THE PUBLISHING HOUSE
OF THE ROMANIAN ACADEMY

PROCEEDINGS OF THE ROMANIAN ACADEMY, Series A,
Volume 18, Number 2/2017, pp. 101–107

SOME NUMERICAL APPLICATION OF THE PSEUDOINVERSES

Alina PETRESCU-NIȚĂ

"Politehnica" University of Bucharest, Depart. of Mathematics, Romania,
E-mail: alina.petrescu@upb.ro

Abstract. After an introduction that recalls the Lagrange interpolation and the Gauss regression, in §2 we introduce what can be called the cylinders and the cones of regression, that indicate the data trend. The main result of this paper is presented in §4, as an application of the pseudoinverses of matrices, namely an algorithm to determine a geometrical pattern which is the nearest to a finite set of points, which can be interpreted as observation points.

Key words: interpolation, regression, pseudoinverse.

1. INTRODUCTION

It is well-known the role either of the interpolations or of the regression curves in processing numerical data, in order to fit a given set of points in a plane and to get some suggestive predictions regarding the trend of those data. This paper treats such problems.

We present now some known notions and results following [2, 3, 5].

Fix integers $n \geq 1$, $r \geq 1$ and a commutative field K of characteristic zero (e.g. $K = \mathbf{Q}, \mathbf{R}, \mathbf{C}, \mathbf{R}(X)$). Let $K_r[X_1,\ldots,X_n]$ be the vector space of all polynomials in n in determinates, of degree at most r, with coefficients in K. Recall that a finite subset $\mathcal{A} = \{a_1, a_2, \ldots, a_N\} \subset K^n$ is a *Lagrange set of degree r* if for any data $b_1,\ldots,b_N \in K$, there is a unique polynomial $P \in K_r[X_1,\ldots,X_n]$ such that

$P(a_p) = b_p$, for any $1 \leq p \leq N$. (1)

Explicitly, if $P = \sum_q c_q X_1^{i_1} \ldots X_n^{i_n}$; $q = (i_1,\ldots,i_n)$, $|q| \leq r$, $c_q \in K$ and if $a_p = (a_p^1,\ldots,a_p^n)$, then (1) becomes $\sum_q c_q (a_p^1)^{i_1} \ldots (a_p^n)^{i_n} = b_p$. The number of coefficients c_q, as well as indices $q = (i_1,\ldots,i_n)$ with $|q| = i_1 + \ldots + i_n \leq r$, is $\binom{n+r}{n}$; all of them could be ordered lexicografically. Therefore, a necessary condition for $\mathcal{A}$ to be Lagrange of degree r is $|\mathcal{A}| = \binom{n+r}{n}$, that will be supposed in what follows. The property of $\mathcal{A}$ to be Lagrange is generic (that means it takes place almost everywhere); indeed, the square matrix $D = (d_{pq})$, where $d_{pq} = (a_p^1)^{i_1} \ldots (a_p^n)^{i_n}$; $1 \leq p \leq N$ and $q = (i_1,\ldots,i_n)$ is nonsingular almost every-where. On the other part, if we denote by $L_1,\ldots,L_N$ the polynomials of degree at most r such that $L_i(a_j) = \delta_{ij}$ for any i, j, then for any table of data (1), the solution of the interpolation problem is $= \sum_{p=1}^{N} b_p L_p$; the polynomials L_p form a basis of the linear space $K_r[X_1,\ldots,X_N]$.

Examples 1.1.

1) If $n = 1$, $r \geq 1$, then $N = r + 1$ and any set $\mathcal{A} = \{a_1, a_2, \ldots, a_{r+1}\}$ is Lagrange of degree r (since the matrix D is Vandermonde, hence nonsingular). For any table (1), through the distinct points (a_p, b_p), $1 \leq p \leq r+1$, passes the graph of the classical Lagrange polynomial of degree $\leq r$.

2) In the case $n = 2$, a Lagrange set of degree r has $N = \frac{1}{2}(r+1)(r+2)$ elments. If $r = 1$, then $N = 3$ and a triangular plate $\{a_1, a_2, a_3\}$ is Lagrange iff the vertexes a_1, a_2, a_3 are not colinear; if $r = 2$, $N = 6$ and hexagonal plate $\{a_1, a_2, \ldots, a_6\}$ is Lagrange iff its vertexes are not situated on the same conic.

In the case $n = 1$, the Lagrange interpolation allowed some remarkable formulas, e.g. the Shannon sampling theorem well known from the signal theory. For $n \geq 2$, the Lagrange interpolation is too rigid and moreover, whenever to a Lagrange set $\mathcal{A}$ one adds a new point a, it is difficult to get conditions to $\mathcal{A} \bigcup \{\alpha\}$ to be Lagrange. Some shortcomings of the Lagrange interpolation, e.g. Runge phenomenon, were repaired by use of spline-functions. Gauss relaxed the rigid conditions like (1); in the case $n = 1$, he changed the point of view. Instead to require a polynomial curve passing through the „observation points" $M_p(a_p, b_p)$, $1 \leq p \leq r+1$, he proposed to find a family of curves $y = \varphi(x, \lambda_1, \ldots, \lambda_M)$ with M parameters and to determine the values $\lambda_1^0, \ldots, \lambda_M^0$ such that the corresponding curve „lies near" all together points M_p. This was the beginning of the L^2 – technics ("least squares method"), see [2, 4].

2. CURVES AND SURFACES OF REGRESSION

In what follows, we will suppose that $K = \mathbf{R}$. Gauss replaced the condition (1), i.e. $\varphi(a_p, \lambda_1, \ldots, \lambda_M) = b_p$ for $1 \leq p \leq r+1$ (an incompatible or overdetermined system), by the condition that all the differences $\varepsilon_p = \varphi(a_p, \lambda_1, \ldots, \lambda_M) - b_p$ to be „simultaneously small". The most convenient for this is the sum $\sum\limits_{p=1}^{M} \varepsilon_p^2$ to be minimum (not the sum of all ε_p or $|\varepsilon_p|$ to be minimum). Therefore, we have to determine λ_p g such that $g(\lambda_1, \ldots, \lambda_M) = \sum\limits_{p=1}^{r+1} (\varphi(x_p, \lambda_1, \ldots, \lambda_M) - y_p)^2$ is minimum. Supposing φ of class C^2, the necessary condition is: $\frac{\partial g}{\partial \lambda_k} = 0$, $1 \leq k \leq M$. Under that hypothesis this system has a solution $(\lambda_1^0, \ldots, \lambda_M^0)$, the corresponding curve $\Gamma : y = \varphi(x, \lambda_1^0, \ldots, \lambda_M^0)$ is called the regression curve which „mediates" among the observation points $M_1, \ldots, M_{r+1}$; see , for exemple, [3, 5].

Examples 2.1.

If $y = \lambda_1 x + \lambda_2$, one obtains the classical Gauss regression right-line; in this case,

$$g(\lambda_1, \lambda_2) = \sum_{p=1}^{r+1} (\lambda_1 x_p + \lambda_2 - y_p)^2$$

and the system $\frac{\partial g}{\partial \lambda_1} = 0$, $\frac{\partial g}{\partial \lambda_2} = 0$ has a unique solution $(\lambda_1^0, \lambda_2^0)$. The right-line $y = \lambda_1^0 x + \lambda_2^0$ passes between and not through the points M_p, $1 \leq p \leq r+1$ and indicates the trend of the respective data. If we add new values a_p at the set $\mathcal{A}$, one obtain some predictive values b_p; for this reason, the term of "regression" is somehow unfortunate and it could be replaced by that of "progression" or "prediction".

We now change the point of view and consider a set $\mathcal{A} = \{a_1, a_2, \ldots, a_N\}$ of points $a_p = (a_p^1, \ldots, a_p^n)$ from $\mathbf{R}^n$, assimilated to "a cloud" of n –dimensional data, called *points of observation* (or surveillance). We introduce two sugestive notions: *cylinders* and *cones of regression* circumscribed to the date cloudes, suggested from [5] Fig. 1.

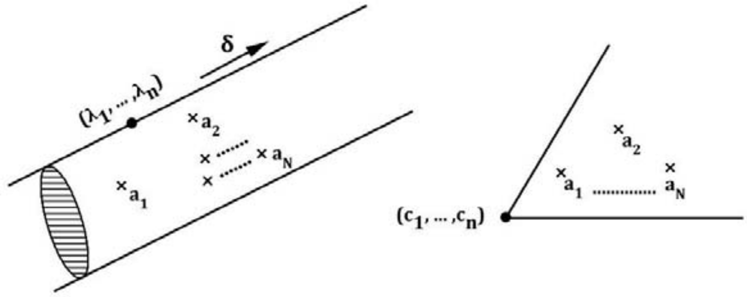

Fig. 1

For any direction δ of versor $\rho=(\rho_1,\ldots,\rho_n)$, one can consider the cylinder Γ_δ, parallel to δ and circumscribed to $\mathcal{A}$; for this, take a right line $\dfrac{x_1-\lambda_1}{\rho_1}=\ldots=\dfrac{x_n-\lambda_n}{\rho_n}$, with $n-1$ equations:

$$\rho_k(x_1-\lambda_1)-\rho_1(x_k-\lambda_k)=0,\ 2\le k\le n.$$

The fact that the points $a_p (1\le p\le N)$ are simultaneously small near Δ means that all

$$\varepsilon_k=\rho_k(x_1^p-\lambda_1)-\rho_1(x_k^p-\lambda_k)$$

are simultaneously small, i.e.

$$\varphi(\lambda_1,\ldots,\lambda_n)=\sum_{p=1}^{N}\sum_{k=2}^{n}[\rho_k(x_1^p-\lambda_1)-\rho_1(x_k^p-\lambda_k)]^2$$

is minimum. From the necessary conditions $\dfrac{\partial\varphi}{\partial\lambda_k}=0$, $1\le k\le n$, one get $\lambda_1,\ldots,\lambda_n$, by making use of the pseudoinverses of the correspoding matrices. Similarly, one can determine a cone with the vertex $c_1,\ldots,c_n$ by determining parameters $\lambda_2,\ldots,\lambda_n$ such that the right line $\dfrac{x_1-c_1}{1}=\dfrac{x_2-c_2}{\lambda_2}=\ldots=\dfrac{x_n-c_n}{\lambda_n}$ is optimal relatively to the set $\mathcal{A}$; thus, all the differences $\lambda_k(x_1-c_1)-(x_k-c_k)$ are simultaneously small, hence the function $\psi(\lambda_2,\ldots,\lambda_n)=\sum\limits_{p=1}^{N}\sum\limits_{k=2}^{n}[\lambda_k(x_1^p-c_1)-(x_k^p-c_k)]^2$ is minimum.

3. ON THE PSEUDOINVERSE OF MATRIX

Fix a matrix $A\in M_{k,n}(\mathbf{R})$ not necessarily square. For a given matrix $B\in M_{k,1}(\mathbf{R})$, Penrose has applied Gauss idea and instead of solving the linear system $A\cdot X=B$, he proved that there is a column vector $\Lambda=(\lambda_1,\ldots,\lambda_n)^T$ such that the norm $\|A\cdot\Lambda-B\|$ is minimum. Moreover, there exists and is unique a matrix $A^+\in M_{k,n}(\mathbf{R})$, called the *pseudoinverse of A*, such that $\Lambda=A^+\cdot B$. If $k = n$ and A is nonsingular, then $A^+=A^{-1}$ (the usual inverse); if $k\le n$ and rank $A = k$, then $A^+=A^T\cdot(A\cdot A^T)^{-1}$; see [7], [8]; for some other applications of the pseudoinverses see [1].

Examples 3.1.

Suppose that A $A\in M_n(\mathbf{R})$ is positively defined, $B\in M_{m,n}(\mathbf{R})$, $b\in M_{m,1}(\mathbf{R})$ and $k\ge 0$ a scalar. Put the problem to minimize the functional $J:\mathbf{R}^m\to\mathbf{R}$, $J(X)=X^T\cdot A+kX^T\cdot A\cdot X$, with the restriction

$B \cdot X = b$. So, $J(X) = X^T \cdot (\mathbf{I}_n + kA)$ and since $\mathbf{I}_n + kA$ is positively defined, there exists an inferior triangular matrix C such that $\mathbf{I}_n + kA = C \cdot C^T$. By denoting $Y = C^T \cdot X$, it follows that

$$J(X) = X^T \cdot C \cdot C^T \cdot X = Y^T \cdot Y = \| Y \|^2 .$$

Thus, we have to minimize $\| Y \|$ with a restriction by the form $D \cdot Y = b$; hence $Y = D^+ \cdot b$.

Let us now fix an open set $U \subset \mathbf{R}^n$ and an integer $k \geq 1$. For a map $f : U \to \mathbf{R}^k$ of class $C^1(U)$, $f = (f_1, \ldots, f_k)^T$ and for any point $x \in U$, we consider the jacobian matrix $J_f(x) = \left(\frac{\partial f_i}{\partial x_j} \right)$; $1 \leq i \leq k$, $1 \leq j \leq n$. The most important case is $k < n$, when a point $x \in U$ is called regular if rank $J_f(x) = k$. We will call *geometrical pattern* any set of type $\Gamma = f^{-1}(0)$, i.e. $\Gamma = \{x \in U \mid f_1(x) = 0, \ldots, f_k(x) = 0\}$, or finite unions of such sets of points, supposed regular. These sets generalize alike plane curves, surfaces, hypersurfaces and differentiable manifolds; see [9].

If $a \in U$, $a = (a_1, \ldots, a_n)^T$ is fixed („point of observation") and Γ is a geometrical pattern as above, following [5; 10]. We define the distance

$$\delta(a, \Gamma) = \| J_f(a)^+ \cdot f(a) \| \text{ (euclidian norm).} \quad (2)$$

Justification. Take $k = 1$ and $c = (c_1, \ldots, c_n)^T$ a nonnull constant vector. If $f(x) = \sum_{i=1}^{n} c_i, x_i$, then for any $a \in \mathbf{R}^n$, $f(a) = c^T \cdot a$, hence $J_f(a) = c^T$ and $J_f(a)^+ = \frac{c}{\| c \|^2}$. If $p = a - J_f(a)^+ \cdot f(a)$, then $p = a - \frac{1}{\| c \|^2}(c - c^T \cdot a)$. Thus $f(p) = c^T \cdot p = 0$ and the vector $a - p$ is normal to the hyperplane $\Gamma = f^{-1}(0)$. To conclude with, $\delta(a, \Gamma) = \| a - p \|$, as expected (Fig. 2).

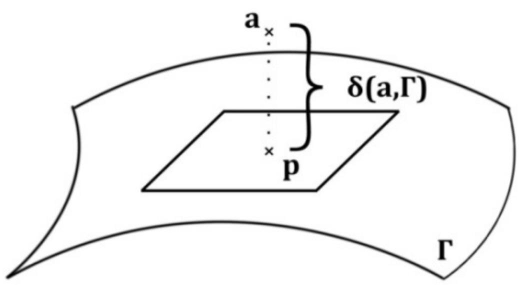

Fig. 2

PROPOSITION 1. *Suppose that* $1 \leq k < n$. *Let* $a \in U$ *be a regular point for a map* $f : U \to \mathbf{R}^k$ *of class* $C^1(U)$. *If* $\Gamma = f^{-1}(0)$, $J = J_f(a)$ *and* $K = J \cdot J^T$, *then* $\delta(a, \Gamma) = (f(a)^T \cdot K^{-1} \cdot f(a))^{1/2}$.

Proof. The rank of the matrix J is k, maximum; thus, K is nonsingular. Moreover, $J^+ = J^T \cdot (J \cdot J^T)^{-1}$. Then by (2),

$$\delta(a, \Gamma)^2 = \| J^+ \cdot f(a) \|^2 = \left\langle J^+ \cdot f(a), J^+ \cdot f(a) \right\rangle = f(a)^T \cdot (J^+) \cdot J^+ \cdot f(a) =$$

$$= f(a)^T \cdot (J \cdot J^T)^{-1} \cdot J \cdot J^T \cdot (J \cdot J^T)^{-1} \cdot f(a).$$

Since $J \cdot J^T \cdot (J \cdot J^T)^{-1} = \mathbf{I}_n$, it will follow that $\delta(a, \Gamma)^2 = f(a)^T \cdot (J \cdot J^T)^{-1} \cdot f(a) = f(a)^T \cdot K^{-1} \cdot f(a)$.

4. GEOMETRICAL PATTERN OF REGRESSION

Let $U \subset \mathbf{R}^n$ be an open set and $\mathcal{A} = \{a_1, \ldots, a_N\} \subset U$ be a finite set of points of observation. Let $f : U \to \mathbf{R}^k$, $1 \le k < n$ be a function of class C^1 such that the points $a_1, \ldots, a_N$ are regular. For $\Gamma = f^{-1}(0)$, define the "distance" [10]

$$\delta(\mathcal{A}, \Gamma) = \sum_{p=1}^{N} \delta(a_p, \Gamma)^2 . \tag{3}$$

The main goal of this paper is *to determine a suitable function f which minimize* $\delta(\mathcal{A}, \Gamma)$; see fig. 3.

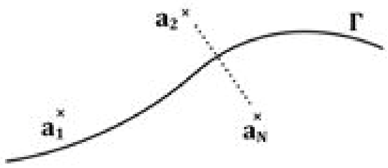

Fig. 3

Such a problem could have applications in Pattern Recognition, Robotics and Computer Vision [5]. The proposed solution is not unique and requires supplementary conditions. Our approximative solution applies the Gauss idea, by checking the unknown *f* in a parametrized family of functions, that could be polynomials or spline-functions.

Let $\Delta = \mathbf{R}^M$ be an open subset in a space of parameters and $F : U \times \Delta \to \mathbf{R}^k$ be a map such that for any $\lambda \in \Delta$, $\lambda = (\lambda_1, \ldots, \lambda_M)^T$, F determines a function of class C^1, $f : U \to \mathbf{R}^k$, $x \mapsto F(x, \lambda)$. Suppose that there is a bijective correspondence between these functions and parameters. This is the case of polynomial functions of degree ≤ *d*, where $\Delta = \mathbf{R}^n$ and $M = \binom{n+d}{n}$, as shown in §1; in that case, the function $\delta(\mathcal{A}, \Gamma)$ becomes a function of coefficients of the polynomials. In the case of linear dependence of parameters, one can impose some suplementary relations between parameters, without modifying solution. We will put the following condition :

$$\sum_{p=1}^{N} J_f(a_p) \cdot J_f(a_p)^T = \mathbf{I}_k . \tag{4}$$

Since the matrixes $J_f(a_p) \cdot J_f(a_p)^T$ are symetrical, positively semi – definite and nonsingular (to points a_p being regular for *f*), they are in fact positively definite. The same thing is valid for the matrix $C = \sum_{p=1}^{N} J_j(a_p) \cdot J_f(a_p)^T$ hence there is an orthogonal matrix $Q \in M_k(\mathbf{R})$ such that $Q^T \cdot C \cdot Q = \mathbf{I}_k$ and Qf will satisfy (4). Suppose now that the points of observation $a_1, \ldots, a_N$ are „sufficiently near" of Γ, in the sense that $\delta(a_p, \Gamma) \cong \| a_p \|$, for any *p*. According to (3), $\delta(\mathcal{A}, \Gamma) \cong \sum_{p=1}^{N} \| a_p \|^2$. Finally, it is reasonable to assume that there are *r* linearly independent functions $\varphi_1, \ldots, \varphi_r : U \to \mathbf{R}^k$ of class $C^2(U)$, such that the unknown $f = (f_1, \ldots, f_k)^T$ of the problem is a linear combination of them. Thus, $f_i = \sum_{j=1}^{r} p_{ij} \varphi_j$, $1 \le i \le k$, where p_{ij} are real numbers to be determined. Matricially, $f = P^T . \varphi$, where $P = (p_{ij}) \in M_{rk}(\mathbf{R})$. The problem reverts to find the function f (i.e. P), such that the sum $\sum_{j=1}^{N} \| f(a_p) \|^2$ is minimum, with the restriction (4). In this case, $\sum_{j=1}^{N} \| f(a_p) \|^2 = \sum_{j=1}^{N} f(a_p)^T \cdot f(a_p) = \sum_{j=1}^{N} \varphi(a_p)^T \cdot P \cdot P^t \cdot \varphi(a_p) = \mathrm{tr}\,(P^T \cdot A \cdot P)$, where $A = \sum_{j=1}^{N} \varphi(a_p) \cdot \varphi(a_p)^T \in M_r(\mathbf{R})$ is a symmetrical and positively semi – definite matrix. On the other hand, $J_f(a_p) \cdot J_f(a_p)^T = P^T \cdot J_\varphi(a_p) \cdot J_\varphi(a_p)^T \cdot P$; moreover, by putting $B = \sum_{p=1}^{N} J_\varphi(a_p) \cdot J_\varphi(a_p)^T$, the matrix $B \in M_r(\mathbf{R})$ is known and the condition (4) becomes $P^T \cdot B \cdot P = \mathbf{I}_k$. One can assume that the matrix $P^T \cdot A \cdot P$ is diagonal [indeed, it is symetrical and positively semi – definite, hence there is an orthogonal matrix $B \in M_k(\mathbf{R})$, i.e. $Q \cdot Q^T = \mathbf{I}_k$, such that $(Q^T) \cdot (P^T \cdot A \cdot P) \cdot Q$ is diagonal; moreover, $\mathrm{tr}\,(P^T \cdot A \cdot P) = = \mathrm{tr}\,((P \cdot Q)^T \cdot A \cdot P \cdot Q)$ and much more, $P^T \cdot B \cdot P = \mathbf{I}_k$ iff $(P \cdot Q)^T \cdot B \cdot P \cdot Q = \mathbf{I}_k$. As such, P can be replaced by $P \cdot Q$]. It remains to apply the following result from linear algebra [3, 10].

The above results can be synthesized in the following, *algorithm* for solving the already formulated problem (see also [9]).

PROPOSITION 2. *Let* $U \subset \mathbf{R}^n$ *be an open set and* $\mathcal{A} = \{a_1, \ldots, a_N\} \subset U$ *be a finite set of observation points. In order to determine a function* $f : U \to \mathbf{R}^k$ *such that the geometrical pattern* $f^{-1}(0)$ *is the nearest to the set* $\mathcal{A}$ *do apply the next steps:*

Step 1. Choose r linear independent functions $\varphi_i : U \subset \mathbf{R}^k$, $1 \le i \le r$ of class $C^2(U)$.

Step 2. Determine the matrices $A, B \in M_r(\mathbf{R})$, $A = \sum_{p=1}^{N} \varphi(a_p) \cdot \varphi(a_p)^T$ and $B = \sum_{p=1}^{N} J_\varphi(a_p) \cdot J_\varphi(a_p)^T$, where $\varphi = (\varphi_1, \ldots, \varphi_r)^T$.

Step 3. Consider k proper versors $x_1, \ldots, x_k$ of the sheaf $A - \alpha B$ and $P = (x_1 | \ldots | x_k)$ verifies the relation $P^T \cdot B \cdot P = \mathbf{I}_k$.

Step 4. The components $f_1, \ldots, f_k$ of the checked function are linear combinations of the functions $\varphi_1, \ldots, \varphi_r$; more exactly, $(f_1, \ldots, f_k) = (\varphi_1, \ldots, \varphi_r).P$.

Note. As soon as the matrices A and B are determined after choosing φ_i, the algoririthm requires $O(r^3)$ operations.

5. CONCLUSIONS

Lagrange interpolation and the Gauss regression are widely used in processing numerical data. In the same time, an important problem is to fit a given set of points and determine a regular geometrical object which is the nearest to some points of observation, (in the sense of a distance defined in terms of the pseudoinverse of jacobian matrices). In this paper we propose such an approach and suggest that this geometrical object could be used to follow the time evolution of the information near the set of observation points. It remains to be seen if this construction is stable and feasible for applications. An algorithm to minimize the distance is formulated in Proposition 2.

REFERENCES

1. R. Gabriel, *Enciphering – maps with pseudoinverses and pseudo-tabulations*, Rev. Roum. Math. Pures et Appl., **LXI**, *1*, pp. 13–18, 2016.
2. G. Glaser, *Etude de qulques algebres tayloriennes*, J. d'Analyse Mathématique, **6**, pp. 1–12, 1958.
3. A.I. Kostrikin, Yu I. Main, *Lineinaya algebra i geometriya*, Nauka Pablishers, Moscow, 1986.
4. G. Marchuk, Méthodes de calcul numerique, Mir, 1980.
5. G.N. Newsam, N.J. Redding, *Fitting the most likely curve through noisy points*, Surveillance Research Lab., Edinborough, Australia, DSTO-RR 0242, 2002.
6. V. Neagoe, O. Stănăşilă, *Teoria recunoaşterii formelor*, Edit. Academiei Române, 1992.
7. A. Niță, *Generalized inverse of a matrix with applications to optimization of some systems* (in Romanian), Ph.D. Thesis, Univ. of Bucharest, Faculty of Mathematics and Computer Sciences, 2004.
8. R. Penrose, *A generalized inverse for matrices*, Proc. Cambridge Phil. Soc, **51**, 1955, pp. 406–413.
9. O. Stănăşilă, *On a geometrical interpolation problem*, Balkan Journal of Geometry, **1**, *2*, pp. 97–103, 1996.
10. G. Taubin, *Nonplanar curve and surface estimation in 3-space,* Proc. IEEE Conf. Robotics-Automation, 1988.

Received July 27, 2016