Journal of Global Optimization (2005) 32:307–323
DOI 10.1007/s10898-004-2707-6

Robust Solution of Nonconvex Global Optimization Problems

HOANG TUY
Institute of Mathematics, 18 Hoang Quoc Viet Road, 10307, Hanoi, Vietnam
(e-mail: htuy@math.ac.vn)

(Received 1 April 2004; accepted in final form 7 April 2004)

Abstract. The concept of ε-approximate optimal solution as widely used in nonconvex global optimization is not quite adequate, because such a point may correspond to an objective function value far from the true optimal value, while being infeasible. We introduce a concept of essential ε-optimal solution, which gives a more appropriate approximate optimal solution, while being stable under small perturbations of the constraints. A general method for finding an essential ε-optimal solution in finitely many steps is proposed which can be applied to d.c. programming and monotonic optimization.

Key words: Essential ε-optimal solution, Incumbent transcending approach, Pitfall in approximate nonconvex global optimization, Robust global optimization, ε-approximate optimal solution.

1. Introduction

A wide class of global optimization problems have the form

$$\min\{f(x)|h(x)\geqslant 0,\ x\in D\}, \tag{P}$$

where f, h are real-valued continuous functions on $\mathbb{R}^n$; and D is a nonempty robust compact set in $\mathbb{R}^n_+$, i.e. such that

$$D=\mathrm{cl}(\mathrm{int}D)\neq\emptyset.$$

Most often $D=\{x|g(x)\leqslant 0\}$, with $g(x)$ being a real-valued continuous function. Two typical cases of this problem that have been studied in the recent literature are the *canonical reverse convex programming problem* [6] (when f, g, h are convex functions) and the *canonical monotonic optimization Problem* [7] (when f, g, h are increasing functions; recall from [7] that a function $f:\mathbb{R}^n_+\to\mathbb{R}$ is said to be increasing if $f(x')\geqslant f(x)$ whenever $x'\geqslant x\geqslant 0$): With rare exceptions, solution methods so far developed for problem (P) compute (in fact) a global optimal solution $\bar{x}(\varepsilon)\in X$ of the ε-relaxed problem

$$\min\{f(x)|h(x)+\varepsilon\geqslant 0,\ x\in D_\varepsilon\}, \tag{P$_\varepsilon$}$$

where $D_\varepsilon=\{x|g(x)\leqslant\varepsilon\}$ and ε is a small positive number (see e.g. [1–4, 6],...). Note that when the constraint $h(x)\geqslant 0$ replaces a system like

$h_i(x) \geqslant 0$ $(i = 1, \ldots, m_1)$, $k_j(x) \leqslant 0$ $(j = 1, \ldots, m_2)$, $l_q(\mathrm{x}) = 0$ $(q = 1, \ldots, m_3)$, then the ε-relaxation is $h_i(x) \geqslant -\varepsilon$, $(i = 1, \ldots, m_1)$, $k_j(x) \leqslant \varepsilon$, $(j = 1, \ldots, m_2)$, $|l_q(x)| \leqslant \varepsilon(q = 1, \ldots, m_3)$. Since $\bar{x}(\varepsilon)$ is almost feasible for $\varepsilon > 0$ small and any cluster point of the sequence $\{\bar{x}(\varepsilon), \varepsilon > 0\}$ gives a global optimal solution of (P), each $\bar{x}(\varepsilon)$ is accepted as an approximate global optimal solution of (P) with tolerance $\varepsilon > 0$ and is usually referred to as an *ε-approximate optimal solution*. Although this has long been a common practice in global optimization, it turns out, as we will show in the sequel, that the concept of ε-approximate optimal solution may be quite inadequate. Actually, an ε-approximate optimal solution may happen to be quite far from the true optimum.

A more appropriate concept of approximate optimal solution is that of *η-optimal solution*. A vector x^* is called an η-optimal solution of problem (P) if it is feasible and satisfies $f(x^*) - f(x) \leqslant \eta$ for all feasible solutions x. [other variations of this concept exist: for instance a feasible solution x^* is called η-optimal if $f(x^*) - f(x) \leqslant \eta(1 + f(x^*))$ for all feasible solutions x.]

Typically, to solve (P) an infinite sequence of feasible solutions $\{x^k\}$ is generated such that $f(x^k) \leqslant \min\{f(x)|h(x) \geqslant 0, x \in D\} + \eta_k$, with $\eta_k \to 0$, so that x^{k} is an η-optimal solution when $\eta_k < \eta$. However, if the optimal solution happens to be an isolated feasible point, i.e. a point x which is the center of a ball containing no feasible point other than x itself, then such a sequence cannot be generated. Therefore, for finding an η-optimal solution in finitely many steps, one usually requires that the feasible set $S = \{x \in D|h(x) \geqslant 0\}$ has no isolated point, i.e.

$$S = S^*, \tag{1}$$

where S^* denotes the *derived set* (the set of cluster points) of S. Clearly this condition is satisfied if $S = \mathrm{cl}(\mathrm{int}S)$($S$ is *robust*):

$$\{x \in D|h(x) \geqslant 0\} = \mathrm{cl}\{x \in \mathrm{int}D|h(x) > 0\},$$

which in view of the continuity of $f(x)$ also amounts to saying that problem (P) is *regular*:

$$\min\{f(x)|x \in S\} = \inf\{f(x)|x \in \mathrm{int}S\}, \tag{2}$$

or equivalently, because D is robust, that

$$\min\{f(x)|h(x) \geqslant 0,\ x \in D\} = \inf\{f(x)|h(x) > 0, \quad x \in \mathrm{int}D\}. \tag{3}$$

Condition (1) (in particular, the regularity condition) rules out the existence of isolated feasible solutions.

The trouble, however, is that quite often checking the regularity assumption is far from being an easy task, so in practice one has to solve the problem without knowing its regularity status. This leads to replacing the given problem by its ε-relaxation (P_ε). Though the latter problem is easily seen to be regular, an optimal solution of it, i.e. an ε-approximate optimal solution of (P), can rarely be computed exactly in finitely many iterations. Therefore, by this approach, the best that one can expect to compute in finitely many iterations is an η-optimal solution of (P_ε), sometimes referred to as an (ε, η)-approximate optimal solution of the original problem.

The above considerations pose the necessity to re-examine the concept of approximate optimal solution in global optimization and to revise accordingly many existing algorithms for problems of the form (P). In any case, from the viewpoint of practical applications, robustness is an important property to demand of numerical methods for nonconvex global optimization.

The organization of the paper is as follows. In Section 2 we show the pitfall connected with the concept of ε-approximate optimal solution. In Sections 3 and 4 we introduce the concept of essential ε-optimal solution and present a general method for finding such an essential ε-optimal solution. In Section 5 this approach is applied to devise robust algorithms called SIT algorithms for d.c. optimization and monotonic optimization. In Section 6, a small monotonic optimization problem is solved to illustrate how a SIT algorithm works. Section 7 closes the paper by some conclusions about the range of applicability of the proposed approach.

2. Pitfall of Approximate Optimality

It seems worthwhile pointing out the pitfall connected with the concept of ε-approximate optimal solution, as defined in the Introduction and widely used, sometimes implicitly, in the literature on global optimization. Although an ε-approximate optimal solution $\{\bar{x}(\varepsilon)\}$ tends to a feasible solution as $\varepsilon \searrow 0$, it may not satisfy the constraint $h(x) \geqslant 0$, i.e. it may be infeasible. Since, on the other hand, $f(\bar{x}(\varepsilon))$ tends to the optimal value γ of (P), the value $f(\bar{x}(\varepsilon))$ should be close to γ when ε is sufficiently small. That is, for any given $\delta > 0$ there exists $\varepsilon_0 = \varepsilon_0(\delta) > 0$ such that $f(\bar{x}(\varepsilon)) < \gamma + \delta$ whenever $0 < \varepsilon < \varepsilon_0$. The difficulty is that, as a rule, ε_0 is far from being easy to determine, so for a given small $\varepsilon > 0$, we can never be sure that $\varepsilon < \varepsilon_0$ and it is quite possible that $f(\bar{x}(\varepsilon)) > \gamma + \delta$. In other words, even for small $\varepsilon > 0$, the value $f(\bar{x}(\varepsilon))$ may be quite far from the actual optimal value γ; and a notable error may be made by accepting $\bar{x}(\varepsilon)$ as an approximate optimal solution.

This is apparent from the example depicted in Figure 1, for a **regular** Problem (P) with the following data:

$$\begin{aligned} f(x) &= x_2 - 2x_1, \qquad h(x) = x_1^2 + x_2^2 - x_1 x_2 - 6x_1 + 4.999999, \\ D &= \{(x_1, x_2) | x_1 + x_2 \leqslant 10, -2x_1 - 3x_2 \leqslant -6, \\ &\quad 2x_2 - x_1 \leqslant 8, x_1 - x_2 \leqslant 4, x_1 \geqslant 0, x_2 \geqslant 0\} \end{aligned} .$$

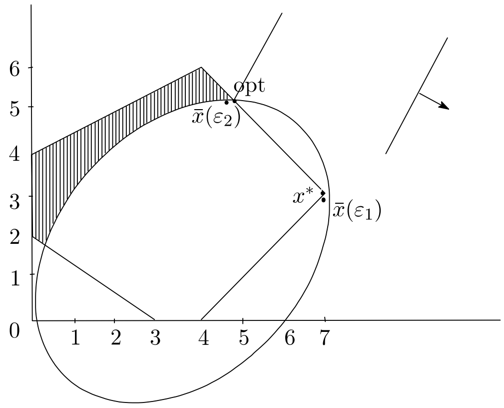

Figure 1. Inadequate ε-approximate optimal solution.

This is a linear program with an additional reverse convex constraint. The point $x^* = (7,\ 3)$ is infeasible but almost feasible: $x^* \in D$, $h(x^*) = -0.000001$. For $\varepsilon_1 = 0.00001$, the ε_1-approximate optimal solution is $\bar{x}(\varepsilon_1) = (7.000003, 2.999994)$, with $f(\bar{x}(\varepsilon_1)) = -11.000013$, far from the true optimal solution and close to x^*. But for $\varepsilon_2 = 0.000001$, an ε_2-approximate optimal solution is $\bar{x}(\varepsilon_2) = (4.996094, 5.003906)$ with $f(\bar{x}(\epsilon_2)) = -4.988281$, quite close to the true optimal solution.

Thus, even for regular problems the ε-relaxation approach may give an incorrect optimal solution if ε is not sufficiently small, while in practice we often do not know what exactly means "sufficiently small".

This situation motivates the following definitions:

Given a problem (P), the set S^* of cluster points of its feasible set (i.e. of its nonisolated feasible solutions) is called its *essential feasible set*, and a nonisolated feasible solution x^* of (P) is called an *essential optimal solution* of it if

$$f(x^*) = \min\{f(x)|x \in S^*\},$$

i.e. if x^* is the minimum of $f(x)$ over the essential feasible set $S^* = \mathrm{cl}\{x \in D|h(x) > 0\}$. Assume $\{x|h(x) > 0, x \in D\} \neq \emptyset$. A nonisolated feasible solution $\bar{x}$ of problem (P) is called an *essential ε-optimal solution* if it satisfies

$$f(\bar{x}) - \varepsilon \leqslant \inf\{f(x)|h(x) > \varepsilon, \quad x \in D\}. \tag{4}$$

Since D is compact, the sequence $\{\bar{x}(\varepsilon)\}$ of essential ε-optimal solutions of (P) has a cluster point x^* which is an essential optimal solution.

REMARK 1. For regular problems an essential optimal solution is necessarily optimal. In the general case it may not be so, but, since it is the best among all nonisolated feasible solutions, hence stable under small perturbations of the constraints, it should be in practice preferred to the true, but unstable, optimal solution.

3. Essential Optimality Criterion

In light of the above discussion, an algorithm for problem (P) that only gives an ε-approximate optimal solution may not be quite correct. In this and the next sections we propose a method to avoid this pitfall and to always guarantee at least an essential ε-optimal solution in finitely many steps.

Let $w \in \mathbb{R}_n$ be any point such that $f(w) - \varepsilon > f(x) \forall\ x \in D$. We first investigate the following subproblem of incumbent transcending:

(*) *Given an* $\bar{x} \in \mathbb{R}^n$, *find a* nonisolated *feasible solution* $\hat{x}$ *of* (P) *such that* $f(\hat{x}) \leqslant f(\bar{x}) - \varepsilon$, *or else establish that none such* $\hat{x}$ *exists.*

Clearly, if $\bar{x} = w$ then an answer to $(*)$ would give a nonisolated feasible solution or else identify essential infeasibility of the problem. If $(\bar{x})$ is the best nonisolated feasible solution currently available then an answer to $(*)$ would give a new nonisolated feasible solution $\hat{x}$ with $f(\hat{x}) \leqslant f(\bar{x}) - \varepsilon$, or else identify $\bar{x}$ as an essential ε-optimal solution.

Since the essential optimal value is upper bounded in view of the compactness of D, by successively solving a finite sequence of subproblems $(*)$ we will finally come up with an essential ε-optimal solution, or an evidence that the problem is essentially infeasible. The key thus reduces to investigating the incumbent transcending subproblem $(*)$. To this end, consider the set

$$\{x \in D | f(x) \leqslant \gamma\}, \tag{5}$$

where $\gamma \in \mathbb{R}$. In many cases of interest, for every fixed γ this set is robust, i.e.

$$\{x \in D | f(x) \leqslant \gamma\} = \mathrm{cl}\{x \in D | f(x) < \gamma\}), \tag{6}$$

so that the problem

$$\max\{h(x) | f(x) \leqslant \gamma, \quad x \in D\} \tag{P$^*/\gamma$}$$

is regular. As we shall show in the sequel, it turns out that under assumption (6) the incumbent transcending subproblem (*) can be solved by solving (P^*/γ) for $\gamma = f(\bar{x}) - \varepsilon$.

Denote the optimal values of (P) and (P^*/γ) by min(P) and $\max(\mathrm{P}^*/\gamma)$, respectively.

PROPOSITION 1. *Under assumption (6):*

(i) *Any feasible solution* x *of* (P^*/γ) *such that* $h(x) > 0$ *is a nonisolated feasible solution of* (P) *with* $f(x) \leqslant \gamma$. *In particular, if* $\max(\mathrm{P}^*/\gamma) > 0$ *then the optimal solution* $\hat{x}$ *of* (P^*/γ) *is a nonisolated feasible solution of* (P) *with* $f(\hat{x}) \leqslant \gamma$.

(ii) *If* $\max(\mathrm{P}^*/\gamma) \leqslant \varepsilon$ *for* $\gamma = f(\bar{x}) - \varepsilon$, *and* $\bar{x}$ *is a nonisolated feasible solution of* (P), *then it is an essential* ε*-optimal solution of* (P). *If*

$\max(\mathrm{P}^*/\gamma) \leqslant \varepsilon$ *for* $\gamma = f(\bar{x}) - \varepsilon$, *and* $\bar{x} = w$, *then the problem* (P) *is essentially infeasible*.

Proof. (i) By (6) x is the limit of an infinite sequence $\{x^k\}$ such that $x^k \in D$, $f(x^k) < \gamma$. Since $h(x) > 0$, one can assume $h(x^k) > 0\ \forall k$, so each x^k is a feasible solution of (P) and therefore, x is a nonisolated feasible solution of (P). That $f(x) \leqslant \gamma$ follows from the fact that x is a feasible solution of (P^*/γ).

(ii) By regularity of (P^*/γ), if $\max(\mathrm{P}^*/\gamma) \leqslant \varepsilon$ then

$$\varepsilon \geqslant \sup\{h(x) | f(x) < \gamma,\ x \in D\},$$

so for every $x \in D$ satisfying $f(x) < \gamma$ we must have $h(x) \leqslant \varepsilon$. Hence $f(x) \geqslant \gamma$ for every $x \in D$ such that $h(x) > \varepsilon$, i.e.

$$\inf\{f(x) | h(x) > \varepsilon,\ x \in D\} \geqslant f(\bar{x}) - \varepsilon.$$

The latter inequality implies, if $\bar{x}$ is a nonisolated feasible solution, that it is an essential ε-optimal solution, and if $\bar{x} = w$, that $\{x | h(x) > \varepsilon, x \in D\} = \emptyset$, i.e. that the problem is essentially infeasible. □

4. Finding an Essential ϵ-optimal Solution

Based on the above Proposition 1 one can develop a procedure for finding an essential ε-optimal solution in finitely many steps. Assume the following condition (A):

(A) *For every fixed* $\gamma \in \mathbb{R}$ *the problem* (P^*/γ) *is regular (i.e. condition (6) holds), and for any* $\bar{x} \in \mathbb{R}^n$ *a procedure called* Procedure $(*, \bar{x})$ *is available which terminates after finitely many steps at either of the following events:*

(a) A nonisolated feasible solution $\hat{x}$ *of* (P) *is found with* $f(\hat{x}) \leqslant f(\bar{x}) - \varepsilon$;
(b) An evidence is produced that no such $\hat{x}$ *exists.*

In the next Section we will show that this assumption is fulfilled for at least the two important special classes of problems (P) mentioned in the Introduction, namely:

I. *Canonical d.c. programming* ($D = \{x | g(x) \leqslant 0\}$, and f, g, h are convex functions) (see e.g. [6]);
II. *Canonical monotonic optimization* ($D = \{x | g(x) \leqslant 0\}$, and f, g, h are increasing functions on $\mathbb{R}^n_+$) (see e.g. [7]).

Under Assumption (A) we can state the next algorithm for finding an essential ε-optimal solution of problem (P). Since the algorithm proceeds basically by *s*uccessive *i*ncumbent *t*ranscending, we will refer to it as a SIT algorithm.

SIT Algorithm

Step 0. If no nonisolated feasible solution for (P) is known, let $\bar{x} = w$; otherwise, let $\bar{x}$ be the best nonisolated feasible solution available.

Step 1. Call procedure $(*, \bar{x})$. Go to Step 2 if this procedure terminates at event (a). Go to Step 3 if it terminates at event (b).

Step 2. Reset $\bar{x} \leftarrow \hat{x}$ and return to Step 0.

Step 3. Terminate: if $\bar{x} = w$; the problem (P) is essentially infeasible; otherwise, $\bar{x}$ is an essential ε-optimal solution.

PROPOSITION 2. *A SIT algorithm terminates after finitely many steps, yielding either an essential ε-optimal solution of* (P) *(which may be an ε-optimal solution), or evidence that* (P) *is essentially infeasible.*

Proof. At every occurrence of event (a) the value $f(\bar{x})$ decreases at least by $\varepsilon > 0$. Since $f(\bar{x})$ is obviously bounded below on the compact set D, it follows that event (a) cannot occur infinitely many times. Therefore the algorithm must terminate after finitely many steps. □

REMARK 2. The algorithm consists of a number of cycles, each involving a procedure $(*, \bar{x})$, with the last $\bar{x}$ being the sought essential ε-optimal solution. Each time the algorithm returns to Step 0, instead of starting the new Procedure $(*, \bar{x})$ from scratch it is often possible to start it from the information already gathered at this stage. In that way the algorithm becomes a unified procedure. This will be illustrated in the robust algorithms for d.c. and monotonic optimization to be discussed in the next Section.

5. Application: Robust d.c. and Monotonic Optimization

As is well known, every d.c. or monotonic optimization problem can be reduced to the canonical form [5, 7]. To apply the above approach to these two classes of problems, we prove in this section that Assumption (A) is fulfilled for both of them.

5.1. D.C. OPTIMIZATION

Consider the canonical reverse convex program

$$\min\{f(x) | h(x) \geqslant 0,\ x \in D\}, \tag{CDC}$$

where $D = \{x | g(x) \leqslant 0\}$ is compact, with $\text{int}\, D \neq \emptyset$, while f, g, h are convex finite functions on an open neighbourhood of D and there exists c such that

$$g(c) < 0, \quad h(c) < 0, \quad f(c) < \min\{f(x) | h(x) \geqslant 0, \quad x \in D\}$$

(so the constraint $h(x) \geqslant 0$ is essential). The incumbent transcending problem associated with any $\bar{x} \in \mathbb{R}^n$ is

$$\max\{h(x)|f(x) \leqslant \gamma, \ x \in D\}. \qquad (\text{CDC}^*/\gamma),$$

with $\gamma = f(\hat{x}) - \varepsilon$. Clearly

$$\{x \in D|f(x) \leqslant \gamma\} = \text{cl}\{x \in \text{int}D, \ f(x) < \gamma\}, \qquad \square$$

so the problem (CDC^*/γ) is regular.

An outer approximation or a branch and bound algorithm for solving the convex maximization problem (CDC^*/γ) as described e.g. in [6] will provide a procedure $(*, \bar{x})$ for (CDC). In fact, such an algorithm generates a sequence of nonisolated feasible solutions x^{k} of (CDC^*/γ) together with a sequence of upper bounds β_k satisfying

$$h(x^k) \leqslant \max(\text{CDC}^*/\gamma) \leqslant \beta_k, \qquad 0 < \beta_k - h(x^k) \to 0 (k \to +\infty).$$

Then either there exists k such that $h(x^k) > 0$, or $h(x^k) \leqslant 0 \ \forall k$. In the former case, by Proposition 1,(i), x^k is a nonisolated feasible solution with $f(x^k) \leqslant f(\bar{x}) - \varepsilon$. In the latter case, since $\beta_k - h(x^k) \to 0$, one must have $\lim_{k \to +\infty} \beta_k \leqslant 0$, hence there exists k such that $\beta_k \leqslant \varepsilon$. Then $\max(\text{CDC}^*/\gamma) \leqslant \varepsilon$, and since $\gamma = f(\bar{x}) - \varepsilon$, it follows from Proposition 1, (ii), that $\bar{x}$ is an essential ε-optimal solution of (CDC) if it is a nonisolated feasible solution, or that the problem is essentially infeasible if $\bar{x} = w$.

For example a SIT algorithm for (CDC) based on an outer approximation algorithm for solving (CDC^*/γ) would read as follows.

5.1.1. *A SIT Algorithm for (CDC)*

Let w be any point where $f(w) - \varepsilon > \min\{f(x)|x \in D\}$.

Step 0. If no nonisolated feasible solution is known, let $\bar{x} = w$; otherwise, let $\bar{x}$ be the best nonisolated feasible solution of (CDC) available. Let $\gamma = f(\bar{x}) - \varepsilon$, $P_1 =$ any simple polytope enclosing $E := \{x \in D|\, f(x) \leqslant \gamma\}$, $V_1 =$ vertex set of P_1. Set $k = 1$.

Step 1. Compute $z^k \in \text{argmax}\{h(x)|x \in V_k\}$.

(a) If $\beta_k := h(z^k) \leqslant \varepsilon$, terminate: if $\bar{x}$ is an isolated feasible solution, it is an essential ε-optimal solution of (CDC); if $\bar{x} = w$, the problem (CDC) is essentially infeasible.

(b) If $h(z^k) > \varepsilon$, go to Step 2.

Step 2. Compute $x^k = c + \lambda_k(z^k - c)$ such that $g_\gamma(x^k) := \max\{f(x^k) - \gamma, g(x^k)\} = 0$.

(a) If $h(x^k) > 0$ then x^k is a nonisolated feasible solution of (CDC^*/γ): go to Step 3.

(b) If $h(x^k) \leqslant 0$, go to Step 4.

Step 3. Compute the point $\bar{x}^k$ where the line segment joining c to x^k meets the surface $h(x)=0$. Reset $\bar{x} \leftarrow \bar{x}^k,\ P_1 \leftarrow P_k,\ V_1 \leftarrow V_k$, and return to Step 0.

Step 4. Take $p^k \in \partial f(x^k)$ if $f(x^k)-\gamma=0$ or $p^k \in \partial g(x^k)$ if $g(x^k)=0$. Compute the vertex set V_{k+1} of the polytope $P_{k+1}=P_k \cap \{x|\langle p^k; x-x^k\rangle \leqslant 0\}$, increment k and return to Step 1.

PROPOSITION 3. *The above algorithm terminates after finitely many steps, yielding either an essential* ε*-optimal solution of (CDC) or an evidence that the problem is essentially infeasible.*

Proof. It is easily seen that the algorithm must terminate at either Step (1a) or Step (2a). Indeed, otherwise the algorithm is an infinite outer procedure for solving the convex maximization problem $\max\{h(x)|f(x)-\gamma \leqslant 0, x \in D\}$. Under the stated conditions, this outer procedure is guaranteed to converge, i.e. $\beta_k - h(x^k) \to 0$ as $k \to +\infty$. Since $h(x^k) \leqslant 0\ \forall k$, we must have $\beta_k \leqslant \beta_k - h(x^k) \to 0$ as $k \to +\infty$, contradicting $\beta_k = h(z^k) > \varepsilon\ \forall k$ (Step 1b)). Therefore, the algorithm must terminate after finitely many steps. □

REMARK 3. If we denote $D(\gamma)=\{x \in D|f(x) \leqslant \gamma\}$, $C=\{x|\,h(x) \leqslant 0\}$, then the incumbent transcending problem (CDC^*/γ) amounts to finding a point $x \in D(\gamma)\backslash C$ or else proving that $D(\gamma) \subset C$. Therefore (CDC^*/γ) is a variant of the problem DC defined in [6], Chapter 5. Actually the above algorithm is a robust variant of the OA Algorithm for (CDC) described in [6] and differs from the latter mainly in that condition $h(z^k) \leqslant -\varepsilon$ in the OA Algorithm is replaced by $h(z^k) \leqslant \varepsilon$.

5.2. MONOTONIC OPTIMIZATION

Monotonic optimization is concerned with mathematical programming problems described by means of monotonic (increasing or decreasing) functions, and more generally, differences of monotonic (d.m.) functions. It includes, in particular, polynomial and posynomial programming. In [7] a general mathematical framework has been developed for the study of monotonic optimization. However, as most existing algorithms of continuous nonconvex global optimization, the algorithms proposed in [7] have been devised without much concern about robustness.

Based on the above discussion, we now present a robust algorithm for the canonical monotonic optimization

$$\min\{f(x)|\,h(x) \geqslant 0,\ g(x) \leqslant 0,\ x \in [a,b]\}, \tag{CMO}$$

where $[a,b] \subset \mathbb{R}^n_+$, and *f, g, h* are increasing functions on $\mathbb{R}^n_+$.

As was demonstrated in [7], this is the general form of any monotonic optimization problem. We shall make the following natural assumptions:

(1) $h(a) < 0$ (which means essentially that the contraint $h(x) \geqslant 0$ cannot be dropped);
(2) $f(a) < \min(\text{CMO}) < f(b)$;
(3) $\{x|g(x) < 0, a < x < b\} \neq \emptyset$.

In view of the last assumption (3), for any fixed γ such that $\min(\text{CMO}) < \gamma < f(b)$ there is $x \in [a,b]$ satisfying $f(x) < \gamma$, $g(x) < 0$, and the set

$$\{x \in [a,b] | f(x) \leqslant \gamma,\ g(x) \leqslant 0,\ x \in [a,b]\}$$

is robust, so the problem

$$\max\{h(x) | f(x) \leqslant \gamma,\ g(x) \leqslant 0,\ x \in [a,b]\} \qquad (\text{CMO}^*/\gamma)$$

is regular.

It is not difficult to derive a procedure $(*, \bar{x})$ for (CMO) from the polyblock approximation algorithm for (CMO^*/γ) [7]. Below we prefer, however, to derive it from a branch and cut algorithm for solving (CMO^*/γ), which is believed to be more efficient than the polyblock approximation method on large scale problems.

5.2.1. *Procedure $(*, \bar{x})$ for (CMO)*

A branch and bound algorithm for solving the monotonic optimization problem (CMO^*/γ) can be outlined as follows.

At a general iteration, we have a collection of nonoverlapping subboxes of $[a,b]$ known to contain at least an optimal solution of (CMO^*/γ) and such that for each subbox M an upper bound $\beta(M)$ of $h(x)$ over M has been estimated. We then select the subbox with largest upper bound and subdivide it into two boxes using an exhaustive rule, i.e. a rule ensuring that any infinite nested sequence of boxes to be generated by it will eventually shrink to a single point. Each of the newly formed boxes is then reduced (replaced by a smaller box) without losing any feasible point currently still of interest. Next an upper bound is computed for $h(x)$ over each of the new boxes and the procedure continues with the new collection of boxes. During the process, the current best value of $h(x)$ is updated and used in the reduction of the new boxes and the pruning of the current collection of boxes. Thus, the algorithm involves, aside from the standard partitioning operation, a reduction and a bounding operation.

The reduction and bounding operations are based on specific cuts that exploit the monotonic structure of the problem.

First observe that if $h(\bar{x}) > 0$, then, since $h(a) < 0$, the halfline from a through $\bar{x}$ meets the surface $h(x) = 0$ at some x' which is a nonisolated feasible solution with $f(x') \leqslant f(\bar{x})$, as $f(x)$ is increasing. Therefore, by replacing $\bar{x}$ with x' if necessary, we can always assume that $\bar{x}$ is a nonisolated feasible solution satisfying $h(\bar{x}) = 0$. With this in mind, we can restrict the search for a nonisolated feasible solution of (CMO^*/γ) such that $f(x) \leqslant \gamma$ to the set

$$B_\gamma := \{x | f(x) - \gamma \leqslant 0,\ g(x) \leqslant 0,\ h(x) = 0\}. \tag{7}$$

5.2.1.1. *Reduction*. Let $[p, q] \subset [a, b]$ be a box generated during the partitioning procedure. The reduction of the box $[p, q]$ aims at replacing $[p, q]$ by a smaller box $[p', q'] \subset [p, q]$ such that every point x of the set B_γ that is contained in $[p, q]$ is still contained in $[p', q']$, i.e. such that

$$B_\gamma \cap [p', q'] = B_\gamma \cap [p, q].$$

Setting

$$\tilde{g}_\gamma(x) = \max\{f(x) - \gamma,\ g(x), h(x)\}$$

we can write

$$B_\gamma = \{x | \tilde{g}_\gamma(x) \leqslant 0 \leqslant h(x)\}.$$

LEMMA 1. (i) *If* $h(q) < 0$ *or* $\tilde{g}_\gamma(p) > 0$, *then* $B_\gamma \cap [p, q] = \emptyset$.

(ii) *If* $h(q) \geqslant 0$, *then the box* $[p', q]$ *where* $p' = q - \sum_{i=1}^n \alpha_i (q_i - p_i) e^i$, *with*

$$\alpha_i = \sup\{\alpha | 0 \leqslant \alpha \leqslant 1,\ h(q - \alpha(q_i - p_i)e^i) \geqslant 0\}, \quad i = 1, \ldots, n, \tag{8}$$

still contains $B_\gamma \cap [p, q]$.

(iii) *If* $\tilde{g}_\gamma(p') \leqslant 0$, *then the box* $[p', q']$ *where* $q' = p' + \sum_{i=1}^n \beta_i (q_i - p'_i) e^i$, *with*

$$\beta_i = \sup\{\beta | 0 \leqslant \beta \leqslant 1,\ \tilde{g}_\gamma(p' + \beta(q_i - p'_i)e^i) \leqslant 0\}, \quad i = 1, \ldots, n, \tag{9}$$

still contains $B_\gamma \cap [p, q]$.

Proof. It suffices to prove (ii), because (iii) can be proved analogously, while (i) is obvious (follows from the fact that $\tilde{g}_\gamma(x)$ and $h(x)$ are increasing). Since $p'_i = \alpha_i p_i + (1 - \alpha_i) q_i$ with $0 \leqslant \alpha_i \leqslant 1$, it follows that $p_i \leqslant p'_i \leq q_i\ \forall i = 1, \ldots, n$, i.e.$[p', q] \subset [p; q]$. Let

$$H := \{x | h(x) \geqslant 0\}.$$

For any $x \in H \cap [p, q]$ we have, because $h(x)$ is increasing, $[x, q] \subset H$, so $x^i = q - (q_i - x_i)e^i \in H$, $i = 1, \ldots, n$. But $x_i \leqslant q_i$, so $x^i = q - \alpha(q_i - p_i)e^i$ with $0 \leqslant \alpha \leqslant 1$. This implies that $\alpha \leqslant \alpha_i$, i.e. $x^i \geqslant q - \alpha_i(q_I - p_i)e^i$, $i = 1, \ldots, n$, and consequently $x \geqslant p'$, i.e. $x \in [p', q]$. Thus, $H \cap [p, q] \subset H \cap [p', q]$, which completes the proof because the converse inclusion is obvious from the fact $[p', q] \subset [p, q]$. □

Clearly the box $[p', q]$ defined in (ii) is obtained from $[p, q]$ by cutting off the set $\cup_{i=1}^{n}\{x|x_i < p'_i\}$, while the box $[p, q']$ defined in (iii) is obtained from $[p, q]$ by cutting off the set $\cup_{i=1}^{n}\{x|x_i > q'_i\}$. The cut $\cup_{i=1}^{n}\{x|x_i > p'_i\}$ is referred to as a *lower γ-valid cut with vertex p'*, and the cut $\cup_{i=1}^{n}\{x|x_i > q'_i\}$ as an *upper γ-valid cut with vertex q'*, applied to the box $[p, q]$. Using these cuts we next define a box $\text{red}_\gamma[p, q]$ referred to as *γ-valid reduction of* $[p, q]$:

$$\text{red}_\gamma[p, q] = \begin{cases} \emptyset & \text{if } h(q) < 0 \text{ or } \tilde{g}_\gamma(p') > 0, \\ [p', q'] & h(q) \geqslant 0 \ \& \ \tilde{g}_\gamma(p') \leqslant 0, \end{cases} \tag{10}$$

where

$$p' = q - \sum_{i=1}^{n} \alpha_i(q_i - p_i)e^i, \qquad q' = p' + \sum_{i=1}^{n} \beta_i(q_i - p'_i)e^i \tag{11}$$

$$\begin{aligned} \alpha_i &= \sup\{\alpha|0 \leqslant \alpha \leqslant 1, \quad h(q - \alpha(q_i - p_i)e^i) \geqslant 0\}, \quad i = 1, \dots, n; \\ \beta_i &= \sup\{\beta_j|0 \leqslant \beta \leqslant 1, \quad \tilde{g}_\gamma(p' + \beta(q_i - p'_i)e^i) \leqslant 0\}, \quad i = 1, \dots, n. \end{aligned} \tag{12}$$

5.2.1.2. *Bounding*. Given a box $M := [p, q]$, we want to compute an upper bound $\beta(M)$ for $h(x)$ over the set

$$B_\gamma \cap [p, q] = \{x \in [p, q] | \ \tilde{g}_\gamma(x) \leqslant 0 \leqslant h(x)\}.$$

After reducing $[p, q]$ as described above, let $[p, q] \leftarrow [p', q'] := \text{red}_\gamma[p, q]$. Since $h(x)$ is increasing, an obvious upper bound is

$$\beta(M) = h(q). \tag{13}$$

Though very simple, this bound suffices to ensure convergence of the algorithm, as we will see shortly. However, for a better performance of the procedure, tighter bounds can be computed using, for instance, the following.

LEMMA 2. *Let $x(M) = p + \lambda(q - p)$ with $\lambda = \max\{\alpha|g_\gamma(p + \alpha(q - p)) \leqslant 0\}$, and let*

$$z^i = q + (x_i(M) - q_i)e^i, \quad i = 1, \dots, n$$

Then an upper bound of $h(x)$ over the set $B_\gamma \cap [p, q]$ is

$$\beta(M) = \max\{h(z^i)| \quad i = 1, \dots, n\}$$

Proof. The function $g_\gamma(x)$ is increasing and satisfies $g_\gamma(p) \leqslant 0 < g_\gamma(q)$ because otherwise $B_\gamma \cap [p, q] = \emptyset$. Therefore $0 \leqslant \lambda < 1$ and $g_\gamma(x(M)) = 0$. If x satisfies $x(M) < x \leqslant q$ then there exists $x' = p + \alpha(q - p)$ such that $\lambda < \alpha$ and $x(M) < x' \leqslant x$. This implies that $0 = g_\gamma(x(M)) < g_\gamma(x') \leqslant g_\gamma(x)$, hence $g_\gamma(x) < 0$ for all x satisfying $x(M) < x \leqslant q$. Let $G_\gamma = \{x \in [p, q] | g_\gamma(x) > 0\}$, $K = \{x| \ x(M) < x \leqslant q, \ K_i = \{x \in [p, q] | \ x_i(M) < x_i\}$. Then $G_\gamma \subset [p, q] \backslash K =$

$[p,q]\setminus\cap_{i=1}^{n} K_i = \cup_{i=1}^{n}([p,q]\setminus K_i)$. Since $[p,q]\setminus K_i = \{x \in [p,q] | x_i \leqslant x_i(M)\} = [p, z^i]$, it follows that, for every $x \in G_\gamma$, there exists i such that $x \leqslant z^i$, hence $h(x) \leqslant h(z^i)$, and consequently, $h(x) \leqslant \max\{h(z^i) | \, i = 1, \ldots, n\} \; \forall x \in G_\gamma$. □

REMARK 4. The set $Q = \cup_{i=1,\ldots,n}[p, z^i]$ is called a *polyblock* generated by $z^1, \ldots, z^n$. Clearly $G_\gamma \subset Q = [p,q]\setminus K$, so Q is what remains from $[p,q]$ after the cut $K = \{x | \, x(M) < x \leqslant q\}$. Since $q \in K$, while $q \notin G_\gamma := \{x | \, g_\gamma(x) \leqslant 0\}$, the polyblock Q separates the point q from G. The cut $K = \{x | \, x(M) < x \leqslant q\}$ is therefore called a *separation cut*.

Using the above bounding and the standard bisection for partitioning, a branch and bound can be developed for solving (CMO). Denote by M_k the partition set (box) selected for partition at iteration k and let $x^k = x(M_k)$, $\beta_k = \beta(M_k)$.

PROPOSITION 4. *x^k is a nonisolated feasible solution of* (CMO$^*/\gamma$) *satisfying*

$$h(x^k) \leqslant \max(\text{CMO}^*/\gamma) \leqslant \beta_k, \quad \beta_k - h(x^k) \to 0 (k \to +\infty).$$

Proof. Since $g_\gamma(x^k) = \max\{f(\bar{x}) - \gamma, \; g(x^k) \leqslant 0\}$, it follows that x^k is a feasible solution of (CMO). This feasible solution is nonisolated because the set $G_\gamma = \{x \in [a,b] | f(x) \leqslant \gamma, g(x) \leqslant 0\}$ is robust. It remains to show that $\beta_k - h(x^k) \to 0$ as $k \to +\infty$. But since the subdivision rule is exhaustive, as $k \to +\infty$, there exists a nested sequence of $M_k = [p^k, q^k]$ such that diam $M_k \to 0$. Then $q^k - x^k \to 0$ and hence, $h(q^k) - h(x^k) \to 0$. Since $h(x^k) \leqslant \max(\text{CMO}^*/\gamma) \leqslant \beta_k \leqslant h(q^k)$, it follows that $\beta_k - h(x^k) \to 0$ as $k \to +\infty$. □

5.2.2. *A SIT Algorithm for (CMO)*

Incorporating the above branch-reduce-and-bound procedure into the SIT scheme we obtain the following robust algorithm for solving (CMO):

Step 0. If no feasible solution is known, let $\bar{x} = w$, otherwise, let $\bar{x}$ be the best nonisolated feasible solution available. Let $\gamma = f(\bar{x}) - \varepsilon$, $\mathcal{P}_1 = \{M_1\}$, $M_1 = [a,b]$, $\mathcal{R}_1 = \emptyset$. Set $k = 1$.

Step 1. For each box $M \in \mathcal{P}_k$:
- Compute its γ-valid reduction;
- Delete M if $\text{red}_\gamma M = \emptyset$;
- Replace M by $\text{red}_\gamma M$ if $\text{red}_\gamma M \neq \emptyset$;
- If $\text{red}_\gamma M = [p', q']$ then compute an upper bound $\beta(M) \leqslant h(q')$ for $f(x)$ over the feasible solutions in M.

Step 2. Let $\mathcal{P}'_k$ be the collection of boxes that results from $\mathcal{P}_k$ after completion of Step 1. Let $\mathcal{R}'_k = \mathcal{R}_k \cup \mathcal{P}'_k$.

Step 3. If $\mathcal{R}'_k = \emptyset$, then terminate: $\bar{x}$ is an essential ε-optimal solution of (CMO) if $\bar{x} \neq w$, or the problem (CMO) is essentially infeasible if $\bar{x} = w$.

Step 4. If $\mathcal{R}'_k \neq \emptyset$, let $[p^k, q^k] := M_k \in \text{argmax}\{\beta(M)|\, M \in \mathcal{R}'_k\}$, $\beta_k = \beta(M_k)$. If $\beta_k \leqslant \varepsilon$ then terminate: $\bar{x}$ is an essential ε-optimal solution of (CMO) if $\bar{x} \neq w$ or the problem (CMO) is essentially infeasible if $\bar{x} = w$. Otherwise, go to Step 5.

Step 5. Compute $x^k = p^k + \lambda_k(q_k - p_k)$ with

$$\lambda_k = \max\{\alpha|\, f(p^k + \alpha(q^k - p^k) - \gamma \leqslant 0,\ g(p^k + \alpha(q^k - p^k) \leqslant 0\}.$$

If $h(x^k) > 0$ then x^k is a new nonisolated feasible solution of (CMO) with $f(x^k) \leqslant \gamma$: compute the point $\bar{x}^k$ where the line segment joining p^k to x^k meets the surface $h(x) = 0$, and reset $\bar{x} \leftarrow \bar{x}^k$.

Step 6. Divide M_k into two subboxes by the standard bisection (or any bisection consistent with the bounding $M \mapsto \beta(M)$). Let $\mathcal{P}_{k+1}$ be the collection of these two subboxes of $M_k, \mathcal{R}_{k+1} = \mathcal{R}'_k \setminus \{M_k\}$. Increment k, and return to Step 1.

PROPOSITION 5. *The above algorithm terminates after finitely many steps, yielding either an essential ε-optimal solution of (CMO), or an evidence that the problem is essentially infeasible.*

Proof. Since any feasible solution x with $f(x) \leqslant \gamma = f(\bar{x}) - \varepsilon$ must lie in some box $M \in \mathcal{R}'_k$ the event $\mathcal{R}'_k = \emptyset$ implies that no such solution exists, hence the conclusion in Step 3. If $\beta_k \leqslant \varepsilon$ in Step 4, then $\max(\text{CMO}^*/\gamma) \leqslant \varepsilon$, hence by Proposition 1, the same conclusion in Step 4. Since $h(p^k) \leqslant 0$, if $h(x^k) > 0$, then the point $\bar{x}^k$ exists and satisfies $g(\bar{x}^k) \leqslant g(x^k) \leqslant 0$, $f(\bar{x}^k) \leqslant f(x^k) \leqslant \gamma$, $h(\bar{x}^k) = 0$, so $\bar{x}^k$ is a nonisolated feasible solution with $f(\bar{x}^k) \leqslant f(\bar{x}) - \varepsilon$. Thus the conclusion is correct if one of the following events occurs: $\mathcal{R}'_k = \emptyset$, $\beta_k \leqslant \varepsilon$, $h(x^k) > 0$. It remains to show that at least one of these events must occur, i.e. that for sufficiently large k Steps 5 and 6 cannot occur. But since each occurrence of Step 5 increases the current best value at least by $\varepsilon > 0$ while $f(x)$ is bounded above it follows that Step 5 cannot occur infinitely often. On the other hand, in Step 6 we have $h(x^k) \leqslant 0$ and since $\beta_k) > \varepsilon$, while $\beta_k - h(x^k) \to 0$ as $k \to +\infty$, it follows that Step 6 cannot occur infinitely often either. Therefore, the algorithm must be finite. □

The complexity of the above algorithm is about the same as that of a branch and cut algorithm for solving directly (CMO) (see [8]), but an advantage of it is that it works its way to the optimum through better and better nonisolated feasible solutions. Therefore, even if for some reason a SIT algorithm has to be stopped prematurely, some reasonably good feasible solution may have been already obtained. This is in contrast with many existing algorithms which may give useless results when stopped prematurely.

6. Illustrative Example

The SIT Algorithm for (CDC) only slightly differs from the OA Algorithm described e.g. in [6]. To illustrate how the SIT Algorithm for (CMO) works, let us consider a small example:

$$\begin{array}{ll} \min & x_1 + x_2^2 \\ \text{s.t.} & 0.75x_1 + x_2 \geqslant 5.5 \\ & (x_1 - 3)^3 + x_2 \leqslant 3 \\ & x_1^2 + x_2^2 \leqslant 36 \\ & 0 \leqslant x_1 \leqslant 6, \ 0 \leqslant x_2 \leqslant 6. \end{array}$$

This is a problem (CMO) with $f(x) = x_2^2 + x_1$, $h(x) = 0.75x_1 + x_2 - 5.5$, $g(x) = \max\{(x_1 - 3)^3 + x_2 - 3, x_1^2 + x_2^2 - 36\}$, a = (0, 6), b = (0, 6). The associated problem (CMO^*/γ) is

$$\begin{array}{ll} \max & 0.75x_1 + x_2 - 5.5 \\ \text{s.t.} & (x_1 - 3)^3 + x_2 \leqslant 3 \\ & x_1^2 + x_2^2 \leqslant 36 \\ & x_1 + x_2^2 \leqslant \gamma \\ & 0 \leqslant x_1 \leqslant 6, 0 \leqslant x_2 \leqslant 6 \end{array}$$

With $\varepsilon = 0.001$ the SIT Algorithm found the essential ε-optimal solution $x^{\text{essopt}} = (1.999408, 4.000444)$ with essential ε-optimal value 18.002961 (see Figure 2) at iteration 35, and needed 4 more iterations to confirm essential ε-optimality. The computation required 0.016 s on a PC Pentium IV 2.53 GHz, RAM 256 MB DDR, and went through 11 cycles of incumbent transcending, with intermediate results as given in Table 1. ($\bar{x}$ is the new incumbent found at the end of the cycle, and Iter indicates the iteration where $\bar{x}$ is found.)

For $\varepsilon = 0.000001$ and $\eta = 0.001$ an η-optimal solution of the ε-relaxed problem is $x^* = (3.500974, 2.874269)$ (see Figure 2). To find such an

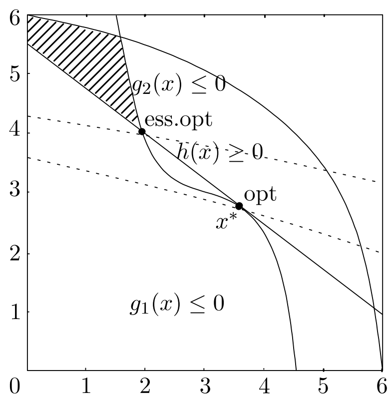

Figure 2. A nonregular (CMO) problem.

Table 1.

Cycle	γ	$\bar{x}$	$f(\bar{x})$	Iter
1	50	(0.965645, 4.775767)	23.773594	1
2	23.772594	(1.555398, 4.333452)	20.334203	2
3	20.333203	(1.793777, 4.154668)	19.055040	3
4	19.054040	(1.900542, 4.074593)	18.502851	5
5	18.501851	(1.951129, 4.036653)	18.245696	7
6	18.244696	(1.975784, 4.018162)	18.121409	9
7	18.120409	(1.988004, 4.008997)	18.060059	10
8	18.059059	(1.994092, 4.004431)	18.029557	14
9	18.028557	(1.997131, 4.002151)	18.014347	18
10	18.013347	(1.998650, 4.001013)	18.006753	26
11	18.005753	(1.999408, 4.000444)	18.002961	35

(ε, η)-approximate optimal solution, the branch and bound algorithm described in [8] would require 591 iterations and 0.203 s (as compared with 0.016 s, on the same computer, by the SIT approach).

This simple example illustrates the potential advantage of the SIT approach over the usual ε-approximation approach: not only a robust solution is obtained, less computational effort may be required for reaching the same accuracy (in the above example, the essential optimal solution and the optimal solution of the 10^{-6}-relaxed problem are both obtained within tolerance 10^{-3}).

REMARK 5. Since the aim of the above example was mainly to illustrate how the SIT Algorithm works, in solving this example we only used the bounds provided by Lemma 2. By exploiting also the d.c. structure of the constraints one could compute much better bounds to speed up the algorithm significantly.

7. Conclusions

It has long been a common practice in nonconvex global optimization to accept an optimal solution of the ε-relaxation (P_ε) of a given problem (P) as an approximate optimal solution of (P). In this paper we demonstrated the incorrectness of this approach which may lead to an infeasible solution very far from the optimum. To overcome the difficulty we proposed a method for computing an essential approximate optimal solution which is the best among all nonisolated feasible solutions. This methodology can be applied to a class of problems that includes d.c. optimization and monotonic programming, i.e. virtually a large majority of continuous nonconvex global optimization problems encountered in practice (see [5, 7]).

8. Acknowledgements

The author would like to thank a referee for his useful remarks and Ng.T. Hoai Phuong for the construction of the illustrative example.

References

1. Floudas, C.A. (2000), *Deterministic Global Optimization: Theory, Methods and Applications,* Kluwer Academic Publishers, Dordrecht.
2. Horst, R. and Tuy, H.(1996), *Global Optimization: Deterministic Approaches*, 3rd ed., Springer-Verlag, Berlin.
3. Konno, H., Thach, P.T. and Tuy, H. (1997), *Optimization on Low Rank Nonconvex Structures*, Kluwer Academic Publishers.
4. Sherali, H.D. and Adams, W.P. (1999), *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Peroblems*, Kluwer Academic Publishers.
5. Tuy, H. (1995), D.C. optimization: Theory, methods and algorithms. In: Horst, R. and Pardalos, P. (eds.), *Handbook of Global Optimization*, pp. 149–216. Kluwer Academic Publishers.
6. Tuy, H. (1998), *Convex Analysis and Global Optimization*, Kluwer Academic Publishers.
7. Tuy, H. (2000), Monotonic optimization: Problems and solution approaches. *SIAM Journal on Optimization* 11(2), 464–494.
8. Tuy, H., Al-Khayyal, F. and Thach, P.T. (2004), Monotonic optimization: branch and cut methods, preprint, Institute of Mathematics, Hanoi.