archives-ouvertes.fr

Asymptotic admissibility of the unity stepsize in exact penalty methods I:Equality-constrained problems

J. Frederic Bonnans

To cite this version:

J. Frederic Bonnans. Asymptotic admissibility of the unity stepsize in exact penalty methods I:Equality-constrained problems. [Research Report] RR-0273, INRIA. 1984. <inria-00076285>

HAL Id: inria-00076285
https://hal.inria.fr/inria-00076285

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INRIA

CENTRE DE ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P. 105
78153 Le Chesnay Cedex
France
Tél. (3) 954 90 20

Rapports de Recherche

N° 273

ASYMPTOTIC ADMISSIBILITY OF THE UNITY STEPSIZE IN EXACT PENALTY METHODS I: EQUALITY-CONSTRAINED PROBLEMS

Joseph Frédéric BONNANS

Mars 1984

ASYMPTOTIC ADMISSIBILITY OF THE UNITY STEPSIZE IN EXACT PENALTY METHODS I : EQUALITY-CONSTRAINED PROBLEMS

Joseph Frédéric BONNANS
INRIA
Domaine de Voluceau
BP 105 - Rocquencourt
78153 LE CHESNAY Cédec (France)

RESUME

Deux points délicats, dans la mise en oeuvre des algorithmes d'optimisation utilisant une fonctionnelle pénalisée exacte associée à des sous-problèmes linéaires quadratiques, sont la prise en compte des problèmes linéaires-quadratiques inconsistants et l'admissibilité du pas unité. Nous montrons que, dans le cas où le problème ne comporte que des contraintes d'égalité, une méthode récente qui résoud de façon satisfaisante le premier problème peut être modifiée d'une manière simple pour que le pas unité soit asymptotiquement admissible.

ABSTRACT

Two difficulties arising when using optimization algorithms based on an exact penalty function associated to quadratic subproblems are the possible inconsistency of the quadratic programs and the admissibility of the unity stepsize after a finite number of iterations. We show that, if no inequality constraints are involved, a recent method, in which the first problem is solved in a satisfactory way, can be modified in a simple way to make the unity stepsize asymtotically admissible.

PAPIER RECUPERE ET RECYCLE

I - INTRODUCTION

We consider a nonlinear programming problem having only equality constraints :

$$(1.1)\quad \begin{cases} \text{Min } f(x), \\ g_i(x) = 0,\ i = 1 \text{ to } m, \end{cases}$$

f and g_i, $i = 1$ to m being smooth (C^3) applications from $\mathbb{R}^n$ onto $\mathbb{R}$. We suppose that $m \leq n$. Let $\bar{x}$ be a local solution of (1.1). We suppose that $\bar{x}$ is a regular point, i.e.

$$(1.2)\quad \nabla g_i(\bar{x}),\ i = 1 \text{ to } m, \text{ are linearly independant.}$$

Then there exists a unique $\bar{\lambda} \in \mathbb{R}^m$ such that

$$(1.3)\quad \begin{cases} \nabla f(\bar{x}) + \Sigma\ \bar{\lambda}_i \nabla g_i(\bar{x}) = 0, \\ g_i(\bar{x}) = 0. \end{cases}$$

Equations (1.3) may be solved by a Newton-type method with unknowns (x,λ) : this is equivalent to the computation of a sequence (x^k, λ^k) with $x^{k+1} = x^k + d^k$, where d^k is a solution of the quadratic program

$$(1.4)\quad \begin{cases} \text{Min } \nabla f(x^k)^t d + \frac{1}{2} d^t H^k d, \\ \text{s.t. } g_i(x^k) + \nabla g_i(x^k)^t d = 0,\ i = 1 \text{ to } m, \end{cases}$$

H^k being an approximation of the Hessian of the lagrangian, and λ^k being the multiplier associated to d^k. We restrict our analysis to the case when H^k is a definite positive approximation of the hessian of the lagrangian. When this approximation is made using the BFGS method associated to some correction, a superlinear convergence rate is obtained (M.J.D. Powell [13]). In order to globalize this algorithm -i.e. to design some globally convergent algorithm that reduce to a Newton-type method in the neighbourhood of a solution- a key idea is to use the non differentiable penalty function

$$\theta_r(x) = f(x) + r\,||g(x)||$$

where $r > 0$ is called the penalization coefficient and $||.||$ is some norm of $\mathbb{R}^m$; we denote by $||.||_{\mathscr{D}}$ the dual norm, i.e.

$$||\lambda||_{\mathscr{D}} = \max\{\sum_{i=1}^{m} \lambda_i\, \mu_i,\ ||\mu|| = 1\}.$$

A key fact is that if (1.4) has a solution d^k associated to a multiplier λ^k, and if $r > ||\lambda||_{\mathscr{D}}$, then d^k is a descent direction of $\theta_r(x)$. This result is due to B. Pschenichny (see [14]) in the case of the L^∞ norm and was rediscovered by S.P. Han [9] in the case of the L^1 norm. The general result can be found in J.F. Bonnans, D. Gabay [3]. A difficulty arises if the m vectors $\{\nabla g_i(x)\}$ are not everywhere linearly independant ; then (1.4) may have no solutions. Some empirical means to take this into account are given in M.J.D. Powell [12], K. Tone [15]. However, the algorithm of J.F. Bonnans, D. Gabay [3] seems, at least from a theorical point of view, give a satisfactory solution to this problem. It is based on the function, defined for each iteration k :

$$\theta_r^k(x) = f(x^k) + \nabla f(x^k)^t(x-x^k) + \frac{1}{2}(x-x^k)^t\, H^k(x-x^k) +$$

$$+ r||g(x^k) + \nabla(x^k)^t(x-x^k)||,$$

which is a simple model of $\theta_r(x)$, having the same behaviour around x^k. A direction $\tilde{d}^k$ is computed as the solution of

$$(1.5) \qquad \mathrm{Min}\{\theta_r^k(x^k + d)\ ;\ d \in \mathbb{R}^n\}.$$

The convex problem (1.5) always has a unique solution $\tilde{d}^k$. Then the iteration is

$$x^{k+1} = x^k + \rho^k\, \tilde{d}^k,$$

where ρ^k is a step size computed according to some line-search rule.

Define

$$\tilde{\theta}_r^k(x) = f(x^k) + \nabla f(x^k)^t(x-x^k) + r||g(x^k) + \nabla g(x^k)^t(x-x^k)||.$$

A convenient line-search rule due to R.M. Chamberlain et al. [4] is the following extension of the L. Armijo rule [1] :

$$(1.6)\quad \begin{cases} \text{Choose } \beta \in]0,1[,\ \sigma \in]0,1/2[\quad \text{(independant of k).} \\ \rho^k = (\beta)^\ell, \text{ where } \ell \text{ is the smallest integer such that} \\ \theta_r(x^k + (\beta)^\ell \tilde{d}^k) - \theta_r(x^k) \le \sigma(\tilde{\theta}_r^k(x^k + (\beta)^\ell \tilde{d}^k) - \theta_r(x)). \end{cases}$$

This means that we reduce the step until the ratio of the achieved decrease on the criterium divided by the decrease predicted by the local model $\tilde{\theta}_r^k$ is superior to σ.

The complexity of problem (1.5) is roughly the same as the one of a quadratic program. In addition, if (1.4) has a solution d^k, the solution of (1.5) will be equal to d^k iff r^k is greater than $||\lambda^k||_{\mathscr{D}}$. Consequently, when the parameter r^k is iteratively modified in a convenient way, the method leads to a globally convergent algorithm, where computed displacements reduce to the solution of (1.4) near a regular solution of (1.1).

One point should be clarified. To compute a descent directions, the algorithm described above relaxes the linearized constraints by penalizing them. This is useful when these linearized constraints are not necessarily compatible at any point. However, if it is a priori known that the linearization of some subset of the constraints are compatible there is no reason to relax the constraints of this subset. A key fact is that this subset must contain the linear constraints (otherwise, the problem has no solution). This means that in practice, the linear constraints should not be relaxed. In this paper, we don't take this remark in account, in order to keep the proofs short. However, this modification of the algorithm should not change essentially the results.

We now turn to the local convergence analysis. We suppose that a sequence $\{x^k\}$

computed by the preceding algorithm converges to a local solution $\bar{x}$ of (1.1), which is a regular point. Let d^k be the solution of (1.4) (d^k is well-defined if k is great enough). Let $||\cdot||_U$ be a norm of $\mathbb{R}^n$. We suppose that

$$(1.7) \quad \begin{cases} \text{The vector } d^k\text{, solution of (1.4), checks} \\ \dfrac{||x^k + d^k - \bar{x}||_U}{||x^k - \bar{x}||_U} \to 0. \end{cases}$$

This hypothesis allows the algorithm to have a superlinear rate of convergence, if $\{\rho^k\}$ converges towards 1. However, this seems not to be the case in general : a counter-example due to N. Maratos [10] shows that even if (1.10) holds, and if r has the same order of magnitude that $||\bar{\lambda}||_{\mathcal{D}}$, $\theta_r(x^k + \rho^k d^k)$ may be superior to $\theta_r(x^k)$ if $\rho^k \to 1$. This may destroy the property of superlinear convergence. Two kind of methods have been proposed to deal with this problem. The first (D. Gabay [8], D.Q. Mayne, E. Polak [11]) needs the computation of the constraints at $x^k + d^k$; then a correction term v^k is computed as the solution of

$$\begin{cases} \text{Min } \sum_{i=1}^{n} (v_i^k)^2 , \\ \text{s.t. } g(x^k + d^k) + \nabla g(x^k)^t v^k \equiv 0. \end{cases}$$

If a line-search is used, it can be performed along the arc

$$x^k + \rho\, d^k + (\rho)^2\, v^k.$$

Then it is shown that, under some convenient assumptions, the stepize $\rho^k = 1$ is admissible if k is great enough. The second method, due to R. Chamberlain et al. [4] is based on the observation that a sufficient decrease of the exact penalty function is obtained (for k great enough) between the iterations k+1 and k-1. Consequently, the line search criterium at step k should use the information of the iteration k-1. However, if the point $x^k + d^k$ is not accepted, one has to come back then, to the point x^{k-1} and to reduce the stepsize at x^{k-1} ; this is to insure the global convergence. The main drawback of these methods is that, in some situations, the can lead to a substantial increase of the amount of computations. This is not the case with the algorithm proposed here. In paragraph 2 we give a technical result on exact penalty functions which is the basis of the subsequent algorithm. Then in paragraph 3 we use this result to formulate a globally convergent method, based on a line-search strategy, for which the unity stepsize is, under some convenient hypothesis, asymptotically admissible.

II - SOME LOCAL PROPERTIES OF A CLASS OF EXACT PENALTY FUNCTIONS

Let $\bar{x}$ be a local solution of (1.1) such that (1.2) holds and $\bar{\lambda}$ be an element of $\mathbb{R}^m$ such that (1.3) holds. Define the augmented lagrangian (for $c \geq 0$)

$$L_c(x,\lambda) = f(x) + \lambda^t g(x) + \frac{c}{2} \sum_{i=1}^{n} g_i(x)^2.$$

Denote

$$H_c = \frac{\partial^2 L(\bar{x},\bar{\lambda})}{\partial x^2} \quad .$$

We suppose that the standard second-order sufficiency conditions hold at $\bar{x}$ (see for instance R. Fletcher [6]) :

$$(2.1) \quad \begin{cases} d^t H_o d > 0, \\ \text{for any } d \in \mathbb{R}^n \text{ such that} \\ \nabla g_i(x)^t d = 0, \ i = 1 \text{ to } m. \end{cases}$$

It is well known that (1.2), (1.3), (2.1) imply

$$(2.2) \quad \exists\, \bar{c} > 0 \ ; \ H_{\bar{c}} \text{ is positive definite.}$$

We now consider the following class of penalty function :

$$\theta_{p,r}(x) = f(x) + p^t g(x) + r||g(x)||,$$

where $p, r \in \mathbb{R}^m \times \mathbb{R}$ are given parameters with $r > 0$. We remind that $||.||$ is a norm of $\mathbb{R}^n$ and $||.||_{\mathcal{D}}$ is its dual norm. We give a sufficient condition for these penalty functions to be exact, i.e. to have a (strict) local minimum at $\bar{x}$.

Proposition 2.1

Let $\bar{x}$ be a local minimum of (1.1) such that, for some $\bar{\lambda}$, (1.2), (1.3), (2.1) hold. Then if

(2.3) $\quad r > ||\bar{\lambda} - p||_{\mathcal{D}}$,

$\bar{x}$ is a strict local minimum of $\theta_{p,r}(x)$. □

Proof

We have

$$\theta_{p,r}(x) - L_{\bar{c}}(x,\bar{\lambda}) = (p-\bar{\lambda})^t\, g(x) + r||g(x)|| - \frac{\bar{c}}{2} \sum_{i=1}^{m} g_i(x)^2,$$

$$\geq (r - ||p-\bar{\lambda}||_{\mathcal{D}})||g(x)|| - \frac{\bar{c}}{2} \sum_{i=1}^{m} g_i(x)^2.$$

As all norms are equivalent in $\mathbb{R}^m$, some $\beta > 0$ exists such that

$$\sum_{i=1}^{m} g_i(x)^2 \leq \beta||g(x)||^2.$$

This implies

$$\theta_{p,r}(x) - L_{\bar{c}}(x,\bar{\lambda}) \geq (r - ||p-\bar{\lambda}||_{\mathcal{D}} - \frac{c}{2}\beta||g(x)||)\, ||g(x)||.$$

If (2.3) holds, the right-hand side is positive in a neighbourhood of $\bar{x}$. As $H_{\bar{c}}$ is positive definite, $\bar{x}$ is a strict local minimum of $L_{\bar{c}}(x,\bar{\lambda})$; this proves the proposition. □

Let $\{x^k\}$ be a sequence converging towards $\bar{x}$ and $\{p^k\}$, $\{r^k\}$ two sequences such that, for k great enough, form some $\gamma > 0$, (2.4) $\quad (3+\gamma)\ ||p^k-\bar{\lambda}||_{\mathcal{D}} < r^k$;

and

(2.5) $$\frac{r^k}{||x^k-\bar{x}||_U} \to +\infty \ .$$

We define

$$\tilde{\theta}^k(x) = f(x^k) + \nabla f(x^k)^t(x-x^k) + (p^k)^t(g(x^k) + \nabla g(x^k)^t(x-x^k))$$

$$+ r||g(x^k) + \nabla g(x^k)^t(x-x^k)||.$$

We remind that the solution d^k of (1.4) is well defined in a neighbourhood of $\bar{x}$.

Theorem 2.1

We suppose that (1.7), (2.4), (2.5) hold. Then there exists $r^o > 0$ such that, for any $\varepsilon > 0$, there exists k_o such that $k > k_o$ and $r^k < r^o$ imply

(2.6) $$\theta_{p^k,r^k}(x^k+d^k) - \theta_{p^k,r^k}(x^k) \leq (\tfrac{1}{2}-\varepsilon)(\tilde{\theta}^k(x^k+d^k) - \theta_{p^k,r^k}(x^k)). \quad \square$$

The proof of the theorem uses two lemmas.

Lemma 2.1

We suppose that (1.7) holds. Then, for any $\varepsilon_1 > 0$, there exists k_1 such that $k > k_1$ implies

$$\tilde{\theta}^k(x^k+d^k) - \theta_{p^k,r^k}(x^k) \geq -(1+\varepsilon_1)(x^k-\bar{x})^t H_{\bar{c}}(x^k-\bar{x}) -$$

$$-(r^k + ||p^k - \bar{\lambda}||_{\mathcal{D}})\ ||g(x^k)||.$$

Proof

We have for k great enough

$$\Delta = \tilde{\theta}^k(x^k+d^k) - \theta_{p^k,r^k}(x^k) = \nabla f(x^k)^t\, d^k + (p^k)^t\, \nabla g(x^k)^t\, d^k - r^k||g(x^k)|| ,$$
$$= \nabla_x\, L_o(x^k,\bar\lambda)^t\, d^k + (p^k-\bar\lambda)^t\, \nabla g(x^k)^t\, d^k - r^k||g(x^k)|| ;$$

From (1.4) we deduce that

$$\Delta = \nabla_x\, L_o(x^k,\bar\lambda)^t\, d^k - (p^k-\lambda)^t\, g(x^k) - r^k||g(x^k)||,$$
$$\geq \nabla_x\, L_o(x^k,\bar\lambda)^t\, d^k - (r + ||p-\lambda||_{\mathscr{D}})\, ||g(x^k)||.$$

We have with (1.3), (1.7)

$$\nabla_x\, L_o(x^k,\bar\lambda)^t = H_o(x^k-\bar x) + o(||x^k-\bar x||_U),$$
$$d^k = \bar x - x^k + o(||x^k - \bar x||_U),$$

where the notation $o(||x^k-\bar x)||_U)$ indicates a term whose ratio to $||x^k-\bar x||_U$ tends to zero as $k \to \infty$. We deduce that

$$\Delta \geq - (x^k-\bar x)\, H_o(x^k-\bar x) - (r^k + ||p^k-\lambda||_{\mathscr{D}})||g(x^k)|| + o(||x^k-\bar x||_U^2).$$

The result is then a consequence of the inequality

$$d^t\, H_{\bar c}\, d \geq d\, H_o\, d, \; \forall d \in \mathbb{R}^n,$$

and of the positive definiteness of $H_{\bar c}$. □

Lemma 2.2

If (1.7) holds, $\exists \beta > 0$ such that for any $\varepsilon_2 > 0$, there exists k_2 such that $k > k_2$ implies

$$(2.8)\quad \begin{cases} \theta_{p^k,r^k}(x^k+d^k) - \theta_{p^k,r^k}(x^k) \leq -\frac{1}{2}(1-\varepsilon_2)(x^k-\bar x)^t\, H_{\bar c}(x^k-\bar x) \\ \qquad + (||p^k-\bar\lambda||_{\mathscr{D}}+r^k)||g(x^k+d^k)|| \\ \qquad - (r^k - ||p^k-\bar\lambda||_{\mathscr{D}} - \beta||g(x^k)||)||g(x^k)||. \end{cases}$$
□

Proof

We have

$$\Delta = \theta_{p^k, r^k}(x^k+d^k) - \theta_{p^k, r^k}(x^k)$$

$$= L_{\bar{c}}(x^k+d^k) - L_{\bar{c}}(x^k) + (p^k-\lambda)^t(g(x^k+d^k) - g(x^k))$$

$$- \bar{c}(\sum_{i=1}^{m} g_i(x^k+d^k)^2 - \sum_{i=1}^{m} g_i(x^k)^2)$$

$$+ r^k(||g(x^k+d^k)|| - ||g(x^k)||).$$

As all norms of $\mathbb{R}^m$ are equivalent, there exists $\beta > 0$ such that

(2.9)
$$\Delta \le L_{\bar{c}}(x^k+d^k) - L_{\bar{c}}(x^k) + (||p^k-\bar{\lambda}||_{\mathscr{D}} + r^k)||g(x^k+d^k)||$$
$$- (r^k - ||p-\bar{\lambda}||_{\mathscr{D}} - \beta||g(x^k)||)||g(x^k)||.$$

We focus on the first terms, using (1.3), (1.7) :

$$L_{\bar{c}}(x^k+d^k) - L_{\bar{c}}(x^k) = L_{\bar{c}}(x^k+d^k) - L_{\bar{c}}(\bar{x}) + L_{\bar{c}}(\bar{x}) - L_{\bar{c}}(x^k)$$

$$\le -\frac{1}{2}(x^k-\bar{x})^t H_{\bar{c}}(x^k-\bar{x}) + o(||x^k-\bar{x}||_U^2).$$

This with (2.9) proves the lemma. □

Proof of theorem 2.1

As by (1.4) :

$$g(x^k) + \nabla g(x^k)^t d^k = 0 ,$$

we deduce from (1.7) the existence of some $a_1 > 0$ such that

$$||g(x^k+d^k)|| \le a_1||x^k-\bar{x}||^2_U.$$

From lemma 2.2 we deduce that for some $a_2 > 0$ and k great enough :

$$\Delta_1 = \theta_{p^k,r^k}(x^k+d^k) - \theta_{p^k,r^k}(x^k) \le -\frac{1}{2}[1 - \varepsilon_2 - a_2(||p^k-\bar{\lambda}||_{\mathcal{D}} + $$
$$+ r^k))](x^k-\bar{x})^t H_{\bar{c}}(x^k-\bar{x}) - (r^k - ||p^k-\bar{\lambda}||_{\mathcal{D}} - \beta||g(x^k)||)||g(x^k)||.$$

Using (2.4), (2.5) we deduce that for k great enough and r_o, small enough

$$\Delta_1 \le -\frac{1}{2}(1 - 2\varepsilon_2)(x^k-\bar{x})^t H_{\bar{c}}(x^k-\bar{x}) - \frac{2}{3} r^k||g(x^k)||.$$

From lemma 2.1 and (2.4), (2.5) we deduce that

$$\tilde{\theta}^k_{p^k,r^k}(x^k+d^k) - \theta_{p^k,r^k}(x^k) \ge -(1+\varepsilon_1)(x^k-\bar{x})^t H_{\bar{c}}(x^k-\bar{x}) - \frac{4}{3} r^k||g(x^k)||.$$

As ε_1 and ε_2 are arbitrarily small, this proves the theorem. □

This result suggests to build algorithms, using a penalty function of type $\theta_{p^k,r^k}(x)$, where p^k and r^k are modified at each iteration in order to

- insure a global convergence,
- check the hypothesis of theorem 2.1 after a finite number of iterations.

This is the subject of the following section.

III - A GLOBALLY AND SUPERLINEARLY CONVERGENT ALGORITHM

We define

(3.1) $\quad \theta^k(x) = \tilde{\theta}^k(x^k) + \frac{1}{2}(x-x^k)^t H^k(x-x^k).$

We consider an algorithm of the following type :

Algorithm 1

0) Choose x^1, p^1, r^1, H^1 such that $r^1 > 0$ and H^1 is positive definite. Set $k = 1$, Choose $\beta \in]0,1[$, $\sigma \in]0,1/2[$.

1) Solve the problem

(3.2) $\quad \text{Min } \theta^k(x^k+d), \ d \in \mathbb{R}^n.$

Let $\tilde{d}^k$ be the unique solution of (3.2).

2) Let ℓ be the smallest integer such that

(3.3) $$\theta_{p^k,r^k}(x^k + (\beta)^\ell d^k) - \theta_{p^k,r^k}(x^k) \le \sigma(\tilde{\theta}^k(x^k + (\beta)^\ell d^k) - \theta_{p^k,r^k}(x^k)).$$

$x^{k+1} = x^k + (\beta)^\ell d^k.$

3) $k = k + 1.$

Set p^k, r^k, H^k.

Go to 1). □

We now proceed to give an explicit adaptation rule for p^k and r^k. This needs some preliminary considerations. Consider

$$Q^k(\lambda) = \frac{1}{2} \lambda^t \nabla g(x^k)^t (H^k)^{-1} \nabla g(x^k)\lambda + \lambda^t(\nabla g(x^k)^t(H^k)^{-1} \nabla f(x^k) - g(x^k)).$$

It is well-known that the quadratic program (1.4) is equivalent to

(3.4) $$\begin{cases} \text{(i)} & \lambda^k = \arg\min Q^k(\lambda), \quad \lambda \in \mathbb{R}^m, \\ \text{(ii)} & d^k = -(H^k)^{-1}(\nabla f(x^k) + \nabla g(x^k)\lambda^k). \end{cases}$$

On the other hand, from J.F. Bonnans, D. Gabay [3] we deduce that (3.2) is equivalent to

$$(3.5)\quad \begin{cases} \mu^k = \arg\min \{Q^k(\mu) + \mu^t(\nabla g(x^k)^t (H^k)^{-1} \nabla g(x^k)\, p^k, \quad ||\mu||_{\mathcal{D}} \le r^k\}, \\ \tilde{\lambda}^k = p^k + \mu^k \\ \tilde{d}^k = -(H^k)^{-1}(\nabla f(x^k) + \nabla g(x^k)\, \tilde{\lambda}^k). \end{cases}$$

Suppose that μ^k checks $||\mu^k||_{\mathcal{D}} < r^k$. Then by (3.5), μ^k is solution of

$$\nabla g(x^k)^t (H^k)^{-1} \nabla g(x^k)\mu^k = \nabla g(x^k)^t (H^k)^{-1} \nabla f(x^k) - g(x^k) - \nabla g(x^k)^t (H^k)^{-1} \nabla g(x^k) p^k,$$

or equivalently

$$\mu^k = \lambda^k - p^k$$

where λ^k is the solution of (3.4i) ; then $\tilde{\lambda}^k$ is equal to λ^k. Consequently :

$$(3.6)\quad \{||\mu^k||_{\mathcal{D}} < r^k\} \implies \{(3.4) \text{ has a solution } (d^k,\lambda^k) \text{ equal to } (\tilde{d}^k,\tilde{\lambda}^k)\}.$$

We define the sequences

$$D^k = ||\nabla f(x^k) + \nabla g(x^k)\, \tilde{\lambda}^k||_U + ||g(x^k)||,$$

$$S^k = \max\{1/D^\ell, \quad \ell = 1 \text{ to } k\},$$

$$\varepsilon^k = S^k - S^{k-1}.$$

The monotonically increasing sequence $\{S^k\}$ has the following property :

<u>Lemma 3.1</u>

If $\{x^k\}$ is bounded, a subsequence of $\{(x^k,\tilde{\lambda}^k)\}$ converges towards $(\bar{x},\bar{\lambda})$ such that (1.3) holds iff $S^k \to +\infty$. □

Let α_i, i = 1 to 4, be some positive constants. The adaptation rule for p^k is

$$(3.7)\quad \begin{cases} \text{Let } \ell \text{ be the index of the last iteration at which } p^k \text{ has been changed.} \\ \text{Then} \\ p^{k+1} = \begin{cases} \lambda^k & \text{if } \varepsilon^k > \alpha_1 \text{ or } S^k \geq (1+\alpha_2)\, S^\ell, \\ p^{k-1} & \text{elsewhere.} \end{cases} \end{cases}$$

We need some tools to define the adaptation law on r^k. Let the function s : $\mathbb{R}^{+*} \to R^+$ be defined by

$$s(a) = \min \{10^q \; ; \; q \in Z \; ; \; a \leq 10^q\} .$$

If $\{a^n\}$ is a sequence of positive numbers, the transformed sequence $\{s(a^n)\}$ has the following properties :

$$(3.8)\qquad \{\{a^n\} \to 0\} \iff \{s(a^n) \to 0\} ,$$

$$(3.9)\qquad \{\limsup a^n = +\infty\} \iff \{\limsup s(a^n) = +\infty\},$$

$$(3.10)\qquad \{a^n \nearrow a,\ 0 < a < +\infty\} \implies \{s(a^n) = s(a) \text{ if } n \text{ is great enough}\}.$$

We define the sequence ϕ^k by

0) $\phi^1 = 0$, $k = 2$, $\ell = 1$.

1) If $S^k - S^\ell > 1$ and $\rho^k \neq 1$, $\phi^k = S^k - S^\ell$ and $\ell = k$.
Else $\phi^k = 0$.

2) k = k+1, go to 1.

We have

$$(3.11)\qquad \sum_k S^k = +\infty \iff \sum_k \phi^k = +\infty ,$$

$$(3.12)\qquad \sum_k S^k < +\infty \implies \phi^k = 0 \text{ for } k \text{ superior to some } k_o .$$

We suppose that

(3.13) $\alpha_4 < 1$.

The adaptation rule for r^k is given by

$$
(3.14)\quad \begin{cases}
\text{(i)} & r_1^{k+1} = (3+\alpha_3)\max(\|\tilde{\lambda}^k - p^k\|_{\mathcal{D}},\ \min(1,(D^k)^{\alpha_4})) \\
\text{(ii)} & r_2^{k+1} = \begin{cases} \max(r_1^{k+1},\ r_2^k) \text{ if } \rho^k = 1, \\ \max(r_1^{k+1},\ r_2^k - \phi^k,\ \frac{1}{2} r_2^k) \text{ if not ;} \end{cases} \\
\text{(iii)} & r^{k+1} = s(r_2^{k+1}).
\end{cases}
$$

We now prove that the resulting algorithm is globally and superlinearly convergent.

<u>Theorem 3.1</u>

Let $\{x^k\}$ be a sequence computed by algorithm 1, p^k and r^k being given by (3.7), (3.14). We suppose that the sequences $\{H^k\}$ and $\{(H^k)^{-1}\}$ are bounded and that $\sigma < 1/2$. Then

a) One of the four following events occurs :

(i) $\liminf_{k\to\infty} (\|\nabla f(x^k) + \nabla g(x^k)\,\tilde{\lambda}^k\|_U + \|g(x^k)\|) = 0.$

(ii) for k great enough, (p^k, r^k) is equal to some (p,r) and

$$
\begin{cases}
r > \|\tilde{\lambda}^k\|_{\mathcal{D}}, \\
\theta_{p,r}(x^k) \to -\infty.
\end{cases}
$$

(iii) The sequence $\{x^k\}$ is bounded and some limit point $\bar{x}$ of $\{x^k\}$ is such that $\{\nabla g_i(\bar{x}),\ i=1 \text{ to } m\}$ are not linearly independant.

(iv) The following relations hold

$$\limsup_{k\to\infty} ||x^k||_U = +\infty \quad ,$$

$$\limsup_{k\to+\infty} r^k = \limsup_{k\to+\infty} ||\tilde{\lambda}^k||_{\mathcal{D}} = +\infty.$$

b) If the sequence converges towards some $\bar{x}$ at which (1.2) and the standard second-order sufficiency conditions hold, and if (1.7) holds, then ρ^k is equal to 1 for k great enough and $x^k \to \bar{x}$ superlinearly. □

Proof

a) We suppose that (i) does not occur : then S^k is bounded, hence p^k is equal to some p for great enough. Suppose now that

$$\limsup_{k\to\infty} ||\lambda^k||_{\mathcal{D}} < +\infty.$$

We then prove that r^k is equal to some r for k great enough. By (3.12), ϕ^k is null for k great enough ; (3.14ii) shows that r_2^k is an increasing sequence (for k great enough) ; as r_1^k is bounded, we get $r_2^k \geq r_2$ with, by (3.14i)

$$r_2 > ||\tilde{\lambda}^k - p^k||_{\mathcal{D}} \quad , k > k_1.$$

Using (3.14iii) and (3.10) we see that

$$r^k = r > ||\tilde{\lambda}^k - p^k||_{\mathcal{D}} \quad , k > k_2, \tag{3.15}$$

hence

$$g(x^k) + \nabla g(x^k)^t d^k = 0, \; k > k_3. \tag{3.16}$$

We now prove that

$$\theta_{p,r}(x^k) - \theta_{p,r}(x^{k+1}) \geq \sigma\, \rho^k (d^k)^t\, H^k\, d^k, \quad k > k_4. \tag{3.17}$$

In fact, we have, using (3.4ii) ($k > k_5$)

$$\Delta = \theta_{p,r}(x^k) - \tilde{\theta}^k(x^k+d^k) = - \nabla f(x^k)^t\, d^k - p^t\, \nabla g(x^k)^t\, d^k + r||g(x^k)||$$

$$= (d^k)^t\, H^k\, d^k + (\lambda^k-p)^t\, \nabla g(x^k)^t\, d^k + r||g(x^k)||,$$

and with (3.15), (3.16)

$$\geq (d^k)^t\, H^k\, d^k + (r-||\lambda^k-p||_{\mathcal{D}})||g(x^k)||,$$

$$\geq (d^k)^t\, H^k\, d^k.$$

The convexity of θ^k implies

$$\theta_{p,r}(x^k) - \theta^k(x^k + \rho^k d^k) \geq \frac{\rho^k}{2}(d^k)^t\, H^k d^k.$$

Then (3.17) is a consequence of the line-search rule (3.3). Because of (3.17), if (ii) does not hold, we have

(3.18) $$\sum_k \rho^k (d^k)^t\, H^k d^k < +\infty ,$$

hence

(3.19) $$\sum_k ||x^{k+1} - x^k||_U\, ||\nabla f(x^k) + \nabla g(x^k)\lambda^k||_U < +\infty.$$

First suppose that for some subsequence

$$||\nabla f(x^k) + \nabla g(x^k)\, \lambda^k||_U \to 0.$$

Then by (3.4ii) $d^k \to 0$ for that subsequence ; we deduce from (3.16) that $S^k \to +\infty$; this contradicts our hypothesis. On the contrary, if

$$\liminf_{k\to\infty} ||\nabla f(x^k) + \nabla g(x^k)\, \lambda^k||_U > 0,$$

then (3.19) implies the convergence of $\{x^k\}$ towards some $\bar{x}$; as p^k and r^k are constant for k great enough, we deduce from the proof of theorem 2.2 of

J.F. Bonnans, D. Gabay [3] that $d^k \to 0$ for some subsequence. This is in contradiction with the fact that S^k is bounded.

We have proved that if (i) does not hold and $\{\lambda^k\}$ is bounded, then (ii) holds. To prove point a), it is sufficient to prove that if $\{x^k\}$ is bounded and $\{\lambda^k\}$ is unbounded, then (iii) holds. As p^k is independant of k for k great enough this is a simple consequence of (3.5).

b) Let $\bar{\lambda}$ be the multiplier associated to $\bar{x}$. We have

$$p^k \to \bar{\lambda},$$
$$r_1^k \to 0.$$

As D^k is of the same order than $||x^k-\bar{x}||_U$ (see [2] for instance) we have by (3.13), (3.14i)

$$\frac{r_1^k}{||x^k-\bar{x}||_U} \to +\infty \quad ,$$

and as $r^k \geq r_1^k$, (2.5) hold. As $p^{k^q} = \lambda^{k^q}$ for some subsequence k^q and as $||\lambda^k-\bar{\lambda}||_{\mathscr{D}}$ is majorized by $a_1||x^k-\bar{x}||_U$ for some $a_1 > 0$, hypothesis (2.4) holds (for q great enough) along the subsequence k^q. If ρ^k is not equal to one for $k > k_6$, the sequence ϕ^k is constructed in such a way that r_2^k converges towards zero ; by (3.8), so does also r^k. Consequently, by theorem 2.1, and as $\sigma < 1/2$, we have

$$x^{k^q+1} = x^{k^q} + d^{k^q}, \text{ q great enough.}$$

We note that S^{k^q} is equal to $(D^{k^q})^{-1}$. By (1.7), for any $\varepsilon > 0$ we have

$$||x^{k^q+1}-\bar{x}||_U < \varepsilon\ ||x^{k^q}-\bar{x}||_U \text{ for q great enough.}$$

As D^k is of the same order than $||x^k-\bar{x}||_U$ this implies that

$$D^{k^q+1} < \frac{1}{1+\alpha_2} D^{k^q} , \quad q > q_1,$$

hence

$$S^{k^q+1} < \frac{1}{1+\alpha_2} S^{k^q} , \quad q > q_1 ;$$

hence by (3.7), p^{k^q+1} will again be equal to λ^{k^q+1}. This means that $k^{q+1} = k^q+1$ for $q > q_2$. This proves the theorem. □

REFERENCES

[1] L. ARMIJO (1966). Minimization of functions having lipschitz continuous first partial derivatives. Pac. J. Math. 16, pp. 1-3.

[2] P.T. BOGGS, J.W. TOLLE, Pyng Wang (1982). On the local convergence of quasi-Newton methods for constrained optimization. SIAM J. Cont. Opt. 20, pp. 161-171.

[3] J.F. BONNANS, D. GABAY (1984). Une extension de la programmation quadratique successive. Sixième Conférence Internationale Analyse et Optimisation des Systèmes, A. Bensoussan, J.L. Lions ed; to appear in the Lecture notes in control and information sciences, Springer Verlag, Berlin.

[4] R.M. CHAMBERLAIN, C. LEMARECHAL, H.C. PEDERSEN, M.J.D. POWELL (1982). The watchdog technique for forcing convergence in algorithms for constrained optimization. Math. Prog. Study 16, pp. 1-17.

[5] T.F. COLEMAN, A.R. CONN (1982). Nonlinear programming via an exact penalty function. Math. Prog. 24, pp. 123-161.

[6] R. FLETCHER (1981). Practical methods of optimization. Vol. 2, J. Wiley, Chichester.

[7] R. FLETCHER (1982). A model algorithm for composite non differentiable optimization problems. Math. Prog. Study 17, pp. 67-76.

[8] D. GABAY (1982). Reduced quasi-Newton methods with feasibility improvement for nonlinearly constrained optimization. Math. Prog. Study 16, pp. 18-44.

[9] S.P. HAN (1977). A globally convergent method for nonlinear programming. J. of Opt. Th . and Appl. 22, pp. 297-309.

[10] N. MARATOS (1978). Exact penalty function algorithms for finite dimensional and control optimization problems. Thesis, University of London.

[11] D.Q. MAYNE, E. POLAK (1982). A superlinearly convergent algorithm for constrained optimization problems. Math. Prog. Study 16, pp. 45-61.

[12] M.J.D. POWELL (1978). A fast algorithm for nonlinearly constrained optimization calculations, in Lecture Notes in Mathematics n° 630, G.A. Watson ed., Springer Verlag, Berlin, pp. 144-157.

[13] M.J.D. POWELL(1978). The convergence of variable metric methods for nonlinearly constrained calculations. Nonlinear Programming 3, O.L. Mangasarian, R.R. Meyer, S.M. Robinson ed., Academic Press, New York, pp. 27-63.

[14] B.N. PSHENICHNYI, Y.D. DANILIN (1977). Numerical methods in extremal problems, Mir Publishers, Moscow.

[15] K. TONE (1983). Revisions of constraint approximations in the successive Q.P. methods for nonlinear programming problems, Math. Prog. 26, pp. 144-152.

Imprimé en France

par

l'Institut National de Recherche en Informatique et en Automatique