RAIRO. RECHERCHE OPÉRATIONNELLE

J. F. BONNANS
M. BOUHTOU

The trust region affine interior point algorithm for convex and nonconvex quadratic programming

RAIRO. Recherche opérationnelle, tome 29, nº 2 (1995), p. 195-217
<http://www.numdam.org/item?id=RO_1995__29_2_195_0>

Recherche opérationnelle/Operations Research
(vol. 29, n° 2, 1995, p. 195 à 217)

THE TRUST REGION AFFINE INTERIOR POINT ALGORITHM FOR CONVEX AND NONCONVEX QUADRATIC PROGRAMMING (*)

by J. F. Bonnans [1] and M. Bouhtou [2]

Communicated by Pierre Tolla

Abstract. – *We study from a theoretical and numerical point of view an interior point algorithm for quadratic QP using a trust region idea, formulated by Ye and Tse. We show that, under a nondegeneracy hypothesis, the algorithm converges globally in the convex case. For a nonconvex problem, under a mild additional hypothesis, the sequence of points converges to a stationary point. We obtain also an asymptotic linear convergence rate for the cost that depends only on the dimension of the problem. Then we show that, provided some modifications are added to the basic algorithm, the method has a good numerical behaviour.*

Keywords: Interior points, affine algorithms, trust region, convergence, quadratic programming.

Résumé. – *Nous étudions du point de vue théorique et numérique un algorithme de points intérieurs pour la programmation quadratique convexe et non convexe. Dans cet algorithme formulé par Ye et Tse, on utilise l'idée de région de confiance. Nous montrons, sous une hypothèse de non dégénérescence, que l'algorithme converge globalement dans le cas convexe. Pour un problème non convexe, sous une hypothèse supplémentaire faible, la suite de points converge vers un point stationnaire. Nous obtenons aussi un taux de convergence du critère asymptotiquement linéaire. Celui-ci ne dépend que de la dimension du problème. Avec quelques modifications de l'algorithme original, nous montrons que la méthode a un bon comportement numérique.*

Mots clés : Points intérieurs, algorithme affine, région de confiance, convergence, optimisation quadratique.

1. INTRODUCTION

Since Karmarkar published his polynomial projective algorithm for linear programming [13], the algorithm known as the affine scaling method attracted interest from several researchers. It consists in minimizing the cost over a sequence of ellipsoids whose shape depends on the distance from the current interior feasible point to the faces of the feasible polyhedron.

(*) Received March 1994.

([1]) INRIA, BP 105, 78153 Rocquencourt, France.
([2]) Service Étude de Réseaux, EDF-DER, 1, avenue du Général de Gaulle, 91141 Clamart, France.

This method was initially developped for linear problems. It has already been proposed in 1967 by Dikin [9] who proved its global convergence under a primal nondegeneracy assumption and with scaled unit displacement stepsize. Independently Barnes [3] and Vanderbei *et al.* [26] rediscovered Dikin's method and proved its convergence for large step size assuming primal and dual nondegeneracy hypothesis. Tseng and Luo [21] proved that the method converges if a step length is close to 2^{-L} where L is the bit size in the input. Tsuchiya [21] proved global convergence with one eighth scaled stepsize, for degenerate problems. Tsuchiya and Muramatsu [24] proved the same result with a two third step length. Tsuchiya and Monteiro [25] proved that a variant of the long-step affine scaling algorithm has a two-step superlinear convergence property and Saigal [19] obtained a three-step quadratically convergent algorithm. The boundary behaviour of the method was studied by Megiddo *et al.* [26]. They showed that the path of the continued version of the method visits the neighborhoods of all vertices of the Klee-Minty cube if the starting point is chosen close to the boundary. This suggest that the complexity might be not polynomial. However, polynomial time complexity of the discrete algorithm is still an open question. Monteiro, Adler and Resende [17] showed polynomial time complexity of the primal-dual version when the starting point is close to the central path, using very short steps. In spite of this, several experimental results [1] show the good practical behavior of the algorithm.

Ye and Tse [27] extended the algorithm to convex and quadratic programs, using a trust-region idea, and assuming the primal nondegeneracy condition proved global convergence. Tsuchiya [23] proved global convergence under only the dual degeneracy hypothesis. Sun [20], using a step-size close to 2^{-L} (L is the bit size of the input), proved global convergence without nondegeneracy assumptions, but his displacement step size makes the algorithm impractical. Recently Ye [28] published some results on the affine scaling algorithm for non-convex quadratic programming and studied in particular the complexity of the computation of the solution of the minimization of a quadratic over an ellipsoid. Gonzaga and Carlos [10] proved global convergence of the first order version of the affine scaling algorithm for linear constrained convex problems under nondegeneracy assumption.

The main aim of this paper is to analyze the convergence of the trust region affine scaling method applied to convex and nonconvex quadratic programs, with a step possibly larger than 1.

The paper is organized as follows. First of all, we prove, for a nonconvex problem, under a mild additional hypothesis, that the sequence of points converges to a stationary point. In the convex case, we prove that all limit points of the sequence are global solution. Convergence of the sequence of the dual estimates is also established. We also obtain an asymptotic linear convergence rate for the cost that depends only on the dimension of the problem. Next, relaxing the choice of step size, and using a linesearch, we prove the same convergence results. Finally, encouraging numerical results for convex problems are presented.

2. THE BASIC ALGORITHM AND ITS THEORETICAL PROPERTIES

Given a quadratic cost

$$f(x) := c^t x + \frac{1}{2} x^t Q x$$

with $c \in \mathbb{R}^n$ and Q a $n \times n$ symmetric matrix, we consider the problem

$$\min f(x); \quad Ax = b; \quad x \geq 0, \tag{P}$$

with A a $p \times n$ matrix. As we do not suppose Q to be positive, problem (P) is in general nonconvex. We denote by $A^{-1}b$ the set $\{x \in \mathbb{R}^n;\ Ax = b\}$ and by F the set of feasible points, that is

$$F := \{x \in \mathbb{R}^n; \quad Ax = b,\ x \geq 0\},$$

and by $\overset{\circ}{F}$ the set of "strictly feasible" points, *i.e.*

$$\overset{\circ}{F} := \{x \in \mathbb{R}^n; \quad Ax = b,\ x > 0\}.$$

In the sequel we assume that F is bounded and $\overset{\circ}{F}$ is non empty. Define $X_k := diag\,\{x^k\}$. We consider the following algorithm, proposed by Ye and Tse [27]:

Algorithm 1

0) Choose $x^0 \in \overset{\circ}{F}$, $\delta \in (0, 1)$; $k \leftarrow 0$.

1) Compute x^{k+1} solution of

$$\min f(x); \quad Ax = b; \quad (x - x^k)^t\, X_k^{-2}\, (x - x^k) \leq \delta^2. \tag{SP}$$

if $(x^{k+1} - x^k)^t\, X_k^{-2}\, (x^{k+1} - x^k) < \delta^2$, stop

2) $k \leftarrow k + 1$. Go to 1).

The non-trivial part of the algorithm consists at each step in solving (SP). The region

$$E_k := \{x \in \mathbb{R}^n;\ (x - x^k)^t\, X_k^{-2}\, (x - x^k) \le \delta^2\}$$

can be interpreted as the Euclidian ball with radius δ after a scaling, which consists in making the change of variable $x \mapsto X_k^{-1} x$ that maps x^k to $e := (1 \dots 1)^t$. As $\delta < 1$ it follows that $E_k \cap \{x \in \mathbb{R}^n;\ Ax = b\}$ is included in $\overset{\circ}{F}$, hence the algorithm generates a sequence of strictly feasible points. Note that as (SP) has a quadratic constraint, it cannot be solved exactly; this will be discussed in section 3. In order to state our main results we need some definitions and hypotheses. Given $x \in F$, let $I(x)$ be the set of active nonnegativity constraints:

$$I(x) := \{i \in \{1, \dots, n\};\ x_i = 0\}.$$

With $I \subset \{1, \dots, n\}$ we associate the optimization problem:

$$\min f(x); \quad Ax = b; \quad x_I = 0. \tag{P_I}$$

We state for future reference the optimality system of (P_I).

$$(OS_{I(x)}) \quad \begin{cases} \nabla f(x) + A^t \lambda - \mu = 0, \\ Ax = b, \\ x_I = 0, \\ \mu_i = 0,\ \forall\, i \notin I(x). \end{cases}$$

As (P_I) is a quadratic problem with only linear equality constraints, its set of solutions is a (possibly empty) affine space.

Some of our results will use two hypotheses. The first one is

$$\text{for all } I \subset \{1, \dots, n\}, \text{ problem } (P_I) \text{ has at most one solution in } F. \tag{$H1$}$$

Note that (H1) is satisfies if Q is positive (or negative) definite.

We will also use an hypothesis of qualification of constraints for the limit-points of $\{x^k\}$: $\bar{x} \in F$ is said to be qualified if the following primal nondegeneracy hypothesis holds:

$$\text{If } (A^t \lambda)_i = 0, \quad \forall\, i \notin I(\bar{x}), \quad \text{then} \quad \lambda = 0. \tag{$H2$}$$

Note that (H2) is equivalent to $\{\lambda = 0$ whenever $A^t \lambda = \sum_{i \in I(\overline{x})} \mu_i e_i\}$. In this case $\mu_i = 0$ for all $i \in I(\overline{x})$; i.e. (H2) is no more that the hypothesis of linear independance of the gradients of active constraints. In the theorem below, hypotheses (H1) and (H2) are used for establishing the convergence to a point satisfying the first order optimality system; in the case of a convex cost, the convergence of the cost to its optimal value can be established by assuming meremy (H2).

We now state the main result of this section:

THEOREM 2.1: *Let $\{x^k\}$ be a sequence generated by Algorithm 1. Then:*

(i) *If at a given step k, it happens that $(x^{k+1} - x^k)^t X_k^{-2} (x^{k+1} - x^k) < \delta$, then x^{k+1} is a global solution of (P) and $x^l = x^{k+1}$ for all $l > k$.*

(ii) *Any limit point $\overline{x}$ of (x^k) is solution of $(P)_{I(\overline{x})}$.*

(iii) *If $(H1)$ holds, the sequence (x^k) converges to some $\overline{x}$. If in addition $(H2)$ holds then $\overline{x}$ satisfies the first-order optimality system of (P)*, i.e.

$$\left.\begin{array}{c} \nabla f(\overline{x}) + A^t \overline{\lambda} - \overline{\mu} = 0, \\ A\overline{x} = b, \qquad \overline{x} \geq 0,\ \overline{\mu} \geq 0, \qquad \overline{x}^t \overline{\mu} = 0. \end{array}\right\} \tag{2.1}$$

(iv) *If f is convex and $(H2)$ holds, then any accumulation point of the sequence (x^k) is an optimal solution of problem (P).*

The proof uses the optimality system of (SP) stated below. It is a simple extension of the known result for problems without equality constraints, see [7], [18].

LEMMA 2.1: *The point x^{k+1} solution of (SP) is characterized by the existence of $\lambda^{k+1} \in \mathbb{R}^p$, $\nu_k \geq 0$ such that*

$$\nabla f(x^{k+1}) + A^t \lambda^{k+1} + \nu_k X_k^{-2} (x^{k+1} - x^k) = 0, \tag{2.2}$$

$$A x^{k+1} = b, \tag{2.3}$$

$$\left.\begin{array}{l} \nu^k \geq 0, \quad (x^{k+1} - x^k)^t X_k^{-2} (x^{k+1} - x^k) \leq \delta^2 \\ \nu_k \left[(x^{k+1} - x^k)^t X_k^{-2} (x^{k+1} - x^k) - \delta^2\right] = 0 \end{array}\right\}, \tag{2.4}$$

$$d^t (Q + \nu_k X_k^{-2}) d \geq 0, \quad \forall d \in \mathcal{N}(A) := \{x \in \mathbb{R}^n;\ Ax = 0\}. \tag{2.5}$$

Remark 2.1: That (2.2), (2.3) and (2.5) hold is equivalent to: the function

$$\varphi_k(x) := f(x) + \frac{\nu_k}{2}(x - x^k)^t X_k^{-2}(x - x^k)$$

is convex on $A^{-1}b$ and attains its minimum on $A^{-1}b$ at x^{k+1}.

The essential ingredient of the proof of Theorem 2.1 is

PROPOSITION 2.1: *Let* $\{x^k\}$ *be a sequence generated by Algorithm* 1, *and* (ν_k, λ^k) *the associated multipliers. Then*

$$\delta^2 \sum_{l=0}^{k} \nu_l \leq 2(f(x^0) - f(x^{k+1})), \tag{i}$$

$$\delta \sum_{l=0}^{k} \|X_l(\nabla f(x^{l+1}) + A^t \lambda^{l+1})\| \leq 2(f(x^0) - f(x^{k+1})), \tag{ii}$$

$$(x^k - x^{k+1})^t Q(x^k - x^{k+1}) \to 0, \tag{iii}$$

$$If\ (\overline{x}, \hat{x})\ is\ a\ limit\ point\ of\ (x^k, x^{k+1})\ then\ I(\overline{x}) = I(\hat{x}). \tag{iv}$$

Proof:

(i) Using Remark 2.1, it follows that

$$\varphi_k(x^{k+1}) \leq \varphi_k(x^k) = f(x^k) \tag{2.6}$$

and

$$f(x^{k+1}) + \nu_k \frac{\delta^2}{2} = \varphi_k(x^{k+1}), \tag{2.7}$$

hence $\delta^2 \nu_k \leq 2(f(x^k) - f(x^{k+1}))$; point (i) follows.

(ii) From (2.2) we deduce

$$X_k[\nabla f(x^{k+1}) + A^t \lambda^{k+1}] = -\nu_k X_k^{-1}(x^{k+1} - x^k).$$

This and (2.4) imply $\|X_k(\nabla f(x^{k+1}) + A^t \lambda^{k+1})\| = \nu_k \delta$. Combining with (i), we obtain (ii).

(iii) As φ_k is quadratic it follows that

$$\varphi_k(x^k) = \varphi_k(x^{k+1}) + \nabla \varphi_k(x^{k+1})^t (x^k - x^{k+1}) + \frac{1}{2}(x^k - x^{k+1})^t \nabla^2 \varphi_k(x^{k+1})(x^k - x^{k+1}).$$

Using (2.2) and (2.4), respectively, we get

$$\nabla \varphi_k (x^{k+1})^t (x^k - x^{k+1}) = -(\lambda^{k+1})^t A (x^k - x^{k+1}) = 0,$$
$$\nu_k (x^k - x^{k+1})^t X_k^{-2} (x^k - x^{k+1}) = \nu_k \delta^2,$$

hence

$$f(x^k) = \varphi_k (x^k) = \varphi_k (x^{k+1}) + \frac{1}{2} (x^k - x^{k+1})^t Q (x^k - x^{k+1}) + \frac{\nu_k}{2} \delta^2.$$

This and (2.7) imply

$$f(x^k) - f(x^{k+1}) - \nu_k \delta^2 = \frac{1}{2} (x^k - x^{k+1})^t Q (x^k - x^{k+1}).$$

By (i), the monotonic decrease of f and the fact that $\inf (P) > -\infty$, the left hand side goes to 0 when $k \to \infty$, the result follows.

(iv) As $x^{k+1} \in E_k$, it follows that $(1 - \delta) x_i^k \leq x_i^{k+1} \leq (1 + \delta) x_i^k$, $i = 1, \ldots, n$, henceforth for a converging subsequence, $x_i^{k+1} \to 0 \text{ iff } x_i^k \to 0$; the result follows. □

Remark 2.2: Assuming Q positive definite, Ye [27] and Tsuchiya [23] used property iii) of Prop. 2.1 to prove the convergence of the sequence $\{x^k\}$. Here, we shall rather use the more general hypothesis (H1).

We need the following result, due to Mangasarian [15].

LEMMA 2.2: *If f is a convex function, then the gradient function $\nabla f(\cdot)$ is constant on the optimal solutions set of $(P_{I(\overline{x})})$.*

Proof of theorem 2.1:

(i) If it happens that $(x^{k+1} - x^k)^t X_k (x^{k+1} - x^k) < \delta$, using Lemma 2.1 and Remark 2.1, we deduce that $\nu_k = 0$ and that $\varphi_k (x) = f(x)$ attains its minimum on $A^{-1} b$ at x^{k+1}. It follows that x^{k+1} is a solution of (P).

(ii) Let $\overline{x}$ be a limit-point of $\{x^{k+1}\}$. Denote $I := I(\overline{x})$, $\overline{I} := \{1, \ldots, n\} - I$. From Prop 2.1 (ii) we get for the given subsequence

$$(\nabla f(x^{k+1}) + A^t \lambda^{k+1})_{\overline{I}} \to 0. \tag{2.8}$$

Define $G := \{(A^t \lambda)_{\overline{I}}, \lambda \in \mathbb{R}^p\}$. Then G is a closed linear subspace; from (2.8) it follows that *dist* $((\nabla f(x^{k+1})_I, G) \to 0$, hence there exist some $\overline{\lambda}$ such that $(\nabla f(\overline{x}) + A^t \overline{\lambda})_{\overline{I}} = 0$. From this it follows that $\overline{x}$ satisfies the first order optimally system of $(P)_{\overline{I}(\overline{x})}$.

Now let $d \in \mathcal{N}(A)$ be such that $d_I = 0$ (the set of such d's is possibly $\{0\}$). Then passing to the limit in (2.5), and reminding that ν_k converges to zero by Prop 2.1 (i), it follows that $\nu^k d^t X_k^{-2} d \to 0$, hence $d^t Q d \geq 0$; now if x is feasible for $(P)_I$, then $d := x - \overline{x}$ is in $\mathcal{N}(A)$ and $d_I = 0$ hence

$$f(x) = f(\overline{x}) + \nabla f(\overline{x})^t d + \frac{1}{2} d^t Q d = f(\overline{x}) + \frac{1}{2} d^t Q d \geq f(\overline{x})$$

which proves (ii).

(iii) From Prop 2.1 (iv), point (ii) and (H1) we deduce that if $(\overline{x}, \hat{x})$ is limite-point of (x^k, x^{k+1}) then $\overline{x} = \hat{x}$. In particular $\|x^{k+1} - x^k\| \to 0$ which implies that the set of limit points of $\{x^k\}$ is connected. Using (ii) and (H1) again we deduce that the all sequence $\{x^k\}$ converges to some $\overline{x}$.

$$(\nabla f(x^{k+1}) + A^t \lambda^{k+1})_i \to 0, \quad i \notin I(\overline{x}),$$

hence with (H2) $\lambda^k \to \overline{\lambda}$ such that $(\nabla f(\overline{x}) + A^t \overline{\lambda})_i = 0$ for all $i \notin I(\overline{x})$, hence $(OS_{I(\overline{x})}$ is satisfied. It remains to prove that $\overline{\mu}_I \geq 0$. From (2.2) and the convergence of $\{\lambda^k\}$ we deduce that $\overline{\mu} = \lim_{\nu_k} X_k^{-2} (x^k - x^{k+1})$. If $i \in I(\overline{x})$ then $x_i^k \to 0$, hence $x_i^{k+1} \leq x_i^k$ at least for a subsequence, and it follows that $\overline{\mu}_i \geq 0$.

(iv) Denote by μ^k the dual estimate term given by (2.2), *i.e.* $\mu^k := \nabla f(x^{k+1}) + A^T \lambda^{k+1}$. We first prove that $\{\mu^k\}$ converges: it is well known (see [10]) that whenever (H2) holds that the matrix $(AX^2 A^T)$, with $X := \operatorname{diag}(x)$, is not singular over F. We note that $\eta^k := X_k \mu^k$ is such that

$$X_k \nabla f(x^{k+1}) + (AX_k)^t \lambda^{k+1} - \eta^k = 0, \text{ and } AX_k \eta^k = 0.$$

It follows that η^k is the orthogonal projection of $X_k \nabla f(x^{k+1})$ onto $\mathcal{N}(AX_k)$, hence $\mu^k = X_k^{-1} \eta^k = [I - A^T (AX_k^2 A^T)^{-1} AX_k^2] \nabla f(x^{k+1})$. As F is bounded, we deduce that the sequence $\{\mu^k\}$ is bounded.

We prove that the set of limit-points of $\{\mu^k\}$ is finite. Let $\overline{x}$ be a limit-point of $\{x^k\}$. If x^* is an other limit-point such that $I(\overline{x}) \subset I(x^*)$, then using Theorem 2.1 (ii), the fact that $f(\overline{x}) = f(x^*)$ and Lemma 2.2, we deduce that $\nabla f(\overline{x}) = \nabla f(x^*)$. Again (H2) implies that for any subsequence $(x^{k'})$ such that $x^{k'+1} \to x^*$, we have that $\lim_k \mu^k = \lim_{k'} \mu^{k'}$. This and the finiteness of the set of faces of F imply that the set of the limit points of $\{\mu^k\}$ is finite.

Now from proposition 2.1 (iv) and proceeding as above with $k' = k+1$, it is easy to see that lim $\nabla f(x^{k'+1}) = \nabla f(\overline{x})$, hence $\{\mu^{k+1} - \mu^k\}$ vanishes. It follows that the set of accumulation points of $\{\mu^k\}$ is connected and finite. This implies that $\{\mu^k\}$ converges to some $\overline{\mu}$.

Now we are ready to prove that, if $\overline{x}$ is an accumulation point of the sequence x^k, and $\overline{\lambda}$ is the limit of the associated subsequence of λ^{k+1}, then $(OS_{I(\overline{x})})$ is satisfied with multipliers $\overline{\lambda}$ and $\overline{\mu}$. As in the proof of (iii) it remains to prove that $\overline{\mu}_I \geq 0$. Assume by contradiction that there exists $i \in I$ such that $\overline{\mu}_i < 0$. Then $\overline{x}_i = 0$ and the convergence of $\{\mu^k\}$ implies that $\nu^k (x_i^k - x_i^{k+1})/(x_i^k)^2 < 0$ for all k large enough. Hence $\overline{x}_i = 0$ and $x_i^k < x_i^{k+1}$ for all k large enough. This is a contradiction and so $\overline{\mu}_I \geq 0$. □

Remark 2.3: In statement of (iii) of Theorem 1, instead of (H1) we can obviously assume only that for any limit point $\overline{x}$ of the sequence $\{x^k\}$, $P_{I(\overline{x})}$ has at most one solution in F.

Note that C. Gonzaga and Carlos [10], under primal nondegeneracy assumption also proved the optimality conditions for the first order version of the affine scaling method applied to linearly constrained convex programs. They did not prove the convergence of the sequence of the dual estimates. However, one can easily proceed as in our proof of (iv) of Theorem 1, in order to also prove the convergence of the dual estimates generated by the first order affine scaling method.

We now analyze the rate of convergence of the algorithm. We note that even in the case of linear cost, the known results deal only with the asymptotic rate of convergence of the cost. We generalize these results in point (i) of Thm 2.2. We also give a result related to the speed of convergence of $\{x^k\}$. We denote

$$\|x\|_k := \sqrt{\sum_{i=1}^{n} (x_i/x_i^k)^2} = \sqrt{x^t X_k^{-2} x}.$$

A solution $(\overline{x}, \overline{\lambda}, \overline{\mu})$ of the optimality system (2.1) is said to be strict complementarity if

$$\overline{x}_i > 0 \quad \text{or} \quad \overline{\mu}_i > 0, \quad \forall i \in \{1, \cdots, n\}. \tag{2.9}$$

THEOREM 2.2: *Let $\{x^k\}$ be generated by Algorithm* 1. *Then:*
(i) *For all $x^* \in F$ such that $f(x^*) < f(x^{k+1})$ then*

$$f(x^{k+1}) - f(x^*) \le \left(1 - \frac{\delta^2}{\|x^k - x^*\|_k^2}\right)(f(x^k) - f(x^*)). \qquad (2.10)$$

In particular, if x^k converges to $\overline{x}$ then for some $\varepsilon_k \to 0$:

$$f(x^{k+1}) - f(\overline{x}) \le \left(1 - \frac{\delta^2}{\text{card}(I(\overline{x})) + \varepsilon_k}\right)(f(x^k) - f(\overline{x})), \qquad (2.11)$$

and

$$f(x^{k+1}) - f(\overline{x}) \le \left(1 - \frac{\delta^2}{n + \varepsilon_k}\right)(f(x^k) - f(\overline{x})). \qquad (2.12)$$

(ii) *In addition, if* (H1), (H2) *and the strict complementarity hypothesis* (2.9) *hold there*

$$\sum_k \|x^k - \overline{x}\| + \sum_k \|\lambda^k - \overline{\lambda}\| < +\infty. \qquad (2.13)$$

Proof:
(i) As φ_k attains its minimum on $A^{-1}b$ at x^{k+1}, we have

$$\varphi_k(x^{k+1}) \le \varphi_k(x^k) = f(x^k),$$
$$\varphi_k(x^{k+1}) \le \varphi_k(x^*) = f(x^*) + \frac{\nu_k}{2}\|x^* - x^k\|_k^2.$$

We obtain that $\varphi_k(x^{k+1}) \le (1-\theta)\varphi_k(x^k) + \theta\varphi_k(x^*)$, for all $\theta \in [0, 1]$. As a consequence $\varphi_k(x^{k+1}) - \varphi_k(x^*) \le (1-\theta)[f(x^k) - \varphi_k(x^*)]$, and therefore

$$f(x^{k+1}) - f(x^*) \le (1-\theta)[f(x^k) - f(x^*)] + \frac{\nu_k}{2}[\theta\|x^* - x^k\|_k^2 - \delta^2].$$

We note that, as $f(x^*) < f(x^{k+1})$, x^* must be outside the ellipsoid E_k. Choosing $\theta := \delta^2/\|x^* - x^k\|_k^2 < 1$, we deduce (2.10). Relations (2.11) and (2.12) easily follow.

(ii) Set $I := I(\overline{x})$. By Thm 2.1 (iii), $\{x^k\}$ converges to some $\overline{x}$. From Prop 2.1 (ii) and (2.9) we get

$$\sum_{k=0}^{\infty} |(\nabla f(x^{k+1}) + A^t\lambda^{k+1})_i| < \infty, \qquad \forall i \notin I. \qquad (2.14)$$

Now from Prop 2.1 (ii) and (2.9) again:

$$\sum_{k=0}^{\infty} x_i^k < \infty \quad \text{for all } i \in I. \tag{2.15}$$

Define $\eta^{k+1} \in \mathbb{R}^n$ by

$$\eta_i^{k+1} := \begin{cases} (\nabla f(x^{k+1}) + A^t \lambda^{k+1})_i & if \quad i \notin I, \\ 0 & if \quad i \notin I. \end{cases}$$

It appears that (x^{k+1}, λ^{k+1}) are primal-dual solution of the optimality system of

$$\min_x f(x) - \eta^t x; \quad Ax = b; \quad x_I = x_I^{k+1},$$

or equivalently $(x^{k+1} - \overline{x}, \lambda^{k+1} - \overline{\lambda})$ is solution of the linear system in (x, λ):

$$\left.\begin{array}{l} (Qx + A^t \lambda)_{\overline{I}} = \eta_{\overline{I}}^{k+1} \\ Ax = 0 \\ x_I = x_I^{k+1}. \end{array}\right\} \tag{2.16}$$

We claim that this system is invertible. Indeed consider a solution of the homogeneous system satisfies $0 = x^t (Qx + A^t \lambda) x = x^t Q x$. Now $\overline{x} + \rho x$ is feasible for $(P)_I$ and

$$f(\overline{x} + \rho x) = f(\overline{x}) + \rho \nabla f(\overline{x})^t x + \frac{\rho^2}{2} x^t Q_x = f(\overline{x});$$

by (H1) x must be null, hence $\lambda = 0$ (H2) which proves the invertibility of this linear system. Now by invertibility of (2.16) and (2.14), (2.15) there exists $K > 0$ such that

$$\sum_{k=0}^{\infty} (\|x^k - \overline{x}\| + \|\lambda^k - \overline{\lambda}\|) \leq K \sum_{k=0}^{\infty} (\|\eta^k\| + \|x_I^k\|) < \infty,$$

as was to be proved. □

Remark 2.4: Ye [28] also proved a similar result on the convergence rate for the cost. Our proof is quite different and seems simpler than the one of Ye.

3. THE EXTENDED ALGORITHM

So far we have analyzed Algorithm 1, assuming that the solution of (SP) could be computed exactly: this is not the case, however, as problem (SP) is nonlinear. However, guessing a value for the multiplier ν associated to the nonlinear constraint, we may solve exactly, whenever it has a solution, the problem of minimizing the associated Lagrangian:

$$\min \psi_\nu^k(x) := f(x) + \frac{\nu}{2}(x - x^k)^t X_k^{-2}(x - x^k) \quad s.t. \quad Ax = b. \quad (Q)_\nu$$

We denote a solution of $(Q)_\nu$ (whenever it exists) by x_ν^k.

As X_k^{-2} is positive definite, there exists a threshold value $\overline{\nu}_k \geq 0$ (we do not consider negative values of ν) such that $\psi_\nu^k(x)$ is convex on $\mathcal{N}(A)$, if and only if $\nu \geq \overline{\nu}_k$. Also by (2.5) this $\overline{\nu}_k$ satisfies $\overline{\nu}_k \leq \nu_k$. By classical argument [7], [18] one can prove that the function $\nu \to (x_\nu^k - x^k)^t X_k^{-2}(x_\nu^k - x^k)$ is strictly decreasing on $]\overline{\nu}_k, \infty[$ when x^k is not a stationary point. Hence ν_k can be computed efficiently within a given precision at least by a simple dichotomic procedure: see section 4.

In order to take into account the fact that (SP) cannot be solved exactly, but that the solution of the trust region subproblem can be computed for a number of values of the trust region close to δ, we allow the possibility for δ to vary at each iteration. Also we add the possibility of a linesearch in the direction computed by the subproblem. This linesearch does not give any new theoretical property, but it proved very effective in our numerical tests. The algorithm is as follows:

Algorithm 2

0) Choose $x^0 \in \overset{\circ}{F}$, $\delta \in (0, 1)$; $k \leftarrow 0$.

1) Compute $\hat{x}^{k+1}$ solution for some $\delta_k > \delta$ of

$$\min f(x); \quad Ax = b; \quad (x - x^k)^t K_k^{-2}(x - x^k) \leq \delta_k^2, \qquad (SP2)$$

the parameter δ_k being such that $\hat{x}^{k+1} > 0$ (hence it may happen that $\delta_k > 1$).

2) Linesearch: denote $d_k := \hat{x}^{k+1} - x^k$.

Fix $\gamma_k \geq 1$ such that $x^k + \gamma_k d^k > 0$.

Compute $\rho_k = \arg\min\{f(x^k + \rho d^k), \rho \in [1, \gamma_k]\}$

3) $k \leftarrow k + 1$, go to 1). □

In the analysis we will see that it is useful to have some bounds on γ_k. Specifically, we assume that there exists $\theta \in (0, 1)$ such that

$$0 < 1 + \rho_k d_i^k / x_i^k \leq \theta^{-1}. \qquad (3.17)$$

Remark 3.1: Indeed (3.17) gives a bound on ρ_k. Excluding the trivial case $(x^k - \hat{x}^{k+1}) X_k^{-2} (x^k - \hat{x}^{k+1}) < \delta_k^2$, we deduce from $\sum_i (d_i^k / x_i^k)^2 = \delta_k^2 > \delta^2$ that $|d_i^k| / x_i^k > \delta / \sqrt{n}$ for at least some i.

Now by (3.17)

$$-1 \leq \rho_k \, d_i^k / x_i^k \leq \theta^{-1} - 1$$

hence

$$\rho_k |d_i^k| / x_i^k \leq \max(\theta^{-1} - 1, \, 1) < \theta^{-1}.$$

It follows that $\rho_k \leq \sqrt{n} / \delta\theta$.

That is, if (3.17) holds we may assume that $\gamma_k \leq \dfrac{1}{\delta\theta} \sqrt{n}$.

Remark 3.2: We can check that δ_k is always bounded from above. Indeed, let $\overline{x}$ be a limit-point of $\{x^k\}$. Assuming that $\{x^k\}$ is not finite, then we find that for the associated subsequence of $\{x^k\}$

$$\delta_k^2 \leq \|\overline{x} - x^k\|_k^2 \to \operatorname{card}(I(\overline{x}))$$

Hence $\limsup\limits_{k \to \infty} \delta_k \leq \sqrt{n}$.

THEOREM 3.1: *Let $\{x^k\}$ be generated by Algorithm* 2. *If* (3.17) *holds, then:*

(i) *If at a given step k, it happens that $(\hat{x}^{k+1} - x^k)^t X_k^{-2} (\hat{x}^{k+1} - x^k) < \delta_k$, then $\hat{x}^{k+1}$ is a global solution of (P).*

(ii) *Any limit point $\overline{x}$ of $\{x^k\}$ is a solution of $(P)_{I(\overline{x})}$.*

(iii) *If* (H1) *holds, the sequence x^k converges to some $\overline{x}$. If in addition* (H2) *holds then $\overline{x}$ satisfies* (2.1), *the first-order optimality system of (P).*

(iv) *If f is convex and* (H2) *holds then any accumulation point of the sequence $\{x^k\}$ is an optimal solution of problem (P).*

For the proof of the Theorem 3.1 we need a statement corresponding to Proposition 2.1.

PROPOSITION 3.1: *Let x^k be a sequence generated by Algorithm* 2, *and (ν_k, λ^k) the associated multipliers. We assume that* (3.17) *holds. Then*

i) $\sum\limits_{l=0}^{k} (\delta_l)^2 \, \nu_l \leq 2 \, (f(x^0) - f(x^{k+1}))$,

ii) $\sum\limits_{l=0}^{k} \delta_l \|X_l (\nabla f(\hat{x}^{k+1}) + A^t \lambda^{l+1})\| \leq 2 \, (f(x^0) - f(x^{k+1})$,

iii) $(x^k - x^{k+1})^t\, Q\, (x^k - x^{k+1}) \to 0$,

iv) *If* $(\overline{x}, \hat{x})$ *is limit-point of* (x^k, x^{k+1}), *then* $I(\overline{x}) \subset I(\hat{x})$.

Proof:

i) Arguing as in the proof of Prop 2.1, we find that $\delta_k^2\, \nu_k \leq 2\,(f(x^k) - f(\hat{x}^{k+1}))$. As $f(\hat{x}^{k+1}) \geq f(x^{k+1})$, we deduce that $\delta_k^2\, \nu_k \leq 2\,(f(x^k) - f(x^{k+1}))$; point (i) follows.

ii) This can be proved as for Prop 2.1.

iii) Arguing as in the proof of Prop 2.1 we find that $(x^k - \hat{x}^{k+1})^t\, Q\, (x^k - \hat{x}^{k+1}) \to 0$. Now by Remark 3.1, ρ_k is bounded and

$$|(x^k - x^{k+1})^t\, Q\, (x^k - x^{k+1})| = (\rho_k)^2 |(x^k - \hat{x}^{k+1})^t\, Q\, (x^k - \hat{x}^{k+1})| \to 0.$$

iv) As $\hat{x}^{k+1} \in E_k$, we have $x_i^{k+1} \leq (1 + \rho_k \delta_k)\, x_i^k$. Using Remarks 3.1 and 3.2, we get $x_i^{k+1} \leq \left(1 + \dfrac{n+1}{\delta\theta}\right) x_i^k$ for k large enough, from which the conclusion follows. □

Proof of Theorem 3.1:

i) The same argument as in the proof of Thm 2.1 gives the result.

ii) Let $\overline{x}$ be a limit-point of x^k. If $\overline{x}$ is not solution of $P_{I(\overline{x})}$, let x^* be feasible for $P_{I(\overline{x})}$ and $f(x^*) < f(\overline{x})$. Arguing as in Thm 2.2 we get

$$f(\hat{x}^{k+1}) - f(x^*) \leq \left(1 - \frac{\delta_k^2}{\|x^k - x^*\|_k^2}\right)(f(x^k) - f(x^*)).$$

As $\delta_k \geq \delta$ and $f(x^{k+1}) \leq f(\hat{x}^{k+1})$ we get

$$f(x^{k+1}) - f(x^*) \leq \left(1 - \frac{\delta^2}{\|x^k - x^*\|_k^2}\right)(f(x^k) - f(x^*)).$$

For the considered subsequence, as $I(x^*) \supset I(\overline{x})$, we get

$$\|x^k - x^*\|_k^2 \to \operatorname{card} I(x^*) + \sum_{i \notin I(x^*)} (1 - x_i^*/\overline{x}_i)^2 < \infty$$

hence $f(x^k) \to f(x^*)$, in contradiction with our hypothesis.

iii) We have by point (ii) and (H1) that if $(\overline{x}, \hat{x})$ is limit point of (x^k, x^{k+1}) then $\overline{x}$ is unique solution of $P_{I(\overline{x})}$ and $\hat{x}$ is unique solution of $P_{I(\hat{x})}$. As $I(\hat{x}) \supset I(\overline{x})$ if follows (by definition of P_I) that $\hat{x}$ is feasible for $P_{I(\overline{x})}$, and $f(\hat{x}) = f(\overline{x}) = \lim f(x^k)$. This implies $\hat{x} = \overline{x}$, and in particular $\|x^{k+1} - x^k\| \to 0$, *i.e.* the set of limit points of $\{x^k\}$ is connected. By

(ii) and (H1) each of these limit-points is isolated. We deduce that $\{x^k\}$ converges towards some $\bar{x}$. Now $\|\hat{x}^{k+1} - x^k\| \leq \|x^{k+1} - x^k\|$ hence $\hat{x}^k$ has also limit $\bar{x}$. We obtain as in the proof of Thm 2.1, that $(OS_{I(\bar{x})})$ is satisfied and, thanks to (H2), $\bar{\mu} = \lim \nu_k X_k^{-2} (x^k - \hat{x}^{k+1})$. For $i \in I(\bar{x})$, $x^k \to 0$ and $x^{k+1} - x^k = \rho_k (\hat{x}^{k+1} - x^k)$ with $\rho_k > 0$ hence $x_i^k \geq \hat{x}_i^{k+1}$ for at least a subsequence, hence $\bar{\mu} \geq 0$, and then (2.1) holds.

iv) Let $\mu^k := \nabla f(\hat{x}^{k+1}) + A^T \lambda^{k+1}$ be the dual estimate. Using again the convexity of f and the nondegeneracy assumption, arguing as in the proof of iv) of theorem 2.1, we prove that $\{\mu^k\}$ is bounded and the set of its limit points is finite.

Using of Theorem 3.1 (ii) and from proposition 3.1 (iv), again as in the proof of Theorem 2.1 (iv), we deduce that $\|\mu^{k+1} - \mu^k\|$ vanishes, hence $\{\mu^k\}$ converges. Thus, it is easy to show the optimality conditions which imply that iv) holds.

Remark 3.3: Theorem 2.2 has an immediate extension to Algorithm 2.

4. A PRACTICAL ALGORITHM AND NUMERICAL RESULTS

Solving problem (SP2) is the hardest stage of algorithm 2. The linesearch is of course easy since the function is quadratic.

Problem (SP2) can be efficiently solved by the classical algorithms used to compute the displacement step in trust region methods (see Moré [18]). These algorithms generally use Newton's method to compute the multiplier ν_k which in our case satisfies $\|x_{\nu_k} - x^k\|_k = \delta_k$, where x_{ν_k} is such that:

$$\begin{pmatrix} Q + \nu_k X^{-2} & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x_{\nu_k} - x^k \\ \lambda^k \end{pmatrix} = - \begin{pmatrix} \nabla f(x^k) \\ 0 \end{pmatrix} \tag{4.18}$$

In our numerical tests, we have used instead an algorithm based on a simple dichotomic procedure. Indeed, on the one hand, we know (see the proof of Prop 2.1 (i)) that $\nu_k \in]\bar{\nu}_k, 2\delta^{-2} (f(x^k) - f(x^{k+1}))]$ where $\bar{\nu}$ and δ are as in section 3.

On the other hand, $\nu \mapsto \|x_\nu - x^k\|_k$ is strictly decreasing on $]\bar{\nu}_k, +\infty[$. Hence, an estimate of $f(x^{k+1})$ and $\bar{\nu}_k$ allows to compute an estimate of the multiplier ν_k using a bisection procedure on $]\bar{\nu}_k, 2\delta^{-2} (f(x^k) - f(x^{k+1}))]$. Unfortunately, to estimate the multiplier $\bar{\nu}_k$ is a hard problem especially when Q is indefinite. We proceeded in the following way:

Algorithm of computation of ν_k

1. $\nu_{\inf} := 0, \quad \nu_{\sup} := \frac{2}{\delta^2}\left[f(x^k) - \gamma_k\right]$ where γ_k is an under estimate of $f(x^{k+1})$

2. $\nu := \frac{1}{2}(v_{\inf} + \nu_{\sup})$.

3. • $H_\nu := Z^T QZ + \nu Z^T X_k^{-2} Z$, where Z is a basis of the null space of the matrix A.

• Solve the reduced system $H_\nu w = -Z^T \nabla f(x^k)$ by an iterative method which controls the positivity of H_ν.

• If H_ν is indefinite, $\nu_{\inf} := \nu$, go to 2.

• Else $x_\nu \leftarrow x^k + Zw$.

 – if $\|x_\nu - x^k\|_k \geq \delta$

  * if $x_\nu > 0$ stop

  * else $\nu_{\inf} \leftarrow \nu$, go to 2.

 – if $\|x_\nu - x^k\|_k < \delta$ then

  * if $\nu_{\sup} - \nu_{\inf} < \varepsilon$ where ε is a given precision then stop

  * else $\nu_{\sup} := \nu$, go to 2.

In the context case, zero is a trivial estimation for $\overline{\nu}_k$ and stage 3 of the above procedures becames simple to implement. We tested the behaviour of algorithm 2 to solve several convex quadratic problems randomly generated. We used a SUN sparc 4/65 computer. The algorithm is written in the language BASILE [4]. All test problems are of the form:

$$\min f(x); \qquad Ax = b; \qquad Bx \leq d; \qquad l \leq x \leq u. \qquad (P')$$

Introducing slack variables, $z := d - Bx$, $\overline{w} := u - x$, $\underline{w} := x - l$ and replacing x by $u - \overline{w}$, one can transform (P') to standard form problem:

$$\left.\begin{array}{cc} \multicolumn{2}{c}{\min \frac{1}{2}\overline{w}^T Q\overline{w} - \nabla f(u)^T \overline{w};} \\ A\overline{w} = Au - b; & B\overline{w} - z = Bu - d \\ \overline{w} + \underline{w} = u - l; & \overline{w} \geq 0,\ \underline{w} \geq 0,\ z \geq 0 \end{array}\right\} \qquad (P'')$$

and apply algorithm (2) to solve it. This has the disadvantage to increase the size of the problem, especially when the number of inequality constraints is large with respect to the number of variables. Indeed, denote $Z_k := \operatorname{diag}(z^k)$,

$\overline{W}_k := \operatorname{diag}(\overline{w}^k)$, $\underline{W}_k := \operatorname{diag}(\underline{w}^k)$, one will have to solve the following sub-problem $(SP3)$

$$\left.\begin{array}{l} \min \dfrac{1}{2}\,\overline{w}^T\,Q\overline{w} - \nabla\,f\,(u)^T\,\overline{w} \\ A\overline{w} = Au - b \\ B\overline{w} - z = Bu - d \\ \overline{w} + \underline{w} = u - l \\ (\overline{w} - \overline{w}^k)^T\,\overline{W}_k^{-2}\,(\overline{w} - \overline{w}^k) + (\underline{w} - \underline{w}^k)^T\,\underline{W}_k^{-2}\,(\underline{w} - \underline{w}^k) \\ \quad + (z - z^k)^T\,Z_k^{-2}\,(z - z^k) \leq \delta_k^2. \end{array}\right\} \quad (SP3)$$

For this, one need to solve as in (4.18) a linear system whose matrix is

$$\begin{bmatrix} Q + \nu_k\,\overline{W}_k^{-2} & 0 & 0 & A^T & B^T & I \\ 0 & \nu_k\,\underline{W}_k^{-2} & 0 & 0 & 0 & I \\ 0 & 0 & \nu_k\,Z_k^{-2} & 0 & -I & 0 \\ A & 0 & 0 & 0 & 0 & 0 \\ B & 0 & -I & 0 & 0 & 0 \\ I & I & 0 & 0 & 0 & 0 \end{bmatrix}$$

Equivalently, one can easily show that the ellipsoid centered at x^k in x space, given by sub-problem (SP3) after eliminating all slack variables, is

$$E_1 = \left\{ x \in \mathbb{R}^n; \quad (x - x^k)^t \operatorname{diag}\left[\left\{\frac{1}{\overline{w}_i^k}\right\}^2 + \left\{\frac{1}{\underline{w}_i^k}\right\}^2\right] \right.$$
$$\left. \times\,(x - x^k) + (x - x^k)^T\,B^T\,Z_k^{-2}\,B\,(x - x^k) \leq \delta_k^2 \right\}$$

Therefore, (SP3) is equivalent to the following subproblem:

$$\min\,f\,(x); \qquad Ax = b, \qquad x \in E_1.$$

Algorithm 2 may be generalized as follows:

Set:

$$\begin{aligned} D_k &:= \operatorname{diag}\{\min\,(x_i^k - l_i,\ u_i - x_i^k)\}, \\ Z_k &:= \operatorname{diag}\{(d - Bx^k)_i\}, \\ E_2 &:= \{x,\ (x - x^k)^t\,D_k^{-2}\,(x - x^k) \\ &\qquad + (x - x^k)^T\,B^T\,Z_k^{-2}\,B\,(x - x^k) \leq \delta_k^2\}. \end{aligned}$$

It is easily checked that E_2 strictly contains E_1 and is included, whenever $\delta^k < 1$, in $\overset{\circ}{F}$.

Hence in our tests, we have taken $M_k = D_k^{-2} + B^T Z_k^{-2} B$. It is easy to show that M_k is positive definite. Now, the linear system to have to solve is similar to (4.18):

$$\begin{bmatrix} Q + \nu_k M_k & A^T \\ A & 0 \end{bmatrix} \begin{pmatrix} x_{\nu_k} - x^k \\ \lambda^k \end{pmatrix} = - \begin{pmatrix} \nabla f \{x^k\} \\ 0 \end{pmatrix} \tag{4.19}$$

Note that when the upper bound's components are infinite, one can easily show that (SP4) is equivalent to sub-problem (SP3).

We briefly prove the convergence of the basic algorithm where the affine scaling is done using matrix D_k. In order to simplify the presentation, we focus on the particular problem

$$\min f(x); \qquad Ax = b, \qquad l \leq x \leq u. \tag{P''}$$

At each iteration of the basic algorithm we compute x^{k+1} solution of

$$\min f(x); \quad Ax = b, \; (x - x^k)^T D_k^{-2} (x - x^k) \leq \delta^2.$$

As in section 2, we check that the algorithm is convergent:

THEOREM 4.1 : i) *Any accumulation point $\overline{x}$ of the sequence (x_k) is an optimal global solution of the reduced problem:*

$$\min f(x); \qquad Ax = b; \qquad x_I = u_I; \qquad x_J = l_J$$

where

$$I := \{i \in \{2, \ldots, n\}; \quad \overline{x}_i = u_i\}$$
$$J := \{i \in \{1, \ldots, n\}; \quad \overline{x}_i = l_i\}$$

ii) *If for any two subset I and J of $\{1, \ldots, n\}$ such that $I \cap J = \emptyset$, we have that problem*

$$\min f(x); \quad Ax = b; \quad x_I = u_I; \quad x_J = l_J$$

has at most one solution, then $\{x^k\}$ converges.

iii) *If* (x^k) *converges to* $\overline{x}$ *and satisfies* (H2), *then* $\overline{x}$ *satisfies the first-order optimality system of* (P'') i.e.,

$$\begin{cases} [\nabla f(\overline{x}) + A^T \overline{\lambda}]_i = 0 & if \quad l_i < \overline{x}_i < u_i \\ [\nabla f(\overline{x}) + A^T \overline{\lambda}]_i \geq 0 & if \quad \overline{x}_i = l_i \\ [\nabla f(\overline{x}) + A^T \overline{\lambda}]_i \leq 0 & if \quad \overline{x}_i = u_i \\ A\overline{x} = b, \qquad l \leq \overline{x} \leq u & \end{cases}$$

iv) *If* f *is convex then any non degenerate limit point of* $\{x^k\}$ *is an optimal solution of* (P'').

Proof: i) Noting that the multiplier ν^k goes to zero such that $\nabla f(x^{k+1}) + A^T \lambda^{k+1} = \nu^k D_k^{-1} (x^k - x^{k+1})$, and from the definition of D_k, we deduce that $(\nabla f(x^{k+1}) + A^T \lambda^{k+1})_i \to 0$ where $i \notin I \cup J$. As in proof of (ii) of theorem 2.1, it is easy to deduce i).

ii) As $|x_i^{k+1} - x_i^k| \leq \delta \min(x_i^k - l_i, x_i^k) \, \forall i = 1, \ldots, n$, we deduce with (i) that $\|x^{k+1} - x^k\| \to 0$; consequently the set of limit points of $\{x^k\}$ is connected. By hypothesis this set is finite. It follows that $\{x^k\})$ converges.

iii) The fact that $\overline{x}$ is not degenerate implies that $\{\lambda^{k+1}\}$ converges to some $\overline{\lambda}$. If $l_i < \overline{x}_i < u_i$, we have that $(\nabla f(x^{k+1}) + A^T \lambda^{k+1})_i \to 0$. If $\overline{x}_i = l_i$, assume by contradiction that $(\nabla f(\overline{x}) + A^T \overline{\lambda})_i < 0$, hence $\nu^k (D_k^{-1} (x^k - x^{k+1}))_i < 0$ for all k large enough. Thus $l_i < x_i^k \leq x_i^{k+1}$ for all k large enough, which contradicts the fact that $\overline{x}_i = l_i$. As for the previous case, it is easy to show that if $\overline{x}_i = u_i$ then $[\nabla f(\overline{x}) + A^T \overline{\lambda}]_i \leq 0$.

iv) The proof is an immediate extension of the proof of (iv) of Theorem 2.1. □

The main aim of our numerical tests is to study the practical behaviour of Algorithm 2. We are not interested in the way to compute the initial point x_0. We also use the optimal value of the tested problem in the stopping test.

Specifically we compute the optimal value γ^* by an active set method (Casas and Pola [8]) and we stop algorithm 2 at iteration k when

$$|f(x^k) - \gamma^*| - 10^{-5} (f(x_0) - \gamma^*)/(f(x_0) - \gamma^* + 1),$$

We have set $\delta := 0.99$ and stopping parameter for ν_k equal to $\varepsilon := 10^{-16}$. The linesearch was exact and done in the direction d^k from $\hat{x}^{k+1}$ to 99% of the way to the boundary of the feasible region.

All our tests were generated such that the point $e = (1, \ldots, 1)^t$ was the initial interior point to start the algorithm. Therefore, the right hand sides b and d in (P') were built such that: $Ae = b$ and $(Be)_j + j = d_j$. Also: $u_i = i + 1$ and $l_i = -(n + 2 - i)$ and $Q = H^T H$ where H is a random (n, n) dense matrix.

At each test we have used two random acces modes to generate entries of A, B, H and C: a uniform random acces in [0, 1] and a normal random acces with mean equal to zero and variance equal to one.

We reporte here the worst results of this statistics. To compute ν_{sup} in the algorithm we need an under estimate γ_k of $f(x^{k+1})$. In our tests, we first

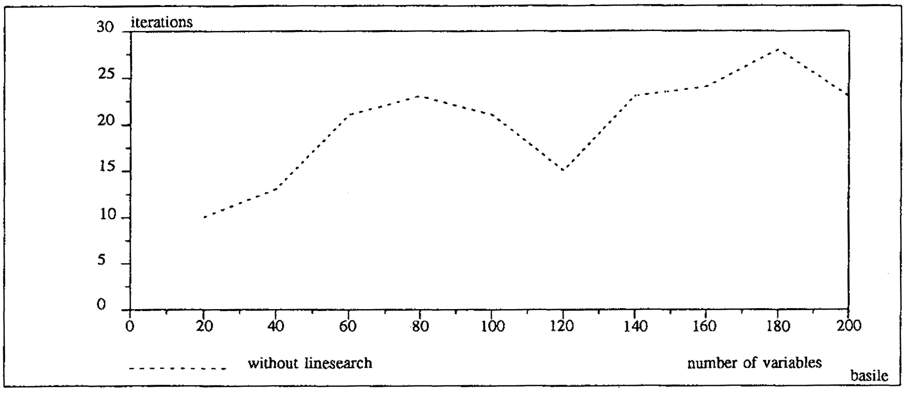

Figure 1. – Inequality constraints number fixed at 100. Number of equality constraints equal to 1. Bound constraints for all variables.

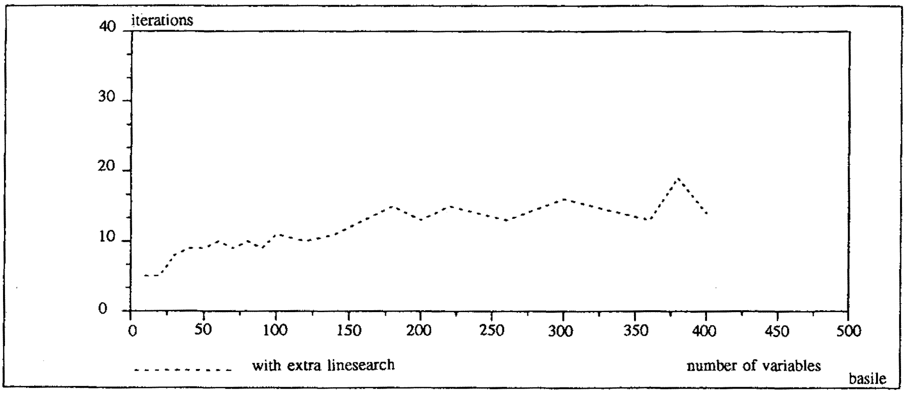

Figure 2. – Same as in figure 1 but with 200 inequality constraints.

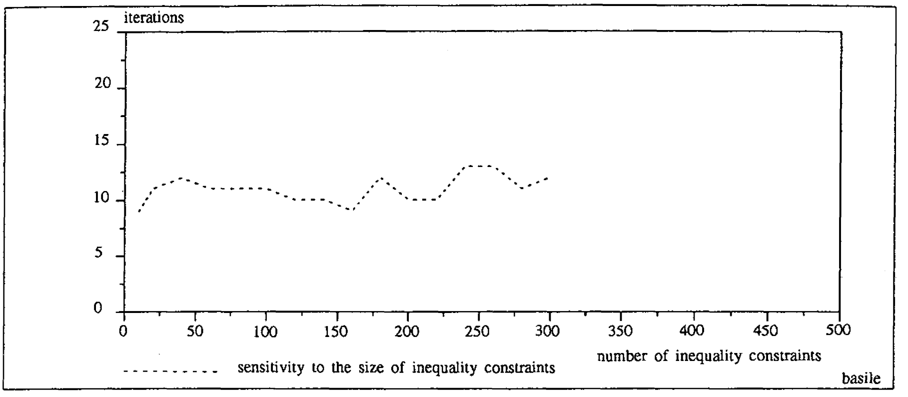

Figure 3. – The number of variables is 100. One equality constraint. Bound constraints for all variables.

computed the optimal solution of the problem without inequality constraints and stopped if it is feasible. Otherwise, we took its function value as γ_k.

Since, it is natural that the number of iterations of the algorithm decreases when the number of equality constraint increases, our tests reported here use only equality constraint. However, we always use inequality and bound constraints. The figures below summarize the principal numerical results obtained:

Result 1: Figure 1 shows that without linesearch, algorithm 2 always converges in a reasonable number of iterations related to the sizes of the tested problems. In Figure 2, we see that the use of the exact linesearch allows to divide the number of iterations roughly by two.

Result 2: In Figure 3, the same function is minimized under a variable number of inequality constraints. We observe that, by contrast to the classical active set methods, the speed of the algorithm is not very sensitive to the variation of the number of inequality constraints.

Result 3: In Figure 4 we show the importance of using ν^{k-1} as an upper estimate for the multiplier ν^k. Indeed, the dotted line in the figure 4, represents the case when we take $2\,\dfrac{f\,(x^k) - \gamma_k}{\delta^2}$ value as the initial (upper) estimate on ν_k, and for the continuous line we compute ν^k in $[0, \nu^{k-1}]$. With this last version, we see that, before the convergence, each iteration usually needs to solve two linear systems and five when close to the convergence. This is very promising and we believe that with the best choice of the upper bound of ν^k and using some preconditioner for the matrix $Q + \nu_k\, M_k$,

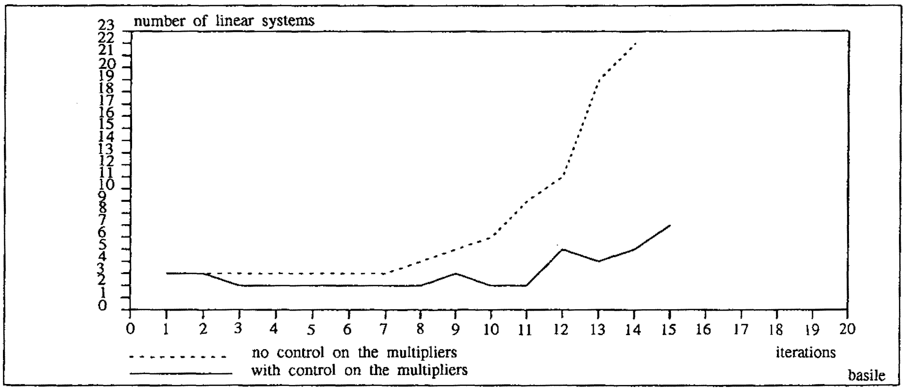

Figure 4. – The number of variables is 100. One equality constraint. Bound constraints for all variables. 100 inequality constraints.

when close to the convergence, one can reduce the number of the linear systems to be solved.

REFERENCES

1. I. Adler, M. G. C. Resende, G. Veiga, An implementation of Karmarkar's algorithm for linear programming, *Math. Programming*, 1989, *44*, pp. 297-335.
2. I. Adler, R. D. C. Monteiro, Limiting behavior of the affine scaling continuous trajectories for linear programming problems, *Math.. Programming*, 1991, *50*, pp. 29-51.
3. E. R. Barnes, A variation on Karmarkar's algorithm for solving linear programming, *Math. Programming*, 1986, *36*, pp. 174-182.
4. F. Delebecque, C. Klimann, S. Steer, *Basile. Un système de CAO pour l'automatique, Version 3.0. Guide d'utilisation*, INRIA, Mars 1989.
5. D. A. Bayer, J. C. Lagarias, The nonlinear geometry of linear programming: I Affine and projective trajectories, *Trans. Amer. Math. Soc.*, 1989, (2), *314*, pp. 499-526.
6. M. Bouthou, *Méthodes de points intérieurs pour l'optimisation des systèmes de grande taille*, Thèse de Doctorat, Université Paris Dauphine, 1993.
7. J.-P. Bulteau, J.-P. Vial, A restricted trust region algorithms for unconstrained optimization, *Optim. Theory and Appl.*, 1985, *47*, n° 4, pp. 413-435.
8. E. Casas, C. Pola, Dealing with degenerated cases in quadratic programs, *RAIRO-Operations Research*, 1994, *27*, n° 4, pp. 401 à 426.
9. I. I. Dikin, Iterative solutions of problems of linear and quadratic programming, *Soviet. Math. Doklady*, 1967, *8*, pp. 674-675.
10. C. C. Gonzaga, L. A. Carlos, *A primal affine scaling algorithm for linearly constrained convex programs*, Report, Federal University of Rio de Janeiro.
11. C. C. Gonzaga, *Convergence of the large step primal affine scaling algorithm for primal non-degenerate linear programs*, Tech. Report, Federal University of Rio de Janeiro, Brazil.
12. M. D. Hebden, *An algorithm for minimization using exact second derivatives. Atomic Energy Research*, Establishment report TP 515, Harwell, England 1973.

13. N. Karmarkar, A new polynomial time algorithm for linear programming, *Combinatorica*, 1984, *4*, pp. 373-395.
14. M. Kojima, S. Mizuno, A. Yoshise, AN $O(\sqrt{n}L)$ iteration potential reduction algorithm for linear complementarity problems, *Math. Programming*, 1991, *50*, pp. 331-342.
15. O. L. Mangasarian, A simple characterization of solution sets of convex programs, *Oper. Res. Lett.*, 1988, *7*, pp. 21-26.
16. N. Megiddo, M. Shub, Boundary behavior of interior point algorithms for linear programming, *Math. Oper. Res.*, 1989, *14*, pp. 97-146.
17. R. D. C. Monteiro, I. Adler, M. G. C. Resende, A polynomial time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extensions, *Math. Oper. Res.*, 1990, *15* (2).
18. J. J. Moré, *Recent developments in algorithms and software for trust region method*, Born 1982, Springer Verlag.
19. R. Saigal, *A three step quadratically convergent implementation of the primal affine scalling method*, Technical Report 93-9, Department of Industrial and Operational Engineering, University of Michigan, Ann-Arbor, MI-48109-2117, USA.
20. J. Sun, *A convergence proof of an affine-scaling algorithm for convex quadratic programming without nondegeneracy assumptions*, Technical Report, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL 60208, USA, 1990.
21. P. Tseng, Z. Q. Luo, *On the convergence of the affine scaling algorithm*, Technical Report, CICS-P-169, Center for Intelligent Control Systems, Massachusetts Institute of Technology (Cambridge, MA, 1989).
22. T. Tsuchiya, Global convergence of the affine scaling methods for degenerate linear programming problem, *Math. Programming*, 1991, *52*, pp. 377-404.
23. T. Tsuchiya, *Global convergence of the affine scaling algorithm for the primal degenerate strictly convex quadratic programming problems*, The Institute of Statistical Mathematics, Tokyo, Japan, Technical report n° 417.
24. T. Tsuchiya, M. Muratmatsu, *Global convergence of a long-step affine scaling algorithm for degenerate linear programming problems*, Research Memorandum 423, the Institute of Statistical Mathematics, 4-6-7 Minami-Azabou, Minoto-Ku, Tokyo 106, Japan, 1992.
25. T. Tsuchiya, R. D. C. Monteiro, *Superlinear convergence of the affine scaling algorithm*, Department of Systems and Industrial Engineering, University of Arizona, Tucson, Technical Report.
26. R. J. Vanderbei, S. M. Meketon, B. A. Freedman, A modification of Karmarkar's linear programming algorithm, *Algorithmica*, 1986, *1*, pp. 395-407.
27. Y. Ye, E. Tse, An extension of Karmarkar's projective algorithm for convex quadratic programming, *Math. Programming*, 1989, *44*, pp. 157-179.
28. Y. Ye, On affine scaling algorithms for nonconvex quadratic programming, *Math. Prog.*, 1992, *56*, 3, pp. 285-301.