Computer Assisted Mechanics and Engineering Sciences, 7: 571–588, 2000.

Numerical modelling and robust parametric estimation of surfaces useful for representation of tooth flanks

Andrzej Kawalec
Faculty of Mechanical Engineering and Aeronautics, Rzeszów University of Technology, ul. W. Pola 2, 35-959 Rzeszów, Poland

(Received November 3, 1999)

This paper concerns modelling of surfaces resulting from measurements or from digital simulation of surfaces e.g., tooth flanks, based on the theory of gearing, with intentionally introduced modification usually defined in discrete form. Computational methods of modelling of curves and surfaces are briefly reviewed. Problems of stability of geometric modelling and related problems of parametric estimation of mathematical models, representing curves and surfaces are discussed. An analysis of multicollinearity of the measurement matrix is performed. A method of regularization of matrix, containing coordinates of nodes from considered surface, is proposed. This method allows to improve the robustness of parametric estimation and is specially helpful for on-line parametric estimation of surface models utilized during measurements.

1. Introduction

Gear transmissions play very important role in almost every machine, esp. in transport equipment: in aviation, automobile and shipbuilding industries, and in many stationary machines, i.e.: machine tools, manufacturing equipment, etc. They must fulfill certain strong criteria, like: transmission of very high loads, low vibrations and noise, high reliability, low weight, low wear, to name only a few. Therefore, modern gear transmissions must be developed in a way, which combines design and technological stages into one consistent algorithm of determination of gear transmission parameters, which enables their optimization with consideration to important criteria. This algorithm should allow modification of structural and technological parameters [26, 40] after every stage of calculations, starting from their initial values, and should include:

- numerical simulation of gear machining, made for the assumed method of machining and tool parameters – to determine tooth profile mapped by the tool;
- numerical simulation of kinematics of mating gears in gear transmission – to check for possible undesired tooth contact disturbances;
- strength calculations and tooth contact analysis to calculate stresses generated in gearing.

All the a.n. stages require precise geometrical modelling of tools and gears in mesh. Tooth flanks and fillets should be modelled on the basis of theoretical data (solution of corresponding equations following the theory of gearing) and measurement data. For these purposes related formulas for generation of gear tooth flanks were created on the basis of the theory of envelopes and the theory of gearing [23, 24, 26, 17]. Considering significant role of the shape of tooth profile (kinematics and transmission of load) as well as the shape of tooth fillet (where concentration of bending stresses occurs), corresponding mathematical formulas for calculation of the whole tooth geometry were developed [17]. Methods described in [17] enable generation of numerical models of spur gear tooth,

including not only tooth flanks but also tooth fillet surfaces machined by gear tools with precisely defined shapes.

Such consistent approach to parametric geometrical modelling of boundary surfaces, representing gear tooth, combined with finite element analysis [3, 20] performed with the use of the ADINA [1] finite element system, allowed to create some modules for optimization of structural and technological parameters for gear design [18].

In an ideal gear transmission, without any deviations, the teeth contact uniformly along the whole tooth width. In real gear transmission tooth contact differs from theoretical one, resulting from the involute gear theory. Tooth contact is disturbed due to manufacturing misalignments [8] and due to deflections of all loaded gear transmission elements (Fig. 1).

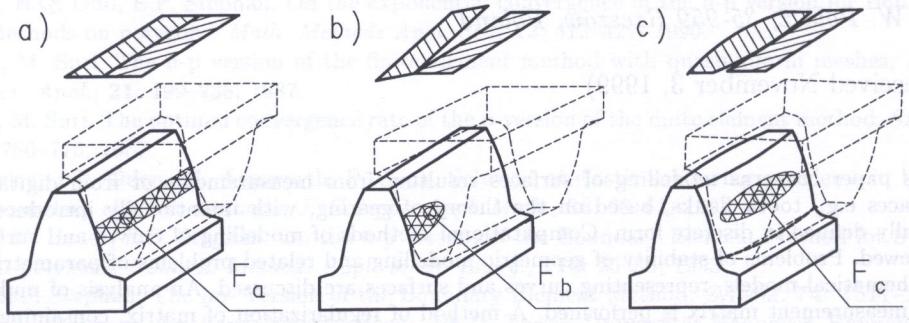

Fig. 1. Possible basic forms and locations of tooth contact area: (a) required form in ideal involute gear transmission, (b) boundary contact of involute tooth flanks, (c) tooth contact after optimal modification of one tooth flank $F_c = F_{\max}$

Both nonuniform distribution of tooth contact area and its incorrect location make kinematic and mechanical properties of gear transmission significantly worse [26, 32, 16]. Therefore, one of the central stages in the a.n. algorithm is proper tooth contact analysis (TCA) of mating gears in gear transmission [32, 42]. This technique enables to predict the level of transmission errors and the shape of the volume, where they develop. It also enables calculation of approximate shift of contact pattern, caused by the errors, which occur in gear manufacturing and assembly [4]. Tooth contact area considered in this analysis is an area of temporary contact of mating gears in gear transmission, which originates under acting load [26, 40].

Tooth contact analysis was developed by The Gleason Works [39] for the analysis of mating gear teeth in spiral bevel and hypoid gear transmissions. The main parameters analyzed were bearing contact and transmission errors, based on machine tool settings for spiral bevel generators. Next, TCA techniques were extended to models of theoretical point contact of tooth flanks [23] and to models of line contact of tooth flanks [24].

Negative influence of misalignments and deflections in gear transmission on its mechanical properties can be compensated by modification of tooth flanks [12, 30, 35, 41, 43]. This method, being one of the most effective methods of improving kinematic and strength properties of gear transmission, requires proper selection of the type and parameters of tooth flank modification (Fig. 2).

Tooth flanks after such modification differ from theoretical tooth flanks based on involutes. In the case of complex three-dimensional modification of gear tooth flanks their shape is obtained by machining on machine tools with Computer Numerical Control (CNC). After machining, tooth flanks are being measured (Fig. 3) on general purpose computer controlled Coordinate Measuring Machines (CMM) or on specialized CMMs, linked with CNC machine tools for gears. After verification, if any corrections are necessary, the gear is machined once again, until required shape of tooth flanks is obtained. Thus, in this manufacturing process, optimization of machine tool settings is done on the basis of measurement data [42]. According to this approach, the settings are changed in such a way that the resulting deviations of the machined gear tooth surface from the ideal one are minimized. These deviations and tooth flanks themselves are determined numerically [25] and compared with real surfaces. Due to flexibility of such approach esp., when CNC machine tool with

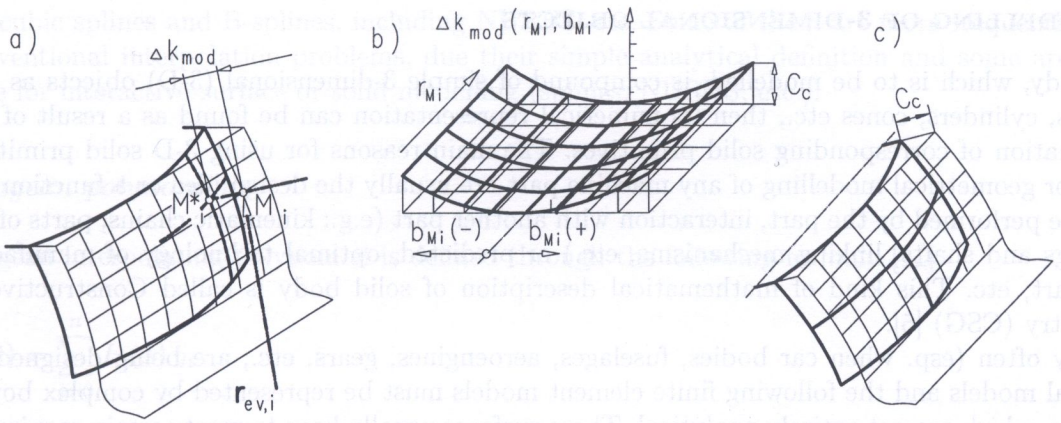

Fig. 2. Modification of tooth flanks by the vector function $\|\Delta \mathbf{k}_{\mathrm{mod}}(r_{Mi}, b_{Mi})\|$ determined in discrete points – nodes of the mesh of curves at the boundary surfaces: (a) crowning tooth flanks along tooth width, (b) modification function $\|\Delta \mathbf{k}_{\mathrm{mod}}(r_{Mi}, b_{Mi})\|$, (c) shapes of modified and non-modified tooth flanks

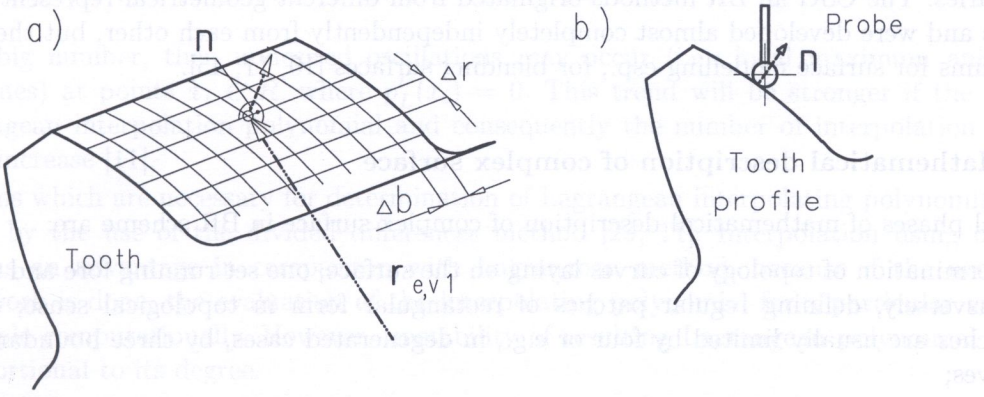

Fig. 3. Measurements of gear tooth flank on the CMM (a), with relationship between the tooth flank and the probe (b)

5-axes control is used, this manufacturing process can be applied to wide range of complex gear tooth surfaces [34, 46].

In order to reconstruct from measurements a real tooth surface of an arbitrary gear, a method was proposed [25], which fitts a second order polynomial surface to the error surface, using linear regression. Another approach, based on fitting a tensor product surface to the measurement data [19] uses the idea of parametric patches [11] and parametric splines. Although large effort has been put into the TCA, there are still uncertainties and opened questions concerning: (i) choice of surface models for suitable description of inspection data representing tooth geometry; (ii) improvement of ill-conditioned data that may occur after performing measurements on CMM gear inspection machines; (iii) effective calculation of relative location of numerically determined modified tooth surfaces or real surfaces measured in discrete points on the CMM gear inspection machines; (iv) accuracy of estimation of contact area and contact stresses over curved real tooth surfaces, represented in discrete form, etc.

Application of interpolation methods to measurement data may lead to highly nonlinear surface models incorporating either high degree polynomials or low degree curves with a set of related control points, which have to be determined in numerical way. They must be calculated with consideration to the fact that interpolating curve must pass through all measured points. These calculations, however, may lead to very complex surface models, hard to be handled in further strength calculations for TCA. Therefore, smoothing techniques, combined with estimation methods are being used to establish less complex surface models, representing important boundaries of the model to be used for TCA, including contact analysis with finite elements.

2. Modelling of 3-dimensional objects

If a body, which is to be modelled, is compound of simple 3-dimensional (3-D) objects as planes, spheres, cylinders, cones etc., then its numerical representation can be found as a result of logical combination of corresponding solid primitives. The main reasons for using 3-D solid primitives as basis for geometrical modelling of any machine part are usually the design idea or a function, which must be performed by the part, interaction with another part (e.g.: kinematic chains, parts of joints, bearings and shafts, linking mechanisms, etc.) or predicted, optimal technology of manufacturing of a part, etc. This kind of mathematical description of solid body is called Constructive Solid Geometry (CSG) [5].

Very often (esp. when car bodies, fuselages, aeroengines, gears, etc., are being designed), geometrical models and the following finite element models must be represented by complex boundary surfaces, which are not entirely analytical. These surfaces usually have to meet certain requirements, resulting from fluid dynamics, strength of materials, specific motion trajectory, etc. In such cases the other method of solid modelling is used – Boundary Representation (BR) [5, 10, 11]. This method allows definition of a solid body by description of basic and blending surfaces, playing the role of its boundaries. The CSG an BR methods originated from different geometrical representations of 3-D objects and were developed almost completely independently from each other, but they use similar algorithms for surface modelling esp., for blending surfaces [10, 11, 45].

2.1. Mathematical description of complex surface

General phases of mathematical description of complex surface in BR scheme are:

1. determination of topology of curves laying on the surface, one set running fore and aft, the other transversely, defining regular patches of rectangular form in topological sense; these surface patches are usually limited by four or e.g., in degenerated cases, by three boundary continuous curves;
2. measuring coordinates of nodes of the mesh on the physical model or calculating them from cross-sectional real or virtual drawings of considered surface;
3. numerical approximation of the a.n. network of curves, meeting all required smoothness specifications;
4. filling the areas of surface patches by defining interpolation curves of constant parameter (e.g.: $u = \text{const}$ or $v = \text{const}$) or implementing known models of elementary surface patches [10, 11] e.g.: Coon's, Gregory's, Bezier, B-Spline, etc.

Very often it is possible and suitable to define the mesh of curves as intersections of considered surface with two families of planes, orthogonal to each other, consisting of parallel planes each. Then, boundary curves of every surface patch are plane curves. Within this approach each surface patch $f(x_1, x_2)$ can be described in terms of two independent parameters x_1 and x_2, measured along the axes of considered plane coordinate system. Directrices of the mesh are plane curves: $g_{x_1} = g_{x_1}(x_2)$ and $h_{x_2} = h_{x_2}(x_1)$. Each curve g_{x_1}, h_{x_2} is represented by equation $y = y(x)$ and can be determined by sequence of points (x_i, y_i), $i = 1(1)n$, where $i = a(b)c$ denotes increment of variable i from value a with step b to value c inclusively. The more precise and suitable approximation of boundary curves for the assumed model of applied patch, the better representation of the whole complex surface can be obtained.

2.2. Practically used methods of approximation of curves

Among many models of curves, used for interpolation of nonlinear functions, the most popular are Lagrangean and Hermite polynomials, complex curves based on piecewise cubic Hermite interpo-

lation, cubic splines and B-splines, including NURBS, etc. Some of them are more frequently used for conventional interpolation problems, due their simple analytical definition and some are more suitable for interactive surface or solid modelling (splines, NURBS, etc.).

Lagrangean polynomials

Lagrangean interpolating polynomial is defined through the following formula [11],

$$p_L(x) = \sum_{i=0}^{n} L_i(x)\, y_i\,, \tag{1}$$

where

$$L_i(x) = \frac{(x-x_0)(x-x_1)\cdots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_n)}{(x_i-x_0)(x_i-x_1)\cdots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_n)}\,. \tag{2}$$

If n is a big number, then undesired oscillations may occur (i.e., local maximum and local minimum values) at points $x_i \in R$, where $p'_L(x_i) = 0$. This trend will be stronger if the degree of the Lagrangean interpolation polynomial and consequently the number of interpolation points considerably increase [11].

Calculations which are necessary for determination of Lagrangean interpolating polynomials can be simplified by the use of the divided differences method [29, 11]. Interpolation using divided differences has an advantage in comparison with Lagrangean method, because if the necessary preliminary work is done, the evaluation of the interpolating polynomial for a particular value of x is very simple computationally. However, unstability of resulting Lagrangean polynomial is still directly proportional to its degree.

Hermite polynomials

The more precise interpolation (especially over the areas between interpolating points) can be achieved when not only coordinates of points $x_i\,,\ y_i\,,\ i = 1(1)n$, but also first derivatives $y'_i\,,\ i = 1(1)n$ are known and used. This set of data forms suitable basis for designing interpolating polynomials of $2n-1$ degree. The following Hermite's formulas [11] can be used in this case,

$$p_H(x) = \sum_{i=1}^{n} H_i(x)\, y_i + \sum_{i=1}^{n} H_i^*(x)\, y'_i\,, \tag{3}$$

where

$$H_i(x) = [1 - 2L'_i(x_i)(x-x_i)]\, L_i^2(x) \qquad \text{and} \qquad H_i^*(x) = (x-x_i)\, L_i^2(x), \tag{4}$$

$L_i(x)$ – Lagrangean interpolation coefficients.

Hermite polynomials enable more precise interpolation than Lagrangean polynomials, because degree of polynomial is $2n-1$, which is distinctly bigger number than degree of Lagrangean interpolating polynomial $n-1$. Although the coordinates $x_i\,,\ y_i\,,\ i = 1(1)n$ can be measured on a physical model or found from cross-sectional drawings of a 3-D object with high accuracy, but tangent vectors (gradients) could have big relative errors. This could lead to oscillations, which might amplify with an increasing degree of interpolating polynomial. This disadvantage causes that in many cases it is more profitable to use so-called piecewise Hermite interpolation based on low-degree polynomials.

Piecewise cubic Hermite interpolation

In the simplest case, taking into consideration available information about the interpolated points: $\mathbf{P}_i = (x_i\,,\, y_i)$, $\mathbf{P}_{i+1} = (x_{i+1}, y_{i+1})\,,\ i = 1(1)n$ and about the first derivatives: y'_i, y'_{i+1}, represented by tangent vectors $\mathbf{P}'_i$ and $\mathbf{P}'_{i+1}$, parametric cubic polynomial $\mathbf{P}(u)$ can be calculated. This polynomial enables joining the sequence of points $\mathbf{P}_i\,,\ i = 1(1)n$ with consideration of C^2 class of continuity. It is based on interpolating Hermite functions (suitable basis for the composite curves) and can be calculated through the following formulas [11],

$$\mathbf{P}(u) = \lfloor h_{00}(u), h_{01}(u), h_{10}(u), h_{11}(u) \rfloor \cdot \lfloor \mathbf{P}_i\,,\, \mathbf{P}_{i+1}\,,\, \mathbf{P}'_i\,,\, \mathbf{P}'_{i+1} \rfloor^T\,, \tag{5}$$

where

$\mathbf{P}_k(u)$ – vector containing coordinates of points; $k = i, i+1$; $u \in \langle 0, 1\rangle$;

h_{ij} – basic Hermite functions : $h_{00}(u) = 2u^3 - 3u^2 + 1$, $h_{01}(u) = -2u^3 + 3u^2$,

$h_{10}(u) = u^3 - 2u^2 + u$, $h_{11}(u) = u^3 - u^2$.

With this method direction and length of tangent vector $\mathbf{P}'_i$ in each interpolating point $(x_i\,,\, y_i)$ as well as continuity of slope at all data points can be forced and stabilized. Unfortunately, continuity of second derivatives and hence, the continuity of curvature at nodal points can not be ensured. One of the advantages of applying this method for curve and surface modelling is that surfaces based on network of composite cubic interpolating polynomials may be smooth at each point. However, the problem of stable and accurate evaluation of the necessary gradients is still unsolved. Unfortunately, even the successive approximations of the gradients may give only rough estimates of tangent vectors and the whole interpolating procedure will not be robust enough.

Cubic spline interpolation

Smooth interpolation of given data points (which are nodal points to the network of curves laying on considered complex surface) can be done on the basis of low-degree polynomials, reducing undesired side effects with oscillations. Such composite curves, spline curves [7, 11], are most frequently based on cubic polynomials, which allow to keep the continuity of first and second derivatives (continuity of tangent vectors and curvature required with consideration of structural and technological aspects) at all nodal points. The cubic spline (a composite curve of cubic polynomials) $f(x)$, which fitts given data points, can be defined by the sequence of $n+1$ points $(x_i\,,\, y_i)$ (called knots), where $i = 0(1)n$. On each interval $\langle x_{i-1}\,,\, x_i\rangle$ (called span) $f(x)$ is a polynomial of a degree not higher than 3. Furthermore, it is assumed that $f(x_i) = y_i$ and first derivative $f'(x_i)$ as well as the second derivative $f''(x_i)$ are continuous at all nodes x_i, where $i = 1(1)n-1$. Briefly, this curve has minimal mean squared curvature. In geometrical sense this is the smoothest curve of all possible curves that pass through the given sequence of fixed points [7]. For completeness of interpolating spline definition some additional information is necessary to make the appropriate system of equations solvable. It depends on different (mostly geometrical) considerations of specific application: (i) built-in ends with a specified first derivatives $f'(x_0) = f'_0$, $f'(x_n) = f'_n$; (ii) free ends – no curvature of the spline at x_0 and x_n, i.e.: $f''(x_0) = f''(x_n) = 0$; (iii) defined curvature at the ends: $f''(x_0) = f''_0$, $f''(x_n) = f''_n$; (iv) continuity of third degree derivative at the end spans (thus, two successive interpolated spans at each end of the spline are the same cubic functions). Splines of degree higher than 3 are used rarely, when derivatives of the third or higher orders are also required to be continuous at each node.

Although the splines have great advantages with continuous curvature, they have also some limitations [7, 11]: (i) problems with stability of computations arise when vertical tangents or near vertical tangents occur; (ii) numerical procedure may fail or lead to undesired oscillations in the case of interpolation of curve with discontinuity in second derivative; (iii) local modification of given data is associated with re-computations of the entire splines. First two problems can be alleviated by local redefinition of axes. The third one can be avoided by the use of fundamental splines (B-splines).

B-spline interpolation

B-splines have been developed [7, 10, 36] and implemented for a number of geometrical modellers in Computer Aided Engineering (CAE) systems, especially for CAD/CAM systems used in automotive engineering and in aviation engineering. B-spline curves and B-spline surfaces are widely used because they exhibit the following advantages: (i) they allow proper description of the curve based on relatively small amount of input data, (ii) each B-spline curve lies within the control polygon (control net) defined by control points (the convex hull property), (iii) displacement of a single control point influences the modelled curve only locally (this is due to local non-zero values of the blending functions defining B-spline), (iv) B-spline representation of curve or surface is invariant under an affine map, which is necessary for modelling of 3D objects, (v) B-spline representation has variation diminishing property.

Parametric representation of a functional B-spline curve is defined in terms of its blending functions by the following formula [7, 10, 36],

$$\mathbf{f}(t) = \lfloor f(t)_x\,,\, f(t)_y\,,\, f(t)_z \rfloor^T = \sum_{i=1}^{n} \lfloor a_{i,x}\,,\, a_{i,y}\,,\, a_{i,z} \rfloor^T N_{i,k}(t), \tag{6}$$

where

$\mathbf{a}_i = \lfloor a_{i,x}\,,\, a_{i,y}\,,\, a_{i,z} \rfloor^T$ – control points defining B-spline control polygon (control coefficients),

$N_{i,k}(t)$ – related blending functions of degree $k-1$,

$\lfloor\,\rfloor$ – row vector, $\lfloor\,\rfloor^T$ – transposition of vector $\lfloor\,\rfloor$.

According to its definition, a spline is of order k or degree $k-1$, when it is defined as a piecewise $k-1$ degree polynomial that is C^{k-2} continuous. This also means that: (i) the degree of the polynomial does not exceed $k-1$ inside each $\langle t_i\,,\, t_{i+1} \rangle$ interval, (ii) the first $k-2$ derivatives of considered spline are continuous.

One way for the calculation of blending functions $N_{i,k}(t)$ is to use the Mansfield–Cox–de Boor formula. Mathematical representation of this recursive formula is [7, 10]

$$N_{i,1}(t) = \begin{cases} 1, & t_i \le t < t_{i+1}\,, \\ 0, & \text{otherwise}, \end{cases} \tag{7}$$

$$N_{i,k}(t) = \frac{t - t_i}{t_{i+k-1} - t_i} N_{i,k-1}(t) + \frac{t_{i+k} - t}{t_{i+k} - t_{i+1}} N_{i+1,k-1}(t), \qquad k > 1. \tag{8}$$

The most commonly used B-splines are of order k=4. They are formed of cubic B-splines with C^2 continuity at the nodal points. B-splines could be of non-rational or rational type. The rational B-splines are more general than non-rational ones and are defined in terms of rational functions. If the elements of knot vector $\lfloor t_i\,,\, \ldots\,,\, t_{i+k} \rfloor$ are spaced equally, the B-spline is called a uniform B-spline. This equal spacing between the knots is easy to use for further calculations, but in the case of highly unevenly spaced data points undesirable oscillations may occur when using uniform B-spline representation. In such case, in order to better control the shape in the design process, a non-uniform knot spacing should be used and the Non-Uniform Rational B-Splines (NURBS) should be generated, based on input data [10]

$$\mathbf{f}(t) = \frac{\sum\limits_i w_i\, \mathbf{a}_i\, N_i^k(t)}{\sum\limits_j w_j\, N_j^k(t)} \tag{9}$$

where

w_i – weights, $w_1 > 0$ and $w_n > 0$, $w_j \ge 0$ for $j = 2(1)n-1$,

$\mathbf{a}_i$ – vector of nodes of control polygon,

$N_{ik}(t)$ – B-spline basis functions of degree k.

Weights w_i introduced in the above equation play the role of additional shape parameters. They affect the modelled curve or surface only locally. As the weight w_i increases, the curve is pulled towards the control point a_i. When w_i decreases, the curve is pushed away from the control point a_i. Therefore, modification of the modelled curve or surface can be done much more precisely, than in the case of non-rational approach. Although free form curve or surface modelling with the use of NURBS gives a lot of flexibility and has advantages in interactive geometric modelling, calculation of suitable knots and weights to obtain the curve or surface passing through a given set of points requires certain computational effort. In addition to this, it is very important to select correctly the order of NURBS for particular set of given data points. The higher the order is, the smoother the modelled curve or surface could be, but at the same time the distance between interpolated points and original measured data points will be greater [6].

The methods of curve modelling, presented above, are useful for interpolation of given data points. However, in the case of measurement data, which by their nature contain statistical or other fluctuations, the above procedures might amplify them. Therefore, rather than using interpolation techniques, certain estimation methods with inherent smoothing properties can be used for creating numerical representation of complex boundary surface. In order to diminish or even to omit undesired oscillations, optimal regularization of input data in combination with estimation techniques can be applied.

Estimation and smoothing of data

Design of interpolating curve, which accurately passes through all the given points (x_i, y_i) is reasonable, if the coordinates of points are reliable. However, when the points (x_i, y_i) represent some experimental data, resulting from non-geometrical calculations, the methods of approximation of given set of points by designing the curve passing through all the data points could lead to undesired side effects. When given points contain statistical or other fluctuations, the procedures that were described above might lead to significant amplification of them.

The most frequently used methods of estimation are the Least Squares (LS) and the Maximum Probability (MP) methods, which are very effective for wide range of problems. Therefore, they are also well known. It should be noticed that in standard conditions, the MP estimators are asymptotically unbiased, normally distributed and effective. At the same time, the MP method requires, before the estimation, all the information about distribution of measured parameters, which is usually not available. Therefore, in many cases, normal distributions of measurements are assumed. Then, the MP method converts to the LS method and MP estimators become LS estimators [9, 33]. It can be proved [22, 33] that LS estimators are the best estimators in the whole class of linear unbiased estimators. The advantage of the LS method lies in its simplicity of practical realization and in possibility of application of corresponding recurrent algorithms, which are particularly useful in the cases of on-line estimation of mathematical models of processes changing in time (e.g., estimation during measurements). On the other hand, the LS method is very sensitive to significant noise in experimental data, especially when the data contain unnormal random errors (positive or negative peaks). It should be noticed that experimental data can contain up to 10%–20% of such peaks. Then, corresponding curves, obtained by the LS method, are significantly shifted in direction to the a.n. peaks. Therefore, the LS method should be applied in combination with the high quality digital filters, which reduce related influence of the peaks on final results of estimation [2, 15]. Unfortunately, these filters are often not as effective as necessary in smoothing or throwing away the a.n. undesired peaks from the experimental data.

The necessity of improvement of stability of parametric estimation caused development of the *robust estimation* methods [9, 13]. The most frequently used robust estimation methods are methods related to the quasi-probable estimators (M-estimators), proposed by P.J. Huber [13], linear combinations of the succeeding statistics (L-estimators) and the estimators based on ordering techniques (R-estimators). For example, Huber [13] proposed the following method of determination of stable estimators of parameter θ. Common probability density function $\varphi(\theta, x)$ of variables θ and x, where

x_i denotes measurements, belongs to the certain class of functions. The principle idea lies in finding such a function $\psi_\theta = \partial\varphi(\theta, x)/\partial\theta$, describing the method of estimation, that asymptotic dispersion $D(\psi_\theta, \varphi)$ does not exceed some minimal value for any arbitrary function φ, i.e.

$$D(\psi_\theta, \varphi) \leq D(\psi_\theta\,, \varphi_\theta), \qquad \text{where} \quad D(\psi_\theta\,, \varphi_\theta) = \min_{\psi} \max_{\varphi} D(\psi, \varphi). \tag{10}$$

This general approach to robust estimation of mathematical models was successfully used for wide range of regression problems [9, 13, 27], including time series analysis, signal processing, filtering of measurement data, etc. Consequently, M-estimators of regression, denoted $\hat{\beta}_M$, can be found as a solution of the following optimization,

$$\min_{\beta_i} \sum_{i=1}^{m} \rho\left(y_i - \sum_{j=1}^{n} \beta_j x_{ij}\right) = \min_{\beta_1,\ldots,\beta_n} \sum_{i=1}^{m} \rho(e_i), \tag{11}$$

where $\rho(e_i) = y_i - \sum_{j=1}^{n} \beta_j x_{ij}$ – residuals, and $\rho(\eta)$ – chosen function. Properties of estimators $\hat{\beta}_M$ depend on function $\rho(\eta)$. In order to fulfil the requirements related to bias of estimators, effectiveness of estimation and the other, additional conditions are considered. The most important is an increment of density function $\rho(\eta)$ sustained smaller than quadratic function η^2 (for $\rho(\eta) = \eta^2$ this method becomes LS method). Depending on the function $\rho(\eta)$ different M-estimators can be obtained. Conditions minimum of the a.n. sum lead to equations expressing that each partial derivative $\partial(\cdot)/\partial\beta_i$ becomes zero, i.e.

$$\sum_{i=1}^{m} \psi\left(y_i - \sum_{j=1}^{n} \beta_j x_{ij}\right) x_{ij} = 0. \tag{12}$$

Main features of the a.n. estimators and the other robust estimators lie in small bias in estimate and in good effectiveness. In addition, they are only slightly sensitive to the change of measured parameters' distributions. The estimators, which are effective for the measured data, characterized by distributions with "heavy ends", are also stable in relation to the measured data with significant occasional random errors. However, in the case of normal distributions of measured data, robust estimators are less effective, than the methods originally oriented to data distributions with "light ends".

With the development of computer aided measurement systems the need to use more versatile methods of estimation increases. An original method of calculations of M-estimators was proposed by W.I. Mudrov and W.A. Kuschko [28] – method of variationally-weighted squared approximations. This method is a good generalization of both standard LS and minimum moduli methods. Based on repeated application of the weighted LS with changing weights, this method allows to obtain significantly probable estimators. This fact confirms the role of the LS method in estimation of mathematical models in the case of standard, normal distributions, as well as in the case of distributions with "heavy ends".

Application of digital filtering and smoothing of measured data partially diminishes the a.n. effects caused by random errors. At the same time, both filtering and smoothing significantly increase collinearities between independent variables (input signals) in regression models. As a result, simultaneously increase both multicollinearity of measurement matrix and multicollinearity of the *Fisher information matrix* (FIM), which becomes in such case ill-conditioned. Application of standard estimation methods to such ill-conditioned data usually leads to unstable results.

In order to diminish influence of multicollinearity on stability of estimation, grouping of different independent parameters is usually being done. Application of such clustering techniques allows to estimate only certain sums of regressors. This is very often not sufficient to create corresponding mathematical model of a surface. Necessity of separation of regressors is specially significant for obtaining proper results when using CNC's or CMM's. Such system must be created on the basis of

correct mathematical model adequately describing real object, which should be measured, modelled or controlled.

Generally, inversion of ill-conditioned FIM is an *incorrect problem*. Effective methods of solution of the incorrect problems were created by A.N. Tikchonov [37] and his co-workers. These so-called *regularization algorithms* make use of certain a priori information about smoothness, monotony and convexity of the solution being sought for. Regularization algorithms, applied to ill-conditioned data, allow to obtain biased estimators, which are more stable than standard LS estimators. However, it should be noticed that in every particular case so-called *regularization parameters* should be chosen individually.

Other methods of stable estimation also have some disadvantages. For example, in the case of *ridge regression estimators* it was not proved, in which way *ridge parameter* should be chosen. For this purpose, so-called trace of the ridge can be analyzed in the area surrounding this value of the ridge trace, for which the ridge regression estimator changes only very slightly. In the case of attracting estimators, optimal value of attracting parameter can be determined in explicit way (James–Stain estimator). However, existing methods of selection of the regularization parameter make the corresponding estimators nonlinear. The a.n. disadvantages limit the use of regularization methods in practice of estimation of mathematical models.

The most believable data and recommendations for proper selection of tooth flank modification and for proper gear design and manufacturing, in general, follow from experimental and operational investigations of real gear transmissions, performed in real working conditions. However, these investigations can be done only after a prototype of gear transmission is made. In addition to this, they are time-consuming and expensive. Therefore, further development in surface modelling of real tooth flanks, combined with finite element analysis, leading to proper TCA of gear transmission under load, is still an up-to-date problem.

In this work analytical and numerical representation of surfaces applicable for modelling of tooth flanks is presented. It is useful for determination of tooth contact area in gear transmission with modified tooth flanks. The advantage of the presented method over the methods given in the papers cited above lies in entirely discrete representation of tooth flanks and puts special attention to the situations, when measurement data is ill-conditioned. It can be used at the design stage of gear transmission and can be helpful in estimation of tooth flanks model, and furthermore – in proper selection of the type and parameters of tooth flank modification defined by discrete data.

3. Problems leading to ill-conditioned matrices

One of the most frequent reasons for generation of ill-conditioned matrices is interpolation with high degree polynomials. Let us assume sequence of data points $(x_i\,,\,y_i)$, $i = 0(1)n$, which should be approximated by the polynomial of n degree, passing through all of them,

$$\phi(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1}x^{n-1} + a_n x^n\,. \tag{13}$$

The following equation must be solved,

$$\mathbf{XA} = \mathbf{Y}, \tag{14}$$

where:

$\mathbf{X}$ – Vandermonde's matrix containing information about x_i^j, $\mathbf{X} \in \mathbf{M}_{(n+1)\times(n+1)}$,

$$\mathbf{X} = \begin{bmatrix} 1 & x_0 & x_0^2 & \cdot & x_0^n \\ 1 & x_1 & x_1^2 & \cdot & x_1^n \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_n & x_n^2 & \cdot & x_n^n \end{bmatrix},$$

$\mathbf{A} \in \mathbf{M}^{n+1}$ and $\mathbf{A} = \lfloor a_0\,,\,a_1\,,\,\ldots\,,\,a_n \rfloor^T$ – vector of unknown coefficients,

$\mathbf{Y} \in \mathbf{M}^{n+1}$ and $\mathbf{Y} = \lfloor y_0, y_1, \ldots, y_n \rfloor^T$ – vector of known values y_i,

$\mathbf{M}_{m\times n}$ – set of matrices with m rows and n columns,

$\mathbf{M}^m$ – set of vectors with dimension m.

In the case of a big number of points and correspondingly – high degree of polynomial $\phi(x)$, matrix $\mathbf{X}$ becomes ill-conditioned. The way to omit influence of the data fluctuations on the result is to design an approximating function, passing close to all the given points and estimate the polynomial coefficients a_i, $i = 0(1)n$ from the Least Squares Method (LSM) [9, 22, 33],

$$\mathbf{A} = \mathbf{F}^{-1}\mathbf{X}^T\mathbf{Y}, \tag{15}$$

where $\mathbf{F} = \mathbf{X}^T\mathbf{X}$ – Fisher information matrix, $\mathbf{F} \in \mathbf{M}_{(n+1)\times(n+1)}$.

In the case of a high degree of approximating polynomial, some of the columns or rows of the matrix $\mathbf{F}$ are almost parallel vectors. This high correlation causes so-called multicollinearity of the matrix $\mathbf{F}$, which is then ill-conditioned.

4. An influence of errors on the solution of matrix equation

Assuming that nonsingular matrix $\mathbf{X} \in \mathbf{M}_{m\times m}$ has elements distorted by $\delta\mathbf{X} \in \mathbf{M}_{m\times m}$ and vector of known values $\mathbf{Y} \in \mathbf{M}^m$ has strict information, then instead of solving $\mathbf{X}\,\mathbf{A} = \mathbf{Y}$ for vector of unknowns $\mathbf{A} \in \mathbf{M}^m$, the following equation must be solved,

$$(\mathbf{X} + \delta\mathbf{X})(\mathbf{A} + \delta\mathbf{A}) = \mathbf{Y}. \tag{16}$$

With Euclid norm of the matrix $\mathbf{X}$ and vector $\mathbf{A}$, respectively,

$$\|\mathbf{X}\| = \max_{\mathbf{A}\neq\mathbf{0}} \frac{\|\mathbf{X}\,\mathbf{A}\|}{\|\mathbf{A}\|}, \qquad \|\mathbf{A}\| = \sqrt{\sum_{k=1}^{n} |a_k|^2}, \tag{17}$$

and considering condition number of matrix $\mathbf{X}$ defined as [38]

$$\mathrm{cond}(\mathbf{X}) = \|\mathbf{X}\| \cdot \|\mathbf{X}^{-1}\|,$$

the following inequality can be obtained,

$$\frac{\|\delta\mathbf{A}\|}{\|\mathbf{A}\|} \leq \mathrm{cond}(\mathbf{X} + \delta\mathbf{X}) \frac{\|\delta\mathbf{X}\|}{\|\mathbf{X} + \delta\mathbf{X}\|}. \tag{18}$$

If vector $\mathbf{Y}$ is also distorted by $\delta\mathbf{Y} \in \mathbf{M}^m$, then Eq. (16) becomes

$$(\mathbf{X} + \delta\mathbf{X})\,(\mathbf{A} + \delta\mathbf{A}) = \mathbf{Y} + \delta\mathbf{Y}. \tag{19}$$

Equation (19) yields

$$\delta\mathbf{A} = \delta\mathbf{A}_1 + \delta\mathbf{A}_2, \tag{20}$$

where

$$\delta\mathbf{A}_1 = (\mathbf{X} + \delta\mathbf{X})^{-1}\,\mathbf{Y} - \mathbf{A} \qquad \text{and} \qquad \delta\mathbf{A}_2 = (\mathbf{X} + \delta\mathbf{X})^{-1}\,\delta\mathbf{Y}.$$

Euclid norm of vector $\delta\mathbf{A}_1$ could be limited by inequality analogous to Ineq. (18). Taking into consideration that $\|\mathbf{X} + \delta\mathbf{X}\| \geq \|\mathbf{X}\|$ and $\|\mathbf{X}\| \cdot \|\mathbf{A}\| = \|\mathbf{X}\,\mathbf{A}\|$ it can be shown that

$$\frac{\|\delta\mathbf{A}_2\|}{\|\mathbf{A}\|} \leq \mathrm{cond}(\mathbf{X} + \delta\mathbf{X}) \frac{\|\delta\mathbf{Y}\|}{\|\mathbf{Y}\|}. \tag{21}$$

From Ineqs. (18) and (21) the following result can be obtained:

$$\frac{\|\delta \mathbf{A}\|}{\|\mathbf{A}\|} \leq \text{cond}(\mathbf{X}+\delta\mathbf{X}) \left(\frac{\|\delta \mathbf{X}\|}{\|\mathbf{X}+\delta\mathbf{X}\|} + \frac{\|\delta \mathbf{Y}\|}{\|\mathbf{Y}\|} \right). \tag{22}$$

For real symmetrical matrix $\mathbf{F} \in \mathbf{M}_{n\times n}$ its condition number can be estimated as [44]

$$\text{cond}(\mathbf{F}) = \frac{\lambda_{\max}[\mathbf{F}]}{\lambda_{\min}[\mathbf{F}]} \tag{23}$$

where λ_i, $i = 1(1)n$ – eigenvalues of matrix $\mathbf{F}$.

From Ineq. (22) it follows that condition number (by definition – not less than 1) of matrix which has to be inverted plays the role of multiplicator for relative errors of computed coefficients of mathematical model representing boundary surface e.g., tooth flank surface (measured surface or surface generated in numerical way with consideration to intentionally introduced modifications of the surface). In addition to this, from Eq. (23) it follows that the most significant factor influencing the distortion of estimated coefficients is the minimal eigenvalue $\lambda_{\max}$ of the considered matrix.

5. An analysis of multicollinearity

For better understanding the problem of multicollinearity, related to high condition number of an arbitrary matrix, which must be inverted, it can be assumed that two vectors (columns $\mathbf{x}_i$, $\mathbf{x}_j$, $i \neq j$, of measurement matrix $\mathbf{X}$) are almost parallel. Then minimal eigenvalue $\lambda_{\min}$ of Fisher matrix $\mathbf{X}^T\mathbf{X}$ is very small.

By definition,

$$\mathbf{X}^T\mathbf{X}\,\mathbf{V}_{\min} = \lambda_{\min}\,\mathbf{V}_{\min}\,, \tag{24}$$

where $\mathbf{V}_{\min}$ – eigenvector associated with $\lambda_{\min}$; $\|\mathbf{V}\|^2 = \mathbf{V}^T\mathbf{V} = 1$.

For very small $\lambda_{\min}$ from Eq. (24) we can obtain $\|\mathbf{X}\,\mathbf{V}\|^2 \simeq 0$, i.e. $\mathbf{X}\,\mathbf{V} \simeq \mathbf{0}$, which could be rewritten in the following form,

$$\sum_{i=1}^{n} \mathbf{X}_i\, v_i \simeq \mathbf{0}, \tag{25}$$

where $\mathbf{X}_i$ – i-th column of matrix $\mathbf{X}$, $\mathbf{X}_i \in \mathbf{M}^m$, and v_i – i-th element of eigenvector $\mathbf{V}_{\min}$ of the matrix $\mathbf{X}^T\mathbf{X}$.

Regarding to the formula (25) for $n = 2$ and considering that vectors $\mathbf{X}_1$, $\mathbf{X}_2$ are normalized, i.e. $\|\mathbf{X}_1\|^2 = 1$, $\|\mathbf{X}_2\|^2 = 1$, we can get

$$\cos(\mathbf{X}_1, \mathbf{X}_2) \simeq -\frac{v_1}{v_2}\,. \tag{26}$$

From Eq. (26) it follows that elements of eigenvector $\mathbf{V}_{\min}$ associated with the minimal eigenvalue $\lambda_{\min} \simeq 0$ of Fisher matrix $\mathbf{X}^T\mathbf{X}$ play the role of weights, informing about the rate of multicollinearity of columns $\mathbf{X}_i$.

6. Stable parametric estimation of mathematical model of boundary curve from the mesh

6.1. Regularization of matrix containing coordinates of the nodes

Let us analyze a set of points (x_i, y_i), $i = 1(1)m$, which are measured on a cross-section of a physical model and represent a curve on the surface – one of the curves defining the mesh, filled by surface patches. Estimation of the curve starts from the equation

$$\mathbf{Y} = \mathbf{X}\,\mathbf{A} + \mathbf{R} \tag{27}$$

where

$\mathbf{Y}$ – vector of measured coordinates y_i, $i = 1(1)m$, $\mathbf{Y} \in \mathbf{M}^m$,

$\mathbf{X}$ – matrix of measured or calculated coordinates x_i^j (measurement matrix), $\mathbf{X} \in \mathbf{M}_{m\times n}$,

$\mathbf{A}$ – vector of unknown coefficients, $\mathbf{A} \in \mathbf{M}^n$,

$\mathbf{R}$ – vector of measurement errors, $\mathbf{R} \in \mathbf{M}^m$.

We assume that $E[\mathbf{R}] = \mathbf{0}$ and $D[\mathbf{R}] = \sigma \mathbf{I}_n$ where $E[\cdot]$ – mathematical expectation, $D[\cdot]$ – mathematical variance, $\mathbf{I}_n \in \mathbf{M}_{n\times n}$ and $\mathbf{I}_n = \mathrm{diag}(1, 1, \ldots, 1)$ – unit matrix; $\sigma \geq 0$.

In order to improve the stability of estimation, we invent regularized matrix $\bar{\mathbf{X}} = \mathbf{X} + \mathbf{B}$, where $\mathbf{B} \in \mathbf{M}_{m\times n}$, which will be used for estimation instead of measurement matrix $\mathbf{X}$. Additional matrix $\mathbf{B}$, designed in a way to diminish multicollinearities, has the following structure,

$$\mathbf{B} = e\,\mathbf{Q}\,\mathbf{S}^T, \tag{28}$$

where

e – small real number, related to the accuracy of measurements and accuracy of computation on particular computer;

$\mathbf{Q}$ – vector of coefficients of regularization, $\mathbf{Q} \in \mathbf{M}^m$;

$\mathbf{S}$ – vector, elements of which are proportional to the collinearity of regressors, $\mathbf{S} \in \mathbf{M}^n$.

Based on Eq. (28) and on LS approach, the vector of estimators $\hat{\mathbf{A}}_B$ can be calculated from the following formula [14],

$$\hat{\mathbf{A}}_B = \left(\bar{\mathbf{X}}^T \bar{\mathbf{X}}\right)^{-1} \bar{\mathbf{X}}^T \mathbf{Y}. \tag{29}$$

Optimal values for regularization matrix $\mathbf{B}$ could be found as a result of minimization of the spherical norm of matrix $\mathbf{B}$ with condition put on the minimal eigenvalue of the Fisher matrix. Therefore, considering $\mathrm{tr}[\mathbf{B}^T\mathbf{B}]$ – trace of matrix $\mathbf{B}^T\mathbf{B}$ – for a spherical norm $\|\mathbf{B}\|^2$ of matrix $\mathbf{B}$, we formulate the following functional,

$$J = \mathrm{tr}\left[\mathbf{B}^T\mathbf{B}\right] - \mu\left\{\bar{\lambda}_{\min}\left[\bar{\mathbf{X}}^T\bar{\mathbf{X}}\right] - \lambda^*\right\}, \tag{30}$$

where

$\mathrm{tr}[\mathbf{F}] = \sum_{i=1}^{n} f_{ii}$ – trace of matrix $\mathbf{F}$, $\mathbf{F} \in \mathbf{M}_{n\times n}$,

μ – Lagrangean multiplicator,

$\bar{\lambda}_{\min}[\bar{\mathbf{X}}^T\bar{\mathbf{X}}]$ – minimal eigenvalue of matrix $\bar{\mathbf{X}}^T\bar{\mathbf{X}}$,

$\lambda^* = \lambda_{\min}[\mathbf{X}^T\mathbf{X}] + \gamma$, $\gamma > 0$,

λ^* – assumed minimal eigenvalue of regularized Fisher matrix.

Matrix $\mathbf{B}$ can be written in the following form

$$\mathbf{B} = e \cdot \mathbf{B}',$$

where e still denotes a small real number.

Considering the theorem of perturbation of eigenvalue [21, 44], we can obtain

$$\bar{\lambda}_{\min}\left[\bar{\mathbf{X}}^T\bar{\mathbf{X}}\right] = \lambda_{\min}\left[\mathbf{X}^T\mathbf{X}\right] + e\,\lambda_{\min}\left[\mathbf{X}^T\mathbf{B}' + \mathbf{B}'^T\mathbf{X}\right] + \mathbf{0}(e^2). \tag{31}$$

From Eq. (31) and from Eqs. (28), (30) we get

$$J = \mathrm{tr}\left[e^2\left(\mathbf{Q}\mathbf{S}^T\right)^T\left(\mathbf{Q}\mathbf{S}^T\right)\right] - \mu\left[\mathbf{V}_{\min}^T\left(\mathbf{X}^T e\,\mathbf{Q}\mathbf{S}^T + \mathbf{S}\mathbf{Q}^T e\,\mathbf{X}\right)\mathbf{V}_{\min} - \gamma\right]. \tag{32}$$

Equation (32) can be rewritten in the following form,

$$J = e^2 \|\mathbf{Q}\|^2 \|\mathbf{S}\|^2 - \mu \left(e\, 2\, \mathbf{V}_{\min}^T \mathbf{X}^T \mathbf{Q} \mathbf{S}^T \mathbf{V}_{\min} - \gamma \right), \tag{33}$$

where: $\mathbf{V}_{\min}$ – eigenvector related to minimal eigenvalue $\lambda_{\min}[\mathbf{X}^T\mathbf{X}]$, $\mathbf{V}_{\min} \in \mathbf{M}^n$.

It could be easily shown that minimum of functional J from Eq. (33) can be achieved for the following values,

$$\mathbf{Q} = \frac{\mu \mathbf{S}^T \mathbf{V}_{\min} \mathbf{X} \mathbf{V}_{\min}}{e \|\mathbf{S}\|^2}, \tag{34}$$

$$\mathbf{S} = \frac{\mu \mathbf{V}_{\min}^T \mathbf{X}^T \mathbf{Q} \mathbf{V}_{\min}}{e \|\mathbf{Q}\|^2}, \tag{35}$$

$$\gamma = 2e\, \mathbf{V}_{\min}^T \mathbf{X}^T \mathbf{Q} \mathbf{S}^T \mathbf{V}_{\min}. \tag{36}$$

From Eqs. (34), (35) we can obtain

$$\mathbf{Q} = \mathbf{X} \mathbf{V}_{\min}. \tag{37}$$

Combining Eqs. (37), (35) and (36) yields

$$\mu = \frac{\gamma}{2\lambda_{\min}[\mathbf{X}^T\mathbf{X}]} \quad \text{and} \quad \mathbf{S} = \frac{\gamma \mathbf{V}_{\min}}{2e\lambda_{\min}[\mathbf{X}^T\mathbf{X}]}. \tag{38}$$

Then, considering Eqs. (28), (37) and (38), we arrive at

$$\mathbf{B} = \frac{\left\{ \lambda^* - \lambda_{\min}\left[\mathbf{X}^T\mathbf{X}\right] \right\} \mathbf{X} \mathbf{V}_{\min} \mathbf{V}_{\min}^T}{2\lambda_{\min}\left[\mathbf{X}^T\mathbf{X}\right]}. \tag{39}$$

In the case when L-iterated minimal eigenvalue of the Fisher information matrix is a very small number, then optimal matrix of regularization can be rewritten in the following form,

$$\mathbf{B}_L = \sum_{i=1}^{L} \frac{\lambda^* - \lambda_i}{2\lambda_i} \mathbf{X} \mathbf{V}_i \mathbf{V}_i^T, \tag{40}$$

where $\lambda_i = \lambda_i[\mathbf{X}^T\mathbf{X}]$.

Spherical norm of regularization matrix $\mathbf{B}$ can be found as

$$\|\mathbf{B}\|^2 = \frac{\operatorname{tr}\left[(\lambda^* - \lambda_{\min})^2 \mathbf{V}_{\min} \mathbf{V}_{\min}^T \lambda_{\min} \right]}{4\lambda_{\min}^2} \tag{41}$$

where $\lambda_{\min} \equiv \lambda_{\min}[\mathbf{X}^T\mathbf{X}]$.

Taking into consideration that $\lambda_{\max}[\mathbf{B}^T\mathbf{B}] \leq \operatorname{tr}[\mathbf{B}^T\mathbf{B}]$, we can obtain

$$\lambda_{\max}\left[\mathbf{B}^T\mathbf{B}\right] \leq \frac{(\lambda^* - \lambda_{\min})^2}{4\lambda_{\min}}. \tag{42}$$

From the theorem [31] limiting the trace of matrix $\mathbf{D}^T\mathbf{D}$, where $\mathbf{D} \in \mathbf{M}_{n\times n}$, i.e. from

$$\operatorname{tr}\left[\mathbf{D}^T\mathbf{D}\right] \geq \sum_{i=1}^{n} \lambda_i^2 \tag{43}$$

and from the known inequality

$$\sum_{i=1}^{n} \lambda_i^2 \geq \lambda_{\max}^2 \tag{44}$$

we can get

$$\lambda_{\max}\left[2\mathbf{X}^T\mathbf{B}\right] \leq \frac{(\lambda^* - \lambda_{\min})\,\mathrm{tr}\left[\mathbf{V}_1\,\mathbf{V}_1^T\,\mathbf{X}^T\,\mathbf{X}\right]}{\lambda_{\min}}. \tag{45}$$

Considering that

$$(\mathbf{a} \in \mathbf{M}^n \wedge \mathbf{b} \in \mathbf{M}^n) \Rightarrow \mathrm{tr}\left[\mathbf{a}\,\mathbf{b}^T\right] = \mathbf{a}^T\,\mathbf{b},$$

we obtain from Ineq. (45) the following inequality,

$$\lambda_{\max}\left[\mathbf{X}^T\mathbf{B} + \mathbf{B}^T\mathbf{X}\right] \leq \lambda^* - \lambda_{\min}\left[\mathbf{X}^T\mathbf{X}\right]. \tag{46}$$

The maximal eigenvalue of the matrix $[\mathbf{X}^T\mathbf{X}]$ from Eq. (29) can be limited by the following inequality,

$$\bar{\lambda}_{\max}\left[\bar{\mathbf{X}}^T\bar{\mathbf{X}}\right] \leq \lambda_{\max}\left[\mathbf{X}^T\mathbf{X}\right] + \lambda_{\max}\left[\mathbf{B}^T\mathbf{B}\right] + \lambda_{\max}\left[\mathbf{X}^T\mathbf{B} + \mathbf{B}^T\mathbf{X}\right]. \tag{47}$$

Therefore, from Eqs. (47), (42), (45) and (46) we can obtain

$$\bar{\lambda}_{\max}\left[\bar{\mathbf{X}}^T\bar{\mathbf{X}}\right] \leq \lambda_{\max}\left[\mathbf{X}^T\mathbf{X}\right] + \lambda_{\min}\left[\frac{\lambda^* - \lambda_{\min}}{\lambda_{\min}} + \left(\frac{\lambda^* - \lambda_{\min}}{2\lambda_{\min}}\right)^2\right]. \tag{48}$$

From Ineq. (48) it can be concluded that

$$\bar{\lambda}_{\max}\left[\bar{\mathbf{X}}^T\bar{\mathbf{X}}\right] \simeq \lambda_{\max}\left[\mathbf{X}^T\mathbf{X}\right].$$

At the same time, minimal eigenvalue $\bar{\lambda}_{\min}\left[\bar{\mathbf{X}}^T\bar{\mathbf{X}}\right]$ could be considerably increased (in relation to it's initial value $\lambda_{\min}\left[\mathbf{X}^T\mathbf{X}\right]$).

The eigenvalue λ^*, which plays the role of control parameter for minimal eigenvalue of the FIM in Eqs. (39), (40), has to be chosen with regard to the following requirements,

$$\lambda^* = \arg \min_{\lambda^* \in \langle \lambda_{\min}, \lambda_{\max} \rangle} \mathrm{cond}\left\{\left[\mathbf{X} + \mathbf{B}(\lambda^*)\right]^T\left[\mathbf{X} + \mathbf{B}(\lambda^*)\right]\right\}. \tag{49}$$

6.2. Limitations to the norm of difference between biased and unbiased LS estimators of polynomials used for surface modelling

In order to find the limit to the norm of difference between biased and unbiased LS estimators of polynomials used for surface modelling it will be assumed that estimators $\hat{\mathbf{A}}$ are computed for initial measurement matrix $\mathbf{X}$ and estimators $\hat{\mathbf{A}}_B$ are computed for regularized measurement matrix $\bar{\mathbf{X}}$ in accordance to the following formulas,

$$\mathbf{A} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y} \quad \text{and} \quad \hat{\mathbf{A}}_B = \left[(\mathbf{X} + \mathbf{B})^T(\mathbf{X} + \mathbf{B})\right]^{-1}(\mathbf{X} + \mathbf{B})^T\,\mathbf{Y}. \tag{50}$$

Denoting $\mathbf{D} = \mathbf{A} - \hat{\mathbf{A}}_B$ and assuming the upper limit for the norm of matrix $\mathbf{A}$ as $\|\mathbf{A}\|^2 \leq r^2$, we can come to the following inequality, limiting the bias of estimators,

$$\|\mathbf{D}\|^2 \leq \frac{\mathrm{tr}\left[\mathbf{H}\right] e^2 \lambda_{\min}^2\left[\mathbf{X}^T\mathbf{X}\right]\left(1 + e^2\right) r^2}{\lambda_{\min}^2\left[\mathbf{H}\right]\left[\det(\mathbf{H})\right]^{\frac{1}{n}}}, \tag{51}$$

where $\mathbf{H} = (\mathbf{X} + \mathbf{B})^T(\mathbf{X} + \mathbf{B})$ and $\mathbf{H} \in \mathbf{M}_{n \times n}$.

From Ineq. (51) it follows that the bias of regularized estimators has upper limit depending, first of all, on the trace and determinant of regularized FIM as well as on the norm of the vector of estimators.

6.3. Numerical example of robust estimation with regularized matrix

Geometrical data obtained from measurements of gear tooth flanks can be interpreted as signals, representing (in three coordinates) the shape of considered tooth flank. Additional regularization matrix $\mathbf{B}$ can be determined on the basis of measurements made on representative piece of tooth flank in direction, in which the whole measurement procedure should be performed.

First, measurement matrix $\mathbf{X}$ and Fisher information matrix $\mathbf{X}^T\mathbf{X}$ with its minimal eigenvalue $\lambda_{\min}[\mathbf{X}^T\mathbf{X}]$ and corresponding eigenvector $\mathbf{V}_{\min}$ should be determined. Then, based on Eqs. (39) or (40), optimal regularization matrix $\mathbf{B}$, for single minimal eigenvalue $\lambda_{\min}[\mathbf{X}^T\mathbf{X}]$ or $\mathbf{B}_L$ for L-iterated minimal eigenvalue, should be computed, respectively. Assumed minimal eigenvalue λ^*, necessary for computation of matrix $\mathbf{B}$ or matrix $\mathbf{B}_L$, should be chosen in accordance with the requirement given by Eq. (49). Columns of matrix $\mathbf{B}$ can be interpreted as supplementary, regularizing noise, added to corresponding columns of measurement matrix $\mathbf{X}$.

For the case of measurement matrix $\mathbf{X}$ with absolute values of the coefficients of correlation [9, 27, 33] between the columns $\mathbf{X}_{*i}$ and $\mathbf{X}_{*j}$, $i \neq j$, bigger than 0.8, an influence of ratio $k = \sigma_{\text{noise}}/\sigma_{\text{signal}}$, where σ^2 – variation of regularizing noise, on condition $\text{cond}(\tilde{X}^T\tilde{X})$ and on determinant $\det(\tilde{X}^T\tilde{X})$ of the regularized FIM was examined on a model and shown in Fig. 4.

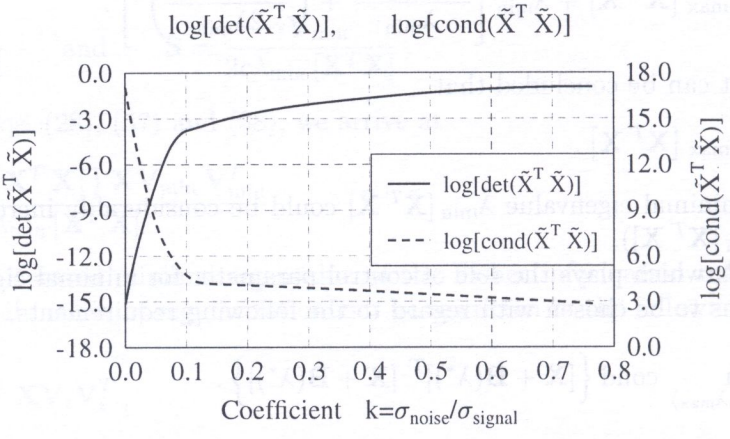

Fig. 4. An influence of ratio $k = \sigma_{\text{noise}}/\sigma_{\text{signal}}$ on stability of parametric estimation with use of the regularized LS method by randomization in space of the measurement matrix; $\tilde{X}$ – randomized measurement matrix

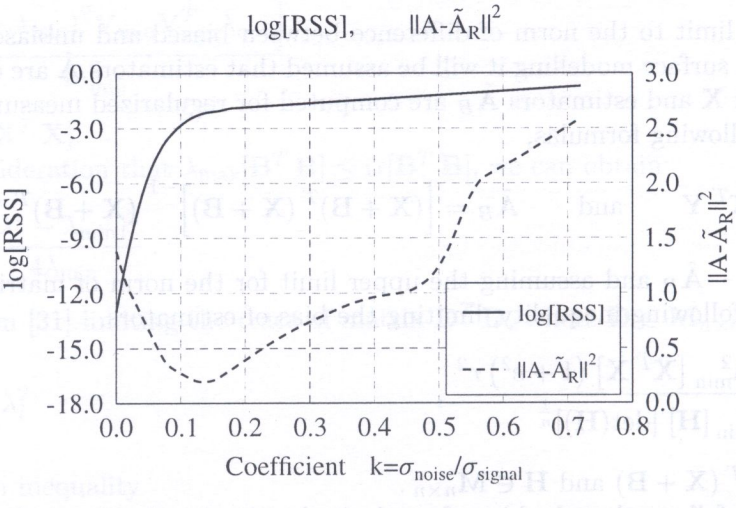

Fig. 5. An influence of ratio $k = \sigma_{\text{noise}}/\sigma_{\text{signal}}$ on accuracy of the parametric estimation in the case of randomized in space measurement matrix; RSS – residual sum of squares; $\mathbf{A}$ – vector of coefficients of considered model; $\tilde{\mathbf{A}}_R$ – vector of estimated coefficients; $\|[]\|$ – Euclidean norm of vector []

Corresponding influence of ratio $k = \sigma_{\text{noise}}/\sigma_{\text{signal}}$ on the Euclidean norm of the bias of estimators $\|\mathbf{A} - \hat{\mathbf{A}}_R\|^2$ and on the residual sum of squares (RSS) [9, 27, 33] was also investigated numerically on the same model with strong collinearities in measurement matrix $\mathbf{X}$ (Fig. 5).

An analysis of the related FIM's eigenvectors and eigenvalues showed that its minimal eigenvalue was very close to zero i.e., $\lambda_{\min} \approx 0$. Optimal randomization of the measurements gave the estimators obtained by randomized LS method app. 5 times more accurate with the condition $\text{cond}(\tilde{X}^T\tilde{X})$ of the randomized FIM app. 10^9 times less, than in the case of conventional standard LS estimation.

7. Conclusions

In the cases, when the smoothness of surface is required, as it takes place with modelling of conjugate surfaces of tooth flanks, and standard procedures lead to undesired fluctuations of curves and associated surfaces, it is profitable to interpolate them by models of curves and surfaces, which do not fit the points but pass very close to given set of points. If standard LS method gives unstable coefficients because of ill-conditioned Fisher matrix (containing strong or almost strong multicollinearities), it is better to regularize coordinates of the nodes from the mesh of curves laying on considered surface with the optimal matrix $\mathbf{B}$. Significant improvement of the stability of estimation, and related surface modelling can be achieved when parameters of regularization are chosen with consideration of the minimal possible condition of the Fisher information matrix. Presented method is useful for determination of tooth contact area in gear transmissions with tooth flanks defined by discrete data e.g., tooth flanks with modified shape.

References

[1] *ADINA Theory and Modeling Guide.* ADINA R & D, Inc., Watertown, MA, 1997.

[2] B.D.O. Anderson, J.B. Moore. *Optimal Filtering.* Prentice Hall, Inc., Englewood Cliffs, NJ, 1979.

[3] K.J. Bathe. *Finite Element Procedures.* Prentice Hall, Inc., Englewood Cliffs, NJ, 1996.

[4] M.L. Baxter. Basic geometry and tooth contact of hypoid gears. *Industrial Mathematics*, **2**: 1–28, 1961.

[5] M.S. Casale, E.L. Stanton. An overview of analytic solid modeling. *IEEE Computer Graphics and Applications* **5**: 45–56, 1985.

[6] M. Chang, P.P. Lin. On-line free form surface measurement via a fuzzy-logic controlled scanning probe. *Int. J. Mach. Tools and Manufacture*, **39**: 537–552, 1999.

[7] C. De Boor. *A Practical Guide to Splines.* Springer-Verlag, New York, 1978.

[8] DIN 3960 *Begriffe und Bestimmungsgrößen für Stirnräder (Zylinderräder) und Stirnradpaare (Zylinderpaare) mit Evolventenverzahnung.*

[9] N.R. Draper, H. Smith. *Applied Regression Analysis.* 2nd ed. J. Wiley, New York, 1981.

[10] G. Farin. *Curves and Surfaces for Computer Aided Geometric Design. A Practical Guide.* Academic Press, Boston, 1990.

[11] I.D. Faux, M.J. Pratt. *Computational Geometry for Design and Manufacturing.* J. Wiley. London. 1983.

[12] T. Hösel. Flankenrichtungskorrekturen für hochbelesteten Stirnradgetriebe, *Textes des 2-éme Congres Mondial des Engrenages, Paris*, **1**: 89–95, 1986.

[13] P.J. Huber. *Robust Statistics.* J. Willey, New York, 1981.

[14] A. Kawalec. *Randomized Methods of Stable Estimation for Aerodynamical Characteristics of Civil Aircrafts.* PhD thesis. Kiev Institute of Civil Aviation Engineers, Kiev, 1988.

[15] A. Kawalec. Modelling of gear tooth flanks based on distorted measurements. *Advances in Technology of Machines and Mechanical Equipment*, **21**(3): 5–28, 1997.

[16] A. Kawalec. An influence of load distribution on stresses and deformations of tooth in straight-tooth gear transmission. *Advances in Technology of the Machines and Equipment*, **22**(3): 53–77, 1998.

[17] A. Kawalec, J. Wiktor. Analytical and numerical method of determination of spur gear tooth form machined by gear tools. *Advances in Technology of the Machines and Equipment*, **23**(2): 5–28, 1999.

[18] A. Kawalec, J. Wiktor. Implementation of CAE techniques in spur gear design and manufacturing. *Proc. 4th Int. Scientific Colloquium "CAx Techniques", Bielefeld, Germany*, 347–359, 1999.

[19] V. Kin. Computerized analysis of gear based on coordinate measurement data. *Trans. ASME, J. Mechanical Design*, **116**(3): 738–744, 1994.

[20] M. Kleiber ed. *Handbook of Computational Solid Mechanics.* Springer, Berlin, 1998.

[21] P. Lancaster. *Theory of Matrices.* Academic Press, New York, 1965.

[22] Ch.L. Lawson, R.J. Hanson. *Solving Least Squares Problems.* Prentice-Hall, Inc., Englewood Cliffs, 1974.
[23] F.L. Litvin, J. Zhang, *et al.* Topology of modified helical gears. *Surface Topography*, **2**: 41–58, 1988.
[24] F.L. Litvin, V. Kin. Simulation of meshing, bearing contact and transmission errors for single-enveloping worm-gear drives. AGMA Paper 90FTM3, AGMA, Alexandria, VA, 1990.
[25] F.L. Litvin, Y. Zhang, *et al.* Computerized inspection of real surfaces and minimization of their deviations. *Int. J. Mach. Tools and Manufacture*, **32**(1/2): 141–145, 1992.
[26] F.L. Litvin. *Gear Geometry and Applied Theory.* Prentice-Hall, Inc., New York. 1994.
[27] J.H. Maindonald. *Statistical Computation.* J. Wiley, New York, 1984.
[28] W.I. Mudrov, W.A. Kuschko. *Methods of Processing of Measurement's Quasiprobable Estimators*, (in Russian). Radio i Sviaz, Moscow, 1983.
[29] J.H. Mathews. *Numerical Methods for Mathematics, Science and Engineering.* Prentice Hall, Inc., Englewood Cliffs, NJ, 1987.
[30] G. Niemann, H. Winter. *Maschinenelemente*, Band II, Springer-Verlag, Berlin, 1985.
[31] C.R. Rao. *Linear Statistical Inference and Its Applications.* J. Wiley, New York, 1966.
[32] J. Rybak, A. Kawalec, J. Wiktor. Analysis of tooth contact in involute cylindrical gear transmissions with modified tooth trace. *Proc. IV World Congress on Gearing and Power Transmissions. Paris*, **1**: 169–181, 1999.
[33] G.A.F. Seber. *Linear Regression Analysis.* J. Wiley, New York, 1977.
[34] H.J. Stadtfeld. *Gleason Bevel Gear Technology – Manufacturing Inspection and Optimization – Collected publications 1994/1995.* Gleason Machine Division, The Gleason Works, Rochester NY, 1995.
[35] A. Seifried, R. Bürkle. Die Berührung der Zahnflanken von Evolventenschraubenrädern, *Werkstatt und Betrieb*, **4**: 183–187, 1968.
[36] I.J. Schoenberg. Contributions to the problem of approximation of equidistant data by analytic functions. *Quarterly of Applied Mathematics*, **4**, Part A: 45–99, 1946.
[37] A.N. Tikchonov, V.Ya. Arsenin. *Methods of Solution of Ill-Conditioned Tasks*, (in Russian). Nauka, Moscow, 1986.
[38] J. Stoer, R. Bulirsch. *Introduction to Numerical Analysis.* Springer-Verlag, New York, 1983.
[39] *Tooth contact analysis. Formulas and calculation procedures.* Gleason Machine Division, The Gleason Works, Rochester NY, 1964.
[40] D.P. Townsend. *Dudley's Gear Handbook.* McGraw–Hill, New York, 1992.
[41] M. Weck, H. Saljé. Leistungssteigerung an hochbelasteten Getrieben durch gezielte Zahnflankenkorrekturen. *Textes des 2-éme Congres Mondial des Engrenages, Paris*, **1**: 161–171, 1986.
[42] M. Weck ed. *Moderne Leistungsgetriebe. Verzahnungsauslegung und Betriebsverhalten.* Springer-Verlag, Berlin. 1992.
[43] J. Wiktor. An influence of type and parameters of tooth flank modification on tooth contact area in spur gear transmission, (in Polish). *Przegląd Mechaniczny*, **57**(15): 15–19, 1998.
[44] J.H. Wilkinson. *The Algebraic Eigenvalue Problem.* Clarendon Press, Oxford, 1965.
[45] J.R. Woodwark. Blends in geometric modelling, In: R.R. Martin, ed., *The Mathematics of Surfaces II. Proc. 2nd IMA Conf. on the Mathematics of Surfaces, Cardiff, Wales*, Oxford University Press, 1987.
[46] Z. Wójcik. *Machine Tools for Machining of Bevel Gears*, (in Polish). WNT, Warsaw, 1993.