Differentiable McCormick relaxations

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

Differentiable McCormick relaxations

Kamil A. Khan · **Harry A. J. Watson** ·
Paul I. Barton

Received: date / Accepted: date

Abstract McCormick's classical relaxation technique constructs closed-form convex and concave relaxations of compositions of simple intrinsic functions. These relaxations have several properties which make them useful for lower bounding problems in global optimization: they can be evaluated automatically, accurately, and computationally inexpensively, and they converge rapidly to the relaxed function as the underlying domain is reduced in size. They may also be adapted to yield relaxations of certain implicit functions and differential equation solutions. However, McCormick's relaxations may be nonsmooth, and this nonsmoothness can create theoretical and computational obstacles when relaxations are to be deployed. This article presents a continuously differentiable variant of McCormick's original relaxations in the multivariate McCormick framework of Tsoukalas and Mitsos. Gradients of the new differentiable relaxations may be computed efficiently using the standard forward or reverse modes of automatic differentiation. Extensions to differentiable relaxations of implicit functions and solutions of parametric ordinary differential equations are discussed. A C++ implementation based on the library MC++ is described and applied to a case study in nonsmooth nonconvex optimization.

Keywords Nonconvex optimization · Convex underestimators · McCormick relaxations · Implicit functions

Mathematics Subject Classification (2000) 49M20 · 90C26 · 65G40 · 26B25

This material was supported by Novartis Pharmaceuticals as part of the Novartis-MIT Center for Continuous Manufacturing, was also supported by Statoil, and was based in part on work supported by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357.

K.A. Khan
Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL, USA.
E-mail: kkhan@anl.gov

H.A.J. Watson
Process Systems Engineering Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA.
E-mail: hwatson@mit.edu

P.I. Barton
Process Systems Engineering Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA.
Tel.: +1 617 253 6526, Fax: +1 617 258 5042
E-mail: pib@mit.edu

1 Introduction

Branch-and-bound methods for deterministic global optimization [27] require the ability to evaluate a lower bound on a nonconvex function on particular classes of subdomains. This bounding information may be generated using a relaxation scheme by McCormick [38], which evaluates convex underestimators of a nonconvex objective function on interval subdomains. McCormick's relaxation method assumes that the objective function can be expressed as a finite composition of known *intrinsic functions*, but assumes no other global knowledge of the function *a priori*. Subgradients may be computed for the McCormick relaxations using dedicated variants [40,6] of automatic differentiation [22,43]. Using this information, a lower bound on a nonconvex objective function on an interval may be supplied by minimizing the corresponding convex McCormick underestimator using a local optimization solver. Other methods for global optimization, such as nonconvex outer approximation [28] and nonconvex generalized Benders decomposition [32], also require the construction and minimization of convex underestimators.

McCormick's relaxation method has several useful properties. Firstly, accurate evaluation of a convex underestimator and a corresponding subgradient is computationally inexpensive and automatable; the C++ library `MC++` [13,40] and the Fortran library `amodMC` [6] use operator overloading to compute these quantities for well-defined user-supplied compositions of the basic arithmetic operations and functions such as $\sin/\cos$ and $\exp/\log$. Secondly, as the width of the interval on which a McCormick relaxation is constructed is reduced to zero, the relaxation approaches the objective function sufficiently rapidly [10] to mitigate a phenomenon called the *cluster effect* [16,67], in which a branch-and-bound method will branch many times on intervals that either contain or are near a global minimum. Since McCormick relaxations are constructed in closed form, they have been used to generate relaxations for implicit functions [68,60,55] and for solutions of parametric ordinary differential equations [57,56]. However, as the following example shows, McCormick's relaxations can be nondifferentiable.

Example 1 Let a function $\mathrm{mid} : \mathbb{R}^3 \to \mathbb{R}$ map to the median of its three scalar arguments, consider the smooth composite function $g : \mathbb{R} \to \mathbb{R} : z \mapsto \exp(z^3)$, and set $z^* := -1+\sqrt{3}$. As shown in [40, Example 2.1], the function $\underline{g}^{\mathrm{C}} : [-1,1] \to \mathbb{R}$ for which

$$\underline{g}^{\mathrm{C}} : z \mapsto \exp(\mathrm{mid}(z^3+3z^2-3, z^3-3z^2+3, -1)) = \begin{cases} \exp(-1), & \text{if } z \le z^*, \\ \exp(z^3+3z^2-3), & \text{if } z > z^*, \end{cases}$$

can be generated from g according to McCormick's rule [40, Section 3] for constructing convex relaxations of a composite function. (In this application of McCormick's rule, αBB relaxations [3] of the inner function $z \mapsto z^3$ have been employed.) Indeed, $\underline{g}^{\mathrm{C}}$ is convex on $[-1,1]$, and $\underline{g}^{\mathrm{C}}(z) \le g(z)$ for each $z \in [-1,1]$. However, even though $\underline{g}^{\mathrm{C}}$ satisfies McCormick's proposed sufficient condition for differentiability of a convex relaxation [38, p. 151], it is in fact nondifferentiable at z^*.

Several factors can introduce failure of differentiability of McCormick relaxations. As illustrated by the above example, the median function used in defining McCormick's composition rule is itself nondifferentiable. Secondly, any nondifferentiability in supplied relaxations of intrinsic functions can propagate to yield nondifferentiability in constructed relaxations of composite functions; this is the case both in McCormick's original rule for handling bivariate products [38,40], and in an improved product rule by Tsoukalas and Mitsos [63]. Thirdly, [54, Example 2.4.1] shows that if McCormick relaxations are weaker than

constant bounds evaluated using interval arithmetic [41,44], then this bounding information may be incorporated via nonsmooth "max" and "min" functions.

A relaxation scheme exhibiting continuous differentiability would be desirable for a number of reasons. In general, minimization of nondifferentiable convex objective functions requires dedicated numerical methods for nondifferentiable problems such as bundle methods [30,31,26,59], which lack the strong convergence rate results of their smooth counterparts. On the other hand, continuously differentiable convex relaxations may be minimized using standard gradient-based algorithms [46,45,12] for local optimization, which generally exhibit at least Q-linear convergence. Gradients of composite differentiable relaxations may be evaluated readily using standard automatic differentiation techniques [22,43], circumventing the need for dedicated theory and methods for subgradient propagation [40, 6,63,48]. A method constructing closed-form differentiable relaxations could be deployed in methods for relaxing implicit functions [68,60,55] and solutions of parametric ordinary differential equations [57,56], to make these relaxations differentiable and to remove theoretical obstacles concerning subgradient evaluation.

Thus, the goal of this article is to present a variant of McCormick's relaxation scheme that produces continuously differentiable relaxations — even if the original function is nonsmooth — while retaining the various theoretical and computational benefits of McCormick's original method. To achieve this, conditions are established under which the general multivariate McCormick relaxations of Tsoukalas and Mitsos [63] are differentiable, circumventing the obstacles listed above. Relaxation rules satisfying both these conditions and the rapid convergence properties of [10,42] are then presented for several intrinsic functions, including univariate functions such as "exp" and "log", the absolute-value function, bivariate sums and products, and the bivariate "max" and "min" functions. Using these rules, closed-form continuously differentiable relaxations may be constructed and evaluated for finite compositions of the considered intrinsic functions; moreover, these relaxations are shown to be twice-continuously differentiable in many cases. Further rules are provided for evaluating gradients of the differentiable relaxations analytically, for use in automatic differentiation methods. Computation of these relaxations and gradients was implemented in C++ by modifying the header library `MC++` [13]; this implementation is applied to several examples for illustration. Extensions of the developed relaxations to implicit functions and solutions of differential equations are also considered.

Several established approaches also construct useful differentiable convex and concave relaxations of functions. Interval arithmetic [41,44] provides constant bounds that are inexpensive to evaluate, but are generally weaker than the McCormick relaxations, and lack the rapid convergence property [10] required to avoid clustering [67,16] in methods for global optimization. Nevertheless, the classical McCormick relaxations [38,40] and the relaxations in this article employ interval arithmetic to glean information about the global behavior of the considered function.

The αBB relaxation scheme [3] also produces closed-form relaxations, and shares several of the benefits of McCormick's method outlined above. The αBB relaxations of twice-continuously differentiable functions are themselves twice-continuously differentiable, and also converge rapidly to the relaxed function as the considered domain is reduced in size. The αBB scheme is particularly well-suited to sums of simple terms; evaluation of αBB relaxations of more complicated compositions typically requires bounding the largest eigenvalue of the considered function's Hessian on the considered domain. Moreover, αBB relaxations cannot generally be applied to nonsmooth functions; in cases where this is possible, the relaxations themselves will also be nonsmooth. The relaxations developed in this article are compared to the αBB relaxations in an example in Section 7.3.

The Auxiliary Variable Method (AVM) of Tawarmalani and Sahinidis [62,61] is used in the state-of-the-art nonconvex solver BARON [51,50], and employs McCormick relaxations in a different manner. Rather than constructing closed-form relaxations for *factorable functions* that are compositions of simple intrinsic functions, the AVM instead relaxes nonconvex nonlinear programs (NLPs) directly, replacing them with convex programs that provide lower-bounding information. Roughly, provided that the objective function and the constraints in an NLP are factorable, the AVM constructs a relaxed optimization problem with auxiliary variables and constraints that correspond to relaxations of each intrinsic function in the original NLP's constraints and objective function. The resulting convex program has differentiable constraints, and is at least as tight a relaxation of the original NLP as a convex program formed by replacing each objective function and each constraint by a corresponding closed-form McCormick relaxation [63,62]. The nonconvex solvers ANTIGONE [39], COUENNE [7], SCIP [1,2], and LINDO Global [34] also employ approaches for constructing differentiable relaxations of NLPs that are similar in some respects to the AVM. When applied to a nonconvex optimization problem, the closed-form differentiable McCormick approach of this article may have an advantage when the AVM involves a large number of auxiliary variables, yielding large subproblems at each node, and when formulation of a nonsmooth nonconvex problem for solution by BARON requires appending many additional variables and constraints. As shown in Section 7, the differentiable McCormick relaxations developed in this article were embedded in simple branch-and-bound solvers, which were found in turn to perform comparably to BARON in a case study problem in nonconvex optimization. Moreover, since the AVM does not produce closed-form differentiable relaxations for factorable functions, it cannot be used directly to produce differentiable relaxations for implicit functions or solutions of parametric ordinary differential equations.

We note, briefly, that the relaxations presented in this article are compatible with the generalized McCormick framework described by Scott et al. [54,68,52,58]; in the language of [68], the relaxations developed in this article are valid *relaxation functions*. For simplicity, we do not pursue this connection further. We also note that twice-continuously differentiable relaxations may be computed reliably for the functions considered in this article, using a method in the first author's Ph.D. thesis [29] that employs smoothing constructions analagous to those in [8,19,17,47]. Again, this approach is not pursued further; the relaxations presented in the current article are generally tighter and are simpler to implement.

This article is structured as follows. Section 2 summarizes key established results concerning McCormick's classical relaxations [38,40], a generalized relaxation scheme by Tsoukalas and Mitsos [63], and an appropriate notion of differentiability on closed sets [69]. Section 3 presents methods for propagating differentiable relaxations through compositions involving known univariate intrinsic functions, including nonsmooth intrinsic functions such as the absolute-value function. Section 4 discusses propagating differentiable relaxations through compositions of multivariate intrinsic functions, and shows that any such rules cannot be straightforward generalizations of their univariate counterparts. Rules are presented for handling addition, multiplication, and the nonsmooth bivariate "max" and "min" functions. Section 5 shows that the generated differentiable relaxations converge rapidly to the relaxed function as the underlying domain is reduced in size. Section 6 demonstrates the utility of differentiable McCormick relaxations in constructing differentiable relaxations for implicit functions and for solutions of parametric ordinary differential equations. Section 7 describes a C++ implementation of the described differentiable relaxation scheme, developed by modifying the C++ header library `MC++` [13]. Examples of relaxations of simple functions are presented for illustration, and compared against the αBB relaxations [3]. A case

study in nonconvex optimization is presented, in which the performance of BARON [51,50] is compared against the approach of this article.

2 Mathematical background

This section presents the mathematical background underlying the results and methods in this article. Section 2.1 describes McCormick's relaxations [38,40] and the multivariate framework of Tsoukalas and Mitsos [63], and Section 2.2 describes notions of differentiability [69] that apply to functions defined on compact boxes.

The following notational conventions are used in this article. Any vector space $\mathbb{R}^n$ is equipped with the standard Euclidean norm $\|\cdot\|$ and inner product $\langle\cdot,\cdot\rangle$. The i^{th} unit coordinate vector in $\mathbb{R}^n$ is denoted as $e_{(i)}$; the dimension n of $e_{(i)}$ will be clear from the context. The i^{th} component of a vector $d\in\mathbb{R}^n$ is denoted as $d_i := \langle d, e_{(i)}\rangle$. Given vectors $x,y\in\mathbb{R}^n$, inequalities such as $x\leq y$ or $x<y$ are to be interpreted componentwise. The interior, closure, and convex hull of a set $S\subset\mathbb{R}^n$ are denoted as $\text{int}(S)$, $\text{cl}(S)$, and $\text{conv}(S)$, respectively; the boundary of S is then $\text{bd}\,(S) := \text{cl}(S)\setminus\text{int}(S)$.

An *interval* is a nonempty compact connected subset of $\mathbb{R}$. The set of all intervals is denoted as $\mathbb{IR}$. Intervals are denoted with boldface lowercase letters (e.g. $\boldsymbol{x}$), with associated bounds denoted as $\underline{x} := \inf\boldsymbol{x}$ and $\overline{x} := \sup\boldsymbol{x}$. Observe that $\underline{x}\leq\overline{x}$ and that $\boldsymbol{x} = [\underline{x},\overline{x}]$. Further details of interval analysis are presented in [44,41,4]. While "$[a,b]$" refers to the interval $\{\xi : a\leq\xi\leq b\}$, "(a,b)" instead refers exclusively to a vector formed by concatenating a and b, and never to an open set $\{\xi : a<\xi<b\}$.

2.1 McCormick's relaxations

McCormick's relaxation scheme [38,40,63] computes convex underestimators and concave overestimators for *factorable functions*, which are well-defined finite compositions of simple known *intrinsic functions*. The discussions of these underestimators and overestimators in this article assume knowledge of established properties of convex sets and convex functions, as presented in [48,25,11].

Definition 1 Consider a set $X\subset\mathbb{R}^n$, a function $h : X\to\mathbb{R}$, and a convex subset $C\subset X$. A function $\underline{h}^{\text{C}} : C\to\mathbb{R}$ is a *convex relaxation* of h on C if $\underline{h}^{\text{C}}$ is convex and $\underline{h}^{\text{C}}(x)\leq h(x)$ for each $x\in C$. A function $\overline{h}^{\text{C}} : C\to\mathbb{R}$ is a *concave relaxation* of h on C if $\overline{h}^{\text{C}}$ is concave and $\overline{h}^{\text{C}}(x)\geq h(x)$ for each $x\in C$.

The *convex envelope* of h on C is the unique convex relaxation of h on C that dominates all other convex relaxations of h on C. The *concave envelope* of h on C is the unique concave relaxation of h on C that is dominated by all other concave relaxations of h on C .

Tsoukalas and Mitsos [63] provide a general scheme for generating convex and concave relaxations of finite compositions of known *intrinsic functions*, by recursive application of the following rule, which generalizes earlier rules by McCormick [38,40]. The continuously differentiable relaxations developed in this article are constructed using this rule.

Theorem 1 (Theorem 2 in [63]) *Consider nonempty compact convex sets $Z\subset\mathbb{R}^n$ and $X_i\subset\mathbb{R}$ for each $i\in\{1,\ldots,m\}$, and define $X := X_1\times\ldots\times X_m$. Consider functions $\phi : X\to\mathbb{R}$ and $f_i : Z\to X_i$ for each $i\in\{1,\ldots,m\}$, and suppose that the following relaxations exist:*

- *a continuous convex relaxation* $\underline{f}_i^{\mathrm{C}} : Z \to X_i$ *of* f_i *on* Z *for each* $i \in \{1,\ldots,m\}$,
- *a continuous concave relaxation* $\overline{f}_i^{\mathrm{C}} : Z \to X_i$ *of* f_i *on* Z *for each* $i \in \{1,\ldots,m\}$,
- *a continuous convex relaxation* $\underline{\phi}^{\mathrm{C}}$ *of* ϕ *on* X, *and*
- *a continuous concave relaxation* $\overline{\phi}^{\mathrm{C}}$ *of* ϕ *on* X.

Then, the following mapping is a continuous convex relaxation of the composite function $g : Z \to \mathbb{R} : z \mapsto \phi(f_1(z),\ldots,f_m(z))$ *on* Z*:*

$$\underline{g}^{\mathrm{C}} : Z \to \mathbb{R} : z \mapsto \min\left\{ \underline{\phi}^{\mathrm{C}}(\xi) : \xi \in \mathbb{R}^m, \quad \underline{f}_i^{\mathrm{C}}(z) \le \xi_i \le \overline{f}_i^{\mathrm{C}}(z),\ \forall i \in \{1,\ldots,m\} \right\}, \quad (1)$$

and the following mapping is a continuous concave relaxation of g *on* Z*:*

$$\overline{g}^{\mathrm{C}} : Z \to \mathbb{R} : z \mapsto \max\left\{ \overline{\phi}^{\mathrm{C}}(\xi) : \xi \in \mathbb{R}^m, \quad \underline{f}_i^{\mathrm{C}}(z) \le \xi_i \le \overline{f}_i^{\mathrm{C}}(z),\ \forall i \in \{1,\ldots,m\} \right\}. \quad (2)$$

Weierstrass's Theorem guarantees that the optimization problems defining $\underline{g}^{\mathrm{C}}$ and $\overline{g}^{\mathrm{C}}$ have solutions for each $z \in Z$. Observe that the optimization problem (1) is a convex program, and the problem (2) is an analogous concave maximization problem on a convex set. Thus, since both problems are bound-constrained for each $z \in Z$, the well-known Karush-Kuhn-Tucker (KKT) conditions (discussed in [5], for example) are necessary and sufficient for any solution of (1) or (2).

Theorem 1 is intended to be applied with ϕ chosen from a library of known *intrinsic functions*, for which the optimization problems (1) and (2) are readily solved. These problems are solved trivially, for example, if ϕ is affine. In addition, Tsoukalas and Mitsos provide closed-form expressions [63, Sections 5 and 6] for the relaxations $\underline{g}^{\mathrm{C}}$ and $\overline{g}^{\mathrm{C}}$ when ϕ is either the bilinear product function $(x,y) \in \mathbb{R}^2 \mapsto xy$ or the bivariate "max" or "min" functions, with $\underline{\phi}^{\mathrm{C}}$ and $\overline{\phi}^{\mathrm{C}}$ chosen to be the convex and concave envelopes of ϕ in each case. These relaxations are generally tighter, respectively, than McCormick's original treatment of the product [38,40] and the treatment of "max" and "min" based on the identities:

$$\max\{x,y\} \equiv \tfrac{1}{2}(x+y+|x-y|) \qquad \text{and} \qquad \min\{x,y\} \equiv \tfrac{1}{2}(x+y-|x-y|).$$

Observe that tighter relaxations $\underline{g}^{\mathrm{C}}$ and $\overline{g}^{\mathrm{C}}$ may be obtained if the supplied sets X_i are chosen to be smaller while still satisfying the demands of the theorem. In practice, appropriate sets X_i are computed using natural interval extensions [41,44].

When $m = 1$ in Theorem 1, $\phi : X_1 \to \mathbb{R}$ is a univariate function. In this case, the relaxations provided by the theorem reduce to McCormick's classical rule [38] for handling univariate intrinsic functions. As the following result from [63] shows, the optimization problems (1) and (2) may then be solved analytically without any additional assumptions on the provided functions and relaxations. Here, the function $\mathrm{mid} : \mathbb{R}^3 \to \mathbb{R}$ returns the median of its three scalar arguments.

Corollary 1 (Proposition 1 in [63]) *Suppose that the conditions of Theorem 1 hold with* $m = 1$, *and set* $f := f_1$. *In this case,* X *is an interval* $\boldsymbol{x} \in \mathbb{IR}$. *Let* $\xi^*_{\min}$ *be a minimum of* $\underline{\phi}^{\mathrm{C}}$ *on* $\boldsymbol{x}$, *and let* $\xi^*_{\max}$ *be a maximum of* $\overline{\phi}^{\mathrm{C}}$ *on* $\boldsymbol{x}$. *Then, the relaxations* $\underline{g}^{\mathrm{C}}$ *and* $\overline{g}^{\mathrm{C}}$ *in Theorem 1 are described equivalently as follows:*

$$\underline{g}^{\mathrm{C}} : Z \to \mathbb{R} : z \mapsto \underline{\phi}^{\mathrm{C}}(\mathrm{mid}(\underline{f}^{\mathrm{C}}(z), \overline{f}^{\mathrm{C}}(z), \xi^*_{\min})),$$

and

$$\overline{g}^{\mathrm{C}} : Z \to \mathbb{R} : z \mapsto \overline{\phi}^{\mathrm{C}}(\mathrm{mid}(\underline{f}^{\mathrm{C}}(z), \overline{f}^{\mathrm{C}}(z), \xi^*_{\max})).$$

In practice, using the above corollary typically requires closed-form expressions for both $\xi^*_{\min}$ and $\xi^*_{\max}$, which can be obtained for many choices of ϕ. These are straightforward to obtain when $\underline{g}^{\mathrm{C}}$ and $\overline{g}^{\mathrm{C}}$ are monotone; if $\underline{g}^{\mathrm{C}}$ is increasing, for example, then $\xi^*_{\min}$ may be set to $\underline{x}$.

2.2 Differentiability on open and closed sets

Since McCormick relaxations are constructed on compact domains, describing differentiability of these relaxations requires some care at the boundaries of these domains. This article considers differentiability in the sense of Whitney [69]. As described in this section, Whitney differentiability coincides with classical (Fréchet) differentiability on the interior of a function's domain, satisfies the classical chain rule even at the domain's boundary, and provides meaningful subgradient information for convex functions and concave functions.

Given an open set $X \subset \mathbb{R}^n$, a function $f : X \to \mathbb{R}^m$ is *(Fréchet) differentiable* at $x \in X$ if there exists a matrix $A \in \mathbb{R}^{m\times n}$ for which

$$0 = \lim_{h\to 0} \frac{f(x+h) - (f(x) + Ah)}{\|h\|}.$$

In this case, the above equation defines A uniquely, and A is the unique *derivative* $Df(x)$ of f at x. If x is partitioned, for example, into $x \equiv (\xi, \zeta)$, then partial derivatives $\frac{\partial f}{\partial \xi}(x)$ and $\frac{\partial f}{\partial \zeta}(x)$ are defined as appropriate submatrices of $Df(x)$. If $m = 1$, in which case f is scalar-valued, then the *gradient* of f at x is the column vector $\nabla f(x) := (Df(x))^{\mathrm{T}} \in \mathbb{R}^m$.

Given an open set $X \subset \mathbb{R}^n$, a function $f : X \to \mathbb{R}^n$ is *continuously differentiable* ($\mathscr{C}^1$) at $x \in X$ if it is differentiable on some neighborhood $N \subset X$ of x and the derivative mapping $y \mapsto Df(y)$ is continuous at x. If $m = 1$, in which case f is scalar-valued, then f is *twice-continuously differentiable* ($\mathscr{C}^2$) on X if f is $\mathscr{C}^1$ on X and there exists a continuous *Hessian* mapping $x \mapsto \nabla^2 f(x) \in \mathbb{R}^{n\times n}$ for which

$$0 = \lim_{h\to 0} \frac{f(x+h) - (f(x) + \langle \nabla f(x), h\rangle + \frac{1}{2}\langle h, \nabla^2 f(x)\, h\rangle)}{\|h\|^2}, \qquad \forall x \in X.$$

A vector-valued function f is $\mathscr{C}^2$ if each of its component functions is $\mathscr{C}^2$. Higher orders of continuous differentiability are defined analogously.

By specializing a classical result by Whitney [69], differentiability on closed sets such as intervals can be defined in a manner that is consistent with the classical chain rule of differentiation, as follows.

Definition 2 (adapted from [69]) Given a closed set $B \subset \mathbb{R}^n$ and some $\nu \in \mathbb{N}$, a function $f : B \to \mathbb{R}^m$ is *Whitney-$\mathscr{C}^\nu$* on B if there exist an open set $X \subset \mathbb{R}^n$ and a *Whitney extension* $\hat{f} : X \to \mathbb{R}^m$ such that $B \subset X$, $\hat{f} \equiv f$ on B, and $\hat{f}$ is $\mathscr{C}^\nu$ (in the classical sense) on X. Given any point x in the boundary of B, define a derivative $Df(x) := D\hat{f}(x)$. If $m = 1$, in which case f is scalar-valued, then define a gradient $\nabla f(x) := Df(x)^{\mathrm{T}}$.

Remark 1 When x lies in the boundary of B, it is possible that $Df(x)$ is not uniquely specified by the above definition, since $\hat{f}$ might not be specified uniquely. For example, if B comprises a single point $\{x_0\} \subset \mathbb{R}^n$, then $Df(x_0)$ may be chosen to be any element of $\mathbb{R}^{m\times n}$, since $\hat{f}$ may be chosen to be any $\mathscr{C}^i$ function for which $\hat{f}(x_0) = f(x_0)$.

The Whitney Extension Theorem [69, Theorem I] provides an equivalent characterization of Whitney-$\mathscr{C}^\nu$ functions. Despite the possible nonuniqueness implied by the previous remark, the following propositions show that the classical chain rule continues to hold for derivatives of compositions of Whitney-$\mathscr{C}^1$ functions. Both propositions are immediate corollaries of the Whitney Extension Theorem.

Proposition 1 *Consider a closed set $B \subset \mathbb{R}^n$ and a Whitney-$\mathscr{C}^1$ function $f : B \to \mathbb{R}^m$. If there exists any sequence $\{x_{(k)}\}_{k\in\mathbb{N}} \to x$ in $B\backslash\{x\}$, then any derivative $Df(x)$ satisfies*

$$0 = \lim_{\substack{h\to 0\\ (x+h)\in B}} \frac{f(x+h) - (f(x) + Df(x)\,h)}{\|h\|}.$$

Proposition 2 *Consider nonempty sets $B \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ such that B is either closed, open, or both, and such that Y is either closed, open, or both. For any fixed $\nu \in \mathbb{N}$, given functions $g : B \to Y$ and $f : Y \to \mathbb{R}^p$ that are $\mathscr{C}^\nu$ in either the classical sense or the Whitney sense, consider the composite function $h \equiv f \circ g : B \to \mathbb{R}^p$. If B is open, then h is $\mathscr{C}^\nu$ on B in the classical sense; if B is closed, then h is Whitney-$\mathscr{C}^\nu$ on B.*

Moreover, for each $x \in B$, $Dh(x) = Df(g(x))\,Dg(x)$. (If B is closed and x lies in the boundary of B, then this construction of $Dh(x)$ satisfies both Definition 2 and Proposition 1 for some Whitney extension $\hat{h}$ of h.)

Corollary 2 *Given a closed convex set $B \subset \mathbb{R}^n$ and a convex Whitney-$\mathscr{C}^1$ function $f : B \to \mathbb{R}$, for each $x \in B$, $\nabla f(x)$ is a* subgradient *of f at x in that*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle, \qquad \forall y \in B.$$

The following proposition shows that local Whitney extensions imply the existence of a single global Whitney extension.

Proposition 3 *Consider a compact convex set $C \subset \mathbb{R}^n$ with nonempty interior, and a function $f : C \to \mathbb{R}^m$. Suppose that, for each $x \in C$, there exists a neighborhood $N_x \subset \mathbb{R}^n$ of x and a $\mathscr{C}^1$ function $\phi_x : N_x \to \mathbb{R}^m$ for which $\phi_x \equiv f$ on $N_x \cap C$. Then f is Whitney-$\mathscr{C}^1$ on C.*

Proof See Appendix A.1. □

3 Differentiable relaxations of univariate intrinsic functions

This section shows that, in Corollary 1, the generated relaxations of $g \equiv \phi \circ f$ are Whitney-$\mathscr{C}^1$ under minimal assumptions on the supplied relaxations of f and the univariate intrinsic function ϕ. If ϕ is Whitney-$\mathscr{C}^1$ on its domain, then its convex and concave envelopes are shown to satisfy these assumptions; if ϕ is nonsmooth, then more care must be taken. Closed-form expressions are provided for the gradients of the generated relaxations. Under certain additional assumptions, the generated relaxations of g are shown to be Whitney-$\mathscr{C}^2$; these additional assumptions are readily satisfied for many typical choices of the univariate intrinsic function ϕ.

3.1 Establishing differentiability

This section shows that the relaxations described by Corollary 1 for univariate composition are in fact Whitney-$\mathscr{C}^1$ when the following assumption is satisfied. When a stronger version of this assumption is satisfied, the relaxations described by Corollary 1 are Whitney-$\mathscr{C}^2$.

Assumption 1 *Consider the conditions of Corollary 1, and suppose that the supplied relaxations $\underline{f}^{\mathrm{C}}$, $\overline{f}^{\mathrm{C}}$, $\underline{\phi}^{\mathrm{C}}$, and $\overline{\phi}^{\mathrm{C}}$ are each Whitney-$\mathscr{C}^1$ on their respective domains.*

Observe that Assumption 1 does not demand any differentiability properties of f or ϕ. Nevertheless, the following proposition shows that the convex and concave envelopes of any unvariate Whitney-$\mathscr{C}^1$ function are themselves Whitney-$\mathscr{C}^1$. In this case, convex and concave envelopes satisfy the above assumption's requirements concerning $\underline{\phi}^{\mathrm{C}}$ and $\overline{\phi}^{\mathrm{C}}$.

Proposition 4 *Consider an interval $\boldsymbol{x} \in \mathbb{IR}$, and a Whitney-$\mathscr{C}^1$ function $f : \boldsymbol{x} \to \mathbb{R}$. The convex envelope $\underline{f}^{\mathrm{C}} : \boldsymbol{x} \to \mathbb{R}$ of f on $\boldsymbol{x}$ is also Whitney-$\mathscr{C}^1$ on $\boldsymbol{x}$, as is the concave envelope $\overline{f}^{\mathrm{C}} : \boldsymbol{x} \to \mathbb{R}$ of f on $\boldsymbol{x}$.*

Proof See Appendix A.2. □

Theorem 2 *Suppose that Assumption 1 holds. The relaxations $\underline{g}^{\mathrm{C}}$ and $\overline{g}^{\mathrm{C}}$ described by Corollary 1 are Whitney-$\mathscr{C}^1$ on Z, with gradients:*

$$\nabla \underline{g}^{\mathrm{C}}(z) = \max\left\{0, \nabla \underline{\phi}^{\mathrm{C}}\left(\underline{f}^{\mathrm{C}}(z)\right)\right\} \nabla \underline{f}^{\mathrm{C}}(z) + \min\left\{0, \nabla \underline{\phi}^{\mathrm{C}}\left(\overline{f}^{\mathrm{C}}(z)\right)\right\} \nabla \overline{f}^{\mathrm{C}}(z),$$

and

$$\nabla \overline{g}^{\mathrm{C}}(z) = \min\left\{0, \nabla \overline{\phi}^{\mathrm{C}}\left(\underline{f}^{\mathrm{C}}(z)\right)\right\} \nabla \underline{f}^{\mathrm{C}}(z) + \max\left\{0, \nabla \overline{\phi}^{\mathrm{C}}\left(\overline{f}^{\mathrm{C}}(z)\right)\right\} \nabla \overline{f}^{\mathrm{C}}(z).$$

Proof The required results concerning $\underline{g}^{\mathrm{C}}$ will be demonstrated; a similar argument yields the results concerning $\overline{g}^{\mathrm{C}}$. Since $\underline{\phi}^{\mathrm{C}}$ is Whitney-$\mathscr{C}^1$, and since any $\mathscr{C}^1$ function is locally Lipschitz continuous, $\underline{\phi}^{\mathrm{C}}$ is Lipschitz continuous on $\boldsymbol{x}$. Thus, proceeding as in the proof of Proposition 4, the following function is a $\mathscr{C}^1$ Whitney extension of $\underline{\phi}^{\mathrm{C}}$ on $\boldsymbol{x}$:

$$\psi : \mathbb{R} \to \mathbb{R} : \xi \mapsto \begin{cases} \underline{\phi}^{\mathrm{C}}(\underline{x}) + (D\underline{\phi}^{\mathrm{C}}(\underline{x}))(\xi - \underline{x}), & \text{if } \xi < \underline{x}, \\ \underline{\phi}^{\mathrm{C}}(\xi), & \text{if } \xi \in \boldsymbol{x}, \\ \underline{\phi}^{\mathrm{C}}(\overline{x}) + (D\underline{\phi}^{\mathrm{C}}(\overline{x}))(\xi - \overline{x}), & \text{if } \overline{x} < \xi. \end{cases}$$

Since $\underline{\phi}^{\mathrm{C}}$ is convex, $\nabla \underline{\phi}^{\mathrm{C}}$ is increasing on $\boldsymbol{x}$. Thus, $\nabla\psi$ is increasing on $\mathbb{R}$, and so ψ is convex. Define a set $H := \{(\ell, u) \in \mathbb{R}^2 : \ell \le u\}$, and a function

$$\gamma : H \to \mathbb{R} : (\ell, u) \mapsto \min\{\psi(\xi) : \ell \le \xi \le u\}.$$

Observing that $\underline{g}^{\mathrm{C}}(z) = \gamma(\underline{f}^{\mathrm{C}}(z), \overline{f}^{\mathrm{C}}(z))$ for each $z \in Z$, it suffices to show that γ is Whitney-$\mathscr{C}^1$ on the set $K := \{(\ell, u) \in \mathbb{R}^2 : \ell \in \boldsymbol{x},\ u \in \boldsymbol{x},\ \ell \le u\}$. Now, if $\nabla\psi(\underline{x}) > 0$, then ψ is increasing on $\mathbb{R}$; thus, $\gamma(\ell, u) = \ell$ for each $(\ell, u) \in K$. This shows that γ is Whitney-$\mathscr{C}^1$ on K, as required. A similar argument shows that γ is Whitney-$\mathscr{C}^1$ on K if $\nabla\psi(\overline{x}) < 0$.

It only remains to consider the case in which $\nabla\psi(\underline{x}) \le 0 \le \nabla\psi(\overline{x})$. By the intermediate-value theorem, there exists $\eta^* \in \boldsymbol{x}$ with $\nabla\psi(\eta^*) = 0$. Since ψ is convex, η^* is a minimum of ψ on $\mathbb{R}$; since $\eta^* \in \boldsymbol{x}$, this implies that $\psi(\eta^*) = \psi(\xi^*_{\min})$. Thus, $\xi^*_{\min}$ is a minimum of ψ on $\mathbb{R}$, in which case $\nabla\psi(\xi^*_{\min}) = 0$. Consider the functions:

$$\psi_{\mathrm{I}} : \mathbb{R} \to \mathbb{R} : \eta \mapsto \psi(\max\{\xi^*_{\min}, \eta\}), \quad \text{and} \quad \psi_{\mathrm{D}} : \mathbb{R} \to \mathbb{R} : \eta \mapsto \psi(\min\{\xi^*_{\min}, \eta\}).$$

Since $\nabla\psi(\xi^*_{\min}) = 0$, both ψ_{I} and ψ_{D} are $\mathscr{C}^1$, as is the function:

$$\tilde{\gamma} : \mathbb{R}^2 \to \mathbb{R} : (a,b) \mapsto \psi_{\mathrm{I}}(a) + \psi_{\mathrm{D}}(b) - \psi(\xi^*_{\min}).$$

Moreover, $\gamma(\ell,u) = \tilde{\gamma}(\ell,u)$ for each $(\ell,u) \in K$. Thus, γ is Whitney-$\mathscr{C}^1$ on K, as required. The gradient of $\underline{g}^{\mathrm{C}}$ may then be evaluated using the chain rule for Whitney-$\mathscr{C}^1$ functions. □

Observe that the gradients of $\underline{g}^{\mathrm{C}}$ and $\overline{g}^{\mathrm{C}}$ provided by Theorem 2 do not necessarily coincide with the corresponding subgradients provided by [40, Theorem 3.2]. As the following two corollaries show, the differentiability results of Theorem 2 can be improved when Assumption 1 is strengthened.

Corollary 3 *Suppose that the conditions of Corollary 1 hold, and suppose that each of the following conditions is satisfied:*

- $\underline{f}^{\mathrm{C}}$, $\overline{f}^{\mathrm{C}}$, $\underline{\phi}^{\mathrm{C}}$, *and* $\overline{\phi}^{\mathrm{C}}$ *are each Whitney-*$\mathscr{C}^2$ *on their respective domains,*
- *if* $\nabla\underline{\phi}^{\mathrm{C}}(\xi^*_{\min}) = 0$, *then* $\nabla^2\underline{\phi}^{\mathrm{C}}(\xi^*_{\min}) = 0$, *and*
- *if* $\nabla\overline{\phi}^{\mathrm{C}}(\xi^*_{\max}) = 0$, *then* $\nabla^2\overline{\phi}^{\mathrm{C}}(\xi^*_{\max}) = 0$.

Then, the relaxations $\underline{g}^{\mathrm{C}}$ *and* $\overline{g}^{\mathrm{C}}$ *described by Corollary 1 are Whitney-*$\mathscr{C}^2$ *on* Z.

Proof The required result concerning $\underline{g}^{\mathrm{C}}$ will be demonstrated; the result concerning $\overline{g}^{\mathrm{C}}$ is analogous. The proof of Theorem 2 is still applicable under the assumptions of this corollary; to extend that proof to demonstrate this corollary, it suffices to show that the function γ is Whitney-$\mathscr{C}^2$ on K. If either $\nabla\psi(\underline{x}) > 0$ or $\nabla\psi(\overline{x}) < 0$, then γ is linear and is therefore Whitney-$\mathscr{C}^2$. If $\nabla\psi(\underline{x}) \leq 0 \leq \nabla\psi(\overline{x})$, then observe that, under the assumptions of this corollary, the functions ψ_{I} and ψ_{D} are $\mathscr{C}^2$ on $\mathbb{R}$. Thus, the function $\tilde{\gamma}$ is $\mathscr{C}^2$ on $\mathbb{R}^2$, which implies that γ is Whitney-$\mathscr{C}^2$ on K. □

Corollary 4 *Suppose that the conditions of Corollary 1 hold, and suppose that, for some* $\nu \in \mathbb{N}$, *each of the following conditions is satisfied:*

- $\underline{f}^{\mathrm{C}}$, $\overline{f}^{\mathrm{C}}$, $\underline{\phi}^{\mathrm{C}}$, *and* $\overline{\phi}^{\mathrm{C}}$ *are Whitney-*$\mathscr{C}^\nu$ *on their respective domains,*
- $\underline{\phi}^{\mathrm{C}}$ *is either increasing on* $\boldsymbol{x}$ *or decreasing on* $\boldsymbol{x}$, *and*
- $\overline{\phi}^{\mathrm{C}}$ *is either increasing on* $\boldsymbol{x}$ *or decreasing on* $\boldsymbol{x}$.

Then, the relaxations $\underline{g}^{\mathrm{C}}$ *and* $\overline{g}^{\mathrm{C}}$ *described by Corollary 1 are Whitney-*$\mathscr{C}^\nu$ *on* Z.

Proof This corollary is a special case of Theorem 3 below. □

When the requirements of Corollary 4 are met, then the expressions for $\nabla\underline{g}^{\mathrm{C}}$ and $\nabla\overline{g}^{\mathrm{C}}$ provided by Theorem 2 become simpler. For example, if $\underline{\phi}^{\mathrm{C}}$ is increasing, then Theorem 2 implies that, for each $z \in Z$,

$$\nabla\underline{g}^{\mathrm{C}}(z) = \nabla\underline{\phi}^{\mathrm{C}}(\underline{f}^{\mathrm{C}}(z))\,\nabla\underline{f}^{\mathrm{C}}(z).$$

Table 1 presents relaxations $\underline{\phi}^{\mathrm{C}}$ and $\overline{\phi}^{\mathrm{C}}$ for various choices of $\phi : \boldsymbol{x} \to \mathbb{R}$; these relaxations are readily verified to satisfy the demands of either Corollary 3 or Corollary 4 with $\nu := 2$. The remainder of this section presents similarly appropriate relaxations for the following choices of ϕ:

- the squaring function $\xi \mapsto \xi^2$,

Table 1 Relaxations $\underline{\phi}^{\mathrm{C}}, \overline{\phi}^{\mathrm{C}}$ of various functions $\phi : \boldsymbol{x} \to \mathbb{R}$ on various intervals $\boldsymbol{x} \in \mathbb{IR}$. These relaxations satisfy the demands of either Corollary 3 or Corollary 4 with $\nu = 2$. Here, $\mathbb{R}_+ := \{\xi \in \mathbb{R} : 0 < \xi\}$ and $\mathbb{R}_- := \{\xi \in \mathbb{R} : \xi < 0\}$.

B	$\phi(\xi)$ for $\xi \in \boldsymbol{x} \subset B$	$\underline{\phi}^{\mathrm{C}}(\xi)$	$\overline{\phi}^{\mathrm{C}}(\xi)$
$\mathbb{R}$	$a\xi + b$ for $a, b \in \mathbb{R}$	$a\xi + b$	$a\xi + b$
$\mathbb{R}$	$\exp\xi$	$\exp\xi$	$\exp\underline{x} + (\exp\overline{x} - \exp\underline{x})\left(\frac{\xi - \underline{x}}{\overline{x} - \underline{x}}\right)$
$\mathbb{R}_+$	$\ln\xi$	$\ln\underline{x} + (\ln\overline{x} - \ln\underline{x})\left(\frac{\xi - \underline{x}}{\overline{x} - \underline{x}}\right)$	$\ln\xi$
$\mathbb{R}$	ξ^{2k+2} for $k \in \mathbb{N}$	ξ^{2k+2}	$\underline{x}^{2k+2} + (\overline{x}^{2k+2} - \underline{x}^{2k+2})\left(\frac{\xi - \underline{x}}{\overline{x} - \underline{x}}\right)$
$\mathbb{R}_+$	$\sqrt{\xi}$	$\sqrt{\underline{x}} + (\sqrt{\overline{x}} - \sqrt{\underline{x}})\left(\frac{\xi - \underline{x}}{\overline{x} - \underline{x}}\right)$	$\sqrt{\xi}$
$\mathbb{R}_+$	$\frac{1}{\xi^k}$ for $k \in \mathbb{N}$	$\frac{1}{\xi^k}$	$\frac{1}{\underline{x}^k} + (\frac{1}{\overline{x}^k} - \frac{1}{\underline{x}^k})\left(\frac{\xi - \underline{x}}{\overline{x} - \underline{x}}\right)$
$\mathbb{R}_-$	$\frac{1}{\xi^{2k}}$ for $k \in \mathbb{N}$	$\frac{1}{\xi^{2k}}$	$\frac{1}{\underline{x}^{2k}} + (\frac{1}{\overline{x}^{2k}} - \frac{1}{\underline{x}^{2k}})\left(\frac{\xi - \underline{x}}{\overline{x} - \underline{x}}\right)$
$\mathbb{R}_+$	$\frac{1}{\xi^{2k-1}}$ for $k \in \mathbb{N}$	$\frac{1}{\underline{x}^{2k-1}} + (\frac{1}{\overline{x}^{2k-1}} - \frac{1}{\underline{x}^{2k-1}})\left(\frac{\xi - \underline{x}}{\overline{x} - \underline{x}}\right)$	$\frac{1}{\xi^{2k-1}}$

- an odd power $\xi \mapsto \xi^{2k+1}$ for $k \in \mathbb{N}$,
- the absolute-value function $\xi \mapsto |\xi|$,
- the mapping $\xi \mapsto \max\{\xi, a\}$ for some $a \in \mathbb{R}$, and
- the mapping $\xi \mapsto \min\{\xi, a\}$ for some $a \in \mathbb{R}$.

If these choices of ϕ, $\underline{\phi}^{\mathrm{C}}$, and $\overline{\phi}^{\mathrm{C}}$ are employed in Corollary 1, then Theorem 2 provides gradients of the obtained relaxations $\underline{g}^{\mathrm{C}}$ and $\overline{g}^{\mathrm{C}}$ of the composite function $g \equiv \phi \circ f$.

3.2 Relaxing squares and odd powers

This section considers the cases in which the function ϕ in Corollary 1 is the power function $\xi \mapsto \xi^a$ with $a \in \{2\} \cup \{2k+1 : k \in \mathbb{N}\}$. In each case, Proposition 4 implies that the demands of Assumption 1 on $\underline{\phi}^{\mathrm{C}}$ and $\overline{\phi}^{\mathrm{C}}$ are satisfied when these functions are chosen to be the convex and concave envelopes of ϕ, respectively. Thus, when $a = 2$, the demands of Theorem 2 on $\underline{\phi}^{\mathrm{C}}$ and $\overline{\phi}^{\mathrm{C}}$ are satisfied when $\underline{\phi}^{\mathrm{C}} := \phi$ and

$$\overline{\phi}^{\mathrm{C}} : \boldsymbol{x} \to \mathbb{R} : \xi \mapsto \underline{x}^2 + (\underline{x} + \overline{x})(\xi - \underline{x}). \tag{3}$$

When $a = 2k+1$ for $k \in \mathbb{N}$, the demands of Theorem 2 on $\underline{\phi}^{\mathrm{C}}$ and $\overline{\phi}^{\mathrm{C}}$ are satisfied when these functions are chosen to be the the odd-power envelopes described in [33].

However, in each of the cases above, the choices of $\underline{\phi}^{\mathrm{C}}$ and $\overline{\phi}^{\mathrm{C}}$ are not necessarily Whitney-$\mathscr{C}^2$, and may fail to satisfy the assumptions of Corollaries 3 or 4. The following propositions provide alternative relaxations of ϕ that are readily verified to satisfy the assumptions of one of these corollaries.

Proposition 5 *Suppose that the conditions of Corollary 1 hold, and that ϕ is the mapping $x \mapsto x^2$. The following choices of $\underline{\phi}^{\mathrm{C}}$ and $\overline{\phi}^{\mathrm{C}}$ satisfy the demands of Corollary 3:*

$$\underline{\phi}^{\mathrm{C}} : \boldsymbol{x} \to \mathbb{R} : \xi \mapsto \begin{cases} \xi^2, & \text{if } 0 \le \underline{x} \text{ or } \overline{x} \le 0, \\ \frac{\xi^3}{(\overline{x})}, & \text{if } \underline{x} < 0 < \overline{x} \text{ and } 0 \le \xi, \\ \frac{\xi^3}{(\underline{x})}, & \text{if } \underline{x} < 0 < \overline{x} \text{ and } \xi < 0, \end{cases}$$

and $\overline{\phi}^{\mathrm{C}}$ *defined as in* (3)*, with* $\xi^*_{\min} := \mathrm{mid}\,(\underline{x},\overline{x},0)$ *and* $\xi^*_{\max} \in \arg\max_{\zeta\in\{\underline{x},\overline{x}\}} \zeta^2$.

Proposition 6 *Suppose that the conditions of Corollary 1 hold, and that* ϕ *is the mapping* $x \mapsto x^{2k+1}$ *for some* $k \in \mathbb{N}$*. The following choices of* $\underline{\phi}^{\mathrm{C}}$ *and* $\overline{\phi}^{\mathrm{C}}$ *satisfy the demands of Corollary 4 when* $\nu \leq 2k+1$*:*

$$\underline{\phi}^{\mathrm{C}} : \boldsymbol{x} \to \mathbb{R} : \xi \mapsto \begin{cases} \underline{x}^{2k+1} + (\overline{x}^{2k+1} - \underline{x}^{2k+1})\left(\frac{\xi-\underline{x}}{\overline{x}-\underline{x}}\right), & \text{if } \overline{x} \leq 0, \\ \underline{x}^{2k+1}\left(\frac{\overline{x}-\xi}{\overline{x}-\underline{x}}\right) + (\max\{0,\xi\})^{2k+1}, & \text{if } \underline{x} < 0 < \overline{x}, \\ \xi^{2k+1}, & \text{if } 0 \leq \underline{x}, \end{cases}$$

$$\overline{\phi}^{\mathrm{C}} : \boldsymbol{x} \to \mathbb{R} : \xi \mapsto \begin{cases} \xi^{2k+1}, & \text{if } \overline{x} \leq 0, \\ \overline{x}^{2k+1}\left(\frac{\xi-\underline{x}}{\overline{x}-\underline{x}}\right) + (\max\{0,\xi\})^{2k+1}, & \text{if } \underline{x} < 0 < \overline{x}, \\ \underline{x}^{2k+1} + (\overline{x}^{2k+1} - \underline{x}^{2k+1})\left(\frac{\xi-\underline{x}}{\overline{x}-\underline{x}}\right), & \text{if } 0 \leq \underline{x}, \end{cases}$$

with $\xi^*_{\min} := \underline{x}$ *and* $\xi^*_{\max} := \overline{x}$*.*

3.3 Relaxing the absolute-value function and variants

This section considers the cases in which the function ϕ in Corollary 1 is either of the nonsmooth univariate intrinsic functions $\xi \mapsto |\xi|$, $\xi \mapsto \max\{\xi,a\}$, or $\xi \mapsto \min\{\xi,a\}$. In each case, relaxations $\underline{\phi}^{\mathrm{C}}$ and $\overline{\phi}^{\mathrm{C}}$ are provided that are readily verified to satisfy the requirements of Theorem 2 and Corollary 3. Since these choices of ϕ are nonsmooth, the assumptions of Proposition 4 are not necessarily met; the convex and concave envelopes of ϕ are thus less useful when obtaining Whitney-$\mathscr{C}^1$ relaxations.

Proposition 7 *Suppose that the conditions of Corollary 1 hold, and that* ϕ *is the absolute-value mapping* $x \mapsto |x|$*. Choose any* $\mu \geq 1$*. The following choices of* $\underline{\phi}^{\mathrm{C}}$ *and* $\overline{\phi}^{\mathrm{C}}$ *satisfy the demands of Assumption 1:*

$$\underline{\phi}^{\mathrm{C}} : \boldsymbol{x} \to \mathbb{R} : \xi \mapsto \begin{cases} |\xi|, & \text{if } 0 \leq \underline{x} \text{ or } \overline{x} \leq 0, \\ \overline{x}\left(\frac{\xi}{(\overline{x})}\right)^{\mu+1}, & \text{if } \underline{x} < 0 < \overline{x} \text{ and } 0 \leq \xi, \\ -\underline{x}\left(\frac{\xi}{(\underline{x})}\right)^{\mu+1}, & \text{if } \underline{x} < 0 < \overline{x} \text{ and } \xi < 0, \end{cases}$$

$$\overline{\phi}^{\mathrm{C}} : \boldsymbol{x} \to \mathbb{R} : \xi \mapsto |\underline{x}| + (|\overline{x}| - |\underline{x}|)\left(\frac{\xi-\underline{x}}{\overline{x}-\underline{x}}\right),$$

with $\xi^*_{\min} := \mathrm{mid}\,(\underline{x},\overline{x},0)$ *and* $\xi^*_{\max} \in \arg\max_{\zeta\in\{\underline{x},\overline{x}\}} |\zeta|$*. Moreover, if* $\mu \geq 2$*, then* $\underline{\phi}^{\mathrm{C}}$ *and* $\overline{\phi}^{\mathrm{C}}$ *satisfy the demands of Corollary 3.*

Proposition 8 *Suppose that the conditions of Corollary 1 hold, and that* ϕ *is the univariate* "max" *function* $x \mapsto \max\{x,a\}$ *for some fixed* $a \in \mathbb{R}$*. Choose any* $\mu \geq 1$*. The following*

choices of $\underline{\phi}^{\mathrm{C}}$ *and* $\overline{\phi}^{\mathrm{C}}$ *satisfy the demands of Corollary 4 when* $\nu \leq \mu$*:*

$$\underline{\phi}^{\mathrm{C}} : \boldsymbol{x} \to \mathbb{R} : \xi \mapsto \begin{cases} a, & \text{if } \overline{x} \leq a, \\ \xi, & \text{if } a \leq \underline{x}, \\ a + (\overline{x} - a)\left(\max\left\{0, \frac{\xi - a}{\overline{x} - a}\right\}\right)^{\mu+1}, & \text{otherwise,} \end{cases}$$

$$\overline{\phi}^{\mathrm{C}} : \boldsymbol{x} \to \mathbb{R} : \xi \mapsto \max\{a, \underline{x}\} + (\max\{a, \overline{x}\} - \max\{a, \underline{x}\})\left(\frac{\xi - \underline{x}}{\overline{x} - \underline{x}}\right),$$

with $\xi^*_{\min} := \underline{x}$ *and* $\xi^*_{\max} := \overline{x}$*.*

The univariate "min" function may then be handled using the identity:

$$\min\{x, a\} \equiv -\max\{-x, -a\}.$$

4 Differentiable relaxations of multivariate intrinsic functions

This section presents conditions under which the multivariate McCormick relaxations provided by Theorem 1 are Whitney-$\mathscr{C}^1$. Satisfaction of these conditions is illustrated in the special cases of bivariate addition, multiplication, "max", and "min".

In general, the approach of the previous section is not transferable to this multivariate case. While Proposition 4 shows that the convex and concave envelopes of a univariate Whitney-$\mathscr{C}^1$ function on an interval are themselves Whitney-$\mathscr{C}^1$, this is not necessarily the case for multivariate Whitney-$\mathscr{C}^1$ functions. For example, the convex and concave envelopes of the bivariate product mapping $(x,y) \mapsto xy$ on a box $X \subset \mathbb{IR}^2$ are both nonsmooth whenever X has nonempty interior. Thus, although Mitsos et al. recommend setting the relaxations $\underline{\phi}^{\mathrm{C}}$ and $\overline{\phi}^{\mathrm{C}}$ in Theorem 1 to be convex and concave envelopes of ϕ on X where possible [63, 42], obtaining Whitney-$\mathscr{C}^1$ relaxations may require employing weaker relaxations of ϕ instead.

4.1 Establishing differentiability

Theorem 3 *Suppose that the conditions of Theorem 1 are satisfied, and that, for some* $\nu \in \mathbb{N}$*, the supplied relaxations* $\underline{f}_i^{\mathrm{C}}$*,* $\overline{f}_i^{\mathrm{C}}$*,* $\underline{\phi}^{\mathrm{C}}$*, and* $\overline{\phi}^{\mathrm{C}}$ *are Whitney-*$\mathscr{C}^\nu$ *on their respective domains. Suppose that, for each* $i \in \{1, \ldots, m\}$*, the partial derivative* $\frac{\partial \underline{\phi}^{\mathrm{C}}}{\partial x_i}(\xi)$ *is either nonnegative for all* $\xi \in X$ *or nonpositive for all* $\xi \in X$*. Then, the convex underestimator* $\underline{g}^{\mathrm{C}}$ *is Whitney-*$\mathscr{C}^\nu$ *on* Z*. Similarly, if, for each* $i \in \{1, \ldots, m\}$*, the partial derivative* $\frac{\partial \overline{\phi}^{\mathrm{C}}}{\partial x_i}(\xi)$ *is either nonnegative for all* $\xi \in X$ *or nonpositive for all* $\xi \in X$*, then the concave overestimator* $\overline{g}^{\mathrm{C}}$ *is Whitney-*$\mathscr{C}^\nu$ *on* Z*.*

Proof The claim regarding $\underline{g}^{\mathrm{C}}$ will be demonstrated; a similar argument yields the claim regarding $\overline{g}^{\mathrm{C}}$. The optimization problem (1) can be solved by inspection to yield:

$$\underline{g}^{\mathrm{C}} : z \mapsto \underline{\phi}^{\mathrm{C}}(f_1^*(z), \ldots, f_m^*(z)),$$

where, for each $i \in \{1, \ldots, m\}$, the mapping $f_i^* : Z \to X_i$ is defined as follows. If $\frac{\partial \underline{\phi}^{\mathrm{C}}}{\partial x_i}$ is nonnegative on X, then set $f_i^* := \underline{f}_i^{\mathrm{C}}$; otherwise, if $\frac{\partial \underline{\phi}^{\mathrm{C}}}{\partial x_i}$ is nonpositive on X, then set $f_i^* := \overline{f}_i^{\mathrm{C}}$. In either case, f_i^* is Whitney-$\mathscr{C}^\nu$, and so $\underline{g}^{\mathrm{C}}$ is as well. □

Observe that the requirements of Theorem 3 concerning $\underline{\phi}^{\mathrm{C}}$ and $\overline{\phi}^{\mathrm{C}}$ are satisfied in the special case where ϕ is affine and $\underline{\phi}^{\mathrm{C}} := \overline{\phi}^{\mathrm{C}} := \phi$. As the following section will show, the supplied convex relaxation $\underline{\phi}^{\mathrm{C}}$ for the product mapping $\phi : (x,y) \mapsto xy$ may be chosen to satisfy the requirements of Theorem 3 and [42, Theorem 6] simultaneously, when X is contained in either the positive or negative quadrant of $\mathbb{R}^2$.

The following theorem demands less stringent requirements of the supplied relaxations of ϕ and f_i than the previous theorem, but can only guarantee continuous differentiability of the obtained relaxations of g.

Theorem 4 *Suppose that the conditions of Theorem 1 are satisfied with $m = 2$, and that X_1 and X_2 have nonempty interior. In addition, suppose that the supplied relaxations $\underline{f}_i^{\mathrm{C}}$, $\overline{f}_i^{\mathrm{C}}$, $\underline{\phi}^{\mathrm{C}}$, and $\overline{\phi}^{\mathrm{C}}$ are each Whitney-$\mathscr{C}^1$ on their respective domains, and that there exist Whitney extensions $\underline{\psi}^{\mathrm{C}}$ and $\overline{\psi}^{\mathrm{C}}$ of $\underline{\phi}^{\mathrm{C}}$ and $\overline{\phi}^{\mathrm{C}}$ that satisfy all of the following conditions:*

1. *$\underline{\psi}^{\mathrm{C}}$ is convex and $\overline{\psi}^{\mathrm{C}}$ is concave on some open convex superset Y of X,*
2. *there exists a nonzero vector $\underline{d} \in \mathbb{R}^2$ for which either:*
   - *$\langle \nabla \underline{\psi}^{\mathrm{C}}(\xi), \underline{d} \rangle > 0$ for each $\xi \in Y$, or*
   - *$Y = \mathbb{R}^2$ and $\langle \nabla \underline{\psi}^{\mathrm{C}}(\xi), \underline{d} \rangle = 0$ for each $\xi \in \mathbb{R}^2$,*
3. *there exists a nonzero vector $\overline{d} \in \mathbb{R}^2$ for which either:*
   - *$\langle \nabla \overline{\psi}^{\mathrm{C}}(\xi), \overline{d} \rangle > 0$ for each $\xi \in Y$, or*
   - *$Y = \mathbb{R}^2$ and $\langle \nabla \overline{\psi}^{\mathrm{C}}(\xi), \overline{d} \rangle = 0$ for each $\xi \in \mathbb{R}^2$.*

Then, the relaxations $\underline{g}^{\mathrm{C}}$ and $\overline{g}^{\mathrm{C}}$ are both Whitney-$\mathscr{C}^1$ on Z.

Proof See Appendix A.3. □

Observe that, unlike [9, Proposition 3.3.3], this theorem does not require the optimization problems (1) or (2) to satisfy second-order sufficient optimality conditions, and does not require these problems to have unique solutions. As the following sections will show, when ϕ is the mapping $(x,y) \in \mathbb{R}^2 \mapsto xy$ on any compact subdomain X, the supplied relaxations $\underline{\phi}^{\mathrm{C}}$ and $\overline{\phi}^{\mathrm{C}}$ can be chosen to satisfy the requirements of Theorem 4.

4.2 Relaxing sums and products

Theorem 5 *Suppose that the conditions of Theorem 1 hold with $m = 2$, and that ϕ is the sum mapping $(x_1, x_2) \mapsto x_1 + x_2$. Suppose that, for some $\nu \in \mathbb{N}$, the supplied relaxations $\underline{f}_i^{\mathrm{C}}$ and $\overline{f}_i^{\mathrm{C}}$ are Whitney-$\mathscr{C}^\nu$ on their respective domains. Then, the following mapping is a Whitney-$\mathscr{C}^\nu$ convex relaxation of $g \equiv f_1 + f_2$ on Z:*

$$\underline{g}_+^{\mathrm{C}} : Z \to \mathbb{R} : z \mapsto \underline{f}_1^{\mathrm{C}}(z) + \underline{f}_2^{\mathrm{C}}(z),$$

and the following mapping is a Whitney-$\mathscr{C}^\nu$ concave relaxation of g on Z:

$$\overline{g}_+^{\mathrm{C}} : Z \to \mathbb{R} : z \mapsto \overline{f}_1^{\mathrm{C}}(z) + \overline{f}_2^{\mathrm{C}}(z).$$

Gradients of the mappings $\underline{g}_+^{\mathrm{C}}$ and $\overline{g}_+^{\mathrm{C}}$ are as follows, for each $z \in Z$:

$$\nabla \underline{g}_+^{\mathrm{C}}(z) = \nabla \underline{f}_1^{\mathrm{C}}(z) + \nabla \underline{f}_2^{\mathrm{C}}(z), \qquad \text{and} \qquad \nabla \overline{g}_+^{\mathrm{C}}(z) = \nabla \overline{f}_1^{\mathrm{C}}(z) + \nabla \overline{f}_2^{\mathrm{C}}(z).$$

Proof The required results can be obtained directly, but also follow immediately from Theorems 1 and 3 when $\underline{\phi}^{\mathrm{C}}$ and $\overline{\phi}^{\mathrm{C}}$ are each chosen to be ϕ. □

The relaxations of sums in Theorem 5 and univariate intrinsic functions in Section 3 may be combined to relax products using the identities:

$$xy \equiv \tfrac{1}{4}((x+y)^2-(x-y)^2) \equiv \tfrac{1}{2}((x+y)^2-x^2-y^2).$$

Though these approaches would lead to valid Whitney-$\mathscr{C}^2$ relaxations with second-order pointwise convergence, they could introduce unnecessary domain violations into a propagated natural interval extension. Moreover, the corresponding relaxations may not be tight in general, due to the repeated appearance of each variable [44]. A dedicated treatment of products can lead to much tighter relaxations.

Note that multiplication by a constant was already considered as a univariate intrinsic function in Table 1. To handle products of two variables, our suggested Whitney-$\mathscr{C}^1$ relaxations of the bilinear product mapping $(x,y) \mapsto xy$ involve the following intermediate functions.

Definition 3 Define $\mathbb{IR}_{\mathrm{prop}} := \{\boldsymbol{x} \in \mathbb{IR} : \underline{x} < \overline{x}\}$. For any $\mu \geq 1$, define the following scalar-valued mappings; in each case, $x,y \in \mathbb{R}$, $\boldsymbol{\zeta},\boldsymbol{\eta} \in \mathbb{IR}_{\mathrm{prop}}$, and $\boldsymbol{\alpha},\boldsymbol{\beta} \in \mathbb{IR}$.

$$\begin{aligned}
\underline{\delta}_{\times,\mathrm{A}} &: (x,y,\boldsymbol{\zeta},\boldsymbol{\eta}) \mapsto (\overline{\zeta}-\underline{\zeta})(\overline{\eta}-\underline{\eta}) \left| \tfrac{y-\underline{\eta}}{\overline{\eta}-\underline{\eta}} - \tfrac{\overline{\zeta}-x}{\overline{\zeta}-\underline{\zeta}} \right|^{\mu+1}, \\
\underline{\delta}_{\times,\mathrm{B}} &: (x,y,\boldsymbol{\zeta},\boldsymbol{\eta}) \mapsto (\overline{\zeta}-\underline{\zeta})(\overline{\eta}-\underline{\eta}) \left(\max\left\{0, \tfrac{y-\underline{\eta}}{\overline{\eta}-\underline{\eta}} - \tfrac{\overline{\zeta}-x}{\overline{\zeta}-\underline{\zeta}}\right\}\right)^{\mu+1}, \\
\underline{\psi}_{\times,\mathrm{A}} &: (x,y;\boldsymbol{\zeta},\boldsymbol{\eta}) \mapsto \tfrac{1}{2}\left(x(\underline{\eta}+\overline{\eta}) + y(\underline{\zeta}+\overline{\zeta}) - (\underline{\zeta}\,\underline{\eta}+\overline{\zeta}\,\overline{\eta}) + \underline{\delta}_{\times,\mathrm{A}}(x,y;\boldsymbol{\zeta},\boldsymbol{\eta})\right), \\
\underline{\psi}_{\times,\mathrm{B}} &: (x,y;\boldsymbol{\zeta},\boldsymbol{\eta}) \mapsto x\underline{\eta} + y\underline{\zeta} - \underline{\zeta}\,\underline{\eta} + \underline{\delta}_{\times,\mathrm{B}}(x,y;\boldsymbol{\zeta},\boldsymbol{\eta}),
\end{aligned}$$

$$\sigma_\mu : x \mapsto \begin{cases} x^{\frac{1}{\mu}}, & \text{if } 0 \leq x, \\ -|x|^{\frac{1}{\mu}}, & \text{if } x < 0, \end{cases}$$

$$\begin{aligned}
x^* &: (y,\boldsymbol{\zeta},\boldsymbol{\eta}) \mapsto \underline{\zeta} + (\overline{\zeta}-\underline{\zeta})\left(\frac{\overline{\eta}-y}{\overline{\eta}-\underline{\eta}} + \sigma_\mu\left(\tfrac{\underline{\eta}+\overline{\eta}}{(\mu+1)(\overline{\eta}-\underline{\eta})}\right)\right), \\
y^* &: (x,\boldsymbol{\zeta},\boldsymbol{\eta}) \mapsto \underline{\eta} + (\overline{\eta}-\underline{\eta})\left(\frac{\overline{\zeta}-x}{\overline{\zeta}-\underline{\zeta}} + \sigma_\mu\left(\tfrac{\underline{\zeta}+\overline{\zeta}}{(\mu+1)(\overline{\zeta}-\underline{\zeta})}\right)\right), \\
\underline{g}^{\mathrm{C}}_{\times,\mathrm{A}} &: (\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\zeta},\boldsymbol{\eta}) \mapsto \min\Big\{\underline{\psi}_{\times,\mathrm{A}}\left(\mathrm{mid}\left(\underline{\alpha},\overline{\alpha},x^*(\underline{\beta},\boldsymbol{\zeta},\boldsymbol{\eta})\right),\underline{\beta},\boldsymbol{\zeta},\boldsymbol{\eta}\right), \\
&\qquad \underline{\psi}_{\times,\mathrm{A}}\left(\mathrm{mid}\left(\underline{\alpha},\overline{\alpha},x^*(\overline{\beta},\boldsymbol{\zeta},\boldsymbol{\eta})\right),\overline{\beta},\boldsymbol{\zeta},\boldsymbol{\eta}\right), \\
&\qquad \underline{\psi}_{\times,\mathrm{A}}\left(\underline{\alpha},\mathrm{mid}\left(\underline{\beta},\overline{\beta},y^*(\underline{\alpha},\boldsymbol{\zeta},\boldsymbol{\eta})\right),\boldsymbol{\zeta},\boldsymbol{\eta}\right), \\
&\qquad \underline{\psi}_{\times,\mathrm{A}}\left(\overline{\alpha},\mathrm{mid}\left(\underline{\beta},\overline{\beta},y^*(\overline{\alpha},\boldsymbol{\zeta},\boldsymbol{\eta})\right),\boldsymbol{\zeta},\boldsymbol{\eta}\right)\Big\}.
\end{aligned}$$

Theorem 6 *Suppose that the conditions of Theorem 1 hold with $m = 2$, that ϕ is the bilinear product mapping $(\xi_1, \xi_2) \mapsto \xi_1 \xi_2$, and that X_1 and X_2 are nondegenerate intervals $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{IR}_{\text{prop}}$, respectively. Suppose that, for some $\nu \in \mathbb{N}$, the supplied relaxations $\underline{f}_i^{\text{C}}$ and $\overline{f}_i^{\text{C}}$ are Whitney-$\mathscr{C}^\nu$ on their respective domains, and that the value of μ employed in Definition 3 satisfies $\mu \geq \nu$. For each $i \in \{1,2\}$ and $z \in Z$, let $\boldsymbol{f}_i^{\text{C}}(z)$ denote the interval $[\underline{f}_i^{\text{C}}(z), \overline{f}_i^{\text{C}}(z)] \in \mathbb{IR}$. For any interval $\boldsymbol{q} \in \mathbb{IR}$, let $-\boldsymbol{q}$ denote the interval $[-\overline{q}, -\underline{q}] \in \mathbb{IR}$.*

Then, the following mapping is a Whitney-$\mathscr{C}^1$ convex relaxation of $g \equiv f_1 f_2$ on Z:

$$\underline{g}_{\times}^{\text{C}} : Z \to \mathbb{R} : z \mapsto \begin{cases} \underline{\psi}_{\times,\text{B}}\left(\underline{f}_1^{\text{C}}(z), \underline{f}_2^{\text{C}}(z), \boldsymbol{x}_1, \boldsymbol{x}_2\right), & \text{if both } 0 \leq \underline{x}_1 \text{ and } 0 \leq \underline{x}_2, \\ \underline{\psi}_{\times,\text{B}}\left(-\overline{f}_1^{\text{C}}(z), -\overline{f}_2^{\text{C}}(z), -\boldsymbol{x}_1, -\boldsymbol{x}_2\right), & \text{if both } \overline{x}_1 \leq 0 \text{ and } \overline{x}_2 \leq 0, \\ \underline{g}_{\times,A}^{\text{C}}\left(\boldsymbol{f}_1^{\text{C}}(z), \boldsymbol{f}_2^{\text{C}}(z), \boldsymbol{x}_1, \boldsymbol{x}_2\right), & \text{otherwise,} \end{cases}$$

and the following mapping is a Whitney-$\mathscr{C}^1$ concave relaxation of g on Z:

$$\overline{g}_{\times}^{\text{C}} : Z \to \mathbb{R} : z \mapsto \begin{cases} -\underline{\psi}_{\times,\text{B}}\left(-\overline{f}_1^{\text{C}}(z), \underline{f}_2^{\text{C}}(z), -\boldsymbol{x}_1, \boldsymbol{x}_2\right), & \text{if both } \overline{x}_1 \leq 0 \text{ and } 0 \leq \underline{x}_2, \\ -\underline{\psi}_{\times,\text{B}}\left(\underline{f}_1^{\text{C}}(z), -\overline{f}_2^{\text{C}}(z), \boldsymbol{x}_1, -\boldsymbol{x}_2\right), & \text{if both } 0 \leq \underline{x}_1 \text{ and } \overline{x}_2 \leq 0, \\ -\underline{g}_{\times,A}^{\text{C}}\left(-\boldsymbol{f}_1^{\text{C}}(z), \boldsymbol{f}_2^{\text{C}}(z), -\boldsymbol{x}_1, \boldsymbol{x}_2\right), & \text{otherwise.} \end{cases}$$

Gradients of $\underline{g}_{\times}^{\text{C}}$ and $\overline{g}_{\times}^{\text{C}}$ are provided in Proposition 15 in Appendix B. If each of the following conditions is also satisfied:

- $\mu \geq \nu \geq 2$,
- $\underline{f}_1^{\text{C}}$, $\overline{f}_1^{\text{C}}$, $\underline{f}_2^{\text{C}}$, *and* $\overline{f}_2^{\text{C}}$ *are Whitney-$\mathscr{C}^\nu$ on their respective domains,*
- *either* $0 \leq \underline{x}_1$ *or* $\overline{x}_1 \leq 0$, *and*
- *either* $0 \leq \underline{x}_2$ *or* $\overline{x}_2 \leq 0$,

then $\underline{g}_{\times}^{\text{C}}$ and $\overline{g}_{\times}^{\text{C}}$ are Whitney-$\mathscr{C}^\nu$ on Z.

Proof It suffices to demonstrate only the claims concerning $\underline{g}_{\times}^{\text{C}}$, for the following reason. Observe that $-\overline{f}_1^{\text{C}}$ is a convex relaxation of $-f_1$ on Z, and that $-\underline{f}_1^{\text{C}}$ is a concave relaxation of $-f_1$ on Z. Thus, if the claims concerning $\underline{g}_{\times}^{\text{C}}$ are demonstrated, then, since f_1 and f_2 were assigned arbitrarily, comparison of the definitions of $\underline{g}_{\times}^{\text{C}}$ and $\overline{g}_{\times}^{\text{C}}$ shows that $-\overline{g}_{\times}^{\text{C}}$ is a Whitney-$\mathscr{C}^\nu$ convex relaxation of the product $z \mapsto (-f_1(z)) f_2(z)$ on Z. Since $g(z) \equiv -((-f_1(z)) f_2(z))$, $\overline{g}_{\times}^{\text{C}}$ is then the required Whitney-$\mathscr{C}^\nu$ concave relaxation of g on Z.

To demonstrate the claims concerning $\underline{g}_{\times}^{\text{C}}$, first, consider the case in which both $0 \leq \underline{x}_1$ and $0 \leq \overline{x}_2$. In this case, it will be shown that the requirements of Theorem 3 concerning $\underline{g}^{\text{C}}$ are met with the substitutions $\underline{g}^{\text{C}} := \underline{g}_{\times}^{\text{C}}$ and $\underline{\phi}^{\text{C}} : (\xi_1, \xi_2) \mapsto \underline{\psi}_{\times,\text{B}}(\xi_1, \xi_2, \boldsymbol{x}_1, \boldsymbol{x}_2)$. Since the mapping $\xi \mapsto (\max\{0, \xi\})^{\mu+1}$ is increasing, convex, and $\mathscr{C}^\mu$ on $\mathbb{R}$ for any $\mu \geq 1$, the mapping $(x, y) \mapsto \underline{\delta}_{\times,\text{B}}(x, y, \boldsymbol{x}_1, \boldsymbol{x}_2)$ is convex and $\mathscr{C}^\mu$ on $\mathbb{R}^2$, and has nonnegative partial derivatives $\frac{\partial \delta_{\times,\text{B}}}{\partial x}$ and $\frac{\partial \delta_{\times,\text{B}}}{\partial y}$. Thus, $\underline{\phi}^{\text{C}}$ is convex and $\mathscr{C}^\mu$ on $\mathbb{R}^2$, and has nonnegative partial derivatives $\frac{\partial \underline{\phi}^{\text{C}}}{\partial x_1}$ and $\frac{\partial \underline{\phi}^{\text{C}}}{\partial x_2}$. Since the mapping $z \mapsto (\max\{0, z\})^{\mu+1}$ is dominated on $\{z \in \mathbb{R} : z \leq 1\}$ by $z \mapsto \max\{0, z\}$, $\underline{\phi}^{\text{C}}$ is dominated on Z by the convex envelope of ϕ on Z, as required.

The case in which both $\overline{x}_1 \leq 0$ and $\overline{x}_2 \leq 0$ is then handled by noting that the previous case applies to the reformulation of g as the "positive-orthant" product $z \mapsto (-f_1(z))(-f_2(z))$ on Z.

Lastly, consider the case in which $\underline{g}^{\mathrm{C}}$ is set to $(x_1,x_2) \mapsto \underline{g}^{\mathrm{C}}_{\times,\mathrm{A}}(\xi_1,\xi_2,\boldsymbol{x}_1,\boldsymbol{x}_2)$. In this case, define $\underline{\phi}^{\mathrm{C}} : (\xi_1,\xi_2) \in \mathbb{R}^2 \mapsto \underline{\psi}_{\times,\mathrm{A}}(\xi_1,\xi_2,\boldsymbol{x}_1,\boldsymbol{x}_2)$. Define $d := (\overline{x}_1 - \underline{x}_1, -(\overline{x}_2 - \underline{x}_2)) \in \mathbb{R}^2$, and observe that, for each $\xi \in \mathbb{R}^2$,

$$\langle \nabla \underline{\phi}^{\mathrm{C}}(\xi), d\rangle = \tfrac{1}{2}(\underline{x}_2 + \overline{x}_2)d_1 + \tfrac{1}{2}(\underline{x}_1 + \overline{x}_1)d_2 + 0 = \overline{x}_1\underline{x}_2 - \underline{x}_1\overline{x}_2. \tag{4}$$

Thus, to show that some case in Theorem 4 applies with either $\underline{d} := d$ or $\underline{d} := -d$, it suffices to show that $\underline{\phi}^{\mathrm{C}}$ is a Whitney-$\mathscr{C}^1$ convex relaxation of ϕ on X, and that, for any intervals $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2 \in \mathbb{IR}_{\mathrm{prop}}$ with $\boldsymbol{\zeta}_1 \subset \boldsymbol{x}_1$ and $\boldsymbol{\zeta}_2 \subset \boldsymbol{x}_2$,

$$\underline{g}^{\mathrm{C}}_{\times,\mathrm{A}}(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2,\boldsymbol{x}_1,\boldsymbol{x}_2) = \min\{\underline{\phi}^{\mathrm{C}}(\xi) :\ \underline{\zeta}_i \leq \xi_i \leq \overline{\zeta}_i,\ \forall i \in \{1,2\}\}. \tag{5}$$

Since the mapping $z \in \mathbb{R} \mapsto |z|^{\mu+1}$ is convex and $\mathscr{C}^1$, it follows that $\underline{\phi}^{\mathrm{C}}$ is both convex and $\mathscr{C}^1$ on $\mathbb{R}^2$. Moreover, since the mapping $z \mapsto |z|^{\mu+1}$ is nonnegative and is dominated on $[-1,1]$ by $z \mapsto |z|$, $\underline{\phi}^{\mathrm{C}}$ is dominated on X by the convex envelope of ϕ. To demonstrate (5), choose any intervals $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2 \in \mathbb{IR}_{\mathrm{prop}}$ with $\boldsymbol{\zeta}_1 \subset \boldsymbol{x}_1$ and $\boldsymbol{\zeta}_2 \subset \boldsymbol{x}_2$. Due to (4), the proofs of Lemmas 3 and Lemma 4 show that there exists a solution of the convex program

$$\min\{\underline{\phi}^{\mathrm{C}}(\xi) :\ \underline{\zeta}_i \leq \xi_i \leq \overline{\zeta}_i,\ \forall i \in \{1,2\}\} \tag{6}$$

in one of the four edges of the box $B := \{(\xi_1,\xi_2) \in \mathbb{R}^2 : \xi_1 \in \boldsymbol{\zeta}_1, \xi_2 \in \boldsymbol{\zeta}_2\}$. Moreover, since this convex program is linearly constrained, the Karush-Kuhn-Tucker (KKT) optimality conditions are necessary and sufficient for any solution of (6). Now, observe that, for any $\eta \in \mathbb{R}$,

$$\frac{\partial \underline{\phi}^{\mathrm{C}}}{\partial x_1}(x^*(\eta,\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2),\eta) = 0 \qquad \text{and} \qquad \frac{\partial \underline{\phi}^{\mathrm{C}}}{\partial x_2}(\eta, y^*(\eta,\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2)) = 0.$$

It follows immediately that any KKT point of (6) in the boundary of B must coincide with one of the four $\underline{\psi}_{\times,\mathrm{A}}$ terms in the definition of $\underline{g}^{\mathrm{C}}_{\times,\mathrm{A}}$. Since the KKT conditions are necessary and sufficient for optimality of (6), (5) is thereby demonstrated.

Next, it will be shown that $\underline{g}^{\mathrm{C}}_{\times}$ is Whitney-$\mathscr{C}^{\nu}$ when the conditions stated at the end of the theorem are satisfied. The cases in which either both $0 \leq \underline{x}_1$ and $0 \leq \underline{x}_2$ or both $\overline{x}_1 \leq 0$ and $\overline{x}_2 \leq 0$ have already been covered above. Consider the case in which $0 \leq \underline{x}_1$ and $\overline{x}_2 \leq 0$. In this case, it is readily verified that, for any intervals $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2 \in \mathbb{IR}_{\mathrm{prop}}$ with $\boldsymbol{\zeta}_1 \in \boldsymbol{x}_1$ and $\boldsymbol{\zeta}_2 \in \boldsymbol{x}_2$,

$$\underline{g}^{\mathrm{C}}_{\times,\mathrm{A}}(\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2,\boldsymbol{x}_1,\boldsymbol{x}_2) = \underline{\psi}_{\times,\mathrm{A}}(\overline{\zeta}_1,\underline{\zeta}_2,\boldsymbol{x}_1,\boldsymbol{x}_2).$$

This equation and (5) imply that, for each $z \in Z$, $\underline{g}^{\mathrm{C}}_{\times}(z) = \underline{\psi}_{\times,\mathrm{A}}(\overline{f}^{\mathrm{C}}_1(z), \underline{f}^{\mathrm{C}}_2(z), \boldsymbol{x}_1, \boldsymbol{x}_2)$. The chain rule for Whitney-$\mathscr{C}^{\nu}$ functions then shows that $\underline{g}^{\mathrm{C}}_{\times}$ is Whitney-$\mathscr{C}^{\nu}$. The case in which $\overline{x}_1 \leq 0$ and $0 \leq \underline{x}_2$ may be handled by a similar argument. □

4.3 Relaxing the bivariate "max" and "min" functions

As was the case with bivariate products, the established treatment of addition and univariate intrinsic functions can be combined to obtain Whitney-$\mathscr{C}^2$ relaxations of the bivariate "max" and "min" functions, since

$$\max\{x,y\} \equiv \tfrac{1}{2}(x+y+|x-y|) \equiv x+\max\{0,y-x\} \equiv y+\max\{0,x-y\}.$$

Observe that if the identity $\max\{x,y\} \equiv x+\max\{0,y-x\}$ is applied, then the resulting Whitney-$\mathscr{C}^1$ convex relaxation is guaranteed to dominate the mapping $(x,y)\mapsto x$. This property will be exploited in Section 6.1 when handling implicit functions.

When this property is not required, however, tighter Whitney-$\mathscr{C}^1$ relaxations can be obtained by treating bivariate "max" and "min" as multivariate intrinsic functions. Our suggested way to do this involves the following intermediate functions.

Definition 4 For any $\mu \geq 1$, define the following scalar-valued mappings; in each case, $x,y \in \mathbb{R}$ and $\boldsymbol{\zeta},\boldsymbol{\eta} \in \mathbb{IR}_{\text{prop}}$.

$$\underline{\psi}_{\max} : (x,y,\boldsymbol{\zeta},\boldsymbol{\eta}) \mapsto \begin{cases} x, & \text{if } \overline{\eta} \leq \underline{\zeta}, \\ y, & \text{if } \overline{\zeta} \leq \underline{\eta}, \\ x+(\overline{\eta}-\underline{\zeta})\left(\max\left\{0,\frac{y-x}{\overline{\eta}-\underline{\zeta}}\right\}\right)^{\mu+1}, & \text{if } \underline{\eta} \leq \underline{\zeta} < \overline{\eta}, \\ y+(\overline{\zeta}-\underline{\eta})\left(\max\left\{0,\frac{x-y}{\overline{\zeta}-\underline{\eta}}\right\}\right)^{\mu+1}, & \text{if } \underline{\zeta} < \underline{\eta} < \overline{\zeta}, \end{cases}$$

$$\begin{aligned} \overline{\theta} : (x,y,\boldsymbol{\zeta},\boldsymbol{\eta}) \mapsto{}& (\max\{\underline{\zeta},\underline{\eta}\}+\max\{\overline{\zeta},\overline{\eta}\}-\max\{\underline{\zeta},\overline{\eta}\}-\max\{\overline{\zeta},\underline{\eta}\}) \\ &\times \left(\max\left\{0,\frac{\overline{\zeta}-x}{\overline{\zeta}-\underline{\zeta}}-\frac{y-\underline{\eta}}{\overline{\eta}-\underline{\eta}}\right\}\right)^{\mu+1}, \end{aligned}$$

$$\overline{\psi}_{\max} : (x,y,\boldsymbol{\zeta},\boldsymbol{\eta}) \mapsto \begin{cases} x, & \text{if } \overline{\eta} \leq \underline{\zeta}, \\ y, & \text{if } \overline{\zeta} \leq \underline{\eta}, \\ \max\{\overline{\zeta},\overline{\eta}\}-(\max\{\overline{\zeta},\overline{\eta}\}-\max\{\underline{\zeta},\overline{\eta}\})\left(\frac{\overline{\zeta}-x}{\overline{\zeta}-\underline{\zeta}}\right) & \\ \quad -(\max\{\overline{\zeta},\overline{\eta}\}-\max\{\overline{\zeta},\underline{\eta}\})\left(\frac{\overline{\eta}-y}{\overline{\eta}-\underline{\eta}}\right)+\overline{\theta}(x,y,\boldsymbol{\zeta},\boldsymbol{\eta}), & \text{otherwise.} \end{cases}$$

The following theorem provides Whitney-$\mathscr{C}^1$ relaxations for the bivariate "max" function. Using these relaxations, similar relaxations may be constructed for bivariate "min" via the identity $\min\{x,y\} \equiv -\max\{-x,-y\}$.

Theorem 7 *Suppose that the conditions of Theorem 1 hold with $m=2$, that ϕ is the bivariate "max" mapping $(\xi_1,\xi_2)\mapsto\max\{\xi_1,\xi_2\}$, and that X_1 and X_2 are nondegenerate intervals $\boldsymbol{x}_1,\boldsymbol{x}_2 \in \mathbb{IR}_{\text{prop}}$, respectively. Suppose that, for some $\nu \in \mathbb{N}$, the supplied relaxations $\underline{f}_i^{\text{C}}$ and $\overline{f}_i^{\text{C}}$ are Whitney-$\mathscr{C}^\nu$ on their respective domains, and that the value of μ employed in Definition 4 satisfies $\mu \geq \nu$. Then, the following mapping $\underline{g}_{\max}^{\text{C}} : Z \to \mathbb{R}$ is a Whitney-$\mathscr{C}^1$ convex*

relaxation of $g \equiv \max\{f_1, f_2\}$ on Z:

$$\underline{g}^{\mathrm{C}}_{\max} : z \mapsto \begin{cases} \underline{\psi}_{\max}(\underline{f}^{\mathrm{C}}_1(z), \underline{f}^{\mathrm{C}}_2(z), \boldsymbol{x}_1, \boldsymbol{x}_2), & \text{if } \overline{x}_2 \leq \underline{x}_1 \text{ or } \overline{x}_1 \leq \underline{x}_2, \\ \underline{\psi}_{\max}\left(\mathrm{mid}\left(\underline{f}^{\mathrm{C}}_1(z), \overline{f}^{\mathrm{C}}_1(z), \underline{f}^{\mathrm{C}}_2(z) - (\overline{x}_2 - \underline{x}_1)(\mu+1)^{-\frac{1}{\mu}}\right), \underline{f}^{\mathrm{C}}_2(z), \boldsymbol{x}_1, \boldsymbol{x}_2\right), & \text{if } \underline{x}_2 \leq \underline{x}_1 < \overline{x}_2, \\ \underline{\psi}_{\max}\left(\underline{f}^{\mathrm{C}}_1(z), \mathrm{mid}\left(\underline{f}^{\mathrm{C}}_2(z), \overline{f}^{\mathrm{C}}_2(z), \underline{f}^{\mathrm{C}}_1(z) - (\overline{x}_1 - \underline{x}_2)(\mu+1)^{-\frac{1}{\mu}}\right), \boldsymbol{x}_1, \boldsymbol{x}_2\right), & \text{if } \underline{x}_1 < \underline{x}_2 < \overline{x}_1, \end{cases}$$

and the following mapping is a Whitney-$\mathscr{C}^\nu$ concave relaxation of g on Z:

$$\overline{g}^{\mathrm{C}}_{\max} : Z \to \mathbb{R} : z \mapsto \overline{\psi}_{\max}(\overline{f}^{\mathrm{C}}_1(z), \overline{f}^{\mathrm{C}}_2(z), \boldsymbol{x}_1, \boldsymbol{x}_2).$$

Gradients of $\underline{g}^{\mathrm{C}}_{\max}$ and $\overline{g}^{\mathrm{C}}_{\max}$ are provided in Proposition 16 in Appendix B.

Proof First, the claims concerning $\underline{g}^{\mathrm{C}}_{\max}$ will be addressed; define a mapping $\underline{\phi}^{\mathrm{C}} : \mathbb{R}^2 \to \mathbb{R} : (\xi_1, \xi_2) \mapsto \underline{\psi}_{\max}(\xi_1, \xi_2, \boldsymbol{x}_1, \boldsymbol{x}_2)$. First, consider the case in which either $\overline{x}_2 \leq \underline{x}_1$ or $\overline{x}_1 \leq \underline{x}_2$. In this case, ϕ is affine on X and $\underline{\phi}^{\mathrm{C}} \equiv \phi$; the requirements of Theorem 3 concerning $\underline{g}^{\mathrm{C}}$ are then satisfied with $\underline{g}^{\mathrm{C}} := \underline{g}^{\mathrm{C}}_{\max}$. Next, consider the case in which $\underline{x}_2 \leq \underline{x}_1 < \overline{x}_2$; an analogous argument applies to the case in which $\underline{x}_1 < \underline{x}_2 < \overline{x}_1$. Observe that the mapping $\xi \in \mathbb{R} \mapsto (\max\{0, \xi\})^{\mu+1}$ is $\mathscr{C}^1$ and convex; it follows immediately that $\underline{\phi}^{\mathrm{C}}$ is $\mathscr{C}^1$ and convex on $\mathbb{R}^2$. Moreover, observe that $\langle \nabla \underline{\phi}^{\mathrm{C}}(\xi), e_{(1)} + e_{(2)} \rangle = 1$ for each $\xi \in \mathbb{R}^2$, and that ϕ dominates $\underline{\phi}^{\mathrm{C}}$ on X; Theorem 4 implies that the mapping $\underline{g}^{\mathrm{C}} : z \in Z \mapsto \min\{\underline{\phi}^{\mathrm{C}}(\xi) : \underline{f}^{\mathrm{C}}(z) \leq \xi \leq \overline{f}^{\mathrm{C}}(z)\}$ is a Whitney-$\mathscr{C}^1$ convex relaxation of g on Z. Now, observe that, for each $\xi \in \mathbb{R}^2$,

$$\langle \nabla \underline{\phi}^{\mathrm{C}}(\xi), e_{(2)} \rangle = (\mu+1)\left(\max\left\{0, \frac{\xi_2 - \xi_1}{\overline{x}_2 - \underline{x}_1}\right\}\right)^{\mu} \geq 0;$$

the mean value theorem then implies that $\underline{g}^{\mathrm{C}}(z) = \min\{\underline{\phi}^{\mathrm{C}}(\xi_1, \underline{f}^{\mathrm{C}}_2(z)) : \underline{f}^{\mathrm{C}}_1(z) \leq \xi_1 \leq \overline{f}^{\mathrm{C}}_1(z)\}$ for each $z \in Z$. Since the mapping $\xi_1 \in \mathbb{R} \mapsto \underline{\phi}^{\mathrm{C}}(\xi_1, \underline{f}^{\mathrm{C}}_2(z))$ is convex and $\mathscr{C}^1$ for each fixed $z \in Z$, Corollary 1 shows that $\underline{g}^{\mathrm{C}} \equiv \underline{g}^{\mathrm{C}}_{\max}$ on Z, which completes this case.

Next, the claims concerning $\overline{g}^{\mathrm{C}}_{\max}$ will be addressed. It will be shown that the conditions of Theorem 3 concerning $\overline{g}^{\mathrm{C}}$ are satisfied with the substitutions $\overline{g}^{\mathrm{C}} := \overline{g}^{\mathrm{C}}_{\max}$ and $\overline{\phi}^{\mathrm{C}} : (\xi_1, \xi_2) \mapsto \overline{\psi}_{\max}(\xi_1, \xi_2, \boldsymbol{x}_1, \boldsymbol{x}_2)$. Observe again that the mapping $\xi \mapsto (\max\{0, \xi\})^{\mu+1}$ is increasing, nonnegative, convex, and $\mathscr{C}^\nu$, and that the coefficient

$$\max\{\underline{\zeta}, \underline{\eta}\} + \max\{\overline{\zeta}, \overline{\eta}\} - \max\{\underline{\zeta}, \overline{\eta}\} - \max\{\overline{\zeta}, \underline{\eta}\}$$

is nonpositive for any $\boldsymbol{\zeta}, \boldsymbol{\eta} \in \mathbb{IR}_{\mathrm{prop}}$. It follows that the mapping $(\xi_1, \xi_2) \mapsto \overline{\theta}(\xi_1, \xi_2, \boldsymbol{x}_1, \boldsymbol{x}_2)$ is concave and $\mathscr{C}^\nu$, and that the mappings $(x, y) \mapsto \frac{\partial \theta}{\partial x}(x, y, \boldsymbol{\zeta}, \boldsymbol{\eta})$ and $(x, y) \mapsto \frac{\partial \theta}{\partial y}(x, y, \boldsymbol{\zeta}, \boldsymbol{\eta})$ are nonnegative on $\mathbb{R}^2$. Thus, $\overline{\phi}^{\mathrm{C}}$ is concave and $\mathscr{C}^\nu$ on $\mathbb{R}^2$, and the partial derivatives $\frac{\partial \overline{\phi}^{\mathrm{C}}}{\partial \xi_1}$ and $\frac{\partial \overline{\phi}^{\mathrm{C}}}{\partial \xi_2}$ are nonnegative on $\mathbb{R}^2$. To show that $\overline{\phi}^{\mathrm{C}}$ dominates ϕ on X, observe that, whenever $x \in \boldsymbol{\zeta}$ and $y \in \boldsymbol{\eta}$,

$$0 \leq \left(\max\left\{0, \frac{\overline{\zeta} - x}{\overline{\zeta} - \underline{\zeta}} - \frac{y - \underline{\eta}}{\overline{\eta} - \underline{\eta}}\right\}\right)^{\mu+1} \leq \max\left\{0, \frac{\overline{\zeta} - x}{\overline{\zeta} - \underline{\zeta}} - \frac{y - \underline{\eta}}{\overline{\eta} - \underline{\eta}}\right\} \leq 1.$$

Thus, $\overline{\phi}^{\mathrm{C}}$ dominates the concave envelope of ϕ on X provided by [63, Lemma 3]. □

5 Convergence order

Intuitively, tighter convex and concave relaxations provide more useful bounding information to a numerical method that employs relaxations. Moreover, to avoid clustering [16] in branch-and-bound methods for global optimization, computed relaxations must converge to the relaxed function sufficiently rapidly as the underlying domain is shrunk [10,67,65, 42]. This section shows that the Whitney-$\mathscr{C}^1$ relaxations provided in this article satisfy this requirement, and converge to the relaxed function as rapidly as McCormick's original relaxations. To show this, we employ the notion of *pointwise convergence order* developed by Bompadre and Mitsos [10], and extended subsequently to the multivariate McCormick relaxations of Theorem 1 [42]. Combined with [42, Theorem 6], this section shows that if the methods of this article are used to relax a finite composition of Whitney-$\mathscr{C}^1$ intrinsic functions, then the obtained relaxations exhibit second-order pointwise convergence. If any employed intrinsic function is nonsmooth, then only first-order pointwise convergence is guaranteed; nevertheless, under certain nonsingularity assumptions, [65, Section 2.3] shows that first-order pointwise convergence suffices in the nonsmooth case to avoid clustering.

In this section, the sets Z and X in Theorem 1 will be varied. Where appropriate, dependence of relaxations on these sets will be denoted explicitly; for example, $\underline{\phi}^{\mathrm{C}}(\xi)$ in Theorem 1 may instead be denoted as $\underline{\phi}^{\mathrm{C}}(\xi;X)$. Notions of convergence of relaxations employ the following definition of the diameter of a set.

Definition 5 The *diameter* of a set $S\subset\mathbb{R}^n$ is $\mathrm{wid}\,S := \sup\{\|x-y\| : x,y\in S\}$.

5.1 Pointwise convergence of differentiable relaxations of intrinsic functions

This section shows that the Whitney-$\mathscr{C}^1$ relaxations $\underline{\phi}^{\mathrm{C}}(\cdot;X)$ and $\overline{\phi}^{\mathrm{C}}(\cdot;X)$ suggested for particular intrinsic functions ϕ in Sections 3 and 4 satisfy the requirements of [42, Theorem 6] concerning the pointwise convergence of supplied relaxations of an outer function $F\equiv\phi$. In each of these cases, if ϕ itself is Whitney-$\mathscr{C}^1$, then these requirements are satisfied with $\gamma_F := 2$; otherwise, they are satisfied with $\gamma_F := 1$.

Noting that the relaxations of sums suggested by Theorem 5 coincide with the classic McCormick treatment of addition, pointwise convergence of these relaxations has already been treated in [10, Theorem 3].

Proposition 9 *Suppose that ϕ is any univariate intrinsic function considered in either Table 1 or Section 3.2, and consider an interval $\boldsymbol{y}\in\mathbb{IR}$. If ϕ was chosen from Table 1, suppose additionally that $\boldsymbol{y}\subset B$. For any interval $\boldsymbol{x}\subset\boldsymbol{y}$, let $\underline{\phi}^{\mathrm{C}}(\cdot;\boldsymbol{x})$ and $\overline{\phi}^{\mathrm{C}}(\cdot;\boldsymbol{x})$ denote the particular convex and concave relaxations for ϕ on $\boldsymbol{x}$ suggested in Table 1 and Section 3.2. There exists a constant $\tau\geq 0$ (dependent on $\boldsymbol{y}$) for which, for each interval $\boldsymbol{x}\subset\boldsymbol{y}$ and each $\xi\in\boldsymbol{x}$,*

$$\phi(\xi)-\underline{\phi}^{\mathrm{C}}(\xi;\boldsymbol{x})\leq\tau(\mathrm{wid}\,\boldsymbol{x})^2,\qquad \textit{and}\qquad \overline{\phi}^{\mathrm{C}}(\xi;\boldsymbol{x})-\phi(\xi)\leq\tau(\mathrm{wid}\,\boldsymbol{x})^2.$$

Proof An appropriate constant $\tau>0$ may be computed directly for each considered choice of ϕ, $\boldsymbol{y}$, and $\boldsymbol{x}$. □

Proposition 10 *Suppose that ϕ is any univariate intrinsic function considered in Section 3.3. For any interval $\boldsymbol{x}\in\mathbb{IR}$, let $\underline{\phi}^{\mathrm{C}}(\cdot;\boldsymbol{x})$ and $\overline{\phi}^{\mathrm{C}}(\cdot;\boldsymbol{x})$ denote the particular convex and concave relaxations for ϕ on $\boldsymbol{x}$ suggested in Section 3.3. Then, for each $\boldsymbol{x}\in\mathbb{IR}$ and each $\xi\in\boldsymbol{x}$,*

$$\phi(\xi)-\underline{\phi}^{\mathrm{C}}(\xi;\boldsymbol{x})\leq\mathrm{wid}\,\boldsymbol{x},\qquad \textit{and}\qquad \overline{\phi}^{\mathrm{C}}(\xi;\boldsymbol{x})-\phi(\xi)\leq\mathrm{wid}\,\boldsymbol{x}.$$

Proof The required results are obtained by inspection. □

Proposition 11 *Consider any $Y \in \mathbb{IR}^2_{\text{prop}}$. For each $X \in \mathbb{IR}^2_{\text{prop}}$ with $X \subset Y$, consider the conditions of Theorem 6 with the substitutions $Z := X$, $f_i : z \in Z \mapsto z_i$, $\underline{f}_i^{\text{C}} : z \mapsto z_i$, and $\overline{f}_i^{\text{C}} : z \mapsto z_i$ for each $i \in \{1,2\}$. There exists a constant $\tau > 0$ (dependent on Y but not X) for which, for each $X \in \mathbb{IR}^2_{\text{prop}}$ with $X \subset Y$ and each $(\xi_1,\xi_2) \in X$,*

$$\xi_1\xi_2 - \underline{g}_{\times}^{\text{C}}(\xi_1,\xi_2;X) \le \tau(\text{wid}X)^2 \qquad \textit{and} \qquad \overline{g}_{\times}^{\text{C}}(\xi_1,\xi_2;X) - \xi_1\xi_2 \le \tau(\text{wid}X)^2.$$

Proof Let $\underline{\phi}_{\times}^{\text{C}}(\cdot;X)$ and $\overline{\phi}_{\times}^{\text{C}}(\cdot;X)$ denote the convex and concave envelopes of $(\xi_1,\xi_2) \mapsto \xi_1\xi_2$ on any $X \in \mathbb{IR}^2_{\text{prop}}$. By [10, Theorem 4], it suffices to show that for each $X = (\boldsymbol{x}_1,\boldsymbol{x}_2) \in \mathbb{IR}^2_{\text{prop}}$ and each $(\xi_1,\xi_2) \in X$,

$$\underline{\phi}_{\times}^{\text{C}}(\xi_1,\xi_2;X) - \underline{g}_{\times}^{\text{C}}(\xi_1,\xi_2;X) \le (\text{wid}X)^2 \text{ and } \overline{g}_{\times}^{\text{C}}(\xi_1,\xi_2;X) - \overline{\phi}_{\times}^{\text{C}}(\xi_1,\xi_2;X) \le (\text{wid}X)^2.$$

The first of these statements will be demonstrated; the second is analogous. Suppose that $0 \le \underline{x}_1$ and $0 \le \underline{x}_2$; a similar argument covers the case in which both $\overline{x}_1 \le 0$ and $\overline{x}_2 \le 0$. In this case, for any $\xi_1 \in \boldsymbol{x}_1$ and $\xi_2 \in \boldsymbol{x}_2$,

$$\begin{aligned}
&\underline{\phi}_{\times}^{\text{C}}(\xi_1,\xi_2;X) - \underline{g}_{\times}^{\text{C}}(\xi_1,\xi_2;X)\\
&= \underline{\phi}_{\times}^{\text{C}}(\xi_1,\xi_2;X) - \underline{\psi}_{\times,\text{B}}(\xi_1,\xi_2,\boldsymbol{x}_1,\boldsymbol{x}_2)\\
&= (\overline{x}_1 - \underline{x}_1)(\overline{x}_2 - \underline{x}_2)\max\left\{0, \frac{\xi_2 - \underline{x}_2}{\overline{x}_2 - \underline{x}_2} - \frac{\overline{x}_1 - \xi_1}{\overline{x}_1 - \underline{x}_1}\right\}\left(1 - \left(\max\left\{0, \frac{\xi_2 - \underline{x}_2}{\overline{x}_2 - \underline{x}_2} - \frac{\overline{x}_1 - \xi_1}{\overline{x}_1 - \underline{x}_1}\right\}\right)^{\mu}\right)\\
&\le (\text{wid}\,\boldsymbol{x}_1)\cdot(\text{wid}\,\boldsymbol{x}_2)\cdot 1 \cdot 1 \le (\text{wid}X)^2.
\end{aligned}$$

Next, suppose that $\min\{\underline{x}_1,\underline{x}_2\} < 0 < \max\{\overline{x}_1,\overline{x}_2\}$. In this case, for any $\xi_1 \in \boldsymbol{x}_1$ and $\xi_2 \in \boldsymbol{x}_2$,

$$\begin{aligned}
&\underline{\phi}_{\times}^{\text{C}}(\xi_1,\xi_2;X) - \underline{g}_{\times}^{\text{C}}(\xi_1,\xi_2;X)\\
&= \underline{\phi}_{\times}^{\text{C}}(\xi_1,\xi_2;X) - \underline{\psi}_{\times,\text{A}}(\xi_1,\xi_2,\boldsymbol{x}_1,\boldsymbol{x}_2)\\
&= \tfrac{1}{2}(\overline{x}_1 - \underline{x}_1)(\overline{x}_2 - \underline{x}_2)\left|\frac{\xi_2 - \underline{x}_2}{\overline{x}_2 - \underline{x}_2} - \frac{\overline{x}_1 - \xi_1}{\overline{x}_1 - \underline{x}_1}\right|\left(1 - \left|\frac{\xi_2 - \underline{x}_2}{\overline{x}_2 - \underline{x}_2} - \frac{\overline{x}_1 - \xi_1}{\overline{x}_1 - \underline{x}_1}\right|^{\mu}\right)\\
&\le \tfrac{1}{2}(\text{wid}\,\boldsymbol{x}_1)\cdot(\text{wid}\,\boldsymbol{x}_2)\cdot 1 \cdot 1 \le (\text{wid}X)^2,
\end{aligned}$$

as required. □

Proposition 12 *Consider any $Y \in \mathbb{IR}^2_{\text{prop}}$. For each $X \in \mathbb{IR}^2_{\text{prop}}$ with $X \subset Y$, consider the conditions of Theorem 7 with the substitutions $Z := X$, $f_i : z \in Z \mapsto z_i$, $\underline{f}_i^{\text{C}} : z \mapsto z_i$, and $\overline{f}_i^{\text{C}} : z \mapsto z_i$ for each $i \in \{1,2\}$. There exists a constant $\tau > 0$ (dependent on Y but not X) for which, for each $X \in \mathbb{IR}^2_{\text{prop}}$ with $X \subset Y$ and each $(\xi_1,\xi_2) \in X$,*

$$\max\{\xi_1,\xi_2\} - \underline{g}_{\max}^{\text{C}}(\xi_1,\xi_2;X) \le \tau\text{wid}X \quad \textit{and} \quad \overline{g}_{\max}^{\text{C}}(\xi_1,\xi_2;X) - \max\{\xi_1,\xi_2\} \le \tau\text{wid}X.$$

Proof Set $X \equiv (\boldsymbol{x}_1,\boldsymbol{x}_2)$. Under the conditions of the proposition, observe that $\underline{g}_{\max}^{\text{C}}(\xi_1,\xi_2;X) = \underline{\psi}_{\max}(\xi_1,\xi_2,\boldsymbol{x}_1,\boldsymbol{x}_2)$ and $\overline{g}_{\max}^{\text{C}}(\xi_1,\xi_2;X) = \overline{\psi}_{\max}(\xi_1,\xi_2,\boldsymbol{x}_1,\boldsymbol{x}_2)$. Now, from Proposition 16, observe that $2(\mu+1)$ is a Lipschitz constant for $\underline{\psi}_{\max}(\cdot,\cdot,\boldsymbol{x}_1,\boldsymbol{x}_2)$ on X, and that the mapping

$(\xi_1,\xi_2)\mapsto\max\{\xi_1,\xi_2\}$ has a Lipschitz constant of 1. Observe also that $\underline{\psi}_{\max}(\underline{x}_1,\underline{x}_2,\boldsymbol{x}_1,\boldsymbol{x}_2)=\max\{\underline{x}_1,\underline{x}_2\}$; thus,

$$\begin{aligned}
&\max\{\xi_1,\xi_2\}-\underline{g}^{\mathrm{C}}_{\max}(\xi_1,\xi_2;X)\\
&\quad=|\max\{\xi_1,\xi_2\}-\underline{g}^{\mathrm{C}}_{\max}(\xi_1,\xi_2;X)|\\
&\quad\le|\max\{\xi_1,\xi_2\}-\max\{\underline{x}_1,\underline{x}_2\}|+|\underline{g}^{\mathrm{C}}_{\max}(\underline{x}_1,\underline{x}_2;X)-\underline{g}^{\mathrm{C}}_{\max}(\xi_1,\xi_2;X)|\\
&\quad\le\|(\xi_1,\xi_2)-(\underline{x}_1,\underline{x}_2)\|+2(\mu+1)\|(\xi_1,\xi_2)-(\underline{x}_1,\underline{x}_2)\|\\
&\quad\le(2\mu+3)\mathrm{wid}\,X.
\end{aligned}$$

Noting that $\overline{\psi}_{\max}(\overline{x}_1,\overline{x}_2,\boldsymbol{x}_1,\boldsymbol{x}_2)=\max\{\overline{x}_1,\overline{x}_2\}$, and that $2+2(\mu+1)(\max\{\underline{y}_1\overline{y}_2+\overline{y}_1\underline{y}_2-\underline{y}_1\underline{y}_2-\overline{y}_1\overline{y}_2\})$ is a Lipschitz constant for $\overline{\psi}_{\max}(\cdot,\cdot,\boldsymbol{x}_1,\boldsymbol{x}_2)$ on X, a similar argument shows that

$$\overline{g}^{\mathrm{C}}_{\max}(\xi_1,\xi_2;X)-\max\{\xi_1,\xi_2\}\le(3+2(\mu+1)(\max\{\underline{y}_1\overline{y}_2+\overline{y}_1\underline{y}_2-\underline{y}_1\underline{y}_2-\overline{y}_1\overline{y}_2\}))\mathrm{wid}\,X,$$

as required. □

5.2 Propagating interval bounds

The results of the previous section show that Theorem 1 produces relaxations with second-order pointwise convergence when the Whitney-$\mathscr{C}^1$ relaxations described in Sections 3 and 4 are employed, subject to two requirements on the set X. Firstly, Theorem 1 requires that each set X_i contains the images of Z under both $\underline{f}_i^{\mathrm{C}}(\cdot;Z)$ and $\overline{f}_i^{\mathrm{C}}(\cdot;Z)$, to ensure that the constructed relaxations of g are well-defined. Secondly, [42, Theorem 6] requires X to exhibit *first-order Hausdorff convergence* as Z shrinks, to ensure second-order pointwise convergence of $\underline{g}^{\mathrm{C}}(\cdot;Z)$ and $\overline{g}^{\mathrm{C}}(\cdot;Z)$. To use this result repeatedly in an inductive argument to describe second-order pointwise convergence of a finite composition of several functions, these two requirements on X must also apply to some computable interval enclosing the images of Z under g, $\underline{g}^{\mathrm{C}}(\cdot;Z)$, and $\overline{g}^{\mathrm{C}}(\cdot;Z)$ simultaneously.

Interval arithmetic [41,44] has been shown [10,53,42] to produce appropriate intervals satisfying both requirements when the traditional McCormick relaxations are employed, or when Theorem 1 is applied using convex and concave envelopes of intrinsic functions. The Whitney-$\mathscr{C}^1$ relaxations of intrinsic functions presented in this article may be weaker than the corresponding convex and concave envelopes, however, which may lead to failure of the first requirement. Thus, this section shows that the first requirement on interval enclosures of g, $\underline{g}^{\mathrm{C}}(\cdot;Z)$, and $\overline{g}^{\mathrm{C}}(\cdot;Z)$ remains satisfied by standard interval arithmetic for all intrinsic functions considered in this article except the bivariate product $(\xi_1,\xi_2)\mapsto\xi_1\xi_2$, whose Whitney-$\mathscr{C}^1$ relaxations are not necessarily bounded by the standard interval product. In this case, a weaker interval product rule is provided that is shown to satisfy both requirements.

Proposition 13 *Suppose that ϕ is any intrinsic function considered in either Table 1, Sections 3.2 or 3.3, or Theorems 5, 6, or 7, and let $\underline{g}^{\mathrm{C}}$ and $\overline{g}^{\mathrm{C}}$ denote the corresponding convex and concave relaxations provided by Theorems 2, 5, 6, or 7. Then,*

$$\min\{\phi(\xi):\xi\in X\}\le\min\{\underline{g}^{\mathrm{C}}(z):z\in Z\}\le\max\{\overline{g}^{\mathrm{C}}(z):z\in Z\}\le\max\{\phi(\xi):\xi\in X\},$$

except possibly in the case where $\phi:(\xi_1,\xi_2)\mapsto\xi_1\xi_2$ and $\min\{\underline{x}_1,\underline{x}_2\}<0<\max\{\overline{x}_1,\overline{x}_2\}$.

Proof In each case, inspection of the expressions provided for $\underline{g}^{\mathrm{C}}$ and $\overline{g}^{\mathrm{C}}$ produces the required results. □

The exceptional case in the previous lemma is accommodated as follows.

Proposition 14 *Consider the conditions of Theorem 6, with* $\phi : (\xi_1,\xi_2) \mapsto \xi_1\xi_2$*, and suppose that* $\min\{\underline{x}_1,\underline{x}_2\} < 0 < \max\{\overline{x}_1,\overline{x}_2\}$*. Define an interval*

$$\begin{aligned}\boldsymbol{\phi}(X) &:= [\min\{\phi(\xi) : \xi \in X\}, \max\{\phi(\xi) : \xi \in X\}]\\ &= [\min\{\underline{x}_1\underline{x}_2, \underline{x}_1\overline{x}_2, \overline{x}_1\underline{x}_2, \overline{x}_1\overline{x}_2\}, \max\{\underline{x}_1\underline{x}_2, \underline{x}_1\overline{x}_2, \overline{x}_1\underline{x}_2, \overline{x}_1\overline{x}_2\}] \in \mathbb{IR},\end{aligned}$$

and an interval $\boldsymbol{\phi}^{\mathrm{B}}(X) \in \mathbb{IR}$*, where, with the convention that* $-\boldsymbol{q} = [-\overline{q}, -\underline{q}]$ *for any* $\boldsymbol{q} \in \mathbb{IR}$*,*

$$\underline{\phi}^{\mathrm{B}}(X) = \underline{g}^{\mathrm{C}}_{\times,A}(\boldsymbol{x}_1,\boldsymbol{x}_2,\boldsymbol{x}_1,\boldsymbol{x}_2), \qquad \textit{and} \qquad \overline{\phi}^{\mathrm{B}}(X) = -\underline{g}^{\mathrm{C}}_{\times,A}(-\boldsymbol{x}_1,\boldsymbol{x}_2,-\boldsymbol{x}_1,\boldsymbol{x}_2).$$

Then $\operatorname{wid}\boldsymbol{\phi}^{\mathrm{B}}(X) - \operatorname{wid}\boldsymbol{\phi}(X) \le 2(\operatorname{wid}X)^2$.

Proof Suppose that Theorem 6 is applied with the substitutions $Z := X$, $f_i : z \in Z \mapsto z_i$, $\underline{f}_i^{\mathrm{C}} : z \mapsto \underline{x}_i$, and $\overline{f}_i^{\mathrm{C}} : z \mapsto \overline{x}_i$ for each $i \in \{1,2\}$. The constructed relaxations then satisfy $[\underline{g}^{\mathrm{C}}_{\times}(z), \overline{g}^{\mathrm{C}}_{\times}(z)] = \phi^{\mathrm{B}}(X)$ for each $z \in Z = X$; it follows from Theorem 6 that $\boldsymbol{\phi}(X) \subset \boldsymbol{\phi}^{\mathrm{B}}(X)$. Thus, it suffices to show that $\underline{\phi}(X) - \underline{\phi}^{\mathrm{B}}(X) \le (\operatorname{wid}X)^2$; a similar argument shows that $\overline{\phi}^{\mathrm{B}}(X) - \overline{\phi}(X) \le (\operatorname{wid}X)^2$, and these two bounds together imply the required result.

Now, as shown in the proof of Theorem 6,

$$\underline{\phi}^{\mathrm{B}}(X) = \min\{\underline{\psi}_{\times,\mathrm{A}}(\xi_1,\xi_2,\boldsymbol{x}_1,\boldsymbol{x}_2) : \xi_1 \in \boldsymbol{x}_1, \xi_2 \in \boldsymbol{x}_2\}.$$

Now, let $\underline{\phi}^{\mathrm{C}}$ denote the convex envelope of ϕ on X. Since the constant mapping $z \mapsto \underline{\phi}(X)$ is a convex underestimator of ϕ on X, it follows that $\underline{\phi}(X) = \min\{\underline{\phi}^{\mathrm{C}}(\xi_1,\xi_2) : \xi_1 \in \boldsymbol{x}_1, \xi_2 \in \boldsymbol{x}_2\}$. So, it suffices to show that $\underline{\phi}^{\mathrm{C}}(\xi_1,\xi_2) - \underline{\psi}_{\times,\mathrm{A}}(\xi_1,\xi_2,\boldsymbol{x}_1,\boldsymbol{x}_2) \le (\operatorname{wid}X)^2$ for each $\xi \in X$; this was accomplished previously in the proof of Proposition 11. □

6 Extensions to implicit functions and differential equations

One of the benefits of McCormick's relaxation approach is its use in constructing relaxations for implicit functions [60,68,58] and solutions of parametric ordinary differential equations (ODEs) [57,56]. This section shows briefly that Whitney-$\mathscr{C}^1$ relaxations of implicit functions and parametric ODE solutions can be obtained by extending the approach of this article analogously.

6.1 Differentiable relaxations of implicit functions

For illustration, this section constructs Whitney-$\mathscr{C}^1$ relaxations for implicit functions in the spirit of [60, Section 3.1], by applying the approach of Sections 3 and 4 to successive fixed-point iterations. The relaxations discussed in [60, Sections 3.2–3.4] may be treated similarly. We suppose that the following assumption is satisfied; notation has been altered from [60] for consistency with the previous sections.

Assumption 2 *Consider intervals $P \in \mathbb{IR}^{n_p}$ and $X \in \mathbb{IR}^{n_x}$, and a Whitney-$\mathscr{C}^1$ function $h : X \times P \to X$ that is a finite composition of intrinsic functions considered in Sections 3 and 4. Suppose that there exists exactly one function $x : P \to X$ (not known explicitly) for which $h(x(p), p) = x(p)$ for each $p \in P$. Define a function $f : P \mapsto X \times P : p \mapsto (x(p), p)$, and known functions $a, b : X \times P \to X$ for which, for each $i \in \{1, \ldots, n_x\}$,*

$$a_i : (\xi, p) \mapsto \xi_i + \max\{0, h_i(\xi, p) - \xi_i\},$$
$$b_i : (\xi, p) \mapsto \xi_i + \min\{0, h_i(\xi, p) - \xi_i\}.$$

Observe that $a(f(p)) = b(f(p)) = x(p)$ for each $p \in P$. Moreover, the approach of the previous sections produces Whitney-$\mathscr{C}^1$ relaxations for each component of a and b, along with each component of $a \circ f$ and $b \circ f$ whenever Whitney-$\mathscr{C}^1$ relaxations are available for f. Hence, the following variant of the method in [60, Section 3.1] may be carried out.

1. Choose vectors $\underline{x}, \overline{x} \in X$ for which $X = [\underline{x}_1, \overline{x}_1] \times \ldots \times [\underline{x}_{n_x}, \overline{x}_{n_x}]$. Define the following constant componentwise relaxations of x on P:

$$\underline{x}^{\mathrm{C}}_{(0)} : P \to X : p \mapsto \underline{x}, \qquad \text{and} \qquad \overline{x}^{\mathrm{C}}_{(0)} : P \to X : p \mapsto \overline{x}.$$

2. For $k = 1, 2, \ldots$, construct Whitney-$\mathscr{C}^1$ componentwise relaxations $\underline{x}^{\mathrm{C}}_{(k)}, \overline{x}^{\mathrm{C}}_{(k)}$ of x on P as follows:
   - Define Whitney-$\mathscr{C}^1$ componentwise relaxations $\underline{f}^{\mathrm{C}}_{(k-1)}, \overline{f}^{\mathrm{C}}_{(k-1)}$ of f on P for which:

$$\underline{f}^{\mathrm{C}}_{(k-1)} : p \mapsto (\underline{x}^{\mathrm{C}}_{(k-1)}(p), p), \qquad \text{and} \qquad \overline{f}^{\mathrm{C}}_{(k-1)} : p \mapsto (\overline{x}^{\mathrm{C}}_{(k-1)}(p), p).$$

   - Construct $\underline{x}^{\mathrm{C}}_{(k)}$ as a Whitney-$\mathscr{C}^1$ componentwise convex relaxation of $a \circ f$, with a handled using the approaches of Sections 3 and 4, and with $\underline{f}^{\mathrm{C}}_{(k-1)}$ and $\overline{f}^{\mathrm{C}}_{(k-1)}$ supplied as Whitney-$\mathscr{C}^1$ relaxations of f.
   - Similarly, construct $\overline{x}^{\mathrm{C}}_{(k)}$ as an analogous Whitney-$\mathscr{C}^1$ componentwise concave relaxation of $b \circ f$.

Observe that, for each $k \in \mathbb{N}$, the functions $\underline{x}^{\mathrm{C}}_{(k)}$ and $\overline{x}^{\mathrm{C}}_{(k)}$ are Whitney-$\mathscr{C}^1$ componentwise relaxations of the unknown implicit function x on P. Each may be evaluated at any $p \in P$ for roughly k times the computational cost of evaluating Whitney-$\mathscr{C}^1$ componentwise relaxations of h. Moreover, inspection of the relaxations of $\xi \mapsto \max\{0, \xi\}$ and $\xi \mapsto \min\{0, \xi\}$ in Section 3.3 and the definitions of a and b show that, for each $k \in \mathbb{N}$, $\underline{x}^{\mathrm{C}}_{(k)}$ dominates $\underline{x}^{\mathrm{C}}_{(k-1)}$ componentwise on P, and $\overline{x}^{\mathrm{C}}_{(k)}$ is dominated by $\overline{x}^{\mathrm{C}}_{(k-1)}$ componentwise on P. Thus, as k increases, $\underline{x}^{\mathrm{C}}_{(k)}$ and $\overline{x}^{\mathrm{C}}_{(k)}$ may become tighter componentwise relaxations of x, and cannot become looser.

6.2 Differentiable relaxations of solutions of ordinary differential equations

Consider intervals $X \in \mathbb{IR}^{n_x}$ and $P \in \mathbb{IR}^{n_p}$, and a system of parametric ordinary differential equations (ODEs):

$$\frac{dx}{dt}(t, p) = f(t, p, x(t, p)), \qquad x(0, p) = x_{(0)}(p), \tag{7}$$

with $f:\mathbb{R}\times P\times X\to\mathbb{R}^{n_x}$ and $x_{(0)}:P\to X$. Suppose that f and $x_{(0)}$ are finite compositions of the particular intrinsic functions ϕ considered in Sections 3 and 4, in which case f and $x_{(0)}$ are both locally Lipschitz continuous. The ODE system (7) then satisfies standard conditions [15] for the local existence and uniqueness of a solution $t\mapsto x(t,p)$. Suppose that, for some interval $P\in\mathbb{IR}^{n_p}$ and each $p\in P$, a solution $t\mapsto x(t,p)$ of (7) exists on a duration $0\le t\le t_f$; local uniqueness implies that $t\mapsto x(t,p)$ is the only such solution for each $p\in P$. Under these assumptions, Scott et al. [57,56] use componentwise closed-form relaxations of f and $x_{(0)}$ to construct auxiliary systems of differential equations that describe componentwise relaxations of the mapping $p\mapsto x(t,p)$ on P at any fixed $t\in[0,t_f]$. However, if McCormick's classical relaxation scheme is used to generate the relaxations supplied for f and $x_{(0)}$, then the generated relaxations may be nonsmooth. As a result, the auxiliary differential equations describing relaxations of $p\mapsto x(t,p)$ may have nonsmooth or discontinuous right-hand sides; this may hinder numerical integration methods that assume differentiability, and may complicate theoretical descriptions of the computed relaxations' subgradients. This section outlines how Whitney-$\mathscr{C}^1$ relaxations of f and $x_{(0)}$ may be employed to circumvent these issues in the method of [57].

Thus, consider an intermediate mapping $h:(t,p)\mapsto(t,p,x(t,p))$, and suppose that Whitney-$\mathscr{C}^1$ relaxations are generated for $f\circ h(t,\cdot)$ and $x_{(0)}$ by repeated application of Theorem 1 and the particular intrinsic function relaxations suggested in Sections 3 and 4, with supplied relaxations $\underline{x}^{\rm C}(t,\cdot)$ and $\overline{x}^{\rm C}(t,\cdot)$ of $x(t,\cdot)$ on P. This approach produces Whitney-$\mathscr{C}^1$ relaxations $\underline{x}^{\rm C}_{(0)}$ and $\overline{x}^{\rm C}_{(0)}$ of $x_{(0)}$ on P, along with Whitney-$\mathscr{C}^1$ functions $u,o:\mathbb{R}\times P\times\mathbb{R}^{n_x}\times\mathbb{R}^{n_x}$ for which $p\mapsto u(t,p,\underline{x}^{\rm C}(t,p),\overline{x}^{\rm C}(t,p))$ and $p\mapsto o(t,p,\underline{x}^{\rm C}(t,p),\overline{x}^{\rm C}(t,p))$ are convex and concave relaxations of $f\circ h(t,\cdot)$ on P, respectively. By the theoretical development in [57, Section 4], the auxiliary ODEs:

$$\begin{aligned}
\frac{da}{dt}(t,p) &= u(t,p,a(t,p),b(t,p)), & a(0,p) &= \underline{x}^{\rm C}_{(0)}(p),\\
\frac{db}{dt}(t,p) &= o(t,p,a(t,p),b(t,p)), & b(0,p) &= \overline{x}^{\rm C}_{(0)}(p),
\end{aligned}$$

have unique solutions $a(\cdot,p),b(\cdot,p)$ for each $p\in P$; moreover, for each fixed $t\in[0,t_f]$, the mappings $p\mapsto a(t,p)$ and $p\mapsto b(t,p)$ are convex and concave relaxations of $p\mapsto x(t,p)$ on P, respectively. Moreover, a classical result from ODE theory [24, Theorem V.3.1] shows that the mappings $p\mapsto a(t,p)$ and $p\mapsto b(t,p)$ are Whitney-$\mathscr{C}^1$ on P. Thus, the derivative mappings $A:(t,p)\mapsto\frac{\partial a}{\partial p}(t,p)$ and $B:(t,p)\mapsto\frac{\partial b}{\partial p}(t,p)$ are the unique solutions of the following auxiliary linear ODE system, with arguments of partial derivatives of u and o suppressed for brevity:

$$\begin{aligned}
\frac{dA}{dt}(t,p) &= \frac{\partial u}{\partial p}+\frac{\partial u}{\partial a}A(t,p)+\frac{\partial u}{\partial b}B(t,p), & A(0,p) &= D\underline{x}^{\rm C}_{(0)}(p),\\
\frac{dB}{dt}(t,p) &= \frac{\partial o}{\partial p}+\frac{\partial o}{\partial a}A(t,p)+\frac{\partial o}{\partial b}B(t,p), & B(0,p) &= D\overline{x}^{\rm C}_{(0)}(p).
\end{aligned}$$

This linear system can be integrated by established efficient ODE solvers with sensitivity evaluation capabilities [36,18], without any need for event detection or for dedicated subgradient propagation theory [14].

We note, briefly, that an analogous approach may be applied to the tighter relaxations of [56].

7 Implementation and examples

This section describes a C++ implementation of the Whitney-$\mathscr{C}^1$ relaxations developed in this article, based on `MC++` [13]. Computational complexity for the corresponding method is discussed briefly. The implementation is used to construct relaxations for several simple functions for illustration. Lastly, a nonsmooth nonconvex optimization problem in chemical engineering is considered as a case study. When applied to this problem, simple branch-and-bound solvers employing the Whitney-$\mathscr{C}^1$ relaxations are shown to perform comparably to the state-of-the-art solver BARON [51].

7.1 Implementation

The suggested Whitney-$\mathscr{C}^1$ relaxations for intrinsic functions ϕ in Sections 3 and 4 were implemented by modifying version 1.0 of `MC++` [13], a C++ header library that uses operator overloading to evaluate classical McCormick relaxations [38,40] and certain improved relaxations of Tsoukalas and Mitsos [63]. Ultimately, the modified implementation evaluates Whitney-$\mathscr{C}^1$ relaxations for finite compositions of these intrinsic functions. The exponent μ employed in several of our suggested intrinsic function relaxations is stored as a static member variable, and may be altered; we recommend setting μ to be 1 or 2. If $\mu = 2$, then the generated relaxations might be Whitney-$\mathscr{C}^2$ even if the corresponding conditions on $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in Theorems 6 and 7 are not met.

`MC++` effectively evaluates subgradients of relaxations using the standard forward mode of automatic differentiation (AD) [22,43]; our implementation modifies this capability to evaluate gradients of Whitney-$\mathscr{C}^1$ relaxations using the various gradient expressions in this article. In our implementation, the interval bounds propagated over bivariate products $(x,y) \mapsto xy$ by `MC++` were weakened in accordance with Proposition 14.

For large problems, the standard reverse mode of AD [22,43,6] may yield computationally cheaper gradients of Whitney-$\mathscr{C}^1$ relaxations; for simplicity, this was not pursued further. We also note that sparsity of a relaxed function's computational graph could be exploited during evaluation of gradients of relaxations via the forward AD mode; again, this was not pursued further.

7.2 Complexity analysis

Roughly, denote the computational cost of evaluating a composition $f : X \subset \mathbb{R}^n \to \mathbb{R}$ of intrinsic functions as "cost(f)". Observe that, when constructing Whitney-$\mathscr{C}^1$ relaxations for f, each intrinsic function is replaced with corresponding operations that propagate convex and concave relaxations according to Theorem 1. Thus, there exists $\gamma_c > 0$ for which the computational cost of evaluating a Whitney-$\mathscr{C}^1$ convex or concave relaxation of f is no greater than $\gamma_c \,\text{cost}(f)$. The parameter γ_c is independent of f, but depends on the library of UIFs considered.

Similarly, using standard complexity results for automatic differentiation [22], it follows that there exist similar library-dependent constants $\gamma_a, \gamma_t > 0$, satisfying the following claim. If the reverse mode of automatic differentiation is used to evaluate a gradient of such a relaxation, then the cost of doing so is bounded above by $\gamma_a \,\text{cost}(f)$; if the forward mode is used instead, then the cost of evaluating this gradient is bounded above by $n\gamma_t \,\text{cost}(f)$, where n denotes the domain dimension of f.

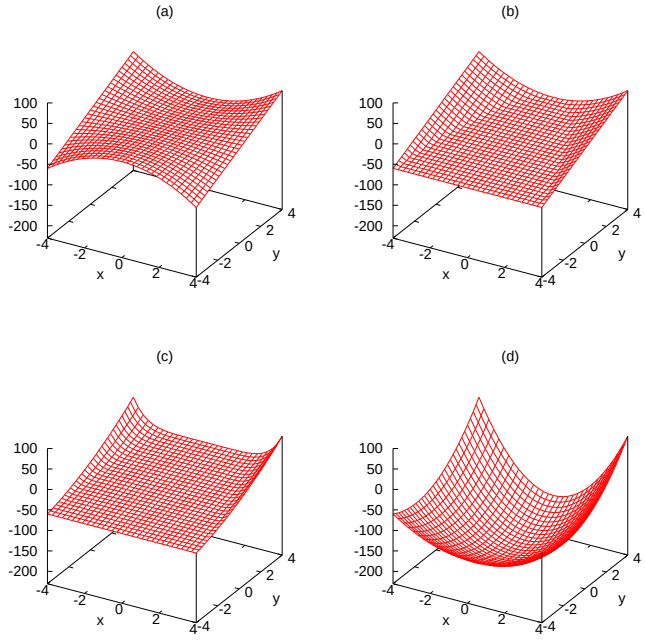

Fig. 1 The function $f : (x,y) \mapsto y(x^2-1)$ and its convex relaxations on $[-4,4]^2$: (a) the function f, (b) the multivariate McCormick relaxation of f, (c) a Whitney-$\mathscr{C}^1$ McCormick relaxation of f, and (d) the αBB relaxation of f that minimizes maximum separation distance.

7.3 Illustrations of differentiable relaxations

Example 2 To illustrate the modified multiplication rule provided by Theorem 6, consider the function $f : \mathbb{R}^2 \to \mathbb{R} : (x,y) \mapsto y(x^2-1)$, which is plotted in Figure 1(a). The function f is (real-)analytic but nonconvex on $Z := [-4,4]^2 \subset \mathbb{R}^2$.

A McCormick convex relaxation of f was constructed with `MC++` [13] on Z using the multivariate product relaxations suggested by Tsoukalas and Mitsos [63], and is plotted in Figure 1(b). This relaxation is not differentiable everywhere; this nondifferentiability is introduced by the rule in [63] for relaxing products of terms whose signs change on the interval of interest. A Whitney-$\mathscr{C}^1$ McCormick relaxation of f on Z was constructed using the approach of this article with $\mu := 2$; this relaxation is plotted in Figure 1(c). Observe that this relaxation is visibly differentiable on the interior of Z, but is otherwise qualitatively similar to the multivariate McCormick relaxation produced using [63]. The multivariate McCormick relaxation dominates its Whitney-$\mathscr{C}^1$ counterpart on Z.

For comparison, the αBB relaxation of f on Z with a nonuniform diagonal shift matrix that minimizes maximum separation distance [3] was computed directly to be:

$$f^{\alpha} : (x,y) \mapsto f(x,y) + 8(x^2-16) + 4(y^2-16),$$

and is plotted in Figure 1(d). The obtained αBB relaxation is analytic, and has a minimum at $(x^*,y^*) := (0, 0.125)$. Observe that $f^{\alpha}(x^*,y^*) = -192.0625$, which is less than the lower bound $\underline{\tilde{f}}(Z) = -60$ provided by the natural interval extension of f on Z. This interval lower

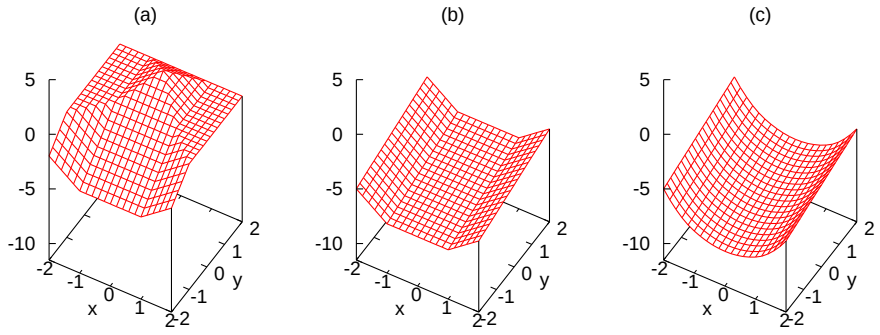

Fig. 2 The function g described in (8) and its convex relaxations on $[-2,2]^2$: (a) the function g, (b) the classical McCormick relaxation of g, and (c) a Whitney-$\mathscr{C}^2$ relaxation of g.

bound coincides with $\min_{(x,y)\in Z} f(x,y)$, and is dominated on Z by both the constructed classical McCormick relaxation and the constructed Whitney-$\mathscr{C}^1$ McCormick relaxation.

Example 3 To illustrate the handling of the absolute-value function according to Section 3.3, consider the function

$$g : \mathbb{R}^2 \to \mathbb{R} : (x,y) \mapsto |x+1| + |x-1| - |x+y-1| - |x-y+1|, \tag{8}$$

which is plotted in Figure 2(a). The function g is piecewise affine, and is nonconvex on $Z := [-2,2]^2 \subset \mathbb{R}^2$.

As in the previous example, the classical McCormick convex relaxation of g on Z was constructed using `MC++`, and is plotted in Figure 2(b); this relaxation is readily verified to be piecewise affine. A Whitney-$\mathscr{C}^2$ McCormick relaxation of g on Z was constructed using our implementation, and is plotted in Figure 2(c).

Example 4 This example illustrates the second-order pointwise convergence results of Section 5. As in [10, Example 7], consider the function

$$f : \mathbb{R}_+ \to \mathbb{R} : x \mapsto (x - x^2)(\log x + e^{-x})$$

on intervals of the form $[0.5-\varepsilon, 0.5+\varepsilon]$ for $0 < \varepsilon \leq 0.2$. The function f is plotted in Figure 3, together with a series of Whitney-$\mathscr{C}^1$ relaxations $\psi_{\boldsymbol{x}(\varepsilon)}$ of f constructed using our implementation, on intervals $\boldsymbol{x} \in \{[0.5-\varepsilon, 0.5+\varepsilon] : \varepsilon = 0.4(2^{-k}), k \in \{1,\ldots,20\}\}$.

For the considered values of ε, Figure 3(b) plots $\sup_{x\in\boldsymbol{x}(\varepsilon)}(f(x) - \psi_{\boldsymbol{x}(\varepsilon)}(x))$ against $\operatorname{wid}\boldsymbol{x}(\varepsilon)$ on a logarithmic scale; the slope of this plot suggests second-order pointwise convergence of the convex relaxation $\psi_{\boldsymbol{x}}(\varepsilon)$ to f as $\varepsilon \to 0^+$ according to the definitions in [10, 42].

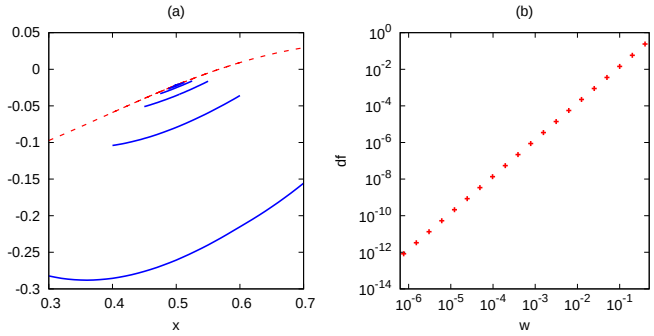

Fig. 3 (a) The function f described in Example 4 (dashed) and its Whitney-$\mathscr{C}^1$ convex relaxations $\psi_{\boldsymbol{x}(\varepsilon)}$ of f on intervals $\boldsymbol{x}(\varepsilon) := [0.5-\varepsilon, 0.5+\varepsilon]$ for various $\varepsilon > 0$ (solid), and (b) a plot of $df := \sup_{x\in\boldsymbol{x}(\varepsilon)}(f(x) - \psi_{\boldsymbol{x}(\varepsilon)}(x))$ vs. $w := \operatorname{wid}\boldsymbol{x}(\varepsilon) = 2\varepsilon$.

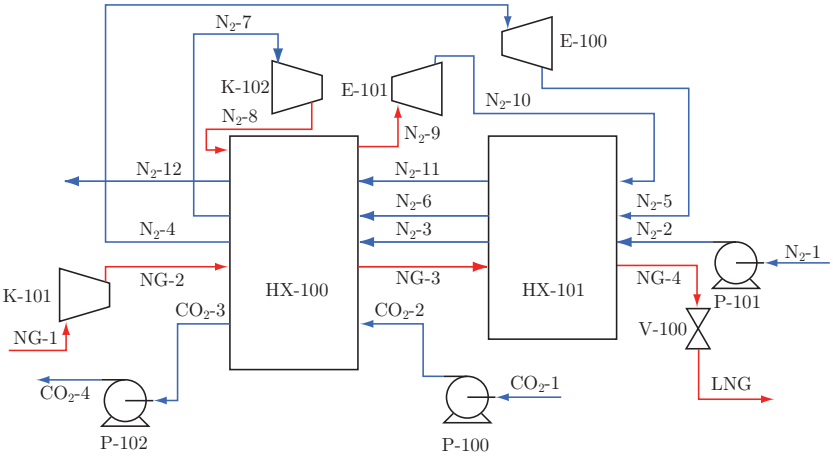

Fig. 4 Flowsheet for the LNG process in Example 5 (from [66])

7.4 Case study: Optimization of a process for offshore LNG production

Example 5 Consider the optimization of an offshore process concept for liquefied natural gas (LNG) production featuring compression and expansion of process streams which was previously studied in [66]. The flowsheet for the process is shown in Figure 4.

Preliminary design work, as described in [66], fixed the flowrates (F), temperature levels (T^{in} and T^{out}), and pressure levels (P^S), of the natural gas streams (NG-x) and carbon dioxide streams (CO_2-x) in the process. Some constraints on the temperature and pressure of the initial pass of the nitrogen stream (N_2-x) were also determined. In place of using physical property calculations and phase detection mechanisms in the simulation, several of the physical process streams are split into substreams of constant heat capacity (c_p), to approximate the real temperature-enthalpy relationship of the process. The natural gas stream is modeled as three separate hot streams (H1-H3), the carbon dioxide stream is modeled as two separate cold streams (C1, C2), and the nitrogen stream is modeled as three cold streams (C3-C5) on its first pass through the heat exchangers. This nitrogen stream is then expanded

through unit E-100 and supplied as an additional cold stream (C6). After this stream runs through both heat exchangers, it is then re-compressed in unit K-102 and passes through the exchangers again as hot stream H4, before being brought to ambient pressure through expander E-101. The ambient nitrogen stream is then used one final time as cold stream C7. Table 2 details the values of the fixed process parameters, as well as the unknown stream variables, which are the decision variables in the optimization problem.

Stream	F [kg/s]	c_p [kJ/kg]	T^{in} [K]	T^{out} [K]	P^s [MPa]
H1 (NG-2–NG-4)	1.00	3.46	319.80	265.15	10.0
H2 (NG-2–NG-4)	1.00	5.14	265.15	197.35	10.0
H3 (NG-2–NG-4)	1.00	3.51	197.35	104.75	10.0
H4 (N_2-8–N_2-9)	F_{N_2}	1.15	$T^{\text{in}}_{\text{H4}}$	$T^{\text{out}}_{\text{H4}}$	P^s_{H4}
C1 (CO_2-2–CO_2-3)	2.46	2.11	221.12	252.55	6.0
C2 (CO_2-2–CO_2-3)	2.46	2.48	252.55	293.15	6.0
C3 (N_2-2–N_2-4)	F_{N_2}	2.48	103.45	171.05	10.0
C4 (N_2-2–N_2-4)	F_{N_2}	1.80	171.05	218.75	10.0
C5 (N_2-2–N_2-4)	F_{N_2}	1.18	218.75	$T^{\text{out}}_{\text{C5}}$	10.0
C6 (N_2-5–N_2-7)	F_{N_2}	1.07	$T^{\text{in}}_{\text{C6}}$	$T^{\text{out}}_{\text{C6}}$	P^s_{C6}
C7 (N_2-10–N_2-12)	F_{N_2}	1.04	$T^{\text{in}}_{\text{C7}}$	$T^{\text{out}}_{\text{C7}}$	0.1

Table 2 Data and unknowns for the offshore LNG production process in Example 5.

In [66], this simultaneous flowsheet simulation and heat integration problem was modeled as an MINLP using the formulation from [23]. This previous work did not distinguish between the two physical heat exchangers in the flowsheet, and instead considered all streams as being part of a single heat integration problem. For consistency of results, this approach is taken here as well. However, in order to make better use of the multivariate relaxations developed in this work, the problem is instead modeled using the framework recently developed by [64], which can be written as follows:

$$Q_H + \sum_{i\in H} F c_{p,i}(T_i^{\text{in}} - T_i^{\text{out}}) = Q_C + \sum_{j\in C} F c_{p,j}(T_j^{\text{out}} - T_j^{\text{in}}), \tag{9}$$

$$\min_{p\in P}\{EBP_H^p - EBP_C^p\} = -Q_c, \tag{10}$$

where Q_H is the heating utility required by the process, Q_C is the cooling utility required by the process, H is the index set of hot streams, C is the index set of cold streams, $P = H \cup C$ is the index set of pinch point candidates, and:

$$\begin{aligned} EBP_H^p = \sum_{i\in H} F c_{p,i}[&\max\{0, T^p - T_i^{\text{out}}\} - \max\{0, T^p - T_i^{\text{in}}\} \\ &- \max\{0, T^{\min} - T^p\} + \max\{0, T^p - T^{\max}\}], \quad \forall p \in P, \end{aligned} \tag{11}$$

$$\begin{aligned} EBP_C^p = \sum_{j\in C} F c_{p,j}[&\max\{0, (T^p - \Delta T_{\min}) - T_j^{\text{in}}\} - \max\{0, (T^p - \Delta T_{\min}) - T_j^{\text{out}}\} \\ &+ \max\{0, (T^p - \Delta T_{\min}) - t^{\max}\} - \max\{0, t^{\min} - (T^p - \Delta T_{\min})\}], \quad \forall p \in P, \end{aligned} \tag{12}$$

$$T^p = \begin{cases} T_i^{\text{in}}, & \forall p = i \in H, \\ T_j^{\text{in}} + \Delta T_{\min}, & \forall p = j \in C, \end{cases} \tag{13}$$

where $\Delta T_{\min}$ is the minimum temperature difference between the hot and cold streams, $T^{\min}$ and $T^{\max}$ are the minimum and maximum hot stream temperatures, and $t^{\min}$ and $t^{\max}$ are the minimum and maximum cold stream temperatures. Note that since some of the decision variables are temperatures, $T^{\min/\max}$ and $t^{\min/\max}$ must be calculated by iteratively using the bivariate min (or max) function on the appropriate temperature set. The compression and expansion operations are modeled as polytropic processes for ideal gases with polytropic exponent $\kappa = 1.352$ as in [66]. The defining relationship for a polytropic process is as follows:

$$(\kappa - 1)\ln P_1^s + \kappa \ln T_2 = (\kappa - 1)\ln P_2^s + \kappa \ln T_1, \tag{14}$$

where T_1 and P_1^s are the temperature and pressure at the inlet and T_2 and P_2^s are the temperature and pressure at the outlet. The work (W) of such a process is given by:

$$W = F_1 c_{p,1} (T_2 - T_1), \tag{15}$$

where a positive value of work indicates work needs to be supplied to the process and a negative value indicates that the process generates work. From [66], compressor K-101 requires 58.69 kW of power, while pumps P-100, P-101 and P-102 use 15.32 kW, 17.61 kW/(kg/s) F_{N_2}, and 40.04 kW, respectively.

As in [66], the process is optimized subject to progressively more stringent sets of constraints that limit the amounts of external utilities and work which the process is allowed to consume. One hot and one cold utility are assumed to be available at 383.15 and 93.15 K, respectively. The objective function in all cases is to minimize the required nitrogen flowrate. The same cases are studied:

- Case I: minimize F_{N_2},
- Case II: minimize F_{N_2} such that $W \leq 0$,
- Case III: minimize F_{N_2} such that $Q_C = 0$ and $W \leq 0$,
- Case IV: minimize F_{N_2} such that $Q_C = Q_H = 0$ and $W \leq 0$.

In all cases, bounds on temperature variables were given by the utility temperatures. The flowrate of nitrogen was allowed to vary between 0.0 and 2.0 kg/s, the pressure of stream C6 was bounded between 0.3 and 1.0 MPa, and the pressure of stream H4 was constrained between 1.0 and 3.5 MPa.

All cases were first resolved in GAMS v24.5 using BARON v15.9 [51] with CPLEX and SNOPT as the LP and NLP solvers, respectively, on an Intel Xeon E5-1650 v2 workstation using six cores at 3.50 GHz and 12 GB RAM under Linux v14.04. Since BARON cannot directly model multivariate max and min functions, the MINLP formulation was used. The model consists of 143 equality constraints, 1101 inequality constraints, 363 binary variables and 173 continuous variables for Case IV, as an example. The relative termination tolerance in GAMS was set to 10^{-4}. The absolute termination tolerance in GAMS, as well as all feasibility tolerances and SNOPT or CPLEX tolerances, were left at their default values. In all cases, an optimal solution was found with objective value within the optimality tolerance of that reported in [66]. The first three rows of Table 3 summarize the computational results for each of the four cases. It is clear when comparing the present solution times to those reported in [66] that BARON has improved significantly on this problem in newer versions.

The four cases were then solved using a basic branch-and-bound method [27] implemented in C++ using the Whitney-$\mathscr{C}^1$ relaxations (with $\mu := 1$) detailed in this work to

construct Whitney-$\mathscr{C}^1$ convex relaxations which were minimized using SNOPT v7.2 [20] to provide lower bounds on the optimal solution. Using the nonsmooth modeling approach, Case IV requires 5 equality constraints, 4 inequality constraints, and 10 continuous variables: a significant reduction compared to the MINLP model. Upper bounds on the solution value were obtained by first finding an approximate solution to the nonsmooth problem with SNOPT in derivative-free mode, and then passing the SNOPT solution to MPBNGC v2.0, a FORTRAN implementation of the proximal bundle method for constrained nonsmooth problems [35]. The bundle method was allowed to take a maximum of five iterations to improve the upper bound before termination. This upper bounding strategy worked well in practice, as the global solution for each of the four cases was found very early. Optimality and feasibility tolerances were set identical to those from GAMS for fair comparison. Branching was performed such that the current box was bisected along the largest current width relative to the original box dimensions, and nodes were selected according to the lowest lower bound heuristic. An optimal solution with objective value within the optimality tolerance of that reported in [66] was found for all cases. The second three rows of Table 3 summarize the computational results for each of these numerical tests.

Noting that BARON was solving the problems more efficiently than the in-house software largely because of the use of range reduction techniques, these features of BARON were turned off completely. With this restriction, BARON was not able to solve any problem except Case I in less than 100 hours. For a better comparison, the cases were solved again, this time allowing BARON to only use dual multiplier-based bounds tightening (DBBT, or option `MDo = 1` in BARON) as described in [49] for range reduction (all preprocessing was also left active). DBBT was also implemented and used in the in-house C++ code using multiplier values calculated by SNOPT in the lower bounding procedure. The results from these experiments are shown in Table 3. Finally, the branch-and-bound code was also augmented with objective function value cuts in addition to DBBT. The cases were resolved in this framework and the results are also shown in Table 3. The CPU cost of each iteration averaged between Cases II, III, and IV for each method is shown in Table 4.

Even without employing range reduction, the Whitney-$\mathscr{C}^1$ relaxations provide tight lower bounds and reasonable solution times for each of the four cases, especially if compared with the original results from [66] or even the current version of BARON running only the standard branch-and-bound algorithm. Case II is somewhat of an exception, as the nonsmooth model appears to be more significantly affected by the degeneracy of the optimal solution than the MINLP method. With just DBBT enabled as a range reduction technique, the performance of the Whitney-$\mathscr{C}^1$ relaxations improves significantly and outperforms BARON running with only DBBT in all four cases. When the in-house code uses both DBBT and objective function value cuts, the solution statistics are very comparable to those of BARON with all features enabled on the most constrained case (IV) studied in this example. Additionally, as Table 4 shows, the average CPU cost per node in full-featured BARON is significantly higher than for any version of the in-house code. For more complicated heat integration problems (those which include real thermodynamic models), this cost could become prohibitive, owing to both the large model size and the dependence on costly range reduction techniques such as probing. Research is ongoing to determine if the Whitney-$\mathscr{C}^1$ relaxations and nonsmooth model presented here will provide an efficient alternative for solving such problems.

Overall, the present results indicate that the multivariate Whitney-$\mathscr{C}^1$ McCormick relaxations developed in this article have a practical application to which they are particularly well-suited and that they provide comparable performance to a state-of-the-art solver us-

Solution method	Statistic	Case I	Case II	Case III	Case IV
BARON v15.9 (all features)	Time (s)	0.52	2.04	15.68	31.91
	Iterations	2	6	102	1099
	Max nodes	2	3	22	74
Whitney-$\mathscr{C}^1$ relaxations (no range reduction)	Time (s)	0.0039	3511.05	831.90	363.96
	Iterations	1	339,207	146,665	59,597
	Max nodes	1	33,343	13,126	6,442
BARON v15.9 (DBBT only)	Time (s)	0.38	783.42	6169.10	5989.74
	Iterations	1	45,477	536,407	452,878
	Max nodes	1	2,551	23,287	15,499
Whitney-$\mathscr{C}^1$ relaxations (DBBT only)	Time (s)	0.0040	141.22	35.67	56.61
	Iterations	1	15,851	3,715	7,647
	Max nodes	1	3,561	431	771
Whitney-$\mathscr{C}^1$ relaxations (DBBT & obj. fun. cuts)	Time (s)	0.0040	93.99	27.52	35.91
	Iterations	1	11,229	1,621	2,095
	Max nodes	1	2,702	317	488

Table 3 Computational results for the LNG process case study.

Solution method	Average time per iteration (ms)
BARON v15.9 (all features)	202.42
Whitney-$\mathscr{C}^1$ relaxations (no range reduction)	6.39
BARON v15.9 (DBBT only)	13.99
Whitney-$\mathscr{C}^1$ relaxations (DBBT only)	8.63
Whitney-$\mathscr{C}^1$ relaxations (DBBT & obj. fun. cuts)	14.16

Table 4 Average time per branch-and-bound iteration for the methods tested in the LNG process case study.

ing only basic range reduction techniques and solution heuristics in an otherwise standard branch-and-bound algorithm.

8 Conclusions

This article has presented conditions under which the multivariate McCormick relaxations of Tsoukalas and Mitsos [63] are continuously differentiable in the sense of Whitney [69], and has provided a corresponding method to construct continuously differentiable relaxations for well-defined finite compositions of the univariate intrinsic functions listed in Table 1, bivariate sums and products, and the bivariate "max" and "min" functions. This method preserves the useful computational properties of McCormick's original method [38, 40], and has been implemented in C++ by modifying version 1.0 of `MC++` [13]. A case study shows that the developed relaxations perform comparably to BARON [51] when embedded in simple

branch-and-bound solvers, for a problem in which the subproblems solved by BARON at each node are large, and in which formulating the problem for BARON requires appending many additional variables and constraints. The developed relaxations were also extended to yield differentiable relaxations for implicit functions defined by fixed-point iterations in the vein of [60], and for solutions of parametric ordinary differential equations using the framework of [57].

Future work will involve using the developed implementation in further nonconvex optimization problems of practical interest. Extensions of the Whitney-$\mathscr{C}^1$ relaxations developed here to the reverse propagation of McCormick relaxations [68] will also be considered, to yield improved differentiable relaxations of implicit functions. Developing the requisite theory would require handling intersections of McCormick objects [68,54] appropriately. Implementation of the Whitney-$\mathscr{C}^1$ relaxations in methods [57,56] for relaxing solutions of ordinary differential equations will also be pursued. Open questions include whether it is possible to strengthen the results in Section 3 and 4 that guarantee continuous differentiability of the relaxations provided by Theorem 1.

Acknowledgements The authors would like to thank Achim Wechsung and Spencer Schaber for several helpful discussions.

Appendices

A Proofs of results

A.1 Proof of Proposition 3

This proof proceeds by showing that the requirements of [69, Theorem I] are met by f on C. Since the components of f may be considered separately, it suffices to consider the case in which $m = 1$. For each $x \in C$, assume that N_x is convex without loss of generality; if this is not true, then redefine N_x to be an open convex subset containing x. Since C is compact, there exists a finite subset $I \subset C$ for which $C \subset \bigcup_{x\in I} N_x$.

Suppose, to obtain a contradiction, that there exist $x,\xi \in I$ and $y \in C$ for which $y \in N_x \cap N_\xi$ but $\nabla\phi_x(y) \neq \nabla\phi_\xi(y)$. Since C is convex and has nonempty interior, either $y \in \operatorname{int}(C)$ or $y \in \operatorname{bd}(\operatorname{int}(C))$. In either case, there exists a sequence $\{z_{(i)}\}_{i\in\mathbb{N}}$ in the nonempty open set $\tilde{N} := N_x \cap N_\xi \cap \operatorname{int}(C)$ that converges to y. Since $\phi_x \equiv \phi_\xi \equiv f$ on $\tilde{N}$, $\nabla\phi_x(z_{(i)}) = \nabla\phi_\xi(z_{(i)})$ for each i. The continuity of $\nabla\phi_x$ and $\nabla\phi_\xi$ on $\tilde{N}$ then yields $\nabla\phi_x(y) = \nabla\phi_\xi(y)$, which contradicts the choices of x, ξ, and y. Thus, there exists a single continuous function $g : C \to \mathbb{R}$ for which, for each $x \in I$, $g \equiv \nabla\phi_x$ on $N_x \cap C$.

To show that f is Whitney-$\mathscr{C}^1$ on C, it suffices in light of [69, Theorem I] to show that, for each $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$ for which

$$\|f(y) - f(x) - \langle g(x), y - x\rangle\| < \varepsilon\|y - x\| \tag{16}$$

whenever $x,y \in C$ and $\|y - x\| < \delta_\varepsilon$. Thus, choose any $\varepsilon > 0$. Since g is continuous on the compact set C, g is uniformly continuous on C; there exists $\tilde{\delta}_\varepsilon > 0$ for which $\|g(y) - g(x)\| < \varepsilon$ whenever $x,y \in C$ satisfy $\|y - x\| < \tilde{\delta}_\varepsilon$. Now, consider any $x,y \in C$ with $\|y - x\| < \tilde{\delta}_\varepsilon$; the bound (16) will be shown to hold for x and y.

Define the line segment $L := \operatorname{conv}\{x,y\}$. Since L is compact and $L \subset C$, choose $J \subset I$ as a set for which $L \subset \bigcup_{\xi\in J} N_\xi$ but $L \not\subset (\bigcup_{\xi\in J} N_\xi)\backslash N_\eta$ for each $\eta \in J$. Using these constructions, choose $k \in \mathbb{N}$, $0 = \lambda_0 < \lambda_1 < \ldots < \lambda_k = 1$, and $\xi_{(1)},\ldots,\xi_{(k)} \in J$ for which:

- $x_{(0)} := x \in N_{\xi_{(1)}}$,
- $x_{(k)} := y \in N_{\xi_{(k)}}$, and
- $x_{(q)} := \lambda_q x + (1-\lambda_q)y \in N_{\xi_{(q+1)}} \cap N_{\xi_{(q)}} \cap L$, for each $q \in \{1,2,\ldots,k-1\}$.

Observe that, for each $q \in \{1,\ldots,k\}$, $x_{(q-1)} \in N_{\xi_{(q)}}$ and $x_{(q)} \in N_{\xi(q)}$. So, the mean-value theorem and the established properties of g yield the following, for some $y_{(q)} \in \operatorname{conv}\{x_{(q-1)},x_{(q)}\} \subset L$:

$$f(x_{(q)}) - f(x_{(q-1)}) = \langle \nabla\phi_{\xi_{(q)}}(y_{(q)}), x_{(q)} - x_{(q-1)}\rangle = \langle g(y_{(q)}), x_{(q)} - x_{(q-1)}\rangle.$$

Hence,

$$\begin{aligned}
&\|f(y)-f(x)-\langle g(x),y-x\rangle\| \\
&\quad=\left\|\sum_{q=1}^{k}\left(f(x_{(q)})-f(x_{(q-1)})-\langle g(x),x_{(q)}-x_{(q-1)}\rangle\right)\right\| \\
&\quad=\left\|\sum_{q=1}^{k}\langle g(y_{(q)})-g(x),x_{(q)}-x_{(q-1)}\rangle\right\| \\
&\quad\leq\sum_{q=1}^{k}\left\|g(y_{(q)})-g(x)\right\|\left\|x_{(q)}-x_{(q-1)}\right\| \\
&\quad\leq\sum_{q=1}^{k}\varepsilon\left\|x_{(q)}-x_{(q-1)}\right\|=\varepsilon\|y-x\|,
\end{aligned}$$

as required; the final equation above follows from the definitions of each $x_{(q)}$ and the inequality chain $0=\lambda_0<\lambda_1<\ldots<\lambda_k=1$. □

A.2 Proof of Proposition 4

This proof employs the following intermediate result.

Lemma 1 *Consider an interval $\boldsymbol{x}\in\mathbb{IR}$, a Lipschitz continuous function $f:\boldsymbol{x}\to\mathbb{R}$, and the convex envelope $\underline{f}^{\mathrm{C}}:\boldsymbol{x}\to\mathbb{R}$ of f on $\boldsymbol{x}$. Then, $\underline{f}^{\mathrm{C}}(\underline{x})=f(\underline{x})$ and $\underline{f}^{\mathrm{C}}(\overline{x})=f(\overline{x})$. Moreover, $\underline{f}^{\mathrm{C}}$ is Lipschitz continuous on $\boldsymbol{x}$, with the same Lipschitz constant as f. Analogous results hold for the concave envelope $\overline{f}^{\mathrm{C}}:\boldsymbol{x}\to\mathbb{R}$ of f on $\boldsymbol{x}$.*

Proof Only the convex envelope of f will be considered here; a similar argument addresses the concave envelope. The required results are trivial if $\underline{x}=\overline{x}$, so assume that $\underline{x}<\overline{x}$. Let k_f denote a Lipschitz constant for f on $\boldsymbol{x}$. Applying the definition of the convex envelope,

$$f(y)\geq\underline{f}^{\mathrm{C}}(y)\geq f(\underline{x})-k_f(y-\underline{x}),\qquad\forall y\in\boldsymbol{x};\tag{17}$$

the first inequality above is due to f dominating $\underline{f}^{\mathrm{C}}$, and the second inequality is due to $\underline{f}^{\mathrm{C}}$ dominating each convex underestimator of f on $\boldsymbol{x}$. Setting y to $\underline{x}$ in the above inequality chain yields $\underline{f}^{\mathrm{C}}(\underline{x})=f(\underline{x})$.

A similar argument yields:

$$f(y)\geq\underline{f}^{\mathrm{C}}(y)\geq f(\overline{x})+k_f(y-\overline{x}),\qquad\forall y\in\boldsymbol{x};\tag{18}$$

setting y to $\overline{x}$ yields $\underline{f}^{\mathrm{C}}(\overline{x})=f(\overline{x})$.

Thus, (17) and (18) become:

$$\begin{aligned}
&\underline{f}^{\mathrm{C}}(y)-\underline{f}^{\mathrm{C}}(\underline{x})\geq-k_f(y-\underline{x}),\qquad\forall y\in\boldsymbol{x},\\
&\underline{f}^{\mathrm{C}}(y)-\underline{f}^{\mathrm{C}}(\overline{x})\geq k_f(y-\overline{x}),\qquad\forall y\in\boldsymbol{x}.
\end{aligned}$$

Defining $D_+\underline{f}^{\mathrm{C}}$ and $D_-\underline{f}^{\mathrm{C}}$ as the right-derivative and left-derivative of $\underline{f}^{\mathrm{C}}$ described in [25, Theorem I.4.1.1], it follows from [25, Proposition I.4.1.3] that $D_+\underline{f}^{\mathrm{C}}(\underline{x})$ and $D_-\underline{f}^{\mathrm{C}}(\overline{x})$ both exist, are finite, and satisfy $D_+\underline{f}^{\mathrm{C}}(\underline{x})\geq-k_f$, and $D_-\underline{f}^{\mathrm{C}}(\overline{x})\leq k_f$. Thus, $\underline{f}^{\mathrm{C}}$ is continuous at $\underline{x}$ and $\overline{x}$. Moreover, [25, Theorem I.4.2.1] implies that for each $y\in\mathrm{int}(\boldsymbol{x})$, each subgradient of $\underline{f}^{\mathrm{C}}$ at y is an element of $[-k_f,k_f]$. This result, combined with the mean-value theorem [25, Theorem I.4.2.4], shows that $\underline{f}^{\mathrm{C}}$ is Lipschitz continuous on $\boldsymbol{x}$, with a Lipschitz constant of k_f. □

Using the above lemma, Proposition 4 may be proved as follows. Only the convex envelope of f will be considered here; a similar argument addresses the concave envelope. The required result is trivial if $\underline{x}=\overline{x}$, so assume that $\underline{x}<\overline{x}$. Theorem 3.2 in [21] implies that $\underline{f}^{\mathrm{C}}$ is $\mathscr{C}^1$ on $\mathrm{int}(\boldsymbol{x})$; it remains to be shown that $\underline{f}^{\mathrm{C}}$ is also $\mathscr{C}^1$ at $\underline{x}$ and $\overline{x}$. Noting that f is Lipschitz continuous on $\boldsymbol{x}$, construct the right-derivative $D_+\underline{f}^{\mathrm{C}}$ and the

left-derivative $D_-\underline{f}^{\mathrm{C}}$ as in the proof of Lemma 1. As in the proof of Lemma 1, $D_+\underline{f}^{\mathrm{C}}(\underline{x})$ and $D_-\underline{f}^{\mathrm{C}}(\overline{x})$ each exist and are finite. Define the following function, which extends the domain of $\underline{f}^{\mathrm{C}}$ to $\mathbb{R}$:

$$\psi : \mathbb{R} \to \mathbb{R} : y \mapsto \begin{cases} \underline{f}^{\mathrm{C}}(\underline{x}) + (D_+\underline{f}^{\mathrm{C}}(\underline{x}))(y-\underline{x}), & \text{if } y < \underline{x}, \\ \underline{f}^{\mathrm{C}}(y), & \text{if } y \in \boldsymbol{x}, \\ \underline{f}^{\mathrm{C}}(\overline{x}) + (D_-\underline{f}^{\mathrm{C}}(\overline{x}))(y-\overline{x}), & \text{if } \overline{x} < y. \end{cases}$$

The function ψ is evidently continuous, and is $\mathscr{C}^1$ at each $y \in \mathbb{R}\backslash\{\underline{x},\overline{x}\}$. Applying the definitions of $D_+\underline{f}^{\mathrm{C}}$ and $D_-\underline{f}^{\mathrm{C}}$, it follows that ψ is differentiable at $\underline{x}$ and $\overline{x}$ as well; thus,

$$\nabla\psi(y) = \begin{cases} D_+\underline{f}^{\mathrm{C}}(\underline{x}), & \text{if } y \leq \underline{x}, \\ \nabla\underline{f}^{\mathrm{C}}(y), & \text{if } y \in \mathrm{int}(\boldsymbol{x}), \\ D_-\underline{f}^{\mathrm{C}}(\overline{x}), & \text{if } \overline{x} \leq y. \end{cases}$$

This equation, together with [25, Theorem I.4.2.1(iii)], shows that ψ is $\mathscr{C}^1$ even at $\underline{x}$ and $\overline{x}$, and is therefore $\mathscr{C}^1$ on $\mathbb{R}$. Hence, $\underline{u}^{\mathrm{C}}$ is Whitney-$\mathscr{C}^1$ on $\boldsymbol{x}$. □

A.3 Proof of Theorem 4

The proof of Theorem 4 uses several intermediate results concerning a generic optimal-value function described by the following assumption.

Assumption 3 *For some $m \in \mathbb{N}$, consider a convex open set $X \subset \mathbb{R}^m$ and a convex set $C \subset X$ with nonempty interior. Define sets:*

$$K := \{(\ell,u) : \ell \in C,\ u \in C,\ \ell \leq u\} \subset \mathbb{R}^{2m},$$
$$\text{and} \qquad H := \{(\ell,u) : \ell \in X,\ u \in X,\ \ell \leq u\} \subset \mathbb{R}^{2m}.$$

Consider a convex $\mathscr{C}^1$ function $\psi : X \to \mathbb{R}$, and an optimal-value function:

$$\gamma : H \to \mathbb{R} : (\ell,u) \mapsto \min\{\psi(\xi) : \ell \leq \xi \leq u\}.$$

Lemma 2 *Suppose that Assumption 3 holds with $m = 2$. Define a function*

$$\omega : H \times X \to \mathbb{R} : ((a,b),c) \mapsto \gamma((c,a),(c,b)).$$

The function ω is convex and $\mathscr{C}^1$ on $\mathrm{int}(H) \times X$.

Proof For each $((a,b),c) \in H \times X$, observe that $\omega((a,b),c) = \min\{\psi(\xi) : \xi_1 = c,\ a \leq \xi_2 \leq b\}$. Hence, according to [48, Section 29], ω is convex. It then suffices by [48, Corollary 25.5.1] to show that ω is differentiable at some arbitrary $((\hat{a},\hat{b}),\hat{c}) \in \mathrm{int}(H) \times X$. Let ($\mathrm{CP}_\omega$) denote the convex program $\min\{\psi(\xi) : \xi_1 = \hat{c},\ \hat{a} \leq \xi_2 \leq \hat{b}\}$. Weierstrass's Theorem guarantees the existence of an optimal solution of (CP_ω); thus, choose a particular solution $\xi^* \in C$. By [48, Theorem 28.3], there exists a Karush-Kuhn-Tucker (KKT) vector $(\lambda,\mu) \in \mathbb{R} \times \mathbb{R}^2$ satisfying the following KKT conditions (among others) for all solutions η^* of (CP_ω) simultaneously:

$$0 = \nabla\psi(\eta^*) + \lambda e_{(1)} + (\mu_2 - \mu_1)e_{(2)},$$
$$\lambda \in \mathbb{R}, \qquad \mu_1 \geq 0, \qquad \mu_2 \geq 0, \qquad \mu_1(\hat{a} - \eta_1^*) = 0, \qquad \mu_2(\eta_2^* - \hat{b}) = 0.$$

Moreover, since $\hat{a} < \hat{b}$, any such vector (λ,μ) is unique; when $\eta^* := \xi^*$, the above KKT conditions imply:

$$\lambda = -\frac{\partial\psi}{\partial\xi_1}(\xi^*), \quad \mu_1 = \begin{cases} \frac{\partial\psi}{\partial\xi_2}(\xi^*), & \text{if } \xi_2^* = \hat{a}, \\ 0, & \text{if } \xi_2^* \neq \hat{a}, \end{cases} \quad \text{and} \quad \mu_2 = \begin{cases} -\frac{\partial\psi}{\partial\xi_2}(\xi^*), & \text{if } \xi_2^* = \hat{b}, \\ 0, & \text{if } \xi_2^* \neq \hat{b}. \end{cases}$$

According to [48, Corollary 29.1.3], this uniqueness shows that ω is differentiable at $((\hat{a},\hat{b}),\hat{c})$, as required.

Lemma 3 *Suppose that Assumption 3 holds, $m = 2$, C is compact, and there exists a vector $d \in \mathbb{R}^2$ such that, for all $\xi \in X$, $\langle \nabla\psi(\xi), d\rangle > 0$. The function γ is Whitney-$\mathscr{C}^1$ on the compact set K.*

Proof Without loss of generality, suppose that $d \geq 0$; the other cases are handled similarly. Observe that K has nonempty interior under Assumption 3; to see this, choose any $x \in \text{int}(C)$ and let $e \in \mathbb{R}^m$ be a vector whose components are all equal to unity. For sufficiently small $\tau > 0$, $(x, x+\tau e) \in \text{int}(K)$, as required.

By Proposition 3, it then suffices to show that, for some arbitrary $(\ell, u) \in K$, there exists a neighborhood $N \in \mathbb{R}^2 \times \mathbb{R}^2$ of (ℓ, u) and a $\mathscr{C}^1$ function $\phi : N \to \mathbb{R}$ for which $\phi \equiv \gamma$ on $N \cap K$. Now,

$$0 < \langle \nabla\psi(\ell), d\rangle = d_1 \frac{\partial \psi}{\partial x_1}(\ell) + d_2 \frac{\partial \psi}{\partial x_2}(\ell).$$

Since $d \geq 0$, the above inequality implies that $\frac{\partial \psi}{\partial x_i}(\ell) > 0$ for some $i \in \{1,2\}$. Suppose that $\frac{\partial \psi}{\partial x_1}(\ell) > 0$; the case in which $\frac{\partial \psi}{\partial x_2}(\ell) > 0$ is handled similarly. Since $\nabla\psi$ is continuous on X, there exists a neighborhood $N_\ell \subset X$ of ℓ for which $\frac{\partial \psi}{\partial x_1}(a) > 0$ for each $a \in N_\ell$. Since N_ℓ is open and $C \subset X$ is compact, choose $\delta > 0$ for which:

- $y \in X$ whenever $x \in C$ and $\|y - x\| < 3\delta$, and
- $a \in N_\ell$ whenever $\|a - \ell\| < 3\delta$.

Define a neighborhood

$$N_{(\ell,u)} := \{(a,b) \in X^2 : \|(a,b) - (\ell,u)\| < \delta\},$$

and a function:

$$\begin{aligned}\phi : N_{(\ell,u)} \to \mathbb{R} : (a,b) \mapsto \gamma((a_1,a_2),(a_1,u_2+2\delta)) \\ + \gamma((a_1,\ell_2-2\delta),(a_1,b_2)) - \gamma((a_1,\ell_2-2\delta),(a_1,u_2+2\delta)).\end{aligned}$$

Since $\ell, u \in C$ and $\ell \leq u$, if $(a,b) \in N_{(\ell,u)}$, then $a_2 < u_2 + 2\delta$ and $\ell_2 - 2\delta < b_2$. Thus, ϕ is indeed well-defined. Lemma 2 shows that each "γ" term in the definition of ϕ is $\mathscr{C}^1$ with respect to (a,b), which in turn shows that ϕ is $\mathscr{C}^1$ on $N_{(\ell,u)}$.

To complete this proof, it will be shown that $\phi \equiv \gamma$ on $N_{(\ell,u)} \cap K$. Consider some arbitrary point $(a,b) \in N_{(\ell,u)} \cap K$, and let ($\text{CP}_\gamma$) denote the convex program:

$$\min\{\psi(\xi) : a \leq \xi \leq b\}.$$

First, it will be shown that the set $\{\xi \in [a,b] : \xi_1 = a_1 \text{ or } \xi_2 = a_2\}$ contains all solutions of (CP_γ). To obtain a contradiction, suppose that this is not so, and choose some solution η^* of (CP_γ) accordingly for which $a_1 < \eta_1^* \leq b_1$ and $a_2 < \eta_2^* \leq b_2$. Since $d \geq 0$, there exists $\tau > 0$ solving the linear program:

$$\max\{\tau \geq 0 : \eta_1^* - \tau d_1 \geq a_1, \quad \eta_2^* - \tau d_2 \geq a_2\}.$$

Thus, with $\zeta := \eta^* - \tau d$, $\zeta \in [a,b]$ and either $\zeta_1 = a_1$ or $\zeta_2 = a_2$. By the mean-value theorem, there exists $s \in [0,\tau]$ for which

$$\psi(\zeta) = \psi(\eta^* - \tau d) = \psi(\eta^*) - \tau\langle \nabla\psi(\eta^* - sd), d\rangle < \psi(\eta^*),$$

which contradicts the definition of η^*.

Thus, all solutions of (CP_γ) lie in the set $\{\xi \in [a,b] : \xi_1 = a_1 \text{ or } \xi_2 = a_2\}$. Now, since the mapping $t \mapsto \psi(a + te_{(1)})$ is convex on $[0, b_1 - a_1]$, the mapping $t \mapsto \frac{\partial \psi}{\partial x_1}(a + te_{(1)})$ is increasing on $[0, b_1 - a_1]$. This implies:

$$0 \leq t\frac{\partial \psi}{\partial x_1}(a) \leq \int_0^t \frac{\partial \psi}{\partial x_1}(a + se_{(1)})\, ds = \psi(a + te_{(1)}) - \psi(a), \qquad \forall t \in [0, b_1 - a_1].$$

Hence, there exists a solution of (CP_γ) in the set $\{\xi \in [a,b] : \xi_1 = a_1\}$, which implies that

$$\gamma(a,b) = \gamma((a_1,a_2),(a_1,b_2)).$$

The convexity of ψ then yields:

$$\gamma(a,b) = \max\{\gamma((a_1,a_2),(a_1,u_2+2\delta)),\ \gamma((a_1,\ell_2-2\delta),(a_1,b_2))\}; \tag{19}$$

to see this, assume, to obtain a contradiction, that (19) does not hold. In this case, since $(a,b) \in N_{(\ell,u)}$ implies that $b_2 < u_2 + 2\delta$ and $\ell_2 - 2\delta < a_2$, the definition of γ implies that the left-hand side of (19) is strictly greater than the right-hand side. Thus, there exist $\underline{\beta} \in [\ell_2 - 2\delta, b_2]$ and $\overline{\beta} \in [a_2, u_2 + 2\delta]$ for which

$$\begin{aligned} & \psi(a_1, \underline{\beta}) < \gamma((a_1,a_2),(a_1,b_2)) = \gamma(a,b) \\ \text{and} \qquad & \psi(a_1, \overline{\beta}) < \gamma((a_1,a_2),(a_1,b_2)) = \gamma(a,b). \end{aligned}$$

The above inequalities and the definition of γ imply that neither $\underline{\beta}$ nor $\overline{\beta}$ are contained in the interval $[a_2, b_2]$; thus, $\underline{\beta} \in [\ell_2 - 2\delta, a_2]$ and $\overline{\beta} \in [b_2, u_2 + 2\delta]$. Since $[a_2, b_2] \subset [\underline{\beta}, \overline{\beta}]$, the convexity of ψ would then imply that

$$\psi(a_1, \tfrac{1}{2}(a_2 + b_2)) < \gamma((a_1,a_2),(a_1,b_2)),$$

which is contradicted by the definition of γ.

Lastly, the following equation follows from the definition of γ:

$$\gamma((a_1, \ell_2 - 2\delta),(a_1, u_2 + 2\delta)) = \min\{\gamma((a_1,a_2),(a_1,u_2+2\delta)),\ \gamma((a_1,\ell_2-2\delta),(a_1,b_2))\}. \tag{20}$$

Along with the definition of ϕ, (19) and (20) show that $\phi(a,b) = \gamma(a,b)$, as required. □

Lemma 4 *Suppose that Assumption 3 holds, $m = 2$, $X = \mathbb{R}^2$, C is compact, and there exists a nonzero vector $d \in \mathbb{R}^2$ such that, for all $\xi \in \mathbb{R}^2$, $\langle \nabla\psi(\xi), d\rangle = 0$. The function γ is Whitney-$\mathscr{C}^1$ on the compact set K.*

Proof Since $d \neq 0$, either $d_1 \neq 0$, $d_2 \neq 0$, or both. Thus, without loss of generality, suppose that $d_1 \geq 0$ and $d_2 > 0$; the other cases are handled similarly. Let π denote the linear transformation $\xi \in \mathbb{R}^2 \mapsto (\xi_1 - (\frac{d_1}{d_2})\xi_2, 0) \in \mathbb{R}^2$, and observe that $\pi(\xi) = \xi - (\frac{\xi_2}{d_2})d$ for each $\xi \in \mathbb{R}^2$. Thus,

$$\psi(\pi(\xi)) = \psi(\xi) - \int_0^{\frac{\xi_2}{d_2}} \langle \nabla\psi(\xi - sd), d\rangle \, \mathrm{d}s = \psi(\xi), \qquad \forall \xi \in \mathbb{R}^2;$$

which implies that, for each $(\ell,u) \in K$,

$$\gamma(\ell,u) = \min\{\psi(\eta) : \eta = \pi(\xi),\ \ell \leq \xi \leq u\}.$$

Since the transformation π is linear and $[\ell,u]$ is convex, the set $\{\pi(\xi) : \ell \leq \xi \leq u\}$ is the convex hull of $\{\pi(\ell_1,\ell_2), \pi(\ell_1,u_2), \pi(u_1,\ell_2), \pi(u_1,u_2)\}$. Since $d \geq 0$, this set is readily evaluated to be:

$$\{\pi(\xi) : \ell \leq \xi \leq u\} = \{(\eta_1, 0) \in \mathbb{R}^2 : \ell_1 - (\tfrac{d_1}{d_2})u_2 \leq \eta_1 \leq u_1 - (\tfrac{d_1}{d_2})\ell_2\}.$$

The following reformulation of γ is thus obtained:

$$\gamma(\ell,u) = \min\{\psi(\eta_1, 0) : \ell_1 - (\tfrac{d_1}{d_2})u_2 \leq \eta_1 \leq u_1 - (\tfrac{d_1}{d_2})\ell_2\}, \qquad \forall (\ell,u) \in H.$$

Since the univariate mapping $\eta_1 \in \mathbb{R} \mapsto \psi(\eta_1, 0)$ is convex and $\mathscr{C}^1$, Theorem 2 implies that γ is Whitney-$\mathscr{C}^1$ on K. □

Theorem 4 is then proved as follows. The claim of the theorem regarding $\underline{g}^{\mathrm{C}}$ will be demonstrated; a similar argument yields the claim regarding $\overline{g}^{\mathrm{C}}$. Define a compact convex set

$$B := \{(\ell,u) : \ell \in X, u \in X, \ell \leq u\} \subset \mathbb{R}^2 \times \mathbb{R}^2.$$

Lemmata 3 and 4 show that the function: $\gamma : B \to \mathbb{R} : (\ell,u) \mapsto \min\left\{\underline{\phi}^{\mathrm{C}}(\xi) : \ell \leq \xi \leq u\right\}$ is Whitney-$\mathscr{C}^1$. Observing that $\underline{g}^{\mathrm{C}}$ is equivalent to the mapping

$$z \mapsto \gamma\left(\underline{f}_1^{\mathrm{C}}(z), \ldots, \underline{f}_m^{\mathrm{C}}(z), \overline{f}_1^{\mathrm{C}}(z), \ldots, \overline{f}_m^{\mathrm{C}}(z)\right)$$

on Z, the chain rule for Whitney-$\mathscr{C}^1$ functions applies to this representation of $\underline{g}^{\mathrm{C}}$. This yields the required result. □

B Gradients for suggested multivariate relaxations

The following two propositions present gradients for the Whitney-$\mathscr{C}^1$ relaxations provided in Theorems 6 and 7. In each case, the provided gradients may be computed directly using the chain rule for Whitney-$\mathscr{C}^1$ functions.

Proposition 15 *Suppose that the conditions of Theorem 6 hold. Gradients for the relaxations $\underline{g}^{\mathrm{C}}_{\times}$ and $\overline{g}^{\mathrm{C}}_{\times}$ are as follows, at any $z \in Z$. Arguments of partial derivatives are suppressed here, and are the same as the analogous function arguments in Theorem 6. The partial derivatives of $\underline{\psi}_{\times,\mathrm{A}}$ may be computed at any argument corresponding to the "min" in the definition of $\underline{g}^{\mathrm{C}}_{\times,\mathrm{A}}$; this follows from a gradient invariance property of Mangasarian [37].*

$$\nabla \underline{g}^{\mathrm{C}}_{\times}(z) = \begin{cases} \frac{\partial \underline{\psi}_{\times,\mathrm{B}}}{\partial x} \nabla \underline{f}^{\mathrm{C}}_1(z) + \frac{\partial \underline{\psi}_{\times,\mathrm{B}}}{\partial y} \nabla \underline{f}^{\mathrm{C}}_2(z), & \text{if both } 0 \le \underline{x}_1 \text{ and } 0 \le \underline{x}_2, \\ -\frac{\partial \underline{\psi}_{\times,\mathrm{B}}}{\partial x} \nabla \overline{f}^{\mathrm{C}}_1(z) - \frac{\partial \underline{\psi}_{\times,\mathrm{B}}}{\partial y} \nabla \overline{f}^{\mathrm{C}}_2(z), & \text{if both } \overline{x}_1 \le 0 \text{ and } \overline{x}_2 \le 0, \\ \max\left\{0, \frac{\partial \underline{\psi}_{\times,\mathrm{A}}}{\partial x}\right\} \nabla \underline{f}^{\mathrm{C}}_1(z) + \min\left\{0, \frac{\partial \underline{\psi}_{\times,\mathrm{A}}}{\partial x}\right\} \nabla \overline{f}^{\mathrm{C}}_1(z) & \\ \quad + \max\left\{0, \frac{\partial \underline{\psi}_{\times,\mathrm{A}}}{\partial y}\right\} \nabla \underline{f}^{\mathrm{C}}_2(z) + \min\left\{0, \frac{\partial \underline{\psi}_{\times,\mathrm{A}}}{\partial y}\right\} \nabla \overline{f}^{\mathrm{C}}_2(z), & \text{otherwise,} \end{cases}$$

$$\nabla \overline{g}^{\mathrm{C}}_{\times}(z) = \begin{cases} \frac{\partial \underline{\psi}_{\times,\mathrm{B}}}{\partial x} \nabla \overline{f}^{\mathrm{C}}_1(z) - \frac{\partial \underline{\psi}_{\times,\mathrm{B}}}{\partial y} \nabla \underline{f}^{\mathrm{C}}_2(z), & \text{if both } \overline{x}_1 \le 0 \text{ and } 0 \le \underline{x}_2, \\ -\frac{\partial \underline{\psi}_{\times,\mathrm{B}}}{\partial x} \nabla \underline{f}^{\mathrm{C}}_1(z) + \frac{\partial \underline{\psi}_{\times,\mathrm{B}}}{\partial y} \nabla \overline{f}^{\mathrm{C}}_2(z), & \text{if both } 0 \le \underline{x}_1 \text{ and } \overline{x}_2 \le 0, \\ \max\left\{0, \frac{\partial \underline{\psi}_{\times,\mathrm{A}}}{\partial x}\right\} \nabla \overline{f}^{\mathrm{C}}_1(z) + \min\left\{0, \frac{\partial \underline{\psi}_{\times,\mathrm{A}}}{\partial x}\right\} \nabla \underline{f}^{\mathrm{C}}_1(z) & \\ \quad - \max\left\{0, \frac{\partial \underline{\psi}_{\times,\mathrm{A}}}{\partial y}\right\} \nabla \overline{f}^{\mathrm{C}}_2(z) - \min\left\{0, \frac{\partial \underline{\psi}_{\times,\mathrm{A}}}{\partial y}\right\} \nabla \underline{f}^{\mathrm{C}}_2(z), & \text{otherwise,} \end{cases}$$

where the required partial derivatives of $\underline{\psi}_{\times,\mathrm{A}}$ and $\underline{\psi}_{\times,\mathrm{B}}$ are as follows.

$$\frac{\partial \underline{\psi}_{\times,\mathrm{A}}}{\partial x}(x,y,\boldsymbol{\zeta},\boldsymbol{\eta}) = \frac{1}{2}\left(\underline{\eta} + \overline{\eta} + (\mu+1)(\overline{\eta}-\underline{\eta})\left(\frac{y-\eta}{\overline{\eta}-\underline{\eta}} - \frac{\overline{\zeta}-x}{\overline{\zeta}-\underline{\zeta}}\right)\left|\frac{y-\eta}{\overline{\eta}-\underline{\eta}} - \frac{\overline{\zeta}-x}{\overline{\zeta}-\underline{\zeta}}\right|^{\mu+1}\right),$$

$$\frac{\partial \underline{\psi}_{\times,\mathrm{A}}}{\partial y}(x,y,\boldsymbol{\zeta},\boldsymbol{\eta}) = \frac{1}{2}\left(\underline{\zeta} + \overline{\zeta} + (\mu+1)(\overline{\zeta}-\underline{\zeta})\left(\frac{y-\eta}{\overline{\eta}-\underline{\eta}} - \frac{\overline{\zeta}-x}{\overline{\zeta}-\underline{\zeta}}\right)\left|\frac{y-\eta}{\overline{\eta}-\underline{\eta}} - \frac{\overline{\zeta}-x}{\overline{\zeta}-\underline{\zeta}}\right|^{\mu+1}\right),$$

$$\frac{\partial \underline{\psi}_{\times,\mathrm{B}}}{\partial x}(x,y,\boldsymbol{\zeta},\boldsymbol{\eta}) = \underline{\eta} + (\mu+1)(\overline{\eta}-\underline{\eta})\left(\max\left\{0, \frac{y-\eta}{\overline{\eta}-\underline{\eta}} - \frac{\overline{\zeta}-x}{\overline{\zeta}-\underline{\zeta}}\right\}\right)^{\mu},$$

$$\frac{\partial \underline{\psi}_{\times,\mathrm{B}}}{\partial y}(x,y,\boldsymbol{\zeta},\boldsymbol{\eta}) = \underline{\zeta} + (\mu+1)(\overline{\zeta}-\underline{\zeta})\left(\max\left\{0, \frac{y-\eta}{\overline{\eta}-\underline{\eta}} - \frac{\overline{\zeta}-x}{\overline{\zeta}-\underline{\zeta}}\right\}\right)^{\mu}.$$

Proposition 16 *Suppose that the conditions of Theorem 7 hold. Gradients for the relaxations $\underline{g}^{\mathrm{C}}_{\max}$ and $\overline{g}^{\mathrm{C}}_{\max}$ are as follows, at any $z \in Z$. The required partial derivatives of $\underline{\psi}_{\max}$ may be evaluated at any argument that yields $\underline{g}^{\mathrm{C}}_{\max}$ in Theorem 7; this follows from a gradient invariance property of Mangasarian [37].*

$$\nabla \underline{g}^{\mathrm{C}}_{\max}(z) = \begin{cases} \nabla \underline{f}^{\mathrm{C}}_1(z), & \text{if } \overline{x}_2 \le \underline{x}_1, \\ \nabla \underline{f}^{\mathrm{C}}_2(z), & \text{if } \overline{x}_1 \le \underline{x}_2, \\ \max\left\{0, \frac{\partial \underline{\psi}_{\max}}{\partial x}\right\} \nabla \underline{f}^{\mathrm{C}}_1(z) + \min\left\{0, \frac{\partial \underline{\psi}_{\max}}{\partial x}\right\} \nabla \overline{f}^{\mathrm{C}}_1(z) & \\ \quad + \max\left\{0, \frac{\partial \underline{\psi}_{\max}}{\partial y}\right\} \nabla \underline{f}^{\mathrm{C}}_2(z) + \min\left\{0, \frac{\partial \underline{\psi}_{\max}}{\partial y}\right\} \nabla \overline{f}^{\mathrm{C}}_2(z), & \text{otherwise,} \end{cases}$$

and the mapping $\overline{g}^{\mathrm{C}}_{\max}$ *has the following gradient, evaluated at any* $z \in Z$*:*

$$\nabla\overline{g}^{\mathrm{C}}_{\max}(z) = \begin{cases} \nabla\overline{f}^{\mathrm{C}}_1(z), & \text{if } \overline{x}_2 \le \underline{x}_1, \\ \nabla\overline{f}^{\mathrm{C}}_2(z), & \text{if } \overline{x}_1 \le \underline{x}_2, \\ \left(\frac{\max\{\overline{x}_1,\overline{x}_2\}-\max\{\underline{x}_1,\overline{x}_2\}}{\overline{x}_1-\underline{x}_1}\right)\nabla\overline{f}^{\mathrm{C}}_1(z) + \left(\frac{\max\{\overline{x}_1,\overline{x}_2\}-\max\{\overline{x}_1,\underline{x}_2\}}{\overline{x}_2-\underline{x}_2}\right)\nabla\overline{f}^{\mathrm{C}}_2(z) \\ \qquad + (\mu+1)\Delta\left(\max\left\{0, \frac{\overline{x}_1-\overline{f}^{\mathrm{C}}_1(z)}{\overline{x}_1-\underline{x}_1} - \frac{\overline{f}^{\mathrm{C}}_2(z)-\underline{x}_2}{\overline{x}_2-\underline{x}_2}\right\}\right)^{\mu}\left(\frac{\nabla\overline{f}^{\mathrm{C}}_1(z)}{\overline{x}_1-\underline{x}_1} + \frac{\nabla\overline{f}^{\mathrm{C}}_2(z)}{\overline{x}_2-\underline{x}_2}\right), \\ \qquad\qquad \text{otherwise,} \end{cases}$$

where $\Delta := \max\{\underline{x}_1,\overline{x}_2\} + \max\{\overline{x}_1,\underline{x}_2\} - \max\{\underline{x}_1,\underline{x}_2\} - \max\{\overline{x}_1,\overline{x}_2\}$*, and where the required partial derivatives of* $\underline{\psi}_{\max}$ *are as follows.*

$$\frac{\partial\underline{\psi}_{\max}}{\partial x}(x,y,\boldsymbol{\zeta},\boldsymbol{\eta}) = \begin{cases} 1-(\mu+1)\left(\max\left\{0, \frac{y-x}{\overline{\eta}-\underline{\zeta}}\right\}\right)^{\mu}, & \text{if } \underline{\eta} \le \underline{\zeta} < \overline{\eta}, \\ (\mu+1)\left(\max\left\{0, \frac{x-y}{\overline{\zeta}-\underline{\eta}}\right\}\right)^{\mu}, & \text{if } \underline{\zeta} < \underline{\eta} < \overline{\zeta}, \end{cases}$$

$$\frac{\partial\underline{\psi}_{\max}}{\partial y}(x,y,\boldsymbol{\zeta},\boldsymbol{\eta}) = \begin{cases} (\mu+1)\left(\max\left\{0, \frac{y-x}{\overline{\eta}-\underline{\zeta}}\right\}\right)^{\mu}, & \text{if } \underline{\eta} \le \underline{\zeta} < \overline{\eta}, \\ 1-(\mu+1)\left(\max\left\{0, \frac{x-y}{\overline{\zeta}-\underline{\eta}}\right\}\right)^{\mu}, & \text{if } \underline{\zeta} < \underline{\eta} < \overline{\zeta}. \end{cases}$$

References

1. Achterberg, T.: SCIP: solving constraint integer programs. Math. Prog. Comp. **1**, 1–41 (2009)
2. Achterberg, T., Berthold, T., Koch, T., Wolter, K.: Constraint integer programming: a new approach to integrate CP and MIP. In: Proceedings of the Fifth International Conference on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems, pp. 6–20. Paris (2008)
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. Computers Chem. Engng **22**, 1137–1158 (1998)
4. Alefeld, G., Mayer, G.: Interval analysis: theory and applications. J. Comput. Appl. Math. **121**, 421–464 (2000)
5. Bazaraa, M.S., Sherali, H.D., Shetty, C.M.: Nonlinear Programming: Theory and Algorithms, 3rd edn. John Wiley & Sons, Inc., Hoboken (2006)
6. Beckers, M., Mosenkis, V., Naumann, U.: Adjoint mode computation of subgradients for McCormick relaxations. In: S. Forth, P. Hovland, E. Phipps, J. Utke, A. Walther (eds.) Recent Advances in Algorithmic Differentiation, pp. 103–113. Springer-Verlag, Berlin (2012)
7. Belotti, P.: COUENNE: A user's manual (2006). Retrieved online from `https://projects.coin-or.org/Couenne`
8. Bertsekas, D.P.: Nondifferentiable optimization via approximation. In: M. Balinski, P. Wolfe (eds.) Mathematical Programming Study 3, pp. 1–25. North-Holland Publishing Company, Amsterdam (1975)
9. Bertsekas, D.P.: Nonlinear Programming, 2nd edn. Athena Scientific, Belmont, MA (1999)
10. Bompadre, A., Mitsos, A.: Convergence rate of McCormick relaxations. J. Glob. Optim. **52**, 1–28 (2012)
11. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
12. Broyden, C.G., Dennis Jr., J.E., Moré, J.J.: On the local and superlinear convergence of quasi-Newton methods. J. Inst. Math. Appl. **12**, 223–245 (1973)
13. Chachuat, B.: MC++: A toolkit for bounding factorable functions, v1.0 (2014). Retrieved online on July 2, 2014, from `https://projects.coin-or.org/MCpp`
14. Clarke, F.H.: Optimization and Nonsmooth Analysis. SIAM, Philadelphia, PA (1990)
15. Coddington, E.A., Levinson, N.: Theory of Ordinary Differential Equations. McGraw Hill Co., Inc., New York, NY (1955)
16. Du, K., Kearfott, R.B.: The cluster problem in multivariate global optimization. J. Glob. Optim. **5**, 253–265 (1994)
17. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, vol. 2. Springer-Verlag New York, Inc., New York, NY (2003)

18. Feehery, W.F., Tolsma, J.E., Barton, P.I.: Efficient sensitivity analysis of large-scale differential-algebraic systems. Appl. Numer. Math. **25**, 41–54 (1997)
19. Gabriel, S.A., Moré, J.J.: Smoothing of mixed complementarity problems. Preprint MCS-P541-0995, Argonne National Laboratory (1995)
20. Gill, P.E., Murray, W., Saunders, M.A.: SNOPT: An SQP Algorithm for Large-Scale Constrained Optimization. SIAM Review **47**(1), 99–131 (2002)
21. Griewank, A., Rabier, P.J.: On the smoothness of convex envelopes. T. Am. Math. Soc. **322**, 691–709 (1990)
22. Griewank, A., Walther, A.: Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation, 2nd edn. Other Titles in Applied Mathematics. SIAM, Philadelphia, PA (2008)
23. Grossmann, I.E., Yeomans, H., Kravanja, Z.: A rigorous disjunctive optimization model for simultaneous flowsheet optimization and heat integration. Computers & Chemical Engineering **22**(98), 157–164 (1998)
24. Hartman, P.: Ordinary Differential Equations, second edn. SIAM, Philadelphia, PA (2002)
25. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms I: Fundamentals. A Series of Comprehensive Studies in Mathematics. Springer-Verlag, Berlin (1993)
26. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods. A Series of Comprehensive Studies in Mathematics. Springer-Verlag, Berlin (1993)
27. Horst, R., Tuy, H.: Global Optimization: Deterministic Approaches, 2nd edn. Springer-Verlag, Berlin (1993)
28. Kesavan, P., Allgor, R.J., Gatzke, E.P., Barton, P.I.: Outer approximation algorithms for separable nonconvex mixed-integer nonlinear programs. Math. Program., Ser. A **100**, 517–535 (2004)
29. Khan, K.A.: Sensitivity analysis for nonsmooth dynamic systems. Ph.D. thesis, Massachusetts Institute of Technology (2015)
30. Kiwiel, K.C.: Methods of Descent for Nondifferentiable Optimization. Lecture Notes in Mathematics. Springer-Verlag, Berlin (1985)
31. Lemaréchal, C., Strodiot, J.J., Bihain, A.: On a bundle algorithm for nonsmooth optimization. In: O.L. Mangasarian, R.R. Meyer, S.M. Robinson (eds.) Nonlinear Programming 4. Academic Press, New York, NY (1981)
32. Li, X., Tomasgard, A., Barton, P.I.: Nonconvex generalized Benders decomposition for stochastic separable mixed-integer nonlinear programs. J. Optim. Theory Appl. **151**, 425–454 (2011)
33. Liberti, L., Pantelides, C.C.: Convex envelopes of monomials of odd degree. J. Glob. Optim. **25**, 157–168 (2003)
34. Lin, Y., Schrage, L.: The global solver in the LINDO API. Optim. Method Softw. **24**, 657–668 (2009)
35. Mäkelä, M.M.: Multiobjective proximal bundle method for nonconvex nonsmooth optimization: Fortran subroutine MPBNGC 2.0. Reports of the Department of Mathematical Information Technology, Series B, Scientific computing B 13/2003, University of Jyväskylä (2003)
36. Maly, T., Petzold, L.R.: Numerical methods and software for sensitivity analysis of differential-algebraic systems. Appl. Numer. Meth. **20**, 57–79 (1996)
37. Mangasarian, O.L.: A simple characterization of solution sets of convex programs. Oper. Res. Lett. **7**(1), 21–26 (1988)
38. McCormick, G.P.: Computability of global solutions to factorable nonconvex programs: Part I - Convex underestimating problems. Math. Program. **10**, 147–175 (1976)
39. Misener, R., Floudas, C.A.: ANTIGONE: Algorithms for coNTinuous/Integer Global Optimization of Nonlinear Equations. J. Glob. Optim. **59**, 503–526 (2014)
40. Mitsos, A., Chachuat, B., Barton, P.I.: McCormick-based relaxations of algorithms. SIAM J. Optim. **20**, 573–601 (2009)
41. Moore, R.E.: Methods and Applications of Interval Analysis. SIAM, Philadelphia (1979)
42. Najman, J., Mitsos, A.: Convergence analysis of multivariate McCormick relaxations. J. Glob. Optim. **in press** (2016)
43. Naumann, U.: The Art of Differentiating Computer Programs. SIAM, Philadelphia (2012)
44. Neumaier, A.: Interval Methods for Systems of Equations. Cambridge University Press, Cambridge (1990)
45. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer Series in Operations Research and Financial Engineering. Springer, New York (2006)
46. Ortega, J.M., Rheinboldt, W.C.: Iterative Solution of Nonlinear Equations in Several Variables. Classics in Applied Mathematics. SIAM, Philadelphia (2000)
47. Qi, L., Sun, D.: Smoothing functions and smoothing Newton method for complementarity and variational inequality problems. J. Optim. Theory App. **113**, 121–147 (2002)

48. Rockafellar, R.T.: Convex Analysis. Princeton Landmarks in Mathematics and Physics. Princeton University Press, Princeton (1970)
49. Ryoo, H.S., Sahinidis, N.V.: Global optimization of nonconvex NLPs and MINLPs with applications in process design. Computers & Chemical Engineering **19**(5), 551–566 (1995)
50. Sahinidis, N.V.: BARON: A general purpose global optimization software package. J. Glob. Optim. **8**, 201–205 (1996)
51. Sahinidis, N.V.: BARON 15.9: Global Optimization of Mixed-Integer Nonlinear Programs, *User's Manual* (2015). Available at https://www.gams.com/help/topic/gams.doc/solvers/baron/index.html
52. Schaber, S.D.: Tools for dynamic model development. Ph.D. thesis, Massachusetts Institute of Technology (2014)
53. Scholz, D.: Theoretical rate of convergence for interval inclusion functions. J. Glob. Optim. **53**, 749–767 (2012)
54. Scott, J.K.: Reachability analysis and deterministic global optimization of differential-algebraic systems. Ph.D. thesis, Massachusetts Institute of Technology (2012)
55. Scott, J.K., Barton, P.I.: Convex and concave relaxations for the parametric solutions of semi-explicit index-one differential-algebraic equations. J. Optim. Theory App. **156**, 617–649 (2013)
56. Scott, J.K., Barton, P.I.: Improved relaxations for the parametric solutions of ODEs using differential inequalities. J. Glob. Optim. **57**, 143–176 (2013)
57. Scott, J.K., Chachuat, B., Barton, P.I.: Nonlinear convex and concave relaxations for the solutions of parametric ODEs. Optim. Control Appl. Meth. **34**, 145–163 (2013)
58. Scott, J.K., Stuber, M.D., Barton, P.I.: Generalized McCormick relaxations. J. Glob. Optim. **51**, 569–606 (2011)
59. Shor, N.Z.: Minimization Methods for Non-Differentiable Functions. Springer Series in Computational Mathematics. Springer-Verlag, Berlin (1985)
60. Stuber, M.D., Scott, J.K., Barton, P.I.: Convex and concave relaxations of implicit functions. Optim. Method Softw. **30**(3), 424–460 (2015)
61. Tawarmalani, M., Sahinidis, N.V.: Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications. Nonconvex Optimization and Its Applications. Springer Science+Business Media, Dordrecht (2002)
62. Tawarmalani, M., Sahinidis, N.V.: Global optimization of mixed-integer nonlinear programs: A theoretical and computational study. Math. Program. A **99**, 563–591 (2004)
63. Tsoukalas, A., Mitsos, A.: Multivariate McCormick relaxations. J. Glob. Optim. **59**, 633–662 (2014)
64. Watson, H.A.J., Khan, K.A., Barton, P.I.: Multistream heat exchanger modeling and design. AIChE Journal **61**(10), 3390–3403 (2015)
65. Wechsung, A.: Global optimization in reduced space. Ph.D. thesis, Massachusetts Institute of Technology (2014)
66. Wechsung, A., Aspelund, A., Gundersen, T., Barton, P.I.: Synthesis of heat exchanger networks at subambient conditions with compression and expansion of process streams. AIChE Journal **57**(8), 2090–2108 (2011)
67. Wechsung, A., Schaber, S.D., Barton, P.I.: The cluster problem revisited. J. Glob. Optim. **58**, 429–438 (2014)
68. Wechsung, A., Scott, J.K., Watson, H.A.J., Barton, P.I.: Reverse propagation of McCormick relaxations. J. Glob. Optim. **63**(1), 1–36 (2015)
69. Whitney, H.: Analytic extensions of differentiable functions defined in closed sets. Trans. Amer. Math. Soc. **36**, 63–89 (1934)