The theoretical and empirical rate of convergence for geometric branch-and-bound methods

Anita Schöbel · Daniel Scholz

Abstract   Geometric branch-and-bound solution methods, in particular the big square small square technique and its many generalizations, are popular solution approaches for non-convex global optimization problems. Most of these approaches differ in the lower bounds they use which have been compared empirically in a few studies. The aim of this paper is to introduce a general convergence theory which allows theoretical results about the different bounds used. To this end we introduce the concept of a ***bounding operation*** and propose a new definition of the ***rate of convergence*** for geometric branch-and-bound methods. We discuss the rate of convergence for some well-known bounding operations as well as for a new general bounding operation with an arbitrary rate of convergence. This comparison is done from a theoretical point of view. The results we present are justified by some numerical experiments using the Weber problem on the plane with some negative weights.

1 Introduction

In global optimization, we are dealing with non-convex optimization problems which are often hard to solve. Nevertheless, a wide range of applications can be found e.g. in Pardalos and Resende (2002). Therefore, facility location problems for instance are often solved using geometric branch-and-bound methods. Note that in the following we consider the case of minimization problems.

A. Schöbel · D. Scholz (✉)
Institut für Numerische und Angewandte Mathematik, Georg-August-Universität Göttingen, Göttingen, Germany
e-mail: dscholz@math.uni-goettingen.de

A. Schöbel
e-mail: schoebel@math.uni-goettingen.de

One of the first geometric branch-and-bound approaches in the area of facility location problems was suggested by Hansen et al. (1985), the ***big square small square*** technique for some facility location problems on the plane. Plastria (1992) generalized this method to the ***generalized big square small square*** technique. In both approaches, lower bounds on the objective function for some smaller rectangles or boxes are required. A general framework for the calculation of such bounds was presented in Plastria (1992).

Drezner and Suzuki (2004) proposed the ***big triangle small triangle*** method which is again a generalization of the big square small square technique. Here, the rectangles are replaced by triangles and the lower bounds are calculated in a more sophisticated way with methods from d.c. programming, see Drezner (2007) and Horst et al. (2000). Recently, Blanquero and Carrizosa (2009) improved the d.c. bounds using a d.c.m decomposition instead. Using again lower bounds from d.c. programming but boxes in higher dimensions, Schöbel and Scholz (2010) suggested the ***big cube small cube*** technique for facility location problems with multiple variables.

The interval branch-and-bound algorithm is a more general global optimization technique which can be applied to problems in every dimension using lower bounds obtained from interval analysis, see Hansen (1992) or Ratschek and Rokne (1988) for a survey on global optimization using interval analysis. In Tóth et al. (2009) and Fernández et al. (2007), some facility location problems were solved using the interval branch-and-bound algorithm. We remark that lower bounds obtained from the natural interval extension are the same ones as those presented in Plastria (1992) for the problems discussed therein.

Since several different approaches for calculating lower bounds in geometric branch-and-bound solution methods for global optimization problems and, in particular, facility location problems are known, our aim in this paper is to introduce a convergence theory with a new definition of the rate of convergence. The main task is to derive theoretical results for the rate of convergence comparing different bounding operations. Apart from the well-known bounding operations in location theory, we will present a general construction which achieves any given rate of convergence.

Moreover, our aim is to analyze the rate of convergence from the empirical point of view. Therefore, we consider the Weber problem with some negative weights on the plane, see Tuy et al. (1995) or Drezner and Suzuki (2004), and we compare both, the empirical rate of convergence using randomly selected boxes and the average number of iterations to solve large scale instances using the same input data but different bounding operations.

The remainder of the paper is organized as follows. In the next section we will summarize notations and basic concepts which we will use throughout the paper. Sect. 3 presents the big cube small cube prototype algorithm before we introduce the convergence theory in Sect. 4. In the following Sects. 5–7, the theoretical and the empirical rates of convergence for several bounding operations are discussed. Finally, we give a brief conclusion in Sect. 8.

2 Notations and basic concepts

Throughout the paper, we will use the following notations.

Notation 1 A compact ***box*** or ***hyperrectangle*** with sides parallel to the axes is denoted by

$$X = [\underline{x}_1, \ \overline{x}_1] \ \times \cdots \times \ [\underline{x}_n, \ \overline{x}_n] \subset \mathbb{R}^n.$$

The ***diameter*** of a box $X \subset \mathbb{R}^n$ is

$$\delta(X) = \max\{\|x - x'\|_2 : x, x' \in X\} = \sqrt{(\overline{x}_1 - \underline{x}_1)^2 + \cdots + (\overline{x}_n - \underline{x}_n)^2}$$

and the **center** of a box $X \subset \mathbb{R}^n$ is defined by

$$c(X) = \left(\frac{1}{2}(\underline{x}_1 + \overline{x}_1), \ldots, \frac{1}{2}(\underline{x}_n + \overline{x}_n) \right).$$

Assuming a feasible area $X \subset \mathbb{R}^n$ which is a box with sides parallel to the axes, our aim is to minimize the objective function $f : X \to \mathbb{R}$. To this end, geometric branch-and-bound algorithms are popular. They result in an ε-optimal solution and require the calculation of lower bounds.

Notation 2 Let $X \subset \mathbb{R}^n$ be a box and consider $f : X \to \mathbb{R}$. A **bounding operation** is a procedure to calculate for any subboxes $Y \subset X$ a **lower bound** $LB(Y) \in \mathbb{R}$ with

$$LB(Y) \le f(x) \quad \text{for all} \quad x \in Y$$

and to specify a point $P(Y) \in Y$.

Formally, we obtain the bounding operation $(LB(Y), P(Y))$ for all subboxes $Y \subset X$.

Note that the choice of $P(Y)$ is important for our theoretical results presented, since it affects the convergence of the proposed algorithm (see also Example 1).

For the discussion of the following bounds, we need the concept of d.c. functions, see e.g. Horst et al. (2000) or Horst and Thoai (1999).

Definition 3 Let $X \subset \mathbb{R}^n$ be a convex set. A function $f : X \to \mathbb{R}$ is said to be a **d.c. function** on X, if there exist two convex functions $g, h : X \to \mathbb{R}$ such that

$$f(x) = g(x) - h(x).$$

Note that every function $f : \mathbb{R}^n \to \mathbb{R}$ whose second partial derivatives are continuous everywhere is a d.c. function, see e.g. Horst and Thoai (1999).

We furthermore need the concept of subgradients, given e.g. in Hiriart-Urruty and Lemaréchal (2004).

Definition 4 Let $X \subset \mathbb{R}^n$ and let $f : X \to \mathbb{R}$ be a convex function. A vector $\xi \in \mathbb{R}^n$ is said to be a **subgradient** of f at $b \in X$ if

$$f(x) \ge f(b) + \xi^T(x - b) \quad \text{for all} \quad x \in X.$$

The set of all subgradients of f at b is called **subdifferential** of f at b and is denoted by $\partial f(b)$.

3 The prototype algorithm

This section briefly describes the proposed **big cube small cube** solution method and can be applied to minimize an arbitrary function

$$f : \mathbb{R}^n \to \mathbb{R},$$

see Schöbel and Scholz (2010) for the complete details. The algorithm is a generalization of the **big square small square** method for higher dimensions. Its main idea is similar to branch-and-bound algorithms using interval analysis or d.c. programming, see Hansen (1992) or Horst and Tuy (1996), respectively. We are bounding subboxes of the feasible area from below and we will split them into smaller ones if the bounds are not sharp enough.

The main differences between the big square small square algorithm, interval analysis, and d.c. programming are the bounding operations which we will discuss in Sects. 5 and 6.

The following big cube small cube algorithm finds an optimal solution $x^* \in C$ within an absolute accuracy of $\varepsilon > 0$. Assume that

$$X = [\underline{x}_1, \overline{x}_1] \times \cdots \times [\underline{x}_n, \overline{x}_n] \subset \mathbb{R}^n.$$

is the feasible area and $f : X \to \mathbb{R}$.

---

(1) Calculate a lower bound $LB(X)$ and set $UB = f(P(X))$ and $\mathcal{L} = \{X\}$.
(2) Choose a box with the lowest lower bound in $\mathcal{L}$, split it into s congruent smaller boxes $Y_1, \ldots, Y_s$, delete the selected box from $\mathcal{L}$, and add $Y_1, \ldots, Y_s$ to $\mathcal{L}$. Calculate lower bounds $LB(Y_1), \ldots, LB(Y_s)$ and update

$$UB = \min\{UB, f(P(Y_1)), \ldots, f(P(Y_s))\}.$$

   Delete all boxes Y from $\mathcal{L}$ with $LB(Y) + \varepsilon \geq UB$.
(3) When there are no boxes left, i.e. $\mathcal{L} = \emptyset$, the algorithm terminates and UB is within the absolute accuracy of ε from the optimum. If there are boxes left, return to step (2).

The main task during the algorithm is the calculation of lower bounds. We will analyze this topic in Sects. 5 and 6.

4 Convergence theory

In this section, we introduce our new convergence theory. A quite general theory concerning this topic can be found in Tuy and Horst (1988) and Horst and Tuy (1996). Therein, general convergence criteria without the explicit use of bounding operations are given. Drezner Drezner (2007) treated a location problem where the objective function is a sum of d.c. functions of the distance. Here, the quality of the proposed bounds is discussed. In Plastria (1992), a convergence theorem based on Lipschitzian optimization is given for planar facility location problems discussed therein, see our special case in Theorem 7.

Our goal in this section is to introduce a convergence theory with a new concept of the rate of convergence. A related concept for the specific use of interval analysis methods was introduced in Csallner and Csendes (1996). Their definition of the rate of convergence does not require the specification of a point $P(Y)$ for all boxes Y, see Notation 2, but they use the exact range $f(Y)$ of f on Y which cannot be derived in general. To the best of our knowledge, this concept was only analyzed from the empirical point of view and not from the theoretical one, see e.g. Tóth and Csendes (2005).

Definition 5 A bounding operation is called ***consistent*** if

$$\lim_{k \to \infty} f(P(Y_k)) - LB(Y_k) = 0$$

holds for all sequences $(Y_k)_{k \in \mathbb{N}}$ of boxes Y_k with $\lim_{k \to \infty} \delta(Y_k) = 0$.

The most important definition for our further considerations is the following one.

Definition 6 Let $X \subset \mathbb{R}^n$ be a box and $f : X \to \mathbb{R}$. Furthermore, consider the minimization problem

$$\min_{x \in X} f(x).$$

We say a bounding operation has the ***rate of convergence*** $p \in \mathbb{N}$ if there exists a fixed constant $C > 0$ such that

$$f(P(Y)) - LB(Y) \leq C \cdot \delta(Y)^p \tag{1}$$

for all boxes $Y \subset X$.

We remark that the rate of convergence does not depend on the norm used for the diameter of a box Y since all norms are equivalent in $\mathbb{R}^n$.

Lemma 1 *Consider a bounding operation which has a rate of convergence of $p \geq 1$. Then the bounding operation is consistent.*

Proof Consider an arbitrary sequences $(Y_k)_{k \in \mathbb{N}}$ of boxes Y_k with

$$\lim_{k \to \infty} \delta(Y_k) = 0.$$

Since the bounding operation has a rate of convergence of $p \geq 1$, it follows that

$$f(P(Y_k)) - LB(Y_k) \leq C \cdot \delta(Y_k)^p$$

for all $k \in \mathbb{N}$. Hence,

$$\lim_{k \to \infty} f(P(Y_k)) - LB(Y_k) \leq \lim_{k \to \infty} C \cdot \delta(Y_k)^p = 0.$$

Furthermore, with $LB(Y_k) \leq f(x)$ for all $k \in \mathbb{N}$ and all $x \in Y_k$, we have

$$\lim_{k \to \infty} f(P(Y_k)) - LB(Y_k) \geq 0.$$

Thus, we obtain

$$\lim_{k \to \infty} f(P(Y_k)) - LB(Y_k) = 0$$

which proves the lemma. $\square$

The following example shows that the choice of $P(Y)$ can have an impact on the rate of convergence.

Example 1 Consider the linear function $f(x) = \alpha x + \beta$ with $\alpha, \beta \in \mathbb{R}$ and for all boxes $Y = [\underline{y}, \overline{y}] \subset \mathbb{R}$ consider the bounding operation

$$LB_1(Y) = \min\{f(\underline{y}), \ f(\overline{y})\} \quad \text{and} \quad P_1(Y) = c(Y).$$

We then easily find

$$f(P_1(Y)) - LB_1(Y) = \frac{|\alpha|}{2}(\overline{y} - \underline{y}) = \frac{|\alpha|}{2} \cdot \delta(Y).$$

On the other hand, the bounding operation

$$LB_2(Y) = \min\{f(\underline{y}), \ f(\overline{y})\} \quad \text{and} \quad P_2(Y) = \arg \min_{x \in \{\underline{y}, \overline{y}\}} f(x).$$

yields

$$f(P_2(Y)) - LB_2(Y) = 0.$$

Both bounding operations are consistent, but the first one has a rate of convergence of not larger than $p = 1$ and the second one $p = \infty$.

In the remainder of this section we apply our concept to the big cube small cube method. Therefore, we restrict ourselves to the case where we split the selected box into $s = 2^n$ congruent smaller boxes. Of course, other subdivision rules are possible, too.

Theorem 2 *Consider the big cube small cube algorithm with a bounding operation which has a rate of convergence of $p \geq 1$. Furthermore, assume a feasible box X and the constants $\varepsilon, C > 0$ as before. Moreover, we assume the division rule which splits the selected box along each side, i.e. into $s = 2^n$ congruent smaller boxes.*

Then the worst case number of iterations for the big cube small cube method can be bounded from above by

$$\sum_{k=0}^{z} 2^{n \cdot k} \quad \text{where} \quad z = \left\lceil \log_2 \frac{\delta(X)}{\left(\frac{\varepsilon}{C}\right)^{1/p}} \right\rceil. \tag{2}$$

Proof Denote the list of boxes throughout the big cube small cube algorithm by $\mathcal{L}$ and note that

$$UB - LB(Y) \leq f(P(Y)) - LB(Y) \leq C \cdot \delta(Y)^p.$$

for all $Y \in \mathcal{L}$ in every iteration of the algorithm. If $\delta(Y)$ is small enough for all boxes $Y \in \mathcal{L}$, we have

$$UB - LB(Y) \leq C \cdot \delta(Y)^p \leq \varepsilon$$

for all $Y \in \mathcal{L}$, since $C > 0$ is a fixed constant. This is a sufficient condition for the termination of the algorithm since the termination rule is satisfied for all $Y \in \mathcal{L}$.

The next step is to calculate how small $\delta(Y)$ need to be in the worst case. From $C \cdot \delta(Y)^p \leq \varepsilon$ we obtain

$$\delta(Y) \leq \left(\frac{\varepsilon}{C}\right)^{1/p} =: K > 0.$$

As a connection between $\delta(X)$ and $\delta(Y)$, consider the smallest value of $N \in \mathbb{N}$ such that

$$\frac{\delta(X)}{N} \leq \delta(Y) \leq K, \quad \text{i.e.} \quad \frac{\delta(X)}{K} \leq N.$$

Thus, the length of each of the n sides of X should be smaller than or equal to the original length divided by N. Hence, if the initial box X is split into N^n congruent smaller boxes, the algorithm terminates.

We now want to count the number of iterations needed. After the first iteration, we split X into 2^n smaller boxes. If $N \leq 2$, the algorithms stops. Otherwise we need 2^n more iterations to split in the worst case all smaller boxes. Again, if $N > 4$, we need $2^n \cdot 2^n$ additional splits and so on. For the general case, define

$$z = \min\{k : k \in \mathbb{N}, N \leq 2^k\} = \lceil \log_2 N \rceil.$$

Table 1 Worst case number of iterations for various values of the rate of convergence p and the dimension n

n	p	# (Iterations)
2	1	1,398,101
2	2	1,365
2	3	341
3	1	1,227,133,513
3	2	37,449
3	3	4,681

Thus, the algorithm stops after

$$1 + 2^n + \left(2^n\right)^2 + \cdots + \left(2^n\right)^z = \sum_{k=0}^{z} 2^{n \cdot k}$$

iterations. This is exactly the same as stated in the theorem. □

Corollary 3 *Consider the big cube small cube algorithm with a bounding operation which has a rate of convergence of $p \geq 1$. Then for any $\varepsilon > 0$ the algorithm stops after finitely many steps with an ε-optimal solution.*

Proof Follows directly from Theorem 2. □

The following examples show the importance of the rate of convergence.

Example 2 Assume an initial box X with $\delta(X) = 1$ and an absolute accuracy of $\varepsilon = 10^{-2}$. Moreover, we assume the constant $C = 10$ for the rate of convergence. Table 1 summarizes the worst case numbers of iterations derived from (2) for various values of the rate of convergence p and the dimension n.

As can be seen, it is quite important to use a bounding operation with the greatest possible rate of convergence.

Example 3 Even if we suppose that every mth box can be discarded throughout the branch-and-bound phase, the worst case runtime can be bounded by

$$1 + \left(2^n - \frac{2^n}{m}\right) + \left(2^n - \frac{2^n}{m}\right)^2 + \cdots + \left(2^n - \frac{2^n}{m}\right)^z.$$

Using again an initial box X with $\delta(X) = 1$, an absolute accuracy of $\varepsilon = 10^{-2}$ and $C = 10$, Table 2 shows the worst case runtimes for the case $n = 3$ and various values of p and m.

5 Discussion of known bounds

Recall that a bounding operation is a rule to calculate a lower bound $LB(Y)$ and to specify a point $P(Y) \in Y$ for all subboxes

$$Y = [\underline{y}_1, \overline{y}_1] \times \cdots \times [\underline{y}_n, \overline{y}_n] \subset X.$$

In this section we will review some known bounding operations and we will discuss their theoretical rate of convergence.

	n	p	m	# (Iterations)
Table 2 Worst case number of iterations for various values of the rate of convergence p if every mth box is discarded	3	1	4	72,559,411
	3	2	4	9,331
	3	3	4	1,555
	3	1	2	1,398,101
	3	2	2	1,365
	3	3	2	341

5.1 D.c. bounding operation

Assume that $f(x)$ is a d.c. function on Y, say

$$f(x) = g(x) - h(x) \quad \text{for all} \quad x \in Y$$

where $g, h : X \to \mathbb{R}$ are convex functions, and set $c = c(Y) \in Y$. Since $g(x)$ is convex, there exists a subgradient ξ of g at c, see e.g. Rockafellar (1970). Hence, we have

$$a(x) := g(c) + \xi^T(x - c) \leq g(x) \quad \text{for all} \quad x \in Y.$$

Together we obtain

$$m(x) := a(x) - h(x) = g(c) + \xi^T(x - c) - h(x) \leq f(x) \quad \text{for all} \quad x \in Y. \qquad (3)$$

Note that $m(x)$ is concave, since the sum of concave functions is concave again and $a(x)$ and $-h(x)$ are concave. Moreover, denote by V the set of the 2^n vertices of Y. The *d.c. bounding operation* is then given by the lower bound

$$LB(Y) := \min_{v \in V} m(v)$$

for $f(x)$ on Y and the point

$$P(Y) \in \arg\min_{v \in V} m(v).$$

D.c. bounding operations have been used for several location problems, see e.g. Drezner and Drezner (2004), Drezner and Suzuki (2004), Schöbel and Scholz (2010), Tuy (1996).

We now discuss the rate of convergence of the d.c. bounding operation.

Lemma 4 *Consider a box $X \subset \mathbb{R}^n$, a subbox $Y \subset X$ with $c = c(Y)$, and some arbitrary functions $f, m : X \to \mathbb{R}$. Moreover, assume that $f(x)$ and $m(x)$ are twice continuously differentiable on X and suppose that*

$$f(c) = m(c) \quad \text{and} \quad \nabla f(c) = \nabla m(c),$$

where ∇ is the gradient. Then there exists a fixed constant $C > 0$ which does not depend on Y with

$$f(x) - m(x) \leq C \cdot \delta(Y)^2 \quad \text{for all} \quad x \in Y.$$

Proof Defining $g(x) := f(x) - m(x)$, the second order Taylor expansion of $g(x)$ at c yields

$$g(x) = g(c) + \nabla g(c)^T \cdot (x - c) + \frac{1}{2} \cdot (x - c)^T \cdot D^2 g(\vartheta(x)) \cdot (x - c)$$

$$= \frac{1}{2} \cdot (x - c)^T \cdot D^2 g(\vartheta(x)) \cdot (x - c),$$

where $\vartheta(x) \in Y$ for all $x \in Y$ and $D^2 g(\vartheta(x))$ is the Hessian matrix of g at $\vartheta(x)$.

Next, consider $D^2 g(\vartheta)$ for an arbitrary $\vartheta \in X$. Since all second partial derivatives of $g(x)$ are continuous on X, the Hessian matrix $D^2 g(\vartheta)$ is symmetric. Hence, for any $\vartheta \in X$ we find a decomposition of the form

$$D^2 g(\vartheta) = Q(\vartheta) \cdot \Lambda(\vartheta) \cdot Q(\vartheta)^T$$

where $Q(\vartheta)$ is orthogonal and $\Lambda(\vartheta) = \mathrm{diag}(\lambda_1(\vartheta), \ldots, \lambda_n(\vartheta))$ is a diagonal matrix which consists of the real eigenvalues of $D^2 g(\vartheta)$. Define

$$\theta := \max\{|\lambda_k(\vartheta)| : k = 1, \ldots, n, \vartheta \in X\}$$
$$= \max\{|\lambda| : \lambda \text{ is an eigenvalue of } D^2 g(\vartheta) \text{ for a } \vartheta \in X\}.$$

Note that $\theta < \infty$ since $D^2 g(x)$ is continuous on X. Moreover, we define

$$y(x) := Q(\vartheta(x))^T \cdot (x - c).$$

Since $\vartheta(x) \in Y$ for all $x \in Y$ and $Q(\vartheta)$ is orthogonal for all $\vartheta \in Y$, we have

$$\|y(x)\|_2 = \|Q(\vartheta(x))^T \cdot (x - c)\|_2 = \|(x - c)\|_2.$$

Using these results, we obtain

$$
\begin{aligned}
f(x) - m(x) &= \frac{1}{2}(x - c)^T \cdot D^2 g(\vartheta(x)) \cdot (x - c) \\
&= \frac{1}{2}(x - c)^T \cdot Q(\vartheta(x)) \cdot \Lambda(\vartheta(x)) \cdot Q(\vartheta(x))^T \cdot (x - c) \\
&= \frac{1}{2} y(x)^T \cdot \Lambda(\vartheta(x)) \cdot y(x) \\
&\leq \frac{1}{2}\theta \cdot y(x)^T y(x) = \frac{1}{2}\theta \cdot \|y(x)\|_2^2 \\
&= \frac{1}{2}\theta \cdot \|(x - c)\|_2^2 \leq \frac{1}{2}\theta \cdot \left(\frac{\delta(Y)}{2}\right)^2 = \frac{\theta}{8} \cdot \delta(Y)^2
\end{aligned}
$$

for all $x \in Y$ which proves the statement. $\square$

Theorem 5 *Consider $f(x) = g(x) - h(x)$ where $g(x)$ and $h(x)$ are convex functions and assume that $g(x)$ is twice continuously differentiable on X.*

Then the d.c. bounding operation has a rate of convergence of $p = 2$.

Proof With $c = c(Y)$ and $\nabla g(c) \in \mathbb{R}^n$ recall that

$$m(x) = g(c) + \nabla g(c)^T (x - c) - h(x),$$

see (3). Hence, we find

$$
\begin{aligned}
m(c) &= g(c) - h(c) = f(c), \\
\nabla m(c) &= \nabla g(c) - \nabla h(c) = \nabla f(c).
\end{aligned}
$$

Therefore, Lemma 4 yields

$$f(P(Y)) - LB(Y) = f(P(Y)) - m(P(Y)) \leq C \cdot \delta(Y)^2$$

for a $C > 0$ which does not depend on Y. This proves the rate of convergence of $p = 2$. $\square$

The following example shows that Theorem 5 cannot be improved, i.e. a rate of convergence of $p = 2$ is the highest we can expect.

Example 4 Consider $f(x) = x^3$ and the initial box $X = [0, 2]$. The function $f(x)$ is convex on X and therefore a d.c. function. For the d.c. bounding operation we hence set $g(x) = f(x)$ and $h(x) = 0$.

For the sequence $Y_\mu = [1 - \mu, 1 + \mu] \subset X$ with $0 < \mu < 1$ and $c = c(Y_\mu) = 1$, the d.c. bounding operation yields

$$m(x) = f(c) + f'(c)(x - c) = 3x - 2,$$

and hence

$$LB(Y_\mu) = 3(1 - \mu) - 2 = 1 - 3\mu,$$

since $P(Y_\mu) = \arg\min_{v \in V} m(v) = 1 - \mu$. Hence, we obtain

$$\frac{f(P(Y_\mu)) - LB(Y_\mu)}{\delta(Y_\mu)^3} = \frac{1 - 3\mu + 3\mu^2 - \mu^3 - (1 - 3\mu)}{8\mu^3}$$

$$= \frac{3\mu^2 - \mu^3}{8\mu^3} = \frac{3}{8\mu} - \frac{1}{8}.$$

Since $3/8\mu$ is unbounded for $\mu \to 0$, we cannot find a fixed constant $C > 0$ such that

$$f(P(Y_\mu)) - LB(Y_\mu) \leq C \cdot \delta(Y_\mu)^3$$

for all $\mu > 0$. Therefore, the d.c. bounding operation does in general not have a rate of convergence of $p \geq 3$.

5.2 D.c.m. bounding operation

Recently, Blanquero and Carrizosa (2009) improved the d.c. bounding operation using a d.c.m. decomposition which we want to analyze in this subsection from the theoretical point of view.

Definition 7 A function $\varphi : [0, \infty) \to \mathbb{R}$ is said to be a ***d.c.m. function***, if there exist two convex and monotone functions $\varphi^1, \varphi^2 : [0, \infty) \to \mathbb{R}$ such that

$$\varphi(x) = \varphi^1(x) - \varphi^2(x).$$

Assume that the objective function can be written as

$$f(x) = \sum_{k=1}^{s} \varphi_k(d_k(a_k, x)),$$

where $a_1, \ldots, a_s \in \mathbb{R}^n$ are some given demand points and $d_k(a_k, x)$ for $k = 1, \ldots, s$ are some convex distance functions. Moreover assume that the functions $\varphi_k : [0, \infty) \to \mathbb{R}$ are d.c.m. functions, i.e. we find

$$\varphi_k(x) = \varphi_k^1(x) - \varphi_k^2(x)$$

for $k = 1, \ldots, s$ and $\varphi_k^1(x)$ and $\varphi_k^2(x)$ are convex and monotone on $[0, \infty)$.

In order to calculate a lower bound for an arbitrary subbox $Y \subset X \subset \mathbb{R}^n$, Blanquero and Carrizosa (2009) distinguish between two cases for both, for $\varphi_k^1(x)$ and for $\varphi_k^2(x)$ for all $k = 1, \ldots, s$.

(1) If $\varphi_k^1(x)$ is non-increasing, define $p_k = d_k(a_k, c)$ with $c = c(Y)$. Then any subgradient $\xi_k^1 \in \mathbb{R}$ of $\varphi_k^1(x)$ at p_k yields

$$\psi_k^1(x) := \varphi_k^1(p_k) + \xi_k^1(x - p_k) \le \varphi_k^1(x).$$

Since $\varphi_k^1(x)$ is non-increasing, we have $\xi_k^1 \le 0$. Hence,

$$\phi_k^1(x) := \psi_k^1(d_k(a_k, x)) = \varphi_k^1(p_k) + \xi_k^1(d_k(a_k, x) - p_k) \le \varphi_k^1(d_k(a_k, x))$$

and $\phi_k^1(x)$ is concave.

(2) If $\varphi_k^1(x)$ is non-decreasing, define $\psi_k^1(x) := \varphi_k^1(d_k(a_k, x))$. Then also $\psi_k^1(x)$ is convex since we assumed $d_k(a_k, x)$ to be convex and the composition of a non-decreasing convex function with a convex function is convex again. Hence, any subgradient $\xi_k^1 \in \mathbb{R}^n$ of $\psi_k^1(x)$ at $c = c(Y)$ yields

$$\phi_k^1(x) := \psi_k^1(c) + (\xi_k^1)^T(x - c) = \varphi_k^1(d_k(a_k, c)) + (\xi_k^1)^T(x - c)$$
$$\le \psi_k^1(x) = \varphi_k^1(d_k(a_k, x))$$

for all $x \in Y$. Note that $\phi_k^1(x)$ is affine linear and therefore concave.

For the functions $\varphi_k^2(x)$, we obtain the following two cases.

(1) If $\varphi_k^2(x)$ is non-increasing, let $\xi_k^2 \in \mathbb{R}^n$ be any subgradient of $d_k(a_k, x)$ at $c = c(Y)$ and define

$$\psi_k^2(x) := d_k(a_k, c) + (\xi_k^2)^T(x - c) \le d_k(a_k, x).$$

Since $\varphi_k^2(x)$ is non-increasing, we find that

$$\phi_k^2(x) := -\varphi_k^2(\psi_k^2(x)) = -\varphi_k^2\left(d_k(a_k, c) + (\xi_k^2)^T(x - c)\right)$$
$$\le -\varphi_k^2(d_k(a_k, x))$$

and $\phi_k^2(x)$ is concave since $\psi_k^2(x)$ is linear and $-\varphi_k^2(x)$ is concave.

(2) If $\varphi_k^2(x)$ is non-decreasing, define

$$\phi_k^2(x) := -\varphi_k^2(d_k(a_k, x)).$$

Since $\varphi_k^2(x)$ is non-decreasing and convex and $d(a_k, x)$ is convex, $\varphi_k^2(d_k(a_k, x))$ is convex again as before. Hence, $\phi_k^2(x)$ is concave.

Summarizing these results, we found

$$m(x) := \sum_{k=1}^{s} (\phi_k^1(x) + \phi_k^2(x)) \le \sum_{k=1}^{s} (\varphi_k^1(d_k, x) - \varphi_k^2(d_k, x)) = f(x)$$

for all $x \in Y$ and $m(x)$ is a sum of concave functions and therefore concave again. Denote by V the set of the 2^n vertices of Y. Then we have the *d.c.m. bounding operation*

$$LB(Y) := \min_{v \in V} m(v),$$
$$P(Y) \in \arg\min_{v \in V} m(v).$$

We now want to show that we again obtain a rate of convergence of $p = 2$.

Theorem 6 *Consider*

$$f(x) = \sum_{k=1}^{s} \left(\varphi_k^1(d_k(a_k, x)) - \varphi_k^2(d_k(a_k, x)) \right)$$

where the functions $\varphi_k^1(x)$ and $\varphi_k^2(x)$ are convex, monotone, and twice continuously differentiable on $[0, \infty)$ and the distance functions $d_k(a_k, x)$ are convex and twice continuously differentiable on X for $k = 1, \ldots, s$.

Then the d.c.m. bounding operation has a rate of convergence of $p = 2$.

Proof Making use of Lemma 4 again it is sufficient to show

$$f(c) = m(c) \quad \text{and} \quad \nabla f(c) = \nabla m(c)$$

for all subboxes $Y \subset X$ where $c = c(Y)$. Therefore, we distinguish again between all possible cases.

(1) If $\varphi_k^1(x)$ is non-increasing, we have

$$\phi_k^1(x) = \varphi_k^1(p_k) + \left(\varphi_k^1 \right)' (p_k) \cdot (d_k(a_k, x) - p_k)$$

with $p_k = d_k(a_k, c)$. Hence, we find

$$\phi_k^1(c) = \varphi_k^1(p_k) = \varphi_k^1(d_k(a_k, c)),$$

$$\nabla \phi_k^1(c) = \left(\varphi_k^1 \right)' (p_k) \cdot \nabla d_k(a_k, c)$$
$$= \left(\varphi_k^1 \right)' (d_k(a_k, c)) \cdot \nabla d_k(a_k, c) = \nabla \varphi_k^1(d_k(a_k, c)),$$

where we used the generalized chain rule.

(2) If $\varphi_k^1(x)$ is non-decreasing, we have

$$\phi_k^1(x) = \varphi_k^1(d_k(a_k, c)) + \nabla \varphi_k^1(d_k(a_k, c))^T \cdot (x - c)$$

and we directly find $\phi_k^1(c) = \varphi_k^1(d_k(a_k, c))$ and $\nabla \phi_k^1(c) = \nabla \varphi_k^1(d_k(a_k, c))$.

For the functions $\phi_k^2(x)$ we find similar results:

(1) If $\varphi_k^2(x)$ is non-increasing, we have

$$\phi_k^2(x) = -\varphi_k^2 \left(d_k(a_k, c) + \nabla d_k(a_k, c)^T \cdot (x - c) \right)$$

and we find

$$\phi_k^2(c) = -\varphi_k^2(d_k(a_k, c)),$$

$$\nabla \phi_k^2(c) = - \left(\varphi_k^2 \right)' (d_k(a_k, c)) \cdot \nabla d_k(a_k, c) = -\nabla \varphi_k^2(d_k(a_k, c))$$

using again the generalized chain rule.

(2) If $\varphi_k^2(x)$ is non-decreasing, we have $\phi_k^2(x) = -\varphi_k^2(d_k(a_k, x))$ and therefore

$$-\varphi_k^2(d_k(a_k, x)) - \phi_k^2(x) = 0 \quad \text{for all} \quad x \in Y.$$

In summary we obtain $m(c) = f(c)$ and $\nabla m(c) = \nabla f(c)$, hence Lemma 4 yields

$$f(P(Y)) - LB(Y) = f(P(Y)) - m(P(Y)) \leq C \cdot \delta(Y)^2$$

for a $C > 0$ which does not depend on Y and the rate of convergence of $p = 2$ is shown. $\square$

Next, we show that the d.c.m. bounding operation does in general not have a rate of convergence of $p > 2$.

Example 5 Consider one demand point $a_1 = 0$, the distance function $d_1(a_1, x) = x^2$, and the d.c.m. functions $\varphi_1^1(x) = x^2$ and $\varphi_1^2(x) = 0$. Note that $\varphi_1^1(x)$ is convex and non-decreasing on $[0, \infty)$. Thus, the objective function is

$$f(x) = \varphi_1^1(d_1(a_1, x)) = x^4.$$

Assume the initial box $X = [0, 2]$ and the sequence $Y_\mu = [1-\mu, 1+\mu] \subset X$ with $0 < \mu < 1$ and $c = c(Y_\mu) = 1$. Hence, the d.c.m. bounding operation with

$$m(x) = \phi_1^1(x) = f(c) + f'(c) \cdot (x - 1) = 1 + 4 \cdot (x - 1) = 4x - 3$$

yields $P(Y_\mu) = 1 - \mu$ and

$$LB(Y_\mu) = 4(1 - \mu) - 3 = 1 - 4\mu.$$

Thus, we obtain

$$\frac{f(P(Y_\mu)) - LB(Y_\mu)}{\delta(Y_\mu)^3} = \frac{(1 - \mu)^4 - (1 - 4\mu)}{8\mu^3} = \frac{3}{4\mu} - \frac{1}{2} + \frac{\mu}{8}$$

and we cannot find a fixed constant $C > 0$ such that

$$f(P(Y_\mu)) - LB(Y_\mu) \leq C \cdot \delta(Y_\mu)^3$$

for all $\mu > 0$. Therefore, the d.c.m. bounding operation does in general not have a rate of convergence of $p \geq 3$.

5.3 Bounding operation for location problems

In this subsection we discuss the bounding operation presented in Plastria (1992) for location problems from the point of view of our new concept. Let the objective be a function of distances between given demand points and a new single facility location.

Consider s demand points $a_1, \ldots, a_s \in \mathbb{R}^n$ and for each demand point a distance function $d_k(a_k, x)$ for $k = 1, \ldots, s$. Furthermore, we assume that $h : \mathbb{R}^s \to \mathbb{R}$ is a Lipschitzian function such that we are in a position to solve problems of the form

$$\min\{h(z) : l_k \leq z_k \leq u_k \text{ for } k = 1, \ldots, s\},$$

where $l_k, u_k \in \mathbb{R}$ are some parameters for $k = 1, \ldots, s$. The objective function for our facility location problem is then given by

$$f(x) = h(d_1(a_1, x), \ldots, d_s(a_s, x)). \tag{4}$$

In order to calculate a lower bound $LB(Y)$ for an arbitrary box $Y \subset \mathbb{R}^n$, suppose that the values

$$d_k^{min}(Y) = \min\{d_k(a_k, x) : x \in Y\},$$
$$d_k^{max}(Y) = \max\{d_k(a_k, x) : x \in Y\}$$

for $k = 1, \ldots, s$ are easily derived. This is the case if d_k are norms or polyhedral gauges, see Plastria (1992).

It is obvious, that now

$$LB(Y) = \min\{h(z) : d_k^{min} \le z_k \le d_k^{max} \quad \text{for } k = 1, \dots, s\},$$
$$P(Y) = c(Y)$$

is a *bounding operation for location problems*.

We remark that this bounding operation results in exactly the same lower bound if we use objective functions as given in (4) and the natural interval extension from interval analysis, see Hansen (1992) or Ratschek and Rokne (1988) for further information on this subject. Recently, Fernández et al. (2007) and Tóth et al. (2009) solved some competitive location problems using the natural interval extension.

The main result of Plastria (1992) can now be reformulated within our new concept as follows.

Theorem 7 *Consider*

$$f(x) = h(d_1(a_1, x), \dots, d_s(a_s, x))$$

where $h : \mathbb{R}^s \to \mathbb{R}$ is a Lipschitzian function with constant L and $d_1, \dots, d_s$ are some norms. Then the bounding operation for location problems has a rate of convergence of $p = 1$.

Finally, the next example shows that the bounding operation for location problems cannot be improved to a rate of convergence of $p > 1$.

Example 6 Consider the two demand points $a_1 = 0$ and $a_2 = 2$, the initial box $X = [0, 2] \subset \mathbb{R}$, and the objective function

$$f(x) = d(a_1, x) + d(a_2, x) = |x| + |2 - x|.$$

Then, for the sequence $Y_\mu = [1 - \mu, 1 + \mu] \subset X$ with $0 < \mu < 1$, the bounding operation for location problems yields the lower bounds

$$LB(Y_\mu) = d_1^{min}(Y_\mu) + d_2^{min}(Y_\mu) = (1 - \mu) + (1 - \mu) = 2(1 - \mu).$$

With $P(Y_\mu) = c(Y_\mu) = 1$, we obtain

$$\frac{f(P(Y_\mu)) - LB(Y_\mu)}{\delta(Y_\mu)^2} = \frac{2 - 2(1 - \mu)}{4\mu^2} = \frac{1}{2\mu}.$$

Since this expression is unbounded for $\mu \to 0$, we cannot find a fixed constant $C > 0$ such that

$$f(P(Y_\mu)) - LB(Y_\mu) \le C \cdot \delta(Y_\mu)^2.$$

6 General bounding operation

Our aim in this section is to present a general bounding operation where the lower bounds can be calculated easily. Therefore, we will first focus on scalar functions before we consider the general case.

6.1 Bounding operations for scalar functions

Consider a scalar function $f : \mathbb{R} \to \mathbb{R}$ and boxes of the type

$$Y = [c - \lambda, \, c + \lambda] \quad \text{for} \quad \lambda > 0.$$

Recall that the center $c(Y)$ of such boxes is c. Assume that $f(x)$ is twice continuously differentiable and that a bounding operation $(L(Y), c(Y))$ of $f''(x)$ with a rate of convergence of 1 is known, i.e. we assume $L(Y) \leq f''(x)$ for all $x \in Y$ with

$$f''(c) - L(Y) \leq D \cdot \delta(Y)$$

for some constant $D > 0$. Note that the lower bound $L(Y)$ depends on Y, but D does not. Defining

$$m(x) = f(c) + f'(c)(x - c) + \frac{1}{2} L(Y)(x - c)^2,$$

we know that

$$f''(x) - m''(x) = f''(x) - L(Y) \geq 0 \quad \text{for all} \quad x \in Y.$$

Thus, $f(x) - m(x)$ is convex on Y and with $f(c) - m(c) = 0$ and $f'(c) - m'(c) = 0$ the subgradient of $f - m$ at c yields

$$f(x) - m(x) \geq (f(c) - m(c)) + (f'(c) - m'(c)) \cdot (x - c) = 0.$$

Hence, we conclude that $f(x) \geq m(x)$ for all $x \in Y$. Setting $d = -f'(c)/L(Y) + c$ we define

$$LB(Y) := \begin{cases} \min\{m(c - \lambda), \, m(c + \lambda)\} & \text{if} \quad L(Y) \leq 0, \\ \min\{m(c - \lambda), \, m(c + \lambda)\} & \text{if} \quad L(Y) > 0 \text{ and } d \notin Y, \\ \min\{m(c - \lambda), \, m(c + \lambda), \, m(d)\} & \text{if} \quad L(Y) > 0 \text{ and } d \in Y, \end{cases} \tag{5}$$

and obtain $LB(Y) \leq f(x)$ for all $x \in Y$ since $m'(d) = 0$. Furthermore, we assign

$$P(Y) \in \arg \min_{x \in Y} m(x). \tag{6}$$

Since $P(Y) \in \{c - \lambda, \, c + \lambda, \, d\}$ we obtain $LB(Y) = m(P(Y))$. Together we have defined a bounding operation $(LB(Y), P(Y))$ for scalar functions.

Theorem 8 *Suppose that $X \subseteq \mathbb{R}$ is a compact set and $f : X \to \mathbb{R}$ is three times continuously differentiable. Let a bounding operation $(L(Y), c(Y))$ for $f''(x)$ with a rate of convergence of 1 be known.*

Then the bounding operation $(LB(Y), P(Y))$ defined in (5) and (6) has a rate of convergence of $p = 3$.

Proof Using the Taylor expansion of $f(x)$ at $c = c(Y)$, we obtain

$$f(x) = f(c) + f'(c)(x - c) + \frac{1}{2} f''(c)(x - c)^2 + \frac{1}{6} f^{(3)}(\vartheta(x))(x - c)^3$$

for a $\vartheta(x) \in Y$. Defining

$$\Gamma = \max\{f^{(3)}(\vartheta) : \vartheta \in X\},$$

we have

$$f(x) - m(x) = \frac{1}{2} \cdot (f''(c) - L(Y)) \cdot (x - c)^2 + \frac{1}{6} \cdot f^{(3)}(\vartheta(x)) \cdot (x - c)^3$$

$$\leq \frac{1}{2} \cdot D \cdot \delta(Y) \cdot \frac{1}{4}\delta(Y)^2 + \frac{1}{6}\Gamma \cdot \frac{1}{8}\delta(Y)^3$$

$$= \left(\frac{1}{8}D + \frac{1}{48}\Gamma\right) \cdot \delta(Y)^3$$

for all $x \in Y$. Thus, with $C = \left(\frac{1}{8}D + \frac{1}{48}\Gamma\right)$ we derived

$$f(P(Y)) - LB(Y) = f(P(Y)) - m(P(Y)) \leq C \cdot \delta(Y)^3$$

for all boxes Y as stated in the theorem. $\qquad\square$

In the remainder of this subsection we consider a more general bounding operation for scalar functions. For an odd $p \geq 3$ assume that $f(x)$ is $p - 1$ times differentiable and that a bounding operation of $f^{(p-1)}(x)$ using $P(Y) = c = c(Y)$ with a rate of convergence of 1 is known, i.e. we assume $L(Y) \leq f^{(p-1)}(x)$ for all $x \in Y$ with

$$f^{(p-1)}(c) - L(Y) \leq D \cdot \delta(Y).$$

Defining

$$m(x) = \sum_{k=0}^{p-2} \frac{1}{k!} \cdot f^{(k)}(c) \cdot (x - c)^k + \frac{1}{(p-1)!} \cdot L(Y) \cdot (x - c)^{p-1}$$

and using the Taylor expansion of $f(x)$ at c of order $p - 2$ with the Lagrange form of the remainder term, i.e.

$$f(x) = \sum_{k=0}^{p-2} \frac{1}{k!} \cdot f^{(k)}(c) \cdot (x - c)^k + \frac{1}{(p-1)!} \cdot f^{(p-1)}(\xi(x)) \cdot (x - c)^{p-1}$$

with $\xi(x) \in Y$, we obtain

$$f(x) - m(x) = \frac{1}{(p-1)!} \cdot \underbrace{\left(f^{(p-1)}(\xi(x)) - L(Y)\right)}_{\geq 0} \cdot \underbrace{(x - c)^{p-1}}_{\geq 0} \geq 0. \qquad (7)$$

Hence, $f(x) \geq m(x)$ for all $x \in Y$ and we derived the bounding operation

$$LB(Y) = \min_{x \in Y} m(x), \qquad (8)$$

$$P(Y) \in \arg\min_{x \in Y} m(x). \qquad (9)$$

We remark that for any subbox $Y \subset X$ in order to obtain $LB(Y)$ and $P(Y)$ we must minimize a polynomial $m(x)$ of degree $p - 1$.

Theorem 9 *Suppose that $X \subseteq \mathbb{R}$ is a compact set and $f : X \to \mathbb{R}$ is p times continuously differentiable for an odd $p \geq 3$. Let a bounding operation $(L(Y), c(Y))$ for $f^{(p-1)}(x)$ with a rate of convergence of 1 be known.*

Then the general bounding operation $(LB(Y), P(Y))$ defined in (8) and (9) has a rate of convergence of p.

Proof Using now the Taylor expansion of $f(x)$ at c of order $p - 1$ with the Lagrange form of the remainder term, we obtain

$$f(x) = \sum_{k=0}^{p-1} \frac{1}{k!} \cdot f^{(k)}(c) \cdot (x - c)^k + \frac{1}{p!} \cdot f^{(p)}(\vartheta(x)) \cdot (x - c)^p$$

for a $\vartheta(x) \in Y$. Defining

$$\Gamma = \max\{f^{(p)}(\vartheta) : \vartheta \in X\},$$

we have

$$f(x) - m(x) = \frac{1}{(p-1)!} \cdot (f^{(p-1)}(c) - L(Y)) \cdot (x - c)^{p-1} + \frac{1}{p!} \cdot f^{(p)}(\vartheta(x)) \cdot (x - c)^p$$

$$\leq \frac{1}{(p-1)!} \cdot D \cdot \delta(Y) \cdot \frac{1}{2^{p-1}} \cdot \delta(Y)^{p-1} + \frac{1}{p!} \Gamma \cdot \frac{1}{2^p} \cdot \delta(Y)^p$$

$$= \left(\frac{D}{(p-1)! \cdot 2^{p-1}} + \frac{\Gamma}{p! \cdot 2^p} \right) \cdot \delta(Y)^p$$

for all $x \in Y$. Thus, with $C = \left(\frac{D}{(p-1)!2^{p-1}} + \frac{\Gamma}{p!2^p} \right)$ we derived

$$f(P(Y)) - LB(Y) = f(P(Y)) - m(P(Y)) \leq C \cdot \delta(Y)^p$$

as stated in the theorem. □

We remark that a similar bounding operation for $Y = [c - \lambda, c + \lambda]$ for an even $p \geq 2$ can be found if the Taylor expansion of $f(x)$ at $\ell = c - \lambda$ is used. Otherwise we cannot guarantee that $m(x) \leq f(x)$ for all $x \in Y$, see (7).

6.2 A general bounding operation

In this subsection we derive a general bounding operation with a rate of convergence of p for a general $p \geq 2$. Therefore, for any box

$$Y = [\underline{y}_1, \overline{y}_1] \times \cdots \times [\underline{y}_n, \overline{y}_n] \subset X$$

it is necessary to minimize a polynomial $m(x) = m(x_1, \ldots, x_n)$ of degree $p - 1$.

For our calculations, define

$$\ell = \ell(Y) = (\ell_1, \ldots, \ell_n) = (\underline{y}_1, \ldots, \underline{y}_n) \in Y$$

and suppose that all partial derivatives of $f(x)$ up to order $p - 1$ exist and are continuous on X. Furthermore, we assume that some constants $L_{(k_1, \ldots, k_n)} = L_{(k_1, \ldots, k_n)}(Y)$ with

$$L_{(k_1, \ldots, k_n)} \leq \frac{\partial^{p-1} f}{\partial^{k_1} x_1 \cdots \partial^{k_n} x_n}(x) \quad \text{for all} \quad x \in Y$$

and for all $k_1, \ldots, k_n \in \mathbb{N}$ with $k_1 + \ldots + k_n = p - 1$ are known. Although these constants $L_{(k_1, \ldots, k_n)}$ depend on Y we will for the sake of simplicity only write $L_{(k_1, \ldots, k_n)}$ in the following.

Next, consider

$$m(x) = \sum_{k=0}^{p-2} \left(\sum_{\substack{k_1,\ldots,k_n \in \mathbb{N} \\ k_1+\ldots+k_n=k}} \frac{1}{\prod_{i=1}^n k_i!} \cdot \frac{\partial^k f}{\partial^{k_1} x_1 \cdots \partial^{k_n} x_n}(\ell) \cdot \prod_{i=1}^n (x_i - \ell_i)^{k_i} \right)$$

$$+ \sum_{\substack{k_1,\cdots,k_n \in \mathbb{N} \\ k_1+\ldots+k_n=p-1}} \frac{1}{\prod_{i=1}^n k_i!} \cdot L_{(k_1,\ldots,k_n)} \cdot \prod_{i=1}^n (x_i - \ell_i)^{k_i}.$$

We obtain using the Taylor expansion of $f(x)$ at ℓ of order $p - 2$ with the Lagrange form of the remainder term for a $\xi(x) \in Y$

$$f(x) - m(x)$$
$$= \sum_{\substack{k_1,\ldots,k_n \in \mathbb{N} \\ k_1+\cdots+k_n=p-1}} \frac{1}{\prod_{i=1}^n k_i!} \cdot \underbrace{\left(\frac{\partial^{p-1} f}{\partial^{k_1} x_1 \cdots \partial^{k_n} x_n}(\xi(x)) - L_{(k_1,\ldots,k_n)} \right)}_{\geq 0} \cdot \underbrace{\prod_{i=1}^n (x_i - \ell_i)^{k_i}}_{\geq 0}$$

for all $x \in Y$. Thus, we conclude that $f(x) \geq m(x)$ for all $x \in Y$ and we have the bounding operation

$$LB(Y) = \min_{x \in Y} m(x), \tag{10}$$

$$P(Y) \in \arg\min_{x \in Y} m(x). \tag{11}$$

Theorem 10 *Suppose that $f : X \to \mathbb{R}$ with a compact set $X \subseteq \mathbb{R}^n$ is p times continuously differentiable. Let bounding operations*

$$(L_{(k_1,\ldots,k_n)}(Y), \ell(Y))$$

for the functions

$$\frac{\partial^{p-1} f}{\partial^{k_1} x_1 \cdots \partial^{k_n} x_n}(x)$$

for all $k_1, \ldots, k_n \in \mathbb{N}$ with $k_1 + \cdots + k_n = p - 1$ be known and assume that all of these bounding operations have a rate of convergence of 1.

Then the general bounding operation $(LB(Y), P(Y))$ as defined in (10) and (11) has a rate of convergence of p.

Proof In the proof we are using the same principles as in our proof of Theorem 9. Since $(L_{(k_1,\ldots,k_n)}(Y), \ell(Y))$ are bounding operations with a rate of convergence of 1 there exist constants $D_{(k_1,\ldots,k_n)}$ such that

$$\frac{\partial^{p-1} f}{\partial^{k_1} x_1 \cdots \partial^{k_n} x_n}(\ell) - L_{(k_1,\ldots,k_n)} \leq D_{(k_1,\ldots,k_n)} \cdot \delta(Y)$$

for all $k_1, \ldots, k_n \in \mathbb{N}$ with $k_1 + \cdots + k_n = p - 1$. The Taylor expansion of $f(x)$ at ℓ of order $p - 1$ with the Lagrange form of the remainder yields

$$
\begin{aligned}
&f(x) - m(x) \\
&= \sum_{\substack{k_1, \ldots, k_n \in \mathbb{N} \\ k_1 + \cdots + k_n = p-1}} \frac{1}{\prod_{i=1}^{n} k_i!} \cdot \underbrace{\left(\frac{\partial^{p-1} f}{\partial^{k_1} x_1 \ldots \partial^{k_n} x_n}(\ell) - L_{(k_1, \ldots, k_n)} \right)}_{\leq D_{(k_1, \ldots, k_n)} \cdot \delta(Y)} \cdot \underbrace{\prod_{i=1}^{n} (x_i - \ell_i)^{k_i}}_{\leq \delta(Y)^{p-1}} \\
&\quad + \sum_{\substack{k_1, \ldots, k_n \in \mathbb{N} \\ k_1 + \cdots + k_n = p}} \frac{1}{\prod_{i=1}^{n} k_i!} \cdot \frac{\partial^p f}{\partial^{k_1} x_1 \ldots \partial^{k_n} x_n}(\vartheta(x)) \cdot \underbrace{\prod_{i=1}^{n} (x_i - \ell_i)^{k_i}}_{\leq \delta(Y)^p} \\
&\leq C \cdot \delta(Y)^p
\end{aligned}
$$

for suitable constants $D_{(k_1, \ldots, k_n)}$, some $\vartheta(x) \in Y$, and a suitable constant C. Hence, we have

$$
f(P(Y)) - LB(Y) = f(P(Y)) - m(P(Y)) \leq C \cdot \delta(Y)^p
$$

which proves the theorem. □

We remark that the general bounding operation requires a solution for the problem

$$
\min_{x \in Y} m(x). \tag{12}
$$

For example, if $p = 3$ then $m(x)$ is a polynomial of degree two. Therefore, the Hessian of $m(x)$ is a symmetric and real valued matrix $H \in \mathbb{R}^{n \times n}$. Hence, $m(x)$ is concave if and only if H is negative semidefinite and $m(x)$ is convex if and only if H is positive semidefinite.

If H is negative semidefinite, we only have to investigate the 2^n vertices of Y. If H is not negative semidefinite, the minimum of $m(x)$ with $x \in Y$ is attained at a point $x \in Y$ with $m'(x) = 0$ or at a point x on the boundary of Y. Note that we obtain all zeros of $m'(x)$ by solving a system of n linear equations with n variables.

Moreover, for the special case $p = 3$ and $n = 2$ we can derive a finite dominating set (FDS) of cardinality nine, i.e. a set of maximal nine points in Y which contains a solution of problem (12). One point is the solution of $m'(x) = 0$, four points are the vertices of Y, and we obtain at most four more points, one on each edge of Y.

We remark that similar calculations for the cases $p = 3$ and $n > 2$ also lead to finite dominating sets, but with a higher cardinality than nine which also increases the runtime of the algorithm.

7 Empirical rate of convergence

In this section our aim is to derive the empirical rate of convergence using three different bounding operations, namely the bounding operation for location problems, the d.c. bounding operation, and the general bounding operation of order three.

Changing the inequality into an equality in (1) and applying the natural logarithm yields

$$
\log(f(P(Y)) - LB(Y)) = \log(C) + p \cdot \log(\delta(Y)).
$$

For a given test function $f(x)$, we can calculate $\log(f(P(Y)) - LB(Y))$ and $\log(\delta(Y))$ for some randomly selected boxes $Y \subset X$. We then obtain p and $\log(C)$ by linear regression.

This strategy to obtain the empirical rate of convergence was also used in Tóth et al. (2007) and Tóth and Csendes (2005) to analyze the related concept of the convergence order for some inclusion function in interval analysis.

As test problem we chose the Weber problem with some negative weights on the plane using Euclidean distances. For a given set of s demand points $a_1, \ldots, a_s \in \mathbb{R}^2$ and weights $w_1, \ldots, w_s \in \mathbb{R}$, the objective function is given by

$$f(x) = \sum_{k=1}^{s} w_k \cdot d_2(a_k, x) = \sum_{k=1}^{s} w_k \cdot \sqrt{(a_{k,1} - x_1)^2 + (a_{k,2} - x_2)^2}.$$

Large scale instances for this global optimization problem were solved in Tuy et al. (1995) using outer approximation methods in d.c. programming and in Drezner and Suzuki (2004) using the big triangle small triangle method.

We now apply the three bounding operations to the Weber problem. To this end, let a box

$$Y = [\underline{y}_1, \overline{y}_1] \times [\underline{y}_2, \overline{y}_2] \subset \mathbb{R}^2$$

be given.

Using the bounding operation for location problems (see Sect. 5.3), we directly obtain the lower bound

$$LB(Y) = \sum_{\substack{k=1 \\ w_k > 0}}^{s} w_k \cdot d_k^{min}(Y) + \sum_{\substack{k=1 \\ w_k < 0}}^{s} w_k \cdot d_k^{max}(Y).$$

For the d.c. bounding operation (see Sect. 5.1), we consider the d.c. decomposition

$$f(x) = \sum_{\substack{k=1 \\ w_k > 0}}^{s} w_k \cdot d_2(a_k, x) - \sum_{\substack{k=1^s \\ w_k < 0}}^{s} (-w_k) \cdot d_2(a_k, x).$$

Note that for this example problem the d.c.m. bounding operation yields exactly the same bounds as the d.c. bounding operation. This fact is, of course, in general not the case, see Blanquero and Carrizosa (2009).

As third bounding operation we applied the general bounding operation defined in (10) and (11) in Sect. 6.2. Since $f(x)$ is not differentiable at the demand points a_k for $k = 1, \ldots, s$, we have to bound $f(x)$ from below by a differentiable function in order to apply the general bounding operation. To this end, consider

$$\tilde{f}(x) = \sum_{\substack{k=1 \\ a_k \notin Y}}^{s} w_k \cdot d_2(a_k, x) + \sum_{\substack{k=1 \\ a_k \in Y, w_k < 0}}^{s} w_k \cdot d_k^{max}(Y)$$

with $\tilde{f}(x) \leq f(x)$ for all $x \in Y$. Notice that $\tilde{f}(x)$ is differentiable on Y and we are calculating a lower bound of $\tilde{f}(x)$ using the general bounding operation of order three. The second partial derivatives of $\tilde{f}(x)$ are bounded from below using methods from interval analysis, namely the natural interval extension, see Hansen (1992) or Ratschek and Rokne (1988) for more details on this subject. Hence, we easily obtain the necessary constants $L_{(0,2)}$, $L_{(1,1)}$, and $L_{(2,0)}$.

In our computational studies, we generated $s = 100$ demand points uniformly distributed in $X = [0, 10] \times [0, 10]$ and weights uniformly distributed in $[-4, 6]$. We randomly selected 40 boxes $Y \subset X$ with different widths for each of five sets of input data. Figure 1
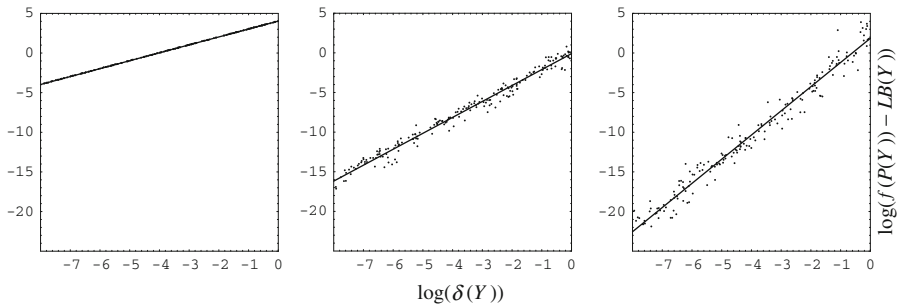
Fig. 1 Empirical results for the location bounding operation (*left*), the d.c. bounding operation (*middle*), and the general bounding operation of order three (*right*)

Table 3 Results for the empirical rate of convergence

Bounding operation	C	p
Location bounding operation	56.26	1.00
D.c. bounding operation	0.90	2.01
General bounding operation	6.36	3.05

shows the results for the three bounding operations. Note that we ignored boxes Y with $f(P(Y)) - LB(Y) = 0$ from our further considerations since $\log(0)$ does not exist. Linear regressions of our results yield the values of C and p summarized in Table 3.

As can be seen in Table 3, the empirical rates of convergence agree fairly well with our theoretical results. The bounds derived from the bounding operation for location problems with a theoretical and empirical rate of convergence of $p = 1$ are the worst ones for all 200 randomly selected boxes. Also the constant $C = 56.26$ is much greater than $C = 0.90$ and $C = 6.36$ for the d.c. bounding operation and the general bounding operation, respectively.

Since in our example the rate of convergence for the general bounding operation is approximately of one order higher than for the d.c. bounding operation but the constant C is also higher, we suggest to use the d.c. bounding operation for larger boxes and the general bounding operation for smaller boxes. The linear regressions show that for boxes Y with $\delta(Y) \geq \exp(-2.30) \approx 0.10$ the d.c. bounding operation yields on average sharper bounds than the general bounding operation. We hence suggest for the Weber problem with some negative weights to combine the d.c. bounding operation $(LB_{dc}(Y), P_{dc}(Y))$ with the general bounding operation $(LB_{gen}(Y), P_{gen}(Y))$ to obtain the ***combined bounding operation***

$$\left(\max\{LB_{dc}(Y), LB_{gen}(Y)\}, \arg\min\{f(P_{dc}(Y)), f(P_{gen}(Y))\}\right).$$

In a second computational study, we compared the d.c. bounding operation and the general bounding operation more precisely. We again chose demand points uniformly distributed in $X = [0, 10] \times [0, 10]$ and weights uniformly distributed in $[-4, 6]$. For each set of input data, the big cube small cube algorithm was run three times: A first run with the d.c. bounding operation, a second run with the general bounding operation of order three, and a third run with the combined bounding operation.

Table 4 shows the average number of iterations for 50 runs for various numbers of demand points. In all cases we used the initial box $X = [2, 8] \times [2, 8]$ and an absolute accuracy of $\varepsilon = 10^{-12}$.

Table 4 Comparison for the average number of iterations using the d.c. bounding operation, the general bounding operation of order three, and the combined bounding operation

Bounding operation	$s = 50$	$s = 100$	$s = 500$	$s = 1,000$
D.c. bounding operation	96.3	114.6	122.1	122.9
General bounding operation	56.7	55.8	78.9	105.7
Combined bounding operation	32.7	35.2	38.3	40.1

In all cases, the average number of iterations increases with the number s of demand points, but much faster for the general bounding operation than for the d.c. bounding operation. Moreover, we remark that the number of iterations for the d.c. bounding operation can be reduced by two-third if we combine both bounding operations.

The overall runtimes for the presented algorithm depend of course not only on the number of iterations, but also on the computational complexity of the selected bounding operation. In our example, we obtained the shortest runtimes for the d.c. bounding operation and the highest runtimes for the general bounding operation.

8 Conclusions

In this paper, our aim was to introduce a convergence theory with a new definition of the rate of convergence for geometric branch-and-bound methods. Note that a related concept but only for the specific use of interval analysis methods was introduced in Csallner and Csendes (1996).

After the general convergence theory, we analyzed the rate of convergence for several bounding operations from the theoretical point of view. Furthermore, we showed how to construct a general bounding operation having any given rate of convergence which is based on the Taylor expansion of the objective function. Apart from the theoretical analysis, we derived the empirical rate of convergence and we showed that the empirical rates of convergence agree fairly well with our theoretical results.

Moreover, our computational studies showed that the general bounding operation of order three requires less iterations to solve the Weber problem with some negative weights up to an absolute accuracy of $\varepsilon = 10^{-12}$ than the d.c. bounding operation. This allowed to observe that the number of iterations can be reduced by two-third if we are using the maximum of these two lower bounds in every iteration.

It should be mentioned that most of the presented results and bounding operations are also valid for more general branch-and-bound algorithms. For example, it is sufficient to assume a more general feasible area which is given by a list of boxes and the technique also works for triangles or simplices instead of boxes.

For some further research, an extensive comparison of the empirical rate of convergence and the number of iterations is an interesting task. In particular, if a d.c. decomposition of the objective function is hard to derive, the big cube small cube method might be much faster if we are using the general bounding operation. Moreover, a further point of interest is to analyze the complexity for the calculation of the lower bounds. Taking this into account, the overall runtime is the complexity for the calculation of the lower bounds times the number of iterations throughout the branch-and-bound algorithm. Furthermore, we will compare the benefit of some further discarding tests for the different bounding operations and we will investigate the meaning of bounding operations for discrete branch-and-bound procedures.

References

Blanquero, R., Carrizosa, E.: Continuous location problems and big triangle small triangle: constructing better bounds. J. Glob. Optim. **45**, 389–402 (2009)

Csallner, A.E., Csendes, T.: The convergence speed of interval methods for global optimization. Comput. Math. Appl. **31**, 173–178 (1996)

Drezner, Z.: A general global optimization approach for solving location problems in the plane. J. Glob. Optim. **37**, 305–319 (2007)

Drezner, T., Drezner, Z.: Finding the optimal solution to the Huff based competitive location model. Comput. Manag. Sci. **1**, 193–208 (2004)

Drezner, Z., Suzuki, A.: The big triangle small triangle method for the solution of nonconvex facility location problems. Oper. Res. **52**, 128–135 (2004)

Fernández, J., Pelegrín, B., Plastria, F., Tóth, B.: Solving a huff-like competitive location and design model for profit maximization in the plane. Eur. J. Oper. Res. **179**, 1274–1287 (2007)

Hansen, E.: Global Optimization Using Interval Analysis, 1st edn. Marcel Dekker, New York (1992)

Hansen, P., Peeters, D., Richard, D., Thisse, J.F.: The minisum and minimax location problems revisited. Oper. Res. **33**, 1251–1265 (1985)

Hiriart-Urruty, J.B., Lemaréchal, C.: Fundamentals of Convex Analysis, 1st edn. Springer, Berlin (2004)

Horst, R., Thoai, N.V.: DC programming: overview. J. Optim. Theory Appl. **103**, 1–43 (1999)

Horst, R., Tuy, H.: Global Optimization: Deterministic Approaches, 3rd edn. Springer, Berlin (1996)

Horst, R., Pardalos, P.M., Thoai, N.V.: Introduction to Global Optimization, 2nd edn. Springer, Berlin (2000)

Pardalos, P.M., Resende, M.G.C.: Handbook of Applied Optimization, 1st edn. Oxford University Press, Oxford (2002)

Plastria, F.: GBSSS: The generalized big square small square method for planar single-facility location. Eur. J. Oper. Res. **62**, 163–174 (1992)

Ratschek, H., Rokne, J.: New Computer Methods for Global Optimization, 1st edn. Ellis Horwood, Chichester (1988)

Rockafellar, R.T.: Convex Analysis, 1st edn. Princeton University Press, Princeton (1970)

Schöbel, A., Scholz, D.: The big cube small cube solution method for multidimensional facility location problems. Comput. Oper. Res. **37**, 115–122 (2010)

Tóth, B., Csendes, T.: Empirical investigation of the convergence speed of inclusion functions in a global optimization context. Reliab. Comput. **11**, 253–273 (2005)

Tóth, B., Fernández, J., Csendes, T.: Empirical convergence speed of inclusion functions for facility location problems. J. Comput. Appl. Math. **199**, 384–389 (2007)

Tóth, B., Fernández, J., Pelegrín, B., Plastria, F.: Sequential versus simultaneous approach in the location and design of two new facilities using planar huff-like models. Comput. Oper. Res. **36**, 1393–1405 (2009)

Tuy, H.: A general D.C. approach to location problems. In: Floudas, C.A., Pardalos, P.M. (eds.) State of the Art in Gloabal Optimization: Computational Methods and Applications, pp. 413–432. Kluwer, Dordrecht (1996)

Tuy, H., Horst, R.: Convergence and restart in branch-and-bound algorithms for global optimization. Application to concave minimization and d.c. optimization problems. Math. Program. **41**, 161–183 (1988)

Tuy, H., Al-Khayyal, F., Zhou, F.: A D.C. optimization method for single facility location problems. J. Glob. Optim. **7**, 209–227 (1995)