Exact Penalty Methods with Constrained Subproblems*

Sílvia M. H. Janesch   *Lúcio Tunes Santos*

Department of Applied Mathematics, IMECC-UNICAMP
University of Campinas
CP 6065, 13081-970 Campinas SP, Brazil
janesch@ime.unicamp.br, lucio@ime.unicamp.br

Abstract
The objective of this paper is twofold. On one hand, the main results of the exact penalty methods for nonlinear programming that are useful for algorithmic purposes are surveyed and the proof of the main result is simplified. On the other hand, it is shown that the main theorem on this subject still holds when one includes only a subset of the constraints in the penalized objective function, so that suitable constrained subproblems are solved at each penalty step. The reasons why this is important from a practical (algorithmic) point of view are explained.

Keywords: Exact penalty function, Penalty methods, Nonlinear programming.

1 Introduction

The Nonlinear Programming problem (NLP) consists in the minimization of a general objective function subject to constraints that are usually given by a set of equalities and inequalities. Nevertheless, the nature of the constraints is not necessarily the same. Many times it is possible to take advantage of the structure of some equality and/or inequality constraints. For example, the geometry of simple bounds, polytopes and Euclidian balls can be conveniently exploited in the development of numerical algorithms for solving (NLP).

Exterior penalty methods, usually referred only as penalty methods, consist in solving (NLP) by means of the resolution of a sequence of subproblems with simple constraints (or unconstrained). The constraints which cannot be easily dealt with are placed into the objective function through a *penalty function*, so that any violation of

*The first author was supported by FAPESP (94/4721-1).
Received december 1996, revised version accepted june 1997.

such constraints is penalized. However, for exact penalty methods, the known result which ensures that the solution to the original problem is also a solution to the subproblems, is stated for unconstrained subproblems only. In this work, we extend such result to allow constrained subproblems in exact penalty methods. In consequence, numerical algorithms might be developed combining the advantages of exploiting the simple constrained set and the usage of exact penalty functions.

Consider the following general (NLP)

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & h(x) = 0 \\ & g(x) \leq 0 \\ & x \in \Omega\,, \end{array} \tag{1}$$

where Ω is a closed set defined by simple constraints, $f : I\!R^n \to I\!R$, $h : I\!R^n \to I\!R^{m_1}$, $g : I\!R^n \to I\!R^{m_2}$ and $f, h, g \in C^2$.

The penalty functions that appear more frequently in the literature are the Quadratic Loss Penalty Function $P_2(x,\rho) = f(x) + \rho\left(\|h(x)\|_2^2 + \|g^+(x)\|_2^2\right)$ and the L_1 Penalty Function $P_1(x,\rho) = f(x) + \rho\left(\|h(x)\|_1 + \|g^+(x)\|_1\right)$, where $\rho > 0$ is the penalty parameter and $g_j^+(x) = \max\{0, g_j(x)\}$. Usually, penalty methods generate a sequence of infeasible points and under very mild conditions any limit point is an optimal solution to the original problem. In most of the methods, the feasibility is achieved only at the solution. However, there are penalty functions that generate a sequence of points that may be interior or exterior to the feasible region (see [12, 24]).

In the theory developed for classical penalty methods the sequence of penalty parameters must be unbounded in order to guarantee global convergence. Computational difficulties of ill-conditioning appear when we have to solve penalty subproblems for large values of ρ. An alternative to remedy this situation is the approach discussed in Martínez & Santos [23] for $P_2(x,\rho)$. The main idea of their work is to determine a stable Newton direction, using a procedure that introduces new variables in the nonlinear system given by the first–order necessary conditions for a local minimizer of $P_2(x,\rho)$, i.e. $\nabla P_2(x,\rho) = 0$. Another weakness of the penalty methods is the assumption that global minimizers of the penalty subproblems have to be computed. In practice, we do not obtain global minimizers: the algorithms usually generate a sequence of local minimizers.

The idea of exact penalty methods is to solve (NLP) by means of a single unconstrained minimization problem. Roughly speaking, an exact penalty function for problem (1) is a function $P(x,\rho)$, where $\rho > 0$ is the penalty parameter, with the property that there exists a lower bound $\bar{\rho} > 0$ such that for $\rho > \bar{\rho}$ any local minimizer of (NLP) is also a local minimizer of the penalty subproblem. A more rigorous definition of an exact penalty function can be found in [8]. Exact penalty functions can be divided into two classes: continuously differentiable and nondifferentiable exact penalty functions. Continuously differentiable exact penalty functions were introduced by

Fletcher [13] for equality constrained problems and by Glad & Polak [16] for problems with inequality constraints; further contributions have been given in Di Pillo & Grippo [8, 9]. Nondifferentiable exact penalty functions were introduced by Zangwill [26] and Pietrzykowski [25], and a lot of research has been developed in this subject (see, e.g. [5, 6, 7, 10, 14]).

A class of nondifferentiable exact penalty functions associated to (1) for $\Omega = I\!R^n$ was analyzed by Charalambous [5] in 1978. It is given by

$$P_p(x,\alpha,\beta) = f(x) + \rho \left(\sum_{i=1}^{m_1} [\alpha_i |h_i(x)|]^p + \sum_{j=1}^{m_2} [\beta_j g_j^+(x)]^p \right)^{1/p} \tag{2}$$

where $p \geq 1$, $\alpha_i, \beta_j > 0$, $i = 1, \ldots, m_1$ and $j = 1, \ldots, m_2$. For $p = 1$ and considering all the penalty parameters equal to ρ, we have the L_1 Penalty Function, introduced by Pietrzykowski [25] in 1969,

$$P_1(x,\rho) = f(x) + \rho \left(\sum_{i=1}^{m_1} |h_i(x)| + \sum_{j=1}^{m_2} g_j^+(x) \right). \tag{3}$$

For $p = \infty$, we have

$$P_\infty(x,\rho) = f(x) + \max_{\substack{1 \leq i \leq m_1 \\ 1 \leq j \leq m_2}} \{\alpha_i |h_i(x)|, \beta_j g_j^+(x)\} \tag{4}$$

which is the exact Minimax Penalty Function of Bandler & Charalambous [1].

Pietrzykowski has shown that function (3) is exact in the sense that there is a finite $\rho > 0$ such that any regular local minimizer of (NLP) is also a local minimizer of the penalized unconstrained problem. In 1970, Luenberger [19] showed that, under convex assumptions, there is a lower bound for ρ, equal to the largest Lagrange multiplier in absolute value, associated to the nonlinear problem. In 1978 Charalambous [5, 6] generalized the result of Luenberger for the L_1 penalty function (3), assuming the second-order sufficient conditions for (NLP).

Other worth mentioning works involving the L_1 penalty function are the ones of Charalambous [6] and Coleman & Conn [7]. These works present the optimality conditions for minimizing the L_1 function. Further contributions involving the L_1 penalty function can be found in [11, 18, 26]. The L_1 penalty function often appears in sequential quadratic programming techniques, used as a merit function (see [4, 21]).

In 1979, Han & Mangasarian [17] introduced a class of exact penalty functions associated to problem (1) with $\Omega = I\!R^n$,

$$P_Q(x,\rho) = f(x) + \rho\, Q(\|h(x), g^+(x)\|) \tag{5}$$

where ρ is the penalty parameter, $\|\cdot\|$ is any fixed vector norm in $I\!R^{m_1+m_2}$, and $Q : I\!R_+ \to I\!R_+$ is any convex function such that

$$Q(t) = 0 \text{ iff } t = 0 \text{ and } 0 < \lim_{t\to 0^+} \frac{Q(t) - Q(0)}{t} < \infty \ .$$

(Throughout the paper, for $u \in I\!R^{m_1}$ and $v \in I\!R^{m_2}$, the notation $\|u, v\|$ means $\|(u_1, \ldots, u_{m_1}, v_1, \ldots, v_{m_2})^T\|$.)

Observe that for $Q(\alpha) = \alpha$ and $\|\cdot\|_1$ we have the L_1 penalty function, $P_1(x, \rho)$; for $Q(\alpha) = \alpha$ and $\|\cdot\|_p$ we have the class of functions studied by Charalambous [5] with equal weights; for $Q(\alpha) = \alpha$ and $\|\cdot\|_\infty$ we have the Minimax penalty function.

For simplicity and without loosing generality, in this work we restrict our attention to the exact penalty function (5) with $Q(\alpha) = \alpha$, that is,

$$P(x, \rho) = f(x) + \rho\, \|h(x), g^+(x)\|. \tag{6}$$

Function $P(x, \rho)$ penalizes only the constraints g and h of the original problem. Keeping the penalty subproblems constrained to the set Ω, we will show that the classical exact penalty function theorem for the unconstrained case is still valid for this new situation. The consequence of such result is that we can apply efficient already known algorithms that take advantage of the structure of the penalty subproblems, as in the recent work of Martínez & Moretti [22].

This work is organized as follows. In Section 2 the results for classical exterior penalty function algorithms are reviewed and the relationship between the solutions to (NLP) and the solutions to the penalty subproblems is commented. In Section 3, a simple proof for the exactness of $P(x, \rho)$ with unconstrained penalty subproblems is presented. Next, this result is extended to constrained subproblems, where it is shown that there exists a finite penalty parameter that depends only on the Lagrange multipliers associated to the penalized constraints. Finally, in Section 4 we discuss the consequences of this extension for practical optimization.

2 General Results

Let us consider the penalty subproblems

$$\begin{array}{ll} \min & P(x, \rho_k) = f(x) + \rho_k \|h(x), g^+(x)\| \\ \text{s.t.} & x \in \Omega \ , \end{array} \tag{7}$$

with $\rho_k > 0$ and $\|\cdot\|$ any given norm in $I\!R^{m_1+m_2}$.

Initially, we recall the classical global convergence theorem.

Theorem 1: *Let $\{x_k\}$, $k = 0, 1, 2, \ldots$, be a sequence of global minimizers of problem (7), and $0 < \rho_k < \rho_{k+1}$, $\rho_k \to +\infty$. Then, every limit point of the sequence $\{x_k\}$ is* *a global minimizer of problem (1).*

Proof: See [2, 3, 20], among others. □

A limitation of the above theorem appears when we assume that the penalty subproblems have a global solution for all ρ_k. The example below shows that there is no guarantee of achieving a minimizer of the original problem if we do not follow global minimizers of the penalty subproblems.

Example 1: Consider the problem $\min 0$ s.t. $x(x^2-1)+1=0$. For the L_1 penalty function, $x_k=0$ is a local minimizer of the penalty subproblem for all $k \geq 0$. However, the unique limit point ($x^*=0$) is infeasible for the original problem.

Example 1 also shows that we might have convergence to a non-stationary point. The next example points out that even if the limit point is feasible, by not following global minimizers of the penalty subproblems, we can have convergence to a maximizer of the original problem.

Example 2: Consider the problem $\min -|x|^{1.5}$ s.t. $0 \leq x \leq 1$. Using the quadratic loss penalty function, $x_k = -9/(16\rho_k^2)$ is a local minimizer of the penalty subproblem. The limit point of the sequence, $x^*=0$, is feasible but it is a maximizer of the original problem.

Another question that might raise is what happens to a sequence of just stationary points of the penalty subproblems. For the quadratic loss penalty function, and assuming regularity at the limit point, it can be proved that this limit point is stationary for the original problem (see, e.g. [3]). For the L_1 penalty function, if we request feasibility and regularity at the limit point, this result remains valid.

In order to avoid the need of a sequence of unbounded penalty parameters, it is possible to construct penalty functions that are exact in the sense that a local solution to the original problem is also a local solution to the penalty subproblem, for a finite value of the penalty parameter. Before proving that (6) is an exact penalty function, we recall the second order sufficient optimality conditions for problem (1) with $\Omega = I\!R^n$. The Lagrangian function is defined by

$$\mathcal{L}(x,\lambda,\mu) = f(x) + h(x)^T\lambda + g(x)^T\mu \ ,$$

and so

$$\nabla_x\mathcal{L}(x,\lambda,\mu) = \nabla f(x) + J_h(x)^T\lambda + J_g(x)^T\mu \ ,$$

$$\nabla^2_{xx}\mathcal{L}(x,\lambda,\mu) = \nabla^2 f(x) + \sum_{i=1}^{m_1}\lambda_i\nabla^2 h_i(x) + \sum_{j=1}^{m_2}\mu_j\nabla^2 g_j(x) \ ,$$

where

$$J_h(x) = [\nabla h_1(x),\ldots,\nabla h_{m_1}(x)]^T \text{ and } J_g(x) = [\nabla g_1(x),\ldots,\nabla g_{m_2}(x)]^T \ .$$

Theorem 2: *If* $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^{m_1} \times \mathbb{R}_+^{m_2}$ *is such that*

$$\nabla_x \mathcal{L}(x^*, \lambda^*, \mu^*) = 0 \ , \tag{8}$$

$$h(x^*) = 0, \quad g(x^*) \leq 0, \quad g(x^*)^T \mu^* = 0 \ , \tag{9}$$

and

$$d^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*, \mu^*) d > 0 \ , \tag{10}$$

for all $d \in \{y \in \mathbb{R}^n \mid J_h(x^*)y = 0, \ \nabla g_j(x^*)^T y = 0, \ j \in J\}$, *where* $J = \{j \mid g_j(x^*) = 0, \ \mu_j^* > 0\}$, *then* x^* *is a strict local minimizer of problem (1) with* $\Omega = \mathbb{R}^n$.

Proof: See, e.g. [20], pp. 316–317. □

3 Exactness of $P(x, \rho)$

To establish a lower bound for the penalty parameter, so that the penalty function $P(x, \rho)$ becomes exact, we need the concept of dual norms. Recall that for any given vector norm $\| \cdot \|$ in $\mathbb{R}^n$ there is a corresponding vector norm $\| \cdot \|'$, called the dual norm, which is defined by

$$\|x\|' = \max_{\|y\|=1} x^T y \ .$$

For example, the dual of $\| \cdot \|_1$ is $\| \cdot \|_\infty$ and $\| \cdot \|_2$ is self-dual. From the above definition, it follows the so-called generalized Cauchy inequality,

$$|x^T y| \leq \|x\|' \cdot \|y\| \text{ for all } x, y \in \mathbb{R}^n \ . \tag{11}$$

Several proofs of the exactness of $P(x, \rho)$ can be found depending on the chosen norm. For $\| \cdot \|_1$ see [3, 5, 6, 20] and for $\| \cdot \|_p$ see [5, 8, 17]. There are also proofs that do not depend on the chosen norm, using either subgradients [14] or the exactness of the L_1 penalty function [17]. In the following we present a simpler version of the proof stated in [17] for the exactness of $P(x, \rho)$.

Theorem 3: *Let* $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^n \times \mathbb{R}^{m_1} \times \mathbb{R}_+^{m_2}$ *satisfy the second-order sufficient conditions for a local minimizer of problem (1) with* $\Omega = \mathbb{R}^n$ *(given by Theorem 2) and let* $\| \cdot \|$ *be any norm in* $\mathbb{R}^{m_1+m_2}$ *with* $\| \cdot \|'$ *its dual norm. Then for* $\rho > \bar{\rho} = \|\lambda^*, \mu^*\|'$, x^* *is a strict unconstrained local minimizer of* $P(x, \rho)$.

Proof: Let us assume that x^* is not a strict local minimizer of $P(x, \rho)$. Therefore, there exists a sequence $\{x^k\}$ converging to x^* such that $P(x^k, \rho) \leq P(x^*, \rho)$ for all $k \geq 0$. Thus,

$$P(x^k, \rho) - P(x^*, \rho) = f(x^k) - f(x^*) + \rho \|h(x^k), g^+(x^k)\| \leq 0 \ . \tag{12}$$

Define the sequence $d^k = (x^k - x^*)/\|x^k - x^*\|$ and take a convergent subsequence such that $\{d^k\}$ converges to d with $\|d\| = 1$. Taking the limit in (12), we obtain

$$\nabla f(x^*)^T d + \rho \| J_h(x^*)d, J_{g^+}(x^*)d\| \leq 0 \ , \tag{13}$$

where

$$J_{g^+}(x^*)d = [\nabla g_1^+(x^*)^T d, \ldots, \nabla g_{m_2}^+(x^*)^T d]^T \ ,$$

with

$$\nabla g_j^+(x^*)^T d = \begin{cases} 0 \quad , \quad g_j(x^*) < 0 \\ \max\{\nabla g_j(x^*)^T d, 0\} \quad , \quad g_j(x^*) = 0 \end{cases} .$$

Using equation (8) in (13), we find

$$\rho \| J_h(x^*)d, J_{g^+}(x^*)d\| - (\lambda^*)^T J_h(x^*)d - (\mu^*)^T J_g(x^*)d \leq 0 \ , \tag{14}$$

which implies,

$$\rho \| J_h(x^*)d, J_{g^+}(x^*)d\| - (\lambda^*)^T J_h(x^*)d - (\mu^*)^T J_{g^+}(x^*)d \leq 0 \ ,$$

since $(\mu^*)^T J_g(x^*)d \leq (\mu^*)^T J_{g^+}(x^*)d$. Now, using the generalized Cauchy inequality (11), we conclude that

$$(\rho - \|\lambda^*, \mu^*\|')\| J_h(x^*)d, J_{g^+}(x^*)d\| \leq 0 \ .$$

Because $\rho > \|\lambda^*, \mu^*\|'$, it follows that $J_h(x^*)d = 0$ and $J_{g^+}(x^*)d = 0$. Moreover, from inequality (14), $\nabla g_j(x^*)^T d = 0$ for $j \in J = \{j \mid g_j(x^*) = 0, \ \mu_j^* > 0\}$. From equation (10) of Theorem 2, it follows that $d^T \nabla^2_{xx} \mathcal{L}(x^*, \lambda^*, \mu^*) d > 0$, which implies that for k sufficiently large

$$\mathcal{L}(x^k, \lambda^*, \mu^*) > \mathcal{L}(x^*, \lambda^*, \mu^*) \ .$$

As a result,

$$\begin{aligned} P(x^k, \rho) &= f(x^k) + \rho \|h(x^k), g^+(x^k)\| \\ &\geq f(x^k) + \|\lambda^*, \mu^*\|' \|h(x^k), g^+(x^k)\| \\ &\geq f(x^k) + h(x^k)^T \lambda^* + (g^+(x^k))^T \mu^* \\ &\geq f(x^k) + h(x^k)^T \lambda^* + g(x^k)^T \mu^* \\ &= \mathcal{L}(x^k, \lambda^*, \mu^*) \\ &> \mathcal{L}(x^*, \lambda^*, \mu^*) \\ &= f(x^*) \\ &= P(x^*, \rho) \\ &\geq P(x^k, \rho) \ , \end{aligned}$$

which is a contradiction. □

Theorem 3 guarantees that, under second-order sufficiency, if x^* is a local minimizer of (NLP) then x^* is also a local minimizer of the penalty subproblem. The practical interest is the converse of this theorem. For the convex case, the characterization of solutions to (NLP) through the solutions to the penalty subproblem is fully satisfactory since, in this case, every local minimum is also global, and the optimal solution sets of the original problem and of the subproblems are identical (see [5, 17]). In the nonconvex case, the study of properties that ensure that local (global) solutions to the penalty subproblem are local (global) solutions to the constrained problem are of great interest.

Theorem 3 remains true if we use different penalty parameters for each constraint. There is no difficulty in reorganizing the theory to account for this change. Thus, assuming that Ω is defined by a set of equalities and inequalities (making sense to refer to the Lagrange multipliers associated to Ω), it is possible to state Theorem 4 below, which illustrates the use of different penalty parameters for the constraints h and g, and for the set Ω.

Theorem 4: *Assume that $x^* \in \mathbb{R}^n$ satisfies the second-order sufficient conditions for a local minimizer of problem (1). Let $\lambda^* \in \mathbb{R}^{m_1}$, $\mu^* \in \mathbb{R}^{m_2}_+$ and η^* be the Lagrange multiplier vectors associated to h, g and the constraint set Ω, respectively. Let $\|\cdot\|$ be any vector norm with $\|\cdot\|'$ its dual norm. Then, for $\rho > \bar{\rho} = \|\lambda^*, \mu^*\|'$, and $\rho_1 > \bar{\rho}_1 = \|\eta^*\|'$, x^* is a strict unconstrained local minimizer of $P(x,\rho,\rho_1) = f(x) + \rho\|h(x), g^+(x)\| + \rho_1 R(x)$, where $R(x)$ is the penalty function for Ω.*

Proof: Analogous to Theorem 3. □

Theorem 3 shows that $P(x,\rho)$ is exact for problem (1) with $\Omega = \mathbb{R}^n$. An analogous result is given by Theorem 4, where the set Ω was included in the penalty function. In the following theorem, we show that there exists a finite lower bound for the penalty parameter in the case that problem (1) is penalized only with respect to h and g, so that the penalty subproblems remain constrained to Ω. Observe that in the classical theory for exact penalty functions the subproblems are unconstrained.

Theorem 5: *Let $x^* \in \mathbb{R}^n$ satisfy the second-order sufficient conditions for a local minimizer of problem (1). Let $\lambda^* \in \mathbb{R}^{m_1}$ and $\mu^* \in \mathbb{R}^{m_2}_+$ be the corresponding Lagrange multipliers of constraints h and g, and $\|\cdot\|$ be any vector norm with $\|\cdot\|'$ its dual norm. There exists $0 < \bar{\rho} < \infty$ such that for $\rho > \bar{\rho}$, x^* is a strict local minimizer of problem (7).*

Proof: Consider the problem

$$\begin{array}{ll} \min & f(x) + \rho\|h(x), g^+(x)\| + \rho_1 R(x) \\ \text{s.t.} & x \in \mathbb{R}^n \ , \end{array} \tag{15}$$

where $\rho, \rho_1 > 0$ and $R(x)$ is as defined in Theorem 4. Since x^* satisfies the second-order sufficient conditions for problem (1), from Theorem 4, there exist $\bar{\rho} = \|\lambda^*, \mu^*\|'$

and $\bar{\rho}_1 = \|\eta^*\|'$ such that for $\rho > \bar{\rho}$ and $\rho_1 > \bar{\rho}_1$, x^* is a strict local minimizer of (15). Therefore, for $\rho > \bar{\rho}$ and $\rho_1 > \bar{\rho}_1$ there exists $\varepsilon \equiv \varepsilon(\rho, \rho_1) > 0$ such that for all $x \in \mathcal{B}(x^*, \varepsilon) \equiv \{x \in I\!R^n \mid \|x - x^*\| \leq \varepsilon\}$,

$$f(x^*) < f(x) + \rho\|h(x), g^+(x)\| + \rho_1 R(x) \text{ for all } x \in \mathcal{B}(x^*, \varepsilon) \cap \Omega \ ,$$

since $R(x) = 0$ for all $x \in \Omega$. Thus, for $\rho > \bar{\rho} = \|\lambda^*, \mu^*\|'$, x^* is a strict local minimizer of

$$\begin{array}{ll} \min & f(x) + \rho\|h(x), g^+(x)\| \\ \text{s.t.} & x \in \Omega \ , \end{array}$$

and the proof is complete. □

The importance of Theorem 5 is to guarantee the existence of a finite penalty parameter that depends on the Lagrange multipliers associated to constraints h and g only, and not to the set Ω. Such result allows us to develop algorithms for solving problem (1) using the function $P(x, \rho)$ when the penalty subproblems remain constrained to Ω. Note that, since the lower bound for the penalty parameter is finite, the sequence of penalty subproblems is also finite.

4 Final Remarks

The result given by Theorem 5 suggests that we can work with exact penalty functions even for constrained subproblems. Although solving constrained problems is more complex than solving unconstrained ones, the approach of dealing with constrained penalty subproblems aims to take advantage of the geometry of the constraint set Ω. This idea makes possible the usage of already established numerical algorithms for nonlinear programming problems with simple constraints (e.g. [15, 22]) to solve the penalty subproblems. Moreover, new algorithms might be developed exploiting particular properties of the chosen exact penalty function and the set Ω.

Exact penalty methods require the solution of a finite sequence of penalty subproblems, contrary to the classical non-exact approach, for which an infinite sequence is needed. Unfortunately, the exact penalty functions are either nondifferentiable or differentiable, but with a considerable degree of complexity. Therefore, it is necessary to develop efficient methods for nonsmooth optimization and/or to propose simpler exact penalty functions. Due to the importance of exact penalty methods in nonlinear programming it is also of great interest to investigate properties relating local (global) minimizers of the penalty function and local (global) solutions to the original problem.

Acknowledgements: We are thankful to Dr. Sandra A. Santos and Professor José Mario Martínez for the careful reading of the manuscript, and to an anonymous referee for the suggestions that improved the paper.

References

[1] J. W. Bandler & C. Charalambous, "Nonlinear programming using minimax techniques". *Journal of Optimization Theory and Applications*, **13**, 679–686, 1974.

[2] M. S. Bazaraa & C. M. Shetty, *Nonlinear Programming Theory and Algorithms*. John Wiley & Sons, 1979.

[3] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, 1982.

[4] R. M. Chamberlain, "Some examples of cycling in variable metric methods for constrained minimization". *Mathematical Programming*, **16**, 378–383, 1979.

[5] C. Charalambous, "A lower bound for the controlling parameters of the exact penalty functions". *Mathematical Programming*, **15**, 278–290, 1978.

[6] C. Charalambous, "On the condition for optimality of the nonlinear L_1 problem". *Mathematical Programming*, **17**, 123–135, 1979.

[7] T. F. Coleman & A. R. Conn, "Second-order conditions for an exact penalty function". *Mathematical Programming*, **19**, 178–185, 1980.

[8] G. Di Pillo & L. Grippo, "A continuously exact penalty functions for nonlinear programming problems with inequality constraints". *SIAM J. Control & Optimization*, **23**, 72–84, 1985.

[9] G. Di Pillo & L. Grippo, "An exact penalty method with global convergence properties for nonlinear programming problems". *Mathematical Programming*, **36**, 1–18, 1986.

[10] G. Di Pillo & L. Grippo, "Exact penalty functions in constrained optimization". *SIAM J. Control & Optimization*, **27**, 1333–1360, 1989.

[11] S. Dolecki & S. Rolewicz, "Exact penalty for local minima". *SIAM J. Control & Optimization*, **17**, 111–124, 1979.

[12] V. R. Evans & F. J. Gould, "Stability and exponential penalty function techniques in nonlinear programming". *Institute of Statistics Mimeo Series*, No. 723, University of North Carolina, Chapel Hill, 1970.

[13] R. Fletcher, "An exact penalty function for nonlinear programming with inequalities". *Mathematical Programming*, **5**, 129–150, 1973.

[14] R. Fletcher, *Practical Methods of Optimization*. Wiley, New York, 1987.

[15] A. Friedlander, J. M. Martínez & S. A. Santos, "A new trust region algorithm for bound constrained minimization". *Applied Mathematics & Optimization*, **30**, 235–266, 1994.

[16] T. Glad & E. Polak, "A multiplier method with automatic limitation of penalty growth". *Mathematical Programming*, **17**, 140–155, 1979.

[17] S–P. Han & O. L. Mangasarian, "Exact penalty functions in nonlinear programming". *Mathematical Programming*, **17**, 251–269, 1979.

[18] S. P. Howe, "New conditions for exactness of a simple penalty function". *SIAM J. Control & Optimization*, **11**, 378–381, 1973.

[19] D. Luenberger, "Control problem with kinks". *IEEE Transaction on Automatic Control*, **15**, 570–574, 1970.

[20] D. Luenberger, *Linear and Nonlinear Programming*. Addison–Wesley, New York, 1984.

[21] N. Maratos, *Exact penalty function algorithms for finite-dimensional and control optimization problems*, Ph.D. Thesis, University of London, England, 1978.

[22] J. M. Martínez & A. C. Moretti, "A trust region method for minimization of nonsmooth functions with linear constraints". *Mathematical Programming*, **76**, 431–449, 1997.

[23] J. M. Martínez & L. T. Santos, *Some new theoretical results on recursive quadratic programming algorithms*. Relatório de Pesquisa – IMECC, 12/95, IMECC – UNICAMP, 1995.

[24] F. H. Murphy, "A class of exponential penalty functions". *SIAM J. Control*, **12**, 679–686, 1974.

[25] T. Pietrzykowski, "An exact potential method for constrained maxima". *SIAM J. Numerical Analysis*, **6**, 299–304, 1969.

[26] W. I. Zangwill, "Nonlinear programming *via* penalty functions". *Management Sciences*, **13**, 344–358, 1967.