ON AUGMENTED LAGRANGIAN METHODS WITH GENERAL LOWER-LEVEL CONSTRAINTS *

R. ANDREANI †, E. G. BIRGIN ‡, J. M. MARTÍNEZ §, AND M. L. SCHUVERDT ¶

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 in which the constraints are only of the lower-level type. 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 a comparison against Ipopt and Lancelot B. 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. The codes are free for download in `www.ime.usp.br/~egbirgin/tango/`

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

1. Introduction.

Many practical optimization problems have the form

$$\text{Minimize } f(x) \text{ subject to } x \in \Omega_1 \cap \Omega_2, \tag{1.1}$$

where the constraint set Ω_2 is such that subproblems of type

$$\text{Minimize } F(x) \text{ subject to } x \in \Omega_2 \tag{1.2}$$

are much easier than problems of type (1.1). By this we mean that there exist efficient algorithms for solving (1.2) that cannot be applied to (1.1). In these cases it is natural to address the resolution of (1.1) by means of procedures that allow one to take advantage of methods that solve (1.2).

Let us mention here a few examples of this situation.

- 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 [37, 55], 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 [61, 64].
- Bilevel problems with "additional" constraints [24]: A basic bilevel problem consists in minimizing $f(x, y)$ with the condition that y solves an optimization problem whose data depend on x. Efficient algorithms for this problem have already been developed (see [24] and references therein). When additional constraints ($h(x, y) = 0, g(x, y) \le 0$) are present the problem is more complicated. Thus, it is attractive to solve these problems using methods that deal with the difficult constraints in an special way and solve iteratively subproblems with the easy constraints.

*FIRST VERSION: OCTOBER 13, 2005

†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

- Minimization with orthogonality constraints [31, 36, 57, 66]: 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 easy set Ω_2.
- Control problems with algebraic constraints: Minimizing an objective function $f(y, u)$ subject to the discretization of $y' = f(y, u)$ is relatively simple using straightforward discrete control methodology. See [47, 52, 53] and references therein. The problem is more difficult if, in addition, it involves algebraic constraints on the control or the state. These constraints are natural candidates to define the set Ω_1, whereas the evolution equation should define Ω_2.
- Problems in which Ω_2 is convex but Ω_1 is not: Sometimes it is possible to take profit of the convexity of Ω_2 in very efficient ways and we do not want to have this structure destroyed by its intersection with Ω_1.

These problems motivated us to revisit Augmented Lagrangian methods with arbitrary lower-level constraints. Penalty and Augmented Lagrangian algorithms seem to be the only methods that can take advantage of the existence of efficient procedures for solving partially constrained subproblems in a natural way. For this reason, many practitioners in Chemistry, Physics, Economy and Engineering rely on empirical penalty approaches when they incorporate additional constraints to models that were satisfactorily solved by pre-existing algorithms.

The general structure of Augmented Lagrangian methods is well known [6, 23, 56]. An Outer Iteration consists of two main steps:

(a) Minimize the Augmented Lagrangian on the appropriate "simple" set (Ω_2 in our case).
(b) Update multipliers and penalty parameters.

However, several decisions need to be taken in order to define a practical algorithm. For example, one should choose a suitable Augmented Lagrangian function. In this paper we use the Powell-Hestenes-Rockafellar PHR definition [45, 58, 62]. So, we pay the prize of having discontinuous second derivatives in the objective function of the subproblems when Ω_1 involves inequalities. We decided to keep inequality constraints as they are, instead of replacing them by equality constraints plus bounds.

Moreover, we need a good criterion for deciding that a suitable approximate subproblem minimizer has been found at Step (a). In particular, one must decide whether subproblem minimizers must be feasible with respect to Ω_2 and which is the admissible level of infeasibility and lack of complementarity at these solutions. Bertsekas [5] analyzed an Augmented Lagrangian method for solving (1.1) in the case in which the subproblems are solved exactly.

Finally, simple and efficient rules for updating multipliers and penalty parameters must be given.

Algorithmic decisions are taken looking at theoretical convergence properties and practical performance. We are essentially interested in practical behavior but, since it is impossible to perform all the possible tests, theoretical results play an important role in algorithmic design. However, only experience tells one which theoretical results have practical importance and which do not. Although we recognize that this point is controversial, we would like to make explicit here our own criteria:

1. External penalty methods have the property that, when one finds the *global* minimizers of the subproblems, every limit point is a global minimizer of the original problem [32]. We think that this property must be preserved by the Augmented Lagrangian counterparts. This is the main reason why, in our algorithm, we will force boundedness of the Lagrange multipliers estimates.
2. We aim feasibility of the limit points but, since this may be impossible (even an empty feasible region is not excluded) a "feasibility result" must say that limit points are stationary points for some infeasibility measure. Some methods require that a constraint qualification holds at all the (feasible or infeasible) iterates. In [15, 70] it was shown that, in such cases, convergence to infeasible points that are not stationary for infeasibility may occur.
3. Feasible limit points must be stationary in some sense. This means that they must be KKT points or that a constraint qualification must fail to hold. The constraint qualification must be as weak as possible (which means that the optimality result must be as strong as possible).

Therefore, under the assumption that all the *feasible* points satisfy the constraint qualification, all the feasible limit points should be KKT.

4. Theoretically, it is impossible to prove that the whole sequence generated by a general Augmented Lagrangian method converges, because multiple solutions of the subproblems may exist and solutions of the subproblems may oscillate. However, since one uses the solution of one subproblem as initial point for solving the following one, the convergence of the whole sequence generally occurs. In this case, under stronger constraint qualifications, nonsingularity conditions and the assumption that the true Lagrange multipliers satisfy the bounds given in the definition of the algorithm, we must be able to prove that the penalty parameters remain bounded.

In other words, the method must have all the good global convergence properties of the External Penalty method. In addition, when everything "goes well", it must be free of the asymptotic instability caused by large penalty parameters. It is important to emphasize that we deal with nonconvex problems, therefore the possibility of obtaining full global convergence properties based on proximal-point arguments is out of question.

The algorithm presented in this paper satisfies those theoretical requirements. In particular, we will show that, if a feasible limit point satisfies the Constant Positive Linear Dependence (CPLD) condition, then it is a KKT point. The CPLD condition was introduced by Qi and Wei [59]. In [3] it was proved that CPLD is a constraint qualification, being weaker than the Linear Independence Constraint Qualification (LICQ) and than the Mangasarian-Fromovitz condition (MFCQ). A feasible point x of a nonlinear programming problem is said to satisfy CPLD if the existence of a nontrivial null linear combination of gradients of active constraints with nonnegative coefficients corresponding to the inequalities implies that the gradients involved in that combination are linearly dependent for all z in a neighborhood of x. Since CPLD is weaker than (say) LICQ, theoretical results saying that *if a limit point satisfies CPLD then it satisfies KKT* are stronger than theoretical results saying that *if a limit point satisfies LICQ then it satisfies KKT*.

These theoretical results indicate what should be observed in practice. Namely:

1. Although the solutions of subproblems are not guaranteed to be close to global minimizers, the algorithm should exhibit a stronger tendency to converge to global minimizers than algorithms based on sequential quadratic programming.
2. The algorithm should find feasible points but, if it does not, it must find "putative minimizers" of the infeasibility.
3. When the algorithm converges to feasible points, these points must be approximate KKT points in the sense of [59]. The case of bounded Lagrange multipliers approximations corresponds to the case in which the limit is KKT, whereas the case of very large Lagrange multiplier approximations announces limit points that do not satisfy CPLD.
4. Cases in which practical convergence occurs in a small number of iterations should coincide with the cases in which the penalty parameters are bounded.

Our plan is to prove the convergence results and to show that, in practice, the method behaves as expected. We will analyze two versions of the main algorithm: with only one penalty parameter and with one penalty parameter per constraint. For proving boundedness of the sequence of penalty parameters we use the reduction to the equality-constraint case introduced in [5].

Most practical nonlinear programming methods published after 2001 rely on sequential quadratic programming (SQP), Newton-like or barrier approaches [1, 4, 14, 16, 19, 18, 34, 35, 38, 39, 50, 54, 65, 68, 69, 71, 72, 73]. None of these methods can be easily adapted to the situation described by (1.1)-(1.2). We selected Ipopt, an algorithm by Wächter and Biegler available in the web [71] for our numerical comparisons. In addition to Ipopt we performed numerical comparisons against Lancelot B [22].

The numerical experiments aim three following objectives:

1. We will show that, in some very large scale location problems, to use a specific algorithm for convex-constrained programming [10, 11, 12, 25] for solving the subproblems in the Augmented Lagrangian context is much more efficient than using general purpose methods like Ipopt and Lancelot B.
2. Algencan is the particular implementation of the algorithm introduced in this paper for the

case in which the lower-level set is a box. For solving the subproblems it uses the code GENCAN [8]. We will show that ALGENCAN tends to converge to global minimizers more often than IPOPT.

3. We will show that, for problems with many inequality constraints, ALGENCAN is more efficient than IPOPT and LANCELOT B. Something similar happens with respect to IPOPT in problems in which the Hessian of the Lagrangian has a "not nicely sparse" factorization.

Finally, we will compare ALGENCAN with LANCELOT B and IPOPT using *all* the problems of the CUTEr collection [13].

This paper is organized as follows. A high-level description of the main algorithm is given in Section 2. The rigorous definition of the method is in Section 3. Section 4 is devoted to global convergence results. In Section 5 we prove boundedness of the penalty parameters. In Section 6 we show the numerical experiments. Applications, conclusions and lines for future research are discussed in Section 7.

Notation.

We denote:

$$\mathbb{R}_+ = \{t \in \mathbb{R} \mid t \geq 0\},$$

$$\mathbb{R}_{++} = \{t \in \mathbb{R} \mid t > 0\},$$

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

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

$[v]_i$ 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 $y \in \mathbb{R}^n$, $y_+ = (\max\{0, y_1\}, \ldots, \max\{0, y_n\})$.

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

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

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

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

2. Overview of the method. We will consider the following nonlinear programming problem:

$$\text{Minimize } f(x) \text{ subject to } h_1(x) = 0, g_1(x) \leq 0, h_2(x) = 0, g_2(x) \leq 0, \tag{2.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}, h_1 : \mathbb{R}^n \to \mathbb{R}^{m_1}, h_2 : \mathbb{R}^n \to \mathbb{R}^{m_2}, g_1 : \mathbb{R}^n \to \mathbb{R}^{p_1}, g_2 : \mathbb{R}^n \to \mathbb{R}^{p_2}$. We assume that all these functions admit continuous first derivatives on a sufficiently large and open domain. We define $\Omega_1 = \{x \in \mathbb{R}^n \mid h_1(x) = 0, g_1(x) \leq 0\}$ and $\Omega_2 = \{x \in \mathbb{R}^n \mid h_2(x) = 0, g_2(x) \leq 0\}$.

For all $x \in \mathbb{R}^n, \rho > 0, \lambda \in \mathbb{R}^{m_1}, \mu \in \mathbb{R}^{p_1}_+$ we define the Augmented Lagrangian with respect to Ω_1 [45, 58, 62] as:

$$L(x, \lambda, \mu, \rho) = f(x) + \frac{\rho}{2} \sum_{i=1}^{m_1} \left([h_1(x)]_i + \frac{\lambda_i}{\rho} \right)^2 + \frac{\rho}{2} \sum_{i=1}^{p_1} \left([g_1(x)]_i + \frac{\mu_i}{\rho} \right)_+^2. \tag{2.2}$$

The main algorithm defined in this paper will consist of a sequence of (approximate) minimizations of $L(x, \lambda, \mu, \rho)$ subject to $x \in \Omega_2$, followed by the updating of λ, μ and ρ. A version of the algorithm with several penalty parameters may be found in [**?**]. Each approximate minimization of L will be called an *Outer Iteration*.

After each Outer Iteration one wishes some progress in terms of *feasibility* and *complementarity*. The *infeasibility* of x with respect to the equality constraint $[h_1(x)]_i = 0$ is naturally represented by $|[h_1(x)]_i|$.

The case of inequality constraints is more complicate because, besides feasibility, one expects to have a null multiplier estimate if $g_i(x) < 0$. A suitable combined measure of infeasibility and non-complementarity with respect to the constraint $[g_1(x)]_i \le 0$ comes from defining $[\sigma(x,\mu,\rho)]_i = \max\{[g_1(x)]_i, -\mu_i/\rho\}$. Since μ_i/ρ is always nonnegative, it turns out that $[\sigma(x,\mu,\rho)]_i$ vanishes in two situations: (a) when $[g_1(x)]_i = 0$; and (b) when $[g_1(x)]_i < 0$ and $\mu_i = 0$. So, roughly speaking, $|[\sigma(x,\mu,\rho)]_i|$ measures infeasibility and complementarity with respect to the inequality constraint $[g_1(x)]_i \le 0$. If, between two consecutive outer iterations, enough progress is observed in terms of (at least one of) feasibility and complementarity, the penalty parameter will not be updated. Otherwise, the penalty parameter is increased by a fixed factor.

The rules for updating the multipliers need some discussion. In principle, we adopt the classical first-order correction rule [45, 58, 63] but, in addition, we impose that the multiplier estimates must be bounded. So, we will explicitly project the estimates on a compact box after each update. The reason for this decision was already given in the introduction: we want to preserve the property of external penalty methods that global minimizers of the original problem are obtained if each outer iteration computes a global minimizer of the subproblem. This property is maintained if the quotient of *the square* of each multiplier estimate over the penalty parameter tends to zero when the penalty parameter tends to infinity. We were not able to prove that this condition holds automatically for usual estimates and, in fact, we conjecture that it does not. Therefore, we decided to force the boundedness condition. The price paid by this decision seems to be moderate: in the proof of the boundedness of penalty parameters we will need to assume that the true Lagrange multipliers are within the bounds imposed by the algorithm. Since "large Lagrange multipliers" is a symptom of "near-nonfulfillment" of the Mangasarian-Fromovitz constraint qualification, this assumption seems to be compatible with the remaining ones that are necessary to prove penalty boundedness.

3. Description of the Augmented Lagrangian algorithm. In this section we provide a detailed description of the main algorithm. Approximate solutions of the subproblems are defined as points that satisfy the conditions (3.1)–(3.4) below. These formulae are relaxed KKT conditions of the problem of minimizing L subject to $x \in \Omega_2$. The first-order approximations of the multipliers are computed at Step 3. Lagrange multipliers estimates are denoted λ_k and μ_k whereas their safeguarded counterparts are $\bar\lambda_k$ and $\bar\mu_k$. At Step 4 we update the penalty parameters according to the progress in terms of feasibility and complementarity.

Algorithm 3.1.

Let $x_0 \in I\!R^n$ be an arbitrary initial point. The given parameters for the execution of the algorithm are: $\tau \in [0,1), \gamma > 1, \rho_1 > 0, -\infty < [\bar\lambda_{\min}]_i \le [\bar\lambda_{\max}]_i < \infty \;\forall\, i = 1,\ldots,m_1,\; 0 \le [\bar\mu_{\max}]_i < \infty \;\forall\, i = 1,\ldots,p_1$, $[\bar\lambda_1]_i \in [[\bar\lambda_{\min}]_i, [\bar\lambda_{\max}]_i] \;\forall\, i = 1,\ldots,m_1$, $[\bar\mu_1]_i \in [0, [\bar\mu_{\max}]_i] \;\forall\, i = 1,\ldots,p_1$. Finally, $\{\varepsilon_k\} \subset I\!R_+$ is a sequence of tolerance parameters such that $\lim_{k\to\infty} \varepsilon_k = 0$.

Step 1. *Initialization*

Set $k \leftarrow 1$. For $i = 1,\ldots,p_1$, compute $[\sigma_0]_i = \max\{0, [g_1(x_0)]_i\}$.

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

$$\|\nabla L(x_k, \bar\lambda_k, \bar\mu_k, \rho_k) + \sum_{i=1}^{m_2} [v_k]_i \nabla [h_2(x_k)]_i + \sum_{i=1}^{p_2} [u_k]_i \nabla [g_2(x_k)]_i\| \le \varepsilon_{k,1}, \tag{3.1}$$

$$[u_k]_i \ge 0 \text{ and } [g_2(x_k)]_i \le \varepsilon_{k,2} \text{ for all } i = 1,\ldots,p_2, \tag{3.2}$$

$$[g_2(x_k)]_i < -\varepsilon_{k,2} \Rightarrow [u_k]_i = 0 \text{ for all } i = 1,\ldots,p_2, \tag{3.3}$$

$$\|h_2(x_k)\| \le \varepsilon_{k,3}, \tag{3.4}$$

where $\varepsilon_{k,1}, \varepsilon_{k,2}, \varepsilon_{k,3} \ge 0$ are such that $\max\{\varepsilon_{k,1}, \varepsilon_{k,2}, \varepsilon_{k,3}\} \le \varepsilon_k$. If it is not possible to find x_k satisfying (3.1)–(3.4), stop the execution of the algorithm.

Step 3. *Estimate multipliers*

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

$$[\lambda_{k+1}]_i = [\bar\lambda_k]_i + \rho_k [h_1(x_k)]_i, \tag{3.5}$$

$$(3.6)\qquad [\bar\lambda_{k+1}]_i \in [[\bar\lambda_{\min}]_i, [\bar\lambda_{\max}]_i].$$

(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 $i = 1, \ldots, p_1$, compute

$$(3.7)\qquad [\mu_{k+1}]_i = \max\{0, [\bar\mu_k]_i + \rho_k [g_1(x_k)]_i\}, \quad [\sigma_k]_i = \max\left\{[g_1(x_k)]_i, -\frac{[\bar\mu_k]_i}{\rho_k}\right\},$$

$$[\bar\mu_{k+1}]_i \in [0, [\bar\mu_{\max}]_i].$$

(Usually, $[\bar\mu_{k+1}]_i = \min\{[\mu_{k+1}]_i, [\bar\mu_{\max}]_i\}$.)
Step 4. *Update the penalty parameter*
If $\max\{\|h_1(x_k)\|_\infty, \|\sigma_k\|_\infty\} \le \tau \max\{\|h_1(x_{k-1})\|_\infty, \|\sigma_{k-1}\|_\infty\}$, then define $\rho_{k+1} = \rho_k$. Else, define $\rho_{k+1} = \gamma\rho_k$.
Step 5. *Begin a new outer iteration*
Set $k \leftarrow k+1$. Go to Step 2.

4. 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 (3.1)-(3.4). Problem-dependent sufficient conditions for this assumption can be given in many cases.

We will also assume that at least a limit point of the sequence generated by Algorithm 3.1 exists. A sufficient condition for this is the existence of $\varepsilon > 0$ such that the set $\{x \in I\!R^n \mid g_2(x) \le \varepsilon, \|h_2(x)\| \le \varepsilon\}$ is bounded. This condition may be enforced adding artificial simple constraints to the set Ω_2.

Global convergence results that use the CPLD constraint qualification are stronger than previous results for more specific problems: In particular, Conn, Gould and Toint [22] and Conn, Gould, Sartenaer and Toint [20] proved global convergence of Augmented Lagrangian methods for equality constraints and linear constraints assuming linear independence of all the gradients of active constraints at the limit points. Their assumption is much stronger than our CPLD assumptions. On one hand, the CPLD assumption is weaker than LICQ (for example, CPLD always holds when the constraints are linear). On the other hand, our CPLD assumption involves only feasible points instead of all possible limit points of the algorithm.

Convergence proofs for Augmented Lagrangian methods with equalities and box constraints using CPLD were given in [2].

We are going to investigate the status of the limit points of sequences generated by Algorithm 3.1. Firstly, we will prove a result on the feasibility properties of a limit point. Theorem 4.1 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 the sum of squares of upper-level infeasibilities.

THEOREM 4.1. *Let $\{x_k\}$ be a sequence generated by Algorithm 3.1. 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) x_* is a KKT point of the problem

$$(4.1)\qquad \text{Minimize } \frac{1}{2}\left[\sum_{i=1}^{m_1} [h_1(x)]_i^2 + \sum_{i=1}^{p_1} \max\{0, [g_1(x)]_i\}^2\right] \text{ subject to } x \in \Omega_2.$$

(ii) x_* does not satisfy the CPLD constraint qualification associated with Ω_2.

Proof. Let K be an infinite subsequence in $I\!N$ such that $\lim_{k\in K} x_k = x_*$. Since $\varepsilon_k \to 0$, by (3.2) and (3.4), we have that $g_2(x_*) \le 0$ and $h_2(x_*) = 0$. So, $x_* \in \Omega_2$.

Now, we consider two possibilities: (a) the sequence $\{\rho_k\}$ is bounded; and (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, $\lim_{k\to\infty} \|h_1(x_k)\| = \lim_{k\to\infty} \|\sigma_k\| = 0$. Thus, $h_1(x_*) = 0$. Now, if $[g_1(x_*)]_j > 0$ then $[g_1(x_k)]_j > c > 0$ for $k \in K$ large enough. This would contradict the fact that $[\sigma_k]_j \to 0$. Therefore, $[g_1(x_*)]_i \le 0 \quad \forall i = 1, \ldots, p_1$.

Since $x_* \in \Omega_2$, $h_1(x_*) = 0$ and $g_1(x_*) \le 0$, x_* is feasible. Therefore, we proved the desired result in the case that $\{\rho_k\}$ is bounded.

Consider now Case (b). So, $\{\rho_k\}_{k\in K}$ is not bounded. By (2.2) and (3.1), we have:

$$\begin{aligned}&\nabla f(x_k) + \textstyle\sum_{i=1}^{m_1}([\bar\lambda_k]_i + \rho_k[h_1(x_k)]_i)\nabla[h_1(x_k)]_i + \sum_{i=1}^{p_1}\max\{0,[\bar\mu_k]_i \\ &+\rho_k[g_1(x_k)]_i\}\nabla[g_1(x_k)]_i + \textstyle\sum_{i=1}^{m_2}[v_k]_i\nabla[h_2(x_k)]_i + \sum_{j=1}^{p_2}[u_k]_j\nabla[g_2(x_k)]_j = \delta_k,\end{aligned}\tag{4.2}$$

where, since $\varepsilon_k \to 0$, $\lim_{k\in K}\|\delta_k\| = 0$.

If $[g_2(x_*)]_i < 0$, there exists $k_1 \in I\!N$ such that $[g_2(x_k)]_i < -\varepsilon_k$ for all $k \ge k_1, k \in K$. Therefore, by (3.3), $[u_k]_i = 0$ for all $k \in K, k \ge k_1$. Thus, by $x_* \in \Omega_2$ and (4.2), for all $k \in K, k \ge k_1$ we have that

$$\begin{aligned}&\nabla f(x_k) + \textstyle\sum_{i=1}^{m_1}([\bar\lambda_k]_i + \rho_k[h_1(x_k)]_i)\nabla[h_1(x_k)]_i + \sum_{i=1}^{p_1}\max\{0,[\bar\mu_k]_i \\ &+\rho_k[g_1(x_k)]_i\}\nabla[g_1(x_k)]_i + \textstyle\sum_{i=1}^{m_2}[v_k]_i\nabla[h_2(x_k)]_i + \sum_{[g_2(x_*)]_j=0}[u_k]_j\nabla[g_2(x_k)]_j = \delta_k.\end{aligned}$$

Dividing by ρ_k we get:

$$\begin{aligned}&\tfrac{\nabla f(x_k)}{\rho_k} + \textstyle\sum_{i=1}^{m_1}\left(\tfrac{[\bar\lambda_k]_i}{\rho_k} + [h_1(x_k)]_i\right)\nabla[h_1(x_k)]_i + \sum_{i=1}^{p_1}\max\left\{0,\tfrac{[\bar\mu_k]_i}{\rho_k} + [g_1(x_k)]_i\right\}\nabla[g_1(x_k)]_i \\ &+\textstyle\sum_{i=1}^{m_2}\tfrac{[v_k]_i}{\rho_k}\nabla[h_2(x_k)]_i + \sum_{[g_2(x_*)]_j=0}\tfrac{[u_k]_j}{\rho_k}\nabla[g_2(x_k)]_j = \tfrac{\delta_k}{\rho_k}.\end{aligned}$$

By Caratheodory's Theorem of Cones (see [6], page 689) there exist $\widehat{I}_k \subset \{1,\dots,m_2\}$, $\widehat{J}_k \subset \{j \mid [g_2(x_*)]_j = 0\}$, $[\widehat{v}_k]_i,\ \ i \in \widehat{I}_k$ and $[\widehat{u}_k]_j \ge 0,\ \ j \in \widehat{J}_k$ 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

$$\begin{aligned}&\tfrac{\nabla f(x_k)}{\rho_k} + \textstyle\sum_{i=1}^{m_1}\left(\tfrac{[\bar\lambda_k]_i}{\rho_k} + [h_1(x_k)]_i\right)\nabla[h_1(x_k)]_i + \sum_{i=1}^{p_1}\max\left\{0,\tfrac{[\bar\mu_k]_i}{\rho_k} + [g_1(x_k)]_i\right\}\nabla[g_1(x_k)]_i \\ &+\textstyle\sum_{i\in\widehat{I}_k}[\widehat{v}_k]_i\nabla[h_2(x_k)]_i + \sum_{j\in\widehat{J}_k}[\widehat{u}_k]_j\nabla[g_2(x_k)]_j = \tfrac{\delta_k}{\rho_k}.\end{aligned}\tag{4.3}$$

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

$$\widehat{J} = \widehat{J}_k \subset \{j \mid [g_2(x_*)]_j = 0\}\tag{4.4}$$

for all $k \in K_1$. Then, by (4.3), for all $k \in K_1$ we have:

$$\begin{aligned}&\tfrac{\nabla f(x_k)}{\rho_k} + \textstyle\sum_{i=1}^{m_1}\left(\tfrac{[\bar\lambda_k]_i}{\rho_k} + [h_1(x_k)]_i\right)\nabla[h_1(x_k)]_i + \sum_{i=1}^{p_1}\max\left\{0,\tfrac{[\bar\mu_k]_i}{\rho_k} + [g_1(x_k)]_i\right\}\nabla[g_1(x_k)]_i \\ &+\textstyle\sum_{i\in\widehat{I}}[\widehat{v}_k]_i\nabla[h_2(x_k)]_i + \sum_{j\in\widehat{J}}[\widehat{u}_k]_j\nabla[g_2(x_k)]_j = \tfrac{\delta_k}{\rho_k},\end{aligned}\tag{4.5}$$

and the gradients

$$\{\nabla[h_2(x_k)]_i\}_{i\in\widehat{I}} \cup \{\nabla[g_2(x_k)]_j\}_{j\in\widehat{J}} \text{ are linearly independent.}\tag{4.6}$$

We consider, again, two cases: (1) the sequence $\{\|(\widehat{v}_k,\widehat{u}_k)\|, k \in K_1\}$ is bounded; and (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, and $\widehat{I} \cup \widehat{J} \ne \emptyset$, there exist $(\widehat{v},\widehat{u}), \widehat{u} \ge 0$ and an infinite set of indices $K_2 \subset K_1$ such that $\lim_{k\in K_2}(\widehat{v}_k,\widehat{u}_k) = (\widehat{v},\widehat{u})$. Since $\{\rho_k\}$ is unbounded, by the boundedness of $\bar\lambda_k$ and $\bar\mu_k$, $\lim[\bar\lambda_k]_i/\rho_k = 0 = \lim[\bar\mu_k]_j/\rho_k$ for all i,j. Therefore, by $\delta_k \to 0$, taking limits for $k \in K_2$ in (4.5), we obtain:

$$\begin{aligned}&\textstyle\sum_{i=1}^{m_1}[h_1(x_*)]_i\nabla[h_1(x_*)]_i + \sum_{i=1}^{p_1}\max\{0,[g_1(x_*)]_i\}\nabla[g_1(x_*)]_i \\ &+\textstyle\sum_{i\in\widehat{I}}\widehat{v}_i\nabla[h_2(x_*)]_i + \sum_{j\in\widehat{J}}\widehat{u}_j\nabla[g_2(x_*)]_j = 0.\end{aligned}\tag{4.7}$$

If $\widehat{I} \cup \widehat{J} = \emptyset$ we obtain $\sum_{i=1}^{m_1} [h_1(x_*)]_i \nabla [h_1(x_*)]_i + \sum_{i=1}^{p_1} \max\{0, [g_1(x_*)]_i\} \nabla [g_1(x_*)]_i = 0$.

Therefore, by $x_* \in \Omega_2$ and (4.4), x_* is a KKT point of (4.1).

Finally, 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 (4.5) by $\|(\widehat{v}_k, \widehat{u}_k)\|$ and taking limits for $k \in K_3$, we deduce that $\sum_{i \in \widehat{I}} \widehat{v}_i \nabla [h_2(x_*)]_i + \sum_{j \in \widehat{J}} \widehat{u}_j \nabla [g_2(x_*)]_j = 0$. But $[g_2(x_*)]_j = 0$ for all $j \in \widehat{J}$. Then, by (4.6), x_* does not satisfy the CPLD constraint qualification associated with the set Ω_2. This completes the proof. □

Roughly speaking, Theorem 4.1 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 4.2 we will prove that, under the CPLD constraint qualification, feasible limit points are stationary (KKT) points of the original problem. 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 4.2. *Let $\{x_k\}_{k \in I\!N}$ be a sequence generated by Algorithm 3.1. 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 (2.1). Moreover, if x_* satisfies the Mangasarian-Fromovitz constraint qualification and $\{x_k\}_{k \in K}$ is a subsequence that converges to x_*, the set*

$$\{\|\lambda_{k+1}\|, \|\mu_{k+1}\|, \|v_k\|, \|u_k\|\}_{k \in K} \text{ is bounded.} \tag{4.8}$$

Proof. For all $k \in I\!N$, by (3.1), (3.3), (3.5) and (3.7), there exist $u_k \in I\!R_+^{p_2}$, $\delta_k \in I\!R^n$ such that $\|\delta_k\| \leq \varepsilon_k$ and

$$\begin{aligned} &\nabla f(x_k) + \textstyle\sum_{i=1}^{m_1} [\lambda_{k+1}]_i \nabla [h_1(x_k)]_i + \sum_{i=1}^{p_1} [\mu_{k+1}]_i \nabla [g_1(x_k)]_i \\ &\quad + \textstyle\sum_{i=1}^{m_2} [v_k]_i \nabla [h_2(x_k)]_i + \sum_{j=1}^{p_2} [u_k]_j \nabla [g_2(x_k)]_j = \delta_k. \end{aligned} \tag{4.9}$$

By (3.7), $\mu_{k+1} \in I\!R_+^{p_1}$ 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 (3.3), $[u_k]_i = 0 \quad \forall k \in K, k \geq k_1$.

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

We consider two cases: (1) $\{\rho_k\}$ is unbounded; and (2) $\{\rho_k\}$ is bounded. In the first case we have that $\lim_{k \in K} \rho_k = \infty$. Since $\{[\bar{\mu}_k]_i\}$ is bounded, there exists $k_3 \geq k_2$ such that, for all $k \in K, k \geq k_3$, $[\bar{\mu}_k]_i + \rho_k [g_1(x_k)]_i < 0$. 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 $\{\rho_k\}$ is bounded. In this case, $\lim_{k \to \infty} [\sigma_k]_i = 0$. Therefore, since $[g_1(x_k)]_i < c < 0$ for $k \in K$ large enough, $\lim_{k \in K} [\bar{\mu}_k]_i = 0$. So, for $k \in K$ large enough, $[\bar{\mu}_k]_i + \rho_k [g_1(x_k)]_i < 0$. By the definition of μ_{k+1}, there exists $k_4 \geq k_2$ such that $[\mu_{k+1}]_i = 0$ for $k \in K, k \geq k_4$.

Therefore, there exists $k_5 \geq \max\{k_1, k_3, k_4\}$ such that for all $k \in K$, $k \geq k_5$,

$$[[g_1(x_*)]_i < 0 \Rightarrow [\mu_{k+1}]_i = 0] \text{ and } [[g_2(x_*)]_i < 0 \Rightarrow [u_k]_i = 0]. \tag{4.10}$$

(Observe that, up to now, we did not use the CPLD condition.) By (4.9) and (4.10), for all $k \in K, k \geq k_5$, we have:

$$\begin{aligned} &\nabla f(x_k) + \textstyle\sum_{i=1}^{m_1} [\lambda_{k+1}]_i \nabla [h_1(x_k)]_i + \sum_{[g_1(x_*)]_i = 0} [\mu_{k+1}]_i \nabla [g_1(x_k)]_i \\ &\quad + \textstyle\sum_{i=1}^{m_2} [v_k]_i \nabla [h_2(x_k)]_i + \sum_{[g_2(x_*)]_j = 0} [u_k]_j \nabla [g_2(x_k)]_j = \delta_k, \end{aligned} \tag{4.11}$$

with $\mu_{k+1} \in I\!R_+^{p_1}, u_k \in I\!R_+^{p_2}$.

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

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

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

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

$$\begin{aligned} &\nabla f(x_k) + \textstyle\sum_{i \in \widehat{I}_k} [\widehat{\lambda}_k]_i \nabla [h_1(x_k)]_i + \sum_{i \in \widehat{J}_k} [\widehat{\mu}_k]_i \nabla [g_1(x_k)]_i \\ &+ \textstyle\sum_{i \in \hat{I}_k} [\widehat{v}_k]_i \nabla [h_2(x_k)]_i + \sum_{j \in \hat{J}_k} [\widehat{u}_k]_j \nabla [g_2(x_k)]_j = \delta_k. \end{aligned} \tag{4.12}$$

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 (4.12),

$$\begin{aligned} &\nabla f(x_k) + \textstyle\sum_{i \in \widehat{I}} [\widehat{\lambda}_k]_i \nabla [h_1(x_k)]_i + \sum_{i \in \widehat{J}} [\widehat{\mu}_k]_i \nabla [g_1(x_k)]_i \\ &+ \textstyle\sum_{i \in \hat{I}} [\widehat{v}_k]_i \nabla [h_2(x_k)]_i + \sum_{j \in \hat{J}} [\widehat{u}_k]_j \nabla [g_2(x_k)]_j = \delta_k \end{aligned} \tag{4.13}$$

and the vectors

$$\{\nabla [h_1(x_k)]_i\}_{i \in \widehat{I}} \cup \{\nabla [g_1(x_k)]_i\}_{i \in \widehat{J}} \cup \{\nabla [h_2(x_k)]_i\}_{i \in \hat{I}} \cup \{\nabla [g_2(x_k)]_i\}_{i \in \hat{J}} \tag{4.14}$$

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

If $\widehat{I} \cup \widehat{J} \cup \hat{I} \cup \hat{J} = \emptyset$, by (4.13) and $\delta_k \to 0$ we obtain $\nabla f(x_*) = 0$. Otherwise, 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; and (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 $\lim_{k \in K_2} [\widehat{\lambda}_k]_i = \widehat{\lambda}_i$, $\lim_{k \in K_2} [\widehat{\mu}_k]_i = \widehat{\mu}_i \geq 0$, $\lim_{k \in K_2} [\widehat{v}_k]_i = \widehat{v}_i$, and $\lim_{k \in K_2} [\widehat{u}_k]_i = \widehat{u}_i \geq 0$. By $\varepsilon_k \to 0$ and $\|\delta_k\| \leq \varepsilon_k$, taking limits in (4.13) 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 (2.1).

Suppose now that $\lim_{k \in K_2} S_k = \infty$. Dividing both sides of (4.13) by S_k we obtain:

$$\begin{aligned} &\tfrac{\nabla f(x_k)}{S_k} + \textstyle\sum_{i \in \widehat{I}} \frac{[\widehat{\lambda}_k]_i}{S_k} \nabla [h_1(x_k)]_i + \sum_{i \in \widehat{J}} \frac{[\widehat{\mu}_k]_i}{S_k} \nabla [g_1(x_k)]_i \\ &+ \textstyle\sum_{i \in \hat{I}} \frac{[\widehat{v}_k]_i}{S_k} \nabla [h_2(x_k)]_i + \sum_{j \in \hat{J}} \frac{[\widehat{u}_k]_j}{S_k} \nabla [g_2(x_k)]_j = \frac{\delta_k}{S_k}, \end{aligned} \tag{4.15}$$

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 (4.15) 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, 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 (4.14). Therefore, the main part of the theorem is proved.

Finally, let us prove that the property (4.8) 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 (4.8) is not true, we have that $\lim_{k\in K} B_k = \infty$. In this case, dividing both sides of (4.11) by B_k and taking limits for an appropriate subsequence, we obtain that x_* does not satisfy the Mangasarian-Fromovitz constraint qualification. □

5. Boundedness of the penalty parameters. When the penalty parameters associated with Penalty or Augmented Lagrangian methods are too large, the subproblems tend to be ill-conditioned and its resolution becomes harder. 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.

5.1. Equality constraints. We will consider first the case $p_1 = p_2 = 0$.

Let $f : I\!R^n \to I\!R, h_1 : I\!R^n \to I\!R^{m_1}, h_2 : I\!R^n \to I\!R^{m_2}$. We address the problem

$$\text{Minimize } f(x) \text{ subject to } h_1(x) = 0,\ h_2(x) = 0. \tag{5.1}$$

The Lagrangian function associated with problem (5.1) is given by $L_0(x, \lambda, v) = f(x) + \langle h_1(x), \lambda\rangle + \langle h_2(x), v\rangle$, for all $x \in I\!R^n$, $\lambda \in I\!R^{m_1}$, $v \in I\!R^{m_2}$.

Algorithm 3.1 will be considered with the following standard definition for the safeguarded Lagrange multipliers.

Definition. *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 $[[\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i]$.*

We will use the following assumptions:

Assumption 1. *The sequence $\{x_k\}$ is generated by the application of Algorithm 3.1 to problem (5.1) and $\lim_{k\to\infty} x_k = x_*$.*

Assumption 2. *The point x_* is feasible ($h_1(x_*) = 0$ and $h_2(x_*) = 0$).*

Assumption 3. *The gradients $\nabla[h_1(x_*)]_1, \ldots, \nabla[h_1(x_*)]_{m_1}, \nabla[h_2(x_*)]_1, \ldots, \nabla[h_2(x_*)]_{m_2}$ are linearly independent.*

Assumption 4. *The functions f, h_1 and h_2 admit continuous second derivatives in a neighborhood of x_*.*

Assumption 5. *The second order sufficient condition for local minimizers ([33], page 211) holds with Lagrange multipliers $\lambda_* \in I\!R^{m_1}$ and $v_* \in I\!R^{m_2}$.*

Assumption 6. *For all $i = 1, \ldots, m_1$, $[\lambda_*]_i \in ([\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i)$.*

PROPOSITION 5.1. *Suppose that Assumptions 1, 2, 3 and 6 hold. Then, $\lim_{k\to\infty} \lambda_k = \lambda_*$, $\lim_{k\to\infty} v_k = v_*$ and $\bar{\lambda}_k = \lambda_k$ for k large enough.*

Proof. The proof of the first part follows from the definition of λ_{k+1}, the stopping criterion of the subproblems and the linear independence of the gradients of the constraints at x_*. The second part of the thesis is a consequence of $\lambda_k \to \lambda_*$, using Assumption 6 and the definition of $\bar{\lambda}_{k+1}$. □

LEMMA 5.2. *Suppose that Assumptions 3 and 5 hold. Then, there exists $\bar{\rho} > 0$ such that, for all $\pi \in [0, 1/\bar{\rho}]$, the matrix*

$$\begin{pmatrix} \nabla^2_{xx} L_0(x_*, \lambda_*, v_*) & \nabla h_1(x_*) & \nabla h_2(x_*) \\ \nabla h_1(x_*)^T & -\pi I & 0 \\ \nabla h_2(x_*)^T & 0 & 0 \end{pmatrix}$$

is nonsingular.

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

LEMMA 5.3. *Suppose that Assumptions 1–5 hold. Let $\bar\rho$ be as in Lemma 5.2. Suppose that there exists $k_0 \in I\!N$ such that $\rho_k \geq \bar\rho$ for all $k \geq k_0$. Define*

$$\alpha_k = \nabla L(x_k, \bar\lambda_k, \rho_k) + \nabla h_2(x_k) v_k, \tag{5.2}$$

$$\beta_k = h_2(x_k). \tag{5.3}$$

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

$$\|x_k - x_*\| \leq M \max\left\{\frac{\|\bar\lambda_k - \lambda_*\|_\infty}{\rho_k}, \|\alpha_k\|, \|\beta_k\|\right\}, \tag{5.4}$$

$$\|\lambda_{k+1} - \lambda_*\| \leq M \max\left\{\frac{\|\bar\lambda_k - \lambda_*\|_\infty}{\rho_k}, \|\alpha_k\|, \|\beta_k\|\right\}. \tag{5.5}$$

Proof. Define, for all $k \in I\!N$,

$$t_k = (\bar\lambda_k - \lambda_*)/\rho_k, \tag{5.6}$$

$$\pi_k = 1/\rho_k. \tag{5.7}$$

By (3.5), (5.2) and (5.3), $\nabla L(x_k, \bar\lambda_k, \rho_k) + \nabla h_2(x_k) v_k - \alpha_k = 0$, $\lambda_{k+1} = \bar\lambda_k + \rho_k h_1(x_k)$ and $h_2(x_k) - \beta_k = 0$ for all $k \in I\!N$.

Therefore, by (5.6) and (5.7), we have that $\nabla f(x_k) + \nabla h_1(x_k)\lambda_{k+1} + \nabla h_2(x_k) v_k - \alpha_k = 0$, $h_1(x_k) - \pi_k \lambda_{k+1} + t_k + \pi_k \lambda_* = 0$ and $h_2(x_k) - \beta_k = 0$ for all $k \in I\!N$. Define, for all $\pi \in [0, 1/\bar\rho]$, $F_\pi : I\!R^n \times I\!R^{m_1} \times I\!R^{m_2} \times I\!R^{m_1} \times I\!R^n \times I\!R^{m_2} \to I\!R^n \times I\!R^{m_1} \times I\!R^{m_2}$ by

$$F_\pi(x, \lambda, v, t, \alpha, \beta) = \begin{pmatrix} \nabla f(x) + \nabla h_1(x)\lambda + \nabla h_2(x) v - \alpha \\ [h_1(x)]_1 - \pi[\lambda]_1 + [t]_1 + \pi[\lambda_*]_1 \\ \cdot \\ \cdot \\ \cdot \\ [h_1(x)]_{m_1} - \pi[\lambda]_{m_1} + [t]_{m_1} + \pi[\lambda_*]_{m_1} \\ h_2(x) - \beta \end{pmatrix}.$$

Clearly,

$$F_{\pi_k}(x_k, \lambda_{k+1}, v_k, t_k, \alpha_k, \beta_k) = 0 \tag{5.8}$$

and, by Assumptions 1 and 2,

$$F_\pi(x_*, \lambda_*, v_*, 0, 0, 0) = 0 \ \ \forall \pi \in [0, 1/\bar\rho]. \tag{5.9}$$

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

$$\begin{pmatrix} \nabla^2_{xx} L_0(x_*, \lambda_*, v_*) & \nabla h_1(x_*) & \nabla h_2(x_*) \\ \nabla h_1(x_*)^T & -\pi I & 0 \\ \nabla h_2(x_*)^T & 0 & 0 \end{pmatrix}.$$

By Lemma 5.2, this matrix is nonsingular for all $\pi \in [0, 1/\bar\rho]$. By continuity, the norm of its inverse is bounded in a neighborhood of $(x_*, \lambda_*, v_*, 0, 0, 0)$ uniformly with respect to $\pi \in [0, 1/\bar\rho]$. 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 $\pi \in [0, 1/\bar{\rho}]$. Therefore, the bounds (5.4) and (5.5) follow from (5.8) and (5.9) by the Implicit Function Theorem and the Mean Value Theorem of Integral Calculus. ∎

THEOREM 5.4. *Suppose that Assumptions 1–6 are satisfied by the sequence generated by Algorithm 3.1 applied to the problem (5.1). In addition, assume that there exists a sequence $\eta_k \to 0$ such that $\varepsilon_k \leq \eta_k \|h_1(x_k)\|_\infty$ for all $k \in I\!N$. Then, the sequence of penalty parameters $\{\rho_k\}$ is bounded.*

Proof. Assume, by contradiction, that $\lim_{k\to\infty} \rho_k = \infty$. Since $h_1(x_*) = 0$, by the continuity of the first derivatives of h_1 there exists $L > 0$ such that, for all $k \in I\!N$, $\|h_1(x_k)\|_\infty \leq L\|x_k - x_*\|$. Therefore, by the hypothesis, (5.4) and Proposition 5.1, we have that $\|h_1(x_k)\|_\infty \leq LM \max\left\{\frac{\|\lambda_k-\lambda_*\|_\infty}{\rho_k}, \eta_k\|h_1(x_k)\|_\infty\right\}$ for k large enough. Since η_k tends to zero, this implies that

$$\|h_1(x_k)\|_\infty \leq LM\frac{\|\lambda_k - \lambda_*\|_\infty}{\rho_k} \tag{5.10}$$

for k large enough.

By (3.6) and Proposition 5.1, we have that $\lambda_k = \lambda_{k-1} + \rho_{k-1}h_1(x_{k-1})$ for k large enough. Therefore,

$$\|h_1(x_{k-1})\|_\infty = \frac{\|\lambda_k - \lambda_{k-1}\|_\infty}{\rho_{k-1}} \geq \frac{\|\lambda_{k-1} - \lambda_*\|_\infty}{\rho_{k-1}} - \frac{\|\lambda_k - \lambda_*\|_\infty}{\rho_{k-1}}. \tag{5.11}$$

Now, by (5.5), the hypothesis of this theorem and Proposition 5.1, for k large enough we have: $\|\lambda_k - \lambda_*\|_\infty \leq M\left(\frac{\|\lambda_{k-1}-\lambda_*\|_\infty}{\rho_{k-1}} + \eta_{k-1}\|h_1(x_{k-1})\|_\infty\right)$. This implies that $\frac{\|\lambda_{k-1}-\lambda_*\|_\infty}{\rho_{k-1}} \geq \frac{\|\lambda_k-\lambda_*\|_\infty}{M} - \eta_{k-1}\|h_1(x_{k-1})\|_\infty$. Therefore, by (5.11), $(1+\eta_{k-1})\|h_1(x_{k-1})\|_\infty \geq \|\lambda_k - \lambda_*\|_\infty\left(\frac{1}{M} - \frac{1}{\rho_{k-1}}\right) \geq \frac{1}{2M}\|\lambda_k - \lambda_*\|_\infty$. Thus, $\|\lambda_k - \lambda_*\|_\infty \leq 3M\|h_1(x_{k-1})\|_\infty$ for k large enough. By (5.10), we have that $\|h_1(x_k)\|_\infty \leq \frac{3LM^2}{\rho_k}\|h_1(x_{k-1})\|_\infty$. Therefore, since $\rho_k \to \infty$, there exists $k_1 \in I\!N$ such that $\|h_1(x_k)\|_\infty \leq \tau\|h_1(x_{k-1})\|_\infty$ for all $k \geq k_1$. So, $\rho_{k+1} = \rho_k$ for all $k \geq k_1$. Thus, $\{\rho_k\}$ is bounded. ∎

5.2. General constraints. In this subsection we address the general problem (2.1). As in the case of equality constraints, we adopt the following definition for the safeguarded Lagrange multipliers in Algorithm 3.1.

Definition. *For all $k \in I\!N$, $i = 1, \ldots, m_1$, $j = 1, \ldots, p_1$, $[\bar{\lambda}_{k+1}]_i$ will be the projection of $[\lambda_{k+1}]_i$ on the interval $[[\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i]$ and $[\bar{\mu}_{k+1}]_j$ will be the projection of $[\mu_{k+1}]_j$ on $[0, [\bar{\mu}_{\max}]_j]$.*

The technique for proving boundedness of the penalty parameter consists of reducing (2.1) to a problem with (only) equality constraints. The equality constraints of the new problem will be the active constraints at the limit point x_*. After this reduction, the boundedness result is deduced from Theorem 5.4. The sufficient conditions are listed below.

Assumption 7. *The sequence $\{x_k\}$ is generated by the application of Algorithm 3.1 to problem (2.1) and $\lim_{k\to\infty} x_k = x_*$.*

Assumption 8. *The point x_* is feasible ($h_1(x_*) = 0$, $h_2(x_*) = 0$, $g_1(x_*) \leq 0$ and $g_2(x_*) \leq 0$.)*

Assumption 9. *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. (LICQ holds at x_*.)*

Assumption 10. *The functions f, h_1, g_1, h_2 and g_2 admit continuous second derivatives in a neighborhood of x_*.*

Assumption 11. *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 $z \in T, z \neq 0$,*

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

Assumption 12. *For all $i = 1, \ldots, m_1, j = 1, \ldots, p_1$, $[\lambda_*]_i \in ([\bar{\lambda}_{\min}]_i, [\bar{\lambda}_{\max}]_i)$, $[\mu_*]_j \in [0, [\bar{\mu}_{\max}]_j)$.*

Assumption 13. *For all i such that $[g_1(x_*)]_i = 0$, we have $[\mu_*]_i > 0$.*

Observe that Assumption 13 imposes strict complementarity related only to the constraints in the upper-level set. In the lower-level set it is admissible that $[g_2(x_*)]_i = [u_*]_i = 0$. Observe, too, that Assumption 11 is weaker than the usual second-order sufficiency assumption, since the subspace T is orthogonal to the gradients of *all* active constraints, and no exception is made with respect to active constraints with null multiplier $[u_*]_i$. In fact, Assumption 11 is not a second-order sufficiency assumption for local minimizers. It holds for the problem of minimizing $x_1 x_2$ subject to $x_2 - x_1 \le 0$ at $(0,0)$ although $(0,0)$ is not a local minimizer of this problem.

THEOREM 5.5. *Suppose that Assumptions 7–13 are satisfied. In addition, assume that there exists a sequence $\eta_k \to 0$ such that $\varepsilon_k \le \eta_k \max\{\|h_1(x_k)\|_\infty, \|\sigma_k\|_\infty\}$ for all $k \in I\!N$. Then, the sequence of penalty parameters $\{\rho_k\}$ is bounded.*

Proof. Without loss of generality, assume that: $[g_1(x_*)]_i = 0$ if $i \le q_1$, $[g_1(x_*)]_i < 0$ if $i > q_1$, $[g_2(x_*)]_i = 0$ if $i \le q_2$, $[g_2(x_*)]_i < 0$ if $i > q_2$. Consider the auxiliary problem:

$$\text{Minimize } f(x) \text{ subject to } H_1(x) = 0, H_2(x) = 0, \tag{5.12}$$

where $H_1(x) = \begin{pmatrix} h_1(x) \\ [g_1(x)]_1 \\ \vdots \\ [g_1(x)]_{q_1} \end{pmatrix}$, $H_2(x) = \begin{pmatrix} h_2(x) \\ [g_2(x)]_1 \\ \vdots \\ [g_2(x)]_{q_2} \end{pmatrix}$.

By Assumptions 7–11, x_* satisfies the Assumptions 2–5 (with H_1, H_2 replacing h_1, h_2). Moreover, by Assumption 8, the multipliers associated to (2.1) are the Lagrange multipliers associated to (5.12).

As in the proof of (4.10) (first part of the proof of Theorem 4.2), we have that, for k large enough: $[[g_1(x_*)]_i < 0 \Rightarrow [\mu_{k+1}]_i = 0]$ and $[[g_2(x_*)]_i < 0 \Rightarrow [u_k]_i = 0]$.

Then, by (3.1), (3.5) and (3.7),

$$\|\nabla f(x_k) + \textstyle\sum_{i=1}^{m_1} [\lambda_{k+1}]_i \nabla [h_1(x_k)]_i + \sum_{i=1}^{q_1} [\mu_{k+1}]_i \nabla [g_1(x_k)]_i$$
$$+ \textstyle\sum_{i=1}^{m_2} [v_k]_i \nabla [h_2(x_k)]_i + \sum_{i=1}^{q_2} [u_k]_i \nabla [g_2(x_k)]_i\| \le \varepsilon_k$$

for k large enough.

By Assumption 9, taking appropriate limits in the inequality above, we obtain that $\lim_{k\to\infty} \lambda_k = \lambda_*$ and $\lim_{k\to\infty} \mu_k = \mu_*$.

In particular, since $[\mu_*]_i > 0$ for all $i \le q_1$,

$$[\mu_k]_i > 0 \tag{5.13}$$

for k large enough.

Since $\lambda_* \in (\bar\lambda_{\min}, \bar\lambda_{\max})^{m_1}$ and $[\mu_*]_i < [\bar\mu_{\max}]_i$, we have that $[\bar\mu_k]_i = [\mu_k]_i$, $i = 1, \ldots, q_1$ and $[\bar\lambda_k]_i = [\lambda_k]_i$, $i = 1, \ldots, m_1$ for k large enough.

Let us show now that the updating formula (3.7) for $[\mu_{k+1}]_i$, provided by Algorithm 3.1, coincides with the updating formula (3.5) for the corresponding multiplier in the application of the algorithm to the auxiliary problem (5.12).

In fact, by (3.7) and $[\bar\mu_k]_i = [\mu_k]_i$, we have that, for k large enough, $[\mu_{k+1}]_i = \max\{0, [\mu_k]_i + \rho_k [g_1(x_k)]_i\}$. But, by (5.13), $[\mu_{k+1}]_i = [\mu_k]_i + \rho_k [g_1(x_k)]_i, i = 1, \ldots, q_1$, for k large enough.

In terms of the auxiliary problem (5.12) this means that $[\mu_{k+1}]_i = [\mu_k]_i + \rho_k [H_1(x_k)]_i, i = 1, \ldots, q_1$, as we wanted to prove.

Now, let us analyze the meaning of $[\sigma_k]_i$. By (3.7), we have: $[\sigma_k]_i = \max\{[g_1(x_k)]_i, -[\bar\mu_k]_i/\rho_k\}$ for all $i = 1, \ldots, p_1$. If $i > q_1$, since $[g_1(x_*)]_i < 0$, $[g_1]_i$ is continuous and $[\bar\mu_k]_i = 0$, we have that $[\sigma_k]_i = 0$ for k large enough. Now, suppose that $i \le q_1$. If $[g_1(x_k)]_i < -\frac{[\bar\mu_k]_i}{\rho_k}$, then, by (3.7), we would have $[\mu_{k+1}]_i = 0$. This would contradict (5.13). Therefore, $[g_1(x_k)]_i \ge -\frac{[\bar\mu_k]_i}{\rho_k}$ for k large enough and we have that $[\sigma_k]_i = [g_1(x_k)]_i$. Thus, for k large enough,

$$H_1(x_k) = \begin{pmatrix} h_1(x_k) \\ \sigma_k \end{pmatrix}. \tag{5.14}$$

Therefore, the test for updating the penalty parameter in the application of Algorithm 3.1 to (5.12) coincides with the updating test in the application of the algorithm to (2.1). Moreover, formula (5.14) also implies that the condition $\varepsilon_k \leq \eta_k \max\{\|\sigma_k\|_\infty, \|h_1(x_k)\|_\infty\}$ is equivalent to the hypothesis $\varepsilon_k \leq \eta_k \|H_1(x_k)\|_\infty$ assumed in Theorem 5.4.

This completes the proof that the sequence $\{x_k\}$ may be thought as being generated by the application of Algorithm 3.1 to (5.12). We proved that the associated approximate multipliers and the penalty parameters updating rule also coincide. Therefore, by Theorem 5.4, the sequence of penalty parameters is bounded, as we wanted to prove. □

Remark. The results of this section provide a theoretical answer to the following practical question: What happens if the box chosen for the safeguarded multipliers estimates is too small? The answer is: the box should be large enough to contain the "true" Lagrange multipliers. If it is not, the global convergence properties remain but, very likely, the sequence of penalty parameters will be unbounded, leading to hard subproblems and possible numerical instability. In other words, if the box is excessively small, the algorithm tends to behave as an external penalty method. This is exactly what is observed in practice.

6. Numerical experiments. For solving unconstrained and bound-constrained subproblems we use GENCAN [8] with second derivatives and a CG-preconditioner [9]. Algorithm 3.1 with GENCAN will be called ALGENCAN. For solving the convex-constrained subproblems that appear in the large location problems, we use the Spectral Projected Gradient method SPG [10, 11, 12]. The resulting Augmented Lagrangian algorithm is 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. The codes are free for download in *www.ime.usp.br/~egbirgin/tango/*. They are written in Fortran 77 (double precision). Interfaces of ALGENCAN with AMPL, CUTEr, C/C++, Python and R (language and environment for statistical computing) are available and interfaces with Matlab and Octave are being developed.

For the practical implementation of Algorithm 3.1, we set $\tau = 0.5$, $\gamma = 10$, $\bar{\lambda}_{\min} = -10^{20}$, $\bar{\mu}_{\max} = \bar{\lambda}_{\max} = 10^{20}$, $\varepsilon_k = 10^{-4}$ for all k, $\bar{\lambda}_1 = 0, \bar{\mu}_1 = 0$ and $\rho_1 = \max\left\{10^{-6}, \min\left\{10, \frac{2|f(x_0)|}{\|h_1(x_0)\|^2 + \|g_1(x_0)_+\|^2}\right\}\right\}$. As stopping criterion we used $\max(\|h_1(x_k)\|_\infty, \|\sigma_k\|_\infty) \leq 10^{-4}$. The condition $\|\sigma_k\|_\infty \leq 10^{-4}$ guarantees that, for all $i = 1, \ldots, p_1$, $g_i(x_k) \leq 10^{-4}$ and that $[\mu_{k+1}]_i = 0$ whenever $g_i(x_k) < -10^{-4}$. This means that, approximately, feasibility and complementarity hold at the final point. Dual feasibility with tolerance 10^{-4} is guaranteed by (3.1) and the choice of ε_k. All the experiments were run on a 3.2 GHz Intel(R) Pentium(R) with 4 processors, 1Gb of RAM and Linux Operating System. Compiler option "-O" was adopted.

All the experiments were run on an 1.8GHz AMD Opteron 244 processor, 2Gb of RAM memory and Linux operating system. Codes are in Fortran 77 and the compiler option "-O" was adopted.

6.1. Testing the theory. In Discrete Mathematics, experiments should reproduce exactly what theory predicts. In the continuous world, however, the situation changes because the mathematical model that we use for proving theorems is not exactly isomorphic to the one where computations take place. Therefore, it is always interesting to interpret, in finite precision calculations, the continuous theoretical results and to verify to what extent they are fulfilled.

Some practical results presented below may be explained in terms of a simple theoretical result that was tangentially mentioned in the introduction: If, at Step 2 of Algorithm 3.1, one computes a global minimizer of the subproblem and the problem (2.1) is feasible, then every limit point is a global minimizer of (2.1). This property may be easily proved using boundedness of the safeguarded Lagrange multipliers by means of external penalty arguments. Now, algorithms designed to solve reasonably simple subproblems usually include practical procedures that actively seek function decrease, beyond the necessity of finding stationary points. For example, efficient line-search procedures in unconstrained minimization and box-constrained minimization usually employ aggressive extrapolation steps [8], although simple backtracking is enough to prove convergence to stationary points. In other words, from good subproblem solvers one expects much more than convergence to stationary points. For this reason, we conjecture that Augmented

Lagrangian algorithms like ALGENCAN tend to converge to global minimizers more often than SQP-like methods. In any case, these arguments support the necessity of developing global-oriented subproblem solvers.

Experiments in this subsection were made using the AMPL interfaces of ALGENCAN (considering all the constraints as upper-level constraints) and IPOPT. Presolve AMPL option was disabled to solve the problems exactly as they are. The ALGENCAN parameters and stopping criteria were the ones stated at the beginning of this section. For IPOPT we used all its default parameters (including the ones related to stopping criteria). The random generation of initial points was made using the function `Uniform01()` provided by AMPL. When generating several random initial points, the seed used to generate the i-th random initial point was set to i.

Example 1: *Convergence to KKT points that do not satisfy MFCQ.*

$$\begin{array}{ll} \text{Minimize} & x_1 \\ \text{subject to} & x_1^2 + x_2^2 \le 1, \\ & x_1^2 + x_2^2 \ge 1. \end{array}$$

The global solution is $(-1, 0)$ and no feasible point satisfies the Mangasarian-Fromovitz Constraint Qualification, although all feasible points satisfy CPLD. Starting with 100 random points in $[-10, 10]^2$, ALGENCAN converged to the global solution in all the cases. Starting from $(5, 5)$ convergence occurred using 14 outer iterations. The final penalty parameter was 4.1649E-01 (the initial one was 4.1649E-03) and the final multipliers were 4.9998E-01 and 0.0000E+00. IPOPT also found the global solution in all the cases and used 25 iterations when starting from $(5, 5)$.

Example 2: *Convergence to a non-KKT point.*

$$\begin{array}{ll} \text{Minimize} & x \\ \text{subject to} & x^2 = 0, \\ & x^3 = 0, \\ & x^4 = 0. \end{array}$$

Here the gradients of the constraints are linearly dependent for all $x \in I\!R$. In spite of this, the only point that satisfies Theorem 4.1 is $x = 0$. Starting with 100 random points in $[-10, 10]$, ALGENCAN converged to the global solution in all the cases. Starting with $x = 5$ convergence occurred using 20 outer iterations. The final penalty parameter was 2.4578E+05 (the initial one was 2.4578E-05) and the final multipliers were 5.2855E+01 -2.0317E+00 and 4.6041E-01. IPOPT was not able to solve the problem in its original formulation because "Number of degrees of freedom is NIND = -2". We modified the problem in the following way

$$\begin{array}{ll} \text{Minimize} & x_1 + x_2 + x_3 \\ \text{subject to} & x_1^2 = 0, \\ & x_1^3 = 0, \\ & x_1^4 = 0, \\ & x_i \ge 0, i = 1, 2, 3, \end{array}$$

and, after 16 iterations, IPOPT stopped near $x = (0, +\infty, +\infty)$ saying "Iterates become very large (diverging?)".

Example 3: *Infeasible stationary points* [18, 46].

$$\begin{array}{ll} \text{Minimize} & 100(x_2 - x_1^2)^2 + (x_1 - 1)^2 \\ \text{subject to} & x_1 - x_2^2 \le 0, \\ & x_2 - x_1^2 \le 0, \\ & -0.5 \le x_1 \le 0.5, \\ & x_2 \le 1. \end{array}$$

This problem has a global KKT solution at $x = (0, 0)$ and a stationary infeasible point at $x = (0.5, \sqrt{0.5})$. Starting with 100 random points in $[-10, 10]^2$, ALGENCAN converged to the global solution in all the cases. Starting with $x = (5, 5)$ convergence occurred using 6 outer iterations. The final penalty parameter was 1.0000E+01 (the initial one was 1.0000E+00) and the final multipliers were 1.9998E+00 and 3.3390E-03. IPOPT found the global solution starting from 84 out of the 100 random initial points. In the other 16 cases IPOPT stopped at $x = (0.5, \sqrt{0.5})$ saying "Convergence to stationary point for infeasibility" (this was also the case when starting from $x = (5, 5)$).

Example 4: *Difficult-for-barrier* [15, 18, 70].

$$\begin{array}{ll} \text{Minimize} & x_1 \\ \text{subject to} & x_1^2 - x_2 + a = 0, \\ & x_1 - x_3 - b = 0, \\ & x_2 \geq 0, \ x_3 \geq 0. \end{array}$$

In [18] we read: "This test example is from [70] and [15]. Although it is well-posed, many barrier-SQP methods ('Type-I Algorithms' in [70]) fail to obtain feasibility for a range of infeasible starting points."

We ran two instances of this problem varying the values of parameters a and b and the initial point x_0 as suggested in [18]. When $(a, b) = (1, 1)$ and $x_0 = (-3, 1, 1)$ ALGENCAN converged to the solution $\bar{x} = (1, 2, 0)$ using 2 outer iterations. The final penalty parameter was 5.6604E-01 (the initial one also was 5.6604E-01) and the final multipliers were 6.6523E-10 and -1.0000E+00. IPOPT also found the same solution using 20 iterations. When $(a, b) = (-1, 0.5)$ and $x_0 = (-2, 1, 1)$ ALGENCAN converged to the solution $\tilde{x} = (1, 0, 0.5)$ using 5 outer iterations. The final penalty parameter was 2.4615E+00 (the initial one also was 2.4615E+00) and the final multipliers were -5.0001E-01 and -1.3664E-16. On the other hand, IPOPT stopped declaring convergence to a stationary point for the infeasibility.

Example 5: *Preference for global minimizers 1*

$$\begin{array}{ll} \text{Minimize} & \sum_{i=1}^{n} x_i \\ \text{subject to} & x_i^2 = 1, i = 1, \ldots, n. \end{array}$$

Solution: $x_* = (-1, \ldots, -1)$, $f(x_*) = -n$. We set $n = 100$ and ran ALGENCAN and IPOPT starting from 100 random initial points in $[-100, 100]^n$. ALGENCAN converged to the global solution in all the cases while IPOPT never found the global solution. When starting from the first random point, ALGENCAN converged using 4 outer iterations. The final penalty parameter was 5.0882E+00 (the initial one was 5.0882E-01) and the final multipliers were all equal to 4.9999E-01.

Problem 6: *Preference for global minimizers 2*

$$\begin{array}{ll} \text{Minimize} & x_2 \\ \text{subject to} & x_1 \cos(x_1) - x_2 \leq 0, \\ & -10 \leq x_i \leq 10, i = 1, 2. \end{array}$$

It can be seen in Figure 1 that the problem has five local minimizers at approx. $(-10, 8.390)$, $(-0.850, -0.561)$, $(3.433, -3.288)$, $(-6.436, -6.361)$ and $(9.519, -9.477)$. Clearly, the last one is the global minimizer. The number of times ALGENCAN found these solutions are 1, 0, 8, 26 and 65, respectively; while the figures for IPOPT are 1, 18, 21, 38 and 22.

6.2. Problems with many inequality constraints. Consider the hard-spheres problem [49]:

$$\begin{array}{ll} \text{Minimize} & z \\ \text{subject to} & \langle v_i, v_j \rangle \leq z, \ i = 1, \ldots, np, \ j = i + 1, \ldots, np, \\ & \|v_i\|_2^2 = 1, \ i = 1, \ldots, np, \end{array}$$

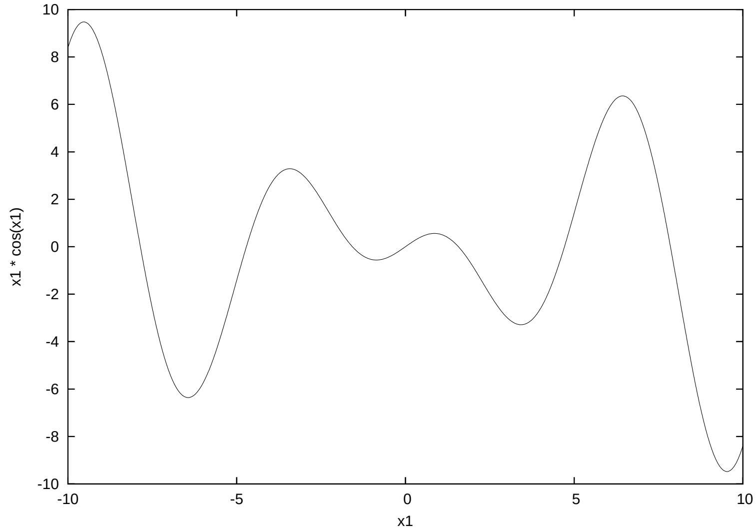

FIG. 6.1. *Problem B.*

where $v_i \in I\!\!R^{nd}$ for all $i = 1, \ldots, np$. This problem has $nd \times np + 1$ variables, np equality constraints and $np \times (np - 1)/2$ inequality constraints.

For this particular problem, we ran ALGENCAN with the Fortran formulation of the problem, LANCELOT B with the SIF formulation and IPOPT with the AMPL formulation. In all cases we used analytic second derivatives and the same random initial point, as we coded a Fortran 77 program that generates SIF and AMPL formulations that include the initial point. The motivation for this choice was that, although the reasons are not clear for us, the combination of ALGENCAN with the AMPL formulation was very slow (this was not the case in the other problems) and the combination of LANCELOT with AMPL gave the error message "[...] failure: ran out of memory" for $np \geq 180$. The combination of IPOPT with AMPL worked well and gave an average behavior almost identical to the combination of IPOPT with the Fortran 77 formulation of the problem. *We are aware that these different computer environments may have some influence in the interpretation of the results.* However, we think that the numerical results indicate the essential correctness of the many-inequalities claim.

For LANCELOT B (and LANCELOT, which will appear in further experiments) we used all its default parameters and the same stopping criterion as ALGENCAN, i.e., 10^{-4} for the feasibility and optimality tolerances measured in the sup-norm. *This may be considered as a rather large feasibility-optimality tolerance. In fact, Augmented Lagrangian methods do not have fast local convergence. If we ask for more strict tolerances, the performance of the Augmented Lagrangian methods deteriorate in comparison to the one of* IPOPT.

We generated 20 different problems fixing $nd = 3$ and choosing $np \in \{10, 20, \ldots, 200\}$. For each problem, we ran ALGENCAN, IPOPT and LANCELOT B starting from 10 different random initial points with $v_i \in [-1, 1]^{nd}$ for $i = 1, \ldots, np$ and $z \in [0, 1]$. The three methods satisfied the stopping criterion in all the cases. Table 1 shows, for each problem and method, the average objective function value found (f), the average CPU time used in seconds (Time), and how many times, over the 10 trials for the same problem, a method found the best functional value (Glcnt). In the table, n and m represent the number of variables and constraints of the original formulation of the problem, respectively, i.e., without considering the slack variables added by IPOPT and LANCELOT B. The number of inequality constraints and the sparsity structure of the Hessian of the Lagrangian favors the application of ALGENCAN for solving this problem.

n	m	Algencan Glcnt	Algencan Time	Algencan f	Ipopt Glcnt	Ipopt Time	Ipopt f	Lancelot B Glcnt	Lancelot B Time	Lancelot B f
31	55	7	0.01	0.404687	9	0.04	0.408676	6	0.04	0.409244
61	210	10	0.03	0.676472	9	0.28	0.676851	10	0.42	0.676477
91	465	10	0.17	0.781551	6	1.08	0.783792	9	1.97	0.782241
121	820	10	0.47	0.837600	6	3.12	0.838449	4	6.79	0.839005
151	1275	10	0.97	0.868312	4	9.26	0.869486	2	14.63	0.870644
181	1830	10	1.65	0.889745	3	13.96	0.891025	2	38.08	0.891436
211	2485	10	3.12	0.905335	5	30.87	0.905975	3	64.36	0.906334
241	3240	9	4.14	0.917323	10	33.91	0.918198	3	109.88	0.918182
271	4095	10	5.66	0.926429	8	44.03	0.927062	10	158.12	0.926889
301	5050	10	7.75	0.933671	9	61.47	0.933892	7	289.92	0.934201
331	6105	10	11.29	0.939487	9	84.18	0.939791	10	287.40	0.940192
361	7260	10	16.48	0.944514	10	115.73	0.944651	10	386.92	0.945134
391	8515	9	20.18	0.953824	9	171.56	0.948912	10	486.04	0.949213
421	9870	10	24.99	0.952265	10	254.16	0.952398	9	842.10	0.952667
451	11325	10	31.00	0.955438	10	259.74	0.955609	10	811.37	0.955943
481	12880	10	35.04	0.958227	10	289.55	0.958325	10	1381.74	0.958654
511	14535	10	42.10	0.960621	10	457.32	0.960682	10	1407.71	0.961051
541	16290	10	48.94	0.962813	10	476.82	0.962837	10	1565.95	0.963113
571	18145	10	57.83	0.964722	10	773.90	0.964795	10	1559.51	0.965159
601	20100	9	68.52	0.969767	10	1155.17	0.966444	10	2480.00	0.966822

Table 6.1
Performance of Algencan, Ipopt *and* Lancelot *B in the hard-spheres problem.*

6.3. Problems with poor Lagrangian Hessian structure. The discretized three-dimensional Bratu-based [26, 48] optimization problem that we consider in this subsection is:

$$\begin{array}{ll} \text{Minimize} & \sum_{(i,j,k)\in S}(u(i,j,k)-u_*(i,j,k))^2 \\ \text{subject to} & \phi_\theta(u,i,j,k)=\phi_\theta(u_*,i,j,k),\ i,j,k=2,\ldots,np-1. \end{array}$$

where u_* was choosed as

$$u_*(i,j,k) = 10\, q(i)\, q(j)\, q(k)\, (1-q(i))\, (1-q(j))\, (1-q(k))\, e^{q(k)^{4.5}}$$

with $q(\ell)=\frac{np-\ell}{np-1}$ for $i,j,k=1,\ldots,np$ and

$$\phi_\theta(v,i,j,k) = -\Delta v(i,j,k) + \theta e^{v(i,j,k)},$$

$$\Delta v(i,j,k) = \frac{v(i\pm 1,j,k)+v(i,j\pm 1,k)+v(i,j,k\pm 1)-6v(i,j,k)}{h^2},$$

for $i,j,k=2,\ldots,np-1$. The number of variables is $n=np^3$ and the number of (equality) constraints is $m=(np-2)^3$. We setted $\theta=-100$, $h=1/(np-1)$, $|S|=7$ and the 3-uples of indices in S were randomly selected in $[1,np]^3$.

Sixteen problems were generated setting $np=5,6,\ldots,20$. They were solved using Algencan, Ipopt and Lancelot (this problem was formulated in AMPL and Lancelot B has no AMPL interface). The initial point was randomly generated in $[0,1]^n$. The three methods found solutions with null objective function value. Table 2 shows some figures that reflect the computational effort of the methods. In the table, "Outit" means number of outer iterations of an augmented Lagrangian method, "It" means number of iterations (or inner iterations), "Fcnt" means number of functional evaluations, "Gcnt" means number of gradient evaluations and "Time" means CPU time in seconds. In the table, n and m represent the number of variables[1] and (equality) constraints of the problem. The poor sparsity structure of the Hessian of the Lagrangian favors the application of Algencan for solving this problem.

[1]The AMPL presolver procedure eliminates some problem variables.

		Algencan					Ipopt		Lancelot			
n	m	OutIt	It	Fcnt	Gcnt	Time	It	Time	It	Fcnt	Gcnt	Time
83	27	6	23	60	39	0.00	12	0.01	145	146	117	0.04
162	64	4	22	49	33	0.02	14	0.02	93	94	77	0.06
277	125	5	24	64	36	0.05	11	0.04	86	87	69	0.13
433	216	5	18	42	30	0.07	13	0.10	86	87	71	0.26
638	343	3	23	37	31	0.19	9	0.17	73	74	59	0.37
897	512	4	33	87	45	0.56	10	0.44	52	53	41	0.45
1216	729	4	19	40	29	0.47	10	0.93	82	83	70	1.44
1601	1000	4	41	113	54	2.37	10	1.86	100	101	82	3.07
2058	1331	1	15	35	19	1.05	10	3.03	166	167	141	7.94
2593	1728	2	38	83	45	4.43	10	6.73	107	108	88	7.81
3212	2197	3	35	80	46	7.24	11	14.71	150	151	130	10.37
3921	2744	2	39	85	46	10.23	11	30.93	295	296	254	34.38
4726	3375	3	55	177	65	19.01	11	59.92	334	335	294	60.98
5633	4096	2	40	86	48	19.78	11	84.57	417	418	369	100.46
6647	4913	3	48	110	57	23.42	11	130.03	426	427	384	172.30
7776	5832	3	62	140	72	36.62	11	305.26	332	333	273	136.76

Table 6.2

Performance of Algencan, Ipopt *and* Lancelot *in the three-dimensional Bratu-based optimization problem.*

The warnings in italics made with respect to the Hard-Spheres problem are also pertinent for this problem. In particular, remember that we use a non-exigent convergence criterion.

6.4. Location problems. Here we will consider a variant of the family of *location* problems introduced in [11]. In the original problem, given a set of np disjoint polygons $P_1, P_2, \ldots, P_{np}$ in $\mathbb{R}^2$ one wishes to find the point $z^1 \in P_1$ that minimizes the sum of the distances to the other polygons. Therefore, the original problem formulation is:

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

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

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

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

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

where $g(x) = (x_1/a)^2 + (x_2/b)^2 - c$, and $a, b, c \in \mathbb{R}$ are positive constants. Observe that the objective function is differentiable in a large open neighborhood of the feasible region.

We generated 36 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 $g(z^1) \le 0$ and $g(z^i) \ge 0, i = 2, \dots, np + nc$ as upper-level constraints, and $z^i \in P_i, i = 1, \dots, np$ and $z^{np+i} \in C_i, i = 1, \dots, nc$ as lower-level constraints. In this way the subproblems can be efficiently solved by the Spectral Projected Gradient method (SPG) [10, 11] as suggested by the experiments in [11]. 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.

The 36 problems are divided in two sets of 18 problems: small and large problems. We first solved the small problems with ALGENCAN (considering all the constraints as upper-level constraints), ALSPG, IPOPT and LANCELOT. All the methods used the AMPL formulation of the problem, except ALSPG which, due to the necessity of a subroutine to compute the projection of an arbitrary point onto the convex set given by the lower-level constraints, used the Fortran 77 formulation of the problem. *We are aware that this is a limitation in terms of a fair comparison, however we do not think that the differences observed in the results are due to the gap between Fortran and AMPL.*

In Table 4 we compare the performance of the four methods for solving this problem. Observe that the four methods obtain feasible points and arrive to the same solutions. Due to the performance of ALSPG and IPOPT, we solved the set of large problems using them. Table 5 shows their performances. For the larger problems IPOPT gave the error message "error running ipopt: termination code 15". Probably this an inconvenient related to memory requirements. However, note that the running times of ALSPG are orders of magnitude smaller than the IPOPT running times. *Again take into account the gap Fortran-AMPL and the lack of plausibility of the fact that the gap explains the large time differences.*

6.5. Problems in the Cuter collection. We have 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). Preliminary experiments showed that the version with a single penalty parameter performed slightly better. So, we compare this version against LANCELOT B and IPOPT. To perform the numerical experiments, we considered all the problems of the CUTEr collection [13]. As a whole, we tried to solve 1023 problems.

We use ALGENCAN, IPOPT and LANCELOT B with all their default parameters. The stopping criterion for ALGENCAN and LANCELOT B is feasibility and optimality (measured in the sup-norm) less than or equal to 10^{-4}, while for IPOPT we use the default stopping criteria. *In general, this implies that the convergence criterion is more exigent for* IPOPT *than for* ALGENCAN *and* LANCELOT. *This fact should be taken into account in the analysis of numerical results.*

We also stop a method if its execution exceeds 5 minutes of CPU time.

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

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

We define

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

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

$$f(x^M_{\text{final}}) \le f_{\text{best}} + 10^{-3}|f_{\text{best}}| + 10^{-6} \quad \text{or} \quad \max\{f_{\text{best}}, f(x^M_{\text{final}})\} \le -10^{20}.$$

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

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

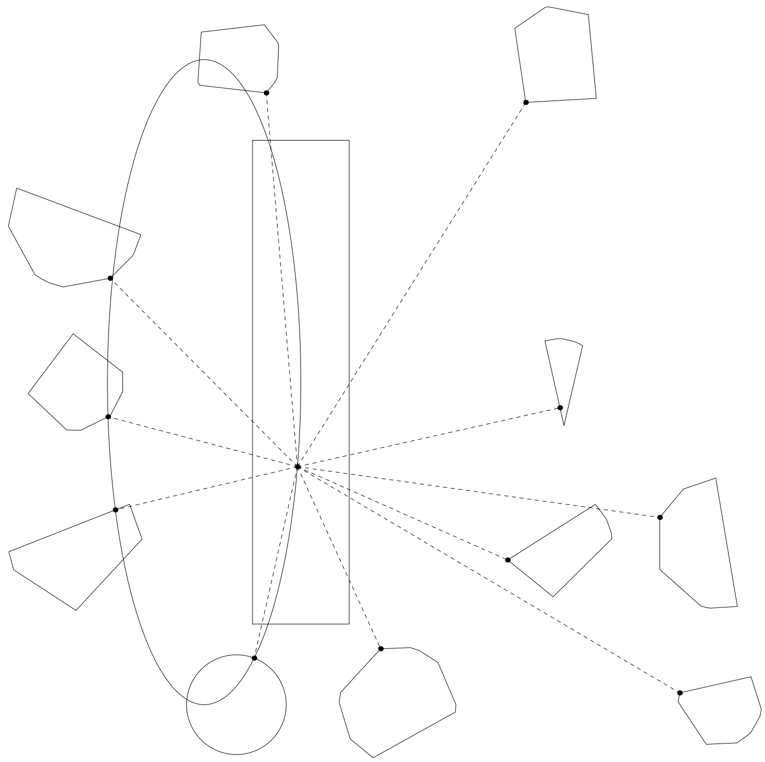

FIG. 6.2. *Twelve-sets very small location problem.*

We use r as a performance measurement. The results of comparing ALGENCAN, IPOPT and LANCELOT B are reported in the form of performance profiles and two small numerical tables. See Figure 3 and Table 6.

There are a few comparison issues that should be noted:

- When LANCELOT B solves a feasibility problem (problem with constant objective function), it minimizes the squared infeasibility instead of addressing the original problem. As a result, it sometimes finishes without satisfying the user required stopping criteria (feasibility and optimality tolerances on the the original problem). In 35 feasibility problems, LANCELOT B stopped declaring convergence but the user-required feasibility tolerance is not satisfied at the final iterate. 16 of the 35 problems seem to be problems in which LANCELOT B converged to a stationary point of the infeasibility (large objective function value of the reformulated problem). In the remaining 19 problems, LANCELOT B seems to have been stopped prematurely. This easy-to-solve inconvenient may slightly deteriorate the robustness of LANCELOT B.
- It is simple to use the same stopping criterion for ALGENCAN and LANCELOT B but this is not the case for IPOPT. So, in IPOPT runs we used its default stopping criterion.
- ALGENCAN and LANCELOT B satisfy the bound constraints exactly, whereas IPOPT satisfies them within a prescribed tolerance $\epsilon > 0$. Consider, for example, the following problem: Min $\sum_i x_i$ subject to $x_i \geq 0$. The solution given by ALGENCAN and LANCELOT B is the origin, while the solution given by IPOPT is $x_i = -\epsilon$. So, the minimum found by ALGENCAN and LANCELOT B is 0 while the minimum found by IPOPT is $-\epsilon n$. This phenomenon occurs for a big family of reformulated complementarity problems (provided by M. Ferris) in the CUTEr collection. For these problems we considered that $f(x_{\text{final}}^{\text{IPOPT}}) = 0$. We do not know if this occurs in other problems of the collection. *It must be awared that the opposite situation may occur in other problems of the collection. Namely, since we use a weak stopping feasibility criterion for the Augmented Lagrangian methods, it is possible that the objective function value of an Augmented Lagrangian method is artificially smaller than the objective function value obtained by* IPOPT*in*

Problem	nc	np	$totnvs$	n	p_1	p_2
1	28	98	295	252	126	323
2	33	108	432	282	141	465
3	33	108	539	282	141	572
4	33	109	652	284	142	685
5	35	118	823	306	153	858
6	35	118	940	306	153	975
7	35	118	1,057	306	153	1,092
8	35	118	1,174	306	153	1,209
9	35	118	1,291	306	153	1,326
10	35	118	1,408	306	153	1,443
11	35	118	1,525	306	153	1,560
12	35	118	1,642	306	153	1,677
13	35	118	1,759	306	153	1,794
14	35	118	1,876	306	153	1,911
15	35	118	1,993	306	153	2,028
16	35	118	2,110	306	153	2,145
17	35	118	2,227	306	153	2,262
18	35	118	2,344	306	153	2,379
19	3,029	4,995	62,301	16,048	8,024	65,330
20	4,342	7,271	91,041	23,226	11,613	95,383
21	6,346	10,715	133,986	34,122	17,061	140,332
22	13,327	22,230	278,195	71,114	35,557	291,522
23	19,808	33,433	417,846	106,482	53,241	437,654
24	29,812	50,236	627,548	160,096	80,048	657,360
25	26,318	43,970	549,900	140,576	70,288	576,218
26	39,296	66,054	825,907	210,700	105,350	865,203
27	58,738	99,383	1,241,823	316,242	158,121	1,300,561
28	65,659	109,099	1,363,857	349,516	174,758	1,429,516
29	98,004	164,209	2,052,283	524,426	262,213	2,150,287
30	147,492	245,948	3,072,630	786,880	393,440	3,220,122
31	131,067	218,459	2,730,798	699,052	349,526	2,861,865
32	195,801	327,499	4,094,827	1,046,600	523,300	4,290,628
33	294,327	490,515	6,129,119	1,569,684	784,842	6,423,446
34	261,319	435,414	5,442,424	1,393,466	696,733	5,703,743
35	390,670	654,163	8,177,200	2,089,666	1,044,833	8,567,870
36	588,251	979,553	12,244,855	3,135,608	1,567,804	12,833,106

TABLE 6.3

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 total number of constraints is $m = p_1 + p_2$. 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.

some (perhaps many) problems. The numerical results presented here should be analyzed with this objection in mind.

- We have good reasons for defining the initial penalty parameter ρ_1 as stated at the beginning of this section. However, in many problems of the CUTEr collection, $\rho_1 = 10$ behaves better. For this reason we include the statistics also for the non-default choice $\rho_1 = 10$.

We detected 73 problems in which both ALGENCAN and IPOPT finished declaring that the optimal solution was found but found different functional values. In 58 of these problems the functional value obtained by ALGENCAN was smaller than the one found by IPOPT. This may confirm the conjecture that

Problem	CPU Time (secs.)				f
	ALGENCAN	ALSPG	IPOPT	LANCELOT	
1	1.53	0.06	0.11	854.82	1.7564E+01
2	2.17	0.11	0.13	319.24	1.7488E+01
3	2.25	0.14	0.14	401.08	1.7466E+01
4	1.71	0.12	0.17	139.60	1.7451E+01
5	1.75	0.11	0.19	129.79	1.7984E+01
6	1.83	0.09	0.21	90.37	1.7979E+01
7	2.18	0.08	0.23	72.14	1.7975E+01
8	1.88	0.08	0.28	92.74	1.7971E+01
9	1.84	0.18	0.29	111.13	1.7972E+01
10	2.06	0.14	0.37	100.23	1.7969E+01
11	2.18	0.13	0.49	86.54	1.7969E+01
12	2.55	0.16	0.50	134.23	1.7968E+01
13	2.39	0.18	0.37	110.28	1.7968E+01
14	2.52	0.20	0.42	223.05	1.7965E+01
15	2.63	0.17	0.61	657.99	1.7965E+01
16	3.36	0.18	0.44	672.01	1.7965E+01
17	2.99	0.17	0.46	505.00	1.7963E+01
18	3.43	0.23	0.48	422.93	1.7963E+01

TABLE 6.4
Performance of ALGENCAN, ALSPG, IPOPT *and* LANCELOT *in the set of small location problems.*

the Augmented Lagrangian method has a stronger tendency towards global optimality than interior-SQP methods but it must also be taken into account the possibility of an unfair functional value comparison due to the different tolerances used.

7. 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 [23], restoration, nonmonotone strategies and advanced sparse linear algebra procedures. See, for example [17, 38, 40, 41, 42, 52] 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. This condition [59] has been proved to be a constraint qualification in [3] where its relations with other constraint qualifications have been given.

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 five minutes of CPU time.

Problem	Alspg					Ipopt		f
	OuIt	InIt	Fcnt	Gcnt	Time	It	Time	
19	8	212	308	220	2.82	60	44.76	4.5752E+02
20	8	107	186	115	2.23	62	61.79	5.6012E+02
21	9	75	149	84	2.37	65	93.42	6.8724E+02
22	7	80	132	87	5.16	66	202.62	4.6160E+02
23	7	71	125	78	6.99	231	1104.18	5.6340E+02
24	8	53	106	61	8.72			6.9250E+02
25	8	55	124	63	8.00			4.6211E+02
26	7	63	127	70	12.58			5.6438E+02
27	9	80	155	89	20.33			6.9347E+02
28	8	67	138	75	22.33			4.6261E+02
29	7	54	107	61	27.57			5.6455E+02
30	9	95	179	104	54.79			6.9382E+02
31	7	59	111	66	38.74			4.6280E+02
32	7	66	120	73	64.27			5.6449E+02
33	9	51	113	60	85.58			6.9413E+02
34	7	58	110	65	78.30			4.6270E+02
35	7	50	104	57	107.28			5.6432E+02
36	10	56	133	66	184.73			6.9404E+02

Table 6.5

Performance of Alspg *and* Ipopt *on set of large location problems. The memory limitation is the only inconvenient for* Alspg *solving problems with higher dimension than problem 36 (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 quite reasonable.*

$\rho_1 = 10$			
	Algencan	Ipopt	Lancelot B
Efficiency	567	583	439
Robustness	783	778	734

Dynamic ρ_1 as stated at the beginning of this section.			
	Algencan	Ipopt	Lancelot B
Efficiency	567	572	440
Robustness	775	777	732

Table 6.6

The total number of considered problems is 1023. Efficiency means number of times that method M *obtained the best* r^M. *Robustness means the number of times in which* $r^M < \infty$.

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 La-

grangian 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 [2] 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.

Acknowledgement

We are indebted to Prof. A. R. Conn, whose comments on a first version of this paper guided a deep revision.

REFERENCES

[1] M. Argáez and R. A. Tapia, *On the global convergence of a modified augmented Lagrangian linesearch interior-point method for Nonlinear Programming*, Journal of Optimization Theory and Applications 114 (2002), pp. 1-25.

[2] 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. To appear in Mathematical Programming.

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

[4] S. Bakhtiari and A. L. Tits, *A simple primal-dual feasible interior-point method for nonlinear programming with monotone descent*, Computational Optimization and Applications 25 (2003), pp. 17-38.

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

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

[7] E. G. Birgin, R. Castillo and J. M. Martínez, *Numerical comparison of Augmented Lagrangian algorithms for non-convex problems*, Computational Optimization and Applications 31 (2005), pp. 31-56.

[8] E. G. Birgin and J. M. Martínez, *Structured minimal-memory inexact quasi-Newton method and secant preconditioners for Augmented Lagrangian Optimization*, submitted (2005).

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

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

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

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

[13] 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 (1995), pp. 123–160.

[14] R. H. Byrd, J. Ch. Gilbert and J. Nocedal, *A trust region method based on interior point techniques for nonlinear programming*, Mathematical Programming 89 (2000), pp. 149-185.

[15] R. H. Byrd, M. Marazzi and J. Nocedal, *On the convergence of Newton iterations to nonstationary points*, Mathematical Programming 99 (2004), pp. 127-148.

[16] R. H. Byrd, J. Nocedal and A. Waltz, *Feasible interior methods using slacks for nonlinear optimization*, Computational Optimization and Applications 26 (2003), pp. 35-61.

[17] 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 (2004), pp. 27–48.

[18] L. Chen and D. Goldfarb. *Interior-Point ℓ_2 penalty methods for nonlinear programming with strong global convergence properties*, CORC Technical Report TR 2004-08, IEOR Department, Columbia University, 2005.

[19] A. R. Conn, N. I. M. Gould, D. Orban and Ph. L. Toint, *A primal-dual trust-region algorithm for nonconvex nonlinear programming*, Mathematical Programming 87 (2000), pp. 215-249.

[20] 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 (1996), pp. 674–703.

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

[22] 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 (1991), pp. 545–572.

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

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

[25] 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 (2004), pp. 497–517.

[26] M. A. Diniz-Ehrhardt, M. A. Gomes-Ruggiero, V. L. R. Lopes and J. M. Martínez, *Discrete Newton's method with local variations for solving large-scale nonlinear systems*, Optimization 52 (2003), pp. 417–440.

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

[28] 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.

[29] 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 (2003), pp. 1120–1140.

[30] 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 (2000), pp. 397–415.

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

[32] A. V. Fiacco and G.P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley & Sons, New York, 1968.

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

[34] R. Fletcher, N. I. M. Gould, S. Leyffer, Ph. L. Toint and A. Wächter, *Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming*, SIAM Journal on Optimization 13 (2002), pp. 635-659.

[35] A. Forsgren, P. E. Gill and M. H. Wright, *Interior point methods for nonlinear optimization*, SIAM Review 44 (2002), pp. 525-597.

[36] 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 (2004), pp. 10863-10878.

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

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

[39] P. E. Gill, W. Murray and M. A. Saunders, *SNOPT: An SQP algorithm for large-scale constrained optimization*, SIAM Review 47 (2005), pp. 99–131.

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

[41] 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 (2000), pp.974-1002.

[42] 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 (2003), pp. 353-372.

[43] N. I. M. Gould, D. Orban and Ph. L. Toint, *An interior point ℓ_1 penalty method for nonlinear optimization*, Computational Science and Engineering Department, Rutherford Appleton Laboratory, Chilton, Oxfordshire, England, 2003.

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

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

[46] W. Hock and K. Schittkowski, *Test examples for nonlinear programming codes*, volume 187 of *Lecture Notes in Economics and Mathematical Systems*, Springer-Verlag, 1981.

[47] Y. Kaya and J. M. Martínez, *Solving control problems with inexact restoration*, Technical Report MCDO050502, Department of Applied Mathematics, State University of Campinas, Brazil, 2005.

[48] C. T. Kelley, *Iterative methods for linear and nonlinear equations*, SIAM, 1995.

[49] N. Krejić, J. M. Martínez, M. P. Mello and E. A. Pilotta, *Validation of an augmented Lagrangian algorithm with a Gauss-Newton Hessian approximation using a set of hard-spheres problems*, Computational Optimization and Applications 16 (2000) pp. 247-263.

[50] X. Liu and J. Sun, *A robust primal-dual interior point algorithm for nonlinear programs*, SIAM Journal on Optimiza-

tion 14 (2004), pp. 1163-1186.

[51] 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 (1967) pp. 37–47.

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

[53] J. M. Martínez and E. A. Pilotta, *Inexact restoration methods for nonlinear programming: advances and perspectives*, in “Optimization and Control with applications”, edited by L. Q. Qi, K. L. Teo and X. Q. Yang. Springer, pp. 271-292, 2005.

[54] J. M. Moguerza and F. J. Prieto, *An augmented Lagrangian interior-point method using directions of negative curvature*, Mathematical Programming 95 (2003), pp. 573-616.

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

[56] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.

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

[58] 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.

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

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

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

[62] R. T. Rockafellar, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, SIAM Journal on Control and Optimization 12 (1974), pp. 268-285.

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

[64] 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 (2000), pp. 611-646.

[65] D. F. Shanno and R. J. Vanderbei, *Interior-point methods for nonconvex nonlinear programming: orderings and high-order methods*, Mathematical Programming 87 (2000), pp. 303-316.

[66] 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 (2004), pp. 16-27.

[67] A. L. Tits, A. Wächter, S. Bakhtiari, T. J. Urban and C. T. Lawrence, *A primal-dual interior-point method for nonlinear programming with strong global and local convergence properties*, SIAM Journal on Optimization 14 (2003), pp. 173-199.

[68] P. Tseng, *A convergent infeasible interior-point trust-region method for constrained minimization*, SIAM Journal on Optimization 13 (2002), pp. 432-469.

[69] M. Ulbrich, S. Ulbrich and L. N. Vicente, *A globally convergent primal-dual interior-point filter method for nonlinear programming*, Mathematical Programming 100 (2004), pp. 379-410.

[70] A. Wächter and L. T. Biegler, *Failure of global convergence for a class of interior point methods for nonlinear programming*, Mathematical Programming 88 (2000), pp. 565-574.

[71] A. Wächter and L. T. Biegler, *On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming*, Technical Report, Optimization Technology Center, Northwestern University, Evanston, IL, 2003, to appear in Mathematical Programming.

[72] R. A. Waltz, J. L. Morales, J. Nocedal and D. Orban, *An interior algorithm for nonlinear optimization that combines line search and trust region steps*, to appear in Mathematical Programming.

[73] H. Yamashita and H. Yabe, *An interior point method with a primal-dual quadratic barrier penalty function for nonlinear optimization*, SIAM Journal on Optimization 14 (2003), pp. 479-499.

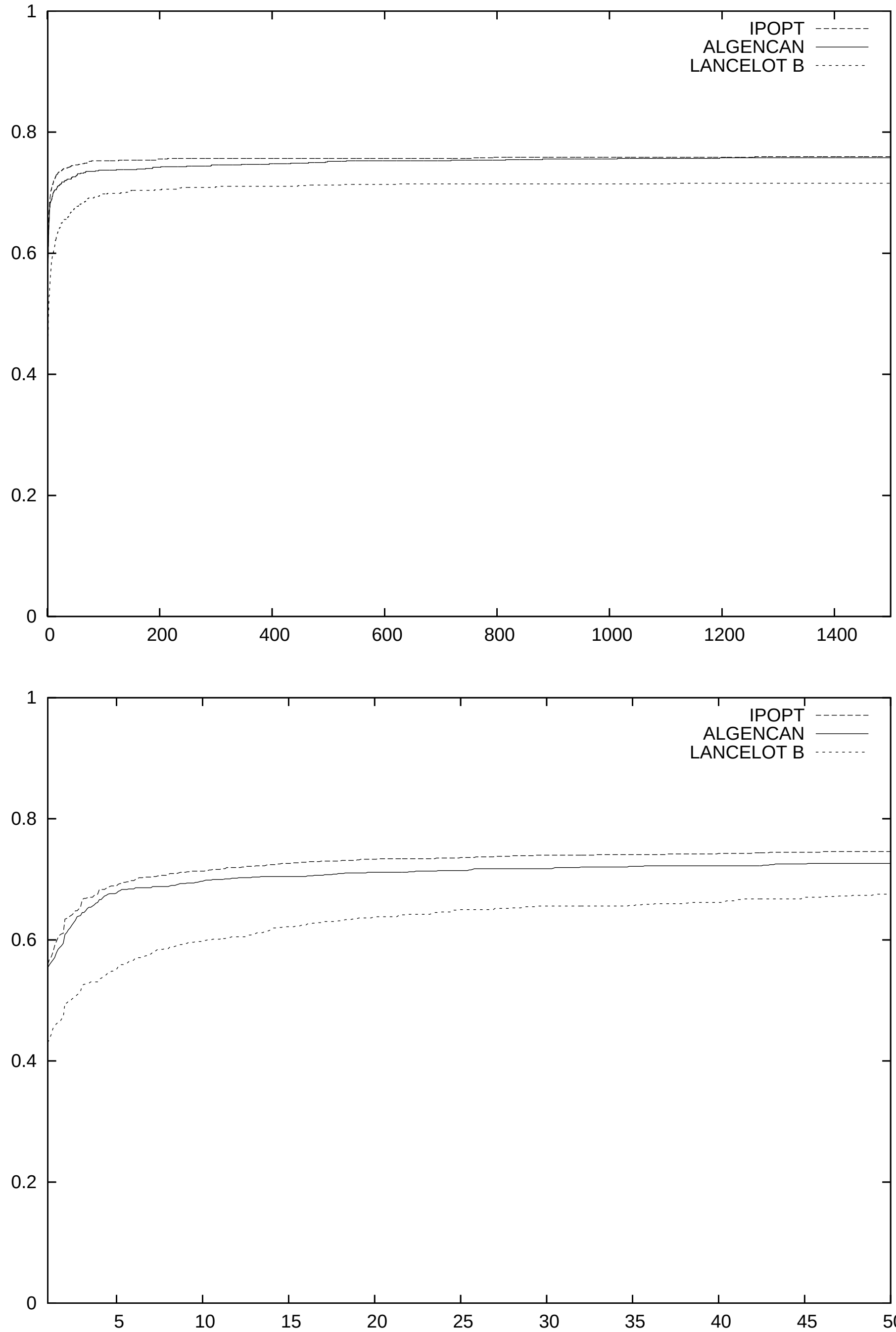

FIG. 6.3. *Performance profiles of* ALGENCAN, LANCELOT *B and* IPOPT *in the problems of the* CUTEr *collection. Note that there is a CPU time limit of 5 minutes for each pair method/problem. The second graphic is a zoom of the left-hand side of the first one. Although in the* CUTEr *test set of problems* ALGENCAN *with $\rho_1 = 10$ performs better (see Table 6) than using the choice of ρ_1 stated at the begining of this section, we used the last option (which is the* ALGENCAN *default option) to build the performance profiles curves in this graphics.*