Riv. Mat. Univ. Parma (4) 5 (1979), 217-219

C. SUTTI (*)

# On quadratic termination of the conjugate gradient method (**)

A GIORGIO SESTINI per il suo 70° compleanno

## 1. - Introduction

In Crowder's and Wolfe's paper [**1**], it is claimed that the conjugate gradient method [**3**] does not present quadratic termination property on a quadratic objective function $f(x)$, $x \in R^n$, in the absence of the standard initial starting condition. That is shown by numerical examples.

This note provides the theoretical reason of this failure. Namely, by the theorem exposed in **2**, it is proved that, when the initial search vector is quite arbitrary, the directions, which are built in $n$ successive iterations by the procedure implemented from Crowder and Wolfe, are not mutually conjugate with respect to the hessian matrix of $f(x)$. Thus is not guaranteed to terminate at the solution in at most $n$ steps.

The thesis of the theorem is proved for whichever conjugate gradient method with standard one-term correction formula.

Therefore it is of primary importance, in order that the termination ensues, to choose the standard start, that is to assume the initial search direction opposed to the gradient, $g_1$, of $f(x)$ in the initial approximation, $x_1$, of the minimum. However, by the proof of the theorem it will be evidentiate an initial condition which is necessary and sufficient for termination in the quadratic case. This condition will be formulated in the corollary presented in **2**.

---

(*) Indirizzo: Istituto di Matematica, Università, 43100 Parma, Italy.

(**) Lavoro eseguito nell'ambito del G.N.I.M. (C.N.R.). Nota presentata al Convegno Nazionale G.N.I.M. (C.N.R.), Rimini 20-22 Settembre 1978. — Ricevuto: 12-X-1978.

## 2. - Theorem on the directions conjugacy

The objective function, that will be considered in the following, is the quadratic

$$f(x) = \tfrac{1}{2} x^T G x + b^T x + c \,, \qquad (x \in R^n) \,, \tag{2.1}$$

where $c$ is a scalar constant, $b$ is an $n$-th vector and $G$ is an $(n \times n)$ symmetric and positive definite matrix.

Moreover the procedure of the conjugate gradient methods, that will be analyzed, is the following. Given $x_1$ and $d_1$, for any $k \geq 1$, let $x_{k+1} = x_k + \lambda_k d_k$, where $\lambda_k$ is such that $f(x_k + \lambda_k d_k) = \min_\lambda f(x_k + \lambda d_k)$. Then, if $g_{k+1} = 0$ stop, if not let

$$d_{k+1} = -g_{k+1} + \beta_k d_k \,, \tag{2.2}$$

where $\beta_k$ is chosen so that

$$d_{k+1}^T G d_k = 0 \,, \tag{2.3}$$

and iterate the procedure.

The conjugate gradient methods, with one-term correction formula (2.2), usually assume $d_1 = -g_1$, while they differ in the rule to compute $\beta_k$. First of all, we mention the rule

$$\beta_k = g_{k+1}^T G d_k / d_k^T G d_k \,,$$

because this one is used in the Crowder's and Wolfe's numerical experiences; but the more commonly used rules are the Hestenes's and Stiefel's one [**3**]

$$\beta_k = g_{k+1}^T (g_{k+1} - g_k) / d_k^T (g_{k+1} - g_k)$$

and the Fletcher's and Reeves's one [**2**]

$$\beta_k = g_{k+1}^2 / g_k^2 \,.$$

In all these methods it has been proved that the successive directions $d_1, d_2, \ldots, d_n$ are all linearly independent and conjugate with respect to $G$, that is

$$d_j^T G d_k = 0 \,, \quad \text{for} \quad j \neq k \,, \qquad (j, k = 1, \ldots, n) \,, \tag{2.4}$$

and $x_k$ minimizes $f(x)$ in the subspace $R^{k-1} = (d_1, \ldots, d_{k-1})$. Consequently the procedure terminates with $g_k = 0$ for some $k \leqslant n$.

On the contrary, Crowder and Wolfe [1] by numerical examples have shown that the conjugate gradient method, applied to the function (2.1), when $d_1 \neq -g_1$, does not possess quadratic termination.

By the following theorem, we shall show that in their case the directions $d_1, d_2, \ldots, d_n$ are not guaranteed to be all mutually conjugate and so the termination can fail.

Theorem. *Let* (2.1) *be the objective function and let us apply to* (2.1) *a conjugate gradient method, with one-term correction formula* (2.2), *but with* $d_1$ *different from* $-g_1$ *and quite arbitrary. The directions* $d_1, d_2, \ldots, d_n$ *are not guaranteed to be all mutually conjugate with respect to* $G$.

Proof. By contradiction, assume that our iterative procedure constructs search directions all mutually conjugate with respect to $G$. This means that $d_k$, $2 \leqslant k \leqslant n$, is conjugate to $d_{k-i}$, $1 \leqslant i \leqslant k-1$, that is $d_2^T G d_1 = 0$ and, by supposing $d_1, d_2, \ldots, d_k$ all mutually conjugate, $d_1, d_2, \ldots, d_{k+1}$ are also all mutually conjugate.

We can soon prove this last thesis. Because $x_{k+1}$ is the minimum point along $d_k$ and $d_1, d_2, \ldots, d_k$ are mutually conjugate, $x_{k+1}$ is the minimum of $f(x)$ in the subspace $R^k = (d_1, \ldots, d_k)$. Then it is

$$g_{k+1}^T d_j = 0 \qquad (1 \leqslant j \leqslant k)\,. \tag{2.5}$$

If $L(a, b, c, \ldots)$ defines a linear combination of $a, b, c, \ldots$, by (2.2) it is

$$d_j = L(g_j, d_{j-1}) \qquad (1 \leqslant j \leqslant k)\,, \tag{2.6}$$

and, because in the quadratic case $g_j = g_{j-1} + \lambda_{j-1} G d_{j-1}$, it is

$$g_j = L(g_{j-1}, G d_{j-1}) \qquad (1 \leqslant j \leqslant k)\,, \tag{2.7}$$

then it is

$$d_j = L(g_{j-1}, G d_{j-1}, d_{j-1}) \qquad (1 \leqslant j \leqslant k)\,. \tag{2.8}$$

But, from (2.5) and (2.6) we have

$$g_{k+1}^T g_j = g_{k+1}^T g_{j-1} = 0 \qquad (1 \leqslant j \leqslant k)\,, \tag{2.9}$$

so, from (2.5), (2.8) and (2.9), it is

$$g_{k+1}^T G d_j = 0 \qquad (1 \leqslant j \leqslant k-1) . \tag{2.10}$$

Therefore, by (2.2) $d_{k+1}$ is conjugate to $d_1, d_2, \dots, d_{k-1}$ and, by (2.3), is conjugate to $d_k$.

We consider now $d_2^T G d_1$; because $G d_1 = \lambda_1^{-1}(g_2 - g_1)$ and $d_2 = -g_2 + \beta_1 d_1$ with $d_1^T g_2 = 0$, we have

$$d_2^T G d_1 = \lambda_1^{-1}(-g_2^2 + g_2^T g_1 - \beta_1 d_1^T g_1) . \tag{2.11}$$

But, if $d_1 \neq -g_1$ or, more generally, $d_1$ is quite arbitrary, $d_2^T G d_1 \neq 0$. Therefore our conjecture on the mutually conjugacy of $d_1, d_2, \dots, d_k, 2 \leqslant k \leqslant n$, is false. Q.E.D.

This theorem justifies the Crowder's and Wolfe's numerical results. From its proof, moreover, some choice of $d_1$, different from $d_1 = -g_1$, but such that the termination is retained, can be deduced. Infact from (2.11) we note that if $d_1$ satisfies to the condition

$$d_1^T g_1 = \beta_1^{-1} \cdot (g_2^T g_1 - g_2^2) \tag{2.12}$$

it is $d_2^T G d_1 = 0$. Thus, in the proof of above theorem, the induction is complete and hence all conjectures are true.

We can conclude with the following corollary.

Corollary. *Let* (2.1) *be the objective function, and let us apply to* (2.1) *a conjugate gradient method with one-term correction formula* (2.2). *If and only if $d_1$ satisfies condition* (2.12), *the n-steps termination is achieved.*

Note that $d_1 = -g_1$ satisfies condition (2.12).

Finally, since (2.12), from $d_1^T g_2 = 0$, it results

$$d_1^T(g_1 - g_2) = \beta_1^{-1}(g_2^T g_1 - g_2^2) ,$$

that is

$$d_1^T G d_1 = \beta_1^{-1}(g_2^T G d_1) = \beta_1^{-1}(d_1^T G g_2) . \tag{2.13}$$

So, from (2.13), it is evident that $g_2$ can be conjugate to $d_1$ with respect to $G$ [**4**] only if $d_1$ is trivial. In this case, however, it is obvious that, by choosing $d_2 = -g_2$, $(n+1)$-steps termination follows.

## References

[1] H. Crowder and P. Wolfe, *Linear convergence of the conjugate gradient method* I.B.M., J. Res. Develop., 1972, 431-433.

[2] R. Fletcher and C. M. Reeves, *Functions minimization by conjugate gradients*, Comput. J. **7** (1964-65), 149-154.

[3] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradient for solving linear systems*, J. Res. Nat. Bur. Standards **49** (1952), 409-436.

[4] M. Y. D. Powell, *Some convergence properties of the conjugate gradient method*, C.S.S. **23** Harwell, 1975.

## Sommario

*Numericamente si è osservato che il metodo a gradiente coniugato non presenta terminazione quadratica per funzioni obiettivo quadratiche, con scelte arbitrarie della direzione di ricerca iniziale.*

*La giustificazione teorica di ciò viene fornita, e scelte iniziali alternative, tali da garantire la proprietà suddetta, vengono messe in evidenza.*

* * *