UNIVERSITATEA Petru Maior TÂRGU-MUREŞ

Scientific Bulletin of the „Petru Maior" University of Târgu Mureş
Vol. 7 (XXIV) no. 1, 2010, ISSN 1841-9267

THE GRADIENT METHOD FOR OVERDETERMINED LINEAR SYSTEMS

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

ABSTRACT

The purpose of this paper is to extend the classical gradient method, known for linear systems of type Cramer, to overdetermined linear systems

Keywords: gradient method, overdetermined linear system, least squares method

1 Introduction

Let us consider the real matrix $A = (a_{ij})_{\substack{i=\overline{1,m}\\ j=\overline{1,n}}}$, and the real, transposed arrays $x = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$ and $b = (b_1, b_2, \dots, b_m)^T \in \mathbb{R}^m$, respectively. The linear system $A \cdot x = b$ is called overdetermined linear system, if $m > n$. Generally, the overdetermined linear system is incompatible, i.e. doesn't exist an array $x^* = (x_1^*, x_2^*, \dots, x_n^*)^T \in \mathbb{R}^n$ such that $A \cdot x^* = b$. For this reason, instead of the classical solution x^*, we consider such array $\overline{x} = (\overline{x}_1, \overline{x}_2, \dots, \overline{x}_n) \in \mathbb{R}^n$ for which the function $f : \mathbb{R}^n \to \mathbb{R}$, $f(x) = \|A \cdot x - b\|_m^2$ takes the minimal value, where $\|\cdot\|_m$ means the Euclidean norm on the space $\mathbb{R}^m$. The array $\overline{x} \in \mathbb{R}^n$, which minimizes the function f, it is accepted like the solution of the overdetermined linear system $A \cdot x = b$ in the sense of the least squares method. We can observe that $f(x) \geq 0$ for all $x \in \mathbb{R}^n$ and the minimal point $\overline{x} \in \mathbb{R}^n$ verifies the following system with partial derivatives $\frac{\partial f}{\partial x_k}(\overline{x}) = 0$ for every $k = \overline{1,n}$. We calculate the partial derivatives and doing the corresponding calculus, we obtain that $\overline{x} \in \mathbb{R}^n$ is the solution of the linear system $(A^T \cdot A) \cdot x = A^T \cdot b$, which is a Cramer's type linear system with n equations and n unknowns, so $\overline{x} \in \mathbb{R}^n$ will be the classical solution of this Cramer's linear system. We also mention the following statement: if $\overline{x} \in \mathbb{R}^n$ is the classical solution of the linear system $(A^T \cdot A) \cdot x = A^T \cdot b$, i.e. $(A^T \cdot A) \cdot \overline{x} = A^T \cdot b$, then $f(\overline{x}) = \|A \cdot \overline{x} - b\|_m^2 \leq \|A \cdot x - b\|_m^2 = f(x)$ for all $x \in \mathbb{R}^n$. So $\overline{x} \in \mathbb{R}^n$ is the solution of the overdetermined linear system $A \cdot x = b$ in the sense of the least squares approache, see [1] and [2].

2 Main part

The linear system $(A^T \cdot A) \cdot x = A^T \cdot b$ can be solved by Cramer's rule from theory of determinants for $n \in \mathbb{N}^*$ small natural numbers, but in other cases, for $n \in \mathbb{N}^*$ great natural numbers we can use numerical methods of linear algebra. We mention here, that one way to solve numerically the linear system $(A^T \cdot A) \cdot x = A^T \cdot b$ is the gradient method, see [1] and [2]. The aim of this paper is not to show the gradient method for the linear system $(A^T \cdot A) \cdot x = A^T \cdot b$, and more generally for a linear system with n equations and n unknowns, which is well known, but to deduce the gradient method for the overdetermined linear system $A \cdot x = b$. We take the functions $g : \mathbb{R}^n \to \mathbb{R}^m$, $g(x) = A \cdot x - b$ and $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(x) = \|A \cdot x - b\|_m^2 = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right)^2 ,$$

and we will apply the gradient method for the function f in order to obtain the minimal point $\overline{x} \in \mathbb{R}^n$. We have the following calculus of gradient:

$$\operatorname{grad} f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix} =$$

$$\begin{pmatrix} \sum_{i=1}^{m}\left[2\cdot\left(\sum_{j=1}^{n}a_{ij}x_j-b_i\right)\cdot a_{i1}\right] \\ \sum_{i=1}^{m}\left[2\cdot\left(\sum_{j=1}^{n}a_{ij}x_j-b_i\right)\cdot a_{i2}\right] \\ \vdots \\ \sum_{i=1}^{m}\left[2\cdot\left(\sum_{j=1}^{n}a_{ij}x_j-b_i\right)\cdot a_{in}\right] \end{pmatrix} = $$
$$= \quad 2\cdot A^T\cdot(A\cdot x-b).$$

Let us choose $x^0=(x_1^0,x_2^0,\dots,x_n^0)\in\mathbb{R}^n$ the start point for the gradient method, and we suppose that we determined the point $x^k=(x_1^k,x_2^k,\dots,x_n^k)\in\mathbb{R}^n$ and we want to find the next point $x^{k+1}=(x_1^{k+1},x_2^{k+1},\dots,x_n^{k+1})\in\mathbb{R}^n$. 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:

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

We calculate:

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

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

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

We denote for every $i=\overline{1,m}$ with $a_i=(a_{i1},a_{i2},\dots,a_{in})$ the rows of the matrix A and with $<\cdot,\cdot>_n$ the Euclidean scalar product on $\mathbb{R}^n$. Then:

$$\begin{aligned} t_k &= \frac{\sum_{i=1}^{m}\left[g_i(x^k)\cdot<a_i,\ \text{grad}f(x^k)>_n\right]}{\sum_{i=1}^{m}\left[<a_i,\ \text{grad}f(x^k)>_n\right]^2}= \\ &= \frac{<g(x^k),A\cdot\ \text{grad}f(x^k)>_m}{<A\cdot\ \text{grad}f(x^k),A\cdot\ \text{grad}f(x^k)>_m}= \\ &= \frac{<A\cdot x^k-b,A\cdot\ \text{grad}f(x^k)>_m}{<A\cdot\ \text{grad}f(x^k),A\cdot\ \text{grad}f(x^k)>_m} \end{aligned}$$

We can substitute in this last formula $\text{grad}f(x^k)$ 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:

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

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

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

Example 1. *Let us consider the following overdetermined linear system:* $\begin{cases} x+y=2 \\ x+2y=3 \\ 2x+y=4 \end{cases}$

If we solve the linear system $\begin{cases} x+y=2 \\ x+2y=3 \end{cases}$ then we receive the solution $x=y=1$, which doesn't verify the last equation: $2x+y=3\neq 4$. We obtain the same conclusion if we calculate the characteristic determinant: $\begin{vmatrix} 1 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 1 & 4 \end{vmatrix}=1\neq 0$. So our overdetermined linear system is incompatible and does not have classical solution. Next we calculate the solution of this system in the sense of the least squares approach. Let us consider the function $f:\mathbb{R}^2\to\mathbb{R}$, $f(x,y)=(x+y-2)^2+(x+2y-3)^2+(2x+y-4)^2$ and for the linear system $\begin{cases} \frac{\partial f}{\partial x}(x,y)=0 \\ \frac{\partial f}{\partial y}(x,y)=0 \end{cases}$ we obtain: $\begin{cases} 6x+5y=13 \\ 5x+6y=12. \end{cases}$ In another way we get $m=3$, $n=2$, $A=\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 2 & 1 \end{pmatrix}$, $b=\begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$, so $A^T\cdot A=\begin{pmatrix} 6 & 5 \\ 5 & 6 \end{pmatrix}$ and $A^T\cdot b=\begin{pmatrix} 1 & 3 \\ 1 & 2 \end{pmatrix}$. Hence our system $A^T\cdot A\cdot x=A^Tb$ is the same: $\begin{pmatrix} 6 & 5 \\ 5 & 6 \end{pmatrix}\cdot\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} 13 \\ 12 \end{pmatrix}$, with classical solution $x=\frac{18}{11}$ and $y=\frac{7}{11}$.

So our overdetermined linear system admits the solution $x=\frac{18}{11}$ and $y=\frac{7}{11}$ in the sense of the least squares method. Next we apply for this overdetermined linear system $A\cdot x=b$ the above shoved gradient method. If $x^0=\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, then $A\cdot x^0-b=\begin{pmatrix} -1 \\ -2 \\ -2 \end{pmatrix}$, $A\cdot A^T\cdot(A\cdot x^0-b)=\begin{pmatrix} -14 \\ -21 \\ -21 \end{pmatrix}$, $\alpha_0=\frac{1}{11}$ and $x^1=x^0-\alpha_0\cdot A^T(A\cdot x^0-b)=\begin{pmatrix} \frac{18}{11} \\ \frac{7}{11} \end{pmatrix}$, so we get the same solution. If $x^0=\begin{pmatrix} 2 \\ 1 \end{pmatrix}$, then $A\cdot x^0-b=\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$,

$A \cdot A^T \cdot (A \cdot x^0 - b) = \begin{pmatrix} 8 \\ 12 \\ 12 \end{pmatrix}$, $\alpha_0 = \frac{1}{11}$ and $x^1 = x^0 - \alpha_0 \cdot A^T(A \cdot x^0 - b) = \begin{pmatrix} \frac{18}{11} \\ \frac{7}{11} \end{pmatrix}$, so we get again the same solution.

3 Conclusions

We saw above that we obtained the same solution for the previous overdetermined linear system by least squares approach and with gradient method using only one step. It is interesting to study, what does happen for arbitrary overdetermined linear systems?

If we choose the least squares method, then we must solve numerically a Cramer's linear system. If we choose the gradient method we must calculate simply and very easy some terms of the iterative sequence $(x^k)_{k\in\mathbb{N}}$ and this last way is more confortable corresponding to the volum of elementary calculus. It is interesting to study and compare the complexity of these two methods.

References

[1] B. Finta, Analiză numerică, Editura Universităţii "Petru Maior", Tg. Mureş, 2004.

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