Revue française d'automatique, informatique, recherche opérationnelle. Recherche opérationnelle

Yves Smeers

A convergence proof of a special version of the generalized reduced gradient method (GRGS)

Revue française d'automatique, informatique, recherche opérationnelle. Recherche opérationnelle, tome 8, n° V3 (1974), p. 105-114

<http://www.numdam.org/item?id=RO_1974__8_3_105_0>

R.A.I.R.O.
(8e année, novembre 1974, V-3, p. 105 à 114)

A CONVERGENCE PROOF OF A SPECIAL VERSION OF THE GENERALIZED REDUCED GRADIENT METHOD (GRGS) (1)

by Yves SMEERS (2)

Résumé. — *Bien que la méthode du gradient réduit généralisé ait fait l'objet de nombreuses expériences numériques, il ne semble pas exister jusqu'à présent de preuve de convergence. L'article considère une version particulière de la méthode et présente des modifications permettant d'obtenir une telle preuve. Il est montré en particulier qu'une reconstruction particulière de la base, toutes les « M » itérations, permet de prouver la convergence.*

1. INTRODUCTION

Among nonlinear programming codes, the « Generalized Reduced Gradient Method » (GRG) [1] stands in a privileged position because of favorable computational experience reported in the literature for different types of problems ([2] [5] [6]). However, to the best of the author's knowledge, no convergence proof for the case of nonlinear constraints has been published so far. The purpose of this paper is to provide such a proof for a special version of GRG, namely the GRGS method. The paper originated from some comments by Abadie [3], who pointed out the relation between the Generalized Reduced Gradient Method, and some previous work by the author [7].

2. THE GENERALIZED REDUCED GRADIENT METHOD

The presentation of the GRGS method given in this section relies heavily on [1], to which the reader is referred for further details.

(1) Core Discussion Paper No. 7341.
(2) Center for Operations Research & Econometrics. Université Catholique de Louvain, Belgium.

We consider the nonlinear program P

$$\text{(1)} \qquad \operatorname{Max} f(X)$$

$$\text{(2)} \qquad s.t.\ g_i(X) = 0 \qquad \text{for } i = 1, \ldots, m$$

$$\text{(3)} \qquad A \leqslant X \leqslant B$$

where X is an N-vector, A and B are bounded, and the functions f and g_i are continuously differentiable.

With every feasible point X we associate a partitioning of the vectors A, B and X, defined as follows.

$$\text{(4)} \qquad X = (x, y)$$

$$\text{(5)} \qquad A = (\alpha, \alpha')$$

$$\text{(6)} \qquad B = (\beta, \beta')$$

with

$$\text{(7)} \qquad \alpha' < y < \beta'$$

$$\text{(8)} \qquad \alpha \leqslant x \leqslant \beta$$

$$\text{(9)} \qquad y \in R^m.$$

Finally, in order to define Kuhn-Tucker points, we introduce the linear system in u and v

$$\text{(10)} \qquad v = \frac{\partial f}{\partial x} - u \frac{\partial g}{\partial x}$$

$$\text{(11)} \qquad 0 = \frac{\partial f}{\partial y} - u \frac{\partial g}{\partial y},$$

which has a unique solution if and only if the matrix $\left(\frac{\partial g}{\partial y}\right)$ is nonsingular. A feasible point $\mathring{X}$ will be called a Kuhn-Tucker point if and only if the solution $(\mathring{u}, \mathring{v})$ of the system (10)-(11) satisfies the conditions

$$\text{(12)} \qquad \begin{aligned} &\mathring{v}_j \leqslant 0 \quad \text{if} \quad \mathring{x}_j = \alpha_j \\ &\mathring{v}_j \geqslant 0 \quad \text{if} \quad \mathring{x}_j = \beta_j \\ &\mathring{v}_j = 0 \quad \text{if} \quad \alpha_j < \mathring{x}_j < \beta_j. \end{aligned}$$

When the relations in (12) are not all satisfied, new feasible points leading to a larger objective function value can be found in the neighborhood of $\mathring{X}$.

The GRGS method selects such a point according to the following rules. Define

(13)
$$\begin{aligned}
\overset{\circ}{v}_{s_1} &= \max \{ [\max_{(j \mid x_j = \alpha_j)} \overset{\circ}{v}_j] ; 0 \} \\
- \overset{\circ}{v}_{s_2} &= - \min \{ [\min_{(j \mid x_j = \beta_j)} \overset{\circ}{v}_j] ; 0 \} \\
|\overset{\circ}{v}_{s_3}| &= \max \{ [\max_{(j \mid \alpha_j < x_j < \beta_j)} |\overset{\circ}{v}_j|] ; 0 \} \\
\Delta_s &= \max [\overset{\circ}{v}_{s_1}, - \overset{\circ}{v}_{s_2}, |\overset{\circ}{v}_{s_3}|],
\end{aligned}$$

where s_1, s_2 and s_3 are the indices of the maximizing and minimizing $\overset{\circ}{v}_j$ (with a proper convention if the max or the min is zero), and s equals s_1, s_2 or s_3 depending on whether $\overset{\circ}{v}_{s_1}$, $- \overset{\circ}{v}_{s_2}$ or $|\overset{\circ}{v}_{s_3}|$ is larger. One can then define the vector

(14)
$$h = e_s \operatorname{sign} \overset{\circ}{v}_s,$$

where e_s is the s^{th} unit vector and sign $\overset{\circ}{v}_s$ equals $+1$ or -1 depending on the sign of $\overset{\circ}{v}_s$. The new point is then found by solving the problem

(15)
$$\begin{aligned}
&\operatorname{Max} f(x, y) \\
&\text{s.t. } g_i(x, y) = 0 \qquad i = 1, \ldots, m \\
&x = \overset{\circ}{x} + \theta h \\
&A \leqq X \leqq B \\
&\quad \theta \geqq 0.
\end{aligned}$$

The reader is referred to [1] for more details about this problem.

The GRGS method can then be stated as follows.

(0) Find a feasible point.

(i) Partition X as indicated in (4), (5), (6), (7), (8) and (9).

(ii) Solve the system (10-11). If the current point satisfies the Kuhn-Tucker conditions, stop. Otherwise go to (iii).

(iii) Select h as in (13) and (14).

(iv) Solve (15), take the solution as the new current point. Return to (i).

Practically there is considerable freedom for partitioning X in step (i), the only constraints being that all components of X, which are equal to one of their bounds, belong to x and that the matrix $\left(\frac{\partial g}{\partial y}\right)$ is invertible in step (ii). The convergence proof, though, will impose some additional conditions on step (i).

3. A NONDEGENERACY ASSUMPTION AND SOME MODIFICATIONS FOR GRGS

As in [1], we make the following nondegeneracy assumption.

Assumption : At every feasible point X there exists a partitioning of X that satisfies (7), (8), (9) and such that $\left(\frac{\partial g}{\partial y}\right)$ is nonsingular.

This assumption, which is closely related to the properties of nondegenerate extreme points in linear programming, guarantees that steps (i) and (ii) in the algorithm can always be performed.

As with linear programming, the y variables will be called *basic* and the x variables *nonbasic*. A basis at point X will be defined by the set L_X of nonbasic variables. We now specify that at every iteration L_X belongs to a set S_X, such that the point to set mapping $X \to S_X$ is closed (see for example Zangwill [10], p. 88), moreover we assume that all the elements of S_X satisfy (7), (8), (9) and are such that $\left(\frac{\partial g}{\partial y}\right)$ is nonsingular. General construction rules of a set S_X satisfying these two requirements can be given for nondegenerate problems (see section 5) but more efficient procedures can often be devised using the special structure of the problem at hand. (See, for instance [10], p. 167 for the convex-simplex method and [8] for a modular design algorithm.) A basis defined by a set L_X belonging to S_X will be called *updated* at X. From now on we shall assume that step (i) has been modified so that an updated basis is constructed at every iteration. The relaxation of this assumption and its practical consequences will be discussed in section 6.

In order to give some more insight into this last notion we consider the following example. Let

$$g_1(X) \equiv X_1 + X_2 + X_3 - 1$$

$$g_2(X) \equiv -X_1 + X_2 + X_4 - \frac{1}{2}$$

$$M \geqslant X_i \geqslant 0 \qquad i = 1, \ldots, 4$$

be a constraint set. Consider the sequence $\{ X^k \}_2^\infty$ where

$$X^k = \left(\frac{1}{k}, \frac{1}{k}, \frac{k-2}{k}, \frac{1}{2}\right).$$

Clearly $\lim X^k = \left(0, 0, 1, \frac{1}{2}\right)$. One can then easily see that the set

$$L^k = (3, 4)$$

defines a basis at every point X^k but that the only admissible basis at X^∞ is $L^\infty = (1, 2)$. Although L^k is a basis of X^k, it is not an updated basis at that point.

Finally, the equations in (13) will be modified to take into account the closeness of the x components to their bounds. Define

$$\mathring{v}_{s_1}(\beta_{s_1} - \mathring{x}_{s_1}) = \max \{ [\max_j \mathring{v}_j(\beta_j - \mathring{x}_j)] ; 0 \}$$

$$(13') \qquad \mathring{v}_{s_2}(\alpha_{s_2} - \mathring{x}_{s_2}) = \max_j \{ [\max \mathring{v}_j(\alpha_j - \mathring{x}_j)] ; 0 \}$$

$$\Delta'_s = \max \{ \mathring{v}_{s_1}(\beta_{s_1} - x_{s_1}) ; \mathring{v}_{s_2}(\alpha_{s_2} - x_{s_2}) \},$$

where s_1, s_2 and s have the same meaning as in section 2.

The GRGS method can now be restated as follows.

(0) Find a feasible point.

(i) Construct an updated basis at the current point.

(ii) Solve the system (10)-(11). If the current point satisfies the Kuhn-Tucker conditions, stop. Otherwise go to (iii).

(iii) Select h by (13′) and (14).

(iv) Solve (15), take the solution as the new current point. Return to (i).

4. CONVERGENCE PROOF

To simplify notation, an iteration in the algorithm will be indicated by superscript k.

The following lemma summarizes all the continuity properties required in the proof of the main theorem.

Lemma : Suppose that $X^k \to X^\infty$, where X^∞ is not a Kuhn-Tucker point. Let $L^k = \bar{L}$ for all k, and H^∞ be the set of h which can be selected at X^∞. Then, there exist a vector $h^\infty \in H^\infty$ and a subsequence $K' \subset K$ such that $x^k \underset{K'}{\to} x^\infty$, $y^k \underset{K'}{\to} y^\infty$, $v^k \underset{K'}{\to} v^\infty$, $u^k \underset{K'}{\to} u^\infty$, $h^k = h^\infty$ for all $k \in K'$.

Proof : 1) $x^k \to x^\infty$ and $y^k \to y^\infty$. This follows immediately from the convergence of $\{ X^k \}_{k \in K}$ and from $L^k = L^\infty$ for all k.

2) $u^k \to u^\infty$ and $v^k \to v^\infty$. By definition of updated bases, $\left(\frac{\partial g}{\partial y}\right)_{y^\infty}$ is non-singular, which proves the convergence of $\{ u^k \}_{k \in K}$ and $\{ v^k \}_{k \in K}$.

3) $h^k = h^\infty$. Since X^∞ is not a Kuhn-Tucker point, $\Delta_s'^{\infty}$ is positive and there exists a subsequence $K' \subset K$ such that h^k is constant for all $k \in K'$.

Because of the continuity of the operations leading to the selection of h, h^k also satisfies (13′) and (14) at X^∞. Hence $h^k = h^\infty \in H^\infty$ for all $k \in K'$.

Before proceeding toward the main proposition of this paper, it is useful to recall the use of the implicit function theorem. Let $\mathring{X}$ be a feasible point and $(\mathring{x}, \mathring{y})$ the partitioning obtained by constructing an updated basis. Because $\left(\frac{\partial g}{\partial y}\right)_{\mathring{y}}$ is nonsingular, the implicit function theorem can be applied. Hence there exists a neighborhood of $\mathring{X}$ such that all points satisfying $g_i(X) = 0$ $(i = 1, \ldots, m)$ can be represented as $(x, y(x))$, where $y(x)$ is continuously differentiable and $y(\mathring{x}) = \mathring{y}$. In this neighborhood of $\mathring{X}$, the objective function of problem (15) can be written as a function

$$f\{x(\theta), y[x(\theta)]\}, \tag{16}$$

where $x(\theta) = \mathring{x} + \theta h$.

Moreover, it can easily be seen that

$$\frac{d}{d\theta} f\{x(\theta), y[x(\theta)]\}\,\Big|_{\theta=0} = |v_s|, \tag{17}$$

where s is the index selected in calculating h. We can now state the following theorem.

Theorem 1 : Every cluster point of $\{X^k\}_{k \in K}$ satisfies the Kuhn-Tucker conditions.

Proof : Let X^∞ be a cluster point which does not satisfy the Kuhn-Tucker conditions. Since the set of possible bases is finite, there exists a subsequence $K_1 \subset K$ such that $L^k = \bar{L}$ for all $k \in K_1$. Applying lemma 1, there exists a subsequence $K_2 \subset K_1$ such that $h^k = h^\infty$ for all $k \in K_2$. Defining the function $x^\infty(\theta) = x^\infty + \theta h^\infty$ and $y[x^\infty(\theta)]$ as discussed before, there exists a neighborhood $N'(x^\infty)$ of x^∞ such that

$$\left|\frac{d}{d\theta} f\{x(\theta), y[x(\theta)]\}\right| \geqslant \frac{1}{2}\left|v_{s^\infty}^\infty\right| > 0 \quad \text{if} \quad x(\theta) \in N'(x^\infty) \tag{T.1}$$

$$\alpha \leqslant y[x(\theta)] \leqslant \beta. \tag{T.2}$$

From now on, we shall assume in order to simplify the presentation that $v_{s^\infty}^\infty$ is positive. It is then possible to find a neighborhood $N''(x^\infty)$ of x^∞ contained in $N'(x^\infty)$ and a positive number δ such that

$$x + \theta h^\infty \in N'(x^\infty) \quad \text{if} \quad x \in N''(x^\infty) \quad \text{and} \quad 0 \leqslant \theta \leqslant \delta. \tag{T.3}$$

Selecting a subsequence K_3 of K_2 such that $x^k \in N''(x^\infty)$ for $k \in K_3$, it is clear by (T.1), (T.2) and (T.3) that the solution of problem (15) will lie outside of $N'(x^\infty)$ for all $k \in K_3$. Hence

$$f(X^{k+1}) \geqq f\{x^k + \delta h^\infty, y[x^k + \delta h^\infty]\},$$

or using Taylor's expansion

$$f(X^{k+1}) \geqq f(X^k) + \delta \frac{d}{d\theta} f\{x^k + \bar{\theta} h^\infty, y[x^k + \bar{\theta} h^\infty]\} \big|_{0<\bar{\theta}<\delta}$$

$$\geqq f(X^k) + \frac{\delta}{2} v_{s^\infty}^\infty \quad \text{for all} \quad k \in K.$$

But this implies $f(X^k) \underset{k \in K_3}{\longrightarrow} \infty$, and hence contradicts $X^k \underset{k \in K_3}{\longrightarrow} X^\infty$.

5. UPDATED BASIS

In this section we show that it is always possible to construct an updated basis at every feasible point of a nondegenerate program. The procedure presented is rather cumbersome. However, as will be discussed in section 6, updated bases do not have to be computed very often in practice.

Consider the mxN matrix $\left(\frac{\partial g}{\partial X}\right)$.

Because of the nondegeneracy assumption, this matrix has rank m. Let L be a subset of $\{1, ..., N\}$. By D_L we shall mean the largest determinant (in absolute value) of all the (mxm) matrices that can be obtained from $\left(\frac{\partial g}{\partial X}\right)$ by crossing out all the columns $\left(\frac{\partial g}{\partial x_l}\right)$, $l \in L$.

Defining the vector ξ by

$$\xi_j = \min\left[|X_j - A_j|, |B_j - X_j|\right] \qquad j = 1, ..., N, \tag{18}$$

we can construct an updated basis as follows.

(0) Set $L = \{j \mid X_j = A_j \text{ or } X_j = B_j\}$.

(i) Select $l^* \in \{1, ..., N\} - L$ such that

$$\frac{\xi_{l^*}}{D_{L\cup\{l^*\}}} = \min_{l \in \{1,...,N\} - L} \frac{\xi_l}{D_{L\cup\{l\}}}.$$

(Any tie can be broken arbitrarily.)

(ii) Include l^* in L. If L contains $N - m$ elements, stop. Otherwise return to (i).

The partitioning obtained by following those steps defines a valid basis; indeed, all the variables equal to one of their bounds are included in L (step 0). Moreover, for nondegenerate programs, step (i) will guarantee that $\left(\frac{\partial g}{\partial y}\right)$ is nonsingular at the current point. We can now state the following theorem.

Theorem 2 : Let S_X be the set of bases that can be constructed at X by following steps (0)-(i)-(ii). The point to set mapping ($X \rightarrow S_X$) is closed, or, in other words, bases constructed by following steps (0)-(i)-(ii) are updated.

Proof : Consider the sequences $\{X^k\}_{k \in K}$ and $\{L^k\}_{k \in K}$, where $\{X^k\}_{k \in K}$ converges to X^∞. Let S^k be the set of bases consistent with (0)-(i)-(ii) at X^k. We shall prove that if $L^k \in S^k$ for all k then there exists $K' \subset K$ such that $L^k \underset{K'}{\rightarrow} L^\infty$ and $L^\infty \in S^\infty$. By proper extraction of subsequences, one can define the sets $K' \subset K$, J and $\bar{J}$ such that

(T. 1) $$X^k \underset{k \in K'}{\rightarrow} X^\infty$$

(T. 2) $$L^k = L^\infty \quad \text{for all} \quad k \in K'$$

(T. 3) $$X_j^k \underset{k \in K'}{\rightarrow} A_j \text{ or } X_j^k \underset{k \in K'}{\rightarrow} B_j \quad \text{for} \quad j \in J$$

(T. 4) $$A_j < \lim_{K'} X_j^k < B_j \quad \text{for} \quad j \in \bar{J}.$$

(T.5) The order in which the indices are selected into L^∞ is the same for all k. Let $\Omega \equiv \{j_1, ..., j_{N-m}\}$ be the set of indices of L^∞ taken in that order.

Because of the nondegeneracy assumption, D_J^∞ computed at X^∞ is positive, and hence

$$D_{L'}^k > 0 \quad \text{for} \quad L' \subset J \text{ and } k \text{ large enough.}$$

One can then write

$$\frac{\xi_j^k}{D_{L' \cup \{j\}}^k} \rightarrow 0 \quad \text{for} \quad j \in J \text{ and } L' \subset J$$

and

$$\lim \inf \frac{\xi_j^k}{D_{L' \cup \{j\}}^k} > 0 \quad \text{for} \quad j \in \bar{J} \text{ and } L' \subset J,$$

which implies that an element of J will always be selected into L^∞ before an element of $\bar{J}$. Moreover, because of nondegeneracy, the number of elements of J will not be larger than $N - m$, and hence all the elements of J will be included in L^∞. Let $\{j_p, ..., j_{N-m}\}$ be the elements of Ω that do not belong to J. Since ξ_j^∞ is positive for those elements, the ratio $\xi_j^\infty / D_{LU\{j\}}^\infty$ is always well determined (0/0 is excluded), and the equality

$$\frac{\xi_{j_p}^k}{D_{JU\{j_p\}}^k} = \min_{\{1,\ldots,N\}-J} \frac{\xi_j^k}{D_{JU\{j\}}^k} \quad \text{for} \quad k \in K'$$

will also hold at X^∞. Hence, j_p can be chosen as the first element of $\bar{J}$ to enter L^∞. Applying the same reasoning successively for all the elements $j_p, \ldots, j_{N-m}$, one concludes that L^∞ is consistent with the construction steps (0)-(i)-(ii) and hence that these steps define an updated basis.

6. CONCLUDING REMARKS

Some modifications of the original version of the GRGS method have been introduced to prove convergence. All of them are closely related to Zangwill's convergence theory [10] and, hence, probably cannot be eliminated completely in a convergence proof. Some of those modifications are, however, cumbersome to implement computationally. So it may be useful to discuss at what price they can be eliminated in practice.

From a theoretical point of view, updated bases need not be computed at every iteration; it is clear that the convergence proof presented in this paper would carry through with some trivial modifications if updated bases were only computed every "M" iteration. A different approach to the convergence of the GRGS method would be to eliminate the construction of updated bases and to assume some « nice » behavior of the sequence $\{ L^k \}_{k \in K}$. A precise statement of this assumption can be found in [9] for the case of linear constraints. It is clear that this *anticycling assumption* can be modified to include the more general case considered here, and the reader can verify that the convergence proof of [9] can be readily adapted for GRGS. In practical problems, one can expect that cycling will not occur and that every cluster point of $\{ X^k \}_{k \in K}$ will satisfy the Kuhn-Tucker conditions. One can then conclude that updated bases need be introduced only when a « limiting point » has been found. Constructing an updated basis can then help check the Kuhn-Tucker conditions (because $\frac{\partial g}{\partial y}$ is then guaranteed to be nonsingular), and the algorithm is reinitiated if these are not satisfied.

The change of equations (13) into (13′) is motivated by lemma 1, and discussion of a similar construction can be found in [9]. One can imagine different selection rules for h which would allow one to prove lemma 1 and which might be more efficient in practice. In particular, if A_i and B_i are very far apart, for some elements $\bar{s}$ of $\{ 1, \ldots, N \}$ $\Delta'_{\bar{s}}$ might be chosen by equation (13′) although $(v_{\bar{s}})$ is small. One can then expect that problem (15) would not lead to a significant improvement in the objective function at such a point. One way to deal with this problem is to construct h according to the following procedure.

1) Let $L^k = \bar{L} \cup \bar{\bar{L}}$, where

$$\bar{L} = \{ j \mid |x_j - \alpha_j| \geqslant \varepsilon \text{ and } |x_j - \beta_j| \geqslant \varepsilon \}.$$

2) Select $\bar{s}$ and $\Delta_{\bar{s}}$ such that

$$\Delta_{\bar{s}} = \max_{j \in \bar{L}} |v_j|.$$

3) Select $\bar{\bar{s}}$ and $\Delta_{\bar{\bar{s}}}$ according to equations (13′) and by restricting j to the elements of $\bar{\bar{L}}$.

4) Take the largest of $\Delta_{\bar{s}}$ and $\Delta_{\bar{\bar{s}}}$.

This type of construction of h can easily be adapted if some of the components A_i and B_i are unbounded. Consider, for instance, the case where some of the β_j are $+\infty$. Let $\bar{L}$ be the set of those j and let $L = \bar{L} \cup \bar{\bar{L}}$. The first relation of (13′) can then be replaced by

$$\Delta_{s'} = \max \{ [\max_{j \in \bar{L}} v_j] \quad [\max_{j \in \bar{\bar{L}}} v_j(\beta_j - x_j)] ; 0 \}.$$

REFERENCES

[1] Abadie J. and Carpentier J., « Generalization of the Wolfe Reduced Gradient Method to the Case of Nonlinear Constraints » in *Optimization*, Academic Press Inc., London, 1969.

[2] Abadie J. and Guigou J., « Numerical Experiments with the GRG Method » in *Integer and Nonlinear Programming*, North-Holland Publishing Company and American Elsevier Publishing Company, 1970.

[3] Abadie J., Private Communication, June 1973.

[4] Berge C., *Espaces Topologiques et Fonctions Multivoques*, Dunod, Paris, 1959.

[5] Colville A. R., A Comparative Study on Nonlinear Programming Code, IBM — New York Scientific Center, Report No. 320.2949, June 1968.

[6] Gochet W., Loute E. and Solow D., Comparative Computer Results of Three Algorithms for Solving Prototype Geometric Programming Problems, (to be published in CERO).

[7] Shaftel T. L., Smeers Y. M. and Thompson G. L., A Simplex-Like Approach for Nonlinear Programs with Nonlinear Constraints, MSR Report No. 285, Carnegie Mellon University, Pittsburgh, July 1972.

[8] Smeers Y. M., A Convergence Proof for Shaftel and Thompson's Modular Design Algorithm, C.O.R.E. Discussion Paper No. 7332, Université Catholique de Louvain, August 1973.

[9] Zangwill W., The Convex-Simplex Method, *Management Science*, Vol. 14, pp. 221-238, 1967.

[10] Zangwill W., *Nonlinear Programming : A Unified Approach*, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1969.