A semifeasible trust-region model algorithm for minimization with inequality constraints

José Mario Martínez *

September 26, 1997

Abstract

In a recent paper, we introduced a new algorithm of inexact-restoration type for solving minimization problems with equality constraints and bounds on the variables. See Reference [12]. The procedure to force global convergence of this algorithm makes use of an augmented Lagrangian merit function. Here, the same ideas are applied to problems where inequality constraints are explicitly given. We prove global convergence of the new method but, perhaps surprisingly, the merit function does not allow us to use arbitrary estimates of Lagrange multipliers.

Key words: Nonlinear programming, trust regions, GRG methods, SGRA methods, projected gradients, SQP methods, global convergence.

*Departamento de Matemática Aplicada, IMECC-UNICAMP, CP 6065, 13081-970 Campinas SP, Brazil (martinez@ime.unicamp.br). Work sponsored by FAPESP (Grant 90-3724-6), CNPq and FAEP-UNICAMP.

1 Introduction

We introduce a model algorithm of trust-region type for solving large-scale minimization problems with inequality constraints:

$$\text{Minimize } f(x)$$
$$\text{subject to} \quad C(x) \leq 0, \quad x \in \Omega, \tag{1}$$

where $f : I\!\!R^n \to I\!\!R$ and $C : I\!\!R^n \to I\!\!R^m$ are continuously differentiable and $\Omega \subset I\!\!R^n$ is compact and convex.

Under the general scope of the new model algorithm many specific methods can be defined. Of course, the efficiency of computer implementations highly depends on these particular choices. In this research we concentrate on proving the global convergence properties that depend only on the most general characteristics of the method. Procedures for solving large-scale subproblems required by the model algorithm are available, so the present contribution defines a general framework under which large-scale nonlinear programming problems can be solved.

As usually, the nonlinear programming algorithm introduced in this paper is iterative. It generates a sequence of points $\{x^k\}$ that are feasible with respect to Ω but not necessarily with respect to $C(x) \leq 0$. Given $x^k \in \Omega$ and the trust region radius $\delta_{k,0} > 0$, the algorithm computes a "trial point" $x^{k,0}_{trial} \in \Omega$ such that $\|x^{k,0}_{trial} - x^k\| \leq \delta_{k,0}$. If $x^{k,0}_{trial}$ is "better" than x^k, according to a merit function that takes into account both progress in feasibility and optimality, we define $x^{k+1} = x^{k,0}_{trial}$. Otherwise, the trust-region radius is reduced, and successive trust-region radius $\delta_{k,i}, i = 0, 1, 2, \ldots$ are tested, until a trial point $x^{k,i}_{trial}$ can be accepted.

The computation of $x^{k,i}_{trial}$ is performed in two phases. In the first phase an intermediate point $x^{k,i}_{nor} \in \Omega$ is computed such that $\|x^{k,i}_{nor} - x^k\| \leq 0.8\delta_{k,i}$ and that $x^{k,i}_{nor}$ is "more feasible" than x^k. In the second phase we consider the auxiliary feasible region defined by

$$\pi_{k,i} = \{x \in \Omega \mid C_j(x^{k,i}_{nor}) + C'_j(x^{k,i}_{nor})(x - x^{k,i}_{nor}) \leq \max\{C_j(x^{k,i}_{nor}), 0\} \quad \text{whenever} \quad C_j(x^{k,i}_{nor}) \geq -p\}.$$

By means of the threshold parameter $p > 0$ we establish that constraints that are "strongly satisfied" at $x^{k,i}_{nor}$ does not need to be considered for the definition of the approximate feasible region. The trial point $x^{k,i}_{trial}$ will be a point in $\pi_{k,i}$ such that $\|x^{k,i}_{trial} - x^k\| \leq \delta_{k,i}$ and $f(x^{k,i}_{trial})$ is "sufficiently smaller" than $f(x^{k,i}_{nor})$. Although $x^{k,i}_{nor}$ is "more feasible" than x^k and $x^{k,i}_{trial}$ is "more optimal" than $x^{k,i}_{nor}$, the trial point $x^{k,i}_{trial}$ can be "worse" than x^k considering feasibility and optimality together. To compare $x^{k,i}_{trial}$ and x^k we use a merit function in which both feasibility and optimality are represented.

In classical "feasible" algorithms for nonlinear programming the nonlinear constraints (most times in the form $C(x) = 0$) are satisfied (in practice up to a small tolerance) at all the iterations. Generalized Reduced Gradient (GRG), Sequential Gradient Restoration (SGRA), projected gradient and some interior point algorithms belong to this family of methods. See [1], [2], [10], [11], [13], [14], [15], [19], [20], [21], [22], [23]. Feasible algorithms are very efficient, even for large-scale problems, when the geometry of the constraint set is not severely nonlinear. In fact, in the presence of strong nonlinearity of $C(x)$, the effort of keeping close to the the feasible region is not worthwhile, especially when x^k is far from the solution.

On the other hand, in "nonfeasible" algorithms the solution of the problem is approximated by means of a sequence points x^k that, in general, do not belong to the feasible set. Penalty, Augmented Lagrangian and Sequential Quadratic Programming (SQP) algorithms belong to this class of methods. See [4], [8] and (many) others. The most important drawback of nonfeasible methods is that their overall performance is affected by the behavior of the objective function f at nonfeasible points.

The algorithm presented in this paper is "semifeasible" in the sense that, given the iterate x^k, a definite effort is made in order to improve the "true feasibility" of the approximation. However, neither a fixed, nor a dynamic (in the sense of [2]) level of feasibility is imposed at each iteration and a trial point $x_{nor}^{k,i}$ (hopefully slightly) less feasible than x^k can be accepted if the objective function exhibits sufficient decrease.

In [12] we presented a model algorithm where similar ideas were developed but the constraints have the form $C(x) = 0$. The constraint set of every nonlinear programming problem can be set in the form $\{x \in \Omega \mid C(x) = 0\}$ by means of introduction of slack variables z_i and incorporating constraints of the form $z_i \geq 0$ in the definition of Ω. However, when the original problem has many inequality constraints and (comparatively) a small number of variables, the above transformation can increase the size of the problem in an unacceptable way. So, it makes sense to develop a similar theory to the one of [12] where inequality constraints are explicitly considered. As expected, many proofs in [12] extend immediately to the present work and will not be repeated here. A main difference between the theory developed in this paper and the one of [12] is the definition of $\pi_{k,i}$. In [12] the trial point $x_{trial}^{k,i}$ has exactly the same level of "approximate linear feasibility" as $x_{nor}^{k,i}$. This seems to be adequate for equality constraints but, when inequalities are present, it is better to permit an improvement of the "linear feasibility" of $x_{trial}^{k,i}$ in relation to that of $x_{nor}^{k,i}$. This feature is reflected in the present definition of $\pi_{k,i}$. A surprising fact that derives from this new definition is the impossibility of maintaining *arbitrary* estimates of the Lagrange multipliers at each iteration, as we did in [12]. More than a technical difficulty, this impossibility perhaps reflects an algorithmic advice: to estimate the Lagrange multipliers should be worthwhile only when active constraints at the solution are identified and the problem can be considered as equality constrained.

The transformation of equality constraints into inequalities is always possible, observing that $C_i(x) = 0$ is equivalent to $C_i(x) \leq 0, -C_i(x) \leq 0$. Sometimes, algorithms for inequality constrained problems require the presence of interior points so that the transformation of an inequality into two inequalities is useless. This is not the case of our model algorithm. In fact, it is merely an exercise to assume that the original problem has equality constraints as well as inequalities and define the algorithm for this case without increasing its complexity.

Many elements of the convergence theory developed in this paper and [12] are present in global convergence theories of SQP algorithms, especially the one introduced in [9]. See also [3], [6], [7], [17], [18]. In this paper we assume that Ω is compact, in order to guarantee that many quantities used in the proofs are bounded. Other convergence theories assume the boundedness of those quantities without mentioning the compactness of Ω. However, we think that the boundedness of Ω is the only reasonable assumption *on the problem* that guarantees boundedness of the continuous functions mentioned above. Other general assumptions on the problem are stated in Section 2. In Section 3 we describe the "restoration phase" that allows us to obtain the intermediate more feasible point $x_{nor}^{k,i}$. Some properties of this phase, that are used in the convergence proof, are proved in Section 4. In Section 5 we describe the "minimization phase", by means of which we compute $x_{trial}^{k,i}$. A rigorous description of the model algorithm is given in Section 6. In Sections 7, 8 and 9 we prove, respectively, that the algorithm is well defined, that it finds feasible points and that an optimality condition is eventually satisfied. Conclusions are given in Section 10.

Notation

We will assume that $\|\cdot\|$ denotes an arbitrary norm on $\mathbb{R}^n$ and its associated matricial norm. We assume that $c > 0$ is a constant such that $\|w\| \leq c\|w\|_2$ and $\|w\|_2 \leq c\|w\|$ whenever w is a vector or a matrix.

The $m \times n$ Jacobian matrix of $C(x)$ is denoted $C'(x)$, consequently, $C_j'(x)$ is the row-vector of first derivatives of $C_j(x)$. So, $C_j'(x) = \nabla C_j(x)^T$.

2 General assumptions on the problem

All along this paper, we suppose that the nonlinear programming problem (1) satisfies the assumptions stated in this section.

(i) Ω is compact and convex. (In practice, Ω is usually defined by a set of linear equations or inequations. Many times, Ω is an $n-$dimensional box. However, only the convex-and-compact assumption is needed in the theory.)

(ii) f and C are continuously differentiable on an open set that contains Ω.

(iii) The Jacobian matrix of $C(x)$ satisfies the Lipschitz condition

$$\|C'(y) - C'(x)\| \leq L_1\|y - x\| \text{ for all } x, y \in \Omega. \tag{2}$$

(iv) The gradient of f satisfies the Lipschitz condition

$$\|\nabla f(y) - \nabla f(x)\| \leq L_2\|y - x\| \text{ for all } x, y \in \Omega. \tag{3}$$

For all $x \in \Omega$, $j = 1, \ldots, m$, we define

$$C_j^+(x) = \max\,\{0, C_j(x)\}$$

and

$$C^+(x) = (C_1^+(x), \ldots, C_m^+(x))^T.$$

Given $x \in \Omega$, the function $\tilde{C}_j(x, y)$ is intended to be a first order approximation of $C_j^+(y)$ in a neighborhood of x. This approximation is defined depending of an additional parameter $p > 0$, as follows:

(a) If $C_j(x) < -p$, then $\tilde{C}_j(x, y) = 0$ for all $y \in \Omega$.

(b) If $C_j(x) \geq -p$, then $\tilde{C}_j(x, y) = \max\,\{0, C_j(x) + C_j'(x)(y - x)\}$ for all $y \in \Omega$.

We define $\tilde{C}(x, y) = (\tilde{C}_1(x, y), \ldots, \tilde{C}_m(x, y))^T$ for all $x, y \in \Omega$. Since Ω is compact, $\|\tilde{C}(x, y)\|$ is bounded for all $x, y \in \Omega$.

The compacity of Ω also implies that there exists $\rho > 0$ (not depending on j) such that $\|y - x\| \geq \rho$ whenever $C_j(x) \leq -p$ and $C_j(y) \geq 0$.

Condition (2) implies that there exists $L_3 > 0$ such that

$$\|C_j'(y) - C_j'(x)\| \leq L_3\|y - x\| \text{ for all } x, y \in \Omega, \;\; j = 1, \ldots, m. \tag{4}$$

Consequently,

$$|C_j(y) - C_j(x) - C_j'(x)(y - x)| \leq \frac{L_3}{2}\|y - x\|^2 \text{ for all } x, y \in \Omega, \;\; j = 1, \ldots, m. \tag{5}$$

Clearly, from (3) we also obtain

$$|f(y) - f(x) - \langle \nabla f(x), y - x \rangle| \leq \frac{L_2}{2}\|y - x\|^2 \text{ for all } x, y \in \Omega. \tag{6}$$

Lemma 2.1. *There exists $L_4 > 0$ such that*

$$\|C^+(y) - \tilde{C}(x, y)\| \leq L_4\|y - x\|^2 \textit{ for all } x, y \in \Omega. \tag{7}$$

Proof. It is sufficient to prove that for all $j = 1, \ldots, m$, there exists $L_{4,j} > 0$ such that

$$|C_j^+(y) - \tilde{C}_j(x,y)| \leq L_{4,j}\|y - x\|^2. \tag{8}$$

For proving (8) we consider two cases.

Case 1: $C_j(x) \geq -p$. We need to analyze four situations:

If $C_j(y) \geq 0$ and $C_j(x) + C_j'(x)(y-x) \geq 0$ the inequality (7) follows directly from (5).

If $C_j(y) \geq 0$ and $C_j(x) + C_j'(x)(y-x) < 0$ then

$$|C_j^+(y) - \tilde{C}_j(x,y)| = C_j(y) \leq C_j(y) - C_j(x) - C_j'(x)(y-x) = |C_j(y) - C_j(x) - C_j'(x)(y-x)|$$

So, (8) follows from (5).

If $C_j(y) < 0$ and $C_j(x) + C_j'(x)(y-x) \geq 0$ then

$$|C_j^+(y) - \tilde{C}_j(x,y)| = C_j(x) + C_j'(x)(y-x) \leq -C_j(y) + C_j(x) + C_j'(x)(y-x) = |C_j(y) - C_j(x) - C_j'(x)(y-x)|.$$

So, (8) also follows from (5).

Finally, if $C_j(y) < 0$ and $C_j(x) + C_j'(x)(y-x) < 0$ we have that $C_j^+(y) = \tilde{C}_j(x,y) = 0$ and (8) holds trivially.

Case 2: $C_j(x) < -p$. In this case $\tilde{C}(x,y) = 0$ for all $y \in \Omega$. If $C_j(y) \leq 0$ the inequality (8) holds trivially. But, if $C_j(y) > 0$ it follows that $\|y - x\| \geq \rho$ and (8) follows for a suitable choice of $L_{4,j}$, using the boundedness of $C_j(y)$ in Ω. $\square$

3 Restoration phase

Assume that the current iterate $x^k \in \Omega$ and the current trust-region radius $\delta_{k,i} > 0$ are given. The procedure by means of which we calculate the "more feasible point" $x_{nor}^{k,i}$ is called *Restoration phase* of the model algorithm iteration.

Let us define

$$\varphi(x) = \frac{1}{2}\|C^+(x)\|_2^2$$

for all $x \in \Omega$. If x^k is feasible, we will define $x_{nor}^{k,i} = x^k$. Otherwise, the restoration algorithm will try to find $x_{nor}^{k,i} \in \Omega$ such that $\varphi(x_{nor}^{k,i}) < \varphi(x^k)$ and $\|x_{nor}^{k,i} - x^k\| \leq 0.8\delta_{k,i}$. Observe that the function φ is differentiable and

$$\nabla\varphi(x) = \sum_{i=1}^{m} C_j^+(x)\nabla C_j(x) = C'(x)^T C^+(x) \tag{9}$$

for all $x \in \Omega$. This implies that $\nabla\varphi(x)$ is bounded in Ω. Moreover, by (9), (4) and the boundedness of $C(x)$, we have that there exists $L_5 > 0$ such that

$$\|\nabla\varphi(y) - \nabla\varphi(x)\| \leq L_5\|y - x\| \text{ for all } x, y \in \Omega. \tag{10}$$

This implies that

$$|\varphi(y) - \varphi(x) - \langle\nabla\varphi(x), y - x\rangle| \leq \frac{L_5}{2}\|y - x\|^2 \text{ for all } x, y \in \Omega. \tag{11}$$

Given $\gamma > 0$, a scaling parameter independent of k, the restoration phase is started by means of the computation of d^k_{nor}, given by

$$d^k_{nor} = P(x^k - \gamma \nabla \varphi(x^k)) - x^k, \tag{12}$$

where $P(z)$ is the orthogonal projection of z on Ω. It is easy to see that, when $d^k_{nor} \neq 0$, this vector defines a feasible descent direction of $\varphi(x)$. See Lemma 4.1 of [12]. Let $\tau_1 > 0$ (small), $\beta > 0$ (large) and $r \in (0,1)$ be algorithmic parameters given independently of the iteration index k. The computation of $x^{k,i}_{nor}$ is performed by the following algorithm.

Algorithm 3.1.

Step 1. If $d^k_{nor} = 0$, we set $x^{k,i}_{nor} = x^k$ and the restoration phase terminates.

Step 2. If $d^k_{nor} \neq 0$ we compute

$$t^{k,i}_{break} = \min \{1, \frac{0.8\delta_{k,i}}{\|d^k_{nor}\|}\} \tag{13}$$

and we calculate $t^{k,i}_{dec}$ as the first number t in the sequence $\{t^k_\ell, \ell = 1, 2, \ldots\}$ that verifies

$$\varphi(x^k + t d^k_{nor}) \leq \varphi(x^k) + 0.1 t \langle \nabla \varphi(x^k), d^k_{nor} \rangle, \tag{14}$$

where $t^k_1 = t^{k,i}_{break}$ and $t^k_{\ell+1} \in [0.1 t^k_\ell, 0.9 t^k_\ell]$ for all $\ell = 1, 2, \ldots$

(Since d^k_{nor} is descent direction, we can see that $t^{k,i}_{dec} > 0$ is well defined. See Lemma 4.2 of [12].)

Step 3. If $\|C^+(x^k)\| > \tau_1$, we define $x^{k,i}_{nor}$ as any point in Ω such that $\|x^{k,i}_{nor} - x^k\| \leq 0.8\delta_{k,i}$ and $\varphi(x^{k,i}_{nor}) \leq \varphi(x^k + t^{k,i}_{dec} d^k_{nor})$. (In particular, the choice $x^{k,i}_{nor} = x^k + t^{k,i}_{dec} d^k_{nor}$ is admissible.)

Step 4.

4.1 If $\|C^+(x^k)\| \leq \tau_1$ we try to find $x^k_{inex} \in \Omega$ such that

$$\|\tilde{C}(x^k, x^k_{inex})\|_2 \leq r \|C^+(x^k)\|_2 \text{ and } \|x^k_{inex} - x^k\| \leq \beta \|C^+(x^k)\| \tag{15}$$

4.2 If a point $x^k_{inex} \in \Omega$ satisfying (15) does not exist, the algorithm is stopped.

4.3 If x^k_{inex} exists and $\|x^k_{inex} - x^k\| \leq 0.8\delta_{k,i}$, we set $\sigma_{k,i} = 1$. Else, we set $\sigma_{k,i} = \frac{0.8\delta_{k,i}}{\|x^k_{inex} - x^k\|}$. In both cases, we define $z^{k,i}_{nor} \in \Omega$ as any point that satisfies $\|z^{k,i}_{nor} - x^k\| \leq \min \{0.8\delta_{k,i}, \beta\|C^+(x^k)\|\}$, and

$$\varphi(z^{k,i}_{nor}) \leq \varphi(x^k + \sigma_{k,i}(x^k_{inex} - x^k)). \tag{16}$$

(Obviously, the choice $z^{k,i}_{nor} = x^{k,i}_{nor} + \sigma_{k,i}(x^k_{inex} - x^k)$ is admissible.)

4.4 Finally, we define

$$x^{k,i}_{nor} = \text{Argmin } \{\varphi(z^{k,i}_{nor}), \varphi(x^k + t^{k,i}_{dec} d^k_{nor})\}. \tag{17}$$

Remarks.

Let us comment briefly the main steps of Algorithm 3.1. If, at Step 1, we detect that $d^k_{nor} = 0$, we cannot obtain descent directions using first order information on $\varphi(x)$. So, the restoration phase is terminated with $x^{k,i}_{nor} = x^k$. Hopefully, this means that x^k is feasible ($C^+(x^k) = 0$). However,

x^k can also be a stationary nonfeasible point of $\varphi(x)$. In any case, to define $x_{nor}^{k,i} = x^k$ is the reasonable decision.

By means of the backtracking procedure described at Step 2 we are able to compute a point $x_{nor}^{k,i}$ with sufficient decrease in relation to x^k. If we are not very close to the feasible region, that sufficient decrease is all we need for defining the "more feasible" point. However, we are free to choose a better point $x_{nor}^{k,i}$ using an auxiliar algorithm for minimizing $\varphi(x)$ on Ω.

If x^k is close to the feasible region ($\|C^+(x^k)\| \le \tau_1$), we compute an inexact-Newton type step (see [5]) at (15). If that step does not exist (and r is close to 1) this probably means that we are close to a stationary nonfeasible point of $\varphi(x)$. Probably, x^k is close to a local but nonglobal minimizer of $\varphi(x)$. There is nothing that can be done in this unfortunate situation and, so, the algorithm is stopped. However, if the inexact Newton point x_{inex}^k exists, it will be shown (in Section 4) that a point along the direction determined by this point has good decreasing properties that can be exploited in the convergence proof. Finally, in the case $\|C(x^k)\| \le \tau_1$, the expression (16) allows us to use an auxiliary algorithm for finding a point $x_{nor}^{k,i}$ which could be better than the one determined by the directions d_{nor}^k and $x_{inex}^k - x^k$.

4 Properties of the restoration step

In this section we prove two technical properties of the restoration step $x_{nor}^{k,i} - x^k$ that will be useful for the convergence proofs. This section can be skipped at a first reading without loss of the main arguments of the paper.

Lemma 4.1. *There exists $c_1 > 0$, a constant that only depends on the problem, such that, whenever x^k and $x_{nor}^{k,i}$ are defined, we have that*

$$\|x_{nor}^{k,i} - x^k\| \le c_1 \|C^+(x^k)\|. \tag{18}$$

Proof. The desired result follows trivially from (9), (12), (15), (16) and (17). □

Lemma 4.2. *For all $\alpha > 0$ there exist $c_2 = c_2(\alpha) > 0$ and $\tau_2 = \tau_2(\alpha) > 0$ such that whenever x^k and x_{inex}^k are defined and $\tau_2 \ge \|C^+(x^k)\| \ge \alpha \delta_{k,i}$, the following inequality holds:*

$$\varphi(x^k) - \varphi(x_{nor}^{k,i}) \ge c_2 \|C^+(x^k)\| \delta_{k,i}. \tag{19}$$

Proof. Let us define $s_{inex}^k = x_{inex}^k - x^k$. By (7) we have that, if x_{inex}^k exists and $\|C^+(x^k)\| \le \tau_1$,

$$\|C^+(x_{inex}^k)\|_2 \le \|\tilde{C}(x^k, x_{inex}^k)\|_2 + c^3 L_4 \|s_{inex}^k\|^2.$$

Therefore, by (15),

$$\|C^+(x_{inex}^k)\|_2 \le r\|C^+(x^k)\|_2 + c^3 L_4 \beta^2 \|C^+(x^k)\|^2.$$

So,

$$\|C^+(x_{inex}^k)\|_2^2 \le r^2 \|C^+(x^k)\|_2^2 + (c^3 L_4 \beta^2)^2 \|C^+(x^k)\|^4 + 2rc^3 L_4 \beta^2 \|C^+(x^k)\|^3.$$

This implies that

$$\|C^+(x^k)\|_2^2 - \|C^+(x_{inex}^k)\|_2^2 \ge [(1-r^2) - (c^3 L_4 \beta^2)^2 \|C^+(x^k)\|_2^2 - 2rc^3 L_4 \beta^2 \|C^+(x^k)\|_2] \|C^+(x^k)\|_2^2.$$

Therefore, if $(c^3 L_4 \beta^2)^2 \|C^+(x^k)\|^2 + 2rc^3 L_4 \beta^2 \|C^+(x^k)\|_2 \leq r^2/2$, we obtain

$$\|C^+(x^k)\|_2^2 - \|C^+(x_{inex}^k)\|_2^2 \geq (1 - \frac{r^2}{2})\|C^+(x^k)\|_2^2.$$

In this case, if $\sigma_{k,i} = 1$ and $\|C^+(x^k)\| \geq \alpha\delta_{k,i}$, we have that

$$\|C^+(x^k)\|_2^2 - \|C^+(x_{inex}^k)\|_2^2 \geq \frac{1}{c^2}(1 - \frac{r^2}{2})\|C^+(x^k)\|\alpha\delta_{k,i}. \tag{20}$$

Let us now consider the case $\sigma_{k,i} < 1$. By the convexity of $\|\tilde{C}\|_2^2$ and (15) we have that

$$\begin{aligned}\|\tilde{C}(x^k, x^k + \sigma_{k,i} s_{inex}^k)\|_2^2 &\leq (1 - \sigma_{k,i})\|\tilde{C}(x^k, x^k)\|_2^2 + \sigma_{k,i}\|\tilde{C}(x^k, x_{inex}^k)\|_2^2 \\ &\leq (1 - \sigma_{k,i})\|C^+(x^k)\|_2^2 + \sigma_{k,i} r^2 \|C^+(x^k)\|_2^2.\end{aligned}$$

Therefore,

$$\begin{aligned}&\|C^+(x^k)\|_2^2 - \|\tilde{C}(x^k, x^k + \sigma_{k,i} s_{inex}^k)\|_2^2 \geq \sigma_{k,i}(1 - r^2)\|C^+(x^k)\|_2^2 \\ &= \frac{0.8\delta_{k,i}}{\|s_{inex}^k\|}(1 - r^2)\|C^+(x^k)\|_2^2 \geq \frac{0.8\delta_{k,i}}{\beta\|C^+(x^k)\|}(1 - r^2)\|C^+(x^k)\|_2^2 \\ &\geq \frac{0.8\delta_{k,i}}{\beta c^2}(1 - r^2)\|C^+(x^k)\| = c_1\|C^+(x^k)\|\delta_{k,i},\end{aligned} \tag{21}$$

where $c_1 = 0.8(1 - r^2)/(\beta c^2)$.

Now, by (7),

$$\|C^+(x^k + \sigma_{k,i} s_{inex}^k) - \tilde{C}(x^k, x^k + \sigma_{k,i} s_{inex}^k)\|_2 \leq c^2 L_4 \sigma_{k,i}^2 \|s_{inex}^k\|^2 = 0.64 c^2 L_4 \delta_{k,i}^2. \tag{22}$$

Moreover, since Ω is compact, the definition of $\tilde{C}(x, y)$ and the condition $\|s_{inex}^k\| \leq \beta\|C^+(x^k)\|$ imply that there exists a constant $c_2 > 0$ such that

$$\|\tilde{C}(x^k, x^k + \sigma_{k,i} s_{inex}^k)\|_2 \leq c_2\|C^+(x^k)\|. \tag{23}$$

So, by (21), (22) and (23),

$$\begin{aligned}&\|C^+(x^k)\|_2^2 - \|C^+(x^k + \sigma_{k,i} s_{inex}^k)\|_2^2 \\ &= \|C^+(x^k)\|_2^2 - \|\tilde{C}(x^k, x^k + \sigma_{k,i} s_{inex}^k) + C^+(x^k + \sigma_{k,i} s_{inex}^k) - \tilde{C}(x^k, x^k + \sigma_{k,i} s_{inex}^k)\|_2^2 \\ &\geq \|C^+(x^k)\|_2^2 - \|\tilde{C}(x^k, x^k + \sigma_{k,i} s_{inex}^k)\|_2^2 - \|C^+(x^k + \sigma_{k,i} s_{inex}^k) - \tilde{C}(x^k, x^k + \sigma_{k,i} s_{inex}^k)\|_2^2 \\ &\quad -2\|\tilde{C}(x^k, x^k + \sigma_{k,i} s_{inex}^k)\|_2 \|C^+(x^k + \sigma_{k,i} s_{inex}^k) - \tilde{C}(x^k, x^k + \sigma_{k,i} s_{inex}^k)\|_2 \\ &\geq c_1\|C^+(x^k)\|\delta_{k,i} - (0.64c^2 L_4)^2 \delta_{k,i}^4 - 1.28c^2 c_2 L_4 \delta_{k,i}^2 \|C^+(x^k)\|.\end{aligned}$$

Therefore, if $\|C^+(x^k)\| \geq \alpha\delta_{k,i}$, we have that

$$\|C^+(x^k)\|_2^2 - \|C^+(x^k + \sigma_{k,i} s_{inex}^k)\|_2^2 \geq c_1\|C^+(x^k)\|\delta_{k,i} - c_3\delta_{k,i}(\frac{\|C^+(x^k)\|}{\alpha})^3 - c_4\|C^+(x^k)\|\delta_{k,i}\frac{\|C^+(x^k)\|}{\alpha},$$

where $c_3 = 0.64c^2 L_4$ and $c_4 = 1.28c^2 c_2 L_4$. So,

$$\begin{aligned}\|C^+(x^k)\|_2^2 - \|C^+(x^k + \sigma_{k,i} s_{inex}^k)\|_2^2 &\geq \|C^+(x^k)\|\delta_{k,i}[c_1 - c_3\frac{\|C^+(x^k)\|^2}{\alpha^3} - c_4\frac{\|C^+(x^k)\|}{\alpha}] \\ &\geq \frac{c_1}{2}\|C^+(x^k)\|\delta_{k,i}\end{aligned} \tag{24}$$

whenever $c_3\frac{\|C^+(x^k)\|^2}{\alpha^3} + c_4\frac{\|C^+(x^k)\|}{\alpha} \leq \frac{c_1}{2}$. The desired result follows from (20) and (24). $\square$

5 Minimization phase

In Section 3 we saw that, given the current iterate x^k and the trust-region radius $\delta_{k,i}$, in the restoration phase the "more feasible" point $x_{nor}^{k,i}$ is computed. In the minimization phase we compute $x_{trial}^{k,i}$, a point that should be "more optimal" than $x_{nor}^{k,i}$ (in the sense that $f(x_{trial}^{k,i}) \leq f(x_{nor}^{k,i})$) and where, hopefully, feasibility is not seriously deteriorated. To keep deterioration of feasibility under control, $x_{trial}^{k,i}$ must belong to the set $\pi_{k,i}$, defined by

$$\pi_{k,i} \equiv \{x \in \Omega \quad | \quad \tilde{C}(x_{nor}^{k,i}, x) \leq \tilde{C}(x_{nor}^{k,i}, x_{nor}^{k,i}) \equiv C^+(x_{nor}^{k,i})\}. \tag{25}$$

It is easy to verify that this definition of $\pi_{k,i}$ coincides with the one given in the Introduction. Clearly, $\pi_{k,i}$ is closed and convex. Moreover, if Ω is a polytope, $\pi_{k,i}$ is a polytope as well. Observe that $d_{tan}^{k,i}$ does not depend on $\delta_{k,i}$ if $d_{nor}^k = 0$ ($x_{nor}^{k,i} = x^k$). So, we write $d_{tan}^k = d_{tan}^{k,i}$ in this case.

The minimization phase begins with the computation of the vector $d_{tan}^{k,i}$, given by

$$d_{tan}^{k,i} = P_{k,i}[x_{nor}^{k,i} - \eta \nabla f(x_{nor}^{k,i})] - x_{nor}^{k,i}, \tag{26}$$

where $\eta > 0$ is an arbitrary scaling parameter and, for all $z \in I\!R^n$, $P_{k,i}(z)$ is the orthogonal projection of z on $\pi_{k,i}$. When $\pi_{k,i}$ is a polytope, the computation of $d_{tan}^{k,i}$ involves the solution of a minimum distance problem that must be easy if η is small. The direction $d_{tan}^{k,i}$ plays the role of a feasible Cauchy-type direction for the objective function f on the set $\pi_{k,i}$. Its norm provides us a measure of optimality on this set.

If $d_{nor}^k = 0$ and $d_{tan}^k = 0$ the model algorithm terminates. In this case, we can neither improve feasibility in the restoration phase nor improve optimality in the minimization phase. So, there is nothing that can be done and, hopefully, the current iterate x^k is a solution of our problem. If $d_{nor}^k \neq 0$ and $d_{tan}^{k,i} = 0$ the minimization phase terminates defining $x_{trial}^{k,i} = x_{nor}^{k,i}$. Finally, if $d_{tan}^{k,i} \neq 0$, the trial point $x_{trial}^{k,i} \in \Omega$ is computed using the following algorithm.

Algorithm 5.1.

Step 1. Define

$$\hat{t}_{break}^{k,i} = \max \{\hat{t} \in [0,1] \quad | \quad \|x_{nor}^{k,i} + t d_{tan}^{k,i}\| \leq \delta_{k,i} \quad \text{for all} \quad t \in [0, \hat{t}_k]\}.$$

Step 2. Define $\hat{t}_{dec}^{k,i}$ as the first number t in the sequence $\{\hat{t}_\ell^k, \ell = 1, 2, \ldots\}$ such that

$$f(x_{nor}^{k,i} + t d_{tan}^{k,i}) \leq f(x_{nor}^{k,i}) + 0.1 t \langle d_{tan}^{k,i}, \nabla f(x_{nor}^{k,i}) \rangle, \tag{27}$$

where $\hat{t}_1^k = \hat{t}_{break}^{k,i}$ and $\hat{t}_{\ell+1}^k \in [0.1 \hat{t}_\ell^k, 0.9 \hat{t}_\ell^k]$ for all $\ell = 1, 2, \ldots$.

Step 3. Choose $x_{trial}^{k,i} \in \pi_{k,i}$ such that $\|x_{trial}^{k,i} - x^k\| \leq \delta_{k,i}$ and

$$f(x_{trial}^{k,i}) \leq f(x_{nor}^{k,i} + \hat{t}_{dec}^{k,i} d_{tan}^{k,i}).$$

As so happens to be with the restoration phase, it is easy to see that the minimization phase is well defined. This follows from the fact that $d_{tan}^{k,i}$ is a descent direction for f. See [12] for details.

Remark on the the definition of $\pi_{k,i}$.

The definition of the approximate feasible region $\pi_{k,i}$ is the main difference between the approach of this paper and that of [12]. Roughly speaking, the technique of [12] should lead us to define $\pi_{k,i}$ as the set of points $x \in \Omega$ such that $\tilde{C}(x_{nor}^{k,i}, x) = C^+(x_{nor}^{k,i})$. This definition should be clearly inconvenient for inequality constrained problems due to several reasons. On one hand, $\pi_{k,i}$ should not be a convex set and, so, the computation of the trial point should be much more complicate. On the other hand, by means of (25) we allow the trial point to be "approximately more feasible" than $x_{nor}^{k,i}$, instead of "approximately equally feasible". This seems to be a clear practical advantage. We will see later that, due to this different definition of $\pi_{k,i}$, the "predicted reduction" of the merit function will be, not an approximation of the "actual reduction" but an "sub-approximation" of that quantity. In other words, the predicted reduction will tend to give us a pessimistic (instead of a realistic) estimation of the actual reduction. Fortunately, this is all we need for proving convergence. In spite of the advantages of the definition of $\pi_{k,i}$ given in this paper, there is a reason why we still do not recommend its use for the purely equality constrained case. In fact, the definition of $\pi_{k,i}$ given in [12] allows us to define a merit function where arbitrary estimates of the Lagrange multipliers are explicitly considered. Conversely, when we define $\pi_{k,i}$ as in (25), and we try to use the Lagrangian in the merit function, the predicted reduction ceases to be an approximation of the actual reduction in any sense. As we mentioned in the introduction, this theoretical fact suggest that estimations of Lagrange multipliers should be used only when active constraints are identified, and we can switch from the algorithm defined in this paper to the one introduced in [12]. However, more theoretical and practical research is required on this topic.

6 Description of the model algorithm

Now we are ready to give a short description of the model algorithm introduced in this paper. Recall that $\gamma > 0$ and $\eta > 0$ are given scaling parameters, and that $\tau_1 > 0$, $\beta > 0$ and $r \in (0,1)$ are also given independently of k and are used in the restoration phase or in the minimization phase. The model algorithm generates iterates that are feasible with respect to Ω, so let $x^0 \in \Omega$ be an initial point. At each iteration, penalty parameters $\theta_k \in [0,1]$ will be defined. We define

$$\theta_k^{min} = \min \{1, \theta_0, \ldots, \theta_{k-1}\} \tag{28}$$

and

$$\theta_k^{large} = \min \{1, \theta_k^{min} + \omega_k\}, \tag{29}$$

where $\{\omega_k\}$ is a given sequence of nonnegative numbers such that $\sum_{k=0}^{\infty} \omega_k < \infty$. The steps for computing x^{k+1} or for deciding to terminate the execution of the algorithm are described below.

Algorithm 6.1.

Step 1. *Initialization of iteration k.*

Set $i \leftarrow 0$, choose $\delta_{k,0} \geq \delta_{min}$, and define $\theta_{k,-1} = \theta_k^{large}$.

Step 2. *Restoration phase.*

2.1 Compute d_{nor}^k.

2.2 Run Algorithm 3.1 for computing $x_{nor}^{k,i}$. If Algorithm 3.1 is stopped at Step 4.2, stop the execution of Algorithm 6.1 with a failure message.

Step 3. *Termination criterion and minimization phase.*

3.1 Compute $d_{tan}^{k,i}$ using (26).

3.2 If $d_{nor}^{k} = 0$ and $d_{tan}^{k} = 0$ the model algorithm *terminates* at iteration k with a message of finite convergence.

3.3 If $d_{tan}^{k,i} = 0$ we define $x_{trial}^{k,i} = x_{nor}^{k,i}$.

3.4 If $d_{tan}^{k,i} \neq 0$ we compute $x_{trial}^{k,i}$ using Algorithm 5.1.

Step 4. *Compute the "predicted reduction" and the penalty parameter.*

Define, for all $\theta \in [0,1]$,

$$Pred_{k,i}(\theta) = \theta[f(x^k) - f(x_{trial}^{k,i})] + (1-\theta)[\varphi(x^k) - \varphi(x_{nor}^{k,i})]. \tag{30}$$

Compute $\theta_{k,i}$, the supremum of the values $\theta \in [0, \theta_{k,i-1}]$ such that

$$Pred_{k,i}(\theta) \geq \frac{1}{2}[\varphi(x^k) - \varphi(x_{nor}^{k,i})]. \tag{31}$$

Define

$$\mathbf{Pred}_{k,i} = Pred_{k,i}(\theta_{k,i}). \tag{32}$$

Step 5. *Compute "actual reduction" and test sufficient decrease of the merit function.*

Compute

$$\mathbf{Ared}_{k,i} = \theta_{k,i}[f(x^k) - f(x_{trial}^{k,i})] + (1-\theta_{k,i})[\varphi(x^k) - \varphi(x_{trial}^{k,i})]. \tag{33}$$

If

$$\mathbf{Ared}_{k,i} \geq 0.1\mathbf{Pred}_{k,i}, \tag{34}$$

define

$$x^{k+1} = x_{trial}^{k,i}, \quad \theta_k = \theta_{k,i},$$
$$\delta_k = \delta_{k,i}, \quad iacc(k) = i, \quad \mathbf{Ared}_k = \mathbf{Ared}_{k,i}, \quad \mathbf{Pred}_k = \mathbf{Pred}_{k,i} \tag{35}$$

and finish iteration k.

If (34) does not hold, choose

$$\delta_{k,i+1} \in [0.1\delta_{k,i}, 0.9\delta_{k,i}], \tag{36}$$

set $i \leftarrow i+1$ and go to Step 2.2.

Remarks.

(i) Observe that, from the definition of $\theta_{k,i}$ at Step 7, it follows that $\theta_{k,i} > 0$ for all k, i. A simple proof of this fact can be made by induction. Moreover, it is easy to prove that the sequence $\{\theta_k\}$ is convergent. See [12].

(ii) The definitions (35) are not used in the algorithm. We state them here because they are used in the convergence proofs.

(iii) We defined the penalty parameter $\theta_{k,i}$ at Step 4 as the maximum admissible value of θ such that the predicted reduction *Pred* is sufficiently positive. As we will see in Section 7, the predicted reduction is an approximate underestimation of the actual reduction **Ared**. This property guarantees that, eventually, the sufficient decrease condition (34) holds.

(iv) The sequence $\{\omega_k\}$ can be defined as being identically null. The advantage of defining it in a different way (with convergence of $\sum \omega_k$) is to permit a nondecreasing behavior of the sequence of penalty parameters. The idea is to prevent a premature decrease of the penalty parameter, which can overweight feasibility in relation to optimality.

7 The model algorithm is well defined

In previous sections we saw that, given $x^k \in \Omega$ and $\delta_{k,i} > 0$, either the algorithm terminates at x^k (Steps 2.2 or 3.2 of Algorithm 6.1) or the intermediate and trial points $x_{nor}^{k,i}$ and $x_{trial}^{k,i}$ and the penalty parameter $\theta_{k,i} \in (0,1]$ are well defined. So, we only need to show that, given a point x^k at which the algorithm does not terminate, a trust region radius $\delta_{k,i} > 0$ that satisfies (34) is necessarily found after a finite number of reductions. That is, if Algorithm 6.1 does not terminate at x^k, the new iterate x^{k+1} can be found in finite time.

In Lemma 7.1 we prove that the predicted reduction $\mathbf{Pred}_{k,i}$ is an "approximate underestimation" of the actual reduction $\mathbf{Ared}_{k,i}$. This lemma will be used several times all along the paper. It will be the main argument that will allow us to show that the actual reduction is positive when the predicted reduction is positive and the trust region is sufficiently small.

Lemma 7.1. *There exist a constant $c_4 > 0$ independent of k and i such that, whenever x^k, $x_{nor}^{k,i}$ and $x_{trial}^{k,i}$ are defined, we have that*

$$\varphi(x_{trial}^{k,i}) \le \varphi(x_{nor}^{k,i}) + c_4(\|x_{nor}^{k,i} - x_{trial}^{k,i}\|^4 + \|x_{trial}^{k,i} - x_{nor}^{k,i}\|^2 \|C^+(x^k)\|) \tag{37}$$

and

$$\mathbf{Ared}_{k,i} \ge \mathbf{Pred}_{k,i} - c_4(1-\theta_{k,i})(\|x_{nor}^{k,i} - x_{trial}^{k,i}\|^4 + \|x_{trial}^{k,i} - x_{nor}^{k,i}\|^2 \|C^+(x^k)\|). \tag{38}$$

Proof. Observe that

$$\varphi(x_{trial}^{k,i}) - \varphi(x_{nor}^{k,i}) = \frac{1}{2}\|C^+(x_{trial}^{k,i})\|_2^2 - \frac{1}{2}\|C^+(x_{nor}^{k,i})\|_2^2$$

$$= \frac{1}{2}\|C^+(x_{trial}^{k,i}) - \tilde{C}(x_{nor}^{k,i}, x_{trial}^{k,i}) + \tilde{C}(x_{nor}^{k,i}, x_{trial}^{k,i})\|_2^2 - \frac{1}{2}\|C^+(x_{nor}^{k,i})\|_2^2$$

$$\le \frac{1}{2}[\|C^+(x_{trial}^{k,i}) - \tilde{C}(x_{nor}^{k,i}, x_{trial}^{k,i})\|_2^2 + \|\tilde{C}(x_{nor}^{k,i}, x_{trial}^{k,i})\|_2^2 + 2\|C^+(x_{trial}^{k,i}) - \tilde{C}(x_{nor}^{k,i}, x_{trial}^{k,i})\|_2 \|\tilde{C}(x_{nor}^{k,i}, x_{trial}^{k,i})\|_2]$$

$$-\frac{1}{2}\|C^+(x_{nor}^{k,i})\|_2^2.$$

But, by (25), $0 \le \tilde{C}_j(x_{nor}^{k,i}, x_{trial}^{k,i}) \le C_j^+(x_{nor}^{k,i})$ for all $j = 1, \ldots, m$. Therefore, $\|\tilde{C}(x_{nor}^{k,i}, x_{trial}^{k,i})\|_2 \le \|C^+(x_{nor}^{k,i})\|_2$. So,

$$\varphi(x_{trial}^{k,i}) - \varphi(x_{nor}^{k,i}) \le \frac{1}{2}[\|C^+(x_{trial}^{k,i}) - \tilde{C}(x_{nor}^{k,i}, x_{trial}^{k,i})\|_2^2 + 2\|C^+(x_{trial}^{k,i}) - \tilde{C}(x_{nor}^{k,i}, x_{trial}^{k,i})\|_2 \|C^+(x_{nor}^{k,i})\|_2].$$

So, by Lemma 2.1, the equivalence of norms on $\mathbb{R}^m$ and the fact that $\|C^+(x_{nor}^{k,i})\|_2 \le \|C^+(x^k)\|_2$, we obtain the inequality (37) for a suitable constant $c_4 > 0$. The inequality (38) follows trivially from (37), and the definitions of $\mathbf{Ared}_{k,i}$ and $\mathbf{Pred}_{k,i}$. □

The following two lemmas state that the decrease of φ from x^k to $x_{nor}^{k,i}$ is proportional to $\delta_{k,i}$ if $d_{nor}^k \ne 0$ and that the decrease of f from $x_{nor}^{k,i}$ to $x_{trial}^{k,i}$ is also proportional to the trust region radius if $d_{tan}^{k,i} \ne 0$. Lemma 7.2 can be proved using the arguments of Lemmas 4.1 and 4.3 of [12] and Lemma 7.3 uses, essentially, the ideas of Lemmas 4.4 and 4.6 of [12]. By this reason their proofs are omitted here.

Lemma 7.2. *Assume that* $d^k_{nor} \neq 0$. *Then,* $\langle d^k_{nor}, \nabla\varphi(x^k)\rangle < 0$ *and there exists* $\bar{t}_k > 0$ *such that*

$$\varphi(x^{k,i}_{nor}) \leq \varphi(x^k) + \min\{0.1, 0.08\frac{\delta_{k,i}}{\|d^k_{nor}\|}, 0.01\bar{t}_k\}\langle d^k_{nor}, \nabla\varphi(x^k)\rangle.$$

Lemma 7.3. *Assume that* $d^{k,i}_{tan} \neq 0$. *Then* $\langle d^{k,i}_{tan}, \nabla f(x^{k,i}_{nor})\rangle < 0$ *and there exists* $\tilde{t}_k > 0$ *such that*

$$f(x^{k,i}_{trial}) \leq f(x^k) + \min\{0.1, 0.1\frac{\delta_{k,i}}{\|d^k_{tan}\|}, 0.01\tilde{t}_k\}\langle d^k_{tan}, \nabla f(x^k)\rangle$$

Now we are ready to prove that (34) holds after a finite number of reductions of $\delta_{k,i}$. In Lemma 7.4 we prove that this is true when $d^k_{nor} \neq 0$ and in Lemma 7.5 we prove that the same result holds when $d^k_{nor} = 0$ but $d^k_{tan} \neq 0$.

Lemma 7.4. *Assume that* $d^k_{nor} \neq 0$. *Then, after a finite number of reductions of the trust region radius at (36), we obtain* $i \geq 0$ *such that (34) holds.*

Proof. By (38) and the boundedness of Ω we have that there exists $c_5 > 0$, independent of k and i, such that

$$\mathbf{Ared}_{k,i} \geq \mathbf{Pred}_{k,i} - c_5\delta^2_{k,i}. \tag{39}$$

On the other hand, by Lemma 7.2 and (31),

$$\mathbf{Pred}_{k,i} \geq \frac{\varphi(x^k) - \varphi(x^{k,i}_{nor})}{2} \geq -\frac{1}{2}\min\{0.1, 0.08\frac{\delta_{k,i}}{\|d^k_{nor}\|}, 0.01\bar{t}_k\}\langle d^k_{nor}, \nabla\varphi(x^k)\rangle > 0.$$

So, if $\frac{0.08\delta_{k,i}}{\|d^k_{nor}\|} \leq \min\{0.1, 0.01\bar{t}_k\}$, we have that

$$\mathbf{Pred}_{k,i} \geq c_k\delta_{k,i} \tag{40}$$

where $c_k = -\frac{1}{2}0.08\frac{\delta_{k,i}}{\|d^k_{nor}\|}\langle d^k_{nor}, \nabla\varphi(x^k)\rangle > 0$.

So, by (39) and (40), we have that, when $\frac{0.08\delta_{k,i}}{\|d^k_{nor}\|} \leq \min\{0.1, 0.01\bar{t}_k\}$,

$$\mathbf{Ared}_{k,i} \geq \mathbf{Pred}_{k,i} - \frac{c_5}{c_k}\mathbf{Pred}_{k,i}\delta_{k,i} = (1 - \frac{c_5}{c_k}\delta_{k,i})\mathbf{Pred}_{k,i}.$$

Therefore, (34) holds if $\delta_{k,i} \leq \frac{0.9c_k}{c_5}$. $\square$

Lemma 7.3. *Assume that* $d^k_{nor} = 0$ *and* $d^k_{tan} \neq 0$. *Then, after a finite number of reductions of the trust region radius at (36), we obtain* $i \geq 0$ *such that (34) holds.*

Proof. Since $d^k_{nor} = 0$ we have that $x^{k,i}_{nor} = x^k$. Therefore, by Lemma 7.3, the inequality (31) holds for all $\theta \in (0,1]$. This implies that $\theta_{k,i} = \theta_{k,i-1} = \theta_{k,-1}$ for all i. Thus, by Lemma 7.3 and the definition of *Pred*, there exists $c_k > 0$ (independent of i) such that, when $0.1\frac{\delta_{k,i}}{\|d^k_{tan}\|} \leq \min\{0.1, 0.01\tilde{t}_k\}$,

$$\mathbf{Pred}_{k,i} \geq c_k\delta_{k,i}.$$

But, by Lemma 7.1, the inequality (39) holds, so the proof can be completed using the same argument as in Lemma 7.4. $\square$

8 The feasibility indicator $\|d_{nor}^k\|$ tends to 0

The vector d_{nor}^k, computed at Step 2.1 of Algorithm 6.1, plays the role of a direction of maximum decrease for the feasibility objective function $\varphi(x)$. The results proved in previous sections show that d_{nor}^k is null when x^k is feasible, or even when x^k is a local minimizer of φ in Ω. By these reasons we say that $\|d_{nor}^k\|$ is a "feasibility indicator" of x^k. Clearly, a desirable characteristic of our algorithm is that $\|d_{nor}^k\|$ can be made as small as desired. Essentially, this is what we are going to prove in this section. Namely, we will show that if an infinite sequence $\{x^k\}$ is generated by Algorithm 6.1 (that is, if the algorithm does not stop at Step 2.2 or Step 3.1) the corresponding sequence of vectors $\{d_{nor}^k\}$ is such that $\lim_{k\to\infty} \|d_{nor}^k\| = 0$. In Assumption C1 we suppose that the opposite situation holds, that is, that there exists a subsequence of $\{x^k\}$ for which the quantities $\{\|d_{nor}^k\|\}$ are bounded away from zero. The foregoing results will show that this assumption leads to a contradiction.

Assumption C1. *Let $\{x^k\}$ an infinite sequence generated by Algorithm 3.1. There exists K_1, an infinite subset of $\{0, 1, 2, \ldots\}$, and $\varepsilon > 0$ such that*

$$\|d_{nor}^k\| \geq \varepsilon \quad \textit{for all} \quad k \in K_1.$$

Lemma 8.1. *If Assumption C1 holds, there exists $c_6 > 0$ such that*

$$\langle d_{nor}^k, \nabla\varphi(x^k)\rangle \leq -c_6. \tag{41}$$

for all $k \in K_1$.

Proof. See the proof of Lemma 5.2 of [12]. □

Lemma 8.2. *Suppose that Assumption C1 holds. Then there exists $\bar{t} > 0$ such that*

$$\varphi(x_{nor}^{k,i}) \leq \varphi(x^k) - c_6 \min\,\{0.1, 0.08\frac{\delta_{k,i}}{\varepsilon}, 0.01\bar{t}\,\}. \tag{42}$$

for all $k \in K_1$, $i = 0, 1, \ldots, iacc(k)$, where $c_6 > 0$ is given in Lemma 8.1.

Proof. See the proof of Lemma 5.3 of [12] □

In the following lemma, we prove that under Assumption C1 the subsequence of trust-region radius must be bounded away from zero.

Lemma 8.3. *Suppose that Assumption C1 holds. Then, there exists $\bar{\delta} > 0$ such that $\delta_k \geq \bar{\delta}$ for all $k \in K_1$.*

Proof. By Lemma 8.2 and (31), if $k \in K_1$ and $\frac{0.08\delta_{k,i}}{\varepsilon} \leq \min\,\{0.1, 0.01\bar{t}\,\}$ we have that

$$\mathbf{Pred}_{k,i} \geq \frac{\varphi(x^k) - \varphi(x_{nor}^{k,i})}{2} \geq \frac{0.04c_6\delta_{k,i}}{\varepsilon}.$$

But, as in Lemma 7.4, we have that

$$\mathbf{Ared}_{k,i} \geq \mathbf{Pred}_{k,i} - c_5\delta_{k,i}^2.$$

So, if $\frac{0.08\delta_{k,i}}{\varepsilon} \leq \min\{0.1, 0.01\bar{t}\}$ we obtain

$$\mathbf{Ared}_{k,i} \geq \mathbf{Pred}_{k,i} - c_5\delta_{k,i}\frac{\varepsilon}{0.04c_6}\mathbf{Pred}_{k,i} = (1 - \frac{c_5\varepsilon}{0.04c_6}\delta_{k,i})\mathbf{Pred}_{k,i}.$$

Therefore, if $\frac{0.08\delta_{k,i}}{\varepsilon} \leq \min\{0.1, 0.01\bar{t}\}$ and $\frac{c_5\varepsilon\delta_{k,i}}{0.04c_6} \leq 0.9$, we have that (34) holds. Therefore, the trust region radius $\delta_{k,i}$ is accepted if it is small enough. This completes the proof of the desired result. □

Finally, Theorem 8.4 shows that Assumption C1 cannot hold. The proof is entirely analogous to the proof of Theorem 5.5 of [12]. Since, under Assumption C1, $\delta_{k,i}$ is bounded away from zero for $k \in K_1$, it follows from Lemma 7.1 that $\mathbf{Ared}_{k,i}$ is bounded away from zero. This contradicts the compacity of Ω.

Theorem 8.4. *If $\{x^k\}$ is an infinite sequence generated by Algorithm 3.1, we have that*

$$\lim_{k\to\infty} d^k_{nor} = 0. \tag{43}$$

Proof. Repeat the proof of Theorem 5.5 of [12] replacing ℓ_k by $f(x^k)$. □.

9 The optimality indicator $\|d^{k,i}_{tan}\|$ tends to 0

Up to now we proved that a sequence $\{x^k\}$ generated by Algorithm 6.1 either terminates at Step 2.2 or Step 3.1 in a finite number of iterations or satisfies $\lim_{k\to\infty}\|d^k_{nor}\| = 0$. Hopefully, this means that we are getting arbitrarily close to the feasible region. However, the possibility of obtaining a sequence such that $d^k_{nor} \to 0$ with $\|C^+(x^k)\|$ bounded away from zero cannot be excluded. In fact, we proved that this necessarily happens when an infinite sequence is generated and the feasible region of problem (1) is empty. Unhappily, in most cases it is impossible to recognize if the situation "$\|d^k_{nor}\|$ small and $\|C^+(x^k)\|$ large" corresponds to the detection of an infeasible problem or to the approximation to a local-nonglobal minimizer of $\varphi(x)$. So, there is nothing that can be done in this case, except to use global optimization of heuristic techniques which, on the other hand, are not applicable to problems with the generality of the ones considered in this paper.

Therefore, in this section we are interested in the case in which an infinite sequence is generated and $d^k_{nor} \to 0$ implies $C^+(x^k) \to 0$. We wish to investigate the behavior of the "tangent directions" $d^{k,i}_{tan}$ which, roughly speaking, can be interpreted as maximum decrease directions of $f(x)$ in the tangent space that passes through $x^{k,i}_{nor}$. Clearly, a method that aims to solve (1) should satisfy, in some sense, $d^{k,i}_{tan} \to 0$. In other words, a practical convergence stopping criterion for Algorithm 6.1 should be

$$\|C^+(x^k)\| \leq \varepsilon_1 \quad \text{and} \quad \|d^{k,i}_{tan}\| \leq \varepsilon_2, \tag{44}$$

for suitable small numbers $\varepsilon_1, \varepsilon_2 > 0$. We hope that, when (44) takes place, x^k is close both to feasibility and optimality. In this section we prove that the criterion (44) is eventually satisfied by x^k for all $\varepsilon_1, \varepsilon_2 > 0$.

The technique of the proof is similar to the one used in Section 8. In Assumption C2 we suppose that there exists $\varepsilon > 0$ such that the criterion $\|d^{k,i}_{tan}\| \geq \varepsilon$ is not satisfied for all $k \geq k_0$. After deriving several consequences of this fact, we arrive to a contradiction.

Assumption C2. *Let $\{x^k\}$ an infinite sequence generated by Algorithm 6.1. Assume that*

$$\lim_{k\to\infty} C^+(x^k) = 0 \tag{45}$$

and there exists $k_0 \in \{0,1,2,\ldots\}$, $\varepsilon > 0$ such that

$$\|d_{tan}^{k,i}\| \geq \varepsilon \tag{46}$$

for all $k \geq k_0$, $i = 0,1,\ldots,iacc(k)$.

In Lemmas 9.1–9.3 it is proved that, under Assumption C2, the decrease of $f(x)$, from $x_{nor}^{k,i}$ to $x_{trial}^{k,i}$, is proportional to $\delta_{k,i}$.

Lemma 9.1. *Suppose that Assumption C2 holds. Then, there exists $c_7 > 0$ such that*

$$\langle d_{tan}^{k,i}, \nabla f(x_{nor}^{k,i})\rangle \leq -c_7 \tag{47}$$

for all $k \geq k_0$, $i = 0,1,\ldots,iacc(k)$

Proof. See the proof of Lemma 6.1 of [12]. □

Lemma 9.2. *Suppose that Assumptions C2 holds. Then, there exists $\tilde{t} > 0$ such that*

$$f(x_{trial}^{k,i}) \leq f(x_{nor}^{k,i}) - \min\{0.1, 0.01\tilde{t}, 0.02\frac{\delta_{k,i}}{\varepsilon}\}c_7. \tag{48}$$

for all $k \geq k_0$, $i = 0,1,\ldots,iacc(k)$.

Proof. The proof of this result is analogous to the proof of Lemma 6.2 of [12], replacing ℓ by f and using (6). □

Lemma 9.3. *Suppose that Assumption C2 holds. Then, there exist $c_8 > 0$, $\tilde{\delta} > 0$ such that*

$$f(x_{nor}^{k,i}) - f(x_{trial}^{k,i}) \geq c_8 \min\{\tilde{\delta}, \delta_{k,i}\} \tag{49}$$

for all $k \geq k_0, i = 0,1,\ldots,iacc(k)$.

Proof. See the proof of Corollary 6.3 of [12]. □

The second term of the predicted reduction $\mathbf{Pred}_{k,i}(\theta)$, defined in (30) is always nonnegative due to the definition of the restoration phase. The first term $\theta[f(x^k) - f(x_{trial}^{k,i})]$ can be negative, however it can be decomposed as

$$\theta[f(x^k) - f(x_{trial}^{k,i})] = \theta[f(x^k) - f(x_{nor}^{k,i})] + \theta[f(x_{nor}^{k,i}) - f(x_{trial}^{k,i})] \tag{50}$$

In the expression above only the term $\theta[f(x^k) - f(x_{nor}^{k,i})]$ can be negative. However, by Lemma 4.1, this term is bounded by a quantity of the same order of $\|\theta C^+(x^k)\|$. Moreover, under Assumption C2, Lemma 9.3 says that the second term on the right hand side of (50) is positive and proportional to $\theta\delta_{k,i}$. It turns out that $\theta[f(x^k) - f(x_{trial}^{k,i})]$ is positive, independently of θ, when $\|C^+(x^k)\|$ is less than a multiple of $\delta_{k,i}$. In Lemma 9.4 it is shown that, under those conditions, it is not necessary to increase the value of the penalty parameter.

Lemma 9.4. *Suppose that Assumption C2 holds. Then, there exist $c_9, c_{10}, c_{11}, \alpha > 0$ such that for all $k \geq k_0, i = 0, 1, \ldots, iacc(k)$, $\theta \in [0, 1]$, we have that*

$$Pred_{k,i}(\theta) - \frac{1}{2}[\varphi(x^k) - \varphi(x^{k,i}_{nor}] \geq \theta[c_8 \min\{\tilde{\delta}, \delta_{k,i}\} - c_9\|C^+(x^k)\|] - c_{10}\|C^+(x^k)\|$$

and

$$Pred_{k,i}(\theta) \geq \theta c_8 \min\{\tilde{\delta}, \delta_{k,i}\} - c_{11}\|C^+(x^k)\| \tag{51}$$

where $\tilde{\delta}$ and c_8 are defined in Lemma 6.3.

Moreover, there exists $k_1 \geq k_0$ such that, whenever $\|C^+(x^k)\| \leq \alpha\delta_{k,i}$, we have

$$Pred_{k,i}(\theta) \geq \frac{1}{2}[\varphi(x^k) - \varphi(x^{k,i}_{nor})]$$

for all $\theta \in [0, 1]$ and

$$\theta_{k,i} = \theta_{k,j} = \theta_{k,-1}$$

for all $k \geq k_1$, $i = 0, 1, \ldots, iacc(k)$, $j = 0, 1, \ldots, i-1$, $\theta \in [0, 1]$.

Proof. Using Lemma 4.1, (11) and replacing ℓ by f at the proper places, the proof of this lemma is entirely analogous to the proof of Lemma 6.4 of [12]. □

In the next Lemma we show that the penalty parameter goes to zero. We wander that this property is obtained under Assumption C2 which, in turn, it will be shown to be false. Therefore, we are not proving a property of the sequences generated by the model algorithm but a consequence of an assumption that cannot be true.

Lemma 9.5. *Suppose that Assumption C2 holds. Then*

$$\lim_{k\to\infty} \theta_k = 0.$$

Proof. By Lemma 7.1 and the boundedness of Ω there exists $c_{12} > 0$ such that

$$\mathbf{Ared}_{k,i} \geq \mathbf{Pred}_{k,i} - c_{12}(\delta^2_{k,i} + \delta^4_{k,i}) \tag{52}$$

This implies that, for all $\bar{\theta}, \tilde{\delta} > 0$ there exists $\bar{\delta} \in (0, \min\{\delta_{min}, \tilde{\delta}\})$ such that, whenever $\delta_{k,i} \leq \bar{\delta}$,

$$\mathbf{Ared}_{k,i} \geq \mathbf{Pred}_{k,i} - \frac{\bar{\theta} c_8 \bar{\delta}}{40}. \tag{53}$$

The proof of the lemma continues as the proof of Lemma 6.5 of [12] using C^+ instead of C and inequalities (52) and (53) instead of (64) and (65) respectively. □

Theorem 9.6 is the final convergence result of this paper. We show that, if $\|d^k_{nor}\|$ is a good measure of feasibility, Assumption C2 must be false. As a consequence, a convergence criterion based on the size of $C^+(x^k)$ and $d^{k,i}_{tan}$ will necessarily be satisfied after a finite number of iterations.

Theorem 9.6. *Suppose that Algorithm 6.1 generates an infinite sequence such that $\|C^+(x^k)\| \to 0$ and that, for all $k \in \{0, 1, 2, \ldots\}$, $d^k_{nor} = 0$ only if $C^+(x^k) = 0$. Then, there exists K_1 an infinite subset of $\{0, 1, 2, \ldots\}$ such that for all $k \in K_1$ there exists $i(k) \in \{0, 1, \ldots, iacc(k)\}$ such that*

$$\lim_{k \in K_1} \|d^{k,i(k)}_{tan}\| = 0.$$

Proof. Without loss of generality, assume that $\|C^+(x^k)\| \leq \tau_1$ for all $k \in \{0,1,2,\ldots\}$. Suppose that Assumption C2 holds. By Lemma 4.2, there exists $c_{12} > 0, \tau_3 \in (0, \tau_1)$ such that, whenever $\tau_3 \geq \|C^+(x^k)\| \geq \frac{\alpha}{10}\delta_{k,i}$,

$$\varphi(x^k) - \varphi(x^{k,i}_{nor}) \geq c_{12}\|C^+(x^k)\|\delta_{k,i},$$

where $\alpha > 0$ is the constant defined in Lemma 9.4.

This implies, by (31), that there exists $k_2 \geq k_1$ such that

$$\mathbf{Pred}_{k,i} \geq \frac{1}{2}[\varphi(x^k) - \varphi(x^{k,i}_{nor})] \geq \frac{1}{2}c_{12}\|C^+(x^k)\|\delta_{k,i}$$

whenever $k \geq k_2$ and $\|C^+(x^k)\| \geq \frac{\alpha}{10}\delta_{k,i}$.

Now, by Lemma 7.1,

$$\mathbf{Ared}_{k,i} \geq \mathbf{Pred}_{k,i} - c_4(\delta^3_{k,i} + 2\delta_{k,i}\|C^+(x^k)\|)\delta_{k,i}.$$

So, if $k \geq k_2$ and $\|C^+(x^k)\| \geq \frac{\alpha}{10}\delta_{k,i}$,

$$\mathbf{Ared}_{k,i} \geq \mathbf{Pred}_{k,i} - c_4[(\frac{10}{\alpha}\|C^+(x^k)\|)^3 + \frac{20}{\alpha}\|C^+(x^k)\|^2]\delta_{k,i}$$

$$\geq \mathbf{Pred}_{k,i} - c_4[(\frac{10}{\alpha}\|C^+(x^k)\|)^3 + \frac{20}{\alpha}\|C^+(x^k)\|^2]\frac{2\mathbf{Pred}_{k,i}}{c_{12}\|C^+(x^k)\|}$$

$$= \mathbf{Pred}_{k,i} - \frac{2c_4}{c_{12}}[(\frac{10}{\alpha})^3\|C^+(x^k)\|^2 + \frac{20}{\alpha}\|C^+(x^k)\|]\mathbf{Pred}_{k,i}.$$

Therefore, since $C^+(x^k) \to 0$, there exists $k_3 \geq k_2$ such that, whenever $k \geq k_3$ and $\|C^+(x^k)\| \geq \frac{\alpha}{10}\delta_{k,i}$, the test (34) is satisfied.

Let $k_4 \geq k_3$ be such that $\frac{10\|C^+(x^k)\|}{\alpha} \leq \delta_{min}$ for all $k \geq k_4$. If, for some $k \geq k_4$, we have that $i = iacc(k)$ and $\delta_{k,i} \leq \frac{\|C^+(x^k)\|}{\alpha}$, then $i \geq 1$ and $\delta_{k,i-1} \leq 10\frac{\|C^+(x^k)\|}{\alpha}$. So, (34) is satisfied by $\delta_{k,i-1}$ and $iacc(k)$ should be $i-1$, This implies that, for all $k \geq k_4$, $\delta_{k,iacc(k)} \geq \frac{\|C^+(x^k)\|}{\alpha}$. So, $\delta_{k,i} \geq \frac{\|C^+(x^k)\|}{\alpha}$ for all $i = 0,1,\ldots,iacc(k)$. Therefore, by Lemma 9.4, $\theta_{k,i} \geq \theta_{k,-1}$ for all $k \geq k_4$, $i = 0,1,\ldots,iacc(k)$ This contradicts the thesis of Lemma 9.5. The contradiction comes from assuming that Assumption C2 holds. Thus, we have proved that Assumption C2 is impossible, which implies the desired result. □

10 Conclusions

Algorithm 6.1 is a model algorithm applicable to large scale nonlinear programming problems of the form (1). When equality constraints are present in the original formulation of the problem, its transformation into two inequality constraints does not increase the complexity of the algorithm. The model algorithm permits a large amount of freedom in the choice of efficient methods for effectively computing the intermediate point $x^{k,i}_{nor}$ and the trial point $x^{k,i}_{trial}$. The efficiency of computer implementations should be mostly dependent from the particular algorithms chosen for computing those points.

This research complements the one reported in [12], where only equality constrained are considered and possible inequalities are reduced to equalities and bounds. As we mentioned before, the transformation that reduces inequalities to equalities can be inefficient in critical cases. However, it is interesting to observe that in the case of [12] we are able to use arbitrary Lagrange multiplier estimates in the definition of the merit function used at each iteration. We cannot complete the convergence theory using estimates of the multipliers in the inequality constraint case, at least using the techniques presented in this paper. If this is not a merely technical difficulty, it can be conjectured that the algorithm introduced in [12] must be used for solving (1) when we it is possible to identify nonlinear inequalities that are active at the solution.

The role of model algorithms in optimization research is to provide general frameworks under which specific methods with global convergence properties can be defined. Most future research in the area considered in this paper will concern the implementation of these particular methods, and the analysis of their further properties, such as local convergence.

Acknowledgement. The author is indebted to Lucio T. Santos for his comments on the first draft of this paper.

References

[1] J. Abadie and J. Carpentier, *Generalization of the Wolfe reduced-gradient method to the case of nonlinear constraints*, Optimization, Edited by R. Fletcher, Academic Press, New York, pp. 37-47, 1968.

[2] R. H. Bielschowsky, *Nonlinear programming algorithms with dynamic definition of near-feasibility: theory and implementations*, Tese de Doutorado, IMECC-UNICAMP, Campinas, 1996.

[3] M. R. Celis, J. E. Dennis and R. A. Tapia, *A trust-region strategy for nonlinear equality constrained optimization*, Numerical Optimization, Edited by P. Boggs, R. Byrd and R. Schnabel, SIAM Publications, Philadelphia, Pennsylvania, pp. 71-82, 1985.

[4] A. R. Conn, N. I. M. Gould and Ph. L. Toint, *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.

[5] R. S. Dembo, S. C. Eisenstat and T. Steihaug, *Inexact Newton methods*, SIAM Journal on Numerical Analysis, Vol. 19, pp. 400-408, 1982.

[6] J. E. Dennis, M. El-Alem and M. C. Maciel, *A global convergence theory for general trust-region-based algorithms for equality constrained optimization*, to appear in SIAM Journal on Optimization.

[7] M. El-Alem, *A robust trust region algorithm with a nonmonotonic penalty parameter scheme for constrained optimization*, SIAM Journal on Optimization, Vol. 5, pp. 348-378, 1995.

[8] P. E. Gill, W. Murray and M. H. Wright, *Practical optimization*, Academic Press, London, England, 1981.

[9] F. M. Gomes, M. C. Maciel and J. M. Martínez, *Nonlinear programming algorithms using trust regions and augmented Lagrangians with nonmonotone penalty parameters*, Relatório Técnico, IMECC-UNICAMP, Campinas, 1995.

[10] J. Herskovits, *An interior point technique for nonlinear optimization*, Rapport de Recherche 1808, INRIA, France, 1992.

[11] L. S. Lasdon, *Reduced gradient methods*, in *Nonlinear Optimization 1981*, edited by M. J. D. Powell, Academic Press, New York, pp. 235-242, 1982.

[12] J. M. Martínez, *A two-phase trust-region model algorithm with global convergence for nonlinear programming*, to appear in *Journal of Optimization Theory and Applications*, Vol. 96, Number 2, Feb. 1998.

[13] A. Miele, H. Y. Huang and J. C. Heideman, *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.

[14] A. Miele, A. V. Levy and E. E. Cragg, *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.

[15] A. Miele, E. M. Sims and V. K. Basapur, *Sequential Gradient-Restoration algorithm for mathematical programming problem with inequality constraints, Part 1, Theory*, Rice University, Aero-Astronautics Report No. 168, 1983.

[16] R. B. Murtagh and M. A. Saunders, *Large-scale linearly constrained optimization*, Mathematical Programming, Vol. 14 , pp. 41-72, 1978.

[17] E. Omojokun, *Trust-region strategies for optimization with nonlinear equality and inequality constraints*, PhD Thesis, Department of Computer Science, University of Colorado, Boulder, Colorado, 1989.

[18] M. J. D. Powell and Y. Yuan, *A trust-region algorithm for equality constrained optimization*, Mathematical Programming, Vol. 49, pp. 190-211, 1991.

[19] M. Rom and M. Avriel, *Properties of the sequential gradient-restoration algorithm (SGRA), Part 1: Introduction and comparison with related methods*, Journal of Optimization Theory and Applications, Vol. 62, pp. 77-98, 1989.

[20] M. Rom and M. Avriel, *Properties of the sequential gradient-restoration algorithm (SGRA), Part 2: Convergence Analysis*, Journal of Optimization Theory and Applications, Vol. 62, pp. 99-126, 1989.

[21] J. B. Rosen, *The gradient projection method for nonlinear programming, Part 1, Linear constraints*, SIAM Journal on Applied Mathematics, Vol. 8, pp. 181-217, 1960.

[22] J. B. Rosen, *The gradient projection method for nonlinear programming, Part 2, Nonlinear constraints*, SIAM Journal on Applied Mathematics, Vol. 9, pp. 514-532, 1961.

[23] J. B. Rosen, *Two-phase algorithm for nonlinear constraint problems*, Nonlinear Programming 3, Edited by O. L. Mangasarian, R. R. Meyer and S. M. Robinson, Academic Press, London and New York, pp. 97-124, 1978.