A STEP-SIZE SELECTION PROCEDURE FOR EQUALITY CONSTRAINED OPTIMIZATION *

Jean Charles GILBERT

International Institute for Applied Systems Analysis
A - 2361 Laxenburg (Austria)

Résumé. Dans cet article, on propose une méthode généralisant à l'optimisation avec contraintes d'égalité, une technique de sélection de pas, dite de Wolfe, qui a fait ses preuves en optimisation sans contrainte. Pour les algorithmes du type quasi-Newtonien, celle-ci semble assez naturelle puisqu'elle permet d'assurer facilement la définie positivité des métriques locales à chaque itération et donc le caractère descendant des directions de recherche. Toutefois, on sait que cette technique ne peut pas être étendue aux méthodes quasi-Newtoniennes en optimisation avec contraintes. On montre ici qu'une généralisation est possible dans le cadre des méthodes sécantes réduites.

Abstract. This paper proposes a generalization to equality constrained optimization of Wolfe's step-size selection procedure, which is used with success in unconstrained optimization. This one appears rather natural for quasi-Newton methods because it allows to maintain easily the positive definiteness of the matrices correcting the steepest descent direction and therefore assures the descent property to search directions. However, this technique is known not to be usable for quasi-Newton methods in constrained optimization. We show here that a generalization can be made in the framework of reduced secant methods.

* Work supported by the FNRS (Fonds National de la Recherche Scientifique) of Belgium.

1. Introduction

In this report, we deal with the following equality constrained optimization problem:

$$\min \{ f(x) : x \in \omega , c(x) = 0 \} , \tag{1.1}$$

where ω is a convex open set in $\mathbb{R}^n$ and the functions $f : \omega \longrightarrow \mathbb{R}$ and $c : \omega \longrightarrow \mathbb{R}^m$, $m < n$, are supposed smooth. We shall suppose that the $m \times n$ Jacobian matrix $A(x) := \nabla c(x)$ has full rank m for all x in ω and that ω contains a local solution x_* of problem (1.1), which with its associated Lagrange multiplier λ_* satisfies the standard sufficient conditions of optimality (see Fletcher (1981)):

$$\left\{ \begin{array}{l} c(x_*) = 0 , \\ \nabla f(x_*) + A(x_*)^T \lambda_* = 0 \end{array} \right. \tag{1.2}$$

and

$$G_* := Z(x_*)^{-T} L_* Z(x_*)^- \text{ is positive definite .} \tag{1.3}$$

In (1.3), L_* is the Hessian according to x of the Lagrangian $l(x,\lambda) := f(x) + c(x)^T\lambda$ at (x_*,λ_*) and $Z(x_*)^-$ is a basis of $N(A(x_*))$, the kernel of $A(x_*)$, i.e. an $n \times (n-m)$ matrix whose columns form a basis of $N(A(x_*))$. We shall suppose that such a basis exists at each point x in ω in such a way that its dependence on x is smooth. We have

$$A(x)\ Z(x)^- = 0 \text{ in } \mathbb{R}^{m\times(n-m)} \text{ for all } x \text{ in } \omega . \tag{1.4}$$

We shall also need to do displacements in the complementary space $R(A(x)^-)$ of $N(A(x))$. We shall also suppose that the right inverse $A(x)^-$ of $A(x)$ is a smooth function of x. We have

$$A(x)\ A(x)^- = I \text{ in } \mathbb{R}^{m\times m} \text{ for all } x \text{ in } \omega . \tag{1.5}$$

This formalism (with the matrix $Z(x)$ introduced later) is due to Gabay (1982).

We shall focus in this paper on the following class of reduced secant methods to compute iteratively a solution of problem (1.1) (see Coleman and Conn (1982)): starting from a point x_k, the next iterate x_{k+1} is obtained by

$$y_k := x_k - A(x_k)^- c(x_k) =: x_k + r_k , \tag{1.6}$$

$$x_{k+1} := y_k - Z(y_k)^- H_k g(y_k) =: y_k + t_k . \tag{1.7}$$

In (1.7), H_k is a matrix of order $n-m$ approximating $H_* := G_*^{-1}$ and g is the *reduced gradient* of f defined by $g(y) := Z(y)^{-T}\nabla f(y)$. The fact that only a matrix of order $n-m$, namely H_k, has to be updated is an advantage of reduced methods over quasi-Newton methods to solve (1.2). Indeed, in the latter case, a matrix L_k of order n approximating the Hessian L_* has to be updated. For a state of the art on quasi-Newton methods for constrained optimization we refer to the paper by Powell (1986).

The convergence of algoritm (1.6)-(1.7) is assured locally, i.e. when (x_0,H_0) is close to (x_*,H_*), and is then superlinearly convergent if the matrix H_k is correctly updated: see Byrd (1984) and Gilbert (1988). Moreover, the algorithm can be globalized to force convergence from poor starting approximations by using as a merit function a penalty function of the form

$$\theta(x) := f(x) + \varphi(c(x)) . \tag{1.8}$$

See Gilbert (1986,a,b,c). In (1.8), φ is a positive real-valued convex function that we shall suppose continuous in a neighborhood of $c(\omega)$. Therefore we can compute the directionnal derivatives of θ. For example, using (1.4), the directionnal derivative of θ at y_k in the direction t_k writes $\theta'(y_k\,;\,t_k) = -\,g(y_k)^T\,H_k\,g(y_k)$. This shows that t_k is a descent direction of θ at y_k if H_k is positive definite.

To maintain the positive definiteness of the matrices H_k, we propose to update them by the *inverse BFGS formula* (see e.g. Dennis and Moré (1977)):

$$H_{k+1} = \left[I - \frac{\delta_k\,\gamma_k^T}{\gamma_k^T\,\delta_k}\right] H_k \left[I - \frac{\gamma_k\,\delta_k^T}{\gamma_k^T\,\delta_k}\right] + \frac{\delta_k\,\delta_k^T}{\gamma_k^T\,\delta_k} . \tag{1.9}$$

Here, δ_k and γ_k are vectors in $\mathbb{R}^{n-m}$. Then, it is well known that H_{k+1} retains the positive definiteness of H_k if and only if

$$\gamma_k^T\,\delta_k > 0 . \tag{1.10}$$

Other update formula may be used if they retain symmetry and positive definiteness under condition (1.10), e.g., formula of the restricted Broyden's class.

Inequality (1.10) must be satisfied at each iteration. Note that it is satisfied locally if we take

$$\gamma_k := g(x_{k+1}) - g(y_k) , \tag{1.11}$$

$$\delta_k := Z(y_k)\,t_k , \tag{1.12}$$

where y_k and x_{k+1} are defined in (1.6)-(1.7) and $Z(y)$ is the unique $(n-m)\times n$ matrix satisfying $Z(y)Z(y)^- = I$ in $\mathbb{R}^{(n-m)\times(n-m)}$ and $Z(y)A(y)^- = 0$ in $\mathbb{R}^{(n-m)\times m}$. This can be seen by expanding $g(x_{k+1})$ about y_k in (1.11), observing with Stoer (1984) that

$$\nabla g(x_*) = Z(x_*)^{-T} L_* . \tag{1.13}$$

This formula shows why γ_k is taken as a difference of reduced gradients and δ_k is the corresponding reduced displacement.

The aim of this paper is to show how the vectors γ_k and δ_k can be calculated in a global framework (when some step-length is introduced) in order to satisfy inequality (1.10) at each iteration and to recover the values (1.11) and (1.12) after a finite number of steps.

Hereafter, many implicit references will be done to Gilbert (1987). In particular, we refer to this paper for the proof of the theorems.

2. The longitudinal displacement

In this section, we show how a step-size procedure attributed to Wolfe (1969) and used in unconstrained optimization can be extended to equality constrained optimization for the longitudinal displacement, which starts from y_k tangently to the manifold $M(y_k) := c^{-1}(c(y_k))$. First, we recall Wolfe's procedure.

In unconstrained minimization, $min \{ \psi(u) : u \in I\!R^p \}$, a large class of methods consist in generating a sequence of approximation (u_k) of a solution such that $u_{k+1} := u_k + \tau_k v_k$, where v_k is a descent direction of ψ at u_k and τ_k is a positive scalar step-size. In this general framework, Wolfe's procedure to determine τ_k, consists in finding a positive τ such that

$$\psi(u_k+\tau v_k) \leq \psi(u_k) + \alpha_1 \tau \nabla\psi(u_k)^T v_k , \tag{2.1}$$

$$\nabla\psi(u_k+\tau v_k)^T v_k \geq \alpha_2 \nabla\psi(u_k)^T v_k , \tag{2.2}$$

where $0 < \alpha_1 < \alpha_2 < 1$ are given. Such a τ exists for example if ψ is bounded from below. In secant methods, the descent direction has the form $v_k := - H_k \nabla\psi(u_k)$, where H_k is the current positive definite approximation of the inverse of the Hessian of ψ at u_k. Usually, H_k is updated by using a formula like (1.9) with

$$\gamma_k := \nabla\psi(u_{k+1}) - \nabla\psi(u_k) ,$$

$$\delta_k := u_{k+1} - u_k = \tau_k v_k .$$

Therefore, if the step-size τ_k is determined by Wolfe's procedure, inequality (2.2) implies the positivity of the scalar product of γ_k and δ_k.

In a first stage, we generalize this to a longitudinal displacement on the manifold $M(y_k)$. For this, we introduce the curve $y_k^M : \tau \in I\!R^1 \longrightarrow y_k^M(\tau) \in M(y_k)$ by the differential equation:

$$\begin{cases} \dot{y}_k^M(\tau) = Z(y_k^M(\tau))^- \, Z(y_k)\, t_k \,, \\ y_k^M(0) = y_k \,. \end{cases} \tag{2.3}$$

Multiplying the first equation of (2.3) to the left by $A(y_k^M(\tau))$ and using (1.4) shows that $y_k^M(\tau)$ belongs to $M(y_k)$. Then, if we define $\psi(\tau) := (\theta \, o \, y_k^M)(\tau)$, with θ given in (1.8), we have $\psi'(\tau) = \theta'(y_k^M(\tau) \,;\, Z(y_k^M(\tau))^- Z(y_k) t_k) = g(y_k^M(\tau))^T Z(y_k) t_k$ (because of (1.4)). So, applying Wolfe's procedure to this ψ with $p = 1$, $u_k = 0$ and $v_k = 1$ shows that, provided $y_k^M(\tau)$ exists for sufficiently large value of τ and f is bounded from below on $M(y_k)$, we can find a positive step-size τ such that

$$\theta(y_k^M(\tau)) \leq \theta(y_k) + \alpha_1 \, \tau \, \nabla f(y_k)^T \, t_k \,, \tag{2.4}$$

$$g(y_k^M(\tau))^T \, Z(y_k) \, t_k \geq \alpha_2 \, g(y_k)^T \, Z(y_k) \, t_k \,. \tag{2.5}$$

Noting τ_k such a τ and defining

$$\gamma_k^M := g(y_k^M(\tau_k)) \; - g(y_k) \,, \tag{2.6}$$

$$\delta_k^M := \tau_k \, Z(y_k) \, t_k \,, \tag{2.7}$$

we obtain from (2.5): $(\gamma_k^M)^T \, \delta_k^M > 0$.

In this framework, (1.11)-(1.12) may be obtained from (2.6)-(2.7) by taking a first order approximation of $y_k^M(\bullet)$ and a unit step-size. We continue with the same point of vue and we build an explicit Euler approximation of the solution of the differential equation (2.3) for the following discretization points:

$$0 =: \tau_k^0 < \tau_k^1 < \cdots < \tau_k^{l_k} \,. \tag{2.8}$$

The problem is to choose them carefully in number and position so that for $l = l_k$, there exists a positive τ with

$$\theta(y_k^l(\tau)) \leq \theta(y_k) + \alpha_1 \, \tau \, \nabla f(y_k)^T \, t_k \,, \tag{2.9}$$

$$g(y_k^l(\tau))^T \, Z(y_k) \, t_k \geq \alpha_2 \, g(y_k)^T \, Z(y_k) \, t_k \,. \tag{2.10}$$

Again, $0 < \alpha_1 < \alpha_2 < 1$ are given. Let us denote by y_k^0 the point y_k and by $y_k^0(\tau)$ the line $y_k + \tau \, t_k$. For $1 \leq l \leq l_k$, let us define recursively the points $y_k^l := y_k^{l-1}(\tau_k^l)$ and the piecewise linear curves

$$y_k^l(\tau) := \begin{cases} y_k^{l-1}(\tau) \quad \text{for} \quad 0 \leq \tau \leq \tau_k^l \,, \\ y_k^l + (\tau - \tau_k^l) \, Z(y_k^l)^- \, Z(y_k) \, t_k \quad \text{for} \quad \tau > \tau_k^l \,. \end{cases} \tag{2.11}$$

Therefore the Euler approximation of $y_k^M(\bullet)$ for the discretization (2.8) is $y_k^{l_k}(\bullet)$. We propose to determine the discretization points τ_k^l by the following algorithm.

Longitudinal search algorithm: (2.12)

1. if $t_k = 0$ then stop
2. set $l := 0$, $\tau_k^0 := 0$; choose $\tau > 0$
3. repeat:
   3.1. if $y_k^l(\tau)$ is not in ω or (2.9) is not satisfied
   3.2. then $\tau := \tau_k^l + J(y_k^l, \tau - \tau_k^l)$
   3.3. else { if (2.10) is satisfied
   3.4. then { $\tau_k := \tau$; $l_k := l$; stop }
   3.5. else { $l := l+1$; $\tau_k^l := \tau$; $\tau := \tau_k^l + E(y_k^l)$ } }

This algorithm is inspired by an algorithm proposed by Lemaréchal (1981) to find a step-size satisfying (2.1)-(2.2) in unconstrained optimization.

The algorithm uses two functions J and E. The *interpolation function* J : $\omega \times \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ is supposed to satisfy the following properties:

$$J \text{ is continuous on } \omega \times \mathbb{R}_+ , \tag{2.13}$$

$$0 < J(y,\tau) < \tau \text{ for all } y \text{ in } \omega \text{ and all positive } \tau , \tag{2.14}$$

$$J_y^p(\tau) := (J_y \, o \, .. \, (p \text{ times}) \, .. \, o \, J_y)(\tau) \longrightarrow 0 \text{ as } p \longrightarrow \infty . \tag{2.15}$$

In (2.15), we have denoted $J_y(\bullet) := J(y,\bullet)$. Therefore the step-size τ is decreased in statement 3.2 (property (2.14)) and tends to τ_k^l if the statement is executed infinitely often (property (2.15)). The *extrapolation function* $E : \omega \longrightarrow \mathbb{R}_+$ is supposed to satisfy the following properties:

$$E \text{ is continuous on } \omega , \tag{2.16}$$

$$0 < E(y) \text{ for all } y \text{ in } \omega . \tag{2.17}$$

Therefore, the step-size τ is increased in statement 3.5.

Let us make some remarks on algorithm (2.12).

1) If the algorithm terminates in statement 3.4 after l loops 3.1-3.5, this means that a step-size τ has been found such that $y_k^l(\tau)$ is in ω, (2.9) is satisfied just as (2.10).

2) If a new discretization point is taken at $\tau = \tau_k^{l+1}$ in statement 3.5, this means that $y_k^l(\tau)$ is in ω, (2.9) is satisfied but not (2.10).

3) The test in statement 3.1 is always rejected after a finite number of loops 3.1-3.2. Indeed, because y_k^l is in ω, $y_k^l(\tau)$ is also in ω for τ in a right neighborhood of τ_k^l. On the other hand, at $y_k^l := y_k^l(\tau_k^l)$, we have $(\theta \, o \, y_k^l)'(\tau_k^l \, ; \, 1) = \theta'(y_k^l \, ; \, Z(y_k^l)^- Z(y_k) t_k) =$

$g(y_k^l)^T Z(y_k) t_k$ (by (1.4)!) $< \alpha_2\, g(y_k)^T Z(y_k) t_k$ (because by construction inequality (2.10) is not satisfied at the point y_k^l) $< \alpha_1\, g(y_k)^T Z(y_k) t_k$ (because $g(y_k)^T Z(y_k) t_k$ is negative) $= \alpha_1\, \nabla f(y_k)^T t_k$. Hence inequality (2.9) is also satisfied in a right neighborhood of τ_k^l because at y_k^l, inequality (2.9) is satisfied and the right derivative of its left hand side is less than the one of its right hand side. Finally, property (2.15) of the interpolation function J shows that after a finite number of loops 3.1-3.2 the value of the step-size τ belongs to the intersection of the above mentioned right neighborhoods of τ_k^l.

Remark 3 shows that either the algorithm terminates in a finite number l of loops 3.1-3.5 or that a sequence $(y_k^l)_{l\geq 0}$ is generated in ω. The following theorem shows that the latter case is, in a way, pathological and that, in general, algorithm (2.12) works well.

Theorem 2.1. *Let θ be the function defined on ω by (1.8) with φ convex, positive and continuous on a neighborhood of $c(\omega)$. Let y_k be a point in ω such that $g(y_k) \neq 0$. Let H_k be a symmetric positive definite matrix of order $n-m$. Then, if the longitudinal search algorithm (2.12) with the definitions (2.8)-(2.11) and the hypotheses (2.13)-(2.17) is applied from y_k, one of the following situations occurs:*

(i) the algorithm terminates in a finite number l_k of loops 3.1-3.5, with a point $x_{k+1} := y_k^{l_k}(\tau_k)$ satisfying both inequalities (2.9) and (2.10),

(ii) the algorithm builds a sequence $(y_k^l)_{l\geq 0}$ in ω and either $(\theta(y_k^l))_{l\geq 0}$ tends to $-\infty$ or $(y_k^l)_{l\geq 0}$ converges to a point on the boundary of the open set ω.

Note that situations in point (ii) do not occur when $\omega = \mathbb{R}^n$ and f is bounded from below. From now on, we shall denote by x_{k+1} the point $y_k^{l_k}(\tau_k)$ found by algorithm (2.12).

3. The algorithm

To insert the longitudinal search algorithm (2.12) in a globally convergent algorithm for problem (1.1), two things remain to be specified: a penalty function, i.e. a function φ in (1.8), and a step-size selection procedure for the transversal displacement (governed by the step r_k in (1.6)).

3.1. A penalty function

The choice of the penalty function is here essentially determined by the longitudinal displacement and the fact that a unit step-size τ_k (with $l_k = 0$) is necessary for not preventing the superlinear convergence from occuring. Like Bonnans (1984), we propose to use the nondifferentiable augmented Lagrangian:

$$l_{p_k}(x,\mu_k) := f(x) + \mu_k^T\, c(x) + p_k\, ||c(x)|| \,, \tag{3.1}$$

where $p_k > 0$, $\mu_k \in I\!R^m$ and $||\bullet||$ is a norm on $I\!R^m$. This penalty function is exact (that is to say that x_* is a local minimizer of it) if $p_k > ||\mu_k - \lambda_*||_D$, where $||\bullet||_D$ is the dual norm of $||\bullet||$: see also Han and Mangasarian (1979). In the algorithm proposed below, p_k is normally modified finitely often to be small enough asymptotically while μ_k normally tends to λ_*. Therefore the penalty function $l_{p_k}(\bullet,\mu_k)$ will resemble more and more the true Lagrangian $l(\bullet,\lambda_*)$. Now, because H_k is updated to be a good approximation of the inverse of the projected Hessian of the Lagrangian, the unit longitudinal step-size will be garanteed asymptotically.

3.2. The transversal displacement

We have $l'_{p_k}(x_k,\mu_k\,;\,r_k) = (\lambda(x_k)-\mu_k)^T\, c(x_k) - p_k\, ||c(x_k)||$, where $\lambda(\bullet)$ is the approximation of the Lagrange multiplier λ_* defined by

$$\lambda(x) := -\, A(x)^{-T}\, \nabla f(x) \,. \tag{3.2}$$

Therefore r_k is a descent direction of the penalty function (3.1) at x_k if

$$p_k \geq ||\lambda(x_k)-\mu_k||_D + \underline{p}_k \,, \tag{3.3}$$

where $\underline{p}_k$ is some positive number. If the algorithn manages to satisfy this inequality at each iteration, the transversal step-size ρ_k can be determined by Armijo's procedure: see Armijo (1966). Let α and β be in $]0,1[$. Then ρ_k is taken in the form

$$\rho_k := \beta^{b_k} \,, \tag{3.4}$$

where b_k is the smallest non negative integer such that:

$$x_k + \rho_k\, r_k \in \omega \tag{3.5}$$

and

$$l_{p_k}(x_k+\rho_k r_k,\mu_k) \leq l_{p_k}(x_k,\mu_k) + \alpha\, \rho_k\, [(\lambda(x_k)-\mu_k)^T\, c(x_k) - p_k\, ||c(x_k)||] \,. \tag{3.6}$$

By expanding $l_{p_k}(x_k+\rho_k r_k,\mu_k)$ in x about x_k, it is not difficult to see that the unit step-size $\rho_k = 1$ is accepted asymptotically if $\underline{p}_k$ in (3.3) is bounded away from zero.

3.3. A reduced secant algorithm

The algorithm we propose hereafter uses

$$\epsilon_k := ||g(y_k)|| + ||c(x_{k+1})|| ,$$

$$\epsilon_k^0 := \min \{ \epsilon_i : 0 \le i \le k \} ,$$

to test the convergence and to decide whether to adapt $\underline{p}_k$, p_k and μ_k.

Reduced secant algorithm: (3.7)

1. Let be given the constants: $0 < \alpha < 1$, $0 < \beta < 1$, $0 < \alpha_1 < 1/2$, $\alpha_1 < \alpha_2 < 1$, $0 < \epsilon$, $1 < a_i$ $(i = 1,2,3)$.
2. Let x_0 be a point in ω and H_0 be a symmetric positive definite matrix of order $n-m$.
3. Calculate $\lambda(x_0)$ by (3.2), choose $\underline{p}_0 > 0$, set $\mu_0 := \lambda(x_0)$ and $p_0 := S(\underline{p}_0)$ and set the indices $k := 0$ (iterations), $i := 0$ (adaptation of $\underline{p}_k$), $j := 0$ (adaptation of p_k and μ_k).
4. Select a transversal step-length ρ_k by Armijo's procedure (3.4)-(3.6) and set $y_k := x_k + \rho_k r_k$.
5. Execute the longitudinal search algorithm (2.12), starting with $\tau = 1$ and using the penalty function $l_{p_k}(\bullet,\mu_k)$ instead of $\theta(\bullet)$ in (2.9) to determine the step-length τ_k and the point $x_{k+1} := y_k^{l_k}(\tau_k)$, if possible.
6. If $\epsilon_k < \epsilon$ then stop.
7. Update H_k by (1.9) with $\gamma_k := g(x_{k+1}) - g(y_k)$ and $\delta_k := \tau_k Z(y_k) t_k$.
8. Adapt $\underline{p}_k$: if $\{ \epsilon_k^0 \le \epsilon_i^0/a_1$ and $(l_k \neq 0$ or $\tau_k \neq 1) \}$ then $\{ i := k$; $\underline{p}_{k+1} := \underline{p}_k/a_2 \}$ else $\underline{p}_{k+1} := \underline{p}_k$.
9. Adapt p_k and μ_k: if $\epsilon_k^0 \le \epsilon_j^0/a_3$ then $\{ j := k$; $\mu_{k+1} := \lambda(x_{k+1})$; $p_{k+1} := S(\underline{p}_{k+1}) \}$ else $\{ \mu_{k+1} := \mu_k$; $p_{k+1} := \max (p_k , S(||\lambda(x_{k+1}) - \mu_{k+1}|| + \underline{p}_{k+1})) \}$.
10. Set $k := k+1$ and go to statement 4.

As function $S :]0,\infty[\longrightarrow]0,\infty[$ in statement 3 and 9, we can take like in Bonnans (1984):

$$S(a) := \min \{\, 10^q : a \leq 10^q \,,\ q \text{ integer} \,\} \,.$$

This non-decreasing function differs from the identity function by the fact that for each $\underline{a} \leq \bar{a}$ in $]0,\infty[$, $S([\underline{a},\bar{a}])$ is finite. This property impedes the penalty factor p_k to change too often and allows to have it constant after a finite number of iterations.

We see that the matrices H_k are updated at each iteration and this is possible because $\gamma_k^T \delta_k$ is positive and therefore formula (1.9) is well defined. In statement 8, $\underline{p}_k$ is decreased when the unit longitudinal step-size ($l_k = 0$ and $\tau_k = 1$) is not admitted and convergence seems to occur (for some subsequence). In statement 9, convergence is also required to adapt p_k and μ_k otherwise only p_k can be increased to preserve inequality (3.3).

4. Convergence results

The following global convergence theorem can be proved.

Theorem 4.1 (global convergence). *Suppose that algorithm (3.7) produces sequences (x_k) and (y_k) in ω and a bounded sequence of matrices (H_k) with bounded inverses. Then, one of the following situations occurs:*

(i) $\liminf (||c(x_k)|| + ||g(y_k)||) = 0$,

(ii) $\mu_k = \mu$ *for* k *large,* (p_k) *is unbounded and* $||\lambda(x_k)||_D \longrightarrow \infty$ *when* $k \longrightarrow \infty$ *in* $\{\, k : p_k > p_{k-1} \,\}$,

(iii) $\mu_k = \mu$ *for* k *large,* $p_k = p$ *for* k *large and either* $l_p(x_k,\mu) \longrightarrow -\infty$ *or for some subsequence* $dist(x_k,\omega^c) \longrightarrow 0$.

Because the theorem assumes that the sequences (x_k) and (y_k) have been generated by the algorithm, it is implicitly supposed that the longitudinal search algorithm (2.12) has succeeded at each iteration k, which, according to theorem 2.1, will usually occur. Situation (ii) of the theorem means that either $\{\, x_k : p_k > p_{k-1} \}$ is unbounded or has accumulation points $\bar{x}$ on the boundary of ω at which $\lambda(\bar{x})$ is not well defined by (1.15), for instance, because $A(\bar{x})$ has not full rank. In (iii), $\text{dist}(x_k,\omega^c)$ is the distance from x_k to the complementary set of ω.

It can also be proved that the transversal and longitudinal unit step-sizes are accepted after a finite number of iterations and, consequently, that the convergence is superlinear.

Theorem 4.2 (superlinear convergence). *Suppose that algorithm (3.7) produces a sequence (x_k) in ω converging in ω to a solution x_* of problem (1.1) and a bounded sequence of positive definite matrices (H_k) with bounded inverses. Suppose, as well, that we have*

$$(G_k - G_*)\, Z(y_k)\, t_k = o(||t_k||)\,. \tag{4.1}$$

Then, $\rho_k = 1$, $l_k = 0$ and $\tau_k = 1$ for k large enough and the sequence (x_k) converges superlinearly.

Condition (4.1) is a sufficient condition of superlinear convergence for the local algorithm (1.6)-(1.7) and is usually satisfied by the update scheme of H_k: see Coleman and Conn (1984), Byrd (1984) and Gilbert (1986,a and 1988).

References

L. Armijo (1966). Minimization of functions having Lipschitz continuous first partial derivatives. Pacific Journal of Mathematics 16/1, 1-3.

J.F. Bonnans (1984). Asymptotic admissibility of the unity stepsize in exact penalty methods I: equality-constrained problems. Rapport de recherche de l'INRIA RR-273. 78153 Le Chesnay, France.

R.H. Byrd (1984). On the convergence of constrained optimization methods with accurate Hessian information on a subspace. Report CU-CS-270-84. Department of Computer Science, University of Colorado, Boulder, Colorado 80309.

T.F. Coleman, A.R. Conn (1982). Nonlinear programming via an exact penalty function: asymptotic analysis. Mathematical Programming 24, 123-136.

T.F. Coleman, A.R. Conn (1984). On the local convergence of a quasi-Newton method for the nonlinear programming problem. SIAM Journal on Numerical Analysis 21/4, 755-769.

J.E. Dennis, J.J. Moré (1977). Quasi-Newton methods, motivation and theory. SIAM Review 19, 46-89.

R. Fletcher (1981). *Practical methods of optimization. Vol. 2: Constrained Optimization.* J. Wiley & Sons.

D. Gabay (1982). Minimizing a differentiable function over a differentiable manifold. Journal of Optimization Theory and its Application 37/2, 171–219.

J.Ch. Gilbert (1986,a). Une méthode à métrique variable réduite en optimisation avec contraintes d'égalité non linéaires. Rapport de recherche de l'INRIA RR-482. 78153 Le Chesnay, France.

J.Ch. Gilbert (1986,b). Une méthode de quasi-Newton réduite en optimisation sous contraintes avec priorité à la restauration. 7th International Conference on Analysis and Optimization of Systems. Lecture Notes in Control and Information Sciences 83, 40-53. A. Bensoussan, J.L. Lions (eds.). Springer.

J.Ch. Gilbert (1986,c). On the local and global convergence of a reduced quasi-Newton method. Rapport de recherche de l'INRIA RR-565. 78153 Le Chesnay, France. (revised version in IIASA Working Papers, WP-87-113, 2361 Laxenburg, Austria)

J. Ch. Gilbert (1987). Maintaining the positive definiteness of the matrices in reduced secant methods for equality constrained optimization. IIASA working Papers, WP-87-123. 2361 Laxenburg, Austria.

J. Ch. Gilbert (1988). Mise à jour de la métrique dans les méthodes de quasi-Newton réduites en optimisation avec contraintes d'égalité. Mathematical Modelling and Numerical Analysis 22/2 (to appear).

S.P. Han, O.L. Mangasarian (1979). Exact penalty functions in nonlinear programming. Mathematical Programming 17, 251-269.

C. Lemaréchal (1981). A view of line-searches. Optimization and optimal control. Lecture Notes in Control and Information Sciences 30, 59-78. A. Auslender, W. Oettli, J. Stoer (eds.). Springer.

M.J.D. Powell (1986). Methods for nonlinear constraints in optimization calculations. Presented at the 1986 IMA/SIAM Meeting on "The State of the Art in Numerical Analysis".

J. Stoer (1984). Principles of sequential quadratic programming methods for solving nonlinear programs. Proceedings of the NATO ASI on Computational Mathematical Programming, Bad Windsheim, Germany.

P. Wolfe (1969). Convergence conditions for ascent methods. SIAM Review 11/2, 226-235.