AOM - Advanced Modeling and Optimization,Volume 14, Number 1, 2012

ABOUT BRANCH AND BOUND METHOD

CH. DAILI

F. ABBAS University. Faculty of Sciences, Department of Mathematics,19000, Sétif - Algeria

ABSTRACT. The Branch and bound method is a rather general algorithm which was applied, first of all, to resolve combinatorial optimization problems. This technique, widely spread, is used to resolve diverse classes of difficult problems of local and global optimization. The branch and bound method knew systematically recent developments and becomes a mattering tool to resolve a senior form of global optimization problems such as the concave minimization, the minimization of the Lipschitz functions and the d-c optimization, ... etc. In this work, a state of art of the branch and bound method is supplied. An algorithm is established followed by a proved theorem of convergence

1. Introduction

The Branch and bound method is a rather general algorithm which was applied first of all to resolve combinatorial optimization problems. Also, the branch and bound algorithm knew systematically recent developments and becomes a mattering tool to resolve a senior form of global optimization problem (with continuous variables) such as the concave minimization, the minimization of the Lipschitz functions and also the d-c optimization (the minimization of the difference of two convex functions), ... etc.

This technique, widely spread, is used to resolve diverse classes of difficult problems of local and global optimization.

The main idea of the method consists in subdividing the feasible set in subsets more and more branching, then in determining uper and lower bounds of the optimal value of the objective function on these subsets. Parts of the feasible set having the lower bounds exceeding the best optimal value found in a certain step of the algorithm will be deleted, because these parts of the domain do not contain the optimum.

In what follows, we present the Branch and bound algorithm in a general frame with its properties.

2. The Prototype of Branch and Bound Method

Consider the following global optimization problem:

$$(\mathcal{P}) \qquad \begin{cases} \min\limits_{x \in X} f(x), \end{cases}$$

where $f : A \to \mathbb{R}$, $X \subset A \subset \mathbb{R}^n$. Let us suppose that the $\min f(X)$ exists.

*AMO - Advanced Modeling and Optimization. ISSN: 1841-4311.
Date: October 10, 2011.
2010 Mathematics Subject Classification. Primary 90C26, 90C57; Secondary 90C90.
Key words and phrases. Branch and bound method, Sudivision, Algorithms.

Definition 1. *Let B be a subset of $\mathbb{R}^n$ and I be a finite set of indices. The $\{ M_i : i \in I \}$ family of subsets is called partition of B, if*

$$B = \bigcup_{i \in I} M_i \quad \text{and} \quad M_i \cap M_j = Fr(M_i) \cap Fr(M_j), \forall i,\ j \in I, i \neq j,$$

where $Fr(M_i)$ is the boundary of M_i.

Definition 2. *Let M be an element of the previous partition.*
.) If $M \cap X = \varnothing$, M is called not feasible.
.) If $M \cap X \neq \varnothing$, M is called feasible.
.) Else M is called uncertain.

3. Branch and Bound Algorithm

Step 1 : Initialization ($k = 0$)
Choose

$$M_0 \supseteq X, \quad S_{M_0} \subset X$$

$$-\infty < \beta_0 \leq \min f(X)$$

Set

$$\mathcal{M}_0 = \{M_0\}$$

$$\alpha_0 = \min f(S_{M_0}), \quad \beta_0 = \beta(M_0).$$

If $\alpha_0 < +\infty$, then find

$$x^0 \in \arg\min f(S_{M_0}) \text{ i.e., } f(x^0) = \alpha_0$$

If $\alpha_0 - \beta_0 \leq \varepsilon$, then stop
Else, go to step 2.
Step 2 : ($k = 1, 2, ...$)
At the beginning of the step k, we have the current partition $\mathcal{M}_{k-1}$ of one subset of M_0 which stays after elimination. Moreover, for any $M \in \mathcal{M}_{k-1}$, we have a set $S_M \subseteq M \cap X$ and bounds $\beta(M)$, $\alpha(M)$ verifying

$$\begin{cases} \beta(M) \leq \inf f(M \cap X) \leq \alpha(M), & \text{if } M \cap X \neq \varnothing \\ \beta(M) \leq \inf f(M), & \text{if } M \cap X \text{ not determined.} \end{cases}$$

We have, also, the bounds $\beta_{k-1}, \alpha_{k-1}$ which verify

$$\beta_{k-1} \leq \inf f(M) \leq \alpha_{k-1}.$$

If $\alpha_{k-1} < +\infty$, then find $x^{k-1} \in X$ such that

$$f(x^{k-1}) = \alpha_{k-1}.$$

($\mathbf{k}_1$) We exclude any subset $M \in \mathcal{M}_{k-1}$ verifying

$$\beta(M) \geq \alpha_{k-1}.$$

Let us consider $\Re_k$ the class of the remaining subsets of $\mathcal{M}_{k-1}$.
($\mathbf{k}_2$) We choose a nonempty class of sets $\Im_k \subset \Re_k$ and we build a partition with every element of $\Im_k$.
Let $\Im_k^{'}$ be the class of all the new elements of the partition.
($\mathbf{k}_3$) Let us put $\mathcal{M}_k^{'}$ the class of all the remaining sets of $\Im_k^{'}$ after exclusion of the sets $M \in \Im_k^{'}$ which verify

$$M \cap X = \varnothing.$$

($\mathbf{k}_4$) Determine for every $M \in \mathcal{M}_k^{'}$, one set $S_M \subseteq M \cap X$ and a number $\beta(M)$ such that

$$\begin{cases} \beta(M) \leq \inf f(M \cap X) \leq \alpha(M), & \text{if } M \cap X \neq \varnothing \\ \beta(M) \leq \inf f(M), & \text{if } M \cap X \text{ not determined.} \end{cases}$$

and

$$S_M \supset S_{M'} \cap M \quad \text{and} \quad \beta(M) \geq \beta(M^{'}),$$

for all subset $M^{'} \supset M$ of $\mathcal{M}_{k-1}$.

Let us put

$$\alpha(M) = \min f(S_M).$$

($\mathbf{k}_5$) Set

$$\mathcal{M}_k = \{\Re_k \backslash \Im_k\} \cup \mathcal{M}_k^{'}.$$

Compute

$$\alpha_k = \min \{\alpha(M) : M \in \mathcal{M}_k\}$$

$$\beta_k = \min \{\beta(M) : M \in \mathcal{M}_k\}.$$

If $\alpha_k < +\infty$, then find $x^k \in X$ such that

$$f(x^k) = \alpha_k.$$

If $\alpha_k - \beta_k \leq \varepsilon$ then stop.
Else, set $k = k + 1$.
Go to step 2.

Remark 1. 1) *In several applications, the upper bound $\alpha(M)$ is simply the minimal value of the objective function f in points known in $M \cap X$, i.e., if f is estimated at points $x_1, x_2, ..., x_l$ of $M \cap X$, we set $\alpha(M) = \min\limits_{1 \leq i \leq l} f(x_i)$.*

2) *If $S_M = \varnothing$, we set $\alpha(M) = +\infty$.*

3) *So that an element M of the partition $\Im_k$ will be deleted, it has to verify the condition $\beta(M) \geq \alpha_{k-1}$. Then the criterion of ruling $\alpha_k = \beta_k$ means that all the elements of the partition are excluded.*

4) *The set M_0 and the subsets M of the partition are chosen among certain families of subsets of $\mathbb{R}^n$ (for example : simplices, or pavements,...).*

In the step ($\mathbf{k}_4$) we can replace whatever $M \in \mathcal{M}_k^{'}$ by a smaller set $\overline{M} \subset M$ such that $\overline{M} \subset \Im_k$, $\overline{M} \cap X = M \cap X$.

5) *It is necessary to impose conditions on S_M and $\beta(M)$ so that $\{\alpha_k\} = \left\{f(x^k)\right\}$ is a decreasing sequence, $\{\beta_k\}_k$ is an increasing sequence, and $\beta_k \leq \min f(M) \leq \alpha_k$, so that the difference $\alpha_k - \beta_k$ measures the nearness of the best optimal solution x^k of the step k.*

Remark 2. 1) *For a given tolerance $\varepsilon > 0$, the algorithm is going to stop as soon as $\alpha_k - \beta_k < \varepsilon$. And because $\{\alpha_k\}_k$ is decreasing monotone and $\{\beta_k\}_k$ is increasing monotone, then the limits $\alpha = \lim\limits_{(k \to \infty)} \alpha_k$ and $\beta = \lim\limits_{(k \to \infty)} \beta_k$ are going to exist, and by recurrence, they are going to satisfy $\beta_k \leq \min f(M) \leq \alpha_k$.*

2) *The algorithm is said finite if $\alpha_k = \beta_k$ in certain step k, whereas it is convergent if $\alpha_k - \beta_k \to 0$, namely,*

$$\alpha = \lim_{(k \to \infty)} f(x^k) = \beta = \min f(M).$$

Remark 3. *The convergence and the efficiency of the Branch and bound method depends on the calculation of the lower and upper bounds of the objective function* f *on the subsets of partitions, as well as of the choice of* $\Im_k$ *and of subdivision of the feasible set and the subsets of partitions.*

Before giving conditions of convergence, we are going to speak about these factors in detail by giving some examples.

4. Some Sets Partitions and their Refinements

In the procedure Branch and bound, the sets of partitions often used are polyhedrons or polytopes as for example: simplices, rectangles (pavements) or polyhedral cones.

For every polytope we have a way of generating her sets partitions and of calculating minimizers of the objective function on these sets.

Let us begin, at first, with the study of the various polytopes used in the algorithm and their processes of refinement.

4.1. **Simplices.** Let us suppose that $X \subset \mathbb{R}^n$ is a compact convex set and let S_0 be a simplex of n dimension and enclosing X. S_0 is the initial set of the algorithm.

Definition 3. *Let* S *be* n*-simplex,* $V(S) = \{v^0, ..., v^n\}$ *the vertex set of* S*. We choose a point* $w \in S$ *such that* $w \notin V(S)$ *with its following unique representation:*

$$w = \sum_{i=0}^{n} \lambda_i v^i, \lambda_i \geq 0, \ (i = 0, ..., n), \ \sum_{i=0}^{n} \lambda_i = 1 \tag{1}$$

For all i *such* $\lambda_i > 0$*, the construction of sub-simplex* $S(i, w)$ *is made from the simplex* S *by replacing the* i^{the} *vertex by* w*, in other words,*

$$S(i, w) = conv \{v^0, ..., v^{i-1}, w, v^{i+1}, ..., v^n\}.$$

This partition is called radial subdivision. It was introduced for the first time by Horst ([8]), then used by several authors.

Proposition 1. *The set of sub-simplices* $S(i, w)$ *can be built from a* n*-simplex* S *according to one arbitrary radial subdivision forming a partition of* S*, namely,*

$$S = \cup S(i, w) \ \text{ and } \ S(i, w) \cap S(j, w) = Fr(S(i, w)) \cap Fr(S(j, w)), \ i \neq j.$$

Proof. We know that, if we have a set of affinely independent points $\{v^0, ..., v^n\}$ and a point w given by the equality (1) then $\{v^0, ..., v^{i-1}, w, v^{i+1}, ..., v^n\}$ is a set of affinely independent points all the times when the condition $\lambda_i > 0$ is verified in (1). So, all the sets $S(i, w)$ generated by the radial subdivision of n-simplices S are n-simplices.

Let $x \in S(i, w)$, namely,

$$x = \sum_{\substack{j=0 \\ j \neq i}}^{n} \mu_j v^j + \mu_i w, \ \ \mu_j \geq 0, \ (j = 0, ..., n), \ \ \mu_i + \sum_{\substack{j=0 \\ j \neq i}}^{n} \mu_j = 1, \tag{2}$$

with equality (1) we have

$$x = \sum_{\substack{j=0 \\ j\neq i}}^{n} \mu_j v^j + \mu_i \sum_{k=0}^{n} \lambda_k v^k, \quad \mu_j \geq 0 \quad (j = 0, ..., n),$$

(3)

$$\lambda_k \geq 0 \quad (k = 0, ..., n), \quad \mu_i + \sum_{\substack{j=0 \\ j\neq i}}^{n} \mu_j = 1, \ \sum_{k=0}^{n} \lambda_k = 1,$$

which can be expressed by

$$x = \sum_{\substack{j=0 \\ j\neq i}}^{n} (\mu_j + \mu_i \lambda_j) v^j + \mu_i \lambda_i v^i. \tag{4}$$

By the equality (3) all the coefficients are ≥ 0. Moreover,

$$\sum_{\substack{j=0 \\ j\neq i}}^{n} (\mu_j + \mu_i \lambda_j) + \mu_i \lambda_i = 1 - \mu_i + \mu_i(1 - \lambda_i) + \mu_i \lambda_i = 1.$$

Consequently x is a convex combination of vertices $v^0, ..., v^n$ of S. It holds that $x \in S$.

It remains to show that $S \subset S(i, w)$. Let $x \in S$ be such that $x \neq w$, we consider the segment $\rho(w, x) = \{\alpha(x - w) + w, \alpha \geq 0\}$ of w to x. The face of S is F the dimension of which is the smallest among those of the faces who contain x and w, and let y be the point where $\rho(w, x)$ joins the relative boundary to the face F. Then we have

$$y = \sum_{i\in I} \mu_i v^i, \ \mu_i \geq 0, \ (i \in I), \ \sum_{i\in I} \mu_i = 1, \ I \subset \mathbb{N}, \ |I| < n + 1. \tag{5}$$

By construction, we have also

$$x = \overline{\alpha} y + (1 - \overline{\alpha}) w, \ \ 0 < \overline{\alpha} < 1, \tag{6}$$

with equality (5) we obtain

$$x = \sum_{i\in I} \overline{\alpha} \mu_i v^i + (1 - \overline{\alpha}) w, \tag{7}$$

where

$$\overline{\alpha}\mu_i \geq 0 \ \ (i \in I), (1 - \overline{\alpha}) \geq 0 \ \text{ and } \ \sum_{i\in I} \overline{\alpha}\mu_i + (1 - \overline{\alpha}) = 1.$$

Which implies $x \in \cup S(i, w)$.

Finally, let $x \in S(i, w) \cap S(k, w)$, then we have

$$x = \sum_{\substack{j=0 \\ j\neq i,k}}^{n} \lambda_j v^j + \lambda_k v^k + \lambda_i w = \sum_{\substack{j=0 \\ j\neq i,k}}^{n} \mu_j v^j + \mu_i v^i + \mu_k w,$$

(8)

$$\lambda_j \geq 0, \ \mu_j \geq 0, \ \ (j = 0, ..., n), \ \sum_{j=0}^{n} \lambda_j = \sum_{j=0}^{n} \mu_j = 1.$$

Thus

$$\sum_{\substack{j=0\\ j\neq i,k}}^{n} (\lambda_j - \mu_j)v^j + \lambda_k v^k + \mu_i v^i + (\lambda_i - \mu_k)w = 0.$$

Let us suppose that $\lambda_k \neq 0$. Then

$$v^k = \sum_{\substack{j=0\\ j\neq i,k}}^{n} \frac{(\lambda_j - \mu_j)}{\lambda_k} v^j + \frac{\mu_i}{\lambda_k} v^i + \frac{(\lambda_i - \mu_k)}{\lambda_k} w,$$

hence $v^k \in S(k, w)$ what is contradictory to the fact that w is replaced to build $S(k, w)$, namely, $v^k \notin S(k, w)$, thus $\lambda_k = 0$.

We show in the same way that $\mu_i = 0$.

Then we have

$$(\lambda_i - \mu_k)w = -\sum_{\substack{j=0\\ j\neq i,k}}^{n} (\lambda_j - \mu_j)v^j.$$

Let us suppose that $(\lambda_i - \mu_k) \neq 0$, then

$$w = \sum_{\substack{j=0\\ j\neq i,k}}^{n} \frac{(\lambda_j - \mu_j)}{(\lambda_i - \mu_k)} v^j\,;$$

but w belongs to the segment $\left[v^i, v^k\right]$, thus it cannot be a linear combination of v^j such that $j \neq i$ and $j \neq k$. Then

$$\lambda_i - \mu_k = 0 \Rightarrow \lambda_i = \mu_k.$$

Then it holds that

$$\sum_{\substack{j=0\\ j\neq i,k}}^{n} (\lambda_j - \mu_j)v^j = 0,$$

or the points v^j, for $j \neq i$ and $j \neq k$, are affinely independent thus $\lambda_j - \mu_j = 0$ $\forall$ $j \neq i$ and $j \neq k$. Then,

$$x = \sum_{\substack{j=0\\ j\neq i,k}}^{n} \lambda_j v^j + \lambda_i w \in Fr(S(i, w)),$$

because the combination does not contain the point v^k ($\lambda_k = 0$), and

$$x = \sum_{\substack{j=0\\ j\neq i,k}}^{n} \mu_j v^j + \mu_k w \in Fr(S(k, w)),$$

and also, the combination does not contain the point v^j ($\mu_k = 0$). Then

$$x \in Fr(S(i, w)) \cap Fr(S(k, w)).$$

□

To establish the convergence in the procedure Branch and bound by the radial subdivision of simplices, we have to make sure that the point w defined in (1) is chosen well so as to have the convergence of any decreasing sequence of simplices produced by the algorithm.

Definition 4. *We call diameter of a polytope P and denote it $\delta(P)$ the following value:*

$$\delta(P) = \max\{\|x - y\| : x, y \in P\}.$$

For a simplex S, $\delta(S)$ is the length of the longest edge of S.

Definition 5. *A subdivision is said exhaustive if $\lim\limits_{(q\to\infty)} \delta(P_q) = 0$ for all decreasing sequence $\{P_q\}_q$ of elements of the partition generated by the used subdivision.*

Remark 4. *The notion of exhaustiveness is introduced by Thoai and Tuy (*[13]*), but it turned out that it is not easy to verify with the process of the radial subdivision. Let us note, that there are procedures of cuttings of simplex which is not exhaustive.*

Example 1. *Let us set $n > 1$ and let $v_q^i, (i = 0, ..., n)$ be the vertices of the simplex S_q. Let us choose the point w defined by* (1) *as the barycenter of S_q, namely,*

$$w = w_q = \frac{1}{n+1}\sum_{i=0}^{n} v_q^i. \tag{9}$$

Let us build the decreasing sequence $\{S_q\}$ of simplices by using the radial subdivision with w given by (9)*, and let us suppose that for all q, S_{q+1} is obtained from S_q by the replacement of v_q^n by the barycenter w_q of S_q. Then, every simplex S_q contains the face conv $\left\{v_1^0, ..., v_1^{n-1}\right\}$ of the initial simplex S_1, and so $\overline{S} = \bigcap\limits_{q=1}^{\infty} S_q$ has a non zero diameter, because $\delta(\overline{S}) = \delta\left\{v_1^0, ..., v_1^{n-1}\right\}$.*

We notice that the number of sub-simplices $S(i, w)$ is $d + 1$ when the smallest face F which contains w is of dimension d. For example if $w \in int(S)$ we obtain $n + 1$ sub-simplices, when w is on a edge of S we obtain 2 sub-simplices. This last partition is called the bisection of S, w is the middle point of the largest edge of S, namely,

$$w = \frac{1}{2}(v^r + v^s), \tag{10}$$

where $[v^r, v^s]$ is the largest edge of S. In that case, S was subdivided into 2 n-simplices. This partition is introduced by Horst ([8]*), it belongs to the exhaustive class of the subdivisions.*

The exhaustiveness of subdivision of simplices by the bisection method is illustrated by the following result:

Proposition 2. *(*[11]*) Let $\{S_q\}_q$ be a decreasing sequence of n-simplices generated by the bissection method. Then we have*

(*i*) $\delta(S_{n+q}) \leq \frac{\sqrt{3}}{2}\delta(S_q) \quad \forall q,$

(*ii*) $\delta(S_q) \to 0$, *as* $q \to \infty$.

4.2. **Rectangles.** For certain classes of problems, n-rectangles R are polyhedrons the most suited to use to assure the convergence towards the global optimum, they are used to resolve certain problems of Lipschitz optimization ([9], [10]). On the other hand, the rectangular sets agree if the functions involved in the problem are separable, namely, the sum of n functions in a single variable, because in that case the appropriate lower bounds are often available.

Let us note that the rectangle R is determined by its vertices $a = (a_1, ..., a_n)^t$ ‹Low to the left›and $b = (b_1, ..., b_n)^t$ ‹upper to the right›, and every 2^n vertices of the rectangle R is as follows:

$$v = a + c,$$

where c is the vector of components equal to 0 or $(b_i - a_i)$, $i \in \{1, ..., n\}$.

Let us consider an n-rectangle R and let $w \in R$ be such that $w \notin V(R)$, where $V(R)$ is the set of vertices of R. Then the radial subdivision of R by using w (defined in the same way in the case of simplices) is not a partition of R in n-rectangles, but a partition of more complicated sets.

Consequently, subdivision of n-rectangles is usually defined through the passing hyperplans by w and parallel to the facets of R, so that R is divided into 2^n rectangles.

For the most part of the algorithms, subdivision must be exhaustive. For example in the method of the bisection, where w is the point middle of the longest edge of R. R is divided into two rectangles having the same volume and w is a vertex for two new rectangles. In the same way as the Proposition 2, we can show that subdivision by the bissection is exhaustive.

4.3. **The Polyhedral Cones.** The polyhedral cones are frequently used in the problems of concave minimization on convex and compact feasible sets ([13]). In the Branch and Bound method the most used conical partition is the radial subdivision by simplices in the following way: let

$$C := \left\{ x \in \mathbb{R}^n : x = \sum_{j=1}^{n} \mu_j d_j,\ \mu_j \geq 0,\ (j = 1, ..., n) \right\}$$

be a polyhedral cone generated by n extreme directions linearly independent d_j $(j = 1, ..., n)$. Let H be an hyperplan getting through all the directions d_j. Let us put

$$H := \left\{ x : x = Q\lambda, \text{ such that } \ e^t\lambda = 1 \right\},$$

where $\lambda \in \mathbb{R}^n$, $e^t = (1, ..., 1) \in \mathbb{R}^n$ and $Q = \{d_1, ..., d_n\}$ the matrix among which columns $d_1, ..., d_n$.

The intersection of H with C is an $(n-1)$-simplex

$$S = H \cap C = \{x : x = Q\lambda, \text{ such that } e^t\lambda = 1, \lambda \geq 0\}$$

$$= conv \{d_1, ..., d_n\}.$$

Let $w \in S$ be such that $w \neq d_j$, $(j = 1, ..., n)$. We consider a radial subdivision $\{S(i, w) : \lambda_i > 0\}$ of S inferred by w. For every $S(i, w)$ of the partition of S, note by $C(i, w)$ the polyhedral cone generated by the vertices of $S(i, w)$. Then, the presentation

$$w = \sum_{j=1}^{n} \lambda_j d_j, \quad \sum_{j=i}^{n} \lambda_j = 1, \quad \lambda_j \geq 0,$$

makes of $C(i,w)$ a cone with extreme directions $d_1,...,d_{i-1},w,\ d_{i+1},...,d_n$.

It ensues that the family of $C(i,w)$ built in this way forms a partition of C in polyhedral cones having the same structure of C.

5. Calculation of the Lower Bounds

Now, we give some examples for the calculations of the lower bounds. Because the calculation of the lower bounds $\beta(M)$ depends on data of the problem of optimization $(\mathcal{P})$, we are going to discuss at first certain ideas connected to classes of problems which we meet frequently.

We have to indicate that bounds built by the controversial operations of calculation below are not inevitably monotonous as required in the step $(\mathbf{k}_4)$ in the algorithm. Indeed, let $M \subset M^{'}$ be an element of one partition of $M^{'}$ such that $\beta(M) < \beta(M^{'})$. Then it is advisable to use

$$\overline{\beta}(M) = \max\left\{\beta(M),\, \beta(M^{'})\right\},$$

on the place of $\beta(M)$ as it is defined above.

5.1. **Lipschitz Optimization.** Let f be a Lipschitzian function on M, namely, there exists $L > 0$ such that

$$|f(y) - f(x)| \leq L\,\|y - x\|\,,\ \ \forall x, y \in M \tag{11}$$

where $\|.\|$ is the euclidean norm. because it is difficult generally to find the exact value of L on every subset

We suppose that an upper bound A of L is known on the whole domain X, then we can use A as a constant of Lipschitz instead of L, because it is difficult generally to find the exact value of L on every subset $M \subset X$. However, there are many problems where L can be easily determined. Let us note that in the algorithm Branch and bound the local bounds of L on sets partitions M must be used on the place of the global bounds on M_0. We admit that the diameter $\delta(M)$ of M is known, such as for a simplex S, $\delta(S)$ is the length of the longest edge of S, and for a rectangle, $\delta(R)$ is the length of the main diagonal.

By (11), we have

$$f(y) \geq f(x) - L\,\|y - x\| \geq f(x) - A\delta(M), \quad \forall x, y \in M. \tag{12}$$

Let $V^{'}(M)$ be a nonempty subset of the vertices set $V(M)$ of M. Then

$$\beta(M) = \max\left\{f(x) : x \in V^{'}(M)\right\} - A\delta(M), \tag{13}$$

is a lower bound.

In (13), $V^{'}(M)$ can be replaced by known subset of M.

For example, if $M = \{x \in \mathbb{R}^n : a \leq x \leq b\}$ is a hyper-rectangle, then

$$\beta(M) = f(\frac{1}{2}(a+b)) - \frac{A}{2}\delta(M),$$

could be a better choice than (13).

5.2. **A Lower on a Vertex.** Let M be a polytope, for certain classes of objective functions, the lower bounds of $\inf f(M \cap X)$ or of $\inf f(M)$ can be simply determined by minimization of certain function connected to f by finite sets of vertices $V(M)$ of M. For example, if f is concave on M, we have

$$\inf f(M \cap X) \geq \min f(M) = \min f(V(M)),$$

then we have to choose

$$\beta(M) = \min f(V(M)). \tag{14}$$

More precise bounds can be obtained by subdividing parts of $M \setminus X$ by a procedure of outside approximation via the cutting of the plans. The resultant polytope P will satisfy

$$M \cap X \subseteq P \subset M.$$

Consequently

$$\inf f(M \cap X) \geq \min f(V(P)) \geq \min f(M), \tag{15}$$

and $\beta(M) = \min f(V(P))$ is, generally, a lower bound more precise than $\min f(V(M))$.

On the other hand, to determine the lower bounds of a d-c function of the type,

$$f(x) = f_1(x) + f_2(x),$$

where f_1 is concave and f_2 is convex on the polytope M, we can take an arbitrary point $v^* \in M$ and define $\partial f_2(v^*)$ a subdifferentiel of f_2 on v^*.

Let $p^* \in \partial f_2(v^*)$ and determine

$$\overline{v} \in \arg\min \left\{ f_1(v) + f_2(v^*) + p^*(v - v^*) : v \in V(M) \right\}. \tag{16}$$

Then, by the definition of subgradient, we have

$$\ell(x) = f_2(v^*) + p^*(x - v^*) \leq f_2(x), \quad \forall x \in \mathbb{R}^n,$$

consequently

$$f_1(x) + \ell(x) \leq f(x), \qquad \forall x \in \mathbb{R}^n.$$

Or $f_1(x) + \ell(x)$ is concave, then it attains its minimum on M one of its vertices. Consequently,

$$\beta(M) = f_1(\overline{v}) + f_2(v^*) + p^*(\overline{v} - v^*) \tag{17}$$

is a lower bound of

$$\min f(M) \leq \inf f(M \cap X).$$

5.3. **Convex Subfunctionals.** An approach usually used to calculate the lower bounds $\beta(M)$ of $\min f(M \cap X)$ or $\min f(M)$ consists in minimizing a certain convex subfunctional of f on $M \cap X$ or on M.

- By definition, a convex subfunctional of f on M is a convex function which does not exceed f on M.
- A convex subfunctional φ is a convex hull, if no other one convex subfunctional of f does not exceed φ.

The convex hull plays an important theoretical role in the global optimization ([8]). Because it represents the best convex lower apoproximation of f on M.

In what follows we give some techniques of construction of the convex hulls of f.

5.3.1. ***Convex Hull of a Lower Semicontinuous (l.s.c.) Function.***

Definition 6. *Let $M \subset \mathbb{R}^n$ be a convex compact subset, and let $f : M \to \mathbb{R}$ be l.s.c. on M.*

A function $\varphi : M \to \mathbb{R}$ is said convex hull of f on M if it satisfies:
(i) $\varphi(x)$ is convex on M,
(ii) $\varphi(x) \leq f(x) \qquad \forall x \in M$,
(iii) there is no function $\psi : M \to \mathbb{R}$ which satisfies (i), (ii) and $\varphi(\overline{x}) < \psi(\overline{x})$ for some point $\overline{x} \in M$.

We can show that a convex hull is unique, if it exists.

Theorem 3. *Let $f : M \to \mathbb{R}$ be a l.s.c. function on the convex and compact set $M \subset \mathbb{R}^n$, and let φ be the convex hull of f on M. Then we have*
a) $\qquad \min \varphi(M) = \min f(M)$,
b) $\qquad \arg\min \varphi(M) \supset \arg\min f(M)$.

Proof. *a)* Let us put

$$f^* = \min f(M) = \min \{f(x) : x \in M\}.$$

By the Definition 6, we have $\varphi(x) \leq f(x)$ for all $x \in M$. Then

$$\min \varphi(M) = \min \{\varphi(x) : x \in M\} \leq \min \{f(x) : x \in M\} = \min f(M).$$

The constant function $\psi(x) = f^*$ is a convex underestimator of f. Still by the definition of the convex hull we have

$$\varphi(x) \geq f^* \text{ for all } x$$

and so

$$\min \varphi(M) = \min \{\varphi(x) : x \in M\} \geq f^* = \min f(M).$$

b) Let us show by the absurd that

$$\arg\min f(M) \subset \arg\min \varphi(M)?$$

Indeed; let $x^* \in \arg\min f(M)$, let us suppose that $x^* \notin \arg\min \varphi(M)$. We take y^* a global minimum of $\varphi(x)$ on M. Then

$$\varphi(y^*) < \varphi(x^*) \leq f^*.$$

Consider the convex function $H(x) = \max \{f^*, \varphi(x)\}$, we have

$$f(x) \leq H(x) \quad \text{and} \quad H(x) \geq \varphi(x) \quad \text{ for all } x \in M.$$

In particular, $H(y^*) \geq \varphi(y^*)$, what contradicts the fact that $\varphi(x)$ is the convex hull of f. Thus $\varphi(y^*) = \varphi(x^*)$ and $x^* \in \arg\min \varphi(M)$. □

By the previous theorem, several nonconvex optimization problems can be replaced by convex optimization problems of convex optimization if the convex hull φ of f is easy to calculate.

The geometrical interpretation of the convex hull $\varphi(x)$ is given by the following lemma:

Lemma 4. *Let $M \subset \mathbb{R}^n$ be a convex compact set and let $f : M \to \mathbb{R}$ be a l.s.c. function on M. Then the function $\varphi : M \to \mathbb{R}$ is a convex hull of f on M if and only if,*

$$epi(\varphi) = conv(epi(f)) \tag{18}$$

or, also

$$\varphi(x) = \inf\{\alpha : (x,\alpha) \in conv(epi(f))\}. \tag{19}$$

Corollary 5. *Let $M \subset \mathbb{R}^n$ be a convex compact set and let $f : M \to \mathbb{R}$ be a l.s.c. function on M. Then φ and f coincide on extreme points of M.*

Proof. Applying Carathéodory's theorem to the equality (19), we notice that φ can be expressed as follows:

$$\varphi(x) = \left\{ \sum_{i=1}^{n+1} \lambda_i\ f(x^i)\ /\ \sum_{i=1}^{n+1} \lambda_i = 1,\ \lambda_i \geq 0,\ \sum_{i=1}^{n+1} \lambda_i x^i = x,\ x^i \in M,\ \text{for } i = 1, ..., n+1 \right\}.$$

In particular for an extreme point x we have $\varphi(x) = f(x)$. □

The convex hull φ can be, so, characterized by the conjugate function defined by Fenchel.

Theorem 6. *([12]) Let $f : M \to \mathbb{R}$ be a l.s.c. function on the convex and compact set $M \subset \mathbb{R}^n$. Then*

$$f^{**}(x) = \varphi(x), \quad \forall x \in \mathbb{R}^n. \tag{20}$$

5.4. Convex Hull of Some Special Classes of Functions.

Other results on the convex hulls are established by using special classes of functions defined on polytopes. We are going to quote two examples due to Falk and Horst ([11]).

Theorem 7. *Let P be a polytope of vertices $v^1, ..., v^k$ and $f : P \to \mathbb{R}$ be a concave function on P. Then the convex hull φ of f on P is expressed as follows:*

$$\begin{cases} \varphi(x) = \min\limits_{\alpha} \sum_{i=1}^{k} \alpha_i f(v^i), \\ \textit{subject to } \sum_{i=1}^{k} \alpha_i v^i = x, \quad \sum_{i=1}^{k} \alpha_i = 1, \ \alpha_i \geq 0 \end{cases} \tag{21}$$

where $\alpha = (\alpha_1, ..., \alpha_k)$.

Proof. Show that the function φ defined by (21) is convex. Indeed, let $0 \leq \lambda \leq 1$, x^1, $x^2 \in M$ and α^1, α^2 coefficients of φ in (21). Then

$$\varphi(\lambda x^1 + (1-\lambda)x^2) \leq \sum_{i=1}^{k} (\lambda\alpha_i^1 + (1-\lambda)\alpha_i^2)\ f(v^i)$$

$$= \lambda \sum_{i=1}^{k} \alpha_i^1 f(v^i) + \sum_{i=1}^{k} (1-\lambda)\alpha_i^2\ f(v^i) = \lambda\varphi(x^1) + (1-\lambda)\varphi(\ x^2)$$

It remains to show that φ is the convex hull of f. By the concavity of f we have

$$\varphi(x) \leq \sum_{i=1}^{k} \alpha_i\ f(v^i) \leq f(x), \quad \forall x \in P.$$

Now, let us suppose that there is another convex function ψ which underestimates f in P and that $\varphi(\overline{x}) < \psi(\overline{x})$ for some $\overline{x} \in P$. Let $\overline{\alpha}$ be a solution of (21) when we replace x by $\overline{x}$, then we have

$$\varphi(\overline{x}) < \psi(\overline{x}) = \psi(\sum_{i=1}^{k} \overline{\alpha}_i v^i) \leq \sum_{i=1}^{k} \overline{\alpha}_i \psi(v^i) \leq \sum_{i=1}^{k} \overline{\alpha}_i\ f(v^i) = \varphi(\overline{x})$$

what contradicts our supposition. □

Theorem 8. *Let $S = conv\left\{v^0, ..., v^n\right\}$ be a n-simplex of vertices $v^0, ..., v^n$, and let $f : S \to \mathbb{R}$ be a concave function on S. Then the convex hull of f on S is the following affine mapping:*

$$\varphi(x) = a^t x + b, \quad a \in \mathbb{R}^n, \quad b \in \mathbb{R}, \tag{22}$$

where a and b are determined by the resolution of the system of linear equations:

$$f(v^i) = a^t v^i + b, \qquad (i = 0, ..., n). \tag{23}$$

Proof. The system (23) is constituted of $(n+1)$ linear equations with $(n+1)$ unknowns $a \in \mathbb{R}^n$, $b \in \mathbb{R}$.

If we deduct the first equation ($i = 0$) from all other equations, we find

$$(v^i - v^0)^t a = f(v^i) - f(v^0), \qquad \text{for } i = 1, ..., n.$$

The matrix V of coefficients among which columns are vectors $v^i - v^0$ $(i = 1, ..., n)$ is not singular, because vectors $v^i - v^0$ $(i = 1, ..., n)$ are linearly independent. What implies that the solution of the system (23) is unique, namely a and b are unique. But $\varphi(x) = a^t x + b$ is affine, consequently it is convex.

Let $x \in S$, then

$$x = \sum_{i=0}^{n} \lambda_i v^i, \qquad \sum_{i=0}^{n} \lambda_i = 1, \quad \lambda_i \geq 0 \quad (i = 0, ..., n).$$

It holds from the concavity of f that we have

$$\varphi(x) = \sum_{i=0}^{n} \lambda_i \varphi(v^i) = \sum_{i=0}^{n} \lambda_i\, f(v^i) \leq f(x).$$

Consequently $\varphi(x) \leq f(x)$, $\forall x \in S$, namely, φ is a subfunctional of f on S.

Now, let us suppose that there is another convex subfunctional ψ of f on S and a point $\overline{x} \in S$ which satisfies $\psi(\overline{x}) > \varphi(\overline{x})$. Then

$$\overline{x} = \sum_{i=0}^{n} \mu_i v^i, \; \sum_{i=0}^{n} \mu_i = 1, \quad \mu_i \geq 0 \quad (i = 0, ..., n)$$

and for $f : S \to \mathbb{R}$, we have

$$\psi(\overline{x}) = \psi(\sum_{i=0}^{n} \mu_i v^i) \leq \sum_{i=0}^{n} \mu_i \psi(v^i) \leq \sum_{i=0}^{n} \mu_i f(v^i) \; = \; \sum_{i=0}^{n} \mu_i \varphi(v^i) = \varphi(\overline{x}),$$

what contradicts our hypothesis. □

By the Theorem 8, the construction of the convex hull is particularly easy in the case of the concave functions of a single variable $f : [a, b] \to \mathbb{R}$. Then the graph of the convex hull is simply a right line passing by points $(a, f(a))$, $(b, f(b))$.

Theorem 9. *Let $R = \prod_{i=1}^{r} R_i$ be the product of r rectangles R_i $(i = 0, ..., r)$ compact of n_i-dimension such that $\sum_{i=0}^{r} n_i = n$. We suppose that $f : R \to \mathbb{R}$ can be decomposed* *under the following shape :* $f(x) = \sum_{i=0}^{r} f_i(x^i)$ *where* $f_i : R_i \to \mathbb{R}$*. Then the convex hull* φ *of* f *on* R *is the sum of convex hulls* φ_i *of* f_i *on* R_i*, namely,*

$$\varphi(x) = \sum_{i=0}^{r} \varphi_i(x^i).$$

5.5. **A Lower Bound by Duality.** Consider the following optimization problem:

$$(24) \qquad (\mathcal{P}) \quad \begin{cases} \min f(x) \\ \text{subject to } g_i(x) \le 0, \quad i = 1, ..., m \\ \qquad x \in C \end{cases}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is l.s.c., $g_i : \mathbb{R}^n \to \mathbb{R}$ are convex functions for $i = 1, ..., m$, and C is a nonempty convex and compact set.

The dual problem of $(\mathcal{P})$ is defined as follows:

$$(25) \qquad (\mathcal{D}) \qquad \max_{\lambda \in \mathbb{R}_+^m} \left\{ \inf_{x \in C} (f(x) + \lambda^t g(x)) \right\},$$

where $g(x) = (g_1(x), ..., g_m(x))$ and

$$\mathbb{R}_+^m = \{(\lambda_1, ..., \lambda_m) \in \mathbb{R}^m : \lambda_i \ge 0, \ i = 1, ..., m\}.$$

The objective function of the dual problem $(\mathcal{D})$ is defined by

$$d(\lambda) := \inf_{x \in C} (f(x) + \lambda^t g(x)).$$

This function is, by definition, concave because it is the infimum of a family of continuous affine functions.

Let $\inf(\mathcal{P})$ and $\sup(\mathcal{D})$ be optimal values of $(\mathcal{P})$ and $(\mathcal{D})$, respectively, then we have the following result:

Lemma 10. *(weak duality) We always have*

$$\inf(\mathcal{P}) \ge \sup(\mathcal{D}).$$

Proof. Let $\overline{\lambda} \ge 0$ and $\overline{x} \in C$ such that $d(\overline{\lambda}) > -\infty$ and $g(\overline{x}) \le 0$. Then

$$d(\overline{\lambda}) = \inf_{x \in C} (f(x) + \lambda^t g(x)) \le f(\overline{x}) + \overline{\lambda}^t g(\overline{x}) \le f(\overline{x}).$$

□

Let us note that this lemma is verified without the convexity of C, also without the continuity of g_i and f. Let

$$M = C \cap \{x : g_i(x) \le 0, \quad (i = 1, ..., m)\}.$$

Then, by the Lemma 10, any feasible point $\overline{\lambda}$ of the dual problem could be a point to divert the lower bounds $\beta(M) = d(\overline{\lambda})$ of $\min f(M)$.

For other results on the duality and the dual bounds, we can see our works (([1]), ([2]), ([3]), ([4]), ([5]) et ([6])).

6. Study of the Convergence

We are going to establish a convergence theorem in the case where the objective function f is continuous and the domain X is a polytope.

If the procedure ends in the iteration k, then obviously, x^k is the optimal solution and α_k is the optimal value of the objective function ($\alpha_k = f(x^k)$). However, generally, we cannot guarantee the stop of the algorithm in a finite number of iteration. We thus have to give conditions which assure that any cluster point of the sequence $\left\{x^k\right\}_k$ is an optimal solution of the problem $(\mathcal{P})$.

At first, let us note that if the algorithm is infinite, then it has to generate at least an extracted infinite sequence $\left\{M_{k_q}\right\}$ of sets M_{k_q} stemming from partitions successively sophisticated and verifying $M_{k_q} \supset M_{k_{q+1}}$. This is an immediate consequence of the fact that, in every iteration k, a class $\Im_k$ contains only a finite number of sets stemming from the partition. In the tree which represents the procedure Branch and bound the extracted infinite sequence corresponds to a branch knots of which are M_{k_q}, $(q = 1, 2, 3, ...)$.

Theorem 11. *If for any infinite sequence $\left\{M_{k_q}\right\}$, $M_{k_q} \supset M_{k_{q+1}}$, $(q = 1, 2, 3, ...)$ of sets partitions successively refined, and for all bounds in the iteration k_q we have*

$$\lim_{(q\to\infty)} (\alpha_{k_q} - \beta_{k_q}) = \lim_{(q\to\infty)} (\alpha_{k_q} - \beta(M_{k_q})) = 0 \tag{26}$$

then

$$\beta = \lim_{(k\to\infty)} \beta_k = \lim_{(k\to\infty)} f(x^k) = \lim_{(k\to\infty)} \alpha_k = \alpha, \tag{27}$$

and every cluster point x^* of the sequence $\left\{x^k\right\}_k$ is an optimal solution of $\min\{f(x) : x \in X\}$.

Proof. Suppose that $\left\{x^k\right\}_k$ is an infinite sequence. Since X is compact and $\left\{x^k\right\}_k \subset X$, then the sequence $\left\{x^k\right\}_k$ has a cluster point. Let x^* be a cluster point of $\left\{x^k\right\}_k$, then there exists a subsequence belonging to $\left\{x^k\right\}_k$ which converges to x^*. For the previous same reasons, we conclude that this subsequence has to contain an infinite subsequence $\left\{x^{k_q}\right\}$ such that subsequences $\left\{M_{k_q}\right\}$ corresponding verify $M_{k_q} \supset M_{k_{q+1}}$, $(q = 1, 2, 3, ...)$.

The continuity of f on the compact X implies that

$$\lim_{(q\to\infty)} f(x^{k_q}) = f(x^*).$$

Denote $f^* = \min\{f(x) : x \in X\}$. The sequence of lower bounds $\{\beta_k\}_k$ satisfy the conditions $\beta_{k+1} > \beta_k$ and $\beta_k < f^*$ then its limit exists. Put $\beta = \lim_{(k\to\infty)} \beta_k$.

Also, the sequence of upper bounds $\{\alpha_k\}_k$ satisfy the conditions $\alpha_k = f(x^k) \geq f^*$ has to have a limit α. Thus it holds that

$$\beta \leq f^* \leq \lim_{(k\to\infty)} f(x^k) = f(x^*) = \alpha,$$

but with the condition (26) which also spells under shape

$$\lim_{(q\to\infty)} \alpha_{k_q} = \lim_{(q\to\infty)} \beta_{k_q},$$

and

$$\begin{cases} \lim\limits_{(q\to\infty)} \alpha_{k_q} = \lim\limits_{(k\to\infty)} \alpha_k \text{ because } \left\{\alpha_{k_q}\right\} \text{ is a subsequence of } \left\{\alpha_k\right\}_k \\ \text{and} \\ \lim\limits_{(q\to\infty)} \beta_{k_q} = \lim\limits_{(k\to\infty)} \beta_k \text{ because } \left\{\beta_{k_q}\right\} \text{ is a subsequence of } \left\{\beta_k\right\}_k . \end{cases}$$

It holds that

$$\lim_{(k\to\infty)} \beta_k = \lim_{(k\to\infty)} \alpha_k.$$

□

References

[1] Ch. Daili, Contributions à la Résolution de Quelques Problèmes d'Optimisation Non Convexes et Applications, Thèse de Doctorat, Université F. ABBAS de Sétif, 2011.

[2] Ch. Daili and A. Ziadi, A New Solution of a Global Optimization Problem of $\mathcal{C}^2$-Class by Branch and Bound Method, Pacific Journal of Applied Mathematics (PJAM), Vol. 2, Isuue ♯1 (2010), pp. 83-90. Nova Sci. Publ. New York.

[3] Ch. Daili, N. Daili and B. Merikhi, Duality Gap and Quadratic Programming, Part *II*: Dual Bound of Quadratic Programming Problem and Applications, Pacific Journal of Applied Mathematics (PJAM), Vol. 3, Issue 4, (2011), pp. 25-51.

[4] Ch. Daili and N. Daili, Duality Gap and Quadratic Programming, Part *III*: Dual Bounds of a D.C. Quadratic Programming Problem and Applications, Global Jour. of Pure and Appl. Maths.(GJPAM), Vol.4, Issue 1 (2008), pp. 25-51.

[5] Ch. Daili and N. Daili, Duality Gap and Quadratic Programming, Part *I*, Far East J. Appl. Math. (FJAM), Vol. 29, Issue 2 (2007), pp. 311-334.

[6] Ch. Daili, N. Daili and B. Merikhi, Duality Gap and Quadratic Programming. Part, *II*: Dual bound of quadratic programming, Report *CRM*- 3242, Centre de Recherches Mathématiques, Montréal, 2007, *http://www.crm.umontreal.ca/cgi/recherche-annee.cgi*

[7] J.-B. Hiriart-Urruty and C. Lemarechal, Convex Analysis and Minimization Algorithms, Grundlehren der Mathematischen Wissenschaften 305-306, Springer-Verlag, Berlin, 1993.

[8] R. Horst, On the characterization on affine Hull functions, Zeitschrift fur Angewandte Mathematik and Mechanik (in German), 52(1976), pp. 347-348.

[9] R. Horst and H. Tuy, On the Convergence of Global Methods in Multiextremal Optimization, Journal of Optimization Theory and Applications, 54 (1987), pp. 253-271.

[10] R. Horst, Outer Cut Methods in global Optimization, Lecture Notes in Economics and Mathematical systems, 304 (1987), pp. 28-40.

[11] R. Horst and H. Tuy, Global Optimization - Deteterminitic Approaches, Springer-Varlag, Berlin, 1990, (2nd edn., 1993).

[12] R.T. Rockafellar, Convex Analysis, Princeton Mathematics Ser. 28, Princeton University Press, Princeton, New Jersey, 1993.

[13] N.V. Thoai and H. Tuy, Convergent Algorithms for minimizing a Concave Function, Mathematics of Operations Research, 1980.

[14] H. Xue and C. Xu, A Branch and bound algorithm for separable concave Programming, J. Computat. Math. J. Comp. Math, 22 (2004), pp. 881-894.

E-mail address: ch_dail@yahoo.fr