UNIVERSITATEA Petru Maior din TÎRGU MUREŞ

Scientific Bulletin of the „Petru Maior" University of Tîrgu Mureş
Vol. 9 (XXVI) no. 2, 2012, ISSN 2285 – 438X (Online), ISSN–L 1841 – 9267

THE GRADIENT METHOD FOR OVERDETERMINED INFINITE LINEAR SYSTEMS

Béla Finta
"Petru Maior" University of Tg. Mureş, Romania
e-mail: fintab@science.upm.ro

ABSTRACT

It is well known the notion of overdetermined finite linear system and its solution in the sense of the least squares method (see for example [1], [7] or [8]). In [5] we presented the gradient method for overdetermined finite linear systems and in [6] we made the comparative efficiencies of the least square method and the gradient method for finite overdetermined linear systems. In [3] we considered the notion of overdetermines infinite linear system and its solution in the sense of the least squares method. In [4] we showed the complex variant of this result. The purpose of this paper is to extend the classical gradient method to overdetermined infinite linear systems.

Keywords: *gradient method, least squares method, infinite linear systems, overdetermined infinite linear systems, space* l^2

1 Introduction

Let $z = (z_i)_{i\in\mathbb{N}}$ be a sequence of real numbers. We remember the real Hilbert space $l^2 = \{z \mid \sum_{i=0}^{\infty} z_i^2$ is finite$\}$, with the standard scalar product $< z, t >_\infty = \sum_{i=0}^{\infty} z_i t_i$, where $z = (z_i)_{i\in\mathbb{N}} \in l^2$ and $t = (t_i)_{i\in\mathbb{N}} \in l^2$. We receive the norm of $z \in l^2$ by the formula $\|z\|_\infty = \sqrt{< z, z >_\infty} = \sqrt{\sum_{i=0}^{\infty} z_i^2}$. The elements of the space l^2 we call vectors. We have $z = t$ iff $z_i = t_i$ for all $i \in \mathbb{N}$.

Next we will consider the vectors $z = (z_i)_{i\in\mathbb{N}} \in l^2$ like column matrices. We take the transpose of z, denoted by z^T, as a row matrix. Now we define for $z, t \in l^2$ the product of the row matrix z^T and the column matrix t by the formula $z^T \cdot t = \sum_{i=0}^{\infty} z_i t_i$. So $< z, t >_\infty = z^T \cdot t$, i.e. the scalar product of two vectors can be obtained as a matrix product.

Let us consider for $j = \overline{1, n}$ the column vectors $a_j = (a_{ij})_{i\in\mathbb{N}} \in l^2$. Using these vectors we can form the infinite matrix $A = (a_1 a_2 \ldots a_n) = (a_{ij})_{i\in\mathbb{N}, j=\overline{1,n}}$ with infinite, but numerable rows and with finite number of columns.

Next we extend in natural and similar way the usual matrix operations from finite matrices to infinite matrices. We define the product of the matrix A with the finite column matrix $x = (x_1 x_2 \ldots x_n)^T \in \mathbb{R}^n$ by $A \cdot x = c$, where $c = (c_0 c_1 \ldots c_n \ldots)^T$ is an infinite column matrix with elements $c_i = \sum_{j=1}^{n} a_{ij} \cdot x_j$ for all $i \in \mathbb{N}$. The transpose of the matrix A is $A^T = \begin{pmatrix} a_1^T \\ a_2^T \\ \vdots \\ a_n^T \end{pmatrix} = (a_{ji})_{j\in\overline{1,n}, i\in\mathbb{N}}$, i.e. a matrix with finite number of rows and infinite, but numerable columns. We define the product of the matrix A^T with the column matrix $z \in l^2$ as $A^T \cdot z = (d_1 d_2 \ldots d_n)^T \in \mathbb{R}^n$, where $d_j = \sum_{i=0}^{\infty} a_{ji} \cdot z_i$ for all $j = \overline{1, n}$. Because $a_j \in l^2$ and $z \in l^2$ we get $d_j = a_j^T \cdot z = < a_j, z >_\infty \in \mathbb{R}$ for every $j = \overline{1, n}$. We define the matrix product $A^T \cdot A = (g_{jk})_{j,k=\overline{1,n}}$, where $g_{jk} = \sum_{i=0}^{\infty} a_{ji} \cdot a_{ik}$ for $j, k = \overline{1, n}$. We mention that there exist the real numbers g_{jk}, because $g_{jk} = a_j^T \cdot a_k = < a_j, a_k >_\infty \in \mathbb{R}$ and $A^T \cdot A$ is a finite matrix with n rows and n columns.

Now we are able to define the overdetermined infinite linear system:

$$\begin{cases} a_{01}x_1 + a_{02}x_2 + \ldots + a_{0n}x_n = b_0 \\ a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \\ \vdots \end{cases} \tag{1}$$

where $x_1, x_2, \ldots, x_n \in \mathbb{R}$ are the unknowns of the linear system and the free term is $b = (b_i)_{i \in \mathbb{N}} \in l^2$. We can take also the matrix form $A \cdot x = b$, where $x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$. We say that $x^* = (x_1^* x_2^* \ldots x_n^*)^T \in \mathbb{R}^n$ is a solution of the overdetermined infinite linear system (1), if we have $A \cdot x^* = b$. We can observe immediately that the system generally doesn't have solution. In order to obtain a solution for the system (1), first we build the functions $g : \mathbb{R}^n \to l^2$, $g(x) = A \cdot x - b$ and $f : \mathbb{R}^n \to \mathbb{R}, f(x) = \|g(x)\|_\infty^2 = \|A \cdot x - b\|_\infty^2$ like in the case of the least squares approach. The above presented functions are well defined, because $A \cdot x - b = \sum_{j=1}^{n} x_j \cdot a_j - b \in l^2$. We consider such array $\overline{x} = (\overline{x}_1 \overline{x}_2 \ldots \overline{x}_n)^T \in \mathbb{R}^n$ for which the function f takes the minimal value. The array $\overline{x} \in \mathbb{R}^n$, which minimizes the function f, it is accepted like the solution of the overdetermined infinite linear system $A \cdot x = b$ in the sense of the least squares method.

In [3] we showed the following result: if $(A^T \cdot A) \cdot \overline{x} = A^T \cdot b$, then $\|A \cdot \overline{x} - b\|_\infty \leq \|A \cdot x - b\|_\infty$ for all $x \in \mathbb{R}^n$. We mention that $A^T \cdot A$ is a finite, well defined matrix of order n, so the system $(A^T \cdot A) \cdot \overline{x} = A^T \cdot b$ is a finite linear system, which can be solved using the numerical methods of linear algebra. In [4] we presented the complex variant of this result. In [5] we realized the extension of the gradient method for finite overdetermined linear systems. In [6] we made the comparative efficiencies of the least square method and the gradient method for finite overdetermined linear systems.

2 Main part

The aim of this paper is to show the gradient method for the function f in order to obtain a minimal point $\overline{x} \in \mathbb{R}^n$. We consider $\overline{x}$ like a solution of the system (1) in the sense of the least squares approach.

First of all we must calculate the gradient of the function $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \|A \cdot x - b\|_\infty^2 = \sum_{i=0}^{\infty} (\sum_{j=1}^{n} a_{ij} \cdot x_j - b_i)^2$, which is a quadratic form in the unknowns $x_1, x_2, \ldots, x_n$. We have the following calculus of gradient:

$$\begin{aligned} \text{grad} f(x) &= \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(x) \\ \dfrac{\partial f}{\partial x_2}(x) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(x) \end{pmatrix} = \\ &\begin{pmatrix} \sum_{i=0}^{\infty} \left[2 \cdot \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right) \cdot a_{i1} \right] \\ \sum_{i=0}^{\infty} \left[2 \cdot \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right) \cdot a_{i2} \right] \\ \vdots \\ \sum_{i=0}^{\infty} \left[2 \cdot \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right) \cdot a_{in} \right] \end{pmatrix} = \\ &= 2 \cdot A^T \cdot (A \cdot x - b). \end{aligned}$$

Let us choose $x^0 = (x_1^0, x_2^0, \ldots, x_n^0) \in \mathbb{R}^n$ the start point for the gradient method, and we suppose that we determined the point $x^k = (x_1^k, x_2^k, \ldots, x_n^k) \in \mathbb{R}^n$ and we want to find the next point $x^{k+1} = (x_1^{k+1}, x_2^{k+1}, \ldots, x_n^{k+1}) \in \mathbb{R}^n$. Let $F_k : [0, +\infty) \to \mathbb{R}$, $F_k(t) = f(x^k - t \cdot \text{grad} f(x^k))$ be such function, for which we calculate the value $t_k \in [0, +\infty)$ in order to obtain the minimal value of the function F_k in the point t_k. We have:

$$\begin{aligned} F_k(t) &= \|A \cdot (x^k - t \cdot \text{grad} f(x^k)) - b\|_\infty^2 = \\ &= \|(A \cdot x^k - b) - t \cdot A \cdot \text{grad} f(x^k)\|_\infty^2 = \\ &= \|g(x^k) - t \cdot A \cdot \text{grad} f(x^k)\|_\infty^2 = \\ &= \sum_{i=0}^{\infty} \left[g_i(x^k) - t \cdot \sum_{j=1}^{n} \left(a_{ij} \cdot \frac{\partial f}{\partial x_j}(x^k) \right) \right]^2. \end{aligned}$$

We calculate:

$$\begin{aligned} F_k'(t) &= \sum_{i=0}^{\infty} 2 \cdot \left[g_i(x^k) - t \cdot \sum_{j=1}^{n} \left(a_{ij} \cdot \frac{\partial f}{\partial x_j}(x^k) \right) \right] \\ &\cdot (-1) \cdot \left[\sum_{j=1}^{n} \left(a_{ij} \cdot \frac{\partial f}{\partial x_j}(x^k) \right) \right]. \end{aligned}$$

From the equation $F_k'(t_k) = 0$ we get:

$$t_k = \frac{\sum_{i=0}^{\infty} \left[g_i(x^k) \cdot \sum_{j=1}^{n} \left(a_{ij} \cdot \frac{\partial f}{\partial x_j}(x^k) \right) \right]}{\sum_{i=0}^{\infty} \left[\sum_{j=1}^{n} \left(a_{ij} \cdot \frac{\partial f}{\partial x_j}(x^k) \right) \right]^2}$$

We denote for every $i \in \mathbb{N}$ with $a_i^* = (a_{i1}, a_{i2}, \ldots, a_{in})$ the rows of the matrix A and with

$< \cdot, \cdot >_n$ the Euclidean scalar product on $\mathbb{R}^n$. Then:

$$
\begin{aligned}
t_k &= \frac{\sum_{i=0}^{\infty} \left[g_i(x^k) \cdot < a_i^*, \ \mathrm{grad} f(x^k) >_n \right]}{\sum_{i=0}^{\infty} \left[< a_i^*, \ \mathrm{grad} f(x^k) >_n \right]^2} = \\
&= \frac{< g(x^k), A \cdot \ \mathrm{grad} f(x^k) >_\infty}{< A \cdot \ \mathrm{grad} f(x^k), A \cdot \ \mathrm{grad} f(x^k) >_\infty} = \\
&= \frac{< A \cdot x^k - b, A \cdot \ \mathrm{grad} f(x^k) >_\infty}{< A \cdot \ \mathrm{grad} f(x^k), A \cdot \ \mathrm{grad} f(x^k) >_\infty}
\end{aligned}
$$

We can substitute in this last formula $\mathrm{grad} f(x^k)$ by $2 \cdot A^T \cdot (A \cdot x^k - b)$, and the scalar 2 we take from the scalar product. We can observe, that $t_k \geq 0$, and we denote:

$$
\alpha_k = \frac{< A \cdot x^k - b, A \cdot A^T \cdot (A \cdot x^k - b) >_\infty}{< A \cdot A^T \cdot (A \cdot x^k - b), A \cdot A^T \cdot (A \cdot x^k - b) >_\infty}
$$

Hence the next point with the gradient method we obtain by the formula:

$$
\begin{aligned}
x^{k+1} &= x^k - t_k \cdot \ \mathrm{grad} f(x^k) = \\
&= x^k - \frac{1}{2} \cdot \alpha_k \cdot 2[A^T \cdot (A \cdot x^k - b)] = \\
&= x^k - \alpha_k \cdot A^T \cdot (A \cdot x^k - b)
\end{aligned}
$$

We mention that the order of convergence of this method we will study in a next paper.

References

[1] Rao, S. S. (2002), Applied Numerical Methods for Engineers and Scientists, Prentice Hall, Upper Saddle River, New Jersey.

[2] Hansen, A. C. , Infinite Dimensional Numerical Linear Algebra; Theory and Applications, http://www.damtp.cam.ac.uk/user/na/people /Anders/Applied3.pdf

[3] Finta, B. (2009), Overdetermined Infinite Linear Systems, Numerical Analysis and Applied Mathematics, International Conference on Numerical Analysis and Applied Mathematics, ICNAAM 2009, Rethymno, Crete, Greece, 18-22 September 2009, Volume 2, AIP Conference Proceedings 1168, ISBN 978-0-7354-0708-4, ISSN 0094-243X, pp. 1035-1037.

[4] Finta, B. (2011), Complex Overdetermined Infinite Linear Systems, International Conference on Numerical Analysis and Applied Mathematics, ICNAAM 2011, Halkidiki, Greece, 19-25 September 2011, AIP Conference Proceedings 1389, ISBN 978-0-7354-0954-5, pp. 1637-1639.

[5] Finta, B. (2010), The Gradient Method for Overdetermined Linear Systems, Scientific Bulletin of the "Petru Maior" University of Târgu Mureş, Romania, 2010, Volume 7 (XXIV), New Series, Number 1, pp. 24-26, Published by Editura Universităţii "Petru Maior", Târgu Mureş, Romania, ISSN 1841-9267.

[6] Finta, B. (2010), Comparative Efficiences of the Least Squares Method and the Gradient Method for Overdetermined Linear Systems, Scientific Bulletin of the "Petru Maior" University of Târgu Mureş, Romania, 2010, Volume 7 (XXIV), New Series, Number 2, pp. 35-36, Published by Editura Universităţii "Petru Maior", Târgu Mureş, Romania, ISSN 1841-9267.

[7] Finta, B. (2004), Numerical Analysis, Editura Universităţii "Petru Maior", Tg. Mureş.

[8] Finta, B. (2004), Numerical Analysis, http:// cs.upm.ro/∼ bela.finta/.files/carti/Numerika.pdf

[9] Stoer, J., Bulirsch, R. (2002), Introduction to numerical analysis, translated from the German by R. Bartels, W. Gautschi and C. Witzgall, 3d ed. Texts in Applied Mathematics, Springer, New York, vol. 12.

[10] Wolfe, M. A. (1978), Numerical methods for unconstrained optimization: An introduction, Van Nostrand Reinhold, New York.

[11] Han, W. and Atkinson, E. K. (2009), Theoretical Numerical Analysis, 3th. ed. Texts in Applied Mathematics, Berlin, Germany; Springer, vol. 39.

[12] Quarteroni, A., Sacco, R., Salieri, F., Numerical Mathematics, 2nd. ed. Texts in Applied Mathematics, Berlin, Germany: Springer, vol. 37.

[13] Bonnans, J. B., Gilbert, C., Lemarchal, C., Sagastizbal, C. A. (2006), Numerical Optimization, Theoretical and Practical Aspects, 2d. ed., Berlin, Germany, Springer.

[14] Sun, W., Yuan, Y. (2006) Optimization Theory and Methods, Nonlinear Programming, Optimization and its Applications, Berlin, Germany, Springer, vol. 1.

[15] Snyman, J. A (2005), Practical Mathematical Optimization, An Introduction to Basic Optimization Theory and Classical and New Gradient Based Algorithms, Applied Optimization, Springer Science + Business Media, Inc..

[16] Wang, Y., Yagola, A. G., Yang, C. (2010), Optimization and Regularization for Computational Inverse Problems and Applications, Higher Education Press, Beijing, Heidelberge, Springer.