A Trust region interior point algorithm for linearly constrained optimization

J. Frederic Bonnans, Cecilia Pola

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Trust Region Interior Point Algorithm for Linearly Constrained Optimization

Joseph Frédéric BONNANS
Cecilia POLA

N° 1948
Juin 1993

PROGRAMME 5

Traitement du signal,
automatique et
productique

Rapport de recherche

1993

A trust region interior point algorithm for linearly constrained optimization

Un algorithme de points intérieurs avec région de confiance pour les problèmes d'optimisation sous contraintes linéaires

J.Frédéric BONNANS [1] Cecilia POLA [2]

June 11, 1993

Abstract. We present an extension, for nonlinear optimization under linear constraints, of an algorithm for quadratic programming using a trust region idea, introduced by Ye and Tse [23] and extended by Bonnans and Bouhtou [2]. Due to the nonlinearity of the cost we use a linesearch in order to reduce the step if necessary. We prove that, under suitable hypotheses, the algorithm converges to a point satisfying the first-order optimality system, and we analyse under which conditions the unit stepsize will be asymptotically accepted.

Résumé. Nous présentons une extension à l'optimisation non linéaire sous contraintes linéaires d'un algorithme pour la programmation quadratique utilisant une idée de région de confiance, introduite par Ye and Tse [23] et étendue par Bonnans et Bouhtou [2]. A cause de la nonlinéarite du coût, nous utilisons une recherche linéaire pour réduire le pas si nécessaire. Nous montrons que, sous des hypothèses convenables, l'algorithme converge vers un point satisfaisant la condition d'optimalité du premier ordre, et nous analysons sous quelles hypothèses le pas unité sera asymptotiquement accepté.

Keywords Trust region, quadratic model, linesearch, interior points.

Mot-clés Région de confiance. modèle quadratique, recherche linéaire, points intérieurs.

1 Introduction

We study in this paper an algorithm for minimizing a nonlinear cost under linear constraints. We consider problems with linear equality constraints and non-negative variables. At each step, a direction is computed by minimizing a convex quadratic model over an ellipsoidal trust region, and then a linesearch of Armijo type is performed in this direction. At

[1]INRIA, Domaine de Voluceau - 78153 Rocquencourt - France
[2]Dpto Mat., Estad. y Comp.-Facultad de Ciencias - 39005 Santander - Spain.

each iteration, the ellipsoid of the quadratic problem is so small as to force the nonnegativity constraints to be satisfied. However the ellipsoid is not necessarily contained in the set of feasible points.

In the case of linear programming (LP) or convex quadratic programming (QP), we may assume the quadratic model to be equal to the cost function. Then the unit step will be accepted by the linesearch. In the case of LP the algorithm reduces then to the celebrated Dikin's algorithm [8]. Ye and Tse [23] have extended this algorithm to convex quadratic programming using the trust region idea. Bonnans and Bouhtou [2] have extended the method to nonconvex quadratic problems, taking a variable size for the trust region. Among the related work dealing with nonlinear costs, we quote Gonzaga and Carlos [10], in which an extension for nonlinear convex optimization of the affine-scaling linear programming algorithm (stated by Dikin [8]) is studied. Interior point algorithms for the solution of linearly constrained convex optimization problems have been studied by many other researchers; see for instance Monteiro and Adler [16], Jarre [12], Mehrotra and Sun [15], Den Hertog, Roos and Terlaky [7]. In [11] Gonzaga explores the shape of the trust regions to generate ellipsoidal regions adapted to the shape of the feasible set. The resulting algorithm generate sequences of points in the interior of the feasible set.

In this paper we obtain some results of global convergence, comparable to those obtained in [2] for quadratic programming; by global convergence we only mean that the limit points of the sequence generated by the algorithm satisfy the first-order optimality system. The main novelty of the paper, however, is in the local analysis in the vicinity of a local solution satisfying some strong second-order sufficient conditions. We check that, if such a point is a limit point of the sequence computed by the algorithm, and under a "sufficient curvature" condition satisfied by the Hessian of the quadratic approximation, then the sequence actually converges to this point and the unit step is asymptotically accepted. In this case, the behaviour of the algorithm seems to be, at least from a theoretical point of view, comparable to the QP case. We note that in the case of convex QP, some favorable results are reported in Bonnans and Bouhtou [2] and Bouhtou [5].

The paper is organized as follows. In Section 2 we present the algorithm and give a result of global convergence, in the sense that, under some convenient hypotheses, the sequence computed by the algorithm converges towards a point satisfying the first-order optimality system. Then in Section 3 we perform the local analysis: we check that if the sequence computed by the algorithm has some regular limit-point $\bar{x}$, and if a condition of "sufficient curvature" holds, then the sequence converges to this point and the unit step is asymptotically accepted.

2 The algorithm

We consider the following problem

$$(P) \qquad \min f(x);\, Ax = b;\, x \geq 0$$

where f is a smooth mapping from $\mathbf{R}^n$ in $\mathbf{R}$. A is a $p \times n$ matrix and $b \in \mathbf{R}^p$. We define the following sets:

$$A^{-1}b := \{x \in \mathbf{R}^n;\, Ax = b\},$$

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

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

so that F is the set of feasible points and $\overset{\circ}{F}$ is the set of "strictly feasible" points. In the sequel we assume that F is bounded and $\overset{\circ}{F}$ is nonempty.

We say that an $n \times n$ matrix M is nonnegative if $d^t M d \geq 0$ for all d in $\mathbb{R}^n$. The algorithm will use at each iteration two nonnegative matrices. One is a scaling matrix that takes care of the positivity constraints, and the other is an approximation to the Hessian of the cost function. To each point x^k generated by the algorithm we shall associate the $n \times n$ matrix X_k defined by $X_k := \mathit{diag}\ (x^k)$. We cons ider the following algorithm:

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

1) Choose δ_k *in* $(\delta, 1/\delta)$ *and an* $n \times n$ *nonnegative symmetric matrix* M_k*. Compute* d^k *such that* $x^k + d^k > 0$ *and* d^k *is solution of*

$$(SP) \qquad \min_d \varphi_k(d) := f(x^k) + \nabla f(x^k)^t d + \frac{1}{2} d^t M_k d;\ Ad = 0;\ d^t X_k^{-2} d \leq \delta_k^2.$$

2) If $\varphi_k(d^k) = f(x^k)$*, stop.*

3) Linesearch: Compute $\rho^k = \beta^{\ell_k}$*, with* ℓ_k *the smallest nonnegative integer such that*

$$f(x^k) - f(x^k + \beta^{\ell_k} d^k) \geq \gamma \beta^{\ell_k} (f(x^k) - \varphi_k(d^k)). \tag{2.1}$$

4) $x^{k+1} = x^k + \rho^k d^k\ ;\ k \leftarrow k+1$*. Go to 1.*

We note that if the algorithm stops at iteration k, then x^k satisfies the first-order optimality condition of (P). To see this, we need the following lemma, that states the optimality system of (SP). This is a simple extension of the known result for problems without equality constraints; see [6].

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

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

$$Ad^k = 0, \tag{2.3}$$

$$\nu_k \geq 0,\ (d^k)^t X_k^{-2} d^k \leq \delta_k^2,\ \nu_k[(d^k)^t X_k^{-2} d^k - \delta_k^2] = 0, \tag{2.4}$$

$$d^t (M_k + \nu_k X_k^{-2}) d \geq 0,\ \forall\, d \in \ker A. \tag{2.5}$$

We now come back to the discussion of Step 2 of the algorithm. Using (2.2), we deduce

$$
\begin{aligned}
f(x^k) - \varphi_k(d^k) &= -\nabla f(x^k)^t d^k - \frac{1}{2}(d^k)^t M_k d^k, \\
&= (\lambda^{k+1})^t A d^k + \nu_k (d^k)^t X_k^{-2} d^k + \frac{1}{2}(d^k)^t M_k d^k.
\end{aligned}
$$

Using (2.3) and (2.4), we get

$$
f(x^k) - \varphi_k(d^k) = \nu_k \delta_k^2 + \frac{1}{2}(d^k)^t M_k d^k. \tag{2.6}
$$

So, if $f(x^k) = \varphi_k(d^k)$, as M_k is a nonnegative matrix, then each of the nonnegative terms on the right hand side is equal 0. We deduce that $M_k^{1/2} d^k = 0$ and so $M_k d^k = 0$, and $\nu_k = 0$. Hence, using again (2.2), we get

$$
\begin{aligned}
&\nabla f(x^k) + A^t \lambda^{k+1} = 0, \\
&A x^k = b, \; x^k > 0.
\end{aligned}
$$

So x^k satisfies the first-order optimality condition of (P).

In the sequel, when studying the convergence of the algorithm, we will assume that it never stops.

Remark 2.1 *¿From Lemma 2.1 it follows that the convex quadratic function*

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

attains its minimum on $A^{-1}b$ *at* $x^k + d^k$.

In Step 3, we see that the linesearch is of Armijo type [1], i.e. it consists simply in testing the unit step, then to reduce the step by a factor $\beta < 1$ until a convenient point is found. We note that this linesearch is well defined because, by nonnegativity of M_k, the function φ_k is convex. It follows that

$$
\nabla f(x^k)^t d^k = \nabla \varphi_k(0)^t d^k \leq \varphi_k(d^k) - \varphi_k(0) = \varphi_k(d^k) - f(x^k),
$$

hence for $\rho > 0$ small enough:

$$
\begin{aligned}
f(x^k) - f(x^k + \rho d^k) &= -\rho \nabla f(x^k)^t d^k + o(\rho), \\
&\geq \rho[f(x^k) - \varphi_k(d^k)] + o(\rho).
\end{aligned}
$$

As $\gamma \in (0,1)$ and $f(x^k) > \varphi_k(d^k)$, condition (2.1) is satisfied whenever ℓ_k is large enough.

The difficult part of the algorithm consists in solving (SP) at each iteration. We will not discuss this point in detail (see Moré [17], Sorensen [20], Bouhtou [5]). Let us just mention that a typical algorithm for solving (SP) consists in guessing a value ν for the Lagrange multiplier associated to the last constraint, and then to solve the problem

$$\min_d \nabla f(x^k)^t d + \frac{1}{2} d^t M_k d + \nu d^t X_k^{-2} d;\ Ad = 0.$$

This is a convex quadratic problem under linear equality constraints whose solution is obtained by solving the associated optimality system, that reduces to a system of linear equations.

For the statement of the result of global convergence, we need some definitions. Given $x \in F$, we denote the set of active constraints by

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

To any $I \subset \{1, \dots, n\}$ we associate the optimization problem

$$(P)_I \qquad \min f(x);\ Ax = b;\ x_I = 0.$$

The first-order optimality system associated to $(P)_I$ is

$$(OS)_I \qquad \left\{ \begin{array}{l} \nabla f(x) + A^t \lambda - \mu = 0, \\ Ax = b, \\ x_I = 0;\ \mu_i = 0, i \notin I. \end{array} \right.$$

We will use the following hypotheses:

$(H1)$ For all $I \subset \{1, \dots, n\}$, system $(OS)_I$ has no nonisolated solutions.

$(H1)'$ For all $I \subset \{1, \dots, n\}$, system $(OS)_I$ has at most one solution.

$(H2)$ There exists $\alpha > 0$; $(d^k)^t (M_k + 2\nu_k X_k^{-2}) d^k \geq \alpha \|d^k\|^2$.

$(H3)$ $\left\{ \begin{array}{l} \text{The constraints of } (P) \text{ are qualified in the sense that} \\ (A^t \lambda)_i = 0, \forall\, i \notin I(\bar{x}) \text{ implies } \lambda = 0. \end{array} \right.$

We briefly discuss these hypotheses. $(H1)'$ will be satisfied in particular if f is strictly convex; $(H1)$ is a weaker condition that may be useful especially for nonconvex problems. Hypothesis $(H2)$ is a means that allows to control the decrease of the cost function at each iteration. Indeed, from (2.6) it follows easily that $(H2)$ is equivalent to

$$\text{there exists } \alpha > 0\ ;\ f(x^k) - \varphi_k(d^k) \geq \frac{\alpha}{2} \|d^k\|^2.$$

Also, $(H3)$ is no more than the hypothesis of linear independence of the gradients of active constraints.

Theorem 2.1 *Let $\{x^k\}$ be computed by Algorithm 1. We assume that $\{M_k\}$ is bounded. Then:*

(i) Any limit point $\bar{x}$ of $\{x^k\}$ is a solution of $(OS)_{I(\bar{x})}$.

(ii) If either $(H1)'$ or $(H1)$-$(H2)$ hold then $\{x^k\}$ converges. If in addition $(H3)$ holds then $\bar{x}$ satisfies the first-order optimality system of (P), i.e.

$$(OS) \qquad \begin{cases} \nabla f(\bar{x}) + A^t\bar{\lambda} - \bar{\mu} = 0, \\ A\bar{x} = b, \\ \bar{x} \geq 0, \bar{\mu} \geq 0, \bar{x}^t\bar{\mu} = 0. \end{cases}$$

The proof of the theorem uses the following lemma.

Lemma 2.2 *The sequence $\{x^k\}$ generated by Algorithm 1 satisfies*

(i) $\sum_k (f(x^k) - \varphi_k(d^k))^2 < \infty$.

(ii) $\nu_k \to 0$,

(iii) $(M_k)^{1/2}d^k \to 0$.

(iv) If in addition $\{M_k\}$ is bounded, then

$$X_k[\nabla f(x^k) + A^t\lambda^{k+1}] \to 0.$$

Proof **(i)** As F is bounded, $\{x^k\}$ and $\{d^k\}$ are bounded too. We deduce that, for some $c_1 > 0$:

$$f(x^k) - f(x^k + \rho d^k) \geq -\rho\nabla f(x^k)^t d^k - c_1\rho^2.$$

Using the convexity of φ_k, we get

$$-\nabla f(x^k)^t d^k \geq f(x^k) - \varphi_k(d^k),$$

so that

$$f(x^k) - f(x^k + \rho d^k) \geq \rho[f(x^k) - \varphi_k(d^k)] - c_1\rho^2.$$

It follows after some algebra that the linesearch test is satisfied whenever

$$\rho \leq \hat{\rho}_k := \min\{1, \frac{1-\gamma}{c_1}[f(x^k) - \varphi_k(d^k)]\}.$$

This implies that $\rho^k \geq \beta\hat{\rho}_k$. Plugging it in the linesearch test, and using the fact that, as F is bounded, $\{f(x^k)\}$ is bounded from below, we deduce that necessarily $(f(x^k) - \varphi_k(d^k))$ vanishes, and for k large enough

$$f(x^k) - f(x^{k+1}) \geq \gamma\beta\frac{1-\gamma}{c_1}(f(x^k) - \varphi_k(d^k))^2.$$

Relation (i) follows.

(ii) (iii) By (i) we get that the left hand side of (2.6) goes to 0. Then each of the nonnegative terms on the right hand side must go to 0, and that proves (ii) and (iii).

(iv) From (2.2) we deduce

$$X_k[\nabla f(x^k) + A^t\lambda^{k+1}] = -\nu_k X_k^{-1}d^k - X_k M_k d^k. \tag{2.7}$$

¿From (2.4) we have $\|\nu_k X_k^{-1}d^k\|_2 = \nu_k\delta_k$. So, using (ii) and the boundedness of $\{\delta_k\}$, it follows that $\|\nu_k X_k^{-1}d^k\| \to 0$. If in addition $\{M_k\}$ is bounded, using the boundedness of $\{X_k\}$ and (iii), we get that $X_k M_k d^k = X_k(M_k)^{1/2}(M_k)^{1/2}d^k \to 0$. Henceforth the left hand side of (2.7) goes to 0. ■

Proof of theorem 2.1

(i) Let us denote by $R(.)$ the range of an operator. Define

$$\dot{I} := \{1,\dots,n\} - I(\bar{x}).$$

¿From point (iv) of Lemma 2.2 it follows that

$$[\nabla f(x^k) + A^t\lambda^{k+1}]_I \to 0,$$

and in particular dist $[(\nabla f(x^k)_I, R\,(A^t)_I] \to 0$. As $R\,(A^t)_I$ is closed we deduce that $\nabla f(\bar{x})_I \in R\,(A^t)_I$, i.e. $(\nabla f(\bar{x}) + A^t\bar{\lambda})_I = 0$ for some $\bar{\lambda} \in \mathbf{R}^p$; system $(OS)_{I(\bar{x})}$ follows.

(ii) We first discuss the convergence of $\{x^k\}$. Note that $x_i^{k+1} = x_i^k(1 + \rho^k d_i^k/x_i^k)$, hence

$$x_i^{k+1} \le (1 + 1/\delta)x_i^k.$$

It follows that if $(x^k, x^{k+1}) \to (\bar{x}, \hat{x})$ for a subsequence, then $I(\bar{x}) \subset I(\hat{x})$.

If $(H1)'$ holds, using point (i) we deduce that $\bar{x} = \hat{x}$ and in particular $\|x^{k+1} - x^k\| \to 0$, hence the set of limit points of $\{x^k\}$ is connex. Using $(H1)'$ again it follows that the set of limit points is finite. Hence all the sequence converges towards the same point.

Now let us analyse the case when $(H1)$ and $(H2)$ hold. We know by Lemma 2.2 (i) that $f(x^k) - \varphi_k(d^k) \to 0$. With (2.6) and $(H2)$ this implies that $d^k \to 0$. As $\|x^{k+1} - x^k\| = \rho^k\|d^k\|$, and $\rho_k \le 1$, the set of limit points of $\{x^k\}$ is connex. By (i) and $(H1)$ each of them is isolated. It follows that the sequence converges.

We now prove that (OS) is satisfied under the additional assumption $(H3)$. If $x^k \to \bar{x}$ then there exists $(\bar{\lambda}, \bar{\mu})$ such that $(\bar{x}, \bar{\lambda}, \bar{\mu})$ verifies the first-order optimality system of $(P)_{I(\bar{x})}$ by (i). We have to show that $\bar{\mu}_{I(\bar{x})} \ge 0$. With Lemma 2.2 (iv) and $(H3)$ we deduce that $\{\lambda^k\}$ converges to $\bar{\lambda}$, hence by (2.2) we have $\mu^{k+1} := -\nu_k X_k^{-2}d^k$ converges to $\bar{\mu}$. Let $i \in I(\bar{x})$ be such that $\bar{\mu}_i < 0$, then $d_i^k = -(x_i^k)^2\mu_i^{k+1}/\nu_k > 0$ for k large enough, and this contradicts the fact that $x_i^k \to \bar{x}_i = 0$. ■

3 Acceptation of the unit stepsize

In this section we perform a local analysis around some point $\bar{x}$, local solution of (P). We seek conditions implying that, if $\bar{x}$ is a limit point of $\{x^k\}$, the sequence x^k converges to $\bar{x}$ and $\rho^k = 1$ is accepted. We need a few definitions.

Assuming that $\bar{x}$ satisfies $(H3)$, it follows that to $\bar{x}$ is associated a unique pair $(\bar{\lambda}, \bar{\mu})$ such that (OS) holds. Define the set of strictly active constraints as

$$J(\bar{x}) := \{i \in \{1, \ldots, n\} \; ; \; \bar{\mu}_i > 0\},$$

and the extended critical cone as

$$T := \{d \in \mathbf{R}^n \; ; \; Ad = 0 \; ; \; d_i = 0, \; i \in J(\bar{x})\}.$$

We say that $\bar{x}$ satisfies the strong second-order condition (see Robinson [19]) whenever

$$(SSOC) \qquad \text{there exists } \alpha_1 > 0 \; ; \; d^t \nabla^2 f(\bar{x}) d \geq \alpha_1 \|d\|^2, \; \forall \, d \in T.$$

This is a sufficient condition for the strong regularity, as defined by [19], of the associated optimality system. It has proved to be useful in sensitivity analysis as well as in the study of convergence properties of algorithms (see e.g. [13], [4], [3]).

Given d in $\ker A$, we now define d_T, d_N as the orthogonal projection of d onto T and N, where N is the orthogonal of T in $\ker A$, i.e.

$$N = \{z \in \ker A \; ; \; z^t d = 0, \; \forall \, d \in T\};$$

of course $d = d_T + d_N$ and $\|d\|^2 = \|d_T\|^2 + \|d_N\|^2$. Similarly, to d^k we associate d_T^k and d_N^k. Last but no least, we define the sufficient curvature condition as

$$(SCC) \qquad \left\{ \begin{array}{l} \text{there exists } \varepsilon_0 > 0, \text{ if } \|x^k - \bar{x}\| \leq \varepsilon_0 \text{ then} \\ \\ (d_T^k)^t M_k d_T^k \geq \dfrac{1}{2-\gamma} \, (d_T^k)^t \nabla^2 f(\bar{x}) d_T^k + \varepsilon_0 \|d_T^k\|^2. \end{array} \right.$$

We briefly discuss this condition. Specifically, we check that if M_k is close in a weak sense to $\nabla^2 f(\bar{x})$, and condition (SSOC) holds, then (SCC) is satisfied. We consider the following condition:

$$(d_T^k)^t M_k d_T^k = (d_T^k)^t \nabla^2 f(\bar{x}) d_T^k + o(\|d_T^k\|^2). \tag{3.1}$$

To see that (3.1) implies (SCC), note that $(2-\gamma)^{-1} \in (0,1)$ and $(d_T^k)^t \nabla^2 f(\bar{x}) d_T^k \geq \alpha_1 \|d_T^k\|^2$ by (SSOC). This and (3.1) imply

$$(d_T^k)^t M_k d_T^k \geq \frac{1}{2-\gamma} (d_T^k)^t \nabla^2 f(\bar{x}) d_T^k + \alpha_1 (1 - (2-\gamma)^{-1}) \|d_T^k\|^2 + o(\|d_T^k\|^2),$$

from which (SCC) follows. In particular, (SCC) is satisfied if (SSOC) holds and $M_k = \nabla^2 f(x^k)$.

Condition (SCC) is similar to a condition recently used in the analysis of successive quadratic programming algorithms [3]. There it is checked that, in the case of unconstrained optimization (then actually d_T^k and d^k coincide), this condition is very weak in the following sense: assuming the second-order sufficient optimality condition to hold ($\nabla^2 f(\bar{x}) > 0$), a necessary condition for the acceptation of the unit step, for x^k close to $\bar{x}$, is

$$(d_T^k)^t M_k d_T^k \geq \frac{1}{2-\gamma} (d_T^k)^t \nabla^2 f(\bar{x}) d_T^k + o(\|d_T^k\|^2).$$

Theorem 3.1 *Assume that $\{M_k\}$ is bounded, $\bar{x}$ satisfies (H3), (SSOC) and that (SCC) is satisfied for x^k close enough to $\bar{x}$. Then there exists $\varepsilon > 0$, if, for some k_0, $\|x^{k_0} - x\| < \varepsilon$ then $d^k \to 0$, $\rho^k = 1$ for all $k \geq k_0$ and $x^k \to \bar{x}$.*

We need a few lemmas (Lemma 3.1 is stated in [3], we give its proof for the reader's convenience):

Lemma 3.1 *For all $\varepsilon > 0$ and for all $n \times n$ symmetric matrix M there exists $K > 0$ such that*

$$d_T^t M d_T \geq d^t M d - \varepsilon\|d_T\|^2 - K\|d_N\|^2, \tag{3.2}$$

$$d^t M d \geq d_T^t M d_T - \varepsilon\|d_T\|^2 - K\|d_N\|^2. \tag{3.3}$$

Proof As $d = d_T + d_N$ it follows that

$$d^t M d = d_T^t M d_T + 2d_T^t M d_N + d_N^t M d_N,$$

hence

$$|d^t M d - d_T^t M d_T| = |2d_T^t M d_N + d_N^t M d_N| \leq \|M\|(2\|d_T\|.\|d_N\| + \|d_N\|^2).$$

Using the inequality $2ab \leq a^2 + b^2$ with $a = \sqrt{\varepsilon}\|d_T\|$ and $b = \|M\|\|d_N\|/\sqrt{\varepsilon}$, we get

$$|d^t M d - d_T^t M d_T| \leq \varepsilon\|d_T\|^2 + \|M\|(1 + \|M\|/\varepsilon)\|d_N\|^2,$$

from which the conclusion follows with

$$K = \|M\|(1 + \|M\|/\varepsilon).$$

■

Lemma 3.2 *There exists $K > 0$ such that*

$$\|z_N\| \leq K \sum_{i \in J(\bar{x})} |z_i|, \ \forall\, z \in \ker A.$$

Proof We have $z_N = z - z_T$ and $(z_T)_i = 0, i \in J(\bar{x})$. Henceforth $z_i = (z_N)_i, i \in J(\bar{x})$, and it suffices to prove that

$$\|z\| \leq K \sum_{i \in J(\bar{x})} |z_i|, \ \forall\, z \in N.$$

As both sides are positively homogeneous it suffices to establish the inequality when $\|z\| = 1$. Then the existence of K amounts to say that the problem

$$\min \sum_{i \in J(\bar{x})} |z_i| \ ; \ z \in N, \ \|z\| = 1;$$

has a positive infimum. If this were not the case, as this problem has a solution by compactness arguments, there would exist $z \in N$, $\|z\| = 1$, with $z_i = 0$, $i \in J(\bar{x})$; hence $z \in T$ (by definition of T), i.e. $z \in T \cap N = \{0\}$, a contradiction. ■

Lemma 3.3 *Assume that $\{M_k\}$ is bounded and $\bar{x}$ satisfies (H3). Then, for all $K \geq 0$, if x^k is sufficiently close to $\bar{x}$, the relation below holds:*

$$\nu_k (d^k)^t X_k^{-2} d^k > K\|d_N^k\|^2. \tag{3.4}$$

Proof Denote $\mu^{k+1} := -\nu_k X_k^{-2} d^k$. ¿From $(H3)$, Lemma 2.2 (iv) and (2.2), we deduce that for any subsequence of $\{x^k\}$ converging to $\bar{x}$, the associated subsequence of $\{\mu^k\}$ converges to $\bar{\mu}$. Hence if x^k is close enough to $\bar{x}$ one has $d_i^k < 0$ and $\mu_i^k > \bar{\mu}_i/2$, $i \in J(\bar{x})$. Denote

$$\theta := \min\{\bar{\mu}_i/2,\ i \in J(\bar{x})\}.$$

It follows that

$$\nu_k (d^k)^t X_k^{-2} d^k \geq \nu_k \sum_{i \in J(\bar{x})} (d_i^k / x_i^k)^2 \geq \frac{1}{2} \sum_{i \in J(\bar{x})} -\bar{\mu}_i d_i^k \geq \theta \sum_{i \in J(\bar{x})} |d_i^k|. \tag{3.5}$$

Also, as $|d_i^k| \leq |x_i^k|/\delta$ it follows that $|d_i^k|$, $i \in J(\bar{x})$ can be made arbitrarily small by taking x^k close to $\bar{x}$. It follows with (3.5) that

$$\nu_k (d^k)^t X_k^{-2} d^k / (\sum_{i \in J(\bar{x})} |d_i^k|)^2 \to \infty. \tag{3.6}$$

We conclude with Lemma 3.2. ■

Lemma 3.4 *Under the hypotheses of Theorem 3.1, for all $K_0 \geq 0$, $\alpha_1 > 0$ being given by (SSOC), if x^k is sufficiently close to $\bar{x}$ then*

$$(d^k)^t (M_k + 2\nu_k X_k^{-2}) d^k \geq \frac{\alpha_1}{2} \|d^k\|^2 + K_0 \|d_N^k\|^2. \tag{3.7}$$

Proof By Lemma 3.1, choosing $\varepsilon = \varepsilon_0$, and (SCC), if k is large enough and x^k is close to $\bar{x}$, then

$$(d^k)^t M_k d^k \geq \frac{1}{2-\gamma} (d_T^k)^t \nabla^2 f(\bar{x}) d_T^k - K \|d_N^k\|^2.$$

As $(2-\gamma)^{-1} \geq 1/2$, using (SSOC) we get

$$\begin{aligned} (d^k)^t M_k d^k &\geq \frac{\alpha_1}{2} \|d_T^k\|^2 - K \|d_N^k\|^2, \\ &= \frac{\alpha_1}{2} \|d^k\|^2 - \left(K + \frac{\alpha_1}{2}\right) \|d_N^k\|^2. \end{aligned}$$

The conclusion is obtained with Lemma 3.3. ■

Proof of Theorem 3.1 a) We first prove that $x^k \to \bar{x}$. We use the fact that $\|d^k\|$ is small whenever x^k is close to $\bar{x}$ as a consequence of Lemma 3.4 and (2.6), and that $\bar{x}$ satisfies (SSOC). The latter fact implies that $\bar{x}$ is an isolated critical point of (P) (see [19]). As $(H3)$ necessarily holds in a neighborhood of $\bar{x}$, it follows that $\bar{x}$ is the only limit point of $\{x^k\}$ in some neighborhood $\mathcal{V}$ of $\bar{x}$. We now just have to prove that x^k remains in $\mathcal{V}$ for k large enough. We can take $\mathcal{V}$ of the form

$$\mathcal{V}_\varepsilon := \{x \in F : \|x - \bar{x}\| \leq \varepsilon\}.$$

Note that $\|d^k\| < \varepsilon/2$ whenever $x^k \in \mathcal{V}_{\varepsilon_1}$ for some $\varepsilon_1 > 0$ small enough. We may assume $\varepsilon_1 < \varepsilon/2$. It follows that, if $x^k \in \mathcal{V}_{\varepsilon_1}$, then $\|x^{k+1} - \bar{x}\| \leq \|x^k - \bar{x}\| + \|d^k\| \leq \varepsilon$. In other words, x^{k+1} is in $\mathcal{V}_\varepsilon$ whenever x^k is in $\mathcal{V}_{\varepsilon_1}$.

On the other hand we also know that $f(x^{k+1}) \leq f(x^k)$. So, define

$$\hat{f} := \inf\{f(x) \; ; \; x \in \mathcal{V}_\varepsilon - \mathcal{V}_{\varepsilon_1}\}.$$

As $\bar{x}$ is a strict local minimum of (P), reducing ε and ε_1 if necessary, we may assume that $\hat{f} > f(\bar{x})$. Now assuming that $f(x^k) \leq \hat{f}$ and $x^k \in \mathcal{V}_{\varepsilon_1}$, it follows that $f(x^{k+1}) < \hat{f}$ and $x^{k+1} \in \mathcal{V}_\varepsilon$; using the definition of $\hat{f}$, we find that x^{k+1} is in $\mathcal{V}_{\varepsilon_1}$ again. This implies that the sequence $\{x^k\}$ remains in $\mathcal{V}_{\varepsilon_1}$, hence that $x^k \to \bar{x}$.

b) We now check that $\rho^k = 1$ for k large enough. Define

$$H_k := 2\int_0^1 (1-\sigma)\nabla^2 f(x^k + \sigma d^k)d\sigma.$$

Then

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

If x^k is close enough to $\bar{x}$, as already observed d^k is then close to 0, hence H_k is close to $\nabla^2 f(\bar{x})$. We deduce that

$$-(d^k)^t H_k d^k \geq -(d^k)^t \nabla^2 f(\bar{x}) d^k - \frac{\varepsilon_0}{2}\|d^k\|^2,$$

with ε_0 given by (SCC). As a consequence

$$\begin{aligned} f(x^k) - f(x^k + d^k) &\geq -\nabla f(x^k)^t d^k - \frac{1}{2}(d^k)^t \nabla^2 f(\bar{x}) d^k - \frac{\varepsilon_0}{4}\|d^k\|^2, \\ &= f(x^k) - \varphi(d^k) + \frac{1}{2}(d^k)^t (M_k - \nabla^2 f(\bar{x})) d^k - \frac{\varepsilon_0}{4}\|d^k\|^2. \end{aligned}$$

So by (2.1) the unit step will be accepted if

$$(1-\gamma)(f(x^k) - \varphi_k(d^k)) + \frac{1}{2}(d^k)^t (M_k - \nabla^2 f(\bar{x})) d^k - \frac{\varepsilon_0}{4}\|d^k\|^2 \geq 0. \tag{3.8}$$

Using (2.6), Lemma 3.1 with $\varepsilon = \varepsilon_0/2$, where ε_0 is given by (SCC), and Lemma 3.3 we get

$$\begin{aligned} f(x^k) - \varphi_k(d^k) &= \frac{1}{2}(d^k)^t M_k d^k + \nu_k \delta_k^2, \\ &\geq \frac{1}{2}(d_T^k)^t M_k d_T^k - \frac{\varepsilon_0}{4}\|d_T^k\|^2 - \frac{K}{2}\|d_N^k\|^2 + \nu_k \delta_k^2, \\ &\geq \frac{1}{2(2-\gamma)}(d_T^k)^t \nabla^2 f(\bar{x}) d_T^k - \frac{K}{2}\|d_N^k\|^2 + \frac{\varepsilon_0}{4}\|d_T^k\|^2 + \nu_k \delta_k^2, \\ &\geq \frac{1}{2(2-\gamma)}(d_T^k)^t \nabla^2 f(\bar{x}) d_T^k + \frac{\nu_k}{2}\delta_k^2. \end{aligned}$$

Similarly

$$\begin{aligned}
\frac{1}{2}(d^k)^t(M_k - \nabla^2 f(\bar{x}))d^k &\geq \frac{1}{2}(d_T^k)^t(M_k - \nabla^2 f(\bar{x}))d_T^k - \frac{\varepsilon_0}{4}\|d_T^k\|^2 - K'\|d_N\|^2, \\
&\geq \frac{\gamma - 1}{2(2-\gamma)}(d_T^k)^t \nabla^2 f(\bar{x}) d_T^k - K'\|d_N^k\|^2 + \frac{\varepsilon_0}{4}\|d_T^k\|^2.
\end{aligned}$$

Combining this and using Lemma 3.3 again, we get

$$\begin{aligned}
(1-\gamma)(f(x^k) - \varphi_k(d^k)) + \frac{1}{2}(d^k)^t(M_k - \nabla^2 f(\bar{x}))d^k - \frac{\varepsilon_0}{4}\|d^k\|^2 \\
\geq (1-\gamma)\frac{\nu_k}{2}\delta_k^2 - (K' + \frac{\varepsilon_0}{4})\|d_N^k\|^2 \geq 0.
\end{aligned}$$

We have proved (3.8) as required. It follows that the unit step is accepted, hence d^k vanishes, as was to be proved. ∎

We now check that, if M_k is close to $\nabla^2 f(\bar{x})$ in a very weak sense (see (3.10) below) then the following holds:

$$\sum_k (\|x^k - \bar{x}\| + \|\lambda^k - \bar{\lambda}\| + \|\mu^k - \bar{\mu}\|) < \infty. \tag{3.9}$$

Theorem 3.2 *Assume that the hypotheses of Theorem 3.1 hold, and in addition that $\bar{x}$ satisfies the strict complementarity condition. If $x^k \to \bar{x}$ (hence $\rho^k = 1$ by Theorem 3.1) then there exists $\varepsilon_1 > 0$ such that*

$$\|(M_k - \nabla^2 f(\bar{x}))d_T^k\| \leq \varepsilon_1 \|d^k\| \tag{3.10}$$

implies (3.9) holds.

Let us note that Newton's method satisfies (3.10). Note that if we assume $M_k \longrightarrow \nabla^2 f(\bar{x})$, since $\nabla^2 f(\bar{x})$ need not be positive definite, then we may violate the positive definiteness requirement on M_k.

Proof Denote

$$I := I(\bar{x}), \; \bar{I} := \{1, \ldots, n\} - I.$$

The proof is based on the mapping

$$\psi(x, \lambda) := \begin{cases} (\nabla f(x) + A^t \lambda)_I, \\ Ax - b, \\ x_{\bar{I}}. \end{cases}$$

It follows easily from (SSOC) and ($H3$) that $\psi(x, \lambda)$ has an invertible derivative at $(\bar{x}, \bar{\lambda})$, hence there exists some $a_1 > 0$ such that

$$\|x^{k+1} - \bar{x}\| + \|\lambda^{k+1} - \bar{\lambda}\| \leq a_1 \|\psi(x^{k+1}, \lambda^{k+1})\|. \tag{3.11}$$

a) Let us prove that

there exists K_1, K_4 ; $\|\psi(x^{k+1}, \lambda^{k+1})\| \leq nK_1\nu_{k+1} + nK_4\nu_k + \|x^{k+1} - x^k\|/(4a_1)$. (3.12)

¿Indeed, from the convergence of $\{x^k\}$ to $\bar{x}$ and $(H3)$, using Lemma 2.2 (iv) and (2.2), it follows that $(\lambda^k, \mu^k) \to (\bar{\lambda}, \bar{\mu})$. Now multiplying (2.2) by X_k, and reminding that $\nu_k\|X_k^{-1}d^k\| = \nu_k\delta_k$, we get

$$\|X_k(\nabla f(x^k) + M_kd^k + A^t\lambda^{k+1})\| = \nu_k\delta_k.$$

Using the strict complementarity hypothesis and the relation $|z_i| \leq \|z\|$, we obtain for some $K_1 > 0$

$$x_i^k \leq K_1\nu_k, \quad i \in I, \tag{3.13}$$

$$|(\nabla f(x^k) + M_kd^k + A^t\lambda^{k+1})_i| \leq K_1\nu_k, \quad i \notin I. \tag{3.14}$$

Now choose ε_1 in (3.10) as $\varepsilon_1 = 1/(8a_1n)$. We have

$$\begin{aligned} \nabla f(x^k) + M_kd^k &= \nabla f(x^k) + \nabla^2 f(\bar{x})d^k + (M_k - \nabla^2 f(\bar{x}))d^k, \\ &= \nabla f(x^{k+1}) + r^k + (M_k - \nabla^2 f(\bar{x}))d^k, \end{aligned} \tag{3.15}$$

where the term r^k, for x^k close to $\bar{x}$, satisfies:

$$\|r^k\| \leq \|d^k\|/(8a_1n). \tag{3.16}$$

Also by (3.10) and as $\{M_k\}$ is bounded we get for some $K_2 > 0$

$$\begin{aligned} \|(M_k - \nabla^2 f(\bar{x}))d^k\| &\leq \|(M_k - \nabla^2 f(\bar{x}))d_T^k\| + K_2\|d_N^k\|, \\ &\leq \|d^k\|/(8a_1n) + K_2\|d_N^k\|. \end{aligned} \tag{3.17}$$

Using (3.15), (3.16), (3.17) and Lemma 3.2 we obtain for some $K_3 > 0$

$$\|\nabla f(x^k) + M_kd^k - \nabla f(x^{k+1})\| \leq \|d^k\|/(4a_1n) + K_3\sum_{j\in I}|d_j^k|. \tag{3.18}$$

Now we prove (3.12). As $\bar{\mu} = -\lim_{k\to+\infty} \nu_kX_k^{-2}d^k$, using the strict complementarity hypothesis, for k large enough and for all $j \in I$ we get $d_j^k < 0$, hence $|d_j^k| \leq x_j^k$. So, combining with (3.13), (3.14) and (3.18) we get for some $K_4 > 0$

$$|(\nabla f(x^{k+1}) + A^t\lambda^{k+1})_i| \leq K_4\nu_k + \|d^k\|/(4a_1n), \quad i \notin I(\bar{x}). \tag{3.19}$$

So, by (3.13) and (3.19), we get (3.12).

b) On the other hand, by (2.6), the linesearch rule and the fact that $\rho^k = 1$, we have

$$f(x^k) - f(x^{k+1}) \geq \gamma(f(x^k) - \varphi_k(d^k)) \geq \gamma\nu_k\delta_k^2,$$

henceforth, as $\delta_k \geq \delta > 0$:

$$\nu := \sum_{k=1}^{\infty} \nu_k < \infty. \tag{3.20}$$

Hence using (3.11) and (3.12) we get

$$\sum_{k=k_0}^{\bar{k}} (\|x^{k+1} - \bar{x}\| + \|\lambda^{k+1} - \bar{\lambda}\|) \leq a_1 n(K_1 + K_4)\nu + \frac{1}{4} \sum_{k=k_0}^{\bar{k}} \|x^{k+1} - x^k\|.$$

Now using

$$\frac{1}{4} \sum_{k=k_0}^{\bar{k}} \|x^{k+1} - x^k\| \leq \frac{1}{4} \sum_{k=k_0}^{\bar{k}} (\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|) \leq \frac{1}{2} \sum_{k=k_0}^{\bar{k}} \|x^{k+1} - \bar{x}\| + \frac{1}{4}\|x^{k_0} - \bar{x}\|,$$

we deduce

$$\sum_{k=k_0}^{\bar{k}} (\|x^{k+1} - \bar{x}\| + \|\lambda^{k+1} - \bar{\lambda}\|) \leq 2a_1 n(K_1 + K_4)\nu + \frac{1}{2}\|x^{k_0} - \bar{x}\|.$$

Finally we obtain (3.9), noticing that by (2.2)

$$\begin{aligned} \mu^{k+1} - \bar{\mu} &= O(\|x^k - \bar{x}\| + \|\lambda^{k+1} - \bar{\lambda}\| + \|d^k\|), \\ &= O(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\| + \|\lambda^{k+1} - \bar{\lambda}\|). \end{aligned}$$

■

References

[1] L. Armijo : Minimization of functions having Lipschitz continuous first partial derivatives. Pacific J. Math 16(1966), 1-3.

[2] J.F. Bonnans, M. Bouhtou : The trust region affine interior point algorithm for nonconvex quadratic programming. INRIA Report, to appear.

[3] J.F. Bonnans, G. Launay : An implicit trust region algorithm for constrained optimization. INRIA Report 1780, 1992 (Revised version)

[4] J.F. Bonnans, A. Sulem: Pseudopower expansion of generalized equations and constrained optimization problems. INRIA Report, to appear.

[5] M. Bouhtou : Méthodes de points interieurs pour l'optimisation des Doctoral dissertation. Université de Dauphine, Paris, 1993, to appear.

[6] J.P. Bulteau, J.P. Vial: A restricted trust region algorithms for unconstrained optimization. Journal of Optim. Theory and App. Vol. 47, n° 4, December 1985.

[7] D. Den Hertog, C. Roos, T. Terlaky : On the classical logarithmic Barrier method for a class of smooth convex programming problems. Journal of Optimization Theory and App. 73(1992), 1, 1-25.

[8] I.I. Dikin : Iterative solutions of problems of linear and quadratic programming. Soviet Math. Doklady 8(1967), 674-675.

[9] R. Fletcher : Practical Methods of Optimization, John Wiley & Sons, Chichester and New York, second edition, 1987.

[10] C.C. Gonzaga, L.A. Carlos : A primal affine-scaling algorithm for linearly constrained convex programs. Rep. ES-238/90, Universidade Federal do Rio de Janeiro, 1990.

[11] C.C. Gonzaga : An interior trust region method for linearly constrained optimization. COAL Bulletin 19(1991), 55-65.

[12] F. Jarre : On the method of analitic centers for solving smooth convex programs. In S. Dolecki, ed., Lecture Notes in Mathematics 1405, Optimization, Springer (1989), 69-85.

[13] K. Jittorntrum : Solution point differentiability without strict complementarity in nonlinear programming. Math. Prog. Study 21(1984), 127-138.

[14] N. Karmarkar : A new polynomial time algorithm for linear programming. Combinatorica 4(1984), 373-395.

[15] S. Mehrotra, J. Sun : An interior point algorithm for solving smooth convex programs based on Newton's method. Mathematical Developments Arising from Linear Programming, J.C. Lagarias and M.J. Todd,des., Contemporary Mathematics 114(1990), 265-284.

[16] R.D.C. Monteiro, I. Adler : An extension of Karmarkar–type algorithms to a class of convex separable programming problems with global rate of convergence. Mathematics of Operations Research 15(1990), 408-422.

[17] J.J. Moré : Recent developments in algorithms and software for trust region method. In "Mathematical Programming, the state of the art". A. Bachem, M. Grötschel, B. Korte eds., Springer Verlag, Berlin, 1983, 258-287.

[18] M.J.D. Powell : Algorithms for nonlinear constraints that use Lagrangian functions. Math. Programming 15(1978), 224-248.

[19] S.M. Robinson : Strongly regular generalized equations. Math. Oper. Res. 5(1980), 43-62.

[20] D.C. Sorensen : Newton's method with a model trust region modification. SIAM J. Numer. Anal. Vol. 19, n° 2, April 1982.

[21] 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. Northwerstern University, Evanston, IL 60208, USA, 1990.

[22] 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.

[23] Y. Ye, E. Tse : An extension of Karmarkar's projective algorithm for convex quadratic programming. Math. Programming 44(1989), 157-179.

Unité de Recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)
Unité de Recherche INRIA Lorraine Technopôle de Nancy-Brabois - Campus Scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 VILLERS LES NANCY Cedex (France)
Unité de Recherche INRIA Rennes IRISA, Campus Universitaire de Beaulieu 35042 RENNES Cedex (France)
Unité de Recherche INRIA Rhône-Alpes 46, avenue Félix Viallet - 38031 GRENOBLE Cedex (France)
Unité de Recherche INRIA Sophia Antipolis 2004, route des Lucioles - B.P. 93 - 06902 SOPHIA ANTIPOLIS Cedex (France)

EDITEUR
INRIA - Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 LE CHESNAY Cedex (France)

ISSN 0249 - 6399

RR.1948