Practical Augmented Lagrangian Methods

E. G. Birgin [*]   J. M. Martínez [†]

June 22, 2007

Keywords: Augmented Lagrangian, nonlinear programming, algorithms, global convergence, constraint qualifications, deterministic global optimization.

MSC2000: 90C30, 49M30, 49M37, 65K05.

1   Introduction

We are concerned with Nonlinear Programming problems defined in the following way:

$$
\begin{aligned}
\text{Minimize} \quad & f(x) \\
\text{subject to} \quad & h(x) = 0 \\
& g(x) \leq 0 \\
& x \in \Omega,
\end{aligned}
\tag{1}
$$

where $h : \mathbb{R}^n \to \mathbb{R}^m, g : \mathbb{R}^n \to \mathbb{R}^p, f : \mathbb{R}^n \to \mathbb{R}$ are continuous and $\Omega \subset \mathbb{R}^n$ is a closed set. From now on $\| \cdot \|$ represents the Euclidean norm and v_+ means $\max\{0, v\}$. The set $\mathbb{R}_+$ will be the set of nonnegative real numbers.

The Powell-Hestenes-Rockafellar (PHR) Augmented Lagrangian [42, 54, 56] is given by:

$$
L_\rho(x, \lambda, \mu) = f(x) + \frac{\rho}{2} \left\{ \sum_{i=1}^{m} \left[h_i(x) + \frac{\lambda_i}{\rho} \right]^2 + \sum_{i=1}^{p} \left[\left(g_i(x) + \frac{\mu_i}{\rho} \right)_+ \right]^2 \right\}
\tag{2}
$$

for all $x \in \mathbb{R}^n, \lambda \in \mathbb{R}^m, \mu \in \mathbb{R}^p_+$.

PHR-based Augmented Lagrangian methods for solving (1) are based on the iterative (approximate) minimization of L_ρ with respect to $x \in \Omega$, followed by the updating of the penalty parameter ρ and the Lagrange multipliers approximations λ and μ. The most popular practical Augmented Lagrangian method gave rise to the LANCELOT package [24, 25, 26]. LANCELOT does not use inequality constraints $g(x) \leq 0$ in its problem formulation. When an inequality

constraint $g_i(x) \leq 0$ appears in a practical problem, it is replaced by $g_i(x) + s_i = 0, s_i \geq 0$. The convergence of the LANCELOT algorithm to KKT points was proved in [24] using regularity assumptions. Under weaker assumptions that involve the Constant Positive Linear Dependence (CPLD) constraint qualification [4, 55], KKT-convergence was proved in [2] for a variation of the LANCELOT method. In the original LANCELOT method Ω was a box. A generalization where Ω is a polytope may be found in [23].

The motivation of (2) comes from the classical External Penalty method [27, 34, 36]. In this method one minimizes the function given by

$$\Phi_\rho(x) = f(x) + \frac{\rho}{2}\left[\sum_{i=1}^{m} h_i(x)^2 + \sum_{i=1}^{p}[g_i(x)_+]^2\right] \tag{3}$$

for successive values of ρ that tend to infinity. If, after minimizing (3) for a given ρ, a satisfactory feasibility is not achieved, the External Penalty philosophy leads to increase the value of ρ. If ρ is very large, the problem of minimizing P_ρ may become very difficult for ordinary minimization solvers.

The Augmented Lagrangian philosophy is different. Assume that the result of minimizing P_ρ, for a given ρ is not satisfactory, in terms of feasibility. Then, instead of increasing ρ (or, perhaps, besides increasing ρ) one modifies the origin with respect to which infeasibility is penalized. For example, suppose that, after the minimization of P_ρ we obtain x such that $h_i(x) = c \neq 0$. A common sense conjecture would be that this "disappointing" result was obtained because we punished the deviation of $h_i(x)$ with respect to 0, whereas the correct strategy would be to punish the deviation with respect to $-c$. This leads to the Shifted Penalty idea, in which, instead of P_ρ, one uses the Shifted Penalty function:

$$\Phi_\rho(x, c, d) = f(x) + \frac{\rho}{2}\left[\sum_{i=1}^{m}(h_i(x) + c_i)^2 + \sum_{i=1}^{p}[(g_i(x) + d_i)_+]^2\right]. \tag{4}$$

Writing $c_i = \lambda_i/\rho$ and $d_i = \mu_i/\rho$ we observe that the Shifted Penalty strategy coincides with the Augmented Lagrangian one. The naive modification of the shifts c_i, d_i sketched above gives rise to the best known (first-order) updating formulae for the Lagrange multipliers in the Augmented Lagrangian method. It is interesting to observe that this intuitive reasoning is independent of the smoothness of f, h and g. In this article we give preference to matrix-free updating procedures, which excludes the consideration of higher order estimates [28, 35].

In [3] a new PHR-like algorithm was introduced that does not use slack variables to complete inequality constraints and admits general constraints in the lower-level set Ω. In the box-constraint case, subproblems are solved using a technique introduced in [13], which improves the GENCAN algorithm [12]. CPLD-based convergence and penalty-parameter boundedness were also proved in [3] under suitable conditions on the problem.

In addition to its intrinsic adaptability to the case in which arbitrary constraints are included in Ω, the following positive characteristics of the Augmented Lagrangian approach for solving (1) must be mentioned:

1. Augmented Lagrangian methods proceed by sequential resolution of simple problems. Progress in the analysis and implementation of simple-problem optimization procedures

produces an almost immediate positive effect on the effectiveness of Augmented Lagrangian algorithms. Box-constrained optimization is a dynamic area of practical optimization from which we can expect Augmented Lagrangian improvements.

2. Global minimization of the subproblems implies convergence to global minimizers of the Augmented Lagrangian method [11]. There is a large field for research on global optimization methods for box-constraint optimization. When the global box-constraint optimization problem is satisfactorily solved in practice, the effect on the associated Augmented Lagrangian method for Nonlinear Programming problem is immediate.

3. Most box-constrained optimization methods are guaranteed to find stationary points. In practice, good methods do more than that. The line-search procedures of [12], for example, include extrapolation steps that are not necessary at all from the point of view of KKT convergence. However, they enhance the probability of convergence to global minimizers. As a consequence, the probability of convergence to Nonlinear Programming global minimizers of a practical Augmented Lagrangian method is enhanced too.

4. The global convergence theory of Augmented Lagrangian methods [11] does not need differentiability of the functions that define the Nonlinear Programming problem. In practice, this indicates that the Augmented Lagrangian approach may be successful in situations were smoothness is dubious.

5. The Augmented Lagrangian approach can be adapted to the situation in which analytic derivatives are not computed. See [47] for a derivative-free version of LANCELOT.

6. In many practical problems the Hessian of the Lagrangian is structurally dense (in the sense that any entry may be different from zero at different points) but generally sparse (given a specific point in the domain, the particular Lagrangian Hessian is a sparse matrix). As an example of this situation, consider the following formulation [18, 19] of the problem of fitting circles of radii r within a circle of radius R without overlapping:

$$\text{Min} \sum_{i<j} \max\{0, 4r^2 - \|p_i - p_j\|_2^2\}^2 \text{ subject to } \|p_i\|_2^2 \leq (R - r)^2.$$

The Hessian of the objective function is structurally dense but sparse at any point such that points p_i are "well distributed" within the big circle, since only pairs of overlapping small circles appear in the expression of the objective function. Newtonian methods usually have difficulties with this situation, both in terms of memory and computer time since the sparsity pattern of the matrix changes from iteration to iteration. This difficulty is almost irrelevant for the Augmented Lagrangian approach if one uses a low-memory box-constraint solver.

7. Independently of the Lagrangian Hessian density, the structure of the KKT system may be very poor for sparse factorizations. This is a serious difficulty for Newton-based methods but not for suitable implementations of the Augmented Lagrangian PHR algorithm.

3

8. If the Nonlinear Programming problem has many inequality constraints the usual slack-variable approach of Interior-Point methods (also used in [2, 24]) may be inconvenient. There are several approaches to reduce the effect of the presence of many slacks, but they may not be as effective as not using slacks at all.

9. Huge problems have obvious inconvenients in terms of storage requirements. The Augmented Lagrangian approach provides a radical remedy: problem data may be computed "on the flight", used when required in the subproblem, and not stored at all. This is not possible if one uses matricial approaches, independently of the sparsity strategy adopted.

10. If, at the solution of the problem, some strong constraint qualification fails to hold, the performance of Newton-like algorithms could be severely affected. The Augmented Lagrangian is not as sensitive to this type of inconvenient.

The amount of research dedicated to Augmented Lagrangian methods decreased in the 21-th century. Modern methods, based on interior-point techniques, sequential quadratic programming, trust regions, restoration, nonmonotone strategies and advanced sparse linear algebra procedures attracted much more attention.

A theoretical reason, and its practical consequence, seems to be behind this switch of interest. Roughly speaking, under suitable assumptions, Interior-Point Newtonian techniques converge quadratically (or, at least, superlinearly) whereas practical Augmented Lagrangian generally converge only linearly. Therefore, if both methods converge to the same point, and the precision required is strict enough, an Interior-Point Newtonian method will require less computer time than an Augmented Lagrangian method, independently of the work per iteration. Several attempts have been made to alleviate both the slow-convergence behavior as the ill-conditioning of the subproblems [14, 21, 33, 32, 39, 49]. Behind these attempts is the fact that the optimality conditions of the Augmented Lagrangian (and Penalty) subproblems may be decomposed in such a way that, for ρ large, resemble the KKT conditions of the original problem. This fact may be exploited in several ways and makes it possible that good implementations of the Augmented Lagrangian method be quite competitive with Interior-Point Newtonian techniques, even when high precision is the main requirement at the solution.

The general form of the Augmented Lagrangian method based on the PHR formula considered in this article is the following.

Algorithm 1.1

Let $\lambda_{\min} < \lambda_{\max}$, $\mu_{\max} > 0$, $\gamma > 1$, $0 < \tau < 1$. Let $\{\varepsilon_k\}$ be a sequence of nonnegative numbers such that $\lim_{k \to \infty} \varepsilon_k = 0$. Let $\lambda_i^1 \in [\lambda_{\min}, \lambda_{\max}]$, $i = 1, \ldots, m$, $\mu_i^1 \in [0, \mu_{\max}]$, $i = 1, \ldots, p$, and $\rho_1 > 0$. Initialize $k \leftarrow 1$.

Step 1. *Solving the subproblem.*

Compute $x^k \in \mathbb{R}^n$ an approximate solution of

$$\text{Minimize } L_{\rho_k}(x, \lambda^k, \mu^k) \text{ subject to } x \in \Omega. \tag{5}$$

Step 2. Define

$$V_i^k = \max\left\{g_i(x^k), -\frac{\mu_i^k}{\rho_k}\right\}, i = 1, \ldots, p.$$

If $k = 1$ or

$$\max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\} \leq \tau \, \max\{\|h(x^{k-1})\|_\infty, \|V^{k-1}\|_\infty\}, \tag{6}$$

define $\rho_{k+1} = \rho_k$. Otherwise, define $\rho_{k+1} = \gamma\rho_k$.

Step 3. Compute $\lambda_i^{k+1} \in [\lambda_{\min}, \lambda_{\max}], i = 1, \ldots, m$ and $\mu_i^{k+1} \in [0, \mu_{\max}], i = 1, \ldots, p$. Set $k \leftarrow k + 1$ and go to Step 1.

In the practical implementation of Algorithm 1.1, we will compute $\lambda_i^{k+1} = \min\{\max\{\lambda_{\min}, \lambda_i^k + \rho_k h_i(x^k)\}, \lambda_{\max}\}$ and $\mu_i^{k+1} = \min\{\max\{0, \mu_i^k + \rho_k g_i(x^k)\}, \mu_{\max}\}$. These are safeguarded first-order estimations of the Lagrange multipliers. The safeguards defined by $\lambda_{\min}, \lambda_{\max}$ and $\mu_{\max}$ are necessary to prove the global convergence results. Some authors prefer to define Augmented Lagrangian algorithms without safeguards for the Lagrange multipliers [9, 26]. However, boundedness of the multiplier estimates are necessary to prove the main convergence results and, if this boundedness is not algorithmically forced, it may be guaranteed only by means of strong problem assumptions.

Different Augmented Lagrangian algorithms will differ only in Step 1. In each case we will need a precise definition for the approximate solution of (5).

2 Cases

2.1 Augmented Lagrangians and Global Optimization

In this section we will only assume continuity of the functions f, h and g. Throughout the section we will assume that a global minimizer of (1) exists. Several versions of the Augmented Lagrangian method generate sequences that converge to global minimizers, provided that global minimizers of the subproblems are available. This property is inherited from the analogous property of the External Penalty method. A practical consequence of this property is the fact that Augmented Lagrangian methods tend to find feasible points with lower objective function values than other nonlinear programming solvers, when clever agressive algorithms are used for solving the subproblems.

Here we will assume we are able to find an approximate global minimizer defined by the tolerance ε_k. At each iteration, x^k will belong to $\Omega \cap P_k$, where P_k is an auxiliary set to which a global minimizer of (1) necessarily belongs. For example, P_k may be a set that contains the feasible region of (1). The presence of the constraints defined by P_k helps in the global resolution of the subproblems. Obviously, in the absence of algorithmic advantages, P_k may be defined as being $\mathbb{R}^n$. Algorithm 2.1 will be Algorithm 1.1, where Step 1 is defined as follows.

Step 1. Let $P_k \subset \mathbb{R}^n$ be a closed set such that a global minimizer z (the same for all k) belongs to P_k. Find an ε_k-global minimizer x^k of the problem Min $L_{\rho_k}(x, \lambda^k, \mu^k)$ subject to $x \in \Omega \cap P_k$. That is $x^k \in \Omega \cap P_k$ is such that:

$$L_{\rho_k}(x^k, \lambda^k, \mu^k) \leq L_{\rho_k}(x, \lambda^k, \mu^k) + \varepsilon_k \tag{7}$$

for all $x \in \Omega \cap P_k$. The ε_k-global minimum can be obtained using a deterministic global optimization approach, such as the αBB method [37].

In most deterministic global optimization methods for solving (7) the point x^{k-1} is not used as "initial approximation" as most local optimization solvers do. In fact, the concept of "initial point" has no meaning at all in this case. The information used by the Outer iteration k is the set of approximate Lagrange multipliers computed after iteration $k-1$, and nothing else.

Theorems 2.1 [11] is the main convergence result related to Algorithm 2.1. Limit points of sequences generated by this algorithm are feasible global minimizers.

Theorem 2.1. *Assume that the sequence $\{x^k\}$ is well defined and admits a limit point x^*. Then, x^* is a global minimizer of (1). If, instead of $\varepsilon_k \to 0$ we assume only that $\varepsilon_k \to \varepsilon \geq 0$, x^* will be feasible and $f(x^*) \leq f(x) + \varepsilon$ for all feasible x.*

The problem of finding $x^k \in \Omega \cap P_k$ satisfying (7) consists of finding an ε_k-global solution of the problem:
$$\text{Minimize } L_{\rho_k}(x, \lambda^k, \mu^k) \text{ subject to } x \in \Omega \cap P_k. \tag{8}$$
When Ω and P_k are defined by linear equality and/or inequality constraints and f, h, g admit continuous second derivatives, this problem has been solved in [11] using the αBB algorithm [37].

The practical results presented in [11] corroborate the theory and give hints on the effectivity of the Augmented Lagrangian method for global optimization.

2.2 Augmented Lagrangian Algorithm with arbitrary lower-level constraints

In [2, 3] safeguarded Augmented Lagrangian methods were defined that converge to KKT points under the CPLD constraint qualification and exhibit good properties in terms of penalty parameter boundedness. ALGENCAN, which is publicly available in the TANGO Project web page *http://www.ime.usp.br/~egbirgin/tango/*, is the application of the main algorithm in [3] to problem (1).

In this section we will assume that f, h, g admit continuous first (and, sometimes, second) derivatives. Observe that the function L_ρ, defined in (2) has continuous first derivatives with respect to x, but second derivatives are not defined at the points such that $g_i(x) + \mu_i/\rho = 0$.

In Algorithm 3.1 we will assume that the set Ω is defined by

$$\Omega = \{x \in I\!\!R^n \mid \underline{h}(x) = 0, \underline{g}(x) \leq 0\}, \tag{9}$$

where $\underline{h} : I\!\!R^n \to I\!\!R^{\underline{m}}, \underline{g} : I\!\!R^n \to I\!\!R^{\underline{p}}$ are as smooth as necessary. The constraints defined by Ω are called *lower-level* constraints. Algorithm 3.1 is identical to Algorithm 1.1 except at Step 1. The subproblem resolution at Algorithm 3.1 is as given below.

Step 1. Compute (if possible) $x^k \in \mathbb{R}^n$ such that there exist $v^k \in \mathbb{R}^{\underline{m}}, u^k \in \mathbb{R}^{\underline{p}}$ satisfying

$$\|\nabla L_{\rho_k}(x^k, \lambda^k, \mu^k) + \sum_{i=1}^{\underline{m}} v_i^k \nabla \underline{h}_i(x^k) + \sum_{i=1}^{\underline{p}} u_i^k \nabla \underline{g}_i(x^k)\| \leq \varepsilon_k, \tag{10}$$

$$u_i^k \geq 0, \underline{g}_i(x^k) \leq \varepsilon_k \text{ for all } i = 1, \ldots, \underline{p}, \tag{11}$$

$$\underline{g}_i(x^k) < -\varepsilon_k \Rightarrow u_i^k = 0 \text{ for all } i = 1, \ldots, \underline{p}, \tag{12}$$

$$\|\underline{h}(x^k)\| \leq \varepsilon_k. \tag{13}$$

$$\tag{14}$$

The conditions (10–13) are approximate KKT conditions for the minimization of L_{ρ_k} subject to the lower level constraints. If $\Omega = \mathbb{R}^n$ these conditions reduce to $\|\nabla L_{\rho_k}(x^k, \lambda^k, \mu^k)\| \leq \varepsilon_k$.

The CPLD (Constant Positive Linear Dependence) condition defined by Qi and Wei [55] is a crucial tool in the convergence theory of Algorithm 3.1. In [4] it has been proved that CPLD is a constraint qualification and its relation with other constraint qualifications have been reported.

A First-Order Constraint Qualification is a property of feasible points of a Nonlinear Programming problem such that, when verified at a local minimizer, implies that the local minimizer is a KKT point. The Linear-Independence Constraint Qualification (LICQ), also called *regularity*, says that the gradients of the active constraints at the feasible point x are linearly independent.

Assume that the feasible set of a nonlinear programming problem is given by $\overline{h}(x) = 0, \overline{g}(x) \leq 0$, where $\overline{h} : \mathbb{R}^n \to \mathbb{R}^{\overline{m}}$ and $\overline{g} : \mathbb{R}^n \to \mathbb{R}^{\overline{p}}$. Let $I(x) \subset \{1, \ldots, \overline{p}\}$ be the set of indices of the active inequality constraints at the feasible point x. Let $I_1 \subset \{1, \ldots, \overline{m}\}$, $I_2 \subset I(x)$. The subset of gradients of active constraints that correspond to the indices $I_1 \cup I_2$ is said to be *positively linearly dependent* if there exist multipliers λ, μ such that

$$\sum_{i \in I_1} \lambda_i \nabla \overline{h}_i(x) + \sum_{i \in I_2} \mu_i \nabla \overline{g}_i(x) = 0, \tag{15}$$

with $\mu_i \geq 0$ for all $i \in I_2$ and $\sum_{i \in I_1} |\lambda_i| + \sum_{i \in I_2} \mu_i > 0$. Otherwise, we say that these gradients are *positively linearly independent*.

The Mangasarian-Fromovitz Constraint Qualification MFCQ says that, at the feasible point x, the gradients of the active constraints are positively linearly independent [48, 57].

The CPLD Constraint Qualification says that, if a subset of gradients of active constraints is positively linearly dependent at the feasible point x (i.e. (15) holds), then there exists $\delta > 0$ such that the vectors

$$\{\nabla \overline{h}_i(y)\}_{i \in I_1}, \{\nabla \overline{g}_j(y)\}_{j \in I_2}$$

are linearly dependent for all $y \in \mathbb{R}^n$ such that $\|y - x\| \leq \delta$.

The main convergence theorems related to Algorithm 3.1 were proved in [3]. Theorem 3.1 says that, if a limit point satisfies the CPLD condition with respect to the lower-level constraints, then this point is stationary with respect to a natural infeasibility measure. In other words, this theorem says that, if the limit point is not feasible, then (very likely) it is a local minimizer of the upper-level infeasibility, subject to lower-level feasibility.

Theorem 3.1. *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

$$\text{Minimize } \frac{1}{2}\left[\sum_{i=1}^{m} h_i(x)^2 + \sum_{i=1}^{p}[g_i(x)_+]^2\right] \text{ subject to } x \in \Omega. \tag{16}$$

(ii) x^* does not satisfy the CPLD constraint qualification associated with Ω.

From the point of view of optimality, we are interested in the status of feasible limit points. Theorem 3.2 says 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 Theorem 3.2 is stronger than results where KKT conditions are proved under MF or regularity assumptions.

Theorem 3.2. *Let $\{x^k\}$ be a sequence generated by Algorithm 3.1. Assume that x^* is a feasible limit point of (1)-(9) that satisfies the CPLD constraint qualification related to set of all the constraints. Then, x^* is a KKT point of the problem (1)-(9).*

Theorems 3.1 and 3.2 are interesting and useful but they do not explain why it is better to use the Augmented Lagrangian instead of a pure penalty method. In fact, if we define $\lambda^k = 0, \mu^k = 0$ for all k these two theorems remain valid and we are in presence of a variation of the External Penalty method. The use of Lagrange multipliers estimates is justified in Theorem 3.3, which is also proved in [3]. Theorem 3.3 says that, under appropriate conditions, the sequence of penalty parameters $\{\rho_k\}$ do not tend to infinity. In practice, this means that the minimization subproblems tend to remain well-conditioned and that minimization algorithms for solving the subproblems will not face difficulties associated to very large values of ρ_k.

Theorem 3.3. *Assume that:*

1. *The sequence $\{x^k\}$ is generated by the application of Algorithm 3.1 to problem (1)-(9) and $\lim_{k\to\infty} x^k = x^*$.*

2. *In Algorithm 3.1 we use the updating rules $\lambda_i^{k+1} = \max\{\lambda_{\min}, \min\{\lambda_i^k + \rho_k h_i(x^k), \lambda_{\max}\}\}$ and $\mu_i^{k+1} = \max\{0, \min\{\mu^k + \rho_k g_i(x^k), \mu_{\max}\}\}$.*

3. *The point x^* is feasible ($h(x^*) = 0$, $\underline{h}(x^*) = 0$, $g(x^*) \leq 0$ and $\underline{g}(x^*) \leq 0$).*

4. *The gradients of the active constraints at x^* are linearly independent. The associated (unique) Lagrange multipliers are $\lambda^*, \mu^*, u^*, v^*$.*

5. *The functions $f, h, g, \underline{h}$ and $\underline{g}$ admit continuous second derivatives in a neighborhood of x^*.*

6. *Define the tangent subspace T as the set of all $z \in I\!R^n$ such that $\nabla h(x^*)^T z = \nabla \underline{h}(x^*)^T z = 0$, $\nabla g_i(x^*)^T z = 0$ for all i such that $g_i(x^*) = 0$ and $\nabla \underline{g}_i(x^*)^T z = 0$ for all i such that $\underline{g}_i(x^*) = 0$. Then, for all $z \in T, z \neq 0$,*

$$z^T[\nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(x^*) + \sum_{i=1}^p \mu_i^* \nabla^2 g_i(x^*) + \sum_{i=1}^m v_i^* \nabla^2 \underline{h}_i(x^*) + \sum_{i=1}^p u_i^* \nabla^2 g_i(x^*)]z > 0.$$

7. *For all $i = 1, \ldots, m, j = 1, \ldots, p$, $\lambda_i^* \in (\lambda_{\min}, \lambda_{\max})$, $\mu_j^* \in [0, \mu_{\max})$.*

8. *For all i such that $g_i(x^*) = 0$, we have $\mu_i^* > 0$. (Strict complementarity in the upper level.)*

9. *There exists a sequence $\eta_k \to 0$ such that $\varepsilon_k \leq \eta_k \max\{\|h(x^k)\|, \|V^k\|\}$ for all $k = 0, 1, 2 \ldots$.*

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

Observe that strict complementarity is imposed only to the constraints in the upper-level set. In the lower-level set it is admissible that $\underline{g}_i(x^*) = u_i^* = 0$. Observe, too, that the assumption on the reduced positive definiteness on the Hessian of the Lagrangian is weaker than the usual second-order sufficiency assumption [36], 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, this is not a second-order sufficiency assumption for local minimizers. It holds for the problem of minimizing $x_1 x_2$ subject to $x_2 - x_1 \leq 0$ at $(0,0)$ although $(0,0)$ is not a local minimizer of this problem.

The last hypothesis of Theorem 3.3 imposes that the precision in which subproblems are solved should tend to zero faster than the measure of infeasibility-noncomplementarity. Some authors [30, 31, 40, 41], in slightly different contexts, also used convergence tolerances that depend on the degree of infeasibility of the current inner iterate.

The Augmented Lagrangian method is the only efficient nonlinear programming algorithm that can take obvious advantage of the existence of case-oriented optimization solvers for problems whose constraints are a subset of the original problem constraints. The partition of the constraints in "easy" and "complicate" is very common in Engineering applications. In the Augmented Lagrangian framework, easy constraints go to the lower level and complicate constraints contribute to the Augmented Lagrangian function. The most common situation is the one in which lower level constraints are linear. Location problems of this type are described in [3]. Problems 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 through the TANGO Project web page *http://www.ime.usp.br/~egbirgin/tango/*. The key for the efficiency of the Augmented Lagrangian method in these problems is the use of the Spectral Projected Gradient method [15, 16, 17] for solving the subproblems.

2.3 Lower level box constraints

In most applications, the definition of the lower level set Ω in (1) is:

$$\Omega = \{x \in I\!R^n \mid a \leq x \leq b\}, \tag{17}$$

where $a, b \in I\!R^n, a \leq b$. In other words, Ω is an n-dimensional box. By the continuity of the Augmented Lagrangian function and the compactness of Ω, this definition guarantees that a

global minimizer of the subproblem exists. Many times one adds bound constraints in the lower level of a nonlinear programming problem in order to guarantee solubility of the subproblems and boundedness of the sequence $\{x^k\}$.

Obviously, the constraints (17) may be written in the form (9) and, so, Algorithm 3.1 may be applied and Theorems 3.1, 3.2 and 3.3 hold. However, many specific algorithms for box-constrained optimization exist that use stronger convergence criteria than the one given in (10-13). Namely, in box-constrained minimization one usually declares convergence when

$$x^k \in \Omega \ \text{ and } \ \|P_\Omega(x^k - \nabla L_{\rho_k}(x^k, \lambda^k, \mu^k))\|_k \leq \varepsilon_k, \tag{18}$$

where P_Ω denotes de Euclidean projection on Ω. The condition (18) implies (10-13). This leads us to define Algorithm 4.1 as Algorithm 1.1 where Step 1 is defined by (18). Theorems 3.1, 3.2 and 3.3 obviously apply to Algorithm 4.1. It must be observed, however, that, since all the points of Ω satisfy CPLD, Theorem 3.1 guarantees that the limit points of the generated sequence $\{x^k\}$ are KKT points of $\sum_{i=1}^m h_i(x)^2 + \sum_{i=1}^p [g_i(x)_+]^2$ subject to $x \in \Omega$.

Algorithm 4.1, with the subproblems solved by the box-constraint solver GENCAN [12], with the modifications introduced in [13], is called ALGENCAN. The code that implements ALGENCAN is free for download in the TANGO Project web page *http://www.ime.usp.br/~egbirgin/tango/*. It is written in Fortran 77 (double precision) and interfaces with AMPL, CUTEr, C/C++, Python and R (language and environment for statistical computing) are available.

The default version of ALGENCAN uses $\tau = 0.5$, $\gamma = 10$, $\lambda_{\min} = -10^{20}$, $\mu_{\max} = \lambda_{\max} = 10^{20}$, $\varepsilon_k = 10^{-4}$ for all k, $\lambda_1 = 0, \mu_1 = 0$ and $\rho_1 = \max\left\{10^{-6}, \min\left\{10, \frac{2|f(x^0)|}{\|h(x^0)\|^2 + \|g(x^0)_+\|^2}\right\}\right\}$. The default convergence criterion is $\max\{\|h(x^k)\|_\infty, \|V^k\|_\infty\} \leq 10^{-4}$. The condition $\|V^k\|_\infty \leq 10^{-4}$ guarantees that, for all $i = 1, \ldots, p$, $g_i(x^k) \leq 10^{-4}$ and that $(\mu_i^k + \rho_k g_i(x^k))_+ = 0$ 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 (18) and the choice of ε_k.

The celebrated package LANCELOT also solves the basic Nonlinear Programming problem with box constraints, but each inequality constraint is completed with a slack variable to become an equality constraint plus a lower-level bound.

A comparison between the default versions of ALGENCAN and LANCELOT B using all the (1023) problems of the CUTEr collection was reported in [3]. The executions were stopped when the precision 10^{-4} was achieved or when the allowed CPU time (5 minutes on an 1.8GHz AMD Opteron 244 processor, 2Gb of RAM memory and Linux operating system) was exhausted. Codes are in Fortran 77 and the compiler option "-O" was adopted.

Given a fixed problem, for each method $M \in \{$LANCELOT, ALGENCAN $\}$, $x_{\text{final}}^{\text{M}}$ was defined in [3] as the final point obtained by M when solving the given problem. The point $x_{\text{final}}^{\text{M}}$ is considered to be *feasible* if

$$\max\{\|h(x_{\text{final}}^{\text{M}})\|_\infty, \|g(x_{\text{final}}^{\text{M}})_+\|_\infty\} \leq 10^{-4}.$$

Define

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

It is said that the method M *found a solution* of the problem if $x_{\text{final}}^{\text{M}}$ is feasible and

$$f(x_{\text{final}}^{\text{M}}) \leq f_{\text{best}} + 10^{-3}|f_{\text{best}}| + 10^{-6} \ \text{ or } \ \max\{f_{\text{best}}, f(x_{\text{final}}^{\text{M}})\} \leq -10^{20}.$$

10

	ALGENCAN	LANCELOT B
Efficiency	692	544
Robustness	809	763

Table 1: Efficiency means number of times that method M obtained the best r^M. Robustness means the number of times in which $r^M < \infty$.

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

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

The quantity r was used as performance measurement in [3]. The results of the comparison are reported in the form of performance profiles [29] and a small numerical table. See Figure 1 and Table 1. [1]

2.4 Alternative Augmented Lagrangians

The main drawback of the PHR formula (2) is that the objective function of the subproblems is not twice continuously differentiable. This is the main motivation for the introduction of many alternative Augmented Lagrangian methods. See, for example, [1, 5, 6, 7, 8, 22, 38, 43, 45, 46, 50, 51, 52, 53, 58]. Most of them have interesting interpretations as proximal point methods for solving the dual problem, when the original nonlinear programming problem is convex [44]. In [10] a comparison of many different Augmented Lagrangian formulae within an algorithmic framework similar to the one of Algorithm 1.1 has been performed using the CUTE collection [20]. In general, the PHR formula seems to be more efficient than the alternative ones for the resolution of the selected problems.

References

[1] R. R. Allran and S. E. J. Johnsen, An algorithm for solving nonlinear programming problems subject to nonlinear inequality constraints, *The Computer Journal* 13, pp. 171–177, 1970.

[2] R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt, Augmented Lagrangian methods under the Constant Positive Linear Dependence constraint qualification, *Mathematical Programming*, to appear (DOI: 10.1007/s10107-006-0077-1).

---

[1] 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 was 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 (less than 2% of the total number of problems), LANCELOT B seems to stop prematurely. This easy-to-solve inconvenient may slightly deteriorate the robustness of LANCELOT B reported here.

[3] R. Andreani, E. G. Birgin, J. M. Martínez and M. L. Schuverdt, On Augmented Lagrangian methods with general lower-level constraints, *SIAM Journal on Optimization*, to appear.

[4] 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, pp. 473–485, 2005.

[5] A. Auslender, M. Teboulle and S. Ben-Tiba, Interior proximal and multiplier methods based on second order homogeneous kernels, *Mathematics of Operations Research* 24, pp. 645–668, 1999.

[6] A. Ben Tal, I. Yuzefovich and M. Zibulevsky, Penalty/barrier multiplier methods for minimax and constrained smooth convex programs, *Research Report 9/92*, Optimization Laboratory, Faculty of Industrial Engineering Management, Technion, Haifa, Israel, 1992.

[7] A. Ben-Tal and M. Zibulevsky, Penalty/barrier multiplier methods for convex programming problems, *SIAM Journal on Optimization* 7, pp. 347–366, 1997.

[8] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods*, Academic Press, New York, 1982.

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

[10] E. G. Birgin, R. Castillo and J. M. Martínez, Numerical comparison of Augmented Lagrangian algorithms for nonconvex problems, *Computational Optimization and Applications* 31, pp. 31–56, 2005.

[11] E. G. Birgin, C. A. Floudas and J. M. Martínez, Global minimization using an Augmented Lagrangian method with variable lower-level constraints, available in Optimization Online, E-Print ID: 2006-12-1544, *http://www.optimization-online.org/DB_HTML/2006/12/1544.html*.

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

[13] E. G. Birgin and J. M. Martínez, Structured minimal-memory inexact quasi-Newton method and secant preconditioners for Augmented Lagrangian Optimization, *Computational Optimization and Applications*, to appear (DOI: 10.1007/s10589-007-9050-z).

[14] E. G. Birgin and J. M. Martínez, Modified Augmented Lagrangian algorithm with fast local convergence, *Optimization Methods & Software*, to appear.

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

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

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

[18] E. G. Birgin, J. M. Martínez and D. P. Ronconi, Optimizing the Packing of Cylinders into a Rectangular Container: A Nonlinear Approach, *European Journal of Operational Research* 160, pp. 19–33, 2005.

[19] E. G. Birgin and F. N. C. Sobral, Minimizing the object dimensions in circle and sphere packing problems, *Computers & Operations Research*, to appear. (DOI: 10.1016/j.cor.2006.11.002).

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

[21] C. G. Broyden and N. F. Attia, Penalty functions, Newton's method, and quadratic programming, *Journal of Optimization Theory and Applications* 58, pp. 377–385, 1988.

[22] R. A. Castillo, Métodos de Lagrangiano Aumentado usando penalidades generalizadas para programação não linear, *Tese de Doutorado*, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil, 1998.

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

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

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

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

[27] R. Courant, Variational methods for the solution of problems of equilibrium and vibration, *Bull. Amer. Math. Soc* 49, pp. 1–23, 1943.

[28] G. Di Pillo and L. Grippo, A new class of Augmented Lagrangians in nonlinear programming, *SIAM J. Control Optim.* 17, pp. 618-628, 1979.

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

[30] Z. Dostál, A. Friedlander and S. A. Santos, Augmented Lagrangians with adaptive precision control for quadratic programming with equality constraints, *Computational Optimization and Applications* 14, pp. 37–53, 1999.

[31] Z. Dostál, A. Friedlander and S. A. Santos, Augmented Lagrangians with adaptive precision control for quadratic programming with simple bounds and equality constraints, *SIAM Journal on Optimization* 13, pp. 1120–1140, 2002.

[32] J-P. Dussault, Numerical stability and efficiency of penalty algorithms, *SIAM Journal on Numerical Analysys* 32, pp. 296–317, 1995.

[33] L. Ferreira-Mendonça, V. L. R. Lopes and J. M. Martínez, Quasi-Newton acceleration for equality constrained minimization, *Computational Optimization and Applications*, to appear.

[34] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming,* John Wiley, 1968.

[35] R. Fletcher, An ideal penalty function for constrained optimization, *J. Inst. Maths. Applns.* 15, pp. 319-342, 1975.

[36] R. Fletcher, *Practical methods of Optimization*, John Wiley, 1987.

[37] C. A. Floudas, *Deterministc global optimization: theory, methods and application*, Kluwer Academic Publishers, 2000.

[38] C. C. Gonzaga and R. A. Castillo, A nonlinear programming algorithm based on non-coercive penalty functions, *Mathematical Programming* 96, pp. 87–101, 2003.

[39] N. I. M. Gould, On the convergence of a sequential penalty function method for constrained minimization, *SIAM Journal on Numerical Analysys* 26, pp. 107–128, 1989.

[40] W. W. Hager, Dual techniques for constrained optimization, *Journal of Optimization Theory and Applications* 55, pp. 33–72, 1987.

[41] W. W. Hager, Analysis and implementation of a dual algorithm for constrained optimization, *Journal of Optimization Theory and Applications* 79, pp. 427–462, 1993.

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

[43] C. Humes and P. S. Silva, Strict convex regularizations, proximal points and Augmented Lagrangians, *RAIRO Operations Research* 34, pp. 283–303, 2000.

[44] A. N. Iusem, Augmented Lagrangian methods and proximal point methods for convex optimization, *Investigación Operativa* 8, pp. 11–50, 1999.

[45] B. W. Kort and D. P. Bertsekas, Multiplier methods for convex programming, in *Proceedings of the IEEE Decision and Control Conference*, San Diego, CA, pp. 260–264, 1973.

[46] B. W. Kort and D. P. Bertsekas, Combined primal–dual and penalty methods for convex programming, *SIAM Journal on Control and Optimization* 14, pp. 268–294, 1976.

[47] R. M. Lewis and V. Torczon, A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds, *SIAM Journal on Optimization* 12, pp. 1075–1089, 2002.

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

[49] J. M. Martínez and L. T. Santos, Some new theoretical results on recursive quadratic programming algorithms, *Journal of Optimization Theory and Applications* 97, pp. 435–454, 1998.

[50] L. C. Matioli, Uma nova metodologia para construção de funções de penalização para algoritmos de Lagrangiano Aumentado, *Tese de Doutorado*, Universidade Federal de Santa Catarina, Florianópolis, Brasil, 2001.

[51] F. H. Murphy, A class of exponential penalty functions, *SIAM Journal on Control* 12, pp. 679–687, 1974.

[52] H. Nakayama , H. Samaya and Y. Sawaragi, A generalized Lagrangian function and multiplier method, *Journal of Optimization Theory and Applications* 17, pp. 211–227, 1975.

[53] R. A. Polyak, Log-sigmoid multiplier method in constrained optimization, *Annals of Operation Research* 101, pp. 427–460, 2001.

[54] 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.

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

[56] R. T. Rockafellar, Augmented Lagrange multiplier functions and duality in nonconvex programming, *SIAM Journal on Control* 12, pp. 268–285, 1974.

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

[58] P. Tseng and D. Bertsekas, On the convergence of the exponential multiplier method for convex programming, *Mathematical Programming* 17, pp. 670–690, 1993.
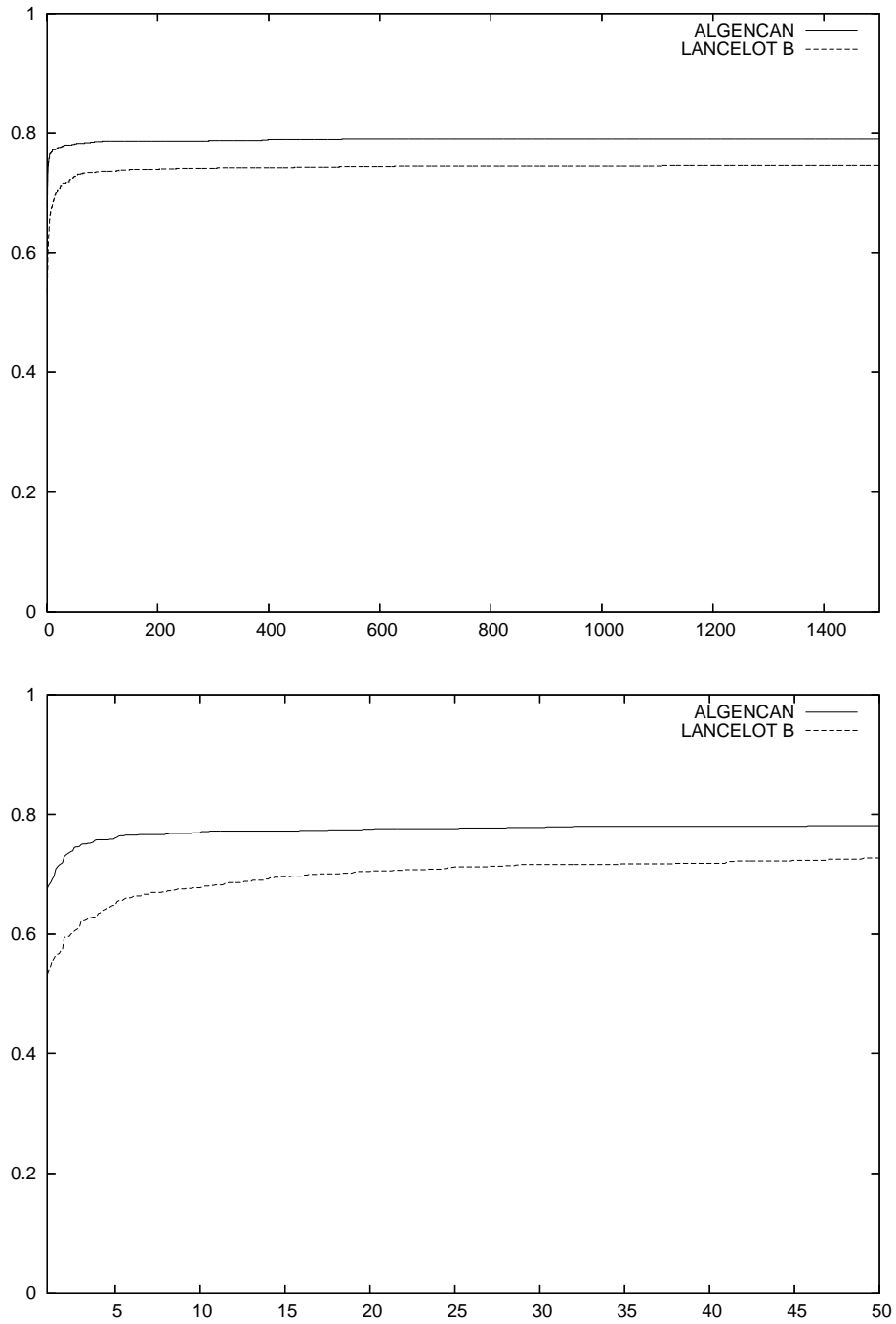
Figure 1: Performance profiles of ALGENCAN and LANCELOT B in the problems of the CUTEr collection. Note that there is a CPU time limit of 5 minutes for each pair method/problem. The second graphic is a zoom of the left-hand side of the first one.