Spectral Projected Gradient Method with Inexact Restoration for Minimization with Nonconvex Constraints

M. A. Gomes-Ruggiero [*] J. M. Martínez [†] S. A. Santos [‡]

September 17, 2007

Abstract

This work takes advantage of the spectral projected gradient direction within the inexact restoration framework to address nonlinear optimization problems with nonconvex constraints. The proposed strategy includes a convenient handling of the constraints, together with nonmonotonic features to speed up convergence. The numerical performance is assessed by experiments with hard-spheres problems, pointing out that the inexact restoration framework provides the adequate environment for the extension of the spectral projected gradient method for general nonlinearly constrained optimization.

Key words: Inexact restoration, nonlinear programming, spectral projected gradients, hard-spheres problems.

1 Introduction

Spectral gradient methods for optimization were originated in the seminal paper of Barzilai and Borwein [3]. The idea is that gradient directions, when coupled with suitable *spectral* steplengths, produce much better results than the traditional Cauchy approach and, many times, are competitive with the ones produced by Newton and quasi-Newton methods, being quite attractive due to their low computational cost. The original method by Barzilai and Borwein was generalized by Raydan [36] for large scale unconstrained minimization. Later, the method was extended for convex constrained problems in [9]. The *Spectral Projected Gradient* method defined in [9] was showed to be astonishingly efficient for solving very large problems in which projections can be computed cheaply [10].

This state of facts motivate the idea of finding a natural generalization of the spectral gradient method for general nonlinearly constrained problems, where projections on the feasible set are

[*]Department of Applied Mathematics, State University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. This author was supported by PRONEX-Optimization (PRONEX - CNPq / FAPERJ E-26 / 171.164/2003 - APQ1), FAPESP (Grants 01/04597-4 and 06/51827-9) and CNPq. e-mail: marcia@ime.unicamp.br

[†]Department of Applied Mathematics, State University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. This author was supported by PRONEX-Optimization (PRONEX - CNPq / FAPERJ E-26 / 171.164/2003 - APQ1), FAPESP (Grants 01/04597-4 and 06/51827-9) and CNPq. e-mail: martinez@ime.unicamp.br

[‡]Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. This author was supported by PRONEX-Optimization (PRONEX - CNPq / FAPERJ E-26 / 171.164/2003 - APQ1), FAPESP (Grants 01/04597-4 and 06/51827-9) and CNPq. e-mail: sandra@ime.unicamp.br

not easy to compute. This is the problem addressed in the present paper. Our point of view is that the Inexact Restoration Algorithm [30] provides an adequate framework for the required extension.

We consider the nonlinear programming problem in the form

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & C(x) \le 0, \;\; x \in \Omega, \end{array} \tag{1}$$

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 closed and convex. In practice, we are mostly interested in the case in which Ω is a polytope.

To address problem (1) we develop upon the inexact restoration method proposed by Martínez and Pilotta [30]. Each equality constraint appearing in the original formulation is assumed to have been transformed into two inequalities.

Generally speaking, inexact restoration iterations are composed of two phases. In the first one, *the restoration phase*, feasibility is improved by the fulfillment of two control conditions. In the second phase, a tangential decrease of a pre-defined optimality measure is produced in the intersection of a convenient linear approximation to the feasible set and a trust region centered at the currently more feasible point. This is *the optimization phase*. The acceptance of the trial point produced in the optimization phase rests upon a sufficient decrease of an elected merit function.

The original method introduced in [30] used the objective function in the optimality phase and the exact penalty as the merit function, with a nonmonotone updating of the penalty parameter proposed in [21]. In [29] the equality constrained nonlinear programming problem with bounds was addressed, with Lagrangian tangent decrease and the sharp Lagrangian as the merit function. Global convergence results were proved for both algorithms. Gonzaga, Karas e Vanti [23] adopted the filter approach instead of a merit function to accept the current trial point, coming up with a globally convergent algorithm as well. Birgin and Martínez [8] analyzed the local features of the algorithm proposed in [29]. Recently, a strategy was devised in [22] to bypass the actual usage of slack variables in the inequality constraints, making it possible a Lagrangian decrease in the optimality phase.

The Spectral Projected Gradient (SPG) method [9, 10, 11] combines the spectral ideas introduced by Barzilai and Borwein [3], and generalized by Raydan [35, 36], with classical projected gradient strategies [5, 20, 28]. The surprisingly effective performance of the spectral gradient for unconstrained problems [36] motivated the extension of the approach, by means of projections, for problems with more general convex constraints [1, 7, 9, 10, 11]. Nonmonotone strategies, like the one proposed by Grippo, Lampariello and Lucidi [24], turned out to be an essential ingredient for the success of the spectral gradients for unconstrained minimization and further extensions.

The SPG method is applicable to convex constrained problems in which the projection on the feasible set is easy to compute. It has been intensively used in distinct applied contexts [4, 12, 16, 26, 34, 38, 40, 44]. Moreover, it has been the object of several spectral-parameter modifications [19, 37, 43], alternative nonmonotone strategies have been suggested [9, 13, 15], convergence and stability properties have been elucidated [14, 41, 42] and it has been combined with other algorithms for particular optimization problems. Linearly constrained optimization were addressed in [1, 31], whereas nonlinear systems were considered in [25, 45]. For general nonlinear programming, a method combining the SPG with the augmented Lagrangian was proposed in [18]. However, in [18] and [31], the SPG method was used as an auxiliary tool in the context of penalty-like approaches. It remained open, therefore, the usage of the SPG for nonconvex constraints within the feasibility paradigm.

In this work we propose a way to plug SPG ideas into the inexact restoration method. As a result, a natural coupling of these two perspectives is devised to solve problems with nonconvex constraints, keeping feasibility under a certain control. The key was to give more freedom to the scaling parameter that defines the tangent direction in the SPG approach, and consequently, to *increase the role* of such direction, previously taken as a generalized Cauchy step and mainly used to ensure global convergence. Here the trust-region scenario is naturally replaced by the linesearch framework along the tangent direction, implying in adaptations in the convergence results.

This paper is organized as follows: in Section 2 we state the main algorithm, and give details about the computation of the spectral parameter, explain the mechanism of the nonmonotone strategy, and provide details on handling equality constraints in the current inexact restoration context. Section 3 is devoted to the theoretical results, stating the necessary adjustments of the global convergence properties of [30] to be valid for Algorithm 2.1. An analysis is given on the effect of replacing each equality constraint by two inequality ones as far as constraint qualifications are concerned. In Section 4, the numerical experiments are presented, prepared with the challenging family of hard-spheres problems. Specific features of the implementation are discussed, together with a complete analysis of the results. Finally, some conclusions are drawn in Section 5.

Notation.

We use two (perhaps different) norms. As in [30], we denote $|\cdot|$ a monotone norm on $I\!R^m$ ($0 \le v \le w \Rightarrow |v| \le |w|$) and $\|\cdot\|$ an arbitrary norm on $I\!R^n$.

We denote $C'(x) \in I\!R^{m\times n}$ the Jacobian matrix of $C(x)$ and $C_j'(x) = \nabla C_j(x)^T$ for all $j = 1,\dots,m$.

We also denote $C_j^+(x) = \max\{C_j(x), 0\}$ and $C^+(x) = (C_1^+(x),\dots,C_m^+(x))^T$.

2 Main Algorithm

In this section we introduce the main algorithm that combines inexact restoration with the spectral gradient approach. Projecting a point on a nonconvex set, as required in the SPG method, is generally as difficult as the original optimization problem. Therefore, the projetions used in SPG must be replaced by partial restoration steps with minimal requirements. After restoration, the computation of the spectral projected gradient direction on a linear approximation of the feasible set takes place. In the Inexact Restoration method of [30] the initial gradient step merely guaranteed theoretical global convergence. Here, due to the spectral choice of the steplength, we also expect efficiency from the gradient step.

Algorithm 2.1 Initialize $k \leftarrow 0$, $\theta_{-1} \in (0,1)$, $r \in [0,1), \beta > 0, \xi > 0$, $0 < \eta_{min} < \eta_{max} < \infty$, ε_{feas}, $\varepsilon_{opt} > 0$. Assume that $\sum_{k=0}^{\infty} \omega_k$ is a convergent series of nonnegative terms. Let $x^0 \in \Omega$ be the initial approximation to the solution of the Nonlinear Programming problem (1).

Step 1. *Restoration.*

Compute $y^k \in \Omega$ such that

$$|C^+(y^k)| \leq r|C^+(x^k)| \tag{2}$$

and

$$\|y^k - x^k\| \leq \beta|C^+(x^k)|. \tag{3}$$

If it is impossible to obtain such y^k, declare 'failure in improving feasibility' and terminate the execution.

Step 2. *Definition of the penalty parameter.*

Compute

$$\theta_{k,-1} = \min\{1, \min\{\theta_{-1}, \ldots, \theta_{k-1}\} + \omega_k\}.$$

Step 3. *Definition of the tangent set.*

$$\pi_k = \{z \in \Omega \ | \ C_j(y^k) + C_j'(y^k)(z - y^k) \leq C_j^+(y^k) \text{ whenever } C_j(y^k) \geq -\xi\}. \tag{4}$$

Step 4. *Compute the search direction and check stopping criteria.*

Compute the spectral parameter $\eta_k \in [\eta_{min}, \eta_{max}]$ and the spectral tangent direction

$$d^{k,tan} = P_k(y^k - \eta_k \nabla f(y^k)) - y^k, \tag{5}$$

where $P_k(z)$ denotes the orthogonal projection of z on the set π_k.

If $|C^+(y^k)| < \varepsilon_{feas}$ and $\|d^{k,tan}\| < \varepsilon_{opt}$ then terminate the execution. Otherwise, go to Step 5.

Step 5. *Obtain tangent decrease.*

Define t_{dec} as the first term t of the sequence $\{t_1, t_2, \ldots\}$ such that

$$f(y^k + td^{k,tan}) \leq f(y^k) + 0.1t\langle \nabla f(y^k), d^{k,tan}\rangle, \tag{6}$$

where $\{t_j\}$ is defined by $t_1 = 1$ and $t_{j+1} \in [0.1t_j, 0.9t_j]$ for all $j = 1, 2, \ldots$.

Set $i \leftarrow 0$ and define $z^{k,i} = y^k + t_{dec}d^{k,tan}$.

Step 6. *Update the penalty parameter.*

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

$$\theta[f(x^k) - f(z^{k,i})] + (1-\theta)[|C^+(x^k)| - |C^+(y^k)|] \geq \frac{1}{2}[|C^+(x^k)| - |C^+(y^k)|]. \tag{7}$$

Step 7. *Acceptance test.*

Set $t_{k,i} = t_{dec}$.

Define

$$\mathbf{Ared}_{k,i} = \theta_{k,i}[f(x^k) - f(z^{k,i})] + (1-\theta_{k,i})[|C^+(x^k)| - |C^+(z^{k,i})|] \tag{8}$$

and

$$\mathbf{Pred}_{k,i} = \theta_{k,i}[f(x^k) - f(z^{k,i})] + (1-\theta_{k,i})[|C^+(x^k)| - |C^+(y^k)|]. \tag{9}$$

If

$$\mathbf{Ared}_{k,i} \geq 0.1\mathbf{Pred}_{k,i}, \tag{10}$$

define $x^{k+1} = z^{k,i}$, set $\theta_k = \theta_{k,i}$ and $iacc(k) = i$, where $iacc$ means the "accepted i". Update $k \leftarrow k+1$ and go to Step 1.

Otherwise, update $i \leftarrow i+1$, $t_{k,i} \leftarrow \dfrac{t_{k,i}}{2}$, $z^{k,i} \leftarrow y^k + t_{k,i}d^{k,tan}$ and go to Step 6.

A key aspect of this and other inexact restoration method is the presence of the conditions (2) and (3). The projection of x on a given set is the point *in the set* that is *closest* to x. The requirements (2) and (3) express the way in which the two basic properties of projections can be relaxed. Our restored point does not need not be *in the set* but must be close to the set in the sense of (2). The restored point does not need to be *the closest* but must not be suitably close to x in the sense of (3). Sufficient conditions for the fulfillment of (2) and (3) may be found in [29].

2.1 Some remarks and elementary properties

Analyzing Steps 3 and 4, it is easy to verify that $d^{k,tan}$ is a descent direction. In fact, since $y^k \in \pi_k$, we have that

$$\|(y^k - \eta_k \nabla f(y^k)) - P_k(y^k - \eta_k \nabla f(y^k))\|_2 \leq \|(y^k - \eta_k \nabla f(y^k)) - y^k\|_2,$$

as illustrated in Figure 1. Therefore,

$$\begin{aligned} \|y^k - P_k(y^k - \eta_k \nabla f(y^k))\|_2^2 + \|\eta_k \nabla f(y^k)\|_2^2 + 2\eta_k \langle P_k(y^k - \eta_k \nabla f(y^k)) - y^k, \nabla f(y^k)\rangle \\ \leq \|\eta_k \nabla f(y^k)\|_2^2, \end{aligned}$$

and so,

$$\langle d^{k,tan}, \nabla f(y^k)\rangle \leq \frac{-1}{2\eta_{max}} \|d^{k,tan}\|_2^2 \leq \frac{-c}{2\eta_{max}} \|d^{k,tan}\|^2, \tag{11}$$

where $c > 0$ is a norm-dependent constant.

Figure 1: Geometric representation of the main elements of Steps 3 and 4 of Algorithm 2.1.

We can use classical arguments for justifying backtracking with Armijo-like conditions (see [17], Chapter 6), to show that t_{dec} is well defined at Step 5. In other words, given the current point x^k it is possible to compute, in finite time, the trial point $z^{k,i}$.

In Steps 6 and 7 a merit function is used to perform the comparison between the new approximation $z^{k,i}$ and the current point x^k. We use the same exact penalty-like nonsmooth merit function proposed in [30]:

$$\psi(x, \theta) = \theta f(x) + (1 - \theta)|C^+(x)| \tag{12}$$

where $\theta \in (0, 1]$ is a penalty parameter used to give different weights to the objective function and to the feasibility objective. This parameter is computed in Step 2, following ideas from [21].

The merit function at $z^{k,i}$ should be less than the merit function at x^k, so that the "actual reduction of the merit function", given in (8), can be rewritten as

$$\mathbf{Ared}_{k,i} = \psi(x^k, \theta_{k,i}) - \psi(z^{k,i}, \theta_{k,i}) > 0.$$

To provide a "sufficient reduction" of the merit function, given by (10), a "predicted reduction" of the merit function between x^k and $z^{k,i}$ was introduced in (9). The quantity $\mathbf{Pred}_{k,i}$ can be nonpositive depending on the value of the penalty parameter. Fortunately, if $\theta_{k,i}$ is small enough, $\mathbf{Pred}_{k,i}$ is arbitrarily close to $|C(x^k)| - |C(y^k)|$ which is necessarily nonnegative. Therefore, we will always be able to choose $\theta_{k,i} \in (0,1]$ such that relation (7) is verified:

$$\mathbf{Pred}_{k,i} \geq \frac{1}{2}[|C^+(x^k)| - |C^+(y^k)|]. \tag{13}$$

When condition (10) is satisfied, we accept $x^{k+1} = z^{k,i}$. Otherwise, we reduce the trial step.

2.2 Computation of the spectral parameter η_k

In order to adapt the spectral projected gradient (SPG) ideas to nonlinear programming problems with nonconvex constraints, we briefly recall some of the ideas from Birgin, Martínez and Raydan [11], starting from a convex problem

$$\begin{array}{ll} \text{minimize} & \phi(x) \\ \text{subject to} & x \in \Omega, \end{array}$$

where $\phi : I\!R^n \to I\!R$ is convex and $\Omega \subset I\!R^n$ is closed and convex.

Their approach rests upon solving approximately a sequence of subproblems which consist of minimizing a quadratic model for the decrease of the objective function within the original convex feasible set. In other words, at each iteration the algorithm finds $\bar{d}$, an approximated solution to

$$\begin{array}{ll} \text{minimize} & Q_k(d) \equiv \frac{1}{2}d^T B_k d + g_k^T d \\ \text{subject to} & x^k + d \in \Omega, \end{array}$$

where $g_k = \nabla\phi(x^k)$ and $B_k \approx \nabla^2\phi(x^k)$. The spectral choice for the Hessian of the model is a key point of the SPG method. Indeed, the Hessian is approximated by the scalar matrix $B_k = \frac{1}{\lambda^k_{SPG}} I$, where

$$\lambda^k_{SPG} = \begin{cases} \min\left\{\lambda_{\max}, \max\left\{\lambda_{\min}, \frac{\langle s^k, s^k\rangle}{\langle s^k, u^k\rangle}\right\}\right\} & \text{if } \langle s^k, u^k\rangle > 0 \\ \lambda_{\max} & \text{otherwise} \end{cases} \tag{14}$$

with $s^k = x^k - x^{k-1}$ and $u^k = g_k - g_{k-1}$. Thus, if the safeguarding scheme is not activated, the expression

$$\frac{\langle s^k, u^k\rangle}{\langle s^k, s^k\rangle} = \frac{\left\langle s^k, \left(\int_0^1 \nabla^2\phi(x^k + ts^k)dt\right) s^k\right\rangle}{\langle s^k, s^k\rangle}$$

encompass a Rayleigh quotient and thus, the scalar matrix $\frac{1}{\lambda^k_{SPG}} I$ is an average Hessian. The safeguarding scheme ensures that the model remains convex, projecting the scalar λ^k_{SPG} into the interval $[\lambda_{\min}, \lambda_{\max}]$, so that the approximated Hessian is bounded.

Now, let us state the problem from which direction $d^{k,tan}$ is computed in Step 4 of Algorithm 2.1. From (5), we have

$$\begin{array}{ll} \text{minimize}_d & \frac{1}{2}\|y^k + d - (y^k - \eta_k \nabla f(y^k))\|_2^2 \\ \text{subject to} & y^k + d \in \pi_k, \end{array}$$

that is,

$$\begin{array}{ll} \text{minimize}_d & \frac{1}{2}\|d + \eta_k \nabla f(y^k))\|_2^2 \\ \text{subject to} & C_j(y^k) + C_j'(y^k)d \leq C_j^+(y^k), \forall j \text{ such that } C_j(y^k) \geq -\xi \\ & y^k + d \in \Omega. \end{array} \tag{15}$$

To accomplish the desired connection, let us focus on the objective function of problem (15). Notice that, in the variable d, the minimization of $d^T d + 2\eta_k d^T \nabla f(y^k) + \eta_k^2 \nabla f(y^k)^T \nabla f(y^k)$ is equivalent to minimizing $\frac{1}{2\eta_k} d^T d + d^T \nabla f(y^k)$, which turns into the SPG quadratic model as long as $\eta_k \equiv \lambda_{SPG}^k$, with $s^k = x^k - x^{k-1}$ and $u^k = \nabla f(x^k) - \nabla f(x^{k-1})$. Therefore, despite the generality of the problem (1), the convexity of problem (15) allows one to adopt the spectral parameter to scale the gradient. It is worth observing that in the original inexact restoration algorithm [30], this scaling parameter was defined as a nonnegative constant: $\eta_k \equiv \eta$.

2.3 A nonmonotone strategy with repeated tangent steps

Another important and desirable feature that is usually present in SPG methods is the usage of some convenient nonmonotone strategy, which generally improves the performance of the algorithm. In our case, the idea is to produce at most a given number of repeated tangent steps as long as the restoration conditions $|C^+(y^k)| \leq r|C^+(x^k)|$ and $\|x^k - y^k\| \leq \beta |C^+(x^k)|$ remain satisfied. In other words, Steps 3 and 4 are repeated $k_{\max}$ times, the updating $y^k \leftarrow y^k + d^{k,tan}$ with stepsize one is taken, provided (2) and (3) remain valid.

Besides allowing longer steps, this strategy favors the method not to be trapped by the feasible set if the optimal point is still far from the current approximation. Moreover, due to the monitoring of conditions (2) and (3), this strategy does not interfere in the global convergence analysis.

2.4 Handling equality constraints

To adjust a general nonlinear programming problem to the format (1), each equality constraint must be turned into two inequalities. In the inexact restoration scenario, this strategy fits better with the feasibility paradigm than keeping the original equalities. This can be followed by assuming, without loss of generality, that $h(y^k) < 0$, and observing that the linear approximation of the single constraint $h(x) = 0$ ($h : I\!R^n \to I\!R$) with respect to the currently more feasible point y^k, given by the set

$$\pi_k^E = \{z \in I\!R^n \mid \nabla h(y^k)^T (z - y^k) = 0\}, \tag{16}$$

with such decoupling, becomes

$$\pi_k^I = \{z \in I\!R^n \mid 0 \leq \nabla h(y^k)^T (z - y^k) \leq -h(y^k)\}. \tag{17}$$

Naturally, the sets (16) and (17) are coincident in case y^k is feasible. At the optimality phase, the set (17) provides more freedom to the algorithm than the set (16), which might help the trial point to conquer feasibility as optimality is improved.

It is also worth mentioning that the parameter ξ, present in definition (4), is in charge of excluding constraints that are strongly satisfied at the current y^k from the linear approximation of the feasible set, so it might be that just one of the inequalities in (17) are actually present in π_k.

3 Theoretical results

In this section we present the theoretical properties of Algorithm 2.1 as far as good definition, achievement of feasibility and convergence to optimality, following the outline of the presentation of [30]. The statements of the results that are analogous to those from [30] are included for completeness, without the proofs. The properties that have suffered some modification, adaptation os simplification are accompanied by the respective demonstrations. The last subsection focuses on constraint qualifications of a general feasible set of nonlinear programming problems as each equality constraint is decoupled into two inequalities. Such properties do not play any role in the global convergence results of Algorithm 2.1. However, they are valuable for analyzing the nature of the generated limit points, according to the specific features of the problem under consideration.

3.1 The algorithm is well defined

Assumptions A1, A2 and A3, stated below, are the only requirements on the nonlinear programming problem (1) that are needed for proving convergence, so they are assumed to hold from now on. In particular, no regularity assumptions are used in the proofs and second derivatives of f and C are not assumed to exist.

A1. The set Ω is convex and compact.
A2. The Jacobian matrix of $C(x)$ exists and satisfies the Lipschitz condition

$$\|C'(y) - C'(x)\| \leq L_1 \|y - x\| \text{ for all } x, y \in \Omega. \tag{18}$$

A3. The gradient of f exists and satisfies the Lipschitz condition

$$\|\nabla f(y) - \nabla f(x)\| \leq L_2 \|y - x\| \text{ for all } x, y \in \Omega. \tag{19}$$

Due to the equivalence of norms on $I\!R^n$, similar conditions to (18) and (19) hold if we consider different norms than $\|\cdot\|$. So, in order to simplify the notation, we can assume that (18) and (19) hold with the same constants L_1 and L_2 for all the norms considered in this work. From these Lipschitz conditions it follows that

$$\|C(y) - C(x) - C'(x)(y - x)\| \leq \frac{L_1}{2} \|y - x\|^2 \tag{20}$$

and

$$|f(y) - f(x) - \langle \nabla f(x), y - x \rangle| \leq \frac{L_2}{2} \|y - x\|^2 \tag{21}$$

for all $x, y \in \Omega$. Again, we can assume, without loss of generality, that (20) and (21) hold for different norms with the same constants and that

$$|C_j(y) - C_j(x) - C_j'(x)(y - x)| \leq \frac{L_1}{2} \|y - x\|^2 \tag{22}$$

for all $j = 1, \ldots, m$.

The assumption on the boundedness of Ω can be replaced by hypotheses that state boundedness of a set of quantities depending on the iterates. This is frequently done in global convergence theories for SQP algorithms. We prefer to state directly Assumption A1 since it seems to be the only reasonable assumption *on the problem* that guarantees boundedness of the required quantities.

The following property is directly deduced from the general assumptions. It states a bounded deterioration result, so that only a second order deterioration of feasibility can be expected for a trial point $z \in \pi_k$ when compared to the feasibility of y^k.

Theorem 3.1. *There exists $c_1 > 0$ (independent of k) such that, whenever $y^k \in \Omega$ is defined and $z \in \pi_k$, we have*

$$|C^+(z)| \leq |C^+(y^k)| + c_1 \|z - y^k\|^2. \tag{23}$$

Proof. See [30, Thm. 3.1]. $\square$

The decrease of the objective function that can be expected when we move from y^k to $z^{k,i}$ is analyzed in the next theorem.

Theorem 3.2. *There exist $c_2 > 0$ (independent of k) such that, whenever $y^k \in \Omega$ is defined and $z^{k,i}$ is computed at Step 5 of Algorithm 2.1, we have that*

$$f(z^{k,i}) \leq f(y^k) - c_2 \|d^{k,tan}\|^2.$$

Proof. By (21) we have that

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L_2}{2} \|y - x\|^2$$

for all $x, y \in \Omega$. Since $y^k + d^{k,tan} \in \Omega$ we have, for all $t \in [0,1]$, that

$$f(y^k + td^{k,tan}) \leq f(y^k) + t\langle \nabla f(y^k), d^{k,tan} \rangle + \frac{t^2 L_2}{2} \|d^{k,tan}\|^2$$

$$= f(y^k) + 0.1t\langle \nabla f(y^k), d^{k,tan} \rangle + 0.9t\langle \nabla f(y^k), d^{k,tan} \rangle + \frac{t^2 L_2}{2} \|d^{k,tan}\|^2.$$

Now, (11) implies that

$$f(y^k + td^{k,tan}) \leq f(y^k) + 0.1t\langle \nabla f(y^k), d^{k,tan} \rangle - \frac{0.9ct\|d^{k,tan}\|^2}{2\eta_{max}} + \frac{t^2 L_2}{2} \|d^{k,tan}\|^2$$

$$= f(y^k) + 0.1t\langle \nabla f(y^k), d^{k,tan} \rangle + \frac{t\|d^{k,tan}\|^2}{2} \left(tL_2 - \frac{0.9c}{\eta_{max}} \right).$$

Therefore, if $t \leq \frac{0.9c}{\eta_{max} L_2}$, we have that

$$f(y^k + td^{k,tan}) \leq f(y^k) + 0.1t\langle \nabla f(y^k), d^{k,tan} \rangle.$$

This implies that $t_{dec} \geq \min \left\{1, \frac{0.09c}{\eta_{max} L_2}\right\}$. It follows that

$$f(y^k + t_{dec} d^{k,tan}) \leq f(y^k) + \min \left\{0.1, \frac{0.009c}{\eta_{max} L_2}\right\} \langle \nabla f(y^k), d^{k,tan} \rangle.$$

So, again by (11), we obtain

$$f(y^k + t_{dec} d^{k,tan}) \leq f(y^k) - \min \left\{ \frac{0.1c\|d^{k,tan}\|^2}{2\eta_{max}}, \frac{0.009c^2\|d^{k,tan}\|^2}{2\eta_{max}^2 L_2} \right\}.$$

Thus,

$$f(y^k + t_{dec}d^{k,tan}) \leq f(y^k) - c_2\|d^{k,tan}\|^2,$$

where $c_2 = \min \left\{ \frac{0.1c}{2\eta_{max}}, \frac{0.009c^2}{2\eta_{max}^2 L_2} \right\}$. $\square$

The following result establishes that Algorithm 2.1 is well defined, that is, for small enough $t_{k,i}$, the inequality (10) is satisfied and, so, the trial point $z^{k,i}$ is accepted as new iterate.

Theorem 3.3. *Algorithm 2.1 is well defined.*

Proof. Observe that

$$\begin{aligned}
&\mathbf{Ared}_{k,i} - 0.1\mathbf{Pred}_{k,i} \\
&= 0.9\theta_{k,i}[f(x^k) - f(z^{k,i})] + (1-\theta_{k,i})[|C^+(x^k)| - |C^+(z^{k,i})|] - 0.1(1-\theta_{k,i})[|C^+(x^k)| - |C^+(y^k)|] \\
&= 0.9\theta_{k,i}[f(x^k) - f(z^{k,i})] + 0.9(1-\theta_{k,i})[|C^+(x^k)| - |C^+(y^k)|] \\
&+(1-\theta_{k,i})[|C^+(x^k)| - |C^+(z^{k,i})|] - (1-\theta_{k,i})[|C^+(x^k)| - |C^+(y^k)|] \\
&= 0.9\mathbf{Pred}_{k,i} + (1-\theta_{k,i})[|C^+(y^k)| - |C^+(z^{k,i})|].
\end{aligned}$$

So, by (7), (9) and (2),

$$\begin{aligned}
\mathbf{Ared}_{k,i} - 0.1\mathbf{Pred}_{k,i} &= 0.9\mathbf{Pred}_{k,i} + (1-\theta_{k,i})[|C^+(y^k)| - |C^+(z^{k,i})|] \\
&\geq 0.45[|C^+(x^k)| - |C^+(y^k)|] - |(|C^+(y^k)| - |C^+(z^{k,i})|)| \\
&\geq 0.45(1-r)|C^+(x^k)| - |(|C^+(y^k)| - |C^+(z^{k,i})|)|.
\end{aligned}$$

Therefore, if $|C^+(x^k)| > 0$, since $\|y^k - z^{k,i}\| = \|t_{k,i}d^{k,tan}\|$ and $|C^+(x)|$ is continuous, it follows that $\mathbf{Ared}_{k,i} - 0.1\mathbf{Pred}_{k,i} \geq 0$ if $\|t_{k,i}d^{k,tan}\| = t_{k,i}\|d^{k,tan}\|$ is small enough. So, we proved that the algorithm is well defined if the current point x^k is infeasible.

If x^k is feasible, (3) implies that $y^k = x^k$ and $|C^+(y^k)| = 0$. If $d^{k,tan} \neq 0$ we have that $f(z^{k,i}) < f(y^k)$ for all $i = 0, 1, 2, \ldots$. So, condition (7) is always satisfied and, consequently, $\theta_{k,i} = \theta_{k,-1}$ for all $i = 0, 1, 2, \ldots$ Thus, in this case, we have

$$\mathbf{Ared}_{k,i} - 0.1\mathbf{Pred}_{k,i} = 0.9\theta_{k,-1}[f(y^k) - f(z^{k,i})] - (1-\theta_{k,-1})|C^+(z^{k,i})|.$$

So, by Theorems 3.1 and 3.2, we obtain that

$$\mathbf{Ared}_{k,i} - 0.1\mathbf{Pred}_{k,i} \geq 0.9\theta_{k,-1}c_2\|d^{k,tan}\|^2 - c_1\|z^{k,i} - y^k\|^2.$$

$$= 0.9\theta_{k,-1}c_2\|d^{k,tan}\|^2 - c_1 t_{k,i}\|d^{k,tan}\|^2 = (0.9\theta_{k,-1}c_2 - c_1 t_{k,i})\|d^{k,tan}\|^2.$$

Therefore, (10) holds if

$$t_{k,i} \leq \left(\frac{0.9\theta_{k,-1}c_2}{c_1}\right)^{1/2}.$$

So, we proved that x^{k+1} is well defined when x^k is feasible and $d^{k,tan} \neq 0$. $\square$

3.2 Convergence to feasibility

The next theorem is an important tool for proving convergence of the model algorithm. It states that the actual reduction effectively achieved at each iteration necessarily tends to zero.

Theorem 3.4. *Suppose that Algorithm 2.1 generates an infinite sequence. Then*

$$\lim_{k\to\infty} \mathbf{Ared}_{k,iacc(k)} = \lim_{k\to\infty} \psi(x^k, \theta_k) - \psi(x^{k+1}, \theta_k) = 0$$

Proof. See [30, Thm. 3.4]. $\square$

An immediate consequence of Theorem 3.4 is that, when Algorithm 2.1 does not stop at Step 1 and generates an infinite sequence, we have that $\lim_{k\to\infty} |C^+(x^k)| = 0$. This means that points arbitrarily close to feasibility are eventually generated.

Corollary 3.5. *If Algorithm 2.1 does not stop at Step 1 for all $k = 0, 1, 2, \ldots$, then*

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

(In particular, every limit point of $\{x^k\}$ is feasible.)

Proof. See [30, Thm. 3.5]. $\square$

3.3 Convergence to optimality

We have proved that, if Algorithm 2.1 does not break down at Step 1, it achieves approximate feasibility up to any desired precision. We are going to prove that, in such case, the optimality indicator $\|d^{k,tan}\|$ cannot be bounded away from zero. In practice, this implies that given arbitrarily small convergence tolerances $\varepsilon_{feas},\ \varepsilon_{opt} > 0$, Algorithm 2.1 eventually finds an iterate x^k such that $\|C^+(x^k)\| \leq \varepsilon_{feas}$ and $\|d^{k,tan}\| \leq \varepsilon_{opt}$. For proving this result, following a reasoning similar to [30], we will proceed by contradiction, assuming that $\|d^{k,tan}\|$ is bounded away from zero for k large enough. From this hypothesis (stated as Hypothesis C below) we will deduce intermediate results that, finally, will lead us to a contradiction.

Hypothesis C. *Algorithm 2.1 generates an infinite sequence $\{x^k\}$ and there exists $\varepsilon > 0$, $k_0 \in \{0, 1, 2, \ldots\}$ such that*

$$\|d^{k,tan}\| \geq \varepsilon \quad \textit{for all} \quad k \geq k_0.$$

Lemma 3.6. *Suppose that Hypothesis C holds. Then, there exists $c_3 > 0$ (independent of k) such that*

$$f(y^k) - f(z^{k,i}) \geq c_3$$

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

Proof. The result follows trivially, with $c_3 = \varepsilon c_2$, from Theorem 3.2 and Hypothesis C. $\square$

Lemma 3.7. *Suppose that Hypothesis C holds. Then, there exists $\varepsilon_1 > 0$, independent of k and i, such that $|C^+(x^k)| \le \varepsilon_1$ implies that $\theta_{k,i} = \theta_{k,i-1}$.*

Proof. Since $\theta_{k,i} \le 1$, let us define $\mathbf{Pred}(1) \equiv \mathbf{Pred}_{k,i}$ with $\theta_{k,i} \equiv 1$. From (9),

$$\mathbf{Pred}(1) = f(x^k) - f(z^{k,i})$$

$$\ge f(y^k) - f(z^{k,i}) - |f(x^k) - f(y^k)| \ge f(y^k) - f(z^{k,i}) - M\|y^k - x^k\|$$

where M is a constant that only depends on the norms and on a bound of $\|\nabla f(x)\|$ on Ω. Therefore, by (3) and Lemma 3.6,

$$\mathbf{Pred}(1) - \frac{1}{2}|C^+(x^k)| \ge c_3 - \left(M\beta + \frac{1}{2}\right)|C^+(x^k)|.$$

Define

$$\varepsilon_1 = \frac{2c_3}{2M\beta + 1}.$$

If $|C^+(x^k)| \le \varepsilon_1$ we have that

$$\mathbf{Pred}(1) - \frac{1}{2}|C^+(x^k)| \ge 0.$$

This implies that condition (13) is valid for any value of $\theta_{k,i}$ in the interval $[0,1]$. In particular $\theta_{k,i-1}$ satisfies (13), as we wanted to prove. □

In the next lemma, we prove that, under Hypothesis C, the penalty parameters $\{\theta_k\}$ are bounded away from zero. It must be warned that this is a property of sequences that satisfy Hypothesis C (which, in turn, will be proved to be non-existent!) and not of *all* the sequences effectively generated by the model algorithm.

Lemma 3.8. *Suppose that Hypothesis C holds. Then, there exists $\bar{\theta} > 0$ such that $\theta_k \ge \bar{\theta}$ for all $k \in \{0, 1, 2, \ldots\}$.*

Proof. We are going to show first that, if $|C^+(x^k)|$ is sufficiently small, a step $t_{k,i}\|d^{k,tan}\|$ that satisfies

$$|C^+(x^k)| \ge \frac{\alpha}{10} t_{k,i}\|d^{k,tan}\| \ge \frac{\alpha}{10} t_{k,i}\varepsilon \equiv \frac{\bar{\alpha}}{10} t_{k,i} \tag{24}$$

is necessarily accepted, where $\alpha > 0$, and $\bar{\alpha} = \alpha\varepsilon$.

In fact, assume that (24) holds. Then, by (13) and (2),

$$\mathbf{Pred}_{k,i} \ge \frac{1}{2}[|C^+(x^k)| - |C^+(y^k)|] \ge \frac{1-r}{2}|C^+(x^k)| \ge \frac{(1-r)\bar{\alpha}}{20} t_{k,i},$$

and so,

$$t_{k,i} \le \frac{20}{(1-r)\bar{\alpha}} \mathbf{Pred}_{k,i}. \tag{25}$$

By (9), (8), and Theorem 3.1,

$$\mathbf{Ared}_{k,i} = \mathbf{Pred}_{k,i} + (1 - \theta_{k,i})[|C^+(y^k)| - |C^+(z^{k,i})|]$$

$$\geq \mathbf{Pred}_{k,i} - (1-\theta_k)c_1\|z^{k,i} - y^k\|^2 \geq \mathbf{Pred}_{k,i} - c_1\|z^{k,i} - y^k\|^2.$$

So, by Hypothesis C:

$$\mathbf{Ared}_{k,i} \geq \mathbf{Pred}_{k,i} - c_1 t_{k,i}^2 \varepsilon^2 \geq \mathbf{Pred}_{k,i} - c_1 t_{k,i}\varepsilon^2 \frac{20}{(1-r)\bar{\alpha}} \mathbf{Pred}_{k,i}.$$

Now, by (24),

$$-t_{k,i} \geq \frac{-10|C^+(x^k)|}{\bar{\alpha}}$$

Therefore,

$$\mathbf{Ared}_{k,i} \geq \left(1 - \frac{200c_1\varepsilon^2}{(1-r)\bar{\alpha}^2}|C^+(x^k)|\right)\mathbf{Pred}_{k,i}.$$

So, if (24) holds and $|C^+(x^k)| \leq \frac{0.9(1-r)\bar{\alpha}^2}{200c_1\varepsilon^2}$, then condition (10) is verified and the trial point $z^{k,i}$ is necessarily accepted.

Let us define

$$\varepsilon_2 = \min \left\{\varepsilon_1, \frac{0.9(1-r)\bar{\alpha}^2}{200c_1}, \bar{\alpha}t_{min}\right\},$$

where ε_1 is defined in Lemma 3.7, and

$$t_{min} = \min\left\{1, \frac{0.09c}{\eta_{max}L_2}\right\}.$$

Let $k_1 \geq k_0$ be such that $|C^+(x^k)| \leq \varepsilon_2$ for all $k \geq k_1$. So, $|C^+(x^k)| \leq \bar{\alpha}t_{min}$, and then, $t_{min} \geq \frac{|C^+(x^k)|}{\bar{\alpha}}$. This implies that $t_{k,0} \geq \frac{|C^+(x^k)|}{\bar{\alpha}}$ for all $k \geq k_1$. Therefore, with the condition $t_{k,0} \geq t_{min}$, a possible steplength $t_{k,i} < \frac{|C^+(x^k)|}{\bar{\alpha}}$ cannot correspond to $i = 0$, so it is preceded by $t_{k,i-1}$ which necessarily verifies

$$t_{k,i-1} \leq 10\frac{|C^+(x^k)|}{\bar{\alpha}},$$

and corresponds to (24). Thus, the trial point $z^{k,i-1}$ is accepted for all $k \geq k_1$. Therefore, $t_{k,i} \geq \frac{|C^+(x^k)|}{\bar{\alpha}}$ for all $k \geq k_1$, $i = 0, 1, \ldots, iacc(k)$. So, by Lemma 3.7, the penalty parameter $\theta_{k,i}$ is never decreased for all $k \geq k_1$, and this implies the desired result. $\square$

Finally, we prove that Hypothesis C cannot hold.

Theorem 3.9. *Let $\{x^k\}$ be an infinite sequence generated by Algorithm 2.1. Then, there exists K, an infinite subset of $\{0, 1, 2, \ldots\}$, such that*

$$\lim_{k \in K} \|d^{k,tan}\| = 0. \tag{26}$$

Proof. Suppose that the thesis of the theorem is not true. Then, there exists $k_0 \in \{0, 1, 2, \ldots\}$, $\varepsilon > 0$ such that Hypothesis C holds.

As in the beginning of the proof of Theorem 3.3, observe that, by Theorem 3.1,

$$\mathbf{Ared}_{k,i} - 0.1\mathbf{Pred}_{k,i}$$

$$= 0.9\{\theta_{k,i}[f(x^k) - f(z^{k,i})] + (1-\theta_{k,i})[|C^+(x^k)| - |C^+(y^k)|]\} + (1-\theta_{k,i})[|C^+(y^k)| - |C^+(z^{k,i})|]$$

$$\geq 0.9\theta_{k,i}[f(y^k) - f(z^{k,i})] + 0.9\theta_{k,i}[f(x^k) - f(y^k)] - (1-r)|C^+(x^k)| - c_1 t_{k,i}^2 \|d^{k,tan}\|^2.$$

Moreover, using Taylor expansion and (3),

$$|f(x^k) - f(y^k)| = |\nabla f(x^k)^t (y^k - x^k) + \mathcal{O}(\|y^k - x^k\|)|$$

$$\leq \|\nabla(f(x^k)\|\|y^k - x^k\| + \mathcal{O}\|y^k - x^k\| \leq M_1\|y^k - x^k\| \leq M_1 \beta |C^+(x^k)|,$$

where M_1 is a norm-dependent constant that also depends on a bound of $\|\nabla f(x)\|$ on Ω.

Therefore,

$$0.9\theta_{k,i}(f(y^k) - f(z^{k,i})) \geq 0.9\bar{\theta} M_1 \beta |C^+(x^k)| \equiv M_2 |C^+(x^k)|.$$

Then, by Lemmas 3.6 and 3.8,

$$\mathbf{Ared}_{k,i} - 0.1\mathbf{Pred}_{k,i} \geq 0.9\bar{\theta} c_3 - M_2 |C^+(x^k)| - c_1 t_{k,i}^2 \varepsilon^2$$

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

Let us define

$$\bar{t} = \left(0.45 \frac{\bar{\theta} c_3}{\varepsilon^2 c_1}\right)^{1/2}.$$

If $t_{k,i} \leq \bar{t}$ we have that

$$c_1 \varepsilon^2 t_{k,i}^2 \leq 0.45\bar{\theta} c_3,$$

so, when $t_{k,i} \leq \bar{t}$, we have that

$$\mathbf{Ared}_{k,i} - 0.1\mathbf{Pred}_{k,i} \geq 0.45\bar{\theta} c_3 - M_2 |C^+(x^k)| \tag{27}$$

for all $k \geq k_0$, $i = 0, 1, iacc(k)$. Let $k_2 \geq k_0$ be such that

$$M_2 |C^+(x^k)| \leq 0.45\bar{\theta} c_3 \tag{28}$$

for all $k \geq k_2$. By (27) and (28) we have that, for all $k \geq k_2$, if $i \in \{0, 1, 2, \ldots\}$ corresponds to the first $t_{k,i}$ less than or equal to $\bar{t}$, then

$$\mathbf{Ared}_{k,i} - 0.1\mathbf{Pred}_{k,i} \geq 0.$$

This means that $t_{k,i} \geq \bar{t}/10$ must be accepted. Therefore, $t_{k,iacc(k)} \geq \bar{t}/10$ for all $k \geq k_2$.

So, if $k \geq k_2$ we have, by Lemma 3.6, Lemma 3.8, (2) and (3), that

$$\mathbf{Pred}_{k,iacc(k)} = \theta_{k,iacc(k)}[f(x^k) - f(z^{k,i})] + (1 - \theta_{k,iacc(k)})[|C^+(x^k)| - |C^+(y^k)|]$$

$$= \theta_{k,iacc(k)}[f(y^k) - f(z^{k,i})] + \theta_{k,iacc(k)}[f(x^k) - f(y^k)] + (1 - \theta_{k,iacc(k)})[|C^+(x^k)| - |C^+(y^k)|]$$

$$\geq \bar{\theta}[f(y^k) - f(z^{k,i})] - |f(x^k) - f(y^k)| - |C^+(x^k)| \geq \bar{\theta} c_3 - M_3 |C^+(x^k)| \tag{29}$$

for all $k \geq k_2$, where M_3 is a constant that depends on the norm and the bound of $\|\nabla f(x)\|$ on Ω. Now, let $k_3 \geq k_2$ be such that

$$M_3 |C^+(x^k)| \leq \frac{\bar{\theta}}{2} c_3$$

for all $k \geq k_3$. By (29), $\mathbf{Pred}_{k,iacc(k)}$ is bounded away from zero for all $k \geq k_3$. This implies, by (10), that $\mathbf{Ared}_{k,iacc(k)}$ is bounded away from zero for all $k \geq k_3$. Clearly, this contradicts Theorem 3.4. This means that Hypothesis C cannot be true, and the desired result is proved. $\square$

So far, we proved that, if the conditions (2) and (3) are fulfilled at every iteration, the algorithm finds a feasible point such that the spectral gradient direction $d^{k,tan}$ is as small as desired. It remains to show that this property is related to optimality. Our strategy will be to show that, under the weak constraint qualification called *Constant Positive Linear Dependence* $d^{k,tan} \to 0$ implies the KKT optimality condition.

3.4 Properties of general feasible sets

Let a feasible set be given in the general form

$$\Xi = \{x \in I\!R^n \mid g(x) \leq 0, h(x) = 0\}, \tag{30}$$

where $g : I\!R^n \to I\!R^p$, $h : I\!R^n \to I\!R^q$, and g, h are continuously differentiable.

Suppose that a single equality constraint is replaced by two inequality constraints, say, $h_1(x) = g_{p+1}(x) \leq 0$ and $-h_1(x) = g_{p+2}(x) \leq 0$, so that the feasible set is restated as

$$\widehat{\Xi} = \{x \in I\!R^n \mid g_1(x) \leq 0, \ldots, g_p(x) \leq 0, h_1(x) \leq 0, -h_1(x) \leq 0, h_2(x) = 0, \ldots, h_q(x) = 0\}. \tag{31}$$

It is easy to see that, if $x \in \Xi$ satisfies the classical constraint qualification of Mangasarian-Fromovitz (MFCQ)[1], such fulfillment is lost with the new description (31): the MFCQ does not hold for $x \in \widehat{\Xi}$. However, this is not the case for the constant positive linear dependence (CPLD) constraint qualification (proved in [2] to be implied by MFCQ), as we show in the sequel.

First, for completeness, we recall the positive-linear dependence definition and the statement of the CPLD condition, introduced by Qi and Wei [33], and proved to be a constraint qualification in [2].

For all $x \in \Xi$, define the set of indices of the active inequality constraints as

$$I(x) = \{i \in \{1, \ldots, p\} \mid g_i(x) = 0\},$$

and let $J = \{1, \ldots, q\}$.

Definition 1. *(Positive-linear dependence - PLD)* Let $x \in \Xi$, $I_0 \subset I(x)$, $J_0 \subset J$. We say that the set of gradients $\{\nabla g_i(x)\}_{i \in I_0} \cup \{\nabla h_j(x)\}_{j \in J_0}$ is *positive-linearly dependent* if there exist scalars $\{\alpha_j\}_{j \in J_0}$, $\{\beta_i\}_{i \in I_0}$ such that $\beta_i \geq 0$ for all $i \in I_0$, $\sum_{j \in J_0} |\alpha_j| + \sum_{i \in I_0} \beta_i > 0$ and

$$\sum_{j \in J_0} \alpha_j \nabla h_j(x) + \sum_{i \in I_0} \beta_i \nabla g_i(x) = 0.$$

Otherwise, we say that the set $\{\nabla g_i(x)\}_{i \in I_0} \cup \{\nabla h_j(x)\}_{j \in J_0}$ is *positive-linearly independent.*

[1] The feasible point $x \in \Xi$ satisfies the MFCQ if the set $\{\nabla h_j(x)\}_{j=1}^q$ is linearly independent and there exists $d \in I\!R^n$ such that $\nabla h_j(x)^T d = 0$, $j = 1, \ldots, m$ and $\nabla g_i(x)^T d < 0$, $i \in I(x)$. Equivalently, $x \in \Xi$ satisfies the MFCQ if $\sum_{j=1}^q \lambda_j \nabla h_j(x) + \sum_{i \in I(x)} \mu_i \nabla g_i(x) = 0$, with $\mu_i \geq 0$, $\forall i \in I(x)$, implies that $\mu_i = 0$, $\forall i \in I(x)$ and $\lambda_j = 0, j = 1, \ldots, q$.

Definition 2. *(Constant positive linear dependence condition - CPLD)* A feasible point $x \in \Xi$ is said to satisfy the CPLD condition if it satisfies the MFCQ or, for any $I_0 \subset I(x)$, $J_0 \subset J$ such that the set of gradients $\{\nabla g_i(x)\}_{i \in I_0} \cup \{\nabla h_j(x)\}_{j \in J_0}$ is positive-linearly dependent, there exists a neighborhood $N(x)$ of x such that, for any $y \in N(x)$, the set $\{\nabla g_i(y)\}_{i \in I_0} \cup \{\nabla h_j(y)\}_{j \in J_0}$ is linearly dependent.

Theorem 3.10. *Let x^* be a point that satisfies the CPLD constraint qualification for the feasible set given by (30). Then x^* also satisfies the CPLD for the description (31).*

Proof. From the hypothesis, $x^* \in \Xi$ and for each $J_0 \subset J$ and each $I_0 \subset I(x^*)$ such that the set

$$\{\nabla g_i(x^*)\}_{i \in I_0} \cup \{\nabla h_j(x^*)\}_{j \in J_0} \tag{32}$$

is positive-linearly dependent, there exists a neighborhood $N(x^*)$ such that for every $y \in N(x^*)$, the set $\{\nabla g_i(y)\}_{i \in I_0} \cup \{\nabla h_j(y)\}_{j \in J_0}$ is linearly dependent. Let $\alpha_j \in I\!R, j \in J_0$ and $\beta_i \geq 0, i \in I_0$ be the scalars of the positive linear combination of the set (32):

$$\sum_{j \in J_0} \alpha_j \nabla h_j(x^*) + \sum_{i \in I_0} \beta_i \nabla g_i(x^*) = 0, \tag{33}$$

with $\sum_{j \in J_0} |\alpha_j| + \sum_{i \in I_0} \beta_i > 0$.

Thus, given I_0, J_0 such that the set (32) is positive-linearly dependent so that (33) holds, two possibilities might happen:

i) $1 \notin J_0$, and by defining $\widehat{I_0} \equiv I_0$ and $\widehat{J_0} \equiv J_0$, the CPLD holds for $x^* \in \widehat{\Xi}$ as a direct consequence of the hypothesis.

ii) $1 \in J_0$, and setting $\widehat{I_0} \equiv I_0 \cup \{p+1, p+2\}$ and $\widehat{J_0} \equiv J_0 \setminus \{1\}$, equation (33) can be restated as

$$\sum_{j \in \widehat{J_0}} \widehat{\alpha}_j \nabla h_j(x^*) + \sum_{i \in \widehat{I_0}} \widehat{\beta}_i \nabla g_i(x^*) = 0,$$

where $\widehat{\alpha}_j = \alpha_j$, $\forall j \in \widehat{J_0}$, $\widehat{\beta}_i = \beta_i$, $\forall i \in I_0$,

$$\widehat{\beta}_{p+1} = \begin{cases} \alpha_1, & \alpha_1 \geq 0 \\ 0, & \alpha_1 < 0 \end{cases} \qquad \text{and} \qquad \widehat{\beta}_{p+2} = \begin{cases} 0, & \alpha_1 \geq 0 \\ -\alpha_1, & \alpha_1 < 0 \end{cases},$$

that is, $\widehat{\beta}_{p+1} - \widehat{\beta}_{p+2} = \alpha_1$, and so $\sum_{j \in \widehat{J_0}} |\widehat{\alpha}_j| + \sum_{i \in \widehat{I_0}} \widehat{\beta}_i > 0$, $\widehat{\beta}_i \geq 0$, $\forall i \in \widehat{I_0}$.

Thus, from the given sets I_0, J_0 for which the CPLD is fulfilled for $x^* \in \Xi$, we can always define $\widehat{I_0}$, $\widehat{J_0}$ such that the same PLD combination is produced, the neighborhood $N(x^*)$ is inherited, and so the CPLD holds for $\widehat{\Xi} \ni x^*$, what completes the proof. $\square$

In the following theorem we prove that, under the CPLD condition, a KKT point is assymptotically found.

Theorem 3.11. *Let $\{x^k\}$ be an infinite sequence generated by Algorithm 2.1, and K be the infinite subset of Theorem 3.9 such that (26) holds. Possibly redefining the index set, assume that $\lim_{k \in K} x^k = x^*$. If x^* satisfies the CPLD for problem (1) then x^* also verifies the KKT condition for (1).*

Proof. For the infinite sequence $\{x^k\}$ of the hypotheses, let $\{y^k\}$ be the corresponding sequence computed in Step 1 (or in the repetitions of Steps 3 and 4, in the nonmonotone strategy), so that y^k verifies (2) and (3). Now, $d^{k,tan}$ is the solution of (15), so there exists $\mu^k \in I\!R^m_+$ such that

$$\begin{aligned} d^{k,tan} + \eta_k \nabla f(y^k) + C'(y^k)^T \mu^k &= 0 && (34)\\ \mu_j^k (C_j'(y^k) d^{k,tan}) &= 0, \ \forall j \text{ such that } C_j(y^k) \geq 0 && (35)\\ \mu_j^k (C_j(y^k) + C_j'(y^k) d^{k,tan}) &= 0, \ \forall j \text{ such that } -\xi \leq C_j(y^k) < 0 && (36)\\ \mu_j^k &= 0, \ \text{otherwise.} && (37) \end{aligned}$$

According to Corollary 3.5, $\lim_{k\in K} |C^+(x^k)| = 0$. Thus, from (2) and (3), it follows respectively that

$$|C^+(y^k)| \to 0 \text{ and } \|y^k - x^k\| \to 0, \ k \in K. \tag{38}$$

As a result, $\lim_{k\in K} y^k = x^*$. Moreover, if $C_j(x^*) < 0$, then $C_j(x^k) < 0$ for k large enough. As $d^{k,tan} \to 0$, it follows from (36) that $C_j(y^k) + C_j'(y^k) d^{k,tan} < 0$ in this case. Therefore, we can assume that

$$\mu_j^k = 0 \text{ whenever } C_j(x^*) < 0. \tag{39}$$

From (34), since $\eta_k \geq \eta_{min} > 0$ for all k, it follows that

$$\nabla f(y^k) + \frac{1}{\eta_k} C'(y^k)^T \mu^k = \frac{-1}{\eta_k} d^{k,tan} \to 0, \ k \in K,$$

where $\eta_k \in [\eta_{\min}, \eta_{\max}], 0 < \eta_{\min} < \eta_{\max} < +\infty$, and as a consequence of (37) and (39), we have

$$\nabla f(y^k) + \sum_{i \in I(x^*)} \frac{\mu_i^k}{\eta_k} \nabla C_i(y^k) = \frac{-1}{\eta_k} d^{k,tan}, \ k \in K.$$

From Caratheodory's theorem ([6, p.689]), there exist $I_k \subset I(x^*)$ and a sequence $\{\alpha^k\} \subset I\!R^m_+$ such that

$$\nabla f(y^k) + \sum_{i \in I_k} \alpha_i^k \nabla C_i(y^k) = \frac{-1}{\eta_k} d^{k,tan}, \ k \in K,$$

and the vectors $\{\nabla C_i(y^k)\}_{i \in I_k}$ are linearly independent.

As the possible number of subsets I_k is finite, there exists an infinite set $K_1 \subset K$ such that for all $k \in K_1$, $I_k = I \subset I(x^*)$,

$$\nabla f(y^k) + \sum_{i \in I} \alpha_i^k \nabla C_i(y^k) = \frac{-1}{\eta_k} d^{k,tan} \to 0, \ k \in K_1, \tag{40}$$

and the vectors $\{\nabla C_i(y^k)\}_{i \in I}$ are linearly independent.

If the sequence $\{\alpha^k\}$ is in a compact set, then there exists $K_2 \subset K_1$ such that $\lim_{k\in K_2} \alpha^k = \alpha \geq 0$. Therefore, taking the limit on (40) along $k \in K_2$ we conclude that

$$\nabla f(x^*) + \sum_{i \in I} \alpha_i \nabla C_i(x^*) = 0,$$

with $\alpha \geq 0$, which proves the desired result in this case.

Now, assume that $\{\alpha^k\}$ is not contained in a compact set. Let us define, for every $k \in K_1$, $M_k = \max\{\alpha_i^k, i \in I\}$, so that $\lim_{k\in K_1} M_k = +\infty$. Then, from (40) we have

$$\frac{\nabla f(y^k)}{M_k} + \sum_{i\in I} \frac{\alpha_i^k}{M_k} \nabla C_i(y^k) = \frac{-1}{M_k \eta_k} d^{k,tan},$$

and so,

$$\lim_{k\in K_1} \sum_{i\in I} \frac{\alpha_i^k}{M_k} \nabla C_i(y^k) = 0.$$

As the number of constraints is finite, there exists an index for which the maximum M_k holds infinitely many times, say, for all $k \in K_2 \subset K_1$. Furthermore, there exists an infinite subset $K_3 \subset K_2$ such that the sequence $\{\frac{\alpha_i^k}{M_k}\}_{k\in K_3}, i \in I$ converges for a set of coefficients not identically zero, as at least one of them is always 1. Hence, taking the limit along $k \in K_3$, there exists $0 \neq \tilde{\alpha} \in I\!R_+^m$ such that

$$\sum_{i\in I} \tilde{\alpha}_i \nabla C_i(x^*) = 0,$$

so the gradients $\{\nabla C_i(x^*)\}_{i\in I}$ are positively linearly dependent. As there are points y^k in a neighborhood of x^* for which $\{\nabla C_i(y^k)\}_{i\in I}$ are linearly independent, this contradicts the CPLD assumption and completes the proof. $\square$

4 Numerical experiments

To assess reliability, we designed a problem-oriented implementation of Algorithm 2.1 focusing on the challenging family of hard-spheres problems. Given a pair of integers (p, dim), where p stands for the number of points and dim for their dimension, the so-called *hard-spheres problem* is to find a solution $\omega^j \in I\!R^{dim}, j = 1, \ldots, p$ to

$$\begin{array}{rl} \text{maximize} \min\limits_{i\neq j} & \|\omega^i - \omega^j\| \\ \text{subject to} & \|\omega^j\| = 1, \forall j, \end{array} \tag{41}$$

which is equivalent to

$$\begin{array}{rl} \text{minimize} \max\limits_{i\neq j} & \langle \omega^i, \omega^j \rangle \\ \text{subject to} & \|\omega^j\| = 1, \forall j. \end{array}$$

The minimax formulation, by its turn, with an additional variable, can be posed as the following nonlinear programming problem, with equality and inequality constraints:

$$\begin{array}{rl} \text{minimize} & \alpha \\ \text{subject to} & \langle \omega^i, \omega^j \rangle \leq \alpha, i \neq j \\ & \|\omega^j\| = 1, j = 1, \ldots, p. \end{array} \tag{42}$$

It is not difficult to see that the Mangasarian-Fromovitz constraint qualification holds for problem (42) (see, e.g. [6], p.325-326). Recalling the analysis of subsection 3.4, as each equality constraint is turned into two inequalities, this property is no longer valid. Since the MFCQ implies the constant positive linear dependence condition [2], as a consequence of Theorem 3.10, the CPLD condition remains valid for the inequality-based reformulation for the hard-spheres problem. As a result, the constraints of the transformed problem remain qualified, so that we may expect to detect Karush-Kuhn-Tucker stationary points with the usual stopping criterion for nonlinear programming for this class of problems.

Problem (41), and in consequence, (42) as well, has a potentially large number of non optimal points satisfying optimality conditions. To reduce the degrees of freedom present in the solution set, some components were kept conveniently fixed, so that rigid motions are avoided in some extent. We adopt the following convention: for the first $dim-1$ points, the first k-th components are set to zero, that is

$$\omega_k^i = 0, \quad \text{for } k = 1, \ldots, dim - i,$$

where $i = 1, 2, \ldots, dim-1$, so that $dim \times (dim+1)/2$ variables are eliminated from the original formulation. Denoting by $[\omega^i]_{j_1}^{j_2}$ the j_1-th to j_2-th components of the i-th point, the vector with the reduced variables of the problem becomes:

$$x^T = \left(\omega_{dim}^1, [\omega^2]_{dim-1}^{dim}, \ldots, [\omega^{dim-1}]_2^{dim}, [\omega^{dim}]_1^{dim}, \ldots, [\omega^p]_1^{dim}, \alpha\right) \in I\!R^N, \tag{43}$$

with dimension $N = p \times dim + 1 - dim \times (dim+1)/2$.

Algorithm 2.1 was implemented in `GnuFortran`, and run in a PC based in X86 machine, with an Intel 3400MHz processor, 2048Mb of main memory and under Fedora Core 6 operating system. Problems (15) were solved using `MINOS` (version 5.5 for `Linux`, see also [32]).

Table 1 contains the features of the problems, namely the chosen pairs (p, dim), corresponding number of variables (dimension N of vector (43)), total number of constraints (the inequalities plus twice the equalities), number of nonzeros in the Jacobian matrix, and density of the Jacobian (ratio between the number of nonzeros and the matrix size).

dim	p	variables	constraints	nonzeros in Jacobian	density of Jacobian
3	11	31	77	387	16%
	12	34	90	464	15%
	13	37	104	548	14%
4	23	87	299	2193	8.4%
	24	91	324	2396	8.1%
	25	95	350	2608	7.8%
5	37	176	740	6976	5.3%
	38	181	779	7373	5.2%
	39	186	819	7781	5.1%
6	58	334	1769	20485	3.5%
	59	340	1829	21221	3.4%
	60	346	1890	21970	3.3%

Table 1: Problem features.

For each pair (p, dim), 50 initial points were randomly generated in the box $[-1, 1]^N$. The structure of the hard-spheres problem allows a natural exact restoration in Step 1. Given the current approximation x^k as in (43), for $i = 1, \ldots, p$, each point ω^i is replaced by its normalized counterpart, and then α is updated with the $\max\{\langle \omega^i, \omega^j \rangle, i \neq j\}$, to obtain y^k satisfying exactly the constraints. If condition (3) is violated by such y^k, it is replaced by

$x^k + \beta[\|C^+(y^k)\|/\|y^k - x^k\|]/(y^k - x^k)$. If condition (2) is not fulfilled by this point, we declare 'failure in improving feasibility' in the restoration phase. We used $r = 0.75$ and $\beta = 10^4$ in our experiments, and such failure never occurred.

An important adaptation was essential for the hard-spheres problems to be solved with an authentic SPG step. As the objective function of the minimax reformulation is linear, its gradient is constant and, as stated in (14), the spectral updating vector u^k would be the null vector, so that the spectral step would be constantly taken as $\eta_k \equiv \eta_{\max}$.

Therefore, instead of considering the pure objective function, we used the modified Lagrangian

$$L(x, y, \mu) = f(x) + C(x)^T \mu - [\nabla C(y)(x - y)]^T \mu$$

and defined the spectral vector as

$$u^k = \nabla L(z, y^k, \mu^k) - \nabla L(x^k, y^{k-1}, \mu^{k-1}),$$

with μ^k the Lagrange multiplier of problem (15). We set $\eta_{\min} = 10^{-3}$ and $\eta_{\max} = 20N$. The parameter $\xi > 0$ that is present in the definition (4) allows strongly satisfied constraints to be excluded from the linearization. For the hard-spheres problems, the usage of $\xi \geq 2$ implies that all the constraints were considered to assemble the set π_k. This was our choice, as preliminary experiments [22] pointed out that for this class of problems, such parameter has to be dynamically set (more demanding in the beginning, and possibly more relaxed as the active set is more or less stabilized) and carefully monitored for not contaminating the results. This tuning was not the focus of the current work and certainly deserves further investigation.

As far as the reasons for Algorithm 2.1 to stop, we declare *stopping 1* whenever the measures for optimality and feasibility are sufficiently small, that is, $\|d^{k,tan}\| < 10^{-4}$ and $|C^+(x^k)| < 10^{-4}$, respectively. The algorithm could also stop due to lack of progress (a too small stepsize: $t_{k,i} < 10^{-8}$), that we denoted by *stopping 2*. Another possibility was reaching the maximum allowed number of outer iterations - five times the problem dimension - declared as *stopping 3*.

To analyze the numerical results, we adopted the *boxplot* statistical tool [39]. Instead of resting upon the raw triple 'minimum-average-maximum' of each list of 50 results, for each distinct generated initial configuration, the boxplot, also known as the five-parameter-graph, provides the visualization of a processed 5-uple 'minimum-first quartile-median-third quartile-maximum'. As a by-product, possible outliers that might come from an unfavorable initial configuration are detected and do not interfere upon the results as much as with the raw triple.

To start the comparative analysis, we have used the boxplot tool to investigate the robustness of the nonmonotonic strategy, described in subsection 2.3, allowing the maximum number of repetitions $k_{\max}$ to be 3, 5 or 7.

Figures 2 and 3 contain the boxplots of the largest minimum distances obtained for the whole set of solved problems, where results for the pure tangential step are indicated with '0', and for the nonmonotone strategy, with the corresponding number of maximum repetitions. As we are maximizing the minimum distance, the higher the values, the better.

We have also included, on the left of the plots, for each choice (p, dim), the information of the triple 'minimum-average-maximum' provided in [30], using the simbols △-◯-▽, respectively. With these triples in perspective, we can see that our algorithm is able to reach a comparable quality of results. For $dim = 6$ we do not have comparative data.

By observing Figures 2 and 3, we notice that as the number of repetitions of the nonmonotone strategy grows, the quality of the results get worse. Allowing at most three repetitions produces larger minimum distances. Therefore, we decided to adopt the maximum of three repetitions in the nonmonotone strategy to proceed the analysis.

Figure 2: Largest minimum distances obtained for problems with $dim = 3$ and 4, with pure tangential step (0) and nomonotone strategies, taking at most 3, 5 or 7 repetitions.

Figure 3: Largest minimum distances obtained for problems with $dim = 5$ and 6, with pure tangential step (0) and nomonotone strategies, taking at most 3, 5 or 7 repetitions.

Table 2 summarizes the reasons for stopping for the whole set of tests, either with a pure tangential step or with at most three repetitions, taking into consideration the three chosen values for p for each value of dim (150 tests). It is worth mentioning that whenever the run ended with *stopping 2*, the largest minimum distance obtained was comparable to the ones found by runs that ended with *stopping 1*, and the optimality criterion was almost satisfied ($\|d^{k,tan}\|$ of the order of 10^{-3}). Thus, comparing the pure tangential step performance with the repeated step one, we observed that the former may get trapped by feasibility, whereas the latter, by looking for optimality with larger steps, is more prone not to stall. Nevertheless, in the great majority of the runs, optimality and feasibility were reached in the end, up to the prescribed tolerances. The exceptions are probably related to nasty starting configurations, with clustered points. Due to our interest in assessing the robustness of the algorithm, we did not exploit any particularly more favorable technique to produce the initial configuration (see, e.g. [27]).

dim	strategy	stopping 1	stopping 2	stopping 3
3	pure	100%	—	—
	repeated(3)	100%	—	—
4	pure	98%	1%	1%
	repeated(3)	98%	2%	—
5	pure	97%	2%	1%
	repeated(3)	100%	—	—
6	pure	93%	6%	1%
	repeated(3)	99%	1%	—

Table 2: Distribution of the reasons for stopping.

In terms of outer iterations, or number of accepted points, Figure 4 displays the results for the four choices for dim (3, 4, 5 and 6), each one in a plot. The six boxplots of each plot must be analyzed in pairs: the first one for the pure (`p`) tangential step, and the second, for the repeated (`r`) tangential step. We have also included the average values, marked with an '`x`', so we can appraise the effect of the outliers upon this measure. Clearly, the algorithm with repeated tangential steps performs better than the one with pure steps, taking fewer outer iterations.

Figure 4: Outer iterations (number of accepted points).

On the other hand, the number of calls to the solver `MINOS`, that was larger for the smaller problems, also decreased as the dimension increased, as can be seen in Figure 5.

Figure 5: Calls to `MINOS` (number of computed $d^{k,tan}$).

The ratio of the number of calls to `MINOS` per the number of outer iterations is depicted in the boxplots of Figure 6, one for each p, separated according to dim. We observe a balance between the median and the average values, and the absence of outliers. The majority of the values are definitely below 3, and this ratio is stable for all dimensions.

Figure 6: Typical number of calls to `MINOS` per outer iteration.

In terms of CPU time, the results are closely related to the calls of `MINOS`. As illustrated in Figure 7, the repeated strategy costs more for smaller problems, and as the dimension increases, the overall demanded time decreases.

Figure 7: CPU time in seconds.

Last but not least, we have tracked as well the percentage of acceptance by the merit function (Step 7) of the point that have satisfied the Armijo condition (Step 5). As can be seen Figure 8, the acceptance improves as the dimension increases.

Figure 8: Same trial point that satisfies Armijo is accepted by merit function.

Summarizing, the nonmonotone strategy shows a sort of deterioration as it becomes too greedy. Taking at most three repetitions of the tangent step generated more robust results as far as largest minimum distances, compared with allowing five or seven repetitions. Moreover, the algorithm that takes repeated tangential steps clearly outperformed the one based on a pure tangential step in terms of outer iterations. As far as calls to the solver `MINOS`, CPU time, and percentage of acceptance by the merit function of the trial point that satisfied the Armijo condition, the performance of the nonmonotone strategy improves as problem dimension increases. This might be related with the Jacobian sparsity, that also increases with the problem dimension, as can be seen in the last column of Table 1. Another possible reason for this good behavior is the increasing number of constraints, relatively larger as dimension increases, which seems to be favored with the larger steps performed by the nonmonotone strategy. In this sense, as the number of constraints increases, the freedom is better exploited.

5 Final remarks

To conclude, we have presented an algorithm that combines inexact restoration with the spectral projected gradient. We believe it is the natural way to extend the SPG philosophy for nonconvex problems. As usual in the SPG context, a nonmonotone strategy was adopted, that showed to be valuable in the numerical experiments, as compared with its monotone counterpart.

We have also discussed the fitting of turning equality constraints into inequalities in the inexact restoration context, together with a detailed analysis as regards as the constraint qualifications, including the MFCQ and the CPLD condition.

Concerning the good definition and the global convergence of Algorithm 2.1, theoretical results with the same flavor of the proved by Martínez and Pilotta for the algorithm presented in [30] are valid.

The proposed strategy showed to be efficient to solve the hard-spheres problem, with improved performance as the problem dimension increases. This may be a consequence of the increasing sparsity of Jacobian matrices as the problem dimension grows, together with a proper exploitation of the freedom provided by the larger steps as the number of constraints gets larger.

References

[1] R. Andreani, E. G. Birgin, J. M. Martínez, J-Y. Yuan. Spectral projected gradient and variable metric methods for optimization with linear inequalities, *IMA Journal of Numerical Analysis* 25 pp. 221-252 (2005).

[2] A. Andreani, J. M. Martínez, M. L. Schuverdt. On the relation between Constant Positive Linear Dependence condition and quasinormality constraint qualification. *Journal of Optimization Theory and Applications* 125, pp. 473-485 (2005).

[3] J. Barzilai and J. M. Borwein. Two point step size gradient method, *IMA Journal of Numerical Analysis* 8, pp. 141-148 (1988).

[4] L. Bello and M. Raydan. Convex constrained optimization for the seismic reflection tomography problem, *Journal of Applied Geophysics*, in press, 2007.

[5] D. P. Bertsekas. On the Goldstein-Levitin-Polyak gradient projection method, *IEEE Transactions on Automatic Control* 21, pp. 174-184 (1976).

[6] D. P. Bertsekas. *Nonlinear Programming*, 2ed. Athena Scientific, Belmont, 1999.

[7] E. G. Birgin and J. M. Martínez. Large-scale active-set box-constrained optimization method with spectral projected gradients, *Computational Optimization and Applications* 23, pp. 101-125 (2002).

[8] E. G. Birgin and J. M. Martínez. Local convergence of an Inexact-Restoration method and numerical experiments, *Journal of Optimization Theory and Applications* 127, pp. 229-247 (2005).

[9] E. G. Birgin, J. M. Martínez, M. Raydan. Nonmonotone spectral projected gradient methods on convex sets, *SIAM Journal on Optimization* 10, pp. 1196-1211 (2000).

[10] E. G. Birgin, J. M. Martínez, M. Raydan. Algorithm 813: SPG - Software for convex-constrained optimization, *ACM Transactions on Mathematical Software* 27, pp. 340-349 (2001).

[11] E. G. Birgin, J. M. Martínez, M. Raydan. Inexact Spectral Projected Gradient methods on convex sets. *IMA Journal on Numerical Analysis* 23, pp. 539-559 (2003).

[12] D. Cores and M. Loreto. A generalized two-point ellipsoidal anisotropic ray tracing for converted waves, *Optimization and Engineering*, in press, 2007.

[13] Y. H. Dai and R. Fletcher. Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming, *Numerische Mathematik* 100, pp. 21-47 (2005).

[14] Y. H. Dai and R. Fletcher. On the asymptotic behaviour of some new gradient methods, *Mathematical Proramming* 103 pp. 541-559 (2005).

[15] Y. H. Dai, W. W. Hager, K. Schittkowski, H. C. Zhang. The cyclic Barzilai-Borwein method for unconstrained optimization, *IMA Journal of Numerical Analysis* 26, pp. 604-627 (2006).

[16] G.P. Deidda, E. Bonomi and C. Manzi, Inversion of electrical conductivity data with Tikhonov regularization approach: some considerations, *Annals of Geophysics* 46, pp. 549-558 (2003).

[17] J. E. Dennis and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Prentice Hall, Englewood Cliffs, NJ, 1983.

[18] M. A. Diniz-Ehrhardt, M. A. Gomes-Ruggiero, J. M. Martínez, S. A. Santos. Augmented Lagrangian algorithms based on the spectral projected gradient for solving nonlinear programming problems. *Journal of Optimization Theory and Applications* 123, pp. 497-517 (2004).

[19] A. Friedlander, J. M. Martínez, B. Molina, M. Raydan. Gradient method with retards and generalizations, *SIAM Journal on Numerical Analysis* 36, pp. 275-289 (1998).

[20] A. A. Goldstein. Convex Programming in Hilbert Space, *Bulletin of the American Mathematical Society* 70, pp. 709-710 (1964).

[21] F. A. M. Gomes, M. C. Maciel, J. M. Martínez. Nonlinear programming algorithms using trust regions and augmented Lagrangians with nonmonotone penalty parameters. *Mathematical Programming* 84, pp. 161-200 (1999).

[22] M. A. Gomes-Ruggiero, J. M. Martínez, S. A. Santos. Inexact-Restoration methods with inequality constraints and Lagrangian tangent decrease, *19th. ISMP*, Book of Abstracts p.137 (2006).

[23] C. C. Gonzaga, E. W. Karas, M. Vanti. A globally convergent filter method for nonlinear programming, *SIAM Journal on Optimization* 14 pp.646-669 (2003).

[24] L. Grippo, F. Lampariello, S. Lucidi. A nonmonotone line search technique for Newton's method, *SIAM Journal on Numerical Analysis* 23, pp. 707-716 (1986).

[25] L. Grippo and M. Sciandrone. Nonmonotone derivative free methods for nonlinear equations, *Computational Optimization and Applications*, to appear.

[26] Z. Jiang. Applications of conditional nonlinear optimal perturbation to the study of the stability and sensitivity of the Jovian atmosphere. *Advances in Atmospheric Sciences* 23, pp. 775-783 (2006).

[27] S. Kucherenko, P. Belotti, L. Liberti, N. Maculan. New formulations for the Kissing Number Problem. *Discrete Applied Mathematics* 155, pp. 1837-1841 (2007).

[28] E. S. Levitin and B. T. Polyak. Constrained Minimization Problems, *USSR Computational Mathematics and Mathematical Physics* 6, pp. 1-50 (1966).

[29] J. M. Martínez. Inexact restoration method with Lagrangian tangent decrease and new merit function for nonlinear programming. *Journal of Optimization Theory and Applications* pp. 39-58 (2001).

[30] J. M. Martínez and E. A. Pilotta. Inexact restoration algorithms for constrained optimization. *Journal of Optimization Theory and Applications* 104, pp. 135-163 (2000).

[31] J. M. Martínez, E. A. Pilotta, M. Raydan. Spectral gradient methods for linearly constrained optimization, *Journal of Optimization theory and Application* 125, pp. 629-651 (2005).

[32] B. A. Murtagh and M. A. Saunders, MINOS 5.4 User's guide, Technical report SOL 83-20R, Stanford University, 1995.

[33] L. Qi and Z. Wei. On the constant positive linear dependence condition and its application to SQP methods, *SIAM Journal on Optimization* 10, pp. 963-981 (2000).

[34] A. Ramirez-Porras and W. E. Vargas-Castro. Transmission of visible light through oxidized copper films: feasibility of using a spectral projected gradient method, *Applied Optics* 43, pp. 1508-1514 (2004).

[35] M. Raydan. On the Barzilai and Borwein choice of steplength for the gradient method, *IMA Journal of Numerical Analysis* 13, pp. 321-326 (1993).

[36] M. Raydan. The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem, *SIAM Journal on Optimization* 7, pp. 26-33 (1997).

[37] M. Raydan and B. F. Svaiter. Relaxed steepest descent and Cauchy-Barzilai-Borwein method, *Computational Optimization and Applications* 21, pp.155-167 (2002).

[38] T. Serafini, G. Zanghirati, L. Zanni. Gradient projection methods for quadratic programs and applications in training support vector machines, *Optimization Methods and Software* 20, pp. 347-372 (2005).

[39] J. W. Tukey. *Exploratory Data Analysis*, Addison-Wesley, Reading, MA. 1977.

[40] W. E. Vargas, D. E. Azofeifa, N. Clark. Retrieved optical properties of thin films on absorbing substrates from transmittance measurements by application of a spectral projected gradient method, *Thin Solid Films* 425, pp.1-8 (2003).

[41] C. Y. Wang and Q. Liu. Convergence properties of inexact projected gradient methods, *Optimization* 55, pp. 301-310 (2006).

[42] C. Y. Wang, Q. Liu, X. M. Yang. Convergence properties of nonmonotone spectral projected gradient methods, *Journal of Computational and Applied Mathematics* 182, pp. 51-66 (2005).

[43] Y-X. Yuan. A new stepsize for the steepest descent method, *Journal of Computational Mathematics* 24, pp. 149-156, 2006.

[44] N. Zeev, O. Savasta, D. Cores. Non-monotone spectral projected gradient method applied to full waveform inversion *Geophisical Prospecting* 54, pp. 525-534 (2006).

[45] L. Zhang and W. J. Zhou. Spectral gradient projection method for solving nonlinear monotone equations, *Journal of Computational and Applied Mathematics* 196, pp. 478-484 (2006).