Global optimization of bounded factorable functions with discontinuities

Citation: Wechsung, Achim, and Paul I. Barton. “Global Optimization of Bounded Factorable Functions with Discontinuities.” Journal of Global Optimization 58.1 (2014): 1–30.

Published Version: http://dx.doi.org/10.1007/s10898-013-0060-3

Publisher: Springer US

Permanent Link: http://hdl.handle.net/1721.1/103613

Version: Author's final manuscript: final author’s manuscript post peer review, without publisher’s formatting or copy editing

Journal of Global Optimization manuscript No.
(will be inserted by the editor)

Global optimization of bounded factorable functions with discontinuities

Achim Wechsung · **Paul I. Barton**

Received: date / Accepted: date

Abstract A deterministic global optimization method is developed for a class of discontinuous functions. McCormick's method to obtain relaxations of nonconvex functions is extended to discontinuous factorable functions by representing a discontinuity with a step function. The properties of the relaxations are analyzed in detail; in particular, convergence of the relaxations to the function is established given some assumptions on the bounds derived from interval arithmetic. The obtained convex relaxations are used in a branch-and-bound scheme to formulate lower bounding problems. Furthermore, convergence of the branch-and-bound algorithm for discontinuous functions is analyzed and assumptions are

A. Wechsung

Process Systems Engineering Laboratory, Department of Chemical Engineering,

Massachusetts Institute of Technology, 77 Massachusetts Ave 66-363, Cambridge, MA 02139

E-mail: awechsun@mit.edu

P. I. Barton

Process Systems Engineering Laboratory, Department of Chemical Engineering,

Massachusetts Institute of Technology, 77 Massachusetts Ave 66-464, Cambridge, MA 02139

Tel.: +1 (617) 253-6526, Fax: +1 (617) 258-5042, E-mail: pib@mit.edu

derived to guarantee convergence. A key advantage of the proposed method over reformulating the discontinuous problem as a MINLP or MPEC is avoiding the increase in problem size that slows global optimization. Several numerical examples for the global optimization of functions with discontinuities are presented, including ones taken from process design and equipment sizing as well as discrete-time hybrid systems.

Keywords Global optimization · Discontinuous functions · Convex relaxations · McCormick relaxations · Nonconvex optimization

Mathematics Subject Classification (2000) 90C26 · 49M20 · 65K05

1 Introduction

Deterministic global optimization methods aim to identify a guaranteed global optimal solution of a nonconvex program of the form

$$\begin{aligned} f^* = \min_{\mathbf{x}} \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in D, \end{aligned} \tag{P}$$

where $f : X \subset \mathbb{R}^n \to \mathbb{R}$ and $D \subset X$ is nonempty compact. One solution method for the nonconvex program (P) is the branch-and-bound algorithm, which constructs a convergent sequence of lower and upper bounds on f^* to find an optimal solution [15, 22, 35]. Typically, the function f is assumed to be continuous, frequently f is also required to be sufficiently smooth. Several methods have been proposed in the literature to find convex relaxations of continuous functions that are given in closed form [1, 26, 38]. Recently, McCormick's method [26], which was originally defined for explicitly stated factorable functions, has been extended to continuous functions that are not known explicitly, e.g., when they result from algorithms [28, 37]. Discontinuities can appear, for example, in algorithms with conditional statements (i.e., `IF-THEN-ELSE`), which have been excluded in a previous paper [28, p. 593]. Hereafter, a class of *discontinuous* functions is considered. Neither are the standard relaxation techniques, which are defined for continuous functions, applicable in this case nor has it been considered in branch-and-bound theory [22].

A closer look at McCormick's composition theorem and its proof [26], however, indicates that the result can be extended to discontinuous factorable functions if they are bounded. It is, in addition, necessary to know valid relaxations of univariate discontinuous functions. Since a discontinuity can be represented by a step function [41], for which relaxations can be constructed easily, this requirement can be met when the factorable representation of the function has a finite number of discontinuous univariate factors. However, it is not clear if the properties of McCormick relaxations shown in [37] hold true. In this paper, the obtained relaxations for bounded factorable functions with discontinuities are analyzed in detail. Such analysis is indispensable in order to establish properties of the proposed relaxation technique required for its use in a branch-and-bound method [22]. Furthermore, branch-and-bound theory must be extended as continuity is a standing assumption throughout [22].

The proposed method is particularly well suited to solve optimization problems with discontinuities depending on continuous variables. Exemplary of this case are discontinuous cost functions in process design: when a certain size is exceeded, two units need to be used instead of one, causing a discontinuity in the investment cost (this starkly contrasts with discrete decisions that require integer variables, e.g., when two exclusive alternatives for one unit exist). Examples for such problems can be found in process synthesis with discontinuous investment costs [39] as well as dynamic optimization problems with discontinuities [2], in particular, hybrid systems [24]. Currently, mixed-integer or complementarity constraint formulations are often used to model discontinuities depending on continuous variables [6, 39]. In the former, binary variables are introduced to model discontinuities whereas in the latter complementarity constraints take on this role. Commercial global optimization algorithms are available for MINLPs [36], however, introducing binary variable to model the discontinuities can increase the number of variables drastically. This can lead to poor performance as branch-and-bound algorithms are known to scale worst-case exponentially with the number of variables. MPECs are usually reformulated as NLPs and are only solved locally at present [6, 16].

While this paper focuses on *global* optimization of discontinuous factorable functions, it should be noted that existing work on discontinuous optimization considered finding *locally* optimal solutions. Aside from using the definition of a local minimum as a point attaining the smallest value of the objective function in a neighborhood, no other characterization, e.g., using gradients, is applicable when the function is not even continuous. In the quest to derive local optimality conditions, the notion of derivatives is generalized to nonsmooth and discontinuous functions by several authors [3, 4, 9, 14, 30, 42]. Recently, Rockafellar generalized derivatives have been used in direct search algorithms for discontinuous functions [40]. Another prevalent idea in the literature is to approximate the discontinuous function by convolving it with an appropriate mollifier resulting in an averaged function. This operation leads to an integration problem, possibly of high dimension, which is computationally expensive and is often evaluated using Monte Carlo schemes [5, 13, 33, 42]. Conn and Mongeau [10] consider piecewise linear optimization problems where the objective function and constraints have discontinuities on a set of hyperplanes and propose an algorithm to identify local minima. In an approach more closely related to the idea proposed in this paper, Zang [41] introduces step functions to express the discontinuities and suggests a family of smooth approximations for these. Similar ideas are used to smooth continuous functions at points of non-differentiability [e.g., 12] and are prone to introduce inaccuracy and numerical instability.

The remainder of this paper is organized as follows. First, McCormick relaxations [26] are studied taking advantage of the formalization provided by [37]. In Section 2, properties such as continuity and convexity of the obtained relaxations of certain bounded functions are proved. Examples of the relaxations of discontinuous functions are provided. Furthermore, the behavior of the relaxations on sequences of intervals is investigated, after the necessary assumptions required for these properties are made explicit. This leads up to the results in Section 3, where it is argued that a branch-and-bound algorithm with finite ε-convergence can be constructed. The paper continues with some examples in Section 4 which showcase the numerical feasibility and provide first promising examples from different applications. The paper is concluded with a summary of the obtained results in Section 5.

Notation The set of closed and bounded real *intervals* is denoted by $\mathbb{IR}$ and the set of n-dimensional intervals (Cartesian products of n intervals) by $\mathbb{IR}^n$. The *lower* and *upper bounds* of $X \in \mathbb{IR}^n$ are denoted as $\underline{\mathbf{x}}$ and $\overline{\mathbf{x}}$, respectively. The set of all interval subsets of $X \subset \mathbb{R}^n$ is denoted by $IX \subset \mathbb{IR}^n$. Convergence of a sequence of intervals $\{X^l\}$, $X^l \in \mathbb{IR}^n$ to $X^* \in \mathbb{IR}^n$ as $l \to \infty$ will be denoted as $X^l \to X^*$ and is equivalent to $\lim_{l\to\infty} d_H(X^l, X^*) = 0$ where $d_H(X,Y)$ denotes the Hausdorff metric for any $X, Y \in \mathbb{IR}^n$. A *convex relaxation* of a function $f : X \to \mathbb{R}$ on a convex set $X \subset \mathbb{R}^n$ is denoted by $\check{f}$, a *concave relaxation* by $\hat{f}$.

2 Relaxations of bounded factorable functions

In this section, the construction of relaxations for factorable and bounded functions is discussed. To this end, the procedure to obtain McCormick relaxations [26, 37] is defined and well-known results for these are extended. Then, discontinuous univariate intrinsic functions are studied more closely and first examples of the constructed relaxations are given. A discussion about the behavior of the relaxations on sequences of intervals is preceded by a collection of necessary assumptions and concludes this section.

The idea of a factorable function is central to this paper and will be formalized below, cf. [26, 37]. Here, the notion of factorable functions will be extended by only requiring boundedness, but not continuity. Such functions will be called *bounded factorable*. The class of bounded factorable functions includes most functions that can be represented finitely on a computer.

Definition 1 A function $\varphi : B \to \mathbb{R}$, $B \subset \mathbb{R}$ is a *univariate intrinsic function* if, for any interval $V \in IB$, an inclusion monotonic interval extension Φ of φ on V, a convex relaxation $\check{\varphi}$ of φ on V and a concave relaxation $\hat{\varphi}$ of φ on V are known and can be evaluated computationally.

Definition 2 Suppose $X \subset \mathbb{R}^n$. A function $f : X \to \mathbb{R}$ is *factorable* if it can be expressed in terms of a finite sequence of factors $v_1, \ldots, v_m$ such that, given $\mathbf{x} \in X$, $v_i = x_i$ for $i = 1, \ldots, n$, and v_k is defined for each k, $n < k \leq m$, as either

(a) $v_k = v_i + v_j$, $i, j < k$, or

(b) $v_k = v_i v_j$, $i, j < k$, or

(c) $v_k = \varphi_k(v_i)$, $i < k$, where $\varphi_k : B_k \to \mathbb{R}$ is any univariate intrinsic function,

and $f(\mathbf{x}) = v_m(\mathbf{x})$.

Note that Definition 2 implies that $v_i \in B_k$ for each k such that $n < k \leq m$ and v_k is defined by Definition 2 (c).

In the literature, a standing assumption is continuity of the univariate functions φ_k and hence f. When X is compact, continuity of each operation in Definition 2 guarantees compactness of $f(X)$ [34]. Hence, continuous factorable functions are always bounded factorable on a compact set X. As shown below, if each univariate function is bounded, then the constructed function is bounded factorable.

Lemma 1 *Suppose $X \subset \mathbb{R}^n$ is bounded. Consider a factorable function $f : X \to \mathbb{R}$. f is bounded factorable if φ_k is bounded on B_k for each k such that $n < k \leq m$ and v_k is defined by Definition 2 (c).*

Proof For $1 \leq k \leq n$, the assertion holds trivially. Suppose the assertion holds for some k where $n < k \leq m$. When v_k is defined by Definition 2 (a) or (b), v_k is certainly bounded. When v_k is defined by Definition 2 (c), v_k is bounded since φ_k is bounded. From finite induction, it follows that v_m is bounded and, hence, f is bounded factorable. □

2.1 Extension of McCormick's result to bounded factorable functions

McCormick [26] presented a recursive procedure to create relaxations of factorable functions f on the interval $X = [\underline{\mathbf{x}}, \overline{\mathbf{x}}]$. While in his exposition, McCormick restricted the result to continuous factorable functions, it can be easily extended to bounded factorable functions.

Theorem 1 *Let $X \subset \mathbb{R}^n$ be a nonempty convex set. Consider the composite function $f_2 \circ f_1$ where $f_1 : X \to \mathbb{R}$ is bounded on X, let $f_1(X) \subset [a,b]$ and $f_2 : [a,b] \to \mathbb{R}$. Suppose that relaxations $\breve{f}_1$ and $\hat{f}_1$ of f_1 on X as well as relaxations $\breve{f}_2$ and $\hat{f}_2$ of f_2 on $[a,b]$ are known. Let $z_{\min}$ be a point at which $\breve{f}_2$ attains its infimum on $[a,b]$, and let $z_{\max}$ be a point at which $\hat{f}_2$ attains its supremum on $[a,b]$. Then*

$$u(\mathbf{x}) = \breve{f}_2[\mathrm{mid}\{\breve{f}_1(\mathbf{x}), \hat{f}_1(\mathbf{x}), z_{\min}\}]$$

is a convex relaxation of $f_2 \circ f_1$ on X, and

$$o(\mathbf{x}) = \hat{f}_2[\mathrm{mid}\{\breve{f}_1(\mathbf{x}), \hat{f}_1(\mathbf{x}), z_{\max}\}]$$

is a concave relaxation of $f_2 \circ f_1$ *on* X*, where* $\text{mid} : \mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ *selects the middle value of the three scalar arguments.*

Proof The original proof [27] remains valid after the continuity hypothesis on f_1 is replaced with a boundedness assumption as only $f_1(X) \subset [a, b]$ is needed. $\square$

Theorem 1 allows the construction of relaxations of complicated functions by decomposing the function into factors for which relaxations are known. A precise definition of this procedure was first given in [37] and is reproduced below for the benefit of the reader.

Definition 3 Denote the *McCormick relaxations* of the bounded factorable function f on an interval $X \in \mathbb{IR}^n$ by the functions $\underset{\smile}{f}, \hat{f} : X \to \mathbb{R}$ where for each $\mathbf{x} \in X$, $\underset{\smile}{f}(\mathbf{x})$ and $\hat{f}(\mathbf{x})$ are defined by the following procedure:

1. Set $\underline{v}_i = \underline{x}_i$ and $\bar{v}_i = \bar{x}_i$ for all $i = 1, \ldots, n$, and denote $V_i \equiv [\underline{v}_i, \bar{v}_i]$.
2. Set $\underset{\smile}{v}_i = \hat{v}_i = x_i$ for all $i = 1, \ldots, n$.
3. Set $k = n + 1$.
4. Compute $\underline{v}_k$, $\bar{v}_k$ and $V_k \equiv [\underline{v}_k, \bar{v}_k]$ according to the definition of v_k, as
   (a) $\underline{v}_k = \underline{v}_i + \underline{v}_j$ and $\bar{v}_k = \bar{v}_i + \bar{v}_j$,
   (b) $\underline{v}_k = \min\left(\underline{v}_i\underline{v}_j, \underline{v}_i\bar{v}_j, \bar{v}_i\underline{v}_j, \bar{v}_i\bar{v}_j\right)$ and $\bar{v}_k = \max\left(\underline{v}_i\underline{v}_j, \underline{v}_i\bar{v}_j, \bar{v}_i\underline{v}_j, \bar{v}_i\bar{v}_j\right)$,
   (c) $\underline{v}_k = \underline{\Phi}_k(V_i)$ and $\bar{v}_k = \overline{\Phi}_k(V_i)$, where $\Phi_k : \mathbb{I}B_k \to \mathbb{IR}$ is an inclusion monotonic interval extension of φ_k on B_k.
5. Compute $\underset{\smile}{\eta}_k$ and $\hat{\eta}_k$ according to the definition of v_k, as
   (a) $\underset{\smile}{\eta}_k = \underset{\smile}{v}_i + \underset{\smile}{v}_j$ and $\hat{\eta}_k = \hat{v}_i + \hat{v}_j$,
   (b) $\underset{\smile}{\eta}_k = \max(\alpha_i + \alpha_j - \underline{v}_i\underline{v}_j, \beta_i + \beta_j - \bar{v}_i\bar{v}_j)$ and $\hat{\eta}_k = \max(\gamma_i + \gamma_j - \underline{v}_i\bar{v}_j, \delta_i + \delta_j - \bar{v}_i\underline{v}_j)$, where

   $\alpha_i = \min(\underline{v}_i\underset{\smile}{v}_j, \underline{v}_i\hat{v}_j)$, $\alpha_j = \min(\underline{v}_j\underset{\smile}{v}_i, \underline{v}_j\hat{v}_i)$, $\beta_i = \min(\bar{v}_i\underset{\smile}{v}_j, \bar{v}_i\hat{v}_j)$, $\beta_j = \min(\bar{v}_j\underset{\smile}{v}_i, \bar{v}_j\hat{v}_i)$,

   $\gamma_i = \max(\underline{v}_i\underset{\smile}{v}_j, \underline{v}_i\hat{v}_j)$, $\gamma_j = \max(\bar{v}_j\underset{\smile}{v}_i, \bar{v}_j\hat{v}_i)$, $\delta_i = \max(\bar{v}_i\underset{\smile}{v}_j, \bar{v}_i\hat{v}_j)$, $\delta_j = \max(\underline{v}_j\underset{\smile}{v}_i, \underline{v}_j\hat{v}_i)$,

(c) $\underset{\smile}{\eta}_k = \underset{\smile}{\varphi}_k(\mathrm{mid}(\underset{\smile}{v}_i, \hat{v}_i, z_k^{\min}))$ and $\hat{\eta}_k = \hat{\varphi}_k(\mathrm{mid}(\underset{\smile}{v}_i, \hat{v}_i, z_k^{\max}))$, where $\underset{\smile}{\varphi}_k, \hat{\varphi}_k : V_i \to \mathbb{R}$ are, respectively, convex and concave relaxations of φ_k on V_i, and where $z_k^{\min}$ is a minimum of $\underset{\smile}{\varphi}_k$ on V_i and $z_k^{\max}$ is a maximum of $\hat{\varphi}_k$ on V_i.

6. Compute $\underset{\smile}{v}_k$ and $\hat{v}_k$ as $\underset{\smile}{v}_k = \max(\underset{\smile}{\eta}_k, \underline{v}_k)$ and $\hat{v}_k = \min(\hat{\eta}_k, \overline{v}_k)$.
7. If $k = m$, go to 8. Otherwise, set $k = k+1$ and go to 4.
8. Set $\underset{\smile}{f}(\mathbf{x}) = \underset{\smile}{v}_m(\mathbf{x})$ and $\hat{f}(\mathbf{x}) = \hat{v}_m(\mathbf{x})$.

Scott et al. also introduced the notions of step and cumulative mappings, which are very helpful in analyzing the constructed functions [37].

Definition 4 Given an interval X, and hence intervals $V_1, \ldots, V_m$, let the *step mapping* be a mapping of the form $v_k : V_i(\times V_j) \to \mathbb{R}$ defined by the expressions given in Definition 2. Let the *cumulative mapping* v_k be the mapping $v_k : X \to \mathbb{R}$, defined for each $\mathbf{x} \in X$ by the value $v_k(\mathbf{x})$ when f is computed at $\mathbf{x}$. Similarly, let the step and cumulative mappings $\underset{\smile}{v}_k, \hat{v}_k$ be mappings of the form $\underset{\smile}{v}_k, \hat{v}_k : V_i(\times V_j) \to \mathbb{R}$ and $\underset{\smile}{v}_k, \hat{v}_k : X \to \mathbb{R}$, respectively, defined in analogous manner by Definition 3.

The following assumption ensures that the domain of each step mapping is a superset of the image of the preceding step mapping. It is discussed in more detail in [37].

Assumption 1 f can be represented on X by a factorization with the property that, for each k such that $n < k \leq m$ and v_k defined by Definition 2 (c), $V_i \subset B_k$ where V_i denotes an interval bound on v_i derived from interval arithmetic beginning with $V_i \in IX_i$, $i = 1, \ldots, n$.

Proposition 2 in [37] remains valid for bounded factorable functions since Theorem 1 shows that the constructed mappings $\underset{\smile}{v}_k, \hat{v}_k$ of the cumulative mapping v_k are indeed valid convex and concave relaxations. Thus, the following result follows.

Theorem 2 *Suppose that Assumption 1 holds. Then, McCormick relaxations of a bounded factorable function* $f : X \to \mathbb{R}$ *are valid convex and concave relaxations.*

Furthermore, Theorem 3 in [37] still holds.

Theorem 3 *Suppose that the convex and concave relaxations of each step mapping v_k are continuous on V_i for each $n < k \leq m$ where v_k is defined by a univariate intrinsic function. Then, $\underset{\smile}{f}(\mathbf{x})$ and $\hat{f}(\mathbf{x})$ are continuous on X.*

2.2 Univariate piecewise continuous functions

From Definition 2, it is apparent that discontinuities of a bounded factorable function must stem from discontinuities in some of the univariate intrinsic functions. When there is only a finite number of discontinuities, these functions can be reduced to products with a generic step function, which incorporates the discontinuity, and continuous factors.

Suppose $\varphi : X \to \mathbb{R}$ is of the form

$$\varphi(x) = \begin{cases} \varphi_1(x) \text{ if } x \in [\underline{x}_1, \overline{x}_1], \\ \varphi_2(x) \text{ if } x \in (\underline{x}_2, \overline{x}_2], \end{cases}$$

where $X, X_1, X_2 \in \mathbb{IR}$, $[\underline{x}_1, \overline{x}_1] = X_1$, $(\underline{x}_2, \overline{x}_2] \subset X_2$, $\varphi_1 : X_1 \to \mathbb{R}$, $\varphi_2 : X_2 \to \mathbb{R}$, $X = X_1 \cup X_2$, and $\overline{x}_1 = \underline{x}_2$ and let φ_1, φ_2 be continuous on their respective domains. Denote the step function as $\pi : \mathbb{R} \to \mathbb{R}$, i.e.,

$$\pi(x) = \begin{cases} 0 \text{ if } x \leq 0, \\ 1 \text{ otherwise.} \end{cases}$$

Then $\varphi(x)$ can be represented by

$$\varphi(x) = \pi(x - \overline{x}_1)\varphi_2(x) + [1 - \pi(x - \overline{x}_1)]\varphi_1(x). \tag{1}$$

As a result, it is sufficient to analyze only $\pi(x)$ in detail. The following result summarizes the information relevant for the construction of McCormick relaxations.

Theorem 4 *Consider $\pi : X \to \mathbb{R}$ on the interval $X = [\underline{x}, \overline{x}]$ as defined above. An inclusion monotonic interval extension Π of π on X is given by*

$$\Pi(X) = \begin{cases} [0,0] \text{ if } \overline{x} \leq 0, \\ [1,1] \text{ if } \underline{x} > 0, \\ [0,1] \text{ otherwise.} \end{cases}$$

Furthermore, convex relaxations are given by

$$\underset{\smile}{\pi}(x) = \begin{cases} 0 \quad \text{if } \overline{x} \leq 0 \vee (\overline{x} > 0 \wedge x \leq 0), \\ 1 \quad \text{if } \underline{x} > 0, \\ x/\overline{x} \text{ otherwise,} \end{cases}$$

and concave relaxations are given by

$$\hat{\pi}(x) = \begin{cases} 0 \qquad \text{if } \overline{x} \leq 0, \\ 1 \qquad \text{if } \underline{x} > 0 \vee (\underline{x} \leq 0 \wedge x \geq 0), \\ 1 - x/\underline{x} \text{ otherwise.} \end{cases}$$

Proof It is easy to check the validity of the bounds and the inclusion monotonicity property. Similarly, the relaxations are easy to check when $0 \notin X$. If $0 \in X$, consider the convex hull of the epigraph of π which yields the convex underestimator given in the result. Similarly, the concave overestimator is given by the convex hull of the hypograph. □

Remark 1 Strictly, φ_1 and φ_2 are defined on X_1 and X_2 only. When one defines $\varphi_i(x) = +\infty$ for $x \notin X_i$ and $0 \cdot +\infty = 0$, the above statement also holds. Furthermore, when φ_1 is defined on X, an alternative to (1) is

$$\varphi(x) = \pi(x - \overline{x}_1)[\varphi_2(x) - \varphi_1(x)] + \varphi_1(x).$$

π is a univariate *intrinsic* function because its inclusion monotonic interval extension Π and its relaxations $\underset{\smile}{\pi}$ and $\hat{\pi}$ are known, cf. Definition 1. Thus, univariate intrinsic discontinuous functions can be incorporated into the McCormick framework discussed in Section 2.1.

2.3 Examples of constructed relaxations

Next, it is demonstrated how more complicated functions with discontinuities can be expressed using the previously introduced function π and the thus computed relaxations are showcased. Example 1 shows how to model a function with multiple discontinuities, including a point where the function attains neither its lower nor its upper limit. Example 2 demonstrates that the discontinuity can depend on a factorable function of the variables. In each case, the calculations are implemented using `libMC` [7, 28] enhanced with functionality for π.

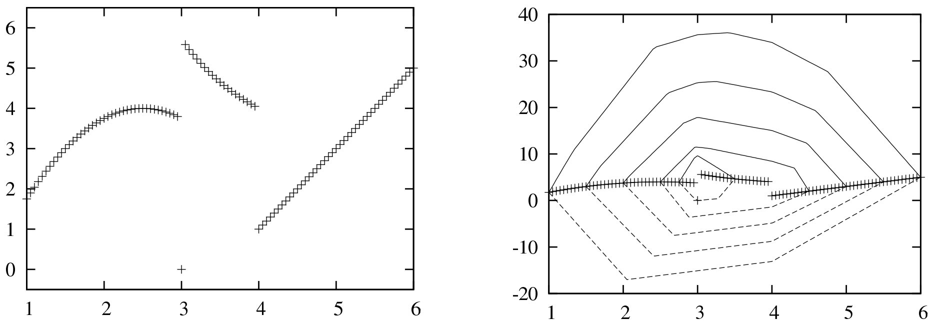
Example 1 Consider the lower semi-continuous function $f_1 : [1,6] \to \mathbb{R}$ with

$$f_1(x) = \begin{cases} -(x-2.5)^2+4 & \text{if } x \in [1,3), \\ 0 & \text{if } x = 3, \\ e^{4-x}+3 & \text{if } x \in (3,4), \\ 2x-7 & \text{if } x \in [4,6]. \end{cases}$$

It can be represented as

$$\begin{aligned} f_1(x) = &\pi(4-x)\left\{\pi(x-3)\left[e^{4-x}+3-\left\{\pi(3-x)(-(x-2.5)^2+4-0)+0\right\}\right]\right. \\ &\left.+\left\{\pi(3-x)(-(x-2.5)^2+4-0)+0\right\}-(2x-7)\right\}+(2x-7). \end{aligned}$$

Its graph and a selection of the constructed relaxations are showcased in Figure 1. It is worth while to point out several observations. First, the example shows that it is possible to model functions with multiple discontinuities, including such where the function does not attain either one-sided limit. Second, the generated relaxations are generally nonsmooth. This is characteristic for McCormick relaxations and has been noted previously [28].



(a) Graph of f_1

(b) Graph of f_1 and its relaxations

Fig. 1: Graph of f_1 (indicated by $+$) as well as convex relaxations (dashed line) and concave relaxations (continuous line) constructed on several intervals. Note that the scales on the vertical axes differ.

Example 2 Consider the lower semi-continuous function $f_2 : [0.5, 1.5]^2 \to \mathbb{R}$ with

$$f_2(x,y) = \begin{cases} 0.5\sin(6y-1)x^2 & \text{if } xy > 1, \\ 2(x+y) - e^{xy+1} & \text{if } xy \leq 1. \end{cases}$$

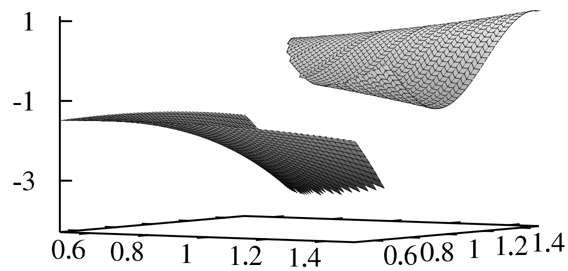
It can be represented as

$$f_2(x,y) = \pi(1-xy)\left[2(x+y) - e^{xy+1} - 0.5\sin(6y-1)x^2\right] + 0.5\sin(6y-1)x^2.$$

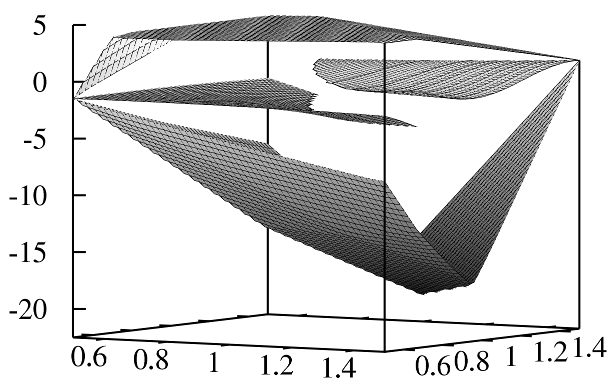
Its graph and a selection of the constructed relaxations are showcased in Figure 2. Note that π can take any arbitrary factor as argument, in this case a bilinear term, and thus the discontinuity can depend on the variables nonlinearly.

2.4 Assumptions on f, on φ_k and on the bounding operation and relaxations of φ_k

In Section 2.5, the convergence properties of McCormick relaxations of bounded factorable functions will be investigated. Prior to this, some assumptions about the interval extensions and the relaxations of the univariate functions will be made. This approach allows a more



(a) Graph of f_2



(b) Graph of f_2 and its relaxations

Fig. 2: Graph of f_2 as well as its convex and concave relaxations on $[0.5, 1.5]^2$. Note that the scales on the vertical axes differ.

general discussion compared to only studying a selection of univariate intrinsic functions or particular factorable functions.

In addition to Assumption 1, three additional assumptions will be made subsequently. While Assumptions 1–3 have been introduced and discussed in [37], Assumption 4 is newly introduced here and will be discussed in more detail below as it can be taken for granted in the setting considered in [37].

Assumption 2 For each k such that $n < k \leq m$ and $v_k : V_i \to \mathbb{R}$ defined by Definition 2 (c), $\check{\varphi}_k$ and $\hat{\varphi}_k$ are continuous functions on V_i.

Assumption 3 Consider two intervals, V_i^1 and V_i^2, such that $V_i^2 \subset V_i^1 \subset V_i$. For each k such that $n < k \leq m$ and v_k defined by Definition 2 (c), denote the convex and concave relaxations of φ_k constructed over V_i^l by $\breve{\varphi}_k^l$ and $\hat{\varphi}_k^l$, respectively. Assume that for each $z \in V_i^2$, $\breve{\varphi}_k^l$ and $\hat{\varphi}_k^l$ are constructed such that $\breve{\varphi}_k^2(z) \geq \breve{\varphi}_k^1(z)$ and $\hat{\varphi}_k^2(z) \leq \hat{\varphi}_k^1(z)$.

Note that is easy to show that the convex and concave relaxations of π satisfy Assumptions 2 and 3.

In order to streamline the presentation, the next assumption will be introduced, which is sufficient to prove convergence of $\breve{f}$ to f. This assumption is discussed in more detail in Appendix A. There, more insight into prerequisites for convergence of the relaxations to the function is given. Lastly, it should be pointed out that this assumption is imposed on a given factorization of a bounded factorable function f, similar to Assumption 1, while the previous assumptions were imposed on the set of considered univariate intrinsic functions φ.

Assumption 4 Consider a nested sequence of intervals $X^l \to X^* = [\mathbf{x}^*, \mathbf{x}^*]$, $X^l \in \mathbb{I}X$, $X^l \neq X^*$ and a factorization $v_1, \ldots, v_m$ of f. For each $n < k \leq m$, let $\underline{v}_k^l$ and $\overline{v}_k^l$ denote the lower and upper bounds of the cumulative mapping $v_k(\mathbf{x})$ on the interval X^l. Assume that for each $n < k \leq m$

$$\lim_{l\to\infty} [\underline{v}_k^l, \overline{v}_k^l] = [\lim_{l\to\infty} \inf_{\mathbf{x}\in X^l} v_k(\mathbf{x}), \lim_{l\to\infty} \sup_{\mathbf{x}\in X^l} v_k(\mathbf{x})]. \tag{2}$$

Note that $\lim_{l\to\infty} \inf_{\mathbf{x}\in X^l} v_k(\mathbf{x})$ does not refer to $\liminf_{\mathbf{x}\to\mathbf{x}^*} v_k(\mathbf{x})$. Assumption 4 states that, as X^l approaches the degenerate interval X^*, the bounds computed for each of the factors become as tight as possible. Since the bounds on the step mappings v_k are obtained from interval arithmetic, this clearly holds when f is composed of continuous factors. When f is discontinuous however, this is not necessarily true. For example, the dependency problem in

i	v_i	V_i^l	$\check{v}_i^l$	$\hat{v}_i^l$
1	x	$[-l^{-1}, l^{-1}]$	x	x
2	$\pi(v_1)$	$[0,1]$	$\begin{cases} 0 & \text{if } x \leq 0 \\ xl & \text{otherwise} \end{cases}$	$\begin{cases} 1+xl & \text{if } x \leq 0 \\ 1 & \text{otherwise} \end{cases}$
3	$v_2 - v_2$	$[-1,1]$	$\begin{cases} -xl-1 & \text{if } x \leq 0 \\ xl-1 & \text{otherwise} \end{cases}$	$\begin{cases} 1+xl & \text{if } x \leq 0 \\ 1-xl & \text{otherwise} \end{cases}$

Table 1: Factorization of $f = \pi(x) - \pi(x)$ on $X^l = [-l^{-1}, l^{-1}]$.

interval arithmetic is exacerbated and bounds do not necessarily converge to the function as the host set converges to a degenerate interval. This is demonstrated in the example below.

Example 3 Consider the *continuous* function $f : [-1,1] \to \mathbb{R}$ with $f(x) = \pi(x) - \pi(x)$. It can be equivalently written as $f(x) = 0$. Consider the nested sequence of intervals $X^l = [-l^{-1}, l^{-1}]$ that converges to $X^* = [x^*, x^*]$ with $x^* = 0$. It can be shown that the relaxations do not converge in this case. Consider this factorization given in Table 1. For all l, the relaxations of f constructed on X^l evaluated at x^* yield $\check{v}_3^l(0) = -1$ and $\hat{v}_3^l(0) = 1$, i.e., $\lim_{l\to\infty} \check{v}_3^l(0) = -1$ and $\lim_{l\to\infty} \hat{v}_3^l(0) = 1$ whereas the relaxations of f constructed on the degenerate interval X^* are $\check{v}_3^*(0) = f(0) = 0$ and $\hat{v}_3^*(0) = f(0) = 0$, also see Figure 3.

Note that, in this case, there exists a factorization that circumvents this dependency problem, namely $f(x) = 0$. Thus, depending on the problem formulation, this limitation may be avoided.

Similarly, applying a univariate function to a discontinuous factor may lead to bounds that do not converge to the infimum/supremum as $X^l \to X^*$. To see this, consider the univariate function $\varphi(x) = \cos(x - 0.75)$, the bounded factorable function $f(x) = \varphi(\pi(x))$ and $X^l = [-l^{-1}, l^{-1}]$, $X^* = [0,0]$. Again, $\Pi(X^l) = [0,1]$ and $F(X^l) = [\cos(-0.75), 1]$ while $\Pi(X^*) = [0,0]$ and $F(X^*) = [\cos(-0.75), \cos(-0.75)]$. Furthermore, $f(x) = \cos(-0.75)$

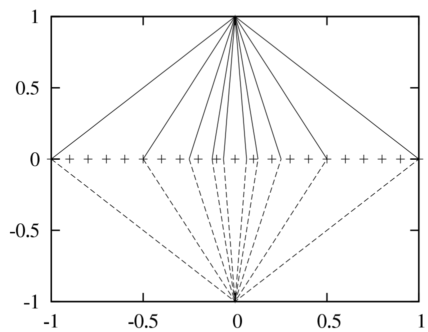


Fig. 3: Graph of f (indicated by $+$) as well as five of its convex and concave relaxations (indicated by dashed and continuous lines, respectively) for $l = 1, 2, 4, 8, 16$.

for all $x \in [-1, 0]$ and $f(x) = \cos(0.25)$ for all $x \in (0, 1]$. Thus, the upper bound does not converge to the supremum as desired. This is due to the fact that there exists a $y \in (0, 1)$ so that $\varphi(y) > \max(\varphi(0), \varphi(1))$. Again, this can be avoided when the problem is recast as $f(x) = \pi(x)(\cos(0.25) - \cos(-0.75)) + \cos(-0.75)$. Similar examples can be constructed so that the lower bound does not converge to the infimum.

The interested reader can find a more detailed discussion in the Appendix A where sufficient criteria for Assumption 4 are proven. In particular, the situations described in the examples above are analyzed in-depth.

2.5 Relaxations on sequences of intervals

The use of the McCormick relaxations in a branch-and-bound algorithm requires further investigation of their behavior with respect to the set on which they are defined. In this section, some properties of the relaxations will be established in such a setting. While the definitions are taken from Scott et al. [37], the facts established hereafter are novel and are not immediate. In the following, it will be assumed that Assumptions 1–3 hold. As noted earlier, Assumption 4 will only be required to show convergence of the bounding operation. It will be pointed out in the statement of the theorem when it is necessary. In the following, a property of the relaxation is first defined and then established by proof. Necessary intermediate results are stated as lemmas.

Definition 5 Let $f : X \to \mathbb{R}$ be bounded on $X \in \mathbb{IR}^n$. An algorithm which generates convex and concave relaxations $\check{f}^l$ and $\hat{f}^l$, respectively, of f on any $X^l \in \mathbb{I}X$ is *partition monotonic* if, for any subintervals $X^{l_2} \subset X^{l_1} \subset X$, $\check{f}^{l_2}(\mathbf{x}) \geq \check{f}^{l_1}(\mathbf{x})$ and $\hat{f}^{l_2}(\mathbf{x}) \leq \hat{f}^{l_1}(\mathbf{x})$, $\forall \mathbf{x} \in X^{l_2}$.

In the literature [29, 31], the result below is stated as follows: the composition of inclusion monotonic Lipschitz interval extensions is an inclusion monotonic Lipschitz interval extension. Here, only the inclusion monotonicity property is available so that the result must be modified, but the proof is straightforward and not given here.

Lemma 2 *Consider $X_1 \in \mathbb{IR}^n$ and $X_2 \in \mathbb{IR}$ and functions $f : X_1 \to X_2$ and $\varphi : X_2 \to \mathbb{R}$. If F is an inclusion monotonic interval extension of f on X_1, Φ is an inclusion monotonic interval extension of φ on X_2, and $F(X_1) \subset X_2$, then $\Phi \circ F$ is an inclusion monotonic interval extension of $\varphi \circ f$ on X_1.*

The following lemma is adapted from [37] so that the conclusion of inclusion monotonicity can be reached without requiring Assumption 3 therein (for each univariate intrinsic function a Lipschitz interval function can be given). This result is needed to use Lemma 5 in [37] in the proof of Theorem 5 below.

Lemma 3 *Choose any K, $1 \leq K \leq m$, and let the interval mapping $H(X^l) \equiv [\underline{v}_K^l, \bar{v}_K^l]$ be defined for any interval $X^l \subset X$ by the procedure in Definition 3 beginning with X^l. Then H is an inclusion monotonic interval extension of the cumulative mapping v_K on X.*

Proof Pick any subintervals of X, $X^2 \subset X^1 \subset X$. First note that $V_i^2 \subset V_i^1 \subset V_i$, $\forall i = 1, \ldots, n$. For an arbitrary k, suppose it is true for all $i < k$. The proof in [37] covers $\underline{v}_k$, $\bar{v}_k$ defined by 4a or 4b in Definition 3. Thus, it remains to show the conclusion for univariate intrinsic functions. If $\underline{v}_k$, $\bar{v}_k$ are defined by 4c, then an inclusion monotonic interval extension of the step mapping v_k on V_i is known by Definition 3. Since the cumulative mapping v_k is a composition of the step mapping v_k with the cumulative mapping v_i and since the cumulative mapping v_i maps any point in X^2 to $V_i^2 \subset V_i^1 \subset V_i$, Lemma 2 can be applied to show that $V_k^2 \subset V_k^1$. By finite induction, this must be true for all k and hence for $k = K$. □

Theorem 5 *McCormick relaxations of bounded factorable functions are partition monotonic.*

Proof Choose any subintervals $X^1 \subset X^2 \subset X$ and any $\mathbf{x} \in X^2$. $V_k^2 \subset V_k^1$ for all $k = 1,\dots,m$ by Lemma 3. For any k such that $1 \le k \le n$, it is easy to see from Definition 3 that $\breve{v}_k^1(\mathbf{x}), \hat{v}_k^1(\mathbf{x}) \in V_k^1$, $\breve{v}_k^2(\mathbf{x}), \hat{v}_k^2(\mathbf{x}) \in V_k^2$, $\breve{v}_k^2(\mathbf{x}) \ge \breve{v}_k^1(\mathbf{x})$ and $\hat{v}_k^2(\mathbf{x}) \le \hat{v}_k^1(\mathbf{x})$. This establishes the hypotheses of Lemma 5 in [37] for $k = n+1$. Hence, the inequalities hold for $k = n+1$ so that, by finite induction, the inequalities are true for all $k = 1,\dots,m$. □

Definition 6 An algorithm which generates convex and concave relaxations of $f : X \to \mathbb{R}$ is *weakly partition convergent* if, for any nested and convergent sequence of subintervals of X, $X^l \to X^*$, $X^l \ne X^*$, the sequences convex and concave relaxations of f on X^l, $\{\breve{f}^l\}$ and $\{\hat{f}^l\}$, converge uniformly to continuous convex and concave relaxations of f on X^*, $\breve{f}^*$ and $\hat{f}^*$, respectively.

Note that this definition deviates from the definition of partition convergent in [37]. Any continuous convex and concave relaxations of f, $\breve{f}^*$ and $\hat{f}^*$, meet the definition while Scott et al. [37] require convergence of $\breve{f}^l$ and $\hat{f}^l$ to the convex and concave relaxations generated on X^*, respectively.

Lemma 4 *Let $\{f^l\}$ be a sequence of functions defined on $X \in \mathbb{IR}^n$ and suppose that $\{f^l\}$ converges pointwise to f on X. If $\{f^l\}$ is nondecreasing, i.e., $f^l(\mathbf{x}) \le f^{l+1}(\mathbf{x})$, $\forall \mathbf{x} \in X$, and each f^l is lower semi-continuous on X, then f is lower semi-continuous on X.*

Proof A function $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is lower semi-continuous on $\mathbb{R}^n$ if and only if the level sets $\{\mathbf{x} \in \mathbb{R}^n : g(\mathbf{x}) \leq \gamma\}$ are closed for all $\gamma \in \mathbb{R}$ [20, p. 148]. This allows to extend the proof of Theorem 5.27 in [19] to $X \in \mathbb{IR}^n$ easily. $\square$

Lemma 5 *Let $f : X \to \mathbb{R}$ be bounded factorable. Suppose $X^l \to X^*$ is a nested sequence of intervals with $X^l \in \mathbb{I}X$, $X^l \neq X^*$ and consider the sequence of convex relaxations of f on X^l, $\{\underset{\smile}{f}^l\}$. Then, $\{\underset{\smile}{f}^l\}$ converges pointwise on X^* to an arbitrary function, denoted as $\underset{\smile}{f}^*$, that is continuous on X^* and a convex relaxation of f on X^*.*

Proof For any $\mathbf{x} \in X^*$ and $l > 0$, $\underset{\smile}{f}^{l+1}(\mathbf{x}) \geq \underset{\smile}{f}^l(\mathbf{x})$ by Theorem 5 and $\underset{\smile}{f}^l(\mathbf{x}) \leq f(\mathbf{x})$ by Theorem 2. Thus, $\underset{\smile}{f}^l$ converges pointwise to some function on X^*. This establishes existence of $\underset{\smile}{f}^*$. $\underset{\smile}{f}^l$ is convex by Theorem 2. Let $\mathbf{x}, \mathbf{y} \in X^*$ and $\lambda \in [0,1]$. Set $\mathbf{z} = \lambda\mathbf{x} + (1-\lambda)\mathbf{y}$. Convexity of $\underset{\smile}{f}^l$ on X^* for all l implies that $\underset{\smile}{f}^*$ is a convex relaxation of f on X^* since

$$\underset{\smile}{f}^*(\mathbf{z}) = \lim_{l\to\infty} \underset{\smile}{f}^l(\mathbf{z}) \geq \lambda \lim_{l\to\infty} \underset{\smile}{f}^l(\mathbf{x}) + (1-\lambda) \lim_{l\to\infty} \underset{\smile}{f}^l(\mathbf{y}) = \lambda \underset{\smile}{f}^*(\mathbf{x}) + (1-\lambda)\underset{\smile}{f}^*(\mathbf{y}).$$

As a result of Theorem 3, which establishes continuity of $\underset{\smile}{f}^l$, Lemma 4 can be applied to find that $\underset{\smile}{f}^*$ is lower semi-continuous on X^*. Lower semi-continuity and convexity of $\underset{\smile}{f}^*$ on X^* imply continuity of $\underset{\smile}{f}^*$ on X^* [32, Theorems 10.2 and 20.5]. $\square$

Theorem 6 *McCormick relaxations of bounded factorable functions are weakly partition convergent.*

Proof Suppose X^l is a nested and convergent sequence of subintervals of X, $X^l \to X^*$, and $X^l \neq X^*$. The intervals X^l are closed and bounded by definition and hence compact. Consider the sequence of $\{\underset{\smile}{f}^l\}$. Lemma 5 and Theorem 5 establish that the relaxations converge pointwise monotonically to a continuous function for each $\mathbf{x} \in X^*$. Rudin [34, Theorem 7.13] shows that this is sufficient for uniform convergence of $\{\underset{\smile}{f}^l\}$ to $\underset{\smile}{f}^*$ on X^*. A similar argument can be made to show $\hat{f}^l \to \hat{f}^*$ uniformly and the theorem follows. $\square$

Definition 7 A procedure such as in Definition 5 is *degenerate perfect* if $X^* = [\mathbf{x}, \mathbf{x}]$ for any $\mathbf{x} \in X$ implies that $\underset{\smile}{f}^\star(\mathbf{x}) = f(\mathbf{x}) = \hat{f}^\star(\mathbf{x})$ where $\underset{\smile}{f}^\star(\mathbf{x})$ and $\hat{f}^\star(\mathbf{x})$ denote the convex and concave relaxations of f on X^*, respectively.

Theorem 7 *McCormick relaxations of bounded factorable functions are degenerate perfect.*

Proof Suppose $X = [\mathbf{x}, \mathbf{x}]$ is degenerate. Note that $\underset{\smile}{v}_k(\mathbf{x}), \hat{v}_k(\mathbf{x}) \in V_k$ for all $k = 1, \ldots, m$ and every $\mathbf{x} \in X$ [37, Lemma 1]. This fact and Lemma 3 with $K = m$ guarantee that McCormick relaxations are degenerate perfect. □

Remark 2 Note that Theorem 6 and Theorem 7 do not imply that, for any nested sequence of subintervals of X with $\{X^l\} \to [\mathbf{x}^*, \mathbf{x}^*]$, $X^l \neq [\mathbf{x}^*, \mathbf{x}^*]$ and $\mathbf{x} \in X$, $\{\underset{\smile}{f}^l(\mathbf{x}^*)\} \to f(\mathbf{x}^*)$ and $\{\hat{f}^l(\mathbf{x}^*)\} \to f(\mathbf{x}^*)$. While this was asserted in [37], the utilized Lipschitz properties of $\underset{\smile}{f}^l$ and $\hat{f}^l$ do not hold here. Example 3 in Section 2.3 demonstrates that there are bounded factorable functions where $\{\underset{\smile}{f}^l(\mathbf{x})\} \to \underset{\smile}{f}^*(\mathbf{x}^*) \neq f(\mathbf{x}^*)$.

Theorem 8 *Assume f is a bounded factorable lower semi-continuous function with a factorization such that Assumption 4 holds. Suppose $\{X^l\}$ is a sequence of nested subintervals of X converging to $X^* = [\mathbf{x}^*, \mathbf{x}^*]$, $X^l \neq X^*$. Let $\underset{\smile}{f}^l : X^l \to \mathbb{R}$ be convex relaxations of $f : X \to \mathbb{R}$ on X^l obtained as described in Definition 3 and let $\mathbf{x}^l_{\min} \in \arg\min_{\mathbf{x} \in X^l} \underset{\smile}{f}^l(\mathbf{x})$. Then, $\underset{\smile}{f}^l(\mathbf{x}^l_{\min}) \to f(\mathbf{x}^*)$.*

Proof Fix $\varepsilon > 0$. Lower semi-continuity of f guarantees that $f(\mathbf{x}^*) \leq \liminf_{\mathbf{x} \to \mathbf{x}^*} f(\mathbf{x})$. Note that $\lim_{l \to \infty} \inf_{\mathbf{x} \in X^l \setminus X^*} f(\mathbf{x}) \geq \liminf_{\mathbf{x} \to \mathbf{x}^*} f(\mathbf{x})$ as, for each l, X^l is a subset of a suitable neighborhood of $\mathbf{x}^*$ referenced in the definition of the lower limit. Therefore, it follows that $f(\mathbf{x}^*) \leq \lim_{l \to \infty} \inf_{\mathbf{x} \in X^l \setminus X^*} f(\mathbf{x})$. Furthermore, it is true that $\lim_{l \to \infty} \inf_{\mathbf{x} \in X^l} f(\mathbf{x}) = f(\mathbf{x}^*)$ since $\lim_{l \to \infty} \inf_{\mathbf{x} \in X^l} f(\mathbf{x}) = \min\left\{ f(\mathbf{x}^*), \lim_{l \to \infty} \inf_{\mathbf{x} \in X^l \setminus X^*} f(\mathbf{x}) \right\}$.

Assumption 4 implies that $\lim_{l \to \infty} \underline{f}^l = \lim_{l \to \infty} \inf_{\mathbf{x} \in X^l} f(\mathbf{x})$. By Lemma 5, $\lim_{l \to \infty} \underset{\smile}{f}^l(\mathbf{x}^*)$ exists and let it be denoted as $\underset{\smile}{f}^*(\mathbf{x}^*)$. Since $\underline{f}^l \leq \underset{\smile}{f}^l(\mathbf{x}^*)$ by Step 6 in Definition 3, it holds that $\lim_{l\to\infty} \underline{f}^l \leq \underset{\smile}{f}^*(\mathbf{x}^*)$. Pointwise convergence of $\underset{\smile}{f}$ implies that there exists $L_1 \in \mathbb{N}$ so that $|\underset{\smile}{f}^l(\mathbf{x}^*) - \underset{\smile}{f}^*(\mathbf{x}^*)| \leq \varepsilon$, $\forall l \geq L_1$. Consequently,

$$f(\mathbf{x}^*) = \lim_{l\to\infty} \inf_{\mathbf{x}\in X^l} f(\mathbf{x}) = \lim_{l\to\infty} \underline{f}^l \leq \underset{\smile}{f}^*(\mathbf{x}^*) \leq \underset{\smile}{f}^l(\mathbf{x}^*) + \varepsilon, \quad \forall l \geq L_1.$$

Continuity of $\underset{\smile}{f}^{L_1}$ guarantees existence of $\delta > 0$ with

$$|\underset{\smile}{f}^{L_1}(\mathbf{x}) - \underset{\smile}{f}^{L_1}(\mathbf{x}^*)| < \varepsilon, \quad \forall \mathbf{x} \in X^{L_1} : \|\mathbf{x} - \mathbf{x}^*\| < \delta.$$

Since $X^l \to X^*$, there exists $L_2 \in \mathbb{N}$ so $\|\mathbf{x} - \mathbf{x}^*\| < \delta$ for all $\mathbf{x} \in X^{L_2}$.

Let $L = \max\{L_1, L_2\}$. Theorem 5 and the previous argument imply that

$$\underset{\smile}{f}^L(\mathbf{x}) \geq \underset{\smile}{f}^{L_1}(\mathbf{x}) > \underset{\smile}{f}^{L_1}(\mathbf{x}^*) - \varepsilon, \quad \forall \mathbf{x} \in X^L.$$

Consequently, $\underset{\smile}{f}^{L_1}(\mathbf{x}^*) - \underset{\smile}{f}^L(\mathbf{x}^L_{\min}) \leq \varepsilon$. As a result,

$$f(\mathbf{x}^*) - \underset{\smile}{f}^L(\mathbf{x}^L_{\min}) = [f(\mathbf{x}^*) - \underset{\smile}{f}^{L_1}(\mathbf{x}^*)] + [\underset{\smile}{f}^{L_1}(\mathbf{x}^*) - \underset{\smile}{f}^L(\mathbf{x}^L_{\min})] \leq 2\varepsilon.$$

Since ε was arbitrary, the theorem follows. □

Remark 3 Note that dropping the assumption of lower semi-continuity of f in Theorem 8 results in a weaker statement. Since $f(\mathbf{x}^*) \leq \lim_{l\to\infty} \inf_{\mathbf{x}\in X^l \setminus X^*} f(\mathbf{x})$ is not necessarily true then, one can only show that $\underset{\smile}{f}^l(\mathbf{x}^l_{\min})$ converges to $\min\left\{\lim_{l\to\infty} \inf_{\mathbf{x}\in X^l\setminus X^*} f(\mathbf{x}), f(\mathbf{x}^*)\right\}$. If $\mathbf{x}^*$ is in the interior of X^l, $\forall l$, then one can prove convergence to $\min\{\liminf_{\mathbf{x}\to\mathbf{x}^*} f(\mathbf{x}), f(\mathbf{x}^*)\}$, a statement that does not depend on the sequence of partition elements $\{X^l\}$. In this sense it is more general, but it is also a weaker result since $\lim_{l\to\infty} \inf_{\mathbf{x}\in X^l\setminus X^*} f(\mathbf{x}) \geq \liminf_{\mathbf{x}\to\mathbf{x}^*} f(\mathbf{x})$.

In this section, fundamental properties of the relaxations of bounded factorable functions with discontinuities have been established and assumptions are clarified when these results hold. These results are important when the relaxations are to be used in a branch-and-bound algorithm.

3 Branch-and-bound for bounded factorable optimization

In this section, it will be shown that McCormick relaxations of bounded factorable functions can be used to obtain a convergent branch-and-bound algorithm under mild assumptions. Branch-and-bound methods can be used to find a global minimum of a nonconvex nonlinear program. The standard reference for this class of algorithms, Horst and Tuy [22], considers *continuous* functions only when a general theoretical framework is constructed and convergence proofs are established. The work rests on several assumptions for the bounding, selection and refining operations [22]. The bounding operation is responsible for generating lower and upper bounds on the optimal objective value on a partition element, while the latter two are responsible for selecting a partition element for further investigation and refining it. In the remainder of this section, the discussion will be focused on the bounding operation.

The reader is referred to Horst and Tuy [22] and Horst [21] for precise definitions of most technical terms. References that are given with the statements in this section guide the reader to similar results in the literature for the case of continuous functions. However, the results are indeed newly established or needed to be verified for the new hypotheses.

The following definition has been adapted to account for discontinuities.

Definition 8 [cf. 21, p. 24] Suppose that for any infinitely decreasing sequence of successively refined partition elements $\{X^l\}$ generated by an exhaustive subdivision and satisfying $\lim_{l\to\infty} X^l \to \{\mathbf{x}^\dagger\}$, there exists a subsequence $\{X^{l_q}\}$ such that

$$\lim_{q\to\infty} \beta(X^{l_q}) \geq \min\left\{\liminf_{\mathbf{x}\to\mathbf{x}^\dagger} f(\mathbf{x}), f(\mathbf{x}^\dagger)\right\}.$$

Then, the lower bounding operation is called *strongly consistent*.

As remarked by Horst and Tuy, finiteness and convergence properties of the branch-and-bound algorithm depend on the behavior of $\alpha(X^l) - \beta(X^l)$ in the limit [22, p. 128].

Whereas favorable behavior of the McCormick relaxations of factorable functions in this spirit has been argued previously [37], it still needs to be established for the case of bounded factorable functions. In the following, f is assumed to be bounded factorable and the lower bound of (P) on a partition element $\tilde{X} \in \mathbb{I}X$, $\beta(\tilde{X})$, is found by constructing the convex McCormick relaxation $\underset{\smile}{f}$ on $\tilde{X}$ and minimizing it, i.e., $\beta(\tilde{X}) = \min_{\mathbf{x}\in\tilde{X}} \underset{\smile}{f}(\mathbf{x})$.

Theorem 9 [cf. 21, p. 28f] *Suppose that Assumption 4 holds and that f is lower semi-continuous. Assume that at every step any unfathomed partition element can be refined. Suppose that the subdivision is exhaustive. Then, the lower bounds of* (P) *obtained by minimizing the McCormick relaxations are strongly consistent.*

Proof It is sufficient to show that, for every decreasing sequence of successively refined partition elements $\{X^l\}$ generated by an exhaustive subdivision such that $\lim_{l\to\infty} X^l = \bigcap_l X^l = \overline{X} = [\mathbf{x}^*, \mathbf{x}^*]$, there is a subsequence $\{X^{l_q}\}$ satisfying $\lim_{q\to\infty} \beta(X^{l_q}) = f(\mathbf{x}^*)$. This is guaranteed by Theorem 8 since $\beta(X^{l_q}) = \min_{\mathbf{x}\in X^{l_q}} \underset{\smile}{f}^{l_q}(\mathbf{x})$. Lower semi-continuity of f implies that $f(\mathbf{x}^*) \leq \liminf_{\mathbf{x}\to\mathbf{x}^*} f(\mathbf{x})$ and hence strong consistency follows. □

Consider the definition of $\beta(X^l)$ and $\beta_k = \min_{l\in I_k}\{\beta(X^l)\}$ and denote as L_k an element of the index set I_k such that $\beta(X^{L_k}) = \beta_k$. Let $\mathbf{x}_{\min}(X^l) \in \arg\min_{\mathbf{x}\in X^l} \underset{\smile}{f}^l(\mathbf{x})$ and define $\mathbf{x}^k_{\min} \equiv \mathbf{x}_{\min}(X^{L_k})$. Similarly, denote as $\mathbf{x}^k \in D$ a point corresponding to α_k, i.e., $\alpha_k = f(\mathbf{x}^k)$. Horst [21, Theorem 2.1] proves that, for a continuous function f, a strongly consistent lower bounding operation in combination with some additional assumptions is sufficient to show that the lower bound β_k converges to the optimal value of (P) and that accumulation points of $\{\mathbf{x}^k_{\min}\}$ solve (P). The argument can also be applied to functions that attain their minimum on D.

Theorem 10 [cf. 21, p. 25f] *Suppose that the subdivision of partition elements is exhaustive, that the selection operations is bound improving, that the lower bounding operation is* *strongly consistent and that the "deletion by infeasibility" rule is certain in the limit. Assume that f attains its minimum on D. Let $X_{\min}$ be the set of accumulation points of $\{\mathbf{x}_{\min}^k\}$. Then, it follows that $\beta = \lim_{k\to\infty} \beta_k = \min_{\mathbf{x}\in D} f(\mathbf{x})$ and $X_{\min} \subset \arg\min_{\mathbf{x}\in D} f(\mathbf{x})$.*

Proof The proof is identical to the argument in [21] assuming that the minimum is attained is sufficient. Also, the modification of the definition of strongly consistent bounding operations is irrelevant for the proof. □

On the other hand, providing an argument to prove consistency of the lower bounds obtained by using the McCormick relaxations for lower semi-continuous functions is more involved and requires an additional assumption. In the case of a continuous function f, it is obvious that $\alpha(X^l)$ approaches $f(\mathbf{x}^*)$ as $X^l \to \{\mathbf{x}^*\}$ for an infinitely decreasing sequence of successively refined partition elements $\{X^l\}$. When the assumption of continuity of f is dropped, the convergence of $\alpha(X^l)$ to $f(\mathbf{x}^*)$ cannot be asserted as $\alpha(X^l)$ is, by definition, the function value at *some* feasible point in X^l. In particular, it cannot be guaranteed that there exists a $\mathbf{x} \in D$ in a neighborhood of a minimizer of (P), denoted as $\mathbf{x}_{\min}$, so that $f(\mathbf{x})$ approximates f^* well. This is demonstrated well in Example 1 in Section 2.3. For a practical implementation, a subset D' of D in the neighborhood of $\mathbf{x}_{\min}$ must exist so that $f(\mathbf{x})$ is close to f^* when $\mathbf{x} \in D'$. Otherwise it may not be possible to identify numerically a sufficiently good approximation of f^*.

Assumption 5 Suppose there exists a $\mathbf{x}_{\min} \in \arg\min_{\mathbf{x}\in D} f(\mathbf{x})$ with the following property: $\mathbf{x}_{\min}$ is not an isolated point of D and for every $\varepsilon > 0$ there is a $\delta > 0$ and a cone C with $\mathbf{x}_{\min}$ at its apex such that $C_\delta = C \cap \{\mathbf{x} \in D : \|\mathbf{x}_{\min} - \mathbf{x}\| < \delta\}$ has nonzero volume and $f(\mathbf{y}) < f^* + \varepsilon$ for all $\mathbf{y} \in C_\delta$.

Remark 4 Note that Assumption 5 implies that f is upper semi-continuous at $\mathbf{x}_{\min}$ when the domain of f is restricted to a feasible subset of a neighborhood of $\mathbf{x}_{\min}$ with nonzero volume, e.g., a sphere with positive radius in $\mathbb{R}^3$, but not a plane in $\mathbb{R}^3$. However, it does not necessarily imply upper semi-continuity of f at $\mathbf{x}_{\min}$.

Theorem 11 *Suppose Assumptions 4 and 5 hold and that f is lower semi-continuous. Assume that at every step any undeleted partition element can be further refined. Suppose that the subdivision is exhaustive. Then, the lower bounds of* (P) *obtained by minimizing the McCormick relaxations, i.e., $\beta(\tilde{X}) = \min_{\mathbf{x}\in\tilde{X}} \underset{\smile}{f}(\mathbf{x})$ for some partition element $\tilde{X} \in \mathbb{I}X$ and for the McCormick relaxation $\underset{\smile}{f}$ constructed on $\tilde{X}$, are consistent.*

Proof Fix $\varepsilon > 0$. If $\alpha_k < f^* + \varepsilon$ at some iteration k, an ε-optimal solution has been found so that, in combination with Theorem 9, consistency of the bounding operation follows.

Otherwise, let $\delta > 0$ and the set C_δ as given by Assumption 5. Denote as $\tilde{X}$ the partition element of partition $\mathscr{P}_{k_1}$ with $\tilde{X} \subset C_\delta$ at some iteration k_1. The existence of such a partition element follows from the assumption of exhaustive subdivision and the fact that $\beta(\tilde{X}) < f^* + \varepsilon \leq \alpha_{k_1-1}$ so that neither $\tilde{X}$ nor a partition element that contains $\tilde{X}$, due to Theorem 5, could have been fathomed previously. By construction, $\alpha(\tilde{X}) < f^* + \varepsilon$. Thus, a feasible point $\tilde{\mathbf{x}} \in D$ has been found so that $f(\tilde{\mathbf{x}})$ is close to f^*, i.e., $\alpha_{k_1} \leq \alpha(\tilde{X}) < f^* + \varepsilon$ or $\alpha_{k_1} - f^* < \varepsilon$ holds. Consider an infinitely decreasing sequence $\{X^l\}$. Since it is infinitely decreasing, it follows that $\beta(X^l) < \alpha_{k_1}$ for all $l > L_{k_1}$ where L_{k_1} corresponds to iteration k_1; otherwise the partition element would be fathomed hereafter contradicting the assumption that $\{X^l\}$ is an infinitely decreasing sequence. Theorem 9 established that $\{\beta(X^l)\}$ converges to f^* so that there exists a k_2 with $f^* - \beta(X^l) < \varepsilon$ for all $l > L_{k_2}$ where L_{k_2} corresponds to iteration k_2. Consequently, $\alpha_{k_l} - \beta(X^l) < 2\varepsilon$ for $l > \max\{L_{k_1}, L_{k_2}\}$. Since ε was arbitrary, the bounding procedure is consistent. $\square$

Horst and Tuy [22] prove the convergence of the sequence of current best points of the branch-and-bound algorithm to an optimal solution. Corollary IV.2 that they present can be extended to lower semi-continuous functions.

Theorem 12 [cf. 22, p. 132] *Let f be lower semi-continuous. Suppose that the bounding operation is consistent and the selection operation is complete. Then every accumulation point of* $\{\mathbf{x}^k\}$ *solves* (P).

Proof Since D is compact, the sublevel set $C(f(\mathbf{x}^0)) = \{\mathbf{x} \in D : f(\mathbf{x}) \leq f(\mathbf{x}^0)\}$ is bounded and, since f is lower semi-continuous, $C(f(\mathbf{x}^0))$ is closed; cf. [20, p. 148]. Thus, $C(f(\mathbf{x}^0))$ is compact. By construction, $f(\mathbf{x}^{k+1}) \leq f(\mathbf{x}^k)$, $\forall k$, so that $\{\mathbf{x}^k\} \subset C(f(\mathbf{x}^0))$. Hence, $\{\mathbf{x}^k\}$ possesses accumulation points. The assertion then follows from [22, Theorem IV.2]. □

In this section it was shown that the branch-and-bound algorithm converges under some mild assumptions to a global optimum even in the presence of discontinuities. The presented results assume either that f is lower semi-continuous or that f attains its minimum on D. A discussion of the case when these hypotheses are not met can be found in Appendix B.

4 Case Studies

In this section, results will be presented from applying the proposed relaxations to some global optimization case studies. First, the discussed method will be applied to a problem from process design and equipment sizing. The section concludes with an example concerning a discrete-time hybrid system.

In the following a simple branch-and-bound algorithm will be used to converge lower and upper bounds and thus find a global optimal solution. At iteration k with partition element $X^l \in \mathscr{P}_k$, upper and lower bounds are found as follows. In general, an upper bound $\alpha(X^l)$ is obtained by evaluating the objective function at the solution of the lower bounding problem (if feasible). To find this solution and a valid lower bound $\beta(X^l)$, different methods are employed. The first method uses only interval arithmetic whereas the other ones use the convex relaxation and a subgradient of the relaxation. The reader is referred to [28] for details on how to construct the subgradient of McCormick relaxations.

Method 1 The bound from interval arithmetic, $\underline{f}$, is used as $\beta(X^l)$. The objective function is evaluated at the midpoint of the interval X^l to find $\alpha(X^l)$. This procedure yields very efficient lower bounds at the expense of tightness.

Method 2 An affine approximation of the convex relaxation of the objective function is constructed sequentially. First, a subgradient of $\underset{\smile}{f}$ is evaluated at the midpoint of X^l and an affine relaxation of $\underset{\smile}{f}$ is thus constructed. Combined with the interval bound, $\underline{f}$, CPLEX is used to find a minimum of the affine relaxations. A subgradient of $\underset{\smile}{f}$ is evaluated at this solution, another affine relaxation is added and CPLEX is used to solve this problem. To balance efficiency and accuracy, a total of five minimization problems are solved with CPLEX. The last solution found is reported as $\beta(X^l)$. $\alpha(X^l)$ is obtained by evaluating the objective function at the last point found by CPLEX.

Method 3 Since CPLEX adds considerable overhead, a simple algorithm is explicitly implemented that mimics Method 2 for one-dimensional problems and constructs only two affine relaxations.

Method 4 A bundle solver [25] with bundle size 15 is used to find the minimum of the convex relaxation of the objective function. Note that the QP routines have been modified to prevent an infinite loop in the inner QP. In this case the bundle solver terminates with $\beta(X^l) = -\infty$. $\alpha(X^l)$ is obtained by evaluating the objective function at the point returned by the bundle solver.

In the remainder of this section the different methods will be referred to by these assigned numerals for brevity. The open source C++ library libMC [7, 28] is used to calculate the necessary convex relaxations, and it relies on the interval library PROFIL [23] with outward rounding. libMC and PROFIL are extended to include π, its bounds and relaxations as well as subgradients. The global optimization problem is considered converged at iteration k when either $\alpha_k - \beta_k \leq \varepsilon_a$ or $\alpha_k - \beta_k \leq \varepsilon_r|\beta_k|$, where $\varepsilon_a = 10^{-5}$ and $\varepsilon_r = 10^{-5}$ (unless noted otherwise). The best bound heuristic is used to determine the next node and the absolute diameter heuristic is used to select on which variable to branch.

In the case of the more involved problems, the behavior of the proposed methods is compared to the commercial global optimization software BARON [36] as part of GAMS 23.9.5 with regard to number of nodes visited and solution times. Results for the following cases will be presented:

BARON1 Literature model with equal branching priority for each variable.

BARON2 Literature model with branching on binary variables and subset of continuous variables only.

BARON3 Literature model reduced by analytically replacing some equality constraints and intermediate variables; equal branching priority for each variable.

BARON4 Literature model reduced by analytically replacing some equality constraints and intermediate variables; branching on binary variables and subset of continuous variables only.

The same tolerances as listed above are used for BARON. The reader should take note that the branch-and-cut algorithm implemented in BARON employs many features (e.g., range reduction, constraint propagation, etc) that are not implemented in the methods proposed above.

Lastly, a note on notation in this section: in tables containing the results, $\mathbf{x}_{\min}$ always denotes the approximate optimal solution, regardless of symbols used in the problem definition, and f^* indicates the objective value at this point.

4.1 Process design and equipment sizing

A specific example from process design in chemical engineering is considered here. Heat exchanger network synthesis problems have been studied extensively, see [17] for a review. A heat exchanger is a device in which two or more fluid streams are brought into energetic contact. Though they cannot exchange mass, the colder stream is heated by the hotter stream and vice versa. The necessary area in the unit for this heat transfer depends on the amount of heat transferred, the temperature difference and the so-called heat transfer coefficient. In the process industry, a common task is to design and size a complex network of heat exchangers to minimize investment and operational cost. Often, heating/cooling utilities such as steam and cooling water are also available. In practice, different device designs are used for different heat transfer areas. As a consequence, the capital cost correlation that links area to cost for these units is not continuous. Also, there are upper limits on the size of a single unit due to the difficulty of transporting large heat exchangers to the plant site. In the present problem, it is assumed that smaller units can be operated in parallel to circumvent this problem.

In the literature, Türkay and Grossmann [39] give a MINLP model that uses disjunctions to model the discontinuity in the cost correlation. An alternative reduced formulation is possible. First, the discontinuous cost correlation can be directly represented without disjunctions or binary variables. Second, equality constraints, in particular energy balances for each heat exchanger, can be used to eliminate variables in the model. Then, one can identify

A [m^2]	investment cost [\$/yr]
$0 \leq A \leq 20$	$670A^{0.83}$+2000
$20 < A \leq 50$	$640A^{0.83}$+8000
$50 < A \leq 100$	$600A^{0.83}$+16000

Table 2: Equipment cost correlation for heat exchangers depending on required area

a small number of temperatures that can be chosen independently. After these temperatures are fixed, all remaining intermediate temperatures can be calculated from energy balances. The area A required for each heat exchanger is determined by $A = \frac{Q}{U \times LMTD}$,

where $Q = Fc_{p,\mathrm{H}}(T_{\mathrm{H,in}} - T_{\mathrm{H,out}})$ is the heat transfered, Fc_p denotes the heat capacity flow rate, T_{in} and T_{out} the in- and outlet temperatures of the hot and cold streams, U the overall heat transfer coefficient, and $LMTD$ the log mean temperature difference. Instead of using the exact expression for $LMTD$, Chen's approximation [8] is used,

$$LMTD = \left((T_{\mathrm{out},H} - T_{\mathrm{in},C})(T_{\mathrm{in},H} - T_{\mathrm{out},C}) \frac{(T_{\mathrm{out},H} - T_{\mathrm{in},C}) + (T_{\mathrm{in},H} - T_{\mathrm{out},C})}{2} \right)^{\frac{1}{3}}.$$

Depending on the heat transfer area, one can then choose from three different available heat exchanger designs with different investment cost correlations, which are given in Table 2. When the necessary area for one heat exchanger exceeds the maximum area of the largest heat exchanger design, the streams will be split and several heat exchangers will be used. At most seven parallel heat exchangers will be allowed to always ensure feasibility of the solution. Lastly, the operating expenses are found by calculating the cost of cooling water (20 \$/kW yr) and the cost of steam (80 \$/kW yr).

Overall, a factorable representation of the objective function can be constructed as outlined above. In the routine to calculate the convex relaxations of the objective function, a priori known bounds on intermediate quantities, e.g., areas need to be nonnegative, temper-

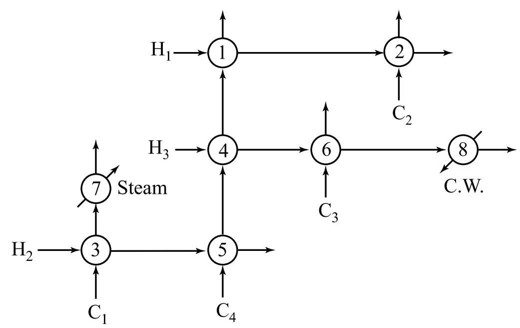


Fig. 4: Structure of heat exchanger network 1

ature differences in the heat exchangers cannot be negative and intermediate temperatures must be between inlet and outlet temperatures of the respective stream, are used to obtain tighter bounds for intermediate expressions.

4.1.1 Heat exchanger network 1

The first case study was taken from [39]. Consider the heat exchanger network depicted in Figure 4 with stream data given in Table 3. Let the overall heat transfer coefficient of the heat exchangers be given by $(U_i) = (1.5, 0.2, 0.06, 1.6, 0.04, 0.3, 0.6, 1.7)\,\mathrm{kW/m^2K}$.

There are seven unknown intermediate stream temperatures and two unknown utility heat loads. Since one can write an energy balance for each of the eight heat exchangers, the problem has one degree of freedom. The temperature of stream H_3 at the outlet of exchanger 6 was selected as the decision variable T. From requirements for feasible heat exchange, i.e., no temperature crossover in the heat exchangers, it follows that $T \in [382.25, 499.36]\,\mathrm{K}$. Once this variable is fixed, the remaining intermediate temperatures, utility heat loads, areas and hence investment costs can be computed by a factorable function as described before.

Stream	Fc_p [kW/K]	T_{in} [K]	T_{out} [K]
H_1	30.0	626	586
H_2	13.5	620	519
H_3	20.0	528	353
C_1	31.0	525	613
C_2	5.0	405	576
C_3	28.0	353	386
C_4	11.0	313	545
steam	—	650	650
cooling water	—	293	308

Table 3: Data for process and utility streams in heat exchanger network 1

The solutions as found with the different methods are compared in Table 4 to the solution obtained with BARON [36] using the MINLP model proposed in [39]. In the case of the reduced model, the energy balances are solved for the intermediate temperatures, which are subsequently substituted in the equation for $LMTD$. The expressions for $LMTD$ have not been substituted since the non-integer exponent is reformulated by GAMS using the exponential function and the natural logarithm. During the model development, GAMS aborted reporting domain violations so that this substitution is not feasible. In the case of selective branching, BARON is instructed to branch on T and the binary variables only.

All methods find the same solution; see Table 4 for more details. Lastly, it is instructive to point out that the full disjunctive model introduces 168 binary and 360 continuous variables.

Method	# LBPs	# UBPs	Runtime [s]	f^*	$\mathbf{x}_{\min}$ [K]
1	399	205	0.2269	$411,809$	418.103
2	63	42	0.2776	$411,809$	418.104
3	67	41	0.076	$411,809$	418.103
4	61	40	0.3334	$411,809$	418.104
BARON1	48 iterations		3.96	$411,809$	418.103
BARON2	91 iterations		4.92	$411,809$	418.103
BARON3	46 iterations		1.97	$411,809$	418.103
BARON4	18 iterations		1.22	$411,809$	418.103

Table 4: Comparison of different methods with BARON for the first heat exchanger case study

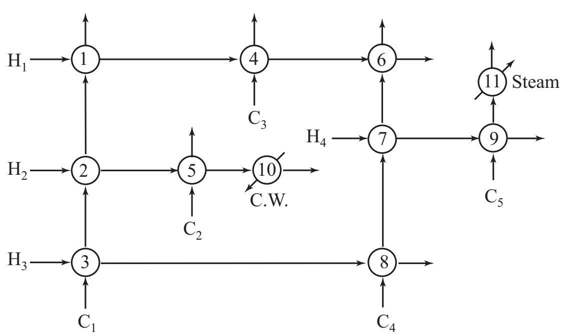


Fig. 5: Structure of heat exchanger network 2

4.1.2 Heat exchanger network 2

Consider the heat exchanger network depicted in Figure 5 with stream data given in Table 5. The goal is to optimize the network and size the equipment so that the combined investment and operational cost is minimized. Let the overall heat transfer coefficient of the heat exchangers be given by $(U_i) = (1.0, 0.1, 2.1, 0.05, 1.0, 0.2, 1.5, 0.7, 4.0, 1.2, 0.1)\,\mathrm{kW/m^2K}$.

Stream	Fc_p [kW/K]	T_{in} [C]	T_{out} [C]
H_1	22.4	400	150
H_2	12.0	180	40
H_3	26.0	150	45
H_4	24.0	135	100
C_1	15.0	105	360
C_2	20.0	40	65
C_3	22.0	90	190
C_4	35.0	25	110
C_5	16.2	30	150
steam	—	400	400
cooling water	—	15	30

Table 5: Data for process and utility streams in heat exchanger network 2

There are eleven unknown intermediate stream temperatures and two unknown utility heat loads. Since one can write an energy balance for each of the eleven heat exchangers, the problem has two degrees of freedom. The temperature of stream H_3 at the outlet of exchanger 3 and the temperature of stream H_2 at the outlet of exchanger 2 were selected as the decision variables T' and T'', respectively. From requirements for feasible heat exchange, i.e., no temperature crossover in the heat exchangers, it follows that $T' \in [129.81, 150.0]$ C and $T'' \in [124.17, 180.0]$ C; furthermore, it needs to hold that

$$26T' + 15T'' \geq 5625$$

$$312T' + 210T'' \leq 84565.$$

The solutions as found with the different methods are compared in Table 6 to the solution obtained with BARON [36] using the model with disjunctions proposed in [39]. The reduced

Method	# LBPs	# UBPs	Runtime [s]	f^*	$\mathbf{x}_{\min}$ [C]
1	$> 100,000$	$> 70,836$	> 92.0	—	—
2	$3,290$	$1,804$	21.6	$599,740$	(130.40, 160.49)
4	$3,661$	$2,000$	25.9	$599,740$	(130.40, 160.49)
BARON1	65 iterations		8.55	$599,740$	(130.40, 160.49)
BARON2	151 iterations		13.80	$599,740$	(130.40, 160.49)
BARON3	389 iterations		33.18	$599,740$	(130.40, 160.49)
BARON4	218 iterations		23.41	$599,740$	(130.40, 160.49)

Table 6: Comparison of different methods with BARON for the second heat exchanger case study

model is constructed as outlined in Section 4.1.2. In the case of selective branching, BARON is instructed to branch on T', T'' and the binary variables only.

A few remarks are in order. First, the interval bounds do not converge to the solution in 100,000 iterations and consequently, the branch and bound procedure in Method 1 fails to terminate with a guaranteed solution. Second, note that BARON requires fewer iterations than Methods 2 and 4 to identify its solution, however, each iteration is significantly more costly; see Table 6 for more details. Also note that the full disjunctive model used in BARON introduces 231 binary and 496 continuous variables.

4.2 Discrete-time hybrid systems

A second class of problems with discontinuous behavior is considered. Hybrid systems combine continuous dynamics, that are described by differential equations, and discrete dynamics, which are discontinuous changes in state variables or switching of the dynamic model triggered by so-called events [11, 18]. In discrete-time systems, the continuous dynamics are discretized and described by difference equations. The problem below, which concerns the optimal control of a linear discrete-time hybrid system, is slightly adapted from [24]. Consider the global optimization problem with an embedded discrete-time hybrid system

$$
\begin{aligned}
\min_{u_0,\dots,u_{N-1}} \quad & \sum_{k=1}^{N} \left(\mathbf{x}_k^{\mathrm{T}}\mathbf{R}\mathbf{x}_k + u_{k-1}Qu_{k-1}\right) \\
\text{s.t.} \quad & \mathbf{x}_{k+1} = \mathbf{A}(m(k))\mathbf{x}_k + \mathbf{B}(m(k))u_k,\ k = 1,\dots,N-1, \\
& m(k) = \begin{cases} 1 \text{ if } x_{k,1} \le x_{k,2}, \\ 2 \text{ otherwise,} \end{cases},\ k = 1,\dots,N-1,
\end{aligned}
$$

with $N = 10$ and, for $k = 1,\dots,N$, $u_{k-1} \in [-1,1]$,

$$
\mathbf{A}(m(k)) = \begin{cases} \begin{bmatrix} 0 & 0.2 \\ -0.4 & -0.06 \end{bmatrix} & \text{if } m(k) = 1, \\ \begin{bmatrix} 0.2 & 0.6 \\ -0.2 & 0.4 \end{bmatrix} & \text{if } m(k) = 2, \end{cases} \qquad \mathbf{B}(m(k)) = \begin{cases} \begin{bmatrix} 0 \\ 0.4 \end{bmatrix} & \text{if } m(k) = 1, \\ \begin{bmatrix} 0.2 \\ 0.1 \end{bmatrix} & \text{if } m(k) = 2, \end{cases}
$$

$$
\mathbf{R} = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.5 \end{bmatrix}, \qquad \text{and } Q = 1.0.
$$

Two cases are considered where the initial conditions differ: Case 1 with $\mathbf{x}_0 = [4,5]^{\mathrm{T}}$ and Case 2 with $\mathbf{x}_0 = [5,4]^{\mathrm{T}}$.

The objective can be calculated using a finite algorithm that takes $u_0,\dots,u_{N-1}$ as input and returns the objective. The detailed results for both cases are shown in Table 7. Here, the relative tolerance is set to $\varepsilon_r = 10^{-1}$ initially. Note that in both cases, Methods 2 and 4 find the optimal solution at the root node while Method 1 does not converge the lower and upper bound within 100,000 iterations, which is indicative of the weakness of only using interval methods for multi-dimensional problems.

It is important to remark that, although both problems are solved at the root node, the lower bound does not converge to the value of the optimal solution. Instead, a small but finite

Case	Method	# LBPs	# UBPs	Runtime [s]	f^*	$\mathbf{x}_{\min}$
	1	$> 100,000$	$> 100,000$	> 30.0	—	—
1	2	1	1	0.0280	7.256	(1,0.2970,0,0,0,0,0,0,0,0)
	4	1	1	0.0114	7.261	(1,0.2479,0,0,0,0,0,0,0,0)
	1	$> 100,000$	$> 69,712$	> 28.2	—	—
2	2	1	1	0.0263	13.077	(-0.7499,-0.2549,0,0,0,0,0,0,0,0)
	4	1	1	0.0112	13.049	(-0.8184,-0.1191,0,0,0,0,0,0,0,0)

Table 7: Comparison of different methods for both cases of the discrete-time hybrid system. Note that Method 1 does not converge in either case after solving 100,000 iterations.

discrepancy will remain indefinitely. This results from the presence of discontinuities; cf. the discussion in Section 2.4. In this example, it does not impact convergence when $\varepsilon_r = 10^{-1}$. However, Methods 2 and 4 do not converge Case 1 within 100,000 iterations when $\varepsilon_r = 10^{-2}$ and Case 2 when $\varepsilon_r = 10^{-4}$.

5 Conclusion

A procedure to construct interval bounds and convex and concave relaxations of factorable functions with discontinuities has been presented. McCormick's composition theorem [26] is extended to bounded, but not necessarily continuous, functions. The crux of the proposed extension lies in the observation that discontinuities can be modeled using a step function [41] and that convex and concave envelopes can be readily constructed for this function. Furthermore, it was shown that most theoretical results developed for the continuous case [37] hold even when the assumption of continuity is dropped. Only establishing convergence of a sequence of relaxations to the function when a sequence of intervals converging to a degenerate interval is considered requires additional assumptions. Currently, some results are established to show when this assumption holds. Nevertheless, this remains an active area of research for the authors as, e.g., examples shown in the previous section indicate that the relaxations converge for problems of practical importance or at least provide sufficiently tight relaxations. Also, these case studies show that the proposed method may provide a very effective means to solve optimization problems with discontinuities to global optimality without introducing binary variables. Thus an increase in size of the global optimization problem can be avoided, which is very desirable since known global optimization algorithms scale exponentially. Also note that so far no range reduction techniques have been employed which considerably improve convergence in BARON. There appears no distinct advantage of reducing the size of the optimization problem in BARON. Also, it is not possible to deduce a general advantage when branching on a subset of the continuous variables only. Lastly, the advantage of convex relaxations over interval bounds is demonstrated for multi-dimensional problems. While one-dimensional problems can be solved efficiently when only interval bounds are available, the convex relaxations are key for efficiently finding the optimal solutions of multi-dimensional problems.

Acknowledgements This work was supported by Statoil as part of the paired Ph. D. research program in gas technologies between MIT and NTNU.

References

1. Adjiman CS, Dallwig S, Floudas CA, Neumaier A (1998) A global optimization method, αBB, for general twice-differentiable constrained NLPs–I. Theoretical advances. Computers & Chemical Engineering 22(9):1137–1158
2. Barton PI, Allgor RJ, Feehery WF, Galán S (1998) Dynamic optimization in a discontinuous world. Industrial & Engineering Chemistry Research 37(3):966–981

3. Batukhtin VD (1993) On solving discontinuous extremal problems. Journal of Optimization Theory and Applications 77:575–589
4. Batukhtin VD (1995) An approach to the solution of discontinuous extremal problems. Journal of Computer and Systems Sciences International 33:30–38
5. Batukhtin VD, Bigil'deev SI, Bigil'deeva TB (1997) Numerical methods for solutions of discontinuous extremal problems. Journal of Computer and Systems Sciences International 36:438–445
6. Baumrucker BT, Renfro JG, Biegler LT (2008) MPEC problem formulations and solution strategies with chemical engineering applications. Computers & Chemical Engineering 32:2903–2913
7. Chachuat B, Mitsos A, Barton PI (2007) libMC - A numeric library for McCormick relaxation of factorable functions. `http://yoric.mit.edu/libMC/`
8. Chen J (1987) Comments on improvements on a replacement for the logarithmic mean. Chemical Engineering Science 42(10):2488–2489
9. Clarke FH (1983) Optimization and nonsmooth analysis. Wiley, New York, NY
10. Conn AR, Mongeau M (1998) Discontinuous piecewise linear optimization. Mathematical Programming 80(3):315–380
11. Cortés J (2008) Discontinuous dynamical systems. IEEE Control Systems Magazine 28(3):36–73
12. Duran MA, Grossmann IE (1986) Simultaneous optimization and heat integration of chemical processes. AIChE Journal 32:123–138
13. Ermoliev YM, Norkin VI (1998) On constrained discontinuous optimization. In: Proceedings of 3rd GAMM/IFIP Workshop, Stochastic optimization: Numerical methods and technical applications, Springer, Berlin, Lecture Notes in Economics and Mathematical Systems, vol 458, pp 128–142
14. Ermoliev YM, Norkin VI, Wets RJB (1995) The minimization of semicontinuous functions: Mollifier subgradients. SIAM Journal on Control and Optimization 33:149–167
15. Falk JE, Soland RM (1969) An algorithm for separable nonconvex programming problems. Management Science 15:550–569
16. Ferris MC, Dirkse SP, Jagla JH, Meeraus A (2009) An extended mathematical programming framework. Computers & Chemical Engineering 33(12):1973–1982
17. Furman KC, Sahinidis NV (2002) A critical review and annotated bibliography for heat exchanger network synthesis in the 20th century. Industrial & Engineering Chemistry Research 41:2335–2370

18. Goebel R, Sanfelice RG, Teel AR (2009) Hybrid dynamical systems. IEEE Control Systems Magazine 29(2):28–93
19. Gordon RA (1994) The integrals of Lebesgue, Denjoy, Perron, and Henstock. American Mathematical Society, Providence, RI
20. Hiriart-Urruty JB, Lemaréchal C (1993) Convex Analysis and Minimization Algorithms. Springer, Berlin
21. Horst R (1988) Deterministic global optimization with partition sets whose feasibility is not known: Application to concave minimization, reverse convex constraints, DC-programming, and Lipschitzian optimization. Journal of Optimization Theory and Applications 58(1):11–37
22. Horst R, Tuy H (1996) Global Optimization: Deterministic Approaches, 3rd edn. Springer, Berlin
23. Knüppel O (1994) PROFIL/BIAS–A fast interval library. Computing 53(3-4):277–287
24. Liu J, Liao Lz, Nerode A, Taylor JH (1993) Optimal control of systems with continuous and discrete states. In: Proceedings of 32nd IEEE Conference on Decision and Control, IEEE, pp 2292–2297
25. Lukšan L, Vlček J (2001) Algorithm 811: NDA: Algorithms for nondifferentiable optimization. ACM Transactions on Mathematical Software 27:193–213
26. McCormick GP (1976) Computability of global solutions to factorable nonconvex programs: Part I – Convex underestimating problems. Mathematical Programming 10:147–175
27. McCormick GP (1983) Nonlinear programming: theory, algorithms, and applications. Wiley, New York, NY
28. Mitsos A, Chachuat B, Barton PI (2009) McCormick-based relaxations of algorithms. SIAM Journal on Optimization 20:573–601
29. Moore RE (1979) Methods and Applications of Interval Analysis. SIAM, Philadelphia, PA
30. Moreau L, Aeyels D (2000) Optimization of discontinuous functions: A generalized theory of differentiation. SIAM Journal on Optimization 11:53–69
31. Neumaier A (1990) Interval methods for systems of equations. Cambridge University Press, Cambridge, UK
32. Rockafellar RT (1996) Convex analysis. Princeton University Press, Princeton, NJ
33. Rubinstein RY (1983) Smoothed functionals in stochastic optimization. Mathematics of Operations Research 8:26–33
34. Rudin W (1976) Principles of Mathematical Analysis, 3rd edn. McGraw-Hill, New York, NY

35. Ryoo HS, Sahinidis NV (1996) A branch-and-reduce approach to global optimization. Journal of Global Optimization 8(2):107–138
36. Sahinidis NV (2012) BARON solver manual. `http://gams.com/dd/docs/solvers/baron.pdf`
37. Scott JK, Stuber MD, Barton PI (2011) Generalized McCormick relaxations. Journal of Global Optimization 51(4):569–606
38. Tawarmalani M, Sahinidis NV (2002) Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming. Kluwer Academic Publishers, Dordrecht
39. Türkay M, Grossmann IE (1996) Disjunctive programming techniques for the optimization of process systems with discontinuous investment costs-multiple size regions. Industrial & Engineering Chemistry Research 35:2611–2623
40. Vicente LN, Custódio AL (2012) Analysis of direct searches for discontinuous functions. Mathematical Programming 133(1-2):299–325
41. Zang I (1981) Discontinuous optimization by smoothing. Mathematics of Operations Research 6:140–152
42. Zheng Q (1990) Robust analysis and global minimization of a class of discontinuous functions (I). Acta Mathematicae Applicatae Sinica 6:205–223

A Discussion of sufficient conditions for convergence of the relaxations

Here, three lemmata will be given that present sufficient conditions for Assumption 4 to hold for a given factor v_k and thus can be used in a finite induction argument to establish Assumption 4. In particular, they formalize the discussion in Section 2.4 and show that, to establish Assumption 4, it is sufficient to exclude these cases from occurring. First, overestimation in binary operations is considered. Here, two reasonably strong results can be given. Then, attention will be directed to univariate functions where more restrictive assumptions need to be made.

Lemma 6 *Consider any k such that $n < k \leq m$ where v_k is defined by a summation or multiplication. Consider a nested sequence of intervals $X^l \to X^* = [\mathbf{x}^*, \mathbf{x}^*]$, $X^l \in \mathbb{I}X$, and $X^l \neq X^*$. Suppose Assumption 4 holds for all $i, j < k$. Suppose that v_i and v_j are discontinuous with respect to $\mathbf{x}$ at $\mathbf{x}^*$ and that these discontinuities are introduced at earlier factors $k_i \leq i$ and $k_j \leq j$, i.e., $v_{k_i} = \pi(v_{r_i})$ and $v_{k_j} = \pi(v_{r_j})$, $r_i < k_i$ and $r_j < k_j$. Assume that v_{k_i} and v_{k_j} are the only discontinuous step mappings. Define subsets of X^l as*

$\Xi_i^l = \{\mathbf{x} \in X^l : v_{r_i}(\mathbf{x}) > 0\}$ and $\Xi_j^l = \{\mathbf{x} \in X^l : v_{r_j}(\mathbf{x}) > 0\}$. If there exists a $L \in \mathbb{N}$ so that for all $l > L$,

$$\Xi_i^l \cap \Xi_j^l \neq \emptyset, \quad \Xi_i^l \cap (X^l \backslash \Xi_j^l) \neq \emptyset, \quad (X^l \backslash \Xi_i^l) \cap \Xi_j^l \neq \emptyset, \quad (X^l \backslash \Xi_i^l) \cap (X^l \backslash \Xi_j^l) \neq \emptyset,$$

then Assumption 4 holds for k.

Proof By assumption, there exist four sequences $\{\mathbf{x}_1^l\}, \ldots, \{\mathbf{x}_4^l\}$ converging to $\mathbf{x}^*$ where $\mathbf{x}_1^l \in \Xi_i^l \cap \Xi_j^l$, $\mathbf{x}_2^l \in \Xi_i^l \cap (X^l \backslash \Xi_j^l)$, $\mathbf{x}_3^l \in (X^l \backslash \Xi_i^l) \cap \Xi_j^l$, and $\mathbf{x}_4^l \in (X^l \backslash \Xi_i^l) \cap (X^l \backslash \Xi_j^l)$.

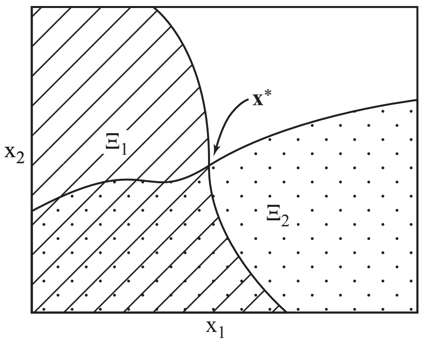
For any X^l with $l > L$, the image of v_{k_i} is $V_{k_i}^l = [0,1]$ as Ξ_i^l is a nonempty strict subset of X^l, $v_{k_i}(\mathbf{x}_q^l) = 1$ for $q = 1,2$ and $v_{k_i}(\mathbf{x}_q^l) = 0$ for $q = 3,4$. Thus, $V_{k_i}^l$ is an exact bound of the range of v_{k_i}. Consider the finite sequence of $s+1$ continuous factors, say $v_{i_1}, \ldots, v_{i_s}, v_i$ with $k_i < i_1 < \ldots < i_s < i$, that maps V_{k_i} to V_i. By assumption, other arguments involved in the definition of the factors $v_{i_1}, \ldots, v_{i_s}, v_i$ are continuous step mappings and, as a result, their corresponding interval bounds converge to degenerate intervals as $l \to \infty$.

Consider factor v_{i_1} and let $[\underline{v}_{i_1}^*, \overline{v}_{i_1}^*] = \lim_{l\to\infty}[\underline{v}_{i_1}^l, \overline{v}_{i_1}^l]$. If this step mapping is a binary operation combining v_{k_i} with a continuous factor, V_{i_1} will converge to a non-degenerate interval and, without loss of generality, $\underline{v}_{i_1}^* = \lim_{l\to\infty} v_{i_1}^l(\mathbf{x}_q^l)$ for $q = 1,2$ and $\overline{v}_{i_1}^* = \lim_{l\to\infty} v_{i_1}^l(\mathbf{x}_q^l)$ for $q = 3,4$. If this step mapping is a univariate operation, Assumption 4 guarantees that V_{i_1} will converge to the exact bounds, i.e., without loss of generality, $\underline{v}_{i_1}^* = \lim_{l\to\infty} v_{i_1}^l(\mathbf{x}_q^l)$ for $q = 1,2$ and $\overline{v}_{i_1}^* = \lim_{l\to\infty} v_{i_1}^l(\mathbf{x}_q^l)$ for $q = 3,4$. Repeating this argument for the factors $v_{i_2}, \ldots, v_{i_s}, v_i$, it follows without loss of generality that $\underline{v}_i^* = \lim_{l\to\infty} v_i^l(\mathbf{x}_q^l)$ for $q = 1,2$ and $\overline{v}_i^* = \lim_{l\to\infty} v_i^l(\mathbf{x}_q^l)$ for $q = 3,4$ where $[\underline{v}_i^*, \overline{v}_i^*] = \lim_{l\to\infty}[\underline{v}_i^l, \overline{v}_i^l]$. It can be argued similarly that, without loss of generality, $\underline{v}_j^* = \lim_{l\to\infty} v_j^l(\mathbf{x}_q^l)$ for $q = 1,3$ and $\overline{v}_j^* = \lim_{l\to\infty} v_j^l(\mathbf{x}_q^l)$ for $q = 2,4$ where $[\underline{v}_j^*, \overline{v}_j^*] = \lim_{l\to\infty}[\underline{v}_j^l, \overline{v}_j^l]$.

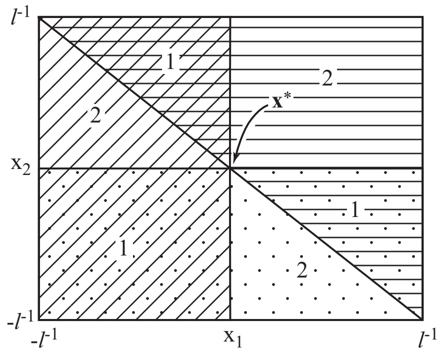
Thus, each combination of the bounds of v_i and v_j is attained in the neighborhood of $\mathbf{x}^*$. In particular, in the case of addition, the sequences $\{v_i(\mathbf{x}_1^l)\}$, $\{v_j(\mathbf{x}_1^l)\}$ and $\{v_i(\mathbf{x}_4^l)\}$, $\{v_j(\mathbf{x}_4^l)\}$ converge to $\underline{v}_i^*$, $\underline{v}_j^*$ and $\overline{v}_i^*$, $\overline{v}_j^*$,respectively. Thus, $[\underline{v}_k^*, \overline{v}_k^*] = [\underline{v}_i^*, \overline{v}_i^*] + [\underline{v}_j^*, \overline{v}_j^*]$ is, in the limit, an exact bound. A similar argument can be presented for the case of multiplication. Here, each combination of lower and upper bounds on v_i and v_j is realized by a different sequence $\{\mathbf{x}_q^l\}$, $q = 1, \ldots, 4$. Thus, Assumption 4 holds for k. □

Remark 5

– Lemma 6 considers the case of adding or multiplying v_i and v_j where v_i and v_j are discontinuous in the limit $\mathbf{x}^*$ and these discontinuities are introduced by exactly one π function each. Then, the dependency problem in interval arithmetic can be mitigated when there exist regions in each interval X^l so that all combination of the lower and upper bounds of the factors v_i and v_j are attained. This can be alternatively



(a) Illustration showing a case where hypothesis of Lemma 6 holds



(b) Counterexample showing when Assumption 4 does not hold. The digits indicate the value of $f(\mathbf{x}) = 1 + \pi(x_1) + \pi(x_2) - \pi(x_1 + x_2)$ in the 6 subsets of $X^l = [-l^{-1}, l^{-1}]^2$, note that $\underline{f} = 0$, $\overline{f} = 3$.

Fig. 6: Illustrations for Assumption 4 when $X \subset \mathbb{R}^2$. The curves indicate discontinuities introduced at previous factors.

expressed as requiring that the intrinsic discontinuities do not coincide in a neighborhood of $\mathbf{x}^*$. A case where this hypothesis of Lemma 6 holds is illustrated in Figure 6 (a).

– A counterexample can be given to show that Lemma 6 cannot be easily extended to the case when more than n intrinsic discontinuities coincide at $\mathbf{x}^* \in \mathbb{R}^n$. To see this, consider $f(\mathbf{x}) = 1 + \pi(x_1) + \pi(x_2) - \pi(x_1 + x_2)$, $X = [-1, 1]^2$ and $X^l = [-l^{-1}, l^{-1}]^2$. As shown in Figure 6 (b), three intrinsic discontinuities coincide at $(0, 0)$. The bounds of f on X^l obtained from interval arithmetic are $\underline{f}^l = 0$ and $\overline{f}^l = 3$. They are not attained for any $\mathbf{x} \in X^l$ and any l and thus Assumption 4 does not hold.
– Also note that, given Assumption 4, the exacerbated dependency problem of interval arithmetic is not acute when there is only one discontinuity present in either v_i or v_j at $\mathbf{x}^*$. This has been exploited in the proof of Lemma 6.
– Lastly, observe that the hypotheses of Lemma 6 cannot be satisfied when $X^l \subset \mathbb{R}$. At most three subsets of X in the vicinity of x^*, $\{x : x < x^*\}$, $\{x : x = x^*\}$ and $\{x : x > x^*\}$, are conceivable where v_i and v_j could attain their lower and upper bounds. To guarantee that Assumption 4 holds for v_k, the interval arithmetic for $v_i + v_j$ or $v_i v_j$ needs to combine the bounds in such a way that v_k attains both its lower and upper bound. However, it is easy to conceive counterexamples where this is not true, e.g., see the discussion prior to Lemma 6.

Though it was pointed out that there are counterexamples restricting the generalization of Lemma 6 when more than 2 intrinsic discontinuities coincide at $\mathbf{x}^*$ in $\mathbb{R}^2$, a generalization is possible to n intrinsic discontinuities coinciding in $\mathbb{R}^n$.

Lemma 7 *Consider any k such that $n < k \leq m$ where v_k is defined by summation or multiplication. Suppose Assumption 4 holds for all $i, j < k$. Consider a nested sequence of intervals $X^l \to X^* = [\mathbf{x}^*, \mathbf{x}^*]$, $X^l \in \mathbb{I}X$, $X^l \neq X^*$. Suppose that v_i and v_j are discontinuous with respect to $\mathbf{x}$ at $\mathbf{x}^*$ and that these discontinuities are introduced by $q \leq n$ earlier factors $k_1, \ldots, k_q$, i.e., $v_{k_{\tilde{q}}} = \pi(v_{r_{\tilde{q}}})$ with $v_{r_{\tilde{q}}}(\mathbf{x}^*) = 0$ for $\tilde{q} = 1, \ldots, q$. Assume that $v_{r_{\tilde{q}}}$ is differentiable with respect to $\mathbf{x}$ at $\mathbf{x}^*$, for all $\tilde{q} = 1, \ldots, q$, and denote the gradient of $v_{r_{\tilde{q}}}$ at $\mathbf{x}^*$ as $\boldsymbol{\nabla} v_{\tilde{q}}$. If $\boldsymbol{\nabla} v_1, \ldots, \boldsymbol{\nabla} v_q$ are linearly independent, then Assumption 4 holds for k.*

Proof Define subsets of X^l as $\Xi^l_{\tilde{q}} = \{\mathbf{x} \in X^l : v_{r_{\tilde{q}}} > 0\}$, $\tilde{q} = 1, \ldots, q$. Requiring linear independence of $\boldsymbol{\nabla} v_1, \ldots, \boldsymbol{\nabla} v_q$ is a sufficient condition for the existence of 2^q nonempty subsets of X^l that realize all combinations of $\Xi^l_{\tilde{q}}$ with $\Xi^l_{\hat{q}}$ or $X^l \backslash \Xi^l_{\hat{q}}$, $\hat{q} = 1, \ldots, q$, $\hat{q} \neq \tilde{q}$, for all $l > L$ for some $L \in \mathbb{N}$. Thus, the argument used in the proof of Lemma 6 can be extended to show that each possible combination of the bounds on intermediate factors is indeed realized. $\square$

Lemma 8 *Consider any k such that $n < k \leq m$ where v_k is defined by $v_k = \pi(v_i)$. Consider a nested sequence of intervals $X^l \to X^* = [\mathbf{x}^*, \mathbf{x}^*]$, $X^l \in \mathbb{I}X$, $X^l \neq X^*$. Suppose either*

1. *that $v_i(\mathbf{x}^*) = 0$ and that for all $l > 0$ there exists a $\mathbf{x}^\dagger_l \in X^l$ and a $\varepsilon_l > 0$ so that $v_i(\mathbf{x}^\dagger_l) = \varepsilon_l$,*
2. *that there exists a $L_1 > 0$ so that $\overline{v}^l_i \leq 0$ for all $l \geq L_1$, or*
3. *that there exists a $L_2 > 0$ so that $\underline{v}^l_i > 0$ for all $l \geq L_2$.*

Then, Assumption 4 holds for k.

Proof Consider Case 1. By assumption, $\underline{v}^l_k = 0$ and $\overline{v}^l_k = 1$, $\forall l$ so that $\lim_{l\to\infty} [\underline{v}^l_k, \overline{v}^l_k] = [0, 1]$. Furthermore, it holds that

$$\begin{aligned}
[\lim_{l\to\infty} \inf_{\mathbf{x}\in X^l} v_k(\mathbf{x}), \lim_{l\to\infty} \sup_{\mathbf{x}\in X^l} v_k(\mathbf{x})] &= [\lim_{l\to\infty} \inf_{\mathbf{x}\in X^l} \pi(v_i(\mathbf{x})), \lim_{l\to\infty} \sup_{\mathbf{x}\in X^l} \pi(v_i(\mathbf{x}))] \\
&= [\pi(v_i(\mathbf{x}^*)), \lim_{l\to\infty} \pi(v_i(\mathbf{x}^\dagger_l))] \\
&= [\pi(0), \lim_{l\to\infty} \pi(\varepsilon_l)] = [0, 1].
\end{aligned}$$

Consider Case 2. By assumption, $[\underline{v}^l_k, \overline{v}^l_k] = [0, 0]$ for all $l \geq L_1$. Thus, $v_k(\mathbf{x}) = 0$ for all $\mathbf{x} \in X^{L_1}$ so that $[\lim_{l\to\infty} \inf_{\mathbf{x}\in X^l} v_k(\mathbf{x}), \lim_{l\to\infty} \sup_{\mathbf{x}\in X^l} v_k(\mathbf{x})] = [0, 0] = \lim_{l\to\infty} [\underline{v}^l_k, \overline{v}^l_k]$.

Consider Case 3. By assumption, $[\underline{v}_k^l, \overline{v}_k^l] = [1,1]$ for all $l \geq L_1$. Thus, $v_k(\mathbf{x}) = 1$ for all $\mathbf{x} \in X^{L_2}$ so that $[\lim_{l\to\infty} \inf_{\mathbf{x}\in X^l} v_k(\mathbf{x}), \lim_{l\to\infty} \sup_{\mathbf{x}\in X^l} v_k(\mathbf{x})] = [1,1] = \lim_{l\to\infty} [\underline{v}_k^l, \overline{v}_k^l]$.

Thus, Eq. (2) and, hence, Assumption 4 hold for factor k. □

Lemma 9 *Consider a nested sequence of intervals $X^l \to X^*$, $X^l \in \mathbb{I}X$, $X^l \neq X^*$ and a continuous function $f : X \to \mathbb{R}$. Then,*

$$\lim_{l\to\infty} \inf_{\mathbf{x}\in X^l} f(\mathbf{x}) = \inf_{\mathbf{x}\in X^*} f(\mathbf{x}) \qquad \textit{and} \qquad \lim_{l\to\infty} \sup_{\mathbf{x}\in X^l} f(\mathbf{x}) = \sup_{\mathbf{x}\in X^*} f(\mathbf{x}).$$

Proof Fix $\varepsilon > 0$. Let $\mathbf{x}^*_{\min} \in \arg\min_{\mathbf{x}\in X^*} f(\mathbf{x})$, the infimum is attained since X^* is compact and f is continuous on X^*. Since $X^l \subset X$ is compact and f is continuous on X, f is uniformly continuous on X^l. Uniform continuity of f implies that $\exists\delta > 0$ so that $|f(\mathbf{x}) - f(\mathbf{y})| < \varepsilon$ for all $\mathbf{x},\mathbf{y} \in X^l$ for which $\|\mathbf{x}-\mathbf{y}\| < \delta\sqrt{n}$. Convergence of X^l to X^* implies that there is a $L > 0$ so that $d_H(X^l, X^*) < \delta$ for all $l > L$. By definition of the Hausdorff metric, $\underline{x}_i^l > \underline{x}_i^* - \delta$ and $\overline{x}_i^l < \overline{x}_i^* + \delta$ for all $l > L$ and $i = 1,\ldots,n$. Thus, $f(\mathbf{x}^\dagger) + \varepsilon > f(\mathbf{x}^\ddagger)$ where $\mathbf{x}^\dagger \in X^l \backslash X^*$ and $\mathbf{x}^\ddagger \in \partial X^*$ with ∂X^* denoting the boundary of X^*. By definition, $f(\mathbf{x}) \geq f(\mathbf{x}^*_{\min})$, $\forall \mathbf{x} \in X^*$ so that $f(\mathbf{x}^\ddagger) \geq f(\mathbf{x}^*_{\min})$. As a result, $f(\mathbf{x}) + \varepsilon > f(\mathbf{x}^*_{\min})$ for all $\mathbf{x} \in X^l$ with $l > L$. Since $X^l \supset X^*$, $\inf_{\mathbf{x}\in X^l} f(\mathbf{x}) \leq f(\mathbf{x}^*_{\min})$ for all l. ε is arbitrary so that $\lim_{l\to\infty} \inf_{\mathbf{x}\in X^l} f(\mathbf{x}) = \inf_{\mathbf{x}\in X^*} f(\mathbf{x})$. An analogous argument can be made to show that $\lim_{l\to\infty} \sup_{\mathbf{x}\in X^l} f(\mathbf{x}) = \sup_{\mathbf{x}\in X^*} f(\mathbf{x})$.

Lemma 10 *Consider any k such that $n < k \leq m$ where v_k is defined by a continuous univariate intrinsic function φ_k. Suppose Assumption 4 holds for all $i < k$. Consider a nested sequence of intervals $X^l \to X^* = [\mathbf{x}^*, \mathbf{x}^*]$, $X^l \in \mathbb{I}X$, $X^l \neq X^*$. Let $[\underline{v}_i^*, \overline{v}_i^*] = \lim_{l\to\infty} [\underline{v}_i^l, \overline{v}_i^l]$. Then, Assumption 4 holds for k if*

$$\min\{\varphi_k(\underline{v}_i^*), \varphi_k(\overline{v}_i^*)\} = \underline{\Phi}_k([\underline{v}_i^*, \overline{v}_i^*]) \quad \textit{and} \quad \max\{\varphi_k(\underline{v}_i^*), \varphi_k(\overline{v}_i^*)\} = \overline{\Phi}_k([\underline{v}_i^*, \overline{v}_i^*]).$$

Proof First, suppose that v_i is continuous with respect to $\mathbf{x}$ at $\mathbf{x}^*$. Then, $[\underline{v}_i^*, \overline{v}_i^*]$ is a degenerate interval. Since Φ_k is an interval extension, $\Phi_k([\underline{v}_i^*, \overline{v}_i^*])$ is also a degenerate interval and, hence, Eq. (2) holds.

Next, suppose that v_i is not continuous with respect to $\mathbf{x}$ at $\mathbf{x}^*$. Since Assumption 4 holds for factor i, it follows that $\lim_{l\to\infty} [\underline{v}_i^l, \overline{v}_i^l] = [\lim_{l\to\infty} \inf_{\mathbf{x}\in X^l} v_i(\mathbf{x}), \lim_{l\to\infty} \sup_{\mathbf{x}\in X^l} v_i(\mathbf{x})]$. Consider the sequence $V_i^l = [\underline{v}^l, \overline{v}^l]$ converging to $V_i^* = [\underline{v}^*, \overline{v}^*]$. According to Lemma 9, it holds that $\lim_{l\to\infty} \inf_{z\in V_i^l} \varphi_k(z) = \inf_{z\in V_i^*} \varphi_k(z)$ and that $\lim_{l\to\infty} \sup_{z\in V_i^l} \varphi_k(z) = \sup_{z\in V_i^*} \varphi_k(z)$. The hypothesis of the lemma imply furthermore that $\underline{\Phi}_k([\underline{v}_i^*, \overline{v}_i^*]) = \inf_{z\in[\underline{v}_i^*, \overline{v}_i^*]} \varphi_k(z)$ and that $\overline{\Phi}_k([\underline{v}_i^*, \overline{v}_i^*]) = \sup_{z\in[\underline{v}_i^*, \overline{v}_i^*]} \varphi_k(z)$. Therefore it follows that

$$\begin{aligned}[\lim_{l\to\infty} \inf_{\mathbf{x}\in X^l} v_k(\mathbf{x}), \lim_{l\to\infty} \sup_{\mathbf{x}\in X^l} v_k(\mathbf{x})] &= [\lim_{l\to\infty} \inf_{z\in V_i^l} \varphi_k(z), \lim_{l\to\infty} \sup_{z\in V_i^l} \varphi_k(z)] \\ &= [\inf_{z\in V_i^*} \varphi_k(z), \sup_{z\in V_i^*} \varphi_k(z)] = \Phi_k([\underline{v}_i^*, \overline{v}_i^*]) = [\underline{v}_k^*, \overline{v}_k^*],\end{aligned}$$

i.e., Eq. (2) holds and, hence, Assumption 4 is established for factor k. □

Remark 6 An example of a class of univariate intrinsic functions φ that can meet the hypotheses of Lemma 10 are monotone functions. However, the specific implementation of Φ will dictate if φ indeed meets the hypotheses of Lemma 10.

B More general convergence results for branch-and-bound algorithm

In Section 3, it was assumed that f is either lower semi-continuous or attains its minimum on D. Results are outlined below that hold even when these assumptions are generalized.

Remark 7 When the assumption that f is lower semi-continuous is dropped in Theorem 9, then one cannot appeal to Theorem 8. However, with Remark 3 in mind, one can argue that $\lim_{q\to\infty}\beta(X^{l_q}) = \lim_{q\to\infty}\inf_{\mathbf{x}\in X^{l_q}} f(\mathbf{x}) \geq \min\{\lim_{q\to\infty}\inf_{\mathbf{x}\in X^{l_q}\setminus\{\mathbf{x}^*\}} f(\mathbf{x}), f(\mathbf{x}^*)\} \geq \min\{\liminf_{\mathbf{x}\to\mathbf{x}^*} f(\mathbf{x}), f(\mathbf{x}^*)\}$, which is sufficient to show that the lower bounding operation is strongly consistent.

Note that Remark 7 does not allow for the argument $\lim_{q\to\infty}\beta(X^{l_q}) = \inf_{\mathbf{x}\in X^*\cap D} f(\mathbf{x})$, and consequently $\lim_{k\to\infty}\beta_k = \inf_{\mathbf{x}\in D} f(\mathbf{x})$, when f is not assumed to be lower semi-continuous. In particular, there may be an infinitely decreasing sequence of nested intervals X^l so that there exists a $\mathbf{y}\in\partial D$ with $\mathbf{y}\in \operatorname{int} X^l$, $\forall l$, i.e., all partition elements contain an element of the boundary of the feasible set in its interior. Suppose that $f(\mathbf{y}) = \inf_{\mathbf{x}\in D} f(\mathbf{x})$. Thus, it is conceivable that there exists a $\varepsilon > 0$ and a sequence $\{\mathbf{z}^l\}$ with $\mathbf{z}^l \notin D$, $\mathbf{z}^l \in X^l$, $\forall l$ so that $f(\mathbf{z}^l) < f(\mathbf{y}) - \varepsilon$. As a result, $\lim_{l\to\infty}\beta(X^l) \leq f(\mathbf{y}) - \varepsilon$.

To avoid this complication, another assumption is introduced.

Assumption 6 Suppose $f(\mathbf{y}) \geq \inf_{\mathbf{x}\in D} f(\mathbf{x})$, $\forall \mathbf{y}\in X : \mathbf{y}\notin D$.

This assumption can be satisfied by reformulating f as a penalty function, e.g., minimizing $\tilde{f}$ with

$$\tilde{f}(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \text{if } \mathbf{x}\in D, \\ \overline{f}(D), & \text{otherwise}, \end{cases}$$

where $\overline{f}(D)$ denotes an upper bound, e.g., derived from interval analysis, of f on D.

Remark 8 When the assumption that f attains its minimum on D in Theorem 10 is removed and Assumption 6 holds, one can still argue that $\beta = \inf_{\mathbf{x}\in D} f(\mathbf{x})$ using Theorem 8 and Remark 3. However, the set of minimizers of f on D, $\arg\min_{\mathbf{x}\in D} f(\mathbf{x})$, is not defined in this case. Instead, consider the set

$$\arg\inf_{\mathbf{x}\in D} f(\mathbf{x}) \equiv \left\{\mathbf{x}\in D : \exists\{\mathbf{z}^l\}\subset D \text{ with } \lim_{l\to\infty}\mathbf{z}^l = \mathbf{x} \text{ and } \lim_{l\to\infty} f(\mathbf{z}^l) = \inf_{\mathbf{z}\in D} f(\mathbf{z})\right\}. \tag{3}$$

In this case $X_{\min} \subset \arg\inf_{\mathbf{x}\in D} f(\mathbf{x})$. This can be shown as follows:

Assume that the algorithm does not terminate after a finite number of steps. Consider the sequence of lower bounds $\{\beta_k\}$ with $\mathbf{x}^k_{\min}$, L_k and X^{L_k} as defined previously. From the construction of the algorithm it follows that $\{\beta_k\}$ is a nondecreasing sequence with $\beta_k \le \inf_{\mathbf{x}\in D} f(\mathbf{x})$. Hence, $\beta = \lim_{k\to\infty}\beta_k$ exists and $\beta \le \inf_{\mathbf{x}\in D} f(\mathbf{x})$. Let $\mathbf{x}^\dagger_{\min}$ denote an element of the set of accumulation points of the sequence $\{\mathbf{x}^k_{\min}\}$ and let $\{\mathbf{x}^{k_r}_{\min}\}$ be a subsequence of $\{\mathbf{x}^k_{\min}\}$ with subsequential limit $\mathbf{x}^\dagger_{\min}$. Since the partition subdivision is exhaustive and the selection operation is bound improving, a finite number of partition elements is visited in each iteration only. Consequently, a decreasing subsequence of successively refined partition elements $\{X^{q'}\} \subset \{X^{L_{k_r}}\}$ exists such that $\lim_{q'\to\infty} X^{q'} = \{\mathbf{x}^\dagger_{\min}\}$. Since the lower bounding operation is strongly consistent, there exists a subsequence $\{X^q\} \subset \{X^{q'}\}$ such that $\lim_{q\to\infty}\beta(X^q) \ge \min\{\liminf_{\mathbf{x}\to\mathbf{x}^\dagger_{\min}} f(\mathbf{x}), f(\mathbf{x}^\dagger_{\min})\}$. The "deletion by infeasibility" rule is certain in the limit so that $\mathbf{x}^\dagger_{\min} \in D$. Thus, $\inf_{\mathbf{x}\in D} f(\mathbf{x}) \ge \beta \ge \min\{\liminf_{\mathbf{x}\to\mathbf{x}^\dagger_{\min}} f(\mathbf{x}), f(\mathbf{x}^\dagger_{\min})\}$. By assumption, $f(\mathbf{y}) \ge \inf_{\mathbf{x}\in D} f(\mathbf{x})$ when $\mathbf{y} \notin D$ so that

$$\inf_{\mathbf{x}\in D} f(\mathbf{x}) = \min\{\liminf_{\mathbf{x}\to\mathbf{x}^\dagger_{\min}} f(\mathbf{x}), f(\mathbf{x}^\dagger_{\min})\} = \beta.$$

Thus, the result follows.

Remark 9 Assumption 5, which implicitly presumes that f attains its minimum on D, is used in Theorem 11. The latter can be modified when the minimum of f on D is not attained: define $\tilde{\mathbf{x}}_{\min} \in D$ as the limit of a sequence $\{\mathbf{x}^l\} \subset D$ with $\lim_{l\to\infty} f(\mathbf{x}^l) = f^*$. Suppose that, for every $\varepsilon > 0$, there exists a $\delta > 0$ and a $\mathbf{x} \in D$ for which $\|\mathbf{x} - \tilde{\mathbf{x}}_{\min}\| < \delta$, $\mathbf{x} \ne \tilde{\mathbf{x}}_{\min}$, $f(\mathbf{x}) \le f^* + \varepsilon$ hold. Under this assumption, consistency of the lower bounding operation can be argued following a proof similar to the one of Theorem 11.

Remark 10 In Theorem 12 it was assumed that f is lower semi-continuous. This assumption was utilized therein to assert that sublevel sets of f are closed. A similar statement is not possible when the assumption of lower semi-continuity of f is dropped as they are equivalent. Consider a discontinuous functions with the following property: there exist two sequences $\{\mathbf{y}^l\}, \{\mathbf{z}^l\} \subset D$ with limits $\mathbf{y}^* \ne \mathbf{z}^*$, respectively, so that $\lim_{l\to\infty} f(\mathbf{y}^l) = f^* = \lim_{l\to\infty} f(\mathbf{z}^l)$ and let $f(\mathbf{y}^*) = f^* \ne f(\mathbf{z}^*)$. The branch-and-bound algorithm is not able to fathom any partition element that contains an infinite number of elements of $\{\mathbf{z}^l\}$. Consequently, $\mathbf{y}^*$ and $\mathbf{z}^*$ are accumulation points of $\{\mathbf{x}^k\}$, whereas, in the strict sense, only $\mathbf{y}^*$ solves (P). However, $\mathbf{z}^*$ is in the set $\arg\inf_{\mathbf{x}\in D} f(\mathbf{x})$ as defined by Equation (3). Using the argument presented in this remark and asserting Assumption 6, one can show that, for any accumulation point $\mathbf{x}^\dagger$ of $\{\mathbf{x}^k\}$, $\mathbf{x}^\dagger \in \arg\inf_{\mathbf{x}\in D} f(\mathbf{x})$ holds.