RAIRO. Recherche opérationnelle

Ana Friedlander
José Mario Martínez

On the numerical solution of bound constrained optimization problems

RAIRO. Recherche opérationnelle, tome 23, n° 4 (1989), p. 319-341

<http://www.numdam.org/item?id=RO_1989__23_4_319_0>

Recherche opérationnelle/Operations Research
(vol. 23, n° 4, 1989, p. 319 à 341)

ON THE NUMERICAL SOLUTION OF BOUND CONSTRAINED OPTIMIZATION PROBLEMS (*)

by Ana FRIEDLANDER [1] and José Mario MÁRTÍNEZ [1]

Abstract. — *This paper considers the problem of maximizing a differentiable concave function subject to bound constraints and a Lipschitz condition on the gradient, using active set strategies. We introduce a general model algorithm for this class of problems. The algorithm includes a procedure for deciding to leave a face of the polytope without having reached a stationary point relative to that face but guaranteing that return is not possible. We prove a global convergence result. Among the many possible applications, we suggest using our algorithm for optimization of external penalization functions on linear programming problems. Some numerical experiments concerning this application are presented.*

Keywords : Optimization; bound constrained problems; numerical methods.

Résumé. — *Dans ce travail on résout le problème de la maximisation d'une fonction différentiable concave soumise à des restrictions de bornes sur les variables, par une méthode de restrictions actives. Un modèle d'algorithme général est proposé pour cette classe de problèmes. L'algorithme proposé contient des critères qui permettent l'abandon des faces du polytope, où un point stationnaire n'est pas nécessairement atteint, tout en garantissant l'impossibilité d'un retour à cette face. On démontre un résultat de convergence globale. Parmi les diverses applications possibles, nous suggérons l'utilisation de cet algorithme pour l'optimisation des fonctions de pénalisation externe dans des problèmes de programmation linéaire. Quelques résultats numériques concernant cette application y sont présentés.*

Mots clés : Optimisation; restrictions de bornes; méthodes numériques.

1. INTRODUCTION

We wish to consider the problem of maximizing a concave function subject to bounds on the variables. This problem (or its equivalent one, minimizing a convex function) arises frequently in applications. For instance, the special case where the objective function is quadratic is applied to finite difference discretization of free boundary problems (*see* [5, 19]), numerical simulation

(*) Received November 1988.
[1] Applied Mathematics Laboratory, IMECC-UNICAMP, CP 6065, 13081 Campinas SP, Brazil.

of friction problems in rigid body mechanics (*see* [17]), image reconstruction from projections (*see* [15]), etc.

Most successful algorithms for solving this type of problems are based on active set strategies (*see* [10, 11, 13, 19, 21]). Briefly speaking, an active set method proceeds generating iterates on a face of the polytope until either a maximum of the objective function on that face or a point on the boundary of the face is reached. In the second case, the algorithm continues working in a face of lower dimension, and only in the first case the iterates are allowed to abandon the current face and go on working in a face of higher dimension. Since function values are strictly increasing, finite convergence is obtained (*see* [19, 21]).

However, these finite convergence results are based on the fact that a finite algorithm is available for finding a stationary point on a given face, when such a point exists. No algorithm with that property exists for general concave functions, and, even in the quadratic case, the use of conjugate gradient algorithms imposes utilization of convergence criteria for inner iterations different from the very exigent stationary point condition. O'Leary [19] suggests using empirically determined tolerance parameters ε_k in order to declare convergence of the inner iteration, but she does not give a theoretical justification for this device.

In this paper we propose an active set algorithm for maximization of a concave function subject to bound constraints with the following characteristics: the criterion for leaving a face going to a higher dimension one does not assume that the current point is stationary relative to that face, but the next point is guaranteed to have a higher function value than the maximum function value on the old face. Therefore, it may be rigorously proved that, after a finite number of iterations, all iterates lie on a face whose closure contains an optimum of the problem. Moreover, inside each face, we are able to use any globally convergent algorithm for unconstrained problems, so that the ultimate rate of convergence is the one of the unconstrained algorithm chosen.

Our ideas may be used to modify existing algorithms in a rather obvious way. However, in this report we preferred to describe a particular implementation which is able to deal with large scale problems. Namely, the internal optimization algorithm is a safeguarded version of Fletcher-Reeves method, whose memory requirements are minima among conjugate gradient procedures (*see* [9, 13]). We applied this implementation to the resolution of Linear Programming problems with an External Penalization approach. We show that, under nondegeneracy conditions, the optimum of the penalized function is obtained in a finite number of steps. We present some numerical experiments.

2. MAIN RESULTS

General hypotheses

We consider the problem of maximizing a continuously differentiable concave function with bound constrained variables:

$$\begin{gathered}\text{Maximize} f(x)\\ \text{s. t. } x\in\Omega,\\ \Omega=\{x\in\mathbb{R}^n \mid l\leqq x\leqq u,\, l<u\}.\end{gathered} \tag{2.1}$$

Let us assume that g, the gradient of f satisfies a Lipschitz condition in Ω:

$$\|g(x)-g(y)\|\leqq L\|x-y\| \quad \text{for all } x, y\in\Omega. \tag{2.2}$$

($\|.\|$ denotes the 2-norm of vectors or matrices).

(2.2) implies that, for all x, $y\in\Omega$.

$$|f(y)-f(x)-g(x)^T(y-x)|\leqq(L/2)\|y-x\|^2 \tag{2.3}$$

(*see* [8]).

Let us define an open face of Ω as a set $F_I\subset\Omega$ such that
I is a (possibly empty) subset of $\{1, 2, \ldots, 2n\}$ such that i and $n+i$ cannot belong to I simultaneously, $i=1, \ldots, n$. (2.4)

$$F_I=\{x\in\Omega \mid x_i=l_i \text{ if } i\in I,\, x_i=u_i \text{ if } n+i\in I,\, l_i<x_i<u_i, \text{otherwise}\}. \tag{2.5}$$

Therefore, the set Ω is divided into 3^n disjoint faces. Let us call $\bar{F}_I$ the closure of each open face, and $\dim(F_I)$ the dimension of the smallest linear manifold which contains F_I. Of course $\dim F_I=n-\# I$.

For each $x\in\Omega$ let us define $g_P(x)$ a real n-vector such that

$$\begin{aligned} g_p(x)_i=0 \quad & \text{if} \quad (x_i=l_i \text{ and } (\partial f/\partial x_i)(x)<0)\\ & \text{or } (x_i=u_i \text{ and } (\partial f/\partial x_i)(x)>0),\\ =(\partial f/\partial x_i)&(x) \text{ otherwise.}\end{aligned} \tag{2.6}$$

Therefore, a necessary and sufficient condition for x being a global optimum of our problem (*see* [13]) is:

$$g_P(x)=0. \tag{2.7}$$

For each $x\in\bar{F}_I$ let us define $g_I(x)$ as

$$\begin{aligned} g_I(x)_i &= 0 \quad \text{if } i\in I \quad \text{or} \quad n+i\in I, \\ &= (\partial f/\partial x_i)(x) \text{ otherwise.} \end{aligned} \tag{2.8}$$

Therefore, $g_I(x)$ is the orthogonal projection of $g(x)$ on the smallest linear manifold which contains F_I. We also define, for $x\in\bar{F}_I$,

$$\begin{aligned} g_I^C(x)_i &= 0 \quad \text{if} \quad i\notin I \quad \text{and} \quad n+i\notin I, \\ &= 0 \quad \text{if} \quad (i\in I \text{ and } (\partial f/\partial x_i)(x)<0) \\ &\qquad \text{or } (n+i\in I \text{ and } \partial f/\partial x_i)(x)>0), \\ &= (\partial f/\partial x_i)(x) \text{ otherwise.} \end{aligned} \tag{2.9}$$

The vector g_I^C places a major role in the main results of this paper.

We shall name it the "chopped gradient" associated to F_I.

LEMMA 2.1: *Assume that $\bar{x}\in\bar{F}_I$ is such that*

$$f(\bar{x})\geqq f(x) \quad \textit{for all} \quad x\in\bar{F}_I. \tag{2.10}$$

Then, the two following statements are equivalent:

$$f(\bar{x})\geqq f(x) \quad \textit{for all} \quad x\in\Omega. \tag{2.11}$$

$$g_I^C(\bar{x})=0. \tag{2.12}$$

Proof: Let us assume (2.11). If $i\in I$, then $\bar{x}_i=l_i$, and so, by (2.6)-(2.7), $(\partial f/\partial x_i)(\bar{x})\leqq 0$. Analogously, if $n+i\in I$, then $\bar{x}_i=u_i$ and so, by (2.6)-(2.7), $(\partial f/\partial x_i)(\bar{x})\geqq 0$. Therefore, by (2.9), $g_I^C(\bar{x})=0$.

Now, assume (2.12). We wish to prove that $g_p(\bar{x})=0$. Thus, for each $i=1,\ldots,n$, let us consider the following three possibilities:

$$\bar{x}_i=l_i, \tag{2.13}$$

$$\bar{x}_i=u_i. \tag{2.14}$$

$$l_i<\bar{x}_i<u_i. \tag{2.15}$$

Let us consider first (2.13). We have two alternatives:

$$i \in I, \tag{2.16}$$

$$i \notin I. \tag{2.17}$$

If (2.16) holds, we have, since $g_I^C(\bar{x})_i = 0$, and using (2.9), that $(\partial f/\partial x_i)(\bar{x}) \leqq 0$. Therefore, by (2.6), $g_P(\bar{x})_i = 0$.

If $i \notin I$, then $x_i > l_i$ for all $x \in F_I$. But, by (2.10), $f(\bar{x}) \geqq f(x)$ for all $x \in F_I$, therefore $(\partial f/\partial x_i)(\bar{x}) \leqq 0$. So, $g_P(\bar{x})_i = 0$.

The same argument leads to $g_P(\bar{x})_i = 0$, when $\bar{x}_i = u_i$.

Now, if (2.15) holds, we have, by (2.10), that $(\partial f/\partial x_i)(\bar{x}) = 0$. Thus, the desired result is proved.

In Lemma 2.1 we proved that a stationary point $\bar{x}$ for $\bar{F}_I$, either is a global optimum in Ω, or has a nonnull $g_I^C(\bar{x})$. Thus, $g_I^C(\bar{x})$ should be a useful direction for escaping from a nonoptimal face. The following lemmas state this assertion more precisely.

LEMMA 2.2: *Let*

$$\bar{\alpha} = \min\{u_i - l_i, i = 1, \ldots, n\}, \tag{2.18}$$

and $x \in \bar{F}_I$ *such that*

$$g_I^C(x) \neq 0. \tag{2.19}$$

Define

$$w_I(x) = g_I^C(x)/\| g_I^C(x) \|. \tag{2.20}$$

Then,

$$x + \alpha w_I(x) \in \Omega \quad \textit{for all} \quad \alpha \in [0, \bar{\alpha}]. \tag{2.21}$$

Proof: It is sufficient to prove that

$$l_i \leqq x_i + \alpha w_I(x)_i \leqq u_i \tag{2.22}$$

for all $i = 1, \ldots, n$, $\alpha \in [0, \bar{\alpha}]$.

If $i \notin I$ and $n + i \notin I$, (2.22) is trivial, since $w_I(x)_i = 0$ by definition (2.9).

If $i \in I$, we have, since $x \in \bar{F}_I$, that $x_i = l_i$. Therefore, by (2.9), either $w_I(x)_i = 0$, or $w_I(x)_i > 0$. In any case, by (2.20),

$$w_I(x)_i \leqq 1. \tag{2.23}$$

Therefore,

$$l_i \leqq l_i + \alpha w_I(x)_i \leqq l_i + \bar{\alpha} \leqq l_i + \min\{u_i - l_i, i = 1, \ldots, n\} \leqq u_i.$$

Thus, (2.22) is proved, if $i \in I$. A similar argument leads to (2.22), if $n+i \in I$. Therefore, the desired result is proved.

LEMMA 2.3: *Let* $\alpha \in [0, \bar{\alpha}]$, D_I *the diameter of* F_I. *Assume that* F_I *does not contain a global optimum in* Ω, *and that* $\bar{x} \in \bar{F}_I$ *satisfies* (2.10). *Let* $x \in F_I$ *be such that* $g_I^C(x) \neq 0$. Then,

$$f(x + \alpha w_I(x)) - f(\bar{x}) \geqq \alpha \| g_I^C(x) \| - (L/2)\alpha^2 - \| g_I(x) \| D_I. \qquad (2.24)$$

Proof. — Since f is a concave C^1-function, we have:

$$f(\bar{x}) \leqq f(x) + \langle g_I(x), \bar{x} - x \rangle. \qquad (2.25)$$

But, using the Cauchy-Schwarz inequality,

$$\langle g_I(x), \bar{x} - x \rangle \leqq \| g_I(x) \| \| \bar{x} - x \| \leqq \| g_I(x) \| D_I.$$

Thus, by (2.25),

$$f(\bar{x}) - f(x) \leqq \| g_I(x) \| D_I. \qquad (2.26)$$

Moreover, by the concavity of f, and (2.3), we have, for all $y \in \Omega$,

$$0 \leqq f(x) + \langle g(x), y - x \rangle - f(y) \leqq (L/2) \| y - x \|^2.$$

In particular, if $y = x + \alpha w_I(x)$,

$$0 \leqq f(x) + \alpha \langle g(x), w_I(x) \rangle - f(x + \alpha w_I(x)) \leqq (L/2)\alpha^2.$$

But, by (2.9), (2.20), $\langle g(x), w_I(x) \rangle = \| g_I^C(x) \|$. Therefore,

$$0 \leqq f(x) + \alpha \| g_I^C(x) \| - f(x + \alpha w_I(x)) \leqq (L/2)\alpha^2.$$

Thus,

$$f(x + \alpha w_I(x)) - f(x) \geqq \alpha \| g_I^C(x) \| - (L/2)\alpha^2. \qquad (2.27)$$

Combining (2.26) and (2.27), we obtain (2.24).

Now, we are able to define the main model algorithm of this paper.

Algorithm 2.1

Let σ, M, θ_1, θ_2 be given constants such that $0<\sigma<(2/L)$, $0<\sigma<M<\infty$, θ_1, $\theta_2 \in (0, 1)$. If x^k is the k-th approximation to the optimum of f in Ω, $l \leqq x^k \leqq u$, $g_P(x^k) \neq 0$, the steps for obtaining x^{k+1} are the following:

Step 1: Let I be such that $x^k \in F_I$. Test the inequality

$$\| g_I^C(x^k) \| / L \geqq \bar{\alpha}. \tag{2.28}$$

If (2.28) holds, go to Step 4.

Step 2: If $(1/(2L)) \| g_I^C(x^k) \|^2 - \| g_I(x^k) \| D_I \leqq 0$, go to Step 5. Else, define $\alpha = \| g_I^C(x^k) \| / L$.

Step 3: $x^{k+1} = x^k + \alpha\, w_I(x^k)$. Stop.

Step 4: If $(L/2)\bar{\alpha}^2 - \| g_I(x^k) \| D_I \leqq 0$, go to Step 5. Else, define $\alpha = \bar{\alpha}$. Go to Step 3.

Step 5. If $x^k + \sigma g_I(x^k) \notin F_I$, go to Step 12.

Step 6. Calculate a direction d_k such that

$$x^k + d_k \in F_I, \tag{2.29}$$

$$\sigma \| g_I(x^k) \| \leqq \| d_k \| \leqq M \| g_I(x^k) \|, \tag{2.30}$$

$$\langle d_k, g_I(x^k) \rangle \geqq \theta_1 \| d_k \| . \| g_I(x^k) \|. \tag{2.31}$$

[Observe that such a direction exists, for instance $\sigma g_I(x^k)$ satisfies (2.29), (2.30), (2.31).]

Step 7: Obtain λ, x^{k+1} performing steps 8 to 11.

Step 8: $\lambda \leftarrow 1$.

Step 9: If

$$f(x^k + \lambda d_k) \geqq f(x^k) + \lambda \theta_2 \langle g_I(x^k), d_k \rangle, \tag{2.32}$$

go to step 11.

Step 10: Let $\lambda_N \in [0.1\lambda, 0.9\lambda]$, $\lambda \leftarrow \lambda_N$. Go to Step 9.

Step 11: $x^{k+1} = x^k + \lambda d_k$. Stop.

Step 12: Let $\bar{\lambda} = \max \{ \lambda \geqq 0 \mid x^k + \lambda g_I(x^k) \in \Omega \}$. $x^{k+1} = x^k + \bar{\lambda} g_I(x^k)$. Stop.

Remark: Though this model algorithm assumes a knowledge of the Lipschitz constant L in order to calculate α at Step 2, which is impractical, a modified algorithm where the steplength never depends on L is given later.

The following lemma is the main "non-returning principle" concerning Algorithm 2.1.

LEMMA 2.4: *If x^{k+1} is defined by Step 3 of Algorithm 2.1, then*

$$x^{k+1} \in \Omega - \bar{F}_I$$

and

$$f(x^{k+1}) > f(x) \quad \text{for all } x \in \bar{F}_I. \tag{2.33}$$

Proof: First, let us observe that, if x^{k+1} is computed at Step 3, we have $g_I^C(x^k) \neq 0$. In fact, if $g_I^C(x^k) = 0$, then, in Step 2,

$$(1/(2L)) \| g_I^C(x^k) \|^2 - \| g_I(x^k) \| D_I \leqq 0,$$

and the control should go to Step 5.

If x^{k+1} is computed by Step 3, one of the two following possibilities holds:

$$(1/(2L)) \| g_I^C(x^k) \|^2 - \| g_I(x^k) \| D_I > 0, \tag{2.34}$$

or

$$(L/2)\bar{\alpha}^2 - \| g_I(x^k) \| D_I > 0. \tag{2.35}$$

Let $\bar{x}$ be the optimum of f over $\bar{F}_I$. If $\alpha = \| g_I^C(x^k) \| / L$, (2.34) holds.

Thus, by (2.24),

$$\begin{aligned} f(x^k + \alpha w_I(x^k)) - f(\bar{x}) &\geqq \| g_I^C(x^k) \|^2 / L \\ &\quad - (L/2) \| g_I^C(x^k) \|^2 / L^2 - \| g_I(x^k) \| D_I \\ &= (1/(2L)) \| g_I^C(x^k) \|^2 - \| g_I(x^k) \| D_I > 0. \end{aligned} \tag{2.36}$$

If $\alpha = \bar{\alpha}$, we may use (2.24), (2.28) and (2.35), to obtain:

$$\begin{aligned} f(x^k + \alpha w_I(x^k)) - f(\bar{x}) &\geqq \bar{\alpha} \| g_I^C(x^k) \| - (L/2)\bar{\alpha}^2 - \| g_I(x^k) \| D_I \\ &\geqq \bar{\alpha}^2 L - (L/2)\bar{\alpha}^2 - \| g_I(x^k) \| D_I = (L/2)\bar{\alpha}^2 - \| g_I(x^k) \| D_I > 0. \end{aligned} \tag{2.37}$$

Therefore, (2.33) is a consequence of (2.36) and (2.37).

Since $\alpha \leqq \bar{\alpha}$, the fact that $x^{k+1} \in \Omega$ follows from (2.21). Thus $x^{k+1} \in \Omega - \bar{F}_I$, and the proof is complete.

Let us now prove that Step 12 provides a way of increasing the function value, decreasing the dimension of the face which contains the current point.

LEMMA 2.5: *If x^{k+1} is defined by Step 12, then*

$$x^{k+1} \in F_J, \text{ where } \dim(F_J) < \dim(F_I), \tag{2.38}$$

$$f(x^{k+1}) > f(x^k). \tag{2.39}$$

Proof: By the definition of g_I and $\bar{\lambda}$ at Step 12, x^{k+1} belongs to the boundary of F_I. Therefore, (2.38) is true.

Let us prove (2.39). Since $x^k + \sigma g_I(x^k) \notin F_I$, we have $g_I(x^k) \neq 0$, and $0 < \bar{\lambda} \leqq \sigma$.

Now, by (2.3), we have:

$$|f(x^k + \bar{\lambda} g_I(x^k)) - f(x^k) - \bar{\lambda} \langle g(x^k), g_I(x^k) \rangle | \leqq (L/2) \bar{\lambda}^2 \| g_I(x^k) \|^2. \tag{2.40}$$

But

$$\langle g(x^k), g_I(x^k) \rangle = \langle g_I(x^k), g_I(x^k) \rangle = \| g_I(x^k) \|^2.$$

Hence, by (2.40),

$$f(x^k + \bar{\lambda} g_I(x^k)) - f(x^k) - \bar{\lambda} \| g_I(x^k) \|^2 \geqq -(L/2) \bar{\lambda}^2 \| g_I(x^k) \|^2.$$

Therefore,

$$f(x^k + \bar{\lambda} g_I(x^k)) \geqq f(x^k) + (\bar{\lambda} - (L/2) \bar{\lambda}^2) \| g_I(x^k) \|^2. \tag{2.41}$$

But $\lambda - (L/2) \lambda^2 > 0$ for all $\lambda \in (0, 2/L)$, and $0 < \bar{\lambda} \leqq \sigma < 2/L$.

Thus, by (2.41),

$$f(x^k + \bar{\lambda} g_I(x^k)) > f(x^k),$$

and the desired result is proved.

As we have seen, both steps 3 and 12 provide ways of leaving the face F_I. On the contrary, when x^{k+1} is computed at step 11, it continues belonging to F_I. As we observed before, there exist directions d_k satisfying (2.29), (2.30), (2.31), since $\sigma g_I(x^k)$ clearly satisfies these three conditions. Now, (2.32) is a sufficient ascent condition of Armijo's rule (*see* [8, 13]), and therefore is satisfied if λ is small enough. Thus, the loop Step 9-Step 10 stops after a finite number of steps, and so, Step 11 is well defined. The following lemma guarantees that the algorithm is able to leave any face whose closure does not contain a global optimum.

LEMMA 2.6: *Assume that $\bar{F}_I$ does not contain an optimum of problem* (2.1), *and $x^k \in F_I$. Then, after a finite number of steps j, x^{k+j} is computed at steps* 3 *or* 12.

Proof: Let us suppose, by contradiction, that x^{k+j} is computed at step 11, for all $j=0, 1, 2, \ldots$ Therefore, $\{x^{k+j}, j=0, 1, 2, \ldots\}$ is an infinite sequence contained in the compact set $\bar{F}_I$. Thus, we may extract a subsequence $\{x^{k+j}, j \in K_1\}$, whose limit is $\bar{x} \in \bar{F}_I$.

Suppose that $g_I(\bar{x}) \neq 0$. Then, there exists $\varepsilon > 0$, such that

$$\| g_I(x^{k+j}) \| \geqq \varepsilon \tag{2.41}$$

for large enough $j \in K_1$ (say, $j \in K_2$).

But g_I is continuous on $\bar{F}_I$, therefore,

$$\| g_I(x^{k+j}) \| \leqq b, \tag{2.42}$$

for some $b > 0$, $j \in K_2$.

Thus, by (2.30), (2.41), (2.42),

$$\sigma\varepsilon \leqq \| d_{k+j} \| \leqq Mb \tag{2.43}$$

for all $j \in K_2$.

Therefore $\{d_{k+j} | j \in K_2\}$ is contained in a compact set of $\mathbb{R}^n$ and so, there exists a nonnull $d \in \mathbb{R}^n$ and an infinite set of indexes K_3 such that:

$$\lim d_{k+j} = d \qquad \text{for } j \in K_3. \tag{2.44}$$

Let us now consider two possibilities:

$$\lim \lambda_{k+j} = 0 \qquad \text{for } j \in K_3. \tag{2.45}$$

There exists $\gamma > 0$, K_4 an infinite subset

$$\text{of } K_3 \text{, such that } \lambda_{k+j} \geqq \gamma \text{ for all } j \in K_4. \tag{2.46}$$

Of course, (2.46) is exactly the opposite of (2.45). If (2.45) holds, by the safeguarded choice of λ_N at Step 10, there exists a sequence $\tilde{\lambda}_{k+j}$, $j \in K_3$ such that $\tilde{\lambda}_{k+j} \leqq 10\,\lambda_{k+j}$, and

$$f(x^{k+j} + \tilde{\lambda}_{k+j} d_{k+j}) < f(x^{k+j}) + \tilde{\lambda}_{k+j} \theta_2 \langle g_I(x^{k+j}), d_{k+j} \rangle$$

for all $j \in K_3$. Therefore,

$$(f(x^{k+j} + \tilde{\lambda}_{k+j} d_{k+j}) - f(x^{k+j}))/\tilde{\lambda}_{k+j} < \theta_2 \langle g_I(x^{k+j}), d_{k+j} \rangle$$

Thus, using the Mean Value Theorem, we may choose $\xi_{k+j} \in [0, 1]$, $j \in K_3$, such that:

$$\langle g_I(x^{k+j} + \xi_{k+j} \tilde{\lambda}_{k+j} d_{k+j}), d_{k+j} \rangle < \theta_2 \langle g_I(x^{k+j}), d_{k+j} \rangle. \tag{2.47}$$

Hence, taking limits on both sides of (2.47), for $j \in K_3$, we obtain:

$$\langle g_I(\bar{x}), d \rangle \leqq \theta_2 \langle g_I(\bar{x}), d \rangle. \tag{2.48}$$

Now, by (2.31),

$$\begin{aligned} \langle g_I(\bar{x}), d \rangle &= \lim_{j \in K_3} \langle g_I(x^{k+j}), d_{k+j} \rangle \\ &\geqq \lim_{j \in K_3} \theta_1 \| g_I(x^{k+j}) \| \, \| d_{k+j} \| \geqq \theta_1 \| g_I(\bar{x}) \| \, \| d \| > 0. \end{aligned} \tag{2.49}$$

Thus, (2.48) implies that $\theta_2 \geqq 1$, contrary to assumptions.

Since (2.45) is impossible, let us consider now the possibility (2.46).

In this case, $\lambda_{k+j} \in [\gamma, 1]$ for all $j \in K_4$. Thus, there exists K_5, an infinite subset of K_4 such that:

$$\lim \lambda_{k+j} = \hat{\lambda} \in [\gamma, 1] \quad \text{for all } j \in K_5.$$

But, by (2.32),

$$f(x^{k+j} + \lambda_{k+j} d_{k+j}) \geqq f(x^{k+j}) + \lambda_{k+j} \theta_2 \langle g_I(x^{k+j}), d_{k+j} \rangle \tag{2.50}$$

for all $j \in K_5$.

Taking limits on both sides of (2.50), we obtain, by (2.49):

$$f(\bar{x} + \hat{\lambda} d) \geqq f(\bar{x}) + \hat{\lambda} \theta_2 \langle g_I(\bar{x}), d \rangle > f(\bar{x}).$$

Therefore,

$$\lim_{j \in K_5} f(x^{k+j+1}) = \lim_{j \in K_5} f(x^{k+j} + \lambda_{k+j} d_{k+j}) = f(\bar{x} + \hat{\lambda} d) > f(\bar{x}).$$

But this is impossible, since $f(x^l)$ is a strictly increasing sequence and $\bar{x}$ is an accumulation point. Thus, we have proved that $g_I(\bar{x}) = 0$.

Since, by hypothesis, $\bar{x}$ is not an optimum of (2.1), we have also that $g_I^C(\bar{x}) \neq 0$. Therefore, by continuity of g_I and g_I^C, we have, for large enough $j \in K_1$,

$$\| g_I(x^{k+j}) \| D_I \leqq (1/(2L)) \| g_I^C(x^{k+j}) \|^2$$

and

$$\| g_I(x^{k+j}) \| D_I \leqq (L/2)\,\bar{\alpha}^2.$$

Thus, both tests at Step 2 or Step 4 indicate that x^{k+j+1} must be calculated at Step 3. Therefore, by Lemma 2.4, $x^{k+j+1} \notin \bar{F}_I$, contradicting the initial assumption in the proof.

So far, we proved that, either Algorithm (2.1) stops after a finite number of iterations k, finding a global solution of (2.1), or it generates an infinite sequence which satisfies the following axioms:

$$f(x^{k+1}) > f(x^k) \quad \text{for all } k = 0, 1, 2, \ldots \tag{2.51}$$

Given $x^k \in F_I$, one of the three following possibilities hold:

$$x^{k+1} \in F_I, \tag{2.52a}$$

$$x^{k+1} \in F_J, \qquad \text{where} \quad \dim F_J < \dim F_I. \tag{2.52b}$$

$$f(x^{k+1}) > f(x) \quad \text{for all } x \in \bar{F}_I. \tag{2.52c}$$

If $x^k \in F_I$, but $\bar{F}_I$ does not contain a global optimum of (2.1), then there exists $l > k$ such that $x^l \notin F_I$. (2.53)

Let us prove now that (2.51), (2.52), (2.53) are the essential properties we need to prove that Algorithm 2.1 identifies the set of active constraints at a solution of (2.1) in a finite number of iterations.

LEMMA 2.7: *If $\bar{F}_I$ does not contain a global optimum of* (2.1), *then there exist k_I such that $x^k \notin F_I$, for all $k \geqq k_I$.*

Proof: The proof is by induction on the dimension of F_I. If dim $F_I = 0$, then F_I is a vertex of Ω and $\bar{F}_I = F_I$. Therefore, if $x^k \in F_I$, we have by (2.53), that $x^{k+l} \notin F_I$ for some $l > 0$. Thus, by (2.51), $x^{k+l+j} \notin F_I$, for all $j = 0, 1, 2, \ldots$

Assume that the thesis is true for all F_J such that dim $F_J < s = \dim F_I$.

Therefore, for each J such that $\bar{F}_J$ does not contain a global optimum of (2.1), and dim $F_J < s$, we may define k_J by:

$$x^k \notin F_J \quad \text{for all } k \geqq k_J. \tag{2.54}$$

Since there exists a finite number of faces with such characteristics, we may define k_0 as the maximum of k_J defined by (2.54).

Assume, by contradiction, that F_I contains an infinite number of iterates. Hence, there exists $k_1 > k_0$ such that $x^{k_1} \in F_I$. Let $l_1 > k_1$ be the first index such that $x^{l_1} \notin F_I$. Its existence is guaranteed by (2.53). Finally, let $k_2 > l_1$ be the first index such that $x^{k_2} \in F_I$.

Consider the finite sequence $\{x^{l_1}, x^{l_1+1}, \ldots, x^{k_2-1}\}$. Define J_{l_1} by $x^{l_1} \in F_{J_{l_1}}$. Since $x^{l_1-1} \in F_I$ and there exist $k_2 > l_1$ such that $x^{k_2} \in F_I$ we must have, by (2.52),

$$\dim F_{J_{l_1}} < \dim F_I.$$

Hence, since $k_1 > k_0$, $\bar{F}_{J_{l_1}}$ contains a global optimum of the problem.

Now, if $x^{l_1+1} \in F_{J_{l_1+1}}$, we necessarily have, since $\bar{F}_{J_{l_1}}$ contains a global optimum, and (2.52), that $\dim F_{J_{l_1+1}} \leqq \dim F_{J_{l_1}} < \dim F_I$. Going on with this reasoning, we find that x^{k_2-1} also belongs to a face F_{I_2} whose dimension is less than s, and whose closure contains a global optimum of the problem. Hence, by (2.52), x^{k_2} should be such that $f(x^{k_2}) > f(x)$ for all $x \in \bar{F}_{I_2}$, which is a contradiction.

THEOREM 2.1: *The sequence generated by Algorithm* 2.1, *either stops at an iterate which is a global optimum of* (2.1), *or, if infinite, satisfies:*

$$x^k \in F_I \quad \textit{for all } k \geqq k_0,\ k_0 \textit{ large enough, and } \bar{F}_I \textit{ contains a global optimum of } (2.1). \quad (2.55)$$

Every accumulation point of (x^k) *is a global optimum of* (2.1). (2.56)

Proof: By Lemma 2.7 and (2.52) there exists k_1 such that $x^k \in \bar{F}_{k_1}$ for all $k \geqq k_1$ and $\bar{F}_{k_1}$ contains a global optimum of the problem. Moreover, if F_k is the face which contains x^k, $k \geqq k_1$, we have, by (2.52), that $(\dim F_k)$ is a decreasing sequence contained in $\{0, 1, 2, \ldots\}$. Therefore, there exists k_0 such that $\dim F_k = \dim F_{k_0}$ for all $k \geqq k_0$. Hence, (2.55) follows from (2.52) and from the fact that $\bar{F}_{k_0}$ contains a global optimum of (2.1). Therefore, for all $k \geqq k_0$, x^{k+1} is computed at Step 11 of Algorithm 2.1. Let $\bar{x}$ be a limit point of (x^k). The same reasoning used in Lemma 2.6, leads to $g_I(\bar{x}) = 0$. Therefore, by the concavity of f,

$$f(\bar{x}) \geqq f(x) \quad \text{for all } x \in \bar{F}_I.$$

In particular, $f(\bar{x}) \geqq f(x^*)$, where $x^* \in \bar{F}_I$ is a global optimum of (2.1).

So, $\bar{x}$ is a global optimum of (2.1).

Remarks: (1) The steplength α at Step 3 of Algorithm 2.1 may be too small if L is very roughly estimated. However, it is easy to verify that it may be replaced by the more practical rule:

$$x^{k+1} = x^k + \tilde{\alpha}\, w_I(x^k) \in \Omega,$$

where (2.57)

$$f(x^k + \tilde{\alpha}\, w_I(x^k)) > f(x^k) + \| g_I(x^k) \| D_I.$$

The theory above guarantees that such an $\tilde{\alpha}$ exists [it is easy to see that $\tilde{\alpha} = \alpha$ verifies (2.57), combining (2.27) and (2.34)] and the convergence proofs depend only on the property (2.52 *c*), which keeps on holding, if (2.57) holds.

(2) The only case in which x^{k+1} can lie on a face of lower dimension than the face which contains x^k is when it is calculated at Step 12. This feature is rather unpractical, since it implies that the boundary of a face may be reached only if the former current point is very close to it. However, it is easy to see that the theoretical properties of the algorithm still hold if we allow decreasing the dimension of the current face (increasing, of course, the current function value) whenever it is judged to be convenient. In order to incorporate this possibility, we introduce the following "Step 0":

Step 0. Either compute x^{k+1} as an arbitrary point satisfying $x^{k+1} \in F_J$, $\dim F_J < \dim F_I$, and $f(x^{k+1}) > f(x^k)$, or perform steps 1 to 12.

The remarks above, and the necessity of allowing natural unconstrained search directions at Step 6 of Algorithm 2.1, lead to the following practical implementation, which, of course, has the same convergence properties as Algorithm 2.1.

Algorithm 2.2

Let σ, M, θ_1, θ_2, x^k be as in Algorithm 2.1, $x^k \in F_I$.

Step 1: If $\| g_I^C(x^k) \| / L \geqq \bar{\alpha}$, go to Step 4.

Step 2: If $(1/(2L)) \| g_I^C(x^k) \|^2 - \| g_I(x^k) \| D_I \leqq 0$, go to Step 5.

Step 3: Compute $\lambda > 0$ such that

$$f(x^k + \lambda\, g_I^C(x^k)) > f(x^k) + \| g_I(x^k) \| D_I,$$

and

$$x^k + \lambda\, g_I^C(x^k) \in \Omega.$$

Set

$$x^{k+1}=x^k+\lambda g_I^C(x^k). \quad \text{Stop.}$$

Step 4: If $(L/2)\,\bar{\alpha}^2-\|g_I(x^k)\|\,D_I\leqq 0$, go to Step 5. Else, go to Step 3.

Step 5: If $x^k+\sigma g_I(x^k)\notin F_I$, go to Step 12.

Step 6.0: Choose a direction d_k satisfying (2.30) and (2.31).

Step 6.1: If $x^k+d_k\in\Omega$, go to Step 7.

Step 6.2: Compute

$$\bar{\lambda}=\max\{\lambda\geqq 0 \mid x^k+\lambda d_k\in\Omega\}. \tag{2.58}$$

Step 6.3: Replace $d_k\leftarrow\bar{\lambda}d_k$.
If d_k satisfies (2.30), go to Step 7. Else, set

$$d_k\leftarrow g_I(x^k).$$

Step 6.4: Compute $\bar{\lambda}$ by (2.58). Replace $d_k\leftarrow\bar{\lambda}d_k$.

Steps 7 *to* 12: The same as in Algorithm 2.1.

It is easy to see that Algorithm 2.2 is a particular case of Algorithm 2.1, except that, at Step 7, a point x^{k+1} may be computed belonging to a face of lower dimension than F_I. This calculation does not modify the axioms (2.51), (2.52), (2.53). Nor does the freedom introduced at Step 3, which allows taking x^{k+1} as a point satisfying (2.33), which certainly exists, since $x^k+\alpha w_I(x^k)$ satisfies (2.33). Therefore, the thesis of Theorem 2.1 is true for Algorithm 2.2.

3. IMPLEMENTATION AND NUMERICAL EXPERIMENTS

The freedom at Step 6.0 of Algorithm 2.2 allows to choose d_k as any safeguarded direction [in the sense of (2.30), (2.31)] derived from unconstrained optimization algorithms (*see* [8,13]). In fact, we may consider the problem inside F_I as an unconstrained problem, with the variables $\{i \mid i\notin I, n+i\notin I\}$ being independent free variables. Newton and Quasi-Newton type directions may be considered, giving strong local quadratic or superlinear convergence results, in addition to the global properties stated in Theorem 2.1. Obviously, a local convergence result (without "order") may be obtained under the sole assumption that $\bar{F}_I$ contains only one global solution of (2.1).

In our implementation of Algorithm 2.2, we decided to use Fletcher-Reeves conjugate-gradient formula, since our main interest is large-scale optimization problems. Therefore, Step 6.0 was decomposed as follows:

Step 6.0.1: If $x^{k-1} \notin F_I$, go to Step 6.0.7.

Step 6.0.2: KON $\leftarrow$ KON + 1. If KON > dim F_I; go to Step 6.0.7.

Step 6.0.3: Set $\hat{d}_k = g_I(x^k) - (\| g_I(x^k) \| / \| g_I(x^{k-1}) \|) \hat{d}_{k-1}$.

Step 6.0.4: If $\hat{d}_k$ does not satisfy (2.31), go to step 6.0.7.

Step 6.0.5: Consider the following problem:

$$\begin{gathered} \text{Maximize} \, f(x^k + \lambda \hat{d}_k) \\ \text{s. t. } x^k + \lambda \hat{d}_k \in \Omega, \\ \sigma \| g_I(x^k) \| \leqq \| \lambda \hat{d}_k \| \leqq M \| g_I(x^k) \|. \end{gathered} \tag{3.1}$$

Compute an approximation $\hat{\lambda}$ to the solution of (3.1) (Use, for instance, GSRCH [20], with small GRHTOL).

Step 6.0.6: $d_k \leftarrow \hat{\lambda} \hat{d}_k$ [Observe that $\hat{d}_k$ authomatically satisfies (2.30)-(2.31)]. Go to Step 7.

Step 6.0.7: KON $\leftarrow$ 0. $\hat{d}_k \leftarrow g_I(x^k)$. Go to Step 6.0.5.

Steps 6.0.1-6.0.7 produce a direction d_k which certainly satisfies (2.29)-(2.31). So, Steps 6.1 to 6.4 are not necessary in this case.

The efficiency of this implementation is determined by the accuracy and economy in the solution of (3.1). A preliminary search based in (3.1) is needed to guarantee a good behavior of the $C-G$ algorithm. Normally, after solving (3.1) with a good accuracy, $\lambda = 1$ is accepted at Step 7 of Algorithm 2.1, specially if a tiny parameter (say $\theta_2 = 10^{-4}$) is used. The counter KON guarantees that the gradient direction is considered when the number of inner iterations reaches the dimension of the face.

Let us now describe the type of problems to which we applied the first version of our algorithm. Consider the Linear Programming Problem:

$$\begin{gathered} \text{Maximize} \, c^T x \\ \text{s. t. } \mathrm{A}x = b, \qquad \mathrm{A} \in \mathbb{R}^{m \times n}, \\ l \leqq x \leqq u. \end{gathered} \tag{3.2}$$

Assume that B is a nonsingular $m \times m$ submatrix of A. Without loss of generality, $A=(B, N)$, and (3.2) may be written as:

$$\begin{gathered}\text{Maximize } c_B^T x_B + c_N^T x_N \\ \text{s. t. } Bx_B + Nx_N = b \\ l_B \leqq x_B \leqq u_B, \qquad l_N \leqq x_N \leqq u_N.\end{gathered} \tag{3.3}$$

Eliminating variables x_B, the problem becomes:

$$\begin{gathered}\text{Maximize } (c_N^T - c_B^T B^{-1} N) x_N \\ \text{s. t. } l_N \leqq x_N \leqq u_N\end{gathered} \tag{3.4}$$

$$l_B \leqq B^{-1}(b - Nx_N) \leqq u_B. \tag{3.5}$$

The constraints (3.5) are the difficult ones, so, we incorporate them to the objective function using a large real parameter ρ, so that a solution of (2.62) may be obtained as a limit, when $\rho \to \infty$, of solutions of (3.6) (*see* [13]).

$$\begin{gathered}\text{Maximize } (c_N^T - c_B^T B^{-1} N) x_N - \rho \sum_{i=1}^{m} (\max\{[B^{-1}(b - Nx_N)]_i - u_i, \\ l_i - [B^{-1}(b - Nx_N)]_i, 0\})^2 \\ \text{s. t. } l_N \leqq x_N \leqq u_N.\end{gathered} \tag{3.6}$$

Problem (3.6) is a particular case of (2.1). The constant L is not difficult to estimate if B is simple enough, and can be estimated, in any case, using, for instance LINPACK estimator [6]. The objective function $f(x_N)$ of (3.6) is a piecewise quadratic function. So, we can expect finite convergence in a small number of steps (*see* [13]) of the $C-G$ version of Algoeithm 2.2, if f is defined by only one quadratic in a neighborhood of a solution. The following theorem states sufficient conditions for that property.

THEOREM 3.1: *Assume that* (3.2) *has a unique nondegenerate solution* x^*, *and that the Lagrange multipliers associated to the constraints* $l \leqq x \leqq u$ *are nonnull. Then, there exists* $\rho_0 > 0$ *such that, for* $\rho \geqq \rho_0$, *the objective function* $f(x_N)$ *of* (3.6) *is defined as a single quadratic function in some neighborhood of* x^*.

Proof: Let us consider the problem in the form (3.4)-(3.5). So $x^* = \begin{pmatrix} x_B^* \\ x_N^* \end{pmatrix}$.

Define

$$K_B^+=\{i\in\{1,\ldots,n\}\,|\,(x_B^*)_i=u_i\}$$
$$K_B^-=\{i\in\{1,\ldots,n\}\,|\,(x_B^*)_i=l_i\}$$
$$K_N^+=\{i\in\{1,\ldots,n\}\,|\,(x_N^*)_i=u_i\}$$
$$K_N^-=\{i\in\{1,\ldots,n\}\,|\,(x_N^*)_i=l_i\}.$$

Therefore, the optimality conditions for (3.4)-(3.5) are:

$$c_N-(B^{-1}N)^T c_B+\sum_{i\in K_B^+}\lambda_i^+(-B^{-1}N)_i^T+\sum_{i\in K_B^-}\lambda_i^-(B^{-1}N)_i^T$$
$$+\sum_{i\in K_N^+}\mu_i^+ e_i+\sum_{i\in K_N^-}\mu_i^-(-e_i)=0, \qquad (3.7)$$
$$\lambda_i^+,\quad \lambda_i^-,\quad \mu_i^+,\quad \mu_i^-<0,$$

where $(B^{-1}N)_i$ denotes the i-th column of $B^{-1}N$ and $\{e_1,\ldots,e_n\}$ is the canonical basis of $\mathbb{R}^n$.

Now, let $x_N(\rho)$ be a solution of (3.6), and define:

$$\bar{K}_N^+(\rho)=\{i\in\{1,\ldots,n\}\,|\,(x_N(\rho))_i=u_i\},$$
$$\bar{K}_N^-(\rho)=\{i\in\{1,\ldots,n\}\,|\,(x_N(\rho))_i=l_i\}.$$

Let us call

$$R_i(x_N)=\max\{[B^{-1}(b-Nx_N)]_i-u_i,\,l_i-[B^{-1}(b-Nx_N)]_i,\,0,\,i=1,\ldots,m\}$$

Hence, the optimality conditions for (3.6) are:

$$c_N-(B^{-1}N)^T c_B-2\rho\sum_{x_i(\rho)>u_i}^{i\leqq m}R_i(x_N(\rho))(-B^{-1}N)_i^T$$
$$-2\rho\sum_{x_i(\rho)<l_i}^{i\leqq m}R_i(x_N(\rho))(B^{-1}N)_i^T$$
$$+\sum_{i\in\bar{K}_N^+(\rho)}\delta_i^+(\rho)e_i+\sum_{i\in\bar{K}_N^-(\rho)}\delta_i^-(\rho)(-e_i)=0, \qquad (3.8)$$
$$\delta_i^+(\rho),\ \delta_i^-(\rho)\leqq 0.$$

But, since the solution of (3.2) is unique, we have:

$$\lim_{\rho \to \infty} x_N(\rho) = x_N^*,$$

$$\lim_{\rho \to \infty} x_B(\rho) = x_B^*.$$

Therefore, there exists $\rho_0 > 0$ such that, for $\rho \geqq \rho_0$,

$$\bar{K}_N^+(\rho) \subset K_N^+, \qquad \bar{K}_N^-(\rho) \subset K_N^-, \tag{3.9}$$

and, if $i \notin K_B^+ \cup K_N^+$, $i = 1, \ldots, m$, then:

$$l_i < x_i(\rho) < u_i. \tag{3.10}$$

Hence, for $\rho \geqq \rho_0$,

$$\bar{K}_B^+(\rho) = \{ i \in \{1, \ldots, m\} \mid x_i(\rho) > u_i \} \subset K_B^+ \tag{3.11}$$

and

$$\bar{K}_B^-(\rho) = \{ i \in \{1, \ldots, m\} \mid x_i(\rho) < l_i \} \subset K_B^-. \tag{3.12}$$

Thus, by (3.8), (3.11), (3.12), we have:

$$\begin{aligned} c_N - (B^{-1} N)^T c_B - 2\rho \sum_{i \in \bar{K}_B^+(\rho)} & R_i(x_N(\rho))(-B^{-1} N)_i^T \\ & - 2\rho \sum_{i \in \bar{K}_B^-(\rho)} R_i(x_N(\rho))(B^{-1} N)_i^T \\ & + \sum_{i \in \bar{K}_N^+(\rho)} \delta_i^+(\rho) e_i + \sum_{i \in \bar{K}_N^-(\rho)} \delta_i^-(\rho)(-e_i) = 0. \end{aligned} \tag{3.13}$$

Therefore, by (3.9), (3.11)-(3.13), the gradient $c_N - (B^{-1} N)^T c_B$ is a linear combination of vectors

$$(-B^{-1} N)_i^T (i \in \bar{K}_B^+(\rho) \subset K_B^+), \qquad (B^{-1} N)_i^T (i \in \bar{K}_B^-(\rho) \subset K_B^-),$$
$$e_i (i \in \bar{K}_N^+(\rho) \subset K_N^+), \qquad \text{and} \qquad -e_i (i \in \bar{K}_N^-(\rho) \subset K_N^-).$$

Now, by the nondegeneracy assumption, and (3.7), we have:

$$\bar{K}_B^+(\rho) = K_B^+, \qquad \bar{K}_B^-(\rho) = K_B^-, \qquad \bar{K}_N^+(\rho) = K_N^+, \qquad \bar{K}_N^-(\rho) = K_N^-. \tag{3.14}$$

We claim that, in a neighborhood of $x_N^*(\rho)$,

$$f(x_N)=(c_N^T-c_B^T B^{-1} N)\,x_N-\rho\Big\{\sum_{i\in K_B^+}[(B^{-1}(b-Nx_N)]_i-u_i)^2 + \sum_{i\in K_B^-}(l_i-[B^{-1}(b-Nx_N)]_i)^2\Big\}, \quad (3.15)$$

But, by (3.10), if $i\notin K_B^+\cup K_N^+$, $i\in\{1,\ldots,m\}$, we have $l_i<x_i<u_i$ in a neighborhood of $x^*(\rho)$. Hence,

$$\max\{[B^{-1}(b-Nx_N)]_i-u_i,\, l_i-[B^{-1}(b-Nx_N)]_i,\, 0\}=0.$$

and the expression (3.15) for $f(x_N)$ follows from (3.6).

We considered the following test problems for our numerical experiments:

$$\begin{aligned} &\text{Maximize} \sum_{i=1}^{n} x_i \\ &\text{s. t. } x_i+2x_{i+1}\leqq 10, \qquad i=1,\ldots,n-1, \\ &\qquad 0\leqq x_i\leqq 20, \qquad i=1,\ldots,n. \end{aligned} \quad (3.16)$$

We may verify that the solution x^* of (3.16) may be obtained setting:

$$x_2^*=0,$$
$$x_i+2x_{i+1}=10, \qquad i=1,\ldots,n-1.$$

Problem (3.16) may be put in the form (3.2) introducing slack variables in the inequality constraints. $x^0=(0,\ldots,0)^T$ is a feasible initial point for (3.16). Moreover, it is a vertex of the feasible region.

However, the only active constraint at x^0 which is still active at x^*, is $x_2=0$. Therefore, the Simplex method should use at least $n-1$ iterations for reaching the optimum, starting from x^0 (*see* [13]). Hence, it is interesting to study the behavior of algorithms like 2.1, 2.2, with the implementation features described at the beginning of this section, for these problems.

We applied our algorithm to (3.16), with $x^0=(0,\ldots,0)^T$, and the following algorithmic parameters: $\rho=10$, $M=10^3$, $\theta_1=10^{-3}$, $\theta_2=10^{-4}$, $L=\sqrt{2mn}$, $\sigma=1.99/L$. The variables x_B were chosen as the slack variables, so that $B=I$. At each iteration of the algorithm we tested the inequality

$$\max\{|x_1^k|,\|x_B^k\|_\infty\}<\min\{|x_2^k|,\ldots,|x_n^k|\}. \quad (3.17)$$

Since, at the solution of (3.16), the left hand side of (3.17) is null, and the right hand side is greater than 0.33, we judge that (3.17) is an indication that the solution is really the vertex of the polytope which is closest to x^k. We call K_1 the first k which verifies (3.17).

Now, after a finite number of steps, all the iterates verify:

$$x_2^k=0, \quad x_i^k>0, \qquad i=,\ldots,n, \qquad i\neq 2. \tag{3.18}$$

(3.18) represents the set of inequations which identify the face where the true solution lies. Therefore, we call K_2, the first k which satisfies (3.18). Table 1 shows the values of K_1 and K_2 detected in our experiments, for different n:

TABLE 1. — *Performance of the algorithms solving a penalized LP-problem.*

n	K_1	K_2
50	3	6
100	14	6
200	13	7
300	12	7
400	11	6
500	11	6

We observe that the performance of the algorithm in terms of number of iterations is relatively independent of the dimension of the problem, a feature which makes it recommendable for large scale situations. For general situations of type (3.2)-(3.6), we recommend to store, at some iterations (say, when k is multiple of a fixed integer q) the indexes of the $n-m$ variables which are closest to their bounds. The m remaining variables are natural candidates to be basic variables at a solution of (3.2). Therefore, a Simplex-type test for verifying if they really determine a solution, is performed. Some heuristic devices are needed in order to avoid repetition of tests and to deal with possibly degenerate problems. In this way, the application of our bound constrained algorithms to LP problems may be viewed as an alternative way of suggesting vertexes which are possible solutions of the problem. According to this point of view, the Simplex method is the classical way to suggest vertexes, and Interior Point Methods may also be interpreted as different ways of suggesting basic solutions (*see* [12, 16]).

4. FINAL REMARKS

Moré ([18], page 6) suggests a procedure for combining gradient projection techniques [1, 4, 7, 18] with active set strategies for solving bound constrained quadratic programming problems. Gradient projection methods are attractive because they are able to add or delete many constraints at each iteration, and because global convergence may be proved without assumptions on the concavity of f.

Moré's recommendation consists in making a suitable number of gradient projection iterations each time a stationary point of the quadratic on the current face is reached. We think that relations and possible combinations between Moré's and our approach for bound constrained problems deserve future research.

ACKNOWLEDGMENTS

The authors are indebted to FINEP, CNPq and FAPESP, for financial support given to their work.

REFERENCES

1. D. P. Bertsekas, *Projected Newton Methods for Optimization problems with simple constraints*, SIAM J. Control Optim., Vol. 20, 1982, pp. 221-246.
2. M. J. Best and K. Ritter, *An Effective Algorithm for Quadratic Minimization Problems*, MRC Tech. Rep. 1691, Mathematics Research Center, University of Wisconsin-Madison, 1976.
3. A. Bjorck, *A Direct Method for Sparse Least-Squares Problems with Lower and Upper Bounds*, Department of Mathematics, Linkoping University, Linkoping, Sweden, 1987.
4. P. H. Calamai and J. J. Moré, *Projected Gradient Methods for Linearly Constrained Problems*, Mathematical Programming, Vol. 39, 1987, pp. 93-116.
5. J. Cea and R. Glowinski, *Sur des méthodes d'optimisation par relaxation*, RAIRO R-3, 1953, pp. 5-32.
6. A. K. Cline, C. B. Moler, G. W. Stewart and J. H. Wilkinson, *An Estimate of the Condition Number of a Matrix*, SIAM J. Numer. Anal., Vol. 16, 1979, pp. 368-375.
7. R. S. Dembo and U. Tulowitzki, *On the Minimization of Quadratic Functions Subject to Box Constraints*, Working paper series B 71, School of Organization and Management, Yale University, New Haven, 1987.
8. J. E. Dennis and R. S. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, NJ, 1983.
9. R. Fletcher and C. M. Reeves, *Function Minimization by Conjugate Gradients*, Computer J., Vol. 2, 1964, pp. 149-153.

10. P. E. Gill and W. Murray, *Minimization Subject to Bounds on the Variables*, NPL report NAC 72, National Physical Laboratory, Teddington, 1976.

11. P. E. Gill and W. Murray, *Numerically Stable Methods for Quadratic Programming*, Mathematical Programming, Vol. 14, 1978, pp. 349-372.

12. P. E. Gill, W. Murray, M. A. Saunders and M. Wright, *A Note on Nonlinear Approaches to Linear Programmong*, TR SOL 86-7, Systems Optimization Laboratory, Stanford University, Stanford, 1986.

13. P. E. Gill, W. Murray and M. Wright, *Practical Optimization*, Academic Press, London-New York, 1981.

14. G. H. Golub and C. F. Van Loan, *Matrix Computations*, The John Hopkins University Press, Baltimore, 1983.

15. G. T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*. Academic Press, New York, 1980.

16. N. Karmarkar, *A New Polynomial-Time Algorithm for Linear Programming*, Combinatorica, Vol. 4, 1984, pp. 373-395.

17. P. Lötstedt, *Solving the Minimal Least Squares Problem Subject to Bounds on the Variables*, BIT, Vol. 24, 1984, pp. 206-224.

18. J. J. Moré, *Numerical Solution of Bound Constrained Problems*, ANL/MCS-TM-96, Math. and Comp. Sci. Div., Argonne National Laboratory, Argonne, Illinois, 1987.

19. D. P. O'Leary, *A Generalized Conjugate Gradient Algorithm for Solving a Class of Quadratic Programming Problems*, Linear Algebra and its Applications, Vol. 34, 1980, pp. 371-399.

20. M. J. D. Powell, *Subroutine GSRCH*, Harwell Subroutine Library, Harwell, Oxfordshire, 1980.

21. B. T. Polyak, *The Conjugate Gradient Method in Extremal Problems*, USSR Computational Mathematics and Mathematical Physics, Vol. 9, 1969, pp. 94-112.