On Augmented Lagrangian methods with general lower-level constraints

R. Andreani [*] E. G. Birgin [†] J. M. Martínez [‡] M. L. Schuverdt [§]

March 3, 2005

Abstract

Augmented Lagrangian methods with general lower-level constraints are considered in the present research. These methods are useful when efficient algorithms exist for solving subproblems where the constraints are only of the lower-level type. Two methods of this class are introduced and analyzed. Inexact resolution of the lower-level constrained subproblems is considered. Global convergence is proved using the Constant Positive Linear Dependence constraint qualification. Conditions for boundedness of the penalty parameters are discussed. The reliability of the approach is tested by means of an exhaustive comparison against LANCELOT . All the problems of the CUTE collection are used in this comparison. Moreover, the resolution of location problems in which many constraints of the lower-level set are nonlinear is addressed, employing the Spectral Projected Gradient method for solving the subproblems. Problems of this type with more than 3×10^6 variables and 14×10^6 constraints are solved in this way, using moderate computer time.

Key words: Nonlinear programming, Augmented Lagrangian methods, global convergence, constraint qualifications, numerical experiments.

1 Introduction

Many practical optimization problems have the form

$$\text{Minimize } f(x) \text{ subject to } x \in \Omega_1 \cap \Omega_2, \tag{1}$$

[*] 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 (Grant 01-04597-4) and CNPq. e-mail: andreani@ime.unicamp.br

[†] Department of Computer Science IME-USP, University of São Paulo, Rua do Matão 1010, Cidade Universitária, 05508-090, São Paulo SP, Brazil. This author was supported by PRONEX-Optimization (PRONEX - CNPq / FAPERJ E-26 / 171.164/2003 - APQ1), FAPESP (Grants 01-04597-4 and 03-09169-6) and CNPq (Grant 300151/00-4). e-mail: egbirgin@ime.usp.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 (Grant 01-04597-4) 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) and FAPESP (Grants 01-04597-4 and 02-00832-1). e-mail: schuverd@ime.unicamp.br

where the constraint set Ω_2 is such that problems of type

$$\text{Minimize } F(x) \text{ subject to } x \in \Omega_2 \tag{2}$$

are considerably easier than problems of type (1). The constraints represented by Ω_1 will be called upper-level constraints and the constraints corresponding to Ω_2 will be lower-level constraints. In these cases, it is natural to solve (1) by means of a sequence of problems of the form (2).

In this paper we consider an Augmented Lagrangian approach for performing this task. At each outer iteration, a subproblem of the form (2) will be (perhaps inexactly) solved and, after each call to the algorithm that solves (2), Lagrange multipliers and penalty parameters will be updated.

Bertsekas [3] analyzed an Augmented Lagrangian method for solving (1) in the case in which the subproblems are solved exactly. Our theoretical contribution in this paper is to address the inexact solution of the subproblems and to prove the following global convergence results:

1. If x_* is a limit point of a sequence generated by the algorithms introduced here, then at least one of the following possibilities hold:
   - The point x_* is feasible. (This is necessarily the case if the penalty parameters are bounded.)
   - x_* is a KKT point of a sum of squares of upper-level infeasibilities, subject to lower-level constraints.
   - x_* does not satisfy the lower-level constant positive linear dependence (CPLD) constraint qualification [2, 38] [1].
2. If a limit point x_* is feasible then at least one of the following possibilities hold:
   - x_* is a KKT point of the original problem.
   - x_* does not satisfy the CPLD constraint qualification related to all the constraints.

We analyze two versions of the main algorithm: with only one penalty parameter and with one penalty parameter per constraint. In the first case we prove boundedness of the sequence of penalty parameters. We use the reduction to the equality-constraint case introduced in [3].

The global convergence results that use the CPLD constraint qualification are stronger than previous results for more specific problems. In particular, Conn, Gould and Toint [14] and Conn, Gould, Sartenaer and Toint [12] proved global convergence of Augmented Lagrangian methods for equality constraints and linear constraints using the assumption of linear independence of all the gradients of active constraints. Their assumption is much stronger than our CPLD-like assumptions. Convergence proofs for Augmented Lagrangian methods with equalities and box constraints using CPLD were given in [1].

[1] A feasible point of a nonlinear programming problem is said to satisfy the CPLD constraint qualification if the positive linear dependence of any set of gradients of active constraints implies the linear dependence of the same gradients in a whole neighborhood of the point.

We describe a particular implementation of the algorithms introduced in the paper for the case in which the lower-level set is a box. For solving the subproblems we use the code GENCAN [6]. We compare the default version of our method with the default version of LANCELOT [13] using *all* the problems of the CUTE collection [10].

The most interesting applications of our approach arises when the lower-level set is arbitrary, perhaps nonconvex or defined by nonlinear constraints. A location problem is defined as an example of this situation. In this case, it is known that the Spectral Projected Gradient method SPG [7, 8, 9] is particularly efficient when only the lower-level constraints are present, but SPG cannot be applied for the problem with the two types of constraints.

This paper is organized as follows. The algorithms are defined in Section 2. Section 3 is devoted to global convergence results. In Section 4 we prove boundedness of the penalty parameters. In Section 5 we show the numerical experiments. Applications, conclusions and lines for future research are discussed in Section 6.

Notation.

We denote:

$$I\!R_+ = \{t \in I\!R \mid t \geq 0\},$$

$$I\!R_{++} = \{t \in I\!R \mid t > 0\},$$

$$I\!N = \{0, 1, 2, \ldots\},$$

$$\| \cdot \| \text{ an arbitrary vectorial norm .}$$

$[v]_i$ is the $i-$th component of the vector v. If there is no possibility of confusion we may also use the notation v_i.

For all $y \in I\!R^n$, $y_+ = (\max\{0, y_1\}, \ldots, \max\{0, y_n\})$.

$[A]_j$ is the $j-$th column of the matrix A.

If $F : I\!R^n \to I\!R^m$, $F = (f_1, \ldots, f_m)$, we denote $\nabla F(x) = (\nabla f_1(x), \ldots, \nabla f_m(x)) \in I\!R^{n \times m}$.

If $v \in I\!R^n$ and $a \in I\!R$, we say that $v > a$ (respectively, $v \geq a$) if $[v]_i > a$ (respectively, $[v]_i \geq a$) for all $i = 1, \ldots, n$.

For all $v \in I\!R^n$ we denote $Diag(v) \in I\!R^{n \times n}$ the diagonal matrix with entries $[v]_i$.

If $K = \{k_0, k_1, k_2, \ldots\} \subset I\!N$ ($k_{j+1} > k_j \forall j$), we denote

$$\lim_{k \in K} x_k = \lim_{j \to \infty} x_{k_j}.$$

2 Augmented Lagrangian algorithms

We will consider the following nonlinear programming problem:

$$\text{Minimize } f(x) \text{ subject to } h_1(x) = 0, g_1(x) \leq 0, h_2(x) = 0, g_2(x) \leq 0, \tag{3}$$

where $f : I\!R^n \to I\!R, h_1 : I\!R^n \to I\!R^{m_1}, h_2 : I\!R^n \to I\!R^{m_2}, g_1 : I\!R^n \to I\!R^{p_1}, g_2 : I\!R^n \to I\!R^{p_2}$. We assume that all these functions admit continuous first derivatives on a sufficiently large and open domain. We define

$$\Omega_1 = \{x \in I\!R^n \mid h_1(x) = 0, g_1(x) \leq 0\}$$

and

$$\Omega_2 = \{x \in I\!R^n \mid h_2(x) = 0, g_2(x) \leq 0\}.$$

For all $x \in I\!R^n, \rho \in I\!R^{m_1+p_1}_{++}, \lambda \in I\!R^{m_1}, \mu \in I\!R^{p_1}_{+}$ we define the Augmented Lagrangian [31, 37, 41]:

$$L(x,\lambda,\mu,\rho) = f(x)+\sum_{i=1}^{m_1} \lambda_i [h_1(x)]_i + \sum_{i=1}^{m_1} \frac{[\rho]_i}{2}([h_1(x)]_i)^2 + \sum_{i=1}^{p_1} \frac{1}{2[\rho]_{m_1+i}}[\max\{0, \mu_i + [\rho]_{m_1+i}[g_1(x)]_i\}]^2. \tag{4}$$

Algorithm 2.1.

Let $x_0 \in I\!R^n$ an arbitrary initial point.

The given parameters for the execution of the algorithm are:

$$\tau \in [0,1), \gamma > 1,$$

$$-\infty < [\bar{\lambda}_{\min}]_i < [\bar{\lambda}_{\max}]_i < \infty \; \forall \; i = 1,\ldots,m_1,$$

$$[\bar{\mu}_{\min}]_i < 0 \leq [\bar{\mu}_{\max}]_i < \infty \; \forall \; i = 1,\ldots,p_1,$$

$$\rho_0, \rho_1 \in I\!R^{m_1+p_1}_{++},$$

$$[\bar{\lambda}_1]_i \in [[\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i] \; \forall i = 1,\ldots,m_1,$$

$$[\bar{\mu}_0]_i, [\bar{\mu}_1]_i \in [0, [\bar{\mu}_{\max}]_i] \; \forall i = 1,\ldots,p_1.$$

Finally, $\{\varepsilon_k\} \subset I\!R_+$ is a sequence of tolerance parameters such that

$$\lim_{k\to\infty} \varepsilon_k = 0. \tag{5}$$

($\{\varepsilon_k\}$ will be called the *tolerance sequence* from now on.)

Step 1. *Initialization*

Set $k \leftarrow 1$. For $j = 1,\ldots,p_1$, compute

$$[\sigma_0]_j = \max\left\{[g_1(x_0)]_j, -\frac{[\bar{\mu}_0]_j}{[\rho_0]_{m_1+j}}\right\}. \tag{6}$$

Step 2. *Solving the subproblem*

Compute (if possible) $x_k \in I\!R^n$ such that there exist $v_k \in I\!R^{m_2}, u_k \in I\!R^{p_2}$ satisfying

$$\|\nabla L(x_k, \bar{\lambda}_k, \bar{\mu}_k, \rho_k) + \textstyle\sum_{i=1}^{m_2} [v_k]_i \nabla [h_2(x_k)]_i + \sum_{i=1}^{p_2} [u_k]_i \nabla [g_2(x_k)]_i\| \leq \varepsilon_k, \tag{7}$$

$$[g_2(x_k)]_j \leq \varepsilon_k \text{ for all } j = 1,\ldots,p_2, \tag{8}$$

$$u_k \geq 0, \tag{9}$$

$$[g_2(x_k)]_j \geq -\varepsilon_k \;\text{ or }\; [u_k]_j = 0 \text{ for all } j = 1,\ldots,p_2, \tag{10}$$

$$\|h_2(x_k)\| \leq \varepsilon_k. \tag{11}$$

If it is not possible to find x_k satisfying (7)–(11), stop the execution of the algorithm.

Step 3. *Estimate multipliers*

For all $i = 1, \ldots, m_1$, compute

$$[\lambda_{k+1}]_i = [\bar{\lambda}_k]_i + [\rho_k]_i [h_1(x_k)]_i \tag{12}$$

and

$$[\bar{\lambda}_{k+1}]_i \in [[\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i]. \tag{13}$$

(Usually, $[\bar{\lambda}_{k+1}]_i$ will be the projection of $[\lambda_{k+1}]_i$ on the interval $[[\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i]$.) For all $j = 1, \ldots, p_1$, compute

$$[\mu_{k+1}]_j = \max\{0, [\bar{\mu}_k]_j + [\rho_k]_{m_1+j}[g_1(x_k)]_j\}, \tag{14}$$

$$[\sigma_k]_j = \max\left\{[g_1(x_k)]_j, -\frac{[\bar{\mu}_k]_j}{[\rho_k]_{m_1+j}}\right\},$$

and

$$[\bar{\mu}_{k+1}]_j \in [0, [\bar{\mu}_{\max}]_j]. \tag{15}$$

(Usually, $[\bar{\mu}_{k+1}]_j$ will be the projection of $[\mu_{k+1}]_j$ on the interval $[[\bar{\mu}_{\min}]_j, [\bar{\mu}_{\max}]_j]$.) Since, by (14), $[\mu_{k+1}]_j$ is always nonnegative, the given value of $[\bar{\mu}_{\min}]_j < 0$ is, in practice, irrelevant. However, we maintain this parameter for theoretical reasons.)

Step 4. *Update the penalty parameters*

Update the penalty parameters according to one of the following rules:

Rule 1. If

$$\max\{\|h_1(x_k)\|_\infty, \|\sigma_k\|_\infty\} \le \tau \max\{\|h_1(x_{k-1})\|_\infty, \|\sigma_{k-1}\|_\infty\},$$

define

$$[\rho_{k+1}]_i = [\rho_k]_i, \quad i = 1, \ldots, m_1 + p_1.$$

Else, define

$$[\rho_{k+1}]_i = \gamma[\rho_k]_i, \quad i = 1, \ldots, m_1 + p_1.$$

Rule 2. For each $i = 1, \ldots, m_1$, if

$$|[h_1(x_k)]_i| \le \tau |[h_1(x_{k-1})]_i|,$$

define

$$[\rho_{k+1}]_i = [\rho_k]_i.$$

Else, define

$$[\rho_{k+1}]_i = \gamma[\rho_k]_i.$$

For each $j = 1, \ldots, p_1$, if

$$|[\sigma_k]_j| \le \tau |[\sigma_{k-1}]_j|,$$

define

$$[\rho_{k+1}]_{m_1+j} = [\rho_k]_{m_1+j}.$$

Else, define

$$[\rho_{k+1}]_{m_1+j} = \gamma[\rho_k]_{m_1+j}.$$

Step 5. *Begin a new outer iteration*
Set $k \leftarrow k + 1$. Go to Step 2.

Remarks

1. The conditions (7)–(11) are approximate KKT conditions of the subproblem
$$\text{Minimize } L(x, \bar{\lambda}_k, \bar{\mu}_k, \rho_k) \text{ subject to } x \in \Omega_2. \tag{16}$$
In the convergence proofs we assume that one is able to compute x_k, v_k, u_k satisfying those conditions. The fulfillment of this practical assumption depends on the characteristics of the set Ω_2 and the objective function of (16). Let us assume that a KKT point $\widehat{x}$ of (16) exists with multipliers $\widehat{v} \in I\!R^{m_2}, \widehat{u} \in I\!R^{p_2}_+$. Assume, moreover, that we are able to approximate $(\widehat{x}, \widehat{v}, \widehat{u})$ by a convergent sequence
$$(\widehat{x}_\ell, \widehat{v}_\ell, \widehat{u}_\ell) \in I\!R^n \times I\!R^{m_1} \times I\!R^{p_1}_+, \ell \in I\!N.$$
(This sequence will be probably generated by a numerical algorithm.) Therefore:
$$\lim_{\ell \to \infty} (\widehat{x}_\ell, \widehat{v}_\ell, \widehat{u}_\ell) = (\widehat{x}, \widehat{v}, \widehat{u}). \tag{17}$$
In spite of (17), it is possible that no element of the sequence $(\widehat{x}_\ell, \widehat{v}_\ell, \widehat{u}_\ell)$ fulfills the condition (10). This happens if, for some j, $[g(\widehat{x})]_j < -\varepsilon_k$ (so $[\widehat{u}]_j = 0$) but $[\widehat{u}_\ell]_j > 0$ for all $\ell \in I\!N$. In other words, although $\lim_{\ell \to \infty} [\widehat{u}_\ell]_j = [\widehat{u}]_j = 0$, $[\widehat{u}_\ell]_j$ might remain strictly positive for all $\ell \in I\!N$.

   Fortunately, this situation can be overcome in the following way: For all $\ell \in I\!N$, if $[g(\widehat{x}_\ell)]_j < -\varepsilon_k$, define $[\widehat{u}'_\ell]_j = 0$. Otherwise, $[\widehat{u}'_\ell]_j = [\widehat{u}_\ell]_j$. Clearly, the condition (10) is fulfilled replacing $\widehat{u}_\ell$ by $\widehat{u}'_\ell$. Moreover,
$$\lim_{\ell \to \infty} (\widehat{x}_\ell, \widehat{v}_\ell, \widehat{u}'_\ell) = (\widehat{x}, \widehat{v}, \widehat{u}).$$
Therefore, the conditions (7), (8) and (11) are also fulfilled by $(\widehat{x}_\ell, \widehat{v}_\ell, \widehat{u}'_\ell)$ if ℓ is large enough.

2. The parameters ρ_0 and $\bar{\mu}_0$ are used only to compute σ_0 in (6). So, they can be eliminated setting σ_0 as a parameter of the algorithm. We prefer the present formulation for theoretical reasons related to the forthcoming equivalence Theorem 4.1.

3. Although there seems to be little motivation to use different updating rules at different iterations, the structure of the algorithm allows one to proceed in this way.

3 Global convergence

In this section we assume that the algorithm does not stop at Step 2. In other words, it is always possible to find x_k satisfying (7)-(11). Problem-dependent sufficient conditions for this assumption can be given in many cases. (However, it must be warned that the algorithm necessarily stops at Step 2 in some situations, for example if the lower-level set is empty.)

We will also assume that at least a limit point of the sequence generated by Algorithm 2.1 exists. A sufficient condition for this is the existence of $\varepsilon > 0$ such that the set

$$\{x \in I\!R^n \mid g(x) \leq \varepsilon, \|h(x)\| \leq \varepsilon\}$$

is bounded. This condition may be enforced adding artificial simple constraints to the set Ω_2. (However, one should be cautions when doing this, due to the danger of destroying the structure of Ω_2.)

We are going to investigate the status of the limit points of sequences generated by Algorithm 2.1. Firstly, we will prove a result on the feasibility properties of a limit point. Theorem 3.1 shows that, either a limit point is feasible or, if the CPLD constraint qualification with respect to Ω_2 holds, it is a KKT point of a weighted sum of squares of upper-level infeasibilities.

Theorem 3.1. *Let $\{x_k\}$ be a sequence generated by Algorithm 2.1. Let x_* be a limit point of $\{x_k\}$. Then, if the sequence of penalty parameters $\{\rho_k\}$ is bounded, the limit point x_* is feasible. Otherwise, at least one of the following possibilities hold:*

(i) There exists $\rho \in [0,1]^{m_1+p_1}$, $\rho \neq 0$, such that x_* is a KKT point of the problem

$$\textit{Minimize } \frac{1}{2}\left[\sum_{i=1}^{m_1}[\rho]_i[h_1(x)]_i^2 + \sum_{i=1}^{p_1}[\rho]_{m_1+i}\max\{0,[g_1(x)]_i\}^2\right] \textit{ subject to } x \in \Omega_2. \tag{18}$$

Moreover, if Rule 1 is used at all the iterations of Algorithm 2.1 and $[\rho_0]_i = [\rho_0]_1$ for all $i = 1,\ldots,m_1+p_1$, the property above holds with $[\rho]_i = 1, i = 1,\ldots,m_1+p_1$.

(ii) x_* does not satisfy the CPLD constraint qualification [2, 38] associated to Ω_2.

Proof. Let K be an infinite subsequence in $I\!N$ such that

$$\lim_{k\in K} x_k = x_*.$$

By (5), (8) and (11), we have that $g_2(x_*) \leq 0$ and $h_2(x_*) = 0$. So,

$$x_* \in \Omega_2. \tag{19}$$

Now, we consider two possibilities:

(a) The sequence $\{\rho_k\}$ is bounded.

(b) The sequence $\{\rho_k\}$ is unbounded.

Let us analyze first Case (a). In this case, from some iteration on, the penalty parameters are not updated. Therefore, both updating rules imply that

$$\lim_{k\to\infty}[h_1(x_k)]_i = \lim_{k\to\infty}[\sigma_k]_j = 0 \ \forall i = 1,\ldots,m_1, \ j = 1,\ldots,p_1.$$

Therefore,

$$h_1(x_*) = 0. \tag{20}$$

Moreover, if $[g_1(x_*)]_j > 0$ then $[g_1(x_k)]_j > 0$ for $k \in K$ large enough. This would contradict the fact that $[\sigma_k]_j \to 0$.

Therefore,

$$[g_1(x_*)]_i \leq 0 \quad \forall i = 1, \dots, p_1. \tag{21}$$

By (19), (20) and (21), x_* belongs to $\Omega_1 \cap \Omega_2$. So, the nonnegative objective function of (18) vanishes at x_*. This implies that x_* is feasible.

Therefore, we proved the thesis in the case that $\{\rho_k\}$ is bounded.

Consider now Case (b). So, $\{\rho_k\}_{k\in K}$ is not bounded. By (4) and (7), we have:

$$\nabla f(x_k) + \sum_{i=1}^{m_1}([\bar\lambda_k]_i + [\rho_k]_i[h_1(x_k)]_i)\nabla[h_1(x_k)]_i + \sum_{i=1}^{p_1}\max\{0, [\bar\mu_k]_i + [\rho_k]_{m_1+i}[g_1(x_k)]_i\}\nabla[g_1(x_k)]_i$$

$$+ \sum_{i=1}^{m_2}[v_k]_i\nabla[h_2(x_k)]_i + \sum_{j=1}^{p_2}[u_k]_j\nabla[g_2(x_k)]_j = \delta_k, \tag{22}$$

where, by (5),

$$\lim_{k\in K}\|\delta_k\| = 0. \tag{23}$$

If $[g_2(x_*)]_j < 0$, there exists $k_1 \in I\!\!N$ such that $[g_2(x_k)]_j < -\varepsilon_k$ for all $k \geq k_1, k \in K$. Therefore, by (10), $[u_k]_i = 0$ for all $k \in K, k \geq k_1$.

Thus, by (19) and (22), for all $k \in K, k \geq k_1$ we have that

$$\nabla f(x_k) + \sum_{i=1}^{m_1}([\bar\lambda_k]_i + [\rho_k]_i[h_1(x_k)]_i)\nabla[h_1(x_k)]_i + \sum_{i=1}^{p_1}\max\{0, [\bar\mu_k]_i + [\rho_k]_{m_1+i}[g_1(x_k)]_i\}\nabla[g_1(x_k)]_i$$

$$+ \sum_{i=1}^{m_2}[v_k]_i\nabla[h_2(x_k)]_i + \sum_{[g_2(x_*)]_j=0}[u_k]_j\nabla[g_2(x_k)]_j = \delta_k.$$

Dividing by $\|\rho_k\|_\infty$ we get:

$$\frac{\nabla f(x_k)}{\|\rho_k\|_\infty} + \sum_{i=1}^{m_1}\left(\frac{[\bar\lambda_k]_i}{\|\rho_k\|_\infty} + \frac{[\rho_k]_i}{\|\rho_k\|_\infty}[h_1(x_k)]_i\right)\nabla[h_1(x_k)]_i + \sum_{i=1}^{p_1}\max\left\{0, \frac{[\bar\mu_k]_i}{\|\rho_k\|_\infty} + \frac{[\rho_k]_{m_1+i}}{\|\rho_k\|_\infty}[g_1(x_k)]_i\right\}\nabla[g_1(x_k)]_i$$

$$+ \sum_{i=1}^{m_2}\frac{[v_k]_i}{\|\rho_k\|_\infty}\nabla[h_2(x_k)]_i + \sum_{[g_2(x_*)]_j=0}\frac{[u_k]_j}{\|\rho_k\|_\infty}\nabla[g_2(x_k)]_j = \frac{\delta_k}{\|\rho_k\|_\infty}.$$

By Caratheodory's Theorem of Cones (see [4], page 689) there exist

$$\widehat{I}_k \subset \{1, \dots, m_2\}, \widehat{J}_k \subset \{j \mid [g_2(x_*)]_j = 0\},$$

$$[\widehat{v}_k]_i, \;\; i \in \widehat{I}_k \;\text{ and }\; [\widehat{u}_k]_j \geq 0, \;\; j \in \widehat{J}_k$$

such that the vectors

$$\{\nabla[h_2(x_k)]_i\}_{i\in\widehat{I}_k}\cup\{\nabla[g_2(x_k)]_j\}_{j\in\widehat{J}_k}$$

are linearly independent and

$$\frac{\nabla f(x_k)}{\|\rho_k\|_\infty}+\sum_{i=1}^{m_1}\left(\frac{[\bar\lambda_k]_i}{\|\rho_k\|_\infty}+\frac{[\rho_k]_i}{\|\rho_k\|_\infty}[h_1(x_k)]_i\right)\nabla[h_1(x_k)]_i+\sum_{i=1}^{p_1}\max\left\{0,\frac{[\bar\mu_k]_i}{\|\rho_k\|_\infty}+\frac{[\rho_k]_{m_1+i}}{\|\rho_k\|_\infty}[g_1(x_k)]_i\right\}\nabla[g_1(x_k)]_i \tag{24}$$
$$+\sum_{i\in\widehat{I}_k}[\widehat{v}_k]_i\nabla[h_2(x_k)]_i+\sum_{j\in\widehat{J}_k}[\widehat{u}_k]_j\nabla[g_2(x_k)]_j=\frac{\delta_k}{\|\rho_k\|_\infty}.$$

Since there exist a finite number of possible sets $\widehat{I}_k$, $\widehat{J}_k$, there exists an infinite set of indices K_1 such that

$$K_1\subset\{k\in K\mid k\geq k_1\},$$
$$\widehat{I}_k=\widehat{I},$$

and

$$\widehat{J}=\widehat{J}_k\subset\{j\mid[g_2(x_*)]_j=0\} \tag{25}$$

for all $k\in K_1$. Then, by (24), for all $k\in K_1$ we have:

$$\frac{\nabla f(x_k)}{\|\rho_k\|_\infty}+\sum_{i=1}^{m_1}\left(\frac{[\bar\lambda_k]_i}{\|\rho_k\|_\infty}+\frac{[\rho_k]_i}{\|\rho_k\|_\infty}[h_1(x_k)]_i\right)\nabla[h_1(x_k)]_i+\sum_{i=1}^{p_1}\max\left\{0,\frac{[\bar\mu_k]_i}{\|\rho_k\|_\infty}+\frac{[\rho_k]_{m_1+i}}{\|\rho_k\|_\infty}[g_1(x_k)]_i\right\}\nabla[g_1(x_k)]_i \tag{26}$$
$$+\sum_{i\in\widehat{I}}[\widehat{v}_k]_i\nabla[h_2(x_k)]_i+\sum_{j\in\widehat{J}}[\widehat{u}_k]_j\nabla[g_2(x_k)]_j=\frac{\delta_k}{\|\rho_k\|_\infty},$$

and the gradients

$$\{\nabla[h_2(x_k)]_i\}_{i\in\widehat{I}}\cup\{\nabla[g_2(x_k)]_j\}_{j\in\widehat{J}} \tag{27}$$

are linearly independent.

We consider, again, two cases:

1. The sequence $\{\|(\widehat{v}_k,\widehat{u}_k)\|,k\in K_1\}$ is bounded.

2. The sequence $\{\|(\widehat{v}_k,\widehat{u}_k)\|,k\in K_1\}$ is unbounded.

If the sequence $\{\|(\widehat{v}_k,\widehat{u}_k)\|\}_{k\in K_1}$ is bounded, there exist $\rho\in[0,1]^{m_1+p_1}$, $\rho\neq0$, $(\widehat{v},\widehat{u}),\widehat{u}\geq0$ and an infinite set of indices $K_2\subset K_1$ such that

$$\lim_{k\in K_2}(\widehat{v}_k,\widehat{u}_k)=(\widehat{v},\widehat{u})$$

and

$$\lim_{k\in K_2}\frac{[\rho_k]_i}{\|\rho_k\|_\infty}=\rho_i,\ i=1,\ldots,m_1+p_1.$$

By (13), (15) and (23), taking limits for $k \in K_2$ in (26), we obtain:

$$\sum_{i=1}^{m_1}[\rho]_i[h_1(x_*)]_i\nabla[h_1(x_*)]_i+\sum_{i=1}^{p_1}[\rho]_{m_1+i}\max\{0,[g_1(x_*)]_i\}\nabla[g_1(x_*)]_i+\sum_{i\in\widehat{I}}\widehat{v}_i\nabla[h_2(x_*)]_i+\sum_{j\in\widehat{J}}\widehat{u}_j\nabla[g_2(x_*)]_j=0.$$

Therefore, by (19) and (25), x_* is a KKT point of (18).

The reasoning above is valid for both updating rules defined in Algorithm 2.1. Now, if Rule 1 is chosen for all k and all the penalty parameters are the same at the first outer iteration, it is obvious that they remain equal for all k. Therefore, all the entries of ρ must be the same and the thesis for Rule 1 holds.

Now, assume that $\{\|(\widehat{v}_k,\widehat{u}_k)\|\}_{k\in K_1}$ is unbounded. Let $K_3 \subset K_1$ be such that $\lim_{k\in K_3}\|(\widehat{v}_k,\widehat{u}_k)\| = \infty$ and $(\widehat{v},\widehat{u}) \neq 0, \widehat{u} \geq 0$ such that

$$\lim_{k\in K_3}\frac{(\widehat{v}_k,\widehat{u}_k)}{\|(\widehat{v}_k,\widehat{u}_k)\|}=(\widehat{v},\widehat{u}).$$

Dividing both sides of (26) by $\|(\widehat{v}_k,\widehat{u}_k)\|$ and taking limits for $k \in K_3$, we get:

$$\sum_{i\in\widehat{I}}\widehat{v}_i\nabla[h_2(x_*)]_i+\sum_{j\in\widehat{J}}\widehat{u}_j\nabla[g_2(x_*)]_j=0.$$

Then, the gradients

$$\{\nabla[h_2(x_*)]_i\}_{i\in\widehat{I}}\cup\{\nabla[g_2(x_*)]_i\}_{i\in\widehat{J}}$$

are positively linearly dependent. But $[g_2(x_*)]_j = 0$ for all $j \in \widehat{J}$. Then, by (27), x_* does not satisfy the CPLD constraint qualification associated with the set Ω_2. This completes the proof. □

In what follows we prove an optimality result. We saw that a limit point of a sequence generated by Algorithm 2.1 may be feasible or not. Roughly speaking, Theorem 3.1 says that, if x_* is not feasible, then (very likely) it is a local minimizer of the upper-level infeasibility, subject to lower-level feasibility. From the point of view of optimality, we are interested in the status of feasible limit points. In Theorem 3.2 we will prove that, under the CPLD constraint qualification, feasible limit points are stationary (KKT) points of the original problem. Let us recall that the CPLD condition was introduced in [38] and its status as a constraint qualification was revealed in [2]. Since CPLD is strictly weaker than the Mangasarian-Fromovitz (MF) constraint qualification, it turns out that the following theorem is stronger than results where KKT conditions are proved under MF or regularity assumptions.

Theorem 3.2. *Let $\{x_k\}_{k\in\mathbb{N}}$ be a sequence generated by Algorithm 2.1. Assume that $x_* \in \Omega_1\cap\Omega_2$ is a limit point that satisfies the CPLD constraint qualification related to $\Omega_1\cap\Omega_2$. Then, x_* is a KKT point of the original problem (3). Moreover, if x_* satisfies the Mangasarian-Fromovitz constraint qualification [33, 42] and $\{x_k\}_{k\in K}$ is a subsequence that converges to x_*, the set*

$$\{\|\lambda_{k+1}\|,\|\mu_{k+1}\|,\|v_k\|,\|u_k\|\}_{k\in K} \text{ is bounded.} \tag{28}$$

Proof. For all $k \in I\!N$, by (7), (9), (12) and (14), there exist $u_k \in I\!R_+^{p_2}$, $\delta_k \in I\!R^n$ such that $\|\delta_k\| \leq \varepsilon_k$ and

$$\nabla f(x_k)+\sum_{i=1}^{m_1}[\lambda_{k+1}]_i\nabla[h_1(x_k)]_i+\sum_{i=1}^{p_1}[\mu_{k+1}]_i\nabla[g_1(x_k)]_i+\sum_{i=1}^{m_2}[v_k]_i\nabla[h_2(x_k)]_i+\sum_{j=1}^{p_2}[u_k]_j\nabla[g_2(x_k)]_j = \delta_k. \tag{29}$$

By (14), $\mu_{k+1} \in I\!R_+^{p_1}$ for all $k \in I\!N$.

Let $K \subset I\!N$ be such that

$$\lim_{k\in K} x_k = x_*.$$

Suppose that $[g_2(x_*)]_i < 0$. Then, there exists $k_1 \in I\!N$ such that $\forall k \in K, k \geq k_1, [g_2(x_k)]_i < -\varepsilon_k$. Then, by (10),

$$[u_k]_i = 0 \quad \forall k \in K, k \geq k_1.$$

Let us prove now that a similar property takes place when $[g_1(x_*)]_i < 0$. In this case, there exists $k_2 \geq k_1$ such that

$$[g_1(x_k)]_i < 0 \;\; \forall k \in K, k \geq k_2.$$

We consider two cases:

1. $\{[\rho_k]_{m_1+i}\}$ is unbounded.
2. $\{[\rho_k]_{m_1+i}\}$ is bounded.

In the first case we have that $\lim_{k\in K}[\rho_k]_{m_1+i} = \infty$. Since $\{[\bar{\mu}_k]_i\}$ is bounded, there exists $k_3 \geq k_2$ such that, for all $k \in K, k \geq k_3$,

$$[\bar{\mu}_k]_i + [\rho_k]_{m_1+i}[g_1(x_k)]_i < 0.$$

By the definition of μ_{k+1} this implies that

$$[\mu_{k+1}]_i = 0 \quad \forall k \in K, k \geq k_3.$$

Consider now the case in which $\{[\rho_k]_{m_1+i}\}$ is bounded. In this case,

$$\lim_{k\to\infty} [\sigma_k]_i = 0.$$

Therefore, since $[g_1(x_k)]_i$ is bounded away from zero for $k \in K$ large enough,

$$\lim_{k\in K} [\bar{\mu}_k]_i = 0.$$

So, for $k \in K$ large enough,

$$[\bar{\mu}_k]_i + [\rho_k]_{m_1+i}[g_1(x_k)]_i < 0.$$

By the definition of μ_{k+1}, there exists $k_4 \in I\!N$ such that $[\mu_{k+1}]_i = 0$ for $k \in K, k \geq k_4$.

Therefore, there exists $k_5 \geq \max\{k_3, k_4\}$ such that for all $k \in K$, $k \geq k_5$,

$$[g_1(x_*)]_i < 0 \Rightarrow [\mu_{k+1}]_i = 0 \text{ and } [g_2(x_*)]_i < 0 \Rightarrow [u_k]_i = 0. \tag{30}$$

(Observe that, up to now, we did not use the CPLD condition. So (30) takes place even without constraint qualification assumptions.)

By (29) and (30), for all $k \in K, k \geq k_5$, we have:

$$\nabla f(x_k) + \sum_{i=1}^{m_1} [\lambda_{k+1}]_i \nabla [h_1(x_k)]_i + \sum_{[g_1(x_*)]_i = 0} [\mu_{k+1}]_i \nabla [g_1(x_k)]_i + \sum_{i=1}^{m_2} [v_k]_i \nabla [h_2(x_k)]_i$$

$$+ \sum_{[g_2(x_*)]_j = 0} [u_k]_j \nabla [g_2(x_k)]_j = \delta_k, \tag{31}$$

with $\mu_{k+1} \in \mathbb{R}_+^{p_1}, u_k \in \mathbb{R}_+^{p_2}$.

By Caratheodory's Theorem of Cones, for all $k \in K, k \geq k_5$, there exist

$$\widehat{I}_k \subset \{1, \ldots, m_1\}, \widehat{J}_k \subset \{j \mid [g_1(x_*)]_j = 0\}, \hat{I}_k \subset \{1, \ldots, m_2\}, \hat{J}_k \subset \{j \mid [g_2(x_*)]_j = 0\},$$

$$[\widehat{\lambda}_k]_i \; \forall i \in \widehat{I}_k, [\widehat{\mu}_k]_j \geq 0 \; \forall j \in \widehat{J}_k, [\widehat{v}_k]_i \; \forall i \in \hat{I}_k, [\widehat{u}_k]_j \geq 0 \; \forall j \in \hat{J}_k$$

such that the vectors

$$\{\nabla [h_1(x_k)]_i\}_{i \in \widehat{I}_k} \cup \{\nabla [g_1(x_k)]_i\}_{i \in \widehat{J}_k} \cup \{\nabla [h_2(x_k)]_i\}_{i \in \hat{I}_k} \cup \{\nabla [g_2(x_k)]_i\}_{i \in \hat{J}_k}$$

are linearly independent and

$$\begin{gathered} \nabla f(x_k) + \textstyle\sum_{i \in \widehat{I}_k} [\widehat{\lambda}_k]_i \nabla [h_1(x_k)]_i + \sum_{i \in \widehat{J}_k} [\widehat{\mu}_k]_i \nabla [g_1(x_k)]_i + \\ \textstyle\sum_{i \in \hat{I}_k} [\widehat{v}_k]_i \nabla [h_2(x_k)]_i + \sum_{j \in \hat{J}_k} [\widehat{u}_k]_j \nabla [g_2(x_k)]_j = \delta_k. \end{gathered} \tag{32}$$

Since the number of possible sets of indices $\widehat{I}_k, \widehat{J}_k, \hat{I}_k, \hat{J}_k$ is finite, there exists an infinite set $K_1 \subset \{k \in K \mid k \geq k_5\}$ such that

$$\widehat{I}_k = \widehat{I}, \widehat{J}_k = \widehat{J}, \hat{I}_k = \hat{I}, \hat{J}_k = \hat{J},$$

for all $k \in K_1$.

Then, by (32),

$$\begin{gathered} \nabla f(x_k) + \textstyle\sum_{i \in \widehat{I}} [\widehat{\lambda}_k]_i \nabla [h_1(x_k)]_i + \sum_{i \in \widehat{J}} [\widehat{\mu}_k]_i \nabla [g_1(x_k)]_i + \\ \textstyle\sum_{i \in \hat{I}} [\widehat{v}_k]_i \nabla [h_2(x_k)]_i + \sum_{j \in \hat{J}} [\widehat{u}_k]_j \nabla [g_2(x_k)]_j = \delta_k \end{gathered} \tag{33}$$

and the vectors

$$\{\nabla [h_1(x_k)]_i\}_{i \in \widehat{I}} \cup \{\nabla [g_1(x_k)]_i\}_{i \in \widehat{J}} \cup \{\nabla [h_2(x_k)]_i\}_{i \in \hat{I}} \cup \{\nabla [g_2(x_k)]_i\}_{i \in \hat{J}} \tag{34}$$

are linearly independent for all $k \in K_1$.

Let us define

$$S_k = \max\{\max\{|[\widehat{\lambda}_k]_i|, i \in \widehat{I}\}, \max\{[\widehat{\mu}_k]_i, i \in \widehat{J}\}, \max\{|[\widehat{v}_k]_i|, i \in \hat{I}\}, \max\{[\widehat{u}_k]_i, i \in \hat{J}\}\}.$$

We consider two possibilities:

(a) $\{S_k\}_{k\in K_1}$ has a bounded subsequence.

(b) $\lim_{k\in K_1} S_k = \infty$.

If $\{S_k\}_{k\in K_1}$ has a bounded subsequence, there exists $K_2 \subset K_1$ such that

$$\lim_{k\in K_2} [\widehat{\lambda}_k]_i = \widehat{\lambda}_i,$$

$$\lim_{k\in K_2} [\widehat{\mu}_k]_i = \widehat{\mu}_i \geq 0,$$

$$\lim_{k\in K_2} [\widehat{v}_k]_i = \widehat{v}_i,$$

$$\lim_{k\in K_2} [\widehat{u}_k]_i = \widehat{u}_i \geq 0.$$

By (5) and the fact that $\|\delta_k\| \leq \varepsilon_k$, taking limits in (33) for $k \in K_2$, we obtain:

$$\nabla f(x_*) + \sum_{i\in\widehat{I}} \widehat{\lambda}_i \nabla[h_1(x_*)]_i + \sum_{i\in\widehat{J}} \widehat{\mu}_i \nabla[g_1(x_*)]_i + \sum_{i\in\hat{I}} \widehat{v}_i \nabla[h_2(x_*)]_i + \sum_{j\in\hat{J}} \widehat{u}_j \nabla[g_2(x_*)]_j = 0,$$

with $\widehat{\mu}_i \geq 0, \widehat{u}_i \geq 0$. Since $x_* \in \Omega_1 \cap \Omega_2$, we have that x_* is a KKT point of (3).

Suppose now that $\lim_{k\in K_2} S_k = \infty$. Dividing both sides of (33) by S_k we obtain:

$$\begin{array}{c} \frac{\nabla f(x_k)}{S_k} + \sum_{i\in\widehat{I}} \frac{[\widehat{\lambda}_k]_i}{S_k} \nabla[h_1(x_k)]_i + \sum_{i\in\widehat{J}} \frac{[\widehat{\mu}_k]_i}{S_k} \nabla[g_1(x_k)]_i \\ + \sum_{i\in\hat{I}} \frac{[\widehat{v}_k]_i}{S_k} \nabla[h_2(x_k)]_i + \sum_{j\in\hat{J}} \frac{[\widehat{u}_k]_j}{S_k} \nabla[g_2(x_k)]_j = \frac{\delta_k}{S_k}, \end{array} \tag{35}$$

where

$$\left|\frac{[\widehat{\lambda}_k]_i}{S_k}\right| \leq 1, \left|\frac{[\widehat{\mu}_k]_i}{S_k}\right| \leq 1, \left|\frac{[\widehat{v}_k]_i}{S_k}\right| \leq 1, \left|\frac{[\widehat{u}_k]_j}{S_k}\right| \leq 1.$$

Therefore, there exists $K_3 \subset K_1$ such that

$$\lim_{k\in K_3} \frac{[\widehat{\lambda}_k]_i}{S_k} = \widehat{\lambda}_i, \lim_{k\in K_3} \frac{[\widehat{\mu}_k]_i}{S_k} = \widehat{\mu}_i \geq 0, \lim_{k\in K_3} \frac{[\widehat{v}_k]_i}{S_k} = \widehat{v}_i, \lim_{k\in K_3} \frac{[\widehat{u}_k]_j}{S_k} = \widehat{u}_j \geq 0.$$

Taking limits on both sides of (35) for $k \in K_3$, we obtain:

$$\sum_{i\in\widehat{I}} \widehat{\lambda}_i \nabla[h_1(x_*)]_i + \sum_{i\in\widehat{J}} \widehat{\mu}_i \nabla[g_1(x_*)]_i + \sum_{i\in\hat{I}} \widehat{v}_i \nabla[h_2(x_*)]_i + \sum_{j\in\hat{J}} \widehat{u}_j \nabla[g_2(x_*)]_j = 0.$$

But the modulus of at least one of the coefficients $\widehat{\lambda}_i, \widehat{\mu}_i, \widehat{v}_i, \widehat{u}_i$ is equal to 1. Then, the gradients

$$\{\nabla[h_1(x_*)]_i\}_{i\in\widehat{I}} \cup \{\nabla[g_1(x_*)]_i\}_{i\in\widehat{J}} \cup \{\nabla[h_2(x_*)]_i\}_{i\in\hat{I}} \cup \{\nabla[g_2(x_*)]_i\}_{i\in\hat{J}}$$

are positively linearly dependent. By the CPLD condition, the gradients

$$\{\nabla[h_1(x)]_i\}_{i\in\widehat{I}} \cup \{\nabla[g_1(x)]_i\}_{i\in\widehat{J}} \cup \{\nabla[h_2(x)]_i\}_{i\in\hat{I}} \cup \{\nabla[g_2(x)]_i\}_{i\in\hat{J}}$$

must be linearly dependent in a neighborhood of x_*. This contradicts (34). Therefore, the main part of the theorem is proved.

Finally, let us prove that the property (28) holds if x_* satisfies the Mangasarian-Fromovitz constraint qualification. Let us define

$$B_k = \max\{\|\lambda_{k+1}\|_\infty, \|\mu_{k+1}\|_\infty, \|v_k\|_\infty, \|u_k\|_\infty\}_{k \in K}.$$

If (28) is not true, we have that $\lim_{k \in K} B_k = \infty$. In this case, dividing both sides of (31) by B_k and taking limits for an appropriate subsequence, we obtain that the set of gradients

$$\{\nabla [h_1(x_*)]_i\}_{i=1}^{m_1} \cup \{\nabla [g_1(x_*)]_i\}_{[g_1(x_*)]_i=0} \cup \{\nabla [h_2(x_*)]_i\}_{i=1}^{m_2} \cup \{\nabla [g_2(x_*)]_i\}_{[g_2(x_*)]_i=0}$$

is positively linearly dependent. So, x_* does not satisfy the Mangasarian-Fromovitz constraint qualification. □

4 Boundedness of the penalty parameters

When the penalty parameters associated to Penalty or Augmented Lagrangian methods are too large, the subproblems tend to be ill-conditioned and its resolution becomes harder. Level sets of the objective function of the subproblems may emulate long narrow valleys, approximately "parallel" to the penalized constraint set. One of the main motivations for the development of the basic Augmented Lagrangian algorithm is the necessity of overcoming this difficulty. Therefore, the study of conditions under which penalty parameters are bounded plays an important role in Augmented Lagrangian approaches. In this section we prove a penalty boundedness result for Algorithm 2.1. The technique will be the reduction to the equality-constrained case, as in [3], Chapter 2, page 143.

4.1 Equivalence result

The following equality-constrained nonlinear programming problem is equivalent to problem (3):

$$\text{Minimize } F(y) \text{ subject to } H_1(y) = 0, G_1(y) = 0, H_2(y) = 0, G_2(y) = 0, \tag{36}$$

where

$$y = (x, w, z) \in I\!R^{n+p_1+p_2},$$
$$F(y) = f(x), H_1(y) = h_1(x), H_2(y) = h_2(x),$$
$$[G_1(y)]_i = [g_1(x)]_i + w_i^2, i = 1, \ldots, p_1,$$
$$[G_2(y)]_i = [g_2(x)]_i + z_i^2, i = 1, \ldots, p_2.$$

In this section we will assume that $\{x_k\}$ is generated by the application of Algorithm 2.1 to problem (3) and we will consider the application of Algorithm 2.1 to problem (36). In this case, the upper-level constraints will be

$$H_1(y) = 0, \ G_1(y) = 0$$

and the lower-level constraints will be

$$H_2(y) = 0,\ G_2(y) = 0.$$

We also consider that, when we apply Algorithm 2.1 to (36), the algorithmic parameters $\tau, \gamma, \rho_0, \rho_1$ are the ones used in the application of the algorithm to the original problem (3). The initial safeguarded vector of Lagrange multipliers estimates for (36) is $(\bar\lambda_1, \bar\mu_1) \in I\!R^{m_1+p_1}$. The vector of lower bounds for the Lagrange multipliers estimates is $(\bar\lambda_{\min}, \bar\mu_{\min}) \in I\!R^{m_1+p_1}$ and the vector of upper bounds is $(\bar\lambda_{\max}, \bar\mu_{\max}) \in I\!R^{m_1+p_1}$. Finally, in the application of Algorithm 2.1 to problem (36), we use the same initial point x_0 as in the application to (3) and we complete the vector $y_0 = (x_0, w_0, z_0)$ setting

$$[w_0]_i = \sqrt{\max\left\{0, -[g_1(x_0)]_i - \frac{[\bar\mu_0]_i}{[\rho_0]_{m_1+i}}\right\}}, i = 1, \ldots, p_1.$$

$$[z_0]_i = \sqrt{\max\{0, -[g_2(x_0)]_i\}}, i = 1, \ldots, p_2.$$

Theorem 4.1. *Assume that the sequence $\{x_k\}$ is generated by the application of Algorithm 2.1 to problem (3). Suppose that $\lim_{k\to\infty} x_k = x_* \in \Omega_1 \cap \Omega_2$ and that at least one of the following two conditions hold:*

1. *There are no inequality constraints in the definition of Ω_2 (so $p_2 = 0$).*
2. *The point x_* satisfies the Mangasarian-Fromovitz constraint qualification. In this case, let $C > 0$ be such that*
$$[u_k]_i \le C \text{ for all } i = 1, \ldots, p_2,\ k \in I\!N.$$
(A constant with this property exists by Theorem 3.2.)

Define

$$[w_k]_i = \sqrt{\max\left\{0, -[g_1(x_k)]_i - \frac{[\bar\mu_k]_i}{[\rho_k]_{m_1+i}}\right\}}, i = 1, \ldots, p_1, k \in I\!N,$$

$$[z_k]_i = \sqrt{\max\{0, -[g_2(x_k)]_i\}}, i = 1, \ldots, p_2, k \in I\!N, \tag{37}$$

$$y_k = (x_k, w_k, z_k) \text{ for all } k \in I\!N.$$

Then, the sequence $\{y_k\}$ may be generated by Algorithm 2.1 applied to problem (36) with the same sequence of penalty parameters and the tolerance sequence $\{\hat\varepsilon_k\}$ defined by

$$\hat\varepsilon_k = C_{nor} \max\{\varepsilon_k, 2C\sqrt{\varepsilon_k}\},$$

where C_{nor} is a constant that depends only of the chosen norm $\|\cdot\|$.

Proof. The rigorous proof of this theorem is by induction on k. The complete verification is rather tedious and follows as in the exact case (see [3], Chapter 3, page 158). Here we give some details on the step that presents some difficulty, due to the inexactness in the solution of the subproblem.

The Augmented Lagrangian associated with problem (36) is:

$$L_E(y,\lambda,\mu,\rho) = F(x) + \sum_{i=1}^{m_1} \lambda_i [H_1(y)]_i + \sum_{i=1}^{m_1} \frac{[\rho]_i}{2} [H_1(y)]_i^2 + \sum_{i=1}^{p_1} \mu_i [G_1(y)]_i + \sum_{i=1}^{p_1} \frac{[\rho]_{m_1+i}}{2} [G_1(y)]_i^2.$$

Using that x_k satisfies (7)-(11), we must prove that y_k satisfies:

(i)

$$\left\| \nabla L_E(y_k, \bar{\lambda}_k, \bar{\mu}_k, \rho_k) + \sum_{i=1}^{m_2} [v_k]_i \nabla [H_2(y_k)]_i + \sum_{i=1}^{p_2} [u_k]_i \nabla [G_2(y_k)]_i \right\| \le \hat{\varepsilon}_k,$$

(ii)

$$\|H_2(y_k)\| \le \hat{\varepsilon}_k,$$

(iii)

$$\|G_2(y_k)\| \le \hat{\varepsilon}_k.$$

The derivative with respect to z_i of L_E is:

$$\frac{\partial}{\partial z_i} L_E(y_k, \bar{\lambda}_k, \bar{\mu}_k, \rho_k) = 2[z_k]_i [u_k]_i.$$

If $[g_2(x_k)]_i < -\varepsilon_k$, by (10), we have that $[u_k]_i = 0$.
If $[g_2(x_k)]_i \ge 0$, by the definition of $[z_k]_i$, we have that $[z_k]_i = 0$.
Consider the case in which $-\varepsilon_k \le [g_2(x_k)]_i < 0$. Since $[u_k]_i, [z_k]_i \ge 0$, we have, by (37), that:

$$0 \le 2[z_k]_i [u_k]_i = 2\sqrt{-[g_2(x_k)]_i}[u_k]_i \le 2\sqrt{\varepsilon_k}[u_k]_i.$$

The remaining conditions are easy to verify. So, the required result follows from the equivalence of norms in $I\!R^n$. □

4.2 Boundedness in equality constrained problems

Due to the identification of the sequence generated by Algorithm 2.1 (applied to problem (3)) with a sequence generated by the same Algorithm applied to a problem that has only equality constraints, for studying boundedness of the penalty parameters we only need to consider equality-constrained problems.

With this purpose in mind, we consider now $f : I\!R^n \to I\!R, c_1 : I\!R^n \to I\!R^{m_1}, c_2 : I\!R^n \to I\!R^{m_2}$ and the problem

$$\text{Minimize } f(x) \text{ subject to } c_1(x) = 0,\ c_2(x) = 0. \tag{38}$$

The Lagrangian function associated with problem (38) is given by

$$L_0(x, \lambda, v) = f(x) + \langle c_1(x), \lambda\rangle + \langle c_2(x), v\rangle,$$

for all $x \in I\!R^n$, $\lambda \in I\!R^{m_1}$, $v \in I\!R^{m_2}$.

We will use the following assumptions:

Assumption 0. *The sequence $\{x_k\}$ is generated by the application of Algorithm 2.1 to the problem (38) and*

$$\lim_{k\to\infty} x_k = x_*.$$

Assumption 1.

1. *The point x_* is feasible ($c_1(x_*) = 0$ and $c_2(x_*) = 0$).*
2. *The gradients $\nabla[c_1(x_*)]_1, \ldots, \nabla[c_1(x_*)]_{m_1}, \nabla[c_2(x_*)]_1, \ldots, \nabla[c_2(x_*)]_{m_2}$ are linearly independent.*
3. *The functions f, c_1 and c_2 admit continuous second derivatives in a neighborhood of x_*.*
4. *The sufficient second order condition for local minimizers holds. This means that, if $(\lambda_*, v_*) \in I\!R^{m_1+m_2}$ is the vector of Lagrange multipliers associated with x_*, then:*

   $$\langle z, \nabla^2_{xx} L_0(x_*, \lambda_*, v_*) z\rangle > 0$$

   for all $z \neq 0$ such that $\nabla c_1(x_*)^T z = 0$ and $\nabla c_2(x_*)^T z = 0$.

Assumption 2. *For all $k \in I\!N$, $i = 1, \ldots, m_1$, $[\bar\lambda_{k+1}]_i$ will be the projection of $[\lambda_{k+1}]_i$ on the interval $[[\bar\lambda_{\min}]_i, [\bar\lambda_{\max}]_i]$.*

Assumption 3. *For all $i = 1, \ldots, m_1$,*

$$[\lambda_*]_i \in ([\bar\lambda_{\min}]_i, [\bar\lambda_{\max}]_i).$$

Lemma 4.1. *Suppose that Assumption 1 (item 4) holds. For all $\rho \in I\!R^{m_1}_{++}$ define $\pi(\rho) = (1/[\rho]_1, \ldots, 1/[\rho]_{m_1})$. Then, there exists $\bar\rho > 0$ such that, for all $\rho \in I\!R^{m_1}$ such that $\rho \geq \bar\rho$, the matrix*

$$\begin{pmatrix} \nabla^2_{xx} L_0(x_*, \lambda_*, v_*) & \nabla c_1(x_*) & \nabla c_2(x_*) \\ \nabla c_1(x_*)^T & -Diag(\pi(\rho)) & 0 \\ \nabla c_2(x_*)^T & 0 & 0 \end{pmatrix}.$$

is nonsingular.

Proof. The matrix is trivially nonsingular for $\pi(\rho) = 0$. So, the thesis follows by the continuity of the matricial inverse. □

Lemma 4.2. *Suppose that Assumptions 0 and 1 hold. Let $\bar{\rho}$ be as in Lemma 4.1. Suppose that there exists $k_0 \in I\!N$ such that $\rho_k \geq \bar{\rho}$ for all $k \geq k_0$. Define*

$$\alpha_k = \nabla L(x_k, \bar{\lambda}_k, \rho_k) + \nabla c_2(x_k) v_k \tag{39}$$

and

$$\beta_k = c_2(x_k). \tag{40}$$

Then, there exists $M > 0$ such that, for all $k \in I\!N$,

$$\|x_k - x_*\| \leq M \max\left\{\frac{|[\bar{\lambda}_k]_1 - [\lambda_*]_1|}{[\rho_k]_1}, \ldots, \frac{|[\bar{\lambda}_k]_{m_1} - [\lambda_*]_{m_1}|}{[\rho_k]_{m_1}}, \|\alpha_k\|, \|\beta_k\|\right\} \tag{41}$$

and

$$\|\lambda_{k+1} - \lambda_*\| \leq M \max\left\{\frac{|[\bar{\lambda}_k]_1 - [\lambda_*]_1|}{[\rho_k]_1}, \ldots, \frac{|[\bar{\lambda}_k]_{m_1} - [\lambda_*]_{m_1}|}{[\rho_k]_{m_1}}, \|\alpha_k\|, \|\beta_k\|\right\}. \tag{42}$$

Proof. Define, for all $k \in I\!N, i = 1, \ldots, m_1$,

$$[t_k]_i = \frac{[\bar{\lambda}_k]_i - [\lambda_*]_i}{[\rho_k]_i}, \tag{43}$$

$$[\pi_k]_i = \frac{1}{[\rho_k]_i}. \tag{44}$$

By (12), (39) and (40),

$$\nabla L(x_k, \bar{\lambda}_k, \rho_k) + \nabla c_2(x_k) v_k - \alpha_k = 0,$$

$$[\lambda_{k+1}]_i = [\bar{\lambda}_k]_i + [\rho_k]_i [c_1(x_k)]_i, i = 1, \ldots, m_1$$

and

$$c_2(x_k) - \beta_k = 0$$

for all $k \in I\!N$.

Therefore, by (43) and (44), we have:

$$\nabla f(x_k) + \nabla c_1(x_k) \lambda_{k+1} + \nabla c_2(x_k) v_k - \alpha_k = 0,$$

$$[c_1(x_k)]_i - [\pi_k]_i [\lambda_{k+1}]_i + [t_k]_i + [\pi_k]_i [\lambda_*]_i = 0, \; i = 1, \ldots, m_1$$

and

$$c_2(x_k) - \beta_k = 0.$$

Define, for all $\pi \in [0, 1/\bar{\rho}]^{m_1}$,

$$F_\pi : I\!R^n \times I\!R^{m_1} \times I\!R^{m_2} \times I\!R^{m_1} \times I\!R^n \times I\!R^{m_2} \to I\!R^n \times I\!R^{m_1} \times I\!R^{m_2}$$

by

$$F_\pi(x,\lambda,v,t,\alpha,\beta) = \begin{pmatrix} \nabla f(x) + \nabla c_1(x)\lambda + \nabla c_2(x)v - \alpha \\ [c_1(x)]_1 - [\pi]_1[\lambda]_1 + [t]_1 + [\pi]_1[\lambda_*]_1 \\ \cdot \\ \cdot \\ \cdot \\ [c_1(x)]_{m_1} - [\pi]_{m_1}[\lambda]_{m_1} + [t]_{m_1} + [\pi]_{m_1}[\lambda_*]_{m_1} \\ c_2(x) - \beta \end{pmatrix}.$$

Clearly,

$$F_{\pi_k}(x_k, \lambda_{k+1}, v_k, t_k, \alpha_k, \beta_k) = 0 \tag{45}$$

and, by Assumption 1, we have that

$$F_\pi(x_*, \lambda_*, v_*, 0, 0, 0) = 0 \ \ \forall \pi \in [0, 1/\bar{\rho}]^{m_1}. \tag{46}$$

Moreover, the Jacobian matrix of F_π with respect to (x, λ, v) computed at $(x_*, \lambda_*, v_*, 0, 0, 0)$ is:

$$\begin{pmatrix} \nabla^2_{xx} L_0(x_*, \lambda_*, v_*) & \nabla c_1(x_*) & \nabla c_2(x_*) \\ \nabla c_1(x_*)^T & -Diag(\pi) & 0 \\ \nabla c_2(x_*)^T & 0 & 0 \end{pmatrix}.$$

By Lemma 4.1, this matrix is nonsingular for all $\pi \in [0, 1/\bar{\rho}]^{m_1}$. By continuity, the norm of its inverse is bounded in a neighborhood of $(x_*, \lambda_*, v_*, 0, 0, 0)$ uniformly with respect to $\pi \in [0, 1/\bar{\rho}]^{m_1}$. Moreover, the first and second derivatives of F_π are also bounded in a neighborhood of $(x_*, \lambda_*, v_*, 0, 0, 0)$ uniformly with respect to $\pi \in [0, 1/\bar{\rho}]^{m_1}$. Therefore, the bounds (41) and (42) follow from (45) and (46) by the Implicit Function Theorem and the Mean Value Theorem of Integral Calculus. □

Lemma 4.3. *Suppose that Assumptions 0 and 1 hold and*

$$\lim_{k\to\infty} [\rho_k]_i = \infty \ \ \forall \ i = 1, \ldots, m_1.$$

Then,

$$\lim_{k\to\infty} \lambda_k = \lambda_*. \tag{47}$$

Proof. Since $\|\alpha_k\| \le \varepsilon_k, \|\beta_k\| \le \varepsilon_k$ and $\{\bar{\lambda}_k\}$ is bounded, the right-hand side of (42) tends to zero. This implies that (47) holds. □

Lemma 4.4. *In addition to the hypotheses of Lemma 4.3, suppose that Assumptions 2 and 3 hold. Then*

$$\bar{\lambda}_k = \lambda_k$$

for k large enough.

Proof. By Lemma 4.3, λ_k tends to λ_* which, by Assumption 3, belongs to the open set

$$([\bar{\lambda}_{\min}]_1, [\bar{\lambda}_{\max}]_1) \times \ldots \times ([\bar{\lambda}_{\min}]_{m_1}, [\bar{\lambda}_{\max}]_{m_1}).$$

Therefore, λ_k belongs to this open set for k large enough. Then, by Assumption 2, $\bar{\lambda}_{k+1} = \lambda_{k+1}$ for k large enough. This completes the proof. □

Theorem 4.2. *Suppose that Assumptions 0, 1, 2 and 3 are satisfied by the sequence generated by Algorithm 2.1 applied to the problem (38). In addition, assume that:*

1. *Rule 1 is used at all the iterations of the algorithm and $[\rho_0]_i = [\rho_0]_1$ for all $i = 1, \ldots, m_1$.*
2. *There exists a sequence $\eta_k \to 0$ such that*

$$\varepsilon_k \leq \eta_k \|c_1(x_k)\|_\infty \tag{48}$$

for all $k \in I\!N$.

Then, the sequence of penalty parameters $\{\rho_k\}$ is bounded.

Proof. In order to simplify the notation, since Rule 1 is the one used here, we write $\rho_k = [\rho_k]_i$ for all $i = 1, \ldots, m_1, k \in I\!N$.

Assume, by contradiction, that

$$\lim_{k\to\infty} \rho_k = \infty. \tag{49}$$

Since $c_1(x_*) = 0$, by the continuity of the first derivatives of c_1 there exists $L > 0$ such that, for all $k \in I\!N$,

$$\|c_1(x_k)\|_\infty \leq L\|x_k - x_*\|.$$

Therefore, by (41), (48) and Lemma 4.4, we have that

$$\|c_1(x_k)\|_\infty \leq LM \max\left\{\frac{\|\lambda_k - \lambda_*\|_\infty}{\rho_k}, \eta_k\|c_1(x_k)\|_\infty\right\}$$

for k large enough. Since η_k tends to zero, this implies that

$$\|c_1(x_k)\|_\infty \leq LM\frac{\|\lambda_k - \lambda_*\|_\infty}{\rho_k} \tag{50}$$

for k large enough.

By (13) and Lemma 4.4, we have that $\lambda_k = \lambda_{k-1} + \rho_{k-1}c_1(x_{k-1})$ for k large enough. Therefore,

$$\|c_1(x_{k-1})\|_\infty = \frac{\|\lambda_k - \lambda_{k-1}\|_\infty}{\rho_{k-1}} \geq \frac{\|\lambda_{k-1} - \lambda_*\|_\infty}{\rho_{k-1}} - \frac{\|\lambda_k - \lambda_*\|_\infty}{\rho_{k-1}} \tag{51}$$

for k large enough.

Now, by (42), Lemma 4.4 and (48),

$$\|\lambda_k - \lambda_*\|_\infty \leq M\left(\frac{\|\lambda_{k-1} - \lambda_*\|_\infty}{\rho_{k-1}} + \eta_{k-1}\|c_1(x_{k-1})\|_\infty\right)$$

for k large enough. So,

$$\frac{\|\lambda_{k-1} - \lambda_*\|_\infty}{\rho_{k-1}} \geq \frac{\|\lambda_k - \lambda_*\|_\infty}{M} - \eta_{k-1}\|c_1(x_{k-1})\|_\infty$$

for k large enough. Therefore, by (51),

$$(1+\eta_{k-1})\|c_1(x_{k-1})\|_\infty \geq \|\lambda_k - \lambda_*\|_\infty \left(\frac{1}{M} - \frac{1}{\rho_{k-1}}\right) \geq \frac{1}{2M}\|\lambda_k - \lambda_*\|_\infty$$

for k large enough. Thus,

$$\|\lambda_k - \lambda_*\|_\infty \leq 3M\|c_1(x_{k-1})\|_\infty$$

for k large enough. By (50), this implies that

$$\|c_1(x_k)\|_\infty \leq \frac{3LM^2}{\rho_k}\|c_1(x_{k-1})\|_\infty.$$

Therefore, since $\rho_k \to \infty$, there exists $k_1 \in I\!\!N$ such that

$$\|c_1(x_k)\| \leq \tau\|c_1(x_{k-1})\|$$

for all $k \geq k_1$. This implies that $\rho_{k+1} = \rho_k$ for all $k \geq k_1$. Thus, (49) is false. □

4.3 Boundedness in the general case

The final boundedness result for the penalty parameters associated to Algorithm 2.1 is given in Theorem 4.3. As in Theorem 4.2 a crucial assumption will be that the precision used to solve subproblems must tend to zero faster than the upper-level feasibility measure. This type of requirement is usual in many Augmented Lagrangian and Multiplier methods [3, 4, 12, 14, 19, 20, 21, 30].

Assumption 5. *We assume that*

1. *The sequence $\{x_k\}$ is generated by the application of Algorithm 2.1 to the problem (3) and*

$$\lim_{k\to\infty} x_k = x_*.$$

2. *The point x_* is feasible ($h_1(x_*) = 0, g_1(x_*) \leq 0, h_2(x_*) = 0, g_2(x_*) \leq 0.$)*

3. *The gradients*

$$\{\nabla[h_1(x_*)]_i\}_{i=1}^{m_1}, \{\nabla[g_1(x_*)]_i\}_{[g_1(x_*)]_i=0}, \{\nabla[h_2(x_*)]_i\}_{i=1}^{m_2}, \{\nabla[g_2(x_*)]_i\}_{[g_2(x_*)]_i=0}$$

are linearly independent.

4. *Strict complementarity takes place at x_*. This means that, if $\mu_* \in I\!\!R_+^{p_1}$ and $u_* \in I\!\!R_+^{p_2}$ are the Lagrange multipliers corresponding to the constraints $g_1(x) \leq 0$ and $g_2(x) \leq 0$, respectively, then:*

$$[g_1(x_*)]_i = 0 \Rightarrow [\mu_*]_i > 0$$

and

$$[g_2(x_*)]_i = 0 \Rightarrow [u_*]_i > 0.$$

5. *The functions f, h_1, g_1, h_2 and g_2 admit continuous second derivatives in a neighborhood of x_*.*

6. *Define the tangent subspace T as the set of all $z \in I\!R^n$ such that*

$$\nabla h_1(x_*)^T z = \nabla h_2(x_*)^T z = 0,$$

$$\langle \nabla [g_1(x_*)]_i, z \rangle = 0$$

for all i such that $[g_1(x_*)]_i = 0$ and

$$\langle \nabla [g_2(x_*)]_i, z \rangle = 0$$

for all i such that $[g_2(x_*)]_i = 0$.
Then, for all $z \in T, z \neq 0$,

$$\langle z, [\nabla^2 f(x_*) + \sum_{i=1}^{m_1} [\lambda_*]_i \nabla^2 [h_1(x_*)]_i + \sum_{i=1}^{p_1} [\mu_*]_i \nabla^2 [g_1(x_*)]_i + \sum_{i=1}^{m_2} [v_*]_i \nabla^2 [h_2(x_*)]_i + \sum_{i=1}^{p_2} [u_*]_i \nabla^2 [g_2(x_*)]_i] z \rangle > 0.$$

Proposition 4.1 *Suppose that Assumption 5 holds. Define*

$$w_* = (\sqrt{-[g_1(x_*)]_1}, \dots, \sqrt{-[g_1(x_*)]_{p_1}}),$$

$$z_* = (\sqrt{-[g_2(x_*)]_1}, \dots, \sqrt{-[g_2(x_*)]_{p_2}})$$

and

$$y_* = (x_*, w_*, z_*).$$

Consider problem (36) with

$$c_1(y) = (H_1(y), G_1(y)), c_2(y) = (H_2(y), G_2(y)).$$

Then, Assumption 1 is satisfied for problem (36).

Proof. See Proposition 3.2 of [3]. □

Theorem 4.3. *Suppose that Assumptions 5 holds. In addition, assume that:*

1. *Rule 1 is used at all the iterations of the algorithm and $[\rho_0]_i = [\rho_0]_1$ for all $i = 1, \dots, m_1 + p_1$.*

2. *There exists a sequence $\eta_k \to 0$ such that*

$$\varepsilon_k \leq \eta_k \max\{\|h_1(x_k)\|_\infty, \|\sigma_k\|_\infty\} \; \forall \; k \in I\!N.$$

3. *$[\lambda_*]_i \in ([\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i) \; \forall \; i = 1, \dots, m_1$ and $[\mu_*]_i \in ([\bar{\mu}_{\min}]_i, [\bar{\mu}_{\max}]_i) \; \forall \; i = 1, \dots, p_1$.*

4. $[\bar\lambda_{k+1}]_i$ is the projection of $[\lambda_{k+1}]_i$ on $[[\bar\lambda_{\min}]_i, [\bar\lambda_{\max}]_i]$ and $[\bar\mu_{k+1}]_j$ is the projection of $[\mu_{k+1}]_j$ on $[[\bar\mu_{\min}]_j, [\bar\mu_{\max}]_j]$ for all $i = 1, \dots, m_1, j = 1, \dots, p_1, k \in I\!N$.

Then, the sequence of penalty parameters $\{\rho_k\}$ is bounded.

Proof. Consider the equivalent problem (36). By the hypotheses of the present theorem and Proposition 4.1, the assumptions of Theorem 4.2 are fulfilled. So, the desired result follows by Theorem 4.2. □

5 Numerical experiments

5.1 Box lower-level constraints

In order to assess the performance of the main algorithm in the widely established Augmented Lagrangian framework in which the subproblems are bound-constrained problems, we compared it against Lancelot [13]. The implementation of the algorithm for the case in which the lower-level set is a box uses the code Gencan [6] for solving the subproblems, so it will be called Algencan (Augmented Lagrangian using Gencan) from now on.

We test two versions of Algencan: with only one penalty parameter and with one penalty parameter per constraint (the penalty parameters are updated using Rules 1 and 2 of Step 4, respectively). We also consider two versions of Lancelot: with all its default options and using true Hessians and Conjugate gradients. To perform the numerical experiments, we considered all the problems of the Cute collection [10]. As a whole, we tried to solve 873 problems.

For both versions of Algencan , based on the numerical experiments reported in [1, 5], we set $\tau = 0.5$, $\gamma = 10$, $\bar\mu_{\min} = \bar\lambda_{\min} = -10^{20}$, $\bar\mu_{\max} = \bar\lambda_{\max} = 10^{20}$, $\varepsilon_k = 10^{-4}$ for all k, and ρ_0 and ρ_1 with all their components equal to 10. For $\bar\mu_0 = \bar\mu_1$ and $\bar\lambda_1$, as well as for the calculation of $\bar\mu_k$ and $\bar\lambda_k$ at Step 3 of Algencan , we consider the projections of μ_k and λ_k onto $[\bar\mu_{\min}, \bar\mu_{\max}]^{p_1}$ and $[\bar\lambda_{\min}, \bar\lambda_{\max}]^{m_1}$, respectively. For Algencan and Lancelot we use the initial guess x_0 provided by Cute and the null vector for the initial Lagrange multipliers approximation. As stopping criterion we used $\max(\|h_1(x_k)\|_\infty, \|\sigma_k\|_\infty) \le 10^{-4}$ and $\|P_{\text{box}}[x_k - \nabla L(x_k, \bar\lambda_k, \bar\mu_k, \rho_k)] - x_k\|_\infty \le 10^{-4}$, where P_{box} represents the projection onto the lower-level box.

All experiments were done in a Sun Fire 880 with 8 900 Mhz UltraSPARC III Processors, 32 Gb of RAM memory, running SunOS 5.8. The codes were written in Fortran 77 and compiled with Forte Developer 7 Fortran 95 7.0 2002/03/09. We used the option -O3 to optimize the code.

We compared four methods: two versions of Lancelot and two versions of Algencan . Given a fixed problem, for each method M, we define x^M_{final} the final point obtained by M when solving the given problem. In this numerical study we say that x^M_{final} *is feasible* if

$$\max\{\|h_1(x^M_{\text{final}})\|_\infty, \|g_1(x^M_{\text{final}})_+\|_\infty\} \le 10^{-4}.$$

(Observe that, by the structure of the lower-level constraints and the solvers of the subproblems, lower-level constraints are always satisfied.)

We define

$$f_{\text{best}} = \min_M \{ f(x_{\text{final}}^M) \mid x_{\text{final}}^M \text{ is feasible} \}.$$

We say that the method M *found a solution* of the problem if x_{final}^M is feasible and

$$f(x_{\text{final}}^M) \le f_{\text{best}} + 10^{-3}|f_{\text{best}}| + 10^{-6}.$$

Finally, let t^M be the computer CPU time that method M used to arrive to x_{final}^M. We define

$$r^M = \begin{cases} t^M, & \text{if method } M \text{ found a solution,} \\ \infty, & \text{otherwise.} \end{cases}$$

The performance profiles for comparing ALGENCAN and LANCELOT use r as a performance measurement. See Figure 1. Readers familiar with performance-profile curves [18] may recognize that ALGENCAN with one penalty parameter is more efficient that the other three Augmented Lagrangian algorithms. With respect to robustness (right-hand side of the performance-profile function), we found that this version of ALGENCAN found a solution in 731 (83.73%) problems, whereas the many-parameter version of ALGENCAN did it in 707 (80.99%) problems, the default LANCELOT solved 712 (81.56%) problems and the Hessian-CG LANCELOT solved 669 (76.63%) problems.

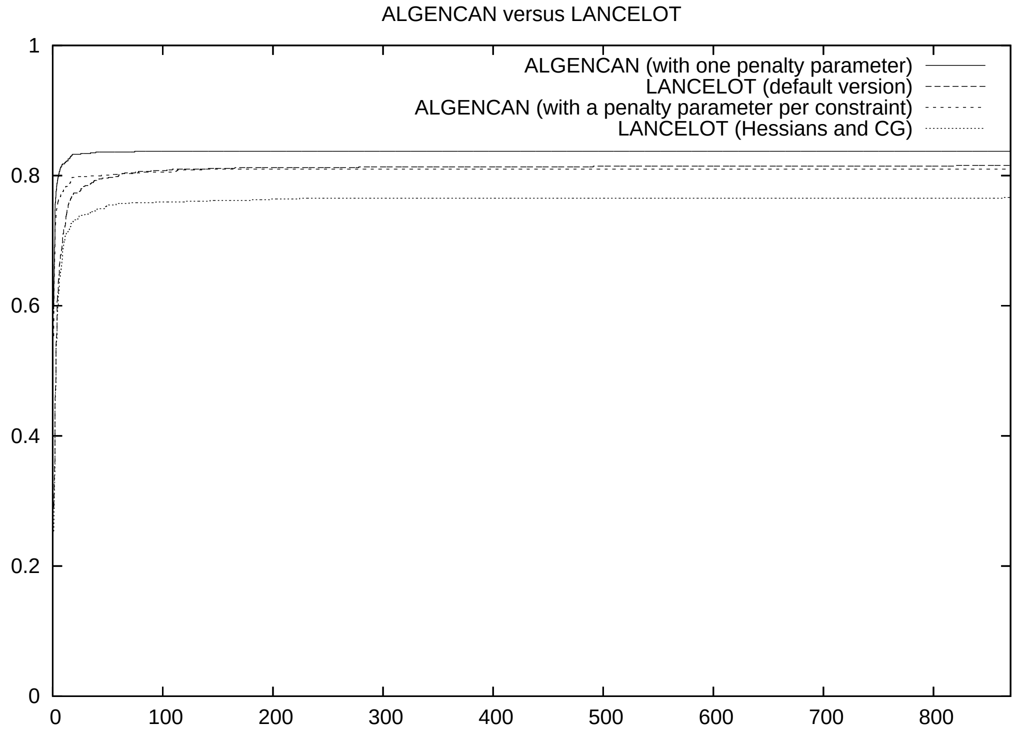

Figure 1: Comparison between ALGENCAN and LANCELOT using performance profiles.

5.2 General lower-level constraints

Here we will consider a variant of the family of *location* problems introduced in [8]. In the original problem, given a set of np disjoint polygons $P_1, P_2, \ldots, P_{np}$ in $I\!\!R^2$ one wishes to find the point $z^1 \in P_1$ that minimizes the sum of the distances to the other polygons. Therefore, the problem was:

$$\min_{z^i,\, i=1,\ldots,np} \quad \frac{1}{np-1} \sum_{i=2}^{np} \|z^i - z^1\|_2$$

$$\text{subject to} \quad z^i \in P_i, \; i = 1, \ldots, np.$$

In the variant considered in the present work, we have, in addition to the np polygons, nc circles. Moreover, there is an ellipse which has a non empty intersection with P_1 and such that z_1 must be inside the ellipse and $z_i, i = 2, \ldots, np + nc$ must be outside. Therefore, the problem considered in this work is

$$\min_{z^i,\, i=1,\ldots,np+nc} \quad \frac{1}{nc+np-1} \left[\sum_{i=2}^{np} \|z^i - z^1\|_2 + \sum_{i=1}^{nc} \|z^{np+i} - z^1\|_2 \right]$$

$$\begin{array}{lrcll} \text{subject to} & g(z^1) & \leq & 0, & \\ & g(z^i) & \geq & 0, & i = 2, \ldots, np + nc, \\ & z^i & \in & P_i, & i = 1, \ldots, np, \\ & z^{np+i} & \in & C_i, & i = 1, \ldots, nc, \end{array}$$

where $g(x) = (x_1/a)^2 + (x_2/b)^2 - c$, and $a, b, c \in I\!\!R$ are positive constants.

We generated 18 problems of this class, varying nc and np and choosing randomly the location of the circles and polygons and the number of vertices of each polygon. The details of the generation, including the way in which we guarantee empty intersections (in order to have differentiability everywhere), are rather tedious but, of course, are available for interested readers. In Table 1 we display the main characteristics of each problem (number of circles, number of polygons, total number of vertices of the polygons, dimension of the problem and number of lower-level and upper-level constraints). Figure 2 shows the solution of a very small twelve-sets problem that has 24 variables, 81 lower-level constraints and 12 upper-level constraints.

To solve this family of problems, we will consider $g(z^1) \leq 0$ and $g(z^i) \geq 0, i = 2, \ldots, np + nc$ as upper-level constraints, and $z^i \in P_i, i = 1, \ldots, np$ and $z^{np+i} \in C_i, i = 1, \ldots, nc$ as lower-level constraints. In this way the subproblems can be efficiently solved by the Spectral Projected Gradient method (SPG) [7, 8] as suggested by the experiments in [8]. So, we implemented an Augmented Lagrangian method that uses SPG to solve the subproblems. This implementation will be called Alspg. In general, it would be interesting to apply Alspg to any problem such that the selected lower-level constraints define a convex set for which it is easy (cheap) to compute the projection of an arbitrary point.

Table 2 shows figures that allow us to have an ideia of the numerical performance of Alspg applied to the problems of Table 1. We used $\epsilon = 10^{-4}$ as a tolerance for feasibility and optimality. Alspg satisfied the convergence criterion in all the problems.

Problem	nc	np	$totnvs$	n	p_1	p_2
1	2,929	4,935	61,755	15,728	7,864	64,684
2	4,403	7,346	91,896	23,498	11,749	96,299
3	6,370	10,732	134,165	34,204	17,102	140,535
4	13,344	22,158	276,927	71,004	35,502	290,271
5	19,924	33,265	415,691	106,378	53,189	435,615
6	29,998	49,754	621,824	159,504	79,752	651,822
7	26,288	43,775	546,943	140,126	70,063	573,231
8	39,459	65,891	823,823	210,700	105,350	863,282
9	59,494	98,875	1,236,031	316,738	158,369	1,295,525
10	65,146	109,360	1,367,104	349,012	174,506	1,432,250
11	98,138	163,854	2,048,239	523,984	261,992	2,146,377
12	147,511	245,983	3,074,109	786,988	393,494	3,221,620
13	130,344	218,296	2,729,187	697,280	348,640	2,859,531
14	195,725	327,418	4,093,497	1,046,286	523,143	4,289,222
15	293,913	490,522	6,130,282	1,568,870	784,435	6,424,195
16	260,705	436,037	5,449,661	1,393,484	696,742	5,710,366
17	391,023	653,802	8,172,819	2,089,650	1,044,825	8,563,842
18	587,182	980,615	12,257,368	3,135,594	1,567,797	12,844,550

Table 1: Location problems and their main features. The problem generation is based on a grid. The number of city-circles (nc) and city-polygons (np) depend on the number of points in the grid, the probability of having a city in a grid point ($procit$) and the probability of a city to be a polygon ($propol$) or a circle ($1 - propol$). The number of vertices of a city-polygon is a random number and the total number of vertices of all the city-polygons together is $totnvs$. Finally, the number of variables of the problem is $n = 2(nc + np)$, the number of upper-level inequality constraints is $p_1 = nc + np$ and the number of lower-level inequality constraints is $p_2 = nc + totnvs$. The central rectangle is considered here a "special" city-polygon. The lower-level constraints correspond to the fact that each point must be inside a city and the upper-level constraints come from the fact that the central point must be inside the ellipse and all the others must be outside.

Problem	OuIt	InIt	Fcnt	Gcnt	Time (secs)	f
1	9	334	488	343	12.28	4.5601D+02
2	9	677	949	686	41.30	5.6228D+02
3	11	245	390	256	23.63	6.8965D+02
4	7	69	120	76	18.11	4.6149D+02
5	9	283	476	292	111.45	5.6461D+02
6	10	88	170	98	58.68	6.9408D+02
7	7	71	122	78	40.29	4.6326D+02
8	7	65	121	72	58.09	5.6419D+02
9	8	105	165	113	136.42	6.9382D+02
10	8	71	135	79	111.41	4.6272D+02
11	7	66	121	73	158.23	5.6425D+02
12	9	87	163	96	308.75	6.9381D+02
13	7	55	98	62	185.13	4.6280D+02
14	7	91	151	98	421.58	5.6440D+02
15	10	83	166	93	602.97	6.9384D+02
16	8	68	138	76	458.96	4.6271D+02
17	7	80	140	87	765.60	5.6433D+02
18	10	85	171	95	1234.53	6.9408D+02

Table 2: Performance of ALSPG on the location problems. The memory limitation is the only inconvenient for ALSPG solving problems with higher dimension than problem 18 (approximately 3×10^6 variables, 1.5×10^6 upper-level inequality constraints, and 1.2×10^7 lower-level inequality constraints), since computer time is very reasonable.

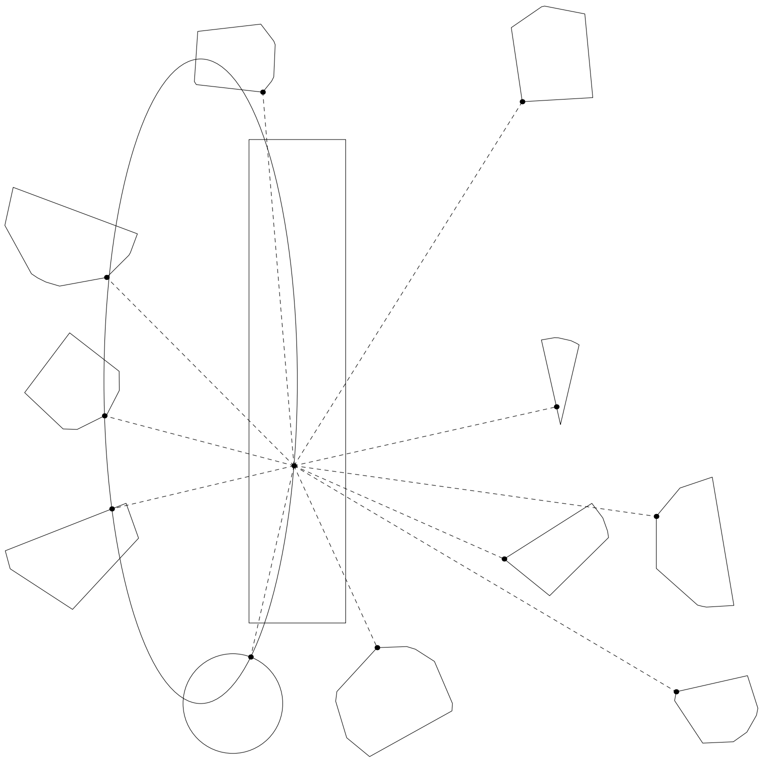

Figure 2: Twelve-sets very small location problem.

6 Final Remarks

In the last few years many sophisticated algorithms for nonlinear programming have been published. They usually involve combinations of interior-point techniques, sequential quadratic programming, trust regions [15], restoration, nonmonotone strategies and advanced sparse linear algebra procedures. See, for example [11], [26], [27], [28], [29], [34] and the extensive reference lists of these papers. Moreover, methods for solving efficiently specific problems or for dealing with special constraints are often introduced. Many times, a particular algorithm is extremely efficient for dealing with problems of a given type, but fails (or cannot be applied) when constraints of a different class are incorporated. Unfortunately, this situation is quite common in engineering applications. In the Augmented Lagrangian framework additional constraints are naturally incorporated to the objective function of the subproblems, which therefore preserve their constraint structure. For this reason, we conjecture that the Augmented Lagrangian approach (with general lower-level constraints) will continue to be used for many years.

This fact motivated us to improve and analyze Augmented Lagrangian methods with arbitrary lower-level constraints. From the theoretical point of view our goal was to eliminate, as much as possible, restrictive constraint qualifications. With this in mind we used, both in the feasibility proof and in the optimality proof, the Constant Positive Linear Dependence (CPLD) condition introduced by Qi and Wei in [38]. This condition has been proved to be a constraint qualification in [2] where its relations with other constraint qualifications have been given.

In addition, we proved a result on the boundedness of penalty parameters for which we used a more restrictive constraint qualification than CPLD.

We felt that the reliability of an Augmented Lagrangian approach with general constraints needed to be assessed in two different ways: On one hand, we felt the necessity of providing a robust and efficient computer algorithm for the version of the method with simple (box) lower-level constraints. Our implementation was based on the robust GENCAN algorithm for box-constrained minimization [6]. We used a large basis of comparison: all the problems of the CUTE collection and we compared our default version with the default version of LANCELOT . We showed that our basic algorithm is at least as robust and efficient as LANCELOT for solving this set of problems. This was rather surprising since the default version of LANCELOT uses second-derivative-based preconditioners, which are not present at all in the implemented version of GENCAN .

On the other hand, we provided a family of examples (Location Problems) where the potentiality of the arbitrary lower-level approach is clearly evidenced. This example represents a typical situation in applications. A specific algorithm (SPG) is known to be very efficient for a class of problems but turns out to be impossible to apply when additional constraints are incorporated. Fortunately, the Augmented Lagrangian approach is able to deal with the additional constraints taking advantage of the efficiency of SPG for solving the subproblems. In this way, we were able to solve nonlinear programming problems with more than 3,000,000 variables and 14,000,000 constraints in less than half an hour of CPU time.

Among the many situations in which the arbitrary lower-level approach may be used, it is worthwhile to mention:

1. Minimizing a quadratic subject to a ball and linear constraints: This problem is useful

in the context of trust-region methods for minimization with linear constraints. In the low-dimensional case the problem may be efficiently reduced to the classical trust-region subproblem [25, 35], using a basis of the null-space of the linear constraints, but in the large-scale case this procedure may be impractical. On the other hand, efficient methods for minimizing a quadratic within a ball exist, even in the large-scale case [40, 43]. Therefore, it is attractive to solve the problem keeping the ball in the lower level and defining the linear constraints as upper-level ones.

2. Bilevel problems with "additional" constraints [16]. A basic bilevel problem consists in minimizing $f(x, y)$ subject to the fact that y solves an optimization problem whose data depend on x. Efficient algorithms for this problem have already been developed (see [16] and references therein). When additional constraints ($h(x, y) = 0, g(x, y) \leq 0$) are present the problem is more complicated. These additional constraints are serious candidates to become upper-level constraints in an Augmented Lagrangian approach since, in that case, the subproblems are "ordinary" bilevel problems. Similar considerations can be made for many mathematical programming problems with equilibrium constraints [32].

3. Minimization with orthogonality constraints [22, 24, 36, 44]. Important problems on this class appear in many applications, such as the "ab initio" calculation of electronic structures. Reasonable algorithms for minimization with (only) orthogonality constraints exist, but they cannot be used in the presence of additional constraints. When these additional constraints appear in an application the most obvious way to proceed is to incorporate them to the objective function, keeping the orthogonality constraints in the lower level.

Many interesting open problems remain:

1. The constraint qualification used for obtaining boundedness of the penalty parameter (regularity at the limit point) is still too strong. We conjecture that it is possible to obtain the same result using the Mangasarian-Fromovitz constraint qualification.

2. An alternative definition of σ_k at the main algorithm seems to be well-motivated: instead of using the approximate multiplier already employed it seems to be natural to use the current approximation to the inequality Lagrange multipliers (μ_{k+1}). It is possible to obtain the global convergence results with this modification but it is not clear how to obtain boundedness of the penalty parameter. Moreover, from the practical point of view it is not clear if such modification produces numerical improvements.

3. The inexact-Newton approach employed by GENCAN for solving box-constrained subproblems does not seem to be affected by the nonexistence of second derivatives of the Augmented Lagrangian for inequality constrained problems. There are good reasons to conjecture that this is not the case when the box-constrained subproblem is solved using a quasi-Newton approach. This fact stimulates the development of efficient methods for minimizing functions with first (but not second) derivatives.

4. The implementation of Augmented Lagrangian methods (as well as other nonlinear programming algorithms) is subject to many decisions on the parameters to be employed.

Some of these decisions are not easy to take and one is compelled to use parameters largely based on experience. Theoretical criteria for deciding the best values of many parameters need to be developed.

5. In [1] an Augmented Lagrangian algorithm with many penalty parameters for single (box) lower-level constraints was analyzed and boundedness of the penalty parameters was proved without strict complementarity assumptions. The generalization of that proof to the general lower-level constraints case considered here is not obvious and the existence of such generalization remains an open problem.

6. Acceleration and warm-start procedures must be developed in order to speed the ultimate rate of convergence and to take advantage of the solution obtained for slightly different optimization problems.

Acknowledgements

We are indebted to the *Advanced Database Laboratory (LABD)* and *Bioinfo* at the Institute of Mathematics and Statistics of the University of São Paulo for computer facilities.

References

[1] R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt. *Augmented Lagrangian Methods under the Constant Positive Linear Dependence Constraint Qualification.* Technical Report MCDO-040806 (see `www.ime.usp.br/∼egbirgin`), Department of Applied Mathematics, UNICAMP, Brazil, 2004.

[2] R. Andreani, J. M. Martínez and M. L. Schuverdt. *On the relation between the Constant Positive Linear Dependence condition and quasinormality constraint qualification.* To appear in Journal of Optimization Theory and Applications.

[3] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, 1982; republished by Athena Scientific, 1996.

[4] D. P. Bertsekas. *Nonlinear Programming*, 2nd edition, Athena Scientific, Belmont, Massachusetts, 1999.

[5] E. G. Birgin, R. Castillo and J. M. Martínez. *Numerical comparison of Augmented Lagrangian algorithms for nonconvex problems.* To appear in Computational Optimization and Applications.

[6] 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).

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

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

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

[10] I. Bongartz, A. R. Conn, N. I. M. Gould and Ph. L. Toint. CUTE *: constrained and unconstrained testing environment*, ACM Transactions on Mathematical Software 21, pp. 123–160 (1995).

[11] R. H. Byrd, N. I. M. Gould, J. Nocedal and R. A. Waltz. *An algorithm for nonlinear optimization using linear programming and equality constrained subproblems*, Mathematical Programming 100, pp. 27–48 (2004).

[12] A. R. Conn, N. I. M. Gould, A. Sartenaer and Ph. L. Toint. *Convergence properties of an Augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints*, SIAM Journal on Optimization 6, pp. 674–703 (1996).

[13] A. R. Conn, N. I. M. Gould and Ph. L. Toint. LANCELOT *: A Fortran package for large scale nonlinear optimization.* Springer-Verlag, Berlin, 1992.

[14] A. R. Conn, N. I. M. Gould and Ph. L. Toint. *A globally convergent Augmented Lagrangian algorithm for optimization with general constraints and simple bounds*, SIAM Journal on Numerical Analysis 28, pp. 545–572 (1991).

[15] A. R. Conn, N. I. M. Gould and Ph. L. Toint. *Trust Region Methods.* MPS/SIAM Series on Optimization, SIAM, Philadelphia, 2000.

[16] S. Dempe. *Foundations of Bilevel Programming*, Kluwer Academic Publishers, Dordrecht, 2002.

[17] M. A. Diniz-Ehrhardt, M. A. Gomes-Ruggiero, J. M. Martínez and 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).

[18] E. D. Dolan and J. J. Moré. *Benchmarking optimization software with performance profiles*, Mathematical Programming 91, pp. 201–213 (2002).

[19] Z. Dostál. *Inexact semi-monotonic Augmented Lagrangians with optimal feasibility convergence for convex bound and equality constrained quadratic programming.* To appear in SIAM Journal on Numerical Analysis.

[20] Z. Dostál, A. Friedlander and S. A. Santos. *Augmented Lagrangian with adaptive precision control for quadratic programming with simple bounds and equality constraints*, SIAM Journal on Optimization 13, pp. 1120–1140 (2003).

[21] Z. Dostál, F. A. M. Gomes and S. A. Santos. *Duality based domain decomposition with natural coarse space for variational inequalities*, Journal of Computational and Applied Mathematics 126, pp. 397–415 (2000).

[22] A. Edelman, T. A. Arias and S. T. Smith. *The geometry of algorithms with orthogonality constraints*, SIAM Journal on Matrix Analysis and Applications 126, pp. 303–353 (1998).

[23] R. Fletcher. *Practical Methods of Optimization*, Academic Press, London, 1987.

[24] J. B. Francisco, J. M. Martínez and L. Martínez. *Globally convergent trust-region methods for self-consistent field electronic structure calculations*, Journal of Chemical Physics 121, pp. 10863-10878 (2004).

[25] D. M. Gay. *Computing optimal locally constrained steps*, SIAM Journal on Scientific and Statistical Computing 2, pp. 186-197 (1981).

[26] E. M. Gertz and P. E. Gill. *A primal-dual trust region algorithm for nonlinear optimization*, Mathematical Programming 100, pp. 49-94 (2004).

[27] C. C. Gonzaga, E. Karas and M. Vanti. *A globally convergent filter method for Nonlinear Programming*, SIAM Journal on Optimization 14, pp. 646–669, 2003.

[28] N. I. M. Gould, D. Orban, A. Sartenaer and Ph. L. Toint. *Superlinear Convergence of Primal-Dual Interior Point Algorithms for Nonlinear Programming*, SIAM Journal on Optimization 11, pp.974-1002 (2000).

[29] N. I. M. Gould, D. Orban and Ph. L. Toint. *GALAHAD: a library of thread-safe Fortran 90 packages for large-scale nonlinear optimization*, ACM Transactions on Mathematical Software 29, pp. 353-372 (2003).

[30] W. W. Hager. *Analysis and implementation of a dual algorithm for constrained optimization*, Journal of Optimization Theory and Applications 79, pp. 37–71 (1993).

[31] M. R. Hestenes. *Multiplier and gradient methods*, Journal of Optimization Theory and Applications 4, pp. 303–320 (1969).

[32] Z-Q. Luo, J-S. Pang and D. Ralph. *Mathematical programs with Equilibrium Constraints.* Cambridge University Press, Cambridge, 1996.

[33] O. L. Mangasarian and S. Fromovitz. *The Fritz-John necessary optimality conditions in presence of equality and inequality constraints*, Journal of Mathematical Analysis and Applications 17, pp. 37–47 (1967).

[34] J. M. Martínez. *Inexact Restoration Method with Lagrangian tangent decrease and new merit function for Nonlinear Programming*, Journal of Optimization Theory and Applications 111, pp. 39–58, (2001).

[35] J. J. Moré and D. C. Sorensen. *Computing a trust region step*, SIAM Journal on Scientific and Statistical Computing 4, pp. 553-572 (1983).

[36] M. D. Plumbey. *Lie group methods for optimization with orthogonality constraints*, Lecture Notes in Computer Science 3195, pp. 1245-1252 (2004).

[37] M. J. D. Powell. *A method for nonlinear constraints in minimization problems*, in *Optimization*, R. Fletcher (ed.), Academic Press, New York, NY, pp. 283–298, 1969.

[38] 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).

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

[40] F. Rendl and H. Wolkowicz. *A semidefinite framework for trust region subproblems with applications to large scale minimization.* Mathematical Programming 77, pp. 273-299 (1997).

[41] R. T. Rockafellar. *The multiplier method of Hestenes and Powell applied to convex programming.* Journal of Optimization Theory and Applications 12, pp. 555–562 (1973).

[42] R. T. Rockafellar. *Lagrange multipliers and optimality*, SIAM Review 35, pp. 183–238 (1993).

[43] M. Rojas, S. A. Santos and D. C. Sorensen. *A new matrix–free algorithm for the large–scale trust–region subproblem*, SIAM Journal on Optimization 11, pp. 611-646 (2000).

[44] L. Thogersen, J. Olsen, D. Yeager, P. Jorgensen, P. Salek and T. Helgaker. *The trust-region self-consistent field method: Towards a black-box optimization in Hartree-Fock and Kohn-Sham theories*, Journal of Chemical Physics 121, pp. 16-27 (2004).