Local Convergence of an Inexact-Restoration Method and Numerical Experiments [1]

E. G. BIRGIN [2] AND J. M. MARTÍNEZ [3]

Communicated by C. T. Leondes

[1]This work was supported by PRONEX-CNPq/FAPERJ Grant E-26/171.164/2003 - APQ1, FAPESP Grants 03/09169-6 and 01/04597-4, and CNPq. The authors are indebted to Juliano B. Francisco and Yalcin Kaya for their careful reading of the first draft of this paper.

[2]Associate Professor, Department of Computer Science, Institute of Mathematics and Statistics, University of São Paulo, São Paulo, SP, Brazil.

[3]Professor, Department of Applied Mathematics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Campinas, SP, Brazil.

Abstract. Local convergence of an Inexact-Restoration method for nonlinear programming is proved. Numerical experiments are performed with the objective of evaluating the behavior of the purely local method against a globally convergent nonlinear-programming algorithm.

Key Words. Inexact-Restoration methods, nonlinear programming, local convergence, numerical experiments.

1 Introduction

Inexact-Restoration (IR) methods (see Ref. 1–3) are modern versions of the classical feasible methods (Ref. 4–12) for nonlinear programming. The main iteration of an IR algorithm consists of two phases: in the restoration phase, infeasibility is reduced and in the optimality phase a Lagrangian function is approximately minimized on an appropriate linear approximation of the constraints. Global convergence is obtained in Ref. 1 by means of a trust-region strategy where the trust balls are centered, not in the current point, as in several sequential quadratic programming algorithms (see, for example, Ref. 13) but in the inexactly restored point. The merit function used in Ref. 1 is a sharp Lagrangian as defined in Ref. 14, Example 11.58.

Merit functions are useful tools in all branches of optimization. However, it has been observed that in many practical situations the performance of optimization algorithms that do not impose merit-function decrease is better than the performance of algorithms whose global convergence is based on merit functions. The reason is that merit-function decrease imposes a restrictive path towards the limit point whereas, sometimes, the purely local algorithm climbs over merit-function valleys in a very efficient way.

In unconstrained optimization, nonmonotone strategies, where decrease of the merit function is not required at every iteration (Ref. 15), became a popular tool in the last decade.

In nonlinear programming, the more consistent strategy for globalizing algorithms without the use of merit functions seems to be the filter technique introduced by Fletcher and Leyffer (Ref. 16). Gonzaga, Karas and Vanti (Ref. 17) applied the filter strategy to an algorithm that resembles Inexact Restoration. Previous attempts of eliminating merit functions as globalization tools for semifeasible methods go back to Ref. 18.

It is not difficult to modify poor algorithms in order to obtain theoretically globally convergent methods. This can be made using both monotone or nonmonotone strategies. In general, the modification of a poor local method leads to a poor global method. A good globally convergent method is usually good even before the global modification and, sometimes, the purely local version is better than the global one. One of the key features that allow one to predict the practical behavior of an optimization algorithm is the presence of a local convergence theorem with order of convergence. In general, the existence of such a theorem indicates that the model used at each iteration to mimic the original problem is adequate. (Other evidences of this adequacy exist but are less susceptible of mathematical formalization.) This was our motivation for developing a local convergence theory for the Inexact-Restoration algorithm. Since our main objective is to explain and test the behavior of methods for solving practical problems, the numerical experiments that complete this paper are directed to evaluate the efficiency and robustness of the purely local algorithm, against globally convergent ones.

The local algorithm and its convergence theory is presented in Section 2. In Section 3 we describe the implementation. Numerical experiments are shown in Section 4 and conclusions are given in Section 5.

2 Local Convergence of Inexact Restoration

In this section we assume that $\Omega \subset I\!R^n$ is closed and convex. We also assume that $f : I\!R^n \to I\!R$ and $h : I\!R^n \to I\!R^m$ admit continuous first derivatives on an open set that contains Ω. The optimization problem to be considered is:

$$\min f(x) \text{ s.t. } h(x) = 0, \quad x \in \Omega. \tag{1}$$

For all $x \in \Omega$, $\lambda \in I\!R^m$, we define the Lagrangian function $L(x, \lambda)$ as

$$L(x, \lambda) = f(x) + \langle h(x), \lambda \rangle.$$

We denote $\nabla h(x) = (\nabla h_1(x), \ldots, \nabla h_m(x))$ and $h'(x) = \nabla h(x)^T$. Therefore, $\nabla L(x, \lambda) = \nabla f(x) + \nabla h(x)\lambda$. The symbol $\|\cdot\|$ will always denote the Euclidian norm along this paper. Let P be the projection operator onto Ω with respect to $\|\cdot\|$. We say that $(x_*, \lambda_*) \in \Omega \times I\!R^m$ is a *critical pair* of the optimization problem (1) if

$$h(x_*) = 0 \text{ and } P(x_* - \nabla L(x_*, \lambda_*)) - x_* = 0. \tag{2}$$

Under suitable constraint qualifications every local minimizer of (1) defines, with its Lagrange multipliers, a critical pair. (See, for example, Ref. 19.) In this section we will analyze a locally convergent algorithm for finding critical pairs, without any mention to the origin of the nonlinear system (2). We will address the resolution of this nonsmooth nonlinear system of equations using a variation of the Inexact-Restoration algorithm introduced in Ref. 1. We denote

$$G(x,\lambda) = P(x - \nabla L(x,\lambda)) - x \;\; \forall \;\; x \in \Omega, \;\; \lambda \in I\!R^m.$$

Therefore, $\|h(x)\|$ is a measure of the feasibility of $x \in \Omega$ and $\|G(x,\lambda)\|$ measures the "optimality" of the pair (x,λ). Given the current iterate $x \in \Omega$, the idea of IR is to find, first, a "more feasible" point $y \in \Omega$, and then, to find a "more optimal" point z such that $z \in \Omega$ and $h'(y)(z-y) = 0$. (This condition will be relaxed in (5).)

The Inexact-Restoration iteration depends on five algorithmic parameters $\theta \in [0,1), \eta \in [0,1)$ and $K_1, K_2, K_3 > 0$. The first two indicate the amount of improvement that we require in the feasibility phase and the optimality phase, respectively. The role of K_1 and K_3 is to maintain the new iterate reasonably close to the current one. (See Ref. 1, 3 for details.) The constant K_2 gives a tolerance for the linear infeasibility of the optimality-phase minimizer.

Given $x \in \Omega$ and $\lambda \in I\!R^m$, we say that an IR iteration starting from (x, λ) can be completed (or is well defined) if we can compute $y, z \in \Omega$, $\mu \in I\!R^m$ such that:

$$\|h(y)\| \leq \theta \|h(x)\|, \tag{3}$$

$$\|y - x\| \leq K_1 \|h(x)\|, \tag{4}$$

$$\|h'(y)(z - y)\| \leq K_2 \|G(y, \lambda)\|^2, \tag{5}$$

$$\|P(z - \nabla L(z, \lambda) - \nabla h(y)(\mu - \lambda)) - z\| \leq \eta \|G(y, \lambda)\| \tag{6}$$

and

$$\|z - y\| + \|\mu - \lambda\| \leq K_3 \|G(y, \lambda)\|. \tag{7}$$

The motivation for the condition (6) comes from considering that, in the optimality phase, one generally minimizes the Lagrangian $L(z, \lambda)$ subject to $z \in \Omega$ and $h'(y)(z - y) = 0$. Writing the optimality conditions for this subproblem and defining $(\mu - \lambda)$ as the vector of Lagrange multipliers corresponding to these conditions, we obtain

$$P(z - \nabla L(z, \lambda) - \nabla h(y)(\mu - \lambda)) - z = 0.$$

So, inequality (6) is an inexact version of this condition. The stability conditions (4) and (7) express the necessity of staying close to the current point if this point is close to feasibility or optimality respectively.

Given the pair $(x, \lambda) \in \Omega \times I\!R^m$, if the IR iteration can be completed giving (z, μ), we denote $N_{[\theta,\eta,K_1,K_2,K_3]}(x, \lambda) = (z, \mu)$. For simplicity, we will always denote $N(x, \lambda) = N_{[\theta,\eta,K_1,K_2,K_3]}(x, \lambda) = (z, \mu)$.

Throughout this section we will assume that ∇f and ∇h are Lipschitz-continuous. To simplify the notation, and without loss of generality, we assume that, for the same Lipschitz constant γ and for all $x, w \in \Omega$, $i = 1, \ldots, m$,

$$\|\nabla f(x) - \nabla f(w)\| \leq \gamma \|x - w\|, \quad \|\nabla h_i(x) - \nabla h_i(w)\| \leq \gamma \|x - w\|, \tag{8}$$

$$\|\nabla h(x) - \nabla h(w)\| \leq \gamma \|x - w\|, \tag{9}$$

and

$$\|h(w) - h(x) - h'(x)(w - x)\| \leq \gamma \|w - x\|^2. \tag{10}$$

We define the following constants, that will be used along this section:

$$c = \max\{K_1, K_2, K_3\}, \quad c_1 = 2c + c\gamma, \quad c_2 = c\gamma,$$

$$c_3 = c + 2c^2 + c^2\gamma, \quad c_4 = c^2\gamma + c.$$

Theorem 2.1. Assume that the IR iteration starting from (x, λ) can be completed and $(z, \mu) = N(x, \lambda)$. Then,

$$\|h(z)\| \leq \theta\|h(x)\| + c_4[\|G(x, \lambda)\| + (c_1 + c_2\|\lambda\|)\|h(x)\|]^2, \tag{11}$$

$$\|G(z, \mu)\| \leq$$
$$\eta[(c_1 + c_2\|\lambda\|)\|h(x)\| + \|G(x, \lambda)\|] + c_4[\|G(x, \lambda)\| + (c_1 + c_2\|\lambda\|)\|h(x)\|]^2, \tag{12}$$

$$\|z - x\| \leq (c_3 + c_4\|\lambda\|)\|h(x)\| + c\|G(x, \lambda)\| \tag{13}$$

and

$$\|\mu - \lambda\| \leq (c_3 + c_4\|\lambda\|)\|h(x)\| + c\|G(x, \lambda)\|. \tag{14}$$

Proof. By (10), $\|h(z) - h(y)\| \leq \|h'(y)(z - y)\| + \gamma\|z - y\|^2$. So, by (3), (5) and (7),

$$\|h(z)\| \leq \theta\|h(x)\| + (\gamma c^2 + c)\|G(y, \lambda)\|^2. \tag{15}$$

Now, by (4) and (8)–(10),

$$\|G(y,\lambda) - G(x,\lambda)\| = \|P(y - \nabla L(y,\lambda)) - y - (P(x - \nabla L(x,\lambda)) - x)\|$$

$$\leq \|y - x\| + \|P(y - \nabla L(y,\lambda)) - P(x - \nabla L(x,\lambda))\|$$

$$\leq \|y - x\| + \|y - x + \nabla L(x,\lambda) - \nabla L(y,\lambda)\|$$

$$\leq 2\|y - x\| + \|\nabla f(y) - \nabla f(x)\| + \|[\nabla h(x) - \nabla h(y)]\lambda\|$$

$$\leq 2c\|h(x)\| + \gamma\|y - x\| + \gamma\|y - x\|\|\lambda\|$$

$$\leq (2c + c\gamma + c\gamma\|\lambda\|)\|h(x)\| = (c_1 + c_2\|\lambda\|)\|h(x)\|.$$

Therefore,

$$\|G(y,\lambda)\| \leq \|G(x,\lambda)\| + (c_1 + c_2\|\lambda\|)\|h(x)\|. \tag{16}$$

So, by (15) and (16),

$$\|h(z)\| \leq \theta\|h(x)\| + (\gamma c^2 + c)[\|G(x,\lambda)\| + (c_1 + c_2\|\lambda\|)\|h(x)\|]^2.$$

Therefore, (11) is proved.

Now,

$$\|P(z - \nabla L(z,\mu)) - z\| = \|P(z - \nabla f(z) - \nabla h(z)\mu) - z\|$$

$$= \|P[z - \nabla f(z) - \nabla h(z)(\mu - \lambda) + \nabla h(z)(\mu - \lambda) - \nabla h(z)\mu] - z\|$$

$$= \|P[z - \nabla f(z) - \nabla h(z)\lambda - \nabla h(z)(\mu - \lambda)] - z\|$$

$$= \|P[z - \nabla L(z, \lambda) - \nabla h(y)(\mu - \lambda) + (\nabla h(y) - \nabla h(z))(\mu - \lambda)] - z\|.$$

Using the property

$$\|P(v + w) - z\| \leq \|P(v + w) - P(v)\| + \|P(v) - z\| \leq \|w\| + \|P(v) - z\|$$

with $v = z - \nabla L(z, \lambda) - \nabla h(y)(\mu - \lambda)$ and $w = (\nabla h(y) - \nabla h(z))(\mu - \lambda)$, by (6), (8)–(10) and (7), we get:

$$\|P(z - \nabla L(z, \mu)) - z\| \leq \|P[z - \nabla L(z, \lambda) - \nabla h(y)(\mu - \lambda)] - z\| + \|\nabla h(y) - \nabla h(z)\| \|\mu - \lambda\|$$

$$\leq \eta \|G(y, \lambda)\| + \gamma \|y - z\| \|\mu - \lambda\| \leq \eta \|G(y, \lambda)\| + \gamma(\|y - z\| + \|\mu - \lambda\|)^2$$

$$\leq \eta \|G(y, \lambda)\| + \gamma c^2 \|G(y, \lambda)\|^2.$$

So, by (16),

$$\|G(z, \mu)\| \leq \eta[\|G(x, \lambda)\| + (c_1 + c_2\|\lambda\|)\|h(x)\|] + \gamma c^2[\|G(x, \lambda)\| + (c_1 + c_2\|\lambda\|)\|h(x)\|]^2.$$

Therefore, (12) is also proved.

Now, by (4), (7) and (16),

$$\|z-x\| \leq \|y-x\| + \|z-y\| \leq c\|h(x)\| + c\|G(y,\lambda)\|$$

$$\leq c\|h(x)\| + c[\|G(x,\lambda)\| + (2c^2 + c^2\gamma + c^2\gamma\|\lambda\|)\|h(x)\|]$$

$$= (c + 2c^2 + c^2\gamma + c^2\gamma\|\lambda\|)\|h(x)\| + c\|G(x,\lambda)\|.$$

So, (13) is proved.

Moreover, by (7) and (16),

$$\|\mu-\lambda\| \leq c\|G(y,\lambda)\| \leq c\|G(x,\lambda)\| + (2c^2 + c^2\gamma + c^2\gamma\|\lambda\|)\|h(x)\|.$$

Thus, (14) is also proved. □

From now on we assume that $(\bar{x},\bar{\lambda}) \in \Omega \times I\!R^m$ is a critical pair. So, $h(\bar{x}) = 0$ and $G(\bar{x},\bar{\lambda}) = 0$.

We also define $M = 2\|\bar{\lambda}\| + 1$, $c_5 = c_1 + c_2 M$, and $H \in I\!R^{2\times 2}$ by

$$H = \begin{pmatrix} \theta & 0 \\ c_5 & \eta \end{pmatrix}.$$

The eigenvalues of H are θ and η. Since both are strictly smaller than 1, given an arbitrary $\varepsilon > 0$, there exists a vector norm $\|\cdot\|_H$ on $I\!R^2$ such that

$$\|H\|_H = \rho \leq \max\{\theta, \eta\} + \varepsilon < 1. \tag{17}$$

Moreover, this norm is monotone in the sense that $0 \leq v \leq w \Rightarrow \|v\|_H \leq \|w\|_H$. From now on, we fix a "contraction" parameter r such that

$$\rho < r < 1. \tag{18}$$

Theorem 2.2. There exist $\varepsilon_1 > 0$, $\delta_1 > 0, \beta > 0$ such that, if r is given by (18), $\|x - \bar{x}\| \leq \varepsilon_1$, $\|\lambda - \bar{\lambda}\| \leq \delta_1$, and the IR iteration starting from (x, λ) is well defined, with $(z, \mu) = N(x, \lambda)$, then

$$\|\lambda\| \leq M,$$

$$\left\| \begin{pmatrix} \|h(z)\| \\ \\ \|G(z, \mu)\| \end{pmatrix} \right\|_H \leq r \left\| \begin{pmatrix} \|h(x)\| \\ \\ \|G(x, \lambda)\| \end{pmatrix} \right\|_H, \tag{19}$$

$$\|z - x\| \leq \beta \left\| \begin{pmatrix} \|h(x)\| \\ \\ \|G(x, \lambda)\| \end{pmatrix} \right\|_H, \tag{20}$$

and

$$\|\mu - \lambda\| \leq \beta \left\| \begin{pmatrix} \|h(x)\| \\ \\ \|G(x,\lambda)\| \end{pmatrix} \right\|_H. \tag{21}$$

Proof. Take $\delta_0 = \|\bar{\lambda}\| + 1$. Then, $\|\lambda - \bar{\lambda}\| \leq \delta_0$. So, $\|\lambda\| \leq \|\bar{\lambda}\| + \delta_0$ and, thus, $\|\lambda\| \leq M$.

By (11) and (12), if $\|\lambda - \bar{\lambda}\| \leq \delta_0$ and the iteration is well defined, we have that

$$\|h(z)\| \leq \theta \|h(x)\| + c_4[\|G(x,\lambda)\| + (c_1 + c_2 M)\|h(x)\|]^2$$

and

$$\|G(z,\mu)\| \leq (c_1 + c_2 M)\|h(x)\| + \eta \|G(x,\lambda)\| + c_4[\|G(x,\lambda)\| + (c_1 + c_2 M)\|h(x)\|]^2.$$

So, since the norm $\|\cdot\|_H$ is monotone,

$$\begin{aligned} &\left\| \begin{pmatrix} \|h(z)\| \\ \\ \|G(z,\mu)\| \end{pmatrix} \right\|_H \\ &\leq \left\| H \begin{pmatrix} \|h(x)\| \\ \\ \|G(x,\lambda)\| \end{pmatrix} \right\|_H + c_4 \left\| \begin{pmatrix} [\|G(x,\lambda)\| + (c_1 + c_2 M)\|h(x)\|]^2 \\ \\ [\|G(x,\lambda)\| + (c_1 + c_2 M)\|h(x)\|]^2 \end{pmatrix} \right\|_H \\ &\leq \rho \left\| \begin{pmatrix} \|h(x)\| \\ \\ \|G(x,\lambda)\| \end{pmatrix} \right\|_H + c_4[\|G(x,\lambda)\| + (c_1 + c_2 M)\|h(x)\|]^2 \times \left\| \begin{pmatrix} 1 \\ \\ 1 \end{pmatrix} \right\|_H. \end{aligned}$$

Now, by the equivalence of norms in $I\!R^2$ there exists $\bar{\alpha} > 0$ such that, for all $a, b > 0$,

$$(c_1 + c_2 M)a + b \leq \bar{\alpha} \left\| \begin{pmatrix} a \\ \\ b \end{pmatrix} \right\|_H,$$

so,

$$\left\| \begin{pmatrix} \|h(z)\| \\ \\ \|G(z,\mu)\| \end{pmatrix} \right\|_H \le \rho \left\| \begin{pmatrix} \|h(x)\| \\ \\ \|G(x,\lambda)\| \end{pmatrix} \right\|_H + c_4\bar{\alpha} \left\| \begin{pmatrix} \|h(x)\| \\ \\ \|G(x,\lambda)\| \end{pmatrix} \right\|_H^2 \left\| \begin{pmatrix} 1 \\ \\ 1 \end{pmatrix} \right\|_H .$$

Since $\|h(x)\|$ and $\|G(x,\lambda)\|$ are continuous and vanish at $\bar{x}, \bar{\lambda}$, taking δ_1 and ε_1 small enough, with $\delta_1 \le \delta_0$ we obtain (19).

Now, let us prove (20) and (21). By (13) and (14), if $\|x - \bar{x}\| \le \varepsilon_1$, $\|\lambda - \bar{\lambda}\| \le \delta_1$ and the iteration is well defined,

$$\max\{\|z - x\|, \|\mu - \lambda\|\} \le (c_3 + c_4 M)\|h(x)\| + c\|G(x,\lambda)\|. \tag{22}$$

But, by the equivalence of norms in $I\!R^2$, there exists $\beta > 0$ such that, for all $a, b > 0$,

$$(c_3 + c_4 M)a + cb \le \beta \left\| \begin{pmatrix} a \\ \\ b \end{pmatrix} \right\|_H .$$

Therefore, taking $a = \|h(x)\|$ and $b = \|G(x,\lambda)\|$, (20) and (21) follow from (22). $\square$

From now on, for all $x \in \Omega$ such that $\|x - \bar{x}\| \le \varepsilon_1$ and $\|\lambda - \bar{\lambda}\| \le \delta_1$, we define:

$$R(x,\lambda) = \left\| \begin{pmatrix} \|h(x)\| \\ \\ \|G(x,\lambda)\| \end{pmatrix} \right\|_H .$$

In the next theorem we prove that, if (x_0, λ_0) is close enough to the critical pair $(\bar{x}, \bar{\lambda})$, the sequence generated by $(x_{k+1}, \lambda_{k+1}) = N(x_k, \lambda_k)$ converges to a critical pair. Uniqueness of the critical pair is not assumed. Convergence at a linear rate can take place to a different critical pair than $(\bar{x}, \bar{\lambda})$.

Theorem 2.3. Let $(\bar{x}, \bar{\lambda})$ be a critical pair. Let ρ and r be given by (17) and (18). Assume that $\varepsilon_2 \in (0, \varepsilon_1]$ and $\delta_2 \in (0, \delta_1]$ are such that the IR iteration starting from (x, λ) can be completed whenever $\|x - \bar{x}\| \leq \varepsilon_2$ and $\|\lambda - \bar{\lambda}\| \leq \delta_2$. For all $k = 0, 1, 2, \ldots$, if $\|x_k - \bar{x}\| \leq \varepsilon_2$ and $\|\lambda_k - \bar{\lambda}\| \leq \delta_2$, we define $(x_{k+1}, \lambda_{k+1}) = N(x_k, \lambda_k)$. Then, there exist $\varepsilon_3 \in (0, \varepsilon_2]$, $\delta_3 \in (0, \delta_2]$ such that, taking $\|x_0 - \bar{x}\| \leq \varepsilon_3$ and $\|\lambda_0 - \bar{\lambda}\| \leq \varepsilon_3$ we have:

(a) The whole sequence $\{x_k\}$, $k = 0, 1, 2, \ldots$ is well defined and

$$\|x_k - \bar{x}\| \leq \varepsilon_2, \;\; \|\lambda_k - \bar{\lambda}\| \leq \delta_2 \;\text{ for all }\; k = 0, 1, 2, \ldots \tag{23}$$

(b) $R(x_{k+1}, \lambda_{k+1}) \leq r\, R(x_k, \lambda_k)$ and $R(x_k, \lambda_k) \leq r^k R(x_0, \lambda_0)$ for all $k = 0, 1, 2, \ldots$

(c) The sequence $\{(x_k, \lambda_k)\}$ is convergent to a critical pair (x_*, λ_*).

(d) For all $k = 0, 1, 2, \ldots$,

$$\|x_k - x_*\| \leq \frac{\beta r^k}{1-r} R(x_0, \lambda_0) \quad \text{and} \quad \|\lambda_k - \lambda_*\| \leq \frac{\beta r^k}{1-r} R(x_0, \lambda_0), \tag{24}$$

where $\beta > 0$ is the constant defined in the thesis of Theorem 2.2.

Proof. Define $\Phi(\varepsilon, \delta) = \max\{R(x, \lambda) \mid \|x - \bar{x}\| \leq \varepsilon, \|\lambda - \bar{\lambda}\| \leq \delta\}$. By the continuity of $R(x, \lambda)$ and the fact that $R(\bar{x}, \bar{\lambda}) = 0$, we have that $\lim_{\varepsilon \to 0, \delta \to 0} \Phi(\varepsilon, \delta) = 0$. Let $\varepsilon_3 \leq \varepsilon_2/2$ and $\delta_3 \leq \delta_2/2$ such that $\Phi(\varepsilon_3, \delta_3) \leq R(x_0, \lambda_0)$ and $\beta\Phi(\varepsilon_3, \delta_3)/(1-r) \leq \min\{\varepsilon_2, \delta_2\}/2$. Let $x_0 \in \Omega, \lambda_0 \in I\!R^m$ be such that $\|x_0 - \bar{x}\| \leq \varepsilon_3, \|\lambda_0 - \bar{\lambda}\| \leq \delta_3$. Then,

$$\varepsilon_3 + \frac{\beta R(x_0, \lambda_0)}{1-r} \leq \varepsilon_2 \text{ and } \delta_3 + \frac{\beta R(x_0, \lambda_0)}{1-r} \leq \delta_2. \tag{25}$$

Let us prove by induction on k that x_k, λ_k are well defined,

$$R(x_k, \lambda_k) \leq r^k R(x_0, \lambda_0), \tag{26}$$

$$\|x_k - \bar{x}\| \leq \varepsilon_3 + \beta R(x_0, \lambda_0) \sum_{j=0}^{k-1} r^j, \tag{27}$$

$$\|\lambda_k - \bar{\lambda}\| \leq \delta_3 + \beta R(x_0, \lambda_0) \sum_{j=0}^{k-1} r^j. \tag{28}$$

For $k = 0$, (26), (27) and (28) are obviously true. Assume, as inductive hypothesis, that (26), (27) and (28) hold for some k. Then, by (25), since $\sum_{j=0}^{k-1} r^j \le \sum_{j=0}^{\infty} r^j = \frac{1}{1-r}$ we have that $\|x_k - \bar{x}\| \le \varepsilon_2 \le \varepsilon_1$, $\|\lambda_k - \bar{\lambda}\| \le \delta_2 \le \delta_1$. Therefore, by the hypothesis of the theorem, x_{k+1} and λ_{k+1} are well defined. Then, by (19), $R(x_{k+1}, \lambda_{k+1}) \le rR(x_k, \lambda_k)$. So, by the inductive hypothesis (26), $R(x_{k+1}, \lambda_{k+1}) \le r^{k+1} R(x_0, \lambda_0)$. Now, by Theorem 2.2 and the inductive hypothesis, $\|x_{k+1} - \bar{x}\| \le \|x_k - \bar{x}\| + \|x_{k+1} - x_k\| \le \varepsilon_3 + \beta R(x_0, \lambda_0) \sum_{j=0}^{k-1} r^j + \beta R(x_k, \lambda_k) \le \varepsilon_3 + \beta R(x_0, \lambda_0) \sum_{j=0}^{k-1} r^j + \beta r^k R(x_0, \lambda_0) \le \varepsilon_3 + \beta R(x_0, \lambda_0) \sum_{j=0}^{k} r^j$. Therefore, (27) holds replacing k by $k + 1$. Analogously, we prove that (28) holds replacing k by $k + 1$. So far, the inductive proof is finished. Thus, the sequence is well defined,

$$\|x_k - \bar{x}\| \le \varepsilon_3 + \beta R(x_0, \lambda_0) \sum_{j=0}^{k-1} r^j \le \varepsilon_3 + \frac{\beta R(x_0, \lambda_0)}{1 - r} \le \varepsilon_2 \tag{29}$$

and

$$\|\lambda_k - \bar{\lambda}\| \le \delta_3 + \beta R(x_0, \lambda_0) \sum_{j=0}^{k-1} r^j \le \delta_3 + \frac{\beta R(x_0, \lambda_0)}{1 - r} \le \delta_2 \tag{30}$$

for all $k = 0, 1, 2, \ldots$ Thus, (a) and (b) are proved.

Now, by Theorem 2.2 and (b), for all $k = 0, 1, 2, \ldots$ we have that $\|x_{k+1} - x_k\| \leq \beta R(x_k, \lambda_k) \leq \beta r^k R(x_0, \lambda_0)$ and $\|\lambda_{k+1} - \lambda_k\| \leq \beta R(x_k, \lambda_k) \leq \beta r^k R(x_0, \lambda_0)$. This means that for all $k, j = 0, 1, 2, \ldots$, $\|x_{k+j} - x_k\| \leq \beta(r^k + \ldots + r^{k+j-1}) R(x_0, \lambda_0) \leq \frac{\beta r^k}{1-r} R(x_0, \lambda_0)$ and $\|\lambda_{k+j} - \lambda_k\| \leq \beta(r^k + \ldots + r^{k+j-1}) R(x_0, \lambda_0) \leq \frac{\beta r^k}{1-r} R(x_0, \lambda_0)$. Therefore, $\{x_k\}$ and $\{\lambda_k\}$ are Cauchy sequences, thus convergent to $x_* \in \Omega$ and $\lambda_* \in I\!R^m$ respectively. Taking limits, we have the error estimates $\|x_k - x_*\| \leq \frac{\beta r^k}{1-r} R(x_0, \lambda_0)$ and $\|\lambda_k - \lambda_*\| \leq \frac{\beta r^k}{1-r} R(x_0, \lambda_0)$. From $R(x_k, \lambda_k) \leq r^k R(x_0, \lambda_0)$ and by the continuity of R we obtain that $R(x_*, \lambda_*) = 0$. Therefore, the theorem is proved. □

Remark 2.1. We used the fact that $\|\cdot\|$ is the Euclidian norm in the theorems above because the properties of the projection operator P are part of the proving arguments. In the particular case in which $\Omega = I\!R^n$ the projection P is the identity. In this case, it is easy to see that the results hold for an arbitrary norm.

Theorem 2.4. In addition to the hypotheses of Theorem 2.3, assume that the parameters θ and η depend on k and tend to zero. Then, $R(x_k, \lambda_k)$ tends to zero Q-superlinearly and (x_k, λ_k) tends to (x_*, λ_*) R-superlinearly.

Proof. The fact that $R(x_k, \lambda_k)$ tends to zero Q-superlinearly follows from Part (b) of Theorem 2.3. By (20) and (21) $\|x_{k+1} - x_k\| + \|\lambda_{k+1} - \lambda_k\|$ is bounded by a sequence that tends superlinearly to zero. This implies that $(x_{k+1} - x_*, \lambda_{k+1} - \lambda_*)$ tends R-superlinearly to zero. □

Theorem 2.5. In addition to the hypotheses of Theorem 2.3, assume that $\theta = \eta = 0$. Then, $R(x_k, \lambda_k)$ converges Q-quadratically to zero and the convergence of (x_k, λ_k) to (x_*, λ_*) is R-quadratic.

Proof. From (11) and (12), $R(x_k, \lambda_k)$ tends to zero Q-quadratically. By (20) and (21) $\|x_{k+1} - x_k\| + \|\lambda_{k+1} - \lambda_k\|$ is bounded by a sequence that tends quadratically to zero. This implies that $(x_{k+1} - x_*, \lambda_{k+1} - \lambda_*)$ tends R-quadratically to zero. □

Remark 2.2. Although somewhat cumbersome, it is not difficult to prove that, under some classical assumptions, the hypothesis of Theorem 2.3 holds. The more simple case is when $\bar{x}$ is an interior point of Ω. In this case the critical pair $\bar{x}, \bar{\lambda}$ is a solution of the nonlinear system

$$h(x) = 0, \ \nabla f(x) + \nabla h(x)\lambda = 0.$$

If the Jacobian of this nonlinear system is nonsingular at $(\bar{x}, \bar{\lambda})$, Brent's generalized method (Ref. 20, 21) defines an admissible iteration for constants that only depend on $(\bar{x}, \bar{\lambda})$. The basic properties of this method guarantee that the iteration is well defined in a neighborhood of the critical pair and that the conditions (3)–(7) are satisfied.

The case in which $\bar{x}$ is not interior can be reduced to the interior case after some manipulations assuming nonsingularity of a reduced nonlinear system.

3 Implementation

From now on, we define $\Omega = \{x \in I\!R^n \mid \ell \leq x \leq u\}$. Algorithm 3.1 is an implementation of (3)–(7). Suppose that the initial pair (x_0, λ_0) is given, $x_0 \in \Omega$, as well as the algorithmic parameters $\theta \in [0,1), \eta \in [0,1)$, $K_1, K_2, \tilde{K}_3 > 0$ and $\varepsilon \geq 0$. The algorithm describes the steps to obtain (x_{k+1}, λ_{k+1}) starting from (x_k, λ_k).

Algorithm 3.1.

Step 1. *Feasibility Phase.* Solve, approximately, the minimization problem

$$\min_y \|h(y)\|^2 \ \text{ s.t. } \ \|y - x_k\|_\infty \leq K_1 \|h(x_k)\|, \ y \in \Omega. \tag{31}$$

The approximate solution y_k is asked to satisfy

$$\|h(y_k)\| \leq \max\{\varepsilon, \ \theta \|h(x_k)\|\}. \tag{32}$$

If we are not able to find such an approximate solution, stop the execution declaring *"Failure at the Feasibility Phase"*.

Step 2. *Test Solution.* If $\|h(y_k)\|_\infty \le \varepsilon$ and $\|G(y_k, \lambda_k)\|_\infty \le \varepsilon$, terminate the execution of the algorithm. The pair (y_k, λ_k) is an approximate solution of the problem (exact if $\varepsilon = 0$).

Step 3. *Optimality Phase.* Obtain an approximate solution of

$$\min_z L(z, \lambda_k) \text{ s.t. } h'(y_k)(z - y_k) = 0,\ \|z - y_k\|_\infty \le \tilde{K}_3 \max\{1, \|y_k\|_\infty\},\ z \in \Omega. \quad (33)$$

Let $(\lambda_{k+1} - \lambda_k) \in I\!R^m$ be the vector of Lagrange multipliers associated to the approximate solution x_{k+1} of (33). This approximate solution is asked to satisfy

$$\|h'(y_k)(x_{k+1} - y_k)\| \le \max\{\varepsilon,\ K_2 \|G(y_k, \lambda_k)\|^2\} \quad (34)$$

and

$$\|\tilde{P}[x_{k+1} - \nabla L(x_{k+1}, \lambda_k) - \nabla h(y_k)(\lambda_{k+1} - \lambda_k)] - x_{k+1}\| \le \max\{\varepsilon,\ \eta \|G(y_k, \lambda_k)\|\}, \quad (35)$$

where $\tilde{P}$ is de Euclidian projection operator onto the box $\Omega \cap \{z \in I\!R^n \mid \|z - y_k\|_\infty \le \tilde{K}_3 \max\{1,\ \|y_k\|_\infty\}$. If we are not able to satisfy these requirements we declare *"Failure* *at the Optimality Phase".*

Approximate solutions that satisfy (34)–(35) exist since the feasible set is nonempty and compact. Therefore, the diagnostic of Failure at the Optimality Phase can only represent lack of success of the algorithm used to solve the linearly constrained optimization problem (33). This failure never occurred in our experiments. The situation is somewhat different in the feasibility phase, because in this case it is possible to incorporate the theoretically required steplength control in the definition of the optimization problem (31). In this case, failure might be a characteristic of the problem. For example, *"Failure at the Feasibility Phase"* necessarily occurs if x_k is a global minimizer of (31) where $h(x_k) \neq 0$.

It is well known by practitioners that, in the process of solving nonlinear systems, locally convergent methods can be improved by the simple device of maintaining the distance between consecutive iterates under control. See Ref. 22. This is the role of the constraint $\|z - y_k\|_\infty \leq \tilde{K}_3 \max\{1, \|y_k\|_\infty\}$ in (33).

We used Gencan (Ref. 23) for solving (31) and Algencan, a straightforward Augmented Lagrangian algorithm based on Gencan for solving (33). Very likely, these are not the best choices from the point of view of efficiency, but they serve for the main questions that we want to be answered by the numerical experiments, which are related with robustness. Nevertheless, we would like to mention that in recent works (Ref. 24, 25) excellent numerical behavior of Augmented Lagrangian algorithms applied to linearly constrained minimization has been reported.

4 Numerical Experiments

The question that we want to answer by means of numerical experiments may be formulated as follows: How bad is the local Inexact-Restoration method when compared to globally convergent nonlinear-programming algorithms? The key point is, of course, robustness. The comparison between local and global methods in nonlinear optimization is, sometimes, surprising. As far as in 1979, Moré and Cosnard (Ref. 22) published a numerical study where Brent's method for solving nonlinear systems (Ref. 26, 27) appeared to be better than globally convergent nonlinear solvers when a suitable control for the steplength was used. The analogy between the local Inexact-Restoration method and the generalized Brown-Brent methods (Ref. 20) as well as the natural way in which steplength controls appear in our implementation increases the motivation for the numerical study.

We selected all the nonlinearly constrained problems with quadratic or nonlinear objective function from the CUTE collection (Ref. 28). Implementation details are given in Ref. 29. A comparison against LANCELOT (Ref. 30) is given in Table 1 of Ref. 29. Surprisingly, Algorithm 3.1 was, at least, as robust as LANCELOT for the set of problems considered.

5 Conclusions

Inexact-Restoration algorithms for nonlinear programming are based on inexact achievement of feasibility at each iteration followed by inexact minimization of the Lagrangian on a linear approximation of the constraints. Different methods can be used at both phases of the IR algorithm. In this paper we proved a local convergence result for an Inexact-Restoration algorithm. Essentially, the theorem says that if the IR iteration is well defined in a neighborhood of the solution, then linear convergence takes places to some solution of the KKT system. Under additional assumptions the convergence is superlinear or quadratic.

Based on the fact that, in Newtonian methods for nonlinear systems of equations, practical convergence can be dramatically improved by means of simple steplength control modifications, we proposed a modification of the optimality phase of IR that implicitly maintains the steplength under control. The proposed modification resembles a trust-region constraint added to the natural constraints of the feasibility phase. However, this trust-region constraint is fixed, and not reduced according to merit-function decrease as in Ref. 1.

Numerical experiments showed that the IR algorithm with this simple modification is at least as robust as a well established globally convergent nonlinear programming method in a set of problems taken from the CUTE collection.

The conclusion of this work is not that one should abandon the project of defining algorithms with the best possible convergence theories, including global convergence, but to put in evidence what kind of practical effects one should expect from globally convergent methods (with or without merit functions). It seems that one should be tolerant with local methods that use a lot of information about the true problem, as Inexact Restoration does, and that the main effect of global modifications should be to maintain the steplength under control. Probably, this reinforces the importance of working with filter strategies and with algorithms that do not force merit function decrease at every iteration.

The results of this paper can be straightforwardly extended to the resolution of

KKT systems of type:

$$h(x) = 0,\ x \in \Omega, P(x - F(x) - \nabla h(x)\lambda) - x = 0.$$

Essentially, the modifications in the proof required to consider these systems consist in the judicious replacement of $\nabla f(x)$ by $F(x)$ in the proper places. This opens the path for the application of Inexact Restoration to variational inequalities, equilibrium problems and other extensions of constrained optimization.

References

1. MARTÍNEZ, J.M., *Inexact-Restoration Method with Lagrangian Tangent Decrease and New Merit Function for Nonlinear Programming*, Journal of Optimization Theory and Applications, Vol. 111, pp. 39–58, 2001.

2. MARTÍNEZ, J.M., and PILOTTA, E.A., *Inexact-Restoration Algorithms for Constrained Optimization*, Journal of Optimization Theory and Applications, Vol. 104, pp. 135–163, 2000.

3. MARTÍNEZ, J.M., and PILOTTA, E.A., *Inexact-Restoration Methods for Nonlinear Programming: Advances and Perspectives*, Optimization and Control with Applications, Edited by L.Q. Qi, K.L. Teo and X.Q. Yang, Kluwer Academic Publishers, 2005.

4. ABADIE, J., and CARPENTIER, J., *Generalization of the Wolfe Reduced-Gradient Method to the Case of Nonlinear Constraints*, Optimization, Edited by R. Fletcher, Academic Press, New York, NY, pp. 37–47, 1968.

5. DRUD, A., *CONOPT: A GRG Code for Large Sparse Dynamic Nonlinear Optimization Problems*, Mathematical Programming, Vol. 31, pp. 153–191, 1985.

6. LASDON, L.S., *Reduced Gradient Methods*, Nonlinear Optimization 1981, Edited by M.J.D. Powell, Academic Press, New York, NY, pp. 235–242, 1982.

7. MIELE, A., HUANG, H.Y., and HEIDEMAN, J.C., *Sequential Gradient-Restoration Algorithm for the Minimization of Constrained Functions: Ordinary

and Conjugate Gradient Version, Journal of Optimization Theory and Applications, Vol. 4, pp. 213–246, 1969.

8. MIELE, A., LEVY, A.V., and CRAGG, E.E., *Modifications and Extensions of the Conjugate-Gradient Restoration Algorithm for Mathematical Programming Problems*, Journal of Optimization Theory and Applications, Vol. 7, pp. 450–472, 1971.

9. MIELE, A., SIMS, E.M., and BASAPUR, V.K., *Sequential Gradient-Restoration Algorithm for Mathematical Programming Problems with Inequality Constraints, Part 1, Theory*, Rice University, Aero-Astronautics Report No. 168, 1983.

10. ROSEN, J.B., *The Gradient Projection Method for Nonlinear Programming, Part 1, Linear Constraints*, SIAM Journal on Applied Mathematics, Vol. 8, pp. 181–217, 1960.

11. ROSEN, J.B., *The Gradient Projection Method for Nonlinear Programming, Part 2, Nonlinear Constraints*, SIAM Journal on Applied Mathematics, Vol. 9, pp. 514–532, 1961.

12. ROSEN, J.B., and KREUSER, J., *A Gradient Projection Algorithm for Nonlinear Constraints*, Numerical Methods for Nonlinear Optimization, Edited by F.A. Lootsma, Academic Press, London, UK, pp, 297–300, 1972.

13. GOMES, F.M., MACIEL, M.C., and MARTÍNEZ, J.M., *Nonlinear Programming Algorithms using Trust Regions and Augmented Lagrangians with Nonmonotone Penalty Parameters*, Mathematical Programming, Vol. 84, pp. 161–200, 1999.

14. ROCKAFELLAR, R.T., and WETS, R.J.-B., *Variational Analysis*, Springer-Verlag, Berlin, Germany, 1998.

15. GRIPPO, L., LAMPARIELLO, F., and LUCIDI, S., *A Class of Nonmonotone Stabilization Methods in Unconstrained Optimization*, Numerische Mathematik, Vol. 59, pp. 779–805, 1991.

16. FLETCHER, R., and LEYFFER, S., *Nonlinear Programming without a Penalty Function*, Mathematical Programming, Vol. 91, pp. 239–269, 2002.

17. GONZAGA, C.C., KARAS, E., and VANTI, M., *A Globally Convergent Filter Method for Nonlinear Programming*, SIAM Journal on Optimization, Vol. 14, pp. 646–669, 2003.

18. BIELSCHOWSKY, R.H., *Nonlinear Programming Algorithms with Dynamic Definition of Near-Feasibility: Theory and Implementations*, University of Campinas, PhD Thesis, 1996.

19. BERTSEKAS, D.P., *Nonlinear Programming*, 2nd Edition, Athena Scientific, Belmont, Massachusetts, 1999.

20. MARTÍNEZ, J.M., *Generalization of the Methods of Brent and Brown for Solving Nonlinear Simultaneous Equations*, SIAM Journal on Numerical Analysis, Vol. 16, pp. 434–448, 1979.

21. MARTÍNEZ, J.M., *Solving Nonlinear Simultaneous Equations with a Generalization of Brent's Method*, BIT, Vol. 20, pp. 501–510, 1980.

22. MORÉ, J.J., and COSNARD, M.Y., *Numerical Solution of Nonlinear Equations*, ACM Transactions on Mathematical Software, Vol. 5, pp. 64–85, 1979.

23. BIRGIN, E.G., and MARTÍNEZ, J.M., *Large-Scale Active-Set Box-Constrained Optimization Method with Spectral Projected Gradients*, Computational Optimization and Applications, Vol. 23, pp. 101–125, 2002.

24. DOSTÁL, Z., FRIEDLANDER, A., and SANTOS, S.A., *Augmented Lagrangians with Adaptive Precision Control for Quadratic Programming with Equality Constraints*, Computational Optimization and Applications, Vol. 14, pp. 37–53, 1999.

25. DOSTÁL, Z., FRIEDLANDER, A., and SANTOS, S.A., *Augmented Lagrangians with Adaptive Precision Control for Quadratic Programming with Simple Bounds and Equality Constraints*, SIAM Journal on Optimization, Vol. 13, pp. 1120–1140, 2003.

26. BRENT, R.P., *Some Efficient Algorithms for Solving Systems of Nonlinear Equations*, SIAM Journal on Numerical Analysis, Vol. 10, pp. 327–344, 1973.

27. MORÉ, J.J., and COSNARD, M.Y., *Algorithm 554 - BRENTM: A FORTRAN Subroutine for the Numerical Solution of Systems of Nonlinear Equations*, ACM Transactions on Mathematical Software, Vol. 6, pp. 240–251, 1980.

28. BONGARTZ, I., CONN, A.R., GOULD, N.I.M., and TOINT, Ph.L., CUTE*: Constrained and Unconstrained Testing Environment*, ACM Transactions on Mathematical Software, Vol. 21, pp. 123–160, 1995.

29. BIRGIN, E.G., and MARTÍNEZ, J.M., *Local Convergence of an Inexact-Restoration Method and Numerical Experiments*, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, Technical Report, 2004 (Available in www.ime.unicamp.br/∼martinez).

30. CONN, A.R., GOULD, N.I.M., and TOINT, Ph.L., *A Globally Convergent Augmented Lagrangian Algorithm for Optimization with General Constraints and Simple Bounds*, SIAM Journal on Numerical Analysis, Vol. 28, pp. 545–572, 1991.