UNIVERSITATEA Petru Maior TÂRGU-MUREȘ

Scientific Bulletin of the „Petru Maior" University of Tîrgu Mureş
Vol. 8 (XXV), no. 1, 2011, ISSN 1841-9267

THE GRADIENT METHOD FOR OVERDETERMINED NONLINEAR 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 nonlinear systems with n equations and n unknowns, to overdetermined nonlinear systems.

Keywords: gradient method, overdetermined nonlinear system

1 Introduction

It is well known the gradient method for nonlinear systems with n equations and n unknowns, see for example [1] and [2]. The notion of overdetermined nonlinear system and its solution as the best least squares approximate is introduced for example in [3], or more recently for example in [4]. The purpose of this paper is to extend the gradient method for overdetermined nonlinear systems. We mention that in [5] we extended the gradient method, known for linear systems of type Cramer, to overdetermined linear systems. In [6] we do the comparative efficiencies of the least squares method and the gradient method for overdetermined linear systems.

2 Main part

Let $G_1, G_2, \ldots, G_m : D \subset \mathbb{R}^n \to \mathbb{R}$, $D \neq \emptyset$ be given functions and $G = (G_1, G_2, \ldots, G_m) : D \subset \mathbb{R}^n \to \mathbb{R}^m$. We can attache to this function G the following nonlinear system:

$$G(x) = \theta, \tag{1}$$

where $x = (x_1, x_2, \ldots, x_n) \in D$ and $\theta \in \mathbb{R}^m$ is the null element. If $m > n$ then we say that the nonlinear system (1) is overdetermined. Generally the overdetermined nonlinear system does not have solution, i.e. doesn't exist $x^* = (x_1^*, x_2^*, \ldots, x_n^*) \in D$ such that $G(x^*) = \theta$. We build the function $f : D \subset \mathbb{R}^n \to \mathbb{R}$ given by the formula $f(x) = \sum_{i=1}^{m} (G_i(x))^2$. It is obvious that $f(x) \geq 0$ for all $x \in D$. We want to determine $\overline{x} = (\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_n) \in D$ such that $f(\overline{x})$ be a local minimal, i.e. $f(\overline{x}) \leq f(x)$, for all $x \in V \subset D$, where V is an appropiate neighborhood of $\overline{x}$. This point $\overline{x}$ we accept like a solution of the overdetermined nonlinear equation (1) in the same sense of the least squares. Next we present the gradient method for the function f in order to obtain $\overline{x} \in D$.

Let us choose the initial point $x^0 = (x_1^0, x_2^0, \ldots, x_n^0) \in V \subset D$. We suppose that we already realized to obtain the point $x^k = (x_1^k, x_2^k, \ldots, x_n^k) \in V$ and we want to get the next point $x^{k+1} = (x_1^{k+1}, x_2^{k+1}, \ldots, x_n^{k+1}) \in V$ by the gradient method. We consider the function $F_k : [0, M_k) \to \mathbb{R}$ given by the formula:

$$F_k(t) = f(x^k - t \cdot \operatorname{grad} f(x^k)) = \\ = \sum_{i=1}^{m} [G_i(x^k - t \cdot \operatorname{grad} f(x^k)]^2,$$

where

$$\operatorname{grad} f(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right)$$

is the gradient of the function f in point x. Here we fixed the value $M_k \in [0, +\infty) \cup \{+\infty\}$ such that the functions G_i and f are well defined, i.e. $x^k - t \cdot \operatorname{grad} f(x^k) \in V$ for all $t \in [0, M_k)$. We suppose that all functions $G_i : D \subset \mathbb{R}^n \to \mathbb{R}$, $i = \overline{1,m}$ are differentiable on the open subset $D \subset \mathbb{R}^n$. So we can consider the Taylor's expansions of the functions G_i, $i = \overline{1,m}$ in the point x^k and we take only the linear term:

$$G_i(x^k - t\cdot \text{grad}\, f(x^k)) = \\ = G_i(x^k) + DG_i(x^k)(-t\cdot \text{grad}\, f(x^k)),$$

where $DG_i(x^k)$ is the differential of the function G_i at point x^k and $DG_i(x^k)(-t\cdot \text{grad}\, f(x^k))$ is the above described differential function in the value $-t\cdot \text{grad}\, f(x^k)$. Now we use the representation of the differential by partial derivatives:

$$G_i(x^k - t\cdot \text{grad}\, f(x^k)) = \\ = G_i(x^k) + \sum_{j=1}^{n} \frac{\partial G_i}{\partial x_j}(x^k)\left(-t\cdot \frac{\partial f}{\partial x_j}(x^k)\right) = \\ = G_i(x^k) - t\cdot < \text{grad}\, G_i(x^k),\ \text{grad}\, f(x^k) >_n,$$

where $< \cdot,\cdot >_n$: $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is the standard scalar product on the space $\mathbb{R}^n$, given by the formula: $< x, y >_n = \sum_{i=1}^{n} x_i y_i$, with $x = (x_i)_{i=\overline{1,n}} \in \mathbb{R}^n$ and $y = (y_i)_{i=\overline{1,n}} \in \mathbb{R}^n$. So we get $F_k(t) = \sum_{i=1}^{m} [G_i(x^k) - t\cdot < \text{grad}\, G_i(x^k), \text{grad}\, f(x^k) >_n]^2$. Now we want to determine such positive value $t \geq 0$ for which F_k takes the minimal value. We calculate:

$$\frac{dF_k(t)}{dt} = \sum_{i=1}^{m} 2\cdot [G_i(x^k) - \\ - t\cdot < \text{grad}\, G_i(x^k), \text{grad}\, f(x^k) >_n]\cdot \\ \cdot [- < \text{grad}\, G_i(x^k), \text{grad}\, f(x^k) >_n].$$

The minimal value for $t \geq 0$ we obtain from the equation $\frac{dF_k(t)}{dt} = 0$. So:

$$\sum_{i=1}^{m} G_i(x^k)\cdot < \text{grad}\, G_i(x^k), \text{grad}\, f(x^k) >_n - \\ - t\cdot \sum_{i=1}^{m} [< \text{grad}\, G_i(x^k), \text{grad}\, f(x^k) >_n]^2 = 0.$$

At the end we get the solution for $t = t_k$:

$$t_k = \frac{\sum_{i=1}^{m} G_i(x^k)\cdot < \text{grad}\, G_i(x^k), \text{grad}\, f(x^k) >_n}{\sum_{i=1}^{m} [< \text{grad}\, G_i(x^k), \text{grad}\, f(x^k) >_n]^2}.$$

From the equality $f(x) = \sum_{i=1}^{m} (G_i(x))^2$ results:

$$\text{grad}\, f(x) = \begin{pmatrix} \sum_{i=1}^{m} 2\cdot G_i(x)\cdot \frac{\partial G_i}{\partial x_1}(x) \\ \sum_{i=1}^{m} 2\cdot G_i(x)\cdot \frac{\partial G_i}{\partial x_2}(x) \\ \vdots \\ \sum_{i=1}^{m} 2\cdot G_i(x)\cdot \frac{\partial G_i}{\partial x_n}(x) \end{pmatrix} =$$

$$= 2\cdot \begin{pmatrix} \frac{\partial G_1}{\partial x_1}(x) & \frac{\partial G_2}{\partial x_1}(x) & \dots & \frac{\partial G_m}{\partial x_1}(x) \\ \frac{\partial G_1}{\partial x_2}(x) & \frac{\partial G_2}{\partial x_2}(x) & \dots & \frac{\partial G_m}{\partial x_2}(x) \\ \vdots & & & \\ \frac{\partial G_1}{\partial x_n}(x) & \frac{\partial G_2}{\partial x_n}(x) & \dots & \frac{\partial G_m}{\partial x_n}(x) \end{pmatrix}\cdot \\ \cdot \begin{pmatrix} G_1(x) \\ G_2(x) \\ \vdots \\ G_m(x) \end{pmatrix} =$$

$$= 2\cdot \begin{pmatrix} \frac{\partial G_1}{\partial x_1}(x) & \frac{\partial G_1}{\partial x_2}(x) & \dots & \frac{\partial G_1}{\partial x_n}(x) \\ \frac{\partial G_2}{\partial x_1}(x) & \frac{\partial G_2}{\partial x_2}(x) & \dots & \frac{\partial G_2}{\partial x_n}(x) \\ \vdots & & & \\ \frac{\partial G_m}{\partial x_1}(x) & \frac{\partial G_m}{\partial x_2}(x) & \dots & \frac{\partial G_m}{\partial x_n}(x) \end{pmatrix}^T\cdot \\ \cdot \begin{pmatrix} G_1(x) \\ G_2(x) \\ \vdots \\ G_m(x) \end{pmatrix} = \\ = 2\cdot (G'(x))^T \cdot G(x),$$

where $G'(x)$ is the Jacobi matrix of the function G in the point x. Consequently:

$$G'(x) = \begin{pmatrix} \text{grad}\, G_1(x) \\ \text{grad}\, G_2(x) \\ \vdots \\ \text{grad}\, G_m(x) \end{pmatrix}.$$

Therefore, if we take the matrix multiplication:

$$G'(x^k)\cdot \text{grad}\, f(x^k) = \\ = \begin{pmatrix} < \text{grad}\, G_1(x^k),\ \text{grad}\, f(x^k) >_n \\ < \text{grad}\, G_2(x^k),\ \text{grad}\, f(x^k) >_n \\ \vdots \\ < \text{grad}\, G_m(x^k),\ \text{grad}\, f(x^k) >_n \end{pmatrix}$$

then

$$t_k = \frac{< G(x^k), G'(x^k)\cdot \text{grad}\, f(x^k) >_m}{< G'(x^k)\cdot \text{grad}\, f(x^k), G'(x^k)\cdot \text{grad}\, f(x^k) >_m}.$$

But $\text{grad}\, f(x^k) = 2\cdot (G'(x^k))^T \cdot G(x^k)$, so

$$t_k = < G(x^k), 2\cdot G'(x^k)\cdot (G'(x^k))^T\cdot G(x^k) >_m / \\ / < G'(x^k)\cdot 2\cdot (G'(x^k))^T\cdot G(x^k), \\ G'(x^k)\cdot 2\cdot (G'(x^k))^T\cdot G(x^k) >_m = \\ = \frac{1}{2} < G(x^k), G'(x^k)\cdot (G'(x^k))^T\cdot G(x^k) >_m / \\ / < G'(x^k)\cdot (G'(x^k))^T\cdot G(x^k), \\ G'(x^k)\cdot (G'(x^k))^T\cdot G(x^k) >_m .$$

From this formula we can see immediately that $t_k \geq 0$, being the quotient of two positive numbers. We suppose in plus that $t_k \in [0, M_k)$, too. This means that:

$$x^{k+1} = x^k - t_k\cdot \text{grad}\, f(x^k) = \\ = x^k - t_k\cdot 2\cdot (G'(x^k))^T\cdot G(x^k).$$

Let us denote by

$$\alpha_k =< G(x^k), G'(x^k) \cdot (G'(x^k))^T \cdot G(x^k) >_m / \\ / < G'(x^k) \cdot (G'(x^k))^T \cdot G(x^k), \\ G'(x^k) \cdot (G'(x^k))^T \cdot G(x^k) >_m$$

and at the end we obtain the iteration

$$x^{k+1} = x^k - \frac{1}{2} \cdot \alpha_k \cdot 2 \cdot (G'(x^k))^T \cdot G(x^k) = \\ = x^k - \alpha_k \cdot (G'(x^k))^T \cdot G(x^k).$$

3 Conclusion

We can observe that the gradient method presented above for overdetermined nonlinear systems $(m > n)$ is valid for the welldetermined nonlinear systems $(m = n)$ and for the underdetermined nonlinear systems $(m < n)$, too.

Next we consider $G : \mathbb{R}^n \to \mathbb{R}^m$, $G(x) = A \cdot x - b$, where A is a matrix with m rows and n columns, x and b are column matrices with n and m rows, respectively. Then $G'(x) = A$ and $G(x) = \theta$ means $A \cdot x = b$. Hence we obtain for the overdetermined linear system $A \cdot x = b$, $(m > n)$ the solution using the gradient method by the iteration:

$$x^{k+1} = x^k - \alpha_k \cdot A^T \cdot (A \cdot x^k - b),$$

where

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

see [5].

Concerning the order of convergence of the gradient method for overdetermined nonlinear systems we will deduce in the next paper.

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.

[3] R. Fletcher, Practical Methods of Optimisation, Volume 1, Unconstrained Optimisation, John Wiley & Sons, 1980, page 110.

[4] Wenyu Sun, Ya-Xiang Yuan, Optimisation Theory and Methods – Nonlinear Programming, Springer, 2006, page 353.

[5] B. Finta, The Gradient Method for Overdetermined Linear Systems, Scientific Bulletin of the „Petru Maior" University of Tg. Mureş, Vol.7 (XXIV), no. 1, 2010, pp. 24-26, ISSN 1841-9267.

[6] B. Finta, Comparative Efficiencies of the Least Squares Method and the Gradient Method for Overdetermined Linear Systems, Scientific Bulletin of the „Petru Maior" University of Tg. Mureş, Vol.7 (XXIV), no. 2, 2010, pp. 35-36, ISSN 1841-9267.