A Flexible Inexact Restoration Method and Application to Optimization with Multiobjective Constraints under Weighted-Sum Scalarization [*]

L. F. Bueno [†]   G. Haeser [‡]   J. M. Martínez [§]

August 23, 2013

Abstract

We introduce a new flexible Inexact-Restoration (IR) algorithm and an application to problems with multiobjective constraints (MOCP) under the weighted-sum scalarization approach. In IR methods each iteration has two phases. In the first phase one aims to improve the feasibility and, in the second phase, one minimizes a suitable objective function. This is done in such a way to ensure bounded deterioration of the improvement obtained in the previous phase. Here we combine the basic ideas of the Fischer-Friedlander approach for IR with the use of approximations of the Lagrange multipliers. We present a new option to obtain a range of search directions in the optimization phase and we employ the sharp Lagrangian as merit function. Furthermore, we introduce a flexible way to handle sufficient decrease requirements and an efficient way to deal with the penalty parameter. We show that with the IR framework there is a natural way to explore the structure of the MOCP in both IR phases. Global convergence of the proposed IR method is proved and examples of the numerical behavior of the algorithm are reported.

Key words: Nonlinear Programming, Inexact Restoration, Lagrange Multipliers, Multi-Objective Optimization, Weighted-Sum Scalarization, Numerical Experiments.

AMS Subject Classification: 90C30, 49K99, 65K05.

1   Introduction

Practical methods for solving nonlinearly constrained optimization problems are iterative. Given an iterate x^k, one tries to find a better approximation to the solution x^{k+1}, taking into account feasibility and optimality criteria. Several of the more traditional nonlinear programming methods preserve feasibility of all the iterates, so that the improvement criterion relies only on objective function values [1, 2, 3, 4, 5, 6, 7, 8, 9]. These methods are very effective in many cases but they tend to be slow in the presence of constraints with high curvature. Modern Inexact-Restoration

(IR) methods, on the other hand, admit infeasibility of iterates and employ inexact unspecific restoration procedures that may be adapted to different classes of problems. See [10, 11], [12, 13], [13, 14, 15] and [16, 17, 18, 19, 20, 21].

The common features to the IR methods mentioned above are the following:

1. Given an iterate x^k a sufficiently more feasible y^{k+1} is computed by any arbitrary method. This is the Feasibility Phase of the IR iteration.

2. At the Optimality Phase of the IR iteration one minimizes (approximately) a suitable objective function (Lagrangian or some approximation) for obtaining a trial point z^k. The trial point is compared with x^k regarding feasibility and optimality (using merit functions as in [10, 11, 22] or filters as in [14, 15]). If the trial point is acceptable, we define $x^{k+1} = z^k$. Otherwise, we find a new trial point z^k "closer to y^{k+1}". The fact that the new trial point should be "closer" to y^{k+1} instead of closer to x^k is crucial in the philosophy of IR methods, since it states that, ultimately, the inexactly restored point y^{k+1} is considered to be better than x^k.

The first IR methods [10, 11] motivated the introduction of a sequential optimality condition called AGP (approximate gradient projection) [23]. During several years it was thought that AGP was equivalent to the Approximate KKT condition (AKKT) implicitly used for defining stopping criteria in nonlinear programming algorithms. Surprisingly, it was recently shown [24] that AGP is stronger than AKKT and, moreover, one of the AGP variations, called L-AGP (Linear AGP) is stronger than AGP. These theoretical facts support one of the decisions taken in our flexible-purpose IR method: When linear constraints are present in the problem, feasibility with respect to these constraints should be satisfied at all the iterations.

Fischer and Friedlander [22] introduced a new line search by means of which an IR method exhibits interesting global convergence properties. The resulting approach turns out to be simpler than more classical approaches based on trust regions or filters. The main convergence result of [22] states that, under proper assumptions, the search direction d^k, used at the optimality phase, tends to zero. This result complements the suggestion of Gomes-Ruggiero, Martínez and Santos [19] who claim that, by means of the spectral gradient choice [25, 26, 27, 28, 29] of the first optimality-phase trial point, IR turns out to be the natural generalization to nonconvex constrained optimization of the Spectral Projected Gradient (SPG) method [30, 31, 32] currently used for large-scale convex constrained minimization.

In [19] spectral gradient tangent directions and the trust-region globalization approach of [11] was invoked for proving convergence. With the aim of getting closer to quadratic convergence [12] we are going to use here a search direction that comes from solving a strictly convex Newton-related quadratic minimization problem. The resulting method will be the natural generalization of the variable metric method [33] to nonlinear programming in the same sense as the method [19] is a generalization of SPG.

In this work we introduce an improved line search IR algorithm based on the Fischer-Friedlander approach. The proposed method is more flexible than the approaches in [10, 11, 22], when dealing with sufficient decrease. Several typical IR algorithmic requirements will be relaxed and, as a consequence, larger steps are more likely to be accepted by the algorithm. We will also employ approximations of Lagrange multipliers in the merit function in order to add relevant information when solving the problem. An optimization procedure for general constrained optimization problems is introduced.

Numerical experiments will be reported. For assessing the reliability of the new algorithm we will consider minimization problems with multiobjective constraints (MOCP) (see [34]) with classical scalarization, so that the reformulation of the MOCP is a Bilevel Problem [35] in the sense

2

of [16].

This paper is organized as follows. In Section 2 we introduce the new IR method and we proved that it is well defined. In Section 3 we state the assumptions that will allow us to prove convergence starting from arbitrary initial points, we prove feasibility and optimality of the limit points. The plausibility of the assumptions used to prove convergence is discussed in Section 4. In Section 5 the application to multiobjective with scalarization is presented. In Section 6 we describe the practical implementation of the algorithm and we report numerical experiments. Conclusions are given in Section 7.

Notation

- The symbol $\|\cdot\|$ denotes the Euclidean norm, although many times it can be replaced by an arbitrary norm on $\mathbb{R}^n$.

- If $h : \mathbb{R}^n \to \mathbb{R}^m$ we denote $\nabla h(x) = (\nabla h_1(x), \ldots, \nabla h_m(x))$.

- The Euclidean projection of x on a non-empty, closed, and convex set Ω is denoted $P_\Omega(x)$.

- $\mathbb{N} = \{0, 1, 2, \ldots\}$.

- If $K \subseteq \mathbb{N}$ is an infinite sequence of indices and $\lim_{k \in K} x^k = x$, we say that x is a cluster point of the sequence $\{x^k\}$.

2 Basic Method

In this section we consider the constrained optimization problem in the form:

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

where $\Omega \subset \mathbb{R}^n$ is a polytope, $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$ and both f and h are smooth. Every smooth finite dimensional constrained optimization problem may be written in the form (1) using slack variables, if necessary.

For all $x \in \Omega, \lambda \in \mathbb{R}^m$, we define the Lagrangian $L(x, \lambda)$ by:

$$L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x).$$

Given a penalty parameter $\theta \in [0, 1]$, we consider, for all $x \in \Omega, \lambda \in \mathbb{R}^m$, the following merit function:

$$\Phi(x, \lambda, \theta) = \theta L(x, \lambda) + (1 - \theta)\|h(x)\|. \tag{2}$$

This merit function is, essentially, the sharp Lagrangian defined in [36] and employed in [37, 38, 39] in the context of nonsmooth optimization and in [10] in the IR framework.

Algorithm 2.1 is a flexible form of the IR method with the line search procedure of [22]. One improvement with respect to [22] is that we are able to use Lagrange multipliers in the algortithm (thus employing (2)) preserving convergence. Moreover, sufficient descent requirements are relaxed so that acceptance of "Pure Newtonian" steps will be more frequent.

Algorithm 2.1. Flexible Inexact Restoration with sharp Lagrangian.

Step 0. Initialization

As initial approximation we choose, arbitrarily, $x^0 \in \Omega$ and $\lambda^0 \in I\!\!R^m$. We initialize $\theta_{-1} \in (0,1)$ and $k \leftarrow 0$.

Step 1. Restoration step

If $\|h(x^k)\| = 0$, define $y^{k+1} = x^k$. Otherwise, try to compute $y^{k+1} \in \Omega$ such that:

$$\|h(y^{k+1})\| < \|h(x^k)\|. \tag{3}$$

If this is not possible we say that Step 1 is deemed unsuccessful and the algorithm stops.

Step 2. Estimation of Lagrange Multipliers

Compute $\lambda^{k+1} \in I\!\!R^m$. (Practical effective ways for this choice will be given later.)

Step 3. Descent direction

Compute $d^k \in I\!\!R^n$ such that $y^{k+1} + d^k \in \Omega$,

$$\nabla L(y^{k+1}, \lambda^{k+1})^T d^k < 0 \quad \text{if} \quad d^k \neq 0, \tag{4}$$

and, for some $c_k > 0$ and all $t > 0$ small enough:

$$\|h(y^{k+1} + td^k)\| \leq \|h(y^{k+1})\| + c_k t^2 \|d^k\|^2. \tag{5}$$

Step 4. Globalization

Step 4.1. Sufficient decrease parameters

If $\|h(x^k)\| = \|h(y^{k+1})\| = 0$, take $r_k \in (0,1)$. Else, define $r_k \neq 0$ such that

$$r_k \in \left[\frac{\|h(y^{k+1})\|}{\|h(x^k)\|}, 1 \right). \tag{6}$$

Set

$$s_k \in [0, r_k). \tag{7}$$

Step 4.2 Penalty parameter computation

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

$$\Phi(y^{k+1}, \lambda^{k+1}, \theta) \leq \Phi(x^k, \lambda^k, \theta) + \frac{1 - s_k}{2}(\|h(y^{k+1})\| - \|h(x^k)\|). \tag{8}$$

Step 4.3. Line search

If $\|d^k\| = 0$ we define $t_k = 1$.

Otherwise, by means of some backtracking procedure, compute $t_k \in [0,1]$ such that

$$L(y^{k+1} + t_k d^k, \lambda^{k+1}) < L(y^{k+1}, \lambda^{k+1}) \tag{9}$$

and

$$\Phi(y^{k+1} + t_k d^k, \lambda^{k+1}, \theta_k) \leq \Phi(x^k, \lambda^k, \theta_k) + \frac{1 - r_k}{2}(\|h(y^{k+1})\| - \|h(x^k)\|). \tag{10}$$

Moreover, the backtracking procedure should be such that either $t_k = 1$ or there exists $\bar{t}_k \in [t_k, 10t_k]$ such that

$$L(y^{k+1} + \bar{t}_k d^k, \lambda^{k+1}) \geq L(y^{k+1}, \lambda^{k+1}) \tag{11}$$

or

$$\Phi(y^{k+1} + \bar{t}_k d^k, \lambda^{k+1}, \theta_k) > \Phi(x^k, \lambda^k, \theta_k) + \frac{1 - r_k}{2}(\|h(y^{k+1})\| - \|h(x^k)\|). \tag{12}$$

Step 4.4. Iteration update
    Set
$$x^{k+1} = y^{k+1} + t_k d^k, \tag{13}$$

update $k \leftarrow k + 1$ and go to Step 1.

Algorithm 2.1 is a simple Inexact Restoration algorithm where even the condition of d^k being in the tangent subspace is relaxed. However, it is interesting to show that this simplified method is well defined in the sense that an iteration can be completed if the inexactly restored point y^{k+1} can be computed at every iteration. In the following theorem we do not make use of Lipschitz conditions at all, neither on the function, nor in the constraints.

Theorem 2.1. *For all $x^k \in \Omega$, if the point y^{k+1} at Step 1 of Algorithm 2.1 is successfully computed, then the iterate x^{k+1} is well defined.*

Proof. By the hypothesis we have that Step 1 is well defined. Steps 2 and 3 can be completed observing that $\lambda^{k+1} = 0 \in I\!\!R^m$ and $d = 0 \in I\!\!R^n$ are always acceptable. Although these option will not be our choice, this ensures that both steps are well defined. We will show now that the other computations required in the iteration are possible, whenever $\|h(y^{k+1})\| < \|h(x^k)\|$ or $y^{k+1} = x^k$.
    Computation of θ_k: If $y^{k+1} = x^k$ we have that $\|h(y^{k+1})\| = \|h(x^k)\| = 0$ and thus $L(x^k, \lambda) = L(y^{k+1}, \mu)$ for all $\lambda, \mu \in I\!\!R^m$. This means that inequality (8) is obvious for all θ and so $\theta_k = \theta_{k-1}$.
    If $y^{k+1} \neq x^k$, by (3), we have that $\|h(y^{k+1})\| < \|h(x^k)\|$ and so, since $0 \leq s_k < 1$,

$$\|h(y^{k+1})\| - \|h(x^k)\| < \frac{1 - s_k}{2} \left(\|h(y^{k+1})\| - \|h(x^k)\| \right).$$

Therefore, for θ small enough we have that

$$\|h(y^{k+1})\| - \|h(x^k)\| \leq \frac{\theta}{1 - \theta}[L(x^k, \lambda^k) - L(y^{k+1}, \lambda^{k+1})] + \frac{1 - s_k}{2(1 - \theta)} \left(\|h(y^{k+1})\| - \|h(x^k)\| \right).$$

This implies that, for $\theta > 0$ small enough, the inequality (8) takes place.
    By direct calculations we note that the parameter θ_k can be computed in the following way: If (8) is verified with $\theta = \theta_{k-1}$ take $\theta_k = \theta_{k-1}$. Otherwise, we compute:

$$\theta_k = \frac{(1 + s_k) \left(\|h(x^k)\| - \|h(y^{k+1})\| \right)}{2[L(y^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) + \|h(x^k)\| - \|h(y^{k+1})\|]}. \tag{14}$$

Computation of t_k: We only need to consider the case in which $\|d^k\| \neq 0$.
    By (4), we have that (9) necessarily holds for t small enough. Let us prove now that (10) also holds for small t.
    Assume first that $y^{k+1} = x^k$. Then, $\|h(y^{k+1})\| = \|h(x^k)\| = 0$ and $L(x^k, \lambda) = f(x^k)$ for all $\lambda \in I\!\!R^m$. Thus

$$\Phi(x^k, \lambda^k, \theta_k) + \frac{1 - r_k}{2}(\|h(y^{k+1})\| - \|h(x^k)\|) = \theta_k L(x^k, \lambda^k) = \theta_k f(y^{k+1}), \tag{15}$$

and, by (5),
$$\|h(y^{k+1} + td^k)\| \leq c_k t^2 \|d^k\|^2. \tag{16}$$
Now, by (4), for t small enough we have that

$$L(y^{k+1} + td^k, \lambda^{k+1}) \leq L(y^{k+1}, \lambda^{k+1}) + \frac{t}{2}\nabla L(y^{k+1}, \lambda^{k+1})^T d^k = f(y^{k+1}) + \frac{t}{2}\nabla L(y^{k+1}, \lambda^{k+1})^T d^k \tag{17}$$

5

and

$$\frac{\theta_k}{2}\nabla L(x^k,\lambda^{k+1})^T d^k + (1-\theta_k)c_k t\|d^k\|^2 < 0. \tag{18}$$

By (15), (16), (17) and (18), for t small enough, we have that

$$
\begin{aligned}
\Phi(y^{k+1}+td^k,\lambda^{k+1},\theta_k) &= \theta_k L(y^{k+1}+td^k,\lambda^{k+1})+(1-\theta_k)\|h(y^{k+1}+td^k)\| \\
&\leq \theta_k(f(y^{k+1})+\frac{t}{2}\nabla L(y^{k+1},\lambda^{k+1})^T d^k)+(1-\theta_k)c_k t^2\|d^k\|^2 \\
&\leq \Phi(x^k,\lambda^k,\theta_k)+\frac{1-r_k}{2}(\|h(y^{k+1})\|-\|h(x^k)\|).
\end{aligned}
$$

So we conclude that (10) holds for t small enough when $y^{k+1}=x^k$.

Now, let us show that t_k is well defined in the case that $\|h(y^{k+1})\| < \|h(x^k)\|$. In this case, by (7), we have that

$$\frac{1-s_k}{2}(\|h(y^{k+1})\|-\|h(x^k)\|) < \frac{1-r_k}{2}(\|h(y^{k+1})\|-\|h(x^k)\|). \tag{19}$$

By (8) and (19),

$$\Phi(y^{k+1},\lambda^{k+1},\theta_k) < \Phi(x^k,\lambda^k,\theta_k)+\frac{1-r_k}{2}(\|h(y^{k+1})\|-\|h(x^k)\|).$$

By continuity,

$$\Phi(y^{k+1}+td^k,\lambda^{k+1},\theta_k) \leq \Phi(x^k,\lambda^k,\theta_k)+\frac{1-r_k}{2}(\|h(y^{k+1})\|-\|h(x^k)\|)$$

for t small enough. This ensures that t_k can be computed in a finite time by means of a backtracking procedure. This completes the proof. $\square$

Observe that the inequality (10) is easier to be fulfilled if r_k is chosen to be close to 1 in (6). This means that the chances of accepting large steps t_k increase if we choose $r_k \approx 1$. Moreover, the gap in (19) is larger as s_k is closer to zero, which also favours the acceptance of large steps. On the other hand, θ_k, the weight of the Lagrangian in the merit function, is larger as s_k is closer to r_k. Since the direction d^k is a descent direction for the Lagrangian, it is natural that large values of θ_k increase the chances to accept large steps. We believe that the best choice of s_k strongly depends on the nonlinearity of the constraints in (1). For simplicity, in our implementation we decide to take the largest gap in (19) setting $s_k = 0$ for all k. This flexible way of handling sufficient decrease allows the algorithm to take larger steps than the ones in [22], were r_k and s_k are equal and fixed for every k.

Theorem 2.1 showed that, for well-definiteness the following conditions are essential: (a) Simple descent of $\|h(y^{k+1})\|$ with respect to $\|h(x^k)\|$; (b) The direction d^k should be quasi-tangent (5) and (c) d^k must be a descent direction for the Lagrangian L.

3 Assumptions and Global Convergence

The global convergence proof for Algorithm 2.1 requires a problem assumption P1 and several algorithmic assumptions.

Assumption P1

The problem should be such that Lipschitz conditions hold both on the gradients of f and h. Namely, there exists $\eta > 0$ such that, for all $x, y \in \Omega$,

$$\|\nabla f(y) - \nabla f(x)\| \leq \eta\|y - x\| \quad \text{and} \quad \|\nabla h(y) - \nabla h(x)\| \leq \eta\|y - x\|. \tag{20}$$

The algorithmic assumptions involve conditions that should be satisfied by the different steps of the algorithm in order to guarantee global convergence. For the introduction of these assumptions, $\{x^k\}, \{y^k\}, \{\lambda^k\}$ will be the sequences generated by Algorithm 2.1. Note that these sequences have infinitely many terms except in the case that the restoration step fails and (3) cannot be obtained for some k.

Assumption A1

For all $k \in I\!N$, Step 1 of the Algorithm is successful and there exists $r \in [0, 1)$ and $\beta > 0$ such that

$$r_k \leq r \tag{21}$$

and

$$L(y^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) \leq \beta\|h(x^k)\|. \tag{22}$$

Conditions (21) and (22) involve the restoration step. Condition (21) states that sufficient uniform feasibility improvement should be obtained by the restoration procedure. In most IR implementations r is an algorithmic parameter and the algorithm stops declaring "restoration failure" if the restoration procedure fails to satisfy (21) after reasonable computer time. Here we adopt the practical point of view that even conservative algorithmic parameters r (say $r = 0.99$) could be excessively strict at some iterations of the algorithm, which, on the other hand, could converge smoothly under single descent requirements as (3). On the other hand, the requirement (21) for some unknown value of r is usually satisfied under regularity assumptions on the constraints. Since $y^{k+1} = x^k$ if $\|h(x^k)\| = 0$ and r_k is defined as in (6) if $y^{k+1} \neq x^k$, condition (21) implies that

$$\|h(y^{k+1})\| - \|h(x^k)\| \leq r_k\|h(x^k)\| - \|h(x^k)\| \leq -(1-r)\|h(x^k)\|. \tag{23}$$

Condition (22) requires that the deterioration of the Lagrangian at (y^{k+1}, λ^{k+1}) should be smaller than a multiple of the infeasibility $\|h(x^k)\|$. If the Lagrange multipliers estimates are bounded, and under suitable Lipschitz assumptions, condition (22) is implied by $\|y^{k+1} - x^k\| = O(\|h(x^k)\|)$. This means that the distance between x^k and the restored point y^{k+1} should be proportional to the infeasibility measure at x^k. Providing suitable safeguarding parameters β is even harder than in the case of (21), since the inequality (22) is scale-dependent. In practical terms, Assumption A1 says that we believe that the restoration algorithm employed is reasonable enough so that sufficient improvement and bounded-distance requirements are automatically satisfied.

Assumption A2

For all $k \in I\!N$ we choose d^k such that $y^{k+1} + d^k \in \Omega$ and $\nabla h(y^{k+1})^T d^k = 0$. Moreover, we assume that there exists $\sigma > 0$ such that

$$\nabla L(y^{k+1}, \lambda^{k+1})^T d^k \leq -\sigma\|d^k\|^2 \tag{24}$$

and

$$\sigma\|P_k(y^{k+1} - \nabla L(y^{k+1}, \lambda^{k+1})) - y^{k+1}\| \leq \|d^k\| \tag{25}$$

for all $k \in I\!N$, where P_k denotes the Euclidean projection on the polytope defined by $y \in \Omega$ and $\nabla h(y^{k+1})^T(y - y^k) = 0$. As we will see in Section 4, there is no loss of generality if we use the same σ in (24) and (25).

In practice, the direction d^k will be obtained as the solution of the following problem:

$$\text{Minimize } \frac{1}{2} d^T H_k d + \nabla L(y^{k+1}, \lambda^{k+1})^T d \tag{26}$$

subject to

$$\nabla h(y^{k+1})^T d = 0, \quad y^{k+1} + d \in \Omega. \tag{27}$$

Since Ω is a polytope, we have that (26)-(27) is a quadratic programming problem.

Note that the requirementn (24) is not necessarily satisfied by the solution of (26)-(27), unless we impose some additional conditions on H_k. We will show in Section 4 that Assumption A2 holds if the eigenvalues of the matrix of $Z_k^T H_k Z_k$ lie in a positive interval $[\sigma_{min}, \sigma_{max}]$, where the columns of Z_k form an orthonormal basis of the null-space of $\nabla h(y^{k+1})^T$. In implementations we will define H_k as the Hessian of the approximate Lagrangian, testing further the descent condition and switching to a safe positive definite matrix if necessary.

Assumption A3

There exists $\gamma > 0$ such that, for all $k \in I\!N$,

$$L(y^{k+1} + t_k d^k, \lambda^{k+1}) \leq L(y^{k+1}, \lambda^{k+1}) - \gamma t_k \|d^k\|^2 \tag{28}$$

Moreover, for all $k \in I\!N$ we have that $t_k = 1$ or there exists $\bar{t}_k \leq 10 t_k$ such that either (12) or

$$L(y^{k+1} + \bar{t}_k d^k, \lambda^{k+1}) > L(y^{k+1}, \lambda^{k+1}) - \gamma \bar{t}_k \|d^k\|^2 \tag{29}$$

takes place.

Assumption (28) states that the Lagrangian, whose quadratic approximation was minimized at (26)-(27), should decrease along the direction d^k. The sufficient decrease condition, which depends on γ, holds for t_k small enough under the choice (26)-(27) with safeguards on the eigenvalues of H. As we saw in Theorem 2.1, condition (10) also holds for t_k small enough. Therefore, Assumption A3 suggests a safeguarding backtracking procedure that aims to satisfy, simultaneously, (10) and (28).

Assumption A4

The Lagrange multipliers estimates $\{\lambda^k\}$ lie in a compact set.

By Assumption P1 we have that $\nabla L(\cdot, \lambda) : \Omega \to I\!R^n$ is Lipschitz continuous in Ω for all fixed $\lambda \in I\!R^m$. By Assumption A4 we can assume that the Lipschitz constant for $\nabla L(\cdot, \lambda^k)$ is the same for all $k \in I\!N$. Without loss of generality we will also denote by η this Lipschitz constant, that is, for all $x, y \in \Omega$ and $k \in I\!N$

$$\|\nabla L(y, \lambda^k) - \nabla L(x, \lambda^k)\| \leq \eta \|y - x\|. \tag{30}$$

Assumption A5

The Lagrange multipliers estimates $\{\lambda^k\}$ are such that

$$\lim_{k \to \infty} P_\Omega(y^k - \nabla f(y^k) - \nabla h(y^k)\lambda^{k+1}) - y^k = 0. \tag{31}$$

As we will see in Section 4, this assumption can be fullfiled if we take λ^{k+1} as the Lagrange multipliers associated to $\nabla h(y^k)$ in the KKT conditions of the problem (26)-(27).

In the next Lemma we will prove that the sequence of the penalty parameters is bounded away from zero. This means that the objective function always has a significative weight in the merit criterion.

8

Lemma 3.1 *Suppose that Assumption A1 holds. The sequence $\{\theta_k\}$ is non-increasing and bounded away from zero.*

Proof. By direct calculations, the inequality (8) is equivalent to

$$\theta[L(y^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) + \|h(x^k)\| - \|h(y^{k+1})\|] \leq \frac{(1+s_k)}{2}(\|h(x^k)\| - \|h(y^{k+1})\|). \quad (32)$$

If $[L(y^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) + \|h(x^k)\| - \|h(y^{k+1})\|] \leq 0$ then (32) holds for all $\theta \geq 0$, thus $\theta_k = \theta_{k-1}$.

If $[L(y^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) + \|h(x^k)\| - \|h(y^{k+1})\|] > 0$, then $y^{k+1} \neq x^k$ and, consequently,

$$\frac{(1+s_k)(\|h(x^k)\| - \|h(y^{k+1})\|)}{2[L(y^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) + \|h(x^k)\| - \|h(y^{k+1})\|]} > 0.$$

In this case we have that

$$\theta_k = \min\left\{\theta_{k-1}, \frac{(1+s_k)(\|h(x^k)\| - \|h(y^{k+1})\|)}{2[L(y^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) + \|h(x^k)\| - \|h(y^{k+1})\|]}\right\}. \quad (33)$$

By the updating rule of the penalty parameter, we have that the sequence $\{\theta_k\}$ is non-increasing. It remains to prove that $\{\theta_k\}$ is bounded away from zero. For this purpose, it suffices to show that θ_k is greater than a fixed positive number when $\theta_k \neq \theta_{k-1}$. In this case, we have:

$$\frac{1}{\theta_k} = \frac{2[L(y^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) + \|h(x^k)\| - \|h(y^{k+1})\|]}{(1+s_k)(\|h(x^k)\| - \|h(y^{k+1})\|)}$$

$$= \frac{2}{1+s_k}\left[\frac{L(y^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k)}{\|h(x^k)\| - \|h(y^{k+1})\|} + 1\right].$$

Thus, by Assumption A1, (7), and (23),

$$\frac{1}{\theta_k} \leq 2\left[\frac{\beta\|h(x^k)\|}{(1-r)\|h(x^k)\|} + 1\right] = 2\left[\frac{\beta}{1-r} + 1\right]. \quad (34)$$

This implies that the sequence $\{1/\theta_k\}$ is bounded. Therefore, the sequence $\{\theta_k\}$ is non increasing and bounded away from zero. $\qquad\square$

The following lemma ensures that any limit point of the sequence generated by Algorithm 2.1 is feasible.

Lemma 3.2 *Suppose that Assumptions A1 and A4 hold. Then the sum $\sum_{k=0}^{\infty} \|h(x^k)\|$ is convergent.*

Proof. By condition (10) and Assumption A1, for all $k \in \mathbb{N}$ one has that

$$\Phi(x^{k+1}, \lambda^{k+1}, \theta_k) \leq \Phi(x^k, \lambda^k, \theta_k) + \frac{1-r}{2}(\|h(y^{k+1})\| - \|h(x^k)\|).$$

Therefore, by (23),

$$\Phi(x^{k+1}, \lambda^{k+1}, \theta_k) \leq \Phi(x^k, \lambda^k, \theta_k) - \frac{(1-r)^2}{2}\|h(x^k)\|. \quad (35)$$

9

Let us define $\rho_k = (1 - \theta_k)/\theta_k$ for all $k \in \{-1, 0, 1, 2 \cdots\}$. By Lemma 3.1 there exists $\bar{\theta} > 0$ such that $\theta_k \geq \bar{\theta}$ for all $k \in \mathbb{N}$. This implies that $\rho_k \leq 1/\bar{\theta} - 1$ for all $k \in \mathbb{N}$. Since $\{\rho_k\}$ is bounded and non-decreasing it follows that

$$\sum_{k=0}^{\infty}(\rho_k - \rho_{k-1}) = \lim_{k \to \infty} \rho_k - \rho_{-1} < \infty. \tag{36}$$

By compactness, the sequence $\{\|h(x^k)\|\}$ is bounded. Therefore, by (36), there exists $c > 0$ such that

$$\sum_{k=0}^{\infty}(\rho_k - \rho_{k-1})\|h(x^k)\| \leq c < \infty. \tag{37}$$

Now, by (35),

$$L(x^{k+1}, \lambda^{k+1}) + \frac{1 - \theta_k}{\theta_k}\|h(x^{k+1})\| \leq L(x^k, \lambda^k) + \frac{1 - \theta_k}{\theta_k}\|h(x^k)\| - \frac{(1-r)^2}{2\theta_k}\|h(x^k)\|.$$

Since $0 < \theta_k < 1$, we have that $\frac{(1-r)^2}{2} < \frac{(1-r)^2}{2\theta_k}$, so

$$L(x^{k+1}, \lambda^{k+1}) + \rho_k\|h(x^{k+1})\| \leq L(x^k, \lambda^k) + \rho_k\|h(x^k)\| - \frac{(1-r)^2}{2}\|h(x^k)\|.$$

Therefore, for all $k \in \mathbb{N}$,

$$L(x^{k+1}, \lambda^{k+1}) + \rho_k\|h(x^{k+1})\| \leq L(x^k, \lambda^k) + \rho_{k-1}\|h(x^k)\| + (\rho_k - \rho_{k-1})\|h(x^k)\| - \frac{(1-r)^2}{2}\|h(x^k)\|.$$

Thus, for all $k \in \mathbb{N}$, we have:

$$L(x^{k+1}, \lambda^{k+1}) + \rho_k\|h(x^{k+1})\| \leq L(x^0, \lambda^0) + \rho_{-1}\|h(x^0)\| + \sum_{j=0}^{k}(\rho_j - \rho_{j-1})\|h(x^j)\| - \frac{(1-r)^2}{2}\sum_{j=0}^{k}\|h(x^j)\|.$$

Therefore, by (37),

$$L(x^{k+1}, \lambda^{k+1}) + \rho_k\|h(x^{k+1})\| \leq L(x^0, \lambda^0) + \rho_{-1}\|h(x^0)\| + c - \frac{(1-r)^2}{2}\sum_{j=0}^{k}\|h(x^j)\|.$$

Thus,

$$\frac{(1-r)^2}{2}\sum_{j=0}^{k}\|h(x^j)\| \leq -[L(x^{k+1}, \lambda^{k+1}) + \rho_k\|h(x^{k+1})\|] + L(x^0, \lambda^0) + \rho_{-1}\|h(x^0)\| + c$$

Since the functions L and h are continuous, by Assumption A4 and the compactness of Ω, we have that the sequences $\{L(x^k, \lambda^k)\}$ and $\{\|h(x^k)\|\}$ are bounded. Since $\{\rho_k\}$ is also bounded, it follows that the series $\sum_{k=0}^{\infty} \|h(x^k)\|$ is convergent. $\qquad\square$

The following Lemma shows that, if the problem is sufficiently smooth, the direction on the tangent subspace ensures an uniform bounded deterioration of the feasibility. The lemma also states that we can guarantee that the step size will not tend to zero. This means that Algorithm 2.1 will not produce steps that are excessively small.

10

Lemma 3.3 *Suppose that Assumptions P1, A1, A2 and A4 hold. Define c_k as in (5) and t_k as in Step 4.3 of Algorithm 2.1. Then there exist $c > 0$ and $\bar{t} > 0$ such that $c_k > c$ and $t_k \geq \bar{t}$ for all $k \in I\!N$.*

Proof. Given any continuously differentiable function $F : I\!R^n \to I\!R^p$ we obtain by Taylor's formula that

$$F(y + td) = F(y) + t\nabla F(y)^T d + t \int_0^1 \left(\nabla F(y + txd) - \nabla F(y)\right)^T d\, dx. \tag{38}$$

Applying (38) with $F(y + td) = h(y^{k+1} + td^k)$, by Assumptions P1 and A2, the triangle inequality and the Cauchy-Schwarz inequality, we have that, for $c = \frac{\eta}{2}$,

$$\|h(y^{k+1} + td^k)\| \leq \|h(y^{k+1})\| + t\int_0^1 \eta\|y^{k+1} + txd^k - y^{k+1}\|\|d^k\|\, dx \leq \|h(y^{k+1})\| + ct^2\|d^k\|^2. \tag{39}$$

Thus (5) holds for $c_k = \frac{\eta}{2}$ independently of k and for all $t \geq 0$.

Now, applying (38) with $F(y + td) = L(y^{k+1} + td^k, \lambda^k)$, by Assumptions P1, A2, A4 and (30),

$$
\begin{aligned}
L(y^{k+1} + td^k, \lambda^{k+1}) &= L(y^{k+1}, \lambda^{k+1}) + t\nabla L(y^{k+1}, \lambda^{k+1})^T d^k + \\
&\quad + t\int_0^1 \left(\nabla L(y^{k+1} + txd^k, \lambda^{k+1}) - \nabla L(y^{k+1}, \lambda^{k+1})\right)^T d^k\, dx \\
&\leq L(y^{k+1}, \lambda^{k+1}) - \sigma t\|d^k\|^2 + t\int_0^1 \eta\|y^{k+1} + txd^k - y^{k+1}\|\|d^k\|\, dx \\
&\leq L(y^{k+1}, \lambda^{k+1}) - \left(\sigma - \frac{\eta t}{2}\right)t\|d^k\|^2.
\end{aligned}
$$

So, for $\gamma = \frac{\sigma}{2}$ and $\tau = \frac{\sigma}{2\eta}$ we have that

$$L(y^{k+1} + td^k, \lambda^{k+1}) \leq L(y^{k+1}, \lambda^{k+1}) - \gamma t\|d^k\|^2, \tag{40}$$

for all $t \leq \tau$.

If $\|d^k\| \neq 0$, condition (40) ensures that (9) holds for any $t \leq \tau$. Let us prove that we can also satisfy (10) for any t in a specific interval. In fact we will prove that any $t \leq \min\left\{\frac{\bar{\theta}\gamma}{c(1-\bar{\theta})}, \tau\right\}$ is acceptable at Step 4.3 of Algorithm 2. Since $r_k \geq 0$, by (8), (19), (39), (40) and Lemma 3.1 we have that

$$\Phi(y^{k+1} + td^k, \lambda^{k+1}, \theta_k) - \Phi(x^k, \lambda^k, \theta_k)$$

$$= \Phi(y^{k+1} + td^k, \lambda^{k+1}, \theta_k) - \Phi(y^{k+1}, \lambda^{k+1}, \theta_k) + \Phi(y^{k+1}, \lambda^{k+1}, \theta_k) - \Phi(x^k, \lambda^k, \theta_k)$$

$$\leq \theta_k \left(L(y^{k+1} + td^k, \lambda^{k+1}) - L(y^{k+1}, \lambda^{k+1})\right) + (1 - \theta_k)\left(\|h(y^{k+1} + td^k)\| - \|h(y^{k+1})\|\right) +$$

$$+ \frac{1 - s_k}{2}\left(\|h(y^{k+1})\| - \|h(x^k)\|\right)$$

$$\leq -\theta_k \gamma t\|d^k\|^2 + (1 - \theta_k)ct^2\|d^k\|^2 + \frac{1 - r_k}{2}\left(\|h(y^{k+1})\| - \|h(x^k)\|\right)$$

$$\leq -\bar{\theta}\gamma t\|d^k\|^2 + (1 - \bar{\theta})ct^2\|d^k\|^2 + \frac{1 - r_k}{2}\left(\|h(y^{k+1})\| - \|h(x^k)\|\right)$$

$$\leq ((1 - \bar{\theta})ct - \bar{\theta}\gamma)t\|d^k\|^2 + \frac{1 - r_k}{2}\left(\|h(y^{k+1})\| - \|h(x^k)\|\right)$$

$$\leq \frac{1 - r_k}{2}\left(\|h(y^{k+1})\| - \|h(x^k)\|\right).$$

11

So, due to the backtracking procedure, we have that $t_k \geq \bar{t} \equiv \min\left\{\frac{\bar{\theta}\gamma}{10c(1-\bar{\theta})}, \frac{\tau}{10}, 1\right\}$ for all $k \in I\!N$.
$\square$

Lemma 3.4 *Suppose that Assumptions P1, A1–A4 hold. Then $\lim_{k\to\infty} \|d^k\| = 0$.*

Proof. By Lemma 3.3 there exists $\bar{t} > 0$ such that $t_k \geq \bar{t}$ for all $k \in I\!N$. By (22) and (28),

$$
\begin{aligned}
L(x^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) &= L(x^{k+1}, \lambda^{k+1}) - L(y^{k+1}, \lambda^{k+1}) + L(y^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k) \\
&\leq -\gamma t_k \|d^k\|^2 + \beta\|h(x^k)\| \\
&\leq -\gamma \bar{t}\|d^k\|^2 + \beta\|h(x^k)\|.
\end{aligned}
$$

By Lemma 3.2 there exists c such that $\sum_{k=0}^{\infty} \|h(x^k)\| = c$. Thus

$$
\begin{aligned}
L(x^{l+1}, \lambda^{l+1}) - L(x^0, \lambda^0) &= \sum_{k=0}^{l}(L(x^{k+1}, \lambda^{k+1}) - L(x^k, \lambda^k)) \\
&\leq -\gamma\bar{t}\sum_{k=0}^{l}\|d^k\|^2 + \beta\sum_{k=0}^{l}\|h(x^k)\| \\
&\leq -\gamma\bar{t}\sum_{k=0}^{l}\|d^k\|^2 + \beta c.
\end{aligned}
$$

Since f and h are bounded below on Ω, and $\{\lambda^{k+1}\}$ remains in a compact set, we have that the series $\sum_{k=0}^{\infty} \|d^k\|^2$ is convergent and, thus, $\{\|d^k\|\}$ converges to zero. $\square$

The next Lemma and its corollary are used to prove optimality of every limit point of the sequence generated by Algorithm 2.1.

Lemma 3.5 *Consider P_k as in Assumpton A2. Given $y^{k+1} \in \Omega$ then*

$$P_k(y^{k+1} - \nabla L(y^{k+1}, \lambda)) = P_k(y^{k+1} - \nabla L(y^{k+1}, \mu)),$$

for all $\lambda, \mu \in I\!R^m$.

Proof. Defining $S_k \equiv \{d : \nabla h(y^{k+1})^T d = 0\}$ we have that $\nabla h(y^{k+1})(\lambda - \mu) \in S_k^\perp$. Since $\nabla L(y^{k+1}, \lambda) = \nabla L(y^{k+1}, \mu) + \nabla h(y^{k+1})(\lambda - \mu)$ the result follows from the Projection Theorem (see [40], Proposition B.11, item (b), page 704). $\square$

Corollary 3.1 *Let x^* be a feasible point of (1) such that*

$$\lim_{k\to\infty} y^{k+1} = x^* \ \text{and} \ \lim_{k\to\infty} P_k(y^{k+1} - \nabla L(y^{k+1}, \lambda^{k+1})) - y^{k+1} = 0,$$

for some $\{y^{k+1}\} \subset \Omega$ and $\{\lambda^{k+1}\} \subset I\!R^m$. Then x^* satisfies the L-AGP optimality condition [24].

Proof. By hypothesis x^* is feasible. Moreover, by Lemma 3.5,

$$
\begin{aligned}
\lim_{k\to\infty} P_k(y^{k+1} - \nabla f(y^{k+1})) - y^{k+1} &= \lim_{k\to\infty} P_k(y^{k+1} - \nabla L(y^{k+1}, 0)) - y^{k+1} \\
&= \lim_{k\to\infty} P_k(y^{k+1} - \nabla L(y^{k+1}, \lambda^{k+1})) - y^{k+1} = 0.
\end{aligned}
$$

12

Thus, x^* satisfies the L-AGP optimality condition. $\qquad\square$

The next lemma will be used to prove the convergence of the Lagrange multipliers estimates $\{\lambda^k\}$.

Lemma 3.6 *Let x^* be a feasible point such that*

$$\lim_{k\to\infty} y^k = x^* \text{ and } \lim_{k\to\infty} P_\Omega(y^k - \nabla L(y^k, \lambda^{k+1})) - y^k = 0,$$

for some $\{y^k\} \subset \Omega$ and $\{\lambda^{k+1}\} \subset \mathbb{R}^m$. Then, if x^* satisfies the Mangasarian-Fromovitz constraint qualification, we have that the KKT conditions hold in x^*. Moreover, all the limit points of $\{\lambda^{k+1}\}$ are Lagrange Multipliers associated with the equality constraints in x^*.

Proof. Let $z^k = P_\Omega(y^k - \nabla L(y^k, \lambda^{k+1}))$, hence $\lim_{k\to\infty} z^k - y^k = 0$ and z^k is the solution of the problem:

$$\text{Minimize } \frac{1}{2}\|z - (y^k - \nabla L(y^k, \lambda^{k+1}))\|^2 \text{ subject to } z \in \Omega.$$

Considering $\Omega = \{x \in \mathbb{R}^n | Ax = b, x \geq 0\}$, for every k, there exists α^k and β^k such that

$$\beta^k \geq 0, \quad (z^k)^T \beta^k = 0, \tag{41}$$

and

$$z^k - y^k + \nabla L(y^k, \lambda^{k+1}) + A^T \alpha^k - \beta^k = 0. \tag{42}$$

Defining $M_k = \max\{\|\lambda^{k+1}\|_\infty, \|\alpha^k\|_\infty, \|\beta^k\|_\infty\}$ we have that $\{M_k\}$ is bounded, otherwise we could divide (42) by M_k and take the limit on a suitable subsequence to get a contradiction with the Mangasarian-Fromovitz condition.

Since $\{M_k\}$ is bounded, if λ^* is a limit point of $\{\lambda^{k+1}\}$, there is K such that $\lim_{k\in K} \lambda^{k+1} = \lambda^*$, $\lim_{k\in K} \alpha^k = \alpha^*$, and $\lim_{k\in K} \beta^k = \beta^*$. Taking limits for $k \in K$ in (41) and (42) we verify that λ^* is a Lagrange multiplier associated with the constraints $h(x) = 0$ in x^*. $\qquad\square$

The following theorem summarizes the convergence properties of Algorithm 2.1 under Assumptions P1 and A1–A5.

Theorem 2.2. *Suppose that Assumptions P1 and A1–A5 hold. Then:*

1. *For all $k \in \mathbb{N}$, x^k is well defined.*

2. *There exists $\bar{\theta} > 0$ such that $\theta_k \geq \bar{\theta}$ for all $k \in \mathbb{N}$.*

3. *$\lim_{k\to\infty} \|h(x^k)\| = \lim_{k\to\infty} \|h(y^k)\| = 0$ and any cluster point of $\{x^k\}$ or $\{y^k\}$ is feasible.*

4. *$\lim_{k\to\infty} d^k = 0$.*

5. *$\lim_{k\to\infty} \|y^k - x^k\| = 0$.*

6. *The sequences $\{x^k\}$ and $\{y^k\}$ admit the same cluster points.*

7. *Limit points of $\{x^k\}$ satisfy the L-AGP optimality condition.*

8. *If a limit point x^* satisfies the Constant Positive Generators (CPG) constraint qualification [41], then the KKT conditions hold at x^**

13

9. *If a limit point x^* satisfies the Mangasarian-Fromovitz constraint qualification, then the sequence $\{\lambda^k\}$ admits a limit point $\{\lambda^*\}$ which is a Lagrange multiplier associated with $\nabla h(x^*)$.*

Proof.

Well-definiteness of x^k has been proved at Theorem 2.1.

The existence of $\bar{\theta}$ follows from Lemma 3.1.

By Lemma 3.2 and (3),

$$\lim_{k \to \infty} \|h(y^{k+1})\| \leq \lim_{k \to \infty} \|h(x^k)\| = 0,$$

and thus, by the continuity of h, any cluster point of $\{x^k\}$ or $\{y^k\}$ is feasible.

The fact that $\lim_{k \to \infty} d^k = 0$ follows from Lemma 3.4.

By item 4 and the triangle inequality we have that

$$\lim_{k \to \infty} \|y^{k+1} - x^{k+1}\| = \lim_{k \to \infty} \|y^{k+1} - (y^{k+1} + t_k d^k)\| \leq \lim_{k \to \infty} \|d^k\| = 0.$$

By item 5 and also by the triangle inequality

$$\lim_{k \in K \underset{\infty}{\subseteq} I\!N} \|x^k - z\| = 0 \Leftrightarrow \lim_{k \in K \underset{\infty}{\subseteq} I\!N} \|y^k - z\| = 0,$$

thus the sequences $\{x^k\}$ and $\{y^k\}$ admit the same cluster points.

Let x^* be a limit point of $\{x^k\}$, so, by item 6, there is $K \underset{\infty}{\subseteq} I\!N$ such that $\lim_{k \in K} y^{k+1} = x^*$. By Assumption A2 and item 4,

$$\lim_{k \in K} \|P_k(y^{k+1} - \nabla L(y^{k+1}, \lambda^{k+1})) - y^{k+1}\| = 0.$$

Since by item 3 we have that x^* is feasible, we conclude, by Lemma 3.1, that x^* satisfies the L-AGP optimality condition.

The statement 8 is a consequence of item 7 and the fact that sequential optimality conditions as L-AGP imply KKT under the CPG constraint qualification [24, 41].

Finally, suppose that $\lim_{k \in K} y^k = x^*$ and the Mangasarian-Fromovitz constraint qualification holds in x^*. Since the sequence $\{\lambda^{k+1}\}$ lie in a compact set, we have that it has at least one limit point. Then, by Lemma 3.6, we conclude that any limit point of $\{\lambda^{k+1}\}$ with $k \in K$ is a Lagrange Multiplier associated with $\nabla h(x^*)$. $\qquad\square$

Remark: The CPG constraint qualification is weaker than CRSC [41], (R)CPLD [42, 43, 44], (R)CRCQ [45] and Mangasarian-Fromovitz constraint qualification.

4 Plausibility of Assumptions

One of our main assumptions is that given an infeasible point x^k, a more feasible point can always be computed at the Restoration phase. Clearly, this assumption is not always true. For example, if the feasible set is empty and x^k is a minimizer of the infeasibility it would be impossible to be successful at the Restoration phase. However, if a good algorithm is used at the Restoration phase, we judge that the improvement on feasibility will be achieved, or the problem is probably infeasible.

Lipschitz conditions on the gradients of f and h, assumed in Assumption P1, are standard in smooth optimization. Convexity conditions on the problem (26)-27) are also usual in quadratic

programming analyses. After solving this problem, a natural way to estimate the Lagrange multipliers is taking λ^{k+1} as the Lagrange multipliers associated to the constraints $\nabla h(y^k)^T d = 0$. In this case it is plausible to assume that λ^{k+1} lies in a compact set if the Mangasarian-Fromovitz constraint qualification holds at y^k. So, if the sequence $\{\|H_k d^k\|\}$ is uniformuly bounded, we have that Assumption A4 holds under the Mangasarian-Fromovitz constraint qualification on Ω. An easy alternative to guarantee that Assumption A4 holds is to use safeguards on the Lagrange multipliers in the implementation.

Assumption A1 asks for a sufficient improvement of feasibility with a bounded deterioration at the Lagrangian. This assumption can be satisfied with typical Lipschitz hypothesis and regularity conditions on the feasible set. In Lemma 2 of [22] it is proved that, if the Mangasarian Fromovitz-constraint qualification holds at all points of Ω then the closest feasible point to x^k fulfills the bounded deterioration for the objective function f, and so any $r_k > 0$ is a possible choice for all k. A practical restoration procedure is shown in Lemma 6.1 of [10] if the Linear Independence constraint qualification holds on Ω. This result asserts that, given $r \in (0, 1)$ there is a neighbourhood N of the feasible set such that if $x^k \in N$ then the linearization of the constraints at x_k $T_k \equiv \{y \in \Omega | h(x^k) + \nabla(x^k)^T(y - x^k) = 0\}$ is not empty, and, denoting by y^{k+1} the point of T_k closest to x^k, one has that $h(y^{k+1}) \leq r h(x^k)$. Moreover under Lipschitz conditions, bounded deterioration for the objective function holds. In both restoration procedures, if the Lagrange multipliers λ_k remain in a compact set (Assumption A4), then the bounded deterioration also holds for the Lagrangian.

The plausibility of Assumption A2 is strictly linked with our proposal in the present paper of choosing d^k as the solution of the quadratic programming problem:

$$\text{Minimize } \nabla L(y^{k+1}, \lambda^{k+1})^T d + \frac{1}{2} d^T H_k d \quad \text{subject to} \quad y^{k+1} + d \in \Omega, \ \nabla h(y^{k+1})^T d = 0, \quad (43)$$

where H_k is symmetric, $Z_k^T H_k Z_k$ is positive definite, and the columns of Z_k form a orthonormal basis of the null-space of $\nabla h(y^{k+1})^T$. In addition, we assume that the eigenvalues of $Z_k^T H_k Z_k$ lie in a positive interval $[\sigma_{min}, \sigma_{max}]$.

Selecting d^k in this way we have that, since $d = 0$ is feasible to (43),

$$\nabla L(y^{k+1}, \lambda^k)^T d^k + \frac{1}{2}(d^k)^T H_k d^k \leq 0.$$

So

$$\nabla L(y^{k+1}, \lambda^k)^T d^k \leq -\frac{\sigma_{min}}{2} \|d^k\|^2.$$

In order to prove condition (25), let us define,

$$D_k = \{d \in \mathbb{R}^n \mid \nabla h(y^{k+1})^T d = 0 \text{ and } y^{k+1} + d \in \Omega\}. \quad (44)$$

Since d^k is the solution of the linear constrained problem (43), then

$$P_{D_k}(d^k - H_k d^k - \nabla L(y^{k+1}, \lambda^{k+1})) - d^k = 0.$$

Changing variables $(y = y^{k+1} + d)$ and using that projections are non-expansive we have:

$$\|P_k(y^{k+1} - \nabla L(y^{k+1}, \lambda^k)) - y^{k+1}\| = \|P_{D_k}(-\nabla L(y^{k+1}, \lambda^{k+1}))\|$$

$$= \|P_{D_k}(-\nabla L(y^{k+1}, \lambda^{k+1})) - P_{D_k}(d^k - H_k d^k - \nabla L(y^{k+1}, \lambda^k)) + P_{D_k}(d^k - \nabla L(y^{k+1}, \lambda^k))\|$$

$$\leq \|P_{D_k}(-\nabla L(y^{k+1}, \lambda^{k+1})) - P_{D_k}(d^k - H_k d^k - \nabla L(y^{k+1}, \lambda^k))\| + \|P_{D_k}(d^k - H_k d^k - \nabla L(y^{k+1}, \lambda^k))\|$$

$$\leq \|d^k - H_k d^k\| + \|d^k\| \leq (2 + \sigma_{max})\|d^k\|$$

So, defining $\sigma = \min \left\{ \frac{\sigma_{min}}{2}, \frac{1}{2+\sigma_{max}} \right\}$, we have that Assumption A2 holds.

From Lemma 3.3 we can see that condition (40) holds for $\gamma = \frac{\sigma_{min}}{4}$ and t sufficiently small. Thus, if σ_{min} is known, we can ensure that Assumption A3 holds testing the sufficient decrease at Step 5 of Algorithm 2.1. Moreover, even if σ_{min} is not known, it is possible to do a double backtracking, in t and σ, to ensure that Assumption A3 holds [46]. However, this does not seem to be neither efficient nor necessary in practice, so we only test simple decrease at Step 5 of Algorithm 2.1.

Finally, to prove the plausibility of assumption A5, we consider that λ^{k+1} is chosen as the vector of Lagrange multipliers associated to $\nabla h(y^k)$ in problem (26)-(27). In this case, denoting $D_k = \{d : y^k + d \in \Omega\}$, we have that

$$P_{D_k}(d^{k-1} - H_{k-1}d^{k-1} - \nabla f(y^k) - \nabla h(y^k)\lambda^{k+1}) - d^{k-1} = 0.$$

So

$$\|P_\Omega(y^k - \nabla f(y^k) - \nabla h(y^k)\lambda^{k+1}) - y^k\| = \|P_{D_k}(-\nabla f(y^k) - \nabla h(y^k)\lambda^{k+1})\|$$

$$\leq \|P_{D_k}(-\nabla f(y^k) - \nabla h(y^k)\lambda^{k+1}) - P_{D_k}(d^{k-1} - H_{k-1}d^{k-1} - \nabla f(y^k) - \nabla h(y^k)\lambda^{k+1})\|+$$

$$+\|P_{D_k}(d^{k-1} - H_{k-1}d^{k-1} - \nabla f(y^k) - \nabla h(y^k)\lambda^{k+1})\|$$

$$\leq \|d^{k-1} - H_{k-1}d^{k-1}\| + \|d^{k-1}\|.$$

Thus, considering that $\|H_k\|$ is uniformuly bounded and that Assumptions P1, A1–A4 hold then, by Lemma 3.4, we have that

$$\lim_{k\to\infty} P_\Omega(y^k - \nabla f(y^k) - \nabla h(y^k)\lambda^{k+1}) - y^k = 0,$$

which means that Assumption 5 holds.

5   Application to Optimization with Multiobjective Constraints under the Weighted-Sum Scalarization

Inexact Restoration methods are useful in problems in which there is a natural way to deal with feasibility and optimality in different phases. This is the case of Bilevel Optimization problems [35], in which Restoration consists of minimizing a function of the state variables for fixed values of the controls [16, 47]. In this section we deal with Bilevel problems coming from the Weighted-Sum Scalarization approach to Multiobjective Optimization. Given q functions $f_1, \ldots, f_q$, nonnegative weights w_i such that $\sum_{i=1}^q w_i = 1$, a set of constraints $h_i(x) = 0, i = 1, \ldots, p$ and a polytope $\Omega \subseteq \mathbb{R}^n$, we define $Z(w)$ as the set of solutions of

$$\text{Minimize (with respect to } x) \sum_{i=1}^q w_i f_i(x) \quad \text{subject to} \quad h(x) = 0, x \in \Omega, \tag{45}$$

and $Z = \bigcup\{Z(w) \mid w \in \mathbb{R}_+^q \text{ and } \sum_{i=1}^q w_i = 1\}$. Given $x_c \in \mathbb{R}^n$, we are interested in the following problem:

$$\text{Minimize } \|x - x_c\|^2 \text{ subject to } x \in Z. \tag{46}$$

Problem (45) represents the Weighted-Sum Scalarization approach for solving Multiobjective problems [48, 49, 50]. Many other scalarizations have been introduced that are better than Weighted-Sum for nonconvex problems [51, 52, 53, 54]. Roughly speaking, the elements of Z are always Multiobjective solutions (Pareto points) but all the Pareto points are solutions of (45) only under

convexity assumptions. Therefore, problem (46) is an instance of the problem of Optimization over the Efficient Set (see [55, 56, 57] and references therein). The objective function of (46) reflects the necessity of taking a decision x with minimal variation with respect to a possible previous decision x_c. For instance, if the goals of a portfolio program are to maximize profit and minimize variance it is natural to establish that the best multiobjective solution is the one that differs less from the present portfolio.

In order to develop an affordable algorithm for solving (46) we rely on Nonlinear Optimization approaches. Problem (46) may be expressed as a standard Nonlinear Programming problem replacing the constraint $x \in Z$ with the KKT conditions of the problem (45). However, the KKT conditions of (45) do not reflect accurately the constraint $x \in Z$. This constraint imposes that x must be a *minimizer* of problem (45), not merely a stationary point.

For simplicity, let us describe here the situation in which (45) is an unconstrained optimization problem. In this case, the Nonlinear Programming reformulation yields:

$$\text{Minimize } \|x - x_c\|^2 \text{ subject to } \sum_{i=1}^{q} w_i \nabla f_i(x) = 0, \ \sum_{i=1}^{q} w_i = 1, \text{ and } w \geq 0. \qquad (47)$$

Problem (47) is a standard constrained optimization problem with $n + q$ variables (x and w), $n + 1$ equality constraints, and bound constraints $w_i \geq 0, i = 1, \ldots, q$. The fulfillment of the constraints of (47) by a pair (x, w) is not sufficient to guarantee that $x \in Z(w)$, except in the case that all the functions are convex. Furthermore, non KKT points of (45) may be attractors for nonlinear programming softwares when solving (47). Consider, for example, that $n = q = 1$ and $f_1(x) = \frac{3}{4}x^4 - \frac{7}{3}x^3 + \frac{5}{2}x^2$. In this case $x^* = 0$ is the only KKT point of (45), although $x = 1$ is a local minimizer of the infeasibility measure $|f_1'(x)|$. For this reason we have that $x = 1$ will be likely founded by a standard nonlinear optimization method when solving (47) with an initial point close to $x = 1$.

When dealing with the constrained case, the situation is even harder. One could have that a solution of problem (45) may not be a KKT point. Consider, for example, that $n = q = 1$, $f_1(x) = x$, and $h(x) = x^2$. In this case $x^* = 0$ is the solution of (45) but the feasible set of the reformulated problem is empty.

On the other hand, the Inexact Restoration approach seems to be adequate for these situations since, essentially, allows one to consider the problem (46) under a formulation that evokes more properly the essence of the problem:

$$\text{Minimize } \frac{1}{2}\|x - x_c\|^2 \quad \text{subject to} \quad w_i \geq 0, \sum_{i=1}^{q} w_i = 1, \text{ and } x \text{ is a minimizer of (45)}.$$

Moreover, assuming that, given a set of weights $w_1, \ldots, w_q$, it is possible to minimize, approximately, $\sum_{i=1}^{q} w_i f_i(x)$, the Inexact Restoration necessarily finds a point that fulfills the L-AGP optimality condition, not requiring constraint qualifications at all. The objective of the following sections of this paper is to show how these theoretical properties are reflected in practical computations.

In order to fully exploit the potentiality of IR we need to define the specific restoration procedure that we want to employ. The natural procedure is, given the (generally infeasible) current point (x^k, w^k), to keep fixed w^k and to obtain the restored point y^{k+1} by inexact minimization of the weighted function $\sum_{i=1}^{q} w_i^k f_i(x)$.

Although the Restoration Phase of the IR algorithm involves only the variables x, the Optimality Phase involves both vectors of variables x and w. Moreover, even in the Optimality Phase,

we are able to maintain the fulfillment of the constraints $\sum_{i=1}^{q} w_i = 1$ and $w \geq 0$ throughout the process by means of the following definition of the polytope Ω in (1):

$$\Omega = \left\{ (x, w) \in I\!\!R^{n+q} \mid \sum_{i=1}^{q} w_i = 1 \text{ and } w \geq 0 \right\}.$$

As a consequence, in the Optimization Phase we will solve Quadratic Programming problems with the constraints defined by Ω and the tangent space (involving both variables x and w) to the manifold defined by $\sum_{i=1}^{q} w_i \nabla f_i(x) = \sum_{i=1}^{q} w_i^k \nabla f_i(x^k)$. Moreover, the presence of Ω as a restriction of all the Optimization Phases guarantees the strength of the condition L-AGP in this case.

In the numerical experiments we used three families of problems.

1. **Portfolios:** This family of problems is related with to the well-known Mean-Variance problem in portfolio optimization. The investor aims to maximize return and to minimize the risk of the investment. In our simulation we used seven shares from the London exchange market: AZN.L, BARC.L, KGF.L, LLOY.L, MKS.L, TSCO.L and VOD.L, plus a risk-free asset. We consider scenarios using the historical data on daily returns from July 16 (2012) to October 10 (2012) available in [59]. The expected profit was defined as $f_1(x) = -v^T x$ and, using the variance as a measure of the risk, we have that $f_2(x) = x^T M x$, where v is the expected return and M is the covariance matrix for the generated scenarios. Combining the data of these assets we can generate as many assets as desired. Therefore, we could address this problem using up to 1000 variables. The constraints of the Multiobjective Problem are $\sum_{i=1}^{n} x_i = 1$ and $x_i \geq 0$ for all $i = 1, \ldots, n$. We defined the current decision x_c as a random point that satisfies the constraints. Note that, in this case, (45) is a linearly constrained optimization problem. This means that the variables of the problem, as well as in the Optimization Phase of IR, are the primal variables x, the optimal weights w, and the Lagrange multipliers of (45).

2. **MGH-generated Problems:** The MGH Collection of Moré, Garbow, and Hillstrom was used to generate 120 different Multiobjective problems. We used the functions 3, 4, 5, 9, 12, 14, 16, 18, 20, 21, 22, 23, 24, 25, 26, and 35 of this collection and, for each pair $(i, j)(i \neq j)$ of this set we defined the problem in which f_1 is the function i and f_2 is the function j. The number of variables n was defined to be the maximum of n_i and n_j, where n_i was the number of variables of the function i in the MGH collection and n_j was the number of variables of the function j. In all the problems the domain was the box with bounds -10 and 10 and we defined $x_c = 0$

3. **Quartic Polynomials:** The third class of problems with multiobjective constraints consisted of considering two objectives where each one is a random fourth-degree polynomial in several variables. For simplicity, we considered only separable polynomials. The objective functions $f_1(x)$ and $f_2(x)$ were defined as

$$f_i(x) = \sum_{j=1}^{n} a_{ij} x_j^4 + b_{ij} x_j^3 + c_{ij} x_j^2 + d_{ij} x_j,$$

where a_{ij} was random in $[0, 10]$ and b_{ij}, c_{ij}, and d_{ij} were randomly chosen in $[-10, 10]$. The current decision x_c was also random, with each coordinate between -10 and 10. We considered three instances of this problem, with $n \in \{1, 10, 20\}$ and 100 different problems in each case.

6 Implementation and Numerical Results

Our numerical tests were destinated to check the effectiveness of the IR algorithms. More precisely, our tests are destinated to corroborate (or not) that the IR framework provides a better alternative than the straightforward Nonlinear Programming approach for minimizing a function over the stationary points of the Weighted-Sum Scalarization problem (45).

The code that implements IR was written in Fortran, employing double precision and the following computer environment: Intel(R) Core(TM) i5-2400 CPU @ 3.10GHz with 4GB of RAM memory. For solving the quadratic subproblems in the Optimization Phase we employed Fletcher's subroutine qlcpd.f [60] and for Restoration steps we minimized the Weighted-Sum Scalarization problem (45) using Fletcher's filterSD.f. The problems (46) were solved using different Inexact Restoration instances and were also tackled using a consolidated Constrained Optimization code using the Nonlinear-Programming reformulation. For this purpose, we also employed Fletcher's filterSD.f [60], which uses Sequential Linearly Constrained Programming with globalization provided by a Trust Region Filter scheme. Linearly Constrained subproblems are solved by glcpd.f, which is limited memory spectral gradient method based on Ritz values. The subroutines qlcpd.f and filterSD.f were always used with their default parameters [60].

At each iteration of the IR algorithm, before the execution of Step 1, we test the stopping criteria

$$\|h(x^k)\| \leq \varepsilon_{feas} \tag{48}$$

and

$$\|P_\Omega(x^k - [\nabla f(x^k) + \nabla h(x^k)\lambda^k]) - x^k\| \leq \varepsilon_{opt}. \tag{49}$$

In the experiments we used $\varepsilon_{feas} = \varepsilon_{opt} = 10^{-8}$.

In the Initialization Step we chose $x^0 = x_c$, $\lambda_0 = 0$, and $\theta_{-1} = 0.99$. The line search in Step 4.3 is performed taking $t_k = 1$ and halving it until conditions (9) and (10) are satisfied.

We tested four instances of the IR algorithm:

- FF: This version corresponds, essentially, to the the Fischer-Friedlander approach to Inexact Restoration. The approximations to the Lagrange multipliers are all null, which means that the merit function is a nonsmooth penalty $\phi(x, \theta) = \theta f(x) + (1 - \theta)\|h(x)\|$. The direction obtained in the optimization phase is the projection of the gradient on the tangent space, which correspond to set H_k equal to the identity in (26) for all the iterations. We also enforce a 50% improvement in the feasibility phase and we set $r_k = s_k = 0.5$ for all k.

- Flex-FF: This is a flexible version of FF. In this version we still do not make use of the Lagrange multipliers and we choose $d^k = P_k(y^{k+1} - \nabla f(y^{k+1})) - y^{k+1}$. However we increase the chance of the optimization step to be accepted by choosing $r_k = 0.99$ if $h(x^k) = 0$ and $r_k = \max\left\{0.99, \frac{h(y^{k+1})}{h(x^k)}\right\}$ otherwise, and taking $s_k = 0$ for all k.

- Q-FF: In this version we use $\phi(x, \theta) = \theta f(x) + (1 - \theta)\|h(x)\|$ as the merit function. On the other hand we use the Lagrange multipliers estimators to set d^k as the solution of the quadratic problem (26)-(27) with $H_k = \nabla^2 L(y^{k+1}, \bar{\lambda}^{k+1})$, where $\bar{\lambda}^{k+1}$ is the vector of Lagrange multipliers of the subproblem solved at the last Optimization Phase subproblem. We set $r_k = 0.5$ and $s_k = 0$.

- New IR: This is the complete version the Algorithm 2.1. Here we use $r_k = 0.5$, we set $s_k = 0$ for all k, and we use the approximations of Lagrange multipliers both in the merit function as in the definition of the quadratic Hessian. In other words, we obtain the direction d^k as the solution of the quadratic problem (26)-(27) with $H_k = \nabla^2 L(y^{k+1}, \lambda^{k+1})$, and we use the sharp Lagrangian $\Phi(x, \lambda, \theta) = \theta L(x, \lambda) + (1 - \theta)\|h(x)\|$ as the merit function.

19

The Nonlinear Programming reformulation of problem (46) consists of replacing the constraint that says that x minimizes $\sum_{i=1}^{q} f_i(x)$ with the KKT conditions of this problem. This approach will be denominated "KKT" form now on.

The results of the comparison of the Nonlinear Programming reformulation against "KKT" are reported below. Note that this is not a comparison between IR and the Nonlinear Programming code described in [60] since, in fact, we used this Nonlinear Programming code, or a subroutine of it, both in our Restoration and in our Optimization Phase. It is a comparison between the approach that tries to solve (46) using the KKT conditions as constraints, without further information, and the IR algorithms that employ the minimization structure of those constraints.

1. **Portfolios:** All these problems were successfully solved both by the IR algorithms as by the Nonlinear Programming reformulation using $n \in \{8, 100, 1000\}$. The execution time for the two first cases was smaller than 1 second and for $n = 1000$ it was smaller than 1.5 seconds.

2. **MGH-generated Problems:** In Figure 1 we show the data profile for the MGH-generated problems. We note that by reformulating the lower level problem with its KKT conditions, only 45% of the problems are solved, however, when the problem is solved, the computer time is much smaller than the one used by the IR algorithms. The original FF approach solves 85.8% of the problems, while Flex-FF solves 79.17%. Therefore, the strategy of accepting more steps aiming to avoid the Maratos effect, resulted in a weaker performance. On the other hand, the use of second order information and Lagrange multipliers in Q-Flex really improves the performance, increasing to 90.8% the number of solved problems. The full IR version of the inexact restoration method including Lagrange multipliers at the merit functions performed best in this test by solving 92.5% of the problems.
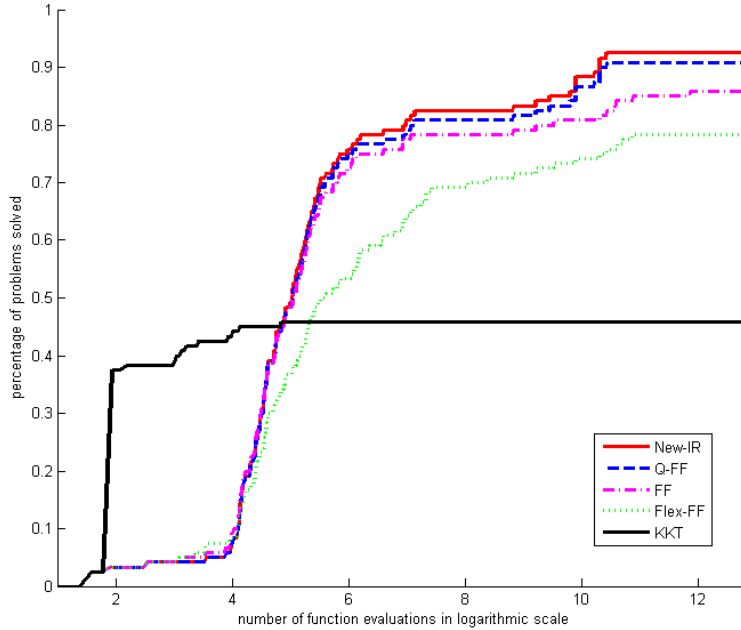

Figure 1: Data profile for the 120 test problems based on the Moré-Garbow-Hillstrom test collection

3. **Quartic Polynomials:** We solved 100 problems for each value of $n \in \{1, 10, 20\}$. In all the 300 problems the IR algorithm found a Pareto point. However, using the KKT approach, we obtained Pareto points only in 87 problems with $n = 1$, 16 problems with $n = 10$ and none of the problems for $n = 20$.

7   Conclusions

Modern Inexact Restoration methods share the following characteristics:

- Very mild assumptions on the method are used for Restoration.

- Approximate minimization, on the tangent set, of a function that resembles the objective function restricted to the feasible region (the function itself, the Lagrangian, or their linear or quadratic approximations)

Different methods differ in two main aspects:

- Mechanism used to accept or reject trial points;

- Functions employed to compare a current iterate with a trial point.

With respect to the first item above we distinguish between methods that use trust regions, filters or line searches. The second item distinguishes between methods that use Lagrange multipliers in the comparison current-trial and those which rely only on functional and constraint values.

With these parameters we are able to construct the following table, where we classify the methods by Martínez and Pilotta [11], Martínez [10], Gonzaga, Karas and Vanti [14], Fischer and Friedlander [22] and the method presented in this paper.

Method	Mechanism	Comparison current-trial
Martínez–Pilotta	Trust regions	Objective function and infeasibility
Martínez	Trust regions	Lagrangian and infeasibility
Gonzaga-Karas-Vanti	Filters	Objective function and infeasibility
Fischer–Friedlander	Line searches	Objective function and infeasibility
New method	Line searches	Lagrangian and infeasibility

The methods of Fischer and Friedlander and the one presented in this paper are, in principle, preferable over the ones of Martínez and Pilotta and Martínez because they allow one to truly prove boundedness of the penalty parameter. Boundedness arguments are also present in the other methods but only in proofs by contradiction. (In [10] and [11] boundedness is proved under the false assumption of non-convergence. Therefore, the penalty parameters could be unbounded for these methods in the case of convergence.) A boundedness proof for the methods given in [10] and [11] remains to be an open problem. Of course this question makes no sense in the case of filter-based methods.

Methods based on Lagrangians should be less prone to Maratos-like effects than methods based only on the true objective functions. Lagrangians on the tangent space are better representations of the objective function on the feasible region than the true objective function, a fact that leads one to conjecture that Newtonian methods based on Lagrangians share the non-Maratos properties of unconstrained Newton methods, for which the unitary step is always acceptable. However, the proof that the methods in which the merit function is the sharp Lagarangian are Maratos-free is still an open problem.

Under reasonable assumptions, we proved the global convergence of the new IR algorithm.

The Inexact Restoration framework fits well to problems that have some structure that can be explored in the Restoration and/or in the Optimization phase. We introduced a new approach to deal with multiobjective optimization problems under the Weighted-Sum Scalarization approach. Those problems have the appropriate structure to be solved with Inexact Restoration algorithms. We tested our algorithms using a a set of problems that includes Portfolio Decisions, 120 problems derived from the MGH collection, and 300 fourth degree polynomial problems. The numerical results show that most theoretical improvements contribute to a better practical performance of the algorithm.

Future research will include the application of alternative IR approaches to wider families of multiobjective problems [34] and the application of the IR techniques introduced in this paper to minimization on the Pareto frontier using more sophisticated scalarizations [52, 53, 54, 51]. Moreover, restoration procedures not based on scalarizations [61] should also be analyzed from the point of view of Inexact Restoration.

References

[1] Abadie, J., Carpentier, J.: Generalization of the Wolfe reduced-gradient method to the case of nonlinear constraints. In: Fletcher, R. (ed.): Optimization, 37–47. Academic Press, New York (1968)

[2] Lasdon, L. S.: Reduced gradient methods. In: Powell, M. J. D. (ed.): Nonlinear Optimization 235–242. Academic Press, New York (1982)

[3] Miele, A., Huang, H. Y., Heideman, J. C.: Sequential gradient-restoration algorithm for the minimization of constrained functions, ordinary and conjugate gradient version. J. Optim. Theory Appl. 4, 213–246 (1969)

[4] Miele, A., Levy, A. V., Cragg, E. E.: Modifications and extensions of the conjugate-gradient restoration algorithm for mathematical programming problems. J. Optim. Theory Appl. 7, 450–472 (1971)

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

[6] Rom, M., Avriel, M.: Properties of the sequential gradient-restoration algorithm (SGRA), Part 1: Introduction and comparison with related methods. J. Optim. Theory Appl. 62, 77–98 (1989)

[7] Rom, M., Avriel, M.: Properties of the sequential gradient-restoration algorithm (SGRA), Part 2: Convergence Analysis. J. Optim. Theory Appl. 62, 99–126 (1989)

[8] Rosen, J. B.: The gradient projection method for nonlinear programming, Part 1, Linear constraints. SIAM J. Applied Math. 8, 181–217 (1960)

[9] Rosen, J. B.: The gradient projection method for nonlinear programming, Part 2, Nonlinear constraints. SIAM J. Applied Math. 9, 514–532 (1961)

[10] Martínez, J. M.: Inexact restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. J. Optim. Theory Appl. 111, 39–58 (2001)

[11] Martínez, J. M., Pilotta, E. A.: Inexact restoration algorithms for constrained optimization. J. Optim. Theory Appl. 104, 135–163 (2000)

[12] Birgin, E. G., Martínez, J. M.: Local convergence of an Inexact-Restoration method and numerical experiments. J. Optim. Theory Appl. 127, 229–247 (2005)

[13] Karas, E. W., Gonzaga, C. C. , Ribeiro, A. A.: Local convergence of filter methods for equality constrained nonlinear programming. Optim. 59, 1153–1171 (2010)

[14] Gonzaga, C. C., Karas, E. W., Vanti, M.: A globally convergent filter method for Nonlinear Programming. SIAM J. Optim. 14, 646–669 (2003)

[15] Karas, E. W., Pilotta, E. A., Ribeiro, A. A.: Numerical comparison of merit function with filter criterion in inexact restoration algorithms using Hard-Spheres Problems. Comput. Optim. Appl. 44, 427–441 (2009)

[16] Andreani, R., Castro, S. L., Chela, J. L., Friedlander, A., Santos, S. A.: An inexact-restoration method for nonlinear bilevel programming problems. Comput. Optim. Appl. 43, 307–328 (2009)

[17] Bueno, L. F., Friedlander, A., Martínez, J. M., Sobral, F.: Inexact Restoration method for Derivative-Free Optimization with smooth constraints. SIAM J. Optim. 23, 1189–1213 (2013)

[18] Francisco, J. B, Martínez, J. M., Martínez, L., Pisnitchenko, F. I.: Inexact Restoration method for minimization problems arising in electronic structure calculations. Comput. Optim. Appl. 50, 555–590 (2011)

[19] Gomes-Ruggiero, M. A., Martínez, J. M., Santos, S. A.: Spectral projected gradient method with inexact restoration for minimization with nonconvex constraints. SIAM J. Sci. Comput. 31, 1628–1652 (2009)

[20] Kaya, C. Y.: Inexact Restoration for Runge-Kutta discretization of Optimal Control problems. SIAM J. Numer. Anal. 48, 1492–1517 (2010)

[21] Kaya, C. Y., Martínez, J. M.: Euler discretization and Inexact Restoration for Optimal Control. J. Optim. Theory Appl. 134, 191–206 (2007)

[22] Fischer, A., Friedlander, A.: A new line search inexact restoration approach for nonlinear programming. Comput. Optim. Appl. 46, 333–346 (2010)

[23] Martínez, J. M., Svaiter, B. F.: A practical optimality condition without constraint qualifications for nonlinear programming. J. Optim. Theory Appl. 118, 117–133 (2003)

[24] Andreani, R., Haeser, G., Martínez, J. M.: On sequential optimality conditions for smooth constrained optimization. Optim. 60, 627–641 (2011)

[25] Barzilai, J., Borwein, J. M.: Two point step size gradient method. IMA J. Numer. Anal. 8, 141–148 (1988)

[26] Dai, Y. H., Fletcher, R.: On the asymptotic behaviour of some new gradient methods. Math. Prog. 103, 541–559 (2005)

[27] Friedlander, A., Martínez, J. M., Molina, B., Raydan, M.: Gradient method with retards and generalizations. SIAM J. Numer. Anal. 36, 275–289 (1998)

[28] Raydan, M.: On the Barzilai and Borwein choice of steplength for the gradient method. IMA J. Numer. Anal. 13, 321–326 (1993)

[29] Raydan, M.: The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. SIAM J. Optim. 7, 26–33 (1997)

[30] Birgin, E. G., Martínez, J. M., Raydan, M.: Nonmonotone spectral projected gradient methods on convex sets. SIAM J. Optim. 10, 1196–1211 (2000)

[31] Birgin, E. G., Martínez, J. M., Raydan, M.: Algorithm 813: SPG - Software for convex-constrained optimization. ACM Trans. Math. Software 27, 340–349 (2001)

[32] Birgin, E. G., Martínez, J. M., Raydan, M.: Inexact Spectral Projected Gradient methods on convex sets. IMA J. Numer. Anal. 23, 539–559 (2003)

[33] Andreani, R., Birgin, E. G., Martínez, J. M., Yuan. J-Y.: Spectral projected gradient and variable metric methods for optimization with linear inequalities. IMA J. Numer. Anal. 25, 221–252 (2005)

[34] Ramirez, V.: Um problema Bilevel-Multiobjetivo com Pontos Propriamente Eficientes. Ph.D. proposal, Departamento de Matemática Aplicada, Universidade Estadual de Campinas (2013)

[35] Dempe, S.: Foundations of Bilevel Programming. Kluwer Academic Publishers, Dordrecht (2002)

[36] Rockafellar, R. T.: Lagrange multipliers and optimality. SIAM Rev. 35, 183–238 (1993)

[37] Burachik, R. S., Gasimov, R. N., Ismayilova, A. N., Kaya, C. Y.: On a modified subgradient algorithm for dual problems via sharp augmented Lagrangian. J. Global Optim. 34, 55–78 (2006)

[38] Burachik, R. S., Iusem, A. N., Melo, J. G.: A primal dual modified subgradient algorithm with sharp Lagrangian. J. Global Optim. 46, 347–361 (2010)

[39] Burachik, R. S., Kaya, C. Y., Mammadov, M.: An inexact modified subgradient algorithm for nonconvex optimization. Comput. Optim. Appl. 45, 1–24 (2010)

[40] Bertsekas, D. P.: Nonlinear Programming, 2nd edition. Athena Scientific, Belmont, Mass. (1999)

[41] Andreani, R., Haeser, G., Schuverdt, M. L., Silva, P. J. S.: Two new constraint qualifications and applications. SIAM J. Optim. 22, 1109–1135 (2012)

[42] Andreani, R., Martínez, J. M., Schuverdt, M. L.: On the relation between the Constant Positive Linear Dependence condition and quasinormality constraint qualification. J. Optim. Theory Appl. 125, 473–485 (2005)

[43] Andreani, R., Haeser, G., Schuverdt, M. L., Silva, P. J. S.: A relaxed constant positive linear dependence constraint qualification and applications. Math. Prog. 135, 255–273 (2012)

[44] Qi, L., Wei, Z.: On the constant positive linear dependence condition and its application to SQP methods. SIAM J. Optim. 10, 963–981 (2000)

[45] Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optim. 60, 429–440 (2011)

[46] Bueno, L. F.: Otimização com restrições LOVO, Restauração Inexata e o Equilíbrio Inverso de Nash. Ph.D. dissertation, Departamento de Matemática Aplicada, Universidade Estadual de Campinas (2011)

[47] Martínez, J. M., Pilotta, E. A.: Inexact restoration methods for nonlinear programming: advances and perspectives. In: Qi, L. Q., Teo, K. L., Yang, X. Q. (eds.): Optimization and Control with applications, pp. 271–292, Springer (2005)

[48] Miettinen, K. M.: Nonlinear Multiobjective Optimization. Kluwer Academic, Boston (1999)

[49] Gass, S., Saaty, T.: The computational algorithm for the parametric objective function. Naval Res. Logis. Quart. 2, 39–45 (1955)

[50] Zadeh, L.: Optimality and non-scalar-valued performance criteria. IEEE Trans. Automat. Control 8, 59–60 (1963)

[51] Pascoletti, A., Serafini, P.: Scalarizing vector optimization problems. J. Optim. Theory Appl. 42, 499–524 (1984)

[52] Burachik, R. S., Kaya, C. Y., Rizvi, M. M.: A new scalarization technique to approximate Pareto fronts on problems with disconnected feasible sets. J. Optim. Theory Appl. (Published online, 23 May 2013: DOI 10.1007/s10957-013-0346-0)

[53] Dutta, J., Kaya, C. Y.: A new scalarization and numerical method for constructing weak Pareto front of multi-objective optimization problems. Optim. 60, 1091–1104 (2011)

[54] Eichfelder, G.: Scalarization Methods in Multiobjective Otimization. Springer, Berlin, Heidelberg (2008)

[55] Bonnel, H., Kaya, C. Y.: Optimization over the efficient set of multi-objective convex optimal control problems. J. Optim. Theory Appl. 147, 93–112 (2010)

[56] Dauer, J. P., Fosnaugh, T. A.: Optimization over the efficient set. J. Global Optim. 7, 261–277 (1995)

[57] Yamamoto, Y.: Optimization over the efficient set: overview. J. Glob. Optim. 22, 285–317 (2002)

[58] Moré, J. J., Garbow, B. S., Hillstrom, K. E.: Testing Unconstrained Optimization Software. ACM Trans. Math. Software 7, 17–41 (1981)

[59] Yahoo Finance website, http://finance.yahoo.com/

[60] Fletcher, R.: A Sequential Linear Constraint Programming Algorithm for NLP. SIAM J. Optim. 22, 772–794 (2012)

[61] Fliege, J., Graña-Drummond, L. F., Svaiter, B. F.: Newton's Method for Multiobjective Optimization, SIAM J. Optim 20, 602–626 (2009).