Riv. Mat. Univ. Parma (4) 9 (1983), 167-172

C. SUTTI (*)

Remarks on the convergence of cyclic coordinate methods (**)

1 - Introduction

In the recent paper [1], Bazaraa, Shetty and Goode presented some convergence conditions, concerning combinations of minimization algorithms and objective functions. The same topic had been treated from the author in the two previous papers [4] and [3].

In this note relations are established among the classes of methods and functions which have been examined in [1], [4] and [3]. Namely they are shown the equality of the algorithms and the equivalence of the functions, from a geometrical point of view. Moreover the conditions, formulated in [1] on the objective functions, are made more precise, evincing in which interpretation they are valid or not. Finally apparent discrepancies among the results presented in [1], [4] and [3] are explained.

In section 2 the previous work in the area is briefly mentioned. In section 3 the discussion follows.

2 - Premise

In [4], the author examined the possibility of nongradient unconstrained minimization algorithms entering the cyclic search type of path, first illustrated by Powell in [2], and proved that the methods considered can generate the above loops, when applied to an objective function $F(x)$, $x \in R^m$, only if they

---

(*) Indirizzo: Istituto di Matematica, Via Arnesano, 73100 Lecce, Italy.
(**) Lavoro eseguito nell'ambito del G.N.I.M. (C.N.R.). — Ricevuto: 8-VII-1981.

construct a sequence of points $\{x^i\}$ $i \geqslant 1$, such that $F(x^{i+1}) \leqslant F(x^i)$ $\forall i$, by spanning m linearly independent directions $d_1^k, d_2^k, \ldots, d_m^k$, $k \geqslant 1$, where $(d_1^1, d_2^1, \ldots, d_m^1) = (d_1^k, d_2^k, \ldots, d_m^k)$ $\forall k$, in a constant order (Theorem 4.6 [**4**]). Moreover the author proved (Theorem 4.6 [**4**]) that this cyclic behaviour cannot occur, when exact minimizations are performed along the search directions, if the objective function satisfies the following assumption

(a_1) F is bounded below and strictly convex.

We shall define such functions as comprising the class Ω_1. In the proof of Theorem 4.6 [**4**], it was shown that the hessian matrix H of the objective function must be singular at all points of any limit cycle, if the line search is assumed to find the first local minimum along the line, so that the following inequalities are satisfied

$$F(x^{i+1}) = F(x^i + \lambda_j^k d_j^k) \leqslant F(x^i + \lambda d_j^k) \leqslant F(x^i) \tag{1}$$

$$\forall i \geqslant 1\,, \quad 1 \leqslant j \leqslant m\,, \quad k \geqslant 1\,, \quad i = j + mk\,, \quad \lambda \in [0, \varepsilon\lambda_j^k]\,, \quad \varepsilon > 0\,.$$

Subsequently, in [**3**], some improved conditions have been provided, namely

(c_1) if $\exists I \subset R^m$: $\det H(x) = 0$, $x \in I$, $\operatorname{meas} I = 0$;

(c_2) if $\exists I \subset R^m$: $\det H(x) = 0$, $x \in I$, and $\operatorname{meas} I \neq 0$, $F(x) \neq \text{const}$, $x \in I$;

(c_3) if $\exists I \subset R^m$: $\det H(x) = 0$, $x \in I$, and $\operatorname{meas} I \neq 0$ and $F(x) = \text{const}$, $x \in I$, $I \not\supset I_1 \cup I_2 \cup \ldots \cup I_m$, where $I_1, I_2, \ldots, I_m$ are segments belonging to a basis of R^m.

Conditions (c_1), (c_2) and (c_3) are sufficient to guarantee the numerical convergence of the cyclic coordinate methods, not only when every step is perfect, in the sense precised by (1), but also when some arbitrary steps are made, provided that these steps are less than or equal to the last optimal step [**3**].

In their recent paper [**1**], Bazaraa, Shetty and Goode consider the class of minimization algorithms, which perform exact line searches along mutually orthogonal directions, where these are interpreted as finding the absolute minimum along the line, that is x^{i+1} such that

$$F(x^{i+1}) = F(x^i + \lambda_j^k d_j^k) \leqslant F(x^i + \lambda d_j^k) \leqslant F(x^i) \tag{2}$$

$$\forall i \geqslant 1\,, \quad 1 \leqslant j \leqslant m\,, \quad k \geqslant 1\,, \quad i = j + mk\,, \quad \forall \lambda > 0\,.$$

For this class of methods, they establish convergence to a stationary point for functions of the class Ω_2 defined as satisfying the following assumption

(a_2) F is continuously differentiable and has a unique minimum along any line.

3 - Discussion

We first note that the classes of n-dimensional minimization algorithms, considered in [4] and in [**1**], coincide, as cyclic loops necessarily imply identical sets of directions and a constant order of searching them.

On the classes of objective functions, we observe that (a_1) is satisfied in assumption (a_2), i.e. $\Omega_1 \subseteq \Omega_2$. Namely, from an analytical point of view, the definition of Ω_1 [4] is more restrictive than the definition of Ω_2 [**1**], because the first requires the positive definiteness of $H(x)$ $\forall x$. However, from a geometrical point of view, for $m \geqslant 2$, Ω_1 is equivalent to Ω_2, in the sense that any function, or of Ω_1 or of Ω_2, defined in R^m, $m \geqslant 2$, has convex and connected level sets. In other words, (a_2) implies the unimodality of F along every line: that results clear from the use of assumption (a_2) in [**1**].

In order to better show that (a_2), as (a_1), concerns the unimodality of P along each line, let us examine the following situation. Namely let us assume that only the uniqueness of the global minimum is required by (a_2) for F. Therefore F could possess one global minimum and one (or more) local minimum along some line. But, in this case, it can exist a line, along which the global minimizer is not unique. Let us examplify that, by simply considering a biconvex function defined in R^2 and having global minimum in $(0, 0)$ and local minimum in $(0, C)$. Its projection along every line presents at most one global minimum, but, for example, along a line passing through the point $(0, C, F(0, C))$ and tangent to the connected component, containing $(0, 0, F(0, 0))$ of the set of level $F(0, C)$, it presents two global minima with equal value, that is two global minimizers. On the other hand, this situation results forbidden from the use of assumption (a_2) in [**1**]. In order to prove that, let us assume that $F \in \Omega_2$, can have many minimizers along some direction d, as figured in the following Fig. 1 and Fig. 2.

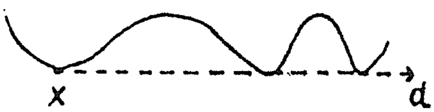

Fig. 1

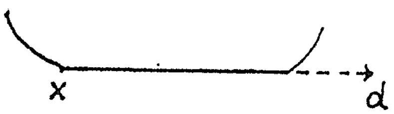

Fig. 2

But in the above cases, an optimum step $s(x, d) > 0$, from x along d, does not guarantee

$$h(e) > 0 \qquad \forall e > 0 \,, \tag{3}$$

where $h(e) = \inf \{F(x) - F(x + s(x, d)d) : s(x, d) \geq e,\ x \in R^m,\ \|d\| = 1\}$, and that invalidates the proof of the lemma given by Bazaraa, Shetty and Goode; therefore it is clear that an unique minimizer is intended in [**1**]. In addition, we shall show that also the conclusion of the lemma is false when the minimizer is not unique. In order to do that, let discuss in detail the case figured in Fig. 2. In this particular case d is a constancy direction for F with respect to ϱ, where $\varrho = F(x)$, in the sense of the following definition [**3**]:

(d_1) if $\varrho \in R$ and $B(\varrho) = \{x \in R^m : F(x) = \varrho\}$, a *constancy direction* for F with respect to ϱ is a direction d s.t. $\|d\| = 1$ and $\exists x \in B(\varrho)$ and $\mu \in R$: $(x + \lambda d) \in B(\varrho)$ $\forall \lambda \in [0, \mu]$.

Definition (d_1) concerns the class of objective function having linear segments on some level surfaces, that is the class Ω_3 of F s.t. $\exists \varrho \in R$ and $(z, y) \in R^m$: $F(z) = F(y) = \varrho$ and $F(z + \lambda(y - z)) = \varrho$ $\forall \lambda \in [0, 1]$ [**3**]. Let further introduce the class Ω_4 of lower bounded functions F s.t. $\exists$ a pair of sequences $\{z_n\}_{n \in N}$ and $\{y_n\}_{n \in N}$: $F(z_n) \neq F(y_n)$ $\forall n$, and $\operatorname{minlim}_n \|z_n - y_n\| > 0$, but for any sequence of scalars $\{\lambda_n\}_{n \in N}$, $\lambda_n \in [0, 1]$ $\forall n$, it is $\lim_n F(z_n) = \lim_n F(y_n) = \lim_n F(z_n + \lambda_n \cdot (y_n - z_n))$. To Ω_4 belong functions having curved segments on some level surfaces, where the curvature radius tends to infinity [**3**]. The functions $F \in \Omega_4$ possess pseudoconstancy directions in the sense of the following definition [**3**]:

(d_2) the direction d, $\|d\| = 1$, is a *pseudoconstancy direction* for F, if $\exists$ $\{x_n\}$ and $\{d_n\}$ and $\mu \in R$: $\lim_n F(x_n + \lambda d_n) = \lim_n F(x_n)$, $\forall \lambda \in [0, \mu]$.

Now, if assumption (a_2) concerns the uniqueness of the minimum, but non of the minimizer of F, there exist functions $F \in \Omega_3 \cap \Omega_2 \neq \emptyset$ (or $\in \Omega_4 \cap \Omega_2 \neq \emptyset$). On the other hand, in [**3**] has been shown that numerical non convergence can occur by applying cyclic coordinate methods to the functions $F \in \Omega_3$ (or $\in \Omega_4$): therefore the conclusion of the lemma in [**1**] is false, when the minimizer is not unique.

The case figured in Fig. 2 occurs in Powell's examples [**2**], when x and d are the asymptotic points and search directions, in fact those functions belong to Ω_4 [**3**]. About the Powell's examples, let now observe that for them, on the limit path

$$\begin{aligned} (-1, 1, -1) &\to (1, 1, -1) \to (1, -1, -1) \to (1, -1, 1) \\ &\to (-1, -1, 1) \to (-1, 1, 1) \to (-1, 1, -1) \end{aligned} \tag{4}$$

it is $\det H = 0$ and $F = \text{const}$, moreover the segments of (4) span all R^3: so none of the conditions (c_1), (c_2) and (c_3) is satisfied. At contrary, for the Bazaraa, Shetty and Goode's function f [1], on the claimed cyclic path (4) it is

$$\begin{aligned}
\det H &= 0 \quad \text{for } -1 \leqslant x_1 \leqslant -\tfrac{1}{2} \text{ or } \tfrac{1}{2} \leqslant x_1 \leqslant 1,\ x_2 = 1,\ x_3 = -1\,,\\
\det H &= 0 \quad \text{for } x_1 = 1,\ \tfrac{1}{2} \leqslant x_2 \leqslant 1 \text{ or } -1 \leqslant x_2 \leqslant -\tfrac{1}{2},\ x_3 = -1\,,\\
\det H &= 0 \quad \text{for } x_1 = 1,\ x_2 = -1,\ -1 \leqslant x_3 \leqslant -\tfrac{1}{2} \text{ or } \tfrac{1}{2} \leqslant x_3 \leqslant 1\,,\\
\det H &= 0 \quad \text{for } \tfrac{1}{2} \leqslant x_1 \leqslant 1 \text{ or } -1 \leqslant x_1 \leqslant -\tfrac{1}{2},\ x_2 = -1,\ x_3 = 1\,,\\
\det H &= 0 \quad \text{for } x_1 = -1,\ -1 \leqslant x_2 \leqslant -\tfrac{1}{2} \text{ or } \tfrac{1}{2} \leqslant x_2 \leqslant 1,\ x_3 = 1\,,\\
\det H &= 0 \quad \text{for } x_1 = -1,\ x_2 = 1,\ \tfrac{1}{2} \leqslant x_3 \leqslant 1 \text{ or } -1 \leqslant x_3 \leqslant -\tfrac{1}{2}\,,
\end{aligned}$$

and

$$f \neq \text{const}.$$

Therefore f satisfies condition (c_2), nevertheless numerical non convergence occurs. For clarifying the above discrepancy, we shall analyze the performance of the onedimensional minimization in this examined case. For this purpose, we note that if f is minimized, for example, from $x^1 = (-1-\varepsilon, 1+\varepsilon/2, -1-\varepsilon/4)$ along $(x^1 + \lambda(1\ 0\ 0))$, $\lambda > 0$, and if the performed line search satisfies assumption (2), the point $m_1 = (1+\varepsilon/8, 1+\varepsilon/2, -1-\varepsilon/4)$ is found; while, if the line search satisfies assumption (1), the point $m_2 = (-1+\varepsilon/8, 1+\varepsilon/2, -1-\varepsilon/4)$ is individuated. In fact $f(m_1) = 1 + \varepsilon/4(\tfrac{1}{2} + \varepsilon 27/16)$ is the absolute minimum and $f(m_2) = 1 + \varepsilon/4(1 + \varepsilon 27/16)$ is the first local minimum along the search direction. Therefore, by applying the axial method to f from x^1 under assumption (2), the sequence giving cyclic behaviour can be obtained; otherwise the procedure should be stable. It follows that the interpretation of the onedimensional minimization is of primary importance if we are to reach conclusions on the convergence of the n-dimensional minimization algorithms. For this purpose, we mention that the convergence theory of descent minimization algorithms applied to non convex objective functions, usually requires that the iterate points are restricted within a neighborhood of the minimum. In other words, it is assumed that inequality (1) is satisfied when optima steps are performed and that, when some arbitrary steps are allowed, they are less than or equal to the last optimum step. Therefore, only a line search, interpreted as in definition (1), is significant from an operative point of view.

Acknowledgment. Thanks are due to G. Resta for the interesting conversation on this topic.

References

[1] M. S. Bazaraa, C. M. Shetty and J. J. Goode, *On the convergence of a class of algorithms using linearly independent search directions*, Math. Programming **18** (1980), 89-93.

[2] M. J. D. Powell, *On search directions for minimization algorithms*, Math. Programming **4** (1973), 193-201.

[3] G. Resta and C. Sutti, *Some safeguards for descent minimization algorithms to avoid numerical non convergence*, L.C.W. Dixon and G. P. Szegö eds., Towards global optimization II, North Holland, Amsterdam (1978), 241-254.

[4] C. Sutti, *Remarks on conjugate directions methods for minimization without derivatives*, L. C. W. Dixon and G. P. Szegö eds., Towards global optimization, North Holland, Amsterdam (1975), 290-304.

Riassunto

In questa nota sono discusse alcune condizioni di convergenza, riguardanti sia gli algoritmi di minimizzazione che le funzioni obiettivo, recentemente formulate nel lavoro [**1**]. *Esse sono rese più precise sulla base dei due precedenti lavori* [**4**] *e* [**3**]. *Dalla discussione viene messo in risalto il ruolo fondamentale che la ricerca di linea riveste nei problemi di convergenza dei metodi a coordinate cicliche.*

* * *