHYSICAL MODEL

The gymnast body is modeled as a planar kinematic structure composed of $N=16$ rigid segments (head, 3 trunk parts, arms, forearms, hands, thighs, legs, feet) interconnected by $k=15$ ideal hinge joints. The motion of the segments is assumed to be represented in the sagittal plane. In the deternministic model of actuation, the interaction between the segments is modeled by means of k resuntant muscle joint torques $\tau=[\tau_1 \cdots \tau_{15}]^T$ and $l=2k$ Lagrange multipliers $\lambda'=[\lambda'_1 \cdots \lambda'_{30}]^T$ that represent the X and Y components of the joint reactions (Fig. 1c). In the non-deterministic model of actuation (Blajer et al., 2007, 2010), the three control torques in the joints of each lower limb are replaced with the action of $m=9$ lower-limb muscles and groups of muscles that actuate the three degrees of freedom (Fig. 1b), $F=[F_1 \cdots F_9]^T$. In this model, due to the control overactuation in the lower limbs, the problem of distribution of the respective muscle torques τ' into the muscle forces F has infinite solutions, and is usually solved using optimization techniques (Winter, 2005; Winters and Woo, 1990; Yamaguchi, 2001). Applying the obtained muscle forces, the joint reactions in the lower limb will involve the tensile muscle forces, and as such λ obtained this way will differ from λ' obtained using the deterministic model.

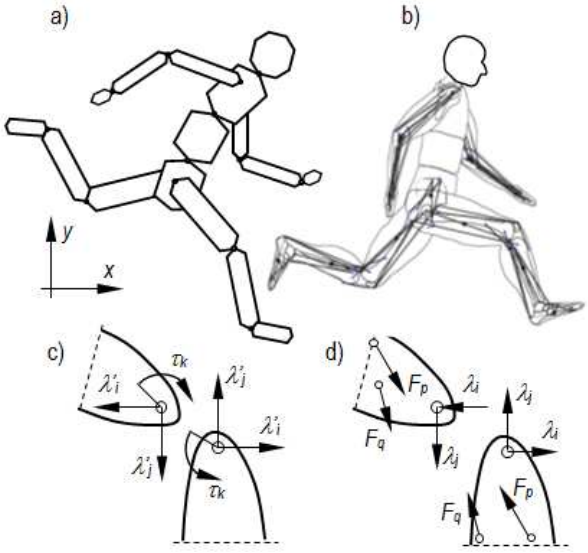


Fig. 1. The deterministic (a) and non-deterministic (b) models of the gymnast's body, and the interaction between the segments in the deterministic (c) and nondeterministic (d) models

The external loads on the gymnast's model are the gravitational forces $f_g = [f_{g1} \cdots f_{g16}]^T$ of the respective segments, and the ground reaction forces reduced to a chosen point on the foot segments, $R = [R_x \;\; R_y \;\; M_z]^T$. A symmetric distribution of the ground reaction between the two legs during the phase of contact with the springboard was assumed.

The inertial and anthropometric data of the subject body, i.e. the segment masses and mass moments of inertia, the segment lengths and mass center locations, the cross-sectional areas of the modelled muscles, the effective origin and insertion points of the muscles and their paths relative the skeleton, etc., were either directly measured of estimated using the guidelines reported in the literature (Winters and Woo, 1990, Zatsiorsky, 2002, Tejszerska et al., 2011).

3. MATHEMATICAL MODEL

The dynamic equations of the gymnast model are formulated in $n = 3N = 48$ coordinates $p = [x_{C1} \; y_{C1} \; \varphi_1 \cdots x_{CN} \; y_{CN} \; \varphi_N]^T$ that specify the location of the segment mass centers and their angular orientations with respect to an inertial (absolute) reference ftame. The generic matrix form of the equation is:

$$M\,\ddot{p} = f_g + B_u u + C_\lambda \lambda^* + C_R R \tag{1}$$

As said, in the deterministic model, the internal loads are modeled by means of the resultant muscle torques in the joints, distributed here into the lower-limb joint torques τ' and the upper body joint torques τ'', and the joint reactions λ' to which the contribution of the tensile muscle forces is not involved:

$$B_u u = B_\tau \tau = \left[B_{\tau'} \mid B_{\tau''}\right]\left[\tau' \mid \tau''\right]^T \quad \lambda^* = \lambda' \tag{2}$$

Then, in the nondeterministic model, the lower-limb joint torques τ' are replaced with the respective muscle forces F, which yields also more realistic joint reactions in the lower limbs (contribution of the tensile muscle forces is involved),

$$B_u u = \left[B_F \mid B_{\tau''}\right]\left[F \mid \tau''\right]^T \quad \lambda^* = \lambda \tag{3}$$

In the deterministic model, using the kinematic characteristics of the analyzed movement, the unknown internal loads and the external reactions can explicitly be determined from:

$$\left[\tau' \mid \tau'' \mid \lambda' \mid R\right]^T = \left[B_\tau \mid C_\lambda \mid C_R\right]^{-1}\left(M\,\ddot{p} - f_g\right) \tag{4}$$

The indeterminate inverse dynamics problem, i.e. the distribution of τ' from the determinate inverse dynamics formulation (4) into the respective muscle forces F, and then the determination of the joint reactions λ that include the influence of the tensile muscle forces in the lower limbs, can then be solved ising the projected dynamic equations. These to aims are achieved by introducing $r = 18$ independent coordinates $q = [x_G \; y_G \; \varphi_1 \cdots \varphi_{16}]^T$, where x_G and y_G are the absolute coordinates of a point on the top of the head segment, and the angular ccordinates are as used in p. Then, using a relationship $p = g(q, z, t)$, where z are the open-constraint coordinates in the directions of λ^*, two matrices can be extracted from:

$$\dot{p} = D\,\dot{q} + E\,\dot{z} + \gamma \tag{5}$$

where the $n \times r$ (48×18) matrix D is an orthogonal complement to the constraint matrix C_λ, $D^T C_\lambda^T = 0$, and the $n \times r$ matrix E its pseudo-inverse, $E^T C_\lambda^T = I$ (identity matrix). The equations (1) projected in q directions can then be represented as:

$$D^T\left(M\,\ddot{p} - f_g - C_R R\right) = D^T \frac{\left[B_{\tau'} \mid B_{\tau''}\right]\left[\tau' \mid \tau''\right]^T}{\left[B_F \mid B_{\tau''}\right]\left[F \mid \tau''\right]^T} \tag{6}$$

with the joint reactions excluded from the evidence.

Using the notation: $\overline{B}_\tau = D^T[B_{\tau'} \vdots B_{\tau''}]$, $\overline{B}_{F\tau} = D^T[B_F \vdots B_{\tau''}]$, and $\overline{M} = D^T M D$, after another projection of equations (6) into the controlled directions, one arrives at:

$$\overline{B}_\tau^T \overline{M}^{-1} \overline{B}_\tau \begin{bmatrix} \tau' \\ \cdots \\ \tau'' \end{bmatrix} = \overline{B}_\tau^T \overline{M}^{-1} \overline{B}_{F\tau} \begin{bmatrix} F \\ \cdots \\ \tau'' \end{bmatrix} \tag{7}$$

from which one can state the following relationship between the resultant muscle torques τ' and the muscle forces F (or stresses σ) in the lower extremities, i.e.

$$S_F\; F = P \quad \text{or} \quad S_F A_0\; \sigma = P \tag{8}$$

where $S_F = (D^T B_{\tau'})^T \overline{M}^{-1} D^T B_F$ is a square matrix of the dimension of F, $P = (D^T B_{\tau'})^T \overline{M}^{-1} D^T B_{\tau'} \tau'$, and A_0 is the diagonal matrix of the muscle cross-sectional areas. The muscular load sharing problem in the lower-limbs can then be stated as the following optimozation procedure

$$\begin{cases} \text{minimize} & J(\sigma) \\ \text{subject to} & S_F A_0 \sigma = P \\ \text{and} & \sigma_{\min} \le \sigma \le \sigma_{\max} \end{cases} \tag{9}$$

where J is an approproate objective function, and $\sigma_{\min}$ and $\sigma_{\max}$ are the physiologically allowable minimal and maximal muscle stresses. In this way the lower-limb muscle torques τ', obtained from the deterministic solution, are distributed into the respective muscle stresses/forces.

Using the evaluated muscle forces in the lower limbs, the joint reactions λ can be determined from the projection of the dynamic equations (1) into the constrained directions specified by E, i.e.

$$E^T M\,\ddot{p} = E^T\left(f_g + \left[B_F \mid B_{\tau''}\right]\begin{bmatrix} F \\ \cdots \\ \tau'' \end{bmatrix} + C_\lambda \lambda + C_R R\right) \tag{10}$$

which, after using $E^T C_\lambda^T = I$, leads to

$$\lambda = E^T\left(M\,\ddot{p} - f_g - \left[B_F \mid B_{\tau''}\right]\begin{bmatrix} F \\ \cdots \\ \tau'' \end{bmatrix} - C_R R\right) \tag{11}$$

4. SIMULATION RESULTS

The analysis was limited to the gymnast (master competitor of age 27, mass 70 kg, height 169 cm) movement during his landing and take-off from the springboard, with some short flying periods before and after the on-board phase, who performed a handspring vault with a front somersault in pike position (Fig. 2). The on-board phase was chosen for the expected high impact/impulsive forces and possible verification of the calculated external reactions from the springboard, which can be compared to the values assessed using the captured board displacements

and its measured elastic behaviors.

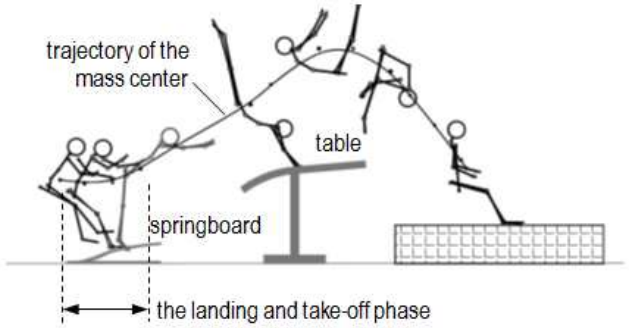


Fig. 2. Distinct phases of the performed handspring vault with a front somersault in pike position

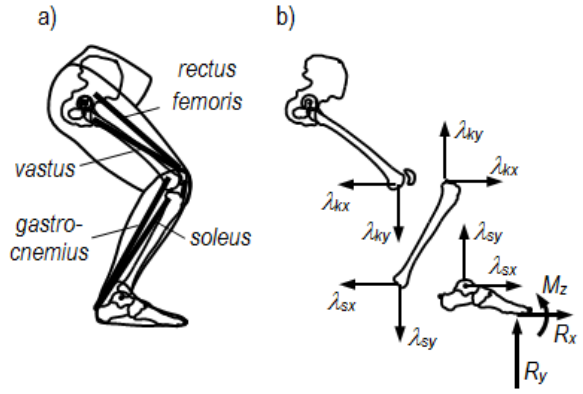


Fig. 3. The selected groups of muscles (a) and the joint reactions in lower limbs together with the external reactions (b)

The actual jump performance was recorded using a set of synchronized digital cameras (100 Hz), together with a separate registration of the springboard displacements. The raw kinematic data were then smoothed (Winter, 2005; Dziewiecki et al., 2011) to obtain the base point trajectories, from which the kinematic characteristics $p_d(t)$, $\dot{p}_d(t)$ and $\ddot{p}_d(t)$, used in the inverse simulation study, were calculated.

In Fig. 4 are reported the simulation results for the muscle forces of four selected group of muscles: *r.fem.* (*rectus femoris*), *vast.* (*vastus*), *gastr.* (*gastrocnemius*), and *sol.* (*soleus*), presented in Fig. 3.

As shown, the results obtained for the right and left limbs are very similar. During the on-board phase, which lasts from 4.48 to 5,58 s, the largest contribution to the movement performance comes from *sol.*, the contribution of *r.fem.* appears meaningless, *vast.* is active mainly during the landing, and *gastr.* during the take-off. Then, in Fig. 5 there are reported the calculations of the knee and ankle reactions during the movement. The estimated total reaction values, respectively in the knee (k) and ankle (s) joints are, are approximately 7 and 10 times higher than the gymnast weight.

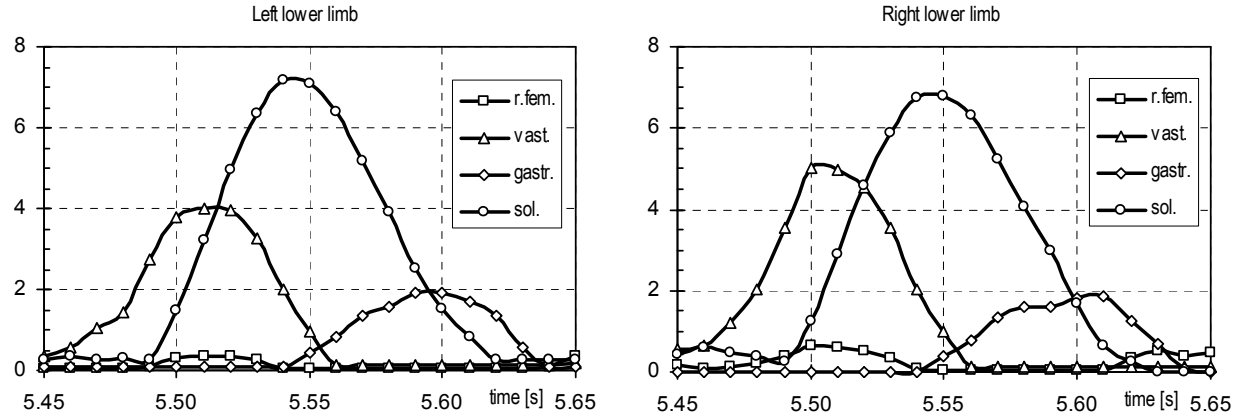


Fig. 4. The forces in selected muscles of the left and right lower limbs during the take-off from the springboard, related to the body weight of the gymnast

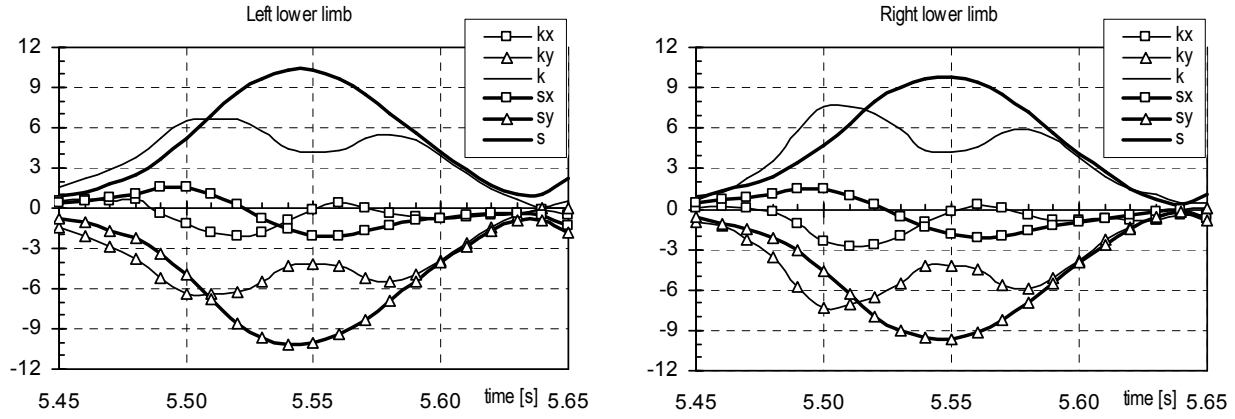


Fig. 5. The reactions in the knee (k) and ankle (s) joints of the left and right lower limbs during the take-off from the springboard, related to the body weight of the gymnast

5. LIMITED VALIDATION OF THE SIMULATION RESULTS

In addition to the calculations, the verical reactions R_y form the springboard during the take-off were estimated using the captured board displacements and its experimentaly measured elastic characteristics (Mazur at al.,2011b). The maximum values of the vertical reaction R_y, calculated from the inverse dynamics analysis and estimated from the recorded board displacements, were in good agreement (Fig. 6). This allows one to have limited confidence to the quality of the developed biomechanical model, correctness of its geometric and inertial parameters, and, finally, accuracy/adequacy of the kinematic characteristics used as an input to the human dynamics model in the inverse dynamics simulation (Erdemir et al., 2007).

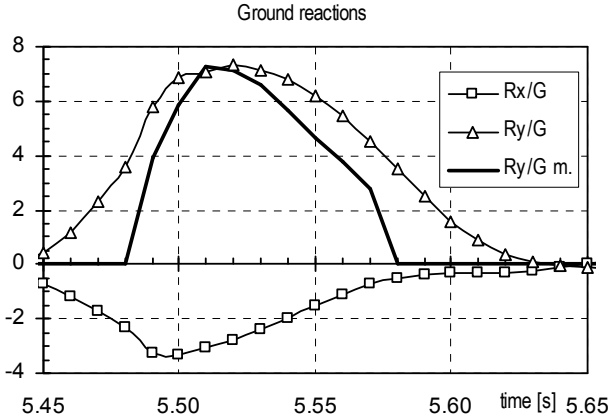


Fig. 6. The horizontal and vertical reactions from the springboard obtained from the inverse dynamics simulation, and the vertical reaction assessed from the recorded board displacements, related to the body weight (G) of the gymnast

6. CONCLUSIONS

The human motion apparatus is extremely complex and, as such, very difficult to model. For these reasons the models used in the inverse dynamics analyses always involve simplifications, according to the aims and expected exactidude of the analysis.

Studies on muscle force prediction usually compare the assessed muscle force loading or activation patterns against the EMG data as an estimate of validity. In this paper we compared the springboard vertical reaction values obtained from the inverse simulation against ithe values estimated from the measured board displacements. A good agrement of the maximum vertical values was reported.

In the literature, see e.g. Erdemir et al. (2007) for a review, advanced analyses exist which incorporate the quantification of muscle force sensitivity on diverse modeling parameters. Some critical model parameters are associated with the assumtions related to the musculotendon paths and the effective attachment points of the tendons (Winters and Woo, 1990; Blajer et al., 2010). The physiological cross-sectional area of muscles are the other parameters that significantly affect the magnitude of muscle force estimates (Mazur et al., 2011a). Of importance is also the way the raw kinematic data are processed (smoothed/filtered) before they are used in the inverse dynamics simulation (Dziewiecki et al., 2011). Finally, the musle force estimates are influenced by muscle decomposition and recruitment ctiteria used in the force sharing optimization process. Nonetheless, though the inverse dynamics simulations are possibly burdened with possible large inaccuracy, they still remain the only prevailing non-invasive method for the assessment of the internal loads during human movements.

The reported evaluations show that in gymnastics, during the dynamic movements like landing and take-off, the internal loads in the lower limbs may be much (a dosen or so) higher compared to the gymnast weight. The situation is reflected in frequent injuries of limb structures of locomotion apparatus, and different diseases after longer sport activity. Knowledge of muscle forces and joint reactions during the sport activities, even if approximate, can be of great importance for the risk assessment.

REFERENCES

1. **Blajer W., Czaplicki A., Dziewiecki K., Mazur Z.** (2010), Influence of selected modeling and computational issues on muscle force estimates, *Multibody System Dynamics*, Vol. 24, No. 4, 473-492.
2. **Blajer W., Dziewiecki K., Mazur Z.** (2007), Multibody modeling of human body for the inverse dynamics analysis of sagittal plane movements, *Multibody System Dynamics*, Vol. 18, No. 2, 217-232.
3. **Blajer W., Dziewiecki K., Mazur Z.** (2010), Remarks on human movement modeling for the inverse dynamics analysis (in Polish), *Acta Mechanica et Automatica*, Vol. 4, No. 2, 17-24.
4. **Dziewiecki K., Blajer W., Mazur Z.** (2011), Remarks on the methods of smoothing raw kinematic data for the inverse dynamics simulation of biomechanical systems (in Polish), *Modelowanie Inżynierskie*, Vol. 10, No. 41, 55-64.
5. **Erdemir A., McLean S., Herzog W., van den Bogert A.**, (2007), Model-based estimation of muscle forces exerted during movements, *Clinical Biomechanics*, Vol. 22, No. 2, 131-154.
6. **Mazur Z., Dziewiecki K., Blajer W**. (2011a), Remarks on the assessment of human body parameters for the inverse dynamics simulation (in Polish), *Aktualne Problemy Biomechaniki*, 5, 89-94
7. **Mazur Z., Dziewiecki K., Drapała K., Blajer W.** (2011b), *Studying elastic behaviorof a high performance springboard* (in Polish), Modelowanie Inżynierskie, (in press)
8. **Tejszerska D., Świtoński E., Gzik M.** (2011) *Biomechanics of human movement system*, Wydawnictwo Naukowe Instytutu Technologii Eksploatacji, Radom (in Polish).
9. **Winter D.A**. (2005), Bi*omechanics and Motor Control of Human Movement*, Wiley, Hoboken.
10. **Winters J.M., Woo S.L.-Y.** (1990), *Multiple Muscle Systems. Biomechanics and Movement Organization*, Springer-Verlag, New York.
11. **Yamaguchi G.T.** (2001), *Dynamic Modeling of Musculoskeletal Motion: a Vectorized Approach for Biomechanical Analysis in Three Dimensions*, Kluwer, Dordrecht.
12. **Zatsiorsky V M.** (2002), *Kinetics of Human Motion*, Human Kinetics, Champaign.

Acknowledgement: The work was supported by the grant No. N N501 156438.

FACTORIZATION OF NONNEGATIVE MATRICES BY THE USE OF ELEMENTARY OPERATION

Tadeusz KACZOREK*

*Faculty of Electrical Engineering, Białystok University of Technology, ul. Wiejska 45D, 15-351 Białystok, Poland

kaczorek@isep.pw.edu.pl

Abstract: A method based on elementary column and row operations of the factorization of nonnegative matrices is proposed. It is shown that the nonnegative matrix $A \in \Re_+^{n\times m}$ $(n \geq m)$ has positive full column rank if and only if it can be transformed to a matrix with cyclic structure. A procedure for computation of nonnegative matrices $B \in \Re_+^{n\times r}$, $C \in \Re_+^{r\times m}$ $(r \leq \text{rank}\,(n,m))$ satisfying $A = BC$ is proposed.

Keywords: Factorization, Nonnegative Matrix, Positive Rank, Procedure, Computation

1. INTRODUCTION

The factorization problem can be stated as: given nonnegative matrix $A \in \Re_+^{n\times m}$, find two nonnegative matrices $B \in \Re_+^{\mathrm{n}\times \mathrm{r}}$ and $C \in \Re_+^{r\times m}$ such that $A = BC$. The problem has been considered in many papers (de Almeida, 2011; Cichocki and Zdunek, 2006; Cohen and Rothblum, 1993; Donoho and Stodden, 2004; Lin, 2007; Lee and Seung, 2001) and it arises in many problems for example signal processing, quantum mechanics, combinational optimization etc. (de Almeida, 2011; Cohen and Rothblum, 1993). The factorization problem is closely related to the positive rank of nonnegative matrices (Cohen and Rothblum, 1993). The positive rank of nonnegative matrices plays important role in control system theory specially in the reachability problem of positive linear systems (Kaczorek, 2001).

In this paper a method based on elementary column and row operations of the factorization of nonnegative matrices will be proposed.

The paper is organized as follows. In section 2 the factorization problem is formulated and some basic definitions are recalled. The main result of the paper is presented in section 3, which is divided in three subsections. In the subsection 3.1 the elementary column and row operations and the elementary operation matrices are defined and their properties are formulated. Matrices with cyclic structures are introduced in subsection 3.2 and it is shown that the nonnegative matrix has positive full column rank if and only if it can be transformed to a matrix with cyclic structure. The proposed method of the factorization of nonnegative matrices is presented in subsection 3.3. The concluding remarks are given in section 4.The following notation will be used: $\Re$ – the set of real numbers, $\Re^{n\times m}$ – the set of $\mathrm{n}\times\mathrm{m}$ real matrices, $\Re_+^{n\times m}$ – the set of $n \times m$ matrices with nonnegative entries and $\Re_+^n = \Re_+^{n\times 1}$, $\mathrm{I_n}$ – *the* $n \times n$ identity matrix.

2. PRELIMINARIES AND PROBLEM FORMULATION

Definition 2.1. (Cohen and Rothblum, 1993) The smallest nonnegative integer r is called the positive rank of the matrix $\mathrm{A} \in \Re_+^{\mathrm{n}\times\mathrm{m}}$ and denoted by $\mathrm{rank}_+ A$ if there exist $b_k \in \Re_+^{\mathrm{n}}$, $k = 1, \ldots, r$ $(r \leq m)$ such that each column $a_i \in \Re_+^{\mathrm{n}}$, $i = 1, \ldots, m$ of A is the linear combination:

$$a_i = \sum_{k=1}^{r} c_{k,i} b_k \quad \text{for } i = 1,\ldots,m \tag{2.1}$$

with nonnegative coefficients $c_{ki} \geq 0$, $k = 1, \ldots, r$; $i = 1, \ldots, m$ of the vectors b_k.

From (2.1) it follows that if $\mathrm{rank}_+ A = r$ then the matrix $A \in \Re_+^{n\times m}$ can be written in the form:

$$A = BC \tag{2.2a}$$

where:

$$B = [b_1 \quad \ldots \quad b_r] \in \Re_+^{n\times r}, \quad C = \begin{bmatrix} c_1 \\ \vdots \\ c_r \end{bmatrix} \in \Re_+^{r\times m}, \tag{2.2b}$$

$$c_k = [c_{k,1} \quad \ldots \quad c_{k,m}], \quad k = 1,\ldots,r.$$

Definition 2.2. The representation of the matrix $A \in \Re_+^{n\times m}$ in the form (2.2) is called factorization of the matrix A.

The standard $\mathrm{rank}\ A$ of $A \in \Re_+^{n\times m}$ and the positive $\mathrm{rank}_+ A$ are related by (Cohen and Rothblum, 1993):

$$\mathrm{rank}\ A \leq \mathrm{rank}_+ A \leq \min(n,m) \tag{2.3}$$

For example the matrix:

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 3 & 3 \end{bmatrix} \tag{2.4}$$

has the standard rank equal to 3 and the positive rank equal to 4. The problem under considerations can be stated as follows.

Given a nonnegative matrix $A \in \Re_+^{n\times m}$, find its factorization (2.2) i.e. the matrices $B \in \Re_+^{n\times r}$ and $C \in \Re_+^{r\times m}$ such that (2.2) holds.

In this paper the factorization problem will be solved by the use of the elementary column and row operations.

3. PROBLEM SOLUTION

3.1. Elementary operations

To solve the factorization problem the following elementary column and row operations will be used:

- Multiplication of the ith column (row) by positive number c. This operation will be denoted by $\mathrm{R}[i \times c]$ ($\mathrm{L}[i \times c]$.
- Addition to the ith column (row) of the jth column (row) multiplied by negative number $-c$ $(c > 0)$. This operation will be denoted by $\mathrm{R}[i + j \times (-c)]$ ($\mathrm{L}[i + j \times (-c)]$.
- Interchange of the ith and jth columns (rows). This operation will be denoted by $\mathrm{R}[i, j]$ ($\mathrm{L}[i, j]$).

Let $R_m[i,c]$, $R_a[i,j,-c]$ and $R_i[i,j]$ be the elementary column operations matrices obtained by applying the elementary column operations $R[i \times c]$, $R[i + j \times (-c)]$ and $R[i,j]$ to the identity matrices respectively. Similarly, are defined the elementary row operations matrices $L_m[i,c]$, $L_a[i,j,-c]$ and $L_i[i,j]$. The elementary column operations are performed by post-multiplication of the matrix by the elementary column operations matrices and the elementary row operations are performed by premultiplication of the matrix by the elementary row operations (Kaczorek, 1993).

It is easy to prove the following lemmas.

Lemma 3.1. The inverse matrices $R_m^{-1}[i,c]$, $R_a^{-1}[i,j,-c]$, $R_i^{-1}[i,j]$ of $R_m[i,c]$, $R_a[i,j,-c]$, $R_i[i,j]$ and the inverse matrices $L_m^{-1}[i,c]$, $L_a^{-1}[i,j,-c]$, $L_i^{-1}[i,j]$ of $L_m[i,c]$, $L_a[i,j,-c]$, $L_i[i,j]$ satisfies the equalities:

$$R_m^{-1}[i,c] = R_m\left[i,\frac{1}{c}\right],\ R_a^{-1}[i,j,-c] = R_a[i,j,c],\ R_i^{-1}[i,j] = R_i[i,j] \tag{3.1a}$$

$$L_m^{-1}[i,c] = L_m\left[j,\frac{1}{c}\right],\ L_a^{-1}[i,j,-c] = L_a[i,j,c],\ L_i^{-1}[i,j] = L_i[i,j] \tag{3.1b}$$

Lemma 3.2. The elementary column operations $R[i \times c]$, $R[i + j \times (-c)]$, $R[i,j]$ and elementary row operations $L[i \times c]$, $L[i + j \times (-c)]$, $L[i,j]$ do not change the positive $\mathrm{rank}_+\mathrm{A}$ of the matrix $\mathrm{A} \in \Re_+^{\mathrm{n}\times\mathrm{m}}$.

Remark 3.1. It is assumed that after performance of any of the elementary column and row operations on a nonnegative matrix $\mathrm{A} \in \Re_+^{\mathrm{n}\times\mathrm{m}}$ the obtained matrix is also nonnegative.

For example performing on the matrix (2.4) the following elementary operations $L[2 \times 1/2]$, $L[4 \times 1/3]$, we obtain:

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \xrightarrow{R[2,4]} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \xrightarrow{L[3,4]} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \xrightarrow{R[2,3]} \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \tag{3.2}$$

The matrices (2.4) and (3.2) have the same positive rank equal to 4.

Let e_i be the ith column of the $n \times n$ identity matrix. The column ae_i for $a > 0$ is called the monomial column (Kaczorek, 2001). The nonnegative matrix consisting of $m(m \le n)$ linearly independent monomial columns has full column positive rank. The positive rank and standard rank of this matrix are the same as the matrix (2.4).

3.2. Matrices with cyclic structure

Definition 3.1. A nonnegative matrix:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1,n} \\ \vdots & \dots & \vdots \\ a_{n,1} & \dots & a_{n,n} \end{bmatrix},\ a_{ij} \ge 0,\ \ i,j = 1,\dots,n \tag{3.3}$$

is called the matrix with cyclic structure if:

$$a_{ii} \ge a_{i+1,i} \ge \dots \ge a_{n,i} \ge a_{1,i} \ge \dots \ge a_{i-1,i} \quad i = 1,\dots,n\,. \tag{3.4}$$

For example the matrix (3.2) has cyclic structure.

Theorem 3.1. [9] The system of linear algebraic equations:

$$a_{i,1}x_1 + a_{i,2}x_2 + \dots + a_{i,n}x_n = 1,\ \ a_{i,j} \ge 0,\ \ i,j = 1,\dots,n \tag{3.5}$$

has a nonnegative solution $x_i \ge 0, i = 1, \dots, n$ if and only if its coefficient matrix has the cyclic structure.

Theorem 3.2. The nonnegative matrix $\mathrm{A} \in \Re_+^{\mathrm{n}\times\mathrm{m}}$, $n \ge m$ has positive full column rank:

$$\mathrm{rank}_+ A = m \tag{3.6}$$

if and only if it can be transformed to a matrix A with cyclic structure, i.e. (3.4) holds for , $i = 1, \dots, m$.

Proof. Consider the matrix equation:

$$\begin{bmatrix} a_{11} & \dots & a_{1,m} & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ a_{m,1} & \dots & a_{m,m} & 0 & \dots & 0 \\ a_{m+1,1} & \dots & a_{m+1,m} & 1 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ a_{n,1} & \dots & a_{n,m} & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \tag{3.7}$$

By Theorem 3.1 the matrix equation has the nonnegative solution $x = [x_1 \dots x_m \ \ 1 \dots 1]^T \in \Re_+^n$ if and only if the matrix A has cyclic structure.

From Theorem 3.2 and Lemma 3.2 we have the following important corollary.

Corollary 3.1. The nonnegative matrix $A \in \Re_+^{n\times m}$ has positive full column rank $\mathrm{rank}_+\mathrm{A} = m$ if and only if it has cyclic structure or can be transformed to this cyclic structure by the elementary column and row operations.

For example the matrix (2.4) has not the cyclic structure but it has been transformed to the matrix (3.2) with cyclic structure by the use of the elementary row and column operations.

3.3. The proposed method

First the proposed method of the factorization of nonnegative matrices will be demonstrated on the following examples.

Examples 3.1. For the nonnegative matrix:

$$A=\begin{bmatrix}2 & 4 & 2\\ 1 & 4 & 5\\ 0 & 1 & 2\end{bmatrix} \tag{3.8}$$

find nonnegative matrices B and C satisfying the condition (2.2a). Using the elementary column operations to (3.8) we obtain:

$$\begin{bmatrix}2 & 4 & 2\\ 1 & 4 & 5\\ 0 & 1 & 2\end{bmatrix}\xrightarrow{R[2+1\times(-2)]}\begin{bmatrix}2 & 0 & 2\\ 1 & 2 & 5\\ 0 & 1 & 2\end{bmatrix}\xrightarrow[R[3+2\times(-2)]]{R[3+1\times(-1)]}\begin{bmatrix}2 & 0 & 0\\ 1 & 2 & 0\\ 0 & 1 & 0\end{bmatrix}=\overline{A} \tag{3.9a}$$

and:

$$\overline{A}=AR \tag{3.9b}$$

where:

$$R=R_a[2,1,-2]R_a[3,1,-1]R_a[3,2,-2]=\begin{bmatrix}1 & -2 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & -1\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & -2\\ 0 & 0 & 1\end{bmatrix}=\begin{bmatrix}1 & -2 & 3\\ 0 & 1 & -2\\ 0 & 0 & 1\end{bmatrix}. \tag{3.10}$$

From (3.9) and (3.10) we have:

$$A=\overline{A}R^{-1}=\begin{bmatrix}2 & 0 & 0\\ 1 & 2 & 0\\ 0 & 1 & 0\end{bmatrix}\begin{bmatrix}1 & -2 & 3\\ 0 & 1 & -2\\ 0 & 0 & 1\end{bmatrix}^{-1}=\begin{bmatrix}2 & 0 & 0\\ 1 & 2 & 0\\ 0 & 1 & 0\end{bmatrix}\begin{bmatrix}1 & 2 & 1\\ 0 & 1 & 2\\ 0 & 0 & 1\end{bmatrix}=BC \tag{3.11}$$

where:

$$B=\begin{bmatrix}2 & 0\\ 1 & 2\\ 0 & 1\end{bmatrix},\quad C=\begin{bmatrix}1 & 2 & 1\\ 0 & 1 & 2\end{bmatrix}. \tag{3.12}$$

Using (3.1) and (3.10) we may compute the inverse matrix R^{-1} as follows:

$$R^{-1}=R_a[3,2,2]R_a[3,1,1]R_a[2,1,2]=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 2\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 1\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 2 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}=\begin{bmatrix}1 & 2 & 1\\ 0 & 1 & 2\\ 0 & 0 & 1\end{bmatrix}. \tag{3.13}$$

The same result we obtain using the elementary row operations to (3.8):

$$\begin{bmatrix}2 & 4 & 2\\ 1 & 4 & 5\\ 0 & 1 & 2\end{bmatrix}\xrightarrow[L[1+2\times(-2)]]{L[2+3\times(-2)]}\begin{bmatrix}0 & 0 & 0\\ 1 & 2 & 1\\ 0 & 1 & 2\end{bmatrix}=\hat{A} \tag{3.14a}$$

and

$$\hat{A}=LA \tag{3.14b}$$

where:

$$L=L_a[1,2,-2]L_a[2,3,-2]=\begin{bmatrix}1 & -2 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & -2\\ 0 & 0 & 1\end{bmatrix}=\begin{bmatrix}1 & -2 & 4\\ 0 & 1 & -2\\ 0 & 0 & 1\end{bmatrix}. \tag{3.15}$$

From (3.14) and (3.15) we have:

$$A=L^{-1}\hat{A}=\begin{bmatrix}1 & -2 & 4\\ 0 & 1 & -2\\ 0 & 0 & 1\end{bmatrix}^{-1}\begin{bmatrix}0 & 0 & 0\\ 1 & 2 & 1\\ 0 & 1 & 2\end{bmatrix}=\begin{bmatrix}1 & 2 & 0\\ 0 & 1 & 2\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}0 & 0 & 0\\ 1 & 2 & 1\\ 0 & 1 & 2\end{bmatrix}=BC \tag{3.16}$$

where the matrices B, C are given by (3.12).

Using (3.1) and (3.15) we may compute the inverse matrix L^{-1} as follows:

$$L^{-1}=L_a[2\quad 3\quad 2]L_a[1\quad 2\quad 2]=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 2\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}1 & 2 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}=\begin{bmatrix}1 & 2 & 0\\ 0 & 1 & 2\\ 0 & 0 & 1\end{bmatrix}. \tag{3.17}$$

In general cases let us consider the nonnegative matrix $A\in\Re_+^{n\times m}$ with $n\geq m$. If $\mathrm{rank}_+\mathrm{A}=m$ then the matrix has trivial factorization (2.2) with B positive full column rank, i.e. $\mathrm{rank}_+\mathrm{B}=m$ and any nonnegative elementary column operations matrix C.

From example for the matrix $A=\begin{bmatrix}0 & 4 & 2\\ 1 & 0 & 3\\ 0 & 0 & 1\end{bmatrix}$ we have:

$$B=\begin{bmatrix}0 & 2 & 0\\ 1 & 0 & 0\\ 0 & 0 & 1\end{bmatrix},\quad C=\begin{bmatrix}1 & 0 & 3\\ 0 & 2 & 1\\ 0 & 0 & 1\end{bmatrix}. \tag{3.18}$$

Let:

$$\mathrm{rank}_+A=r<\min(n,m)\,. \tag{3.19}$$

If $n>m$ the following elementary column operations procedure is recommended.

Procedure 3.1.

Step 1. Using a suitable sequence of elementary column operations reduce the matrix $A\in\Re_+^{n\times m}$ to the form:

$$\overline{A}=AR=[B\quad 0]\in\Re_+^{n\times m},\quad B\in\Re_+^{n\times r} \tag{3.20}$$

where:

$$R=R_1R_2\ldots R_q \tag{3.21}$$

And $R_k, k=1,\ldots,q$ are the elementary column operation matrices defined in 3.1.

Step 2. Performing the elementary column operations on the identity matrix I_n and using (3.1) compute the inverse matrix:

$$R^{-1} = R_q^{-1} ... R_2^{-1} R_1^{-1} = \begin{bmatrix} \overline{R}_1 \\ \overline{R}_2 \end{bmatrix}, \quad . \qquad (3.22)$$
$$\overline{R}_1 \in \Re_+^{r \times m}, \quad \overline{R}_2 \in \Re_+^{(m-r) \times m}$$

Step 3. Using (3.20) and (3.22) find the desired matrices $B \in \Re_+^{n \times r}$ and $C = \bar{R}_1 \in \Re_+^{r \times m}$ satisfying (2.2).

Justification of Procedure 3.1 follows from (3.20) and (3.22) since:

$$A = \overline{A} R^{-1} = [B \quad 0] \begin{bmatrix} \overline{R}_1 \\ \overline{R}_2 \end{bmatrix} = B \overline{R}_1 = BC \, . \qquad (3.23)$$

Remark 3.2. If $n > m$ and $\mathrm{rank}_+ \mathrm{A} = m$ then the elementary row operations procedure is recommended or we may apply Procedure 3.1 to the transpose matrix A^T and use the equality $A^T = (BC)^T = C^T B^T$.

Example 3.2. Find the factorization (2.2) of the nonnegative matrix:

$$A = \begin{bmatrix} 3 & 2 & 1 & 0 \\ 0 & 2 & 1 & 0.5 \\ 6 & 8 & 4 & 1 \end{bmatrix} \in \Re_+^{3 \times 4} \, . \qquad (3.24)$$

In this case we apply the elementary row operations approach since $m = 4 > n = 3$. Using Procedure 3.1 we obtain the following.

Step 1. Using the following row operations we obtain:

$$A \xrightarrow[L[3+2\times(-2)]]{L[3+1\times(-2)]} \begin{bmatrix} 3 & 2 & 1 & 0 \\ 0 & 2 & 1 & 0.5 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \overline{A} = \begin{bmatrix} C \\ 0 \end{bmatrix}, \quad . \qquad (3.25)$$
$$C = \begin{bmatrix} 3 & 2 & 1 & 0 \\ 0 & 2 & 1 & 0.5 \end{bmatrix}$$

Step 2. Performing the elementary row operations $L[3 + 2 \times 2]$ $L[3 + 1 \times 2]$ on the identity matrix I_3 we obtain:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow[L[3+1\times 2]]{L[3+2\times 2]} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 2 & 1 \end{bmatrix} = L^{-1} = [B \quad \overline{B}], \quad . \qquad (3.26)$$
$$B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 2 \end{bmatrix}, \quad \overline{B} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Step 3. The desired matrices B and C satisfying $A = BC$ have the forms:

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 3 & 2 & 1 & 0 \\ 0 & 2 & 1 & 0.5 \end{bmatrix}. \qquad (3.27)$$

4. CONCLUDING REMARKS

The factorization problem of nonnegative real matrices has been addressed. A method based on elementary column and row operations of the factorization of nonnegative matrices has been proposed. It has been shown that the nonnegative matrix $\mathrm{A} \in \Re_+^{n \times m}$ $(n \geq m)$ has positive full column rank if and only if it can be transformed to a matrix A with cyclic structure (Theorem 3.2). A procedure based on the elementary operations for computation of nonnegative matrices $B \in \Re_+^{n \times r}$, $C \in \Re_+^{r \times m}$ $(r \leq rank(n, m))$ satisfying the condition (2.2a) has been proposed and illustrated by numerical examples.

REFERENCES

1. **Cichocki A., Zdunek R.**, (2006), Multilayer Nonnegative Matrix Factorization, *Electronics Letters*, Vol. 42, No. 16, 947-948.
2. **Cohen J.E., Rothblum U.G.** (1993), Nonnegative ranks, decompositions, and factorizations of nonnegative matrices, *Linear Algebra Appl.*, 190, 149–168.
3. **de Almeida A.** (2011), About Nonnegative Matrix Factorization: on the posrank approximation, *Proceedings of International Conference on Adaptive and Natural Computing Algorithms*, ICANNGA'11, Ljubljana, Slovenia.
4. **Donoho D., Stodden V.** (2004), When does nonnegative matrix factorization give a correct decomposition into parts?, S. Thrun et al. (eds.) *Proceedings of Advances in Neural Information Processing*, NIPS 2003, Vol. 16, MIT Press Cambridge.
5. **Kaczorek T.** (1993), *Linear Control Systems*, Vol. 1, J. Wiley, New York.
6. **Kaczorek T.** (2001), *Positive 1D and 2D systems*, Springer Verlag, London.
7. **Lee D.D., Seung H.S.** (2001), Algorithms for nonnegative matrix factorization, *Advances in Neural Information Processing*, Vol. 13, MIT Press Cambridge.
8. **Lin C.** (2007), Projected gradient methods for nonnegative matrix factorization, *Neural Computing*, Vol. 19, 2756-2779.
9. **Merzyakov J.I.** (1963), On the existence of positive solutions of a system of linear equations, *Uspekhi Matematicheskikh Nauk*, Vol. 18, No. 3, 179–186 (in Russian).

Acknowledgment: This work was supported by National Science Centre in Poland under work S/WE/1/11.

BIOMECHANICAL ANALYSIS OF TWO-POINT ASYMMETRIC SCREW FIXATION WITH IMPLANT FOR FEMORAL NECK FRACTURE

Dmitrij B. KAREV[*], Vladimir G. BARSUKOV[**]

[*]The Grodno State Medical University, 230009 ul. Gorkogo 80, Grodno, Belarus,
[**]The Yanka Kupala State University of Grodno, 230023, ul. Orzeshko 22,Grodno, Belarus,

bkarev@gmail.com, v.g.barsukov@grsu.by

Abstract: Stressed state peculiarities of cortical and trabecular bones by two-point asymmetric screw fixation with implant for femoral neck fracture are studied. Layer construction mechanic methods are used for analysis of stresses in cortical and trabecular bones. Biomechanical conditions for non-opening of the junction of the bone parts being joined are determined. It has been found that the total tightness of the broken parts when they rest against each other is secured over the whole fracture section without junction opening under condition that fixing screws are positioned in the trabecular bone without penetration of the thread side surface into cortical bone.

Key words: Femoral Neck Fracture, Screw Fixation, Stressed State, Cortical Bone, Trabecular Bone

1. INTRODUCTION

The main task of screw fixation for femoral neck fracture is securing the tightness (compression) of broken parts when they rest against each other. This tightness can be secured by means of different connections: one-point (central and eccentric); two-point (symmetric or asymmetric); three-point (symmetric, asymmetric with two bearing and one auxiliary fastening elements); four-point, etc. Extensive review of the world scientific and technical achievements in this field is given in the research papers (Manniger et al., 2007; Mow and Huiskes, 2005; Booth et al., 1998). The analysis of the published works shows that the major disadvantage of one-point fixation is the difficulty in preventing the parts being joined from possible rotation. In addition, one-point fixation stipulates the centric fixing element position which deteriorates blood circulation and fosters avascular necrosis of femoral head in case of femoral neck stabilization. Two-point, three-point and other multiple-point fixation secures broken parts from mutual rotation but a big number of fixation points makes osteosynthesis operation more traumatic and labour-consuming. Therefore, two-point and three-point fixation is considered to be preferable (Fig.1).

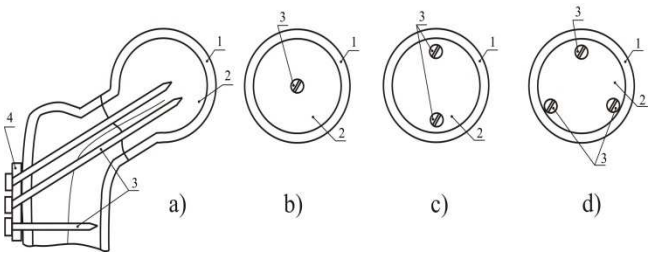


Fig. 1. Conventional screw fixation for femoral neck fractures: a – side view; b – one-point; c – two-point symmetric; d – three-point symmetric fixation: 1 – cortical bone; 2 – trabecular bone; 3 – fixing screw; 4 – plate

Moreover, the three-point fixation under certain conditions (strain-retention loss of one of the screws) may work as a two-point asymmetric fixation.

In spite of intensive research and development work, a number of issues of deflected mode of the broken femoral neck parts being joined by means of different types of fixation have been studied so far insufficiently. Approaches described in the literature (Booth et al.,1998; Akulich and Denisov, 2008; Yeremin, 2010) are based on simplified biomechanical models which consider bone tissue as homogeneous material. This introduces significant errors at determining the tension in the parts being joined.

On the other hand, modified fixation methods have been used increasingly in recent medical practice wherein one of the fixing screws is substituted with an implant of definite shape which is made from the bone of the person being operated upon (Karev, 2012) (Fig.2).

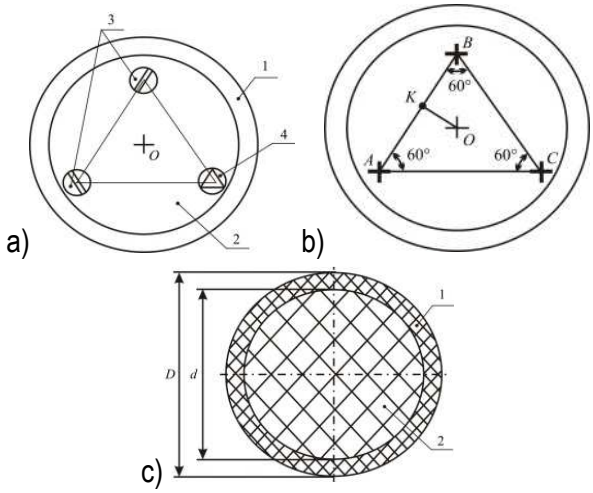


Fig. 2. Graphic representation of the basic scheme (a) and characteristic dimensional parameters (b), (c) for two-point asymmetric screw fixation with an implant: $OK = e$, $OB = OA = R$: 1 – cortical bone; 2 – trabecular bone; 3 – fixing screw; 4 – implant

These methods make it possible to decrease the proportion of foreign objects (fixing screws), facilitate drainage and thereby improve recovery of the femoral head in postoperative period by reducing the time of medical treatment. However, the biomechanical aspect of such approach is still not fully investigated which makes it difficult to analyze the potentialities of this osteosynthesis method.

The objective of this work is development of improved calculation methods based on taking into account differences of deformation-stress properties of the outer and inner layers of the bone and the performance of bones tension analysis in transfer from three-point to two-point asymmetric fixation type with an implant.

2. BIOMECHANICAL ANALYSIS OF FIXATION TYPES

From the mechanics of materials perspective, the femoral neck represents a two-layer material consisting of solid outer cortical layer 1 of diameter *D* and relatively less solid inner trabecular layer 2 of diameter *d* (Fig. 2c). Specifically, according to the data from literary sources the breaking stress of cortical bone under longitudinal tension is 133 MPa (Mow and Huiskes, 2005), under longitudinal compression – 193 MPa; the breaking stress of trabecular bone is from 3.65 to 9.1 MPa (Yeremin, 2010); modulus of elongation: 17–25 GPa for cortical bone and 0.2–2.5 GPa for trabecular bone (Mow and Huiskes, 2005). Indicator values of mechanical properties depend on age-related and pathological changes of bone tissue as a consequence of past medical history as well as on the loading speed while testing (Mow and Huiskes, 2005).

Significant (by order of magnitude greater) difference in physical and mechanical properties of bone tissue of cortical and trabecular layers results in nonuniform distribution of compression stress in bone section when fixing screws are tightened. Even if the screws are tightened uniformly in case of three-point symmetric fixation (Fig.1c), the tension, being uniform within each layer, differs at the layer borders proportionally to the differences in elasticity modulus (Fig. 3).

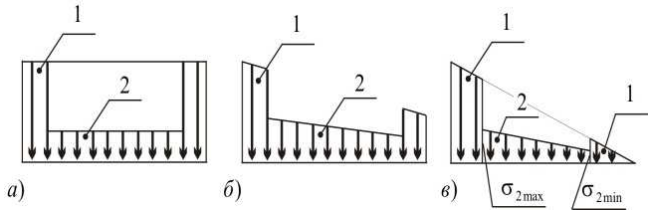


Fig. 3. Tension distribution pattern from the screw tightening over the section: a) – uniform tightening; b) – nonuniform tightening; c) – threshold case (junction opening); 1 – cortical bone; 2 – trabecular bone

In case of strain-retention loss of one of the screws, section tension distribution within each layer becomes nonuniform (Fig. 3b). At that, nonuniformity grows proportionally to the strain-retention loss degree. Maximum permissible is the case when the pressure (compression) in the outer layer becomes equal to zero (Fig. 3c). Further release of the screw strain is inadmissible as it is followed by the junction opening of the bone parts being joined.

It should be noted that the same compression distribution relevant to our case can be achieved by implementation of two-point asymmetric fixation illustrated in Fig. 2a. Let us analyze this type of fixation. Under the condition of uniform tightening of the fixing screws 1 and 2 with equal strain *V* their resultant (integral force) *F*=2*V* will be applied to the point *K*. Since from mechanical point of view this type of fixation corresponds to eccentrical compression of the layered structure by the applied effort which is positioned at the distance of eccentricity *e*=*OK* from the longitudinal axis, such loading can be considered as combination of centric compression with effort *F* and bending moment *M*=*F***e* (Vinokurov et al., 1998; Minenkov and Stasenko, 1977).

Corresponding formulas of mechanics of layered structures (Vinokurov et al., 1998; Minenkov and Stasenko, 1977), adapted to the calculation model under consideration are used for the evaluation of compression in each layer within femoral neck.

Specifically, for compression stresses in the cortical bone:

$$\sigma_{1c} = \frac{FE_1}{A_1E_1 + A_2E_2}, \quad (1)$$

and for the trabecular bone:

$$\sigma_{2c} = \frac{FE_2}{A_1E_1 + A_2E_2}, \quad (2)$$

where: σ_{1c} , σ_{2c} are correspondingly compression stress in cortical (1) and trabecular (2) layers of femoral neck; E_1 and E_2 – elasticity moduli of cortical and trabecular layers; A_1 and A_2 – cross-sectional area of cortical and trabecular layers; *F* – integral force of screw strain.

Bending stress for layered material (Vinokurov et al., 1998; Minenkov and Stasenko, 1977) in any point at the distance *r* from centroid of section:
in cortical bone:

$$\sigma_{1b} = \frac{M \cdot E_1 \cdot r}{E_1I_1 + E_2I_2} \quad (3)$$

and in trabecular bone:

$$\sigma_{2b} = \frac{M \cdot E_2 \cdot r}{E_1I_1 + E_2I_2} \quad (4)$$

Here σ_{1b} is the bending stress in the outer (cortical) layer; σ_{2b} is bending stress in the trabecular layer; I_1 and I_2 – correspondingly axial moment of inertia of cortical and trabecular layers of bone section.

Calculation values of areas of cortical A_1 and trabecular A_2 layers and corresponding axial moments of inertia of section I_1 and I_2 can be expressed in initial approximation through outer *D* and inner *d* diameters (fig.2c) by means of the following correspondences

$$A_1 = \frac{\pi D^2}{4}(1-\alpha^2) \quad (5)$$

$$A_2 = \frac{\pi D^2}{4}\alpha^2 \quad (6)$$

$$I_1 = \frac{\pi D^4}{64}\left(1-\alpha^4\right) \quad (7)$$

$$I_2 = \frac{\pi D^4}{64}\alpha^4 \quad (8)$$

where α=*d*/*D*.

The formulae given above (5)–(8) are approximate because they do not take into account the weakening of sections by the openings intended for fixing screws and the implant due to its little effect.

With regard to (5)–(8) formulae (1)–(4) for stresses calculation are brought to the following form:

Compression stress in a cortical bone:

$$\sigma_{1c} = \frac{4FE_1}{\pi D^2\left[\left(1-\alpha^2\right)E_1 + \alpha^2 E_2\right]} \tag{9}$$

compression stress in a trabecular bone:

$$\sigma_{2c} = \frac{4FE_2}{\pi D^2\left[\left(1-\alpha^2\right)E_1 + \alpha^2 E_2\right]} \tag{10}$$

bending stress in a cortical bone:

$$\sigma_{1b} = \frac{64 \cdot M \cdot E_1 \cdot r}{\pi D^4\left[\left(1-\alpha^4\right)E_1 + \alpha^4 E_2\right]} \tag{11}$$

bending stress in a trabecular bone:

$$\sigma_{2b} = \frac{64 \cdot M \cdot E_2 \cdot r}{\pi D^4\left[\left(1-\alpha^4\right)E_1 + \alpha^4 E_2\right]} \tag{12}$$

Maximum bending stress values are obtained if:

$$M = F * e_{max,} \tag{13}$$

where e_{max} – maximum eccentricity of integral force application F.

The maximum resulting stresses magnitude will be observed in the outermost point from centroid of section of the corresponding layer from the side where the compression stress and bending stress coincide, and the minimum ones – at the same kind of point from the side where the compression stress and bending stress signs are opposite.

3. BIOMECHANICAL CONDITION FOR NON-OPENING OF THE JUNCTION OF THE BONE PARTS BEING JOINED

The criterion for non-opening of the junction is the absence of tensile stress in the zone where the signs of compression and bending stress are opposite. Having equated the maximum bending stress to compression stress after transformation we get the value of maximum permissible level of eccentricity e_{max} of screw strain resultant application:

for cortical bone:

$$e_{1\max} = \frac{D\left(1-\alpha^4\right)E_1 + \alpha^4 E_2}{8\left[\left(1-\alpha^2\right)E_1 + \alpha^2 E_2\right]} \tag{14}$$

for trabecular bone:

$$e_{2\max} = \frac{D\left(1-\alpha^4\right)E_1 + \alpha^4 E_2}{8\alpha\left[\left(1-\alpha^2\right)E_1 + \alpha^2 E_2\right]} \tag{15}$$

Maximum permissible relative eccentricity value of screw strain resultant application:

for cortical bone:

$$\frac{e_{1\max}}{D} = \frac{1}{8}\frac{\left(1-\alpha^4\right)E_1 + \alpha^4 E_2}{\left[\left(1-\alpha^2\right)E_1 + \alpha^2 E_2\right]} \tag{16}$$

for trabecular bone:

$$\frac{e_{2\max}}{D} = \frac{1}{8}\frac{\left(1-\alpha^4\right)E_1 + \alpha^4 E_2}{\alpha\left[\left(1-\alpha^2\right)E_1 + \alpha^2 E_2\right]} \tag{17}$$

For illustrative purposes the tables (1, 2, 3) contain calculated magnitudes of the maximum permissible eccentricity value by condition of non-opening of the junction in cortical and trabecular bone layers. Calculations were made through the example of widely occurring diameter of femoral neck D=40 mm for different values of elasticity moduli of layer materials in case of two-point fixation with uniform screw strains.

Tab. 1. The calculated magnitudes of dimensional parameters for value of cortical bone elasticity modulus E_1=17 GPa

Name and identifier of dimensional parameter	Elasticity modulus of trabecular bone E_2, GPa					
	0.25	0.5	1.0	1.5	2.0	2.5
Maximal relative eccentricity for cortical bone e_{1max}/D	0.24	0.24	0.22	0.21	0.21	0.20
Maximal relative eccentricity for trabecular bone e_{2max}/D	0.27	0.26	0.25	0.24	0.23	0.22
Maximal value of eccentricity for cortical bone $e_{1max,}$ mm	9.72	9.4	8.94	8.54	8.20	7.9
Maximal value of eccentricity for trabecular bone $e_{2max,}$ mm	10.8	10.4	9.93	9.48	9.11	8.78

Tab. 2. The calculated magnitudes of dimensional parameters for value of cortical bone elasticity modulus E_1=20 GPa

Identifier of dimensional parameter	Elasticity modulus of trabecular bone E_2,GPa					
	0.25	0.5	1.0	1.5	2.0	2.5
Maximal relative eccentricity for cortical bone e_{1max}/D	0.24	0.23	0.23	0.22	0.21	0.20
Maximal relative eccentricity for trabecular bone e_{2max}/D	0.27	0.26	0.25	0.24	0.23	0.23
Maximal value of eccentricity for cortical bone $e_{1max,}$ mm	9.77	9.5	9.08	8.71	8.40	8.12
Maximal value of eccentricity for trabecular bone $e_{2max,}$ mm	10.9	10.6	10.1	9.68	9.3	9.0

Tab. 3. The calculated magnitudes of dimensional parameters for value of cortical bone elasticity modulus E_1=25 GPa

Name and identifier of dimensional parameter	Elasticity modulus of trabecular bone E_2, GPa					
	0.25	0.5	1.0	1.5	2.0	2.5
Maximal relative eccentricity for cortical bone e_{1max}/D	0.25	0.24	0.23	0.23	0.23	0.21
Maximal relative eccentricity for trabecular bone e_{2max}/D	0.27	0.27	0.26	0.25	0.24	0.23
Maximal value of eccentricity for cortical bone e_{1max}, mm	9.83	9.62	9.25	9.07	8.64	8.40
Maximal value of eccentricity for trabecular bone e_{2max}, mm	10.9	10.7	10.28	10.1	9.6	9.3

While analyzing the findings, we consider that $OK = e$; $OB = OA = R = 2e$ as it appears in Fig. 2b; i.e. maximum permissible distance from the center of section to the fixing screws installation points is equal to the doubled amount of eccentricity.

Since the fixing screws diameter is usually 8 mm (radius is 4 mm), maximum geometrically permissible distance from the center of bone cross-section to the fixing screws installation point without injury of the inner part of cortical layer with the screw thread (parameter *R*max) is *R*max =40/2-2-4=14 mm, providing that cortical layer thickness is 2 mm. Collation of this value with the maximum permissible values calculated on the basis of the data given in the Tabs. 1–3 points to secured provision of conditions for non-opening of junction in case of two-point asymmetric fixation with implant throughout the studied range of bone layers elasticity modules values variation.

The obtained date are based on approximate method of calculation and as result are approximate and estimate. But the accuracy of calculation used is sufficient for position determination of fixing screws' arrangement. More precise analysis of cortical and trabecular bones' stressed state is possible with using of numerical methods, e.g. FEM.

4. CONCLUSION

Methodology is suggested for calculation assessment of cortical and trabecular layer tension parameters in case of two-point asymmetric fixation with an implant for femoral neck fracture. It has been found that solid tightness of the broken parts against each other is secured over the whole fracture section without junction opening under condition that fixing screws are positioned in the trabecular layer without penetration of the thread side surface into cortical layer. This concerns the studied range of mechanical properties change of cortical and trabecular layers of bone tissue.

REFERENCES

1. **Akulich Yu.V., Nyashin Yu.I., Podgayets R.M., Dmitrenok I.A.** (2001), The bone tissue remodeling in the proximal femur under the variations of hip joint daily loading history, *Russian Journal of Biomechanics,* Vol.5, No.1, 12-23.
2. **Akulich, Yu.V. Denisov A.S.** (2008), Adaptivnyje izmenenija svojstv kostnoy tkani fragmentov kosti posle osteosynteza shejki bedra zhestkimi rezbovymi fixatorami, *Mehanika kompositsionnyh materialov i konstruktsyj*, Vol. 14, No.3, 313-331. (in Russian).
3. **Booth KC, Donaldson TK, Dai QG.** (1998), Femoral neck fracture fixation: a biomechanical study of two cannulated screw placement techniques / KC Booth, TK Donaldson, QG Dai, *Orthopedics*, 21, 1173–1176.
4. **Karev D.B.** (2012), Kontseptsyja osteosynteza v lechenii patsyjentov trudosposobnogo vozrasta s perelomom shejki bedrennoj kosti, *ARS MEDICA*, 4, 103-107 (in Russian).
5. **Manniger J., Bosch U., Cserhati P., Fekete K., Kazar G**. (2007), *Internal fixation of femoral neck fractures. An atlas*, Springer, Wien - New York.
6. **Minenkov B.V**., **Stasenko I.V**.(1977), *Prochnost detalej iz plastmass*. Moskva, Mashinostrojenije (in Russian).
7. *Spravochnik po soprotivleniju materialov* (1988) **Vinokurov Ye.F., Balykin M.K., Golubev A.I., et als**. – Minsk, Nauka i Tekhnika (in Russian).
8. **Van C.Mow, Rik Huiskes**. (2005), *Basic Ortopaedic Biomechanics and Mechano-Biology*, 3rd ed., Lipincott Williams @ Vilkins.
9. **Yeremin A.V.** (2010) Prochnostnyje harakteristiki gubchatoj kostnoj tkani krestsa i Lv pozvonka cheloveka //*Ukrainskij medychnyj almanah*, Vol. 13, No. 4, 83-84. (in Russian).

STATIONARY ACTION PRINCIPLE FOR VEHICLE SYSTEM WITH DAMPING

Witold KOSIŃSKI*, Wiera OLIFERUK**

*Polish-Japanese Institute of Information Technology, Computer Science Department, ul. Koszykowa 86, 02-008 Warsaw, Poland
*Kazimierz Wielki University of Bydgoszcz, Institute of Mechanics and Applied Computer Science, ul. Chodkiewicza 30, 85-064 Bydgoszcz, Poland
*, **Białystok University of Technology, Faculty of Mechanical Engineering, Department of Mechanics and Applied Computer Science, ul. Wiejska 45C, 15-351 Białystok, Poland

wkos@pjwstk.edu.pl, w.oliferuk@pb.edu.pl

Abstract: The aim of this note is to show possible consequences of the principle of stationary action formulated for non-conservative systems. As an example, linear models of vibratory system with damping and with one and two degrees of freedom are considered. This kind of models are frequently used to describe road and rail vehicles. There are vibrations induced by road profile. The appropriate action functional is proposed with the Lagrangian density containing: the kinetic and potential energies as well as dissipative one. Possible variations of generalized coordinates are introduced together with a noncommutative rule between operations of taking variations of the coordinates and their time derivatives. The stationarity of the action functional leads to the Euler-Lagrange equations.

Key words: Non-Conservative System, Principle of Stationary Action, Lagrangian, Non-Commutative Rule, Damping

1. INTRODUCTION

Historically, the classical Lagrange and Hamilton's formalisms were formulated for the point mechanics problems. Accordingly, if a dynamical system is described by the vector-valued generalized coordinate $\boldsymbol{q}$ and the Lagrangian $L = T - V$, where T and V are, respectively, the kinetic and potential energy, then one formulates the variational principle of the dynamical system requiring that between all curves $\boldsymbol{q} = \boldsymbol{q}(t)$ in a configuration space $\mathcal{V}$ the actual path (i.e. the solution of the system) is that which makes the action integral

$$I[\boldsymbol{q}] = \int_{t_0}^{t_1} L(\boldsymbol{q}, \dot{\boldsymbol{q}}, t)dt \tag{1}$$

stationary. Taking the first variation $\delta\boldsymbol{q}$ subject to the conditions $\delta\boldsymbol{q}(t_0) = \delta\boldsymbol{q}(t_1) = 0$ the stationarity of the action requires $\delta I = 0$, which is equivalent to the Euler-Lagrange's equation:

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{q}}} - \frac{\partial L}{\partial \boldsymbol{q}} = 0 \tag{2}$$

provided the classical commutative rule:

$$\delta\dot{\boldsymbol{q}} = \frac{d}{dt}\,\delta\boldsymbol{q}, \text{ or written as } \left[\left|\delta, \frac{d}{dt}\right|\right]\boldsymbol{q} = 0 \tag{3}$$

holds. Here the bracket $\left[\left|\delta, \frac{\mathrm{d}}{\mathrm{dt}}\right|\right]$ defines the difference between two compositions of operators $\delta\frac{\mathrm{d}}{\mathrm{dt}} - \frac{\mathrm{d}}{\mathrm{dt}}\delta$, not a vector.

It is well known that the governing equations for non-conservative, mechanical systems, i.e. when dissipative phenomena occur, at the present time cannot be derived from Hamilton's variational principle understood as the requirement: $\delta I[\boldsymbol{q}] = 0$ occupied with (3) (Schechter, 1967). In the present paper we will show that neglecting the commutativity law (3) governing equations of non-conservative system is possible to derive. First in Sec. 2 short review of different approaches is presented together with the main assumption about non-commutativity. As an example, linear models of vibratory system with damping and with one and two degrees of freedom are considered in Sec. 3. This kind of models are frequently used to describe road and rail vehicles under vibrations induced kinematically by road profile. The appropriate action functionals are proposed with the Lagrangian density containing: the kinetic and potential energies as well as dissipative one. Possible variations of generalized coordinates are introduced together with a non-commutative rule between operations of taking variations of the coordinates and their time derivatives. The stationarity of the action functional leads to the Euler-Lagrange equations.

2. NON-COMMUTATIVITY RULE

In order to derive governing equations describing irreversible phenomena using the variational technique some artificial restrictions must be made, concerning the basic rules of variational calculus. A good example is served by the variational principle formulation made in: Biot (1970), Chambers (1956), Kotowski (1989, 1992), Marsden and Hughes (1983), Prigogine and Glandsdorff (1965), Rosen (1954), Schechter (1967), Vujanović and Djuković (1972) and Yang (2010).

Different procedure was given by Vujanović (1971) and applied to governing hyperbolic equation of heat conduction (with finite wave speed), in which the new Lagrangian was proposed with an explicit dependence on time in the form of the exponential term $\exp t/\tau$ appearing as a factor. This term has the power t/τ, where τ is the thermal relaxation time. The corresponding transition to the case of infinite speed of thermal disturbance (parabolic equation) is performed by setting the relaxation time equal to zero.

After his first paper Vujanović has proposed in Vujanović (1974, 1975) the new method for deriving the class of equations describing some physical irreversible processes and based on the

variational principle which has a Hamiltonian structure, and in most cases its form does not differ from that known for conservative systems. However, the crucial assumption of the new method is a non-commutative rule between operations of taking variations of the generalized coordinate (field) and their time derivatives.

In dissipative systems the loss of energy is a crucial effect. Because of this effect a physical process cannot be reversed without change in the environment. Such process is irreversible. Irreversibility means that variation of the quantity involved in description of irreversible process and the variation of its time derivative are related to dissipative mechanism governing the process considered, not the time differentiation. It means that the time derivative $\frac{d}{dt}\delta\boldsymbol{q}$ from one side must be different from the variation of the time derivative of the quantity, i.e. $\delta\frac{d}{dt}\boldsymbol{q}$. From the other side this variation is the dynamic quantity and it should depend on the non-conservative forces (according to Vuhanović (1975)) acting upon the system. We are not going to tamper with the usual notation of the variation of $\delta\boldsymbol{q}$ and the velocity of variation $\frac{d}{dt}\delta\boldsymbol{q}$. These two vectors are regarded as purely kinematic in nature. The vector $\delta\boldsymbol{q}$ means that we consider the *infinitesimal transformation* (i.e. first variation) replacing $\boldsymbol{q}(t,x)$, for example by $\boldsymbol{q}(t,x)+s\boldsymbol{h}(t,x)$, where $\boldsymbol{h}(t,x)$ is an arbitrary differentiable function of t and x, and s is a small parameter passing through zero. Then, from the definition:

$$\delta\boldsymbol{q}(t)=\frac{\partial}{\partial s}\big(\boldsymbol{q}(t)+s\boldsymbol{h}(t)\big)\delta s=\boldsymbol{h}(t)\delta s \tag{4}$$

Hence in this notation we have:

$$\frac{d}{dt}\delta\boldsymbol{q}=\frac{d}{dt}\boldsymbol{h}(t)\delta s \tag{5}$$

Since the vector $\delta\frac{d}{dt}\boldsymbol{q}$ has a purely dynamic character and its form depends on the nature of dissipative (non-conservative) phenomena and forces acting on the body, the infinitesimal transformation replaces $\frac{d}{dt}\boldsymbol{q}(t)$ by $\frac{d}{dt}\boldsymbol{q}(t)+s\boldsymbol{k}(t)$, where $\boldsymbol{k}(t)$ is not arbitrary differentiable function of t. This function, however, may differ from $\frac{d}{dt}\boldsymbol{h}(t,x)$ by a part that is related to the function $\boldsymbol{h}(t,x)$ via some relationship depending on the irreversible phenomena of the system. Hence we can write:

$$\delta\frac{d}{dt}\boldsymbol{q}=\frac{\partial}{\partial s}\left\{\frac{d}{dt}\boldsymbol{q}(t)+s\boldsymbol{k}(t)\right\}\delta s=\boldsymbol{k}(t)\delta s \tag{6}$$

together with:

$$\boldsymbol{k}(t)=\frac{d}{dt}\boldsymbol{h}(t)+\mathbf{H}\left(\boldsymbol{q}(t),\frac{d}{dt}\boldsymbol{q}(t),t\right)\boldsymbol{h}(t) \tag{7}$$

Notice that here we admit the tensor function $\mathbf{H}$. The appearing of the tensor $\mathbf{H}$ is a new fact (cf. Grochowicz and Kosiński (2011), Kosiński and Perzyna, 2012)).

Comparing (4) and (6) with (7) we end up with the following non-commutative rule:

$$\left[\left|\delta,\frac{d}{dt}\right|\right]\boldsymbol{q}=\mathbf{H}\left(\boldsymbol{q}(t),\frac{d}{dt}\boldsymbol{q}(t),t\right)\delta\boldsymbol{q} \tag{8}$$

We can see that the case when the function $\mathbf{H}=0$ corresponds to a commutative rule. This non-commutative rule will be crucial in developing new variation principle for deformable body made of the dissipative material with internal state variables.

In the previous papers (Grochowicz and Kosiński (2011), Kosiński and Perzyna, 2012)) the simpler version of the variational technique developed here was applied to 3 continuous irreversible systems. They were: long-line (i.e. telegraph) equation, hyperbolic model of heat conduction as well as governing equations of deformable body with internal state variables. In the next paper we will generalize the last derivation to the case of thermomechanics.

3. VARIATIONAL PRINCIPLE FOR DYNAMIC SYSTEM WITH DAMPING

Let us consider two linear models of a vehicle (Grzyb, 2012) which moves on a road or on a track. We will consider its vertical movement characterized by the displacement $Z(t)$ and produced by its horizontal motion along the road (track) profile, see Fig. 1. The vertical motion will be induced by kinematic excitation described by function $u(t)$. The car body of the vehicle is modeled by a rigid body of mass m. Vibrations of the mass are forced by a spring of an elastic constant k. The damping of the vehicle is realized by a linear, viscose damper characterized by a viscosity constant B. The first model has one degree of freedom.

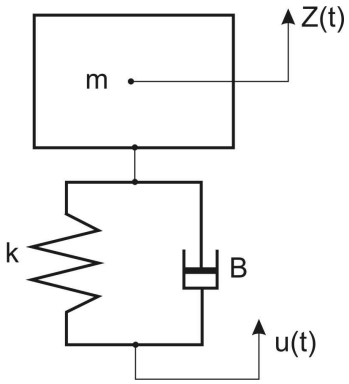


Fig. 1. Physical model of vehicle with one degree of freedom

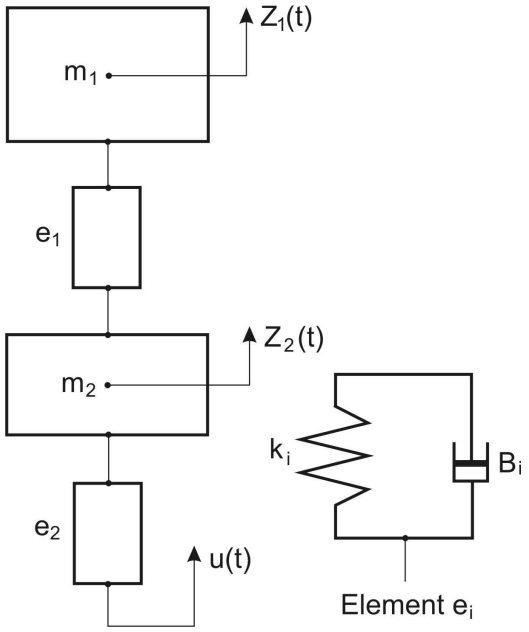


Fig. 2. Physical model of vehicle with two degrees of freedom

In order to take into account two parts of the vehicle, namely the vehicle body and the boogie, the model with two degrees of freedom, see Fig. 2, will be considered. The vibrations of the vehicle body of mass m_1 and the boogie of mass m_2 are characterized by displacements: $Z_1(t)$ and $Z_2(t)$, respectively. The vertical vibration of the both parts are induced by two separate springs with elastic constants: k_1 and k_2. The damping of considered vibrations is realized by linear dampers with viscosity constants B_1 and B_2, compare the element e_i on Fig. 2.

3.1. One degree of freedom model

In the first model the motion of the system (i.e. the car) is governed by the second order ODE (cf. Grzyb (2012)):

$$m\ddot{Z}(t) + B\dot{Z}(t) + kZ(t) = B\dot{u}(t) + ku(t) \tag{9}$$

According to the author of Grzyb (2012) this equation can be derived from the following Lagrange equation of the second kind:

$$\frac{d}{dt}\left(\frac{\partial E_k}{\partial \dot{q}}\right) + \frac{\partial R}{\partial \dot{q}} + \frac{\partial E_p}{\partial q} = Q \tag{10}$$

with the appropriate forms of the kinetic energy E_k, the potential energy E_p, the dissipation function of the system R and the generalized force Q. In Grzyb (2012) in the derivation the author put $q = Z$, $Q = 0$ and suitable forms of other functions. It will be shown, that in the two degrees of freedom system the governing of equations are Euler-Lagrange equations derivable from a stationary action principle.

3.2. Two degree of freedom model

Having the derivation of the first model, let us start with the definition of the kinetic and potential energies for the system together with its dissipative part. Then, assuming the index 1 for the quantities describing the car body, we have the kinetic energy for the whole system: the car body and the boogie $T(\dot{Z}_1, \dot{Z}_2)$:

$$T\left(\dot{Z}_1, \dot{Z}_2\right) = \frac{1}{2}\left[m_1 {\dot{Z}_1}^2 + m_2 {\dot{Z}_2}^2\right] \tag{11}$$

the potential energy $V(Z_1, Z_2)$:

$$V(Z_1, Z_2, u) = \frac{1}{2}\left[k_1(Z_2 - Z_1)^2 + k_2(u - Z_2)^2\right] \tag{12}$$

and the dissipative part $D(\dot{u}, Z_2)$:

$$D(\dot{u}, Z_2) = B_2 \dot{u} Z_2 \tag{13}$$

Let us postulate the following non-commutative rule (cf. (8)):

$$\left[\left|\delta, \frac{d}{dt}\right|\right]\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} H^{11} & H^{12} \\ H^{21} & H^{22} \end{bmatrix}\begin{bmatrix} \delta Z_1 \\ \delta Z_2 \end{bmatrix} = \begin{bmatrix} -\frac{B_1}{m_1} & \frac{B_1}{m_1} \\ \frac{B_1}{m_2} & -\frac{B_1+B_2}{m_2} \end{bmatrix}\begin{bmatrix} \delta Z_1 \\ \delta Z_2 \end{bmatrix} \tag{14}$$

Let us define the action functional with its Lagrangian density $L = T + D - V$:

$$I(Z_1, Z_2, u) = \int_{t_0}^{t_1}\left(T\left(\dot{Z}_1, \dot{Z}_2\right) + D(\dot{u}, Z_2) - V(Z_1, Z_2)\right)dt \tag{15}$$

Now we formulate the following stationary action principle.

Postulate: Let a system will be exited by the kinematic loading u. Along all curves $(Z_1(t), Z_2(t))$ in the configuration space the actual path is that which makes the action integral (15) stationary provided the both variations $\delta Z_1(t)$ and $\delta Z_2(t)$ vanish at the end points t_0 and t_1 and the non-commutativity load (14) holds.

Let us take the first variation of (15). We get:

$$\delta I(Z_1, Z_2, u) = \int_{t_0}^{t_1}\left(\frac{\partial T}{\partial \dot{Z}_1}\delta\dot{Z}_1 + \frac{\partial T}{\partial \dot{Z}_1}\delta\dot{Z}_2\right)dt - \int_{t_0}^{t_1}\left(\frac{\partial (V-D)}{\partial Z_1}\delta Z_1 + \frac{\partial (V-D)}{\partial Z_2}\delta Z_2\right)dt \tag{16}$$

Now we use the rule (14) to $\delta\dot{Z}_1$ and $\delta\dot{Z}_1$, to get:

$$\begin{aligned}\delta I = &\int_{t_0}^{t_1}\left(\frac{\partial T}{\partial \dot{Z}_1}\left(\frac{d\delta Z_1}{dt}\right) + H^{11}\delta Z_1 + H^{12}\delta Z_2\right)dt \\ &+ \int_{t_0}^{t_1}\left(\frac{\partial T}{\partial \dot{Z}_2}\left(\frac{d\delta Z_2}{dt}\right) + H^{21}\delta Z_1 + H^{22}\delta Z_2\right)dt \\ &- \int_{t_0}^{t_1}\left(\frac{\partial (V-D)}{\partial Z_1}\delta Z_1 + \frac{\partial (V-D)}{\partial Z_2}\delta Z_2\right)dt\end{aligned} \tag{17}$$

Taking the time derivative on some terms and using formula of product differentiation as well as grouping similar terms, we obtain:

$$\begin{aligned}\delta I = &\int_{t_0}^{t_1}\left(-\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{Z}_1}\right)\delta Z_1 - \frac{d}{dt}\left(\frac{\partial T}{\partial \dot{Z}_2}\right)\delta Z_2\right)dt \\ &+ \left[\frac{\partial T}{\partial \dot{Z}_1}\delta Z_1 + \frac{\partial T}{\partial \dot{Z}_2}\delta Z_1\right]\Big|\begin{matrix} t = t_1 \\ t = t_0 \end{matrix} \\ &+ \int_{t_0}^{t_1}\left(\frac{\partial T}{\partial \dot{Z}_1}H^{11} + \frac{\partial T}{\partial \dot{Z}_2}H^{21}\right)\delta Z_1 dt \\ &+ \int_{t_0}^{t_1}\left(\frac{\partial T}{\partial \dot{Z}_1}H^{12} + \frac{\partial T}{\partial \dot{Z}_2}H^{22}\right)\delta Z_2 dt \\ &- \int_{t_0}^{t_1}\left(\frac{\partial (V-D)}{\partial Z_1}\delta Z_1 + \frac{\partial (V-D)}{\partial Z_2}\delta Z_2\right)dt\end{aligned} \tag{18}$$

Since our Postulate requires vanishing $\delta I[Z_1, Z_2, u] = 0$ and the variations of Z_1 and Z_2 vanish at the end points, then inside the interval grouping terms appearing in the front of the variations δZ_1 and δZ_2, we obtain, in view of their arbitrariness and independence, two ODE's:

$$\begin{aligned}\frac{d}{dt}\frac{\partial T}{\partial \dot{Z}_1} + \frac{\partial (V-D)}{\partial Z_1} - \frac{\partial T}{\partial \dot{Z}_1}H^{11} - \frac{\partial T}{\partial \dot{Z}_2}H^{21} = 0 \\ \frac{d}{dt}\frac{\partial T}{\partial \dot{Z}_2} + \frac{\partial (V-D)}{\partial Z_1} - \frac{\partial T}{\partial \dot{Z}_1}H^{12} - \frac{\partial T}{\partial \dot{Z}_2}H^{22} = 0\end{aligned} \tag{19}$$

The above equations form the Euler-Lagrange equations of our stationary action functional Postulate.

We substitute the expressions for the ***H***'s coefficients from (14). We obtain:

$$m_1\ddot{Z}_1 + B_1\left(\dot{Z}_1 - \dot{Z}_2\right) + k_1(Z_1 - Z_2) = 0$$
$$m_2\ddot{Z}_2 - B_1\dot{Z}_1 + (B_1 + B_2)\dot{Z}_2 - k_1(Z_1 - Z_2) + k_2Z_2 = k_2u + B_2\dot{u} \quad (20)$$

The above method can be applied to the previous model with one degree of freedom system (9) by putting:

$$V(Z) = \frac{1}{2}m\dot{Z}^2$$
$$V(Z,u) = \frac{1}{2}k(u - Z)^2 \quad (21)$$
$$D(\dot{u}, Z) = B\dot{u}Z$$

and the non-commutativity rule:

$$\left[\left|\delta, \frac{d}{dt}\right|\right] Z = -\frac{B}{M} \quad (22)$$

4. CONCLUSIONS

In this paper starting with the action functional in which kinetic energy, potential energy and dissipative part of energy related to both elastic and irreversible phenomena of the system appear, the principle of its stationarity is formulated. Making the first variation of the functional compatible with the assumed initial and boundary conditions and postulating the appropriate non-commutative laws between the variation and time differentiation operators, the consequence in the form of Euler-Lagrange equations is obtained. The derived equations form the governing system of equations of the model with two degree of freedom. The passage to the system with one degrees of freedom is also shown.

The formulated method may be very helpful in other derivations and the investigation of particular and approximate forms of solutions of non-conservative systems.

REFERENCES

1. **Biot M. A.** (1970), *Variational principles in heat transfer*, Oxford Mathematical Monographs, Oxford Univ.
2. **Chambers I. G.** (1956), A variational principle for the conduction of heat., *Q. J. Mech. Appl. Math*., IX (2), 234–235.
3. **Grochowicz B., Kosiński W.** (2011), Lagrange's method for derivation of long line equations, *Acta Technica*, 56 (1), 331–341.
4. **Grzyb A.** (2012), Optimal parameters selection of vibration dampers of dynamic vehicles system, *Polioptymization and computer aided design* (in Polish), Mielno, Kiczkowiak T., Tarnowski W. (red.), Wydaw. Uczelniane Politechniki Koszalińskiej, Koszalin, 20–31.
5. **Kosiński W., Perzyna P.** (2012), On consequences of the principle of stationary action for dissipative bodies, *Arch. Mech.*, 64 (1), 95–106.
6. **Kotowski R.** (1989), On the Lagrange functional for dissipative processes, *Arch. Mech.*, 41 (4), 571–587.
7. **Kotowski R.** (1992), Hamilton's principle in thermodynamics, *Arch. Mech*., 44 (2), 203–215.
8. **Lebon G., Lambermont J.** (1973), Generalization of Hamilton's principle to continuous dissipative systems, *J. Chem. Phys*., 59, 2929–2936.
9. **Marsden J. E., Hughes T. J. R**. (1983), *Mathematical Theory of Elasticity*, Prentice-Hall, Englewood Cliffs, New York,
10. **Prigogine I., Glansdorff P.** (1965), Variational properties and fluctuation theory, *Physica*, 31 (8), 1242–1256.
11. **Rosen P.** (1954), Use of restricted variational principles for the solution of differential equations, *J. Appl. Phys*., 25, 336–338.
12. **Schechter R. S.** (1967), *The Variational Methods in Engineering*, McGraw-Hill, New York.
13. **Vujanović B.** (1971), An approach to linear and non-linear heat transfer problem using a Lagrangian, *A. I. A. A. Journal*, 9(1), 131–134.
14. **Vujanović B.** (1974), On one variational principle for irreversible phenomena, *Acta Mechanica*, 19, 259–275.
15. **Vujanović B.** (1975), A variational principle for nonconservative dynamical systems, *ZAMM – Zeitschrift f¨ur Angewandte Mathematik und Mechanik*, 55(6), 321–331.
16. **Vujanović B., Djukiić D.** (1972), On one variational principle of Hamilton's type for nonlinear heat transfer problem, *International Journal of Heat and Mass Transfer*, 15, 1111–1123.
17. **Yang Q.** (2010), Hamilton's principle for Green-inelastic bodies, *Mechanical Research Communications*, 37, 696–699.

FREQUENCY ANALYSIS WITH CROSS-CORRELATION ENVELOPE APPROACH

Adam KOTOWSKI*

*Faculty of Mechanical Engineering, Bialystok University of Technology, ul. Wiejska 45C, 15-351 Bialystok, Poland

a.kotowski@pb.edu.pl

Abstract: A new approach for frequency analysis of recorded signals and readout the frequency of harmonics is presented in the paper. The main purpose has been achieved by the cross-correlation function and Hilbert transform. Using the method presented in the paper, there is another possibility to observe and finally to identify single harmonic apart from commonly used Fourier transform. Identification of the harmonic is based on the effect of a straight line of the envelope of the cross-correlation function when reference and signal harmonic have the same frequency. This particular case is the basis for pointing the value of the frequency of harmonic component detected.

Key words: Frequency Analysis, Cross-Correlation, Hilbert Transform, Envelope

1. INTRODUCTION

It is common knowledge that spectrum analysis using fast Fourier transform (FFT) presents the amplitudes of all harmonics the fast way. This method of showing the frequency profile of the signal is applied both during the post-processing and as real-time processing.

There is many engineering applications of correlation function (Bendat and Piersol, 1980). To provide for a new application, the cross-correlation function has been utilized to correlate real-measured signal and a single harmonic signal generated by a software. Also, the Hilbert transform has been used for obtaining the envelope of the cross-correlation function (Thrane, 1984) where the envelope removes the oscillations (Thrane et al., 1999). In particular cases, experimental results have shown a linear shape of the envelope. It is observed when correlated signals have a common frequency value (Kotowski, 2010). This effect is well noted and very sensitive to generated single harmonic signal frequency. Thus, the paper presents the method of reading the particular frequency harmonic developed on the base of cross-correlation function and its envelope.

It is obviously known that after signal recording there is no way to have the longer one. This case couses the fixed frequency resolution as inverse of period of signal duration when using FFT. This case is especially noted for very short-time signals, e.g., from impulse tests. For avoiding that limitation, Cawley and Adams (1979) investigated the problem mentioned aboved and showed to be possible to obtain frequency resolution of one-tenth of the spacing between the frequency points produced by the Fourier transform. Also, it is commonly used zero-padding for improving frequency estimation (Quinn, 2009; Dunne, 2002). Zero-padding means that an array of zeros is appended to the end (or beginning) of analysed signal. Using the method presented in the paper there also is possible to obtain different frequency resolution than that fixed using FFT. Frequency resolution can be variable adjusted by user of the method starting from reference value of 1 Hz to up or down.

2. METHODOLOGY

The cross-correlation function $R_{xy}(\tau)$ between two processes, $x(t)$ and $y(t)$, is calculating by the expression (Bendat, Piersol, 1980):

$$R_{xy}(\tau)=\lim_{T\to\infty}\frac{1}{T}\int_0^T x(t)y(t+\tau)\,dt \tag{1}$$

where: T – signal record length, τ – argument of cross-correlation function (time delay).

Then, the cross-correlation function $R_{xy}(\tau)$ is transformed into the envelope by Hilbert transform. The Hilbert transform of a real time signal, $x(t)$, is defined as follows (Thrane, 1984)

$$\mathrm{H}\left[x(t)\right]=\widetilde{x}(t)=\frac{1}{\pi}\int_{-\infty}^{\infty}x(\tau)\frac{1}{t-\tau}\,d\tau \tag{2}$$

Thus, the Hilbert transform of the cross-correlation $R_{xy}(\tau)$ is given by:

$$\mathrm{H}\left[R_{xy}(\tau)\right]=\widetilde{R}_{xy}(\tau)=\frac{1}{\pi}\int_{-\infty}^{\infty}R_{xy}(\tau')\frac{1}{\tau-\tau'}\,d\tau' \tag{3}$$

The Hilbert transform enables calculation of the envelope of the signal $x(t)$ as follows (Thrane et al., 1999):

$$|\overset{\nabla}{x}(t)|=\sqrt{x^2(t)+\widetilde{x}^2(t)} \tag{4}$$

where $|\overset{\nabla}{x}(t)|$ is the envelope. Similarly, we can calculate the envelope of the function $R_{xy}(\tau)$ as:

$$|\overset{\nabla}{R}_{xy}(\tau)|=\sqrt{{R_{xy}}^2(\tau)+\widetilde{R}_{xy}{}^2(\tau)} \tag{5}$$

The method also needs series of harmonic signals generated as follows:

$$y_i = \sin(2\pi \cdot (f_s + w \cdot i) \cdot t) \tag{6}$$

where: i - an integer value (index), f_s – starting frequency, w – factor as frequency resolution parameter.

Frequency f_s is fixed as a start point value and is increasing by i = 1, 2, 3, … , n. Also, the factor w is applied for changing the resolution of the harmonic frequency reading. This way, a form of the envelope points the case of detection and finally identification of harmonic. The harmonic frequency value equals one of the harmonics existing within the signal y_i(t). Preliminary studies have shown that envelope of the cross-correlation function is in the form of a straight line when input signal x(t) and the signal y(t) have in common one frequency determined as (f_s+w·i). This phenomenon is easy to detect and determination of the common frequency is fast. For that reason, plot of the envelope can be effectively used to identify harmonics included in recorded signals without Fourier transform.

There are a hundred of plots of the cross-correlation function envelope to illustrate four particular cases of the straight line effect mentioned above in Fig. 1. It has been used the four-harmonic signal x(t) generated as follows

$$x(t) = \sum_{k=1}^{4} \sin(2\pi \cdot f_k \cdot t) \tag{7}$$

where: f_1=64 Hz, f_2=85 Hz, f_3=130 Hz, f_4=150 Hz. The signal y_i is calculated in the way determined in Eq. 6, where f_s and w are constant and equal 60.0 and 1.0 respectively. The index i varies in the range from 1 to 100.

The value of frequency of harmonic included in the input signal x(t) is determined on the base of the plot of the envelope. When observing straight-line effect, we know the f_s value, w value and the i index value of the signal y_i(t) which was used for calculations (Eq. 6). This way, a formula (f_s +w·i) indicates the frequency of recognized harmonic.

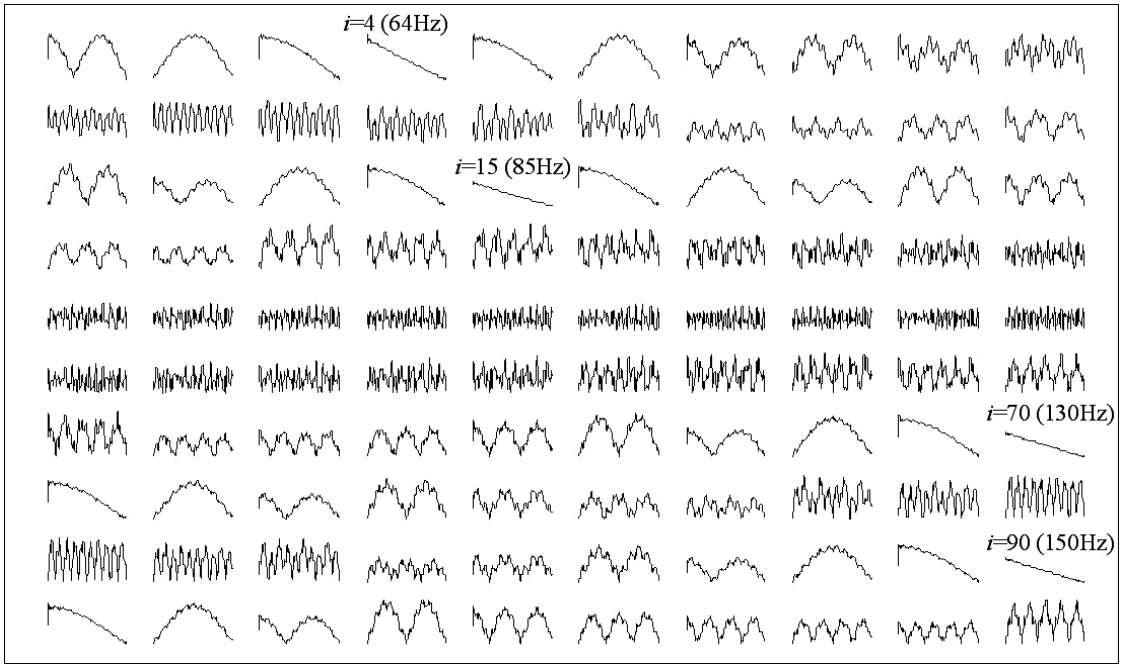


Fig. 1. The cross-correlation envelopes

3. RESULTS FOR STATIONARY SIGNAL

The exemplary experimental results have been based on signal of vibration presented in Fig. 2. The signal has been recorded by sampling frequency of 4096 Hz and over time of one second. The spectrum shown in Fig. 3 has a lot of well-observed harmonics. As shown in Fig. 4, two cases have been detected between 210 Hz and 310 Hz where envelopes of the cross-correlation function are almost in the form of a straight line. Thise situation has occurred for i=17 and i=92 by f_s=210Hz and w=1 (Eq. 6). Hence, it has been for 227 and 302 Hz with frequency resolution fixed by w as 1Hz (w=1).

Apart from detection based on cross-correlation envelope image, an indicator L_e has been used to express in numbers deviation of cross-correlation envelope from linearity. This way, it was possible to present a plot of changes in straight line overlay for all frequency span of recorded vibration signal. L_e is described as follows

$$L_e = \sum_{n=1}^{N} \left| y_{ref} - y_{env} \right| \tag{8}$$

where: y_{ref} – reference straight line, y_{env} – cross-correlation envelope, N – number of points for calculation.

This way, a plot of changes of indicator L_e has been prepared and presented in Fig. 5. It seems to be no difference between spectrun presented in Fig. 3 and the plot of L_e but if zooming the plot there are local minimas in places of dominant frequency appearance (spectrum peaks). If having the plots of L_e, it is possible to readout frequencies being under consideration (Figs. 6-9).

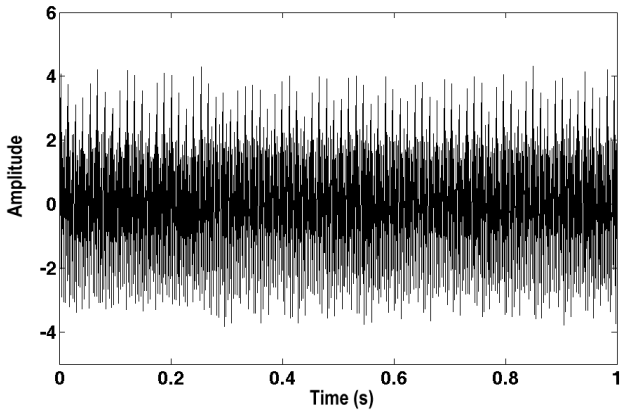


Fig. 2. Signal of vibration

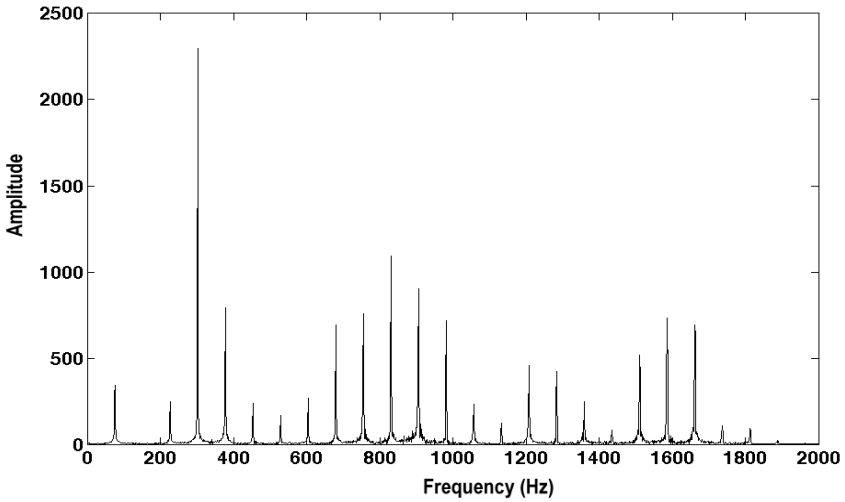


Fig. 3. Signal spectrum

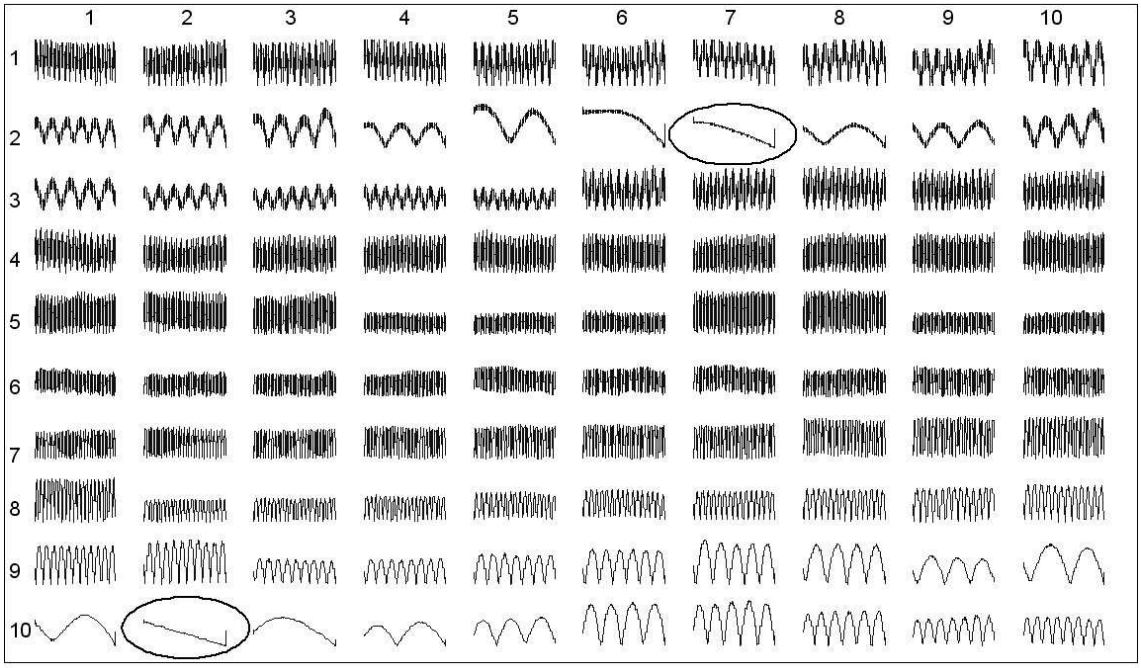


Fig. 4. Envelopes of the cross-correlation functions

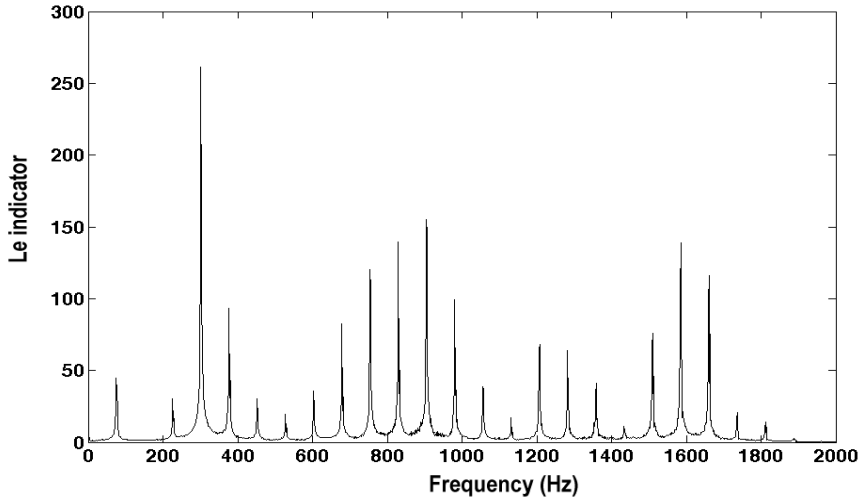


Fig. 5. L_e indicator plot for all frequency span of recorded vibration signal

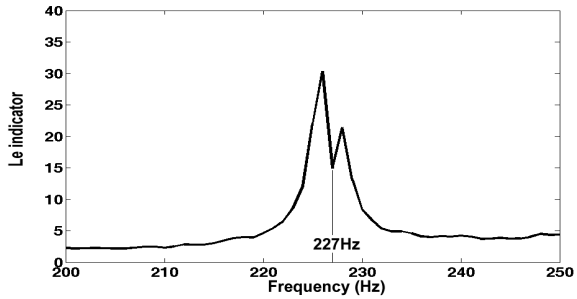


Fig. 6. Enlargement of L_e indicator plot around 227 Hz

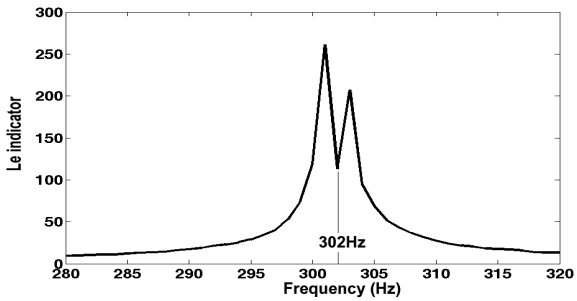


Fig. 7. Enlargement of L_e indicator plot around 302 Hz

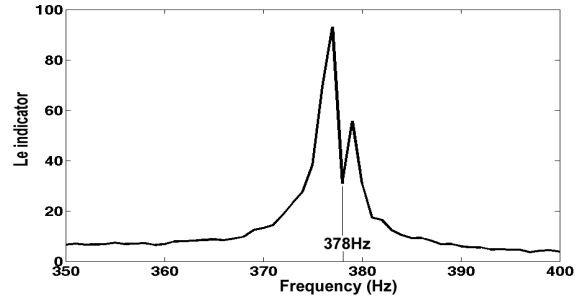


Fig. 8. Enlargement of L_e indicator plot around 378 Hz

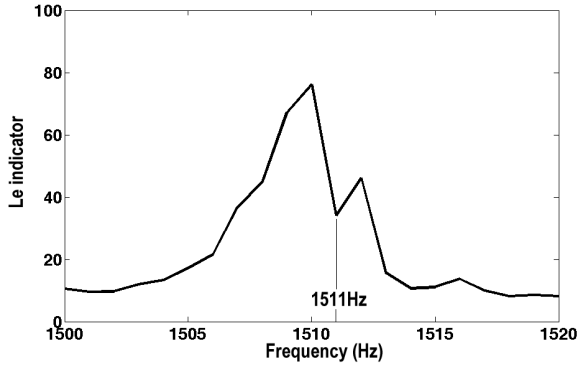


Fig. 9. Enlargement of L_e indicator plot around 1511 Hz

4. RESULTS FOR NONSTATIONARY SIGNAL

Frequency identification presented in section 3 can be applied for nonstationary signals where commonly used Fourier transform relies on a stationarity assumption and it is difficult to guarantee, in practice, the stationarity over a long signal time horizon (Benko and Juričić, 2008). A typical nonstationary signal can be the signal of response from impulse test. Exemplary impulse response under consideration is shown in Fig. 10. Laboratory software tool using curvefitting procedure have been utilized to obtain values of two frequencies at two highest amplitudes. It have resulted the frequency of 3794 and 13714 Hz.

In this case, impulse response analysis have shown that frequency readout is based on some different form of L_e indicator plot than obtained for vibration signal in section 3. It is well-observed an push-up effect presented in Figs. 11-12. This effect revealed the frequency of harmonics really existing in impulse response, i.e., 3794 and 13714 Hz.

Results presented previously have been obtained by the frequency resolution of 1 Hz. By changing the w parameter, it is possible to get greater resolution e.g. 0.2 Hz (w=0.2). As show in Fig. 13, the plot of L_e indicator has the same character but readout of frequency is more exact. In this case it is 3794.2 Hz – frequency at maximum of push-up effect.

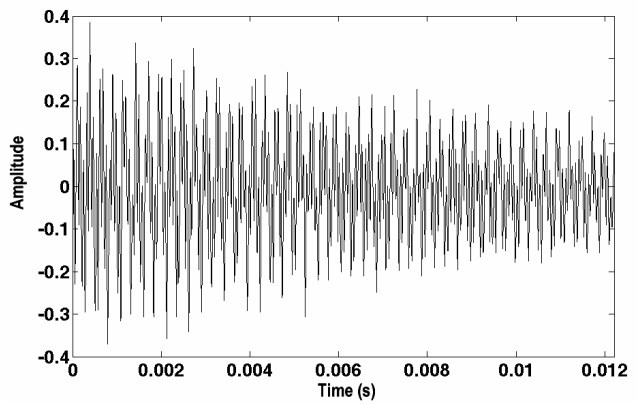


Fig. 10. Impulse response

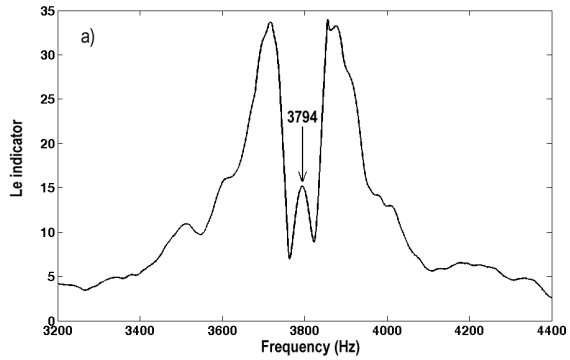


Fig. 11. L_e indicator plot for impulse response around 3794 Hz

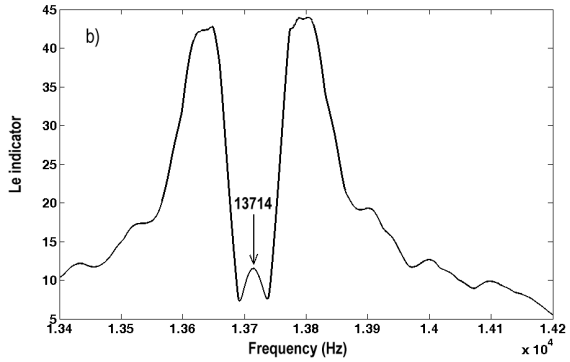


Fig. 12. L_e indicator plot for impulse response around 13714 Hz

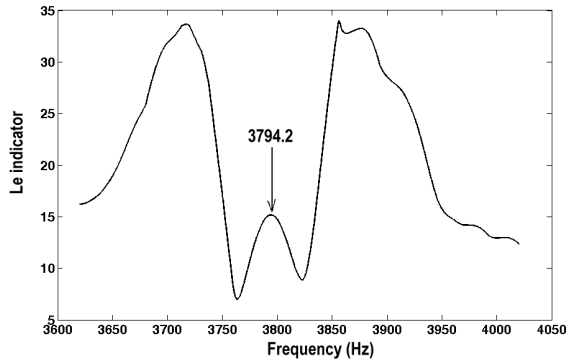


Fig. 13. L_e indicator plot for impulse response around 3794 Hz by the frequency resolution of 0.2 Hz

It is also able to obtain quasi-stationary signal from nonstationary by deviding the nonstationary signal into several sections and then use FFT. But this way, the Fourier spectrum resolution is going down. By deviding the signal presented in Fig. 10 into two parts, the spectrum resolution equals 164Hz (duration is 6.10 milisecond). After splitting into four sections, the spectrum resolution equals 328Hz (duration is 3.05 milisecond). However using the indicator L_e the method have its own spectrum resolution independent of duration of analysed signal or a fragment of analysed signal.

5. CONCLUSIONS

A general view of the use of cross-correlation function and its envelope for frequency analysis has been presented in the paper. That approach brings in the method for reading the frequency both for stationary and for nonstationary signals. For stationary signals, new possibility is based on the cross-correlation envelope straight-line effect observed for two signals (input signal and reference signal) when having one harmonic in common. The approach proposed in the paper shows a possibility to detect and finally to identify frequencies being within the input signal without use of Fourier transform, thus, without limitation in frequency resolution. The frequency resolution of proposed frequency analysis is determined over the factor used for generating reference signal. The method proposed in the paper gives a possibility to have the spectrum resolution controlled and independent of period of signal recording, e.g. signals lasting much less than one second always have Fourier spectrum resolution much over than 1 Hz and using the proposed methos it is able to obtain spectrum resolution 1Hz or even below 1Hz.

The cross-correlation function and its envelope can be a complementary method for frequency analysis, e.g. for accurate detection of natural frequencies using impulse tests.

REFERENCES

1. **Bendat J.S., Piersol A.G.** (1980), *Engineering applications of correlation and spectral analysis*, Wiley-Interscience, New York.
2. **Benko U., Juričić Đ.** (2008), Frequency analysis of noisy short-time stationary signals using filter-diagonalization, *Signal Processing*, 88, 1733–1746.
3. **Cawley P., Adams R.D.** (1979), Improved frequency resolution from transient tests with short record lengths, *Journal of Sound and Vibration*, 64 (1), 123-132.
4. **Dunne J.F.** (2002), A fast time-domain integration method for computing non-stationary response histories of linear oscillators with discrete-time random forcing, *Journal of Sound and Vibration*, 254 (4), 635-676.
5. **Kotowski A.** (2010), Reading the frequency of harmonics by cross-correlation function and its envelope, *Proc. 6th Int. Conf. Mechatronic Systems and Materials*.
6. **Quinn B.G.** (2009), Recent advances in rapid frequency estimation, *Digital Signal Processing*, 19, 942-948.
7. **Thrane N.** (1984), The Hilbert Transform, *Technical Review*, 3, Brüel&Kjær, Naerum, Denmark.
8. **Thrane N., Wismer J., Konstantin-Hansen H., Gade S.** (1999), *Practical use of the "Hilbert transform"*, Application Note, 0437, Brüel&Kjær, Naerum, Denmark.

This work was supported by Bialystok University of Technology under work No. S/WM/1/2012.

ROTOR CRACK DETECTION APPROACH USING CONTROLLED SHAFT DEFLECTION

Zbigniew KULESZA*

*Department of Automatic Control and Robotics, Faculty of Mechanical Engineering, Bialystok University of Technology, ul. Wiejska 45C, 15-351 Białystok, Poland

z.kulesza@pb.edu.pl

Abstract: Rotating shafts are important and responsible components of many machines, such as power generation plants, aircraft engines, machine tool spindles, etc. A transverse shaft crack can occur due to cyclic loading, creep, stress corrosion, and other mechanisms to which rotating machines are subjected. If not detected early, the developing shaft crack can lead to a serious machine damage resulting in a catastrophic accident. The article presents a new method for shaft crack detection. The method utilizes the coupling mechanism between the bending and torsional vibrations of the cracked, non-rotating shaft. By applying an external lateral force of a constant amplitude, a small shaft deflection is induced. Simultaneously, a harmonic torque is applied to the shaft inducing its torsional vibrations. By changing the angular position of the lateral force application, the position of the deflection also changes opening or closing of the crack. This changes the way the bending and torsional vibrations are being coupled. By studying the coupled lateral vibration response for each angular position of the lateral force one can assess the possible presence of the crack. The approach is demonstrated with a numerical finite element model of a rotor. The results of the numerical analysis demonstrate the potential of the suggested approach for effective shaft crack detection.

Key words: Rotordynamics, Shaft Crack, Structure Health Monitoring, Diagnosis

1. INTRODUCTION

One of the most dangerous malfunctions of rotating machines are shaft cracks. Transverse cracks occur due to cyclic loading, thermal stresses, creep, corrosion, and other mechanisms to which rotating shafts are subjected. Once a crack has appeared, high stresses develop at its edge and allow the crack to propagate deeper, even if external loads are not changing. When the crack has propagated to a relevant depth, the propagation speed increases dramatically and the shaft may fail in a very short time, what usually leads to a catastrophic accident. That is why an early detection of the potential shaft cracks inside the rotating machine components is so important.

The problems of early shaft crack detection and warning have been in the limelight of many research centers for over 40 years. Different methods have been analyzed, tested and validated experimentally. Generally, the developed approaches can be divided into the vibration based methods and other methods (e.g. ultrasonic, eddy current testing, dye penetrant testing, etc.) (Bachschmid et al., 2010).

Usual crack detection methods are based on vibration signal analysis (Bently and Muszynska, 1986; Gasch, 1993; Grabowski, 1982) for which dynamic signal analyzers, evaluating the fast Fourier transform (FFT) are utilized. By studying the changes in the vibration spectra, the appearance of the possible shaft crack can be easily assessed. The frequently discussed changes in frequency spectra induced by a crack are: a considerable increase of the amplitude of the synchronous frequency 1X and an appearance of its second multiple 2X, especially for a rotor speed near the half of the critical frequency (Bachschmid et al. 2010). However, such symptoms are characteristic not only for cracked rotors, but can be induced by other faults such, as: bearing malfunctions, misalignment, thermal sensitivity, etc. (Bently and Muszynska, 1986).

Other vibration based methods include changes in rotor modal parameters, such as its natural frequencies and mode shapes, which appear in the presence of the developing shaft crack (Bachschmid et al., 2000, 2010).

Nowadays, model-based methods are gaining a special interest. A mathematical model of the analyzed rotor is extensively used here for designing state observers, Kalman filters or the so called robust fault detection filters, which have proved their efficiency not only for shaft crack detection, but also for the determination of its location along the shaft axis (Bachschmid et al., 2000; Isermann, 2005; Kulesza and Sawicki, 2010).

Methods utilizing new signal processing algorithms, such as neural networks, genetic algorithms, wavelets, Huang-Hilbert transform, etc. are also progressing quickly (Guo and Peng, 2007; He at al., 2001; Litak and Sawicki, 2009).

A relatively new approach employs the use of a specially designed diagnostic force applied to the rotating shaft (Ishida and Inoue, 2006; Mani et al., 2005; Sawicki and Lekki, 2008). If the force is harmonic, then the presence of the crack generates responses containing frequencies at combinations of the angular speed, applied forcing frequency, and the rotor natural frequencies. It has been shown, that the appearance of the combinational frequencies is a very strong signature of the shaft crack (Sawicki et al., 2011). However, the research conducted so far has focused on applying the harmonic force, acting in one, fixed direction only.

A well known feature of the cracked shaft is the coupling between the lateral and torsional vibrations. The appearance of coupled bending and torsional vibrations can be utilized as a possible shaft crack indicator, which has been reported by several authors (Darpe et al., 2004; Kiciński, 2005).

Similarly to the previous methods, the present paper recommends the use of an additional diagnostic force applied perpen-

dicularly to the shaft axis. However, the shaft is not rotating, but excited by an additional torque inducing its torsional vibrations. The proposed method is based on vibration signal analysis, namely on the coupling mechanism between the lateral and torsional vibrations.

2. THE CONCEPT OF THE NEW METHOD FOR ROTOR CRACK DETECTION

Schematic diagram explaining the concept of the proposed method is shown in Fig. 1.

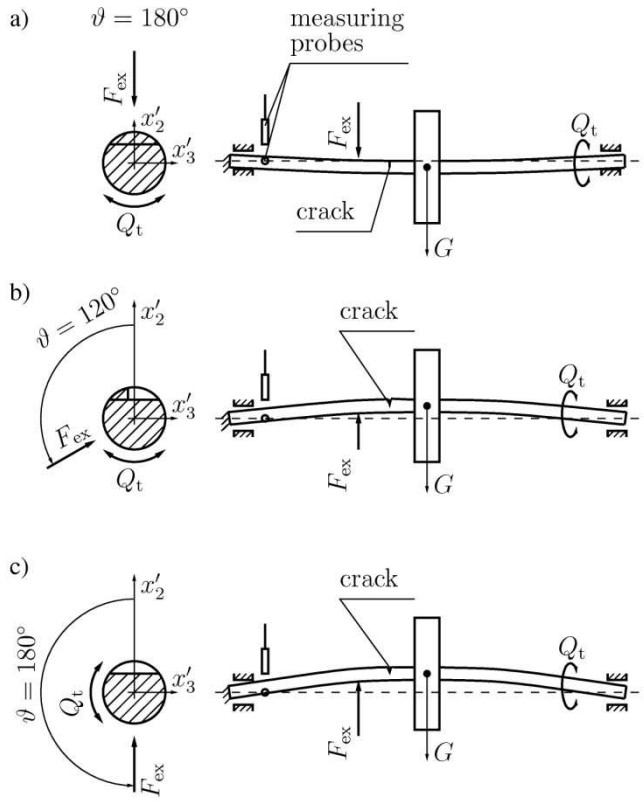


Fig. 1. Schematic diagram of the method for different angular positions ϑ of the external force F_{ex}: a) $\vartheta = 0°$ – fully closed crack, b) $\vartheta = 120°$ – partially open crack, c) $\vartheta = 180°$ – fully open crack

The rotor supported by bearings is not rotating, as one of its ends is fixed to an unmovable base, removing its rotational degree of freedom. The other end is twisted by torque Q_t acting around the axis of the shaft. The amplitude of the torque changes harmonically inducing forced torsional vibrations of the shaft. Simultaneously, an external force F_{ex} of a constant amplitude is applied perpendicularly to the shaft. The force is applied at different angles ϑ, inducing some small deflections of the shaft. By changing the angular position of the force, the position of the deflection also changes opening or closing the crack. This changes the stiffness of the shaft and the way the bending and torsional vibrations are being coupled. It is supposed, that by studying the coupled bending vibration response for each angular position of the external force one will be able to assess the possible presence of the crack.

The suggested method will be tested numerically. For this, the following mathematical models will be formulated: the finite element (FE) model of the rotor, the model of the shaft element with the crack, and the model of crack opening/closing. Based on these models the vibration responses of the cracked rotor for different values and angles of the lateral force as well as for different amplitudes and frequencies of the torsional excitation will be calculated. The Fourier spectra of the vibration responses obtained for both the cracked and uncracked rotors will be used for the comparative study assessing the possible employment of the proposed method for an efficient shaft crack detection.

3. FINITE ELEMENT MODEL OF THE ROTOR

Fig.2 presents the finite element model of the tested rotor.

The rotor consists of a shaft of diameter 16 mm and length 600 mm, and a rigid disk of diameter 120 mm and widtt 30 mm. Two ball bearings located 30 mm from both ends of the shaft are used to support the rotor. Radial stiffness and damping coefficients of the bearings are assumed as $k_b = 3.4 \times 10^6$ N/m and $d_b = 10$ Ns/m. Furthermore, the torsional stiffness and damping coefficients at the left bearing are chosen to be $k_t = 4 \times 10^4$ Nm/rad and $d_t = 20$ Nms/rad, as the left end of the shaft is fixed (Fig. 1). The rotor is made of steel of Young's modulus $E = 2.08 \times 10^{11}$ Pa, Poisson's ratio $v = 0.3$ and density $\rho = 7850$ kg/m^3.

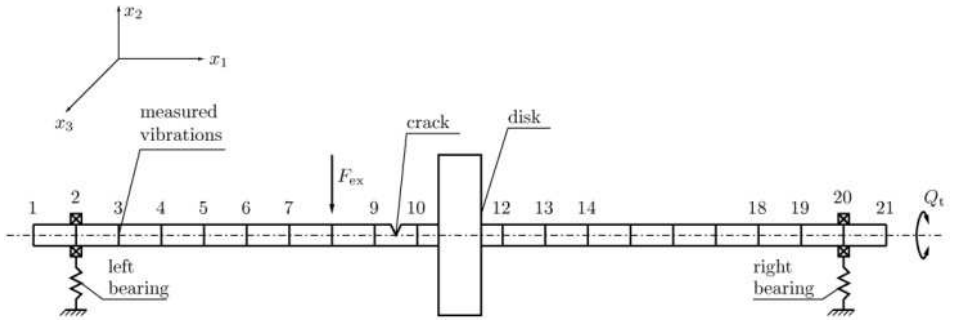


Fig. 2. Finite element model of the tested rotor

The shaft has been divided into 20 finite beam elements (Fig. 2). The 9th element has been assumed as cracked (see: section 4). The bearings are located at the 2nd and 20th node. The external force F_{ex} deflecting the shaft and the additional torque Q_t inducing the torsional vibrations of the rotor are applied at the 8th and the 21st nodes, respectively. The vibration response of the rotor is measured at the 3rd node; bending (along axes x_2 and x_3) and torsional (around axis x_1) vibrations are registered.

Usually, the motion of the rotor is considered in two separate coordinate systems: global (stationary) and local (rotating with a constant angular speed Ω. For the non-rotating rotor fixed with its end to the basis and oscillating around its axis (Fig. 1), only one stationary coordinate system $x_1x_2x_3$ has been assumed, as it is shown in Fig. 3.

Using the finite element method, the motion equations of the rotor can be presented in the following form (Gawroński et al., 1984):

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{D}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{G} + \mathbf{F}_{\text{ex}} + \mathbf{Q}_{\text{t}} \tag{1}$$

where $\mathbf{M}$ is the mass matrix including the masses and mass moments of inertia of shaft finite elements, rigid disks, etc., $\mathbf{D}$ is the damping matrix and $\mathbf{K}$ is the stiffness matrix (including the stiffness of the cracked shaft finite element). The gyroscopic matrix is not included, as the rotor is not rotating.

Vector $\mathbf{q}$ defines the generalized coordinates of the nodes of the finite element mesh discretizing the shaft. This vector consists of N 6-element sub-vectors, where N is the number of nodes. First three components of each sub-vector are displacements along axes x_1, x_2, x_3, the next three are rotation angles around these axes.

$\mathbf{G}$, $\mathbf{F}_{\text{ex}}$ and $\mathbf{Q}_t$ are vectors of the following generalized forces: gravity, external force perpendicular to the rotor axis, and external torque inducing the oscillations of the rotor.

Mass and stiffness matrices are assembled using the corresponding mass and stiffness sub-matrices of the shaft finite elements, rigid disks, bearings, etc. The damping matrix is usually calculated as a linear combination of the mass and stiffness matrices (the Rayleigh damping). The sub-matrices for rotor elements are given in Appendix A.1. The stiffness matrix for the cracked shaft finite element is discussed in the next sections of this paper.

4. MODEL OF THE CRACK

Usually the crack is modeled by local shaft stiffness changes resulting from the constant opening and closing of the crack. This periodic opening and closing of the crack due to the rotation of the shaft is called the *breathing mechanism*. The first models of the crack accounted for the breathing behavior with only two states, i.e., fully open and fully closed at certain angular position (Gasch, 1993; Grabowski, 1982). These models are defined as *hinge models*. Mayes and Davies (1984) developed a similar model except that the transition from fully open to fully closed is governed by a cosine function depending on shaft rotation angle. Progressive development of the finite element method and its application for rotor dynamics (Nelson and McVaugh, 1976) resulted in more or less complicated models of a variable stiffness cracked shaft finite element. Dimarogonas and Paipetis (1983) derived a full stiffness matrix for a transverse open surface crack on a shaft. Darpe et al. (2004) provided more detail and complete derivations of the flexibility matrix of a cracked rotor segment starting from Castigliano's theorem. They introduced an original model of the crack breathing mechanism, in which the extent of crack opening is determined by calculating the values of compressive stresses at the crack edge.

In the model introduced by Mayes and Davies (1984) shaft stiffness reduction $\Delta\mathbf{K}_{\text{c}}$ for the fully open crack is represented by reductions ΔJ_2, ΔJ_3 of the second moments of area of the shaft cross section around axes x_2 and x_3 at the location of the crack. Different authors (Mayes and Davies, 1984; Sinou and Lees, 2005) provide different formulas for ΔJ_2, ΔJ_3 as the functions of crack depth μ. Here, the relative crack depth μ is defined as $\mu = a/(2R)$, where a is the absolute crack depth and R is the shaft radius (Fig. 3b)).

Consider, for example, the paper of Sinou and Lees (Sinou and Lees, 2005) where:

$$\Delta J_3 = \frac{R^4}{4}\left[(1-\mu)(1-4\mu+2\mu^2)\sqrt{2\mu-\mu^2} + \cos^{-1}(1-\mu)\right]$$
$$\Delta J_2 = \Delta\tilde{J}_2 - \tilde{A}\tilde{X}^2 \tag{2}$$

and

$$\Delta\tilde{J}_2 = \frac{\pi R^4}{4} + R^4\left[\frac{2}{3}(1-\mu)(2\mu-\mu^2)^{3/2} + \frac{1}{4}(1-\mu)(1-4\mu+2\mu^2)\sqrt{2\mu-\mu^2} + \sin^{-1}(\sqrt{2\mu-\mu^2})\right]$$

$$\tilde{A} = R^2\left[(1-\mu)\sqrt{2\mu-\mu^2} + \cos^{-1}(1-\mu)\right]$$

$$\tilde{X} = \frac{2}{3\tilde{A}}R^3(2\mu-\mu^2)^{3/2}$$

After shaft stiffness reduction $\Delta\mathbf{K}_{\text{c}}$ is determined, stiffness matrix $\mathbf{K}_{\text{c}}$ of the cracked element is calculated, as follows (Mayes and Davies, 1984):

$$\mathbf{K}_{\text{c}} = \mathbf{K}_0 - f(\psi)\Delta\mathbf{K}_{\text{c}}, \tag{3}$$

where $\mathbf{K}_0$ is the stiffness matrix of the shaft element with no crack, and $f(\psi)$ is the so called *crack steering function*. Depending on the crack model assumed, the crack steering function $f(\psi)$ takes different forms, e.g. for the hinge model:

$$f(\psi) = \begin{cases} 0, \text{ for } \psi < 0 \\ 1, \text{ for } \psi \geq 0 \end{cases} \tag{4}$$

and for the Mayes and Davies model:

$$f(\psi) = \frac{1}{2}(1-\cos\psi) \tag{5}$$

The argument of these functions is the so called *shaft torsional angle* ψ, or for the simplified models, for which *weight dominance* is assumed, it is the *shaft rotation angle* $\Phi = \Omega t$.

For $f(\psi) = 0$ the crack is fully closed and the stiffness of the cracked element is the same as the stiffness of the uncracked element, i.e. $\mathbf{K}_{\text{c}} = \mathbf{K}_0$. For $f(\psi) = 1$ the crack is fully open, i.e. $\mathbf{K}_{\text{c}} = \mathbf{K}_0 - \Delta\mathbf{K}_{\text{c}}$. For other values the stiffness of the cracked element is somewhere in between these two extreme values.

As can be seen the value of the crack steering function depends only on shaft rotation angle (or on shaft torsional angle). It is sufficient for most cases, where the breathing mechanism of the rotating cracked shaft should be included. However, for the non-rotating shaft, which oscillates harmonically around its axis and is deflected in different angular directions, the presented concept of the crack steering function is insufficient. The extent of crack opening should depend not only on shaft rotation /torsional angle, but also on internal loads at the crack location and resulting internal stresses. As mentioned previously, the method for calculating the extent of crack opening on the

basis of compressive stresses at the crack edge has been introduced by Darpe et al. (2004). Similar approach is used in the present article and is discussed in detail in the next section.

4.1. Stiffness matrix of the cracked shaft element

Figure 3a) presents a shaft element of radius R and length l containing a transverse crack of depth a, located at distance z_c from the i th node. The element is modeled as the finite beam element of six degress of freedom at each node, and loaded with shear forces P_2, P_3, P_8, P_9 bending moments P_5, P_6, P_{11}, P_{12}, torsional moments P_4, P_{10} and axial forces P_1, P_7. According to the Saint-Venant principle, the crack affects the stress field only in the region adjacent to the crack, i.e. only the stiffness matrix of the given finite element is considered.

The cross-section of the shaft element at the location of the crack is presented in Fig. 3b). The uncracked area as well as the closed area of the cracked portion of the cross-section are hatched. The area of the open cracked portion of the cross-section is marked as A_c. The crack is considered as an infinitely thin notch of a half-penny shape. This shape can be limited from the left (or from the right) with the crack left (or right) limit line resulting from its breathing action. This is described in more details in the next section. The positions of the limits are given by b_l and b_r (Fig. 4). The elemental strip of width $d\beta$ and height h, at distance β from shaft axis x_2' is marked on the cross-section. Heights h and α can be calculated, as follows:

$$h = 2\sqrt{R^2 - \beta^2} \qquad \alpha = h - R + a \tag{6}$$

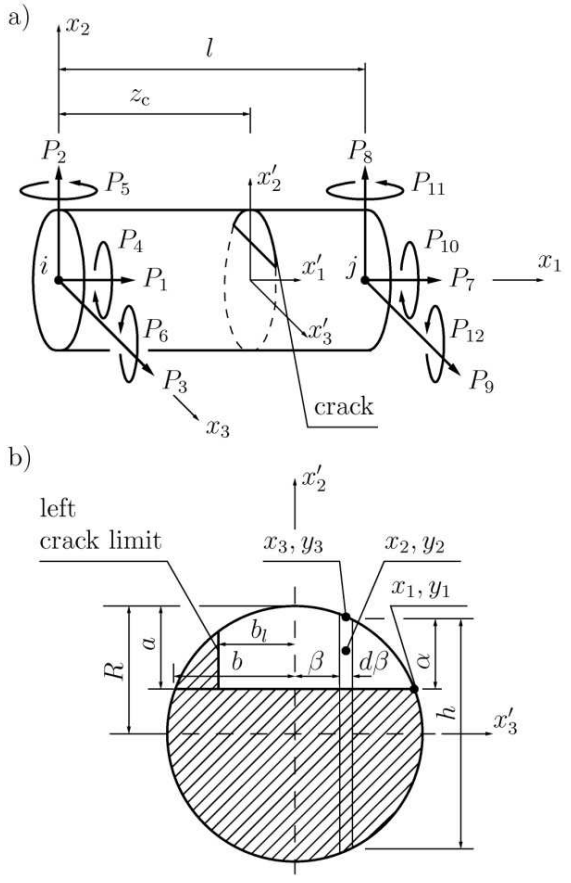


Fig. 3. Cracked shaft finite element: a) acting forces and coordinate systems, b) cracked cross-section

Using Castigliano theorem, the total node displacement q_i in the direction of load P_i can be calculated, as follows (Darpe et al., 2004):

$$q_i = \frac{\partial U_0}{\partial P_i} + \frac{\partial U_c}{\partial P_i} \tag{7}$$

where U_0 is the elastic strain energy of the uncracked element and U_c is the additional strain energy due to the crack. The elastic strain energy U_0 can be presented, as (Darpe et al., 2004):

$$U_0 = \frac{1}{2}\left(\frac{P_1^2 l}{AE} + \frac{\kappa P_2^2 l}{GA} + \frac{P_2^2 l^3}{3EJ_3} + \frac{\kappa P_3^2 l}{GA} + \frac{P_3^2 l^3}{3EJ_2} + \frac{P_4^2 l}{GJ_1} + \frac{P_5^2 l}{EJ_2} + \frac{P_6^2 l}{EJ_3} - \frac{P_2 P_6 l^2}{EJ_3} + \frac{P_3 P_5 l^2}{EJ_2} \right) \tag{8}$$

where E is Young's modulus, G is modulus of rigidity J_1, J_2, J_3 are area moments of inertia around axes x_1, x_2 and x_3, and κ is the shear coefficient.

The additional strain energy due to the crack U_c is given by the following expression (Tada et al., 1973):

$$U_c = \frac{1-\nu}{E} \int_{A_c} \left[\left(\sum_{i=1}^{6} K_{Ii} \right)^2 + \left(\sum_{i=1}^{6} K_{IIi} \right)^2 + (1+\nu) \left(\sum_{i=1}^{6} K_{IIIi} \right)^2 \right] dA_c \tag{9}$$

where A_c is the area of the open cracked portion of the shaft cross-section (Fig. 4), ν is the Poisson's ratio, and K_{Ii}, K_{IIi}, K_{IIIi} are stress intensity factors (SIFs) corresponding to three different modes of crack displacement: opening (I), sliding (II) and shearing (III).

The nonzero SIFs take the following forms (Tada et al., 1973):

$$K_{II} = \frac{P_1}{\pi R^2}\sqrt{\pi\alpha}F_1, \quad K_{I5} = \frac{4(P_5 + P_3 z_c)\beta}{\pi R^4}\sqrt{\pi\alpha}F_1$$

$$K_{I6} = \frac{2(P_2 z_c - P_6)h}{\pi R^4}\sqrt{\pi\alpha}F_2, \quad K_{II2} = \frac{\kappa P_2}{\pi R^2}\sqrt{\pi\alpha}F_{II} \tag{10}$$

$$K_{III3} = \frac{\kappa P_3}{\pi R^2}\sqrt{\pi\alpha}F_{III}, \quad K_{III4} = \frac{P_4 h}{\pi R^4}\sqrt{\pi\alpha}F_{III}$$

where the correction functions F_1, F_2, F_{II}, F_{III} are defined, as follows (Tada et al., 1973):

$$F_1 = F_{III}\frac{0.752 + 2.02\mu + 0.37(1-\sin\lambda)^3}{\cos\lambda}$$

$$F_2 = F_{III}\frac{0.923 + 0.199(1-\sin\lambda)^4}{\cos\lambda}$$

$$F_{II} = \frac{1122 - 0561\mu + 0085\mu^2 + 018\mu^3}{\sqrt{1-\mu}}, \quad F_{III} = \sqrt{\frac{\tan\lambda}{\lambda}}$$

where: $\mu = \frac{\alpha}{h}, \lambda = \frac{\pi\alpha}{2h}$.

Integrating Eqs. (8) and (9) with Eqs. (7) and (10), the generalized coordinates q_i can be presented in the following matrix form

$$\mathbf{q} = \mathbf{G}_c \mathbf{P} \tag{11}$$

where $\mathbf{q} = [q_1 \quad q_2 \quad \cdots \quad q_6]^T$, $\mathbf{P} = [P_1 \quad P_2 \quad \cdots \quad P_6]^T$ and $\mathbf{G}_c$ is the symmetric 6×6 flexibility matrix. The nonzero elements

of this matrix are given in Appendix 2. As can be seen the non-zero elements are located not only at the main diagonal, but also above and below it (e.g. $g_{2,4}$, $g_{3,4}$, $g_{3,1}$). It is obvious that these elements will couple the bending, axial and torsional vibrations. However, the off-diagonal, nonzero elements are present only in the flexibility matrix of the cracked shaft element. The other shaft finite elements do not contain the nonzero elements beyond the main diagonal (Appendix 1).

Considering the static equilibrium condition, 12 generalized coordinates of the cracked shaft finite element can be obtained (Przemieniecki, 1968):

$$\begin{bmatrix} q_1 & q_2 & \dots & q_{12} \end{bmatrix}^T = \mathbf{T}\begin{bmatrix} q_1 & q_2 & \dots & q_6 \end{bmatrix}^T \tag{12}$$

where $\mathbf{T} = [\mathbf{I}, \mathbf{T}_s]^T$ is the 12×6 transformation matrix, I is the identity matrix, and the nonzero elements of the 6×6 matrix $\mathbf{T}_s$ are, as follows:

$$t_{s1,1} = t_{s2,2} = t_{s3,3} = t_{s4,4} = t_{s5,5} = t_{s6,6} = -1\,,\; t_{s2,6} = -t_{s3,5} = l$$

The flexibility matrix $\mathbf{G}_c$ can be used to find the stiffnes matrix $\mathbf{K}_c$ of the cracked shaft finite element:

$$\mathbf{K}_c = \mathbf{T}\mathbf{G}_c\mathbf{T}\,. \tag{13}$$

4.2. Crack breathing mechanism

The changes in the extent of crack opening can be presented in terms of changes of circular segment area A_c inside the cross section of the cracked element (Fig. 4). Depending on external loads, this area changes from zero (for the fully closed crack) to its maximum value (for the fully open crack). Thus, the limits b_l and b_r separating cracked and uncracked portions of this area from the left and from the right, change from $-b$ to b (for left limit b_l) and from b to $-b$ (for right limit b_r). Here, b denotes half of the crack edge width. As can be seen from the lower part of Fig. 4, only one limit (left or right) can change at the same time, but not both. This way, the integration limits for the flexibility matrix $\mathbf{G}_c$ (Eq. 11) change in time. Consequently, the stiffness matrix K_c (Eq. 13) also changes in time, simulating the breathing behavior of the crack.

To determine the locations of the left b_l and right b_r limits, the generalized forces $\mathbf{P}_w$ acting at the nodes of the cracked shaft element should be be evaluated at each time step. These forces can be calculated using the generalized coordinates $\mathbf{q}_w$ and the stiffness matrix $\mathbf{K}_c$ of the cracked element

$$\mathbf{P}_w = \mathbf{K}_c\mathbf{q}_w \tag{14}$$

Vector of nodal coordinates $\mathbf{q}_w$ can be obtained from the vibration response $\mathbf{q}$ of the rotor by solving the motion equations (1). The nodal forces $\mathbf{P}_w$ are used in Eq. 10 to calculate stress intensity factors along the crack edge. For this, the crack edge is divided into a given number of equally spaced points at which the SIFs are evaluated. In practice only K_{I1}, K_{I5}, K_{I6} stress intensity factors are accounted for, as only they are responsible for the opening mode crack displacement influencing the extent of crack opening. To simplify, not separate SIFs are analyzed, but their sum K_s, where:

$$K_s = K_{I1} + K_{I5} + K_{I6} \tag{15}$$

A negative sign of K_s indicates compressive stress and the closed crack at a given point of the crack edge. Similarly, a positive sign of K_s indicates tensile stress and the open state of the crack at a given point of the crack edge. Thus, analyzing the sign of the overall stress intensity factor K_s at each point of the crack edge, the locations of the left b_l or right b_r crack limit can be determined. Once the crack limits are ascertained the flexibility $\mathbf{G}_c$ and stiffness $\mathbf{K}_c$ matrices are updated (Eqs. 11) and 13), and the global stiffness matrix $\mathbf{K}$ is assembled.

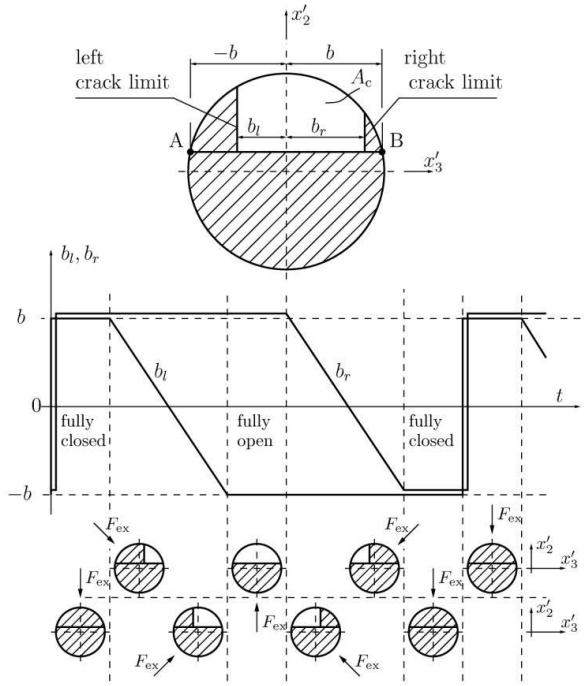


Fig. 4. Crack breathing mechanism

Next, from Eq. 1 the rotor response $\mathbf{q}$ is evaluated for the new time step, and the vector of nodal coordinates is extracted from it. Again, using Eq. 14, the vector of nodal forces is obtained, and the overall SIF K_s at several points along the crack edge is calculated. Based on the sign of K_s the new locations b_l and b_r of crack limits are evaluated and stiffness matrix $\mathbf{K}_c$ is updated. This way, at every iteration step, the overall stiffness matrix $\mathbf{K}$ of the rotor is updated by reevaluating the stiffness matrix $\mathbf{K}_c$ of the cracked finite element.

5. RESULTS

During the numerical analysis, three different models of the rotor have been considered: the first with no crack, the second with a 25% deep crack and the third with a 40% deep crack. In all cases the value of the lateral force was $F_{ex} = 100$ N, while the form of the external torque $Q_t = A_Q \sin(2\pi f_Q t)$, where the amplitude $A_Q = 500$ Nm. Two different frequencies of the exciting torque have been considered: $f_Q = 60$ Hz and $f_Q = 80$ Hz.

Using stiffness $\mathbf{K}$, damping $\mathbf{D}$, and mass $\mathbf{M}$ matrices (Eq. (1)), the natural frequencies of the rotor have been evaluated. The first two bending frequencies are located at $f_n = 40.6$ Hz and $f_n = 166.1$ Hz, while the first torsional frequency is at $f_t = 612.3$ Hz.

Motion equations (1) are solved using the Newmark integration scheme (Newmark, 1959), which is more efficient for large systems. The equations are integrated until a steady state has been established and then the FFT is calculated.

Figs. 5-12 present frequency responses for different angles ϑ of the lateral force $\mathrm{F_{ex}}$. Bending response is shown only for the vertical $\mathrm{x_2}$ axis, as the vibrations along axes $\mathrm{x_2}$ and $\mathrm{x_3}$ are much the same.

Figs. 5 and 6 present torsional and bending responses of the uncracked rotor. As expected, the torsional spectrum contains only one component of the exciting torque frequency $f_Q = 60$ Hz. In the bending response only the first natural frequency $f_n = 40.6$ Hz is slightly induced. Such characteristics are typical for the linear model of the rotor.

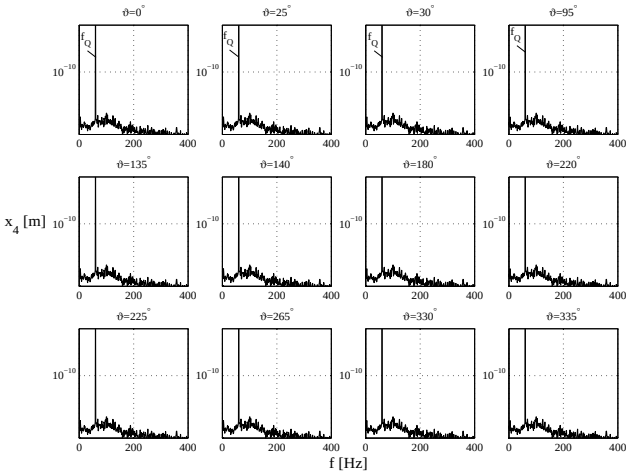


Fig. 5. Torsional response for different angles ϑ; uncracked shaft; $f_Q = 60$ Hz

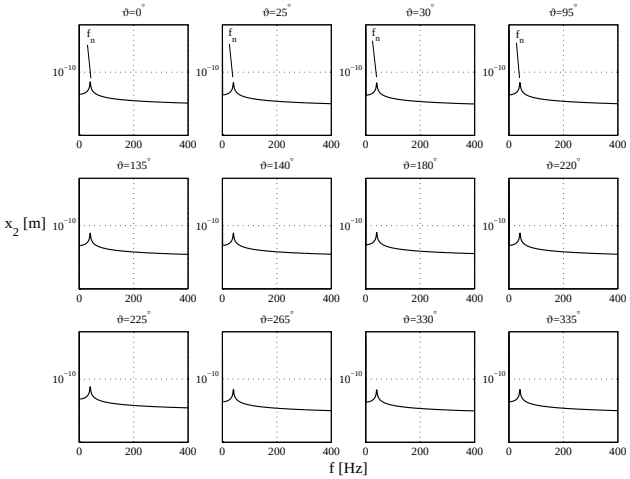


Fig. 6. Bending response for different angles ϑ; uncracked shaft; $f_Q = 60$ Hz

Figs. 7, 8 and 9 present responses of the 25% cracked rotor. Due to the nonlinearities introduced by the crack subsequent integer multiples of the exciting torque frequency $f_Q = 60$ Hz denoted as 2X (120 Hz), 3X (180 Hz), 4X (240 Hz), 5X (300 Hz), and several other frequencies of the same high amplitudes of 10^{-7} rad appear in the torsional response (Fig. 7). However, all these frequencies are observed only for particular angles ϑ, i.e. for ϑ from $30°$ to $135°$ and for ϑ from $225°$ to $330°$. It should be noticed, that such angle ranges correspond to the situations, when the crack is partially open. For other ranges, only one component is present in the vibration spectra. This is the frequency of the exciting torque $f_Q = 60$ Hz. In this case, the angles are near $0°$ and $180°$, what corresponds to the (almost) fully open and (almost) fully closed crack.

The similar, yet more important situation, is in the bending spectra (Figs. 8 and 9), where for the same angle ranges the same frequency components can be observed (including the multiples 2X, 3X, 4X, 5X, and so on). For other angle ranges, the bending frequency spectrum contains only slightly induced: natural frequency $f_n = 40.6$ Hz and exciting torque frequency $f_Q = 60$ Hz (or $f_Q = 80$ Hz).

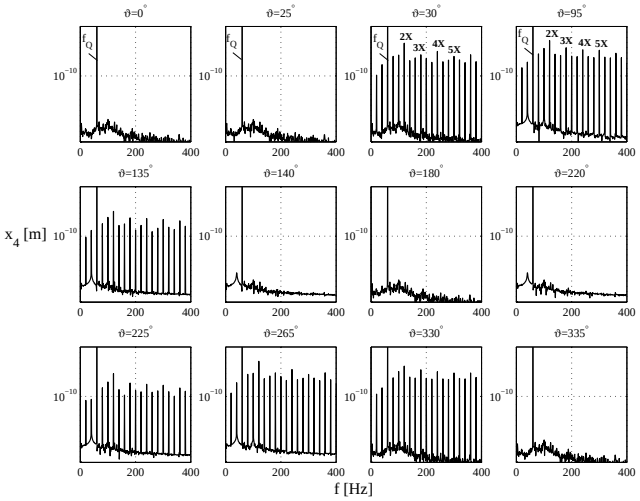


Fig. 7. Torsional response for different angles ϑ; 25% crack; $f_Q = 60$ Hz

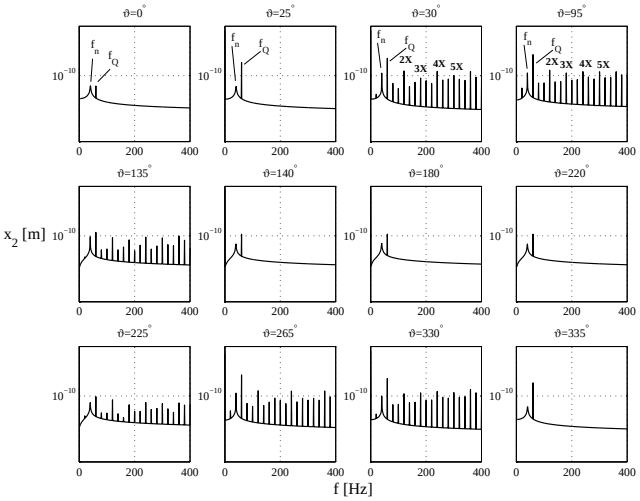


Fig. 8. Bending response for different angles ϑ; 25% crack; $f_Q = 60$ Hz

The rotor with a 40% deep crack behaves similarly (Figs. 10, 11, and 12), yet the angle ranges for which additional bending frequencies are induced are wider: from $\vartheta = 20°$ to $\vartheta = 140°$ and from $\vartheta = 210°$ to $\vartheta = 340°$. This would suggest, that for deeper cracks it is more difficult to completely close (or completely open) the crack and consequently not to induce the additional bending frequencies. Nevertheless, for the 40% deep crack the angle ranges with the differences in the frequency responses

are evident. Presumably, such crack signatures can be used for the efficient diagnosis of the health of the machine.

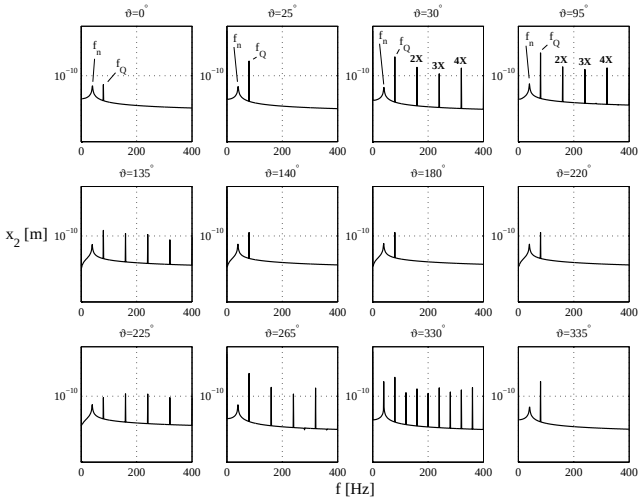


Fig. 9. Bending response for different angles ϑ; 25% crack; $f_Q = 80$ Hz

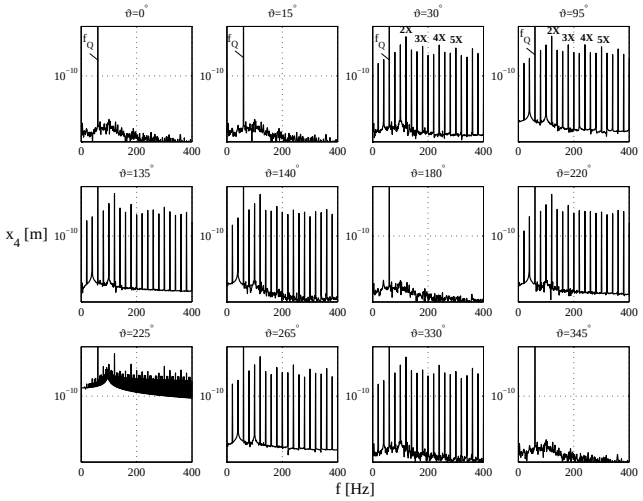


Fig. 10. Torsional response for different angles ϑ; 40% crack; $f_Q = 60$ Hz

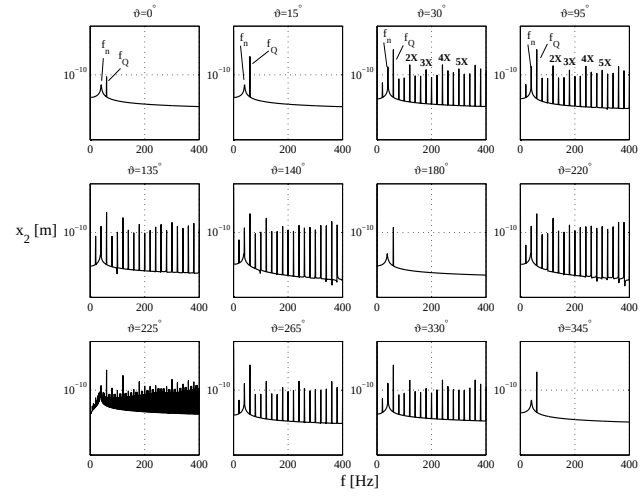


Fig. 11. Bending response for different angles ϑ; 40% crack; $f_Q = 60$ Hz

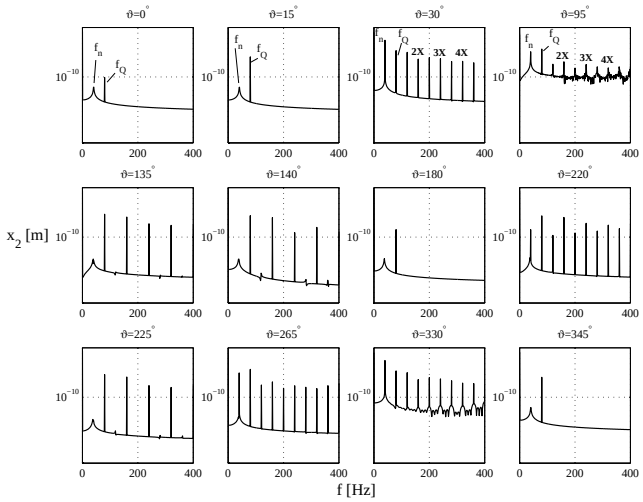


Fig. 12. Bending response for different angles ϑ; 40% crack; $f_Q = 80$ Hz

6. CONCLUSIONS

Early crack detection is a serious problem, as small shaft stiffness changes due to the crack have little influence on the rotor vibration response. During the normal machine operation the changes in the rotor response are small and practically unmeasurable. Hence, the methods amplifying the rotor sensitivity to the crack appearance and propagation should be applied.

One of these methods is suggested in the present article. Inducing the deflection of the non-rotating shaft excited by the forced torsional vibrations, the coupled bending vibrations are induced. The maximum amplification and the appearance of the multiples of the torsional frequency in the bending spectrum are observed if the deflection is induced in a direction opening the crack partially. On the other hand, the minimum coupled bending amplitudes are observed if the deflection is directed in a way ensuring the fully opening or closing of the crack. Such behavior can be explained by the fact, that in a case of a partially open crack, the multiples of the forced frequency appear quite naturally in the torsional spectrum. These frequencies are transformed by the off-diagonal non-zero elements of the stiffness matrix to the coupled bending vibrations resulting in the same multiples in the bending vibration spectra. The coupling between the bending and torsional vibrations takes place only if the cracked shaft is considered, as only then the off-diagonal non-zero elements appear in the stiffness matrix.

Numerical results confirm the potential of the proposed method. The changes in coupled bending vibrations are observed only for the cracked shaft. However, further analysis is needed to determine the required value of the external force inducing the shaft deflection, the amplitude and frequency of the exciting torque generating the forced torsional vibrations, the location of these forces along the shaft length, the location of the measuring probes, etc. At the same time, the experimental verification of the proposed method should also be conducted.

Future extension of the proposed method may involve its application for the rotating shafts. This would enable the continuous monitoring of the rotor's health, without the need to switch the machine off its normal operation.

REFERENCES

1. **Bachschmid N., Pennacchi P., Tanzi E.** (2010), *Cracked rotors: a survey on static and dynamic behaviour including modelling and diagnosing*, Springer-Verlag Berlin Heidelberg.
2. **Bachschmid N., Pennacchi P., Tanzi E., Vania A.** (2000), Identification of transverse crack position and depth in rotor systems, *Meccanica*, 35, 563-582.
3. **Bently D. E., Muszynska A.** (1986), Detection of rotor cracks, *Proceedings of Texas A&M University 15th Turbomachinery Symposium and Short Courses*, Corpus Christi, TX, 129-139.
4. **Darpe A. K., Gupta K., Chawla A.** (2004), Coupled bending, longitudinal and torsional vibrations of a cracked rotor, *Journal of Sound and Vibration*, 269, 33-60.
5. **Dimarogonas A.D., Paipetis S.A.** (1983), *Analytical Methods in Rotor Dynamics,* Applied Science Publishers, London.
6. **Gasch R.** (1993), A survey of the dynamic behavior of a simple rotating shaft with a transverse crack, *Journal of Sound and Vibration*, 160 (2), 313-332.
7. **Gawroński W., Kruszewski J., Ostachowicz W., Tarnowski J., Wittbrodt E.** (1984), *Finite element method in dynamics of structures*, Arkady, Warsaw (in Polish).
8. **Grabowski B.** (1982), Shaft vibrations in turbomachinery excited by cracks, In proceedings of the 2nd Workshop on Rotordynamic Instability Problems in High-performance Turbomachinery, Texas A&M University, *NASA Conference Publication*, 2250, 81-97.
9. **Guo D., Peng Z. K.** (2007), Vibration analysis of a cracked rotor using Hilbert-Huang transform, *Mechanical Systems and Signal Processing*, 21, 3030-3041.
10. **He Y., Guo, D., Chu, F.** (2001), Using genetic algorithms to detect and configure shaft crack for rotor-bearing system, *Computer Methods in Applied Mechanics and Engineering*, 190, 5895-5906.
11. **Isermann R.** (2005), Model-based fault detection and diagnosis – status and applications, *Annual Reviews in Control*, 29, 71-85.
12. **Ishida Y., Inoue T.** (2006), Detection of a rotor crack using a harmonic excitation and nonlinear vibration analysis, *ASME Journal of Vibration and Acoustics*, 128, 741-749.
13. **Kiciński J.** (2005), Dynamics of rotors and slide bearings, *Fluid Flow Machinery Series*, Vol. 28, IMP PAN, Gdansk (in Polish).
14. **Kulesza Z., Sawicki J. T.** (2010), Auxiliary state variables for rotor crack detection, *Journal of Vibration and Control*, 17 (6), 857-872.
15. **Litak G., Sawicki J. T.** (2009), Intermittent behaviour of a cracked rotor in the resonance region, *Chaos, Solitions and Fractals*, 42, 1495-1501.
16. **Mani G., Quinn D. D., Kasarda M.** (2005), Active health monitoring in a rotating cracked shaft using active magnetic bearings as force actuators, *Journal of Sound and Vibration*, 294, 454-465.
17. **Mayes, I. W. and Davies, W. G. R.** (1984), Analysis of the response of a multi-rotor-bearing system containing a transverse crack in a rotor, *Journal of Vibration, Acoustics, Stress and Reliability in Design*, 83, DET 84, 139-145.
18. **Nelson H. D., McVaugh J. M.** (1976), The dynamics of rotor bearing systems using finite elements, *ASME Journal of Engineering for Industry*, 98, 593-600.
19. **Newmark N. M.** (1959), A method of computation for structural dynamics, *ASCE Journal of Engineering Mechanics Division*, 85, 67-94.
20. **Przemieniecki J. S.** (1968), *Theory of matrix structural analysis*, Mc Graw-Hill, New York.
21. **Sawicki J. T., Friswell M. I., Kulesza Z., Wroblewski A., Lekki J. D.** (2011), Detecting cracked rotors using auxiliary harmonic excitation, *Journal of Sound and Vibration,* 330, 1365-1381.
22. **Sawicki J. T., Lekki J. D.** (2008), Smart structural health monitoring of rotating components using active magnetic force actuators, *Proceedings of NASA Aviation Safety Technical Conference*, Denver, Colorado, 1-23.
23. **Sinou J-J., Lees A.W.** (2005), The influence of cracks in rotating shafts, *Journal of Sound and Vibration*, 285, 1015-1037.
24. **Tada H., Paris P. C., Irwin G. R.** (1973), *The stress analysis of cracks handbook*, Del Research Corporation, Hellertown, PA.

This study is supported by the research project No S/WM/1/2012.

APPENDIX 1

Elemental matrices of the finite element model of the rotor have been obtained on the basis of (Gawroński et al., 1984).

Mass matrix of shaft finite element is, as follows:

$$\mathbf{M} = \rho l \begin{bmatrix} m_{1,1} & m_{1,2} & m_{1,3} & m_{1,4} & \dots & m_{1,11} & m_{1,12} \\ & m_{2,2} & m_{2,3} & m_{2,4} & \dots & m_{2,11} & m_{2,12} \\ & & m_{3,3} & m_{3,4} & \dots & m_{3,11} & m_{3,12} \\ & & & \dots & \dots & \dots & \dots \\ & & & & & m_{11,11} & m_{11,12} \\ symm. & & & & & & m_{12,12} \end{bmatrix}$$

where the nonzero elements lying on and above the main diagonal are, as follows:

$$m_{1,1} = \frac{A}{3},\ m_{1,7} = \frac{A}{6},\ m_{2,2} = \frac{13A}{35} + \frac{6J_3}{5l^2}$$

$$m_{2,6} = \frac{11Al}{210} + \frac{J_3}{10l},\ m_{2,8} = \frac{9A}{70} - \frac{J_3}{5l^2}$$

$$m_{2,12} = \frac{-13Al}{420} + \frac{J_3}{10l},\ m_{3,3} = \frac{13A}{35} + \frac{6J_2}{5l^2}$$

$$m_{3,5} = \frac{-11Al}{210} - \frac{J_2}{10l},\ m_{3,9} = \frac{9A}{70} - \frac{6J_2}{5l^2}$$

$$m_{3,11} = \frac{13Al}{420} - \frac{J_2}{10l},\ m_{4,4} = \frac{J_1}{3},\ m_{4,10} = \frac{J_1}{6}$$

$$m_{5,5} = \frac{Al^2}{105} + \frac{2J_2}{15},\ m_{5,9} = -m_{3,11}$$

$$m_{5,11} = \frac{-Al^2}{140} - \frac{J_2}{30},\ m_{6,6} = \frac{Al^2}{105} + \frac{2J_3}{15},\ m_{6,8} = -m_{2,12}$$

$$m_{6,12} = \frac{-Al^2}{140} - \frac{J_3}{30},\ m_{7,7} = m_{1,1},\ m_{8,8} = m_{2,2}$$

$$m_{8,12} = -m_{2,6},\ m_{9,9} = m_{3,3},\ m_{9,11} = -m_{3,5}$$

$$m_{10,10} = m_{4,4},\ m_{11,11} = m_{5,5},\ m_{12,12} = m_{6,6}$$

Stiffness matrix of shaft finite element takes the following form:

$$\mathbf{K} = \frac{E}{l} \begin{bmatrix} k_{1,1} & k_{1,2} & k_{1,3} & k_{1,4} & \dots & k_{1,11} & k_{1,12} \\ & k_{2,2} & k_{2,3} & k_{2,4} & \dots & k_{2,11} & k_{2,12} \\ & & k_{3,3} & k_{3,4} & \dots & k_{3,11} & k_{3,12} \\ & & & \dots & \dots & \dots & \dots \\ & & & & & k_{11,11} & k_{11,12} \\ sym. & & & & & & k_{12,12} \end{bmatrix}$$

where the nonzero elements lying on and above the main diagonal are, as follows:

$$k_{1,1} = A\ ,\ k_{1,7} = -k_{1,1}\ ,\ k_{2,2} = \frac{12J_3}{l^2}\ ,\ k_{2,6} = \frac{6J_3}{l}$$

$$k_{2,8} = -k_{2,2}\ ,\ k_{2,12} = k_{2,6}\ ,\ k_{3,3} = \frac{12J_2}{l^2}\ ,\ k_{3,5} = \frac{-6J_2}{l}$$

$$k_{3,9} = -k_{3,3}\ ,\ k_{3,11} = k_{3,5}\ ,\ k_{4,4,} = \frac{J_1}{2(1+\nu)}\ ,\ k_{4,10} = -k_{4,4}$$

$$k_{5,5} = 4J_2\ ,\ k_{5,9} = -k_{3,5}\ ,\ k_{5,11} = 2J_2\ ,\ k_{6,6} = 4J_3$$

$$k_{6,8} = -k_{2,6}\ ,\ k_{6,12} = 2J_3\ ,\ k_{7,7} = k_{1,1}\ ,\ k_{8,8} = k_{2,2}$$

$$k_{8,12} = -k_{2,6}\ ,\ k_{9,9} = k_{3,3}\ ,\ k_{9,11} = k_{3,5}\ ,\ k_{10,10} = k_{4,4}$$

$$k_{11,11} = k_{5,5}\ ,\ k_{12,12} = k_{6,6}$$

Damping matrix $\mathbf{D}$ of shaft finite element is calculated, as: $\mathbf{D} = \alpha_{\mathrm{d}}\mathbf{K} + \beta_{\mathrm{d}}\mathbf{M}$, where the following values have been assumed: $\alpha_{\mathrm{d}} = 1 \times 10^{-5}, \beta_{\mathrm{d}} = 0$.

Mass matrix of a disk takes the following form: $\mathbf{M} = \mathrm{diag}(m, m, m, J_{m1}, J_{m2}, J_{m3})$, where m is the mass of the disk, and J_{m1}, J_{m2}, J_{m3} are mass moments of inertia of the disk around x_1, x_2, and x_3 axes.

Stiffness matrix of a bearing takes the following form: $\mathbf{K} = \mathrm{diag}(k_a, k_b, k_b, k_t, 0, 0)$, where k_a, k_b, k_t are stiffness coefficients for axial, bending and torsional displacements.

Damping matrix of the bearing takes the following form: $\mathbf{D} = \mathrm{diag}(d_a, d_b, d_b, d_t, 0, 0)$, where d_a, d_b, d_t are damping coefficients for axial, bending and torsional speeds.

APPENDIX 2

Flexibility matrix G_{c} of the cracked shaft element can be presented, as:

$$\mathbf{G}_{\mathrm{c}} = \begin{bmatrix} g_{1,1} & g_{1,2} & \cdots & g_{1,6} \\ & g_{2,2} & \cdots & g_{2,6} \\ & & \cdots & \cdots \\ sym. & & & g_{6,6} \end{bmatrix}$$

where the nonzero elements lying on and above the main diagonal are, as follows:

$$g_{1,1} = \frac{l}{AE} + \frac{2}{\pi E' R^4}\int_{A_{\mathrm{c}}} \alpha F_1^2 dA_{\mathrm{c}}\ ,\quad g_{1,2} = \frac{4z_{\mathrm{c}}}{\pi E' R^6}\int_{A_{\mathrm{c}}} \alpha h F_1 F_2 dA_{\mathrm{c}}$$

$$g_{1,3} = \frac{8z_{\mathrm{c}}}{\pi E' R^6}\int_{A_{\mathrm{c}}} \alpha\beta F_1^2 dA_{\mathrm{c}}$$

$$g_{1,5} = \frac{8}{\pi E' R^6}\int_{A_{\mathrm{c}}} \alpha\beta F_1^2 dA_{\mathrm{c}}\ ,\ g_{1,6} = \frac{-4}{\pi E' R^6}\int_{A_{\mathrm{c}}} \alpha h F_1 F_2 dA_{\mathrm{c}}$$

$$g_{2,2} = \left(\frac{\kappa l}{GA} + \frac{l^3}{3EJ_2}\right) + \frac{8z_{\mathrm{c}}^2}{\pi E' R^8}\int_{A_{\mathrm{c}}} \alpha h^2 F_2^2 dA_{\mathrm{c}} + \frac{2\kappa^2}{\pi E' R^4}\int_{A_{\mathrm{c}}} \alpha F_{\mathrm{II}}^2 dA_{\mathrm{c}}$$

$$g_{2,3} = \frac{16z_{\mathrm{c}}^2}{\pi E' R^8}\int_{A_{\mathrm{c}}} \alpha\beta h F_1 F_2 dA_{\mathrm{c}}$$

$$g_{2,4} = \frac{4\kappa}{\pi E' R^6}\int_{A_{\mathrm{c}}} \alpha\beta F_{\mathrm{II}}^2 dA_{\mathrm{c}}$$

$$g_{2,5} = \frac{16z_{\mathrm{c}}}{\pi E' R^8}\int_{A_{\mathrm{c}}} \alpha\beta h F_1 F_2 dA_{\mathrm{c}}$$

$$g_{2,6} = -\frac{l^2}{2EJ_3} - \frac{8z_{\mathrm{c}}}{\pi E' R^8}\int_{A_{\mathrm{c}}} \alpha h^2 F_2^2 dA_{\mathrm{c}}$$

$$g_{3,3} = \left(\frac{\kappa l}{GA} + \frac{l^3}{3EJ_2}\right) + \frac{32z_{\mathrm{c}}^2}{\pi E' R^8}\int_{A_{\mathrm{c}}} \alpha\beta^2 F_1^2 dA_{\mathrm{c}} + \frac{2\kappa^2(1+\nu)}{\pi E' R^4}\int_{A_{\mathrm{c}}} \alpha F_{\mathrm{III}}^2 dA_{\mathrm{c}}$$

$$g_{3,4} = \frac{2\kappa(1+\nu)}{\pi E' R^6}\int_{A_{\mathrm{c}}} \alpha h F_{\mathrm{III}}^2 dA_{\mathrm{c}}$$

$$g_{3,5} = \frac{l^2}{2EJ_2} + \frac{32z_{\mathrm{c}}}{\pi E' R^8}\int_{A_{\mathrm{c}}} \alpha\beta^2 F_1^2 dA_{\mathrm{c}}$$

$$g_{3,6} = \frac{-16z_{\mathrm{c}}}{\pi E' R^8}\int_{A_{\mathrm{c}}} \alpha\beta h F_1 F_2 dA_{\mathrm{c}}$$

$$g_{4,4} = \frac{l}{GJ_1} + \frac{8}{\pi E' R^8}\int_{A_{\mathrm{c}}} \alpha\beta^2 F_{\mathrm{II}}^2 dA_{\mathrm{c}} + \frac{2(1+\nu)}{\pi E' R^8}\int_{A_{\mathrm{c}}} \alpha h^2 F_{\mathrm{III}}^2 dA_{\mathrm{c}}$$

$$g_{5,5} = \frac{l}{EJ_2} + \frac{32}{\pi E' R^8}\int_{A_{\mathrm{c}}} \alpha\beta^2 F_1^2 dA_{\mathrm{c}}$$

$$g_{5,6} = \frac{-16}{\pi E' R^8}\int_{A_{\mathrm{c}}} \alpha\beta h F_1 F_2 dA_{\mathrm{c}}$$

$$g_{6,6} = \frac{l}{EJ_3} + \frac{8}{\pi E' R^8}\int_{A_{\mathrm{c}}} \alpha h^2 F_2^2 dA_{\mathrm{c}}$$

DESIGN OF FRACTIONAL ORDER CONTROLLER SATYSFYING GIVEN GAIN AND PHASE MARGIN FOR A CLASS OF UNSTABLE PLANT WITH DELAY

Tomasz NARTOWICZ*

*Białystok University of Technology, Faculty of Electrical Engineering, ul. Wiejska 45 D, 15-351 Białystok, Poland

tomek.nartowicz@gmail.com

Abstract: The paper describes the design problem of fractional order controller satisfying gain and phase margin of the closed loop system with unstable plant with delay. The proposed method is based on using Bode's ideal transfer function as a reference transfer function of the open loop system. Synthesis method is based on simplify of the object transfer function. Fractional order of the controllers is relative with gain and phase margin only. Computer method for synthesis of fractional controllers is given. The considerations are illustrated by numerical example and results of computer simulation with MATLAB/Simulink.

Key words: Fractional Order Controller, Stability, Delay, Bode's Ideal Transfer Function

1. INTRODUCTION

In recent years considerable attention has been paid to fractional calculus and its application in many areas in science or engineering (see, e.g. (Kilbas et al., 2006; Das, 2008; Ostalczyk, 2008; Kaczorek, 2011)).

In control system fractional order controllers are used to improve the performance of the feedback control loop. One of the most developed approaches in science to design robust and fractional order controllers is CRONE control methodology (French acronym of "Commande Robuste d'Ordre Non Entier" - non-integer order robust control; Oustaloup, 1991, 1995, 1999).

The fractional order PID controllers, namely $\mathrm{PI}^{\lambda}\mathrm{D}^{\mu}$ controllers, where λ integrator order and μ differentiator order were proposed in (Podlubny, 1994, 1999). Several design methods based on the mathematical description of the process of tuning the $\mathrm{PI}^{\lambda}\mathrm{D}^{\mu}$ controllers were presented in (Monje et al., 2004; Valerio, 2005; Valerio and Costa, 2006).

Also knows in science are approaches based on optimalization methods (Monje et al., 2004), and classic Zieglera-Nicholsa metod (Valerio and da Costa, 2006). Methods based on the first order-plant with time delay, is the most frequently used model for tuning fractional and integral controllers (O'Dwyer, 2003).

In this paper a simple method of determining the fractional order controller satisfying given gain ang phase margin of the closed loop system with unstable plant with delay is given.

Transfer function of the controller follows from the use of Bode's ideal transfer function as a reference transfer function for the open loop system (Barbosa et al., 2004; Busłowicz and Nartowicz, 2009, Nartowicz 2010). Approach submit In the paper was proposed in (Barbosa et al., 2004) for a class of natural order controller, and (Busłowicz and Nartowicz, 2009) for a fractional order controller synthesis.

The considerations are illustrated by numerical example and results of computer simulation with MATLAB/Simulink.

2. METHOD

Consider the feedback control system show In Fig. 1 in which the process to be controlled is described by (1):

$$G(s) = \frac{k}{1 - s\tau} e^{-sh} \tag{1}$$

where: k, π, h are positive real numbers, and C(s) is fractional order controller.

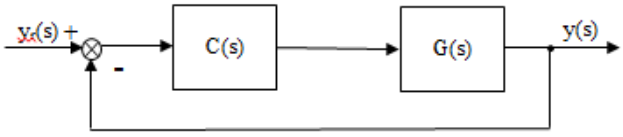


Fig. 1. Feedback control system structure

The paper presents the simple synthesis method of the fractional order controller satisfying given A_m gain and $\phi_m$phase margin of the of the closed loop system with unstable plant with delay. Transfer function of the controller follows directly from the use of Bode's ideal transfer function as a reference transfer function for the open loop system:

$$K(s) = \left(\frac{\omega_c}{s}\right)^{\beta} \tag{2}$$

where ω_c is gain crossover frequency ($|K(j\omega_c)| = 1$) and β is real number. Transfer function (2) describe derivative plant for $\beta < 0$ and integral plant for $\beta > 0$. The open loop system (2) has constant value of phase margin $\phi_m = (1 - 0.5\beta)\pi$ hence such a system is unsensitive to gain changes in open loop system. For a detailed analysis of the considered system, including time domain, see paper (Ostalczyk P., 2008).

To obtain an open loop system in the form (2), simplify the plant transfer function:

$$G(s) = \frac{k}{s(1 - s\tau)} e^{-sh} \approx -\frac{k}{s^2\tau} e^{-sh} \tag{3}$$

The controller transfer function should have a structure:

$$C(s) = -k_c s^{2-\alpha} \tag{4}$$

where α is real number.

Open loop transfer function is given:

$$K(s) = C(s)G(s) = \frac{kk_c}{\tau} \frac{e^{-sh}}{s^{\alpha}} \tag{5}$$

Note that open loop transfer function of control system shown in Fig.1. is different than Bode's ideal transfer function (2) with coefficient $\exp(-sh)$. It takes differences while Bode's diagram drawing.

Consider synthesis of the fractional order controller (4). For a given A_m gain and ϕ_m phase margins, the controller parameters k_c and real number α are searching.

Using $(j\omega)^{\alpha} = |\omega|^{\alpha} e^{j\alpha\pi/2}$ wrote gain and phase for a transfer function (5):

$$|K(j\omega)| = \frac{kk_c}{\tau} \frac{1}{\omega^{\alpha}} \qquad \phi(\omega) = \arg K(j\omega) = -h\omega - \alpha\frac{\pi}{2} \tag{6}$$

For a gain ω_g and ω_p phase crossover frequency terms can be written:

$$|K(j\omega_g)| = 1 \qquad \phi(\omega_p) = \arg K(j\omega_p) = -\pi \tag{7}$$

Using (6) the equations (7) can be rewritten as:

$$\frac{kk_c}{\tau\omega_g^{\alpha}} = 1 \qquad -h\omega_p - \alpha\frac{\pi}{2} = -\pi \tag{8}$$

By solving equations (8) gain ω_g and ω_p phase crossover frequency:

$$\omega_g^{\alpha} = \frac{kk_c}{\tau} \qquad \omega_p = \frac{(2-\alpha)\frac{\pi}{2}}{h} \tag{9}$$

Considering the second od (9) equations, we can said that ω_p is positive number while $\alpha < 2$.

For a given A_m gain and ϕ_m phase margins:

$$\frac{kk_c}{\tau\omega_p^{\alpha}} = \frac{1}{A_m} \qquad \phi_m = \pi - h\omega_g - \alpha\frac{\pi}{2} \tag{10}$$

By solving (10) we can written:

$$\omega_p = \left(\frac{A_m kk_c}{\tau}\right)^{1/\alpha} \qquad \omega_g = \frac{(2-\alpha)\frac{\pi}{2} - \phi_m}{h} \tag{11}$$

Simply using first equations of (9) and (10):

$$A_m = \frac{\omega_p^{\alpha}}{\omega_g^{\alpha}} \tag{12}$$

Using second equations (9) and (11), and solving with (12) gain margin A_m is given by:

$$A_m = \left(\frac{(2-\alpha)\frac{\pi}{2}}{(2-\alpha)\frac{\pi}{2} - \phi_m}\right)^{\alpha} \tag{13}$$

Nonlinear equation (13) is handling A_m gain and ϕ_m phase margins and fractional order of the considerable controller (4) α.

Parameter α can be determined by solving equation (13) using computer methods.

By solving one of the first equations of (8) or (10):

$$k_c = \frac{\tau\omega_g^{\alpha}}{k} = \frac{\tau\omega_p^{\alpha}}{kA_m} \tag{14}$$

where k – gain of the transfer function (1).

Gain and phase fraquency crossover are determined from second equation of (9) or (11).

Note that fractional order of the controllers is relative with gain and phase margin only. Gain controller k_c is relative with gain or phase crossover frequency, gain k and time τ of the considerable object.

Method of the synthesis fractional order controller satisfying gain A_m and phase ϕ_m margin of the closed loop system with unstable plant with delay is given.

Synthesis method:

1. Solving nonlinear equation (13) for a given gain A_m and phase ϕ_m margins
   Real number α is given.
2. Solving phase crossover frequency from equation (9) or gain crossover frequency from equation (11)
   Parameter k_c of the controller is given with (14)
   Stability margin for a real object is smaller because of simplify used in (3).

3. SYNTHESIS METHOD FOR A UNSTABLE PLANT WITH INTEGRAL TERM WITH DELAY

Consider proposed synthesis method of the fractional controller in feedback control system shown in Fig. 1 in which the process to be controlled is described by transfer function:

$$G_1(s) = \frac{k}{s(1 - s\tau)} e^{-sh} \tag{15}$$

To obtain an open loop system in the form (2), simplify the plant transfer function:

$$G(s) = \frac{k}{s(1 - s\tau)} e^{-sh} \approx -\frac{k}{s^2\tau} e^{-sh} \tag{16}$$

The controller transfer function should have a structure:

$$C(s) = -k_c \frac{s^2}{s^{\alpha}} = -k_c s^{2-\alpha} \tag{17}$$

where α is real number.

Open loop transfer function is given:

$$K(s) = C(s)G(s) = \frac{kk_c}{\tau} \frac{e^{-sh}}{s^{\alpha}} \tag{18}$$

the same as given in (5), so for considered class main results are as given in. Determining controller parameters for a transfer function (15) we use equations (5-14) and given synthesis method.

4. SYNTHESIS METHOD FOR A SECOND ORDER UNSTABLE PLANT WITH DELAY

Consider proposed synthesis method of the fractional controller in feedback control system shown in Fig.1 in which the process to be controlled is described by transfer function:

$$G_1(s) = \frac{k}{(1-s\tau)(1+s\tau_2)} e^{-sh} \tag{19}$$

To obtain an open loop system in the form (2), simplify the plant transfer function:

$$G(s) = \frac{k}{(1-s\tau)(1+s\tau_2)} e^{-sh} \approx -\frac{k}{s\tau(1+s\tau_2)} e^{-sh} \tag{20}$$

The controller transfer function should have a structure:

$$C(s) = -k_c \frac{s(1+sT)}{s^{\alpha}} = -k_c s^{1-\alpha}(1+sT) \tag{21}$$

where α is real number, and while $\tau_2 = \mathrm{T}$, we can said that open loop transfer function is given:

$$K(s) = C(s)G(s) = \frac{kk_c}{\tau} \frac{e^{-sh}}{s^{\alpha}} \tag{22}$$

the same as given in (5), so for considered class main results are as given. Determining controller parameters for a transfer function (23) we use equations (5-14) and given synthesis method.

5. RESULTS

Example 1:

Consider the feedback control system shown in Fig.1 in which the process to be controlled is described by transfer function:

$$G(s) = \frac{0.55}{1-62s} e^{-10s} \tag{23}$$

Using synthesis method determine controller parameters for a given gain $A_m = 4$ (ab. 12dB) and phase $\phi_m = 55°$ (ab. 0.96 rad) margins for a closed loop system. In that case: $k = 0.55$, $\tau = 62$, $h = 10$.

Using synthesis method:

1. By solving equation (13) we get $\alpha = 1.13385$.
2. By solving (11) we get $\omega_g = 0.0401$.

Parameter k_c is given by (14): $k_c = 2.9358$

So transfer function of the controller (4) can be written:

$$C(s) = \frac{2.9358}{s^{0.13385}} \tag{24}$$

Fig. 2 shows step response for control system presented in Fig. 1, where the process to be controlled is described by (23) and fractional order controller (24). Step response is drawn for a few values of parameter k.

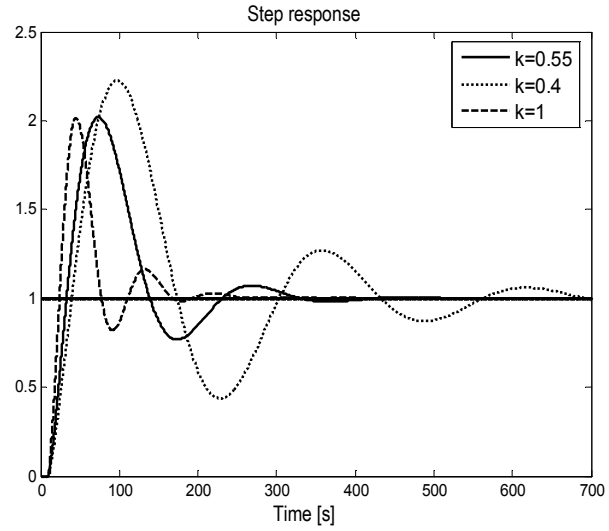


Fig. 2. Step response of the closed loop system with object (15) and controller (16) for a few different values of parameter k

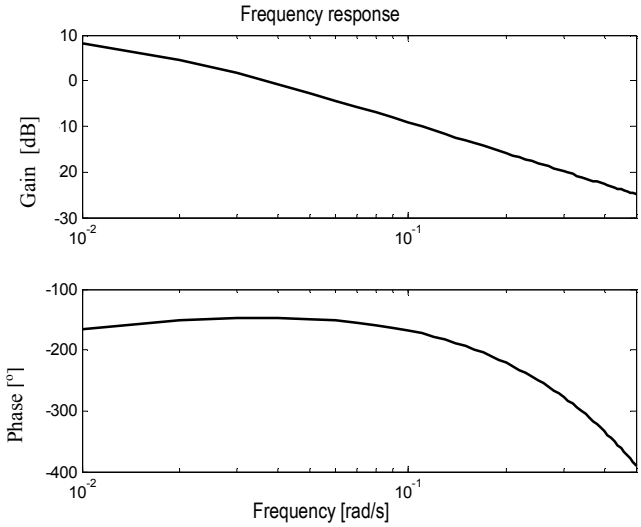


Fig. 3. Frequency response of the open loop system with object (15) and controller (16)

Overshoot of step response drawn for $k = 0.55$ is 100%. Overshoot is growing for a values less than 0.55. Drawing step response for a values k more that 0.55 (simulations were drawing to 1) no changing of overshoot were note.

Note that parameter k_c of fractional controller (24) is negative, so with simplify transfer function (3) we finally get negative feedback. Simulations in matlab are drawn for a original transfer function of the considerable transfer function, note that feedback is positive.

Fig. 3 shows Frequency response of the open loop system with object (15) and controller (16). Measured stability margin for a designed control system:

$$A_m = 3.5956, \phi_m = 32.9208°.$$

Stability margin measured is smaller because of simplify used in (3).

Example 2:

Consider the feedback control system shown in Fig.1 in which the process to be controlled is described by transfer function:

$$G(s)=\frac{0.55}{s(1-62s)}e^{-10s} \tag{25}$$

Using synthesis method determine controller parameters for a given gain $A_m = 4$ (ab. 12dB) and phase $\phi_m = 55°$ (ab. 0.96 rad) margins for a closed loop system.

In that case: $k = 0.55, \tau = 62, h = 10$.

Using synthesis method::

1. By solving equation (13) we get $\alpha = 1.13385$.
2. By solving (11) we get $\omega_g = 0.0401$.

Parameter k_c is given by (14): $k_c = 2.9358$.

So transfer function of the controller (4) can be written:

$$C(s)=-\frac{2.9358}{s^{0.13385}} \tag{26}$$

Fig. 4 shows step response for control system show in Fig. 1, where the process to be controlled is described by (25) and fractional order controller (26). Step response is drawn for a few values of parameter k.

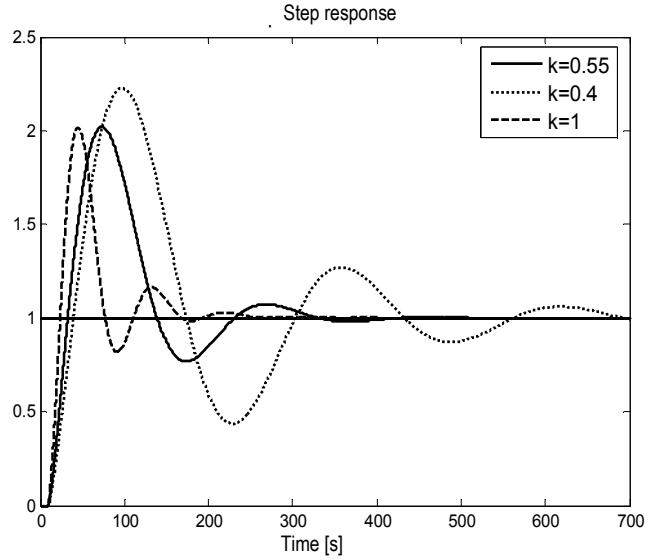


Fig. 4. Step response of the closed loop system with plant (21) and controller (22) for a few different values of parameters k

Overshoot of step response drawn for $k = 0.55$ is 100%. Overshoot is growing for a values less than 0.55. Drawing step response for a values k more that 0.55 (simulations were drawing to 1) no changing of overshoot were note. In that case we can also said that parameter k_c of fractional controller (26) is negative, and because of simplify (3) we finally get negative feedback. Simulations in matlab are drawn for a original transfer function of the considerable transfer function, note that feedback is positive.

Measured stability margin for a designed control system:

$$A_m = 3.5956, \phi_m = 32.9208°.$$

Stability margin measured is smaller because of simplify used in (3), and the same as given in example 1.

Example 3:

Consider the feedback control system shown in Fig.1 in which the process to be controlled is described by transfer function:

$$G(s)=\frac{0.55}{(1-62s)(1+36)}e^{-10s} \tag{27}$$

Using synthesis method determine controller parameters for a given gain $A_m = 4$ (ab. 12dB) and phase $\phi_m = 55°$ (ab. 0.96 rad) margins for a closed loop system.

In that case: $k = 0.55, \tau = 62, \tau_2 = 36, h = 10$.

Using synthesis method transfer function of the controller (25) can be written:

$$C(s)=-\frac{2.9358}{s^{0.13385}}-105.6888s^{0.8661}. \tag{28}$$

Fig. 5 shows step response for control system show in Fig. 1, where the process to be controlled is described by (27) and fractional order controller (28). Step response is drawn for a few values of parameter k.

Overshoot of step response drawn for $k = 0.55$ is 100%. Overshoot is growing for a values less than 0.55. Drawing step response for a values k more that 0.55 (simulations were drawing to 1) no changing of overshoot were note.

Measured stability margin for a designed control system:

$$A_m = 3.5956, \phi_m = 32.9208°.$$

Stability margin measured is the same as in example 2 and example 2 because of transfer function of the open loop system given by the same transfer function

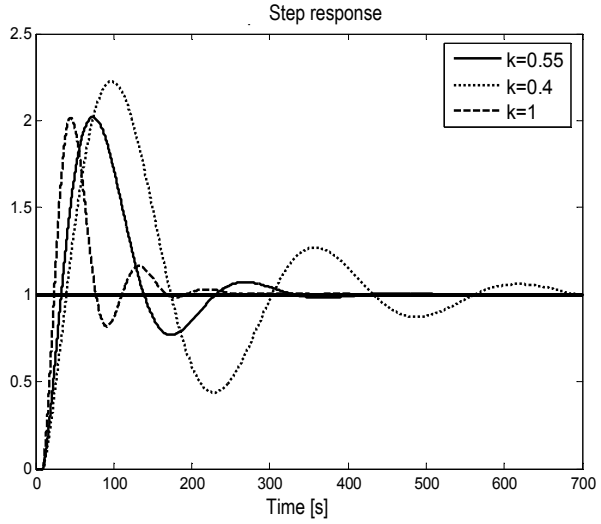


Fig. 5. Step response of the closed loop system with object (27) and controller (28) and a few different values of parameters k

6. CONCLUSIONS

The paper considers the design problem of fractional order controller satisfying gain and phase margin of the closed loop system. The proposed method is based on using Bode's ideal transfer function as a reference transfer function of the open loop system. Synthesis method is based on simplify of the object transfer function. Fractional order of the controllers is relative with gain and phase margin only. Method is based on using Bode's ideal transfer function as a reference transfer function of the open loop system. Synthesis method is based on simplify of the object transfer function. Fractional order of the controllers is relative with gain

and phase margin only. Open loop transfer function of control system shown in Fig.1. is different than Bode's ideal transfer function (2) with coefficient $\exp(-sh)$. It takes differences while Bode's diagram drawing, phase and gain margin is different than given while synthesis the controller.

Parameter k_c of fractional controller is negative in each example considered in the paper, so with simplify transfer function (3) we finally get negative feedback. Simulations in matlab are drawn for a original transfer function of the considerable transfer function, the feedback is positive this example.

Computer method for synthesis of fractional controllers is given. The considerations are illustrated by numerical example and results of computer simulation with MATLAB/Simulink.

REFERENCES

1. **Barbosa R. S., Machado J. A., Ferreira I. M.** (2004), Tuning of PID controllers based on Bode's ideal transfer function, *Nonliner Dynamics*, Vol. 38, 305-321.
2. **Boudjehem B., Boudjehem D., Tebbikh H.** (2008), Simple analytical design method for fractional-order controller, *Proc. 3-rd IFAC Workshop on Fractional Differentiation and its Application*s, Ankara, Turkey (CD-ROM).
3. **Busłowicz M.** (2008a), Frequency domain method for stability analysis of linear continuous-time fractional systems, in: Malinowski K., Rutkowski L.: *Recent Advances in Control and Automation*, Academic Publishing House EXIT, Warszawa, 83-92.
4. **Busłowicz M.** (2008b), Robust stability of convex combination of two fractional degree characteristic polynomials, *Acta Mechanica et Automatica*, Vol. 2, No. 2, 5-10.
5. **Busłowicz M.** (2009), Stability analysis of linear continuous-time fractional systems of commensurate order, *Journal of Automation, Mobile Robots and Intelligent Systems*, Vol. 3, 15-21.
6. **Busłowicz M., Nartowicz T**. (2009), Fractional order contro ller far a class of inertial plant with delay, *Pomiary Automatyka Robotyka*, 2/2009, 398-405.
7. **Das. S.** (2008), *Functional Fractional Calculus for System Identification and Controls*, Springer, Berlin.
8. **Kilbas A. A., Srivastava H. M., Trujillo J. J.** (2006), *Theory and Applications of Fractional Differential Equations*, Elsevier, Amsterdam.
9. **Oustaloup A., Sabatier J., Lanusse P., Malti R., Melchior P., Moreau X., Moze M.** (2008), An overview of the CRONE approach in system analysis, modeling and identification, observation and control, *Proc. 17th World Congress IFAC*, Soul, 14254-14265.
10. **Podlubny I.** (1994), Fractional order systems and fractional order controllers, *The Academy of Sciences Institute of Experimental Physics*, Kosice, Slovak Republic.
11. **Podlubny I.** (1999a), *Fractional Differential Equations*, Academic Press, San Diego.
12. **Podlubny I.** (1999b), Fractional-order systems and PID-controllers, *IEEE Trans. Autom. Control*, Vol. 44, No. 1, 208-214.
13. **Skogestad S.** (2001), *Probably the best simple PID tuning rules in the world*, AIChE Annual Meeting, Reno, Nevada.
14. **Valerio D.** (2005), *Fractional Robust Systems Control. PhD Dissertation*, Technical University of Lisbona.
15. **Valerio D., da Costa J. S.** (2006), Tuning of fractional PID controllers with Ziegler-Nichols type rules, *Signal Processing*, Vol. 86, 2771-2784.

ANALITICAL METHOD OF PID CONTROLLER TUNING FOR A CLASS OF UNSTABLE PLANT

Tomasz NARTOWICZ*

*Białystok University of Technology, Faculty of Electrical Engineering, ul. Wiejska 45 D, 15-351 Białystok, Poland

tomek.nartowicz@gmail.com

Abstract: The aim of the paper is to present the synthesis method of classic PID controller for a class of unstable plant. The proposed method based directly on Skogestad paper, where analitical synthesis of PID controllers in described. This paper is generalization of that approach on a class of unstable plants with delay. Analitical method for synthesis of fractional controllers is given. The considerations are illustrated By numerical example and results of computer simulation with MATLAB/Simulink.

Key words: PID Controller, Skogestad, Stability, Delay

1. INTRODUCTION

PID controllers are so far widely used in practice, because of well known simple structure. Many methods of tuning PID controllers for satisfactory behavior have been proposed in the literature (Johnson andMoradi, 2005). The methods are based on knowledge of mathematical descriptionof process (O'Dwyer, 2003).

This paper is generalization of Skogestad method for a class of unstable plant. The starting point has been the IMC PID tuning rules of Rivera (1986). Furthermore Skogestad start by approximating the process by afirst-order plus delay processes. He proposed analytic tuning rules, simply but still result in a good closed-loop behavior.

2. METHOD

Consider the feedback control system show in Fig.1:

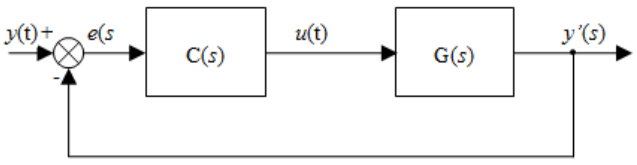


Fig. 1. Feedback control system structure

The process to be controlled is described by (1):

$$G(s)=\frac{k}{1+T_1 s \quad 1+T_2 s} e^{-sh} \tag{1}$$

where: $T_1 \geq 0$, $T_2 \geq 0$, $h \geq 0$, and $C(s)$ – transfer function of the controller.

The method proposed by Skogestad is based on approximating the process by a first- or second-order plus delay model, where: k – plant gain,τ_1 – dominant time constant,h –effective time delay.

Skogestad method is based directly on analytical set of the controller parameters (Skogestad, 2001):

$$C(s) = K_c \frac{(1+\tau_I s)}{\tau_I s}(1+\tau_D s) \tag{2}$$

with fallowing rules:

– Controller gain K_c:

$$K_c = \frac{1}{k}\frac{T_1}{\tau_C + h} \tag{3}$$

– Time constansτ_I:

$$\tau_I = \min\{T_1, 4(\tau_C + h)\} \tag{4}$$

– Time constansτ_D:

$$\tau_D = T_2 \tag{5}$$

In equation (3) parameter τ_c is recomended as follows:

$$\tau_C \geq h \tag{6}$$

In the paper the synthesis method of PID controller for an unstable plant described with transfer function (7) is proposed.

$$G(s)=\frac{k}{1 \quad T_{,1} s \quad 1+T_2 s} e^{sh}, \tag{7}$$

where$T_1 \geq 0$, $T_2 \geq 0$, $h \geq 0$.

To make generalization of Skogestad method for a class of unstable plant (7), the controller gain $K_c$is described with:

$$K_c = \frac{1}{k}\frac{-T_1}{\tau_C + h} \tag{8}$$

Equations (4)-(6) take their form unchanged.

Compare proposed method with synthesis method of fractional order controller proposed in Nartowicz (2011).

In that paper transfer function of the controller follows directly

from the use of Bode's ideal transfer function as a reference transfer function for the open loop system:

$$K(s)=\left(\frac{\omega_c}{s}\right)^{\beta} \tag{9}$$

To obtain an open loop system in the form (9) the controller transfer function should have a structure:

$$C(s)=-k_c\frac{s(1+sT)}{s^{\alpha}}=-k_c s^{1-\alpha}(1+sT) \tag{10}$$

where α is real number, and $\tau_2 = T$.

Open loop transfer function is given by:

$$K(s)=C(s)G(s)=\frac{kk_c}{\tau}\frac{e^{-sh}}{s^{\alpha}} \tag{11}$$

Knowing$(j\omega)^{\alpha}=|\omega|^{\alpha}e^{j\alpha\pi/2}$ gain and phase for a transfer function (7) was written:

$$|K(j\omega)|=\frac{kk_c}{\tau}\frac{1}{\omega^{\alpha}}\qquad \phi(\omega)=\arg K(j\omega)=-h\omega-\alpha\frac{\pi}{2} \tag{12}$$

For a gain ω_g and ω_p phase crossover frequency terms was given:

$$|K(j\omega_g)|=1\qquad \phi(\omega_p)=\arg K(j\omega_p)=-\pi. \tag{13}$$

Using (12) the equations (13) was rewritten as:

$$\frac{kk_c}{\tau\omega_g^{\alpha}}=1\qquad -h\omega_p-\alpha\frac{\pi}{2}=\pi \tag{14}$$

By solving equations (14) gain ω_g and ω_p phase crossover frequency:

$$\omega_g^{\alpha}=\frac{kk_c}{\tau}\qquad \omega_p=\frac{(2-\alpha)\frac{\pi}{2}}{h} \tag{15}$$

For a given $A_m$gain and ϕ_m phase margins:

$$\frac{kk_c}{\tau\omega_p^{\alpha}}=\frac{1}{A_m}\qquad \phi_m=\pi-h\omega_g-\alpha\frac{\pi}{2} \tag{16}$$

Transformed equation (16) were written in a form:

$$\omega_p=\left(\frac{A_m kk_c}{\tau}\right)^{1/\alpha}\qquad \omega_g=\frac{(2-\alpha)\frac{\pi}{2}-\phi_m}{h} \tag{17}$$

Using given equationsgain margin $A_m$was given by:

$$A_m=\left(\frac{(2-\alpha)\frac{\pi}{2}}{(2-\alpha)\frac{\pi}{2}-\phi_m}\right)^{\alpha} \tag{18}$$

Nonlinear equation (18) is handling$A_m$gainand ϕ_m phase margins and fractional order of the considerable controller (4)α.

Parameter α can be determined by solving equation (18) using computer methods.

By solving one of the first equations of (14) or (16):

$$k_c=\frac{\tau\omega_g^{\alpha}}{k}=\frac{\tau\omega_p^{\alpha}}{kA_m} \tag{19}$$

where:k – gain of the transfer function (1).

3. RESULTS

Example 1: $T_1 \gg h$

Consider the feedback control system shown in Fig. 1 in which the process to be controlled is described by transfer function:

$$G(s)=\frac{4}{(1-6s)(1+1.2s)}e^{-0.25s} \tag{20}$$

Determine controller parameters for a given transfer function (20) to stabilize the closed loop control system. In that case:$k = 5, T_1 = 6, T_2 = 1.2; h = 0.25$.

Using equations (4)-(6) and (8) transfer function of the controller is given:

$$c(s)=-3\frac{(1+2s)}{2s}(1+1.2s) \tag{21}$$

Fig. 2 shows step response for control system presented in Fig.1, where the process to be controlled is described by (20), controller (21), and fractional controller proposed using Bode transfer function (Nartowicz, 2011), where:

$C(s) = -2.56/s^{0.13385} - 3.07s^{0.8661}$.

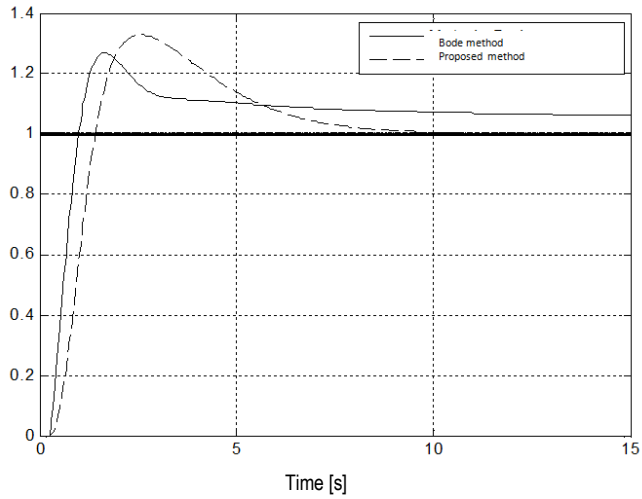
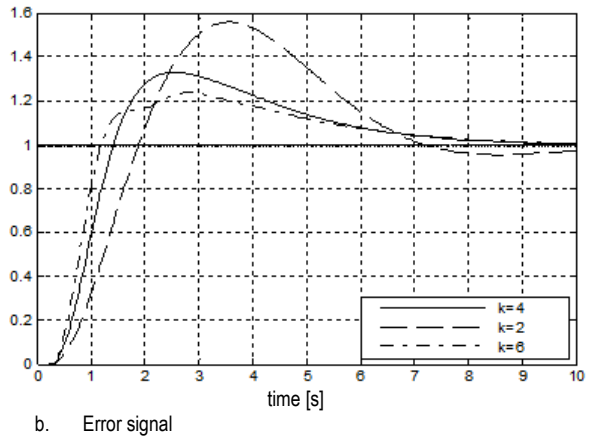


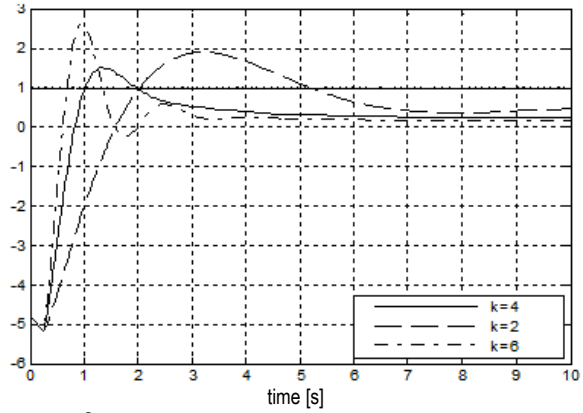
Fig. 2. Step response of the closed loop system with object (20) and controller (21), and fractional order controller synthesis with Bode method

Overshoot of step response drawn with method proposed in the paper is 10% bigger than drawn using Bode's method. Settling time is 8 sec. The time is much longer while Bode method using. Fig 3 shows step response for control system with controller (21), fractional order controller proposed with Bode method for a few values of parameter k (Fig.3a), and few values of delay h for a controlled process described with transfer function (20)

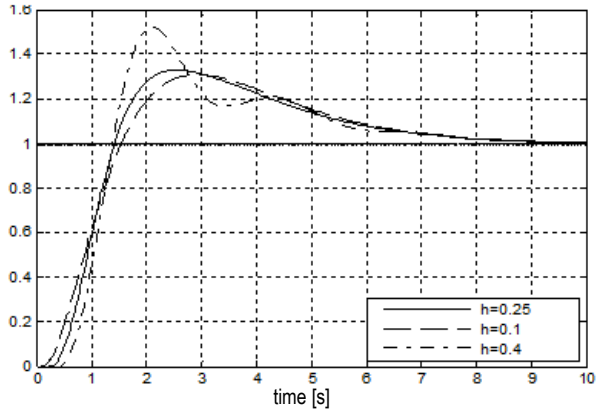
a. Step response



b. Error signal



c. Step response



d. Error signal

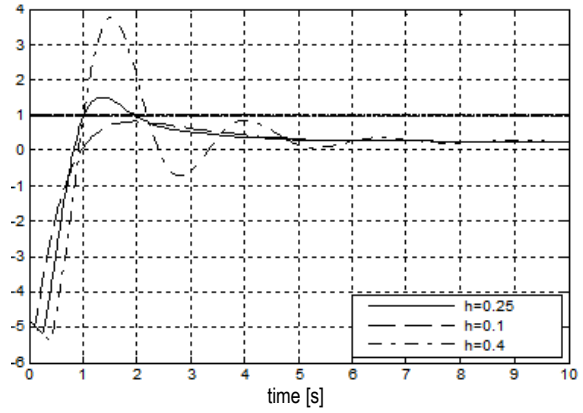


Fig. 3. Step response of the closed loop system with object (20) and controller (21) for a few different values of parameters *k* and *h*: a), b) step response; c), d) error signal

Fig 3b and 3d shows error signal of the controlled system. Overshoot for $k = 2$ is 55% bigger than for original value ($k = 4$), and moreover settling time is growing. There is no overshoot change while k is bigger, settling time also the same.

For a few values of time delay h settling time is const, but when the time is growing, the overshoot is also growing, for $h = 0.4$ even to 50%.

Example 2: $\mathbf{T_1 > h}$

Consider the feedback control system shown in Fig.1 in which the process to be controlled is described by transfer function:

$$G(s) = \frac{0.55}{(1-62s)(1+36s)} e^{-10s} \tag{22}$$

Determine controller parameters for a given transfer function (22) to stabilize the closed loop control system.

In that case: $k = 0.55, T_1 = 62, T_2 = 36; h = 10.$

Using equations (4)-(6) and (8) transfer function of the controller is given:

$$c(s) = -5.64 \frac{(1+62s)}{62s}(1+36s) \tag{23}$$

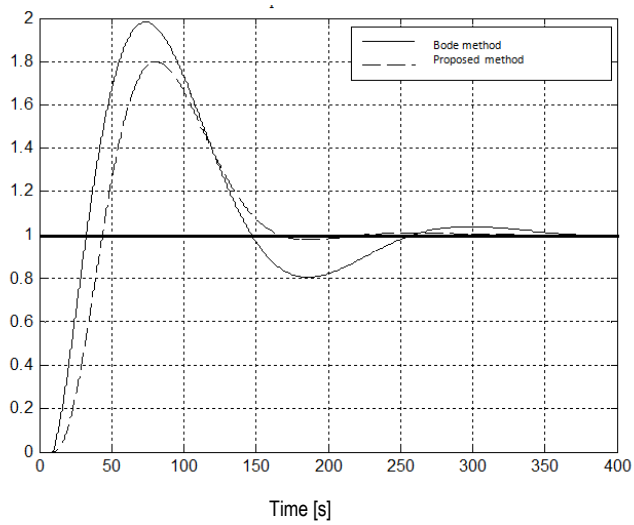


Fig. 4. Step response of the closed loop system with object (20) and controller (21), and fractional order controller syntheses with Bode method

Fig. 4 shows step response for control system presented in Fig. 1, where the process to be controlled is described by (22), controller (23), and fractional controller proposed using Bode transfer function:

$$C(s) = -2.9358/s^{0.13385} - 105.6888s^{0.8661}.$$

Overshoot of step response drawn with method proposed in the paper is 20% bigger than drawn using Bode method. Fig. 5 shows step response for control system with controller (23), fractional order controller proposed with Bode method for a few values of parameter k (Fig. 5a), and few values of delay h for a controlled process described with transfer function (22). Settling time is ab. 160 sec, while the time for Bode method is 250 sec.

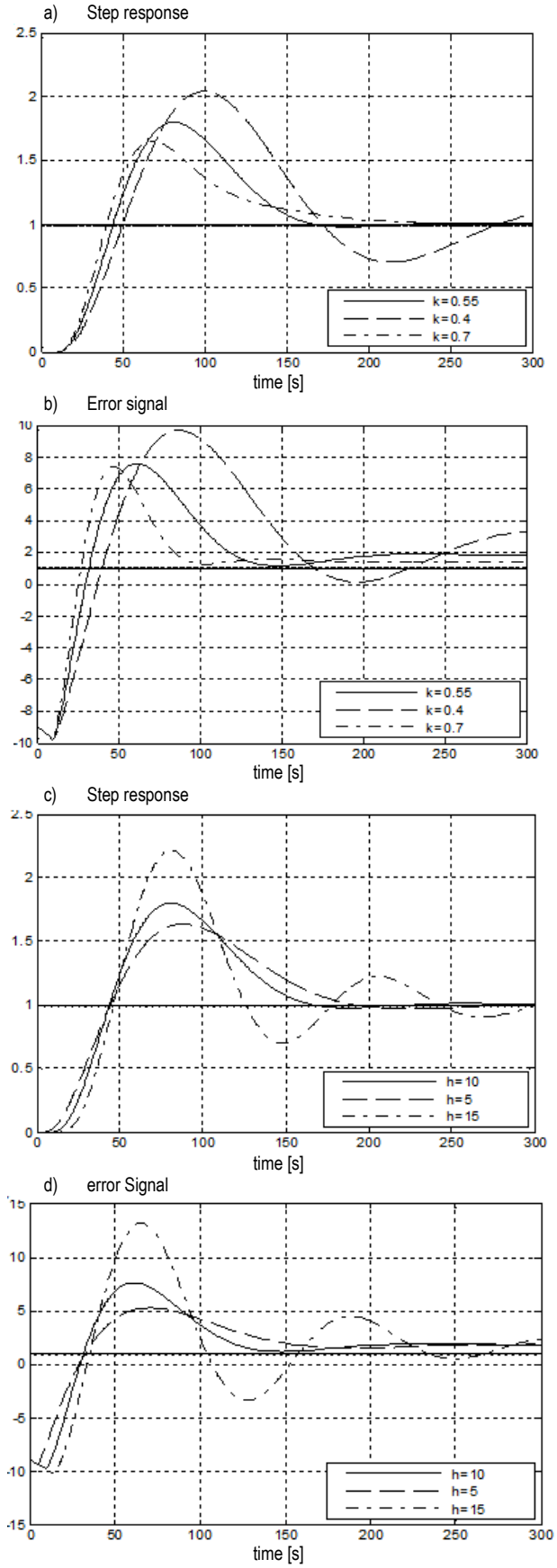


Fig. 5. Step response of the closed loop system with object (22) and controller (23) for a few different values of parameters *k* and *h*: a), b) step response; c), d) error signal

Fig. 5b and 5d show error signal of the controlled system. Overshoot for $k = 0.4$ is 100%, and moreover settling time is growing. For bigger value k overshoot value is decreasing, for $k = 0.7$even to 60%. For a few values of time delay h settling time is const when time delay is decreasing.

4. CONCLUSIONS

Skogestad method is well knows in the literature. That paper is generalization of Skogestad method for a class of unstable plant described with transfer function (7). For an unstable plant Skogestadanalytical method is extended for an negative gain of the controller, what makes positive feedback while simulating.As in Skogestad method, in that paper simulation for different time T_1, T_2, and time delayh was drawing. Note that for simulation $\tau_C$was posited as:$\tau_C = h$ In the paper two cases was presented: $T_1 \gg h$, and$T_1 > h$. Proposed method was compared with synthesis method of fractional order controller proposed in other author papers.Computer method for synthesis of fractional controllers is given. The considerations are illustrated by numerical exampleandresults of computer simulation with MATLAB/ Simulink.

REFERENCES

1. **Barbosa R. S., Machado J. A., Ferreira I. M.** (2004), Tuning of PID controllers based on Bode's ideal transfer function, *Nonliner Dynamics*, Vol. 38, 305-321.
2. **Bode H.** (1945), *Network Analysis and Feedback Amplifier*, Van Nostrand, New York.
3. **Busłowicz M., Nartowicz T.** (2009), Fractional controllel synthesis for a class of inertail plant with delay, *Pomiary Automatyka Robotyka*, 398-405.
4. **Farkh R., Laabidi K., Ksouri M.** (2009), Computation of All Stabilizing PID Gain for Second-Order Delay System, *Mathematical Problems in Engineering*, Vol. 2009.
5. **Johnson M. A., Moradi M. H.** (2005), *PID Control: New Identification and Design Methods*,Springer.
6. **O'Dwyer A.** (2003), *Handbook of PI and PID Controller Tuning Rules*, Imperial College Press, World Scientific, New Jersey, London, Singapore, Hong Kong.
7. **Oustaloup A., Sabatier J., Lanusse P., Malti R., Melchior P., Moreau X., Moze M.** (2008), An overview of the CRONE approach in system analysis, modeling and identification, observation and control, *Proc. 17th World Congress IFAC*, Soul, 14254-14265.
8. **Rivera D.E., Morari M., Skogestad S.** (1986), Internal model contro, 4. PID controller design, *Ind. Eng. Chem. Res.*, 25(1).
9. **Silva G. J., Datta A., Bhattacharyya S. P.** (2005), *PID Controllers for Time Delay Systems*, Birkhäuser, Boston.
10. **Skogestad S.** (20001), *Probably the best simple PID tuning rules in the world*, AIChE Annual Meeting, Reno, Nevada.
11. **Tan K. K., Wang Q. G., Hang C. C.** (1999), *Advances in PID control*, Springer Verlag, London.
12. **Xiang C., Wang Q.G., Lu X., Nguyen L.A., Lee T.H.** (2007), Stabilization of second-order unstable delay processes by simple controllers, *Journal of Process Control*, 17, 675–682

NULL-CONTROLLABILITY OF LINEAR SYSTEMS ON TIME SCALES

Ewa PAWŁUSZEWICZ*

*Faculty of Mechanical Engineering, Bialystok University of Technology, ul. Wiejska 45C, 15-351 Białystok, Poland

e.paluszewicz@pb.edu.pl

Abstract: The purpose of the paper is to study the problem of controllability of linear control systems with control constrains, defined on a time scale. The obtained results extend the existing ones on any time domain. The set of values of admissible controls is a given closed and convex cone with nonempty interior and vertex at zero or is a subset of R^m containing zero.

Key words: Linear Control System, Control Constrains, Null-Controllability, Time Scale, Delta Derivative

1. INTRODUCTION

The paper deals with linear control systems $x^{\Delta}(t) = A(t)x(t) + B(t)u(t)$ defined on a time scale T. We assume that $u: T \to U$, where U is a subset of R^m. In systems theory linear systems have (by definition) $U = R^m$, but in many practical situations the set U should be bounded, see for example Abel (2010). A restriction on controls brings some difficulties with controllability conditions. For example (The example comes from Sontag (1998)), let us take a system $dx(t)/dt = -\mathrm{x} + \mathrm{u}$ and let $\mathrm{u}: [0, +\infty) \to (-1,1)$. It is easy to see that the pair (A, B) is controllable, but the system with restricted controls is not, since it is impossible to transfer the state $\mathrm{x}_0 = 0$ to $\mathrm{x}_\mathrm{f} = 2$ (we have $dx(t)/dt < 0$ whenever $x(t) \in (1,2)$).

For continuous-time linear systems, the problem of controllability with control constrains has been sudied for example in Ahmed (1985), Chukwu and Lenhart (1991), Klamka (1991), Path et al., (2000), Schmitendorf and Barmish (1981), Sontag (1998). For discrete-time case – in Benzaid and Lutz (1980), Path et al., (2000).

Analysis on time scales is nowadays recognized as the right tool to unify and extend the existing results for continuous- and discrete-time dynamical systems to the nonhomegonous time domains, see for example Bartosiewicz and Pawłuszewicz (2006), Bartosiewicz and Pawłuszewicz (2008), Davis et al., (2009), DaCunha and Davis (2011), Gravagne et al., (2009), Ferreira and Torres (2010), Pawłuszewicz and Torres (2010).

A time scale is a model of time. Besides the standard cases of the whole real line (continuous-time case) and all integers (discrete-time case) there are many other models of time included that can be partially continuous and partially discrete, q-scales, quantum time scales (objects with non-uniform domains), and many others – see Bohner and Peterson (2001). However, discrete-time systems on time scales are based on the difference operator and not on the more conventional shift operator. One of the main concepts in the time scale analysis is the delta derivative, which is a generalization of the classical (time) derivative in the continuous time and the finite forward difference in the discrete time. Similarly, the integral of a real function defined on a time scale is an extension of the Riemann integral in the continuous time and the finite sum in the discrete time. As a consequence, differential equations as well as difference equations are naturally accommodated in this theory.

The goal of this paper is to study conditions under which a linear system defined on a time scale with control constrains is controllable. For this aim, in Section 2 gives general information about solution of considered class of systems. Section 3 is devoted to the investigation of the problem of null-controllability of time-varying systems with control constrains. It also presents the necessary and sufficient conditions for global null controllability for the systems with control constrains on homogenous time scale. In Section 4 linear time-invariant systems with control constrains are studied. The main result of this Section is that such a system is controllable if and only if the Kalmann rank condition is satisfied.

The necessary elements of delta-measurability and nonlinear theory on time scales are presented in Appendix. At this moment we only introduce the following notation: if $a, b \in T, a \le b$, then $[a, b]_T$ denotes the intersection of the real closed interval $[a, b]$ with T. A similar notation is used for open, half-open, or infinite intervals.

2. LINEAR SYSTEMS ON TIME SCALES

Let T be any time scale and let $A \subset T$. Recall (see Cabada and Vivero (2005)) that a function $f: T \to R$ is absolutely continuous on a time scale T if and only if f is continuous and of bounded variation on T and f maps every Δ-null subset of T into a null set. Let L^p_Δ denote spaces linked to the Lebesgue Δ-measure and absolutely continuous function on arbitrary closed interval of time scale T. We say that $f \in L^p_\Delta(E)$ provided that $\int |f(t)|^p \Delta t < \infty$ if $p \in R, p < \infty$, Agrawal et al (2006), Cabada and Vivero (2006).

Let I be the identity $n \times n$ - matrix and $Z \in R^{n \times n}$. Recall that the matrix Δ-differential system defined on time scale T:

$$X^{\Delta}(t) = Z(t)X(t) \qquad X(t_0) = I \tag{1}$$

for any $X \in R^n$, $\mathrm{t} \in [\mathrm{t}_0, \mathrm{sup}\mathrm{T})_\mathrm{T}$, has a unique solution $\mathrm{X}(t) = \Phi_{Z(t)}(t, t_0)$. Using the same arguments as in Bartosiewicz and

Pawłuszewicz (2006) for time-invariant case, we can show that for every $t, s, r \in T$ such that $t \leq s \leq r$ the following hold:

- $\Phi_0(t,s) = I, \Phi_{Z(t)}(t,t) = I;$
- If $Z(t)$ is an regressive matrix, i.e. if matrix $\big(I + \mu(t)\big)Z(t)$ is invertible, then $\Phi_{Z(t)}(t,s) = (\Phi_{Z(t)}(s,t))^{-1}$;
- $\Phi_{Z(t)}(t,s)\Phi_{Z(t)}(s,r) = \Phi_{Z(t)}(t,r)$.

If Z is time-invariant, the solution of the equation (1) is given by an exponential matrix function on time scale T $\mathrm{X(t)} = \mathrm{e_Z(t,t_0)}$, see: Bartosiewicz and Pawluszewicz (2006), Jackson (2007).

Let us consider a linear control system defined on T:

$$x^{\Delta}(t) = A(t)x(t) + B(t)u(t), \;\; x(t_0) = x_0 \qquad (2)$$

where $A(t) \in R^{n\times n}$ and $B(t) \in R^{n\times m}$ are rd-continuous matrices on T, i.e. each entry of these matrices is an rd-continuous function on T. Also $x(t) \in \Sigma \subset R^n$ and $u(t) \in U \subset R^m$. Let us choose a control u. The trajectory of system (2) is a function $\psi(\cdot, t_0, x_0, u)\colon [t_0, supT)_T \to \Sigma$ that is the unique solution of (2), provided it is defined on $[t_0, supT)_T$ and for all $t \in [t_0, supT)_T$, $x(t) \in \Sigma$. This solution for all $t \in [t_0, supT)_T$ is given by (see Bartosiewicz and Pawłuszewicz (2006)):

$$x_f = \Phi_A(t_f, t_0)x_0 + \int_{t_0}^{t_f} \Phi_A\left(t_f, \sigma(s)\right) B(s)u(s)\Delta s \qquad (3)$$

If *A* is a regressive matrix, i.e. if the matrix $\big(I + \mu(t)\big)A(t)$ is invertible, then (3) describes both forward and backward trajectories of (2).

We say that a control u is admissible for $x_0 \in R^n$ if there exists a trajectory of system (2) from $\mathrm{x_0}$ corresponding to u. The set of all admissible controls (for x_0) will be denoted by U_{ad}.

Let $E = [t_0, t_f]_T$. Assume that the set of the values of admissible controls U is a given closed and convex cone with nonempty interior and vertex at zero. Thus the set of admissible controls U_{ad} for system (2) has the form $L^2_{\Delta}(E,U)$, i.e. is a Banach space endowed with the norm defined for every $u\colon E \to U$ as:

$$||u||_{L^2_{\Delta}} := \left[\int_E |u|^2(t)\Delta t\right]^{1/2}$$

3. TIME-VARYING SYSTEMS WITH CONTROL CONSTRAINS

Let $\Sigma \subseteq \mathrm{R^n}$. We say that system (2) is:

- *U*-controllable on a time interval $[t_0, t_f]_T$ if, for any $x_0 \in \Sigma$ and any x_f there exists a control $u \in L^2_{\Delta}(E,U)$ such that $\psi\big(t_f, t_0, x_0, u\big) = x_f, x_f \in \Sigma$.
- *U-controllable* if it is *U*-controllable on every time interval $[t_0, t_f]_T$.
- locally *U-controllable* on $[t_0, t_f]_T$ if, for the given trajectory $\psi(\cdot, t_0, x_0, u) = x(\cdot)$ of (2) with $u_0 \in L^2_{\Delta}(E,U)$ and $x(t_0) = x_0 \in \Sigma$ there exists a neighborhood V_{x_0} of x_0 such that, for any $z \in V_{x_0}$ there exists an admissible control u_0 such that $\psi\big(t_f, t_0, x_0, u\big) = x_f \in V_{x_0}$.

If $x_f = 0$, then we have respectively null *U*-controllability on a time interval $[t_0, t_f]_T$ null *U*-controllability, local null *U*-controllability.

Our goal is to show certain properties characterizing the null U-controllability. Let us assume that there exists a unique evolution operator ϑ defined as $\{\bar{\Phi}_A(t,s)\colon t_0 \leq s \leq t \leq t_f;\ t_0, s,\ t, t_f \in T\}$ and corresponding to the $\mathbb{A} = \{A(t)\colon t \in [t_0, supT)_T\}$ in (2). The ideas of proofs of next two propositions come from Chuwku and Lenhart (1991).

Proposition 1. Let us assume that system (2) is null U-controllable on $[t_0, t_f]_T$. Then there exists a bounded operator $H\colon \Sigma \to L^2_{\Delta}(E,U)$ such that, with the admissible control $u = Hx_0$, the solution of (2) satisfies $x\big(t_f\big) = \psi\big(t_f, t_0, x_0, Hx_0\big) = 0$.

Proof: For arbitrary initial state x_0, let $T_t\colon \Sigma \times U_{ad} \to \Sigma$ be a map defined as $T_t(x_0, u) := \psi(t, t_0, x_0, u)$ for any $t \in [t_0, t_f]_T$. Then T_t is the continuous linear map with respect to u. Let us consider also a map $S_t\colon L^p_{\Delta}(E) \to \Sigma$ defined as:

$$S_t(u) := \int_{t_0}^{t} \Phi_A\big(t, \sigma(s)\big)B(s)\Delta s$$

for any $t \in [t_0, t_f]_T$. Note that S_t is linear, bounded and $T_t(x_0, u) = \Phi_A(t, t_0)x_0 + S_t(u)$. Since for all $t_f \in T$ $\Phi_A\big(t_f, t_0\big)\Sigma \subset S_{t_f}\left(L^2_{\Delta}(E,U)\right)$ then, from definition, this condition is equivalent to the null *U*-controllability of (2).

Let us consider a map $\varsigma\colon \mathrm{N^{\perp}} \to S_{t_f}\left(L^2_{\Delta}(E,U)\right)$, where $\mathrm{N^{\perp}}$ denotes the orthogonal complement of the null space of S_{t_f}. Define $Hx_0 := -\varsigma^{-1}\Phi_A(t, t_0)x_0$. Note that by Banach Theorem and closed graph theorem this operator is bounded (see Musielak (1989)). Moreover:

$$\psi\big(t_f, t_0, x_0, Hx_0\big) =$$
$$\Phi_A\big(t_f, t_0\big)x_0 + S_{t_f}(-\varsigma^{-1})\Phi_A\big(t_f, t_0\big)x_0 = 0$$

Proposition 2: Suppose that zero belongs to the interior of the set of admissible controls. If the system (2) is null U-controllable, then it is locally null *U*-controllable.

Proof follows from the fact that map H defined in Proposition 1 is continuous at 0. This implies the existence of an open set $\mathrm{W_0}$ containing 0 and such that $H(W_0) \subset V \subset intU_{ad}$. Hence, the state 0 can be achieved from any $x_0 \in W_0$ using $u = Hx_0$.

Other conditions for null U-controllability can be obtained under exponential stability assumption. Recall that system (1) defined on unbounded time scale T with bounded graininess function $\mu\colon T \to R_+ \cup \{0\}$ is exponentially stable if there exists a constant $\alpha > 0$ such that for every $t_0 \in T$ there exists $K = K(t_0) \geq 1$ with

$$||\Phi_{Z(T)}(t, t_0)\mathrm{x}(t_0)|| \leq Ke^{-\alpha(t-t_0)}||\,\mathrm{x}(t_0)||$$

for any $\mathrm{t} \in [\mathrm{t_0}, \mathrm{sup})_{\mathrm{T}}$; $||.||$ denotes the classical Euclidian norm.

Proposition 3. If system (2) is null U-controllable on each time interval $[t_0, t + \sigma(t))_T$, $[t_0, supT)_T$, and the system $x^{\Delta}(t) = A(t)x(t)$, $x(t_0) = x_0$, is exponentially stable, then the system (2) is null U-controllable.

Proof: By Proposition 2, null U-controllable of the given system implies local null U-controllability of this system. Then there exists a neighborhood V_{x_0} of x_0 such that all states from V_{x_0} can be steered to 0 with $u \in U_{ad}$.

Let $z \in V_{x_0}$. Exponential stability of the system $x^{\Delta}(t) = A(t)x(t)$, $x(t_0) = z$, implies existence t_{f_1} such that the solution of this equation satisfies $x(t_{f_1}) = x_{f_1} \in V_{x_0}$. If we take as an initial data (t_{f_1}, x_{f_1}), then there exists $t_{f_2} \in (t_{f_1}, supT)_T$, such that, for some $\bar{u} \in U_{ad}$ holds $\psi\big(t_{f_1}, t_{f_1}, x_0, \bar{u}\big) = x_{f_1}$ and

$\psi(t_{f_2}, t_{f_1}, x_0, \bar{u}) = 0$. Taking as a control multifunction $\upsilon(t) \coloneqq u(t)$ for $t \in [t_0, t_{f_1}]_T$ and $\upsilon(t) := \bar{u}(t)$ for $t \in [t_{f_1}, t_{f_2}]_T$ (with swithching time t_{f_1}), state z can be transferred to state 0 in time $t \in [t_0, t_{f_2}]_T$.

Let $X = \{x \in R^n : Lx = c\}$ where L is a given $p \times n$-matrix of the rank p and $c \in R^n$ is a given vector. For any vector $a \in R^p$ let $H_U(a) = \sup\{w^T L^T a : w \in U\}$ denote the support function of a set U.

Theorem 1: Let T be a time scale with a constant graininess μ. The system (2) is globally null U-controllable if and only if for every admissible control u holds

$$\int_{t_0}^{\infty} H_U\left(B^T(s)\Phi_A^T(t_0,\sigma(s))L^T a\right)\Delta s = +\infty \qquad (4)$$

where M^T denotes the transposition matrix of M.

Proof: For $T = R$ the theorem was proved in Schmitendorf and Barmish (1981), Klamka (1991) and, using more general approach, in Path et al., (2000). The proof for $T = hZ,\ h > 0$, mimics the one given in Path et al., (2000) for discrete-time systems.

In continuous-time case, relation (4) can be formulated in terms of the solution of the adjoint equation Ahmed (1985), Path et al., (2000). For time scale system (2) such reformulation requires an assumption that matrix A is regressive for all $t \in T^{\kappa}$.

4. LINEAR TIME-INVARIANT SYSTEMS WITH CONTROL CONSTRAINS

Let us consider a linear time-invariant control system defined on a time scale T:

$$x^{\Delta}(t) = Ax(t) + Bu(t) \qquad x(t_0) = x_0 \qquad (5)$$

where: $A \in R^{n\times n}$, $B \in R^{n\times m}$, $x(t) \in \Sigma \subset R^n$, $u(t) \in U$. As previous, we assume that the set of values of admissible controls $\mathrm{U_{ad}}$ is a given closed and convex cone with nonempty interior and vertex at zero. The matrix:

$$Q_{t_f} = \int_{t_0}^{t_f} e_A(t_0, s)BB^T e_A^T(t_0, s)\Delta s$$

is called the controllability gramian. If there exists $t_f \in T$ such that the matrix Q_{t_f} is nonsingular and A is a regressive matrix, then using control:

$$\bar{u}(t) = -B^T e_A^T(t_0, \sigma(s))\, Q_{t_f}^{-1}[e_A(t_f, t_0)x_0 - x_f]$$

every state $\mathrm{x_f = x(t_f)}$ can be achieved from an initial state $\mathrm{x_0}$.

Proposition 4. Let $0 \in intU_{ad}$. If the system (5) is controllable and matrix A is regressive, then it is locally null U-controllable.

Proof: If $t_f \in T$ is arbitrary, then there exists a control $\bar{u}(s) = -B^T e_A^T(t_0, \sigma(s))Q_{t_f}^{-1} e_A(t_f, t_0)x, s \in [t_0, t_f]_T$, such that state x can be steered to 0 in a finite time. Since map $t \to e_A(t, t_0)$ is rd-continuous, then there exists a constant K such that $||\bar{u}(s)|| \le K||x||, s \in [t_0, t_f]_T$. Hence the thesis.

Let T be an unbounded time scale. Recall that system (5) is stabilizable (see Bartosiewicz et al.,, (2007)) if there exists a state feedback $u(t) = Fx(t)$, for $F \in R^{m\times n}$, such that the closed loop system $x^{\Delta}(t) = (A + BF)\mathrm{x}(t)$ is exponentially stable. The set of exponential stability on a time scale T is defined as (see Pötzsche et al., (2003)):

$$S(T) \coloneqq S_C(T) \cup S_R(T)$$

where: $S_C(T) =$
$\left\{\lambda \in C : \lim_{t\to\infty} sup\, \frac{1}{\tau - t_0}\int_{t_0}^{\tau} \lim_{s\to\mu(\xi)} \frac{\log|1+s\lambda|}{s} \Delta\xi < 0\right\}$
$S_R(T) = \{\lambda \in R : \forall \tau \in T\ \exists \xi \in T, \xi > \tau : 1 + \mu(\xi)\lambda = 0\}$

For the arbitrary time scale T it holds that $S_C(T) \subseteq \{\lambda \in C : Re\lambda < 0\}$ and $S_R(T) \subset (-\infty, 0)$.

Theorem 2. (Pötzsche et al., (2003)) The following holds:

a) If (5) is exponentially stable then $spec(A) \subset S(T)$.
b) If A is diagonalizable, then (5) is exponentially stable if and only if $spec(A) \subset S(T)$

where: $spec(A)$ denotes the set of all eigenvalues of A.

Since the null U-controllability is a particular case of U-controllability, we can reformulate the result from Bartosiewicz et al., (2007) as follows:

Theorem 3. Assume that $\mu(t)$ is bounded. If system (5) is null U-controllable, then it is stabilizable.

Lemma 1. If system (5) is stabilizable then it is U-controllable.

Proof: The idea of the proof is based on Zabczyk (1995). Using classical arguments one can easily deduce that if the pair (A, B) is controllable, then there exists a matrix $F \in R^{m\times n}$ and a vector $v \in R^m$ such that the pair $(A + BF, Bv)$ is controllable.

Let $P \in R^{n\times n}$ be a nonsingular matrix such that $PAP^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$, $PB = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}$ and the pair (A_{11}, B_1), $A_{11} \in R^{l\times l}, B_1 \in R^{l\times m}$, is controllable.

Since system (5) is stabilizable, then there is a matrix $F \in R^{m\times n}$, such that the closed loop system $x^{\Delta}(t) = (A + BF)x(t) + (Bv)u(t)$ is exponentially stable. The characteristic polynomial of $A + BF$ is of the form:

$$\begin{aligned} p_{A+BF}(\lambda) &= \det[\lambda I - (A + BF)] \\ &= \det(\lambda I - PAP^{-1} - PBFP^{-1}) \\ &= \det[\lambda I - (A_{11}P^{-1} + B_1 F)]\det(\lambda I \\ &\quad - A_{22}), \quad \lambda \in C. \end{aligned}$$

So, for any F, $spec(A_{22}) \subset spec(A + BF) \subset S(T)$ and, if there exists $\alpha > 0$ such that for every $t_0 \in T$ there is $K \ge 1$ then $\alpha \le -sup\{Re\lambda : \lambda \in spec(A_{22})\}$. Hence the contradiction with stabilizability of (5).

Exponential stability and Proposition 2 imply the following.

Proposition 5. If $0 \in U_{ad}$ system (5) is U-controllable and exponentially stable, then it is null U-controllable.

Let $A_{x_0,U_{ad}}(t_0, t_f)$ be a *reachability set* of system (2), i.e. a set of all points that can be reached at time t_f starting from $x_0 = x(t_0)$. The set of all points that can be reached from x_0 at t_0 in a finite time will be denoted as $A_{x_0,U_{ad}}(t_0)$. The image of the map $u \mapsto \psi(t_f, t_0, x_0, u)$, i.e. the set $A_{x_0,U_{ad}}(t_0, t_f)$ is a linear subspace of R^n and:

$$A_{x_0,U_{ad}}(t_0, t_f) = \Phi_A(t_f, t_0)x_0 + A_{0,U_{ad}}(t_0, t_f)$$

Using classicall arguments, similarly as in Sontag (1998) we can show the following:

- if U is convex, then $A_{o,U_{ad}}(t_0)$ is a convex subset of R^n;
- suppose that A is regressive. If system (5) is U-controllable and U_{ad} is a neighborhood of $0 \in R^n$ then $A_{o,U_{ad}}(t_0)$ is an open subset of R^n .

Collorary 1. Suppose that $0 \in intU$ system (5) is stabilizable and matrix A is diagonalizable. Then system (5) is null U-controllable.

For each eigenvalue λ of the matrix A, let $J_{k,\lambda} \coloneqq \ker(\lambda I - A)^k$ and $J^R_{k,\lambda} \coloneqq \{Rev : v \in J_{k,\lambda}\}$. Let $L = \cup_{Re\lambda \geq 0} J^R_{k,\lambda}$ and $M = \cup_{Re\lambda < 0} J^R_{k,\lambda}$. If C is an open convex subset of R^n, L is a subset of R^ncontained in C, then $C + T = C$, see Sontag (1998).

Lemma 2. Let A be an $n \times n$ −matrix with eigenvalues $\lambda_1, \dots, \lambda_n$. If system (5) is U-controllable and U_{ad} is a neighborhood of 0, then $L \subseteq A_{o,U_{ad}}(t_0)$.

Proof: Without loss of generality we assume that U_{ad} is a convex neighborhood of 0. Using mathematical induction with respect to k, we show that $J^R_{k,\lambda} \subseteq A_{o,U_{ad}}(t_0)$.

For $\mathrm{k} = 0$, the case is trivial. Let us assume that $J^R_{k-1,\lambda} \subseteq A_{o,U_{ad}}(t_0)$, $\lambda = \alpha + i\beta, \alpha, \beta \in R$, and take any $\bar{v} \in J_{k,\lambda}$, $\bar{v} = \bar{v}_1 + i\bar{v}_2$. Then for any $\mathrm{t} \in [\mathrm{t}_0, \sup)_T$, $e_\lambda(t, t_0) = e_\alpha(t, t_0)$. Since $0 \in \mathrm{intU_{ad}}$ we can choose any $\delta > 0$ such that $v_1 \coloneqq \delta\bar{v}_1 \in A_{o,U_{ad}}(t_0)$. Moreover, since $v \in ker(\lambda I - A)^k$ and $e_A(t, t_0) = \sum_{k=0}^{\infty} A^k h_k(t, t_0)$, where $h_0(t, t_0) \equiv 1$, $h_{k+1}(t, t_0) = \int_{t_0}^{t} h_k(\tau, t_0)\Delta\tau$ (see Mozyrska and Pawłuszewicz (2008)), then:

$$e_A(t, t_0)v = \left[\sum_{k=0}^{\infty} A^k h_k(t, t_0)\right] v = v + w$$

with $w \in J_{k-1,\lambda}$. So,

$$e_\alpha(t, t_0) = e_\lambda(t, t_0) = e_\lambda(t, t_0)e_{(A-\lambda I)}(t, t_0)v - e_\lambda(t, t_0)w = e_\lambda(t, t_0)e_{(A-\lambda I)}(t, t_0)v - e_\alpha(t, t_0)w$$

Moreover, since $w = w_1$+iw_2, then:

$$Re(e_\alpha(t, t_0)v_1) = e_A(t, t_0)v_1 + e_{\alpha I}(t, t_0)w_1 \in A_{x_0,U_{ad}}(t_0, t) + J^R_{k,\lambda} \subseteq A_{0,U_{ad}}(t_0)$$

L is the sum of the spaces $J^R_{k,\lambda}$ over all eigenvalues λ with the real part nonnegative, and each of these spaces is included $A_{o,U_{ad}}(t_0)$, so the sum of the L's is included in $A_{o,U_{ad}}(t_0)$.

The ideas of the next Lemma and next Theorem come from Sontag (1998).

Lemma 3. Let $supT = \infty$. If system (5) is U-controllable, U_{ad} is a convex, bounded neighborhood of 0, then there exists a set N such that $A_{o,U_{ad}}(t_0) = N + L$ and N is bounded, convex and open relative to M.

Proof: Note that (see Sontag (1998)):

$$(A_{o,U_{ad}}(t_0) \cap M) + L \subseteq A_{o,U_{ad}}(t_0) + L = A_{o,U_{ad}}(t_0)$$

and $A_{o,U_{ad}}(t_0) \supseteq (A_{o,U_{ad}}(t_0) \cap M) + L$

Let $N \coloneqq A_{o,U_{ad}}(t_0) \cap M$. Then N is open and convex. Let $\pi : R^n \to R^n$, $\pi(x + y) = x$ for $x \in M$, $y \in L$. If $v = x + y$, $Ax \in M$, $Ay \in L$, then $PAv = Ax = APv$. Let $x \in A_{o,U_{ad}}(t_0) \cap M$. Since $x \in A_{o,U_{ad}}(t_0)$, then there exists an admissible control u and $t \geq t_0$ such that $x = \int_{t_0}^{t} e_A\big(t, \sigma(s)\big)Bu(s)\Delta s$. On the other hand, since $x \in M$, $x = Px$, then:

$$x = Px = \int_{t_0}^{t} Pe_A\big(t, \sigma(s)\big)Bu(s)\Delta = \int_{t_0}^{t} e_A\big(t, \sigma(s)\big)x(s)\Delta s$$

where $x(s) = PBu(s) \in M \cap PB(U)$ for all $s \in T$. Since the restriction of A to M has all eigenvalues with a negative real part, then there are positive constants $\alpha, K > 0$ such that (see Pötzsche et al., (2003)): $|e_\lambda(t, t_0)| \cdot \big||x|\big| = ||e_A(t, t_0)x|| \leq Ke^{-\alpha(t-t_0)}||x||$ for $t \geq t_0$ and $x \in M$. Since $PB(U)$ is bounded, there is a constant C such that if x is also in $PB(U)$, then:

$$|e_\lambda(t, t_0)| \cdot \big||x|\big| = \big||e_A(t, t_0)x|\big| \leq Ce^{-\alpha(t-t_0)}\big||x|\big|, t \geq t_0$$

So, (see Sontag (1998)):

$$\big||x|\big| \leq C\int_{t_0}^{t} Ce^{-\alpha(t-t_0)}ds \leq \frac{C}{\alpha}(1 - e^{-\alpha t}) \leq \frac{C}{\alpha}$$

Hence N is bounded.

Theorem 4. Let $supT = \infty$ and U_{ad} be bounded a neighborhood of zero. Then $A_{o,U_{ad}}(t_0) = R^n$ if and only if:

- system (5) is controllable;
- matrix A has no eigenvalues with a negative real part.

Proof: If $A_{o,U_{ad}}(t_0) = R^n$ then (i) is obvious (see Bartosiewicz and Pawłuszewicz (2006)). If (ii) doesn't hold then L should be a proper subspace of R^n and $K \neq 0$.We may assume that U_{ad} is convex and bounded. Lemma 3 implies that $R^n = A_{o,U_{ad}}(t_0)$ is a subset of $L + N$ and N is bounded, hence the contradiction. If (i) and (ii) hold, then by Lemma 2, $R^n = L \subseteq A_{o,U_{ad}}(t_0)$.

Theorem 4 and Kalman controllability rank condition imply the following.

Collorary 2. Let $supT = \infty$ and U_{ad} be a bounded neighborhood of zero. Then $A_{o,U_{ad}}(t_0) = R^n$ if and only if $rank[\mathrm{B}, \mathrm{AB}, \dots \mathrm{A^{n-1}B}] = \mathrm{n}$.

5. CONCLUSIONS

The paper extends the conditions for constrained relative controllability for linear time-varying and time-invariant systems to the systems defined on different time models, also on nonhomogenous time domains. A calculus on time scales is used to achieve this goal. The existing necessary and sufficient conditions for null controllability of time varying systems were unify. The Kalman rank condition for time-invariant systems with control constrains was extended on systems defined to any unbounded from above time scale.

REFERENCES

1. **Abel D.L.** (2010), Constrains vs Controls, *The Open Cybernetics & Systemics Journal*, 4, 14-27.
2. **Agrawal R.P., Otero-Espinar V., Perera K., Vivero D.** (2006), Basic properties of Sobolev's spaces on time scales, *Advances in Difference Equations,* article ID 38121, 1-14.
3. **Ahmed N.U** (1985), Finite-time null controllability for a class of linear evolution equations on a Banach space with control constraints, *Journal of Optimization Theory and Applications*, 47(2), 129-158.
4. **Bartosiewicz Z., Piotrowska E., Wyrwas M.** (2007), Stability, stabilization and observers of linear control systems on time scales, *Proceedings of the 46th IEEE Conference on Decision and Control*, New Orleans, LA, USA, 2803-2808.
5. **Bartosiewicz Z., Pawluszewicz E.** (2006), Realizations of linear control systems on time scales, *Control & Cybern*, 35(4), 769–786.
6. **Bartosiewicz Z., Pawluszewicz E.** (2008), Realizations of nonlinear control systems on time scales, *IEEE Transactions on Automatic Control* , 53(2), 571-575.
7. **Benzaid Z., Lutz D.A.** (1988), Constrained controllability of perturbed discrete-time systems, *Int. J. Control*, 48(2), 655-673.

8. **Bohner M., Petrson A.** (2001), *Dynamic equations on time scales,* Birkhäuser.
9. **Bohner M., Petrson A.** (2003), *Advances in dynamic equations on time scales,* Birkhäuser.
10. **Cabada A, Vivero D.R. (**2005), Criterions for absolute continuity on times cales, *Journal of Difference Equations and Applications*, Vol. 11(11), 1013–1028.
11. **Cabada A., Vivero, D.R.** (2006), Expression of the Lebesgue - integral on time scales as a usual Lebesgue integral: application to the calculus of *Δ*-antiderivatives. *Math. Comput. Model,* 43(1–2), 194–207.
12. **Chukwu E.N., Lenhart S.M.** (1991), Controllability questions for nonlinear systems in abstract spaces, *Journal of Optimization Theory and Applications*, 68(3), 432-462.
13. **DaCunha J.J., Davis J.M.** (2011), A unified Floquet theory for discrete, continuous, and hybrid periodic linear systems, *Journal of Difference Equations,* 251, 2987-3027.
14. **Davis J.M., Gravagne I.A., Jackson B.J., Marks II R.J.** (2009), Controllability, observability, realizability, and stability of dynamic linear systems, *Electronic Journal of Differential Equation*, No. 37, 1-32.
15. **Deniz A.** (2009), Measure theory on time scales. MSc thesis, Graduate School of Engineering and Sciences of Izmir Institute of Technology, Turkey.
16. **Ferreira R.A.C., Torres D.F.M.** (2010), Isoperimetric Problems of the Calculus of Variations on Time Scales, *Nonlinear Analysis and Optimization Ii: Optimization*, 514, 123-131 .
17. **Gravagne I.A., Davis J.M., DaCunha J.J.** (2009), A unified approach to high-gain adaptive controllers, *Abstract and Applied Analysis,* 2009, 1-13.
18. **Jackson B.J.** (2007), A General Linear Systems Theory on Time Scales: Transforms, Stability, and Control, PhD Thesis, Baylor University.
19. **Klamka J.** (1991), Controllability dynamic systems, Kluwer.
20. **Mozyrska D., Pawłuszewicz E.** (2008), Functional series on time scales, *International Journal of Mathematics and Statisctics,* 2(S08), 95 -106
21. **Musielak J.** (1989), Wstęp do analizy funkcjonalnej, PWN (in Polish)
22. **Path V.N., Park J.Y., Jung I.H.** (2000), Stability and constrained controllability of linear systems in Banach spaces, *J.Korean Math. Soc*., 37(4), 593-611.
23. **Pawłuszewicz E., Torres D.F.M.** (2010), Avoidance control on time scales, *J. Optim. Theory Appl*., 145, 527-542.
24. **Pötzsche P., Siegmund S., Wirth F.** (2003), A spectral characterization of expnential stability for linear time-invariant systems on time scales. *Discrete and Continuous Dynamical Systems* 9(5), 1223 - 1241.
25. **Schmitendorf W.E., Barmish B.R.** (1981), Controlling a constrained linear systems to an affine target, *IEEE Transactions on Automatic Control,* AC-25(3), 761-763.
26. **Sontag E.** (1998), *Mathematical Control Theory*, Springer – Verlag.
27. **Zabczyk J.** (1995), *Mathematical Control Theory*, Birkhäser.

Acknowledgements: The work is supported by the Białystok University of Technology grant S/WM/2/08.

APPENDIX

A1. BASICS ON TIME SCALES CALCULUS

Let us recall that a *time scale* T is an arbitrary nonempty closed subset of the set R of real numbers. The standard cases comprise $T = R, T = Z$ and $T = hZ$ for $h > 0$. We assume that T is a topological space with the topology induced from R. For $t \in T$ we define the *forward jump operator* $\sigma: T \to T$ by $\sigma(t) := \inf\{s \in T: s > t\}$, the *backward jump operator* $\rho: T \to T$ by $\rho(t) := \sup\{s \in T: s < t\}$, the *graininess function* $\mu: T \to [0, \infty)$ by $\mu(t) := \sigma(t) - t$. Using these operators we can classify the points of the time scale as follows:

- If $\sigma(t) > t$, then t is called *right-scattered* and if $\rho(t) < t$, then t is called *left-scattered*;
- if $t < supT$ and $\sigma(t) = t$, then t is called *right-dense* and if $t > infT$ and $\rho(t) = t$, then t is *left-dense*.

Function $f: T \to R$ is called *rd-continuous* provided it is continuous at right-dense points in T and its left-sided limits exist (finite) at left-dense points in T. Function f is called *regulated* provided its right-sided limits exist (finite) at all right-dense points of T and its left-sided limits exist (finite) at all left-dense points in T. Function f is *piecewise rd-continuous*, if it is regulated and if it is rd-continuous at all, except possibly at finitely many, right-dense points $t \in T$.

Let $T^\kappa := T - \big((supT), supT\big]$ if $supT < \infty$ and $T^\kappa := \infty$ if $supT = \infty$.

Definition A1. Let $f: T \to R$ and $t \in T^\kappa$. The *delta derivative* of f at t, denoted by $f^\Delta(t)$, is the real number (provided it exists) with the property that given any $\varepsilon > 0$, there is a neighborhood V of t such that:

$$|\big[f\big(\sigma(t)\big) - f(s)\big] - f^\Delta(t)(\,\sigma(t) - s)| \le\ \varepsilon|\sigma(t) - s|$$

for all $s \in V$.

We say that f is delta differentiable on T^κ provided $f^\Delta(t)$ exists for all $t \in T^\kappa$. In general, the function σ may not be delta differentiable. Delta derivatives of higher order are defined in the standard way: $f^{[k]}(t) = f^\Delta(f^{\Delta^{k-1}}(t))$ for $k \ge 1$.

Remark A2. [Bohner and Peterson (2001)] If $T = R$, then $f: R \to R$ is delta differentiable at $t \in R$ if and only if f is differentiable in the classical sense at t. If $T = Z$, then $f: Z \to R$ is always delta differentiable at every $t \in Z$ with $f^\Delta(t) = f(t + 1) - f(t)$.

Let $f: T \to R$ be a bounded function on $[a, b]_T$ and let P be a partition of $[a, b]_T$ such that $a = t_0 < t_1 < \cdots < t_n = b$. In each interval $[t_{i-1}, t_i)_T$, $i = 1, \dots, n$, choose an arbitrary ξ_i and form the sum:

$$S = \sum_{i=1}^{n} f(\xi_i)(\,t_i - t_{i-1})$$

We say that f is *Riemann* Δ*-integrable* (or Δ*-integrable*) from a to b if there exists a number I with the following property: for each $\varepsilon > 0$ there exists $\delta > 0$ such that:

$$|S - I| <\ \varepsilon$$

for every S corresponding to any partition P of $[a, b]_T$ and independent of the choice of $\xi_i \in [t_{i-1}, t_i)_T$ ξ_i, $i = 1, \dots, n$. Such a number I is unique, see Bohner and Peterson (2003).

A function $F: T \to R$ is called a Δ-*antiderivative* of $f: T \to R$ provided F is Δ-differentiable on T^{κ} and $F^{\Delta}(t) = f(t)$ for all $t \in T^{\kappa}$. *F* is called *a* Δ-*prederivative* of f provided F is Δ -predifferentiable with region of differentiation D and $F^{\Delta}(t) = f(t)$ for all $t \in D$.

Theorem A3. [Bohner and Peterson (2003)] Let f be a Δ-integrable function on$[a, b]_T$. If f has a Δ-prederivative $F: [a, b]_T \to R$ with region of differentiation D, then:

$$\int_s^t f(t)\Delta t \coloneqq F(t) - F(s)$$

A.2. ELEMENTS OF Δ-MEASURES ON TIME SCALES

The notions of Δ-measurable set and Δ-measurable function are studied in Cabada and Vivero (2006), Deniz (2007). Let us consider a set $\mathrm{F} = \{[a, b)_T: a, b \in T, a \leq b\}$ The interval $[a, a)_T$ is understood as the empty set. Let $m_1: F \to [0, \infty)$ be a set of functions that assigns to each interval $[a, b)_T \in F$ its length: $m_1([a, b)_T) = b - a$. Using the pair (F, m_1) one can generate an outer measure m_1^* on the family of all subsets of T as follows. Let $E \subseteq T$. If there exists at least one finite or countable system of intervals $V_j \in F, j \in N$, such that $E \subset \bigcup_j V_j$, then we put $m_1^*(E) = inf \sum_j m_1(V_j)$, where the infinum is taken over all coverings of E by a finite or countable system of intervals $V_j \subseteq F$. If there is no such covering of E, then we put $m_1^*(E) = \infty$. A subset A of a time scale T is Δ-measurable if $m_1^*(E) = m_1^*(E \cap A) + m_1^*(E \cap (T - A))$ holds true for any $E \subset T$. Defining a family:

$$M(m_1^*) = \{\Lambda \subset T: \Lambda \; is \; \Delta\text{-measurable}\}$$

the *Lebesgue* Δ-*measure*, denoted by μ_{Δ}, is the restriction of $m_1^*$to $M(m_1^*)$. If se*t* E is Lebesgue measurable, then set $E \cap T$ is Δ-measurable, see Deniz (2007).

A function $f: T \to [-\infty, \infty]$ is Δ-measurable if for every real α the set $f^{-1}([-\infty, \alpha)) = \{t \in T: f(t) < \alpha\}$ is Δ-measurable. If f is rd-continuous, then f is Δ-measurable, see Deniz (2007).

Properties of rd-continuous and continuous functions on a time scales implies that if f is a continuous function defined on T, then it is Δ-measurable. Moreover, if an rd-continuous function f is defined on a Δ-measurable set $E \subseteq T$, then f is a Δ -measurable function.

Proposition A4 [Deniz (2007)] Let f be defined on a Δ-measurable subset E of T. Function f is Δ -measurable if the set of all right-dense points of E, where f is discontinuous, is a set of Δ-measure zero.

Proposition A5 [Pawłuszewicz and Torres (2010)] Assume that $f: T \to [-\infty, \infty]$. Then f is Δ-measurable if and only if, given $\varepsilon > 0$, there is a rd-continuous function $\varphi: [a, b]_R \to R$ such that the Δ-Lebesgue measure of the set $\{x: f(x) \neq \varphi(x)\}$ is strictly less than ε.

A NEW STRESS CRITERION FOR HOT-TEARING EVALUATION IN SOLIDIFYING CASTING

Norbert SCZYGIOL*

*Institute of Computer and Information Sciences Czestochowa University of Technology, ul. Dąbrowskiego 73, 42-201 Częstochowa, Poland

sczygiol@icis.pcz.pl

Abstract: This work concerns a new criterion for hot-tearing evaluation in castings. Algorithm describing the conduction of computer simulations of phenomena accompanying the casting formation, which performing is the preparation stage for using of this criterion, is also described. According to the low recurrence of phenomena occurring during solidification (e.g. grained structure parameters, stresses distribution) the casting's hot-tearing inclination can be estimated only in approximated manner. Because of still following at present rapid computer processors development, and techniques of its programming, enables to suppose that in short time the efficiency of computer simulations will arise so much, the problem of hot-tearing evaluation newly became interesting for the team working on computer simulations at the Institute for Computer and Information Sciences at Częstochowa University of Technology.

Key words: Casting, Computer Simulation, Hot Tearing, Solidification Processing

1. INTRODUCTION

The production of casts is a technology that involves many significant factors which have an impact on the end result, that is to say, on the quality of the cast.

In shape casting, an equiaxial structure is formed (Fig. 1). During solidification processes various types of defects may appear in the solid-liquid areas. These areas solidify as the last batches of material. The shrinkage leads to micro-porosity effect. Looking at it the other way, the stress effect reveals the commonly named hot tearing in the casting.

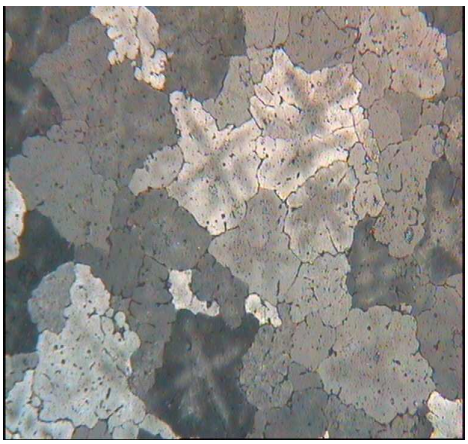


Fig. 1. Equiaxial grains in Al-2% Cu alloy

Hot tearing of solid-liquid areas occurs when the stresses acting on them are able to break the backbone of solid phase, filed with the liquid phase (Sczygiol and Szwarc, 2003b).

Founders and scientists were and still are interested in the problem of hot tearing of castings.

At the beginning, the problem of hot tearing formations was solved by experimental estimation of the hot tearing susceptibility of foundry alloy. Then, the mathematical models describing hot tearing have been developed. The review and analysis of this work can be found in paper Parkitny and Sczygiol (1987). Research on this field focused mostly on the formation of a single crack and were not relevant to industrial practice. The next step in the development of methods for testing susceptibility to hot tearing was the use of advanced numerical methods, through the computer simulation (Rappaz et al., 1999).

There are two groups of works which uses computer simulations. The first group concentrates on the analysis of a single crack development. The second group involves a comprehensive analysis of thermo-mechanical phenomena, accompanying the production process of castings.

On the basis of this analysis, the degree of risk of the appearance of defects in continuity in the entire casting or in its selected parts, is attempted to be drawn (Szwarc, 2003; Sczygiol and Szwarc, 2005). The use of such approach is also possible while performing simulations with use of commercial engineering programs. Usually, such programs do not provide any criteria for hot tearing evaluation in castings. Users of such kind of software have to choose which of the available values characterizing the state of stress and/or deformation should be used to the rupture-susceptibility assessment. It should be mentioned, that such an analysis requires good knowledge of the phenomena in casting formation and skills in simulation of these phenomena. Specialized engineering software, usually based on a finite element method is also required. In turn, simulations performed with use of such software are very time consuming.

This paper concerns research on the analysis of the susceptibility to hot tearing during an equiaxial structure casting. A new stress criterion to evaluate the level of risk of rupture in selected fragments of the casting is proposed. According to the algorithm described in this paper, the assessment of castings hot tearing susceptibility with the use of this criterion is possible only after conducting a series of simulation calculations. The information about the degree of rupture risk in selected areas of the casting is obtained as a result of such evaluation. Studying the suscepti-

bility to hot tearing by using the method proposed here is time-consuming. However, continuous development of computational processors, such as GPUs, as well as effective methods of programming such processors, let us believe that accelerations reached nowadays (Michalski and Sczygiol, 2010; Michalski, 2011) give us ability to use proposed solutions in foundry practice.

2. THE CRITERION FOR EVALUATION OF SUSCEPTIBILITY TO HOT TEARING

Metal alloys often solidify by increasing equiaxial dendrites. It can be assumed that initially each dendrite grows individually. As the dendrites are in contact with each other they form the backbone of the solid phase. Dendrite arms are intertwined with the arms of their neighbors. From this moment, in the solidifying solid-liquid area appears tension. It is carried by each entangled dendrite arm. Dendrites are separated by layers largely filled with the liquid phase.

The two-phase region described in this way consists of growing equiaxial grains and layers of the liquid phase which separate these grains. Such area in the numerical modeling is represented by hexagonal solid phase grain and the surrounding layer of the liquid phase. These hexagonal grains can be also divided into smaller hexagons. The solid phase is presented using regular hexagons and the liquid phase using flattened. The corresponding parts of these two types of hexagons are placed on the border area, see Fig. 1. The size of both areas (solid and liquid phases) is characterized by participation of the solid phase, calculated at the stage of solidification simulation.

This way of modeling used for area of solidifying cast enables operating at the micro level of analysis separately for growing grains and for narrowing layers of the liquid phase. Finite elements used in calculations and in the macroscopic stress analysis are almost always much bigger than solidifying metal grains. In the macroscopic analysis, two-phase area is treated as isotropic. The grain nature of the casting construction is ignored. Nevertheless, because the solidification simulation is conducted on the basis of the coupled model, i.e. macro-microscopic, so after the simulation of solidification the accumulation of grains in two-phase areas can be easily reconstructed. In combination with the analysis of stress at a microscopic level, it enables to analyze the phenomena leading to hot tearing.

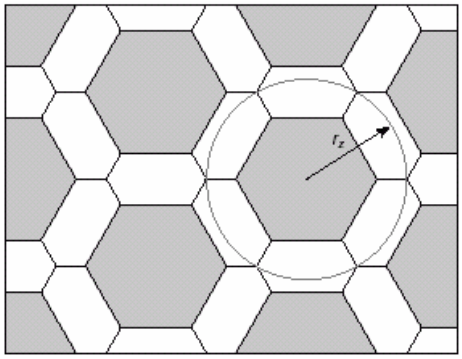


Fig. 2. Model of two-phase area for the alloy solidifying in the form of equiaxial grains

Different temperature gradients and the resistance posed by the wall of the mold to the shrinking casting are the most significant causes of stress in the cast. Conditions of heat evacuation from the casting to the mold and to the environment determine the speed of the alloy solidification, i.e. the equiaxial grains growth speed, but also the speed of the stress generated in the casting.

A new stress criterion is proposed to assess the hot tearing susceptibility of solidifying cast. The criterion takes into account the stress-speed ratio of effective stress in the layers separating the congealed particles to the speed of effective strain in these grains. The proposed criterion is expressed by the local coefficient of susceptibility to hot tearing and it is marked as Θ. This criterion assumed, that stress states are considered in micro scale, but these conditions are obtained under the stress states in a macro scale. During the solidification, the changes in geometry (size) of grains and separating layers is obtained from microscopic analysis conducted on the basis of macroscopic modeling. This is possible because in the macroscopic modeling the growth of equiaxial grains is represented by the connection of diffusion phenomena (micro scale) with thermal phenomena (macro scale).

The calculation of the local coefficient of susceptibility to hot tearing proceeds in the following time steps, beginning with the participation of the solid phase, in which the backbone of solid phase is formed, until complete solidification.

Effective strain rate can be written as:

$$\dot{\bar{\sigma}} = \frac{|\Delta\bar{\sigma}|}{\Delta t} \tag{1}$$

where: $\Delta\bar{\sigma}$ is the effective stress increment in the time step Δt Δt .

Nevertheless, conducted research show, that much better results are obtained if the relative effective stress is introduced to the criterion. Consequently, the criterion can be described as follows:

$$\Theta = \frac{|\Delta\bar{\sigma}_l|}{|\Delta\bar{\sigma}_g|}\frac{\bar{\sigma}_g}{\bar{\sigma}_l} \tag{2}$$

where: l is a sub-layer separation, while g denotes the sub-grain-solidified parts. Because the quotient of relative increment of effective stress in the layers and the grains tends to zero with increasing share of the solid phase, in the sake of clarity, the hot tearing criterion can be transformed to the following form:

$$\Theta = -\ln\left(\frac{|\Delta\bar{\sigma}_l|}{|\Delta\bar{\sigma}_g|}\frac{\bar{\sigma}_g}{\bar{\sigma}_l}\right) \tag{3}$$

Application of the criterion (3) requires a computer simulation in the macro scale, and afterwards in the micro scale. At the macro level standard macroscopic finite elements are used. Conversely, at the micro level – microscopic elements are used. These microscopic elements cover the macroscopic element area. The finite element method, for both types of simulation, is formulated in slightly different way. A traditional formulation e.g., based on the method of weighted residuals is used at the macro level. At the micro level was used a hybrid formulation (Ghosh and Moorthy, 1995; Parkitny et al., 2001).

The equation (3) describes the local susceptibility to hot tearing of a small macroscopic area, corresponding to one finite macroscopic element, subdivided into two areas, i.e. grains and layers separating them. Stress values and their increments used in equation (3) are determined for the subdivisions of layers and grains, receiving two tensors which describe the resultant state

of stress in all the grains and the resultant state of stress in all the layers of separation, which belong to the analyzed area. Tensors are obtained as a result of the so-called homogenization, based on the integration of the stress function in the above-mentioned subdivision, and then dividing the resulting value by the area of integrated subdivision Θ.

High susceptibility to hot tearing is indicated by large values of factor Θ. Nevertheless, the criterion Θ does not indicate a specific limit value, above which the casting will crack. This follows from the fact that, that the value of Θ increases with increasing equiaxial grain, as a result of stress growing with an increasing solidification area. Θ factor values are used to indicate the areas of analyzed casting, where most likely appears a damage, i.e. the rupture. Θ criterion can also be used to determine the conditions most conducive to the production of a given type of cast.

2.1. Description of algorithm used for preparatory calculations

Computation of factor Θ is possible after a complex computer simulations. Data provided by these simulations are used to estimate the susceptibility to hot tearing of the cast. A number of preparatory tasks must be performed at this stage. The steps leading to determining the casting susceptibility to hot tearing cover the following:

1. Simulation of solidification. For succeeding time steps the temperature field, the distribution of the solid phase participation and the mean radii of equiaxial grains, are determined.
2. Calculating distributions of stress in consecutive time steps.
3. Identification (selection) of subdivided areas for the hot tearing analysis.
4. Division of the macroscopic finite elements into the microscopic-hexagon-hybrid-finite elements in order to obtain the solid-liquid areas. These solid-liquid areas are part of subdivisions which are to be analyzed if terms of susceptibility to hot tearing
5. Calculation of the stress in all solid-liquid areas corresponding to the macroscopic finite elements,
6. Calculation of the value of coefficient Θ for each macroscopic finite element in the selected areas
7. Preparation of the scale of susceptibility to hot-tearing based on the simulations and calculations carried out for all the analyzed variants of the task.
8. Execution of a local distribution coefficient diagrams for susceptibility to hot tearing for different variants of the task.
9. Drawing conclusions.

Since the first two steps are standard and well described in literature (Desbiolles et al., 1987; Sczygiol, 2000), only the remaining steps will be shortly described below.

2.2. Identification of subdivisions which are to be analyzed

It makes no sense to carry out the analysis of the susceptibility to hot tearing for all the macroscopic finite elements of the casting, because experience (the practice) shows that, the cracks appear only in selected, fairly easy to identify parts of the cast. The identification of such fragments requires selecting a group of finite elements and the area around them. This selection should be done on the basis of the probability of the hot tearing localization. The group of selected macroscopic elements should be slightly greater than the area of the analysis.

The second group of macroscopic elements (of a similar number of the elements) selected for analysis should be chosen in the area least subjected to hot tearing. If there is a suspicion about the possibility of appearance of hot tearing in other parts of the casting, then another group of elements should be created and analyzed.

Fig. 3 shows the cast and three groups of elements selected for analysis. The main group is placed in the central part of the cast and consists of elements collected under infusion and forming a notch around the bottom of the casting. This group is most at high risk of hot tearing. The second group is located on the left, in the casting arm. The risk of occurring hot tearing in this area is very low. Rupture should not occur either, by design, in the third group, comprising the area around a 'notch' connecting the right shoulder with the casting 'head' located at its end.

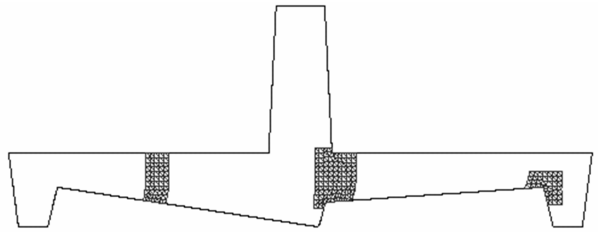


Fig. 3. Location of selected groups of macroscopic elements for the analysis of hot tearing

Comparision of the sizes of areas selected for the analysis with the size of the entire area of the casting (Fig. 3) it can be easily remarked, that the number of elements selected for analysis is relatively small in comparison with the number of finite elements in the whole casting.

2.3. Division of the macroscopic finite elements into the microscopic – hybrid elements

Each of macroscopic finite elements, which belong to the group of the elements analyzed from the point of view of hot tearing susceptibility, is divided into hybrid, microscopic finite elements (Parkitny et al., 2002; Szwarc and Sczygiol, 2002). The mesh of hybrid finite element is generated on the basis of the characteristic dimension of the grains (grain radius). This value is obtained during the simulation of solidification. The mesh created in this way ca be also taken into account in further analysis of two areas of material properties: densely tangled dendrites (solid phase) and the layers separating them in a solid-liquid state (Sczygiol and Szwarc, 2003a).

The surface area of a macroscopic finite element determines the number of microscopic finite elements that belongs to the given macro element. Regardless of the original shape of this element, the hybrid elements mesh is always built on a rectangular plan (similar to a square) with an area equal to or close to the macro element area.

Such an approximation of projecting a macroscopic element to microscopic elements is dictated by the polygonal shape of hybrid elements.

In the areas of separating layers there is not only the liquid

phase, but also the solid phase in the form of dendrite arms. Therefore, the participation of the grains area in the region of the whole solid-liquid area can be written as: Szwarc and Sczygiol, (2004):

$$q = \frac{A_g}{A} = \frac{f_s(1-u)}{1-uf_s} \quad \text{where } u \in \langle 0,1 \rangle \tag{4}$$

where: u is the part of the solid phase share in the separating layers, while f_s is the share of the solid phase. The course of the function q, depending on the participation of the solid phase, including a displacement of half participation of the solid phase to the area of layers ($u = 0.5$) and assuming that all of the solid phase is in the grain area ($u = 0$), is shown in Fig. 4.

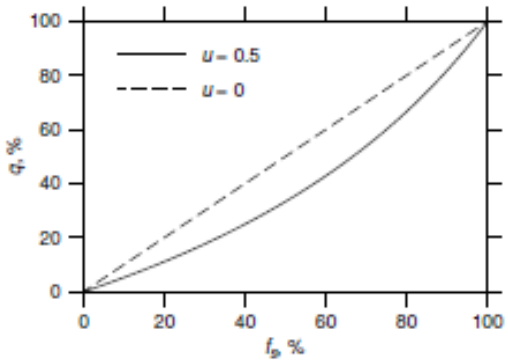


Fig. 4. The course of the function q for different u values

$$x = x_c - \sqrt{\frac{q}{q'}}(x_c - x') \tag{5}$$

where: x is the coordinate of the node, x_c is the coordinate of the so-called measure of the solid phase increase, while the symbol' denotes the current location of the node and the output share of the grain area.

2.4. Stress calculation in microscopic areas

As a result of the macroscopic calculation a number of instantaneous fields is obtained. These are: the temperature profile, liquid phase participation, stress, strain and deformation. The temperature of the end of solidification, the characteristic grain's size and the stress field are relevant for further simulations.

In the selected area the macroscopic finite elements are isolated from the rest of the elements mesh of the casting. The parameters describing the state of the macro elements are used as input for further calculations leading to the determination of the susceptibility to hot tearing. The dimension of the hybrid finite elements is determined on the basis of the equiaxial grain radius assigned to the macro element. The growth of the grains area in successive time steps is controlled with use of the solid phase participation function $f_s(t)$. The temperature profile $T(t)$ is used to control the change in material properties.

The stress tensor $\sigma(t)$ constitute the basis for the formulation of the appropriate boundary conditions, see Fig. 5.

Because of there is no symmetry in loading the system, the stress tensor is converted into an equivalent tensor of main stresses. As the result of this approach it is possible to analyze only a quarter of the system, suitably mounted on symmetry axes and charged by the main stress.

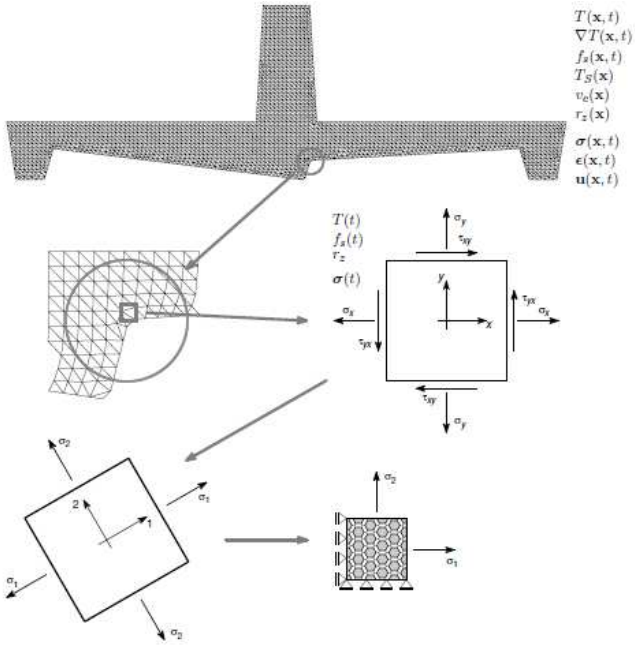


Fig. 5. The division of macroscopic finite element into hybrid microscopic finite elements and the way of support and load of the analyzed area

For the purpose of numerical modeling of the solid-liquid center cracking it is necessary to separate the macroscopic properties as the properties of the two subdivisions. These properties are determined in an experimental way. The participation of the area surface q in the whole solid-liquid area determines the "amount" of the subdivision. Therefore, the value of material property W can be described as:

$$W = qW_g + (1-q)W \tag{6}$$

where: W_g is the value of material properties for the solid phase area (grains), while W_l is the value of material properties for the area of separating layers. Furthermore, it has been assumed that the material properties in the subdivisions are in a relationship expressed as:

$$\frac{W_l}{W_g} = p \quad \text{where} \quad p \in (0,1\rangle \tag{7}$$

Equation (7) describes the distribution of material properties related from temperature ($p = p(T)$). Substituting (7) to (6) we obtain the relationship describing material properties for the solid phase:

$$W_g = \frac{W}{p + (1-p)q} \tag{8}$$

One of the possibilities of determining function p consists in making it relevant of the range of the solidification temperatures:

$$p(T) = \frac{T_L - T}{T_L - T_S} \tag{9}$$

where: T_L and T_S are the liquidus and the solidus temperatures, respectively. Function of participation of solid phase (Sczygiol, 2000) can be also used as the function of distribution of material properties.

The analyzed solid-liquid area is covered by the microscopic finite element mesh. This mesh is charged by the macroscopic state of stress. The boundary conditions are updated on grains arising from the simulation at the macro level.

On the basis of the current temperature values material properties of sub-grains and separating layers are determined. The calculations are carried out from the 'appearance' of stress, i.e., when the share of the solid phase exceeds a critical value (e.g. 25%) until complete solidification.

2.5. Calculation of the value of Θ for the macroscopic finite element

The equation (3) allows for calculation of the values of the local coefficient of susceptibility to hot-tearing Θ. Since different areas of the casting solidify at different time intervals it is convenient, due to further analysis, to present the course of Θ in the function of the solid phase share. Fig. 6 presents sample graph representing such a course.

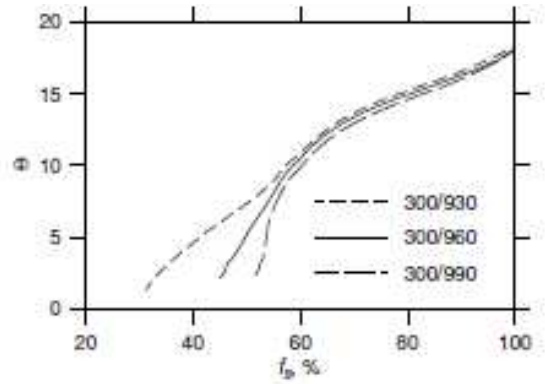


Fig. 6. Sample courses of Θ for given casting conditions (temperature of the mold/casting temperature)

Presentation of results in the function of the solid phase share enables a direct comparison of the coefficient value of all the solved task variants.

2.6. Drawing up the scale of Θ

In order to compare values Θ for different tasks, different conditions for pouring and solidification have been drawn up the scale of susceptibility to hot tearing, based on the critical value Θ_{cr}. It was assumed that the scale is dependent on the participation of the solid phase. The critical value Θ_{cr} is determined from the maximum values Θ for all the variants of the simulation for the solid phase participation, ranging from 50 to 95%, in steps of 5%. On the basis of the received values the function determining critical values of Θ in the function of the solid phase may be constructed. This function is the basis for determining the degrees of the susceptibility to hot tearing.

Thus one should decide whether further analysis of susceptibility to hot tearing will run for four degrees. As high (the highest) degree adopted values Θ larger and equal to Θ_{cr}, as the average – values from $0.9\Theta_{cr}$ to Θ_{cr}, as low degree – values from $0.8\Theta_{cr}$ to $0.9\Theta_{cr}$. For values Θ below $0.8\Theta_{cr}$ the lack of susceptibility to hot tearing is accepted.

2.7. Execution of diagram Θ distribution

Proposed in the previous section, the scale is the basis for drawing up diagrams (maps) of the coefficient Θ distribution for the main group of elements and for the control groups (Fig. 3). The maps are drawn up for certain selected values of the solid phase participation (Fig. 7).

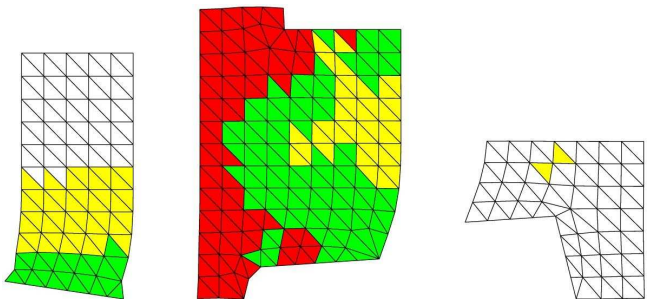


Fig. 7. Sample map of the coefficient Θ distribution of susceptibility to hot-tearing (a darker color means a greater susceptibility to hot tearing)

Since the coefficient Θ distribution maps are only comparative, there are compared elements with the same share of the solid phase in a single casting. So they do not represent any real situation, i.e. those which may occur in the solidifying casting. Such maps are made to indicate that while the main group values Θ indicate the possibility of hot tearing, in the control groups the coefficient values Θ are so small, that they are not at risk from cracking.

2.8. Conclusions from the simulations

After preparing maps of the coefficient Θ distribution some relevant, for the casting practice, conclusions arise. These conclusions may involve the casting hot tearing at different stages of solidification. What is also important, is the evaluation of infusion conditions, which determine the temperature of the mold or flooding temperature, to ensure obtaining sound castings.

3. THE CRITERION FOR EVALUATION OF SUSCEPTIBILITY TO HOT TEARING

Application of the proposed criterion for the hot tearing evaluation has been illustrated by the simulations and analysis of the casting made of Al-2% Cu alloy, solidifying in a metal form. A shape of the metal form and cast are presented in Fig. 8.

The material properties are taken from the work Sczygiol, 2000. Three groups of the macroscopic finite elements (indicated in Fig. 3) were selected for analysis of hot tearing susceptibility. Three series of computer simulations of solidifications were per-

formed. For all simulations, the initial mold temperature was set to 300 K.

The variable parameter was the pouring temperature, that was equal to: 930 K, 960 K and 990 K, respectively.

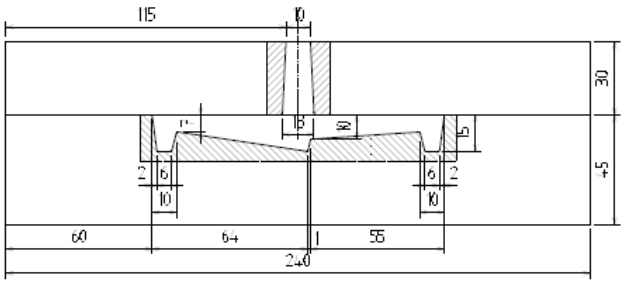


Fig. 8. The analyzed cast in the casting mold

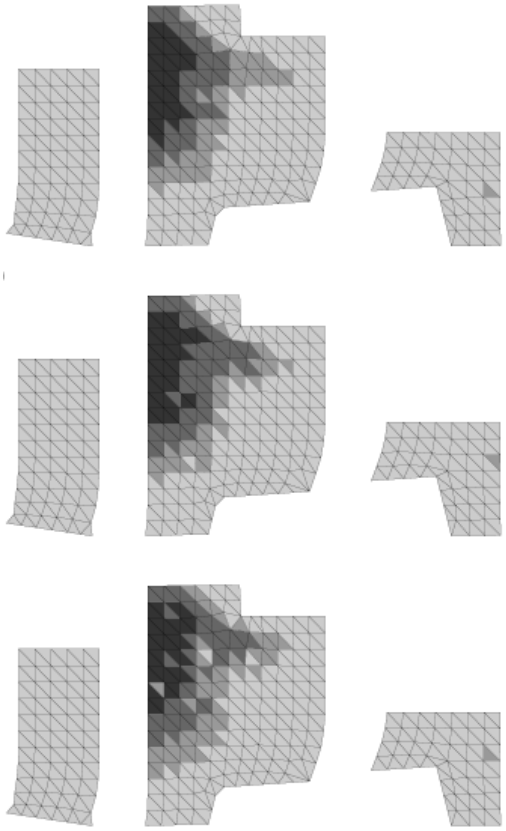


Fig. 9. The distribution of Θ for the solid phase share of 60% in each macroscopic finite element

Distributions of the local coefficient of susceptibility to hot tearing for the major group and control groups are presented in Fig. 9, 10, 11. The upper distributions were made for the pouring temperature 930 K, the middle – for 960 K and the lower distributions for 990 K.

The analysis shows that in all the cases, there is a high risk of the rupture of hot casting. It is therefore concluded that the initial mold temperature is too low. The obtained results were confirmed by experimental research. The hot tearing occurred for an initial mold temperature of 300 K (Fig. 12), while raising the temperature to 600 K guaranteed to receive a sound casting.

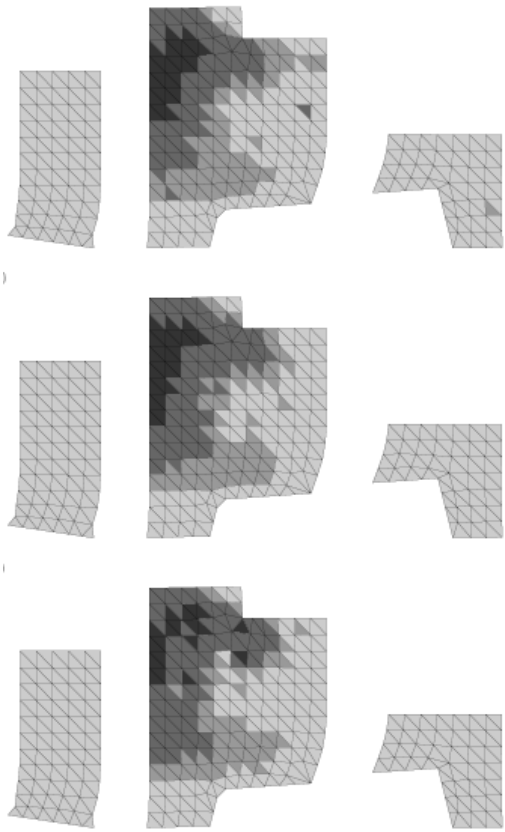


Fig. 10. The distribution of Θ for the solid phase share of 75% in each macroscopic finite element

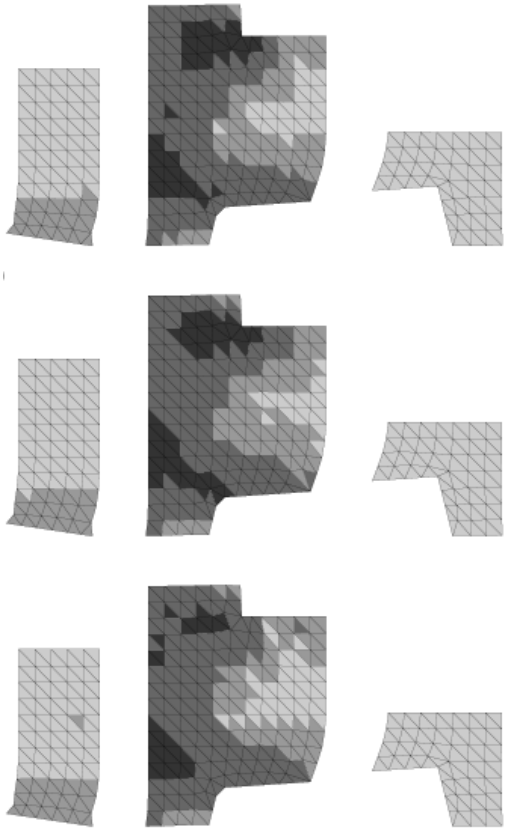


Fig. 11. The distribution of Θ for the solid phase share of 95% in each macroscopic finite element

Fig. 12. Hot tearing in Al-2% Cu alloy

4. SUMMARY

The new stress criterion, proposed in this paper, is a local criterion used for the evaluation of hot tearing susceptibility in casts. It covers the area of a single macroscopic finite element.

A global evaluation of the casting susceptibility to hot tearing is possible with application of this criterion for compact groups of finite elements, in selected casting areas. The analysis can be carried out jointly for several ranges of the initial and the boundary conditions. As a result of computer simulations and the analysis of the susceptibility to hot tearing, the most advantageous (from the point of view of the rupture risk) variant of the casting can be chosen.

Application of proposed criterion requires a lot of preparatory work and computer simulations. However, due to the increasing computing performance of new processors, the use proposed solutions in foundry practice becomes more real.

REFERENCES

1. **Desbiolles J.-L., Droux J.-J., Rapapaz J., Rappaz M.** (1987), Simulation of solidification of alloys by the finite element method, *Computer Physics Reparts* , Vol. 6, 371-383.
2. **Ghosh S., Moorthy S.** (1995), Elastic-plastic analysis of arbitrary heterogeneous materials with the Voronoi Cell finite element method, *Comput. Methods Appl. Mech. Engrg*, Vol. 121, 372-409.
3. **Michalski G.** (2011), *The analysis of multi-/manycore architectures property in selected engineering simulations*, Doctor thesis (adviser N. Sczygiol, Czestochowa University of Technology (in Polish).
4. **Michalski G., Sczygiol N.** (2010), Assembling of the global stiffness matrix in the finite element method for the multi-core processors, Metody Informatyki Stosowanej, *Metody Informatyki Stosowanej*, Vol. 23, No. 2, 97-104 (in Polish).
5. **Parkitny R., Sczygiol N.** (1987), Evaluation of the hot tearing susceptibility of castings, *Krzepnięcie metali i stopów*, Vol. 12, 5-28 (in Polish).
6. **Parkitny R., Sczygiol N., Szwarc G.** (2001), Cracking modelling of brittle materials with grained microstructure by the use of hybrid finite elements, *Zeszyty Nauk. Polit. Białostockiej*, Vol. 138, No 24, 329-336 (in Polish).
7. **Parkitny R., Sczygiol N., Szwarc G.** (2002), Application of the hybrid finite element formulation to numerical modelling of hot tearing of castings, *International Symposium ABDM*, Kraków-Przegorzały.
8. **Rappaz M., Drezet J.-M., Gremaud M.** (1999), A new hot-tearing criterion, *Metallurgical and materials Transactions A*, Vol. 30A, 449-455.
9. **Sczygiol N.** (2000), *Numerical modelling of thermo-mechanical phenomena in a solidifying casting and mould*, Wydawnictwo Politechniki Częstochowskiej, seria Monografie nr 71 (in Polish).
10. **Sczygiol N., Szwarc G.** (2003a), Modelling of elastoplastic thermal stresses in castings in semi-solid state with microstructure taken into consideration, *International Conference CMM-2003*, Gliwice.
11. **Sczygiol N., Szwarc G.** (2003b), Numerical analysis of hot tearing susceptibility in castings with equiaxed inner structure, *IV Krajowa Konferencja MSK'03*, Materiały Konferencyjne, 787-792 (in Polish).
12. **Sczygiol N., Szwarc G.** (2005), Fracture modelling of grained medium in semi-solid domains, *Informatyka w Technologii Materiałów*, Vol. 5, No 3, 103-118 (in Polish).
13. **Szwarc G.** (2003), *Numerical fracturing modelling of alloys with equiaxed microstructure in semi-solid state*, Doctor thesis (adviser N. Sczygiol, Czestochowa University of Technology (in Polish).
14. **Szwarc G., Sczygiol N.** (2002), Numerical analysis of the stress state in semi-solid castings region, *Archives of Foundary*, Vol. 2, No 4, 280-287 (in Polish).
15. **Szwarc G., Sczygiol N.** (2004),), A new criterion for castings hot tearing susceptibility estimation, *Materiały 11. Konferencji „Informatyka w Technologii Metali"*, 331-338 (in Polish).

NUMERICAL ANALYSIS OF RESIDUAL STRESS IN A GRADIENT SURFACE COATING

Wiesław SZYMCZYK*

*Faculty of Mechanical Engineering, Department of Mechanics and Applied Informatics, Military University of Technology, ul. Kaliskiego 2, 00-908 Warsaw, Poland

wszymczyk@wat.edu.pl

Abstract: There were conducted numerical analyses of thermal residual stress in a double gradient surface coating consisting of porous ZrO_2 of 500 μm thickness, placed on Inconel 718 substrate, with an interlayer of NiCoCrAlY of 200 μm thickness. It was assumed that the coating was deposited by plasma spraying. It showed out roughness of the free surface and borders between material phases. Numerical model took into consideration the real geometry of material structure. The results were analysed after transforming them into discrete distributions of particular components of stress. Distributions were presented for particular zones of the surface coating. This allowed to obtain signals originated by distinct features of the material structure such as particular material phases, their contact borders, pores, roughness of the external surface, roughness of internal borders between layers of the gradient system and others.

Key words: Graded Surface Coating, Numerical Simulations, Residual Stress, Results as Discrete Distributions

1. INTRODUCTION

During cooling down from manufacturing to the room temperature an interaction between materials which have different thermo mechanical properties takes place which leads to origination of spatially complex state of residual stresses. They may cause coating durability lowering or even spallation. If the thermal expansion coefficient of the substrate is greater than for the surface coating than, after the cooling process, the substrate is under tension and the coating is compressed (Khor and Gu, 2000).

Stress distribution depends on microstructure features like material phases composition, inclusions, voids and pores, substrate surface geometry. Residual stress may induce phase change in ZrO2 phase which has advantageous influence on fracture toughness of the coating (Szymczyk, 2002, 2005).

It is possible to achieve internal energy redistribution in substrate/coating system which improves durability by establishing it as gradient material. Gradient coating is a special composite in which material properties change along the specific directions. If the change is discrete the gradient system consists of several layers of materials of different properties and the borders between them are distinguishable. If the change is continuous the term functionally graded material (FGM) may be used.

Gradient materials allows for spatial optimisation of material properties in chosen directions in surface coatings. Their definition differs from classic microstructural composites, where it is expect that properties and distributions of particular phases are homogeneous. Gradation of properties of the surface coating in the depth direction is achieved by changing of material phases volume fractions during the deposition process (Grujicic and Zhao, 1998).

2. DISCRETE MODEL OF SURFACE COATING

Numerical FEM modelling and simulations allow for determining of constructional and technological factors influence on residual stress distribution in surface coatings. Numerical methods are effective tool for extending knowledge concerning mechanisms of stress redistribution, especially in cases when it is impossible to conduct experimental research.

Different types of discrete models are used for investigations of surface coatings systems. In the paper are presented methods of microstructure modelling and results analysis which allowed for assuming residual stress state in a surface coating based on ZrO_2.

Numerical model took into consideration the real geometry of material structure. As an example pattern of the real geometry of material microstructure was taken into consideration the structure presented in (Widjaja, 2003)[6]. There was elaborated a discrete raster 2D model (Fig. 1), consisted of uniform square elements.

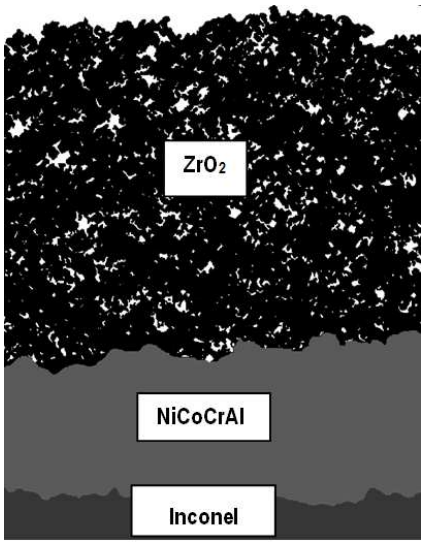


Fig. 1. Model structure of the porous ZrO2 surface coating corresponding with an example presented in Widjaja S. et al., (2003)

There were conducted numerical analyses of thermal residual stress in a double gradient surface coating consisting of porous ZrO_2 of 500 μm thickness, placed on Inconel 718 substrate, with an interlayer of NiCoCrAlY of 200 μm thickness.

It was assumed that the coating was deposited by plasma spraying. The substrate was heated during deposition process and kept in stabile temperature of 700 K. Surface coating showed out roughness of the free surface and borders between material phases.

3. MATERIAL PROPERTIES

Material properties (Tab. 1) for particular phases of modelled graded coating system were found in Widjaja S. et al., (2003). Authors declared that the surface system was in temperature of 700 K. For the needs of numerical simulations the temperature decrease ΔT = -400 K from the deposition stage to the room temperature was assumed. Characteristics presented in (Tab. 1) are the averaged values for the assumed temperature decrease.

Tab. 1. Material properties averaged for temperature decrease ΔT = -400K (Widjaja S. et al., 2003) (E – Young modulus, α – thermal expansion coefficient, ν – Poisson ratio)

Material	E [GPa]	α [10^{-6}/K]	ν
ZrO_2	48	7.72	0.25
Inconel 718	189.5	14.4	0.30
NiCoCrAlY	193.5	12.8	0.30

4. RESULTS ANALYSIS METHOD

Contour plots are not sufficient for residual stress state analysis. They provide qualitative evaluations, allow for determining stress concentrations areas. They show out that features of microstructure like roughness and porosity are effective residual stress concentrators. These stress concentrations are strong enough to initialize microcracking. More detailed quantitative comparison research with the use of contour plots is impossible.

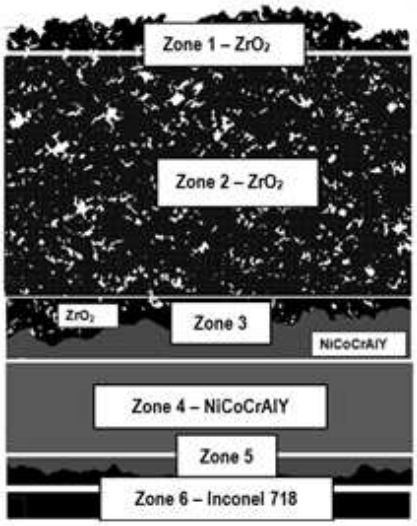


Fig. 2. Zones in the model for stress components discrete distributions establishing

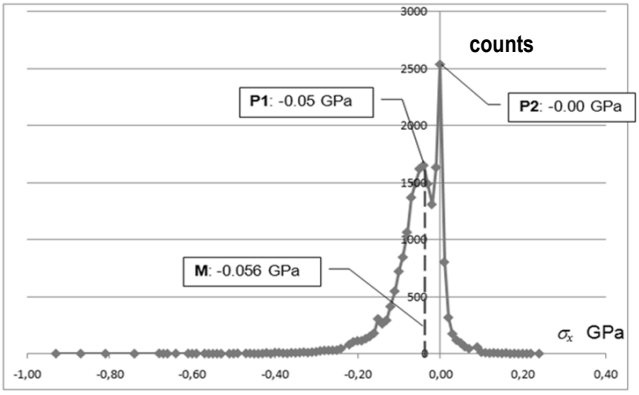


Fig. 3. Distribution of the stress component σ_x in Zone 1

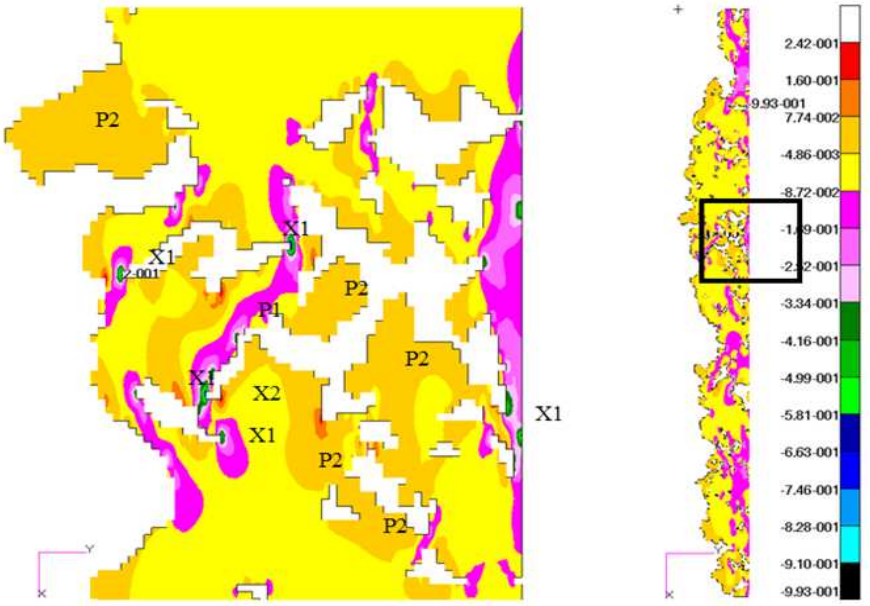


Fig. 4. Zone 1 – areas responsible for forming of the P1 and P2 peaks in the discrete distribution (compare Fig. 3)

For the needs of quantitative research the numerical calculations results were transformed to discrete distributions of particular stress components values. The distributions were obtain for particular zones of the surface coating (Fig. 2). In Fig. 3÷10 there are presented discrete distributions for distinguished zones. Positions of distribution peaks may be treated as average values of residual stress component for each zone.

In Fig. 3 the distribution of the stress component σ_x in Zone 1 is presented (roughed external surface of ZrO_2 layer), weighted average value of the P1 and P2 peaks: M = -0.056 GPa. Peaks P1 and P2 are from the areas that are marked in Fig. 4.

In Fig. 5 the distribution of the stress component σ_x in Zone 2 is presented (pure ZrO_2). Weighted average value of the P1 and P2 peaks: M = -0.110 GPa.

In Fig. 6 the distribution of the stress component σ_x in Zone 3 is presented (transition area between ZrO_2 and NiCoCrAlY phases). Weighted average value of the P1, P2 and P3 peaks: M = -0.183 GPa.

In Fig. 7 the distribution of the stress component σ_x in Zone 4 is presented (pure NiCoCrAlY). Weighted average value of the P1, P2 and P3 peaks: M = -0.174 GPa.

In Fig. 8 the distribution of the stress component σ_x in Zone 5 is presented (transition area between NiCoCrAlY and Inconel 718). Weighted average value of the P1, P2 and P3 peaks: M = -0.167 GPa.

In Fig. 9 the distribution of the stress component σ_x in Zone 6 is presented (substrate Inconel 718). Weighted average value of the distribution: M = 0.0 GPa.

In Fig. 10 distributions are presented of the stress component σ_x in all zones are presented together. All the peaks are in the range from -0.450 to 0.150 GPa.

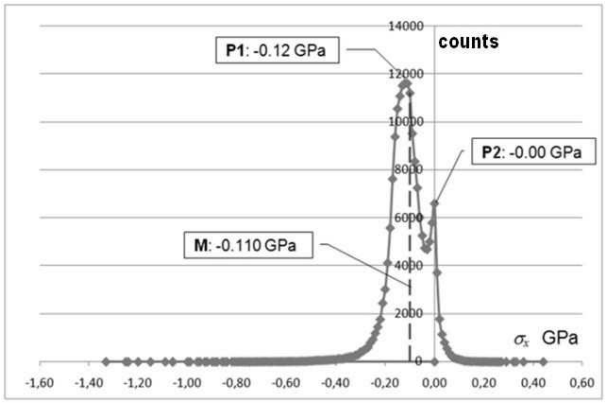


Fig. 5. Distribution of the stress component σ_x in Zone 2

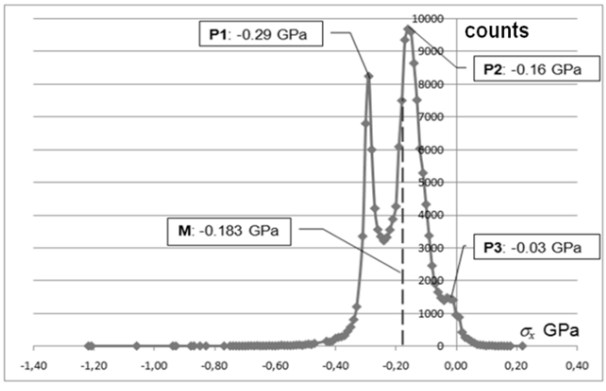


Fig. 6. Distribution of the stress component σ_x in Zone 3

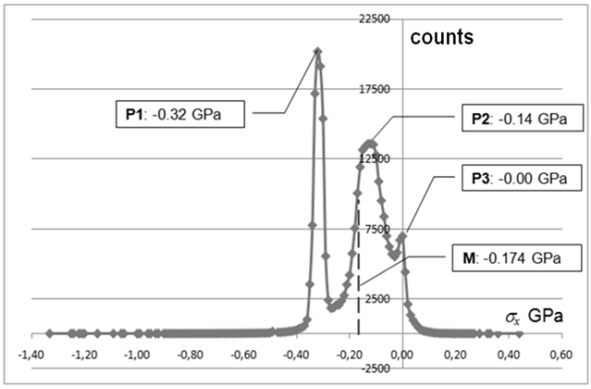


Fig. 7. Distribution of the stress component σ_x in Zone 4

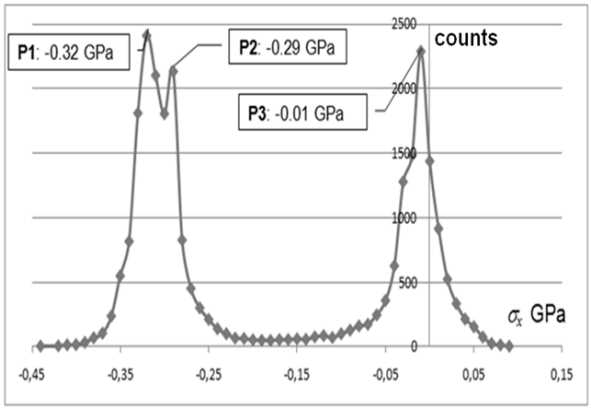


Fig. 8. Distribution of the stress component σ_x in Zone 5

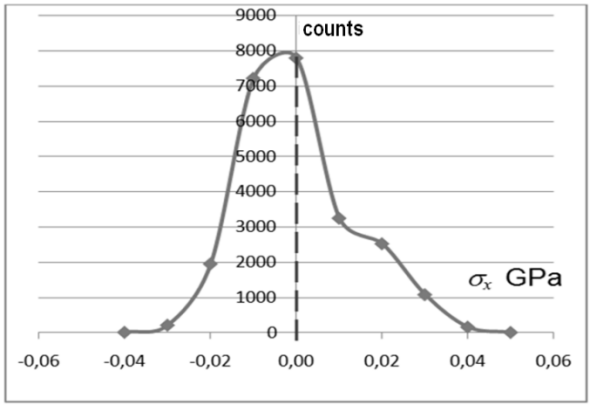


Fig. 9. Distribution of the stress component σ_x in Zone 6

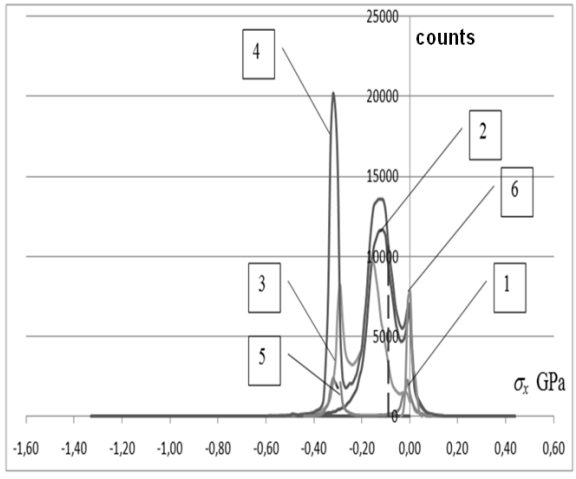


Fig. 10. Distributions of the stress component σ_x in all zones together

5. CONCLUSIONS

Numerical calculations results transformed into the shape of discrete distributions allow to analyse the influence of particular features of material microstructure on residual stress field forming. Each feature like pores, phases borders, roughness of external surface of the coating, roughness of layer borders affects the shape of distribution. Each material phase layer of surface coating produce characteristic peak (Fig. 11, 12).

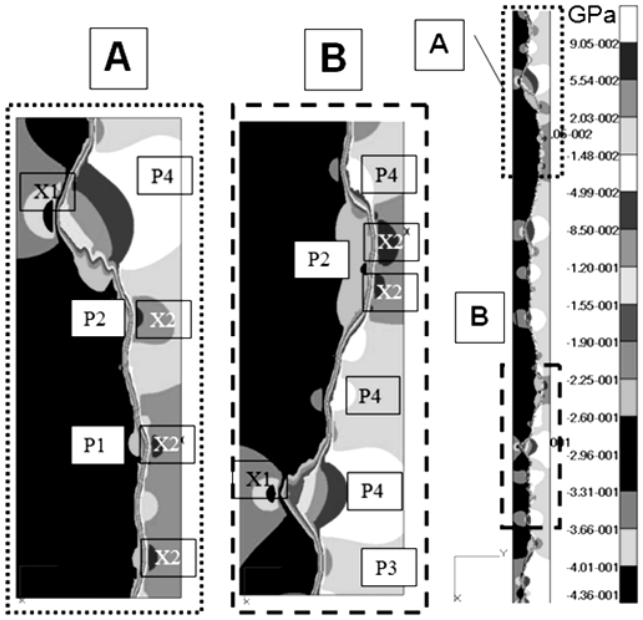


Fig. 11. An example of contour plots of residual stress component σ_x in Zone 5

In Fig. 11 an example of contour plots of residual stress component σ_x in Zone 5 is presented. Zone 5 contains the border between NiCoCrAlY interlayer and substrate. There are marked areas which are responsible for particular parts of discrete distribution forming (compare Fig. 12).

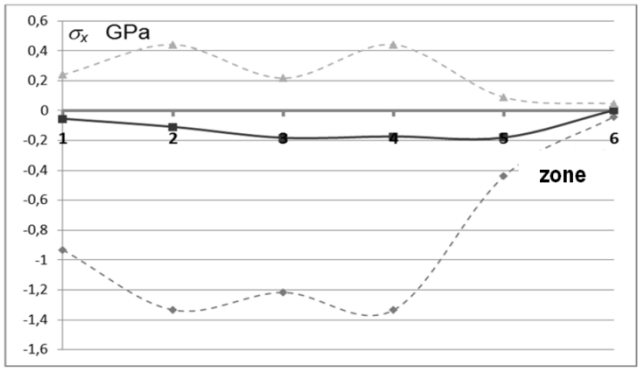


Fig. 12. Average and stress concentration values in subsequent zones of investigated surface coating

In Fig. 12 average and stress concentration values in subsequent zones of investigated surface coating are presented. Average values are negative and reach -0.2 GPa. In each zone extreme values of stress concentrations are negative or positive reaching -1.36 or 0.43 GPa. Stress concentrations of positive values may cause microcracking initiation.

If we assume that materials in all the layers are homogeneous and borders between them are smooth the planar state of stress is the consequence and the formula (1) may be used for residual stress value approximation (Szymczyk, 2005):

$$\sigma_{tc} = E_c \Delta\alpha \Delta T (1+\nu)/(1-\nu^2) \quad (1)$$

where: E_c – Young modulus of the layer, $\Delta\alpha$ – difference between thermal expansion coefficients of the layer and substrate, ΔT – difference between temperatures of deposition stage and room conditions, ν – Poisson ratio.

Accordingly to the formula (1), average thermal residual stress value approximation should reach the value of -0.171 GPa for ΔT= -400 K. Real materials are not homogeneous, though spatial state of stress exists in them. We can expect lower values of σ_x stress component. From numerical calculations it was obtained -0.110 GPa for the pure porous ZrO_2 layer and weighted average value for all layers in the coating -0.151 GPa (Fig. 10). Numerical result are also affected by raster mesh shape itself.

REFERENCES

1. **Grujicic M., Zhao H**. (1998), Optimization of 316 stainless/alumina functionally graded material for reduction of damage induced by thermal residual stress, *Mat. Sci. Eng*., A252, 117-132.
2. **Khor K. A., Gu Y. W**. (2000), Effects of residual stress on the performance of plasma sprayed functionally graded ZrO_2/NiCoCrAlY coatings, *Mat. Sci. Eng*., A277, 64-76.
3. **Szymczyk W.** (2005), Numerical simulation of composite surface coating as a functionally graded material, *Materials Science and Engineering*, A 412, 61-65.
4. **Szymczyk W. et al. (2002),** A concept of phase change numerical simulation in analysis of stress in the area around crack in a ceramic composite, *Przegląd Mechaniczny*, 7-8, Warsaw, 37-40.
5. **Tanaka K., Tanaka Y., Enomoto K., Poterasu V.F., Sugano Y**. (1993), Design of thermoelastic materials using direct sensitivity and optimization methods. Reduction of thermal stresses in functionally gradient materials, *Comp. Meth. Appl. Mech. Engng*., 106, 271-284.
6. **Widjaja S. et al.** (2003), Modeling of residual stresses in a plasma-sprayed zirconia/alumina functionally graded thermal barrier coating, *Thin Solid Films,* 434, 216-227.

APPLICATION OF BOUNDARY ELEMENT METHOD TO SOLUTION OF TRANSIENT HEAT CONDUCTION

Anna Justyna WERNER-JUSZCZUK*, Sławomir Adam SORKO*

*Faculty of Civil and Environmental Engineering, Department of Heat Engineering, Bialystok University of Technology, ul. Wiejska 45 E, 15-351 Bialystok, Poland

a.juszczuk@pb.edu.pl, s.sorko@pb.edu.pl

Abstract: The object of this paper is the implementation of boundary element method to solving the transient heat transfer problem with nonzero boundary condition and particularly with periodic boundary condition. The new mathematical BEM algorithm for two dimensional transient heat conduction problem with periodic boundary condition is developed and verified. The results of numerical simulation of transient heat conduction in two dimensional flat plate under non zero initial condition are compared with results obtained with analytical method. Then the practical application of developed algorithm is presented, that is the solution of ground temperature distribution problem with oscillating temperature of ambient. All results were obtained with a new authoring computer program for solving transient heat conduction problem, written in Fortran.

Keywords: Boundary Element Method, Transient Heat Conduction Problem, Periodic Boundary Condition

1. INTRODUCTION

The heat transfer in solids, with the changes of temperature in time on physical boundaries of analysed objects, occur in many engineering mechanisms (engines, compressors), heating and cooling systems and hydraulic networks (Zhang et al., 2009; Lu and Viljanen, 2006). The analysis of basic mechanism of heat transfer in solids, that is heat conduction problem, is significant for process of designing and optimization mechanical systems and devices. Accordingly, the heat conduction equations with conditions of variable temperature or heat flux on boundaries become an important instrument for mathematical description of many engineering, geothermal and biological problems. As a result, there is a need to develop effective computational methods and tools for solving transient heat conduction problem (Mansur et al., 2009; Yang and Gao, 2010).

Two groups of method are applied to obtain transient heat conduction problem solution: analytical and numerical. In the literature, many analytical methods have been proposed, inter alia based on orthogonal and quasi-orthogonal expansion technique, Laplace transform method, Green's function approach or finite integral transform technique, but they are feasible only for problems with simple geometries (Singh et al., 2008).

Monte et al. (2012) presented very accurate analytical solutions modeling transient heat conduction processes in 2D Cartesian finite bodies, such as rectangle and two layer objects, for small values of the time. In the paper, the geometry criterion was provided that permit to use 1D semi-infinite solutions for solving 2D finite single- and multi-layer transient heat conduction problems. Yumrutas (Yumrutas et al., 2005) developed new method based on complex finite Fourier transform (CFFT) technique for calculation of heat flux, through multilayer walls and flat roofs, and the temperature on the inner surface. The periodic boundary conditions were assumed, that is hourly changeable values of external air temperature and solar radiation. Lu et al. (Lu et al., 2006; Lu and Viljanen, 2006) adopted the Laplace transform to solve the multidimensional heat conduction in composite circular cylinder and multilayer sphere, with time-dependent temperature changes on boundary, which were approximated as Fourier series. Singh et al. (2008) applied separation of variables method to obtain analytical solution, in the form of transient temperature distribution, to the 2D transient heat conduction problem in polar coordinates with multiple layers in the radial direction. Rantala (2005) proposed a new semi-analytical method for the calculation of temperature distribution along the fill layer underneath a slab-on-ground structure subjected to periodic external and internal temperature.

In spite of development of analytical techniques, this methods still cannot be employed for solving most practical heat transfer problems, such as heat conduction in anisotropic materials, objects of complex geometries or complex boundary conditions (Rantala, 2005; Johansson and Lesnic, 2009). Hence, for last few decades, the numerical methods have been strongly developed, as more universal computational tool.

The most popular are mesh methods, such as the finite element method (FEM) and the finite difference method (FDM). Although this methods are well established and commonly applied to transfer heat analysis, in many problems, mesh generation can be very laborious and constitutes the most expensive and difficult part of numerical simulations. Moreover, in objects of complex geometries, generated meshes can be distorted, what contributes to increase of computational error (Li, 2011).

The drawback of mesh generation is overcome in the mesh free (meshless) methods, that use a set of scattered nodal points in considered object (no connectivity among nodes), instead of meshes (Cheng and Liew, 2012; Ochiai et al., 2006). Some of this methods have been recently applied to transient heat conduction analysis in 2D objects, like meshless element free Galerkin (EFGM) method (Zhang et al., 2009), meshless local Petrov-Galerkin (MLPG) method (Li et al., 2011), method of fundamental solutions MFS (Johansson and Lesnic, 2008, 2009), meshless local radial basis function-based differential quadrature (RBF-DQ) method (Soleimani et al., 2010), and in 3D objects, like meshless reproducing kernel particle (RKPM) method (Cheng and Liew, 2012). The disadvantage of this methods, is that, in some cases, like transient heat conduction, they are more time-consuming than mesh methods, such as FEM, because of the larger dimensions of generated matrices (Zhang et al., 2009).

The alternative for above mentioned mesh and mesh free

methods is boundary element method (BEM). Compared to grid methods (FDM, FEM), the great advantage of BEM is the possibility of determination of the solution (both the function and the derivative of this function) at any point of the domain without necessity of construction of grids in considered 2D or 3D space. The discretization is performed only over the boundary, not over the whole analyzed domain hence the size of system of equations, that need to be solved, is reduced by one. In BEM, the fully populated coefficient matrices are generated, what is the opposite of banded and symmetric matrices in FEM. However, the small dimensions of BEM matrices counterbalance this disadvantage (Katsikadelis, 2002; Majchrzak, 2001; Pozrikidis, 2000). Application of the BEM requires the knowledge of fundamental solution of the governing differential operator, but at the same time, the use of fundamental solution stabilize the numerical commutations (Ochiai et al., 2006).

The BEM is successfully applied to steady and unsteady heat conduction problems. As opposed to steady problem, in mathematical description of transient heat conduction, the domain integrals occur. In order to keep the boundary character of the method, many different techniques have been developed, but the most popular are: method using the Laplace transformation to eliminate the time derivative, the dual reciprocity method, and the convolution scheme (employing time-dependent fundamental solutions).

Erhart (Erhart et al., 2006) implemented the Laplace transformation for solution of transient heat transfer in multi-region objects. As a result the time-independent boundary integral equation was produced, solved further with a steady BEM approach. The last step was numerical inversion of the solution, done with the use of Stehfest method. The derived algorithm was applied to heat conduction in a bar, laminar airfoil with three cooling passages and non-symmetric airfoil. The results were compared with those obtained with finite volume method (FVM).

Sutradhar and Paulino (2004) also used the Laplace transformation, both with Galerkin approximation, for analysis of the non-homogenous transient heat conduction problem in functionally graded materials FGM of variable thermal conductivity and specific heat. The three kinds of material variation, that is quadratic, exponential and trigonometric, were assumed for verifying the accuracy of presented method. The practical example for the functionally graded rotor problem was carried out.

Another approach is Fourier transform, applied by Simoes (Simoes et al., 2012) and Godinho (Godinho et al., 2004), consist in three general steps: converting analyzed domain into frequency domain, solving the heat conduction problem with BEM and obtaining the final solution in time domain with the use of inverse Fourier transform. Simoes tested method in 2D object with unit initial temperatures and with non-constant temperature distribution in domain. Godinho analyzed transient heat conduction around a cylindrical irregular inclusion of infinite length, inserted in a homogeneous elastic medium and subjected to heat point sources placed at some point in the host medium.

Mohammadia (Mohammadia et al., 2010) solved 2D nonlinear transient heat conduction problems with non-uniform and nonlinear heat sources, with the new BEM approach, using time-dependent fundamental solutions. In this method temperature is computed on the boundary and in internal points at every time step, and the results constitute the initial values for the next time step. However, for 3D and large problems, the storage of coefficient matrices for every time step can be problematic (Erhart et al., 2006).

Tanaka et al. (2006) applied dual reciprocity boundary element method (DRBEM) for analysis of 3D transient heat conduction problem in nonlinear temperature-dependent materials. In proposed method, domain integral is transformed into boundary integrals with the use of radial basis functions. To entertain the material nonlinearity, the iterative solution procedure was employed. Białecki et al. (2002) proposed the DRBEM without matrix inversion for linear and non-linear transient heat conduction problem, that reduce the time of computations. The method was applied to solve heat transfer problem in a turbine rotor blade. Ochiai (Ochiai et al., 2006; Ochiai and Kitayama, 2009) developed the triple-reciprocity BEM to solve 2D and 3D transient heat conduction problems. One of the recent methods is radial integration boundary element method RIBEM applied to transient heat conduction problem by Yang and Gao (2010), which can be employed to analysis of functionally graded material problems.

In this paper, BEM is applied to solve the unsteady heat conduction problem in 2D area of arbitrary shape of boundary line in particular case of periodic changes of temperature on boundary line. The new mathematical BEM algorithm for periodic boundary condition was developed, both with a new authoring computer program, written in Fortran, applied to verifying the accuracy of presented algorithm and to solving practical example.

2. TRANSIENT HEAT CONDUCTION

The thermal processes, in which the heat conduction is the main mechanism, are described by Fourier-Kirchhoff equation.

The unsteady heat conduction in homogeneous solid substance with constant material properties without inner heat sources, is described by the heat conduction equation (also named thermal diffusion equation)

$$\left(\nabla^2 - \frac{1}{\alpha}\frac{\partial}{\partial t}\right)T(x,y,t) = 0 \tag{1}$$

In the above equation $\alpha=\lambda/c$ is the thermal diffusivity, in which λ is the thermal conductivity and c is the volumetric specific heat; and $\partial / \partial t$ is the local time derivative.

In order to find the solution of this equation, it is necessary to introduce the boundary conditions (1a) and (1b), and initial condition (1c) that take the following form:

$$T(x,y,t) = T_L(x,y,t) \quad , \qquad (x,y) \in L_q \tag{1a}$$

$$q(x,y,t) = -\lambda\frac{\partial T(x,y,t)}{\partial n} = q_L(x,y,t) \, , \quad (x,y) \in L_T \tag{1b}$$

$$T(x,y,0) = T_0(x,y) \quad , \qquad (x,y) \in \Lambda \tag{1c}$$

The boundary conditions (1a) and (1b) assume respectively the value of temperature at point $\mathbf{p}(x_\mathbf{p},y_\mathbf{p})$ on boundary (L_q) (Dirichlet boundary condition), and the value of heat flux at any point $\mathbf{p}(x_\mathbf{p},y_\mathbf{p})$ on boundary (L_T) (Neumann boundary condition). The initial condition (1c) assumes the value of temperature at point $\mathbf{v}(x_\mathbf{v},y_\mathbf{v})$ inside the domain at initial time t=0.

Particular form of the boundary problem for transient heat equation (1) is the formulation with the condition of periodic changes of the temperature on the boundary, which takes the following form:

$$T(x,y,t) = \tilde{T}\cos(\omega t) \quad , \qquad (x,y) \in (L) \tag{1d}$$

where $\tilde{T}$ is the amplitude of the temperature oscillations.

The sketch for two dimensional boundary problem analysis of Fourier equation (1) is shown in Fig. 1.

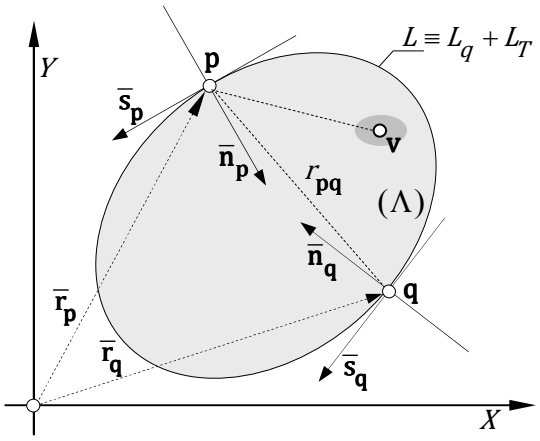


Fig. 1. Sketch for the two dimensional boundary problem analysis of Fourier equation

3. PROBLEM FORMULATION

The fundamental solution of heat conductivity equation (1), also called Green function for heat equation, and its normal derivative for two dimensional problems are given by:

$$T^*(\mathbf{p},\mathbf{q};t,\tau)=\frac{1}{4\pi\alpha(t-\tau)}\exp\left[-\frac{r_{\mathbf{pq}}^2}{4\alpha(t-\tau)}\right] \qquad (2a^1)$$

$$Q^*(\mathbf{p},\mathbf{q};t,\tau)=\frac{\left|r_{\mathbf{pq}}\right|\cos(r_{\mathbf{pq}},\bar{\mathbf{n}}_{\mathbf{q}})}{8\pi\alpha^2(t-\tau)^2}\exp\left[-\frac{r_{\mathbf{pq}}^2}{4\alpha(t-\tau)}\right] \qquad (2a^2)$$

The solution of the Fourier first problem in closed domain (Λ) is described by the sum of double layer heat potential and Poisson-Weierstrass integral:

$$T(\mathbf{p},t)+\alpha\int_{t_0}^{t}\int_{(L_q)} T(\mathbf{q},\tau)Q^*(\mathbf{p},\mathbf{q};t,\tau)dL_{\mathbf{q}}d\tau+$$

$$\iint_{(\Lambda)} T_0(\mathbf{v})T^*(\mathbf{p},\mathbf{v};t,\tau_0)d\Lambda_{\mathbf{v}}=0 \qquad (3)$$

Density $T(\mathbf{q},\tau)$ of double layer potential satisfies integral equation on boundary line (L):

$$-\frac{1}{2}T(\mathbf{p},t)+\alpha\int_{t_0}^{t}\int_{(L_q)} T(\mathbf{q},\tau)Q^*(\mathbf{p},\mathbf{q};t,\tau)dL_{\mathbf{q}}d\tau=g(\mathbf{p},t) \qquad (3a)$$

where:

$$g(\mathbf{p},t)=T_L(\mathbf{p},t)-\iint_{(\Lambda)} T_0(\mathbf{v})T^*(\mathbf{p},\mathbf{v};t,\tau)d\Lambda_{\mathbf{v}}. \qquad (3b)$$

The solution of the Fourier second problem in closed domain: (Λ) is described by the sum of single layer heat potential and Poisson-Weierstrass integral:

$$q(\mathbf{p},t)+\frac{1}{c}\int_{t_0}^{t}\int_{(L_T)} q(\mathbf{q},\tau)T^*(\mathbf{p},\mathbf{q};t,\tau)dL_{\mathbf{q}}d\tau+$$

$$\iint_{(\Lambda)} T_0(\mathbf{v})T^*(\mathbf{p},\mathbf{v},t;\tau_0)d\Lambda_{\mathbf{v}}=0 \qquad (4)$$

Density $q(\mathbf{q},\tau)$ of single layer potential satisfies integral equation on boundary line (L):

$$\frac{1}{2}q(\mathbf{p},t)-\frac{1}{c}\int_{t_0}^{t}\int_{(L_T)} q(\mathbf{q},\tau)T^*(\mathbf{p},\mathbf{q},t,\tau)dL_{\mathbf{q}}\,d\tau=h(\mathbf{p},t) \qquad (4a)$$

where:

$$h(\mathbf{p},t)=q_L(\mathbf{p},t)-\iint_{(\Lambda)} T_0(\mathbf{p})T^*(\mathbf{p},\mathbf{v},t,\tau)d\Lambda_{\mathbf{v}}\,. \qquad (4b)$$

3.1. Boundary integral equation for heat conduction equation

The mixed internal Fourier problem for differential equation (1) with conditions (1a,1b) and (1c) in two dimensional area (Λ) has the general solution of the integral form (Brebbia et al, 1984)

$$\chi(\mathbf{p})T(\mathbf{p},t_k)+\alpha\int_{t_0}^{t_k}\int_{(L_q)} T(\mathbf{q},\tau)Q^*(\mathbf{p},\mathbf{q};t_k,\tau)dL_{\mathbf{q}}d\tau+$$

$$\frac{1}{c}\int_{t_0}^{t_k}\int_{(L_T)} q(\mathbf{q},\tau)T^*(\mathbf{p},\mathbf{q};t_k,\tau)dL_{\mathbf{q}}d\tau+$$

$$\iint_{(\Lambda)} T_0(\mathbf{v})T^*(\mathbf{p},\mathbf{v};t_k,t_0)d\Lambda_{\mathbf{v}}=0\,, \qquad (5)$$

where $\mathbf{p}$ and $\mathbf{q}$, $\mathbf{v}$ are respectively source and field points within the domain (Λ) or on the boundary (see Fig.1) and $[t_0,\ t_K]$ is the analyzed time interval. Coefficient $\chi(\mathbf{p})$ is related to the local geometry of the boundary at point $(\mathbf{p})$. For smooth boundary point $\chi(\mathbf{p})=1/2$ and for an internal point $\chi(\mathbf{v})=1$.

Unknown functions in integral equation (5) are: temperature $T(\mathbf{q},\tau)$ on the part (L_q) of boundary line and heat flux on the part (L_T) of boundary line, whereat $L= L_q\cup L_T$.

In the simplest method of discretization the integral equation (5) in relation to time, it is supposed that the time variations of the functions $T(\mathbf{q},\tau)$ and $q(\mathbf{q},\tau)$ are small as compared to functions $T^*(\mathbf{p},\mathbf{q};t_k,\tau)$ and $Q^*(\mathbf{p},\mathbf{q};t_k,\tau)$. It can be reasonably assumed that the functions $T(\mathbf{q},\tau)$ and $q(\mathbf{q},\tau)$ are also constant in small time period $[t_{k-1},t_k]$ (Wrobel, 2002; Kythe, 2005).

Accordingly to the above assumption the integral equation (5) can be denoted in the form:

$$\chi(\mathbf{p})T(\mathbf{p},t_k)+\alpha\int_{(L_q)}T(\mathbf{q},\tau)\tilde{Q}^*(\mathbf{p},\mathbf{q};t_k,\tau)dL_{\mathbf{q}}+$$

$$\frac{1}{c}\int_{(L_T)}q(\mathbf{q},\tau)\tilde{T}^*(\mathbf{p},\mathbf{q};t_k,\tau)dL_{\mathbf{q}}+$$

$$\iint_{(\Lambda)}T_0(\mathbf{v})\hat{T}^*(\mathbf{p},\mathbf{v};t_k,t_0)d\Lambda_{\mathbf{v}}=0, \qquad (6)$$

where the kernels $\tilde{T}^*(\mathbf{p},\mathbf{q};t_k,\tau)$ and $\tilde{Q}^*(\mathbf{p},\mathbf{q};t_k,\tau)$ are given by expression:

$$\tilde{T}^*(\mathbf{p},\mathbf{q};t_k,\tau)=\frac{1}{c}\int_{(L_T)}T^*(\mathbf{p},\mathbf{q};t_k,\tau)dL_{\mathbf{q}}=$$

$$=\frac{1}{4\pi\lambda}\int_{t_0}^{t_k}\frac{1}{(t_k-\tau)}\exp\left[-\frac{r_{\mathbf{pq}}^2}{4\alpha(t_k-\tau)}\right]d\tau=$$

$$=\frac{1}{4\pi\lambda}Ei\left(-\frac{r_{\mathbf{pq}}^2}{4\alpha(t_k-t_0)}\right), \qquad (6a)$$

where $Ei(.)$ is the exponential integral function:

$$\tilde{Q}^*(\mathbf{p},\mathbf{q};t_k,\tau)=\alpha\int_{(L_T)}Q^*(\mathbf{p},\mathbf{q};t_k,\tau)dL_{\mathbf{q}}=$$

$$=\frac{d}{8\pi\alpha}\int_{t_0}^{t_k}\frac{1}{(t_k-\tau)^2}\exp\left[-\frac{r_{\mathbf{pq}}^2}{4\alpha(t_k-\tau)}\right]d\tau=$$

$$=\frac{d}{2\pi r_{\mathbf{pq}}{}^2}\exp\left(-\frac{r_{\mathbf{pq}}^2}{4\alpha(t_k-t_0)}\right), \qquad (6b)$$

where $d=r_{x\mathbf{pq}}\cdot\left|\partial y/\partial l\right|_{\mathbf{q}}-r_{y\mathbf{pq}}\cdot\left|\partial x/\partial l\right|_{\mathbf{q}}$.

3.2. Boundary integral equation for heat conduction equation with periodic boundary condition

The unsteady heat conduction problem in two dimensional object with condition of periodical changes of temperature on boundary line is described by the equation (1) with periodic boundary condition (1d).

In this case, the temperature may be treated as the function:

$$T(x,y,t)=U(x,y)\exp(-i\omega t) \qquad (7)$$

where only the real part of the above expression has physical meaning as consequence of boundary condition (1d) and the basic relation for complex functions: $\exp(-iz)=\cos(z)-i\sin(z)$.

Inserting space and time derivatives of the temperature, expressed by relation (7), to the equation (1), leads to the modified Helmholtz differential equation for function $U\equiv U(x,y)$ (Sorko and Karpowich, 2007).

$$\nabla^2U-\hat{k}^2U=0;\quad \text{where}\quad \hat{k}=\sqrt{-i\,\omega/\alpha} \qquad (8)$$

The integral solution of differential equation (8) has the form:

$$\int_{(L)}G(\mathbf{p},\mathbf{q})\frac{\partial}{\partial n_{\mathbf{q}}}U(\mathbf{q})\,dL_{\mathbf{q}}=$$

$$-\chi(\mathbf{p})U(\mathbf{p})+\int_{(L)}U(\mathbf{q})\frac{\partial}{\partial n_{\mathbf{q}}}G(\mathbf{p},\mathbf{q})dL_{\mathbf{q}} \qquad (9)$$

where Green function $G(\mathbf{p},\mathbf{q})$ of the Helmholtz equation is given by a modified Bessel function:

$$G(\mathbf{p},\mathbf{q})=\frac{1}{2\pi}K_0(\hat{k}r_{\mathbf{pq}}) \qquad (10)$$

Modified Bessel function of order (0) of complex argument $(\hat{k}r_{\mathbf{pq}})$ can be expressed by the Kelvin functions of real argument $(kr_{\mathbf{pq}})$ in which $k=\sqrt{\omega/\alpha}$, then the Green function and its derivative take the form:

$$G(\mathbf{p},\mathbf{q})=\frac{1}{2\pi}K_0(\hat{k}r_{\mathbf{pq}})=\frac{1}{2\pi}\left[\ker_0(kr_{\mathbf{pq}})-ikei_0(kr_{\mathbf{pq}})\right] \qquad (10a)$$

$$\frac{\partial}{\partial n_{\mathbf{q}}}G(\mathbf{p},\mathbf{q})=\left(\frac{\partial}{\partial r}G(\mathbf{p},\mathbf{q})\right)_{\mathbf{q}}\left(\frac{\partial r}{\partial n}\right)_{\mathbf{q}}=H(\mathbf{p},\mathbf{q})$$

$$=k\left[\ker_0{}'(kr_{\mathbf{pq}})-ikei_0{}'(kr_{\mathbf{pq}})\right]\left(\frac{\partial r}{\partial n}\right)_{\mathbf{q}} \qquad (10b)$$

Taking the limit as the source point $\mathbf{p}$ approaching contour (L), where function $U(\mathbf{q})$ is equal to the amplitude of the temperature oscillations and expressing the limit of the double layer potential in equation (9) in terms of its principal value (when $\mathbf{p}\equiv\mathbf{q}$), one obtains integral equation of the first kind for the normal derivative of the function $U(\mathbf{q})$

$$\int_{(L)}G(\mathbf{p},\mathbf{q})\frac{\partial}{\partial n_{\mathbf{q}}}U(\mathbf{q})\,dL_{\mathbf{q}}=\tilde{T}\left[\int_{(L)}^{PV}H(\mathbf{p},\mathbf{q})dL_{\mathbf{q}}-\frac{1}{2}\right] \qquad (11)$$

Separating the real part and imaginary part of kernels in integral equation (10), one receives the system of two integral equations in relation to functions $F(\mathbf{q})=(F_{Re}(\mathbf{q}),F_{Im}(\mathbf{q}))$, which are the derivatives of function $U(\mathbf{q})$.

$$\int_{(L)}G_{Re}(\mathbf{p},\mathbf{q})F_{Re}(\mathbf{q})\,dL_{\mathbf{q}}=\tilde{T}\left[\int_{(L)}^{PV}H_{Re}(\mathbf{p},\mathbf{q})dL_{\mathbf{q}}-\frac{1}{2}\right] \qquad (11a)$$

$$\int_{(L)}G_{Im}(\mathbf{p},\mathbf{q})F_{Im}(\mathbf{q})\,dL_{\mathbf{q}}=\tilde{T}\left[\int_{(L)}^{PV}H_{Im}(\mathbf{p},\mathbf{q})dL_{\mathbf{q}}-\frac{1}{2}\right] \qquad (11b)$$

After determination of discrete values $F_{Re}(\mathbf{q})$, $F_{Im}(\mathbf{q})$ on the boundary (L), the values $U_{Re}(\mathbf{q})$, $U_{Im}(\mathbf{q})$ of the function $U(\mathbf{q})$ at points of domain (Λ) are obtained from the system of equations

$$U_{Re}(\mathbf{p}) = \int_{(L)} F_{Re}(\mathbf{q}) G_{Re}(\mathbf{p},\mathbf{q}) dL_{\mathbf{q}} + \tilde{T} \int_{(L)} H_{Re}(\mathbf{p},\mathbf{q}) dL_{\mathbf{q}} \quad (12a)$$

$$U_{Im}(\mathbf{p}) = \int_{(L)} F_{Im}(\mathbf{q}) G_{Im}(\mathbf{p},\mathbf{q}) dL_{\mathbf{q}} + \tilde{T} \int_{(L)} H_{Im}(\mathbf{p},\mathbf{q}) dL_{\mathbf{q}} \quad (12b)$$

4. NUMERICAL SOLUTION OF INTEGRAL EQUATION OF HEAT CONDUCTION WITH PERIODIC BOUNDARY CONDITION

Numerical solution of integral equations in two dimensional problems consists in discretization of the boundary line into straight or arc elements with constant or linear distributed value of investigated function and consequently, the integral equation transforms to the system of algebraic linear equations in relation to the unknown integrand.

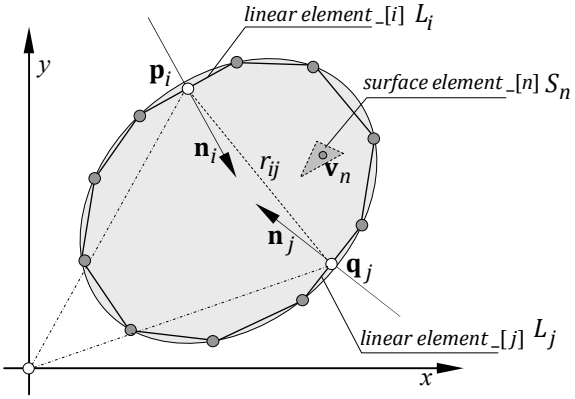


Fig. 2. Discretization of area (Λ)

Discrete solution of integral equation (6) can be obtained dividing the boundary line (L) into I (i=1,2,3,...,I) ,straight line elements, domain (Λ) into N (n=1,2,3,...,N) surface elements and time interval $[t_0,t_K]$ into K (k=1,2,3,...,K) subintervals (Fig. 2).

Using the nodal values, with assumption that the functions $T(\mathbf{q}_j,\tau)$ and $q(\mathbf{q}_j,\tau)$ are constant on each linear element L_j, function $T_0(\mathbf{v}_n)$ is constant on each surface element S_n and also they areconstant at any subintervals $[t_{k-1},t_k]$, the boundary integral equation is obtained in discrete form as follows:

$$\chi(\mathbf{p}_i)T(\mathbf{p}_i,t_K) + \alpha \sum_{\substack{j=1 \\ j\neq i}}^{J} \sum_{k=1}^{K} T(\mathbf{q}_j,t_k)\tilde{Q}^*(\mathbf{p}_i,\mathbf{q}_j;t_k,t_0)\Delta L_j + \frac{1}{c} \sum_{\substack{j=1 \\ j\neq i}}^{J} \sum_{k=1}^{K} q(\mathbf{q}_j,t_k)\tilde{T}^*(\mathbf{p}_i,\mathbf{q}_j;t_k,t_0)\Delta L_j + \sum_{n=1}^{N} T_0(\mathbf{v}_n) \sum_{k=1}^{K} \hat{T}^*(\mathbf{p}_i,\mathbf{v}_n;t_k,t_0)\Delta S_n \quad , \quad (13)$$

where:

$$\tilde{Q}^*(\mathbf{p}_i,\mathbf{q}_j;t_k,t_0) = \frac{d_{ij}}{2\pi r_{ij}^2} \exp\left(-\frac{r_{ij}^2}{4\alpha(t_k-t_0)}\right) \quad (13a)$$

$$d_{ij} = \Delta x_{ij} \cdot \left|\Delta y / \Delta l\right|_j - \Delta y_{ij} \cdot \left|\Delta x / \Delta l\right|_j \quad (13a^*)$$

$$\tilde{T}^*(\mathbf{p}_i,\mathbf{q}_j;t_k,t_0) = \frac{1}{4\pi\lambda} Ei\left(\frac{r_{ij}^2}{4\alpha(t_k-t_0)}\right) \quad (13b)$$

$$\hat{T}^*(\mathbf{p}_i,\mathbf{v}_n;t_k,t_0) = \frac{1}{4\pi\lambda} Ei\left(\frac{r_n^2}{4\alpha(t_k-t_0)}\right) \quad (13c)$$

Similarly, the integral equation (10) expressed with the system of two integral equations (10a) and (10b), describing properly the real part and the imaginary part of the function, by discretization of the boundary line moves to two systems of linear equations:

$$\sum_{j=1}^{j} F_{Re}(\mathbf{q}_j)\tilde{G}_{Re}(\mathbf{p}_i,\mathbf{q}_j) = \tilde{T}\left[\sum_{j=1}^{J} \tilde{H}_{Re}(\mathbf{p}_i,\mathbf{q}_j) - \frac{1}{2}\right] \quad (14a)$$

$$\sum_{j=1}^{j} F_{Im}(\mathbf{q}_j)\tilde{G}_{Im}(\mathbf{p}_i,\mathbf{q}_j) = \tilde{T}\left[\sum_{j=1}^{J} \tilde{H}_{Im}(\mathbf{p}_i,\mathbf{q}_j) - \frac{1}{2}\right] \quad (14b)$$

where:

$$\left.\begin{aligned} G_{Re}(\mathbf{p}_i,\mathbf{q}_j) &= \int_{(L_j)} G_{Re}(\mathbf{p}_i,\mathbf{q}_j) dL_j \\ \tilde{H}_{Re}(\mathbf{p}_i,\mathbf{q}_j) &= \int_{(L_j)} H_{Re}(\mathbf{p}_i,\mathbf{q}_j) dL_j \end{aligned}\right\} \quad (14a^*)$$

$$\left.\begin{aligned} \tilde{G}_{Im}(\mathbf{p}_i,\mathbf{q}_j) &= \int_{(L_j)} G_{Im}(\mathbf{p}_i,\mathbf{q}_j) dL_j \\ \tilde{H}_{Im}(\mathbf{p}_i,\mathbf{q}_j) &= \int_{(L_j)} H_{Im}(\mathbf{p}_i,\mathbf{q}_j) dL_j \end{aligned}\right\} \quad . \quad (14b^*)$$

5. EXAMPLES

Basing on developed BEM algorithm, a new authoring computer program was written in Fortran, which was applied to the following examples.

Example 1:

The accuracy of the formulation was tested by computing the heat field in a finite square a=1.0 m, when non zero initial temperatures are prescribed inside the domain and variable temperatures are assumed on the boundaries.

The thermal properties of the homogeneous medium are assumed to be: thermal conductivity λ=200.0 W/(mK), volumetric

specific heat c=2.0 10^6 J/(m^3K), which defines a thermal diffusivity α=1.0 10^{-4} m^2/s.

The temperature distribution on the boundary at t_0=0 is described by the relations:

$$\left.\begin{aligned} T(1,y,0) &= 100.0\cdot(1.0-\sin(0.5\pi y)\\ T(x,1,0) &= 100.0\cdot(1.0-\cos(0.5\pi x)\\ T(0,y,0) &= 100.00\\ T(x,0,0) &= 100.00 \end{aligned}\right] \quad (15a)$$

and the initial temperature distribution, satisfying the boundary conditions described above, is given by the relation:

$$T(x,y,0) = 100.0\cdot(1.0-\cos(0.5\pi x)\sin(0.5\pi y) \quad (15b)$$

and is presented on the sketch (Fig. 3b.)

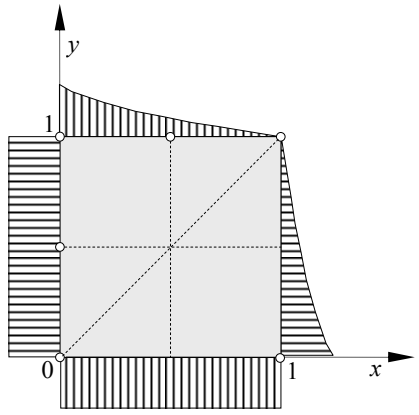


Fig. 3a. The unit square and boundary conditions

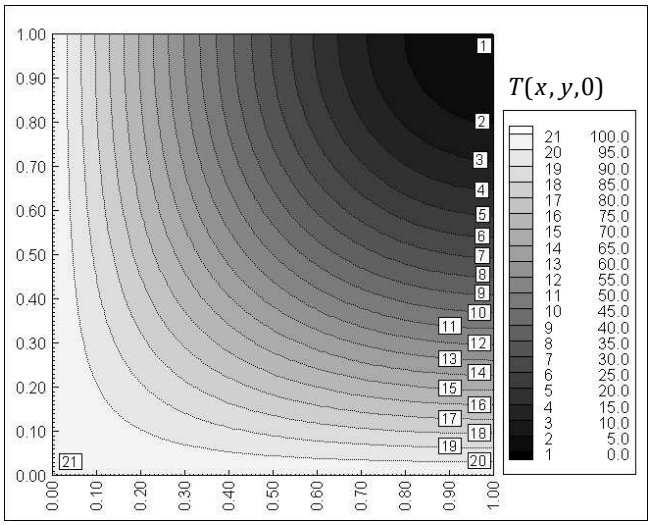


Fig. 3b. Temperature distribution in the unit square

The field of the temperature is symmetrical in relation to the diagonal of the square, so the time changes of the temperature can conveniently be presented along the line x=y.

Temporal evolution of the temperature $T=f(t)$ along diagonal of the square is shown on sketch (Fig. 4.) and the changes of field temperature are presented on sketch (Fig. 5.)

The presented above problem has the analytical solution as follows

$$T(x,y,t) = 100\cdot\left[1-\exp(-2\pi^2\alpha t)\cos\left(\frac{\pi x}{2}\right)\sin\left(\frac{\pi y}{2}\right)\right] \quad (15c)$$

In numerical solution of considered problem with the boundary element method one assumes: 400 similar line elements on boundary, 100 square elements with collocation points in the geometrical center of every area the temporary step Δt =1.0 at estimated time t_{max}=3600.

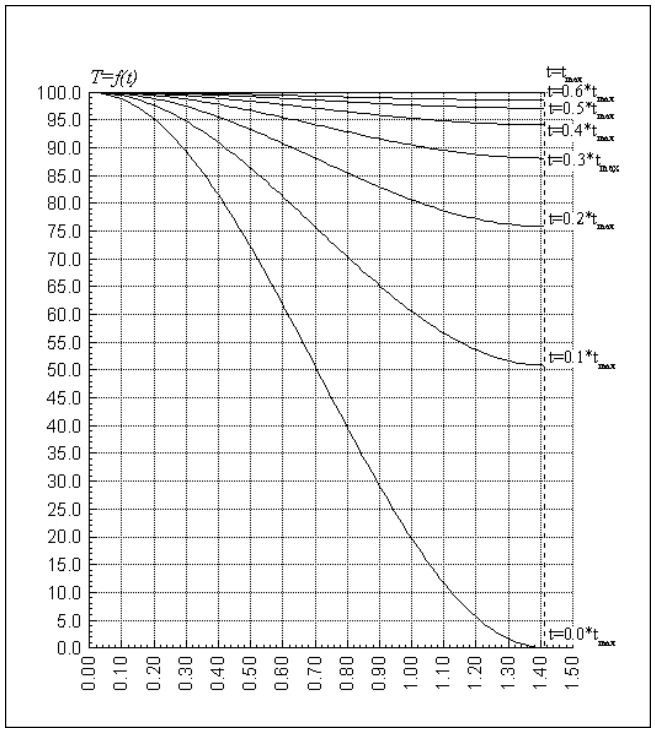


Fig. 4. Temperature distribution $T=f(t)$ along diagonal of the square

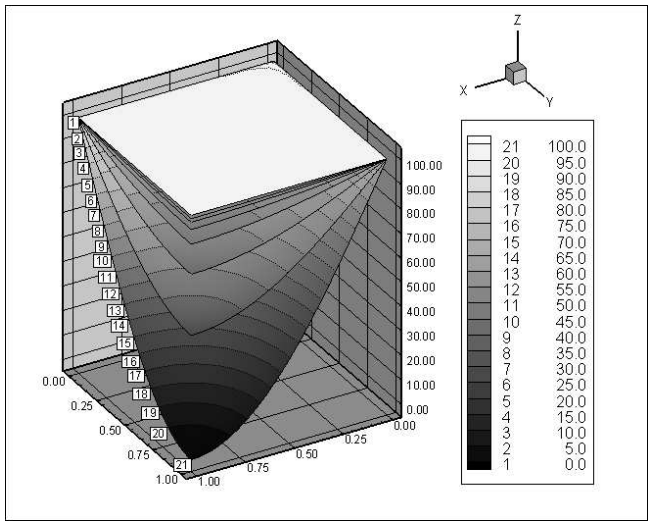


Fig. 5. Temperature distribution $T=f(t)$ in the square

The maximum error of numeric solution, estimated from relation:

$$\delta_{max}=100(T(x,y,t)_{th}-T(x,y,t)_{num})/T(x,y,t)_{th}, \quad (16)$$

does not exceed the value 0.1%.

Example 2:

In technical problems of optimization the devices using renewable thermal energy, that for designing ground heat exchangers (horizontal and vertical) of heat pump systems, it is necessary to determine the annual ground temperature distribution for various values of ground thermal conductivity coefficient. This problem is the subject of many empirical studies, leading to formulation of complex empirical formulas describing the annual temperature propagation.

The mathematical description of ground temperature distribution problem consists in solving the transient heat conduction problem in homogeneous or heterogeneous area with constant thermal conductivity coefficient and with boundary conditions assuming the heat flux on the boundaries of value equal 0 (Fig. 6). On the ground surface the boundary periodic condition is assumed, that is the changeable annual temperature of ambient in the form:

$$T_u = T_{sa} + \Delta T_a \cos(\omega t) \tag{17}$$

The problem can be considered in heterogeneous area composed of layers of known thickness and known values of thermal conductivity coefficient.

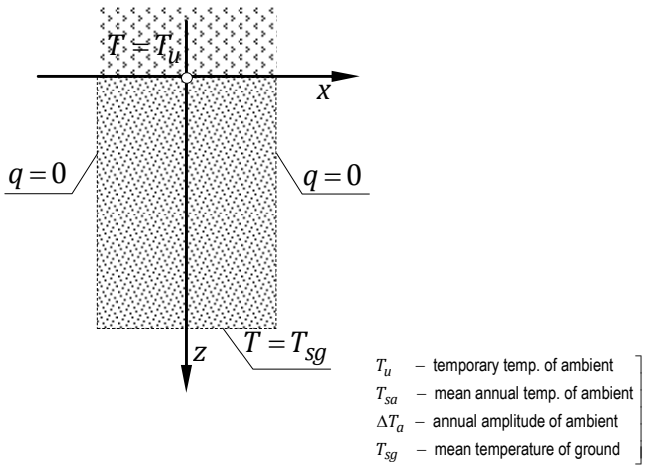


Fig. 6. Area with boundary temperature and heat flux

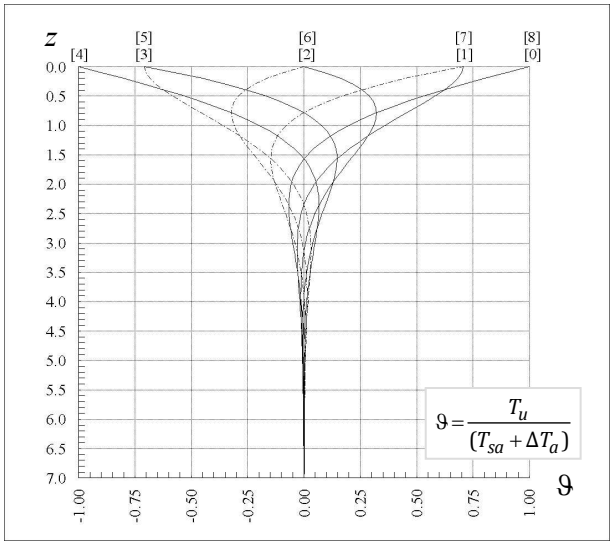


Fig. 7. Temperature profile in the ground

The Fig. 6 shows the sketch of area with boundary conditions of considered problem. The solution of the problem, in the form of unified temperature distribution (8) from the surface layer to layer of constant temperature at every moment of cycle of annual changes of temperature, is presented in the Fig. 7.

6. CONCLUSIONS

In this paper the utility of the boundary element method for solving the transient heat conduction problem with periodic boundary condition is proved. The general solutions of Fourier equation with initial and boundary value problems are introduced, on the assumption that temperature changes periodically on the boundary. The new mathematical algorithm is developed, which is further verified by solving transient heat conduction problem in two dimensional area. The comparison between analytical and numerical solution of test problem proves the great accuracy of proposed BEM algorithm. Finally, the method is applied to solve the ground temperature distribution problem with the boundary condition of the oscillating temperature of ambient. All numerical computations were made with the use of a new computer program, written by authors, in Fortran.

Although, the boundary element method is not so widely applied, as an efficient numerical method and computational tool, constitutes the great alternative to popular mesh methods (FEM, FDM), and can be successfully employed for analysis of many engineering problems.

REFERENCES

1. **Białecki R.A., Jurgaś P., Günther K.** (2002), Dual reciprocity BEM without matrix inversion for transient heat conduction, *Engineering Analysis with Boundary Elements*, 26, 227–236.
2. **Brebbia C.A.** (ed) (1984) Topics in Boundary Element Research. Vol1. Basic Principles and Applications Springer-Verlag
3. **Brebbia C.A.,Telles J.C.F., Wrobel L.C.** (1984), Boundary Element Techniques. Theory and Applications in Engineering, Springer-Verlag
4. **Cheng R.J., Liew K.M.** (2012), A meshless analysis of three-dimensional transient heat conduction problems, *Engineering Analysis with Boundary Elements 36,* 203–210.
5. **Erhart K., Divo E., Kassab A.J.** (2006), A parallel domain decomposition boundary element method approach for the solution of large-scale transient heat conduction problems, *Engineering Analysis with Boundary Elements*, 30, 553–563.
6. **Godinho L., Tadeu A., Simoes N.** (2004), Study of transient heat conduction in 2.5D domains using the boundary element method, *Engineering Analysis with Boundary Elements 28,* 593–606.
7. **Johansson B.T., Lesnic D**. (2008), A method of fundamental solutions for transient heat conduction, *Engineering Analysis with Boundary Elements*, 32, 697–703.
8. **Johansson T., Lesnic D.** (2009), A method of fundamental solutions for transient heat conduction in layered materials, *Engineering Analysis with Boundary Elements*, 33, 1362–1367.
9. **Katsikadelis, J.T.** (2002), *Boundary Elements. Theory and Applications*, Elsevier Science Ltd.
10. **Kythe P.K.** (2005), *Introduction to Boundary Element Methods*, CRC Press.
11. **Li Q.-H., Chen S.-S., Kou G.-X.** (2011), Transient heat conduction analysis using the MLPG method and modified precise time step integration method, *Journal of Computational Physics*, 230, 2736–2750.
12. **Lu X., Tervola P., Viljanen M.** (2006), Transient analytical solution to heat conduction in composite circular cylinder, *International Journal of Heat and Mass Transfer*, 49, 341–348.

13. **Lu X., Viljanen M.** (2006), An analytical method to solve heat conduction in layered spheres with time-dependent boundary conditions. *Physics Letters A 351,* 274–282.
14. **Majchrzak E.** (2001*), Boundary element method in heat transfer***,** Częstochowa University of Technology (in Polish).
15. **Mansur W.J., Vasconcellos C.A.B., Zambrozuski N.J.M., Rotunno Filho O.C.** (2009), Numerical solution for the linear transient heat conduction equation using an Explicit Green's Approach. *International Journal of Heat and Mass Transfer*, 52, 694–701.
16. **Mohammadia M., Hematiyan M.R., Marin L.** (2010), Boundary element analysis of nonlinear transient heat conduction problems involving non-homogenous and nonlinear heat sources using time-dependent fundamental solutions, *Engineering Analysis with Boundary Elements*, 34*,* 655–665.
17. **Monte F., Beck J.V., Amos D.E.** (2012), Solving two-dimensional Cartesian unsteady heat conduction problems for small values of the time*, International Journal of Thermal Sciences 60,* 106– 113.
18. **Ochiai Y., Kitayama Y.** (2009) Three-dimensional unsteady heat conduction analysis by triple-reciprocity boundary element method, *Engineering Analysis with Boundary Elements*, 33*,* 789–795.
19. **Ochiai Y., Sladek V., Sladek J.** (2006) Transient heat conduction analysis by triple-reciprocity boundary element method*, Engineering Analysis with Boundary Elements*, 30*,* 194–204.
20. **Pozrikidis C.A.** (2000), *Practical Guide to Boundary Element Methods with the software*, Library BEMLIB Chapman&Hall/CRC.
21. **Rantala J.** (2005), A new method to estimate the periodic temperature distribution underneath a slab-on-ground structure, *Building and Environment*, 40*,* 832–840.
22. **Simoes N., Tadeu A., Antonio J., Mansur W.** (2012), Transient heat conduction under nonzero initial conditions: A solution using the boundary element method in the frequency domain*, Engineering Analysis with Boundary Elements*, 36*,* 562–567.
23. **Singh S., Jain P.K., Rizwan-uddin** (2008), Analytical solution to transient heat conduction in polar coordinates with multiple layers in radial direction*, International Journal of Thermal Sciences*, 47*,* 261–273.
24. **Soleimani S., Jalaal M., Bararnia H., Ghasemi E., Ganji D.D., Mohammadi F.** (2010), Local RBF-DQ method for two-dimensional transient heat conduction problems*, International Communications in Heat and Mass Transfer*, 37, 1411–1418.
25. **Sorko S.A., Karpovich S.** (2007), Solving the unsteady heat transfer problem with periodic boundary condition by the boundary integral equations method, *Teoretičeskaă i Prikladnaă Mehanika*, Vol. 22.
26. **Sutradhar A., Paulino G.H.** (2004), The simple boundary element method for transient heat conduction in functionally graded materials, *Computer Methods in Applied Mechanics and Engineering*, 193*,* 4511–4539.
27. **Tanaka M., Matsumoto T., Takakuwa S.** (2006), Dual reciprocity BEM for time-stepping approach to the transient heat conduction problem in nonlinear materials, *Computer Methods in Applied Mechanics and Engineering*, 195*,* 4953–4961.
28. **Wrobel L.C.** (2002), *The Boundary Element Method Vol I. Applications in Thermo-Fluids and Acoustic*, Willey.
29. **Yang K., Gao X.-W.** (2010), Radial integration BEM for transient heat conduction problems, *Engineering Analysis with Boundary Elements*, 34*,* 557–563.
30. **Yumrutas R., Unsal M., Kanoglu M.** (2005), Periodic solution of transient heat flow through multilayer walls and flat roofs by complex finite Fourier transform technique, *Building and Environment*, 40*,* 1117–1125.
31. **Zhang X.H., Ouyang J., Zhang L.** (2009), Matrix free meshless method for transient heat conduction problems, *International Journal of Heat and Mass Transfer*, 52, 2161–2165.

Acknowledgement: The work was supported by Bialystok University of Technology Research Project S/WBiIŚ,5/11.

ABSTRACTS

Anna Demianiuk, Sławomir Adam Sorko
Analysis of Flow and Thermal Phenomena In Evacuated Tube Collectors

The subject of this case study is an issue of optimisation of flat tube solar collectors. Basic elements of energy analysis of per-formance parameters described by Hottel-Whillier equation are presented in the article. It is considered to be crucial to precisely analyse fluid flow through flow elements in evacuated tube collectors. It is especially important in the case of systems with channels of cross-sections shapes different from circular and for the use of detailed mathematical description of complex film conduction phenomena. It is presented that the advanced analysis of the flow and thermal phenomena in complex heat transfer systems, represented by evacuated tube collectors, enables engineering rationalisation of technical solutions for these devices.

Krzysztof Dziewiecki, Zenon Mazur, Wojciech Blajer
Assessment of Muscle Forces and Joint Reaction in Lower Limbs During the Take-Off from the Springboard

Computer simulation methods, based on the biomechanical models of human body and its motion apparatus, are commonly used for the assessment of muscle forces, joint reactions, and some external loads on the human body during its various activities. In this paper a planar musculoskeletal model of human body is presented, followed by its application to the inverse simulation study of a gymnast movement during the take-off from the springboard when performing the handspring somersault vault on the table. Using the kinematic data of the movement, captured from optoelectronic photogrammetry, both the internal loads (muscle forces and joint reactions) in the gymnast's lower limbs and the external reactions from the springboard were evaluated. The calculated vertical reactions from the springboard were then compared to the values assessed using the captured board displacements and its measured elastic behaviors.

Tadeusz Kaczorek
Factorization of Nonnegative Matrices by the Use of Elementary Operation

A method based on elementary column and row operations of the factorization of nonnegative matrices is proposed. It is shown that the nonnegative matrix $A \in \Re_+^{n\times m}$ $(n \geq m)$ has positive full column rank if and only if it can be transformed to a matrix with cyclic structure. A procedure for computation of nonnegative matrices $B \in \Re_+^{n\times r}$, $C \in \Re_+^{r\times m}$ $(r \leq \mathrm{rank}\,(n,m))$ satisfying $A = BC$ is proposed.

Dmitrij B. Karev, Vladimir G. Barsukov
Biomechanical Analysis of Two-Point Asymmetric Screw Fixation with Implant for Femoral Neck Fracture

Stressed state peculiarities of cortical and trabecular bones by two-point asymmetric screw fixation with implant for femoral neck fracture are studied. Layer construction mechanic methods are used for analysis of stresses in cortical and trabecular bones. Biomechanical conditions for non-opening of the junction of the bone parts being joined are determined. It has been found that the total tightness of the broken parts when they rest against each other is secured over the whole fracture section without junction opening under condition that fixing screws are positioned in the trabecular bone without penetration of the thread side surface into cortical bone.

Witold Kosiński, Wiera Oliferuk
Stationary Action Principle for Vehicle System with Damping

The aim of this note is to show possible consequences of the principle of stationary action formulated for non-conservative systems. As an example, linear models of vibratory system with damping and with one and two degrees of freedom are considered. This kind of models are frequently used to describe road and rail vehicles. There are vibrations induced by road prole. The appropriate action functional is proposed with the Lagrangian density containing: the kinetic and potential energies as well as dissipative one. Possible variations of generalized coordinates are introduced together with a non-commutative rule between operations of taking variations of the coordinates and their time derivatives. The stationarity of the action functional leads to the Euler-Lagrange equations.

Adam Kotowski
Frequency Analysis with Cross-Correlation Envelope Approach

A new approach for frequency analysis of recorded signals and readout the frequency of harmonics is presented in the paper. The main purpose has been achieved by the cross-correlation function and Hilbert transform. Using the method presented in the paper, there is another possibility to observe and finally to identify single harmonic apart from commonly used Fourier transform. Identification of the harmonic is based on the effect of a straight line of the envelope of the cross-correlation function when reference and signal harmonic have the same frequency. This particular case is the basis for pointing the value of the frequency of harmonic component detected.

Zbigniew Kulesza
Rotor Crack Detection Approach Using Controlled Shaft Deflection

Rotating shafts are important and responsible components of many machines, such as power generation plants, aircraft engines, machine tool spindles, etc. A transverse shaft crack can occur due to cyclic loading, creep, stress corrosion, and other mechanisms to which rotating machines are subjected. If not detected early, the developing shaft crack can lead to a serious machine damage resulting in a catastrophic accident. The article presents a new method for shaft crack detection. The method utilizes the coupling mechanism between the bending and torsional vibrations of the cracked, non-rotating shaft. By applying an external lateral force of a constant amplitude, a small shaft deflection is induced. Simultaneously, a harmonic torque is applied to the shaft inducing its torsional vibrations. By changing the angular position of the lateral force application, the position of the deflection also changes opening or closing of the crack. This changes the way the bending and torsional vibrations are being coupled. By studying the coupled lateral vibration response for each angular position of the lateral force one can assess the possible presence of the crack. The approach is demonstrated with a numerical finite element model of a rotor. The results of the numerical analysis demonstrate the potential of the suggested approach for effective shaft crack detection.

Tomasz Nartowicz
Design of Fractional Order Controller Satisfying Given Gain and Phase Margin for a Class of Unstable Plant with Delay

The paper describes the design problem of fractional order controller satisfying gain and phase margin of the closed loop system with unstable plant with delay. The proposed method is based on using Bode's ideal transfer function as a reference transfer function of the open loop system. Synthesis method is based on simplify of the object transfer function. Fractional order of the controllers is relative with gain and phase margin only. Computer method for synthesis of fractional controllers is given. The considerations are illustrated by numerical example and results of computer simulation with MATLAB/Simulink.

Tomasz Nartowicz
Analitical Method of PID Controller Tuning for a Class of Unstable Plant

The aim of the paper is to present the synthesis method of classic PID controller for a class of unstable plant. The proposed method based directly on Skogestad paper, where analitical synthesis of PID controllers in described. This paper is generalization of that approach on a class of unstable plants with delay. Analitical method for synthesis of fractional controllers is given. The considerations are illustrated By numerical example and results of computer simulation with MATLAB/Simulink.

Ewa Pawłuszewicz
Null-Controllability of Linear Systems on Time Scales

The purpose of the paper is to study the problem of controllability of linear control systems with control constrains, defined on a time scale. The obtained results extend the existing ones on any time domain. The set of values of admissible controls is a given closed and convex cone with nonempty interior and vertex at zero or is a subset of $\boldsymbol{R^m}$ containing zero.

Norbert Sczygiol
A New Stress Criterion for Hot-Tearing Evaluation in Solidifying Casting

This work concerns a new criterion for hot-tearing evaluation in castings. Algorithm describing the conduction of computer simu-lations of phenomena accompanying the casting formation, which performing is the preparation stage for using of this criterion, is also described. According to the low recurrence of phenomena occurring during solidification (e.g. grained structure parameters, stresses distribution) the casting's hot-tearing inclination can be estimated only in approximated manner. Because of still following at present rapid computer processors development, and techniques of its programming, enables to suppose that in short time the efficiency of computer simulations will arise so much, the problem of hot-tearing evaluation newly became interesting for the team working on computer simulations at the Institute for Computer and Information Sciences at Częstochowa University of Technology.

Wiesław Szymczyk
Numerical Analysis of Residual Stress in a Gradient Surface Coating

There were conducted numerical analyses of thermal residual stress in a double gradient surface coating consisting of porous ZrO2 of 500 μm thickness, placed on Inconel 718 substrate, with an interlayer of NiCoCrAlY of 200 m thickness. It was assumed that the coating was deposited by plasma spraying. It showed out roughness of the free surface and borders between material phases. Numerical model took into consideration the real geometry of material structure. The results were analysed after transforming them into discrete distributions of particular components of stress. Distributions were presented for particular zones of the surface coating. This allowed to obtain signals originated by distinct features of the material structure such as particular material phases, their contact borders, pores, roughness of the external surface, roughness of internal borders between layers of the gradient system and others.

Anna Justyna Werner-Juszczuk, Sławomir Adam Sorko
Application of Boundary Element Method to Solution of Transient Heat Conduction

The object of this paper is the implementation of boundary element method to solving the transient heat transfer problem with nonzero boundary condition and particularly with periodic boundary condition. The new mathematical BEM algorithm for two dimensional transient heat conduction problem with periodic boundary condition is developed and verified. The results of numerical simulation of transient heat conduction in two dimensional flat plate under non zero initial condition are compared with results obtained with analytical method. Then the practical application of developed algorithm is presented, that is the solution of ground temperature distribution problem with oscillating temperature of ambient. All results were obtained with a new authoring computer program for solving transient heat conduction problem, written in Fortran.