J Glob Optim (2012) 52:1–28
DOI 10.1007/s10898-011-9685-2

Convergence rate of McCormick relaxations

Agustín Bompadre · Alexander Mitsos

Received: 8 July 2010 / Accepted: 1 February 2011 / Published online: 18 February 2011

Abstract Theory for the convergence order of the convex relaxations by McCormick (Math Program 10(1):147–175, 1976) for factorable functions is developed. Convergence rules are established for the addition, multiplication and composition operations. The convergence order is considered both in terms of pointwise convergence and of convergence in the Hausdorff metric. The convergence order of the composite function depends on the convergence order of the relaxations of the factors. No improvement in the order of convergence compared to that of the underlying bound calculation, e.g., via interval extensions, can be guaranteed unless the relaxations of the factors have pointwise convergence of high order. The McCormick relaxations are compared with the αBB relaxations by Floudas and coworkers (J Chem Phys, 1992, J Glob Optim, 1995, 1996), which guarantee quadratic convergence. Illustrative and numerical examples are given.

Keywords Nonconvex optimization · Global optimization · Convex relaxation · McCormick · AlphaBB · Interval extensions

1 Introduction

Some of the most successful methods for global optimization of nonconvex programs, e.g., [38], rely on the construction of convex/concave relaxations of the objective and constraints. To ensure finite termination, these relaxations must converge to the nonconvex functions in the limit, i.e., as the diameter of the host sets vanishes (reduction to singleton), e.g., through

A. Bompadre
Graduate Aerospace Laboratories, California Institute of Technology, Pasadena, CA 91125, USA
e-mail: abompadr@alum.mit.edu

A. Mitsos (✉)
Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
e-mail: amitsos@alum.mit.edu

branching or subdivision. To limit branching/subdivision and therefore computational time it is very important to have as tight as possible relaxations. Constructing the convex envelope is in general as hard as solving the global optimization problem. Therefore, systematic methods for the construction of underestimators have been proposed.

One of the most established techniques are the McCormick relaxations [24], which consider a factorization of the functions to a finite set of addition, multiplication and composition operations. Using relaxations for these factors, the relaxation of the function is constructed via a small set of rules. A similar method is to introduce additional optimization variables for each factor; this technique is used among others by the well-known solver BARON by Tawarmalani and Sahinidis [33,38]. Note also the recent solver COUENNE [10]. An alternative approach is used by the αBB and γBB relaxations, developed by Floudas and coworkers [1–6,9,18–23]. These latter methods estimate the Hessian of the nonconvex functions and add a known convex/concave term to the function.

In addition to ensuring convergence in the limit, it is important to consider also the order of convergence. This concept from interval extensions [8,28] essentially compares the rate of convergence of the estimation error to the rate of the decrease of the range of the function. To motivate the importance of the order of convergence, consider the classical branch-and-bound (B&B) algorithm applied to minimization problems. B&B employs local solutions as upper bounds and relaxations as lower bounds. The B&B algorithm is inherently worst-case exponential in the number of variables and to perform well in practice it must fathom the majority of the nodes as early as possible. Nodes in the B&B tree are fathomed when their lower bound is higher than the best upper bound. Since the lower bounds are lower than the optimal objective value confined to this node, nodes with objective value close to the optimal objective value can only be fathomed when the relaxations are very tight. The order of convergence provides a criterion on how small the diameter of the node needs to be to achieve convergence within a prescribed tolerance. Note also that since the relaxations often become weaker with the number of variables, worst-case complexity is worse than exponential, and thus fast convergence is even more important. See also the discussion on the so-called cluster effect [15,30] and a very recent article [34] considering convergence of geometric B&B methods. In the following, the convergence order is formally defined, extending the well-known results from interval arithmetic. The main focus of this article is to consider the McCormick relaxations and determine how the convergence order propagates through addition, multiplication and composition.

In Sect. 2 basic concepts are repeated for the sake of completeness, followed by the formalization of convergence order in Sect. 3. In Sect. 4 lower and upper bounds are established for the convergence order for the McCormick relaxations. The basic assumption made for the bounds developed is Lipschitz continuity of the factors. Moreover, simple illustrative examples are given, demonstrating that the developed bounds are sharp. In Sect. 5 the known (quadratic) convergence order of the αBB relaxations is formalized. Section 6 uses the results from Sect. 5 to prove a positive result for the convergence order of envelopes. Finally, in Sect. 7 numerical examples are presented comparing the two alternative methods in terms of convergence, and in Sect. 8 conclusions and potential for future work are discussed.

2 Basic concepts

Definition 1 (*Relaxation of Functions*) Given a convex set $Z \subset \mathbb{R}^{n_z}$ and a function $h : Z \to \mathbb{R}$, a convex function $h^u : Z \to \mathbb{R}$ is a *convex relaxation* (or *convex underestimator*) of h on Z if

$$h^u(\mathbf{z}) \leq h(\mathbf{z}), \quad \forall \mathbf{z} \in Z$$

and a concave function $h^o : Z \to \mathbb{R}$ is a *concave relaxation* (or *concave overestimator*) of h on Z if

$$h^o(\mathbf{z}) \geq h(\mathbf{z}), \quad \forall \mathbf{z} \in Z.$$

The *convex envelope* $h^{u,env} : Z \to \mathbb{R}$ of h on Z is a convex relaxation of h on Z such that for any convex relaxation h^u of h on Z

$$h^u(\mathbf{z}) \leq h^{u,env}(\mathbf{z}), \quad \forall \mathbf{z} \in Z.$$

Similarly, the *concave envelope* $h^{o,env} : Z \to \mathbb{R}$ of h on Z is a concave relaxation of h on Z such that for any concave relaxation h^o of h on Z

$$h^{o,env}(\mathbf{z}) \leq h^o(\mathbf{z}), \quad \forall \mathbf{z} \in Z.$$

Definition 2 (*Diameter of a Set*) Let $Z \subset \mathbb{R}^{n_z}$. The *diameter* of Z, denoted $w(Z)$ is the maximal distance between two points in Z

$$w(Z) = \sup_{\mathbf{z}_1,\mathbf{z}_2 \in Z} \|\mathbf{z}_1 - \mathbf{z}_2\|,$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^{n_z}$.

Definition 3 (*Lipschitz Continuous Function*) A function $f : Z \to \mathbb{R}$ is a *Lipschitz continuous function* with *Lipschitz constant* M if, for any two points $\mathbf{z}_1$, $\mathbf{z}_2$ of Z, it follows that $|f(\mathbf{z}_1) - f(\mathbf{z}_2)| \leq M\|\mathbf{z}_1 - \mathbf{z}_2\|$, and M is the smallest value for which the inequality holds.

2.1 Interval extensions

Let $\mathbb{IR}$ denote the set of closed intervals of $\mathbb{R}$. We define the Hausdorff metric between intervals of $\mathbb{IR}$ as follows.

Definition 4 Let $X = \left[x^L, x^U\right]$ and $Y = \left[y^L, y^U\right]$ be two bounded intervals in $\mathbb{IR}$. The *Hausdorff metric* $q(X, Y)$ is given by:

$$q(X, Y) \equiv \max\left\{\left|x^L - y^L\right|, \left|x^U - y^U\right|\right\}.$$

The following is an equivalent definition of the Hausdorff metric between intervals (see, e.g., [29]).

Proposition 1 *Let* $X = \left[x^L, x^U\right]$ *and* $Y = \left[y^L, y^U\right]$ *be two bounded intervals in* $\mathbb{IR}$*. The Hausdorff metric* $q(X, Y)$ *is equal to*

$$q(X, Y) = \max\left\{\sup_{x\in X}\inf_{y\in Y}|x - y|, \sup_{y\in Y}\inf_{x\in X}|x - y|\right\}. \tag{1}$$

Definition 5 (*Image and Inclusion Function*) Consider a continuous function $h : Z \to \mathbb{R}$, where Z is an n_z-dimensional interval defined as

$$Z = \left[z_1^L, z_1^U\right] \times \cdots \times \left[z_{n_z}^L, z_{n_z}^U\right] = \left[\mathbf{z}^L, \mathbf{z}^U\right].$$

The image of Z under h is denoted by the scalar interval $\bar{h}(Z) = \left[h^L(Z), h^U(Z)\right]$. Consider also an interval-valued function

$$\begin{aligned} H : Z &\to \mathbb{IR} \\ Y &\mapsto \left[H^L(Y), H^U(Y)\right]. \end{aligned}$$

H is an *inclusion function* for h on Z if the following relation holds

$$\left[h^L(Y), h^U(Y)\right] = \bar{h}(Y) \subset H(Y) = \left[H^L(Y), H^U(Y)\right], \quad \forall Y \in \mathbb{IR}^{n_z} : Y \subset Z.$$

The natural interval extension is an example of such an inclusion function [28,31]. The tightness of inclusion functions can be quantified using the Hausdorff metric $q(\bar{h}(Z), H(Z))$. It is well-known that the natural interval extensions have first-order convergence rate (linear convergence), while there are different methods, such as Taylor models (standard or optimally-centered forms) with second-order convergence rate (quadratic convergence) [8]. The latter schemes are typically more expensive to evaluate. To achieve convergence of the inclusion function over the entire range of the function, subdivision can be employed. A formal definition of convergence order is given in the following.

2.2 Convex relaxation

Many deterministic global optimization algorithms rely on the construction of convex relaxations. Given a nonlinear program involving nonconvex functions $\mathbf{g} : Z \to \mathbb{R}^m$, with $Z = [\mathbf{z}^L, \mathbf{z}^U] \subset \mathbb{R}^n$ the goal is to construct a convex relaxation, i.e., a program with convex constraints and a convex objective function, whose optimal objective value underestimates the optimal solution value of the nonconvex NLP. Convex and concave envelopes or tight relaxations are known for a variety of simple nonlinear terms [3,35,37] and these can be used for the construction of convex and concave relaxations. Several methods have been proposed, e.g, [4,16,24,35], which all rely on a few key ideas and elements. McCormick's theorems [24] enables the relaxation of factorable functions. Floudas and coworkers [1–6,9,18–23] have proposed convex relaxations for twice continuously differentiable functions by the addition of a simple, sufficiently negative function that is known to be convex.

3 Convergence rate of estimators

We are interested in studying the convergence of different convex and concave relaxations. To do so, we formalize the concept of approximating functions on intervals.

Definition 6 Let $Z \subset \mathbb{R}^n$ be a nonempty convex set, and let $f : Z \to \mathbb{R}$ be a continuous function. Assume that, for every interval $Y \in \mathbb{IR}^n, Y \subset Z$, we know two functions $f_Y^u, f_Y^o : Y \to \mathbb{IR}$ such that

1. the function f_Y^u is a convex underestimator of f in Y,
2. the function f_Y^o is a concave overestimator of f in Y.

We call the set of functions $\left(f_Y^u, f_Y^o\right)_{Y \subset Z}$ a *scheme of estimators* of f in Z. We call such a scheme continuous if f_Y^u, f_Y^o are continuous for all Y.

A scheme of estimators of a function f defines an inclusion function for f in a natural way.

Definition 7 Let $Z \subset \mathbb{R}^n$ be a nonempty convex set, and let $f : Z \to \mathbb{R}$ be a continuous function. Let $\left(f_Y^u, f_Y^o\right)_{Y \subset Z}$ be a scheme of estimators of f in Z. The inclusion function H_f associated to this scheme is as follows.

$$H_f : Y \in \mathbb{IR}^n, \quad Y \subset Z \to \mathbb{IR}$$
$$H_f(Y) = \left[\inf_{\mathbf{z} \in Y} f_Y^u(\mathbf{z}), \sup_{\mathbf{z} \in Y} f_Y^o(\mathbf{z})\right].$$

We next define the order of Hausdorff convergence of an inclusion function (cf. [31]).

Definition 8 Let $f : Z \subset \mathbb{R}^n \to \mathbb{R}$ be a continuous function, and let H_f be an inclusion function of f on Z. The inclusion function H_f has *Hausdorff convergence of order* $\beta > 0$ if there exists a constant $\tau > 0$ such that, for any interval $Y \in \mathbb{IR}^n$, $Y \subset Z$,

$$q\left(\bar{f}(Y), H_f(Y)\right) \leq \tau w(Y)^{\beta}.$$

Note that the constants τ and β depend on Z but not on the intervals Y. An equivalent definition of order of Hausdorff convergence (also given in [31]) is the following.

Definition 9 Let $f : Z \subset \mathbb{R}^n \to \mathbb{R}$ be a continuous function, and let H_f be an inclusion function of f on Z. The inclusion function H_f has Hausdorff convergence of order $\beta > 0$ if there exists a constant $\hat{\tau} > 0$ such that, for any interval $Y \in \mathbb{IR}^n$, $Y \subset Z$,

$$w(H(Y)) - w(\bar{f}(Y)) \leq \hat{\tau} w(Y)^{\beta}.$$

We note that this second definition of Hausdorff convergence order appears with a typo in recent articles [11,12,27].

We say that a scheme of estimators $\left(f_Y^u, f_Y^o\right)$ of f has Hausdorff convergence of order β when its associated inclusion function has Hausdorff convergence of order β. This notion of convergence bounds the distance between the infima of f and f_Y^u on Y and the suprema of f and f_Y^o on Y. However, it does not give much information about the difference of f with f_Y^u and f_Y^o for given points in Y. We next introduce a stronger notion of convergence based on the maximum difference of f with f_Y^u and f_Y^o on all points of Y.

Definition 10 Let $f : Z \subset \mathbb{R}^n \to \mathbb{R}$ be a continuous function. Let $\left(f_Y^u, f_Y^o\right)_{Y \subset Z}$ be a scheme of estimators of f in Z. The scheme has *pointwise convergence of order* γ if there exists a constant $\tau > 0$ such that, for any interval $Y \in \mathbb{IR}^n$, $Y \subset Z$,

$$\sup_{\mathbf{z} \in Y} |f(\mathbf{z}) - f_Y^u(\mathbf{z})| \leq \tau w(Y)^{\gamma},$$

and

$$\sup_{\mathbf{z} \in Y} |f(\mathbf{z}) - f_Y^o(\mathbf{z})| \leq \tau w(Y)^{\gamma}.$$

In the rest of the paper we will analyze the behavior of schemes of estimators under these two definitions of convergence (Definitions 8 and 10). The next theorem shows that pointwise convergence is indeed stronger than Hausdorff convergence, that is, the inclusion function associated with a scheme with pointwise convergence of a certain order has also Hausdorff convergence of the same order.

Theorem 1 *Let $\left(f_Y^u, f_Y^o\right)_{Y \subset Z}$ be a scheme of estimators of a continuous function f in Z. If the scheme has pointwise convergence of order γ, its associated inclusion function H_f has Hausdorff convergence of order $\beta \geq \gamma$.*

Proof Let $\tau > 0$ be the constant of Definition 10. Let Y be an interval of Z. Let $\mathbf{z}_Y^* \in Y$ and $\hat{\mathbf{z}}_Y^* \in Y$ be points where f and f_Y^u attain their minimum in Y respectively.

Since f attains its minimum at $\mathbf{z}_Y^*$

$$f\left(\mathbf{z}_Y^*\right) \leq f\left(\hat{\mathbf{z}}_Y^*\right),$$

and therefore

$$\inf f(Y) - \inf f^u(Y) = f\left(\mathbf{z}_Y^*\right) - f^u\left(\hat{\mathbf{z}}_Y^*\right) \leq f\left(\hat{\mathbf{z}}_Y^*\right) - f^u\left(\hat{\mathbf{z}}_Y^*\right).$$

The scheme has pointwise convergence of order γ. Therefore, by Definition 10,

$$f\left(\hat{\mathbf{z}}_Y^*\right) - f^u\left(\hat{\mathbf{z}}_Y^*\right) \leq \tau w(Y)^\gamma.$$

Combining the last two inequalities we obtain

$$\inf f(Y) - \inf f^u(Y) \leq \tau w(Y)^\gamma.$$

Note also that $0 \leq \inf f(Y) - \inf f^u(Y)$ since f^u underestimates f. Then,

$$0 \leq \inf f(Y) - \inf f^u(Y) \leq \tau w(Y)^\gamma.$$

A similar argument shows that $0 \leq \sup f^o(Y) - \sup f(Y) \leq \tau w(Y)^\gamma$, and thus H_f has at least Hausdorff convergence of order $\beta = \gamma$. □

On the other hand, pointwise convergence is more restrictive than Hausdorff convergence: a scheme of estimators of a smooth, nonlinear function, cannot have pointwise convergence of order greater than two over the whole domain of the function. The reason is that, for a smooth function, pointwise convergence of order greater than two implies that both the convex underestimator and the concave overestimator must have the same curvature as the function, and that is not possible if the function has nonzero curvature. The next theorem formalizes this argument. Note also the minimal convergence order of envelopes for smooth functions, Theorem 10.

Theorem 2 *Let $Z \subset \mathbb{R}^n$ be a nonempty open interval, and let $f : Z \to \mathbb{R}$ be a nonlinear, $\mathcal{C}^2$ function. Let $\left(f_Y^u, f_Y^o\right)_{Y \subset Z}$ be a scheme of estimators of f in Z. Then, the pointwise convergence order of the scheme is at most 2.*

Proof Recall that the Hessian H of f is given by $\left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)_{1 \leq i,j \leq n}$. Since f is a nonlinear $\mathcal{C}^2$ function, there exists a point $\mathbf{x}_0 \in Z$ such that the Hessian of f evaluated at $\mathbf{x}_0$ is nonzero: $H(\mathbf{x}_0) \neq 0$. In particular, there exists a vector $\mathbf{v}$ such that $\mathbf{v}^T H(\mathbf{x}_0)\mathbf{v} = \sum_{1 \leq i,j \leq n} \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\mathbf{x}_0} v_i v_j \neq 0$. Without loss of generality, we can assume that $\|\mathbf{v}\| = 1$.

Let us assume that $\mathbf{v}^T H(\mathbf{x}_0)\mathbf{v} > 0$; the proof for the negative case is similar and thus omitted. Since f is $\mathcal{C}^2$, there exists an interval $Z' \subset Z$ such that $\mathbf{x}_0 \in Z'$, and so that $M := \inf_{\mathbf{y} \in Z'} \mathbf{v}^T H(\mathbf{y})\mathbf{v} > 0$.

Let $\mathbf{x}$ be the function

$$\mathbf{x} : \mathbb{R} \to \mathbb{R}^n,$$
$$\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{v}.$$

Let $I \subset \mathbb{R}$ be an interval containing 0, and small enough so that the image of I under $\mathbf{x}$ is contained in Z'. By applying the Taylor theorem of order 1 to the composite function $f \circ \mathbf{x}$, for any $t \in I$ the following equality holds

$$f(\mathbf{x}(t)) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0) t\mathbf{v} + \frac{1}{2} t^2 \mathbf{v}^T H(\mathbf{x}(\xi_t))\mathbf{v}, \quad \text{for some } \xi_t \in (0, t).$$

For any $t \in I$, the point $\mathbf{x}(\xi_t)$ belongs to Z'. Therefore, the following inequality holds.

$$f(\mathbf{x}(t)) = f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)t\mathbf{v} + \frac{1}{2}t^2\mathbf{v}^T H(\mathbf{x}(\xi_t))\mathbf{v} \geq f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)t\mathbf{v} + \frac{1}{2}t^2 M. \quad (2)$$

Let $\varepsilon > 0$ be small enough such that

$$Y_\varepsilon := \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y} - \mathbf{x}_0\|_\infty \leq \varepsilon\} \subset Z', \text{ and}$$
$$[-\varepsilon, \varepsilon] \subset I.$$

By definition, $w(Y_\varepsilon) = 2\varepsilon$, and the values $\mathbf{x}(t) = \mathbf{x}_0 + t\mathbf{v}$ belong to Y_ε for any $t \in [-\varepsilon, \varepsilon]$. The function $f^o_{Y_\varepsilon}$ of the scheme $\left(f^u_Y, f^o_Y\right)_{Y \subset Z}$ is a concave overestimator of f on Y_ε; in particular, the following inequalities hold.

$$f^o_{Y_\varepsilon}(\mathbf{x}(-\varepsilon)) \geq f(\mathbf{x}(-\varepsilon)) \geq f(\mathbf{x}_0) - \varepsilon\nabla f(\mathbf{x}_0)\mathbf{v} + \frac{1}{2}\varepsilon^2 M, \text{ and}$$
$$f^o_{Y_\varepsilon}(\mathbf{x}(\varepsilon)) \geq f(\mathbf{x}(\varepsilon)) \geq f(\mathbf{x}_0) + \varepsilon\nabla f(\mathbf{x}_0)\mathbf{v} + \frac{1}{2}\varepsilon^2 M.$$

Combining these two inequalities, and since $f^o_{Y_\varepsilon}$ is concave, the following inequalities hold.

$$f^o_{Y_\varepsilon}(\mathbf{x}_0) = f^o_{Y_\varepsilon}(\mathbf{x}(0)) \geq \frac{f^o_{Y_\varepsilon}(\mathbf{x}(-\varepsilon)) + f^o_{Y_\varepsilon}(\mathbf{x}(\varepsilon))}{2} \geq f(\mathbf{x}_0) + \frac{1}{2}\varepsilon^2 M.$$

In other words, for any $\varepsilon > 0$ such that $[-\varepsilon, \varepsilon] \subset I$,

$$\sup_{\mathbf{y} \in Y_\varepsilon} |f^o_{Y_\varepsilon}(\mathbf{y}) - f(\mathbf{y})| \geq |f^o_{Y_\varepsilon}(\mathbf{x}_0) - f(\mathbf{x}_0)| \geq \frac{1}{2}\varepsilon^2 M = M\frac{w(Y_\varepsilon)^2}{8}.$$

Since this inequality holds for any small enough $\varepsilon > 0$, the pointwise convergence of the scheme $\left(f^u_Y, f^o_Y\right)_{Y \subset Z}$ is at most quadratic. □

As the previous theorem shows, a smooth, nonlinear function cannot be approximated over its domain of definition by a scheme with pointwise convergence greater than two. However, this result does not rule out the existence of schemes with better than quadratic convergence on particular points of the domain. For example, the function $f : [-1, 1] \to \mathbb{R}$, $f(z) = z^4$ is smooth, nonlinear, and can be approximated by a scheme $\left(f^u_Y, f^o_Y\right)_{Y \subset [-1,1]}$ defined as follows. For each interval $Y = [z^L, z^U] \subset [-1, 1]$, let $f^u_Y(z) = f(z)$, and let $f^o_Y(z) = \max\left\{f(Y^L), f(Y^U)\right\}$. This scheme has convergence of order four on the origin (which is also the minimizer of f on $[-1, 1]$). However, at the point $z = 1$, the convergence order is not fourth order, as can be seen by the interval $[1 - \varepsilon, 1]$.

4 McCormick relaxations

McCormick [24,25], introduced a method to generate convex underestimators and concave overestimators of factorable functions. In this section, we restate this method in terms of schemes of estimators and study its convergence rate. Note that in deviation of the original proposal we do not require the envelopes of the factors.

4.1 Relaxation of sum of two functions

We state McCormick's relaxation of the sum of two functions in terms of scheme of estimators.

Proposition 2 (Relaxation of Sum) *Let $Z \subset \mathbb{R}^n$ be a nonempty convex set, and let $g_1, g_2, g : Z \to \mathbb{R}$ be functions so that $g(\mathbf{z}) = g_1(\mathbf{z}) + g_2(\mathbf{z})$. Let $(g^u_{1,Y}, g^o_{1,Y})_{Y\subset Z}$ and $(g^u_{2,Y}, g^o_{2,Y})_{Y\subset Z}$ be schemes of estimators of g_1 and g_2 respectively in Z. For each interval $Y \subseteq Z$, let $g^u_Y, g^o_Y : Y \to \mathbb{R}$ be the functions $g^u_Y(\mathbf{z}) = g^u_{1,Y}(\mathbf{z}) + g^u_{2,Y}(\mathbf{z})$ and $g^o_Y(\mathbf{z}) = g^o_{1,Y}(\mathbf{z}) + g^o_{2,Y}(\mathbf{z})$. Then, $\left(g^u_Y, g^o_Y\right)_{Y\subset Z}$ is a scheme of estimators of g on Z.*

Having schemes of estimators of g_1 and g_2 with Hausdorff convergence order β does not imply that the McCormick relaxation scheme for addition has the same convergence order, as the next example shows.

Example 1 Let $Z = [-1, 1]$, and let $g_1(z) = z,\ g_2(z) = -z$, and $g(z) = g_1(z) + g_2(z) \equiv 0$. For each interval $Y = \left[z^L_Y, z^U_Y\right] \subset Z$, let $g^u_{1,Y}(z) = z^L_Y$ and let $g^o_{1,Y}(z) = z^U_Y$. Then, $(g^u_{1,Y}, g^o_{1,Y})_{Y\subset Z}$ is a scheme of estimators of g_1 of arbitrarily high Hausdorff convergence order. Similarly, for each interval $Y = [z^L_Y, z^U_Y] \subset Z$, let $g^u_{2,Y}(z) = -z^U_Y$ and let $g^o_{2,Y}(z) = -z^L_Y$. The scheme $(g^u_{2,Y}, g^o_{2,Y})_{Y\subset Z}$ of estimators of g_2 also has arbitrarily high Hausdorff convergence order. For each interval $Y = [-\varepsilon, \varepsilon] \subset Z$, the corresponding McCormick's estimators g^u_Y and g^o_Y of g are as follows:

$$g^u_Y(z) = g^u_{1,Y}(z) + g^u_{2,Y}(z) = -\varepsilon - \varepsilon = -2\varepsilon,$$

and

$$g^o_Y(z) = g^o_{1,Y}(z) + g^o_{2,Y}(z) = \varepsilon + \varepsilon = 2\varepsilon.$$

Note that the Hausdorff distance between $\bar{g}(Y) = [0, 0]$ and $H_g(Y) = [\left(g^u_Y\right)^L(Y), \left(g^o_Y\right)^U(Y)] = [-2\varepsilon, 2\varepsilon]$ is $q(\bar{g}(Y), H_g(Y)) = 2\varepsilon = w(Y)$. In other words, the measure of the error between $\bar{g}(Y)$ and $H_g(Y)$ is linear on $w(Y)$. Since this relationship holds for any $0 < \varepsilon < 1$, it then follows that the Hausdorff convergence order of $\left(g^u_Y, g^o_Y\right)_{Y\subset Z}$ is at most linear in this example.

On the other hand, the McCormick sum estimator preserves pointwise convergence order, as the following Theorem shows.

Theorem 3 *Let $Z \subset \mathbb{R}^n$ be a nonempty convex set, and let $g_1, g_2, g : Z \to \mathbb{R}$ be functions so that $g(\mathbf{z}) = g_1(\mathbf{z}) + g_2(\mathbf{z})$. Let $(g^u_{1,Y}, g^o_{1,Y})_{Y\subset Z}$ and $(g^u_{2,Y}, g^o_{2,Y})_{Y\subset Z}$ be schemes of estimators of g_1 and g_2 respectively in Z and assume that the schemes have pointwise convergence of order $\gamma_1, \gamma_2 > 0$ respectively. Then, the scheme $(g^u_Y, g^o_Y)_{Y\subset Z}$ of estimators of g on Z constructed from $(g^u_{1,Y}, g^o_{1,Y})$ and $(g^u_{2,Y}, g^o_{2,Y})$ as in Proposition 2 has pointwise convergence of order* $\min\{\gamma_1, \gamma_2\}$.

Proof Let $Y \subset Z$ be an interval, and let $\mathbf{z}$ be a point of Y. Using the triangle inequality, we can bound the distance between $g(\mathbf{z})$ and $g^u_Y(\mathbf{z})$ as follows.

$$\begin{aligned}\left|g(\mathbf{z}) - g^u_Y(\mathbf{z})\right| &= \left|(g_1(\mathbf{z}) + g_2(\mathbf{z})) - \left(g^u_{1,Y}(\mathbf{z}) + g^u_{2,Y}(\mathbf{z})\right)\right| \\ &\le \left|g_1(\mathbf{z}) - g^u_{1,Y}(\mathbf{z})\right| + \left|g_2(\mathbf{z}) - g^u_{2,Y}(\mathbf{z})\right| \\ &\le \tau_{g_1} w(Y)^{\gamma_1} + \tau_{g_2} w(Y)^{\gamma_2},\end{aligned}$$

for two constants τ_{g_1} and τ_{g_2} that do not depend on $\mathbf{z}$ or Y.

Consider now $\gamma = \min\{\gamma_1, \gamma_2\}$ and two constants $\hat{\tau}_{g_1} = \tau_{g_1} \max\left\{1, w(Z)^{\gamma_1-\gamma}\right\}$ and $\hat{\tau}_{g_2} = \tau_{g_2} \max\left\{1, w(Z)^{\gamma_2-\gamma}\right\}$. We obtain directly

$$\tau_{g_1} w(Y)^{\gamma_1} \le \hat{\tau}_{g_1} w(Y)^{\gamma} \text{ and } \tau_{g_2} w(Y)^{\gamma_2} \le \hat{\tau}_{g_2} w(Y)^{\gamma},$$

and thus

$$\left|g(\mathbf{z}) - g^u_Y(\mathbf{z})\right| \leq \hat{\tau}_{g_1} w(Y)^\gamma + \hat{\tau}_{g_2} w(Y)^\gamma = \left(\hat{\tau}_{g_1} + \hat{\tau}_{g_2}\right) w(Y)^\gamma.$$

A similar bound holds for the distance between $g(\mathbf{z})$ and $g^o_Y(\mathbf{z})$; since $\hat{\tau}_{g_1} + \hat{\tau}_{g_2}$ does not depend on $\mathbf{z}$ or Y, it then follows that the scheme $\left(g^u_Y, g^o_Y\right)_{Y \subset Z}$ has pointwise convergence of order γ. □

4.2 Relaxation of product of two functions

We state McCormick's relaxation of the product of two functions in terms of scheme of estimators, see also [7,32].

Proposition 3 (Relaxation of Products) [24,26] *Let $Z \subset \mathbb{R}^n$ be a nonempty convex set, and let $g_1, g_2 : Z \to \mathbb{R}$ be two continuous functions. Let $(g^u_{1,Y}, g^o_{1,Y})$ and $(g^u_{2,Y}, g^o_{2,Y})$ be schemes of estimators of g_1 and g_2 respectively in Z. For each interval $Y \subseteq Z$, let $g^L_{1,Y}, g^U_{1,Y}, g^L_{2,Y}, g^U_{2,Y} \in \mathbb{R}$ such that*

$$g^L_{1,Y} \leq g_1(\mathbf{z}) \leq g^U_{1,Y}, \quad \forall \mathbf{z} \in Y,$$
$$g^L_{2,Y} \leq g_2(\mathbf{z}) \leq g^U_{2,Y}, \quad \forall \mathbf{z} \in Y.$$

Consider the following intermediate functions $g_{a1,Y}, g_{a2,Y}, g_{b1,Y}, g_{b2,Y}, g_{c1,Y}, g_{c2,Y}, g_{d1,Y}, g_{d2,Y} : Y \to \mathbb{R}$,

$$g_{a1,Y}(\mathbf{z}) = \min\left\{g^L_{2,Y} g^u_{1,Y}(\mathbf{z}), g^L_{2,Y} g^o_{1,Y}(\mathbf{z})\right\}, \quad g_{a2,Y}(\mathbf{z}) = \min\left\{g^L_{1,Y} g^u_{2,Y}(\mathbf{z}), g^L_{1,Y} g^o_{2,Y}(\mathbf{z})\right\}$$
$$g_{b1,Y}(\mathbf{z}) = \min\left\{g^U_{2,Y} g^u_{1,Y}(\mathbf{z}), g^U_{2,Y} g^o_{1,Y}(\mathbf{z})\right\}, \quad g_{b2,Y}(\mathbf{z}) = \min\left\{g^U_{1,Y} g^u_{2,Y}(\mathbf{z}), g^U_{1,Y} g^o_{2,Y}(\mathbf{z})\right\}$$
$$g_{c1,Y}(\mathbf{z}) = \max\left\{g^L_{2,Y} g^u_{1,Y}(\mathbf{z}), g^L_{2,Y} g^o_{1,Y}(\mathbf{z})\right\}, \quad g_{c2,Y}(\mathbf{z}) = \max\left\{g^U_{1,Y} g^u_{2,Y}(\mathbf{z}), g^U_{1,Y} g^o_{2,Y}(\mathbf{z})\right\}$$
$$g_{d1,Y}(\mathbf{z}) = \max\left\{g^U_{2,Y} g^u_{1,Y}(\mathbf{z}), g^U_{2,Y} g^o_{1,Y}(\mathbf{z})\right\}, \quad g_{d2,Y}(\mathbf{z}) = \max\left\{g^L_{1,Y} g^u_{2,Y}(\mathbf{z}), g^L_{1,Y} g^o_{2,Y}(\mathbf{z})\right\}.$$

Then, $g_{a1,Y}, g_{a2,Y}, g_{b1,Y}$ and $g_{b2,Y}$ are convex on Y while $g_{c1,Y}, g_{c2,Y}, g_{d1,Y}$ and $g_{d2,Y}$ are concave on Y. Moreover, $g^u_Y, g^o_Y : Y \to \mathbb{R}$ such that

$$g^u_Y(\mathbf{z}) = \max\left\{g_{a1,Y}(\mathbf{z}) + g_{a2,Y}(\mathbf{z}) - g^L_{1,Y} g^L_{2,Y}, g_{b1,Y}(\mathbf{z}) + g_{b2,Y}(\mathbf{z}) - g^U_{1,Y} g^U_{2,Y}\right\},$$
$$g^o_Y(\mathbf{z}) = \min\left\{g_{c1,Y}(\mathbf{z}) + g_{c2,Y}(\mathbf{z}) - g^U_{1,Y} g^L_{2,Y}, g_{d1,Y}(\mathbf{z}) + g_{d2,Y}(\mathbf{z}) - g^L_{1,Y} g^U_{2,Y}\right\},$$

are respectively a convex and concave relaxation of $g_1 g_2$ on Y. In other words, $\left(g^u_Y, g^o_Y\right)_{Y \subset Z}$ is a scheme of estimators of $g_1 g_2$ on Z.

Having schemes of estimators of g_1 and g_2 with Hausdorff convergence of order β_1, β_2 does not guarantee that the relaxation of the product has the same convergence order. In fact, it may not have better than linear convergence. The next example is an extreme case for the product rule: in it, g_1 and g_2 are such that their minima and maxima on an interval Y centered at 0 are attained at the extreme points of the interval, but the maximum of the product is obtained at 0. The schemes that approximate g_1 and g_2 are chosen so that they have arbitrarily high convergence order, but they do not approximate well points other than the minima and maxima of g_1 and g_2, and in particular they do a poor job at 0, where the maximum of $g_1 g_2$ is attained. In Proposition 4, we show that, in this particular example, the scheme of McCormick's product rule has Hausdorff convergence order at most linear.

Example 2 Let Z be the interval $[-1, 1]$. Let $g_1, g_2, g : [-1, 1] \to \mathbb{R}$ be the functions $g_1(z) = 1 + z$, $g_2(z) = 1 - z$, and $g(z) = g_1(z)g_2(z) = (1 + z)(1 - z) = 1 - z^2$. For each interval $Y = [z_Y^L, z_Y^U] \subset [-1, 1]$, let $g_{1,Y}^u, g_{1,Y}^o, g_{2,Y}^u, g_{2,Y}^o : Y \to \mathbb{R}$ be the functions

$$\begin{aligned} g_{1,Y}^u(z) &= 1 + z_Y^L, \quad g_{1,Y}^o(z) = 1 + z_Y^U, \\ g_{2,Y}^u(z) &= 1 - z_Y^U, \quad g_{2,Y}^o(z) = 1 - z_Y^L. \end{aligned}$$

We note that, since g_1 and g_2 are linear functions, we could have taken $g_{1,Y}^u(z) = g_{1,Y}^o(z) = g_1(z)$ and $g_{2,Y}^u(z) = g_{2,Y}^o(z) = g_2(z)$; instead, we chose an extreme case of bad relaxations.

Similarly, for each interval $Y = [z_Y^L, z_Y^U] \subset [-1, 1]$, let $g_{1,Y}^L, g_{1,Y}^U$ ($g_{2,Y}^L, g_{2,Y}^U$) be the minimum and maximum values of g_1 (g_2) in Y, namely,

$$\begin{aligned} g_{1,Y}^L &= g_1^L(Y) = 1 + z_Y^L, \quad g_{1,Y}^U = g_1^U(Y) = 1 + z_Y^U, \\ g_{2,Y}^L &= g_2^L(Y) = 1 - z_Y^U, \quad g_{2,Y}^U = g_2^U(Y) = 1 - z_Y^L. \end{aligned} \tag{3}$$

We next prove that, for this example, the McCormick scheme for the product g_1g_2 obtained from the functions $g_{1,Y}^u, g_{1,Y}^o, g_{2,Y}^u, g_{2,Y}^o$ and the bounds $g_{1,Y}^L, g_{1,Y}^U, g_{2,Y}^L, g_{2,Y}^U$ has Hausdorff convergence order at most linear.

Proposition 4 *In Example 2, $g_{1,Y}^L, g_{1,Y}^U, g_{2,Y}^L, g_{2,Y}^U$, the lower and upper bounds of g_1 and g_2, define the following inclusion functions of g_1 and g_2:*

$$\begin{aligned} &H_1, H_2 : Y \in \mathbb{IR}^n, Y \subset [-1, 1] \to \mathbb{IR} \\ &H_1(Y) = \left[g_{1,Y}^L, g_{1,Y}^U\right], \\ &H_2(Y) = \left[g_{2,Y}^L, g_{2,Y}^U\right]. \end{aligned}$$

These inclusion functions are also the ones associated with the scheme of estimators $(g_{1,Y}^u, g_{1,Y}^o)_{Y\subset[-1,1]}$ and $(g_{2,Y}^u, g_{2,Y}^o)_{Y\subset[-1,1]}$ of g_1 and g_2 respectively. These inclusion functions have Hausdorff convergence of order β_1, β_2, with β_1, β_2 arbitrarily high. Let $\left(g_Y^u, g_Y^o\right)_{Y\subset[-1,1]}$ be the scheme of estimators of $g(z) = g_1(z)g_2(z)$ defined according to McCormick's relaxation of products. The inclusion function H_g associated with $\left(g_Y^u, g_Y^o\right)_{Y\subset[-1,1]}$ has at most linear Hausdorff convergence.

Proof Since for each interval $Y \subseteq [-1, 1]$, the intervals $H_1(Y), H_2(Y)$ are equal to $[g_{1,Y}^L, g_{1,Y}^U]$ and $[g_{2,Y}^L, g_{2,Y}^U]$ respectively, it follows that the inclusion functions $H_1(Y), H_2(Y)$ have essentially Hausdorff convergence of order infinity.

On the other hand, to show that the Hausdorff convergence order of H_g is at most one, it is enough to analyze what happens on intervals centered at 0. Let $Y = [-\varepsilon, \varepsilon] \subset [-1, 1]$ be such an interval. For this interval, the functions $g_{1,Y}^u, g_{1,Y}^o, g_{2,Y}^u, g_{2,Y}^o$, and $g_{1,Y}^L, g_{1,Y}^U, g_{2,Y}^L, g_{2,Y}^U$, the lower and upper bounds of g_1 and g_2, are equal to

$$g^u_{1,Y}(z) = g^L_{1,Y} = 1-\varepsilon, \quad g^o_{1,Y}(z) = g^U_{1,Y} = 1+\varepsilon,$$
$$g^u_{2,Y}(z) = g^L_{2,Y} = 1-\varepsilon, \quad g^o_{2,Y}(z) = g^U_{2,Y} = 1+\varepsilon.$$

We note that the z-dependence of the functions $g^u_{1,Y}, g^o_{1,Y}, g^u_{2,Y}, g^o_{2,Y}$ is lost due to the chosen relaxations.

The corresponding intermediate functions $g_{a1,Y}, g_{a2,Y}, g_{b1,Y}, g_{b2,Y}, g_{c1,Y}, g_{c2,Y}, g_{d1,Y}, g_{d2,Y} : Y \to \mathbb{R}$ of Proposition 3 are equal to

$$\begin{aligned}
g_{a1,Y}(z) &= \min\{(1-\varepsilon)(1-\varepsilon), (1-\varepsilon)(1+\varepsilon)\} = (1-\varepsilon)^2,\\
g_{a2,Y}(z) &= \min\{(1-\varepsilon)(1-\varepsilon), (1-\varepsilon)(1+\varepsilon)\} = (1-\varepsilon)^2,\\
g_{b1,Y}(z) &= \min\{(1+\varepsilon)(1-\varepsilon), (1+\varepsilon)(1+\varepsilon)\} = 1-\varepsilon^2,\\
g_{b2,Y}(z) &= \min\{(1+\varepsilon)(1-\varepsilon), (1+\varepsilon)(1+\varepsilon)\} = 1-\varepsilon^2,\\
g_{c1,Y}(z) &= \max\{(1-\varepsilon)(1-\varepsilon), (1-\varepsilon)(1+\varepsilon)\} = 1-\varepsilon^2,\\
g_{c2,Y}(z) &= \max\{(1+\varepsilon)(1-\varepsilon), (1+\varepsilon)(1+\varepsilon)\} = (1+\varepsilon)^2,\\
g_{d1,Y}(z) &= \max\{(1+\varepsilon)(1-\varepsilon), (1+\varepsilon)(1+\varepsilon)\} = (1+\varepsilon)^2,\\
g_{d2,Y}(z) &= \max\{(1-\varepsilon)(1-\varepsilon), (1-\varepsilon)(1+\varepsilon)\} = 1-\varepsilon^2.
\end{aligned}$$

Again, these functions are constant because of the chosen relaxations. Then, the convex and concave relaxations of g on Y are as follows:

$$\begin{aligned}
g^u_Y(z) &= \max\left\{(1-\varepsilon)^2 + (1-\varepsilon)^2 - (1-\varepsilon)^2, (1-\varepsilon^2) + (1-\varepsilon^2) - (1+\varepsilon)^2\right\}\\
&= \max\left\{1-2\varepsilon+\varepsilon^2, 1-2\varepsilon-3\varepsilon^2\right\} = 1-2\varepsilon+\varepsilon^2,\\
g^o_Y(z) &= \min\left\{(1-\varepsilon^2)+(1+\varepsilon)^2-(1+\varepsilon)(1-\varepsilon), (1-\varepsilon^2)+(1+\varepsilon)^2-(1-\varepsilon)(1+\varepsilon)\right\}\\
&= \min\left\{(1+\varepsilon)^2, (1+\varepsilon)^2\right\} = (1+\varepsilon)^2.
\end{aligned}$$

The inclusion function $H_g = [\inf g^u_Y(Y), \sup g^o_Y(Y)]$ evaluated on $Y = [-\varepsilon, \varepsilon]$ is equal to $\left[1-2\varepsilon+\varepsilon^2, 1+2\varepsilon+\varepsilon^2\right]$. The interval $\bar{g}([-\varepsilon, \varepsilon])$ is equal to $[1-\varepsilon^2, 1]$, and so

$$\begin{aligned}
q\left(\bar{g}([-\varepsilon,\varepsilon]), H_g([-\varepsilon,\varepsilon])\right) &= \max\left\{(1-\varepsilon^2)-(1-2\varepsilon+\varepsilon^2), (1+2\varepsilon+\varepsilon^2)-1\right\}\\
&= \max\left\{2\varepsilon-2\varepsilon^2, 2\varepsilon+2\varepsilon^2\right\} \geq 2\varepsilon = w(Y).
\end{aligned}$$

In other words, the measure of the error between $\bar{g}$ and H_g on intervals $Y = [-\varepsilon, \varepsilon]$ is at least the size of Y. Since this lower bound on the error holds for any $0 \leq \varepsilon \leq 1$, the inclusion function H_g has linear Hausdorff convergence at best. □

In the example just analyzed, the relaxation functions $g^u_1, g^o_1, g^u_2, g^o_2$ do not satisfy pointwise convergence, and that causes the loss of the convergence rate. In the next theorem we show a positive convergence result of the relaxation of products with pointwise convergence. If the schemes of estimators $(g^u_{1,Y}, g^o_{1,Y})$, $(g^u_{2,Y}, g^o_{2,Y})$ of g_1 and g_2 respectively have pointwise convergence of order $\gamma_1, \gamma_2 \geq 1$ respectively, and if the schemes of constant estimators $(g^L_{1,Y}, g^U_{1,Y})$, $(g^L_{2,Y}, g^U_{2,Y})$ converge to g_1 and g_2 respectively with order $\beta_1 \geq 1$ and $\beta_2 \geq 1$, then $\left(g^u_Y, g^o_Y\right)$, the scheme of estimators of $g_1 g_2$ constructed on Proposition 3 has pointwise convergence of order $\gamma = \min\{\gamma_1, \gamma_2, 2\}$. This result has some loss of the convergence rate; even if the pointwise convergence of $(g^u_{1,Y}, g^o_{1,Y})$, $(g^u_{2,Y}, g^o_{2,Y})$ is greater than 2, the pointwise convergence order of the product scheme is quadratic. In Example 4 we will show that.

Because of Theorem 1, under the hypothesis of the following theorem, the convergence of the product scheme in the Hausdorff metric (in the sense of Definition 8) is of order $\beta = \gamma = \min\{\gamma_1, \gamma_2, 2\}$ too, even when the image estimates have convergence order 1. In other words, the product scheme can improve on the convergence order of the estimates if the schemes for the factors have high pointwise convergence order.

We first prove a Lemma that bounds the size of the range of a Lipschitz continuous function.

Lemma 1 *Let $f : Z \subset \mathbb{R}^n \to \mathbb{R}$ be a Lipschitz continuous function with Lipschitz constant $M \in \mathbb{R}$. Let Y be a subset of Z. Then,*

$$w(\bar{f}(Y)) \leq M w(Y).$$

Proof By definition of the Lipschitz constant we obtain for any $\mathbf{y}_1, \mathbf{y}_2 \in Y$

$$|f(\mathbf{y}_1) - f(\mathbf{y}_0)| \leq M \|\mathbf{y}_1 - \mathbf{y}_0\|$$

and since $\|\mathbf{y}_1 - \mathbf{y}_0\| \leq w(Y)$ we also have $|f(\mathbf{y}_1) - f(\mathbf{y}_0)| \leq M w(Y)$. It then follows that

$$w\left(\bar{f}(Y)\right) = \sup_{\mathbf{y}_0, \mathbf{y}_1 \in Y} |f(\mathbf{y}_1) - f(\mathbf{y}_0)| \leq M w(Y).$$

□

Theorem 4 *Let $Z \subset \mathbb{R}^n$ be a nonempty convex set, and let $g_1, g_2 : Z \to \mathbb{R}$ be two Lipschitz continuous functions with Lipschitz constants M_1 and M_2 respectively. Let $(g^u_{1,Y}, g^o_{1,Y})_{Y \subset Z}$ and $(g^u_{2,Y}, g^o_{2,Y})_{Y \subset Z}$ be schemes of estimators of g_1 and g_2 respectively in Z. Let $(g^L_{1,Y}, g^U_{1,Y})_{Y \subset Z}$ and $(g^L_{2,Y}, g^U_{2,Y})_{Y \subset Z}$ be schemes of constant estimators of g_1 and g_2 respectively in Z. Assume that $(g^u_{1,Y}, g^o_{1,Y})_{Y \subset Z}$ and $(g^u_{2,Y}, g^o_{2,Y})_{Y \subset Z}$ have pointwise convergence of order $\gamma_1, \gamma_2 \geq 1$ respectively. Furthermore, assume that $(g^L_{1,Y}, g^U_{1,Y})_{Y \subset Z}$ and $(g^L_{2,Y}, g^U_{2,Y})_{Y \subset Z}$ have convergence of order at least* 1. *Let $\left(g^u_Y, g^o_Y\right)_{Y \subset Z}$ be the scheme of estimators of $g_1 g_2$ constructed in Proposition 3. Then, the scheme $\left(g^u_Y, g^o_Y\right)_{Y \subset Z}$ has also pointwise convergence of order* $\min\{\gamma_1, \gamma_2, 2\}$.

Proof Let $Y \subseteq Z$ be an interval. We will show that, for any $\mathbf{z} \in Y$, the distance between $g_1(\mathbf{z}) g_2(\mathbf{z})$ and $g_{a1,Y}(\mathbf{z}) + g_{a2,Y}(\mathbf{z}) - g^L_{1,Y} g^L_{2,Y}$ is at most $C_0 w(Y)^2 + C_1 w(Y)^{\gamma_1} + C_2 w(Y)^{\gamma_2}$ for constants $C_0, C_1, C_2 > 0$ that do not depend on Y. We note that such a bound can be further estimated as $C w(Y)^{\min\{\gamma_1, \gamma_2, 2\}}$ for a constant C that does not depend on Y. Since g^u_Y underestimates $g_1 g_2$, and since by definition g^u_Y is at least $g_{a1,Y}(\mathbf{z}) + g_{a2,Y}(\mathbf{z}) - g^L_{1,Y} g^L_{2,Y}$, it will then follow that

$$\begin{aligned} |g_1(\mathbf{z}) g_2(\mathbf{z}) - g^u_Y(\mathbf{z})| &\leq \left| g_1(\mathbf{z}) g_2(\mathbf{z}) - \left(g_{a1,Y}(\mathbf{z}) + g_{a2,Y}(\mathbf{z}) - g^L_{1,Y} g^L_{2,Y} \right) \right| \\ &\leq C_0 w(Y)^2 + C_1 w(Y)^{\gamma_1} + C_2 w(Y)^{\gamma_2}. \end{aligned} \tag{4}$$

In order to bound $\left| g_1(\mathbf{z}) g_2(\mathbf{z}) - \left(g_{a1,Y}(\mathbf{z}) + g_{a2,Y}(\mathbf{z}) - g^L_{1,Y} g^L_{2,Y} \right) \right|$, it is convenient to rewrite $g_1(\mathbf{z}) g_2(\mathbf{z})$ as follows.

$$g_1(\mathbf{z}) g_2(\mathbf{z}) = \left(g_1(\mathbf{z}) - g^L_{1,Y} \right) \left(g_2(\mathbf{z}) - g^L_{2,Y} \right) + g^L_{1,Y} g_2(\mathbf{z}) + g^L_{2,Y} g_1(\mathbf{z}) - g^L_{1,Y} g^L_{2,Y}.$$

Then, the distance between $g_1(\mathbf{z}) g_2(\mathbf{z})$ and $(g_{a1,Y}(\mathbf{z}) + g_{a2,Y}(\mathbf{z}) - g^L_{1,Y} g^L_{2,Y})$ satisfies the following inequalities.

$$\begin{aligned}
&\left|g_1(\mathbf{z})g_2(\mathbf{z}) - \left(g_{a1,Y}(\mathbf{z}) + g_{a2,Y}(\mathbf{z}) - g^L_{1,Y}g^L_{2,Y}\right)\right| \\
&= \left|\left(g_1(\mathbf{z}) - g^L_{1,Y}\right)\left(g_2(\mathbf{z}) - g^L_{2,Y}\right) + \left(g^L_{1,Y}g_2(\mathbf{z}) - g_{a2,Y}(\mathbf{z})\right) + \left(g^L_{2,Y}g_1(\mathbf{z}) - g_{a1,Y}(\mathbf{z})\right)\right| \\
&\leq \left|\left(g_1(\mathbf{z}) - g^L_{1,Y}\right)\left(g_2(\mathbf{z}) - g^L_{2,Y}\right)\right| + \left|g^L_{1,Y}g_2(\mathbf{z}) - g_{a2,Y}(\mathbf{z})\right| + \left|g^L_{2,Y}g_1(\mathbf{z}) - g_{a1,Y}(\mathbf{z})\right|.
\end{aligned} \tag{5}$$

We next bound the three terms of the right hand side of Inequality (5).

We claim that the term $\left|(g_1(\mathbf{z}) - g^L_{1,Y})(g_2(\mathbf{z}) - g^L_{2,Y})\right|$ is at most $C_0 w(Y)^2$ for a constant $C_0 > 0$ that does not depend on Y. Let $g^L_1(Y) = \inf g_1(Y)$. Since g_1 is a Lipschitz function with Lipschitz constant M_1, the following bound holds: $\|g_1(\mathbf{z}) - g^L_1(Y)\| \leq w(\bar{g_1}(Y)) \leq M_1 w(Y)$ (Lemma 1). Moreover, since $(g^L_{1,Y}, g^U_{1,Y})$ converges to g_1 with order 1, the following inequalities hold:

$$\begin{aligned}
\left|g_1(\mathbf{z}) - g^L_{1,Y}\right| &= \left|(g_1(\mathbf{z}) - g^L_1(Y)) + \left(g^L_1(Y) - g^L_{1,Y}\right)\right| \\
&\leq \left|g_1(\mathbf{z}) - g^L_1(Y)\right| + \left|g^L_1(Y) - g^L_{1,Y}\right| \\
&\leq M_1 w(Y) + \tau_1 w(Y) = (M_1 + \tau_1) w(Y),
\end{aligned} \tag{6}$$

where $M_1 > 0$ and $\tau_1 > 0$ do not depend on Y. Similarly, $\left|g_2(\mathbf{z}) - g^L_{2,Y}\right| \leq (M_2 + \tau_2) w(Y)$ for constants $M_2 > 0$ and $\tau_2 > 0$. Therefore, the first term of the right hand side of Inequality (5) has second order, namely,

$$\left|\left(g_1(\mathbf{z}) - g^L_{1,Y}\right)\left(g_2(\mathbf{z}) - g^L_{2,Y}\right)\right| \leq (M_1 + \tau_1)(M_2 + \tau_2) w(Y)^2 = C_0 w(Y)^2. \tag{7}$$

We next bound $\left|g^L_{1,Y}g_2(\mathbf{z}) - g_{a2,Y}(\mathbf{z})\right|$. We note that, by definition, the function $g_{a2,Y}(\mathbf{z})$ is either $g^L_{1,Y}g^u_{2,Y}(\mathbf{z})$ or $g^L_{1,Y}g^o_{2,Y}(\mathbf{z})$. In either case, we can bound $\left|g^L_{1,Y}g_2(\mathbf{z}) - g_{a2,Y}(\mathbf{z})\right|$ as follows.

$$\left|g^L_{1,Y}g_2(\mathbf{z}) - g_{a2,Y}(\mathbf{z})\right| \leq |g^L_{1,Y}| \max\left\{|g_2(\mathbf{z}) - g^u_{2,Y}(\mathbf{z})|, |g_2(\mathbf{z}) - g^o_{2,Y}(\mathbf{z})|\right\}. \tag{8}$$

Since the scheme of constants $(g^L_{1,Y}, g^U_{1,Y})$ converges to g_1 with order at least 1, the value $|g^L_{1,Y}|$ is at most $\|g_1\|_\infty + \tau_1 w(Y) \leq \|g_1\|_\infty + \tau_1 w(Z)$. We remark that this bound does not depend on Y or on $\mathbf{z}$.

Since the scheme $(g^u_{2,Y}, g^o_{2,Y})$ converges to g_2 pointwise with order γ_2, the term $\max\left\{|g_2(\mathbf{z}) - g^u_{2,Y}(\mathbf{z})|, |g_2(\mathbf{z}) - g^o_{2,Y}(\mathbf{z})|\right\}$ is at most $\hat{\tau}_2 w(Y)^{\gamma_2}$. Combining these bounds with Inequality (8), we have that

$$\left|g^L_{1,Y}g_2(\mathbf{z}) - g_{a2,Y}(\mathbf{z})\right| \leq C_2 w(Y)^{\gamma_2}, \tag{9}$$

for a constant $C_2 > 0$ that does not depend on Y. A similar bound holds for $\left|\left(g^L_{2,Y}g_1(\mathbf{z}) - g_{a1,Y}(\mathbf{z})\right)\right|$, the third term of the right hand side of Inequality (5):

$$\left|g^L_{2,Y}g_1(\mathbf{z}) - g_{a1,Y}(\mathbf{z})\right| \leq C_1 w(Y)^{\gamma_1}. \tag{10}$$

Combining Eqs. 5, 7, 9, and 10, we proved that the distance between $g_1(\mathbf{z})g_2(\mathbf{z})$ and $g_{a1,Y}(\mathbf{z}) + g_{a2,Y}(\mathbf{z}) - g^L_{1,Y}g^L_{2,Y}$ is at most $C_0 w(Y)^2 + C_1 w(Y)^{\gamma_1} + C_2 w(Y)^{\gamma_2} \leq C w(Y)^{\min\{\gamma_1,\gamma_2,2\}}$ for a constant $C > 0$ that does not depend on the point $\mathbf{z} \in Y$ or Y

itself. By Inequality (4), this also implies that the distance between $g_1(\mathbf{z})g_2(\mathbf{z})$ and $g^u_Y(\mathbf{z})$ is at most $Cw(Y)^{\min\{\gamma_1,\gamma_2,2\}}$.

A similar argument shows that the distance between $g_1(\mathbf{z})g_2(\mathbf{z})$ and $g^o_Y(\mathbf{z})$ is at most $C'w(Y)^{\min\{\gamma_1,\gamma_2,2\}}$ for a constant $C' > 0$ that does not depend on the point $\mathbf{z} \in Y$ or Y itself, and therefore, the scheme $\left(g^u_Y, g^o_Y\right)$ converges to $g_1 g_2$ pointwise with order $\min\{\gamma_1, \gamma_2, 2\}$. □

The next two examples show that the bounds on the convergence order in Theorem 4 are sharp, i.e., the pointwise convergence order of a product may be as low as $\min\{\gamma_1, \gamma_2, 2\}$. Example 3 shows that the convergence order of the relaxations to a product function are determined by the function with the weaker relaxations, i.e., the relaxations with lower convergence order.

Example 3 Let Z be the interval $[0, 1]$. Let $g_1 : [0, 1] \to \mathbb{R}$ be the constant function $g_1(z) = 1$. For each interval $Y = \left[z^L_Y, z^U_Y\right] \subset [0, 1]$, let $g^u_{1,Y}, g^o_{1,Y} : Y \to \mathbb{R}$ be a convex (concave) underestimator (overestimator) of g_1 defined as

$$\forall z \in Y : g^u_{1,Y}(z) = g^o_{1,Y}(z) = 1.$$

For each interval $Y = \left[z^L_Y, z^U_Y\right] \subset [0, 1]$, let $g^L_{1,Y}, g^U_{1,Y} \in \mathbb{R}$ be 1.

Let $g_2 : [0, 1] \to \mathbb{R}$ be the function $g_2(z) = z$. For each interval $Y = \left[z^L_Y, z^U_Y\right] \subset [0, 1]$, let $g^u_{2,Y}, g^o_{2,Y} : Y \to \mathbb{R}$ be a convex (concave) underestimator (overestimator) of g_2 defined as

$$g^u_{2,Y}(z) = z^L_Y, \quad g^o_{2,Y}(z) = z^U_Y.$$

For each interval $Y = \left[z^L_Y, z^U_Y\right] \subset [0, 1]$, let $g^L_{2,Y}, g^U_{2,Y} \in \mathbb{R}$ be $g^L_{2,Y} = z^L_Y$, $g^U_{2,Y} = z^U_Y$.

The function g_1 and the functions of the scheme $(g^u_{1,Y}, g^o_{1,Y})_{Y\subset Z}$ are all equal to 1, so the scheme $(g^u_{1,Y}, g^o_{1,Y})_{Y\subset Z}$ has pointwise convergence order γ_1 arbitrarily high. The bounds of $(g^L_{1,Y}, g^U_{1,Y})_{Y\subset Z}$ are also equal to 1, and therefore the corresponding scheme has convergence arbitrarily high.

On the other hand, the scheme $(g^u_{2,Y}, g^o_{2,Y})_{Y\subset Z}$ has pointwise convergence $\gamma_2 = 1$, since, for each $Y = [z^L_Y, z^U_Y] \subset Z$:

$$\sup_{z\in Y} \left|g_2(z) - g^u_{2,Y}\right| = \left|z^U_Y - z^L_Y\right| = w(Y),$$

$$\sup_{z\in Y} \left|g_2(z) - g^o_{2,Y}\right| = \left|z^L_Y - z^U_Y\right| = w(Y).$$

The product function $g(z) = g_1(z)g_2(z)$ is equal to $g(z) = z$. By Proposition 3 (McCormick's relaxation of product), it follows that, for each interval $Y \subset Z$, the resulting approximation functions $g^u_Y, g^o_Y(z)$ are equal to $g^u_Y(z) = g^L_{2,Y}$ and $g^o_Y(z) = g^U_{2,Y}$ respectively. In other words, the function g is equal to g_2, and the product scheme $\left(g^u_Y, g^o_Y\right)_{Y\subset Z}$ is the scheme $(g^u_{2,Y}, g^o_{2,Y})_{Y\subset Z}$ which has linear pointwise convergence.

In the next example the pointwise convergence of McCormick's product scheme is quadratic although the estimation schemes for the composing functions are exact. Recall that by Theorem 4 the scheme of estimators for smooth nonlinear functions have at most quadratic pointwise convergence.

Example 4 Let Z be the interval $[-1, 1]$. Let $g_1, g_2, g : [-1, 1] \to \mathbb{R}$ be the functions $g_1(z) = g_2(z) = z$, and $g(z) = g_1(z)g_2(z) = z^2$. For each interval $Y = \left[z^L_Y, z^U_Y\right] \subset [-1, 1]$, let $g^u_{1,Y}, g^o_{1,Y}, g^u_{2,Y}, g^o_{2,Y} : Y \to \mathbb{R}$ be convex (concave) underestimators (overestimators) of g_1 and g_2

$$g^u_{1,Y}(z) = g^o_{1,Y}(z) = g^u_{2,Y}(z) = g^o_{2,Y}(z) = z.$$

Similarly, for each interval $Y = \left[z^L_Y, z^U_Y\right] \subset [-1, 1]$, let $g^L_{1,Y}, g^U_{1,Y}$ $\left(g^L_{2,Y}, g^U_{2,Y}\right)$ be the lower and upper bounds respectively, of g_1 (g_2) in Y, namely,

$$g^L_{1,Y} = z^L_Y, \quad g^U_{1,Y} = z^U_Y, \quad g^L_{2,Y} = z^L_Y, \quad g^U_{2,Y} = z^U_Y.$$

The estimators $g^u_{1,Y}, g^o_{1,Y}, g^u_{2,Y}, g^o_{2,Y}$ are equal to g_1, g_2, the functions they estimate. Then, the schemes of estimators $(g^u_{1,Y}, g^o_{1,Y})$ and $(g^u_{2,Y}, g^o_{2,Y})$ converge to g_1 and g_2 pointwise with order γ_1 and γ_2, for any $\gamma_1, \gamma_2 > 0$. For any interval $Y \subset Z$, the lower and upper bounds $g^L_{1,Y}, g^U_{1,Y}, g^L_{2,Y}, g^U_{2,Y}$ agree with the minima and maxima of g_1 and g_2 on Y.

We will show that the scheme of estimators of Proposition 3 converge to g pointwise with convergence order at most quadratic. To show this, it is enough to prove a quadratic lower bound of the distance between g and $g^U_{1,Y}$ on intervals centered at the origin.

Let $Y = [-\varepsilon, \varepsilon]$ be an interval of Z. For this interval, the intermediate functions $g_{a1,Y}, g_{a2,Y}, g_{b1,Y}, g_{b2,Y}$ of Proposition 3 are equal to

$$\begin{aligned}
g_{a1,Y}(z) &= \min\left\{g^L_{2,Y} g^u_{1,Y}(z), g^L_{2,Y} g^o_{1,Y}(z)\right\} = \min\{-\varepsilon z, -\varepsilon z\} = -\varepsilon z,\\
g_{a2,Y}(z) &= \min\left\{g^L_{1,Y} g^u_{2,Y}(z), g^L_{1,Y} g^o_{2,Y}(z)\right\} = \min\{-\varepsilon z, -\varepsilon z\} = -\varepsilon z,\\
g_{b1,Y}(z) &= \min\left\{g^U_{2,Y} g^u_{1,Y}(z), g^U_{2,Y} g^o_{1,Y}(z)\right\} = \min\{\varepsilon z, \varepsilon z\} = \varepsilon z,\\
g_{b2,Y}(z) &= \min\left\{g^U_{1,Y} g^u_{2,Y}(z), g^U_{1,Y} g^o_{2,Y}(z)\right\} = \min\{\varepsilon z, \varepsilon z\} = \varepsilon z.
\end{aligned} \tag{11}$$

The convex underestimator of g in Y of Proposition 3 is equal to

$$\begin{aligned}
g^u_Y(z) &= \max\left\{g_{a1,Y}(z) + g_{a2,Y}(z) - g^L_{1,Y} g^L_{2,Y}, g_{b1,Y}(z) + g_{b2,Y}(z) - g^U_{1,Y} g^U_{2,Y}\right\}\\
&= \max\left\{-2\varepsilon z - \varepsilon^2, 2\varepsilon z - \varepsilon^2\right\}.
\end{aligned}$$

The value of the underestimator function on the origin is $g^u_Y(0) = -\varepsilon^2$. Since $g(0) = 0$, it then follows that the pointwise convergence of g^u_Y to g is at most quadratic on the size of Y:

$$\sup_{z\in Y} |g(z) - g^u_Y(z)| \geq |g(0) - g^u_Y(0)| = \varepsilon^2 = \frac{w(Y)^2}{4}.$$

4.3 McCormick's relaxation of composition of functions

McCormick [24,25] provided a relaxation result for composition of functions. In what follows, we use the following function: given three numbers $a \leq b \leq c \in \mathbb{R}$, we define $\text{mid}\{a, b, c\} = b$.

Theorem 5 (McCormick's Composition Theorem) [24,26]. *Let $Z \subset \mathbb{R}^n$ and $X \subset \mathbb{R}$ be two nonempty convex sets. Consider the composite function $g = F \circ f$ where $f : Z \to \mathbb{R}$ is continuous, $F : X \to \mathbb{R}$, and let $f(Z) \subset X$. Suppose that a convex underestimator $f^u : Z \to \mathbb{R}$ and a concave overestimator $f^o : Z \to \mathbb{R}$ of f on Z are known. Let $F^u : X \to \mathbb{R}$ be a convex underestimator of F on X, let $F^o : X \to \mathbb{R}$ be a concave overestimator of F on X.*

Let $x^{\min} \in X$ be a point at which F^u attains its minimum on X, and let $x^{\max} \in X$ be a point at which F^o attains its maximum on X. Then, $g^u : Z \to \mathbb{R}$,

$$g^u(\mathbf{z}) = F^u\left(\mathrm{mid}\left\{f^u(\mathbf{z}), f^o(\mathbf{z}), x^{\min}\right\}\right)$$

is a convex underestimator of g on Z. Moreover, $g^o : Z \to \mathbb{R}$

$$g^o(\mathbf{z}) = F^o\left(\mathrm{mid}\left\{f^u(\mathbf{z}), f^o(\mathbf{z}), x^{\max}\right\}\right)$$

is a concave overestimator of g on Z.

We now restate McCormick's Composition Theorem in terms of schemes of estimators.

Theorem 6 *Let $Z \subset \mathbb{R}^n$ and $X \subset \mathbb{R}$ be two nonempty convex sets. Consider the composite function $g = F \circ f$ where $f : Z \to \mathbb{R}$ is continuous, $F : X \to \mathbb{R}$, and let $f(Z) \subset X$. Let $\left(f_Y^u, f_Y^o\right)_{Y \subset Z}$ be a scheme of continuous estimators of f in Z, and let H_f be the inclusion function associated with this scheme. Assume that an inclusion function $T : Y \subset Z \to Q \subset X$ of f is known. Moreover, assume that T is conservative enough so as to estimate $H_f(Y)$ also, that is, $H_f(Y) \subset T(Y)$ for any $Y \in \mathbb{IR}^n$, $Y \subset Z$.*

Let $\left(F_Q^u, F_Q^o\right)_{Q \subset X}$ be a scheme of estimators of F on X. For each interval $Q \subset X$, let $x_Q^{\min} \in Q$ be a point at which F_Q^u attains its minimum on Q, and let $x_Q^{\max} \in Q$ be a point at which F_Q^o attains its maximum on Q.

For each $Y \in \mathbb{IR}^n$, $Y \subset Z$ we define the following functions $g_Y^u : Y \to \mathbb{R}$ and $g_Y^o : Y \to \mathbb{R}$:

$$g_Y^u(\mathbf{z}) = F_{T(Y)}^u\left(\mathrm{mid}\left\{f_Y^u(\mathbf{z}), f_Y^o(\mathbf{z}), x_{T(Y)}^{\min}\right\}\right),$$

and

$$g_Y^o(\mathbf{z}) = F_{T(Y)}^o\left(\mathrm{mid}\left\{f_Y^u(\mathbf{z}), f_Y^o(\mathbf{z}), x_{T(Y)}^{\max}\right\}\right).$$

Then, for each $Y \in \mathbb{IR}^n$, $Y \subset Z$, the function g_Y^u (g_Y^o) is a convex underestimator (concave overestimator resp.) of g in Y. In other words, the set of functions $\left(g_Y^u, g_Y^o\right)_{Y \subset Z}$ is a scheme of estimators of g in Z.

The assumption that the inclusion function T also estimates H_f is required to exclude a domain violation of g_Y^u or g_Y^o, see Example 3.1 of [26]. Moreover, we note that this assumption also implies that the Hausdorff convergence order of H_f is at least the Hausdorff convergence order of T.

In order to associate an inclusion function to the scheme of estimators $\left(g_Y^u, g_Y^o\right)_{Y \subset Z}$, we note that, for each interval Y and each point $\mathbf{z} \in Y$, the points $\mathrm{mid}\left\{f_Y^u(\mathbf{z}), f_Y^o(\mathbf{z}), x_{T(Y)}^{\min}\right\}$ and $\mathrm{mid}\left\{f_Y^u(\mathbf{z}), f_Y^o(\mathbf{z}), x_{T(Y)}^{\max}\right\}$ are between $f_Y^u(\mathbf{z})$ and $f_Y^o(\mathbf{z})$, and therefore they belong to the interval $H_f(Y)$. Then, we can associate an inclusion function H_g to the scheme of estimators $\left(g_Y^u, g_Y^o\right)_{Y \subset Z}$ as follows.

$$H_g : Y \in \mathbb{IR}^n, \quad Y \subset Z \to \mathbb{IR}$$
$$H_g(Y) = H_{F_{T(Y)}}\left(H_f(Y)\right).$$

In Theorems 7 and 8 we analyze the convergence order of the schemes defined by McCormick's composition method. In their proofs we make use of the following two Lemmata.

Lemma 2 *Let* $f : Z \subset \mathbb{R} \to \mathbb{R}$ *be a Lipschitz continuous function with Lipschitz constant* $M \in \mathbb{R}$. *Let* $A, B \in \mathbb{IR}$ *be two subsets of* Z. *Then*

$$q\left(\bar{f}(A), \bar{f}(B)\right) \leq Mq\,(A, B)\,,$$

where $q\,(X, Y)$ *denotes the Hausdorff metric.*

Proof In what follows we use the equivalent definition of Hausdorff metric between two sets X and Y given in Proposition 1, Eq. (1). We can express the distance between $\bar{f}(A)$ and $\bar{f}(B)$ as follows.

$$\begin{aligned} q\left(\bar{f}(A), \bar{f}(B)\right) &= \max\left\{ \sup_{\mathbf{c}\in\bar{f}(A)} \inf_{\mathbf{d}\in\bar{f}(B)} |\mathbf{c}-\mathbf{d}|\,,\ \sup_{\mathbf{d}\in\bar{f}(B)} \inf_{\mathbf{c}\in\bar{f}(A)} |\mathbf{c}-\mathbf{d}| \right\} \\ &= \max\left\{ \sup_{\mathbf{a}\in A} \inf_{\mathbf{b}\in B} |f(\mathbf{a})-f(\mathbf{b})|\,,\ \sup_{\mathbf{b}\in B} \inf_{\mathbf{a}\in A} |f(\mathbf{a})-f(\mathbf{b})| \right\}. \end{aligned} \tag{12}$$

Since f is Lipschitz continuous on Z with constant M, for any $\mathbf{a}\in A$ and $\mathbf{b}\in B$, $|f(\mathbf{a})-f(\mathbf{b})| \leq M|\mathbf{a}-\mathbf{b}|$. Applying this bound to the right-hand side of Equality (12), we bound $q\left(\bar{f}(A), \bar{f}(B)\right)$ as follows.

$$\begin{aligned} q\left(\bar{f}(A), \bar{f}(B)\right) &= \max\left\{ \sup_{\mathbf{a}\in A} \inf_{\mathbf{b}\in B} |f(\mathbf{a})-f(\mathbf{b})|\,,\ \sup_{\mathbf{b}\in B} \inf_{\mathbf{a}\in A} |f(\mathbf{a})-f(\mathbf{b})| \right\} \\ &\leq M \max\left\{ \sup_{\mathbf{a}\in A} \inf_{\mathbf{b}\in B} |\mathbf{a}-\mathbf{b}|\,,\ \sup_{\mathbf{b}\in B} \inf_{\mathbf{a}\in A} |\mathbf{a}-\mathbf{b}| \right\} = Mq(A, B). \end{aligned}$$

□

Lemma 3 *Let* $Z \in \mathbb{IR}^n$ *be a bounded interval. Let* $f : Z \subset \mathbb{R}^n \to \mathbb{R}$ *be a Lipschitz continuous function with Lipschitz constant* $M \in \mathbb{R}$. *Let* $H_f : Y \subset Z \to \mathbb{IR}$ *be an inclusion function of* f *on* Z *with Hausdorff convergence order* $\beta_f \geq 1$. *Then, there exists a constant* $C > 0$ *such that, for any interval* $Y \subset Z$, $w(H_f(Y)) \leq Cw(Y)$.

Proof Since H_f has Hausdorff convergence of order β_f, there exists a constant $\tau_f > 0$ such that, for any interval $Y \subset Z$, $q(H_f(Y), \bar{f}(Y)) \leq \tau_f w(Y)^{\beta_f}$. In particular, $H_f(Y) \subset \bar{f}(Y) + [-\tau_f w(Y)^{\beta_f}, \tau_f w(Y)^{\beta_f}]$. This inclusion between intervals implies that

$$\begin{aligned} w(H_f(Y)) &\leq w\left(\bar{f}(Y) + \left[-\tau_f w(Y)^{\beta_f}, \tau_f w(Y)^{\beta_f}\right]\right) \\ &\leq w\left(\bar{f}(Y)\right) + w\left(\left[-\tau_f w(Y)^{\beta_f}, \tau_f w(Y)^{\beta_f}\right]\right) \\ &\leq w\left(\bar{f}(Y)\right) + 2\tau_f w(Y)^{\beta_f}. \end{aligned} \tag{13}$$

Since f is Lipschitz continuous on Z with constant M, Lemma 1 implies that $w(\bar{f}(Y)) \leq Mw(Y)$. Combining this bound with Inequality (13), we obtain

$$\begin{aligned} w(H_f(Y)) &\leq Mw(Y) + 2\tau_f w(Y)^{\beta_f} \\ &= \left(M + 2\tau_f w(Y)^{\beta_f-1}\right) w(Y) \leq \left(M + 2\tau_f w(Z)^{\beta_f-1}\right) w(Y), \end{aligned}$$

where in the last inequality $\beta_f - 1 \geq 0$ is used. We note that the constant $C \equiv \left(M + 2\tau_f w(Z)^{\beta_f-1}\right)$ depends on Z but not on Y, and therefore the inequality

$$w\left(H_f(Y)\right) \leq Cw(Y)$$

holds for any interval $Y \subset Z$. □

4.3.1 Convergence in the Hausdorff metric

We next show that the Hausdorff convergence order of the composite inclusion function H_g is equal to the Hausdorff convergence order of the range estimator T or of the outer estimator H_F, whichever is smaller. In this result, the Hausdorff convergence order of H_g does not depend on the Hausdorff convergence order of the inner estimator H_f; this is so since, under the assumption that the inclusion function T also estimates H_f, the Hausdorff convergence order of T dominates, i.e., is weaker than, the Hausdorff convergence order of H_f.

Theorem 7 *Let $Z \subset \mathbb{R}^n$ and $X \subset \mathbb{R}$ be two nonempty convex sets. Let $f : Z \to X$ and $F : X \to \mathbb{R}$ be two Lipschitz continuous functions with Lipschitz constants M_f and M_F respectively, and such that $f(Z) \subset X$. Let $\left(f_Y^u, f_Y^o\right)_{Y\subset Z}$ be a scheme of continuous estimators of f in Z. Let T be an inclusion function of f in Z. Assume furthermore that T also overestimates the range of the scheme $\left(f_Y^u, f_Y^o\right)_{Y\subset Z}$, namely, $H_f(Y) \subset T(Y)$ for any $Y \subset Z$. Let $\left(F_Q^u, F_Q^o\right)_{Q\subset X}$ be a scheme of continuous estimators of F on X. Furthermore, assume the following:*

1. *The inclusion function T of f in Z has a Hausdorff convergence order $\beta_{f,T} \geq 1$.*
2. *The inclusion function H_F of F in X associated to $\left(F_Q^u, F_Q^o\right)_{Q\subset X}$ has a Hausdorff convergence order β_F.*

Let $g = F \circ f$ be the composite function of F and f. Let $\left(g_Y^u, g_Y^o\right)_{Y\subset Z}$ be the scheme of McCormick's estimators of g in Z constructed from the schemes $\left(f_Y^u, f_Y^o\right)_{Y\subset Z}$ and $\left(F_Q^u, F_Q^o\right)_{Q\subset X}$, and the inclusion function T as in Theorem 6, and let H_g be its corresponding inclusion function. Then, H_g has a Hausdorff convergence order $\min\left\{\beta_{f,T}, \beta_F\right\}$.

Proof Let $Y \subset Z$ be an interval. We are interested in bounding the term $q\left(H_g(Y), \bar{g}(Y)\right) = q\left(H_{F_{T(Y)}}(H_f(Y)), \bar{F}(\bar{f}(Y))\right)$.

We first observe that, since $H_f(Y) \subset T(Y)$, it follows that $H_{F_{T(Y)}}(H_f(Y)) \subset H_{F_{T(Y)}}(T(Y))$. Moreover, since $\bar{F}(\bar{f}(Y)) \subset H_{F_{T(Y)}}(H_f(Y))$, it follows that

$$q\left(H_{F_{T(Y)}}(H_f(Y)), \bar{F}(\bar{f}(Y))\right) \leq q\left(H_{F_{T(Y)}}(T(Y)), \bar{F}(\bar{f}(Y))\right).$$

We next bound $q\left(H_{F_{T(Y)}}(T(Y)), \bar{F}(\bar{f}(Y))\right)$. Applying the triangle inequality to the Hausdorff metric q,

$$q\left(H_{F_{T(Y)}}(T(Y)), \bar{F}(\bar{f}(Y))\right) \leq q\left(H_{F_{T(Y)}}(T(Y)), \bar{F}(T(Y))\right) + q\left(\bar{F}(T(Y)), \bar{F}(\bar{f}(Y))\right). \tag{14}$$

We next bound the two terms of the right hand side of this inequality.

The inclusion function H_F converges to F with order β_F, and therefore

$$q\left(H_{F_{T(Y)}}(T(Y)), \bar{F}(T(Y))\right) \leq \tau_F\left(w(T(Y))\right)^{\beta_F}. \tag{15}$$

By hypothesis, T converges to f on Z with order $\beta_{f,T}$. In particular, Lemma 3 implies that $w(T(Y)) \leq C_T w(Y)$ for some constant C_T that does not depend on Y. Therefore, we can bound Inequality (15) as follows

$$q\left(H_{F_{T(Y)}}(T(Y)), \bar{F}(T(Y))\right) \leq \tau_F C_T^{\beta_F}(w(Y))^{\beta_F}. \tag{16}$$

Now we bound the second term of the right hand side of Inequality (14). Since F is a Lipschitz function with constant M_F, Lemma 2 implies that

$$q\left(\bar{F}(T(Y)), \bar{F}(\bar{f}(Y))\right) \leq M_F q\left(T(Y), \bar{f}(Y)\right). \tag{17}$$

Since T converges to f on Z with order $\beta_{f,T}$, it then follows that

$$q\left(\bar{F}(T(Y)), \bar{F}(\bar{f}(Y))\right) \leq M_F \tau_T (w(Y))^{\beta_{f,T}}. \tag{18}$$

Combining Inequality (14) with the bounds of Inequalities (16) and (18), we prove that H_g has Hausdorff convergence of order $\min\{\beta_{f,T}, \beta_F\}$, namely,

$$q\left(H_{F_{T(Y)}}(T(Y)), \bar{g}(Y)\right) \leq \tau_F C_T^{\beta_F}(w(Y))^{\beta_F} + M_F \tau_T (w(Y))^{\beta_{f,T}}. \qquad \square$$

In other words, McCormick's composition method is determined by the smallest of the Hausdorff convergence orders of the outer scheme and of the inclusion function T, and the convergence order of the inner scheme is irrelevant. In particular, having inner and outer schemes with high convergence order is not enough to guarantee a high convergence order of the composition scheme when the convergence order of the inclusion function T is not high enough.

The following example, based on a well-known result from interval extensions [8] shows that indeed, the convergence order of the inclusion function T determines the convergence order of the composition scheme.

Example 5 Consider $Z = [0, 1]$ and $g : Z \to \mathbb{R}$, $g(z) = |z - z^2 - 0.25|$. A natural decomposition is $g(z) = F(f(z))$ with $f : Z \to \mathbb{R}$, $f(z) = z - z^2 - 0.25$ and $F : X \to \mathbb{R}$, $F(x) = |x|$, with $X \subset \mathbb{R}$. We will consider the convergence rate for the intervals $Y \subset Z$, s.t., $Y = [0.5 - \varepsilon_1, 0.5 + \varepsilon_1]$. The image of the inner function is $\bar{f}(Y) = \left[-\varepsilon_1^2, 0\right]$ and therefore the image of the composite function is $\bar{g}(Y) = \left[0, \varepsilon_1^2\right]$.

The inner function f is a variation of a well-known example [8] for which natural interval extensions have a convergence of first order and give the inclusion function $T(Y) = \left[-2\varepsilon_1 - \varepsilon_1^2, 2\varepsilon_1 - \varepsilon_1^2\right]$. In contrast, natural interval extensions for the centered form $f_{cen}(z) = -(z - 0.5)^2$ have a convergence of second order, giving $T'(Y) = \left[-\varepsilon_1^2, \varepsilon_1^2\right]$. As a consequence, natural interval extension gives linear convergence for the composite function, whereas the centered form gives the exact range.

Consider now the McCormick relaxations. The inner function is itself a sum of monomials, and therefore the McCormick relaxation is the sum of the relaxations. The concavity is identified, and the relaxations are the envelopes. For $Y = [0.5 - \varepsilon_1, 0.5 + \varepsilon_1]$ these are given by $f^{cv}(z) = -\varepsilon_1^2$ and $f^{cc}(z) = z - z^2 - 0.25$. As a consequence the scheme of McCormick estimators is exact in the Hausdorff metric and has a quadratic pointwise convergence. The outer function $|\cdot|$ is convex and univariate. Therefore, the convex underestimator is the function itself, and the concave overestimator the secant. The associated scheme of estimators is exact in the Hausdorff metric but has only linear pointwise convergence for intervals containing zero.

In accordance with Theorem 7 the convergence order of the scheme of McCormick estimators for the composite function is linear when natural interval extensions are used for the calculation of the inclusion function of the inner function f. When the centered form is used for the inner function f, the McCormick estimators of the composite function g are exact.

4.3.2 Pointwise convergence

We next analyze the pointwise convergence order of McCormick's composition scheme.

Theorem 8 *Let $Z \subset \mathbb{R}^n$ and $X \subset \mathbb{R}$ be two nonempty convex sets. Let $f : Z \to X$ and $F : X \to \mathbb{R}$ be two Lipschitz continuous functions with Lipschitz constants M_f and M_F respectively, such that $f(Z) \subset X$. Let $\left(f_Y^u, f_Y^o\right)_{Y\subset Z}$ be a scheme of continuous estimators of f in Z as defined in Definition 6. Let T be an inclusion function of f in Z. Assume furthermore that T also overestimates the range of the scheme $\left(f_Y^u, f_Y^o\right)_{Y\subset Z}$, namely, $H_f(Y) \subset T(Y)$ for any $Y \subset Z$. Let $\left(F_Q^u, F_Q^o\right)_{Q\subset X}$ be a scheme of continuous estimators of F on X as in Definition 6. Furthermore, assume the following:*

1. *The scheme $\left(f_Y^u, f_Y^o\right)_{Y\subset Z}$ has pointwise convergence of order γ_f.*
2. *The inclusion function T of f on Z has a Hausdorff convergence order $\beta_{f,T} \geq 1$.*
3. *The scheme $\left(F_Q^u, F_Q^o\right)_{Q\subset X}$ has pointwise convergence of order γ_F.*

Let $g = F \circ f$ be the composite function of F and f. Let $\left(g_Y^u, g_Y^o\right)_{Y\subset Z}$ be the scheme of estimators of g in Z constructed from the schemes $\left(f_Y^u, f_Y^o\right)_{Y\subset Z}$ and $\left(F_Q^u, F_Q^o\right)_{Q\subset X}$, and from T, as in Theorem 6. Then, $\left(g_Y^u, g_Y^o\right)_{Y\subset Z}$ has also pointwise convergence of order $\min\{\gamma_f, \gamma_F\}$.

Proof We will show that there exists a constant $\tau_g > 0$ such that, for any interval $Y \subset Z$ and any point $\mathbf{z} \in Y$, $|g(\mathbf{z}) - g_Y^u(\mathbf{z})| \leq \tau_g w(Y)^{\min\{\gamma_f,\gamma_F\}}$. A similar bound will hold for $|g(\mathbf{z}) - g_Y^o(\mathbf{z})|$, and therefore the proof of this case is omitted.

Let $Y \subset Z$. Let $\mathbf{z}$ be a point of Y. To simplify notation, we denote by $x^{\text{mid}}(\mathbf{z})$ the point $x^{\text{mid}}(\mathbf{z}) \equiv \text{mid}\left\{f_Y^u(\mathbf{z}), f_Y^o(\mathbf{z}), x_{T(Y)}^{\min}\right\}$, and so $g_Y^u(\mathbf{z}) = F_{T(Y)}^u\left(x^{\text{mid}}(\mathbf{z})\right)$. We note that $x^{\text{mid}}(\mathbf{z})$ also depends on $T(Y)$ and $F_{T(Y)}^u$, but to keep the notation simple we omit these dependencies.

We bound the distance between $g(\mathbf{z})$ and $g_Y^u(\mathbf{z})$ as follows.

$$\begin{aligned}|g(\mathbf{z}) - g_Y^u(\mathbf{z})| &= \left|\left(g(\mathbf{z}) - F\left(x^{\text{mid}}(\mathbf{z})\right)\right) + \left(F(x^{\text{mid}}(\mathbf{z})) - g_Y^u(\mathbf{z})\right)\right| \\ &\leq \left|g(\mathbf{z}) - F\left(x^{\text{mid}}(\mathbf{z})\right)\right| + \left|F\left(x^{\text{mid}}(\mathbf{z})\right) - g_Y^u(\mathbf{z})\right|. \end{aligned} \tag{19}$$

We next bound each term of the right hand side of this inequality.

We note that, by definition, $x^{\text{mid}}(\mathbf{z})$ is between $\min\{f_Y^u(\mathbf{z}), f_Y^o(\mathbf{z})\}$ and $\max\{f_Y^u(\mathbf{z}), f_Y^o(\mathbf{z})\}$. In particular,

$$|x^{\text{mid}}(\mathbf{z}) - f(\mathbf{z})| \leq \max\left\{|f_Y^u(\mathbf{z}) - f(\mathbf{z})|, |f_Y^o(\mathbf{z}) - f(\mathbf{z})|\right\}.$$

Since the scheme $\left(f_Y^u, f_Y^o\right)$ has pointwise convergence of order γ_f, there exists a constant $\tau_f > 0$ that does not depend on Y such that

$$\begin{aligned}|f_Y^u(\mathbf{z}) - f(\mathbf{z})| &\leq \tau_f w(Y)^{\gamma_f}, \text{ and} \\ |f_Y^o(\mathbf{z}) - f(\mathbf{z})| &\leq \tau_f w(Y)^{\gamma_f}.\end{aligned}$$

Therefore,

$$|x^{\text{mid}}(\mathbf{z}) - f(\mathbf{z})| \leq \tau_f w(Y)^{\gamma_f}. \tag{20}$$

Since F is a Lipschitz continuous function with Lipschitz constant M_F, and because of Inequality (20), the first term of the right hand side of (19) is at most

$$\begin{aligned}\left|g(\mathbf{z}) - F\left(x^{\text{mid}}(\mathbf{z})\right)\right| &= \left|F(f(\mathbf{z})) - F\left(x^{\text{mid}}(\mathbf{z})\right)\right| \\ &\leq M_F \left|f(\mathbf{z}) - x^{\text{mid}}(\mathbf{z})\right| \leq M_F \tau_f w(Y)^{\gamma_f}. \end{aligned} \tag{21}$$

Since the scheme $\left(F_Q^u, F_Q^o\right)$ has pointwise convergence of order γ_F, we can bound the second term of the right hand side of (19) as follows.

$$\left|F\left(x^{\text{mid}}(\mathbf{z})\right) - g_Y^u(\mathbf{z})\right| = \left|F\left(x^{\text{mid}}(\mathbf{z})\right) - F_{T(Y)}^u\left(x^{\text{mid}}(\mathbf{z})\right)\right| \le \tau_F(w(T(Y)))^{\gamma_F}.$$

The inclusion function T converges to f with order $\beta_{f,T} \ge 1$, and therefore, by Lemma 3, $w(T(Y)) \le Cw(\bar{f}(Y))$ for a constant $C > 0$ that does not depend on Y. Moreover, since f is a Lipschitz continuous function with Lipschitz constant M_f, the term $w(\bar{f}(Y))$ is at most $M_f w(Y)$ (see Lemma 1), and therefore

$$\left|F\left(x^{\text{mid}}(\mathbf{z})\right) - g_Y^u(\mathbf{z})\right| \le \tau_F(CM_f)^{\gamma_F} w(Y)^{\gamma_3}. \tag{22}$$

Combining the bounds of (21) and (22), it then follows that $|g(\mathbf{z}) - g_Y^u(\mathbf{z})|$ is at most

$$M_F\tau_f w(Y)^{\gamma_f} + \tau_F(CM_f)^{\gamma_F} w(Y)^{\gamma_F} = \tau_g w(Y)^{\min\{\gamma_f,\gamma_F\}},$$

for a constant τ_g that does not depend on $\mathbf{z}$ or on Y. A similar bound holds for $|g(\mathbf{z}) - g_Y^o(\mathbf{z})|$, and therefore, $\left(g_Y^u, g_Y^o\right)_{Y\subset Z}$ has pointwise convergence of order $\min\{\gamma_f, \gamma_F\}$. □

The following example is a variation of Example 5 in which the McCormick estimators result in an improvement of the convergence order compared to natural interval extensions. The distinguishing difference is the quadratic pointwise convergence order of the outer function estimators.

Example 6 Consider $Z = [0, 1]$ and $g : Z \to \mathbb{R}$, $g(z) = (z - z^2 - 0.25)^2$. A natural decomposition is $g(z) = F(f(z))$ with $f : Z \to \mathbb{R}$, $f(z) = z - z^2 - 0.25$ and $F : X \to \mathbb{R}$, $F(x) = x^2$, with $X \subset \mathbb{R}$. Similarly to Example 5 we will consider the convergence rate for the intervals $Y \subset Z$, s.t., $Y = [0.5 - \varepsilon_1, 0.5 + \varepsilon_1]$. The image of the inner function is $\bar{f}(Y) = [-\varepsilon_1^2, 0]$ and therefore the image of the composite function is $\bar{g} = [0, \varepsilon_1^4]$.

Recall the discussion on the inner function and its estimators based on interval arithmetic and McCormick relaxations. The outer function $(\cdot)^2$ is convex and univariate. Therefore, the convex underestimator is the function itself, and the concave overestimator the secant. The associated scheme of estimators is exact in the Hausdorff metric and has quadratic pointwise convergence for symmetric intervals around zero.

In accordance with Theorem 8 the convergence order of the scheme of McCormick estimators for the composite function is superlinear even when natural interval extensions are used for the calculation of the inclusion function of the inner function f. When the centered form is used for the inner function f, the McCormick estimators of the composite function g are exact in the Hausdorff metric.

5 αBB relaxations

Floudas and coworkers [1–4,9,18–23] introduced the αBB relaxations of a function, which is an alternative method for constructing convex and concave estimators of functions. Given a $\mathcal{C}^2$ function $f : Z \subset \mathbb{R}^n \to \mathbb{R}$, its αBB relaxation consists of adding a (nonpositive) quadratic term to get a convex underestimator of f. Note also the extension by Zlobec [39] to Lipschitz continuous functions. For a sufficiently large $\alpha^u \ge 0$, the Hessian of the function

$$f^{u,\alpha}(\mathbf{z}) = f(\mathbf{z}) + \alpha^u \sum_{i=1}^{n} \left(z_i - z_i^L\right)\left(z_i - z_i^U\right)$$

is positive semidefinite on Z and therefore $f^{u,\alpha}(\mathbf{z})$ is a convex function. In particular, since the added terms are nonpositive on Z, it is a convex underestimator of f in Z. Similarly, for a sufficiently large $\alpha^o \geq 0$, the function

$$f^{o,\alpha}(\mathbf{z}) = f(\mathbf{z}) - \alpha^o \sum_{i=1}^{n} \left(z_i - z_i^L\right)\left(z_i - z_i^U\right)$$

is a concave overestimator of f in Z. For the theory developed, we assume $\alpha^u = \alpha^o = \alpha$, without loss of generality. For the numerical examples separate constants are calculated.

Here, the simpler variant of αBB is assumed using the same α for all components of $\mathbf{z}$. This results in general in weaker relaxations than possible. Also, the function f is taken as a whole without decomposing it in terms such as bilinear, univariate concave, etc., as done in the original work by Floudas and coworkers.

Note that the Hessian of the relaxations $f^{u,\alpha}(\mathbf{z})$ and $f^{o,\alpha}(\mathbf{z})$ do not explicitly depend on the values of $\mathbf{z}^L$ and $\mathbf{z}^U$, but typically depend implicitly by the calculation of α. We can restate the αBB relaxations in terms of scheme of estimators as follows. Note that we consider a constant α and thus the analysis is conservative.

Definition 11 Let $f : Z \subset \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^2$ function. Let I_{ij} denote the elements of the identity matrix in $\mathbb{R}^n \times \mathbb{R}^n$. Let $\alpha > 0$ be such that the matrices $\left(\frac{\partial^2 f}{\partial z_i \partial z_j} + 2\alpha I_{ij}\right)_{1\leq i,j\leq n}$ and $\left(\frac{\partial^2 f}{\partial z_i \partial z_j} - 2\alpha I_{ij}\right)_{1\leq i,j\leq n}$ are positive semidefinite and negative semidefinite respectively for any $\mathbf{z}$ in Z. For each interval $Y = [\mathbf{z}_Y^L, \mathbf{z}_Y^U]$ of Z, let the αBB relaxations $f_Y^{u,\alpha}(\mathbf{z}) : Y \to \mathbb{R}$ and $f_Y^{o,\alpha}(\mathbf{z}) : Y \to \mathbb{R}$ be given by:

$$f_Y^{u,\alpha}(\mathbf{z}) = f(\mathbf{z}) + \alpha \sum_{i=1}^{n} \left(z_i - z_{Y,i}^L\right)\left(z_i - z_{Y,i}^U\right),$$

and

$$f_Y^{o,\alpha}(\mathbf{z}) = f(\mathbf{z}) - \alpha \sum_{i=1}^{n} \left(z_i - z_{Y,i}^L\right)\left(z_i - z_{Y,i}^U\right).$$

Let $\left(f_Y^{u,\alpha}, f_Y^{o,\alpha}\right)_{Y\subset Z}$ be the scheme of estimators of f in Z defined by these functions.

In [22] it is stated that the αBB relaxation has quadratic convergence. In the next theorem, we prove a similar result, namely that the αBB scheme of estimators has pointwise quadratic convergence.

Theorem 9 *Let $f : Z \subset \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^2$ function. Let $\left(f_Y^{u,\alpha}, f_Y^{o,\alpha}\right)_{Y\subset Z}$ be a scheme of estimators defined as in Definition 11. Then, the scheme $\left(f_Y^{u,\alpha}, f_Y^{o,\alpha}\right)$ converges to f pointwise with order 2.*

Proof Let $Y = [\mathbf{z}_Y^L, \mathbf{z}_Y^U]$ be a fixed interval of Z. For any vector $\mathbf{z} \in Y$, we can bound the distance between $f(\mathbf{z})$ and $f_Y^{u,\alpha}(\mathbf{z})$ as follows.

$$\begin{aligned}\left|f(\mathbf{z}) - f_Y^{u,\alpha}(\mathbf{z})\right| &= \left|\alpha \sum_{i=1}^{n} \left(z_i - z_{Y,i}^L\right)\left(z_i - z_{Y,i}^U\right)\right| \leq \alpha \sum_{i=1}^{n} \left|z_i - z_{Y,i}^L\right|\left|z_i - z_{Y,i}^U\right| \\ &\leq \alpha n \max_{1\leq i\leq n} \left\{\left|z_{Y,i}^U - z_{Y,i}^L\right|\right\}^2 \leq \alpha n w(Y)^2.\end{aligned}$$

The constant $\alpha n > 0$ does not depend on $\mathbf{z}$ or Y, and therefore, $\left(f_Y^{u,\alpha}, f_Y^{o,\alpha}\right)$ converges to f pointwise with order 2. □

Since pointwise convergence implies Hausdorff convergence (Theorem 1), we have the following Corollary.

Corollary 1 *Let $f : Z \subset \mathbb{R}^n \to \mathbb{R}$ be a $\mathcal{C}^2$ function. Let $\left(f_Y^{u,\alpha}, f_Y^{o,\alpha}\right)_{Y \subset Z}$ be a scheme of αBB estimators defined as in Definition 11. Then, the associated inclusion function H_f of f has Hausdorff convergence of order 2.*

Note also that the estimates on α become tighter for decreasing diameter of Y and therefore the Hausdorff convergence can have higher than quadratic order. Note however, that by Theorem 2 the order of pointwise convergence cannot be higher than quadratic.

6 Convergence order of envelopes

The following Theorem uses the properties of αBB relaxations to give a positive result for the convex envelopes.

Theorem 10 *Let $Z \subset \mathbb{R}^n$ be a convex set and $f : Z \to \mathbb{R}$. Assume that f is a $\mathcal{C}^2$ function. Then, the scheme associated to the convex and concave envelopes has (at least) quadratic pointwise convergence.*

Proof Since f is a $\mathcal{C}^2$ function we can construct the αBB relaxations, Definition 11. By Theorem 9 the scheme associated with these relaxations has quadratic pointwise convergence order, i.e., there exists $\tau > 0$, such that for any $Y \subset Z$:

$$\sup_{\mathbf{z} \in Y} f(\mathbf{z}) - f^{u,\alpha}(\mathbf{z}) \leq \tau(w(Y))^2, \quad \sup_{\mathbf{z} \in Y} f^{o,\alpha}(\mathbf{z}) - f(\mathbf{z}) \leq \tau(w(Y))^2.$$

By definition, the convex/concave envelopes are at least as tight as the αBB relaxations

$$\forall \mathbf{z} \in Z : f^{u,env}(\mathbf{z}) \geq f^{u,\alpha}(\mathbf{z}), \quad f^{o,env}(\mathbf{z}) \leq f^{o,\alpha}(\mathbf{z})$$

and therefore for the chosen τ and any $Y \subset Z$

$$\sup_{\mathbf{z} \in Y} f(\mathbf{z}) - f^{u,env}(\mathbf{z}) \leq \tau(w(Y))^2, \quad \sup_{\mathbf{z} \in Y} f^{o,env}(\mathbf{z}) - f(\mathbf{z}) \leq \tau(w(Y))^2,$$

and thus the scheme associated to the convex and concave envelopes has quadratic pointwise convergence. □

Consequently, envelopes of smooth functions have at least quadratic convergence order in the Hausdorff metric. This implies that for factorable functions involving only smooth factors, one can ensure that the McCormick relaxations have quadratic convergence by utilizing the convex/concave envelopes of each factor. Note that this is the original proposal by McCormick [24]. As shown in Example 5 nonsmooth functions such as the absolute function can lead to linear pointwise convergence. Recall also, that by Theorem 2 the pointwise convergence order of a smooth nonlinear function cannot be higher than quadratic.

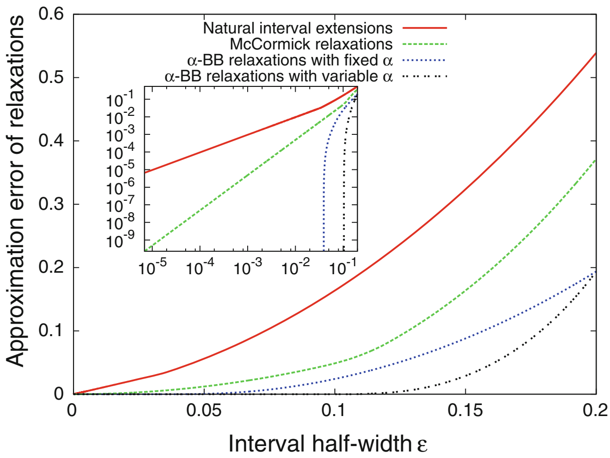

Fig. 1 Convergence in the Hausdorff metric of estimators for Example 7 in regular and double logarithmic scale

7 Numerical examples

In this section the convergence of the McCormick and αBB relaxations is studied numerically for a few small-scale problems. The McCormick relaxations are calculated by libMC [13,26], now superseded by MC++ [14]. The required constant α in the αBB relaxations is calculated via interval extensions of the Hessian using Gerschgorin's theorem, also via libMC. The entries of the Hessian are calculated analytically in Maple. Two subcases for the αBB relaxations are considered, namely for a uniform α, calculated for the entire host set Z, and for α calculated as a function of $Y \subset Z$. The alternative definition of convergence order, Definition 9 is used, i.e., using $w(H(Y)) - w(\bar{f}(Y))$ as a metric. The numerical examples illustrate the results developed and provide insight into the properties of the relaxation methods, but are not meant as a definitive comparison between these.

Example 7 Let $Z = [0.3, 0.7]$ and $Y \subset Z$, s.t. $Y = [0.5 - \varepsilon, 0.5 + \varepsilon]$ and consider $f : Z \to \mathbb{R}$, $f(z) = (z - z^2)(\log(z) + \exp(-z))$. Recall the discussion in the preceding examples and in [8] on the convergence order for the term $z - z^2$. Figure 1 shows the convergence in the Hausdorff metric of the estimators to the true function as $\varepsilon \to 0$. The αBB relaxation is tighter than the McCormick relaxations, which is not the case for larger Z. This suggests that the αBB relaxations converge faster. The double logarithmic plot suggests a linear convergence for the interval extension, quadratic for the McCormick relaxations and higher than quadratic for αBB.

Example 8 Let $Z = [-1, 1]$ and $Y \subset Z$, s.t. $Y = [-\varepsilon, \varepsilon]$ and consider $f : Z \to \mathbb{R}$, $f(z) = \exp(1 - z^2)$. To mimic the weak propagation of relaxations through complicated expressions the function is coded as $f(z) = \exp((1 - x)(1 - x))$. See also the discussion on relaxations of products of univariate functions in Maranas and Floudas [23]. As shown in Fig. 2, the αBB relaxations with a fixed α are weaker than the McCormick relaxations. The αBB relaxations with a variable α become exact for $\varepsilon_1 \lesssim 0.35$ since concavity is recognized. This is a well-known advantage of αBB relaxations, e.g., [16]. Moreover, for $0.35 \lesssim \varepsilon \lesssim 0.5$ the αBB relaxations are not exact, but tighter than the McCormick estimators, suggesting a

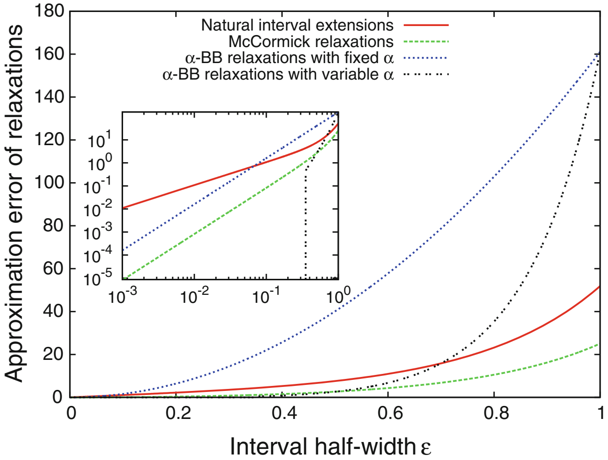

Fig. 2 Convergence in the Hausdorff metric of estimators for Example 8 in regular and double logarithmic scale

higher convergence order. The double logarithmic plot suggests a linear convergence in the Hausdorff metric for the interval extension, quadratic for the McCormick relaxations and the αBB with fixed α, but higher than quadratic for αBB with variable α.

Example 9 This example is problem 4 in Gatzke et al. [16], originally from Goldstein and Price [17]. Let $Z = [-1, 1]^2$ and $Y \subset Z$, s.t., $Y = [-\varepsilon, \varepsilon]^2$ and consider $f : Z \to \mathbb{R}$, s.t.,

$$\begin{aligned} f(\mathbf{z}) = &\left(1 + (z_1 + z_2)^2 \left(19 - 14z_1 + 3z_1^2 - 14(z_2 - 1) + 6z_1(z_2 - 1) + 3(z_2 - 1)^2\right)\right) \\ &\times \left(30 + (2z_1 - 3(z_2 - 1))^2 \left(18 - 32z_1 + 12z_1^2 + 48(z_2 - 1)\right.\right. \\ &\left.\left. - 36z_1(z_2 - 1) + 27(z_2 - 1)^2\right)\right). \end{aligned}$$

Note that x is replaced by z_1 and y by $z_2 - 1$, i.e., the second variable is shifted to be symmetric around 0. Recall that the simpler variant of αBB is assumed using the same α for all components of $\mathbf{z}$.

The behavior is very similar to Example 8. As shown in Fig. 3 for large host sets, the αBB relaxations are much weaker than McCormick, but they converge faster. If α is calculated for $\varepsilon \lesssim 0.0375$ then the αBB relaxations are tighter than the McCormick relaxations. Note that the values for α are positive, suggesting that the tightness is due to a higher convergence order in the Hausdorff metric as opposed to recognizing convexity/concavity. The double logarithmic plot suggests a linear convergence for the interval extension and quadratic for the McCormick relaxations and the αBB with fixed α. For αBB with variable α the convergence is higher than quadratic for $\varepsilon \in [0.001, 1]$.

8 Conclusions and future work

Theory for the convergence order is developed for the well-known convex relaxations by McCormick [24], based on the corresponding theory of interval inclusion functions. Pointwise convergence and convergence in the Hausdorff metric are considered. The framework is

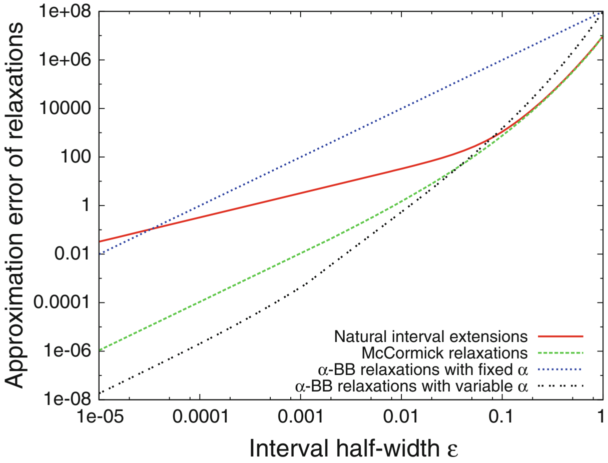

Fig. 3 Convergence in the Hausdorff metric of estimators for Example 9 in double logarithmic scale

also used to formalize the quadratic order of convergence for the αBB relaxations. Moreover, it is shown that the convergence order of the envelopes is at least quadratic pointwise. However, it is also demonstrated that any convex relaxations of nonsmooth nonlinear functions cannot have higher than quadratic pointwise convergence.

The convergence order of the McCormick relaxations depends on the convergence order of the relaxations of the factors, as well as the convergence order of the inclusion functions used to propagate the range of the functions. Table 1 summarizes the rules for the convergence order of the McCormick schemes. To achieve a high order of convergence for the estimators of a function it is necessary to have either interval inclusions of high convergence order, or relaxations to the factors with *pointwise* convergence of high order, e.g., use the envelopes. In contrast, it is not sufficient to have relaxations of the factors with high order of *convergence in the Hausdorff metric*. In a sense the order of relaxations is as weak as the underlying interval inclusions. This is proved theoretically and demonstrated in examples.

Schemes combining the αBB and the McCormick relaxations can achieve quadratic convergence. In particular, in the work by Gatzke et al. [16] the "simple hybrid reformulation" achieves quadratic convergence, because it considers the two relaxations simultaneously, thus selecting the tighter of the two. In other words, this hybrid method converges as fast as the fastest of αBB and McCormick (lowest curve in Figs. 1–3). Moreover, the "advanced hybrid reformulation" of Gatzke et al., in which αBB relaxations are used in the factors of the McCormick relaxations, also achieves quadratic convergence, since αBB has quadratic pointwise convergence.

It would be interesting to formally consider the convergence order of the relaxations involving auxiliary variables, [36] as well as for the γBB relaxations [5,6]. Furthermore, the implications of the convergence order to the global optimization of standard test problems should be studied.

The theory developed herein considers convergence in the limit. In the practice of global optimization it is also important to have tight estimators of nonconvex functions over big host sets, thus limiting branching early on. It is therefore of interest to "reverse" the convergence in the limit and consider how the estimation error grows with increasing diameter of the variable host set.

Table 1 Convergence order of McCormick schemes assuming Lipschitz continuity of the factors

Convergence order of factors	Resulting convergence order
Addition $g(\mathbf{z}) = f_1(\mathbf{z}) + f_2(\mathbf{z})$	
Scheme for f_i has β_i	No order propagation, $\beta = 1$ possible for $\beta_i \to \infty$
Scheme for f_i has $\gamma_i > 0$	$\gamma = \min\{\gamma_1, \gamma_2\}$
Multiplication $g(\mathbf{z}) = f_1(\mathbf{z}) \times f_2(\mathbf{z})$	
Scheme for f_i has β_i	No order propagation, $\beta = 1$ possible for $\beta_i \to \infty$
Scheme for f_i has $\gamma_i \geq 1$	$\gamma = \min\{\gamma_1, \gamma_2, 2\}$
Inclusions for f_i have $\beta_{i,T} \geq 1$	
Composition $g(\mathbf{z}) = F(f(\mathbf{z}))$	
Inclusion of f has $\beta_{f,T} \geq 1$	$\beta = \min\{\beta_{f,T}, \beta_F\}$
Scheme for F has β_F	
Scheme for f has γ_f	$\gamma = \min\{\gamma_f, \gamma_F\}$
Inclusion for f has $\beta_{f,T} \geq 1$	
Scheme for F has γ_F	

The factors are characterized by the convergence order in the Hausdorff metric β_i and/or the pointwise convergence order γ_i of the corresponding schemes for $i = 1, 2, f, F$. The additional subscript T denotes the inclusion function used to overestimate the range. The convergence order of the resulting scheme is characterized by the convergence order in the Hausdorff metric β and/or the pointwise convergence order γ. The expressions for β, γ are the smallest that are guaranteed; these bounds are sharp. Pointwise convergence is stronger than convergence in the Hausdorff metric

Acknowledgments We would like to thank Corey J. Noone for useful suggestions. We are grateful to Benoît Chachuat for providing libMC/MC++. We appreciate the thorough comments provided by the anonymous reviewers and editors which resulted in a significantly improved manuscript.

References

1. Adjiman, C.S., Androulakis, I.P., Floudas, C.A.: A global optimization method, αBB for general twice-differentiable constrained NLPs—II. Implementation and computational results. Comput. Chem. Eng. **22**(9), 1159–1179 (1998)
2. Adjiman, C.S., Androulakis, I.P., Maranas, C.D., Floudas, C.A.: A global optimization method, αBB for process design. Comput. Chem. Eng. **20**(Suppl A), S419–S424 (1996)
3. Adjiman, C.S., Dallwig, S., Floudas, C.A., Neumaier, A.: A global optimization method, αBB for general twice-differentiable constrained NLPs—I. Theoretical advances. Comput. Chem. Eng. **22**(9), 1137–1158 (1998)
4. Adjiman, C.S., Floudas, C.A.: Rigorous convex underestimators for general twice-differentiable problems. J. Glob. Optim. **9**(1), 23–40 (1996)
5. Akrotirianakis, I.G., Floudas, C.A.: Computational experience with a new class of convex underestimators: Box-constrained NLP problems. J. Glob. Optim. **29**(3), 249–264 (2004)
6. Akrotirianakis, I.G., Floudas, C.A.: A new class of improved convex underestimators for twice continuously differentiable constrained NLPs. J. Glob. Optim. **30**(4), 367–390 (2004)
7. Al-Khayyal, F.A., Falk, J.E.: Jointly constrained biconvex programming. Math. Oper. Res. **8**(2), 273–286 (1983)
8. Alefeld, G., Mayer, G.: Interval analysis: Theory and applications. J. Comput. Appl. Math. **121**(1–2), 421–464 (2000)
9. Androulakis, I.P., Maranas, C.D., Floudas, C.A.: αBB: A global optimization method for general constrained nonconvex problems. J. Glob. Optim. **7**(4), 337–363 (1995)
10. Belotti, P., Lee, J., Liberti, L., Margot, F., Wachter, A.: Branching and bounds tightening techniques for non-convex MINLP. Optim. Methods Softw. **24**(4–5), 597–634 (2009)

11. Bhattacharjee, B., Green, W.H. Jr.., Barton, P.I.: Interval methods for semi-infinite programs. Comput. Optim. Appl. **30**(1), 63–93 (2005)
12. Bhattacharjee, B., Lemonidis, P., Green, W.H. Jr., Barton, P.I.: Global solution of semi-infinite programs. Math. Program. Ser. B **103**(2), 283–307 (2005)
13. Chachuat, B.: libMC: A numeric library for McCormick relaxation of factorable functions. Documentation and Source Code available at: http://yoric.mit.edu/libMC/ (2007)
14. Chachuat, B.: MC++: A versatile library for McCormick relaxations and Taylor models. Documentation and Source Code available at: http://www3.imperial.ac.uk/people/b.chachuat/research (2010)
15. Du, K.S., Kearfott, R.B.: The cluster problem in multivariate global optimization. J. Glob. Optim. **5**(3), 253–265 (1994)
16. Gatzke, E.P., Tolsma, J.E., Barton, P.I.: Construction of convex function relaxations using automated code generation techniques. Optim. Eng. **3**(3), 305–326 (2002)
17. Goldstein, A.A., Price, J.F.: Descent from local minima. Math. Comput. **25**(115), 569–574 (1971)
18. Gounaris, C.E., Floudas, C.A.: Tight convex underestimators for C-2-continuous problems: I. Univariate functions. J. Glob. Optim. **42**(1), 51–67 (2008)
19. Gounaris, C.E., Floudas, C.A.: Tight convex underestimators for C-2-continuous problems: II. Multivariate functions. J. Glob. Optim. **42**(1), 69–89 (2008)
20. Maranas, C.D., Floudas, C.A.: A global optimization approach for Lennard-Jones microclusters. J. Chem. Phys. **97**(10), 7667–7678 (1992)
21. Maranas, C.D., Floudas, C.A.: Global optimization for molecular conformation problems. Ann. Oper. Res. **42**(3), 85–117 (1993)
22. Maranas, C.D., Floudas, C.A.: Global minimum potential energy conformations of small molecules. J. Glob. Optim. **4**, 135–170 (1994)
23. Maranas, C.D., Floudas, C.A.: Finding all solutions of nonlinearly constrained systems of equations. J. Glob. Optim. **7**(2), 143–182 (1995)
24. McCormick, G.P.: Computability of global solutions to factorable nonconvex programs: Part I. Convex underestimating problems. Math. Program. **10**(1), 147–175 (1976)
25. McCormick, G.P.: Nonlinear Programming: Theory, Algorithms and Applications. Wiley, New York (1983)
26. Mitsos, A., Chachuat, B., Barton, P.I.: McCormick-based relaxations of algorithms. SIAM J. Optim. **20**(2), 573–601 (2009)
27. Mitsos, A., Lemonidis, P., Lee, C.K., Barton, P.I.: Relaxation-based bounds for semi-infinite programs. SIAM J. Optim. **19**(1), 77–113 (2008)
28. Moore, R.: Methods and Applications of Interval Analysis. SIAM, Philadelphia, PA (1979)
29. Munkres, J.: Topology, 2nd edn. Prentice Hall, Englewood Cliffs (1999)
30. Neumaier, A.: Complete search in continuous global optimization and constraint satisfaction. Acta Numer. **13**, 271–369 (2004)
31. Ratschek, H., Rokne, J.: Computer Methods for the Range of Functions. Ellis Horwood Series, Mathematics and its Applications, New York (1984)
32. Ryoo, H.S., Sahinidis, N.V.: Analysis of bounds for multilinear functions. J. Glob. Optim. **19**(4), 403–424 (2001)
33. Sahinidis, N., Tawarmalani, M.: BARON. http://www.gams.com/solvers/baron.pdf (2005)
34. Schöbel, A., Scholz, D.: The theoretical and empirical rate of convergence for geometric branch-and-bound methods. J. Glob. Optim. **48**(3), 473–495 (2010)
35. Smith, E.M.B., Pantelides, C.C.: A symbolic reformulation/spatial branch-and-bound algorithm for the global optimisation of nonconvex MINLPs. Comput. Chem. Eng. **23**(4–5), 457–478 (1999)
36. Tawarmalani, M., Sahinidis, N.V.: Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming. Nonconvex Optimization and its Applications. Kluwer Academic Publishers, Boston (2002)
37. Tawarmalani, M., Sahinidis, N.V.: Global optimization of mixed-integer nonlinear programs: A theoretical and computational study. Math. Program. **99**(3), 563–591 (2004)
38. Tawarmalani, M., Sahinidis, N.V.: A polyhedral branch-and-cut approach to global optimization. Math. Program. **103**(2), 225–249 (2005)
39. Zlobec, S.: On the Liu-Floudas convexification of smooth programs. J. Glob. Optim. **32**(3), 401–407 (2005)