Geodesic survey and modernization of a route as the task of optimization

Edward Nowak

Warsaw University of Technology
Faculty of Geodesy and Cartography
Pl. Politechniki 1, 00-661 Warsaw, Poland
e-mail: e.nowak@gik.pw.edu.pl; phone No.: +48 22 2347751

Abstract: A geodesic survey of an existing route requires one to determine the approximation curve by means of optimization using the total least squares method (TLSM). The objective function of the LSM was found to be a square of the Mahalanobis distance in the adjustment field v. In approximation tasks, the Mahalanobis distance is the distance from a survey point to the desired curve. In the case of linear regression, this distance is codirectional with a coordinate axis; in orthogonal regression, it is codirectional with the normal line to the curve. Accepting the Mahalanobis distance from the survey point as a quasi-observation allows us to conduct adjustment using a numerically exact parametric procedure.

Analysis of the potential application of splines under the NURBS (non-uniform rational B-spline) industrial standard with respect to route approximation has identified two issues: a lack of the value of the localizing parameter for a given survey point and the use of vector parameters that define the shape of the curve. The value of the localizing parameter was determined by projecting the survey point onto the curve. This projection, together with the aforementioned Mahalanobis distance, splits the position vector of the curve into two orthogonal constituents within the local coordinate system of the curve. A similar system corresponds to points that form the control polygonal chain and allows us to find their position with the help of a scalar variable that determines the shape of the curve by moving a knot toward the normal line.

1. Introduction

A geodesic survey of an existing route requires us to determine the approximation curve by means of optimization using the total least squares method (TLSM). The solution is obtained through a numerically complicated process called the generalized adjustment (i.e., total least squares) and involves an orthogonal projection that was investigated in detail by C. R. Rao (1982). In this article, we will investigate how the properties of this projection relate to distances from the observed points to the

approximated curve and whether the calculations can be streamlined by reducing them to a parametric procedure, as in linear regression. Furthermore, we will conduct route approximation using splines under the NURBS industrial standard. We will also investigate how survey point distribution affects our ability to detect gross errors in the observation system.

2. Formulating the objective function

Optimization involves identifying the minimum of a function in an area defined by certain constraints. We begin the identification by specifying the objective function. The least squares method (LSM), developed for independent observations by Gauss and Lagrange, is a statistical algorithm used to assess congruity between the model and observation.

$$\min \left[\frac{vv}{mm}\right] = \min \sum_{i=1}^{n} \frac{v_i^2}{\sigma_i^2} \quad \text{gdzie } v_i = L_i^{adj} - L_i^{obs} \tag{1}$$

The general formulation (Aitken, 1935; Knight et al., 2010) involves the minimization of the objective function in its square form

$$min\ E = min\ v^T C^{-1} v$$

where
$$C = Cov(L^{obs}, L^{obs}), \quad v = L^{obs} - L^{adj} \tag{2}$$

and is derived from estimating parameters of random variables using the maximum likelihood method and type II regression (Fisz, 1969; Rao, 1982).

Modeling a route involves a particular type of observation that forms vectors of the observed positions of points. Today, such observations are most frequently obtained via GPS technology. In this case, the observation vector comprises two-dimensional subvectors that include an inseparable pair of coordinates of the k-th survey point $r_k^T = \left[x_k^{obs}, y_k^{obs}\right]^T$ and results in the division of the covariance matrix of observations C into 2×2 block matrices that describe covariance pairs of survey points

$$Cov(r_k, r_j) = \begin{bmatrix} cov(x_k, x_j) & cov(x_k, y_j) \\ cov(y_k, x_j) & cov(y_k, y_j) \end{bmatrix} \tag{3}$$

The accuracy with which individual coordinates are determined may vary significantly. Theoretically, one of the coordinates may be determined without error, which results in the two-dimensional random variable becoming degenerate. For an error-free measurement of the x coordinate (in practice, $\sigma_x < 0.1\ \sigma_y$ is sufficient), the block diagonal matrix is

$$Cov(r_k, r_k) = \begin{bmatrix} 0 & 0 \\ 0 & \sigma^2(y_k) \end{bmatrix} \tag{3a}$$

which leads to the linear regression **y|x**.

$$min \ \sum_{k=1}^{n} \left(\frac{v_{y_k}}{\sigma_{y_k}}\right)^2 \tag{1a}$$

Similarly, for an error-free measurement of the *y* coordinate, the diagonal block matrix leads to the linear regression **x|y**.

3. Formulating constraints

Conditions limit the freedom of movement within the adjustment space and may bind adjustments directly or through parameters.

3.1. Equation of a survey point on a curve

We will formulate the condition that binds adjustments of survey point coordinates and that binds parameters p of the curve given by the implicit equation $F(x, y, p_1, ..., p_u) = 0$. Each observed point $\left[x_k^{obs}, y_k^{obs}\right]^T for \ k = 1, 2, ..., n \geq u$ provides one restriction equation that binds the parameters we wish to determine. These parameters describe the model and the observed coordinates of the survey point (vector-based observation).

$$F\left(x_k^{obs} + v_{x_k}, y_k^{obs} + v_{x_k}, p_1, ... p_u\right) = 0 \text{ for } k = 1,2, ..., n \tag{4}$$

Equations (4) include adjustments and parameters. They lead to a numerically complicated generalized adjustment formulation (i.e., total least squares, also known as rigorous least squares) (Golub and Van Loan, 1980). Our task becomes easier if the curve is defined by an explicit function of parameters and an error-free coordinate; for instance, for $\sigma^2(x_k) = 0$

$$y_k^{obs} + v_{y_k} = f\left(x_k^{obs}, p_1, ..., p_u\right) \tag{4a}$$

which leads to the equation for the adjustment of the observational parametric LSM procedure (1a)

$$v_{y_k} = f\left(x_k^{obs}, p_1, ..., p_u\right) - y_k^{obs} \tag{5}$$

that determines the simple regression curve **y|x**.

3.2. Observational equation of a survey point as the Mahalanobis distance

The Mahalanobis distance (also called the weighted Euclidean distance) between two points with position vectors x, y is given by

$$v_{y_k} = f(x_k^{obs}, p_1, \dots, p_u) - y_k^{obs} \tag{6}$$

where C is a covariane matrix

For vectors that represent the observed and adjusted values

$$x - y = L^{obs} - L^{adj} = v$$

the square of the Mahalanobis distance in the adjustment field v constitutes the objective function of the LSM. If observations of survey point positions are independent, the objective function is the sum of the squares of Mahalanobis distances from individual survey points to the curve

$$E = \sum_{j=1}^{n} D_j^2$$

$$\text{where} \quad D_j^2 = \begin{bmatrix} v_{x_j} \\ v_{y_j} \end{bmatrix}^T \begin{bmatrix} \text{cov}(x_j, x_j) & \text{cov}(x_j, y_j) \\ \text{cov}(y_j, x_j) & \text{cov}(y_j, y_j) \end{bmatrix}^{-1} \begin{bmatrix} v_{x_j} \\ v_{y_j} \end{bmatrix} \tag{7}$$

The Mahalanobis distance from a survey point to the desired curve is a quasi-observation that allows us to conduct adjustments using a parametric procedure.

3.3. Compound curves

The curves analyzed above were sets of points for which the coordinates are consistent with the appropriate equations. A curve is a one-dimensional geometric object composed of points in a multidimensional space. A curve can be described in the parametric form using a function (relation) F that relates the scalar parameter t to the vector-point $\mathbf{P}$ in space. Thus, the **geometric relation** scalar→vector generates plane curves in a given two-dimensional space R→R^2 and space curves in a three-dimensional space R→R^3.

$$\mathbf{P} = F(t) \tag{10}$$

The t parameter determines the position on the curve. By calculating the differential of the curve segment length, we obtain the natural parameter

$$s = \int \sqrt{1 + F'^2} \, dt \tag{11}$$

that provides the exact position of a point on the curve. In respect to the route, this parameter constitutes the length of the route.

Curves composed of curve segments must meet additional conditions related to parameters that ensure continuity and smoothness.

Curve continuity means that the start point of the subsequent curve segment $\boldsymbol{Q}(s)$ overlaps the end point of the preceding curve segment.

$$\mathbf{P}(s = s_{max}) = \mathbf{Q}(s = 0) \tag{12}$$

Curve smoothness means that tangents of curve segments overlap at points of connection between these segments. In other words, the derivative of the curve is continuous with respect to the natural parameter

$$\left(\frac{d\mathbf{P}}{ds}\right)_{s=s_{max}} = \left(\frac{d\mathbf{Q}}{ds}\right)_{s=0} \tag{13}$$

Similarly, we may require that second order derivatives and higher order derivatives, or even their functions (e.g., the radii of the curvature), be continuous. Conditions related to parameters most frequently apply in free adjustment and lead to the boundarization procedure (Rao and Mitra, 1971) or the apparent connection procedure (Nowak, 2000).

4. Geometric parametrization

The geometric projection scalar $t \rightarrow$ vector **P** generates curves. We will now consider parametrization that uses vectors in the analyzed space. The simplest curve is a line on which any point described by the vector **P** meets the equation

$$\mathbf{P}(t) = \mathbf{W_0} * (1 - t) + \mathbf{W_1} * t \tag{14}$$

where t is the current parameter of the line and $\mathbf{W_0}$ and $\mathbf{W_1}$ are point-vectors of the parameters.

Therefore, the point $\boldsymbol{P}(t)$ on the line is a vector given by a linear combination of vectors of the parameters. In the end, we obtain a segment of the line W_0-W_1 for a given interval of the variable $0 < t < 1$. This is important because in practical applications, we usually work with finite curve segments (Kiciak, 2000). Thus, formal parameters have become the same vectors as survey points and the calculated points, which facilitates transformations when algorithms are carried out as well as sharing results with the user.

Further examples include:
− a polynomial curve (Figure 1)

$$\mathbf{P}(t) = \sum_{i=0}^{n} \mathbf{W}_i * t^i \tag{15}$$

− a curve defined by basis functions

$$\mathbf{P}(t) = \sum_{i=0}^{n} \mathbf{W}_i * f_i(t) \tag{16}$$
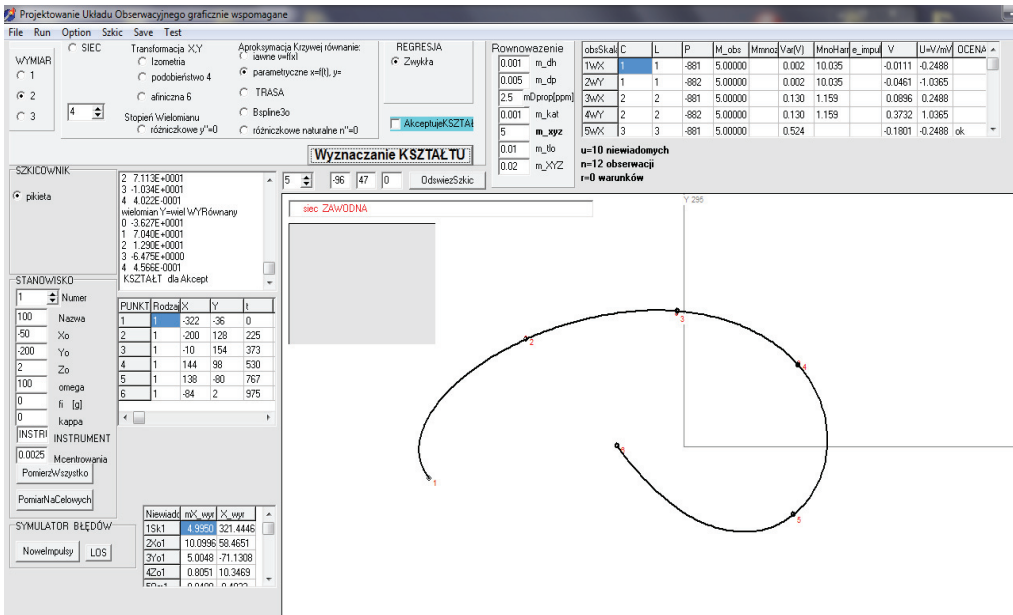

Fig. 1. Approximation with a parametric polynomial using the PUOgraf software, developed by the author of this article

An interesting case of basis functions are Bernstein polynomials

$$B_i^n(t) = \binom{n}{i} t^i (1-t)^{n-i} \qquad i = 0,1,\dots,n; \quad 0 \le t \le 1 \tag{17}$$

that describe analytic Bézier curves

$$\mathbf{P}(t) = \sum_{i=0}^{n} \mathbf{W}_i * B_i^n(t) \tag{18}$$

which follow a control polygonal chain defined by a sequence of vertices $\mathbf{W_0, W_1, \dots, W_n}$.

The derivative of a Bézier curve (Kiciak, 2000)

$$\mathbf{P}'(t) = n \sum_{i=0}^{n-1}(\mathbf{W_{i+1}} - \mathbf{W_i}) * B_i^{n-1}(t) \tag{19}$$

is a Bézier curve with a control polygonal chain that follows subtractions of vectors of parameters. It follows that the tangents of the Bézier curve at end points overlap with the first and last segments of the control polygonal chain. This enables us to easily connect Bézier curve segments into a smooth spline. Derivatives of higher order have control polygonal chains defined over subtractions of consecutive orders. In particular, second order subtractions enable us to determine the curvature.

A set of curve segments is characterized by a vector of parameters composed of subvectors, that is, points on control polygonal chains that correspond to appropriate curve segments. These parameters are bound by conditions derived from the continuity and smoothness requirements at join points between curve segments. Defining smoothness conditions frequently proves difficult. A breakthrough took place when parametrization of a polynomial Bézier curve in the form of a control polygonal chain was applied due to the fact that the last segment of the polygonal chain is also the tangent to the curve at the end point. A smooth connection between Bézier curves can therefore be attained simply by extending the last segment of the control polygonal chain into the first segment of the control polygonal chain for the neighboring curve segment (B-spline). The ease with which curve segments can be smoothly connected leads to the definition of the NURBS (non-uniform rational B-spline) industrial standard, used to design complicated shapes (Humienny, 2009).

Let us consider two Bézier curve segments characterized by control polygonal chains, $(\mathbf{p_0}, \mathbf{p_1}, \ldots, \mathbf{p_n})$ and $(\mathbf{r_0}, \mathbf{r_1}, \ldots, \mathbf{r_n})$

– the continuity of connection condition is $\mathbf{p_n} - \mathbf{r_0} = 0$ (20)

– the smoothness of connection condition is $\mathbf{p_{n-1}} + \mathbf{r_1} - 2\mathbf{r_0} = 0$ (21)

– the continuity of curvature condition is $\mathbf{r_2} - \mathbf{p_{n-2}} - 4(\mathbf{r_1} - \mathbf{r_0}) = 0$ (22)

We will analyze the procedure of an approximation task in a case when the shape of the approximation curve determines the set of vectors that constitutes a geometric parametrization. For instance, a B-spline composed of three curve segments of the third degree requires:

– 3*4-2=10 vectors that form 3 connected control polygonal chains; knot vectors at join points between curve segments overlap, which replaces two continuity conditions (20);

– those vectors of parameters must satisfy two conditions for a smooth connection between curve segments (21).

Parametric equations (10) enable a direct generation of coordinates of any point on the curve for a given value of the parameter t. It is a perfect tool for the designer and the developer (*viz.* the NURBS standard); however, in approximation tasks, the value of the parameter t is unknown. We will write the vector equation (10) in the form of two scalar equations

$$x = g(t) \tag{10a}$$

$$y = h(t) \tag{10b}$$

The desired value of the parameter t for a survey point can be calculated by solving the equation (10a)

$$t = g^{-1}(x^{obs}) \tag{23}$$

Substituting into (10b), we obtain

$$y = h(g^{-1}(x^{obs})) \tag{24}$$

The value of the y coordinate calculated in this manner is a special case of equation (4a) that is present in linear regression (1a).

The above considerations prove that the coordinates of a survey point do not constitute two separate observations; for this to be true, we would need to know the value of the parameter that localizes the survey point on the curve. As far as orthogonal regression is concerned, the localizing parameter is obtained by projecting the survey point onto the curve. Such projection at the same time allows us to identify curve segments in the case of compound curves.

5. Excessive parametrization

To connect curve segments, we need to introduce conditions for the continuity of the function and its derivatives. This clearly indicates that the number of parameters in a spline is too great; in the case of geometric parametrization, such excess involves vectors rather than scalars. Geometric parametrization, while very elegant in form, frequently creates a latent excess of parameters by replacing scalars with vectors. We will demonstrate this issue on the basis of a line described by Equation (14). Equation (14) includes two vector parameters, and the equation for a line in the normal form can be written using two scalar parameters (the azimuth of the normal line and the distance to the origin).

The tangent component of the vector of parameters determines the location on the curve, while the normal component determines the shape of the curve. Therefore, moving the knot towards the second vector derivative will have the greatest effect on the shape of the curve (Nowak, 2001).

5.1. Replacing point parametrization with directional parametrization

The normal component has for a long time been regarded as a means to determine shape. As a result, the normal component exists within theoretical considerations in the form of a natural equation of a curve that describes curvature as a function of curve segment length. In practical terms, it is used to adjust train tracks by aligning sagittas. I have applied the method of manipulating knots by moving them along a given direction (moving the normal line closer) using a scalar parameter for modernizing train routes in Katowice Ironworks by means of orthogonal regression (Distinction of the President of the Head Office of Land Surveying and Cartography from 1980). The above solution can applied to approximate a route through splines under the NURBS standard. We may indirectly obtain the required parameterization in the form of vectors that constitute the control curve by using intermediate scalar variables u_i and second order subtractions that constitute the approximation of the direction of the normal line

$$\mathbf{W_i} = \mathbf{W_i^{app}} + \left(\mathbf{W_{i+1}^{app}} - 2\mathbf{W_i^{app}} + \mathbf{W_{i-1}^{app}}\right) u_i \tag{25}$$

5.2. Constructing unconditional parametrization of a route composed of circular and line segments

Route modeling is a typical task of engineering geodesy. The route is usually composed of line segments and circular segments, which ensures smoothness (and constitutes a spline). Sometimes, transitional curves are used to ensure continuity of curvature. The aforementioned geometric parametrization of polynomial splines uses sets of vectors of parameters related to individual curve segments; it also makes use of conditions for connections between them. Such a procedure involves working with an excessive number of parameters, as can best be seen in the continuity condition. We will attempt to construct an unconditional parametrization of a route composed of circular and line segments. Circular segments connected through tangents create a smooth route. Therefore, the crucial part of the desired spline are circular segments, which can be easily parametrized by providing the coordinates of the center of the circle together with the radius (x, y, r). The same convention allows us to write the start and end points of the route (as well as sharp corners if needed) by treating these points as the boundary case of a null-radius circle. The last step is to determine the manner in which the circles can connect through tangents. Because the route is a linear object, we begin from the start point $(x_0, y_0, r_0 = 0)$. The first line segment is the tangent that runs from the start point to the first circle (x_1, y_1, r_1). We need to choose between two cases: the left-turn or the right-turn. We can write the direction of the turn in the form of the binary orientation of the radius of the circle: a right-turn corresponds to a positive radius $+r$, while a left-turn corresponds to a negative radius $-r$. Thus, we have constructed a route parametrization limited to indispensable parameters only,

which greatly facilitates describing the route and reduces approximation tasks to the simplest case of the least squares method. Figure 2 shows a route composed of two circular segments.
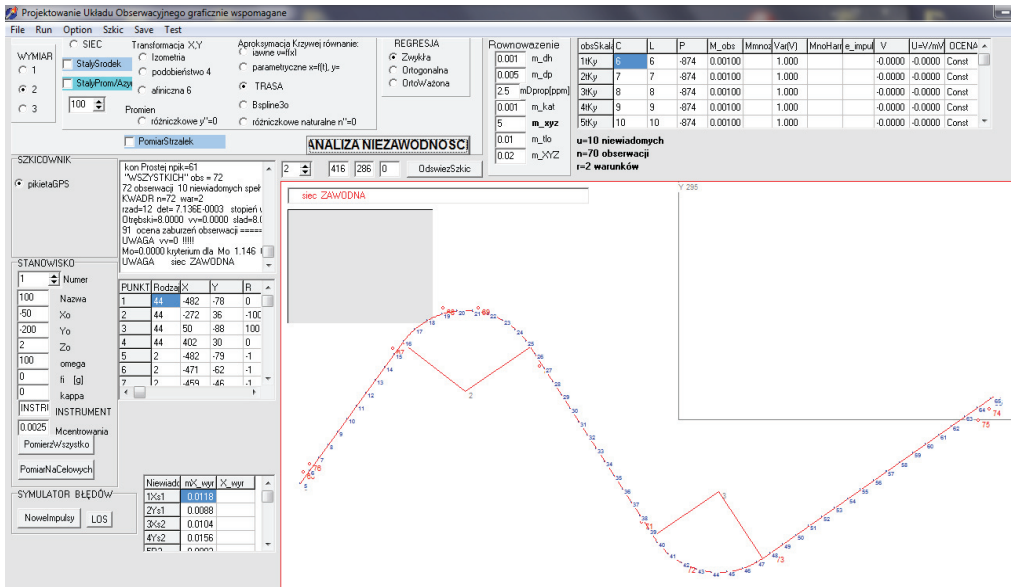

Fig. 2. A route described by four points (the start and end points have $r = 0$; other points have either a positive radius, which corresponds to a right-turn, or a negative radius, which corresponds to a left-turn; the radius is written as the z coordinate)

Such parametrization not only allows us to obtain the values of parameters of the approximation curve for a set of survey points, but also to adjust these values so that the route crosses a given point or has a particular radius of the circular segment; this enables us to modernize existing routes, rather than only adjusting them.

6. Designing geodesic surveys of routes

The object of accuracy- and reliability-based design of regressive systems is the number and distribution of points. In other words, we only manipulate the observations, while the variables are determined by our choice of the regression function, similar to classical networks. Usually, all survey points meet accuracy requirements for route surveying. In such cases, the analysis of parameter accuracy is conducted for formal, rather than technical, reasons. Detecting flaws in the route geometry in the form of unacceptable deviations of survey points from the approximated route requires us to compare the survey points to the standard deviation of the adjustment rather than to

the observation. Furthermore, to unambiguously detect a flaw, the observation system has to meet reliability requirements, just as in observations that constitute a network (Prószyński, 1994). The **PUOgraf** software, developed by the author of this article, can facilitate the procedures described above. The software was created to assist in computer-aided design (CAD) of various observation (network and regression) systems.

It has been shown that in the case of a closed curve, we may achieve reliability through an appropriately dense distribution of survey points (Nowak, 2007); in the case of an open curve, we also need to strengthen the end points with additional survey points. The considerations above lead to the conclusion that the shape of the curve in a vector space is unimportant for reliability-based design. Instead, the determining factor is the distribution of observations in the domain of a scalar parameter. This means that we can create typical projects for use in production instead of individually designing every observation system.

7. Conclusions

The objective function of the LSM was found to be a square of the Mahalanobis distance in the adjustment field v. In approximation tasks, the Mahalanobis distance is the distance from a survey point to the desired curve. Accepting this distance as a quasi-observation allows us to conduct adjustment using a numerically exact parametric procedure.

Analysis of the potential application of splines under the NURBS industrial standard with respect to route approximation has identified two issues: a lack of the value of the localizing parameter for a given survey point and the use of vector parameters that define the shape of the curve. The value of the localizing parameter was determined by projecting the survey point onto the curve. This projection, together with the aforementioned Mahalanobis distance, splits the position vector of the curve into two orthogonal constituents within the local coordinate system of the curve. This system allows us to obtain the position of points that create the control polygonal chain of the Bézier function using a scalar variable that determines the shape of the curve by moving the knot in the direction of the normal line.

Acknowledgments

References

Fisz, M. (1969). *Rachunek prawdopodobieństwa i statystyka matematyczna* [*Probability and Mathematical Statistics*]. Warsaw: PWN Warszawa.

Golub, G. H. & Van Loan C. F. (1980). An analysis of the total least squares problem. *SIAM J. on Numer. Anal.*, 17, 883–893.

Humienny, Z. (2009). *Specyfikacje geometri wyrobów* [*Geometrical Product Specifications*]. Course for Technical Universities. Warsaw: WNT Warszawa.

Kiciak, P. (2000). *Podstawy modelowania krzywych i powierzchni* [*Basics of modeling curves and surfaces*]. Warsaw: WNT Warszawa.

Knight, N., Wang J. & Rizos C. (2010): Generalised measures of reliability for multiple outliers. *J Geod,* (84), 625-635.

Nowak, E. & Prószyński W. (1990): An interactive system for solving horizontal positioning problems in engineering surveys. FIG XIX International Congress, pp. 507. 1-7 May 1990, Helsinki Finland.

Nowak, E. (2000). *Algorytmy numeryczne w geodezji.* [*Digital alhorithms in geodesy*] Warsaw: Oficyna Wydawnicza PW Warszawa.

Nowak, E. (2001). *Algorytmy wygładzania i generalizacji* łuków [*Algorithms of smoothing and generalization curves*]. Zesz. Nauk. Akademii Rolniczej we Wrocławiu, Geodezja i Urządzenia Rolne series no. 394, Wrocław. pp.317.

Nowak, E. (2007). Quality analysis testing for quantifying reliability properties. *Reports on Geodesy*, 1(82). 207

Prószyński, W. (1994). Criteria for internal reliability of linear least squares models. *Bull. Geodesique, 68*,162-167.

Rao, C. R. (1982). *Modele liniowe statystyki matematycznej* [*Linear models of mathematical statistics*]. Warsaw: PWN Warszawa.

Rao, C.R. & Mitra S.K. (1971). *Generalized Inverse of Matrices and its Applications*. New York: John Wiley.

Inwentaryzacja i modernizacja trasy jako zadanie optymalizacji

Edward Nowak

Politechnika Warszawska
Wydział Geodezji i Kartografii
Pl. Politechniki 1, 00-661 Warszawa
e-mail: e.nowak@gik.pw.edu.pl; phone No.: +48 22 2347751

Streszczenie

Inwentaryzacja istniejącej trasy wymaga wyznaczenia krzywej aproksymującej w wyniku optymalizacji realizowanej metodą najmniejszych kwadratów (TLSM). Analiza funkcji celu LSM wykazała, że jest ona kwadratem odległości Mahalanobisa w przestrzeni poprawek. W zadaniach aproksymacyjnych odległość Mahalanobisa jest miarą odstępu pikiety od wyznaczanej krzywej, w przypadku regresji zwykłej odstęp ten ma kierunek osi układu współrzędnych a w przypadku regresji ortogonalnej odstęp ma kierunek normalnej do krzywej. Uznanie odległości Mahalanobisa pikiety od wyznaczanej krzywej za quasi-obserwację pozwala na wykonanie wyrównania dopracowaną numerycznie procedurą parametryczną.

Badanie możliwości zastosowania funkcji sklejanych w przemysłowym standardzie NURBS do aprosymacji przebiegu trasy wykazało dwa problemy: brak wartości parametru lokalizującego dla pikiety oraz operowanie parametrami wektorowymi defilującymi kształt krzywej. Wartość parametru lokalizującego wyznaczono przez rzut pikiety na krzywą – łącznie z opisaną wyżej odległością Mahalanobisa stanowi on rozkład wektora wodzącego pikiety na dwie składowe podłużną i poprzeczną w lokalnym układzie krzywej. Analogiczny układ w punktach tworzących łamaną kontrolną funkcji Beziera pozwala na wyznaczenie ich położenia za pośrednictwem niewiadomej skalarnej modelującej kształt krzywej po-przez przesunięcia węzła w kierunku normalnej.