SIAM J. OPTIMIZATION
Vol. 5, No. 4, pp. 000–000, November 1995

SEQUENTIAL QUADRATIC PROGRAMMING WITH PENALIZATION OF THE DISPLACEMENT*

J. F. BONNANS† AND G. LAUNAY†

Abstract. In this paper we study the convergence of a sequential quadratic programming algorithm for the nonlinear programming problem. The Hessian of the quadratic program is the sum of an approximation of the Lagrangian and of a multiple of the identity that allows us to penalize the displacement. Assuming only that the direction is a stationary point of the current quadratic program we study the local convergence properties without strict complementarity. In particular, we use a very weak condition on the approximation of the Hessian to the Lagrangian. We obtain some global and superlinearly convergent algorithm under weak hypotheses. As a particular case we formulate an extension of Newton's method that is quadratically convergent to a point satisfying a strong sufficient second order condition.

Key words. nonlinear programming, Newton's method, quasi-Newton algorithms, exact penalization, trust region

AMS subject classifications. 90C30, 49M37, 65K05

1. Introduction.

1.1. The family of Newton-type algorithms. In this paper we present a new algorithm for solving the standard nonlinear programming problem

$$(P) \qquad \min f(x)\ ;\ g(x) \ll 0,$$

with f, g smooth mapping from $\mathbb{R}^n$ to $\mathbb{R}$ and $\mathbb{R}^p$, and, given a partition (I, J) of $\{1, \dots, p\}$, by $z \ll 0$ we mean $z_i \leq 0$, $i \in I$, $z_j = 0$, $j \in J$. Occasionally for $K \subset I$ we will denote

$$z \overset{K}{\ll} 0 \Leftrightarrow \begin{cases} z_i \leq 0, & i \in K, \\ z_j = 0, & j \in J. \end{cases}$$

With (P) is associated the first-order optimality system

$$(1) \qquad \begin{aligned} &\nabla f(x) + g'(x)^t \lambda = 0, \\ &g(x) \ll 0,\ \lambda_I \geq 0,\ \lambda^t g(x) = 0. \end{aligned}$$

If (x, λ) satisfies (1), then we say that λ is a multiplier associated to x. By extension we say that x is solution of (1) if there exists λ such that (x, λ) satisfies (1).

We define the quadratic problem

$$Q(x, M) \qquad \min_d \nabla f(x)^t d + \frac{1}{2} d^t M d\ ;\ g(x) + g'(x) d \ll 0,$$

with which is associated the optimality system

$$(2) \qquad \begin{aligned} &\nabla f(x) + Md + g'(x)^t \mu = 0, \\ &g(x) + g'(x)d \ll 0,\ \mu_I \geq 0,\ \mu^t (g(x) + g'(x)d) = 0. \end{aligned}$$

*Received by the editors November 3, 1992; accepted for publication(in revised form) March 13, 1994.

†Institut National de Recherche en Informatique et en Automatique, Domaine de Voluceau, 78153 Rocquencourt, France (Frederic.Bonnans@inria.fr).

Denote by $L(x,\lambda) := f(x) + \lambda^t g(x)$ the Lagrangian associated with (P). It has been observed by Wilson [26] that, when no inequality is present, the computation of the Newton step in (1) amounts to solve $Q(x, M)$ with $M = \nabla_x^2 L(x,\lambda)$, and that this allows a natural generalization for problems with inequality constraints. In order to deal with the case when second derivatives are not available, a larger class of interest is the following.

ALGORITHM 0 (Newton-type algorithms).
0. Choose $x^0 \in I\!R^n$, M^0 an $n \times n$ symmetric matrix ; $k \leftarrow 0$.
1. Compute (d^k, μ^k) solution of the optimality system of $Q(x^k, M^k)$.
2. Linesearch: choose ρ_k in $[0,1]$.
3. $x^{k+1} \leftarrow x^k + \rho_k d^k$
Choose M^{k+1}.
$k \leftarrow k+1$, go to 1.

1.2. Local study. Let $\bar{x}$ be a local solution of (P) with which is associated a unique Lagrange multiplier $\bar{\lambda}$. The local analysis typically assumes that (x^0, M^0) is close to $(\bar{x}, \nabla^2 L(\bar{x}, \bar{\lambda}))$ and that $\rho_k = 1$. The question is to determine if convergence occurs, and at which rate. It happens that in this case d^k should not, in general, be taken as the global minimum of $Q(x^k, M^k)$.

Indeed, let us consider the simple example

$$\min_x \ell n(1+x) \; ; \; -x \leq 0, \; x \leq 10.$$

This problem has a unique solution $\bar{x} = 0$ associated to the unique multiplier $\bar{\lambda} = (1,0)^t$ and the strongest regularity hypothesis and sufficient second order condition (see (8) and (29) below) are satisfied by $(\bar{x}, \bar{\lambda})$. Now let us start Newton's method at the solution. We get the quadratic problem

$$\min_d d - d^2/2 \; ; \; 0 \leq d \leq 10,$$

whose unique solution is $d = 10$, the worst possible displacement! As the Newton step is obtained by linearizing the data, is it clear that the quadratic program is meaningful only if the displacement is not too large. Indeed, in our example, the "good" displacement $d = 0$ is a local solution of the quadratic program.

Of course if $M^k \geq 0$, which is the case for some quasi-Newton algorithms based on positive definite updates, and also for Newton's method when (P) is convex, i.e., has convex cost and inequality constraints and linear equality constraints, then $Q(x^k, M^k)$ is itself convex, and local and global minima coincide. We now quote some recent results about the speed of convergence of Newton-type algorithms. For this purpose, we need to define the set of active inequality constraints:

$$I(x) := \{i \in I \; ; \; g_i(x) = 0\},$$

the set of active constraints

$$I(x) \cup J,$$

the extended critical cone

$$C(x) := \{d \in I\!R^n \; ; \; g'(x)d \overset{I(\bar{x})}{\ll} 0 \; ; \; g_i'(x)d = 0 \text{ if } \bar{\lambda}_i > 0, \; i \in I\}. \tag{3}$$

Note that when $x = \bar{x}$ we recover the usual critical cone, or cone of critical directions:

$$C(\bar{x}) := \{d \in I\!R^n \ ; \ g'(\bar{x})d \overset{I(\bar{x})}{\ll} 0 \ ; \ g_i'(\bar{x})d = 0 \text{ if } \bar{\lambda}_i > 0, \ i \in I\}. \tag{4}$$

We also define the (standard) second-order sufficient condition

$$d^t \nabla_x^2 L(\bar{x}, \bar{\lambda}) d > 0 \quad \text{for all } d \in C(\bar{x}), d \neq 0, \tag{5}$$

and the orthogonal projection onto $C(x^k)$, denoted by P^k.

Note that usually the critical cone is defined as

$$C(\bar{x}) := \{d \in I\!R^n \ ; \ \nabla f(\bar{x})^t d \leq 0 \ ; \ g'(\bar{x})d \overset{I(\bar{x})}{\ll} 0\}.$$

Both definitions coincide (as is easy to check using (2)) because we assume the existence of a Lagrange multiplier. We now quote two results of Bonnans [8].

THEOREM 1.1. *Let $\bar{x}$ be a local solution of* (1) *such that the gradients of active constraints are linearly independent, $\bar{\lambda}$ be the unique multiplier associated with $\bar{x}$, and the second-order sufficient condition holds. Then if (x^k, μ^k) computed by Algorithm* 0 *converge to $(\bar{x}, \bar{\lambda})$, then $\{x^k\}$ converges superlinearly if and only if (iff)*

$$P^k[(\nabla_x^2 L(\bar{x}, \bar{\lambda}) - M^k)d^k] = o(d^k).$$

THEOREM 1.2. *Assume that $\bar{x}$ is a local solution of* (1), *$\bar{\lambda}$ is the unique Lagrange multiplier associated to $\bar{x}$, and the second-order sufficiency condition holds. Then there exists $\varepsilon > 0$ such that if $\|x^0 - \bar{x}\| + \|\lambda^0 - \bar{\lambda}\| < \varepsilon$, and (x^{k+1}, λ^{k+1}) is chosen so that $\|x^{k+1} - x^k\| + \|\lambda^{k+1} - \lambda^k\| < 2\varepsilon$, then Algorithm* 0 *with $M^k = \nabla_x^2 L(x^k, \lambda^k)$ and $\rho_k = 1$, i.e., Newton's method, is well defined and converges at a quadratic rate to $(\bar{x}, \bar{\lambda})$.*

We note that the existence of a unique multiplier is a qualification hypothesis slightly weaker than the linear independence of gradients of active constraints (see Fletcher [14]). Note also that if the following strict complementarity hypothesis holds:

$$\bar{\lambda}_i > 0 \quad \text{for all } i \text{ in } I(\bar{x}),$$

then, for (x^k, M^k) close to $(\bar{x}, \nabla^2 L(\bar{x}, \bar{\lambda}))$, λ^k is close to $\bar{\lambda}$; hence if $i \in I(\bar{x})$, the corresponding inequality in $Q(x^k, M^k)$ is active and everything goes as if we were analyzing the problem

$$\min f(x) \ ; \ g_i(x) = 0, \ i \in I(\bar{x}) \cup J.$$

Then Theorem 1.1 reduces to a result of Boggs, Tolle, and Wang [6], whereas Theorem 1.2 reduces to the application of the general result on quadratic convergence of Newton's method for a system of equations. The novelty in the theorems above lies in the fact that no strict complementarity hypothesis holds and only the standard (weak) sufficient condition is assumed.

1.3. Globalization. The local results that we just presented insure a superlinear or quadratic convergence, provided that the data at the starting point are sufficiently close to the optimum. When these hypotheses are not satisfied, the algorithm must be modified, for different reasons.

(i) It may happen that the optimality system of $Q(x^k, M^k)$ has no feasible solution; a possible remedy is to solve a modified quadratic program. This has been discussed by Fletcher [14]. We will not address this point.

(ii) The point $x^k + d^k$ may be farther from any local solution than x^k. For this reason it is safe to introduce a linesearch on some potential function; the most popular potential is the so-called exact penalty function (see Eremin [13], Zangwill [28], Han [16], Pschenichny and Danilin [22])

$$\theta_r(x) := f(x) + r\|g(x)^\sharp\|$$

with $r > 0$ (the penalty parameter) and $.^\sharp$ defined as follows:

$$z_i^\sharp = \begin{cases} z_i^+ & \text{if} \quad i \in I, \\ z_i & \text{if} \quad i \in J. \end{cases}$$

Here $\|.\|$ stands for an arbitrary norm in $I\!R^p$, although we note that most often the ℓ^1 norm is chosen for practical reasons. The dual norm $\|.\|_*$ is defined as

$$\|\mu\|_* := \max\{z^t\mu \ ; \ \|z\| \le 1\}.$$

Usually r is chosen so that $r > \|\mu^k\|_*$, where μ^k is the multiplier associated to d^k.

However this potential suffers from the Maratos effect (Maratos [19], Mayne and Polak [20]). Even when x is close to $\bar{x}$ and $x + d - \bar{x} = O(x - \bar{x})^2$, and r close to $\|\bar{\lambda}\|_*$, it may happen that $\theta_r(x + d) > \theta_r(x)$, and in the context of composite optimization it has been shown that this may occur an infinite number of times. See Yuan [27].

Various remedies have been proposed, the first of them being to make an additional restoration step (Mayne and Polak [20], Gabay [15]), i.e., denoting $\|.\|$ an arbitrary norm in $I\!R^n$, different from the one in $I\!R^p$, to compute v^k solution of

$$\min_v \|v\| \ ; g_i(x^k + d^k) + g'(x^k)v = 0, \ i \in I_k^*,$$

where I_k^* is some prediction of the set of active constraints, obtained as a byproduct of the computation of d^k, and to perform a linesearch along the arc

$$\rho \to x^k + \rho d^k + \rho^2 v^k.$$

Other possible remedies are to modify the potential, specifically to use a nondifferentiable augmented Lagrangian [7], and to compare the value of $\theta_r(x^{k+1})$ to the value of θ_r not only at x^k, but also at x^{k-1}, $x^{k-2}, \ldots$ (see, Chamberlain et al. [12], Panier and Tits [21], Bonnans et al. [9]). To our knowledge, all published papers concerning the Maratos effect assume that the strict complementarity hypothesis holds.

1.4. Our contribution. In this paper we present an algorithm that has global and local properties under weak hypotheses on the sequence $\{M^k\}$ of approximations of the Hessian of the Lagrangian. At step k of the algorithm, a parameter $\alpha_k \ge 0$ is set and a direction d^k is computed as a stationary point (if any) of the quadratic problem

$$Q_{\alpha_k}(x^k, M^k) \qquad \min_d \nabla f(x^k)^t d + \frac{1}{2} d^t M^k d + \frac{\alpha_k}{2}\|d\|_2^2 \ ; \ g(x^k) + g'(x^k)d \ll 0,$$

where $\|.\|_2$ is the Euclidean norm. This technique was first introduced by Bell [2].

We note for future reference that the first order optimality system of $Q_{\alpha_k}(x^k, M^k)$ is, denoting by μ^k the Lagrange multiplier,

$$\begin{aligned} &\nabla f(x^k) + M^k d^k + \alpha_k d^k + g'(x^k)^t \mu^k = 0, \\ &g(x^k) + g'(x^k)d^k \ll 0; \quad \mu_I^k \geq 0; \quad (\mu^k)^t(g(x^k) + g'(x^k)d^k) = 0. \end{aligned} \tag{6}$$

The parameter $r = r_k$ of the exact penalty function $\theta_r(x)$ is adapted at each iteration in order to allow a linesearch; however, null steps may happen and in this case α_k is increased. We prove that $\{r_k\}$ and $\{\alpha_k\}$ are bounded and that any limit point of $\{x^k\}$ satisfies (1). Our hypotheses are as follows. First we assume

$$\{M^k\}, \{x^k\}, \text{ and } \{d^k\} \text{ are bounded.} \tag{7}$$

Note that if upper and lower bounds on x are present, then $\{x^k\}$ and $\{d^k\}$ are necessarily bounded. Second, we assume that

$$\text{the linearized constraints } g(x) + g'(x)d \ll 0 \text{ are feasible and qualified,} \tag{8}$$

which means that for any x, (8) is satisfied for at least one d, and for all (x, d) such that $g(x) + g'(x)d \ll 0$, the gradients of active constraints of this system are linearly independent.

Hypothesis (8) may seem excessively strong. If a nonlinear optimization problem is solved with a random starting point, it might not be satisfied in the neighborhood of the starting point. We have in mind large-scale real-world applications where, in order to solve the problem in a reasonable time, the initial point is the result of some heuristics so that in the region in which the sequence $\{x^k\}$ lies, (8) is satisfied, although a linesearch may be useful. This is, in particular, the case in the optimal load flow problem (see [5]).

We show also how to avoid the Maratos effect using a second order correction; there we use a very weak hypothesis on the approximation of the Hessian of the Lagrangian. We show also how to combine this result with Theorem 1.1 in order to obtain a superlinearly convergent algorithm.

If second-order derivatives are available we show how to formulate a globally convergent algorithm that reduces locally to Newton's method, and this seems to be the first globally convergent extension of Newton's method for nonconvex constrained optimization. Other globally convergent algorithms have been published, e.g., Han [16] and Fletcher [14], but they assume the approximation to the Hessian to be bounded. The difficulty is that there is no a priori bound for the estimate of the multiplier. We give a device that overcomes this difficulty. We note that Bell [2] has a global convergence result comparable to ours, but he assumes the penalization coefficient r_k to be fixed. By contrast, we deal with the more difficult question of adapting this parameter.

It may seem surprising that the algorithm includes a penalization of the displacement as well as a linesearch; this is due to the presence of constraints. For fixed x, when $\alpha \to \infty$, d solution of $Q_\alpha(x, M)$ converges to $\pi(x)$ solution of

$$\min_d \|d\|_2 \; ; \; g(x) + g'(x)d \ll 0,$$

and (if $\pi(x)$ is nonzero) it may happen that $f(x + \pi(x)) > f(x)$ and $\|g(x + \pi(x))^\sharp\| > \|g(x)^\sharp\|$; in this case the step $\rho_k = 1$ cannot be accepted whenever α_k is large enough.

2. A globally convergent algorithm with fixed penalty parameter. In this section we will present some properties of the exact penalty function that allow the design of a linesearch that extends the one due to Armijo [1] for unconstrained minimization. The ideas that we present here are classical (see [16]) and this section must be considered mainly as a way to prepare the more sophisticated algorithms of §§3 and 4. We note however two specific features. The first is that our hypothesis on the norm is as follows:

$$z \to \|z^\sharp\| \quad \text{is a convex mapping.} \tag{9}$$

This hypothesis is easy to check for the ℓ_p norms, $1 \leq p \leq \infty$, and in that case $\|.^\sharp\|$ coincides with the distance to the cone generating the partial order $x \ll y$. (The property is not true for all norms, e.g., in $I\!R^2$ consider $\|x\| = |x_1| + |x_2 - x_1|$. If $J = \emptyset$ then $\|x^\sharp\| = \|x^+\| = x_1^+ + |x_2^+ - x_1^+|$. We compute $\|(1,0)^+\| = 2 > \frac{1}{2}\|(1,-1)^+\| + \frac{1}{2}\|(1,1)^+\| = \frac{3}{2}$).

The second hypothesis is the choice of directions of sufficient descent. For this we use relation (10) below.

We define the directional derivative of θ_r at x in direction d as $\theta_r'(x,d)$. This is well defined, even if (9) does not hold, because $\rho \mapsto g(x+\rho d)^\sharp$ has a directional derivative $w(x,d)$ (that can easily be computed explicitly) and $z \to \|z\|$ is convex and Lipschitz, hence

$$\begin{aligned} \theta_r(x+\rho d) &= f(x) + \rho f'(x)d + r\|g(x)^\sharp + \rho w(x,d)\| + o(\rho) \\ &= \theta_r(x) + \rho[f'(x)d + r\mu^t w(x,d)] + o(\rho), \end{aligned}$$

where μ is some element of the subdifferential of $\|.\|$ at $g(x)^\sharp$.

We define the "linearized" (at point x^k) exact penalty function as follows:

$$\theta^k(d) = f(x^k) + f'(x^k)d + r_k\|(g(x^k) + g'(x^k)d)^\sharp\|.$$

For any d feasible for $Q_{\alpha_k}(x^k, M^k)$, we note that the decrease of the linearized exact penalty function when step $\rho_k = 1$ is accepted is equal to $\Delta_{r_k}(x^k, d)$, where

$$\Delta_r(x,d) := r\|g(x)^\sharp\| - f'(x)d.$$

We say that $\Delta_r(x,d)$ is feasible if

$$\Delta_r(x,d) \geq \|d\|^3. \tag{10}$$

By Δ_k we denote $\Delta_{r_k}(x^k, d^k)$.

LEMMA 2.1. *Let d be a stationary point of $Q_\alpha(x,M)$ and μ the associated Lagrange multiplier. Then*

(i) *if (9) holds, then*

$$\theta_r'(x,d) \leq -\Delta_r(x,d). \tag{11}$$

(ii) *The following relations hold:*

$$\Delta_r(x,d) \geq (r - \|\mu\|_*)\|g(x)^\sharp\| + \alpha\|d\|_2^2 + d^t M d + \mu^t(g(x)^\sharp - g(x)), \tag{12}$$

$$\Delta_r(x,d) \geq (r - \|\mu\|_*)\|g(x)^\sharp\| + \alpha\|d\|_2^2 + d^t M d. \tag{13}$$

Proof. (i) From (9) we deduce that

$$\theta'_r(x,d) = f'(x)d + r\eta^t g'(x)d,$$

where η is some subgradient of $\|.^\sharp\|$ at $g(x)$, i.e.,

$$\|z^\sharp\| \geq \|g(x)^\sharp\| + \eta^t(z - g(x)) \quad \forall z \in I\!R^p.$$

Choosing $z = g(x) + g'(x)d$, and noting that $z^\sharp = 0$, we deduce that $\eta^t g'(x)d \leq -\|g(x)^\sharp\|$, from which (11) follows.

(ii) From (6) we deduce

$$0 = f'(x)d + d^t M d + \alpha\|d\|_2^2 + \mu^t g'(x)d.$$

¿From the complementarity condition we get that $\mu^t g'(x)d = -\mu^t g(x)$, hence

$$-f'(x)d = d^t M d + \alpha\|d\|_2^2 - \mu^t g(x),$$

and so

$$\begin{aligned}\Delta_r(x,d) &= \alpha\|d\|_2^2 + d^t M d + r\|g(x)^\sharp\| - \mu^t g(x),\\ &= \alpha\|d\|_2^2 + d^t M d + r\|g(x)^\sharp\| - \mu^t g(x)^\sharp + \mu^t(g(x)^\sharp - g(x))\\ &\geq \alpha\|d\|_2^2 + d^t M d + (r - \|\mu\|_*)\|g(x)^\sharp\| + \mu^t(g(x)^\sharp - g(x)).\end{aligned}$$

Thus (12) is proved. Now, as $\mu_I \geq 0$, we get from the definition of $g(x)^\sharp$ that $\mu^t(g(x)^\sharp - g(x)) \geq 0$, and so (13) holds. □

Let x^k be the current point of the algorithm and d^k a stationary point of $Q_{\alpha_k}(x^k, M^k)$. From (13) it follows that, at least if $r_k > \|\mu^k\|_*$ and α_k is large enough, then Δ_k is feasible (note that for α_k sufficiently large, $\|d^k\| \approx \|\pi(x^k)\|$, hence (10) is satisfied).

¿From (11) it follows that d^k is a descent direction of θ_{r_k} if $\Delta_k > 0$. This allows us to define a linesearch in the following way.

Linesearch rule. LS1. Parameters $\gamma \in (0, 1/2), \beta \in (0,1)$. If Δ_k is feasible then compute $\rho_k = (\beta)^\ell$, with ℓ smallest integer such that

$$\begin{aligned}&\theta_{r_k}(x^k + (\beta)^\ell d^k) \leq \theta_{r_k}(x^k) - (\beta)^\ell \gamma \Delta_k,\\ &x^{k+1} \leftarrow x^k + \rho_k d^k.\end{aligned} \tag{14}$$

We note that (11) and the relation $\gamma < \frac{1}{2}$ imply that (14) is satisfied for ℓ large enough. Hence the linesearch is well defined. In order to analyse the global properties associated with this linesearch we deal in this section with the simple case when r_k is equal to some constant r.

We can now formulate a conceptual algorithm.

ALGORITHM 1

0. Data: $\alpha_0 \geq 0$, M^0 an $n \times n$ symmetric matrix, $x^0 \in I\!R^n$; $k \rightarrow 0$.
1. Computation of (d^k, μ^k) satisfying the optimality system of $Q_{\alpha_k}(x^k, M^k)$.
2. If Δ_k is not feasible, i.e., (10) not satisfied for Δ_k, stop.
3. Perform the linesearch LS1.
4. Choose α_{k+1} and M^{k+1};
   $k \rightarrow k+1$,
   go to 1.

THEOREM 2.1. *Assume that* (7) *and* (8) *hold. Let* x^k *be computed by Algorithm* 1 *in which* Δ_k *is assumed to be feasible at each step. Assume that* (α_k, M^k, d^k) *are bounded,* $r_k = r > 0$. *Then* $d^k \to 0$ *and the set of limit points of* (x^k, μ^k) *is a connected subset of the set of solutions of the first-order optimality system* (1).

Proof. We prove that $d^k \to 0$. We note that $\theta_r(x^k)$ decreases, hence converges, so that by (14) $\rho_k \Delta_k \to 0$. Assume that for some subsequence k', we have $(x^{k'}, \alpha_{k'}, M^{k'}, d^{k'}) \to (\hat{x}, \hat{\alpha}, \hat{M}, \hat{d})$ with $\hat{d} \neq 0$. We observe that $\Delta_{k'} \to \hat{\Delta} := \Delta_r(\hat{x}, \hat{d}) > 0$ by (10) and that $\hat{d}$ satisfies the first-order optimality system of $Q_{\hat{\alpha}}(\hat{x}, \hat{M})$; hence $\theta_r'(\hat{x}, \hat{d}) \le -\hat{\Delta}$ by (11), which implies for ρ small enough

$$\begin{aligned}\theta_r(\hat{x} + \rho \hat{d}) &\le \theta_r(\hat{x}) - \rho\hat{\Delta} + o(\rho) \\ &\le \theta_r(\hat{x}) - \frac{2\rho}{3}\hat{\Delta},\end{aligned}$$

hence for k' large enough by continuity (as $\Delta_{k'} \to \hat{\Delta} > 0$)

$$\theta_r(x^{k'} + \rho d^{k'}) \le \theta_r(x^{k'}) - \frac{\rho}{2}\Delta_r(x^{k'}, d^{k'}),$$

which proves that $\rho_{k'}$ cannot converge to 0, hence we get $\hat{\Delta} = \lim \Delta_{k'} = 0$, from $\rho_k \Delta_k \to 0$, contradicting $\hat{\Delta} > 0$ obtained from our assumption $\hat{d} \neq 0$.

Now as $d^k \to 0$ for any converging subsequence of $(x^k, \alpha_k, M^k, d^k)$, we can pass to the limit in (6), deducing the boundedness of $\{\mu^k\}$ from (7) and (8), and so that any limit point of (x^k, μ^k) is solution of (1). Now as $d^k \to 0$, the set of limit points of $\{x^k\}$ is connected; by (8) the Lagrange multiplier of (1) (whenever it exists) must depend continuously on x; the conclusion follows. □

In the next section we relax the restrictive hypothesis on r^k and on the a priori feasibility of Δ_k.

3. A general globally convergent algorithm. This section is devoted to the statement and analysis of a globally convergent algorithm, more precisely an algorithm computing a sequence $\{x^k, \mu^k\}$ such that any of its limit-points satisfy the first-order optimality conditions (1). In this algorithm we must update the two parameters r_k and α_k.

For r_k the idea is the following: take $r_k = r_{k-1}$ whenever it is possible, i.e., if $\Delta_{r_{k-1}}(x^k, d^k)$ is feasible and $\rho_k = 1$ is accepted by the linesearch; otherwise choose r_k satisfying $r_k > \|\mu^k\|_*$. In order to make the sequence r_k constant after a finite number of steps we choose $r_k = \max(r_{k-1}, \mathrm{int}(\|\mu^k\|_* + 2))$. Finally the update rule for r_k is as follows:

$$r_k = \begin{cases} r_{k-1} \text{ if } \Delta_{r_{k-1}}(x^k, d^k) \text{ is feasible and} \\ \theta_{r_{k-1}}(x^k + d^k) \le \theta_{r_{k-1}}(x^k) - \gamma \Delta_{r_{k-1}}(x^k, d^k), \\ \max(r_{k-1}, \mathrm{int}(\|\mu^k\|_* + 2)) \text{ if not.} \end{cases} \tag{15}$$

For α_k the idea is the following. If Δ_k is not feasible or ρ_k is close to 0, then choose $\alpha_{k+1} > \alpha_k + \varepsilon_1$, with $\varepsilon_1 > 0$ (because of Lemma 2.1 this will eventually yield the feasibility of Δ_k). On the other hand, if Δ_k is feasible and $\rho_k = 1$, then α_{k+1} will be taken smaller than α_k.

Finally we mention the possibility of null steps, i.e., when Δ_k is not feasible then x^{k+1} is taken equal to x^k (or equivalently $\rho_k = 0$) and α_k is increased. We now state the algorithm.

ALGORITHM 2

0. Data: $\alpha_0 \geq 0$, M^0 $n \times n$ symmetric matrix, $x^0 \in \mathbb{R}^n$. Parameters $0 < \varepsilon_1 < \varepsilon_2$, $0 < \varepsilon_3 < 1$; $k \leftarrow 0$.
1. Computation of (d^k, μ^k), satisfying the optimality system of $Q_{\alpha_k}(x^k, M^k)$.
2. If $k = 0$, set $r_{-1} \leftarrow \|\mu^0\|_* + 1$.
3. Choice of r_k using the rule (15).
4. If Δ_k is not feasible (null step):
   $\rho_k \leftarrow 0$,
   $x^{k+1} \leftarrow x^k$,
   go to 6.
5. If Δ_k is feasible: perform the linesearch LS1.
6. Update of α_k:
   If $\rho_k = 1$, choose $\alpha_{k+1} \leq \alpha_k/2$.
   If $\rho_k \in (\varepsilon_3, 1)$, choose $\alpha_{k+1} \leq \alpha_k + \varepsilon_2$.
   If $\rho_k \leq \varepsilon_3$ choose $\alpha_{k+1} \in [\alpha_k + \varepsilon_1, \alpha_k + \varepsilon_2]$.
   Choose M^{k+1}.
7. $k \leftarrow k + 1$,
   go to 1.

Remark 3.1. We observe that $\{r_k\}$ increases, and $\{r_k\}$ is bounded iff there exists $r > 0$ such that $r_k = r$ for $k \geq k_0$.

THEOREM 3.1. *Let x^k be computed by Algorithm* 2. *We assume that* (7) *and* (8) *hold. Then* (i) *the sequences $\{r_k\}$, $\{\alpha_k\}$, and $\{\mu^k\}$ are bounded;*

(ii) *the set of limit-points of $\{x^k\}$ is connected, and with each limit point is associated a Lagrange multiplier.*

We give a proof that makes use of some lemmas below.

Proof. (a) We prove that $\{r_k\}$ is bounded. If not, then there exists a subsequence k' with $r_{k'} > r_{k'-1}$, and by (15) $\|\mu^{k'}\|_* \to \infty$. This, and (6)–(8) imply that $\alpha_{k'}\|d^{k'}\| \to \infty$. Now by Lemma 3.1, we obtain $\|g(x^k)^\sharp\| \to 0$ and Lemma 3.2 ensures that for k' large enough, $r_{k'} = r_{k'-1}$, contrary to the definition of $\{k'\}$.

(b) We prove that $\{\alpha_k\}$ is bounded. As $\{r_k\}$ is bounded, we know from Remark 3.1 that r_k is constant, say equal to r for $k \geq k_0$. Lemma 3.3 says that there exists $\hat{\alpha} \geq 0$ such that Δ_k is feasible if $\alpha_k \geq \hat{\alpha}$ and $k \geq k_0$.

¿From step 6 of Algorithm 2, it follows that $\alpha_{k+1} \leq \alpha_k + \varepsilon_2$ for all k. By Lemma 3.3, if $\alpha_k \geq \hat{\alpha}$ and $k \geq k_0$, then $\rho_k = 1$ and $\alpha_{k+1} \leq \alpha_k/2$; hence $\alpha_{k+1} \leq \max(\hat{\alpha}, \alpha_{k_0}/2) + \varepsilon_2$ whenever $k \geq k_0$.

(c) We now prove (ii). Let $\hat{\alpha}$ be given by Lemma 3.3. By step 6 of Algorithm 2, after at most $\hat{\alpha}/\varepsilon_1$ successive null steps, one has $\alpha_k \geq \hat{\alpha}$; by Lemma 3.3 the next step is not a null step. This means that $\mathbb{K} := \{k \in \mathbb{N}; \rho_k > 0\}$ is not finite. The sequence $\{x^k\}_{k\in\mathbb{K}}$, can be viewed as generated by Algorithm 1, and we deduce from Theorem 2.1 that $\{d^k\}_{k\in\mathbb{K}} \to 0$ and that with each limit-point of $\{x^k\}$ is associated a Lagrange multiplier. As $\{x^k\}_{k\in\mathbb{N}}$ and $\{x^k\}_{k\in\mathbb{K}}$ obviously have the same limit-points, point (ii) follows. □

We now state and prove the three lemmas used in the proof of Theorem 3.1.

LEMMA 3.1. *Let $\{x^k\}$ be computed by Algorithm* 2. *Under hypotheses* (7) *and* (8), *if $r_k \nearrow \infty$ then $\|g(x^k)^\sharp\| \to 0$.*

Proof. (a) Let us verify that $\|g(x^k)^\sharp\|$ converges. Let $m := \inf\{f(x^k), k \in \mathbb{N}\}$. Note that $m > -\infty$ as $\{x^k\}$ is bounded. Then, as $\{r_k\}$ increases $\theta_{r_k}(x^{k+1}) \leq \theta_{r_k}(x^k)$

and so we deduce

$$\|g(x^{k+1})^\sharp\| + \frac{f(x^{k+1}) - m}{r_k} \le \|g(x^k)^\sharp\| + \frac{f(x^k) - m}{r_k} \le \|g(x^k)^\sharp\| + \frac{f(x^k) - m}{r_{k-1}},$$

hence $\{\|g(x^k)^\sharp\| + (f(x^k) - m)/r_{k-1}\}$ is a decreasing sequence, and so converges since it is bounded. As $r_k \nearrow \infty$ and $\{f(x^k)\}$ is bounded since $\{x^k\}$ is bounded, it follows that $\|g(x^k)^\sharp\|$ converges.

(b) It suffices now to get a contradiction when assuming that $\lim \|g(x^k)^\sharp\|$ is positive. Let us note that by (6)–(8), if $\{\alpha_k\}$ is bounded, so is $\{\mu^k\}$ hence $\{r_k\}$ cannot go to ∞. Hence we may extract a subsequence k' such that $\alpha_{k'} \to \infty$ and $x^{k'} \to \bar{x}$. It is easily checked that $d^{k'} \to \bar{d} := \pi(\bar{x})$. Now since $\|.^\sharp\|$ is a Lipschitz mapping:

$$\begin{aligned}\theta_{r_{k'}}(x^{k'}) - \theta_{r_{k'}}(x^{k'} + \rho d^{k'}) &= r_{k'}\|g(x^{k'})^\sharp\| - r_{k'}\|g(x^{k'} + \rho d^{k'})^\sharp\| + O(1)\\ &= r_{k'}\|g(\bar{x})^\sharp\| - r_{k'}\|g(\bar{x} + \rho\bar{d})^\sharp\| + o(r_{k'}),\end{aligned}$$

with $o(r_{k'})/r_{k'} \to 0$ uniformly on $\rho \in [0,1]$.

As $g(x^{k'}) + g'(x^{k'})d^k \ll 0$ and (7) holds, it follows that

$$\|g(\bar{x} + \rho\bar{d})^\sharp\| \le (1-\rho)\|g(\bar{x})^\sharp\| + a_0\rho^2 \quad \text{for some } a_0 > 0,$$

hence since $\|.^\sharp\|$ is a Lipschitz mapping and (7) holds:

$$\begin{aligned}\theta_{r_{k'}}(x^{k'}) - \theta_{r_{k'}}(x^{k'} + \rho d^{k'}) &\ge \rho r_{k'}\|g(\bar{x})^\sharp\| - r_{k'}a_0\rho^2 + o(r_{k'}),\\ &\ge \rho r_{k'}\|g(x^{k'})^\sharp\| - r_{k'}a_0\rho^2 + o(r_{k'}),\\ &= \rho\Delta_{k'} - r_{k'}a_0\rho^2 + o(r_{k'}).\end{aligned}$$

We note that $\Delta_{k'}/r_{k'} \to \|g(\bar{x})^\sharp\|$ which is assumed to be positive. Using this we get for some $a_1 > 0$

$$\theta_{r_{k'}}(x^{k'}) - \theta_{r_{k'}}(x^{k'} + \rho d^{k'}) \ge \Delta_{k'}[\rho - a_1\rho^2 + o(1)]$$

and it follows that $\rho_{k'} \ge \hat{\rho}$ for some $\hat{\rho} > 0$. Then this implies that for some $a_2 > 0$

$$\overline{\lim}\|g(x^{k'+1})^\sharp\|/\|g(x^{k'})^\sharp\| \le 1 - a_2\hat{\rho},$$

in contradiction with our hypothesis. □

LEMMA 3.2. *Let x^k be computed by Algorithm 2. Under the hypotheses (7) and (8), if a subsequence $\{x^{k'}\}$ satisfies $\|g(x^{k'})^\sharp\| \to 0$ and $\alpha_{k'}\|d^{k'}\| \to \infty$, then*

(i) $\|d^{k'}\|_2/\|\pi(x^{k'})\|_2 \to 1$.

(ii) *For k' large enough, $r_{k'} = r_{k'-1}$.*

Proof. Denote by

$$q_k(d) := \nabla f(x^k)^t d + \frac{1}{2}d^t M^k d + \frac{1}{2}\alpha_k d^t d,$$

the cost function of $Q_{\alpha_k}(x^k, M^k)$. As $\|d^k\|$ is bounded it follows from the unboundedness of $\alpha_{k'}\|d^{k'}\|$ that $\alpha_{k'} \to \infty$. So we see that for $k' \ge k_0'$, $q_{k'}(d)$ is convex, hence $d^{k'}$ is a global solution of $Q_{\alpha_{k'}}(x^{k'}, M^{k'})$. In particular, denoting $\pi^k := \pi(x^k)$, we have

$$q_{k'}(d^{k'}) \le q_{k'}(\pi^{k'}). \tag{16}$$

From the definition of π^k we have $\|\pi^k\|_2 \leq \|d^k\|_2$. On the other hand, dividing (16) by $\alpha_{k'}\|d^{k'}\|_2^2$, remembering that $\alpha_{k'} \to \infty$ we obtain $1 \leq \underline{\lim}\|\pi^{k'}\|_2/\|d^{k'}\|_2$ and point (i) follows.

We now prove (ii). We may assume that $r_k \nearrow \infty$, for otherwise r_k is constant for k large enough (see Remark 3.1) and then the conclusion holds trivially. The idea of the proof is that the penalization term dominates in the linesearch. Indeed,

$$\begin{aligned}\Delta_{r_{k'-1}}(x^{k'}, d^{k'}) &= r_{k'-1}\|g(x^{k'})^\sharp\| - f'(x^{k'})d^{k'}, \\ &= r_{k'-1}\|g(x^{k'})^\sharp\| \left(1 - \frac{1}{r_{k'-1}} \frac{f'(x^{k'})d^{k'}}{\|\pi^{k'}\|_2} . \frac{\|\pi^{k'}\|_2}{\|g(x^{k'})^\sharp\|}\right).\end{aligned}$$

We claim that the term between parentheses converges to 1. By point (i), $f'(x^{k'})d^{k'}/\|\pi^{k'}\|_2$ is bounded. As $r_{k-1} \nearrow \infty$ it suffices to prove that $\|\pi^{k'}\|_2/\|g(x^{k'})^\sharp\|$ is bounded. If this is not the case, extracting if necessary a subsequence we may assume that $x^{k'} \to \hat{x}$. As $\|g(x^{k'})^\sharp\| \to 0$, $\hat{x}$ is feasible.

Let $D(x)$ be the set $\{d \in I\!R^n \ ; \ g(x) + g'(x)d \ll 0\}$. As (8) holds we may apply to the feasible sets of $Q_\alpha(x, M)$ a theorem of Robinson [23] that asserts that for x in a neighborhood of $\bar{x}$, $d = 0$ is at a distance of $D(x)$ of order $\|g(x)^\sharp\|$. It follows that the element of minimum norm $\pi(x)$ satisfies $\|\pi(x)\| = 0(\|g(x)^\sharp\|)$, and this proves our claim.

Now let us prove that if $\Delta_{r_{k'-1}}(x^{k'}, d^{k'})$ is feasible by (i) and the boundedness of $\|\pi^{k'}\|_2/\|g(x^{k'})^\sharp\|$ proved above it follows that $\|d^{k'}\|/\|g(x^{k'})^\sharp\|$ is bounded. Using (7) and $r_{k'-1} \nearrow \infty$ we deduce that $r_{k'-1}\|g(x^{k'})^\sharp\|/\|d^{k'}\|^3 \to \infty$. This and our claim above imply (10), i.e., feasibility of $\Delta_{r_{k'-1}}(x^{k'}, d^{k'})$.

On the other hand

$$\theta_{r_{k'-1}}(x^{k'}) - \theta_{r_{k'-1}}(x^{k'} + d^{k'}) = r_{k'-1}(\|g(x^{k'})^\sharp\| - \|g(x^{k'} + d^{k'})^\sharp\|) + f(x^{k'}) - f(x^{k'} + d^{k'})$$

and so

(17)
$$\theta_{r_{k'-1}}(x^{k'}) - \theta_{r_{k'-1}}(x^{k'} + d^{k'}) = \Delta_{r_{k'-1}}(x^{k'}, d^{k'}) - r_{k'-1}\|g(x^{k'} + d^{k'})^\sharp\| + O(d^{k'})^2.$$

But $\|.^\sharp\|$ is a Lipschitz mapping, and from (6) $(g(x^{k'}) + g'(x^{k'})d^{k'})^\sharp = 0$, hence $\|g(x^{k'} + d^{k'})^\sharp\| = O(d^{k'})^2$. Also $d^{k'} = O(g(x^{k'})^\sharp)$ hence, with (17),

$$\theta_{r_{k'-1}}(x^{k'}) - \theta_{r_{k'-1}}(x^{k'} + d^{k'}) = \Delta_{r_{k'-1}}(x^{k'}, d^{k'}) + o(\Delta_{r_{k'-1}}(x^{k'}, d^{k'})).$$

As the rule (15) is used in Algorithm 2, the two previous results imply $r_{k'-1} = r_{k'}$ for any $k' \geq k'_0$, in contradiction with the hypothesis $r_{k'} \nearrow \infty$. □

LEMMA 3.3. *Let x^k be computed by Algorithm 2. Under hypotheses* (7) *and* (8), *if $\{r_k\}$ is bounded, then there exists $\hat{\alpha} > 0$ and k_0 such that $\rho_k = 1$ whenever $\alpha_k \geq \hat{\alpha}$ and $k \geq k_0$.*

Proof. Since r_k is bounded, there exists r such that $r_k = r$ for $k \geq k_0$ (cf. Remark 3.1). Using (13) we know that

$$\Delta_k \geq (r - \|\mu^k\|_*)\|g(x^k)^\sharp\| + \alpha_k\|d^k\|_2^2 + d^{kt}M^kd^k,$$

and so, as from (7) $\{M^k\}$ is bounded, we obtain for some $a_3 > 0$

$$\Delta_k \geq (r - \|\mu^k\|_*)\|g(x^k)^\sharp\| + (\alpha_k - a_3)\|d^k\|_2^2,$$

for k large enough. If $\rho_k \neq 1$ then $r_k = r > \|\mu^k\|_*$; hence

$$\Delta_k \geq (\alpha_k - a_3)\|d^k\|_2^2. \tag{18}$$

As $\{d^k\}$ is bounded, we deduce that for α_k large enough, Δ_k is feasible. Now

$$\theta_r(x^k) - \theta_r(x^k+d^k) = r(\|g(x^k)^\sharp\| - \|g(x^k+d^k)^\sharp\|) + f(x^k) - f(x^k+d^k),$$

so since (7) holds and f, g are smooth, we get for some $a_4 > 0$

$$\theta_r(x^k) - \theta_r(x^k+d^k) \geq \Delta_k - a_4\|d^k\|_2^2, \tag{19}$$

hence using (18), for α_k large enough

$$\theta_r(x^k) - \theta_r(x^k+d^k) \geq \frac{1}{2}\Delta_k.$$

As $\gamma < \frac{1}{2}$, the rule (15) ensures that the two previous results imply $\rho_k = 1$, in contradiction with the hypothesis $\rho_k \neq 1$. □

4. A globally and superlinearly convergent algorithm. Let $\bar{x}$ be a local solution of (P) and $\bar{\lambda}$ its associated Lagrange multiplier. We know that Algorithm 2 is not generally superlinearly convergent, even if $x^k \to \bar{x}$ and $M^k \to \nabla_x^2 L(\bar{x}, \bar{\lambda})$. This is due to the Maratos effect (Maratos [19], Mayne and Polak [20]). In this section we show how to adapt the idea of a restoration step in order to accept the unit stepsize. We define

$$I^* := \{i \in I\ ;\ \bar{\lambda}_i > 0\} \cup J,$$
$$I_k^* := \{i \in I\ ;\ \mu_i^k > 0\} \cup J.$$

We first perform a local analysis in which our hypotheses are as follows:

$$\begin{array}{l} \{M^k\},\ \{x^k\},\ \{\alpha_k\},\ \{d^k\} \text{ are given such that } x^k \to \bar{x}, \\ \{M^k\} \text{ and } \{\alpha_k\} \text{ are bounded}, \\ d^k \text{ is stationary point of } Q_{\alpha_k}(x^k, M^k) \text{ and } d^k \to 0. \end{array} \tag{20}$$

We define v^k as the solution of

$$\min_v \|v\| \quad \left\{ \begin{array}{l} g(x^k+d^k) + g'(x^k)v \ll 0, \\ g_i(x^k+d^k) + g_i'(x^k)v = 0 \text{ for any } i \in I_k^*, \end{array} \right. \tag{21}$$

where $\|.\|$ is an arbitrary norm in $\mathbb{R}^n$. Under some reasonable assumptions we show in Proposition 4.1 below that the point $x^k + d^k + v^k$ insures a significant decrease of the exact penalty function. It could be argued that the computation of v^k may be expensive. A possibility ([20], [15]) is to compute v^k solution of

$$\min_v \|v\|\ ;\ g_i(x^k+d^k) + g_i'(x^k)v = 0 \quad \text{for any } i \in I_k^*. \tag{22}$$

If the strict complementarity hypothesis holds, the two corrections are, for k large enough, identical. This indicates that a reasonable way to solve (21) might be to solve (22) first and to check if its solution is also the solution of (21). We start with a technical lemma.

LEMMA 4.1. *Assume that* (8), (20), *and* (21) *hold. Then one has for some* $a > 0$, $k_0 \in \mathbb{N}$

$$I^* \subset I_k^* \text{ for } k > k_0 \ , \tag{23}$$

$$\|v^k\| \leq a\|d^k\|^2, \tag{24}$$

$$g(x^k + d^k + v^k)^\sharp = o(d^k)^2, \tag{25}$$

$$g_{I^*}(x^k + d^k + v^k) = o(d^k)^2. \tag{26}$$

Proof. (a) It follows from (20), (6), and (8) that $\mu^k \to \bar{\lambda}$. So for k large enough, $\{i \in I\ ;\ \bar{\lambda}_i > 0\} \subset \{i \in I\ ;\ \mu_i^k > 0\}$ and thus (23) is proved.

(b) Since (8) holds and by definition of I_k^*:

$$\begin{aligned} &g(x^k) + g'(x^k)d^k \ll 0,\\ &g_i(x^k) + g_i'(x^k)d^k = 0,\ i \in I_k^*, \end{aligned}$$

it follows that

$$\begin{aligned} &g(x^k + d^k) \ll O(d^k)^2,\\ &g_i(x^k + d^k) = O(d^k)^2,\ i \in I_k^*, \end{aligned}$$

hence using again (8), $v^k = O(d^k)^2$.

(c) Expanding $g(x^k + d^k + v^k)$ and using (24) we get

$$g(x^k + d^k + v^k) = g(x^k) + g'(x^k)(d^k + v^k) + \frac{1}{2}(d^k)^t g''(x^k)d^k + o(d^k)^2. \tag{27}$$

Moreover, since (21) implies $(g(x^k + d^k) + g'(x^k)v^k)^\sharp = 0$, expanding $g(x^k + d^k)$ and using $z \to \|z^\sharp\|$ Lipschitz, we obtain

$$\|(g(x^k) + g'(x^k)d^k + \frac{1}{2}(d^k)^t g''(x^k)d^k + g'(x^k)v^k)^\sharp\| = o(d^k)^2.$$

Then, as $z \to \|z^\sharp\|$ is Lipschitz, we have (25).

(d) Since v^k is solution of (21), the expansion of $g_{I_k^*}(x^k + d^k)$ yields

$$g_i(x^k) + g_i'(x^k)d^k + \frac{1}{2}(d^k)^t g_i''(x^k)d^k + g_i'(x^k)v^k = o(d^k)^2 \text{ for any } i \in I_k^*.$$

Hence (26) follows from (23) and (27). □

Then we compute x^{k+1} along the path $\rho \to x^k + \rho d^k + \rho^2 v^k$. The first trial point is $x^k + d^k + v^k$ and if it appears to be necessary to test a small value for ρ_k, then the contribution of v^k is small with respect to the one of d^k, and this allows us to preserve the descent property on θ_r. Specifically the linesearch is as follows.

Linesearch rule LS2. Parameters $\gamma \in (0, 1/2)$, $\beta \in (0, 1)$. Compute v^k solution of (21).

If Δ_k is feasible, i.e., (10) holds for Δ_k, then compute $\rho_k = (\beta)^\ell$ with ℓ smallest integer such that

$$\theta_{r_k}(x^k + (\beta)^\ell d^k + (\beta)^{2\ell} v^k) \le \theta_{r_k}(x^k) - (\beta)^\ell \gamma \Delta_k, \tag{28}$$
$$x^{k+1} \leftarrow x^k + \rho_k d^k + (\rho_k)^2 d^k.$$

In order to perform a local analysis we are led to assume that $(\bar{x}, \bar{\lambda})$ (local solution (P) and associated multiplier) satisfies the following strong second-order sufficient condition (Robinson [24]):

$$\text{for any} \quad d \in \ker g'_{I^*}(\bar{x}) \backslash \{0\}, \quad d^t \nabla_x^2 L(\bar{x}, \bar{\lambda}) d > 0. \tag{29}$$

Recalling (8) we see that (29) is stronger than the standard sufficient condition (5), and that both coincide if the strict complementarity hypothesis holds at $\bar{x}$.

The next proposition insures that the new linesearch rule accepts the step $\rho^k = 1$, if x^k is close to $\bar{x}$ satisfying (28). Define

$$d_T^k \text{ orthogonal projection of } d^k \text{ onto } \ker g'_{I^*}(x^k),$$
$$d_N^k := d^k - d_T^k,$$
$$H := \nabla_x^2 L(\bar{x}, \bar{\lambda}),$$

and for $z \in I\!R^p$, $\tilde{z}$ by

$$\tilde{z}_i = \begin{cases} z_i & \text{if } i \in I^*, \\ z_i^+ & \text{if not.} \end{cases}$$

PROPOSITION 4.1. *Assume* $\{M^k\}$, $\{x^k\}$, $\{\alpha_k\}$, $\{r_k\}$, $\{d^k\}$ *given such that* (8), (20), (21), *and* (29) *hold and* $r_k = r$ *with* $r > \|\bar{\lambda}\|_*$. *If there exists* $\varepsilon_0 > 0$ *such that for* x^k *close enough to* $\bar{x}$,

$$(d_T^k)^t M^k d_T^k + \alpha_k \|d_T^k\|^2 \ge \frac{1}{2(1-\gamma)} (d_T^k)^t H d_T^k + \varepsilon_0 \|d_T^k\|^2 \tag{30}$$

then LS2 *accepts step* $\rho_k = 1$ *for* k *large enough.*

We call (30) the condition of sufficient curvature . A typical condition for the unit step to be accepted is that M^k is close to H, or maybe in some direction only in some sense. Our condition is of a somewhat different nature, as we require the curvature in the tangent direction, i.e., $(d_T^k)^t M^k d_T^k$ to be sufficiently positive. This condition is minimal in the following sense: in the framework of unconstrained optimization, so that $d^k = d_T^k$, then it can be checked that a necessary condition for the unit step to be accepted is (30) in which we change $+\varepsilon_0 \|d_T^k\|^2$ into $-\varepsilon_0 \|d_T^k\|^2$, as shown in Lemma 4.2.

LEMMA 4.2. *Let* $\bar{x}$ *be such that* $\nabla f(\bar{x}) = 0$ *and* $\nabla^2 f(\bar{x}) > 0$. *Let* $x^k \to \bar{x}$ *and* $\{M^k\}$ *be such that* $d^k = -(M^k)^{-1} \nabla f(x^k)$ *vanishes, and*

$$f(x^k + d^k) \le f(x^k) + \gamma f'(x^k) d^k.$$

Then for any $\varepsilon_0 > 0$ *we have for* k *large enough*

$$(d^k)^t M^k d^k \ge \frac{1}{2(1-\gamma)} d^t \nabla^2 f(\bar{x}) d - \varepsilon_0 \|d\|^2.$$

Proof. Choose $\varepsilon_0 > 0$. Set $H^k := \int_0^1 (1-\sigma)\nabla^2 f(x^k + \sigma d^k)d\sigma$. We have

$$f(x^k + d^k) = f(x^k) + f'(x^k)d^k + \frac{1}{2}(d^k)^t H^k d^k$$

with $\|M^k - \nabla^2 f(\bar{x})\| \leq 2(1-\gamma)\varepsilon_0$ for k large enough. It follows that

$$\begin{aligned} 0 &\leq f(x^k) + \gamma f'(x^k)d^k - f(x^k + d^k) \\ &= (\gamma - 1)f'(x^k)d^k - \frac{1}{2}(d^k)^t H^k d^k \\ &= (1-\gamma)\left[(d^k)^t H^k d^k - \frac{1}{2(1-\gamma)}(d^k)^t H^k d^k\right], \end{aligned}$$

so that

$$\begin{aligned} (d^k)^t H^k d^k &\geq \frac{1}{2(1-\gamma)}(d^k)^t H^k d^k \\ &\geq \frac{1}{2(1-\gamma)}(d^k)^t \nabla^2 f(\bar{x}) d^k - \varepsilon_0 \|d^k\|^2, \end{aligned}$$

as was to be proved. □

Before giving the proof we set some preliminary results.

LEMMA 4.3. *For any $n \times n$ symmetric matrix M and for any $\varepsilon > 0$ one has*

$$(d^k)^t M d^k \geq (d_T^k)^t M d_T^k - \varepsilon^2 \|d_T^k\|_2^2 - \|M\|(1 + \|M\|/\varepsilon^2)\|d_N^k\|_2^2, \tag{31}$$

$$(d_T^k)^t M d_T^k \geq (d^k)^t M d^k - \varepsilon^2 \|d_T^k\|_2^2 - \|M\|(1 + \|M\|/\varepsilon^2)\|d_N^k\|_2^2. \tag{32}$$

Proof. Since $d^k = d_T^k + d_N^k$ we get

$$(d^k)^t M d^k = (d_T^k)^t M d_T^k + 2(d_T^k)^t M d_N^k + (d_N^k)^t M d_N^k,$$

hence the following relation holds:

$$|(d^k)^t M d^k (d_T^k)^t M d_T^k| \leq 2\|M\|\|d_T^k\|_2 \|d_N^k\|_2 + \|M\|\|d_N^k\|_2^2. \tag{33}$$

As for all $\varepsilon > 0$, $a > 0$, $b > 0$, one has $2ab = 2(\varepsilon a)(b/\varepsilon) \leq \varepsilon^2 a^2 + b^2/\varepsilon^2$, it comes for $a = \|d_T^k\|_2$ and $b = \|d_N^k\|_2 \|M\|$:

$$2\|M\|\|d_T^k\|_2 \|d_N^k\|_2 \leq \varepsilon^2 \|d_T^k\|_2^2 + \|M\|^2 \|d_N^k\|_2^2/\varepsilon^2,$$

which with (33) gives the conclusion. □

LEMMA 4.4. *Under the hypotheses of Proposition* 4.1, *for k large enough, Δ_k is feasible and the following holds:*

$$\exists a_5 > 0; \qquad \Delta_k \geq a_5 \|d^k\|_2^2 \tag{34}$$

and (γ being the constant involved in LS2, *i.e., $\gamma \in (0, \frac{1}{2})$):*

$$\exists \varepsilon_1 > 0; \qquad \Delta_k \geq \frac{1}{2(1-\gamma)}(d^k)^t H d^k + \varepsilon_1 \|d^k\|^2. \tag{35}$$

Proof. We restrict our attention to k such that x^k is close to $\bar{x}$.

(a) Preliminaries. It was already noticed (cf. proof of Lemma 4.1(a)) that under our hypotheses $\mu^k \to \bar{\lambda}$. From this result and the hypothesis $r > \|\bar{\lambda}\|_*$ one has for k large enough

$$r - \|\mu^k\|_* \geq (r - \|\bar{\lambda}\|_*)/2,$$

and also it comes for $\zeta := \min\{\bar{\lambda}_i \; ; \; i \in I^* \cap I\}$ (and so $\zeta > 0$) that for k large enough

$$\min\{\mu_i^k \; ; \; i \in I^* \cap I\} > \frac{\zeta}{2}.$$

Hence, as $\mu^k \geq 0$ and $g(x^k)^\sharp \geq g(x^k)$,

$$\begin{aligned}(\mu^k)^t(g(x^k)^\sharp - g(x^k)) &\geq \frac{\zeta}{2} \sum_{i \in I^* \cap I} (g_i(x^k)^\sharp - g_i(x^k)) \\ &= \frac{\zeta}{2} \sum_{i \in I^* \cap I} \max(0, -g_i(x^k)).\end{aligned}$$

¿From the definition of $g(x^k)^\sharp$ and $\tilde{g}(x^k)$ we finally get with (12) that there exists $\xi > 0$ such that for k large enough

$$\Delta_k \geq \xi \|\tilde{g}(x^k)\| + \alpha_k \|d^k\|^2 + (d^k)^t M^k d^k.$$

Now from (32) with $M = M^k$ it follows that for all $\varepsilon > 0$

$$\Delta_k \geq \xi \|\tilde{g}(x^k)\| + \alpha_k \|d^k\|_2^2 + (d_T^k)^t M^k d_T^k - \varepsilon^2 \|d_T^k\|_2^2 - \|M^k\|(1 + \|M^k\|/\varepsilon^2)\|d_N^k\|_2^2.$$

As $\{M^k\}$ is bounded, (8) holds, and d_N^k is solution of

$$\min_d \|d\|_2 \; ; \; g_{I^*}(x^k) + g'_{I^*}(x^k)d = 0,$$

we have

$$d_N^k = O(g_{I^*}(x^k)) = O(\tilde{g}(x^k)), \tag{36}$$

hence for k large enough, since $\|d^k\|_2^2 = \|d_T^k\|_2^2 + \|d_N^k\|_2^2 \geq \|d_T^k\|_2^2$, we get

$$\Delta_k \geq \frac{\xi}{2} \|\tilde{g}(x^k)\| + (\alpha_k - \varepsilon^2)\|d_T^k\|_2^2 + (d_T^k)^t M^k d_T^k. \tag{37}$$

(b) *Proof of* (34). Since (8) and (29) hold, there exists $\delta > 0$ such that for x^k close enough to $\bar{x}$

$$\text{for any} \quad d \in \ker g'_{I^*}(x^k), \quad d^t H d \geq \delta \|d\|^2. \tag{38}$$

¿From (30), (37), and (38) one has for k large enough

$$\begin{aligned}\Delta_k &\geq \frac{\xi}{2} \|\tilde{g}(x^k)\| + \chi (d_T^k)^t H d_T^k - \varepsilon^2 \|d_T^k\|_2^2 + o(d_T^k)^2, \\ &\geq \frac{\xi}{2} \|\tilde{g}(x^k)\| + (\chi\delta - \varepsilon^2)\|d_T^k\|_2^2 + o(d_T^k)^2.\end{aligned}$$

Hence for k large enough, taking $\varepsilon = \sqrt{\chi\delta/3}$ we get

$$\Delta_k \geq \frac{\xi}{2}\|\tilde{g}(x^k)\| + \chi\frac{\delta}{2}\|d_T^k\|_2^2.$$

Using (36) we deduce (34).

Hence, as we assume that $d^k \to 0$, it follows that Δ_k is asymptotically feasible.

(c) We now prove (35). ¿From (30) and (37) we have for k large enough

$$\Delta_k \geq \frac{\xi}{2}\|\tilde{g}(x^k)\| + \chi(d_T^k)^t H d_T^k + \varepsilon^2\|d_T^k\|_2^2.$$

Then using (32) of Lemma 4.2 with $M = H$, we obtain for all $\varepsilon > 0$

$$\Delta_k \geq \frac{\xi}{2}\|\tilde{g}(x^k)\| + \chi(d^k)^t H d^k - 2\varepsilon^2\|d_T^k\|_2^2 - \|H\|(1 + \|H\|/\varepsilon^2)\|d_N^k\|_2^2.$$

Hence one has from (36) for k large enough,

$$\Delta_k \geq \chi(d^k)^t H d^k - 2\varepsilon^2\|d_T^k\|_2^2. \tag{39}$$

Take θ in (0,1) such that $\theta\chi = \dfrac{1}{2(1-\gamma)}$. It follows with (34), (39), and the relation $\|d_T^k\| \leq \|d^k\|$ that

$$\begin{aligned}\Delta_k &= \theta\Delta_k + (1-\theta)\Delta_k \\ &\geq \frac{1}{2(1-\gamma)}(d^k)^t H^k d^k + [(1-\theta)a_5 - 2\varepsilon^2]\|d^k\|^2.\end{aligned}$$

We now choose $\varepsilon_1 = (1-\theta)a_5/2$ and $\varepsilon = \sqrt{\varepsilon_1}$, so that $\varepsilon_1 = (1_\theta)a_5 - 2\varepsilon^2$; relation (35) follows. □

LEMMA 4.5. *Assume that the hypothesis of Proposition* 4.1 *holds. Define* $\bar{x}^k := x^k - \bar{x}$. *Then* $\bar{x}^k = O(d^k)$.

Proof. ¿From the optimality system of $Q_{\alpha_k}(x^k, M^k)$ we deduce that x^k satisfies the optimality system of

$$\min_x f(x) + x^t c^k \ ; \ g(x) + e^k \ll 0$$

with $c^k := M^k d^k + \alpha_k d^k$ and $e^k := g'(x^k)d^k$ and so $c^k = O(d^k)$ and $e^k = O(d^k)$.

Consider the family of perturbed problems

$$(P_{c,e}) \qquad \min_x f(x) + x^t c \ ; \ g(x) + e \ll 0.$$

For $\bar{c} = 0$, $\bar{e} = 0$, $\bar{x}$ is a local solution of $P_{\bar{c},\bar{e}}$ satisfying the regularity hypothesis (the linearized constraints are qualified) and the strong second-order sufficient condition. It follows that for c^k, e^k close to 0, any local solution x^k of the first-order optimality system of (P_{c^k,e^k}) which is in a given neighbourhood of $\bar{x}$ is such that $\bar{x}^k = O(c^k) + O(e^k) = O(d^k)$ (see Robinson [25]). □

Proof of Proposition 4.1. We know from Lemma 4.4 that, for k large enough, Δ_k is feasible; so it remains to check that (28) holds with $\ell = 0$. Define

$$\begin{aligned}\hat{x}^{k+1} &:= x^k + d^k + v^k, \\ \check{x}^{k+1} &:= \hat{x}^{k+1} - \bar{x}, \\ a &:= \theta_r(x^k) - \theta_r(\hat{x}^{k+1}).\end{aligned}$$

We must prove that $a \geq \gamma\Delta_k$. Indeed

$$a = L(x^k, \bar{\lambda}) - L(\hat{x}^{k+1}, \bar{\lambda}) + \bar{\lambda}^t(g(\hat{x}^{k+1}) - g(x^k)) + r(\|g(x^k)^\sharp\| - \|g(\hat{x}^{k+1})^\sharp\|). \tag{40}$$

Expanding $L(., \bar{\lambda})$ at $\bar{x}$ one obtains

$$L(x^k, \bar{\lambda}) - L(\hat{x}^{k+1}, \bar{\lambda}) = \frac{1}{2}(\bar{x}^k)^t H \bar{x}^k - \frac{1}{2}(\check{x}^{k+1})^t H \check{x}^{k+1} + o(\bar{x}^k)^2 + o(\check{x}^{k+1})^2. \tag{41}$$

Moreover one has

$$\begin{aligned}(\bar{x}^k)^t H \bar{x}^k - (\check{x}^{k+1})^t H \check{x}^{k+1} &= (\bar{x}^k - \check{x}^{k+1})^t H (\bar{x}^k + \check{x}^{k+1}) \\ &= -(d^k + v^k)^t H (2\bar{x}^k + d^k + v^k).\end{aligned}$$

So using (24) we get

$$(\bar{x}^k)^t H \bar{x}^k - (\check{x}^{k+1})^t H \check{x}^{k+1} = -2(d^k)^t H \bar{x}^k - (d^k)^t H d^k + o(d^k)^2,$$

then, since (24) yields $\check{x}^{k+1} = \bar{x}^k + d^k + o(d^k)$ and using Lemma 4.4, we obtain from (41)

$$L(x^k, \bar{\lambda}) - L(\hat{x}^{k+1}, \bar{\lambda}) = -(d^k)^t H \bar{x}^k - \frac{1}{2}(d^k)^t H d^k + o(d^k)^2.$$

Then from (25), (26), and Lemma 4.4 we get from (40)

$$a = -(d^k)^t H \bar{x}^k - \frac{1}{2}(d^k)^t H d^k - \bar{\lambda}^t g(x^k) + r\|g(x^k)^\sharp\| + o(d^k)^2. \tag{42}$$

On the other hand we have

$$\begin{aligned}\Delta_k &= r\|g(x^k)^\sharp\| - f'(x^k)d^k \\ &= r\|g(x^k)^\sharp\| - \nabla_x L(x^k, \bar{\lambda})^t d^k + \bar{\lambda}^t g'(x^k) d^k.\end{aligned}$$

So expanding $\nabla_x L(x^k, \bar{\lambda})$ at $\bar{x}$ and using Lemma 4.4

$$\Delta_k = r\|g(x^k)^\sharp\| - (\bar{x}^k)^t H d^k + \bar{\lambda}^t g'(x^k) d^k + o(d^k)^2.$$

Using (23) and the complementarity condition in (6), we get for any $i \in I^*$

$$g_i(x^k) + g_i'(x^k) d^k = 0,$$

hence $-\bar{\lambda}^t g(x^k) = \bar{\lambda}^t g'(x^k) d^k$ and so

$$\Delta_k = r\|g(x^k)^\sharp\| - (\bar{x}^k)^t H d^k - \bar{\lambda}^t g(x^k) + o(d^k)^2.$$

Plugging this in (42) we obtain

$$a = -\frac{1}{2}(d^k)^t H d^k + \Delta_k + o(d^k)^2.$$

We want $a \geq \gamma\Delta_k$, i.e.,

$$(1 - \gamma)\Delta_k \geq \frac{1}{2}(d^k)^t H d^k + o(d^k)^2$$

which is a consequence of (35). □

According to §1.4, we now present an algorithm that is globally convergent (as in §3) and that converges superlinearly when we assume that $\{M^k\}$ approximates in some sense the Hessian of the Lagrangian of problem (P) (using §4 and properties of Newton type algorithms quoted in §1.2). We now state the algorithm.

ALGORITHM 3.
Perform the same steps as in Algorithm 2, replacing LS1 by LS2.

THEOREM 4.1. *Let x^k be computed by Algorithm* 3. *We assume that* (7) *and* (8) *hold. Then*

(i) $\{r_k\}$ *and* $\{\alpha_k\}$ *are bounded.*

(ii) *The set of limit points of* $\{x^k\}$ *is connected and to each of them is associated a Lagrange multiplier.*

(iii) *Assume that the algorithm computes the solution d^k of minimal norm of the optimality system of $Q_{\alpha_k}(x^k, M^k)$. If to some $\bar{x}$ limit-point of $\{x^k\}$ is associated a multiplier $\bar{\lambda}$ such that* (29) *and* (30) *hold, then $x^k \to \bar{x}$ and $\rho_k = 1$ for k large enough. If in addition $P^k[(\nabla_x^2 L(\bar{x}, \bar{\lambda}) - M^k)d^k] = o(d^k)$, then the convergence is superlinear.*

Proof. The arguments for proving (i), (ii) are essentially the same as for Theorem 3.1. As they are rather long we do not reproduce them in detail but rather analyse where the differences are.

Proof of (i). This proof relies on extension of Lemmas 3.1–3.3 for Algorithm 3. Lemma 3.1 is proved by checking that $\|g(x^k)^\sharp\|$ converges if $r_k \searrow \infty$, and on a first-order expansion (in ρ) of $\|g(x^k + \rho d^k)^\sharp\|$. These last arguments have immediate extensions as the paths $\rho \to x^k + \rho d^k$ and $\rho \to x^k + \rho d^k + \rho^2 v^k$ have the same first-order expansion, the term v^k being uniformly bounded. Simple considerations allow an immediate extension of Lemma 3.2. For the extension of Lemma 3.3, estimate (18) on Δ_k is still valid, and (19) also holds, but with a possibly different constant a_4 (because of the additional term v^k) and the conclusion follows. Now the same discussion of points (a), (b) of proof of Theorem 3.1 can be used in order to check that (i) holds.

Proof of (ii). The mechanism of adaptation of $\{x^k\}$ and Lemma 3.3 imply that $\rho_{k'} > 0$ for an infinite subsequence $\{k'\}$, and we may suppose that $\{x^{k'}\} \to \hat{x}$. If $\Delta_{k'} \to 0$ it follows that $d^{k'} \to 0$, hence $\hat{x}$ is a stationary point of (P). If not, assuming $d^{k'} \to \hat{d} \neq 0$ and $v^{k'} \to \hat{v}$ (note that $v^{k'}$ is bounded by (24) hence has limit-points) expanding $\rho \to \theta_r(\hat{x} + \rho\hat{d} + \rho^2\hat{v})$ as in the proof of Theorem 2.1 we deduce that $\rho_{k'}$ cannot converge to 0, hence $\theta_r(x^{k'}) \to \infty$, which is impossible. Henceforth $\hat{d} = 0$ and point (ii) follows. Using (29), (30), and applying the sensitivity result of Robinson [25] to $\bar{d} = 0$ solution of $Q(\bar{x}, \nabla_x^2 L(\bar{x}, \lambda))$ we deduce that $d^k \to 0$ for the considered subsequence.

Proof of (iii). That $\rho_{k'} = 1$ asymptotically for the subsequence $\{x^{k'}\} \to \bar{x}$ is then a consequence of Proposition 4.1. Indeed $\mu^{k'} \to \bar{\lambda}$ as $d^{k'} \to 0$ and (M^k, α_k) are bounded. If $\rho_{k'} < 1$ for a subsequence then (for k' large enough) $r_{k'+1} > \|\bar{\lambda}\|_*$, hence $r > \|\bar{\lambda}\|_*$ and the hypotheses of Proposition 4.1 are satisfied: it follows that $\rho_{k'} = 1$ for k' large enough, hence $\alpha_k \searrow 0$ at a geometric rate.

Now by (29), $\bar{x}$ is an isolated stationary point (see Robinson [25]), and by point (ii) is an isolated limit-point of $\{x^k\}$. As the set of limit points of $\{x^k\}$ is connected it follows that all the sequence converges to $\bar{x}$.

If in addition $P^k[(\nabla_x^2 L(\bar{x}, \bar{\lambda}) - M^k)d^k] = o(d^k)$, then as $\alpha_k \searrow 0$, $P^k[(\nabla_x^2 L(\bar{x}, \bar{\lambda}) - (M^k + \alpha_k I)d^k] = o(d^k)$ hence by Theorem 1.2, $x^k + d^k - \bar{x} = o(x^k - \bar{x})$. As $v^k =$

$O(d^k)^2 = o(x^k - \bar{x})$ we get $x^{k+1} - \bar{x} = o(x^k - \bar{x})$, as desired. □

We now formulate an algorithm that, assuming that the second derivatives of f and g are known, is an extension of Newton's method in the sense that, when x^k is close to some $\bar{x}$ satisfying (29), it computes d^k using $M^k = \nabla_x^2 L(x^k, \mu^{k-1})$ where μ^{k-1} is the multiplier associated to d^{k-1}, and $x^k \to \bar{x}$ with a quadratic rate. The rule is as follows:

$$\text{choose } M^{k+1} = \nabla_x^2 L(x^{k+1}, \lambda^{k+1}) \text{ with}$$

$$\lambda^{k+1} := \begin{cases} \mu^k & \text{if } \alpha_k \|d^k\| + \|M^k d^k\| \leq 1, \\ \mu^k/(1 + \alpha_k \|d^k\| + \|M^k d^k\|) & \text{if not.} \end{cases} \tag{43}$$

THEOREM 4.2. (a) *Let* $\{x^k\}$ *be computed by Algorithm* 3 *with* $\{M^k\}$ *computed by* (43). *We assume that* $\{x^k\}$, $\{d^k\}$ *are bounded, that* (8) *holds and that* $\alpha_{k+1} = 0$ *if* $\rho_k = 1$. *Then points* (i), (ii) *of Theorem* 4.1 *still hold.*

(b) *In addition, if* $\bar{x}$ *satisfying* (29) *is limit-point of* x^k *and* d^k *is the solution of minimal norm of the optimality system of* $Q_{\alpha_k}(x^k, M^k)$, *then all the sequence* $\{x^k\}$ *converges to* $\bar{x}$ *with a quadratic rate.*

Proof. (a) In order to get point (i), (ii) of Theorem 4.1 we must just check that $\{M^k\}$ is bounded; indeed λ^{k+1} is bounded by (8) and (42) hence so is $\{M^k\}$.

Now as $d^k \to 0$ and (M^k, α_k) are bounded, it follows that $\mu^k \to \bar{\lambda}$ and $\lambda^{k+1} = \mu^k$ by (43), hence $M^k \to \nabla_x^2 L(\bar{x}, \bar{\lambda})$ and point (iii) of Theorem 4.1 implies that $\rho_k = 1$ since (30) obviously holds which implies the convergence of all the sequence to $\bar{x}$ at a quadratic rate by Theorem 1.2. □

Acknowledgments. Thanks are due to M. J. D Powell, P. Terpolilli, and an anonymous referee for their remarks that improved a preliminary version of this paper.

REFERENCES

[1] L. ARMIJO, *Minimization of function having Lipschitz continuous first partial derivatives*, Pacific J. Math., 16 (1966), pp. 1–3.

[2] B. M. BELL, *Global convergence of a semi-infinite optimization method*, Appl. Math. Optim., 21 (1990), pp. 69–88.

[3] A. BEN-TAL, *Second-order and related extremality conditions in nonlinear programming*, J. Optimi. Theory Appl., 21 (1980), pp. 143–165.

[4] D. P. BERTSEKAS, *Constrained optimization and Lagrange multipliers methods*, Academic Press, New York, 1982.

[5] G. BLANCHON, J. F. BONNANS, AND J. C. DODU, *Optimisation des réseaux électriques de grande taille*, in Lecture Notes Inf. Control Sci. no 144, A. Bensoussan and J. L. Lions, eds., 1990, pp. 423–431.

[6] J. W. BOGGS, P. W. TOLLE, AND P. WANG, *On the local convergence of quasi-Newton methods for constrained optimization*, SIAM J. Control Optim., 20 (1982), pp. 161–171.

[7] J. F. BONNANS, *Asymptotic admissibility of the unit stepsize in exact penalty methods*, SIAM J. Control Optim., 27 (1989), pp. 631–641.

[8] ———, *Local analysis of Newton type methods for variational inequalities and nonlinear programming*, Applied Math. Optim., 29 (1994), pp. 161–186.

[9] J. F. BONNANS, E. R. PANIER, A. L. TITS, AND J. L. ZHOU, *Avoiding the Maratos effect by means of a nonmonotone line search* II: *Inequality constrained problems–feasible iterates*, SIAM J. Numer. Anal., 29 (1992), pp. 1187–1202.

[10] J. F. BONNANS AND C. POLA, *A trust region interior point algorithm for linearly constrained optimization*, INRIA Report no 1948, 1993.

[11] E. CASAS AND C. POLA, *A sequential generalized quadratic programming algorithm using exact* L_1 *penalty functions*, Optim. Methods Software, to appear.

[12] R. M. CHAMBERLAIN, C. LEMARÉCHAL, H. C. PEDERSEN, AND M. J.D. POWELL, *The watchdog technique for forcing convergence in algorithm for constrained optimization*, Math. Programming Stud., 16 (1982), pp. 1–17.

[13] I. I. EREMIN, *The penalty method in convex programming*, Dokl. Akad. Nauk. SSSR 8 (1966); English translation, Soviet Math. Dokl., 8(1966), pp. 459–462.

[14] R. FLETCHER, *Practical Methods of Optimization*, 2nd edition, Wiley, Chichester, 1987.

[15] D. GABAY, *Reduced quasi-Newton methods with feasibility improvement for nonlinearly constrained optimization*, Math. Programming Stud., 16 (1982), pp. 18–44.

[16] P. HAN *A globally convergent method for nonlinear programming*, J. Optim. Theory Appl., 22 (1977), pp. 297–309.

[17] P. HAN AQND O. L. MANGASARIAN, *Exact penalty functions in nonlinear programming*, Math. Programming, 17 (1979), pp. 251–269.

[18] N. M. JOSEPHY, *Newton's method for generalized equations*, Tech. Summary Report 1965, Mathematics Research Center, University of Wisconsin-Madison, 1979.

[19] N. MARATOS, *Exact Penalty Function Algorithms for Finite Dimensional and Control Optimization Problems*, Ph. D. thesis, University of London, UK, 1978.

[20] D. Q. MAYNE AND E. POLAK *A superlinearly convergent algorithm for constrained optimization problems*, Math. Programming Stud., 16 (1982), pp. 45–61.

[21] E. R. PANIER AND A. L. TITS, *Avoiding the Maratos effect by means of a nonmonotone line search* I *: General constrained problems*, SIAM J. Numer. Anal., 28 (1991), pp. 1183–1195.

[22] B. N. PSCHENICHNY AND Y. M. DANILIN, *Numerical methods in extremal problems*, MIR, Moscow, 1975. (English Translation 1978.)

[23] S. M. ROBINSON, *Stability theory for systems of inequalities, Part* II: *Differentiable nonlinear systems*, SIAM J. Numer. Anal., 13 (1976), pp. 497–513.

[24] ———, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), pp. 43–62.

[25] ———, *Generalized equations and their applications, Part* II *: application to nonlinear programming*, Math. Programming Study, 19 (1982), pp. 200–221.

[26] R. B. WILSON, *A Simplicial Algorithm for Concave Programming,* Ph.D. thesis, Harvard University, Cambridge, MA, 1963.

[27] Y. YUAN, *An example of only linear convergence of trust region algorithms for non-smooth optimization*, IMA J. Numer. Anal., 4 (1984), pp. 327–335.

[28] W. I. ZANGWILL, *Nonlinear programming via penalty functions*, Man. Science, 13 (1967), pp. 344–358.