# Some New Theoretical Results on Recursive Quadratic Programming Algorithms[1]

J. M. Martínez[2] and L. T. Santos[3]

**Abstract.** Recursive quadratic programming is a family of techniques developed by Bartholomew-Biggs and other authors for solving nonlinear programming problems. The first-order optimality conditions for a local minimizer of the augmented Lagrangian are transformed into a nonlinear system where both primal and dual variables appear explicitly. The inner iteration of the algorithm is a Newton-like procedure that updates simultaneously primal variables and Lagrange multipliers. In this way, as observed by N. I. M. Gould, the implementation of Newton's method becomes stable, in spite of the possibility of having large penalization parameters. In this paper the inner iteration is analyzed from a different point of view. Namely, the size of the convergence region and the speed of convergence of the inner process are considered and it is shown that, in some sense, both are independent of the penalization parameter when an adequate version of Newton's method is used. In other words, classical Newton-like iterations are improved, not only in relation to stability of the linear algebra involved, but also with regards to the overall convergence of the nonlinear process. Some numerical experiments suggest that, in fact, practical efficiency of the methods is related to these theoretical results.

**Key Words.** Recursive quadratic programming, penalty methods, Newton's method.

## 1 Introduction

A family of techniques for solving constrained minimization problems, called generically *Recursive Quadratic Programming* (RQP), have been developed by Bartholomew-Biggs and other authors in the last twenty years (see Refs. [1], [2] and [3], among others). Consider the nonlinear programming problem

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & h(x) = 0 \\ & g(x) \leq 0 \end{array} , \tag{1}$$

where $f, h, g$ are smooth functions, $f : I\!R^n \to I\!R$, $h : I\!R^n \to I\!R^m$ and $g : I\!R^n \to I\!R^p$.

The term *Sequential Quadratic Programming* (SQP) will be used in this paper for all the optimization techniques that deal with (1) solving a quadratic programming problem per iteration, while the specific approach in Ref. [1] is called RQP. Given a "current approximation" $x^k$ to the solution of (1), most popular sequential quadratic programming techniques (see, for example, Ref. [4]) solve, at least approximately, a quadratic programming problem where the constraints are local linearizations of the constraints of (1). Unfortunately, the feasible region of this quadratic program can be empty. When inequality constraints are not present, emptyness of the feasible region of the quadratic program usually occurs if $h'(x^k)$ is not a full-rank matrix. In these cases, the domain of the quadratic problem must be modified. The RQP modification is based on the quadratic penalty (or augmented Lagrangian) merit function associated to (1). A straightforward interpretation of this modification is to see the quadratic programming problem as a Newton-like iteration related to the minimization of the augmented Lagrangian of (1). It will be explained in Section 2 (see comments that follow

[1]Work supported by FAPESP (Grant 90-3724-6), CNPq and FAEP / UNICAMP. The authors are indebted to two referees, whose comments were very useful to improve this paper.

[2]Full Professor, Department of Applied Mathematics, IMECC, State University of Campinas, Campinas (SP), Brazil

[3]Assistant Professor, Department of Applied Mathematics, IMECC, State University of Campinas, Campinas (SP), Brazil

formula (10)) the sense in which the RQP approach helps to avoid empty feasible regions in quadratic programming subproblems.

However, as is well-known, large penalty parameters have a harmful influence on the behavior of constrained minimization methods. In fact, the condition number of the associated Hessian matrix, as a function of the penalty parameter, tends to infinity, so the associated linear systems are very hard to solve accurately. Gould (Ref. [5]) introduced a procedure that, by means of the addition of a new artificial "multiplier-like" variable, produces systems where the condition number is essentially independent of the penalty parameter. The implementations in Refs. [1] and [2] have a similar aim.

In this paper it is observed (see Ref. [6], pp. 335–336) that large penalty parameters affect negatively not only the condition of the system but also the size of the convergence region of Newton's (inner) iteration. The remedy for this situation is to extend Gould's procedure in order to define a Newton method applied to the pair primal variable – new variable, instead of a stable implementation of the primal Newton method. In this way, it can be proved, essentially, that the convergence region of the inner iteration is independent of the penalty parameter, as is the condition number of the linear system. This procedure is implicit in Ref. [1] and other implementations of RQP and so, the results presented here explain theoretically the good numerical behavior reported for these methods. In other words, the essential difference between the approach in Ref. [5] and the RQP philosophy is that, in the first case, a stable procedure for the Newtonian linear system applied to the annihilation of the gradient of the penalized objective function is suggested, while, in RQP, the same nonlinear system is decoupled and the Newtonian method is applied to this expanded version. As a result, not only stable implementations (from the linear algebra point of view), but also convergence properties independent of the penalty parameter are obtained.

As with other RQP papers, the developments here will be based on the equality constrained minimization problem:

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & h(x) = 0 \end{array} , \tag{2}$$

where $f : I\!R^n \to I\!R$, $h : I\!R^n \to I\!R^m$, $f, h \in C^2$. The version of RQP analyzed in this paper is, basically, the one presented in Ref. [1]. It consists of an Outer Iteration, where the penalty parameter and the Lagrange multiplier estimators are updated and an Inner Iteration directed to the approximate minimization of the Augmented Lagrangian. In this work the inner iteration is analyzed. Given $x^0 \in I\!R^n$, an estimate to the solution of (2), $\lambda \in I\!R^m$ a vector of estimators for the Lagrange multipliers, and $\mu > 0$ a penalty parameter, the Lagrangian function is defined as

$$L(x, \lambda) = f(x) + \lambda^T h(x). \tag{3}$$

In order to simplify the notation, let us write

$$\ell(x) = L(x, \lambda) \text{ for all } x \in I\!R^n.$$

At a very high level, the inner iteration is

$$\text{Minimize (approximately) } P_\mu(x) \tag{4}$$

where $P_\mu(x) = \ell(x) + \frac{1}{2\mu}\|h(x)\|_2^2$ for all $x \in I\!R^n$. Under suitable conditions, for $\mu$ small, the minimizer of $P_\mu$ is an approximation of the solution of (2) (see Refs. [4], [6] and [7]).

Newton's method for minimizing $P_\mu$ is defined by

$$x^{k+1} = x^k - [\nabla^2 P_\mu(x^k)]^{-1} \nabla P_\mu(x^k) \ , \tag{5}$$

where

$$\nabla P_\mu(x) \ = \ \nabla \ell(x) \ + \ \frac{1}{\mu} h'(x)^T h(x) \ ,$$

$$\nabla^2 P_\mu(x) \;=\; \nabla^2 \ell(x) \;+\; \frac{1}{\mu} h'(x)^T h'(x) \;+\; \frac{1}{\mu}\sum_{i=1}^{m} h_i(x)\nabla^2 h_i(x)$$

and $h'(x) = [\nabla h_1(x), \ldots, \nabla h_m(x)]^T$.

Since the condition number of $\nabla^2 P_\mu$ tends to infinity as $\mu$ goes to zero, the Newton iteration can be unstable. Gould's (Ref. [5]) proposal consists of replacing (5) by

$$\begin{pmatrix} B(x^k) & h'(x^k)^T \\ h'(x^k) & -\mu I \end{pmatrix} \begin{pmatrix} s^k \\ r^k \end{pmatrix} = \begin{pmatrix} -\nabla \ell(x^k) \\ -h(x^k) \end{pmatrix}, \tag{6}$$

where $B(x) = \nabla^2 \ell(x) + \frac{1}{\mu}\sum_{i=1}^{m} h_i(x)\nabla^2 h_i(x)$ and $x^{k+1} = x^k + s^k$. (The intermediate variables $r^k$ are not used at all in further computations.) The condition number of the matrix of the system (6) turns out to be bounded independently of $\mu$, and so, the Newton step can be computed without major stability problems. In fact, calling $G(\mu)$ the matrix of the system (6) for $\mu > 0$, $G(0)$ is the Jacobian matrix of the nonlinear system associated with the Lagrange optimality conditions, and assuming that $G(0)$ is nonsingular, continuity ensures that there exists $\mu' > 0$ such that the condition number of $G(\mu)$ is less than (say) $2\|G(0)\|\,\|G(0)^{-1}\|$ for $\mu < \mu'$. Since this quantity is independent of $\mu$, the stability of (6) is also independent of the penalty parameter.

As mentioned above, another reason for practical inefficiency of the penalty approach in many problems is related to the convergence properties of the iteration (5). Newton's method has local quadratic convergence if $\nabla^2 P_\mu$ is nonsingular at $\tilde{x}(\mu)$, the minimizer of $P_\mu$, but classical convergence analysis shows that the size of the region of fast convergence is proportional to $1/(2\overline{\beta}(\mu)\overline{\gamma}(\mu))$ and that the speed of convergence is given by

$$\|x^{k+1}(\mu) - \tilde{x}(\mu)\| \le \overline{\beta}(\mu)\overline{\gamma}(\mu)\|x^k(\mu) - \tilde{x}(\mu)\|^2 ,$$

where $\overline{\beta}(\mu) = \|[\nabla^2 P_\mu(\tilde{x}(\mu))]^{-1}\|$ and $\overline{\gamma}(\mu)$ is a Lipschitz constant associated to the variation of $\nabla^2 P_\mu$ (see Ref. [8], pp. 90–91). It is easy to see that $\overline{\beta}(\mu)$ is bounded independently of $\mu$, but this is not the case for $\overline{\gamma}(\mu)$, which tends to infinity as $\mu$ tends to zero, when $h$ is nonlinear. Therefore, iterations based on (5) or (6) tend to be very slow or not convergent when $\mu$ is very small, independently of the condition number of $\nabla^2 P_\mu$.

In the RQP approach both the stability and the convergence drawbacks of the classical approach are overcome. Instead of transforming the Newton iteration into a stable augmented linear system, the optimality condition $\nabla P_\mu(x) = 0$ is converted into an equivalent decomposed *nonlinear* system. It will be proved that for this system the convergence properties of Newton's method do not deteriorate as $\mu$ goes to zero. The Newton–Gould iteration described in Ref. [5] can be interpreted as a Newton iteration on the augmented system, followed by a *restoration* step that forces one of the equations of the system to be satisfied at each iteration. So, the RQP approach consists, simply, in eliminating the restoration step.

In Section 2 of this work the augmented system is defined and convergence results are proved. In Section 3 a numerical example is shown that illustrates well the theoretical properties. In Section 4 the relation with homotopic approaches are discussed. Finally, in Section 5 the consequences of this approach for practical optimization are commented.

## 2 The Augmented Nonlinear System

Introducing the new variables $y \equiv h(x)/\mu$, the nonlinear system $\nabla P_\mu(x) = 0$ can be written as

$$\left\{ \begin{array}{lcll} \nabla \ell(x) & + & h'(x)^T y & = & 0 \\ h(x) & - & \mu y & = & 0 \end{array} \right. . \tag{7}$$

The system (7) has the form $F_\mu(x,y)=0$, where $F_\mu : I\!R^{n+m} \to I\!R^{n+m}$. Observe that $F_0(x,y)=0$ is the system of optimality conditions for (2). Let us define

$$H(x,y) \;\equiv\; \nabla^2\ell(x) \;+\; \sum_{i=1}^{m} y_i \nabla^2 h_i(x)\;. \tag{8}$$

Therefore, the Jacobian of $F_\mu$ is given by

$$F'_\mu(x,y) \;=\; \begin{pmatrix} H(x,y) & h'(x)^T \\ h'(x) & -\mu I \end{pmatrix}\;. \tag{9}$$

After some algebraic manipulation, it can be shown that the Newton iteration for (7) turns to be

$$\begin{pmatrix} x^{k+1} \\ y^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ 0 \end{pmatrix} - \begin{pmatrix} H(x^k,y^k) & h'(x^k)^T \\ h'(x^k) & -\mu I \end{pmatrix}^{-1} \begin{pmatrix} \nabla\ell(x^k) \\ h(x^k) \end{pmatrix}\;. \tag{10}$$

Note that $x^{k+1}$ satisfies the optimality conditions for the quadratic problem

$$\begin{array}{ll} \text{Minimize} & \frac{1}{2}(x-x^k)^T H(x^k,y^k)(x-x^k) + \nabla\ell(x^k)(x-x^k) \\ \text{subject to} & h'(x^k)(x-x^k) + h(x^k) = \mu y^{k+1} \end{array}\;. \tag{11}$$

The interpretation (11) justifies the denomination "Recursive Quadratic Programming" for the methods based on (10). A peculiarity of the subproblem (11), which is not present in Sequential Quadratic Programming methods, is that the variable $y^{k+1}$ appears in the definition of the feasible region. Clearly, the solvability of (10) is not related to the rank of $h'(x^k)$, so that (11) can be solvable even if $h'(x^k)$ is not a full-rank matrix (when, very likely, the SQP feasible set, $h'(x^k)(x-x^k)+h(x^k)=0$, is empty).

As in (6), if $F'_0$ is nonsingular at a limit point, the condition number of the matrix that appears in (10) does not deteriorate when $\mu$ goes to zero and so, the Newton iteration (10) is stable from the linear algebra point of view.

Observe that if $x^{k+1}$ is computed from (10) but the replacement $y^{k+1}=h(x^{k+1})/\mu$ is performed, i.e. the current iterate is "projected" on the last equation of (7), the resulting iteration is equivalent to (6). This is the restoration step mentioned in Section 1. The calculation of $x^{k+1}$ and $y^{k+1}$ in (10) depends on the solution of a linear system with coefficient matrix $F'_\mu(x^k,y^k)$. Extensive numerical experimentation on the solution of this type of systems within the RQP framework has been reported in Refs. [2], [3] and other RQP papers. Good results have been obtained using the Harwell sparse symmetric subroutine MA27 and present active research concerns dealing efficiently with the indefinite case.

The aim of the analysis presented in this paper is to show that convergence properties of (10) can be found which are independent of the penalty parameter $\mu$. This means that, if the restoration step $y^{k+1}=h(x^{k+1})/\mu$ is eliminated, "all" the subproblems (7) have (essentially) the same degree of difficulty, as far as a Newtonian procedure for its solution is considered. In the following, the main results are Theorems 2.3, 2.4 and 2.5. In Theorem 2.3 it is proved that a convergence region for (10) can be defined with size independent of $\mu$. In Theorem 2.4, bounds (independent of $\mu$) are given for the errors in $x^{k+1}$ and $y^{k+1}$ with respect to the corresponding errors in $x^k$ and $y^k$. Finally, in Theorem 2.5 a result concerning uniform (independent of $\mu$) Q-superlinear convergence is proved.

For all $x \in I\!R^n$ and $y \in I\!R^m$ let us define $\|(x,y)\| = \max\{\|x\|,\|y\|\}$, where $\|\cdot\|$ is an arbitrary norm. Below the main local assumptions for the convergence analysis are stated.

**Main Assumptions.** For $\mu = 0$ the system (7) has a unique solution $(x^*, y^*)$ with $F_0'(x^*, y^*)$ nonsingular. Moreover, there exist constants $L_i > 0$, $i = 0, \ldots, m+1$, and $\varepsilon > 0$, such that

$$\|\nabla^2 \ell(x) - \nabla^2 \ell(z)\| \leq L_0 \, \|x - z\| \, , \tag{12}$$

$$\|\nabla^2 h_i(x) - \nabla^2 h_i(z)\| \leq L_i \, \|x - z\| \tag{13}$$

and

$$\|h'(x) - h'(z)\| \leq L_{m+1} \, \|x - z\| \, , \tag{14}$$

for all $x, z \in \mathcal{B}_\varepsilon(x^*)$, where $\mathcal{B}_\varepsilon(x^*) = \{x \in I\!R^n \mid \|x - x^*\| < \varepsilon\}$. This implies (see Ref. [8], pp. 75–76) that for all $x, z \in \mathcal{B}_\varepsilon(x^*)$, $i = 1, \ldots, m$,

$$\|\nabla \ell(x) - \nabla \ell(z) - \nabla^2 \ell(z)(x - z)\| \leq \frac{L_0}{2} \, \|x - z\|^2 \, , \tag{15}$$

$$\|\nabla h_i(x) - \nabla h_i(z) - \nabla^2 h_i(z)(x - z)\| \leq \frac{L_i}{2} \, \|x - z\|^2 \, , \tag{16}$$

and

$$\|h(x) - h(z) - h'(z)(x - z)\| \leq \frac{L_{m+1}}{2} \, \|x - z\|^2 \, . \tag{17}$$

**Theorem 2.1.** There exist $\overline{\mu} > 0$, $0 < r_1 < \varepsilon$, $r_2 > 0$ such that for all $\mu \in [0, \overline{\mu}]$, $\|x - x^*\| \leq r_1$ and $\|y - y^*\| \leq r_2$, the system (7) has a unique solution $(\tilde{x}(\mu), \tilde{y}(\mu))$, with $F_\mu'(\tilde{x}(\mu), \tilde{y}(\mu))$ nonsingular. Moreover, $\tilde{x}(\mu)$ and $\tilde{y}(\mu)$ are continuous functions of $\mu$.

**Proof.** Use the main assumptions and the implicit function theorem. □

From now on let us denote

$$S = \{(x, y) \in I\!R^{n+m} \mid \|x - x^*\| \leq r_1 \, , \; \|y - y^*\| \leq r_2\},$$

where $r_1$ and $r_2$ are given in Theorem 2.1. The constant $\overline{\mu}$ and the functions $\tilde{x}(\mu)$ and $\tilde{y}(\mu)$ will be the ones defined in that theorem.

**Theorem 2.2.** There exist $L, M > 0$ such that, for all $\mu \in [0, \overline{\mu}]$ and $(x, y) \in S$,
(i) $\|F_\mu'(x, y) - F_\mu'(\tilde{x}(\mu), \tilde{y}(\mu))\| \leq L \, \|(x, y) - (\tilde{x}(\mu), \tilde{y}(\mu))\|$ .
(ii) $\|[F_\mu'(\tilde{x}(\mu), \tilde{y}(\mu))]^{-1}\| \leq M$.

**Proof.** The two inequalities follow from (8), (9), (12), (13), (14) and the continuity of $\tilde{x}(\mu)$ and $\tilde{y}(\mu)$ . □

**Theorem 2.3.** There exist $\widehat{\mu}, \varepsilon_1, \varepsilon_2 > 0$ such that for all $\mu \in [0, \widehat{\mu}]$, whenever

$$\|x^0(\mu) - \tilde{x}(\mu)\| \leq \varepsilon_1 \text{ and } \|y^0(\mu) - \tilde{y}(\mu)\| \leq \varepsilon_2 \, ,$$

the sequences $\{x^k(\mu)\}$ and $\{y^k(\mu)\}$ generated by (10) are well-defined and converge to $\tilde{x}(\mu)$ and $\tilde{y}(\mu)$, respectively. Moreover, for all $\mu \in [0, \widehat{\mu}]$, $k = 0, 1, 2, \ldots$,

$$\|(x^k(\mu), y^k(\mu)) - (\tilde{x}(\mu), \tilde{y}(\mu))\| \leq \frac{\max\{\varepsilon_1, \varepsilon_2\}}{2^k}. \tag{18}$$

**Proof.** By Theorem 2.1, there exists $\widehat{\mu} \in (0, \overline{\mu})$ such that, for all $\mu \in [0, \widehat{\mu}]$,

$$\|\tilde{x}(\mu) - x^*\| \leq \frac{r_1}{2} \quad \text{and} \quad \|\tilde{y}(\mu) - y^*\| \leq \frac{r_2}{2} \, .$$

Therefore, for $0 \le \mu \le \widehat{\mu}$, if

$$\|x - \tilde{x}(\mu)\| \le \frac{r_1}{2} \quad \text{and} \quad \|y - \tilde{y}(\mu)\| \le \frac{r_2}{2}, \tag{19}$$

then $(x, y) \in S$. Thus, the inequalities (i) and (ii) of Theorem 2.2 hold when $(x, y)$ satisfy (19) and $\mu \in [0, \widehat{\mu}]$. Let us define

$$\varepsilon_1 = \min\left\{\frac{r_1}{2}, \frac{1}{2LM}\right\} \quad \text{and} \quad \varepsilon_2 = \min\left\{\frac{r_2}{2}, \frac{1}{2LM}\right\}.$$

By the convergence theorem of Newton's method (Ref. [8], pp. 90–91), for all $\mu \in [0, \widehat{\mu}]$, if

$$\|x^0(\mu) - \tilde{x}(\mu)\| \le \varepsilon_1 \quad \text{and} \quad \|y^0(\mu) - \tilde{y}(\mu)\| \le \varepsilon_2$$

then the Newton iteration (10) is well-defined,

$$\|[F'_\mu(x^k(\mu), y^k(\mu))]^{-1}\| \le 2M \ ,$$

the sequence $\{(x^k(\mu), y^k(\mu))\}$ converges to $(\tilde{x}(\mu), \tilde{y}(\mu))$ and satisfies

$$\|(x^{k+1}(\mu), y^{k+1}(\mu)) - (\tilde{x}(\mu), \tilde{y}(\mu))\| \le \frac{1}{2}\, \|(x^k(\mu), y^k(\mu)) - (\tilde{x}(\mu), \tilde{y}(\mu))\| \ ,$$

for $k = 0, 1, 2, \ldots$ . Therefore, the inequality (18) also holds and the proof is complete. □

**Theorem 2.4.** Assume that $\{x^k(\mu)\}$ and $\{y^k(\mu)\}$ are generated by (10) under the conditions of Theorem 2.3. Then, there exist $\alpha, \beta, \gamma, \delta > 0$, independent of $\mu$, such that, for all $\mu \in [0, \widehat{\mu}]$, $k = 0, 1, 2, \ldots$,

$$\|x^{k+1}(\mu) - \tilde{x}(\mu)\| \le \alpha \|x^k(\mu) - \tilde{x}(\mu)\| \cdot \|y^k(\mu) - \tilde{y}(\mu)\| + \beta \|x^k(\mu) - \tilde{x}(\mu)\|^2 \tag{20}$$

and

$$\|y^{k+1}(\mu) - \tilde{y}(\mu)\| \le \gamma \|x^k(\mu) - \tilde{x}(\mu)\| \cdot \|y^k(\mu) - \tilde{y}(\mu)\| + \delta \|x^k(\mu) - \tilde{x}(\mu)\|^2 \ . \tag{21}$$

If $h$ is linear ($\nabla^2 h_i(x) \equiv 0$ for all $i$) then $\alpha = \gamma = 0$.

**Proof.** For simplicity, let us write $\tilde{y} \equiv \tilde{y}(\mu)$, $\tilde{x} \equiv \tilde{x}(\mu)$, $x^k \equiv x^k(\mu)$, $y^k \equiv y^k(\mu)$, $u^k \equiv x^k(\mu) - \tilde{x}$ and $v^k \equiv y^k(\mu) - \tilde{y}$. By (10), we have that

$$\begin{aligned} \begin{pmatrix} u^{k+1} \\ v^{k+1} \end{pmatrix} &= \begin{pmatrix} u^k \\ -\tilde{y} \end{pmatrix} - [F'_\mu(x^k, y^k)]^{-1} \begin{pmatrix} \nabla \ell(x^k) \\ h(x^k) \end{pmatrix} \\ &= [F'_\mu(x^k, y^k)]^{-1} \left[ F'_\mu(x^k, y^k) \begin{pmatrix} u^k \\ -\tilde{y} \end{pmatrix} - \begin{pmatrix} \nabla \ell(x^k) \\ h(x^k) \end{pmatrix} \right] \ . \end{aligned} \tag{22}$$

But, by (15) and (17),

$$\nabla \ell(\tilde{x}) = \nabla \ell(x^k) - \nabla^2 \ell(x^k) u^k + w_1(x^k, \tilde{x}) \tag{23}$$

and

$$h(\tilde{x}) = h(x^k) - h'(x^k) u^k + w_2(x^k, \tilde{x}) \ , \tag{24}$$

where $w_1 : \mathbb{R}^{2n} \to \mathbb{R}^n$ and $w_2 : \mathbb{R}^{2n} \to \mathbb{R}^m$ are such that

$$\|w_1(x^k, \tilde{x})\| \le \frac{L_0}{2} \|u^k\|^2 \text{ and } \|w_2(x^k, \tilde{x})\| \le \frac{L_{m+1}}{2} \|u^k\|^2. \tag{25}$$

Using (23) and (24) and remembering that $F_\mu(\tilde{x},\tilde{y})=0$, the recurrence (22) takes the form

$$\begin{pmatrix} u^{k+1} \\ v^{k+1} \end{pmatrix} = [F'_\mu(x^k,y^k)]^{-1} \begin{pmatrix} [h'(\tilde{x})-h'(x^k)]^T\tilde{y} + \left[\sum_{i=1}^m y_i^k \nabla^2 h_i(x^k)\right] u^k + w_1(x^k,\tilde{x}) \\ w_2(x^k,\tilde{x}) \end{pmatrix} . \tag{26}$$

Now, by (16),

$$\nabla h_i(\tilde{x}) = \nabla h_i(x^k) - \nabla^2 h_i(x^k)u^k + w_3^i(x^k,\tilde{x}) ,$$

where $w_3^i : I\!R^{2n} \to I\!R^n$ is such that

$$\|w_3^i(x^k,\tilde{x})\| \le \frac{L_i}{2}\|u^k\|^2 , \ i=1,\ldots,m .$$

Therefore, we can write

$$[h'(\tilde{x})-h'(x^k)]^T\tilde{y} = -\left[\sum_{i=1}^m \tilde{y}_i \nabla^2 h_i(x^k)\right] u^k + w_3(x^k,\tilde{x},\tilde{y}) , \tag{27}$$

where $w_3(x^k,\tilde{x},\tilde{y}) = \sum_{i=1}^m \tilde{y}_i w_3^i(x^k,\tilde{x})$ is such that

$$\|w_3(x^k,\tilde{x},\tilde{y})\| \le \sum_{i=1}^m \frac{L_i\ \tilde{y}_i}{2}\ \|u^k\|^2 \le \sum_{i=1}^m \frac{L_i(r_2+\|y^*\|)}{2}\ \|u^k\|^2 \le \overline{L}\ \|u^k\|^2 , \tag{28}$$

with

$$\overline{L} = \sum_{i=1}^m \frac{L_i\ (r_2+\|y^*\|^2)}{2} .$$

Finally, substituting (27) in (26) we obtain

$$\begin{pmatrix} u^{k+1} \\ v^{k+1} \end{pmatrix} = [F'_\mu(x^k,y^k)]^{-1} \begin{pmatrix} \left[\sum_{i=1}^m v_i^k \nabla^2 h_i(x^k)\right] u^k + w_1(x^k,\tilde{x}) + w_3(x^k,\tilde{x},\tilde{y}) \\ w_2(x^k,\tilde{x}) \end{pmatrix} . \tag{29}$$

Let us write now

$$[F'_\mu(x^k,y^k)]^{-1} = \begin{pmatrix} P_\mu^k & Q_\mu^k \\ R_\mu^k & S_\mu^k \end{pmatrix} .$$

It can be proved that $R_\mu^k = (Q_\mu^k)^T$. Since by the proof of Theorem 2.3, we have

$$\|[F'_\mu(x^k,y^k)]^{-1}\| \le 2M,$$

there exists a constant $c>0$ (which only depends on $M$ and $\|\cdot\|$), such that $\|P_\mu^k\|$, $\|Q_\mu^k\|$, $\|R_\mu^k\|$, $\|S_\mu^k\| \le c$. Therefore, the desired result follows by taking norms in (29), using the continuity of the $\nabla^2 h_i$'s and the bounds for $w_1$, $w_2$ and $w_3$ given by (25) and (28). If $h$ is linear, by (29) we have $\alpha=\gamma=0$. □

**Remarks.** The inequalities (20)–(21) imply that for all $\mu \in [0,\widehat{\mu}]$ and $k=0,1,2,\ldots$,

$$\|(u^{k+1}(\mu),v^{k+1}(\mu))\| \le \max\{\alpha+\beta,\gamma+\delta\}\|(u^k(\mu),v^k(\mu))\|^2. \tag{30}$$

This is a restatement of the well-known quadratic convergence of Newton's method on the pair $(x,y)$. Therefore, the sequences $\{\|x^k(\mu)-\tilde{x}(\mu)\|\}$ and $\{\|y^k(\mu)-\tilde{y}(\mu)\|\}$ are bounded by a $Q$-quadratically

convergent sequence and so, they converge $R$-quadratically (see Ref. [9]). What is important here is to emphasize that $\alpha, \beta, \gamma$ and $\delta$ do not depend on $\mu$ and thus *$R$-quadratic convergence independent of $\mu$* is established. Recall that in the classical Newton approach (5)–(6), $Q$-quadratic convergence takes place but the constants involved depend on $\mu$ and tend to infinity as $\mu$ goes to zero.

Let us now give a precise meaning to the intuitive expression "$Q$-superlinear convergence independent of $\mu$". The sequences $\{\{x^k(\mu)\}, \mu \in [0, \widehat{\mu}]\}$ are said to converge to $\{\tilde{x}(\mu), \mu \in [0, \widehat{\mu}]\}$ *$Q$-superlinearly and uniformly with respect to* $\mu \in [0, \widehat{\mu}]$, if

$$\lim_{k \to \infty} x^k(\mu) = \tilde{x}(\mu) \text{ for all } \mu \in [0, \widehat{\mu}]$$

and there exists a sequence of positive real numbers $\{c_k\}$ converging to 0, such that

$$\|x^{k+1}(\mu) - \tilde{x}(\mu)\| \leq c_k \|x^k(\mu) - \tilde{x}(\mu)\| \text{ for all } \mu \in [0, \widehat{\mu}] \ .$$

**Theorem 2.5.** Let $\widehat{\mu} > 0$, $\varepsilon_1$, $\varepsilon_2$ be defined as in Theorem 2.3. Then, $\{\{x^k(\mu)\}, \mu \in [0, \widehat{\mu}]\}$ converge to $\{\tilde{x}(\mu), \mu \in [0, \widehat{\mu}]\}$ $Q$-superlinearly and uniformly with respect to $\mu \in [0, \widehat{\mu}]$.

**Proof.** Define $\varepsilon = \max\{\varepsilon_1, \varepsilon_2\}$. By (18) and (20) we have

$$\begin{aligned} \|x^{k+1}(\mu) - \tilde{x}(\mu)\| &\leq \alpha(\|y^k(\mu) - \tilde{y}(\mu)\| + \beta\|x^k(\mu) - \tilde{x}(\mu)\|)\|x^k(\mu) - \tilde{x}(\mu)\| \\ &\leq \frac{\varepsilon}{2^k}(\alpha + \beta)\|x^k(\mu) - \tilde{x}(\mu)\| \ . \end{aligned}$$

So, the uniform superlinear convergence of $\{\{x^k(\mu)\}, \mu \in [0, \widehat{\mu}]\}$ is obtained defining $c_k = \frac{\varepsilon}{2^k}(\alpha + \beta)$ .□

# 3 Numerical Example

In this section, a simple example illustrates the convergence results proved in Section 2. Consider the problem

$$\begin{array}{ll} \text{Minimize} & x_1^2 + 2x_2^2 \\ \text{subject to} & (x_1 - 2)^2 + 2(x_2 - 1)^2 = 1 \end{array} \ .$$

The solution is $x^* \approx (1.1835, 0.59175)^T$ and the associated Lagrange multiplier is $y^* \approx 1.4495$. Classical penalization (i.e. $\lambda = 0$) will be used here.

| $\mu$ | $\tilde{x}_1(\mu)$ | $\tilde{x}_2(\mu)$ | $\tilde{y}(\mu)$ |
|---|---|---|---|
| $10^0$ | 0.8993576 | 0.44967879 | 0.8171206 |
| $10^{-1}$ | 1.1319029 | 0.56595143 | 1.3038896 |
| $10^{-2}$ | 1.1776775 | 0.58883877 | 1.4321359 |
| $10^{-3}$ | 1.1829126 | 0.59145630 | 1.4477186 |
| $10^{-4}$ | 1.1834443 | 0.59172213 | 1.4493123 |
| $10^{-5}$ | 1.1834975 | 0.59174875 | 1.4494720 |

**Table 3.1.** *Solution of $F_\mu(x, y) = 0$ for each $\mu$.*

| $k$ | $\mu = 1$ | $\mu = 10^{-1}$ | $\mu = 10^{-2}$ | $\mu = 10^{-3}$ | $\mu = 10^{-4}$ | $\mu = 10^{-5}$ |
|---|---|---|---|---|---|---|
| 0 | 0.6409 | 0.6471 | 0.6483 | 0.6484 | 0.6484 | 0.6484 |
| 1 | 0.6270 | 0.6370 | 0.6388 | 0.6390 | 0.6391 | 0.6391 |
| 2 | 0.6049 | 0.6215 | 0.6245 | 0.6248 | 0.6249 | 0.6249 |
| 3 | 0.5698 | 0.5977 | 0.6026 | 0.6032 | 0.6032 | 0.6032 |
| 4 | 0.5151 | 0.5618 | 0.5698 | 0.5707 | 0.5708 | 0.5708 |
| 5 | 0.4369 | 0.5134 | 0.5258 | 0.5272 | 0.5273 | 0.5273 |
| 6 | 0.3341 | 0.4864 | 0.5052 | 0.5071 | 0.5073 | 0.5074 |
| 7 | 0.1621 | 0.5789 | 0.6821 | 0.6935 | 0.6946 | 0.6948 |
| 8 | 0.0249 | 0.4461 | 0.9009 | 0.9827 | 0.9915 | 0.9924 |
| 9 | 0.0006 | 0.1106 | 0.5499 | 0.9354 | 1.0008 | 1.0078 |
| 10 | 0.0000 | 0.0348 | 0.3509 | 0.2965 | 0.7458 | 0.9676 |
| 11 | | 0.0007 | 0.2688 | 0.9719 | 0.9566 | 0.7733 |
| 12 | | 0.0000 | 0.0592 | 0.2666 | 0.3061 | 0.9957 |
| 13 | | | 0.0075 | 0.6004 | 1.0238 | 0.3373 |
| 14 | | | 0.0000 | 0.0354 | 0.5636 | 1.0307 |
| 15 | | | | 0.0645 | 0.7831 | 0.9294 |
| 16 | | | | 0.0000 | 0.2253 | 0.5577 |
| 17 | | | | | 0.7010 | 0.9812 |
| 18 | | | | | 0.0187 | 0.0623 |
| 19 | | | | | 0.0869 | 1.0362 |
| 20 | | | | | 0.0000 | 0.0607 |
| 21 | | | | | | 0.2038 |
| 22 | | | | | | 0.0007 |
| 23 | | | | | | 0.0000 |

**Table 3.2.** *Convergence rates for Newton's method (Gould's approach).*

| $k$ | $\mu = 1$ | $\mu = 10^{-1}$ | $\mu = 10^{-2}$ | $\mu = 10^{-3}$ | $\mu = 10^{-4}$ | $\mu = 10^{-5}$ |
|---|---|---|---|---|---|---|
| 0 | 0.6409 | 0.6471 | 0.6483 | 0.6484 | 0.6484 | 0.6484 |
| 1 | 0.5931 | 0.6041 | 0.6061 | 0.6064 | 0.6064 | 0.6064 |
| 2 | 0.5534 | 0.5732 | 0.5768 | 0.5772 | 0.5772 | 0.5772 |
| 3 | 0.4992 | 0.5355 | 0.5418 | 0.5425 | 0.5426 | 0.5426 |
| 4 | 0.4158 | 0.4823 | 0.4933 | 0.4945 | 0.4946 | 0.4946 |
| 5 | 0.2887 | 0.4300 | 0.4495 | 0.4516 | 0.4518 | 0.4518 |
| 6 | 0.1020 | 0.4763 | 0.5928 | 0.6065 | 0.6078 | 0.6080 |
| 7 | 0.0083 | 0.3160 | 0.8499 | 0.9686 | 0.9819 | 0.9832 |
| 8 | 0.0000 | 0.0351 | 0.3687 | 0.8750 | 0.9932 | 1.0066 |
| 9 | | 0.0017 | 0.0645 | 0.1192 | 0.4789 | 0.9062 |
| 10 | | 0.0000 | 0.0022 | 0.0346 | 0.1030 | 0.1528 |
| 11 | | | 0.0000 | 0.0009 | 0.0057 | 0.0179 |
| 12 | | | | 0.0000 | 0.0002 | 0.0004 |
| 13 | | | | | 0.0000 | 0.0000 |

**Table 3.3.** *Convergence rates for RQP approach with* $y^0 = h(x^0)/\mu$.

| $k$ | $\mu = 1$ | $\mu = 10^{-1}$ | $\mu = 10^{-2}$ | $\mu = 10^{-3}$ | $\mu = 10^{-4}$ | $\mu = 10^{-5}$ |
|---|---|---|---|---|---|---|
| 0 | 0.4758 | 0.4851 | 0.4868 | 0.4870 | 0.4870 | 0.4870 |
| 1 | 0.4601 | 0.4818 | 0.4856 | 0.4861 | 0.4861 | 0.4861 |
| 2 | 0.3539 | 0.4073 | 0.4162 | 0.4172 | 0.4173 | 0.4173 |
| 3 | 0.1454 | 0.2741 | 0.2962 | 0.2986 | 0.2989 | 0.2989 |
| 4 | 0.1096 | 0.1679 | 0.2039 | 0.2082 | 0.2087 | 0.2087 |
| 5 | 0.0381 | 0.0692 | 0.0667 | 0.0673 | 0.0673 | 0.0673 |
| 6 | 0.0003 | 0.0110 | 0.0208 | 0.0221 | 0.0222 | 0.0223 |
| 7 | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0001 | 0.0001 |
| 8 | | | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

**Table 3.4.** *Convergence rates for RQP approach with* $y^0 = 0$.

The minimization of $P_\mu(x)$ for $\mu = 10^{-j}, j = 0, 1, \ldots, 5$ using (6) (classical Newton applied to $\nabla P_\mu(x) = 0$) and the approach based on (7) and (10) was considered. The initial approximations were $x^0 = (-10, -10)^T$ and $y^0 = h(x^0)/\mu$ or $y^0 = 0$, for all $\mu$, and the iterative process was terminated when $\|x^k(\mu) - \tilde{x}(\mu)\|_2 < 10^{-8}$.

The values of $\tilde{x}(\mu)$ for this problem were obtained with 16 correct digits using a polynomial solver as shown in Table 3.1. In Tables 3.2, 3.3 and 3.4 the convergence rates $\dfrac{\|x^{k+1}(\mu) - \tilde{x}(\mu)\|_2}{\|x^k(\mu) - \tilde{x}(\mu)\|_2}$ for Newton's classical method (with Gould's system) and for the RQP approach are shown. With regards to the results on Table 3.2, observe that Newton's method automatically imposes the choice $y^k = h(x^k)/\mu$, and in particular for $k = 0$.

# 4 The Homotopic Approach

The discussion displayed in this section was motivated on comments by N. I. M. Gould, who also called the attention of the authors to the references Refs. [10], [11], [12], and [13].

The iteration (5), applied to the solution of (4), can be written in the following way:

$$[\nabla^2 \ell(x^k) + \sum_{k=1}^{m} y_i^k \nabla^2 h_i(x^k) + \frac{1}{\mu} h'(x^k)^T h'(x^k)](x^{k+1} - x^k) = -\nabla P_\mu(x^k) \ , \tag{31}$$

where

$$y^k = \frac{h(x^k)}{\mu} \ , \tag{32}$$

for $k = 0, 1, 2, \ldots$. On the other hand, the RQP–Newton iteration (10) obeys the formula (31) for the updating of $x$ and

$$y^{k+1} = \frac{h(x^k) + h'(x^k)(x^{k+1} - x^k)}{\mu} \tag{33}$$

for the updating of $y$. In (33) nothing is said about the initial estimate of $y^0$, while in (32) the first Newton iteration imposes that $y^0 = h(x^0)/\mu$. In many implementations of penalty methods a sequence of problems of the form (5) for different values of a monotone decreasing sequence $\mu_\nu \to 0$ must be solved. In fact, penalty methods can be interpreted as particular homotopic methods for solving the nonlinear system determined by the Lagrangian equations and the feasibility requirement. As in other homotopic methods, the (approximate) solution for some value of the homotopic parameter provides the necessary elements for the computation of an estimate of the solution for the "next" homotopic parameter. Therefore, if the difference between consecutive penalty parameters "is not very large", it can be guaranteed that few iterations are needed to obtain an approximate solution corresponding to the new penalty parameter (see Ref. [7], p. 285).

Gould (Ref. [12]) observed that, since the $y$-solution of problem (7) defined by $\mu_\nu$ is $h(\tilde{x}(\mu_\nu))/\mu_\nu$, then the initial $y^0$ that is reasonable to use in (31)–(32) for $\mu_{\nu+1}$ is $h(x^0)/\mu_\nu$ instead of $h(x^0)/\mu_{\nu+1}$, where $x^0 = \tilde{x}(\mu_\nu)$. It turns out that Gould's new estimate is of the correct order while the traditional estimate is a factor $\mu_\nu/\mu_{\nu+1}$ too large. So, in the implementation of Newton's method for solving (4), the formulae (31)–(32) should be used, but $y^0$ should be replaced by $h(\tilde{x}(\mu_\nu))/\mu_\nu$. The analysis of Gould (Ref. [12]) shows that, proceeding in this way, if the successive penalty parameters are chosen conveniently, only one "modified Newton" iteration is necessary for each value of the penalty parameter and overall convergence of the algorithm to the solution of the original problem is global and two-step quadratic.

Related results, for different homotopies, were obtained in Refs. [10], [11] and [13]. The predictor–corrector idea underlies most theoretically justified versions of penalty methods. In the predictor phase, a "tangent" step to the homotopy curve is taken, which is able to provide a good initial point

for Newton's method. The corrector phase consists of the application of Newton's method, or some other iterative nonlinear solver, to the problem defined by the new penalty parameter. Clever implementations take advantage of the special structure of nonlinear programming. (The exploitation of structure, connected with interior-point homotopies for linear programming problems has been surveyed in Ref. [14].) The homotopic idea for general nonlinear systems has a long tradition in numerical analysis, having being recognized as one of the best practical ways to improve global convergence of local nonlinear solvers (see Refs. [9], [15], [16] and [17]).

Many times (for example, in the case of nonlinear programming) the only objective is to solve the nonlinear system (optimality conditions) and not to follow closely the trajectory. In this cases it is interesting to permit large variations between $\mu_\nu$ and $\mu_{\nu+1}$ with the expectancy of arriving quickly to a solution of (1). An extreme case is the "shortcut" strategy (see Ref. [7], p. 285) in which only one small penalty parameter is used. Other "nonparametric" variants are discussed in Lootsma (Ref. [18]). Of course, when $\mu_\nu/\mu_{\nu+1}$ is large, it can no longer be expected that the solution of the $\nu$-th problem provides accurate information to compute an initial estimate for the solution of the $(\nu+1)$-th problem. In such case, several Newtonian iterations will be necessary to attain suitable approximate solutions for (4) (or (7)) and so, to have convergence rates and asymptotic speed independent of $\mu$ becomes very important. Nevertheless, it can be conjectured that large-step penalty methods (in the sense that $\mu_\nu/\mu_{\nu+1} \gg 1$) can be developed associated to (10) as an inner iteration, with fixed number of internal steps, taking advantage of the uniform convergence properties proved in Section 2. In fact, Gould's updating procedure is a theoretically justified way of allowing rapid decrease of the penalty parameter. Probably, many heuristic decreasing strategies can also be effective when they are associated to the system (10).

It is also important, in practice, to consider the globalization of the local scheme defined by (31) and (33). This has been done in current implementations of RQP methods (Refs. [1], [2] and [3]). If $x^{k+1} - x^k$ is replaced by $s^k$ in (31) then $s^k$ can be considered a search direction for a quasi-Newton-like method oriented to the minimization of $P_\mu(x)$. Of course, in the case of line-search globalizations, safeguards are needed in order to guarantee positive-definiteness of the matrix, and boundedness of the Lagrangian-Hessian is required for trust-region modifications. This implies that global modifications of (31) must be implemented with safeguards for guaranteeing boundedness of the vectors $y^k$. Fortunately, it has been proved in Ref. [1] that, under suitable conditions on the problem, near an exact solution of (4) the unitary step is accepted (so the global method reduces to the local one) and safeguards on $y$ are not necessary.

# 5 Final Remarks

Practical efficiency of an optimization algorithm is related not only to global and local convergence results, but also to the theoretical properties of inner algorithms used in the computations. In general, inner algorithms are linear iterative methods for which assymptotic convergence properties are well-known. In other situations, as is the case of the methods studied in this paper, the inner algorithm is itself nonlinear, and its convergence analysis is more involved.

In this paper, the inner algorithm used in the Newtonian version of RQP is interpreted as a pure Newton method on the pair $(x, y)$ for the augmented nonlinear system (7). Moreover, if $y^0 = h(x^0)/\mu$, the Newton iteration for the system $\nabla P_\mu(x) = 0$ can be interpreted as a Newton iteration for (7) followed by a restoration step $y^{k+1} = h(x^{k+1})/\mu$ (this is a particular case of a property that was proved for arbitrary nonlinear systems by Griewank (Ref. [19])). Both the Newton iteration for (7) and the restored Newton iteration for $\nabla P_\mu(x) = 0$, proposed originally by Gould, are stable from the linear algebra point of view. However, the Newtonian iteration applied to (7) exhibits additional properties which, roughly speaking, can be described as *convergence radius and assymptotic speed independent of* $\mu$. These theoretical results explain, at least partially, the excellent practical performance of RQP algorithms. A similar independence property is exhibited by the method introduced in Ref. [10].

Quasi-Newton implementations of RQP use suitable secant updates for the Hessian approximations, instead of true Hessians. In these implementations, the role of the auxiliary variable $y^k$ is not evident, since some linear algebra manipulations (see Ref. [1]) make it disappear on explicit calculations. However, a more careful analysis shows that the non-restored variable $y^{k+1}$ is used to define the secant equation. In fact, the quasi-Newton version of (10) consists of replacing $H(x^k, y^k)$ by a suitable symmetric matrix $H_k$, where $H_{k+1}$, for $k = 0, 1, 2, \ldots$ is chosen in order to satisfy the secant equation

$$H_{k+1}(x^{k+1} - x^k) = [\nabla \ell(x^{k+1}) + \sum_{i=1}^{m} y_i^{k+1} \nabla h_i(x^{k+1})] - [\nabla \ell(x^k) + \sum_{i=1}^{m} y_i^{k+1} \nabla h_i(x^k)], \tag{34}$$

minimizing some norm of $H_{k+1} - H_k$. In other words, the Hessian secant approximation used in Ref. [1] corresponds to a structured quasi-Newton update for the system (7) and so, it can be conjectured that *uniformity* results also hold for quasi-Newton versions of RQP. Classical theories on structured quasi-Newton updates (Refs. [20], [21], [22] and [23]) can be useful for a rigorous proof of this conjecture.

In practical implementations of RQP, damped inner iterations ($x^{k+1} = x^k + t_k s_k$) are used instead of the pure Newton or quasi-Newton iterations (where $t_k = 1$), with the aim of producing monotone reduction of a suitable merit function (the augmented Lagrangian). In this context, uniformity results can be interpreted as establishing uniform lower bounds on the size of the regions where the undamped iteration can be used and have a uniform high rate of convergence. The simple example presented in Section 3 is intended to give some numerical insight on how the theoretical results link to numerical performance.

The decomposition process that leads to (7) admits yet another interpretation, when one deals with the general purpose of solving nonlinear systems $F(x) = 0$, where $F : \mathbb{R}^n \to \mathbb{R}^n$. In fact, assume that the nonlinear system can be written as

$$F(x) = G(x, \varphi(x)) = 0 \ , \tag{35}$$

where $\varphi : \mathbb{R}^n \to \mathbb{R}^m$. Therefore, solving $F(x) = 0$ is equivalent to solving the augmented nonlinear system

$$\begin{cases} G(x, y) = 0 \\ \varphi(x) - y = 0, \end{cases} \tag{36}$$

where the structure of (7) can be recognised. It is easy to prove (see Ref. [19]) that the Newton iteration applied to $F(x) = 0$ corresponds to the Newton iteration applied to (36), starting with $y^0 = \varphi(x^0)$ and followed by the restoration step $y^{k+1} = \varphi(x^{k+1})$. The process that leads to (7) is a particular case of this decomposition procedure, where it could be proved that using Newton's method on (36), without restoration, is better than using Newton's method on (35). This model structure can be discovered in other typical situations of numerical analysis. For example, the relation between (35) and (36) generalizes the one existing between single shooting and multiple shooting methods for solving two-point boundary-value problems. Independently of convergence results, numerical analysts prefer to use (36) instead of $F(x) = 0$ when good estimators of the variables $y$ are available, so that the restoration step, which ignores such estimation, is not recommendable. These observations suggest that the relations between the Newtonian decomposition strategy that leads to (7) and more general decomposition schemes of nonlinear systems should be exploited. As one of the referees observed, the application of the logarithmic barrier function to inequality constrained optimization with slack variables also generates a system, whose (primal–dual) decomposition can be useful both from the stability and from the uniform convergence point of view. A lot of research on this subject has been produced in the last ten years, especially in connection to the linear programming problem. Moreover, growing research in the last few years (see Refs. [10] and [24]) has been devoted to the efficient implementation of interior point methods for nonlinearly constrained minimization.

## References


1. Bartholomew-Biggs, M. C., *Recursive Quadratic Programming Methods Based on the Augmented Lagrangian*, Mathematical Programming Study, Vol. 31, pp. 21–41, 1987.

2. Bartholomew-Biggs, M. C., and Hernández, M. F. G., *Some Improvements of the Subroutine OPALQP for Dealing with Large Problems*, Journal of Economic Dynamics and Control, Vol. 18, pp. 185–203, 1994.

3. Bartholomew-Biggs, M. C., and Hernández, M. F. G., *Using the KKT Matrix in an Augmented Lagrangian SQP Method for Sparse Constrained Optimization*, Journal of Optimization Theory and Applications, Vol. 85, pp. 201–220, 1995.

4. Gill, P. E., Murray, W., and Wright, M. H., *Practical Optimization*, Academic Press, London, 1981.

5. Gould, N. I. M., *On the Accurate Determination of Search Directions for Simple Differentiable Penalty Functions*, IMA Journal of Numerical Analysis, Vol. 6, pp. 357–372, 1986.

6. McCormick, G. P., *Nonlinear Programming*, John Wiley and Sons, New York, 1983.

7. Fletcher, R., *Practical Methods of Optimization*, John Wiley and Sons, New York, 1987.

8. Dennis, J. E., and Schnabel, R. B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, New Jersey, 1983.

9. Ortega, J. M., and Rheinboldt, W. C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

10. Conn, A. R., Gould, N. I. M., and Toint, Ph. L., *A Note on Using Alternative Second-Order Models for the Subproblems Arising in Barrier Function Methods for Minimization*, Numerische Mathematik, Vol. 68, pp. 17–33, 1994.

11. Dussault, J-P., *Numerical Stability and Efficiency of Penalty Algorithms*, SIAM Journal on Numerical Analysis, Vol. 32, pp. 296–317, 1995.

12. Gould, N. I. M., *On the Convergence of the Sequential Penalty Function Method for Constrained Minimization*, SIAM Journal on Numerical Analysis, Vol. 26, pp. 107–128, 1989.

13. McCormick, G. P., *The Superlinear Convergence of a Nonlinear Primal–Dual Algorithm*, Technical Report OR T-550/91, Department of Operations Research, George Washington University, 1991.

14. Gonzaga, C. C., *Path Following Methods for Linear Programming*, SIAM Review, Vol. 34, pp. 167–224, 1992.

15. Forster, W., *Homotopy Methods*, to appear in Handbook of Global Optimization, Kluwer, 1995.

16. Rheinboldt, W. C., *Numerical Analysis of Parameterized Nonlinear Equations*, John Wiley and Sons, New York, 1986.

17. Watson, L. T., Billups, S. C., and Morgan, A. P., *Algorithm 652 - HOMPACK: A Suite of Codes for Globally Convergent Homotopy Algorithms*, ACM Transactions Mathematical Software, Vol. 13, pp. 281–310, 1987.

18. Lootsma, F. A., *A Survey of Methods for Solving Constrained Minimization Problems via Unconstrained Minimization*, Numerical Methods for Nonlinear Optimization, Edited by F. A. Lootsma, Academic Press, New York, New York, pp. 313–347, 1972.

19. Griewank, A., *Direct Calculation of Newton Steps without Accumulating Jacobians*, Large-Scale Numerical Optimization, Edited by T. F. Coleman and Y. Li, SIAM, Philadelphia, pp. 115–137, 1990.

20. Dennis, J. E., and Walker, H. F., *Convergence Theorems for Least-Change Secant Update Methods*, SIAM Journal on Numerical Analysis, Vol. 18, pp. 949–987, 1981.

21. Martínez, J. M., *Local Convergence Theory of Inexact Newton Methods Based on Structured Least Change Updates*, Mathematics of Computation, Vol. 55, pp. 143–167, 1990.

22. Martínez, J. M., *Fixed-Point Quasi-Newton Methods*, SIAM Journal on Numerical Analysis, Vol. 29, pp. 1413–1434, 1992.

23. Tapia, R. A., *Diagonalized Multiplier Methods and Quasi-Newton Methods for Constrained Optimization*, Journal of Optimization Theory and Applications, Vol. 22, pp. 135–194, 1977.

24. Wright, M. H., *Why a Pure Primal Newton Barrier Step May Be Infeasible*, SIAM Journal on Optimization, Vol. 5, pp. 1–13, 1995.

## References


[1] Bartholomew-Biggs, M. C., *Recursive Quadratic Programming Methods Based on the Augmented Lagrangian*, Mathematical Programming Study, Vol. 31, pp. 21–41, 1987.

[2] Bartholomew-Biggs, M. C., and Hernández, M. F. G., *Some Improvements of the Subroutine OPALQP for Dealing with Large Problems*, Journal of Economic Dynamics and Control, Vol. 18, pp. 185–203, 1994.

[3] Bartholomew-Biggs, M. C., and Hernández, M. F. G., *Using the KKT Matrix in an Augmented Lagrangian SQP Method for Sparse Constrained Optimization*, Journal of Optimization Theory and Applications, Vol. 85, pp. 201–220, 1995.

[4] Gill, P. E., Murray, W., and Wright, M. H., *Practical Optimization*, Academic Press, London, 1981.

[5] Gould, N. I. M., *On the Accurate Determination of Search Directions for Simple Differentiable Penalty Functions*, IMA Journal of Numerical Analysis, Vol. 6, pp. 357–372, 1986.

[6] McCormick, G. P., *Nonlinear Programming*, John Wiley and Sons, New York, 1983.

[7] Fletcher, R., *Practical Methods of Optimization*, John Wiley and Sons, New York, 1987.

[8] Dennis, J. E., and Schnabel, R. B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, New Jersey, 1983.

[9] Ortega, J. M., and Rheinboldt, W. C., *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

[10] Conn, A. R., Gould, N. I. M., and Toint, Ph. L., *A Note on Using Alternative Second-Order Models for the Subproblems Arising in Barrier Function Methods for Minimization*, Numerische Mathematik, Vol. 68, pp. 17–33, 1994.

[11] Dussault, J-P., *Numerical Stability and Efficiency of Penalty Algorithms*, SIAM Journal on Numerical Analysis, Vol. 32, pp. 296–317, 1995.

[12] Gould, N. I. M., *On the Convergence of the Sequential Penalty Function Method for Constrained Minimization*, SIAM Journal on Numerical Analysis, Vol. 26, pp. 107–128, 1989.

[13] McCormick, G. P., *The Superlinear Convergence of a Nonlinear Primal–Dual Algorithm*, Technical Report OR T-550/91, Department of Operations Research, George Washington University, 1991.

[14] Gonzaga, C. C., *Path Following Methods for Linear Programming*, SIAM Review, Vol. 34, pp. 167–224, 1992.

[15] Forster, W., *Homotopy Methods*, to appear in Handbook of Global Optimization, Kluwer, 1995.

[16] Rheinboldt, W. C., *Numerical Analysis of Parameterized Nonlinear Equations*, John Wiley and Sons, New York, 1986.

[17] Watson, L. T., Billups, S. C., and Morgan, A. P., *Algorithm 652 - HOMPACK: A Suite of Codes for Globally Convergent Homotopy Algorithms*, ACM Transactions Mathematical Software, Vol. 13, pp. 281–310, 1987.

[18] Lootsma, F. A., *A Survey of Methods for Solving Constrained Minimization Problems via Unconstrained Minimization*, Numerical Methods for Nonlinear Optimization, Edited by F. A. Lootsma, Academic Press, New York, New York, pp. 313–347, 1972.

[19] Griewank, A., *Direct Calculation of Newton Steps without Accumulating Jacobians*, Large-Scale Numerical Optimization, Edited by T. F. Coleman and Y. Li, SIAM, Philadelphia, pp. 115–137, 1990.

[20] Dennis, J. E., and Walker, H. F., *Convergence Theorems for Least-Change Secant Update Methods*, SIAM Journal on Numerical Analysis, Vol. 18, pp. 949–987, 1981.

[21] Martínez, J. M., *Local Convergence Theory of Inexact Newton Methods Based on Structured Least Change Updates*, Mathematics of Computation, Vol. 55, pp. 143–167, 1990.

[22] Martínez, J. M., *Fixed-Point Quasi-Newton Methods*, SIAM Journal on Numerical Analysis, Vol. 29, pp. 1413–1434, 1992.

[23] Tapia, R. A., *Diagonalized Multiplier Methods and Quasi-Newton Methods for Constrained Optimization*, Journal of Optimization Theory and Applications, Vol. 22, pp. 135–194, 1977.

[24] Wright, M. H., *Why a Pure Primal Newton Barrier Step May Be Infeasible*, SIAM Journal on Optimization, Vol. 5, pp. 1–13, 1995.