On the Role of Continuously Differentiable Exact Penalty Functions in Constrained Global Optimization

STEFANO LUCIDI
Dipartimento di Informatica e Sistemistica, Università di Roma "La Sapienza", via Buonarroti 12, 00185 Roma, Italy.

(Received: 20 July 1992; accepted: 7 November 1993)

Abstract. The aim of this paper is to show that the new continuously differentiable exact penalty functions recently proposed in literature can play an important role in the field of constrained global optimization. In fact they allow us to transfer ideas and results proposed in unconstrained global optimization to the constrained case.

First, by drawing our inspiration from the unconstrained case and by using the strong exactness properties of a particular continuously differentiable penalty function, we propose a sufficient condition for a local constrained minimum point to be global.

Then we show that every constrained local minimum point satisfying the second order sufficient conditions is an "attraction point" for a particular implementable minimization algorithm based on the considered penalty function. This result can be used to define new classes of global algorithms for the solution of general constrained global minimization problems. As an example, in this paper we describe a simulated annealing algorithm which produces a sequence of points converging in probability to a global minimum of the original constrained problem.

Key words. Constrained global optimization, global optimality, strict local minima, simulated annealing algorithm, exact penalty function.

1. Introduction

Let us consider the following nonlinear programming problem:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0\,, \\ & h(x) = 0\,, \end{aligned} \tag{P}$$

where $f: \mathbb{R}^n \to \mathbb{R}$, $g: \mathbb{R}^n \to \mathbb{R}^m$, $h: \mathbb{R}^n \to \mathbb{R}^q$, $q \leq n$ are three times continuously differentiable functions. We denote by

$$\mathcal{F} := \{x \in \mathbb{R}^n \colon g(x) \leq 0,\ h(x) = 0\}$$

feasible set of Problem (P).

In this paper we consider the problem of finding a global minimizer of Problem (P), that is a point $x^* \in \mathcal{F}$ such that:

$$f(x^*) \leq f(x)\,, \quad \text{for all} \quad x \in \mathcal{F}\,.$$

Journal of Global Optimization **5**: 49–68, 1994.

In the literature a big variety of results and algorithms have been proposed in the field of unconstrained global optimization (see [1–6]).

The situation is very different for the constrained case. In fact, apart from special classes (see, for instance, [7–9]), so far only few works have been done in the field of general constrained global minimization problems [9–13].

A natural way to treat problems with constraints is to get rid of the constraints by redefining the objective function [10]. Recently, in [14] a continuously differentiable exact penalty function has been introduced which, under suitable assumptions, presents very strong exactness properties. The merit function of [14] has been further investigated in [15] where its expression has been changed so as to obtain a new continuously differentiable penalty function which maintains the same properties of exactness under weaker assumptions.

By making use of these exact penalty functions, many ideas and techniques applicable to unconstrained global optimization can be transferred to the constrained case. Therefore these penalty functions provide us with a direct and powerful way of tackling constrained global optimization problems. In this paper we begin to exploit the properties of these new merit functions and the results obtained confirm the goodness of this approach.

In particular, as first result, we present a sufficient condition for a local constrained minimum point to be global. This global optimality condition derives from an analogous result given in [6] for the unconstrained case and from the exactness properties of the penalty function proposed in [15]. It states that, under quite mild assumptions, if every Kuhn–Tucker point of the problem is a strict local minimum then the problem has only one Kuhn–Tucker point which is a global minimum.

As second result, we show that every constrained local minimum point satisfying the second order sufficient conditions is an "attraction point" for a particular implementable minimization algorithm based on the penalty function of [15]. This result can be the starting point to define new classes of algorithms for the solution of general constrained global minimization problems. As an example, in this paper we describe a simulated annealing algorithm which generalizes the approach described in [16] and it consists of several local unconstrained minimizations of the considered penalty function. Any of this minimization is started from points chosen at random (by using an acceptance-rejection technique) according to a probability density function which is updated during the algorithm so as to be concentrated around the global minimizers of the original problem. The exactness property of the considered penalty function and the way of generating the starting points of the local minimizations ensure that, under mild assumptions, the sequence of points produced by the algorithm converges in probability to the solution of the problem.

The paper is organized as follows. In Section 2 we briefly describe the penalty function proposed in [15] and its exactness properties. In Section 3 we derive the sufficient condition for the global optimality. In Section 4 we consider the problem to define methods to solve constrained global optimization problems.

2. The Exact Penalty Function

In this section we briefly describe the exact penalty function proposed in [15] and some of its properties which will be used in this paper. We refer to [15] and [14] for a more detailed analysis of these results (see also [17] for a similar treatment for nonlinear programming problems with equality and inequality constraints).

The Lagrangian function associated with Problem (P) is the function $L: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q \to \mathbb{R}$ defined by

$$L(x, \lambda, \mu) := f(x) + \lambda' g(x) + \mu' h(x) .$$

A point $\bar{x} \in \mathbb{R}^n$ is called a *Kuhn–Tucker point* for Problem (P) if there exist multiplier vectors $(\bar{\lambda}, \bar{\mu}) \in \mathbb{R}^m \times \mathbb{R}^q$ such that

$$\nabla_x L(\bar{x}, \bar{\lambda}, \bar{\mu}) = 0 , \quad G(\bar{x})\bar{\lambda} = 0 , \quad \bar{\lambda} \geq 0 , \quad g(\bar{x}) \leq 0 , \quad h(\bar{x}) = 0 ,$$

where $G(x) := \text{diag}(g_i(x))$ and, furthermore, $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is called *Kuhn–Tucker triple*. We say that *the strict complementarity* holds at K–T triple $(\bar{x}, \bar{\lambda}, \bar{\mu})$ if, for any index i such that $g_i(\bar{x}) = 0$, we have $\bar{\lambda}_i > 0$. Then, for any $x \in \mathbb{R}^n$, we define the index set

$$I_0(x) := \{i: g_i(x) = 0\} .$$

We say that a local minimum point $\bar{x}$ for Problem (P) satisfies *the strict complementary and the second order sufficiency conditions* if there exist vectors $\bar{\lambda} \in \mathbb{R}^m$, $\bar{\mu} \in \mathbb{R}^q$ such that the point $(\bar{x}, \bar{\lambda}, \bar{\mu})$ is a Kuhn–Tucker triple for Problem (P) satisfying the strict complementarity condition and

$$z' \nabla_x^2 L(\bar{x}, \bar{\lambda}, \bar{\mu}) z > 0 ,$$

for all $z \neq 0$ such that $\nabla g_i(\bar{x})' z = 0$, for $i \in I_0(\bar{x})$ and $\nabla h_j(\bar{x})' z = 0$, for $j = 1, \ldots, q$.

We denote by $g^+(x)$ the vector with components $g_i^+(x) := \max[0, g_i(x)]$, $i = 1, \ldots, m$.

Let now α, $p \in \mathbb{R}$ be given scalars such that $\alpha > 0$ and $p \geq 2$.

In connection with these two scalars we consider an open perturbation of the feasible set $\mathcal{F}$ defined by:

$$\mathcal{A}_{\alpha p} := \left\{ x \in \mathbb{R} : \sum_{i=1}^{m} g_i^+(x)^p + \sum_{j=1}^{q} h_j(x)^2 < \alpha \right\}$$

and we denote by $\bar{\mathcal{A}}_{\alpha p}$ the closure of $\mathcal{A}_{\alpha p}$ and by $\partial \mathcal{A}_{\alpha p}$ the boundary of $\mathcal{A}_{\alpha p}$. Then we introduce the function

$$a(x) := \alpha - \sum_{i=1}^{m} g_i^+(x)^p - \sum_{j=1}^{q} h_j(x)^2 , \tag{1}$$

which assumes positive values on $\mathcal{A}_{\alpha p}$ and it is null on its boundary.

In the sequel we shall make use of the following hypotheses:

ASSUMPTION A1. *The set $\mathcal{A}_{\alpha p}$ is compact.* □

ASSUMPTION A2. *For every $x \in \mathcal{F}$ the gradients $\nabla g_i(x)$, $i \in I_0(x)$, $\nabla h_j(x)$, $j = 1, \dots, q$ are linearly independent.* □

ASSUMPTION A3. *For every $x \in \mathcal{A}_{\alpha p}$*

$$\sum_{i=1}^{m} v_i(x) g_i^+(x) \nabla g_i(x) + \sum_{j=1}^{q} w_j(x) h_j(x) \nabla h_j(x) = 0 ,$$

where

$$v_i(x) = 1 + \frac{p}{2} \frac{(\|g^+(x)\|^2 + \|h(x)\|^2)}{a(x)} g_i^+(x)^{p-2}$$

$$w_j(x) = 1 + \frac{\|g^+(x)\|^2 + \|h(x)\|^2}{a(x)} ,$$

implies that $g_i^+(x) = 0$, for $i = 1, \dots, m$, and $h_j(x) = 0$, for $j = 1, \dots, q$. □

The preceding assumptions seem to be mild conditions on the problem and, so far, they are the weakest possible assumptions which imply a total equivalence between the solution of a constrained problem and the unconstrained minimization of a differentiable function.

Assumption A1 ensures that the proposed penalty function has compact level sets and, in particular, it is satisfied if all the variables are bounded or if there exist at least a function g_i or a function h_j that is radially unbounded.

Assumption A2 is needed to define an exact penalty function which is continuously differentiable. In particular it is the weakest assumption allowing the definition a continuously differentiable multiplier functions that yield an estimate of the multiplier vectors associated with Problem (P) as functions of the variable x. The definition of such multiplier functions is an essential element to define a continuously differentiable exact penalty function.

Assumption A3 is required when a feasible point is not known "a priori". In this case, besides minimizing the objective function, we must also solve the problem of finding a feasible point. This is a difficult problem because it is equivalent to find the global minimum of a (usually nonconvex) function which measures the violation of the constraints. Therefore in order to ensure the attainment of feasibility we must assume some "good behaviour" of the constraints at nonfeasible points. In fact all the papers dealing with global convergent methods for constrained minimization problem require a (global) regularity condition on the constraints. Assumption A3 appears to be at least comparable with the other assumptions proposed in literature. In particular, in [15] it is shown that Assumption A3 is implied by the typical assumptions used in sequential quadratic programming methods and in the methods based on exact continuously differentiable penalty functions. Furthermore, it is possible to prove (see [15]) that, in general, Assumption A3 is a sufficient condition for the feasible set to be not empty but, in particular, it is also necessary in the case of compact feasible sets given by convex inequalities and linear equalities. Therefore, at least for this class of feasible sets, this assumption is the weakest possible condition that ensures that the original constrained problem is well defined.

We refer to [15] (see also [14]) for a more detailed discussion of Assumptions A1–A3.

Here we recall a proposition, proposed in [14], that will be used later on.

PROPOSITION 1. *Assume that:*
(i) there exists scalars $\hat{\alpha}$ and $p \geq 2$ such that the set $\bar{\mathscr{A}}_{\hat{\alpha}p}$ satisfies Assumption A1;
(ii) Assumption A2 holds.

Then, there exists an $\bar{\alpha} \leq \hat{\alpha}$ such that for all $\alpha \in (0, \bar{\alpha}]$ the corresponding set $\mathscr{A}_{\alpha p}$ satisfies: $\mathscr{F} \subseteq \mathscr{A}_{\alpha p}$ and for every $x \in \bar{\mathscr{A}}_{\alpha p}$ the gradients $\nabla g_i(x)$, $i \in \{i': g_{i'}(x) \geq 0\}$, $\nabla h_j(x)$, $j = 1, \ldots, q$ are linearly independent. □

In the sequel of this section we shall assume that Assumptions A1 and A2 hold; Assumption A3 will be invoked explicitly when needed.

Now we describe the multiplier functions $(\lambda(x), \mu(x))$ introduced in [15] and needed in the expression of the new penalty function.

PROPOSITION 2. *Let $(\lambda(x), \mu(x))$ be the functions given by:*

$$\begin{bmatrix} \lambda(x) \\ \mu(x) \end{bmatrix} = -M^{-1}(x) \begin{bmatrix} \nabla g(x)' \\ \nabla h(x)' \end{bmatrix} \nabla f(x) ;$$

where

$$M(x) := \begin{bmatrix} \nabla g(x)'\nabla g(x) + G^2(x) + r(x) I_m & \nabla g(x)'\nabla h(x) \\ \nabla h(x)'\nabla g(x) & \nabla h(x)'\nabla h(x) + r(x) I_q \end{bmatrix},$$

$r(x) := \Sigma_{i=1}^m g_i^+(x)^p + \Sigma_{j=1}^q h_j(x)^2$ and I_m (I_q) indicates the $m \times m$ ($q \times q$) identity matrix. Then if $(\bar{x}, \bar{\lambda}, \bar{\mu}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q$ is a K–T triple for Problem (P) we have that $\lambda(\bar{x}) = \bar{\lambda}$ and $\mu(\bar{x}) = \bar{\mu}$. □

Now, we can define on $\mathscr{A}_{\alpha p}$ the exact penalty function proposed in [15]:

$$Z(x; \varepsilon) := f(x) + \lambda(x)'c(x; \varepsilon) + \mu(x)'h(x) + \frac{1}{\varepsilon a(x)} [\|c(x; \varepsilon)\|^2 + \|h(x)\|^2] , \tag{3}$$

where:

$$c_i(x;\varepsilon) := \max\left[g_i(x), -\frac{\varepsilon a(x)}{2}\lambda_i(x)\right], \quad i = 1, \ldots, m .$$

Given a point $\tilde{x} \in \mathcal{A}_{\alpha p}$ we can introduce the following level set:

$$\Omega_{\alpha p}(\tilde{x}) := \{x \in \mathcal{A}_{\alpha p}: Z(x;\varepsilon) \leq Z(\tilde{x};\varepsilon)\} .$$

Then we report some properties of the function $Z(x;\varepsilon)$ which are necessary to prove the results of this paper. We refer to [15] for the proofs of Proposition 3, Theorem 1 and Theorem 2. While the proof of Proposition 4 can be obtained, with minor modifications, from the one given in [18] for a similar result.

PROPOSITION 3. *For any* $\varepsilon > 0$,
(a) $Z(x;\varepsilon)$ *is continuously differentiable for all* $x \in \mathcal{A}_{\alpha p}$, *with gradient*

$$\begin{aligned}\nabla Z(x;\varepsilon) &= \nabla f(x) + \nabla g(x)\lambda(x) + \nabla h(x)\mu(x) \\ &\quad + \nabla\lambda(x)c(x;\varepsilon) + \nabla\mu(x)h(x) + \frac{2}{\varepsilon a(x)}[\nabla g(x)c(x;\varepsilon) + \nabla h(x)h(x)] \\ &\quad + \frac{\|c(x;\varepsilon)\|^2 + \|h(x)\|^2}{\varepsilon a(x)^2}\left[p\sum_{i=1}^{m}\nabla g_i(x)g_i^+(x)^{p-1} + 2\nabla h(x)h(x)\right];\end{aligned} \tag{4}$$

(b) $Z(x;\varepsilon) \leq f(x)$ *for all* $x \in \mathcal{F}$;
(c) $Z(x;\varepsilon)$ *admits a global minimum point on* $\mathcal{A}_{\alpha p}$. □

THEOREM 1.
(a) *Let* $(\bar{x}, \bar{\lambda}, \bar{\mu})$ *be a K–T triple for Problem (P); then, for any* $\varepsilon > 0$, *we have:* $\nabla Z(\bar{x};\varepsilon) = 0$, $c(\bar{x};\varepsilon) = 0$ *and* $Z(\bar{x};\varepsilon) = f(\bar{x})$.
(b) *Assume that either* $\tilde{x} \in \mathcal{F}$ *or Assumption A3 holds. Then, there exists an* $\varepsilon^* > 0$ *such that, for all* $\varepsilon \in (0, \varepsilon^*]$, *if* $x_\varepsilon \in \Omega_{\alpha p}(\tilde{x})$ *is a stationary point of* $Z(x;\varepsilon)$, *we have that* $(x_\varepsilon, \lambda(x_\varepsilon), \mu(x_\varepsilon))$ *is a K–T triple for Problem (P).* □

THEOREM 2. *Suppose that Problem (P) is well defined, namely the feasible set* $\mathcal{F}$ *is not empty. Then there exists an* ε^* *such that for all* $\varepsilon \in (0, \varepsilon^*]$, *any global minimum point of Problem (P) is an unconstrained global minimum point of* $Z(x;\varepsilon)$ *on* $\mathcal{A}_{\alpha p}$ *and conversely.* □

PROPOSITION 4. *Assume that* $p \geq 3$ *in the definitions of* $a(x)$ *and* $\mathcal{A}_{\alpha p}$. *Let* (x^*, λ^*, μ^*) *be a K–T triple for Problem (P) that satisfies the strict complementarity and the second order sufficiency conditions. Then, there exist an* ε^* *such that for all* $\varepsilon \in (0, \varepsilon^*]$, x^* *is an isolated local minimum point for* $Z(x;\varepsilon)$ *and the Hessian matrix* $\nabla^2 Z(x^*;\varepsilon)$ *is positive definite.* □

3. A Global Optimality Condition for Constrained Minimization Problems

In [6] an interesting global optimality condition is described. Under very weak assumptions, it states that if a function has the property that its stationary points are strict local minimum points then this function has only one stationary point which is a global minimum point. In this section we extend this condition to the constrained case. In order to obtain this result, we need two new propositions which make deeper the analysis of the exactness properties of Z. The first proposition is quite technical. It shows that in part (b) of Theorem 1 it is possible to replace the level set $\Omega_{\alpha p}(\tilde{x})$ by the set $\mathscr{A}_{\alpha p}$. In others words, it states that, if the penalty parameter is sufficiently small, all the stationary points of Z belonging to $\mathscr{A}_{\alpha p}$ are K–T points for Problem (P) (whereas part (b) of Theorem 1 considered only the stationary points of Z belonging to $\Omega_{\alpha p}(\tilde{x})$).

PROPOSITION 5. *Assume that*
(i) Assumptions A1 and A2 hold;
(ii) for every $x \in \bar{\mathscr{A}}_{\alpha p}$

$$p \sum_{i=1}^{m} \nabla g_i(x) g_i^+(x)^{p-1} + 2\nabla h(x) h(x) = 0$$

implies that $g_i^+(x) = 0$, *for* $i = 1, \ldots, m$, *and* $h_j(x) = 0$, *for* $j = 1, \ldots, q$.

Then, there exists an $\varepsilon^* > 0$ *such that, for all* $\varepsilon \in (0, \varepsilon^*]$, *if* $x_\varepsilon \in \mathscr{A}_{\alpha p}$ *is a stationary point of* $Z(x; \varepsilon)$, *we have that* $(x_\varepsilon, \lambda(x_\varepsilon), \mu(x_\varepsilon))$ *is a K–T triple for Problem (P).*

Proof. Let $\hat{\varepsilon}$ be any positive constant. First we show that there exists a point $\tilde{x} \in \mathscr{A}_{\alpha p}$ such that, for every $\varepsilon \in (0, \hat{\varepsilon}]$, all the stationary points of $Z(x; \varepsilon)$ in $\mathscr{A}_{\alpha p}$ belong to $\Omega_{\alpha p}(\tilde{x})$. We proceed by contradiction. Assume that the assertion is false. Then for every sequence $\{x_k\}$ of points $x_k \in \mathscr{A}_{\alpha p}$ there exist two sequences $\{\varepsilon_k\}$ and $\{x_k^*\}$ such that $\varepsilon_k \in (0, \hat{\varepsilon}]$ and every x_k^* is a stationary point of $Z(x; \varepsilon_k)$ in $\mathscr{A}_{\alpha p}$ and

$$Z(x_k; \varepsilon_k) < Z(x_k^*; \varepsilon_k) . \tag{5}$$

Now we can choose the sequence $\{x_k\}$ such that $x_k \to \hat{x} \in \partial\mathscr{A}_{\alpha p}$. By (1) and the continuity assumptions, we have $\lim_{k\to\infty} a(x_k) = a(\hat{x}) = 0$. Then recalling (3) and the compactness of $\bar{\mathscr{A}}_{\alpha p}$, we obtain $\lim_{k\to\infty} \varepsilon_k Z(x_k; \varepsilon_k) = \infty$. Now (5) yields:

$$\lim_{k\to\infty} \varepsilon_k Z(x_k^*; \varepsilon_k) \geq \lim_{k\to\infty} \varepsilon_k Z(x_k; \varepsilon_k) = \infty ,$$

which, since $x_k^* \in \mathscr{A}_{\alpha p}$ and $\bar{\mathscr{A}}_{\alpha p}$ is compact, implies (by recalling (3)):

$$\lim_{k\to\infty} a(x_k^*) = 0 . \tag{6}$$

Since the sequence $\{\varepsilon_k\}$ is bounded we also have:

$$\lim_{k\to\infty} \varepsilon_k a(x_k^*)^2 = 0\,. \tag{7}$$

By using again the compactness of $\bar{\mathcal{A}}_{\alpha p}$, we have that there exists a convergent subsequence, which we relabel $\{x_k\}$, such that $\lim_{k\to\infty} x_k^* = x^* \in \bar{\mathcal{A}}_{\alpha p}$. Furthermore, (6) yields $a(x^*) = 0$.

Taking into account that every x_k^* is a stationary point of $Z(x; \varepsilon_k)$ and recalling (4) and (7), we obtain:

$$\begin{aligned} 0 &= \lim_{k\to\infty} \varepsilon_k a(x_k^*)^2 \nabla Z(x_k^*; \varepsilon_k) \\ &= (\|g^+(x^*)\|^2 + \|h(x^*)\|^2)\left[p \sum_{i=1}^{m} \nabla g_i(x^*) g_i^+(x^*)^{p-1} + 2\nabla h(x^*) h(x^*)\right], \end{aligned}$$

which, by assumption (ii), implies that $x^* \in \mathcal{F}$. Therefore we obtain a contradiction with $a(x^*) = 0$.

Now, since there exists a point $\tilde{x} \in \mathcal{A}_{\alpha p}$ such that, for every bounded $\varepsilon > 0$, all the stationary points of $Z(x; \varepsilon)$ in $\mathcal{A}_{\alpha p}$ belong to $\Omega_{\alpha p}(\tilde{x})$, the result follows directly from part (b) of Theorem 1. □

The next proposition completes the correspondence between the isolated local minimum points of Problem (P) and those of the function Z.

PROPOSITION 6. *Assume that the feasible set $\mathcal{F}$ is not empty and that Assumptions A1 and A2 hold. Then there exists an $\bar{\varepsilon}$ such that for all $\varepsilon \in (0, \bar{\varepsilon}]$, every strict local minimum point of Problem (P) is an unconstrained strict local minimum point of $Z(x; \varepsilon)$ on $\mathcal{A}_{\alpha p}$ and, if moreover Assumption A3 holds, also the converse is true.*

Proof. First we assume that $\bar{x}$ is a strict local minimum point of Problem (P). Therefore there exists an open set $E \subseteq \mathcal{A}_{\alpha p}$ such that $\bar{x} \in E$ and for every point $x \in \mathcal{F} \cap \bar{E}$ (where $\bar{E}$ is the closure of E), with $x \neq \bar{x}$, then $f(x) > f(\bar{x})$. By part (a) of Theorem 1 we have $Z(\bar{x}; \varepsilon) = f(\bar{x})$. Now, by contradiction, assume that the proposition is false. Then, for every integer k, there must exist an $\varepsilon_k \leq 1/k$ such that $\bar{x}$ is not an unconstrained strict local minimum point of $Z(x; \varepsilon_k)$. Therefore the global minimum point x_k of $Z(x; \varepsilon)$ on $\bar{E}$ satisfies:

$$Z(x_k; \varepsilon_k) \leq Z(\bar{x}; \varepsilon_k) = f(\bar{x}) \tag{8}$$

Since $\bar{E}$ is a compact set, there exists a convergent subsequence which we relabel $\{x_k\}$, such that $\lim_{k\to\infty} x_k = \hat{x}$. By (8) we obtain:

$$\limsup_{k\to\infty} Z(x_k; \varepsilon) \leq f(\bar{x})\,,$$

which, since $\varepsilon_k \to 0$, recalling (3) and Assumption A1, implies

$$c(\hat{x}; 0) = 0\,, \quad h(\hat{x}) = 0 \quad \text{and} \quad f(\hat{x}) \leq f(\bar{x})\,,$$

and, hence, since $\hat{x} \in \mathcal{F} \cap \bar{E}$, it must result that $\hat{x} = \bar{x}$. Therefore, since $x_k \to \hat{x} = \bar{x} \in E$, we have that, for sufficiently large values of k, the points x_k are unconstrained global minima and, hence, we have also that $\nabla Z(x_k; \varepsilon_k) = 0$.

Now (8) implies that $x_k \in \Omega_{\alpha p}(\bar{x})$ and, hence, for k large enough, part (b) of Theorem 1 ensures that $(\bar{x}, \lambda(\bar{x}), \mu(\bar{x}))$ is a K–T triple for Problem (P). Then part (a) of Theorem 1 implies in turn that, for sufficiently large values of k,

$$Z(x_k; \varepsilon_k) = f(x_k) \quad \text{and} \quad x_k \in \mathcal{F}\,,$$

and, by using (8) it follows that $f(x_k) \leq f(\bar{x})$ which contradicts the assumption that $x_k \in \bar{E}$.

Now assume that $\bar{x}$ is a strict local minimum point of $Z(x; \varepsilon)$ on $\mathcal{A}_{\alpha p}$ (and, hence, a stationary point of $Z(x; \varepsilon)$). We note that, by using part (b) of Theorem 1, we have that there exists an $\bar{\varepsilon}$ such that, for all $\varepsilon \in (0, \bar{\varepsilon}]$, $(\bar{x}, \lambda(\bar{x}), \mu(\bar{x}))$ is a K–T triple for Problem (P) and, by recalling part (a) of Theorem 1, $Z(\bar{x}; \varepsilon) = f(\bar{x})$. Since $\bar{x}$ is an isolated local minimum point of $Z(x; \varepsilon)$ there exists a neighborhood $\mathcal{B}$ of $\bar{x}$ such that:

$$f(\bar{x}) = Z(\bar{x}; \varepsilon) < Z(x; \varepsilon)\,, \quad \text{for all} \quad x \in \mathcal{B}\,.$$

Then, recalling part (b) of Proposition 3, we obtain:

$$f(\bar{x}) < Z(x; \varepsilon) \leq f(x)\,, \quad \text{for all} \quad x \in \mathcal{B} \cap \mathcal{F}$$

which completes the proof of the proposition. □

Now we are ready to state the main result of this section. Roughly speaking, this theorem allows to identify a class of constrained global optimization problems that are "easy" to solve.

THEOREM 3. *Assume that:*

(i) the feasible set $\mathcal{F}$ is not empty and connected;
(ii) there exist scalars $\hat{\alpha} > 0$ and $p \geq 2$ such that the set $\mathcal{A}_{\hat{\alpha} p}$ is bounded;
(iii) for every $x \in \mathcal{F}$ the gradients $\nabla g_i(x)$, $i \in I_0(x)$, $\nabla h_j(x)$, $j = 1, \ldots, q$ are linearly independent;
(iv) every KKT point of Problem (P) is a strict local minimum point.

Then there exists only one KKT point which is a global minimum point for Problem (P).

Proof. First of all, we note that assumption (iii) coincides with Assumption A2. Then, by using assumptions (ii) and (iii), and by recalling Proposition 1 we can conclude that there exists an $\alpha \leq \hat{\alpha}$ such that the corresponding set $\bar{\mathcal{A}}_{\alpha p}$ satisfies Assumptions A1, Assumption A3 and assumption (ii) of Proposition 5.

Now, on $\mathcal{A}_{\alpha p}$, we introduce the exact penalty function $Z(x; \varepsilon)$ and we show that, for sufficiently small values of the penalty parameter ε, the function $Z(x; \varepsilon)$ and the set $\mathcal{A}_{\alpha p}$ satisfy the assumptions of Theorem 2.1 of [6]. More precisely, we show that they satisfy the assumptions of Theorem A of Appendix which is an equivalent version of the result of [6].

First of all we note that assumption (ii) of Theorem A follows from part (a) of Proposition 3. Since $Z(x; \varepsilon) \rightarrow \infty$ for $x \rightarrow \partial\mathcal{A}_{\alpha p}$ we have that also assumption (iii) holds. Then part (a) of Theorem 1, Proposition 5 and Proposition 6 state that, for sufficiently small values of the penalty parameter ε, every stationary point (every strict local minimum point) of $Z(x; \varepsilon)$ on $\mathcal{A}_{\alpha p}$ is a KKT point (a strict local minimum point) of Problem (P) and vice versa. Therefore assumption (iv) of this theorem implies, for sufficiently small ε, assumption (iv) of Theorem A.

Finally we have to show that also assumption (i) of Theorem A holds, namely we have to prove that the set $\mathcal{A}_{\alpha p}$ is connected. We prove this property by proceeding by contradiction. We assume that $\mathcal{A}_{\alpha p}$ is disconnected. Since the feasible set $\mathcal{F}$ is connected there exist two open sets A_1, A_2 such that $\mathcal{A}_{\alpha p} \cap A_1$ and $\mathcal{A}_{\alpha p} \cap A_2$ are disjoint, nonempty sets whose union is $\mathcal{A}_{\alpha p}$ and such that $\mathcal{F} \subseteq \mathcal{A}_{\alpha p} \cap A_1$. Now, the compactness of $\bar{\mathcal{A}}_{\alpha p}$ implies that also the set $\bar{\mathcal{A}}_{\alpha p} \cap \bar{A}_2$ is a compact set. Now we introduce the function $s(x) = 1/a(x)$ (where $a(x)$ is defined in (1)). Since $s(x) \rightarrow \infty$ if $x \rightarrow \partial\mathcal{A}_{\alpha p}$, this function admits a global minimum point $\hat{x}$ on $\mathcal{A}_{\alpha p} \cap \bar{A}_2$. Moreover, if $\hat{x} \in \mathcal{A}_{\alpha p} \cap \partial\bar{A}_2$ this would imply, since $\mathcal{A}_{\alpha p} \subseteq A_1 \cup A_2$, that $\hat{x}$ belongs to the open set $\mathcal{A}_{\alpha p} \cap A_1$ this would contradict the fact that $\mathcal{A}_{\alpha p} \cap A_1$ and $\mathcal{A}_{\alpha p} \cap A_2$ are disjoint. Therefore we can conclude that $\hat{x}$ is an unconstrained minimum point of the function $s(x)$ on $\mathcal{A}_{\alpha p} \cap A_2$ and, hence, recalling (1), we must have:

$$\nabla s(\hat{x}) = \frac{1}{a(\hat{x})^2}\left[p \sum_{i=1}^{m} \nabla g_i(\hat{x}) g_i^+(\hat{x})^{p-1} + 2\nabla h(\hat{x}) h(\hat{x})\right] = 0 .$$

which, by recalling that $\bar{\mathcal{A}}_{\alpha p}$ satisfy assumption (ii) of Proposition 5, implies that $\hat{x} \in \mathcal{F}$ and this establishes a contradiction with the fact that $\mathcal{F} \subseteq \mathcal{A}_{\alpha p} \cap A_1$.

At this point we can apply Theorem A of the Appendix and we can conclude that function $Z(x; \varepsilon)$ has only one stationary point on $\mathcal{A}_{\alpha p}$ which is a global minimum point. The proof of the theorem follows from Theorem 2 and again from part (a) of Theorem 1, Proposition 5 and Proposition 6. □

Assumptions (i)–(iii) appear to be mild requirements on the constraint functions. In particular, as regards the less familiar assumption (ii), Proposition 2.1 of [15] shows some conditions under which this assumption holds for every $\alpha > 0$ and $p \geq 2$, namely:
- if all the variables are bounded;

– if there exists either a function $g_i(x)$ such that $\lim_{\|x\|\to\infty} g_i(x) \to \infty$ or a function $h_i(x)$ such that $\lim_{\|x\|\to\infty} \|h_i(x)\| \to \infty$;
– if there exists an index set J such that the functions $g_i(x)$, $i \in J$ are convex and the set $\{x \in \mathbb{R}^n: g_i(x) \leq 0, i \in J\}$ is compact.

From the proof of Theorem 3 we can derive also the following result.

COROLLARY 1. *Assume that:*
(i) *there exist scalars $\alpha > 0$ and $p \geq 2$ such that the set $\mathcal{A}_{\alpha p}$ is bounded, nonempty and connected;*
(ii) *Assumption A3 and assumption (ii) of Proposition 5 hold on $\mathcal{A}_{\alpha p}$;*
(iii) *for every $x \in \mathcal{F}$ the gradients $\nabla g_i(x)$, $i \in I_0(x)$, $\nabla h_j(x)$, $j = 1, \ldots, q$ are linearly independent;*
(iv) *every KKT point of Problem (P) is a strict local minimum point.*

Then, there exists only one KKT point which is a global minimum point for Problem (P).

4. Constrained Global Minimization Algorithms

In this section we describe one of the possible uses of the penalty function Z to define algorithms to solve constrained global minimization problems.

Apart from very particular and simple cases, we cannot find the global minimizer of Problem (P) by applying directly an unconstrained global minimization algorithm to the penalty function $Z(x; \varepsilon)$. In fact, as we have seen in Section 2, Problem (P) is equivalent to the unconstrained minimization of function $Z(x; \varepsilon)$ only if the value of the penalty parameter ε is less or equal to a threshold value ε^*. The problem is that the value ε^* is not known *a priori*. Therefore, when we set a value for ε we can choose a wrong value (namely $\varepsilon > \varepsilon^*$) and, hence, the global minimizer of $Z(x; \varepsilon)$ can have no connection with the global solution of Problem (P). In order to increase the likelihood of using a right value of the penalty parameter (that is $\varepsilon \leq \varepsilon^*$) we should choose very small values of ε. However, such values of the penalty parameters would make the penalty function $Z(x; \varepsilon)$ so ill-conditioned that any minimization algorithm would not be able, in practice, to locate the minimizers of this function.

Therefore we must define *ad hoc* algorithms that overcome this problem of the choice of the penalty parameter (without assuming the knowledge of the threshold value ε^*).

Many of the most efficient algorithms in the field of the unconstrained global optimization (see, e.g., tunneling methods, clustering methods, multistart algorithms, simulated annealing algorithms and so on) consist of a global strategy, which tries to locate the region of attraction of a global minimum point x^*, and a local strategy, which tries to determine the point x^* by using a local minimization algorithm.

As regards the local strategy, in the unconstrained case it is possible to exploit the fact that, under suitable assumptions, every local minimum point x^* (and hence also the global minimum point) is an attraction point for a local minimization algorithm, namely that there exists a neighborhood B of x^* such that the sequence of points produced by a local minimization algorithm, starting from any point of B, converges to x^* (see, for example, Proposition 1.12 of [19] which states that unconstrained strong local minima tend to attract gradient-related methods).

In the constrained case, it is not immediate to prove a similar property. In fact if we want to use the penalty function $Z(x; \varepsilon)$ to determine the local minima of Problem (P), since the threshold value ε^* is not known, we have to recourse to algorithms which include automatic adjustment rules for the penalty coefficient which appears in the function $Z(x; \varepsilon)$ (see, for example, the algorithms proposed in [14] and [15]). These algorithms change the objective function during the minimization procedure and hence, in general, it is not possible to show that they are attracted by global minima of Problem (P). However, following the approach proposed in [14] and [15], we can describe a particular algorithm model that enjoys this property. This algorithm model employs a general unconstrained minimization algorithm described by the iteration map $\mathcal{M}$: $\mathcal{A}_{\alpha p} \rightarrow 2^{\mathcal{A}_{\alpha p}}$ and such that it satisfies the following assumption.

ASSUMPTION M. *For every fixed value of ε and every starting point $x_0 \in \mathcal{A}_{\alpha p}$ the iteration map $\mathcal{M}$ produces a sequence of points $\{x_k\}$ such that:*

(i) all the points x_k belong to $\Omega_{\alpha p}(\tilde{x})$;

(ii) there exists a positive constant γ such that for all k

$$\|x_{k+1} - x_k\| \leq \gamma \|\nabla Z(x_k; \varepsilon)\| ;$$

(iii) all the limit points of the sequence $\{x_k\}$ are stationary points of $Z(x; \varepsilon)$. □

Furthermore, as stopping criterion, the proposed algorithm model makes use of the following function:

$$\phi(x) := (\|\nabla L(x, \lambda(x), \mu(x))\|^2 + \|G(x)\lambda(x)\|^2 + \|h(x)\|^2 + \|g^+(x)\|^2 + \|\lambda^-(x)\|^2)^{1/2}, \tag{9}$$

where $\lambda^-(x)$ is the vector with component $\lambda_i^-(x) := \min[0, \lambda_i(x)]$.

ALGORITHM LA.

Data: $\tilde{x} \in \mathbb{R}^n$, $\eta_1 > 0$, $\eta_2 > 0$, $\eta_3 > 0$ and $p \geq 2$.

Step 0: If $\phi(\tilde{x}) = 0$ stop; else set $\varepsilon = \eta_1 \min[1, \phi(\tilde{x})]$, $\alpha = \eta_2 \phi(\tilde{x}) + \Sigma_{i=1}^m g_i^+(\tilde{x})^p + \Sigma_{j=1}^q h_j(\tilde{x})^2$ and $z = \tilde{x}$.

Step 1: Set $k = 0$. If $Z(\tilde{x}; \varepsilon) \leq Z(z; \varepsilon)$ set $x_0 = \tilde{x}$ else set $x_0 = z$.

Step 2: If

$$\|\nabla Z(x_k;\varepsilon)\|^2+\|\nabla g(x_k)'\nabla Z(x_k;\varepsilon)\|^2+\|\nabla h(x_k)'\nabla Z(x_k;\varepsilon)\|^2$$
$$<\eta_3(\|c(x_k;\varepsilon)\|^2+\|h(x_k)\|^2)$$

choose $\varepsilon \in (0, \varepsilon)$, set $z = x_k$ and go to step 1.

Step 3: Compute $x_{k+1} \in \mathcal{M}[x_k]$ and set $k = k+1$.

Step 4: If $\phi(x_k) = 0$ stop; else go to step 2.

The convergence properties of Algorithm LA are stated in the following proposition.

PROPOSITION 7.

(i) If Algorithm LA updates the penalty parameter ε a finite number of times then, either the algorithm terminates at some $x_\nu \in \mathcal{A}_{\alpha p}$ and $(x_\nu, \lambda(x_\nu), \mu(x_\nu))$ is a K–T triple for Problem (P), or the algorithm produces an infinite sequence $\{x_k\} \subseteq \mathcal{A}_{\alpha p}$ such that every limit point x^* yields a K–T triple $(x^*, \lambda(x^*), \mu(x^*))$ for Problem (P).

(ii) If Algorithm LA updates the penalty parameter ε an infinite number of times then there exists, at least, a limit point x^* of the produced sequence $\{x_k\}$ where Assumption A3 is not satisfied.

(iii) If either $\tilde{x}$ or Assumption A3 holds then Algorithm LA updates the penalty parameter ε a finite number of times.

Proof. The proof follows with minor modifications from those of Proposition 5.1, Theorem 5.2 and Proposition 5.3 of [15]. □

As regards Assumption M on the map M, it is easy to see that it can be satisfied by almost all globally convergent algorithm for the unconstrained minimization of Z. In fact it is enough to ensure, by simple device, that the trial points (produced along the search direction) remain in $\Omega_{\alpha p}(\tilde{x})$.

Algorithm LA is very similar to the one proposed in [15], in particular we have defined the choices of the initial penalty parameter and the parameter α. These particular choices are very important because they allow us to ensure that Algorithm LA, unlike the algorithms proposed in [14] and [15], is "attracted" by a strong local minimum of Problem (P). In fact we can state the following result.

PROPOSITION 8. *Assume that:*

(i) there exist scalars $\hat{\alpha} > 0$ and $p \geq 3$ such that the set $\mathcal{A}_{\hat{\alpha}p}$ is bounded;

(ii) for every $x \in \mathcal{F}$ the gradients $\nabla g_i(x)$, $i \in I_0(x)$, $\nabla h_j(x)$, $j = 1, \ldots, q$ are linearly independent.

Then for every local minimum point x^* of Problem (P) that satisfies the strict complementarity and the second order sufficiency conditions there exists an open set S containing x^* such that, if $\tilde{x} \in S$, Algorithm LA, starting from $\tilde{x}$, produces a sequence of points $\{x_k\}$ which remains in S and converges to x^*.

Proof. First of all, Proposition 1 implies that there exists an $\bar{\alpha} \leq \hat{\alpha}$ such that for all $\alpha \in (0, \bar{\alpha}]$ the corresponding set $\mathcal{A}_{\alpha p}$ satisfies Assumptions A1 and A3. Then, by continuity, there exists a $\sigma_0 > 0$ such that if $\tilde{x} \in B(x^*; \sigma_0) := \{x: \|x - x^*\| < \sigma_0\}$ then Step 0 of Algorithm LA chooses

$$\alpha = \eta_2 \phi(\tilde{x}) + \sum_{i=1}^{m} g_i^+(\tilde{x})^p + \sum_{j=1}^{q} h_i(\tilde{x})^2 \leq \bar{\alpha} ,$$

and, hence, the chosen set $\mathcal{A}_{\alpha p}$ satisfies Assumptions A1 and A3.

Then, by Lemma 4.2 of [15], we have that there exist numbers $\varepsilon(x^*) > 0$, $\sigma(x^*) > 0$ (with $\sigma(x^*) \leq \sigma_0$) and $\rho(x^*) > 0$ such that, for all $\varepsilon \in (0, \varepsilon(x^*)]$ and for all $x \in \mathcal{A}_{\alpha p}$ satisfying $\|x - x^*\| \leq \sigma(x^*)$, the following formula holds:

$$\|\nabla g(x)'\nabla Z(x; \varepsilon)\|^2 + \|\nabla h(x)'\nabla Z(x; \varepsilon)\|^2 \geq \frac{\rho(x^*)}{\varepsilon^2} (\|c(x; \varepsilon)\|^2 + \|h(x)\|^2) . \tag{10}$$

By the continuity assumptions, it follows that there exist a $\sigma_1 > 0$, with $\sigma_1 \leq \sigma(x^*)$, such that for all $x \in B(x^*; \sigma_1) := \{x: \|x - x^*\| < \sigma_1\}$ the following inequality holds:

$$\phi(x) < \frac{1}{\eta_1} \min[\varepsilon^*, \varepsilon(x^*), \hat{\varepsilon}] , \tag{11}$$

where

$$\hat{\varepsilon} = \left(\frac{\rho(x^*)}{\eta_3}\right)^{1/2} , \tag{12}$$

and $\phi(x)$ is defined by (9), η_1 and η_2 are the constants used in Algorithm LA, and ε^* is the number considered in Proposition 4.

Now let us suppose that $x_0 = \tilde{x} \in B(x^*; \sigma_1)$. In this case, at Step 1, Algorithm LA chooses an initial penalty parameter $\varepsilon = \tilde{\varepsilon}$ such that, by using (11) and Proposition 4, the point x^* is an isolated local minimum point for $Z(x; \tilde{\varepsilon})$ and the Hessian matrix $\nabla^2 Z(x^*; \tilde{\varepsilon})$ is positive definite. Then (10), (11) and (12) imply that the test at Step 2 is not satisfied by x_0 and $\tilde{\varepsilon}$. More in general, if the points x_k produced by Algorithm LA starting from x_0 remain in $B(x^*; \sigma_1)$ then the penalty parameter is never updated.

At this point, the proof of the proposition can be carried out by reasoning as in the proof of Proposition 1.12 of [19] where a similar result is proposed for the case of unconstrained minimization algorithms.

We can find a $\sigma_2 > 0$, with $\sigma_2 < \sigma_1$, such that, for all $x \in \bar{B}(x^*; \sigma_2) := \{x: \|x - x^*\| \leq \sigma_2\}$, the function $Z(x; \tilde{\varepsilon})$ is twice differentiable and the matrix $\nabla^2 Z(x; \tilde{\varepsilon})$ is positive definite. Then we define:

$$\theta_1 = \min_{x \in \bar{B}(x^*;\sigma_2)} \sigma_m(\nabla^2 Z(x;\bar{\varepsilon})) , \quad \theta_2 = \max_{x \in \bar{B}(x^*;\sigma_2)} \sigma_M(\nabla^2 Z(x;\bar{\varepsilon})) ,$$

where $\sigma_m(\nabla^2 Z(x;\bar{\varepsilon}))$ and $\sigma_M(\nabla^2 Z(x;\bar{\varepsilon}))$ are respectively the smallest and the largest eigenvalue of $\nabla^2 Z(x;\bar{\varepsilon})$.

Now we introduce the open set:

$$S := \left\{ x: \|x - x^*\| < \sigma_2, \; Z(x;\bar{\varepsilon}) < Z(x^*;\bar{\varepsilon}) + \frac{\theta_1}{2}\left(\frac{\sigma_2}{1+\gamma\theta_2}\right)^2 \right\},$$

and we prove that if $x_0 \in S$ then Algorithm LA, starting from x_0, produces points x_k that belong to open set S.

First we recall that, since $x_0 \in S \subseteq B(x^*;\sigma_1)$, the test at Step 2 is not satisfied and, hence, the penalty parameter is not updated. Then, by using Taylor's theorem, we obtain:

$$\|\nabla Z(x_0;\bar{\varepsilon})\| \le \theta_2 \|x_0 - x^*\| ,$$

$$\|x_0 - x^*\| \le \left[\frac{2}{\theta_1}(Z(x_0;\bar{\varepsilon}) - Z(x^*;\bar{\varepsilon}))\right]^{1/2} \le \frac{\sigma_2}{1+\gamma\theta_2} ,$$

which, recalling Assumption M, implies:

$$\begin{aligned}\|x_1 - x^*\| &\le \|x_0 - x^*\| + \|x_1 - x_0\| \le \|x_0 - x^*\| + \gamma\|\nabla Z(x_0;\bar{\varepsilon})\| \\ &\le (1+\gamma\theta_2)\|x_0 - x^*\| \le \sigma_2 .\end{aligned}$$

Moreover, by assumption, we have

$$Z(x_1;\bar{\varepsilon}) \le Z(x_0;\bar{\varepsilon}) < Z(x^*;\bar{\varepsilon}) + \frac{\theta_1}{2}\left(\frac{\sigma_2}{1+\gamma\theta_2}\right)^2 ,$$

and, therefore, we obtain that $x_1 \in S$.

Suppose now that for a given k we have that $x_k \in S$. Using the same arguments used for the first iteration of Algorithm LA and taking into account that, by assumption, $Z(x_{k+1};\bar{\varepsilon}) \le Z(x_0;\bar{\varepsilon})$, we obtain that $x_{k+1} \in S$. Therefore we can conclude that $x_k \in S$ for all k.

The preceding part of the proof implies that every limit points of sequence $\{x_k\}$ belongs to $\bar{S}$ (the closure of S). Moreover, by using the assumptions made on the iteration map $\mathcal{M}$ of Step 3, we have that every limit points of $\{x_k\}$ is stationary point of $Z(x;\bar{\varepsilon})$. Since $\bar{S} \subseteq \bar{B}(x^*;\sigma_2)$, $Z(x;\bar{\varepsilon})$ is a strictly convex function within $\bar{S}$ and, hence, the point x^* is the unique critical point of $Z(x;\bar{\varepsilon})$ within $\bar{S}$. Therefore we can conclude that the sequence $\{x_k\}$ converges to x^*. □

The preceding proposition shows that, for every strong global minimum, there exists a neighborhood such that Algorithm LA, starting from any point in this neighborhood, converges to global minimum. This property ensures that Algorithm LA can be successfully used, as local strategy, in a constrained global minimization algorithm.

As regards the problem to locate the region of attraction of the global minimum, we can again draw our inspiration to the unconstrained case and we can define several different global strategies based on the penalty function $Z(x; \varepsilon)$. However we must take into account that, also for the global strategy, we have the problem of the choice of the penalty parameter ε.

Here, as an example, we study the possibility to define an algorithm which combines the simulated annealing approach with Algorithm LA. It consists of several local minimizations and its aim, as usual, is to let the local minimizations start only when the point generated at random appears to be sufficiently promising and, hence, to spare useless local minimizations. In particular we follow the approach proposed in [16] which was inspired by the works on the simulated annealing approach to continuous global optimization [20–22].

ALGORITHM GA.

Data: $\bar{x} \in \mathbb{R}^n$, $\bar{\varepsilon} > 0$, $\bar{T} > 0$.

Step 0: Choose $\hat{\alpha} > 0$ and $p \geq 3$ such that $\bar{x} \in \mathcal{A}_{\bar{\alpha}p}$, set $k = 1$, $\hat{x}_k = x_k^0 = \bar{x}$, $\varepsilon_k = \bar{\varepsilon}$ and $T_k = \bar{T}$.

Step 1: Compute a new point x_k by using Algorithm LA starting from x_k^0.

Step 2: If $(x_k, \lambda(x_k), \mu(x_k))$ is a K–T triple for Problem (P) and $f(x_k) < f(\hat{x}_k)$ set $\hat{x}_{k+1} = x_k$; else set $\hat{x}_{k+1} = \hat{x}_k$.

Step 3: Choose $\varepsilon_{k+1} \in (0, \varepsilon_k]$ and $T_{k+1} \in (0, T_k]$, set $k = k + 1$.

Step 4: Set $i = 1$.

Step 5: Generate the independent random variables Y_k^i and W_k^i uniformly distributed in $\mathcal{A}_{\alpha p}$ and $[0, 1]$ respectively, independently of all those previously generated.

Step 6: If

$$W_k^i \leq e^{-[Z(Y_k^i; \varepsilon_k) - Z(\hat{x}_k; \varepsilon_k)]^+ / T_k}$$

set $i_k = i$, $x_k^0 = Y_k^{i_k}$ and go to Step 1; else set $i = i + 1$ and go to Step 5. □

Steps 4–6 of the preceding algorithm constitute the Von Neumann's acceptance-rejection method that generates (see Theorem 3.4.1 of [23] or Theorem 2.2 of [16]) a sample $Y_k^{i_k}$ distributed according the following probability density function:

$$\pi(x; \varepsilon_k, T_k) := \frac{e^{-[Z(x; \varepsilon_k) - Z(\hat{x}_k; \varepsilon_k)]^+ / T_k}}{\int_{\mathcal{A}_{\alpha p}} e^{-[Z(x; \varepsilon_k) - Z(\hat{x}_k; \varepsilon_k)]^+ / T_k} \, dx}.$$

Therefore as T_k becomes small, Algorithm GA tends to generate points $Y_k^{i_k}$ such that $Z(Y_k^{i_k}; \varepsilon_k)$. Therefore this acceptance-rejection technique produces a kind of random tunnelling effect on the penalty function $Z(x; \varepsilon_k)$.

Now we want to prove that Algorithm GA converges to a global minimum of Problem (P). Since Algorithm GA is strongly based on the penalty function $Z(x; \varepsilon_k)$, intuitively it would seem necessary to assume that $\varepsilon_k \leq \varepsilon^*$, where ε^* is the threshold value which ensures the correspondence between problem (P) and the unconstrained minimization of $Z(x; \varepsilon_k)$. Therefore the problems described for the local strategies should arise again. On the contrary, quite surprisingly, the following theorem shows that Algorithm GA converges in probability to a global minimum point of Problem (P) for every bounded sequence $\{\varepsilon_k\}$ (note, in particular, that it is possible to choose $\varepsilon_k = \bar{\varepsilon}$ for all k).

THEOREM 4. *Suppose that:*

(i) the set $\mathcal{A}_{\hat{\alpha}p}$ is bounded;

(ii) for every $x \in \mathcal{F}$ the gradients $\nabla g_i(x)$, $i \in I_0(x)$, $\nabla h_j(x)$, $j = 1, \dots, q$ are linearly independent;

(iii) every global minimum point of Problem (P) satisfies the strict complementarity and the second order sufficiency conditions.

Then the sequence of points $\{\hat{x}_k\}$ produced by Algorithm GA converges in probability to a global minimum point of Problem (P).

Proof. Again Proposition 1 ensures that there exists an $\alpha \leq \hat{\alpha}$ such that the corresponding set $\mathcal{A}_{\alpha p}$ satisfies Assumptions A1 and A3. Now let S be the set of global minimum points of Problem (P), that is:

$$S := \{x^* \in \mathcal{F}\colon f(x^*) \leq f(x), \text{ for all } x \in \mathcal{F}\} .$$

By (i) we have that S is a compact set and by using also (iii) we obtain that S has a finite number of elements.

Recalling Corollary 2.1 of [16] (see also [24]), in order to prove the theorem we have only to show that if $\hat{x}_i \notin S$, for $i = 1, \dots, k$, then, for any $\delta > 0$ sufficiently small there exists a positive constant γ_1 such that:

$$e^{-[Z(x;\, \varepsilon_k) - Z(\hat{x}_k;\, \varepsilon_k)]^+ / T_k} \geq \gamma_1 \tag{13}$$

for all x such that $Z(x; \varepsilon_k) \leq Z(x^*; \varepsilon_k) + \delta$, with $x^* \in S$.

Let $\bar{\varepsilon}$ be a sufficiently small value of the penalty parameter such that Theorem 1, Theorem 2 and Proposition 4 hold.

By Theorem 2 we have:

$$S = \{x^* \in \mathcal{A}_{\alpha p}\colon Z(x^*; \bar{\varepsilon}) \leq Z(x; \bar{\varepsilon}), \text{ for all } x \in \mathcal{A}_{\alpha p}\} .$$

By using assumption (iii) and Proposition 4 we obtain that, for every $x^* \in S$,

$Z(x; \bar{\varepsilon})$ is strictly convex within a neighborhood of x^*, namely there exists a positive constant σ_{x^*} such that:

$$\nabla Z(x; \bar{\varepsilon}) \neq 0 , \quad \text{for all} \quad x : \|x - x^*\| < \sigma_{x^*} , \quad x \neq x^* . \tag{14}$$

Now, we define the following set:

$$\mathcal{T} := \bigcap_{x^* \in S} \{x : \|x - x^*\| \geq \sigma_{x^*}\} .$$

Recalling that $\hat{x}_k$ is a KKT point, by Theorem 1 we have that $\nabla Z(\hat{x}_k ; \bar{\varepsilon}) = 0$ and hence, recalling (14), we conclude that $\hat{x}_k \in \mathcal{T}$.

Since the set $\bar{\mathcal{A}}_{\alpha p} \cap \mathcal{T}$ is compact and $Z(x; \bar{\varepsilon}) \to \infty$ as $x \to \partial \mathcal{A}_{\alpha p}$ there exists $\tilde{x}$ such that:

$$Z(\tilde{x}; \bar{\varepsilon}) = \min_{x \in A_{\alpha p} \cap \mathcal{T}} Z(x; \bar{\varepsilon}) .$$

Now, the fact that $\mathcal{T} \cap S = 0$ implies that $Z(x^*; \bar{\varepsilon}) < Z(\tilde{x}; \bar{\varepsilon})$ for all $x^* \in S$.

Taking into account that $\hat{x}_k \in \mathcal{T}$ we have

$$Z(x^*; \bar{\varepsilon}) < Z(\tilde{x}; \bar{\varepsilon}) \leq Z(\hat{x}_k ; \bar{\varepsilon})$$

and recalling again Theorem 1 we obtain:

$$Z(x^*; \varepsilon_k) = f(x^*) = Z(x^*; \bar{\varepsilon}) < Z(\tilde{x}; \bar{\varepsilon}) \leq Z(\hat{x}_k ; \bar{\varepsilon}) = f(\hat{x}_k) = Z(\hat{x}_k ; \varepsilon_k) .$$

Therefore, finally, for any $\delta \leq Z(\tilde{x}; \bar{\varepsilon}) - Z(x^*; \bar{\varepsilon}) \leq Z(\hat{x}_k ; \varepsilon_k) - Z(x^*; \varepsilon_k)$ we have:

$$e^{-[Z(x; \varepsilon_k) - Z(\hat{x}_k; \varepsilon_k)]^+ / T_k} = 1$$

for all x such that $Z(x; \varepsilon_k) \leq Z(x^*; \varepsilon_k) + \delta$, with $x^* \in S$. □

The preceding theorem confirms that the Algorithm LA and the strong exactness properties of the function Z allow to obtain the same properties of convergence obtained in [16] for the unconstrained case.

Finally we conclude this section by pointing out that the use of an exact penalty function could be essential also to define stopping criteria for constrained global optimization methods. In fact, since these functions transform the original problem into an unconstrained one, they could allow to use directly the many interesting stopping criteria proposed in literature (see, e.g. [25–30]) for the unconstrained global minimization problems. With regard to Algorithm LA, since it is based on an acceptance-rejection technique which produces uniformly distributed points, these points could be used for any kind of sophisticated statistical analysis. Therefore, various stopping rules proposed in literature could be easily modified and used in Algorithm LA.

Acknowledgement

I thank Dr. A. De Santis and Dr. M. Roma for helpful discussions and the anonymous referees for their constructive comments and remarks, which led to improve very much the paper.

Appendix

For the sake of completeness, in this appendix we recall Theorem 2.1 of [6]. In particular we report a slightly modified version of this theorem. The proof of this version follows immediately, with minor modification from the proof of the original result.

THEOREM A. *Let $\hat{f}$ be a function from $\mathcal{S} \to \mathbb{R}$ and assume that:*

(i) $\mathcal{S}$ is open, connected, and not empty;

(ii) $\hat{f}$ is continuously differentiable in $\mathcal{S}$;

(iii) for every $\sigma > 0$, there exists a compact subset K_σ of $\mathcal{S}$, so that $\hat{f}(x) \geq \sigma$ for all $x \notin K_\sigma$;

(iv) every stationary point of $\hat{f}$ on $\mathcal{S}$ is a strict local minimum point.

Then, there exists only one stationary point which is a global minimum point of $\hat{f}$ on $\mathcal{S}$.

References

1. L. C. W. Dixon and G. P. Szegö (1975), *Towards Global Optimization 1*, North-Holland, New York.
2. L. C. W. Dixon and G. P. Szegö (1978), *Towards Global Optimization 2*, North-Holland, New York.
3. F. Archetti and F. Shoen (1984), A Survey on the Global Optimization Problems: General Theory and Computational Approaches, *Annals of Operations Research* **1**, 87–110.
4. A. Törn and A. Žilinskas (1989), *Global Optimization*, Springer-Verlag, Lecture Notes in Computer Science 350.
5. A. H. G. Rinnooy Kan and G. H. Timmer (1989), Global Optimization, in G. L. Nemhauser, A. H. G. Rinnooy Kan, and M. J. Todd (eds.), *Optimization*, North-Holland, Amsterdam.
6. G. Gabrielsen (1986), Global Maxima of Real-Valued Functions, *Journal of Optimization Theory and Applications* **50**, 257–266.
7. P. M. Pardalos and J. B. Rosen (1986), Methods for Global Concave Optimization: A Bibliografic Survey, *SIAM Review* **28**, 367–379.
8. P. M. Pardalos and J. B. Rosen (1987), *Constrained Global Optimization*, Springer-Verlag, Lecture Notes in Computer Science 268, 198.
9. R. Horst and H. Tuy (1990), *Global Optimization. Deterministic Approaches*, Springer-Verlag, Berlin.
10. G. T. Timmer (1984), Global Optimization: A Stochastic Approach, Ph.D. Dissertation, Erasmus University Rotterdam.
11. A. V. Levy and S. Gomez (1984), The Tunneling Algorithm Applied to Global Optimization, in P. T. Boggs, R. H. Byrd, and R. B Schnabel (eds.), *Numerical Optimization 1984* (SIAM Philadelphia), pp. 213–244.
12. S. Palosaari, S. Parviainem, J. Hiironen, J. Reunanen, and P. Neittaanmäki (1986), A Random

Search Algorithm for Constrained Global Optimization, *Acta Polytech. Scand., Ch. Techn. and Metallurgy Ser.* **172**, 1–45.
13. V. N. Neifiodov (1987), The Search of the Global Maximum of a Function of Several Variables on a Set Defined by Inequality Bounds, *Comp. Math. and Math. Phys.* **27**(1), 35–51. (In Russian)
14. G. Di Pillo and L. Grippo (1986), An Exact Penalty Method with Global Convergence Properties for Nonlinear Programming Problems, *Math. Prog.* **36**, 1–18.
15. S. Lucidi (1992), New Results on a Continuously Differentiable Exact Penalty Function, *SIAM J. on Optimization* **2**, 558–574.
16. S. Lucidi and M. Piccioni (1989), Random Tunneling by Means of Acceptance-Rejection Sampling for Global Optimization, *J. Optim. Theory and Appl.* **62**, 255–278.
17. S. Lucidi (1989), New Results on a Continuously Differentiable Exact Penalty Function, Tech. Rep. R. 277, Istituto di Analisi dei Sistemi ed Informatica del CNR, Roma, Italy.
18. G. Di Pillo and L. Grippo (1985), A Continuously Differentiable Exact Penalty Function for Nonlinear Programming Problems with Inequality Constraints, *SIAM J. Control and Optimization* **23**, 72–84.
19. D. P. Bertsekas (1982), *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York.
20. F. Aluffi-Pentini, V. Parisi, and F. Zirilli (1985), Global Optimization and Stochastic Differential Equations, *J. Optim. Theory Appl.* **47**, 1–16.
21. M. Piccioni (1987), *A Combined Multistart-Annealing Algorithm for Continuous Global Optimization*, Systems Research Center, University of Maryland, Tech. Res. Rep. 87-45.
22. S. German and C.-R. Hwang (1986), Diffusions for Global Optimization, *SIAM J. Control and Optimization* **24**, 1031–1043.
23. R. Y. Rubinstein (1981), *Simulation and the Monte Carlo Method*, John Wiley & Son, New York.
24. F. J. Solis and R. J. B. Wets (1981), Minimizations by Random Search Techniques, *Math. of Oper. Res.* **6**, 19–30.
25. D. C. Karnopp (1963), Random Search Techniques for Optimization Problems, *Automatica* **1**, 111–121.
26. A. Törn (1978), A Search Clustering Approach to Global Optimization, in L. C. W. Dixon and G. P. Szegö (eds.), *Towards Global Optimization 2*, North-Holland, New York.
27. R. Zieliński (1981), A Statistical Estimate of the Structure of Multiextremal Problems, *Math. Prog.* **21**, 348–356.
28. B. Betrò (1984), Bayesian Testing of Nonparametric Hypotheses and its Application to Global Optimization, *J. Optim. Theory and Appl.* **42**, 31–50.
29. B. Betrò and F. Schoen (1987), Sequential Stopping Rules for the Multistart Algorithm in Global Optimization, *Math. Prog.* **38**, 271–280.
30. C. G. E. Boender and A. H. G. Rinnooy Kan (1987), Bayesian Stopping Rules for Multistart Global Optimization Methods, *Math. Prog.* **37**, 59–80.