Aplikace matematiky

Fridrich Sloboda

An imperfect conjugate gradient algorithm

Aplikace matematiky, Vol. 27 (1982), No. 6, 426–432

Persistent URL: http://dml.cz/dmlcz/103989

AN IMPERFECT CONJUGATE GRADIENT ALGORITHM

FRIDRICH SLOBODA

(Received January 20, 1981)

1. INTRODUCTION

Let us consider the system

$$Ax = b \tag{1}$$

where A is a symmetric, positive definite n by n matrix and b is an n-vector. The biorthogonalization process defined by Hestenes and Stiefel [1] consists of constructing two sequences $g_0, g_1, \ldots,$ and $h_0, h_1, \ldots$ in the n-dimensional Euclidean space R_n such that

$$(g_i, g_j) = (Ah_i, h_j) = 0 \quad i \neq j\,.$$

Let $x_0 \in R_n$ be arbitrary. Let $g_i = Ax_i - b$ and let $g_0 = h_0$. For $i = 0, 1, \ldots, n-1$ we have the following process:

$$x_{i+1} = x_i - \alpha_i h_i\,, \tag{2}$$

$$\alpha_i = \frac{(h_i, g_i)}{(Ah_i, h_i)} = \frac{(g_i, g_i)}{(Ah_i, g_i)}\,, \tag{3}$$

$$g_{i+1} = g_i - \alpha_i A h_i\,, \tag{4}$$

$$h_{i+1} = g_{i+1} - \frac{(Ah_i, g_{i+1})}{(Ah_i, h_i)} h_i = g_{i+1} + \frac{(g_{i+1}, g_{i+1})}{(g_i, g_i)} h_i\,, \tag{5}$$

and there exists $m \leqq n$ with $g(x_m) = 0$, i.e., x_m is the solution of (1). The function $f(x) := \frac{1}{2}x^{\mathrm{T}}Ax^{\mathrm{T}} - bx$ is a strictly convex quadratic function. Then α_i defined by (3) has the following geometrical interpretation

$$f(x_i - \alpha h_i) = \min! \quad \text{for} \quad \alpha = \alpha_i = \frac{(h_i, g_i)}{(Ah_i, h_i)}\,, \tag{6}$$

i.e., $(g_{i+1}, h_i) = 0$. Because of (6) we say that the biorthogonalization process (2)–(5) is defined by the minimization principle, i.e., it is a perfect algorithm with

respect to the step size. Algorithm (2)–(5) was generalized for minimization of nonquadratic functions by many authors [2, 3, 4]. These algorithms were also based on exact line searches. Unfortunately, many algorithms of the conjugate gradient type are sensitive to the exactness of the line searches and this phenomenon may destroy the global efficiency of these methods. The larger the dimension of the problem is, the bigger is the influence of this phenomenon. According to the choice of the recurrent formula for h_{i+1} we obtain various algorithms

$$h_{i+1} = g_{i+1} + \beta_i h_i \,, \tag{7}$$

$$\beta_i = \varphi(g_{i+1}, g_i, h_i) = \begin{cases} \dfrac{(g_{i+1}, g_{i+1})}{(g_i, g_i)} \\ -\dfrac{(g_{i+1} - g_i, g_{i+1})}{(g_i, g_i)} \\ -\dfrac{(g_{i+1} - g_i, g_{i+1})}{(g_{i+1} - g_i, h_i)} \end{cases}.$$

The first one is the most sensitive formula with respect to the exactness of the step-size.

In order to improve the efficiency and the local rate of convergence of conjugate gradient algorithms several techniques have been suggested [5, 6]. The common idea of these techniques is the generation of mutually conjugate direction vectors independently of the step-size used. By help of these vectors it is possible to define an extra step which then enables us to define the minimum of a quadratic function without using perfect step-sizes. This extra step is identical to a quasi-Newton step. In [5] the conjugate direction vectors and the extra step is defined in terms of the quantities

$$\frac{(g_{i+1}, s_i)}{(g_{i+1} - g_i, s_i)} s_i \,,$$

where $s_i = x_{i+1} - x_i$. In [6] the extra step was also defined by help of these terms. By this approach it is possible to define the conjugate direction vectors and the extra step only if all steps are sufficiently inaccurate, i.e., $(g_{i+1}, s_i) \neq 0$.

In [7] it is shown that by an extra step is is possible also to improve the theoretical rate of convergence of minimization algorithms. Under some assumptions the rate of convergence is n-step cubic. But this result holds only for the projection gradient method because the important assumption is that the conjugate direction vectors are uniformly linearly independent. In general this is not possible to show for all conjugate gradient algorithms.

In the next part a new biorthogonalization process is described which does not depend on the step-size. The idea of the new biorthogonalization algorithm lies in the fact that the asymptotically perfect algorithms with the quadratic termination

property have the same rate of convergence as the perfect algorithms [8, 9]. Therefore they converge at least n-step quadratically also for a strictly convex quadratic function, whereas the perfect algorithms are finite. The new algorithm does not use an extra step and it is devised so as to minimize the total error after n steps if imperfect steps are used. It is easy to show that for a quadratic function this occurs if the direction vectors are mutually conjugate independently of the step-sizes used. The stability of the new algorithm does not depend on the accuracy or inaccuracy of the step-sizes.

2. THE ALGORITHM

Let $x_0 \in R_n$ be arbitrary. Let us denote $g_i = Ax_i - b$, where A is a symmetric, positive definite n by n matrix. Set $\tilde{g}_0 = g_0 = \tilde{h}_0$. Let us consider for $i = 0, 1, 2, \ldots, \ldots, n-1$ following biorthogonalization process:

$$x_{i+1} = x_i - \tilde{\alpha}_i \tilde{h}_i \,, \tag{9}$$

$$\tilde{g}_{i+1} = (g_{i+1} - g_i) - \frac{(g_{i+1} - g_i, \tilde{g}_i)}{(\tilde{g}_i, \tilde{g}_i)} \tilde{g}_i \,, \tag{10}$$

$$\tilde{h}_{i+1} = \tilde{g}_{i+1} - \frac{(g_{i+1} - g_i, \tilde{g}_{i+1})}{(g_{i+1} - g_i, \tilde{h}_i)} \tilde{h}_i \,, \tag{11}$$

where $\tilde{\alpha}_i$ is an arbitrary positive real number.

Theorem 1. *Suppose that m is the largest number $\leqq n-1$ such that all vectors $h_0, h_1, \ldots, h_m$ and $g_0, g_1, \ldots, g_m$ constructed according to (2)–(5) are nonzero and let $\tilde{h}_0, \tilde{h}_1, \ldots,$ and $\tilde{g}_0, \tilde{g}_1, \ldots$ be vectors constructed according to (9)–(11). Then*

$$\tilde{g}_i = c_i g_i \,, \quad \tilde{h}_i = c_i h_i \quad \text{for} \quad 0 \leqq i \leqq m$$

where $c_i \neq 0$ is a constant and

$$\tilde{g}_i = g_i = \tilde{h}_i = h_i = 0 \quad \text{for all} \quad i > m \,, \quad \text{with} \quad m \leqq n-1 \,.$$

Proof. According to the definition $\tilde{g}_0 = g_0$. For $i = 1$ from (2)–(5) we obtain

$$g_1 = g_0 - \alpha_0 Ah = -\alpha_0 \left(Ah_0 - \frac{(Ah_0, g_0)}{(g_0, g_0)} g_0\right) \tag{12}$$

and

$$h_1 = g_1 - \frac{(Ah_0, g_1)}{(Ah_0, h_0)} h_0 \,. \tag{13}$$

From (9)–(11) we have

$$\tilde{g}_1 = -\tilde{\alpha}_0 A h_0 - \frac{(-\tilde{\alpha}_0 A h_0, g_0)}{(g_0, g_0)} g_0 = -\tilde{\alpha}_0 \left(A h_0 - \frac{(A h_0, g_0)}{(g_0, g_0)} g_0\right) \tag{14}$$

since $g_1 - g_0 = -\tilde{\alpha}_0 A h_0$ and $\tilde{h}_0 = h_0 = \tilde{g}_0 = g_0$.

According to (12) and (14) we obtain

$$\tilde{g}_1 = c_1 g_1 \quad \text{with} \quad c_1 = \tilde{\alpha}_0/\alpha_0 , \tag{15}$$

and from (11) according to (15)

$$\tilde{h}_1 = c_1 g_1 - \frac{(A h_0, c_1 g_1)}{(A h_0, h_0)} h_0 = c_1 \left(g_1 - \frac{(A h_0, g_1)}{(A h_0, h_0)} h_0\right). \tag{16}$$

From (16), (13) we then have

$$\tilde{h}_1 = c_1 h_1 .$$

Suppose that

$$\tilde{g}_j = c_j g_j , \quad \tilde{h}_j = c_j h_j \quad \text{for} \quad j \in \{0, 1, \ldots, m-1\} .$$

Then according to (2)–(5) we have

$$g_m = g_{m-1} - \alpha_{m-1} A h_{m-1} = -\alpha_{m-1} \left(A h_{m-1} - \frac{(A h_{m-1}, g_{m-1})}{(g_{m-1}, g_{m-1})} g_{m-1}\right), \tag{17}$$

$$h_m = g_m - \frac{(A h_{m-1}, g_m)}{(A h_{m-1}, h_{m-1})} h_{m-1} . \tag{18}$$

From (9)–1(1) we obtain

$$\tilde{g}_m = -\tilde{\alpha}_{m-1} A \tilde{h}_{m-1} - \frac{(-\tilde{\alpha}_{m-1} A \tilde{h}_{m-1}, \tilde{g}_{m-1})}{(\tilde{g}_{m-1}, \tilde{g}_{m-1})} \tilde{g}_{m-1} =$$

$$= -\tilde{\alpha}_{m-1} \left(A \tilde{h}_{m-1} - \frac{(A \tilde{h}_{m-1}, \tilde{g}_{m-1})}{(\tilde{g}_{m-1}, \tilde{g}_{m-1})} \tilde{g}_{m-1}\right) =$$

$$= -\tilde{\alpha}_{m-1} c_{m-1} \left(A h_{m-1} - \frac{(A h_{m-1}, g_{m-1})}{(g_{m-1}, g_{m-1})} g_{m-1}\right)$$

since $g_m - g_{m-1} = -\tilde{\alpha}_{m-1} A \tilde{h}_{m-1}$.

According to (17) we have

$$\tilde{g}_m = c_m g_m \quad \text{with} \quad c_m = (\tilde{\alpha}_{m-1}/\alpha_{m-1})\, c_{m-1} ,$$

so that from (11), (18)

$$\tilde{h}_m = c_m h_m .$$

Let $g_{m+1} = 0$. Then according to (4) we obtain

$$g_m = \alpha_m A h_m$$

so that

$$\tilde{g}_{m+1} = -\tilde{\alpha}_m A\tilde{h}_m - \frac{(-\tilde{\alpha}_m A\tilde{h}_m, \tilde{g}_m)}{(\tilde{g}_m, \tilde{g}_m)}\tilde{g}_m =$$

$$= -\tilde{\alpha}_m A\tilde{h}_m + \tilde{\alpha}_m \frac{(A\tilde{h}_m, c_m g_m)}{(c_m g_m, c_m g_m)} c_m g_m =$$

$$= -\tilde{\alpha}_m c_m A h_m + \tilde{\alpha}_m c_m A h_m = 0 .$$

Let $\tilde{g}_{m+1} = 0$. Then according to (10),

$$0 = A\tilde{h}_m - \frac{(A\tilde{h}_m, \tilde{g}_m)}{(\tilde{g}_m, \tilde{g}_m)}\tilde{g}_m = (Ah_m - \frac{(Ah_m, g_m)}{(g_m, g_m)} g_m) c_m. \tag{19}$$

According to (19) we obtain

$$0 = g_m \frac{(g_m, g_m)}{(Ah_m, g_m)} Ah_m = g_m - \alpha_m Ah_m,$$

i.e., according to (4),

$$g_{m+1} = 0 .$$

The vectors $\tilde{g}_0, \tilde{g}_1, \ldots, \tilde{g}_m$ are orthogonal to each other, and since they are nonzero, their total number cannot exceed n. According to (11) and (5) for $m \leqq n - 1$ and for all $i > m$

$$\tilde{g}_i = g_i = \tilde{h}_i = h_i = 0 .$$

Corollary. *According to Theorem 1 for $\alpha_i = \tilde{\alpha}_i$ the biorthogonalization process (9)–(11) is equivalent to the biorthogonalization process (2)–(5).*

3. APPLICATION

Suppose that $f : R_n \to R_1$ is strictly convex and twice continuously differentiable and satisfies

$$\lim_{\|x\| \to \infty} f(x) = +\infty .$$

Let us denote the gradient $g_i = g(x_i)$. The algorithm for minimization of $f(x)$ is defined as follows:

(20) **Algorithm**

Step 0. Select $x_0 \in R_n$. If $g_0 = 0$, stop, else set $i = 0$ and go to Step 1.

Step 1. Set $\tilde{g}_i = g_i = h_i$.

Step 2. Compute

$$x_{i+1} = x_i - \alpha_i h_i$$

such that

$$(g_{i+1}, h_i) = \mu_i (g_i, h_i)$$

where $|\mu_i| \leqq \bar{\mu} < 1$ and

$$|\mu_i| \leqq \min \{\bar{\mu}, c\|g_i\|\}, \quad c > 0 .$$

Step 3. If $g_{i+1} = 0$, stop, else go to Step 4.

Step 4. If $i \equiv -1 \bmod n$ then set $i = i + 1$ and go to Step 1, else set

$$\tilde{g}_{i+1} = (g_{i+1} - g_i) - \frac{(g_{i+1} - g_i, \tilde{g}_i)}{(\tilde{g}_i, \tilde{g}_i)} \tilde{g}_i .$$

Step 5. If $\tilde{g}_{i+1} = 0$ then set $i = i + 1$ and go to Step 1, else set

$$h_{i+1} = \tilde{g}_{i+1} - \frac{(g_{i+1} - g_i, \tilde{g}_{i+1})}{(g_{i+1} - g_i, h_i)} h_i$$

set $i = i + 1$ and go to Step 2.

Algorithm (20) uses restarts. The method does not use an extra step and is devised so as to eliminate the zigzagging phenomenon if an imperfect step-size is used. An extra step after n regular steps is interesting if this step improves the theoretical rate of convergence. The need for a step along a special direction every n iterations is avoided by forcing the linear search to become more accurate as the solution is approached.

4. CONCLUSION

In this paper a new biorthogonalization process is defined. All known perfect conjugate gradient algorithms use exact line searches in order to obtain mutually conjugate directions. The inexactness of the line searches destroys the conjugacy of the directions generated and thereby the global efficiency of these methods. In order to eliminate this influence of the inexactness of the line searches a new biorthogonalization process is defined which does not depend on the step-size used. In the perfect case the new biorthogonalization process is equivalent to the cg-algorithms of Hestenes and Stiefel. The stability of the new algorithm does not depend on the accuracy or inaccuracy of the step sizes.

Acknowledgement. The author is greatly indebted to Professor J. Stoer for many useful discussions during the preparation of this paper. This work was supported by the Alexander von Humboldt Stiftung during the stay of the author at the Institut für Angewandte Mathematik und Statistik in Würzburg.

References

[1] *M. R. Hestenes, E. Stiefel:* The method of conjugate gradients for solving linear systems. J. Res. Nat. Bur. Standards, 49 (1952), 409—436.
[2] *R. Fletcher, C. M. Reeves:* Function minimization by conjugate gradients. Comp. J., 2 (1964), 149—154.
[3] *E. Polak, G. Ribiere:* Note sur le Convergence des Methods de Directions Conjuges, Reone Fr. Int. Rech. Oper. 16R1 (1969), 35—43.
[4] *J. W. Daniel:* The conjugate gradient method for linear and nonlinear operator equations, SIAM J. Numer. Anal. 4 (1967), 10—26.
[5] *L. C. W. Dixon:* Conjugate Gradient algorithms: Quadratic termination properties without line searches. J. of Inst. of Math. and Applics, 15 (1975), 9—18.
[6] *L. Nazareth:* A conjugate direction algorithm without line searches, JOTA, 3 (1977), 373—387.
[7] *M. J. Best:* A Method to Accelerate the Rate of Convergence of a Class of Optimization Algorithms. Math. Programming, 9 (1975) 139—160.
[8] *J. Stoer:* On the Relation between Quadratic Termination and Convergence Properties of Minimization Algorithms, Part I. Theory, Num. Math., 28 (1977) 343—366.
[9] *P. Baptist, J. Stoer:* On the Relation between Quadratic Termination and Convergence Properties of Minimization Algorithms, Part II, Applications, Num. Math., 28 (1977), 367—391

Súhrn

IMPERFEKTNÝ ALGORITMUS ZDRUŽENÝCH GRADIENTOV

FRIDRICH SLOBODA

V článku je popísaný nový biortogonalizačný algoritmus, ktorý nezávisí od dĺžky kroku. Algoritmus je navrhnutý za účelom minimalizácie chyby spôsobenej nepresnou realizáciou kroku. Väčšina doteraz známych algoritmov združených smerov je citlivá na presnosť realizácie perfektného kroku.

Author's address: Dr. *Fridrich Sloboda*, CSc., Ústav technickej kybernetiky SAV, Dúbravská cesta 1, 842 37 Bratislava.