



Generalized derivatives of optimal-value functions with parameterized convex programs embedded

Yingkai Song¹ · Paul I. Barton¹

Received: 27 June 2022 / Accepted: 23 December 2023 / Published online: 25 January 2024
© The Author(s) 2024

Abstract

This article proposes new practical methods for furnishing generalized derivative information of optimal-value functions with embedded parameterized convex programs, with potential applications in nonsmooth equation-solving and optimization. We consider three cases of parameterized convex programs: (1) partial convexity—functions in the convex programs are convex with respect to decision variables for fixed values of parameters, (2) joint convexity—the functions are convex with respect to both decision variables and parameters, and (3) linear programs where the parameters appear in the objective function. These new methods calculate an LD-derivative, which is a recently established useful generalized derivative concept, by constructing and solving a sequence of auxiliary linear programs. In the general partial convexity case, our new method requires that the strong Slater conditions are satisfied for the embedded convex program’s decision space, and requires that the convex program has a unique optimal solution. It is shown that these conditions are essentially less stringent than the regularity conditions required by certain established methods, and our new method is at the same time computationally preferable over these methods. In the joint convexity case, the uniqueness requirement of an optimal solution is further relaxed, and to our knowledge, there is no established method for computing generalized derivatives prior to this work. In the linear program case, both the Slater conditions and the uniqueness of an optimal solution are not required by our new method.

Keywords Perturbation analysis · Generalized derivatives · Convex programs

Mathematics Subject Classification 49J52 · 90C25 · 90C31

This work was funded by the OCP Group, Morocco.

✉ Paul I. Barton
pib@mit.edu

Yingkai Song
yingkais@mit.edu

¹ Process Systems Engineering Laboratory, Department of Chemical Engineering, MIT, 77 Massachusetts Ave, Cambridge, MA 02139, USA

1 Introduction

This article considers parameterized convex programs of the form

$$\begin{aligned} \phi(\mathbf{y}) &:= \min_{\mathbf{x} \in X} f(\mathbf{x}, \mathbf{y}) \\ \text{s.t. } & \mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0}, \\ & \mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{0}, \end{aligned} \quad (1)$$

where the functions f , \mathbf{g} , \mathbf{h} are continuously differentiable, ϕ is an optimal-value function in the sense of Thibault [51], and for each fixed \mathbf{y} , the optimization problem is convex. It is well-known that even if the optimization problem in (1) is convex and if the functions f , \mathbf{g} , and \mathbf{h} are sufficiently smooth, the function ϕ is still nonsmooth and nonconvex in general, or even discontinuous without certain regularity conditions (c.f. [4]). Due to such nonsmoothness, algorithms for equation-solving [10, 15, 40, 52] and optimization [31, 32, 53, 55] with ϕ embedded typically require furnishing ϕ 's generalized derivatives, which are sensitivity concepts for nonsmooth functions generalized from smooth functions. This article will propose new implementable methods for furnishing the generalized derivatives of ϕ , under mild conditions that can be satisfied in practice.

(Clarke's) *generalized Jacobian* [6] is a point-to-set generalization of the classical derivative to locally Lipschitz continuous functions that are not differentiable everywhere, and is useful in methods for nonsmooth equation-solving or optimization. Semismooth Newton methods [40, 52] for equation-solving replace the usual derivatives employed in the classical Newton method with elements of the generalized Jacobian, and can exhibit Q-quadratic local convergence properties [39] similarly to their smooth counterparts [36]. Recently, a linear program (LP)-based Newton method [10, 15] even relaxes the non-singularity requirement of sensitivities in both classical and semismooth Newton methods, while keeping the same local convergence properties. Besides, bundle methods [31, 32] for nonsmooth optimization utilize generalized Jacobian elements to approximate the original nonsmooth functions locally at each iteration. Elements of the generalized Jacobian were also supplied to the smooth optimization solver IPOPT [54] to solve nonsmooth programs [53, 55], which turned out to outperform the bundle methods in some cases.

Though the generalized Jacobian elements are useful, they are typically difficult to compute in practice. One reason is that the generalized Jacobian only satisfy calculus rules as inclusions (see [5, Theorem 2.6.6]); often strict inclusions [1]. *Nesterov's lexicographic derivative (L-derivative)* [35] and the *lexicographic directional derivative (LD-derivative)* [27] are useful for mitigating this problem. The L-derivative is another generalized derivative concept for nonsmooth functions. It has been shown in [26] that the L-derivatives are elements of the plenary hull of the generalized Jacobian whenever they exist. As a result, the L-derivatives are no less useful than generalized Jacobian elements in methods for equation-solving or optimization, in the sense that using L-derivatives in place of these elements does not affect the desirable convergence properties of these methods (c.f. [27, Sect. 2.2]). Moreover, for piecewise-differentiable functions (in the sense of Scholtes [47]) or scalar-valued functions such as ϕ above, L-derivatives are guaranteed to comprise a subset of the generalized Jacobian (c.f. [27, Sect. 3] and [35, Theorem 11], respectively). The LD-derivative [27] is a nonsmooth extension of the classical directional derivative, which provides a practical way for calculating the L-derivatives. From an LD-derivative, an L-derivative can be computed by solving a simple linear equation system. Unlike the generalized Jacobian, the LD-derivatives follow a sharp chain rule for composite functions, which enables a vector for-

ward mode of automatic differentiation (AD) [27] or a reverse AD mode via branch-locking [25].

The goal of this article is to propose implementable methods for furnishing useful generalized derivative information for ϕ in (1). Unfortunately, established methods for estimating, characterizing, or computing generalized derivatives of optimal-value functions are of limited scope for parameterized nonlinear programs (NLPs). For parameterized NLPs whose parameters only appear in the objective function, sufficient conditions under which the optimal-value function is differentiable were given in [4, 38]. When the optimal-value function is not differentiable, Clarke [5] gave a characterization of the generalized Jacobian. For the situation where parameters only appear at constraints' right-hand sides, Gauvin [16] estimated the generalized Jacobian as a subset of the convex hull of all KKT multipliers. If assuming parameterized convex NLPs and that the strong Slater conditions ([20, Definition 2.3.1]) hold, then the corresponding optimal-value function is convex, and the generalized Jacobian is identical to the set of all KKT multipliers (follows from [5, Proposition 1.2] and [20, Theorems 3.3.2 and 3.3.3]). For linear programs (LPs) parameterized by the constraints' right-hand-side vectors, Höffner et al. [21] proposed a practical method for computing an LD-derivative by solving a sequence of auxiliary LPs. This method was extended to compute LD-derivatives for lexicographic LPs [18] and dynamic systems with LPs embedded [21]. The situation where parameters appear at the left-hand side of constraints, such as in (1), is in general more difficult to handle (and can readily cover the right-hand-side perturbations). Some results for estimating the Fréchet subgradients can be found in [34]. Rockafellar [43, 44] estimated the generalized Jacobian based on Lagrange multipliers (treat active inequality constraints as equalities) and the concept of proximal subgradients. When the decision space is closed and convex for fixed parameters, Thibault [51] estimated the generalized Jacobian of the optimal-value function using generalized Jacobians of the support functions of the decision space. De Wolf and Smeers [9] characterized the generalized Jacobian of LPs parameterized by the matrix coefficients in constraints, which were then applied to solve nonsmooth NLPs arising in the applications of natural gas transmission networks. Other studies (e.g. [4, 7, 8, 22, 41, 45], among many) focus on directional differentiability of optimal-value functions. However, such directional derivatives are not enough for furnishing an L-derivative or an element in the generalized Jacobian [1], which are in turn useful in algorithms for nonsmooth equation-solving and optimization.

There are also extensive studies (refer to [4, 11, 12, 14, 29]) focusing on generalized sensitivity information of the optimal-solution mapping of parameterized programs. Given such information, generalized derivatives of the optimal-value function may be easily computed. Recent works computed an LD-derivative of the optimal-solution mapping of parametric NLPs by solving a nonsmooth and nonlinear equation system [49], or by solving a sequence of quadratic programs [48], both under the linear independence constraint qualification (LICQ) and strong second-order sufficient condition. Under these conditions, the optimal solution is guaranteed to be unique around the parameter values of interest. Then, an LD-derivative of the optimal-value function follows from the sharp chain rule for LD-derivatives [27]. These methods could in principle be used to compute an LD-derivative of ϕ in (1). However, these methods cannot deal with parameterized convex NLPs with non-unique optimal solutions, and the LICQ cannot be verified unless the original NLP is solved, and is also somewhat stringent in convex optimization. For example, LICQ is not satisfied at any degenerate vertex (c.f. [3, Sect. 2.4]) of a polyhedral decision space. However, to the authors' knowledge, these are the only established methods that may be used to achieve the goal of this article.

In this article, we propose new practical methods for evaluating useful LD-derivatives of the optimal-value function ϕ with parameterized convex NLPs embedded of the form (1). As mentioned, LD-derivatives can be used for computing L-derivatives, and the L-derivatives are guaranteed to be valid generalized Jacobian elements for optimal-value functions, which are in turn useful for nonsmooth equation-solving and optimization. Moreover, such LD-derivatives follow a sharp chain rule for composite functions, and thus allowing treatment of complex problems with ϕ embedded (in contrary to methods only computing generalized Jacobian elements e.g. [5, 9]). We consider three cases of parameterized convex NLPs of the form (1):

- *Partial convexity* – the functions f , \mathbf{g} , and \mathbf{h} are convex with respect to \mathbf{x} , for each fixed \mathbf{y} .
- *Joint convexity* – the functions f , \mathbf{g} , and \mathbf{h} are jointly convex with respect to both \mathbf{x} and \mathbf{y} .
- *Parameterized LP's objective function* – the optimization problem in (1) is an LP wherein \mathbf{y} only appears in the objective function.

We propose new methods for computing the LD-derivatives for each of the three cases mentioned above. In each case, the new method computes the LD-derivatives by solving a sequence of LPs, of which the computational complexity grows polynomially with the size of the problem. In the partial convexity case, our new method is in general computationally preferable over the established methods [48, 49], which solve a nonsmooth and nonlinear equation system or a sequence of quadratic programs. The new method requires that the strong Slater conditions hold for the parameterized convex NLPs and requires that there exists exactly one optimal solution at the parameter value of interest. These conditions are essentially less stringent than the regularity conditions in [48, 49] (as will be discussed in Sect. 3.1), and thus our new method is applicable to a broader class of problems with partial convexity. We nevertheless note that [48, 49] primarily aim at calculating LD-derivatives of the optimal-solution mapping for general nonconvex parameterized NLPs, which intuitively needs more restrictive assumptions than our purpose. In the joint convexity case, the uniqueness requirement of an optimal solution is further relaxed; our new method only assumes a nonempty and bounded optimal solutions set for the parameters' values of interest. Due to such non-uniqueness of optimal solutions, the methods in [48, 49] are no longer applicable. To the authors' knowledge, prior to this article, there is no established method for furnishing computationally relevant generalized derivatives in this case. In the case of parameterized LP's objective function, both the Slater conditions and the uniqueness requirement of an optimal solution are not required, and the new method necessarily yields an element of the *B-subdifferential* (see Definition 1 below). It has been shown [27] that the B-subdifferential-based Newton method [30] exhibits local Q-quadratic convergence under less stringent conditions than the semismooth Newton method [40] based on the generalized Jacobian.

The rest of this article is structured as follows. Section 2 presents mathematical preliminaries including notational convention, Clarke generalized Jacobian, lexicographic derivative, and set-valued mapping. Sections 3.1, 3.2, and 3.3 respectively present the new methods for computing an LD-derivative of ϕ in the partial convexity, joint convexity, and parameterized LP's objective function cases.

2 Preliminaries

2.1 Notation

Throughout this article, scalars are denoted as lowercase letters (e.g. $x \in \mathbb{R}$), vectors are denoted as boldface lowercase letters (e.g. $\mathbf{x} \in \mathbb{R}^n$), and the i^{th} component of a vector \mathbf{x} is denoted as x_i . $\mathbf{0}$ denotes a vector with all elements to be 0, whose dimensions are clear from the context. Matrices are denoted as boldface uppercase letters (e.g. $\mathbf{M} \in \mathbb{R}^{n \times m}$). The j^{th} column of a matrix \mathbf{M} is denoted as $\mathbf{m}_{(j)}$. Sets are denoted as uppercase letters (e.g. $X \subset \mathbb{R}^n$). The interior of a set X is denoted as $\text{int}(X)$. Inequalities involving vectors are to be interpreted componentwise. The standard Euclidean norm $\|\cdot\|$ is considered on \mathbb{R}^n . For a function $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ that is (Fréchet) differentiable at $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{Jf}(\mathbf{x}) \in \mathbb{R}^{m \times n}$ denotes the Jacobian of \mathbf{f} at \mathbf{x} . For a scalar-valued differentiable function f of (\mathbf{x}, \mathbf{y}) , $\nabla_{\mathbf{x}} f \in \mathbb{R}^{n_x}$ and $\nabla_{\mathbf{y}} f \in \mathbb{R}^{n_y}$ denote the partial derivative of f with respect to \mathbf{x} and \mathbf{y} , respectively. A vector-valued function \mathbf{f} is said to be convex if each component f_i is convex.

2.2 Generalized Jacobian

Clarke’s generalized Jacobian [6] defined below is a pertinent generalized derivative concept for locally Lipschitz continuous functions that are not differentiable everywhere.

Definition 1 (from [27], originally from [6]) Given an open set $X \subset \mathbb{R}^n$ and a locally Lipschitz continuous function $\mathbf{f} : X \rightarrow \mathbb{R}^m$, let $S \subset X$ be the set on which \mathbf{f} is differentiable. The *B-subdifferential* of \mathbf{f} at $\mathbf{x} \in X$ is defined as

$$\partial_B \mathbf{f}(\mathbf{x}) := \{ \mathbf{H} \in \mathbb{R}^{m \times n} : \mathbf{H} = \lim_{i \rightarrow \infty} \mathbf{Jf}(\mathbf{x}^{(i)}), \mathbf{x} = \lim_{i \rightarrow \infty} \mathbf{x}^{(i)}, \mathbf{x}^{(i)} \in S, \forall i \in \mathbb{N} \}.$$

The (Clarke) generalized Jacobian of \mathbf{f} at \mathbf{x} , denoted as $\partial \mathbf{f}(\mathbf{x})$, is the convex hull of $\partial_B \mathbf{f}(\mathbf{x})$, and is nonempty, compact, and convex.

As introduced in Sect. 1, elements of the generalized Jacobian are used in the semismooth Newton method [40] and LP-Newton method [10, 15] for nonsmooth equation-solving, and in bundle methods [31, 32] for nonsmooth optimization.

2.3 Lexicographic derivatives

This section introduces Nesterov’s lexicographic derivatives [35], which have been shown to be no less useful than the generalized Jacobian for nonsmooth equation-solving and optimization [26, 35].

Definition 2 (from [47]) Given an open set $X \subset \mathbb{R}^n$, a function $\mathbf{f} : X \rightarrow \mathbb{R}^m$, some $\mathbf{x} \in X$, and some $\mathbf{d} \in \mathbb{R}^n$, the limit

$$\lim_{\alpha \downarrow 0} \frac{\mathbf{f}(\mathbf{x} + \alpha \mathbf{d}) - \mathbf{f}(\mathbf{x})}{\alpha}$$

is called the *directional derivative* of \mathbf{f} at \mathbf{x} if and only if it exists, and is denoted as $\mathbf{f}'(\mathbf{x}; \mathbf{d})$. The function \mathbf{f} is *directionally differentiable* at \mathbf{x} if and only if $\mathbf{f}'(\mathbf{x}; \mathbf{d})$ exists and is finite for each $\mathbf{d} \in \mathbb{R}^n$. The function \mathbf{f} is directionally differentiable if and only if it is directionally differentiable at each point in its domain.

Definition 3 (from [27], originally from [35]) Let $X \subset \mathbb{R}^n$ be open and $\mathbf{f} : X \rightarrow \mathbb{R}^m$ be locally Lipschitz continuous and directionally differentiable at $\mathbf{x} \in X$. The function \mathbf{f} is *lexicographically smooth (L-smooth)* at \mathbf{x} if and only if for any $p \in \mathbb{N}$ and any matrix $\mathbf{M} := [\mathbf{m}_{(1)} \ \mathbf{m}_{(2)} \ \cdots \ \mathbf{m}_{(p)}] \in \mathbb{R}^{n \times p}$, the following homogenization sequence of functions is well defined:

$$\begin{aligned} \mathbf{f}_{\mathbf{x},\mathbf{M}}^{(0)} : \mathbb{R}^n &\rightarrow \mathbb{R}^m : \mathbf{d} \mapsto \mathbf{f}'(\mathbf{x}; \mathbf{d}), \\ \mathbf{f}_{\mathbf{x},\mathbf{M}}^{(j)} : \mathbb{R}^n &\rightarrow \mathbb{R}^m : \mathbf{d} \mapsto [\mathbf{f}_{\mathbf{x},\mathbf{M}}^{(j-1)}]'(\mathbf{m}_{(j)}; \mathbf{d}), \quad \forall j \in \{1, \dots, p\}. \end{aligned}$$

The function \mathbf{f} is L-smooth if and only if it is L-smooth at each point in its domain. When the columns of \mathbf{M} span \mathbb{R}^n , $\mathbf{f}_{\mathbf{x},\mathbf{M}}^{(p)}$ is guaranteed to be linear. If $p = n$ and $\mathbf{M} \in \mathbb{R}^{n \times n}$ is non-singular, then a *lexicographic derivative (L-derivative)* of \mathbf{f} at \mathbf{x} in the directions \mathbf{M} , denoted as $\mathbf{J}_L \mathbf{f}(\mathbf{x}; \mathbf{M})$, is defined as

$$\mathbf{J}_L \mathbf{f}(\mathbf{x}; \mathbf{M}) := \mathbf{J} \mathbf{f}_{\mathbf{x},\mathbf{M}}^{(n)}(\mathbf{0}) \in \mathbb{R}^{m \times n},$$

and the *lexicographic subdifferential* $\partial_L \mathbf{f}(\mathbf{x})$ is defined as

$$\partial_L \mathbf{f}(\mathbf{x}) := \{\mathbf{J}_L \mathbf{f}(\mathbf{x}; \mathbf{M}) \in \mathbb{R}^{m \times n} : \forall \mathbf{M} \in \mathbb{R}^{n \times n} \text{ and non-singular}\}.$$

It has been shown that $\partial_L \mathbf{f}(\mathbf{x}) \subset \partial \mathbf{f}(\mathbf{x})$ if \mathbf{f} is piecewise differentiable in the sense of Scholtes [47] ([27, Sect. 3]), or if \mathbf{f} is a scalar-valued function ([35, Theorem 11]). Moreover, for vector-valued functions but not necessarily piecewise differentiable, Khan and Barton [26] showed that the L-derivatives are elements of the plenary hull [50] of the generalized Jacobian whenever they exists. These elements are no less useful than the generalized Jacobian in methods for nonsmooth equation-solving and optimization, in the sense that using these in place of the generalized Jacobian elements will not affect the convergence results of these methods (c.f. [26, Sect. 2.4]).

The following definition of *lexicographic directional derivative* provides a convenient way for evaluating an L-derivative.

Definition 4 (from [27]) Given an open set $X \subset \mathbb{R}^n$, a locally Lipschitz continuous function $\mathbf{f} : X \rightarrow \mathbb{R}^m$ that is L-smooth at $\mathbf{x} \in X$, and a matrix $\mathbf{M} := [\mathbf{m}_{(1)} \ \cdots \ \mathbf{m}_{(p)}] \in \mathbb{R}^{n \times p}$, define the *lexicographic directional derivative (LD-derivative)* as

$$\mathbf{f}'(\mathbf{x}; \mathbf{M}) := [\mathbf{f}_{\mathbf{x},\mathbf{M}}^{(0)}(\mathbf{m}_{(1)}) \ \mathbf{f}_{\mathbf{x},\mathbf{M}}^{(1)}(\mathbf{m}_{(2)}) \ \cdots \ \mathbf{f}_{\mathbf{x},\mathbf{M}}^{(p-1)}(\mathbf{m}_{(p)})].$$

As shown in [27], such LD-derivatives are particularly useful for evaluating the L-derivatives. Firstly, it holds that

$$\mathbf{f}'(\mathbf{x}; \mathbf{M}) \equiv [\mathbf{f}_{\mathbf{x},\mathbf{M}}^{(p)}(\mathbf{m}_{(1)}) \ \cdots \ \mathbf{f}_{\mathbf{x},\mathbf{M}}^{(p)}(\mathbf{m}_{(p)})].$$

Furthermore, since $\mathbf{f}_{\mathbf{x},\mathbf{M}}^{(n)}$ is linear, if $\mathbf{M} \in \mathbb{R}^{n \times n}$ is non-singular, $\mathbf{f}'(\mathbf{x}; \mathbf{M})$ and $\mathbf{J}_L \mathbf{f}(\mathbf{x}; \mathbf{M})$ are related by

$$\mathbf{f}'(\mathbf{x}; \mathbf{M}) \equiv \mathbf{J}_L \mathbf{f}(\mathbf{x}; \mathbf{M}) \mathbf{M}. \tag{2}$$

Secondly, the LD-derivatives follow a sharp chain rule (c.f. [27, Proposition 2.2] and [18, Theorems 2.1 and 2.2]) for L-smooth composite functions, which enables a vector forward mode of automatic differentiation [27].

2.4 Set-valued mappings

This section summarizes some closedness, boundedness, and Lipschitzian properties of set-valued mappings.

Definition 5 (from [46]) Given $Y \subset \mathbb{R}^n$ and $X \subset \mathbb{R}^m$, a *set-valued mapping* $\Omega : Y \rightrightarrows X$ maps each element of Y to a subset of X .

Definition 6 (from [23]) Given $Y \subset \mathbb{R}^n$ and $X \subset \mathbb{R}^m$, a set-valued mapping $\Omega : Y \rightrightarrows X$ is *closed* at $\bar{y} \in Y$ if for some sequence $\{y^k\} \subset Y$ such that $\bar{y} = \lim_{k \rightarrow \infty} y^k$, and some sequence $\{x^k\}$ such that $x^k \in \Omega(y^k)$ and $\bar{x} = \lim_{k \rightarrow \infty} x^k$, it follows that $\bar{x} \in \Omega(\bar{y})$.

Definition 7 (from [23]) Given $X \subset \mathbb{R}^m$, an open $Y \subset \mathbb{R}^n$, and a $\bar{y} \in Y$, a set-valued mapping $\Omega : Y \rightrightarrows X$ is *uniformly bounded near \bar{y}* , if there exists a neighborhood $N_{\bar{y}} \subset Y$ of \bar{y} such that $\cup_{y \in N_{\bar{y}}} \Omega(y)$ is bounded.

Definition 8 (from [28]) For some $x \in \mathbb{R}^n$ and $Z \subset \mathbb{R}^n$, let

$$d(x, Z) := \inf\{\|x - z\| : z \in Z\}.$$

Then, given $X^A, X^B \subset \mathbb{R}^n$, define the *Hausdorff distance* between X^A and X^B as

$$d_H(X^A, X^B) := \inf\{k : d(x^A, X^B) \leq k, d(x^B, X^A) \leq k, x^A \in X^A, x^B \in X^B\}.$$

Definition 9 (from [28]) Given $X \subset \mathbb{R}^m$, an open $Y \subset \mathbb{R}^n$, and $\bar{y} \in Y$, a set-valued mapping $\Omega : Y \rightrightarrows X$ is *Lipschitzian near \bar{y}* if there exist a neighborhood $N_{\bar{y}}$ of \bar{y} and a $l > 0$ such that

$$d_H(\Omega(y^A), \Omega(y^B)) \leq l \|y^A - y^B\|, \quad \forall y^A, y^B \in N_{\bar{y}}.$$

3 Computing generalized derivatives

This section presents our new methods for computing an LD-derivative of ϕ of the form (1), which is specialized to the partial convexity, joint convexity, and parameterized LP’s objective function cases as introduced in Sect. 1. These new methods compute an LD-derivative by solving a sequence of easily-constructed LPs. From an LD-derivative, an L-derivative of ϕ is readily computed via (2), provided a non-singular directions matrix $M \in \mathbb{R}^{n_y \times n_y}$ (the simplest is to employ an identity matrix $M := I$). Since ϕ is scalar-valued, [35, Theorem 11] indicates that the L-derivatives are elements of Clarke generalized Jacobian, which are in turn useful in nonsmooth equation-solving and optimization. Moreover, the sharp chain rule of LD-derivatives allows to evaluate generalized derivatives for more complex problems with ϕ embedded.

With a slightly abuse of notation, the same letters may be used to represent different quantities in different subsections. Nevertheless, these letters will have consistent meanings in each subsection.

3.1 Partial convexity

The following assumption formalizes a *parameterized convex NLP with partial convexity* considered in this subsection.

Assumption 1 Let $X \subset \mathbb{R}^{n_x}$ be open and convex, and let $Y \subset \mathbb{R}^{n_y}$ be open. Consider functions $f : X \times Y \rightarrow \mathbb{R}$, $\mathbf{g} : X \times Y \rightarrow \mathbb{R}^m$, and $\mathbf{h} : X \times Y \rightarrow \mathbb{R}^l$. Consider the following parameterized NLP:

$$\begin{aligned} \phi(\mathbf{y}) &:= \min_{\mathbf{x} \in X} f(\mathbf{x}, \mathbf{y}) \\ \text{s.t. } &\mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0}, \\ &\mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{0}. \end{aligned} \quad (3)$$

For some $\bar{\mathbf{y}} \in Y$ and some neighborhood $N_{\bar{\mathbf{y}}} \subset Y$ of $\bar{\mathbf{y}}$, define the parameterized decision space as a set-valued mapping $\Omega : N_{\bar{\mathbf{y}}} \rightrightarrows X$ such that for each $\mathbf{y} \in N_{\bar{\mathbf{y}}}$,

$$\Omega(\mathbf{y}) := \{\mathbf{x} \in X : \mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \text{ and } \mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{0}\}.$$

Suppose that the following conditions hold:

- I.1 the functions f , \mathbf{g} , and \mathbf{h} are continuously differentiable on $X \times N_{\bar{\mathbf{y}}}$,
 - I.2 the optimization problem in (3) for $\mathbf{y} := \bar{\mathbf{y}}$ has exactly one optimal solution, denoted as $\bar{\mathbf{x}}^*$,
 - I.3 there exists a closed set $\tilde{X} \subset X$ such that $\cup_{\mathbf{y} \in N_{\bar{\mathbf{y}}}} \Omega(\mathbf{y}) \subset \tilde{X}$,
- and for each $\mathbf{y} \in N_{\bar{\mathbf{y}}}$,
- I.4 $\phi(\mathbf{y})$ is finite,
 - I.5 the mappings $f(\cdot, \mathbf{y})$ and $\mathbf{g}(\cdot, \mathbf{y})$ are convex,
 - I.6 the function $\mathbf{h}(\cdot, \mathbf{y})$ is affine,
 - I.7 the strong Slater conditions hold for $\Omega(\mathbf{y})$: since Condition I.6 holds, for each $i \in \{1, \dots, l\}$, let $\mathbf{a}_y^i := \nabla_{\mathbf{x}} h_i(\mathbf{x}, \mathbf{y})$, $\forall \mathbf{x} \in X$, suppose that the vectors $\mathbf{a}_y^i, \forall i \in \{1, \dots, l\}$ are linearly independent and there exists a $\xi_{\mathbf{y}} \in X$ (can be dependent on \mathbf{y}) such that $\mathbf{g}(\xi_{\mathbf{y}}, \mathbf{y}) < \mathbf{0}$ and $\mathbf{h}(\xi_{\mathbf{y}}, \mathbf{y}) = \mathbf{0}$.

Observe that for each $\mathbf{y} \in N_{\bar{\mathbf{y}}}$, the optimization problem in (3) is convex. Then, (3) is called a *parameterized convex NLP with partial convexity*.

Condition I.3 can often be satisfied in practice, and if $X \equiv \mathbb{R}^{n_x}$, this condition is always satisfied with $\tilde{X} := \mathbb{R}^{n_x}$. Condition I.2 can be satisfied if the objective function $f(\cdot, \mathbf{y})$ is strictly convex, or if certain regularity conditions are satisfied (e.g. [41, 48, 49]). This condition can be relaxed in both the joint convexity and parameterized LP cases, as will be seen in Sects. 3.2 and 3.3.

For a special case where the parameters \mathbf{y} only appear in the objective function, [4, Remark 4.14] indicates that the optimal-value function ϕ in Assumption 1 is differentiable, and gives a practical method for calculating the gradients. In other general nonsmooth cases, [48, 49] are the only established methods that can be used for calculating a generalized derivative of ϕ . These methods compute an LD-derivative of a unique optimal-solution function, and the LD-derivative of ϕ follows from a sharp chain rule. We note that in the partial convexity setting (i.e., Conditions I.4–I.6 hold), Conditions I.1, I.2 and I.7 are essentially less stringent than the required regularity conditions in [48, 49] for the following reasons. [48, 49] require that the linear independence constraint qualification and the strong second-order sufficient condition hold at an optimal solution, and that the functions f , \mathbf{g} , and \mathbf{h} to be second-order continuously differentiable. Under these conditions, the parameterized primal optimal solution and dual optimal solution are guaranteed to be unique near $\bar{\mathbf{y}}$, and to be piecewise-differentiable functions on $N_{\bar{\mathbf{y}}}$. Notably, this directly implies Conditions I.1, I.2, and I.7 (see [20, Theorem 2.3.2] for satisfaction of Condition I.7), but not the converse. Interestingly,

our method requires the optimal solution at $\bar{\mathbf{y}}$ to be unique, but not necessarily around $\bar{\mathbf{y}}$. We nevertheless note that [48, 49] primarily aim at calculating LD-derivatives of the optimal-solution mapping for general nonconvex parameterized NLPs, which intuitively needs more restrictive assumptions than our purpose.

The following proposition establishes the locally Lipschitz continuity of ϕ in Assumption 1, which is essential for ϕ to be L-smooth.

Proposition 1 Under Assumption 1, the function ϕ is locally Lipschitz continuous at $\bar{\mathbf{y}}$.

Proof Let $W \subsetneq X$ be a bounded open convex set such that $\bar{\mathbf{x}}^* \in W$. Let $\tilde{N}_{\bar{\mathbf{y}}} \subsetneq N_{\bar{\mathbf{y}}}$ be a neighborhood of $\bar{\mathbf{y}}$. Since Condition I.1 holds, it follows that the functions f , \mathbf{g} , and \mathbf{h} are (globally) Lipschitz continuous on $W \times \tilde{N}_{\bar{\mathbf{y}}}$. Since Condition I.2 holds, the optimal solutions set of the NLP for $\mathbf{y} := \bar{\mathbf{y}}$ in (3) is nonempty and bounded. Combining these with Conditions I.3, I.5, I.6, and I.7 in Assumption 1, according to [28, Theorem 1], we obtain that the infimum function

$$\begin{aligned} \tilde{\phi}(\mathbf{y}) &:= \inf_{\mathbf{x} \in X} f(\mathbf{x}, \mathbf{y}) \\ \text{s.t. } &\mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0}, \\ &\mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{0}, \end{aligned} \tag{4}$$

is locally Lipschitz continuous at $\bar{\mathbf{y}}$. Since $\tilde{\phi}(\mathbf{y}) \equiv \phi(\mathbf{y}), \forall \mathbf{y} \in N_{\bar{\mathbf{y}}}$ due to Condition I.4, it follows that ϕ is locally Lipschitz continuous at $\bar{\mathbf{y}}$. \square

Remark 1 Note that [28, Theorem 1] used in the proof above requires that the set $X \equiv \mathbb{R}^{n_x}$, while here X is an open set. However, since the decision variables \mathbf{x} never leave X , relevant properties of f , \mathbf{g} , and \mathbf{h} only need to be held on X . Moreover, since Condition I.3 holds, [28, Theorem 1] is indeed applicable here with the justification above.

The following Lemmata establish important regularity properties of the primal and dual optimal solutions mappings of the optimization problem near $\bar{\mathbf{y}}$ in (3).

Lemma 1 Suppose that Assumption 1 holds. Define $M : N_{\bar{\mathbf{y}}} \rightrightarrows \mathbb{R}^{n_x}$ so that for each $\mathbf{y} \in N_{\bar{\mathbf{y}}}$, $M(\mathbf{y})$ is the set of optimal solutions of the optimization problem in (3), i.e.,

$$M(\mathbf{y}) := \{\mathbf{x} \in \Omega(\mathbf{y}) : \phi(\mathbf{y}) = f(\mathbf{x}, \mathbf{y})\}.$$

Then, M is nonempty and uniformly bounded near $\bar{\mathbf{y}}$.

Proof Condition I.4 implies that M is nonempty near $\bar{\mathbf{y}}$.

For some $\epsilon > 0$, define $M_\epsilon : N_{\bar{\mathbf{y}}} \rightrightarrows \mathbb{R}^{n_x}$ such that for each $\mathbf{y} \in N_{\bar{\mathbf{y}}}$, $M_\epsilon(\mathbf{y})$ is the set of all ϵ -optimal solutions of the NLP in (3), i.e.,

$$M_\epsilon(\mathbf{y}) := \{\mathbf{x} \in \Omega(\mathbf{y}) : f(\mathbf{x}, \mathbf{y}) \leq \phi(\mathbf{y}) + \epsilon\}.$$

Following the arguments in proof of Proposition 1 and applying [28, Theorem 1], we obtain that M_ϵ is Lipschitzian near $\bar{\mathbf{y}}$ as in Definition 9. Moreover, since $M(\bar{\mathbf{y}})$ is bounded due to Condition I.2, according to [42, Corollary 8.7.1], $M_\epsilon(\bar{\mathbf{y}})$ is bounded. Now, according to Definition 9, there exist a neighborhood $\hat{N}_{\bar{\mathbf{y}}} \subset N_{\bar{\mathbf{y}}}$ of $\bar{\mathbf{y}}$, a $l > 0$, and a $\sigma > 0$ such that

$$d_H(M_\epsilon(\mathbf{y}), M_\epsilon(\bar{\mathbf{y}})) \leq l \|\mathbf{y} - \bar{\mathbf{y}}\| \leq l\sigma, \quad \forall \mathbf{y} \in \hat{N}_{\bar{\mathbf{y}}}.$$

This implies that for each $\mathbf{y} \in \hat{N}_{\bar{\mathbf{y}}}$ and each $\mathbf{x}^B \in M_\epsilon(\mathbf{y})$,

$$d(\mathbf{x}^B, M_\epsilon(\bar{\mathbf{y}})) \equiv \inf\{\|\mathbf{x}^B - \mathbf{x}^A\| : \mathbf{x}^A \in M_\epsilon(\bar{\mathbf{y}})\} \leq l\sigma.$$

Since $M_\epsilon(\bar{y})$ is bounded, define $\tilde{M}_\epsilon(\bar{y})$ to be the closure of $M_\epsilon(\bar{y})$. Then, we have

$$d(\mathbf{x}^B, M_\epsilon(\bar{y})) \equiv \min\{\|\mathbf{x}^B - \tilde{\mathbf{x}}^A\| : \tilde{\mathbf{x}}^A \in \tilde{M}_\epsilon(\bar{y})\} \leq l\sigma.$$

Now, consider an arbitrary $\mathbf{y} \in \hat{N}_{\bar{y}}$ and an arbitrary $\mathbf{x}^B \in M_\epsilon(\mathbf{y})$. We may find a $\tilde{\mathbf{x}}^{A*} \in \tilde{M}_\epsilon(\bar{y})$ such that

$$d(\mathbf{x}^B, M_\epsilon(\bar{y})) \equiv \|\mathbf{x}^B - \tilde{\mathbf{x}}^{A*}\|.$$

Moreover, since $\tilde{M}_\epsilon(\bar{y})$ is bounded, there exists an $\alpha > 0$ such that $\|\tilde{\mathbf{x}}^A\| \leq \alpha, \forall \tilde{\mathbf{x}}^A \in \tilde{M}_\epsilon(\bar{y})$. Then, it follows that

$$\begin{aligned} \|\mathbf{x}^B\| - \|\tilde{\mathbf{x}}^{A*}\| &\leq \|\mathbf{x}^B - \tilde{\mathbf{x}}^{A*}\| \leq l\sigma \\ \|\mathbf{x}^B\| &\leq l\sigma + \|\tilde{\mathbf{x}}^{A*}\| \leq l\sigma + \alpha. \end{aligned}$$

Since both the choices of $\mathbf{y} \in \hat{N}_{\bar{y}}$ and $\mathbf{x}^B \in M_\epsilon(\bar{y})$ are arbitrary above, it implies that M_ϵ is uniformly bounded on $\hat{N}_{\bar{y}}$. Since $M(\mathbf{y}) \subset M_\epsilon(\mathbf{y}), \forall \mathbf{y} \in \hat{N}_{\bar{y}}$, it follows that M is uniformly bounded on $\hat{N}_{\bar{y}}$ too. □

Remark 2 The proof of Lemma 1 implies that in general, if a set-valued mapping is bounded at a point, and if the mapping is also locally Lipschitzian at this point, then the mapping must also be locally bounded.

Lemma 2 Under Assumption 1, the set-valued mapping M in Lemma 1 is closed at \bar{y} as in Definition 6.

Proof Since Condition I.3 holds, for each $\mathbf{y} \in N_{\bar{y}}$,

$$\Omega(\mathbf{y}) \equiv \{\mathbf{x} \in \tilde{X} : \mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0} \text{ and } \mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{0}\}.$$

Moreover, since the functions \mathbf{g} and \mathbf{h} are continuous by assumption, the set-valued mapping Ω is closed at \mathbf{y} due to [22, Theorem A.5].

Since it has been established that M is nonempty near \bar{y} in Lemma 1, we may choose a sequence $\{\mathbf{y}^k\} \subset N_{\bar{y}}$ such that $\lim_{k \rightarrow \infty} \mathbf{y}^k = \bar{y}$ and sequence $\{\mathbf{x}^k\}$ such that $\mathbf{x}^k \in M(\mathbf{y}^k)$ and $\lim_{k \rightarrow \infty} \mathbf{x}^k = \bar{\mathbf{x}}$. Then, we have

$$f(\bar{\mathbf{x}}, \bar{y}) = \lim_{k \rightarrow \infty} f(\mathbf{x}^k, \mathbf{y}^k) = \lim_{k \rightarrow \infty} \phi(\mathbf{y}^k).$$

Since ϕ is continuous near \bar{y} due to Proposition 1 and since $\lim_{k \rightarrow \infty} \mathbf{y}^k = \bar{y}$, we have $\lim_{k \rightarrow \infty} \phi(\mathbf{y}^k) = \phi(\bar{y})$, which implies that $f(\bar{\mathbf{x}}, \bar{y}) = \phi(\bar{y})$. Moreover, since Ω is closed at \bar{y} , it follows that $\bar{\mathbf{x}} \in \Omega(\bar{y})$. Thus, $\bar{\mathbf{x}} \in M(\bar{y})$ and M is closed at \bar{y} . □

Lemma 3 Suppose that Assumption 1 holds. Define the Lagrange dual function $\ell : X \times N_{\bar{y}} \times \mathbb{R}^m \times \mathbb{R}^l \rightarrow \mathbb{R}$ of the parameterized convex NLP in (3) such that for each $(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \in X \times N_{\bar{y}} \times \mathbb{R}^m \times \mathbb{R}^l$,

$$\ell(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}) := f(\mathbf{x}, \mathbf{y}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}, \mathbf{y}) + \boldsymbol{\mu}^T \mathbf{h}(\mathbf{x}, \mathbf{y}).$$

Define $U : N_{\bar{y}} \rightrightarrows \mathbb{R}^{m+l}$ such that for each $\mathbf{y} \in N_{\bar{y}}$, $U(\mathbf{y})$ is the set of dual optimal solutions of the optimization problem in (3), i.e.,

$$U(\mathbf{y}) := \arg \max_{\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathbb{R}^l} o(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \text{ s.t. } \boldsymbol{\lambda} \geq \mathbf{0},$$

where

$$o(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}) := \inf_{\mathbf{x} \in X} \ell(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}).$$

Then, U is closed at $\bar{\mathbf{y}}$, and is nonempty and uniformly bounded near $\bar{\mathbf{y}}$.

Proof This proof proceeds similarly to the proof of [22, Lemma 2].

Since it has been shown in Lemma 1 that the optimal-solution set-valued mapping M is nonempty and uniformly bounded near $\bar{\mathbf{y}}$ and since X is assumed to be open, we may choose a compact set $\hat{X} \subset X$ such that $M(\mathbf{y}) \subset \text{int}(\hat{X})$ for all \mathbf{y} in some neighborhood $\hat{N}_{\bar{\mathbf{y}}} \subset N_{\bar{\mathbf{y}}}$ of $\bar{\mathbf{y}}$. Then, define the following optimal-value function on $\hat{N}_{\bar{\mathbf{y}}}$:

$$\begin{aligned} \hat{\phi}(\mathbf{y}) &:= \min_{\mathbf{x} \in \hat{X}} f(\mathbf{x}, \mathbf{y}) \\ \text{s.t. } &\mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0}, \\ &\mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{0}. \end{aligned} \tag{5}$$

It is straightforward that for each $\mathbf{y} \in \hat{N}_{\bar{\mathbf{y}}}$, $\phi(\mathbf{y}) \equiv \hat{\phi}(\mathbf{y})$, and the optimization problems in (3) and (5) have identical primal optimal solutions. Define $\hat{U} : \hat{N}_{\bar{\mathbf{y}}} \rightrightarrows \mathbb{R}^{m+l}$ such that for each $\mathbf{y} \in \hat{N}_{\bar{\mathbf{y}}}$, $\hat{U}(\mathbf{y})$ is the set of dual optimal solutions of the optimization problem in (5), that is

$$\hat{U}(\mathbf{y}) := \arg \max_{\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathbb{R}^l} \hat{o}(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \text{ s.t. } \boldsymbol{\lambda} \geq \mathbf{0}, \tag{6}$$

where

$$\hat{o}(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}) := \inf_{\mathbf{x} \in \hat{X}} \ell(\mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}). \tag{7}$$

Now, we show that for each $\mathbf{y} \in \hat{N}_{\bar{\mathbf{y}}}$, $U(\mathbf{y}) \equiv \hat{U}(\mathbf{y})$. We will show later that \hat{U} is uniformly bounded near $\bar{\mathbf{y}}$ and is closed at $\bar{\mathbf{y}}$. Since Conditions 1.4, 1.5, 1.6, and 1.7 hold, [20, Theorem 2.3.2] shows that U is bounded at $\bar{\mathbf{y}}$, is nonempty near $\bar{\mathbf{y}}$, and strong duality holds near $\bar{\mathbf{y}}$. Thus, for each $\mathbf{y} \in \hat{N}_{\bar{\mathbf{y}}}$,

$$\phi(\mathbf{y}) \equiv \left(\sup_{\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathbb{R}^l} o(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \text{ s.t. } \boldsymbol{\lambda} \geq \mathbf{0} \right). \tag{8}$$

Similarly to [28, Theorem 1] mentioned in the proof of Proposition 1, [20, Theorem 2.3.2] also requires $X \equiv \mathbb{R}^{n_x}$. However, since Condition 1.3 holds, this result is applicable here. The weak duality theorem [2, Proposition 5.1.3] indicates that for each $\mathbf{y} \in \hat{N}_{\bar{\mathbf{y}}}$,

$$\phi(\mathbf{y}) \equiv \hat{\phi}(\mathbf{y}) \geq \left(\sup_{\boldsymbol{\lambda} \in \mathbb{R}^m, \boldsymbol{\mu} \in \mathbb{R}^l} \hat{o}(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \text{ s.t. } \boldsymbol{\lambda} \geq \mathbf{0} \right). \tag{9}$$

Consider an arbitrary fixed $\mathbf{y} \in \hat{N}_{\bar{\mathbf{y}}}$. Since $\hat{X} \subset X$, we have that

$$\hat{o}(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \geq o(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}), \quad \forall \boldsymbol{\lambda} \geq \mathbf{0}, \forall \boldsymbol{\mu} \in \mathbb{R}^l.$$

Moreover, since (8) holds, for each $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \in U(\mathbf{y})$,

$$\hat{o}(\mathbf{y}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \geq o(\mathbf{y}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \phi(\mathbf{y}).$$

Combining this with (9), we have that $\sup\{\hat{o}(\mathbf{y}, \boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\mu} \in \mathbb{R}^l\} = \phi(\mathbf{y})$ and $U(\mathbf{y}) \subset \hat{U}(\mathbf{y})$.

To arrive a contradiction, assume that $U(\mathbf{y}) \subsetneq \hat{U}(\mathbf{y})$. This implies that there exists a $(\hat{\lambda}, \hat{\mu}) \in \hat{U}(\mathbf{y})$ such that $o(\mathbf{y}, \hat{\lambda}, \hat{\mu}) < \phi(\mathbf{y}) = \hat{o}(\mathbf{y}, \hat{\lambda}, \hat{\mu})$, i.e.,

$$\inf\{\ell(\mathbf{x}, \mathbf{y}, \hat{\lambda}, \hat{\mu}) : \mathbf{x} \in X\} < \inf\{\ell(\mathbf{x}, \mathbf{y}, \hat{\lambda}, \hat{\mu}) : \mathbf{x} \in \hat{X}\}.$$

Thus, there must exist a point $\mathbf{x}^A \in (X \setminus \hat{X})$ such that $\ell(\mathbf{x}^A, \mathbf{y}, \hat{\lambda}, \hat{\mu}) < \ell(\mathbf{x}^*, \mathbf{y}, \hat{\lambda}, \hat{\mu})$, where $\mathbf{x}^* \in (M(\mathbf{y}) \subset \text{int}(\hat{X}))$. Since the mapping $\ell(\cdot, \mathbf{y}, \hat{\lambda}, \hat{\mu})$ is convex, it follows that for any $0 < \gamma < 1$, with $\mathbf{x}^B := \gamma\mathbf{x}^A + (1 - \gamma)\mathbf{x}^*$, we have that $\ell(\mathbf{x}^B, \mathbf{y}, \hat{\lambda}, \hat{\mu}) < \ell(\mathbf{x}^*, \mathbf{y}, \hat{\lambda}, \hat{\mu})$. Moreover, since \mathbf{x}^* minimizes the NLP in (5), [2, Proposition 5.1.1] indicates that \mathbf{x}^* also minimizes $\ell(\cdot, \mathbf{y}, \hat{\lambda}, \hat{\mu})$ on \hat{X} . This implies that \mathbf{x}^B must be not in \hat{X} , which contradicts the fact that $\mathbf{x}^* \in (M(\mathbf{y}) \subset \text{int}(\hat{X}))$. Thus, Combining this with $U(\mathbf{y}) \subset \hat{U}(\mathbf{y})$, we have that $U(\mathbf{y}) \equiv \hat{U}(\mathbf{y})$.

Since \hat{X} is compact, applying [22, Theorem A.1] to (7) yields that \hat{o} is continuous. Since $\hat{U}(\bar{\mathbf{y}}) \equiv U(\bar{\mathbf{y}})$, \hat{U} is also nonempty near $\bar{\mathbf{y}}$ and bounded at $\bar{\mathbf{y}}$. Then, applying [22, Theorem A.4] to (6) yields that \hat{U} is uniformly bounded near $\bar{\mathbf{y}}$. Finally [22, Theorem A.2] shows that \hat{U} is closed at $\bar{\mathbf{y}}$. All these properties also hold for U . □

The following theorem shows that the function ϕ in Assumption 1 is directionally differentiable at $\bar{\mathbf{y}}$, and the directional derivative can be computed by constructing and solving an LP.

Theorem 1 *Under Assumption 1, the directional derivative $\phi'(\bar{\mathbf{y}}; \mathbf{d})$ exists and is finite for all $\mathbf{d} \in \mathbb{R}^{n_y}$, and moreover,*

$$\phi'(\bar{\mathbf{y}}; \mathbf{d}) \equiv [\nabla_{\mathbf{y}} f(\bar{\mathbf{x}}^*, \bar{\mathbf{y}})]^T \mathbf{d} + \psi(\mathbf{d}), \tag{10}$$

where the function $\psi : \mathbb{R}^{n_y} \rightarrow \mathbb{R}$ is defined as follows:

$$\begin{aligned} \psi(\mathbf{d}) &:= \min_{\mathbf{w} \in \mathbb{R}^{n_x}} [\nabla_{\mathbf{x}} f(\bar{\mathbf{x}}^*, \bar{\mathbf{y}})]^T \mathbf{w} \\ \text{s.t. } &[\nabla_{\mathbf{x}} g_i(\bar{\mathbf{x}}^*, \bar{\mathbf{y}})]^T \mathbf{w} \leq -[\nabla_{\mathbf{y}} g_i(\bar{\mathbf{x}}^*, \bar{\mathbf{y}})]^T \mathbf{d}, \quad \forall i \text{ such that } g_i(\bar{\mathbf{x}}^*, \bar{\mathbf{y}}) = 0, \\ &[\nabla_{\mathbf{x}} h_j(\bar{\mathbf{x}}^*, \bar{\mathbf{y}})]^T \mathbf{w} = -[\nabla_{\mathbf{y}} h_j(\bar{\mathbf{x}}^*, \bar{\mathbf{y}})]^T \mathbf{d}, \quad \forall j \in \{1, \dots, l\}. \end{aligned} \tag{11}$$

Proof Under Assumption 1, Lemmata 1, 2, and 3 above show that both the primal optimal solutions mapping M and the dual optimal solutions mapping U are closed at $\bar{\mathbf{y}}$ and are nonempty and uniformly bounded near $\bar{\mathbf{y}}$. Then, we reformulate (3) to a maximization problem and reformulate each equality constraint $h_i(\mathbf{x}, \mathbf{y}) = 0$ as two inequalities. Since X is open, it must be that $M(\bar{\mathbf{y}}) \subset \text{int}(X)$. Then, following the arguments in the proof of [22, Theorem 2 and Corollary 2.1], we obtain that $\phi'(\bar{\mathbf{y}}; \mathbf{d})$ exists and is finite for all $\mathbf{d} \in \mathbb{R}^{n_y}$ and

$$\begin{aligned} \phi'(\bar{\mathbf{y}}; \mathbf{d}) &\equiv \min_{\substack{\mathbf{x} \in M(\bar{\mathbf{y}}), \mathbf{w} \in \mathbb{R}^{n_x}, \\ \mathbf{z}^g \in \mathbb{R}^m, \mathbf{z}^h \in \mathbb{R}^{2l}}} [\nabla_{\mathbf{y}} f(\mathbf{x}, \bar{\mathbf{y}})]^T \mathbf{d} + [\nabla_{\mathbf{x}} f(\mathbf{x}, \bar{\mathbf{y}})]^T \mathbf{w} \\ \text{s.t. } &z_i^g g_i(\mathbf{x}, \bar{\mathbf{y}}) + [\nabla_{\mathbf{x}} g_i(\mathbf{x}, \bar{\mathbf{y}})]^T \mathbf{w} \leq -[\nabla_{\mathbf{y}} g_i(\mathbf{x}, \bar{\mathbf{y}})]^T \mathbf{d}, \quad \forall i \in \{1, \dots, m\}, \\ &z_j^h h_j(\mathbf{x}, \bar{\mathbf{y}}) + [\nabla_{\mathbf{x}} h_j(\mathbf{x}, \bar{\mathbf{y}})]^T \mathbf{w} \leq -[\nabla_{\mathbf{y}} h_j(\mathbf{x}, \bar{\mathbf{y}})]^T \mathbf{d}, \quad \forall j \in \{1, \dots, l\}, \\ &\quad -z_{j+i}^h h_j(\mathbf{x}, \bar{\mathbf{y}}) + [-\nabla_{\mathbf{x}} h_j(\mathbf{x}, \bar{\mathbf{y}})]^T \mathbf{w} \\ &\leq -[-\nabla_{\mathbf{y}} h_j(\mathbf{x}, \bar{\mathbf{y}})]^T \mathbf{d}, \quad \forall j \in \{1, \dots, l\}. \end{aligned} \tag{12}$$

Since $M(\bar{\mathbf{y}}) := \{\bar{\mathbf{x}}^*\}$ as in Condition I.2, (12) reduces to (10) and (11). □

Remark 3 The proof above utilizes arguments of [22, Theorem 2 and Corollary 2.1], and these results’ statements require X to be closed and convex, and require a parameterized convex program with only inequality constraints and the Slater conditions must hold. However, these requirements are only to guarantee that the mappings M and U in [22]’s case have the properties as in Lemmata 1, 2, and 3. Since these properties have already been established in our case without satisfying [22]’s conditions, it is verified that the arguments in the proof of [22, Theorem 2 and Corollary 2.1] are sufficient to obtain (12).

Remark 4 Theorem 1 provides a characterization of the directional derivative of parameterized convex NLPs with both inequality and equality constraints. This is a nontrivial generalization of [22, Theorem 2 and Corollary 2.1] which only deal with parameterized convex NLPs with inequality constraints, and deal with a closed X . We allow an open X by imposing Condition 1.3, which opens up the theory to a broader class of functions in practice. The main difficulties for including affine equality constraints is to establish the uniform boundedness of M near $\bar{\mathbf{y}}$ and the closedness of M at $\bar{\mathbf{y}}$. In [22], the justifications of these properties (c.f. Theorems A.2 and A.4) rely heavily on the fact that the decision space mapping Ω is open at $\bar{\mathbf{y}}$ (in the sense of Hogan [22, 23]). This is true for Ω only defined by inequality constraints (see [23, Theorem 12]), but does not necessarily hold if also considering equality constraints. In our work, based on a later result [28, Theorem 1], Lemmata 1 and 2 establish these properties without requiring Ω to be open at $\bar{\mathbf{y}}$.

Observe that the function ψ in (10) is an LP parameterized by the constraints’ right-hand side. To proceed, we reformulate the optimization problem of $\psi(\mathbf{d})$ for each $\mathbf{d} \in \mathbb{R}^{n_y}$ to a standard form of LP (c.f. [3]). Let $n_{\bar{c}}$ be the number of active constraints of (3) for $\mathbf{y} := \bar{\mathbf{y}}$ at $\bar{\mathbf{x}}^*$, and then we have for each $\mathbf{d} \in \mathbb{R}^{n_y}$ that

$$\psi(\mathbf{d}) \equiv q(\boldsymbol{\beta}(\mathbf{d})) \tag{13}$$

where $q : \mathbb{R}^{n_{\bar{c}}} \rightarrow \mathbb{R} \cup \{+\infty, -\infty\}$ such that

$$\begin{aligned} q(\mathbf{b}) &:= \inf_{\mathbf{v}} \mathbf{c}^T \mathbf{v} \\ \text{s.t. } &\mathbf{A}\mathbf{v} = \mathbf{b}, \\ &\mathbf{v} \geq \mathbf{0}, \end{aligned}$$

with appropriate dimensions for \mathbf{c} , \mathbf{v} and \mathbf{A} , and $\boldsymbol{\beta} : \mathbb{R}^{n_y} \rightarrow \mathbb{R}^{n_{\bar{c}}} : \mathbf{d} \mapsto \mathbf{G}\mathbf{d}$ is a linear function where $\mathbf{G} \in \mathbb{R}^{n_{\bar{c}} \times n_y}$ is the stacked $[-\nabla_{\mathbf{y}} g_i(\bar{\mathbf{x}}^*, \bar{\mathbf{y}})]^T$ for all i such that $g_i(\bar{\mathbf{x}}^*, \bar{\mathbf{y}}) = 0$ and all $[-\nabla_{\mathbf{y}} h_j(\bar{\mathbf{x}}^*, \bar{\mathbf{y}})]^T$. With this preparation, the following proposition shows that ψ is L -smooth whose LD-derivatives can be computed by solving a sequence of LPs.

Proposition 2 Consider the functions ψ , q , $\boldsymbol{\beta}$ as in (13). Then, the function ψ is L -smooth, and for any $\mathbf{d} \in \mathbb{R}^{n_y}$ and any $\mathbf{K} \in \mathbb{R}^{n_y \times r}$, the following sequence of functions is well-defined: $\forall j \in \{0, \dots, r\}$,

$$\begin{aligned} \psi_{\mathbf{d}, \mathbf{K}}^{(j)} : \mathbb{R}^{n_y} \rightarrow \mathbb{R} : \tilde{\mathbf{d}} \mapsto \max_{\boldsymbol{\lambda} \in \mathbb{R}^{\bar{c}}} [\mathbf{G}\tilde{\mathbf{d}}]^T \boldsymbol{\lambda} \\ \text{s.t. } \begin{bmatrix} \mathbf{A}^T \\ -[\boldsymbol{\beta}(\mathbf{d})]^T \\ -[\mathbf{G}\mathbf{k}_{(1)}]^T \\ \vdots \\ -[\mathbf{G}\mathbf{k}_{(j)}]^T \end{bmatrix} \boldsymbol{\lambda} \leq \begin{bmatrix} \mathbf{c} \\ -\psi(\mathbf{d}) \\ -\psi_{\mathbf{d}, \mathbf{M}}^{(0)}(\mathbf{k}_{(1)}) \\ \vdots \\ -\psi_{\mathbf{d}, \mathbf{M}}^{(j-1)}(\mathbf{k}_{(j)}) \end{bmatrix}, \end{aligned} \tag{14}$$

where for $j = 0$, the constraint $-\mathbf{[Gk}_{(j)}]^T \boldsymbol{\lambda} \leq -\psi_{\mathbf{d}, \mathbf{K}}^{(j-1)}(\mathbf{k}_{(j)})$ is understood to be vacuous. Moreover, associated to the sequence (14), define a sequence of sets $D_{\mathbf{d}, \mathbf{K}}^{(j)} \subset \mathbb{R}^{n_c}$ such that for each $j \in \{0, \dots, r - 1\}$, $D_{\mathbf{d}, \mathbf{K}}^{(j)}$ is the optimal solution set of the LP for evaluating $\psi_{\mathbf{d}, \mathbf{K}}^{(j)}(\mathbf{k}_{(j+1)})$ in (14). Then, it follows that

$$D_{\mathbf{d}, \mathbf{K}}^{(0)} := \arg \max_{\boldsymbol{\lambda} \in \Lambda(\mathbf{Gd})} [\mathbf{Gk}_{(1)}]^T \boldsymbol{\lambda}$$

where $\Lambda(\mathbf{Gd})$ is the optimal solution set of the dual problem (c.f. [3]) of $q(\mathbf{Gd})$, and for each $j \in \{1, \dots, r - 1\}$,

$$D_{\mathbf{d}, \mathbf{K}}^{(j)} := \left(\arg \max_{\boldsymbol{\lambda} \in D_{\mathbf{d}, \mathbf{K}}^{(j-1)}} [\mathbf{Gk}_{(j+1)}]^T \boldsymbol{\lambda} \right) \subset D_{\mathbf{d}, \mathbf{K}}^{(j-1)}. \tag{15}$$

Furthermore, the LD-derivative of ψ at \mathbf{d} in the directions \mathbf{K} is given by

$$\psi'(\mathbf{d}; \mathbf{K}) := \boldsymbol{\lambda}^T \mathbf{GK}, \quad \text{for any } \boldsymbol{\lambda} \in D_{\mathbf{d}, \mathbf{K}}^{(r-1)}.$$

Proof From Theorem 1, ψ is finite for any $\mathbf{d} \in \mathbb{R}^{n_y}$, i.e., $F \equiv \mathbb{R}^{n_y}$ where $F := \{\mathbf{d} \in \mathbb{R}^{n_y} : -\infty < \psi(\mathbf{d}) < +\infty\}$, and thus $\mathbf{d} \in \text{int}(F)$. Moreover, since $\boldsymbol{\beta}$ is linear and (13) holds, the claimed results follow from [21, Proposition 3.7]. \square

Remark 5 Analogously to [21, Remark 3.4], evaluating $\psi'(\mathbf{d}; \mathbf{K})$ requires solving at most n_y LPs, because $\psi_{\mathbf{d}, \mathbf{K}}^{(k)}$ is guaranteed to be linear if the columns $\{\mathbf{k}_{(1)}, \dots, \mathbf{k}_{(k)}\}$ span \mathbb{R}^{n_y} . However, if at a $i < n_y$ such that $D_{\mathbf{d}, \mathbf{K}}^{(i)}$ is a singleton, then (15) implies that $D_{\mathbf{d}, \mathbf{K}}^{(i)} \equiv D_{\mathbf{d}, \mathbf{K}}^{(i+1)} \equiv \dots \equiv D_{\mathbf{d}, \mathbf{K}}^{(r-1)}$; no need to solve more LPs. Moreover, since the computational complexity of solving each LP grows polynomially with the size of the problem, computational complexity for evaluation of the LD-derivative also grows polynomially with the size of the problem.

Motivated by Remark 5, we employ the following result from [33] to examine whether an obtained optimal solution for an LP is unique or not. This result examines LP solutions' uniqueness by constructing and solving an auxiliary LP. This result provides a practical way to terminate solving the sequence of LPs in Proposition 2 effectively when one LP in the sequence turns out to have a unique solution.

Proposition 3 (Adapted from [33]) *Consider an LP of the general form:*

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{p}^T \mathbf{x} \\ \text{s.t.} \quad & \tilde{\mathbf{A}}\mathbf{x} = \tilde{\mathbf{b}}, \quad \leftarrow \text{multipliers } \mathbf{u} \\ & \tilde{\mathbf{C}}\mathbf{x} \geq \tilde{\mathbf{d}}, \quad \leftarrow \text{multipliers } \mathbf{v}. \end{aligned} \tag{16}$$

Let \mathbf{x}^* be a primal optimal solution and $(\mathbf{u}^*, \mathbf{v}^*)$ be a dual optimal solution of (16). Let $\tilde{\mathbf{c}}_i$ denote the i^{th} row of $\tilde{\mathbf{C}}$. Define index sets

$$\begin{aligned} \mathcal{K} &:= \{i : \tilde{\mathbf{c}}_i \mathbf{x}^* = \tilde{d}_i \text{ and } v_i^* > 0\}, \\ \text{and } \mathcal{L} &:= \{i : \tilde{\mathbf{c}}_i \mathbf{x}^* = \tilde{d}_i \text{ and } v_i^* = 0\}. \end{aligned}$$

Let $\tilde{\mathbf{C}}_{\mathcal{K}}$ and $\tilde{\mathbf{C}}_{\mathcal{L}}$ denote matrices whose rows are $\tilde{\mathbf{c}}_i, \forall i \in \mathcal{K}$ and are $\tilde{\mathbf{c}}_i, \forall i \in \mathcal{L}$, respectively. Consider the following LP:

$$\begin{aligned} \max_{\mathbf{x}} \quad & \mathbf{e}^T \tilde{\mathbf{C}}_{\mathcal{L}} \mathbf{x} \\ \text{s.t.} \quad & \tilde{\mathbf{A}} \mathbf{x} = \mathbf{0}, \\ & \tilde{\mathbf{C}}_{\mathcal{K}} \mathbf{x} = \mathbf{0}, \\ & \tilde{\mathbf{C}}_{\mathcal{L}} \mathbf{x} \geq \mathbf{0}, \end{aligned} \tag{17}$$

where \mathbf{e} is a vector of appropriate dimension with all components being one. Then, \mathbf{x}^* is a unique optimal solution of (16) if and only if the optimal objective value of (17) is zero. For any LP of the form (16), (17) will be called its (optimal) solutions’ uniqueness examination LP.

Proof This result follows from [33, Theorem 2 and Remark 2]. □

The following theorem shows that the function ϕ defined in (3) is L-smooth at $\bar{\mathbf{y}}$, whose LD-derivatives can be evaluated by solving a sequence of LPs.

Theorem 2 Suppose that Assumption 1 holds, and consider the formulation of the directional derivative $\phi'(\bar{\mathbf{y}}; \mathbf{d})$ established in Theorem 1 and the sequence of functions $\psi_{\mathbf{d}, \mathbf{K}}^{(j)}$ and sets $D_{\mathbf{d}, \mathbf{K}}^{(j)}$ defined in Proposition 2. Then, the function ϕ is L-smooth at $\bar{\mathbf{y}}$, and the following statements hold for any $\mathbf{M} \in \mathbb{R}^{n_y \times p}$:

II.1 The following sequence of functions

$$\begin{aligned} \phi_{\bar{\mathbf{y}}, \mathbf{M}}^{(0)} : \mathbb{R}^{n_y} &\rightarrow \mathbb{R} : \mathbf{d} \mapsto \phi'(\bar{\mathbf{y}}; \mathbf{d}), \\ \phi_{\bar{\mathbf{y}}, \mathbf{M}}^{(j)} : \mathbb{R}^{n_y} &\rightarrow \mathbb{R} : \mathbf{d} \mapsto [\phi_{\bar{\mathbf{y}}, \mathbf{M}}^{(j-1)}]'(\mathbf{m}_{(j)}; \mathbf{d}), \quad \forall j \in \{1, \dots, p\} \end{aligned}$$

are well-defined and given by:

$$\begin{aligned} \phi_{\bar{\mathbf{y}}, \mathbf{M}}^{(0)}(\mathbf{d}) &:= \phi'(\bar{\mathbf{y}}; \mathbf{d}) \text{ is given in (10),} \\ \phi_{\bar{\mathbf{y}}, \mathbf{M}}^{(j)}(\mathbf{d}) &:= [\nabla_{\mathbf{y}} f(\bar{\mathbf{x}}^*, \bar{\mathbf{y}})]^T \mathbf{d} + \psi_{\mathbf{m}_{(1)}, \mathbf{M}_{(2):(p)}}^{(j-1)}(\mathbf{d}), \quad \forall j \in \{1, \dots, p\}, \end{aligned}$$

where $\mathbf{M}_{(2):(p)}$ denotes $[\mathbf{m}_{(2)} \ \dots \ \mathbf{m}_{(p)}]$.

II.2 The LD derivative $\phi'(\bar{\mathbf{y}}; \mathbf{M})$ is given by

$$\phi'(\bar{\mathbf{y}}; \mathbf{M}) := [\phi'(\bar{\mathbf{y}}; \mathbf{m}_{(1)}) (\nabla_{\mathbf{y}} f(\bar{\mathbf{x}}^*, \bar{\mathbf{y}})]^T + \lambda^T \mathbf{G} \mathbf{M}_{(2):(p)} \tag{18}$$

for any $\lambda^T \in D_{\mathbf{m}_{(1)}, \mathbf{M}_{(2):(p)}}^{(p-2)}$.

Proof Observe that the term $[\nabla_{\mathbf{y}} f(\bar{\mathbf{x}}^*, \bar{\mathbf{y}})]^T \mathbf{d}$ in (10) is affine w.r.t. \mathbf{d} , and then Statement II.1 follows from Theorem 1 and Proposition 2. Moreover, since Proposition 1 holds and according to Definition 3, ϕ is L-smooth at $\bar{\mathbf{y}}$. Then, Statement II.2 again follows from Theorem 1 and Proposition 2. □

With the theoretical results established above, the following algorithm computes an LD-derivative of ϕ at $\bar{\mathbf{y}}$ in Assumption 1 by solving a sequence of LPs.

Remark 6 In the best case, the algorithm above requires solving two LPs to evaluate $\phi'(\bar{\mathbf{y}}; \mathbf{M})$, namely the LP in (13) for evaluating $\psi(\mathbf{m}_{(1)})$ and the solutions’ uniqueness examination LP

Algorithm 1 Computes an LD-derivative $\phi'(\bar{\mathbf{y}}; \mathbf{M})$ of ϕ at $\bar{\mathbf{y}}$ in Assumption 1

Require: $\phi(\bar{\mathbf{y}})$ and $\bar{\mathbf{x}}^*$ as in Assumption 1, $\mathbf{M} := [\mathbf{m}_{(1)} \cdots \mathbf{m}_{(p)}] \in \mathbb{R}^{n_y \times p}$.
 Compute $\psi(\mathbf{m}_{(1)})$ by solving the LP in (13) with $\mathbf{d} := \mathbf{m}_{(1)}$,
 and set $\mathbf{v}^* \leftarrow$ a primal optimal solution, $\lambda^* \leftarrow$ a dual optimal solution
 Use $(\lambda^*, \mathbf{v}^*)$ to construct and solve the solutions’ uniqueness examination LP of the dual problem of this LP,
 and set $t \leftarrow$ the optimal objective value
 Compute $\phi'(\bar{\mathbf{y}}; \mathbf{m}_{(1)})$ defined in (10)
if $t \equiv 0$ **then**
 Compute $\phi'(\bar{\mathbf{y}}; \mathbf{M})$ by (18) with $\lambda := \lambda^*$
 return $\phi'(\bar{\mathbf{y}}; \mathbf{M})$
else
 continue
end if
 Set $j \leftarrow 1$.
while $j < p$ **do**
 Compute $\psi_{\mathbf{d}, \mathbf{K}}^{(j-1)}(\mathbf{k}_{(j)})$ by solving the LP in (14) with $\mathbf{d} := \mathbf{m}_{(1)}$, $\mathbf{K} := \mathbf{M}_{(2):(p)}$, and
 $\bar{\mathbf{d}} := \mathbf{k}_{(j)}$, and set $\lambda^* \leftarrow$ a primal optimal solution, $\eta^* \leftarrow$ a dual optimal solution
 if $j < p - 1$ **then**
 Use (λ^*, η^*) to construct and solve the solutions’ uniqueness examination LP of
 the LP defined for $\psi_{\mathbf{d}, \mathbf{K}}^{(j-1)}(\mathbf{k}_{(j)})$, and set $t \leftarrow$ the optimal objective value
 if $t \equiv 0$ **then**
 break
 else
 continue
 end if
 else
 continue
 end if
 Set $j \leftarrow j + 1$
end while
 Compute $\phi'(\bar{\mathbf{y}}; \mathbf{M})$ by (18) with $\lambda := \lambda^*$
return $\phi'(\bar{\mathbf{y}}; \mathbf{M})$

of the dual problem of this LP. If this solutions’ uniqueness examination LP has a maximum zero, i.e., when the dual problem of the LP in (13) has a unique solution, then Proposition 2 implies that this solution must solve all LPs in the sequence defined in (14). Thus, this solution can be used directly to evaluate $\phi'(\bar{\mathbf{y}}; \mathbf{M})$ using (18), as the algorithm above shows. We also note that Algorithm 1 requires solving at most $(2p - 1)$ LPs, of which the computational cost grows polynomially with p .

We believe that the best case mentioned above may happen frequently in practice for the following reasons. Since ϕ is locally Lipschitz continuous, Rademacher’s theorem ([19, Theorem 3.1]) implies that ϕ is differentiable almost everywhere in the Lebesgue sense. Note that the non-differentiable points of measure-zero are indeed reachable during problem-solving on a finite precision machine (c.f. [1, Sect. 4]). If ϕ is differentiable at $\bar{\mathbf{y}}$, then [37, Proposition 1] implies that the function ψ in (10) must be linear. Since ψ is an optimal-value function with embedded LP parameterized by its constraints’ right-hand side vector, [3, Theorem 5.2] shows that the subdifferential (in the sense of convex analysis [42]) of ψ at any $\mathbf{d} \in \mathbb{R}^{n_y}$ is identical to the dual optimal solutions set of the corresponding LP. Thus, for any $\mathbf{m}_{(1)} \in \mathbb{R}^{n_y}$, the dual optimal solutions set of the LP in (13) with $\mathbf{d} := \mathbf{m}_{(1)}$ must be a singleton, and Algorithm 1 computes an LD-derivative by solving two LPs as discussed.

Then, an L-derivative, which reduces to the Jacobian of ϕ at $\bar{\mathbf{y}}$ in this case, can be readily computed via (2).

To implement Algorithm 1, an LP solver that can return both primal and dual optimal solutions is favorable; this is hardly a problem for prevalent LP solvers such as CPLEX [24]. If such solver is unavailable, then extra efforts need to be taken to compute a dual optimal solution. Given a primal optimal solution, a practical approach may be to identify an extreme point of the dual optimal solutions set represented by the KKT conditions.

3.2 Joint convexity

In this subsection, we show that if the functions f , \mathbf{g} , and \mathbf{h} in Assumption 1 are jointly convex with respect to (\mathbf{x}, \mathbf{y}) . Then, with Condition I.2 being relaxed, we can still calculate an LD-derivative of ϕ using Algorithm 1 with easy modifications. Such *parameterized convex NLPs with joint convexity* are formalized as follows.

Assumption 2 Consider the setup in Assumption 1, suppose that Condition I.5 is strengthened to:

III.1 f , \mathbf{g} , and \mathbf{h} are convex on $X \times N_{\bar{\mathbf{y}}}$,
and Condition I.2 is relaxed to

III.2 $M(\bar{\mathbf{y}})$ is nonempty and bounded, where M is the primal solution mapping defined in Lemma 1.

Then, (3) is a *parameterized convex NLP with joint convexity*.

Roughly, under Assumption 2, Theorem 2 still holds with the unique optimal solution $\bar{\mathbf{x}}^*$ replaced by any $\hat{\mathbf{x}} \in M(\bar{\mathbf{y}})$ for evaluating ψ , and thus Algorithm 1 is valid for this joint convexity case with $\bar{\mathbf{x}}^*$ replaced by $\hat{\mathbf{x}}$. This result requires significantly simpler justification than in the previous section, due to the joint convexity of f , \mathbf{g} , and \mathbf{h} . When the optimal solutions are non-unique, this result (which is formalized below) is particularly beneficial for implementation, since any optimal solution of (3) obtained by numerical solver can be used directly to evaluate an LD-derivative. Note that since the optimal solution is non-unique here, the methods in [48, 49] cannot be applied to compute an LD-derivative of ϕ .

Theorem 3 Suppose that Assumption 2 holds, and consider an arbitrary $\hat{\mathbf{x}} \in M(\bar{\mathbf{y}})$. Then, Algorithm 1 can be used for computing an LD-derivative $\phi'(\bar{\mathbf{y}}; \mathbf{M})$ with $\hat{\mathbf{x}}$ in place of $\bar{\mathbf{x}}^*$ wherever $\bar{\mathbf{x}}^*$ appears.

Proof Since Condition III.2 holds, following the same arguments in the proof of Proposition 1, we have ϕ is locally Lipschitz continuous at $\bar{\mathbf{y}}$.

Define a set

$$V := \{\mathbf{y} \in N_{\bar{\mathbf{y}}} : \exists \mathbf{x} \in X, \mathbf{g}(\mathbf{x}, \mathbf{y}) < 0 \text{ and } \mathbf{h}(\mathbf{x}, \mathbf{y}) = 0\}.$$

Then, Condition I.7 implies that $\bar{\mathbf{y}} \in \text{int}(V)$. Since Assumption 2 holds, by reformulating each equality constraint $h_i(\mathbf{x}, \mathbf{y}) = 0$ as two inequalities and applying [22, Corollaries 3.2 and 3.1], we have that for each $\mathbf{d} \in \mathbb{R}^{n_y}$, $\phi'(\bar{\mathbf{y}}; \mathbf{d})$ is given by (10) and (11) with $\hat{\mathbf{x}}$ in place of $\bar{\mathbf{x}}^*$.

Then, Proposition 2 and Theorem 2 follow immediately with $\hat{\mathbf{x}}$ in place of $\bar{\mathbf{x}}^*$. Consequently, Algorithm 1 is valid for this joint convexity case with $\hat{\mathbf{x}}$ in place of $\bar{\mathbf{x}}^*$. □

Remark 7 In the joint convexity case, ϕ is convex on $N_{\tilde{y}}$ due to [13, Corollary 2.1]. Thus, an L-derivative is a subgradient in the context of convex analysis [42].

3.3 Parameterized LP’s objective function

In this subsection, we propose a new method for computing LD-derivatives of optimal-value functions with embedded an LP whose parameters appear in the objective function. Similarly to the methods in the previous subsections, this new method also computes an LD-derivative by solving a sequence of LPs. The following assumption formalizes such parameterized LP.

Assumption 3 Let $Y \subset \mathbb{R}^{n_y}$ be open and consider a function $\mathbf{c} : Y \rightarrow \mathbb{R}^{n_x}$. Consider a parameterized LP of the form:

$$\begin{aligned} \phi(\mathbf{y}) &:= \min_{\mathbf{x} \in \mathbb{R}^{n_x}} [\mathbf{c}(\mathbf{y})]^T \mathbf{x} \\ \text{s.t. } &\mathbf{Ax} = \boldsymbol{\alpha}, \\ &\mathbf{Bx} \leq \boldsymbol{\beta}, \end{aligned} \tag{19}$$

with appropriate dimensions for the matrices (\mathbf{A}, \mathbf{B}) and the vectors $(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Suppose that the following conditions hold:

- IV.1 the decision space $X := \{\mathbf{x} \in \mathbb{R}^{n_x} : \mathbf{Ax} = \boldsymbol{\alpha} \text{ and } \mathbf{Bx} \leq \boldsymbol{\beta}\}$ is compact, and
- IV.2 the function \mathbf{c} is continuously differentiable.

Unlike the partial and joint convexity cases in the previous subsections, here we do not assume the Slater conditions for the decision space X , and do not assume uniqueness of an optimal solution of the LP. Instead, X is assumed to be compact. This requirement is guaranteed to be satisfied if the decision variables \mathbf{x} have known lower and upper bounds. Moreover, for a well-posed problem where $\phi(\mathbf{y})$ is finite for each $\mathbf{y} \in Y$, even if X is not compact, one can always impose sufficiently large bounds on \mathbf{x} to make the decision space compact, yet without affecting the value of $\phi(\mathbf{y})$.

We first introduce a classical result by Danskin [8] for describing directional derivatives of optimal-value functions with parameter-independent decision space.

Proposition 4 (Originally from [8], summarized in [22]) *Let $\tilde{X} \subset \mathbb{R}^n$ and $\tilde{Y} \subset \mathbb{R}^m$ be open, and let $\Xi \subset \tilde{X}$ be compact. Consider a function $f : \tilde{X} \times \tilde{Y} \rightarrow \mathbb{R}$ of (\mathbf{x}, \mathbf{y}) , for which suppose that f and the partial derivative mapping $\nabla_{\mathbf{y}} f : \tilde{X} \times \tilde{Y} \rightarrow \mathbb{R}^m$ are continuous. Consider an optimal-value function $v : \tilde{Y} \rightarrow \mathbb{R}$ such that for each $\mathbf{y} \in \tilde{Y}$,*

$$v(\mathbf{y}) := \min_{\mathbf{x} \in \Xi} f(\mathbf{x}, \mathbf{y}).$$

Then, v is directionally differentiable on \tilde{Y} , and for each $\mathbf{y} \in \tilde{Y}$ and $\mathbf{d} \in \mathbb{R}^m$,

$$v'(\mathbf{y}; \mathbf{d}) := \min_{\mathbf{x} \in M(\mathbf{y})} [\nabla_{\mathbf{y}} f(\mathbf{x}, \mathbf{y})]^T \mathbf{d}, \tag{20}$$

where $M(\mathbf{y}) := \{\mathbf{x} \in \Xi : v(\mathbf{y}) = f(\mathbf{x}, \mathbf{y})\}$ is the optimal solutions set of the optimization problem in (20).

The following result provides a method for computing the directional derivatives of ϕ in Assumption 3, by applying Theorem 4.

Proposition 5 *Consider the setup in Assumption 3. Then, for each $\mathbf{y} \in Y$ and each $\mathbf{d} \in \mathbb{R}^{n_y}$, the directional derivative $\phi'(\mathbf{y}; \mathbf{d})$ is well-defined and given by*

$$\begin{aligned} \phi'(\mathbf{y}; \mathbf{d}) &:= \min_{\mathbf{x} \in \mathbb{R}^{n_x}} [\mathbf{Jc}(\mathbf{y})\mathbf{d}]^T \mathbf{x} \\ \text{s.t. } \mathbf{Ax} &= \boldsymbol{\alpha}, \\ \mathbf{Bx} &\leq \boldsymbol{\beta}, \\ [\mathbf{c}(\mathbf{y})]^T \mathbf{x} &= \phi(\mathbf{y}). \end{aligned} \tag{21}$$

Proof Under Assumption 3, this result follows immediately from Theorem 4. □

Observe that the decision space of the optimization problem in (21) is \mathbf{d} -independent and is compact, and the mapping $\mathbf{d} \mapsto \mathbf{Jc}(\mathbf{y})\mathbf{d}$ is linear. This enables using Proposition 4 to compute the directional derivatives of the mapping $\mathbf{d} \mapsto \phi'(\mathbf{y}; \mathbf{d})$. Based on this intuition, the following result shows that the function ϕ in Assumption 3 is L-smooth, and moreover, is piecewise differentiable. Besides, for any $\mathbf{y} \in Y$ and directions matrix $\mathbf{M} \in \mathbb{R}^{n_y \times p}$, the LD-derivative $\phi(\mathbf{y}; \mathbf{M})$ can be computed by solving a sequence of LPs.

Theorem 4 *Consider the setup in Assumption 3. Then, the function ϕ is piecewise differentiable and is L-smooth on Y . Moreover, for each $\mathbf{y} \in Y$ and each $\mathbf{M} \in \mathbb{R}^{n_y \times p}$, the following sequence of functions*

$$\begin{aligned} \phi_{\mathbf{y}, \mathbf{M}}^{(0)} &: \mathbb{R}^{n_y} \rightarrow \mathbb{R} : \mathbf{d} \mapsto \phi'(\mathbf{y}; \mathbf{d}), \\ \phi_{\mathbf{y}, \mathbf{M}}^{(j)} &: \mathbb{R}^{n_y} \rightarrow \mathbb{R} : \mathbf{d} \mapsto [\phi_{\mathbf{y}, \mathbf{M}}^{(j-1)}]'(\mathbf{m}_{(j)}; \mathbf{d}), \quad \forall j \in \{1, \dots, p\} \end{aligned}$$

are well-defined and given by: for each $j \in \{0, \dots, p\}$,

$$\begin{aligned} \phi_{\mathbf{y}, \mathbf{M}}^{(j)}(\mathbf{d}) &:= \min_{\mathbf{x} \in \mathbb{R}^{n_x}} [\mathbf{Jc}(\mathbf{y})\mathbf{d}]^T \mathbf{x} \\ \text{s.t. } \mathbf{Ax} &= \boldsymbol{\alpha}, \\ \mathbf{Bx} &\leq \boldsymbol{\beta}, \\ [\mathbf{c}(\mathbf{y})]^T \mathbf{x} &= \phi(\mathbf{y}), \\ [\mathbf{Jc}(\mathbf{y})\mathbf{m}_{(1)}]^T \mathbf{x} &= \phi_{\mathbf{y}, \mathbf{M}}^{(0)}(\mathbf{m}_{(1)}), \\ &\vdots \\ [\mathbf{Jc}(\mathbf{y})\mathbf{m}_{(j)}]^T \mathbf{x} &= \phi_{\mathbf{y}, \mathbf{M}}^{(j-1)}(\mathbf{m}_{(j)}), \end{aligned} \tag{22}$$

where the constraint $[\mathbf{Jc}(\mathbf{y})\mathbf{m}_{(1)}]^T \mathbf{x} = \phi_{\mathbf{y}, \mathbf{M}}^{(0)}(\mathbf{m}_{(1)})$ is vacuous for $j = 0$. Associated to the sequence (22), define a sequence of sets $D_{\mathbf{y}, \mathbf{M}}^{(j)} \subset \mathbb{R}^{n_x}$ such that for each $j \in \{0, \dots, p - 1\}$, $D_{\mathbf{y}, \mathbf{M}}^{(j)}$ is the optimal solution set of the LP for evaluating $\phi_{\mathbf{y}, \mathbf{M}}^{(j)}(\mathbf{m}_{(j+1)})$ in (22). Then, it follows that

$$D_{\mathbf{y}, \mathbf{M}}^{(0)} := \arg \max_{\mathbf{x} \in M(\mathbf{y})} [\mathbf{Jc}(\mathbf{y})\mathbf{m}_{(1)}]^T \mathbf{x} \tag{23}$$

where $M(\mathbf{y})$ is the optimal solution set of the LP in (19), and for each $j \in \{1, \dots, p - 1\}$,

$$D_{\mathbf{y}, \mathbf{M}}^{(j)} := \left(\arg \max_{\mathbf{x} \in D_{\mathbf{y}, \mathbf{M}}^{(j-1)}} [\mathbf{Jc}(\mathbf{y})\mathbf{m}_{(j+1)}]^T \mathbf{x} \right) \subset D_{\mathbf{y}, \mathbf{M}}^{(j-1)}. \tag{24}$$

Furthermore, an LD-derivative $\phi'(\mathbf{y}; \mathbf{M})$ is given by

$$\phi'(\mathbf{y}; \mathbf{M}) := \mathbf{x}^T \mathbf{Jc}(\mathbf{y})\mathbf{M}, \quad \text{for any } \mathbf{x} \in D_{\mathbf{y}, \mathbf{M}}^{(p-1)}. \tag{25}$$

Proof Consider a function $\tilde{\phi} : \mathbb{R}^{n_x} \rightarrow \mathbb{R}$:

$$\begin{aligned} \tilde{\phi}(\tilde{\mathbf{c}}) &:= \min_{\mathbf{x} \in \mathbb{R}^{n_x}} [\tilde{\mathbf{c}}]^T \mathbf{x} \\ \text{s.t. } &\mathbf{Ax} = \boldsymbol{\alpha}, \\ &\mathbf{Bx} \leq \boldsymbol{\beta}. \end{aligned}$$

[3, Sect. 5.4] shows that $\tilde{\phi}$ is a piecewise linear concave function. Since \mathbf{c} is continuously differentiable and $\phi \equiv \tilde{\phi} \circ \mathbf{c}$, ϕ is piecewise differentiable according to [47, Sect. 4.1]. Then, it follows that ϕ is L-smooth according to [27, Proposition 3.1].

The function $\phi_{\mathbf{y}, \mathbf{M}}^{(1)}$ is obtained by applying Proposition 4 to $\phi_{\mathbf{y}, \mathbf{M}}^{(0)} \equiv \phi'(\mathbf{y}; \cdot)$ established in Proposition 5. Then, the rest functions in the sequence (22) are obtained by applying Proposition 4 inductively.

The relations (23) and (24) follow immediately from the structure of the sequence of functions (22).

Consider an arbitrary $\mathbf{x} \in D_{\mathbf{y}, \mathbf{M}}^{(p-1)}$. Since (24) holds and following from [27, Definition 2.1],

$$\begin{aligned} \phi'(\mathbf{y}; \mathbf{M}) &= [\phi_{\mathbf{y}, \mathbf{M}}^{(0)}(\mathbf{m}_{(1)}) \phi_{\mathbf{y}, \mathbf{M}}^{(1)}(\mathbf{m}_{(2)}) \cdots \phi_{\mathbf{y}, \mathbf{M}}^{(p-1)}(\mathbf{m}_{(p)})] \\ &= [\mathbf{x}^T \mathbf{Jc}(\mathbf{y})\mathbf{m}_{(1)} \mathbf{x}^T \mathbf{Jc}(\mathbf{y})\mathbf{m}_{(2)} \cdots \mathbf{x}^T \mathbf{Jc}(\mathbf{y})\mathbf{m}_{(p)}] \\ &= \mathbf{x}^T \mathbf{Jc}(\mathbf{y})\mathbf{M}. \end{aligned}$$

□

Theorem 4 shows that under Assumption 3, an LD-derivative $\phi'(\mathbf{y}; \mathbf{M})$ can be computed by solving a sequence of LPs defined in (22). Since (24) holds, similarly to Remark 5, if one LP in the sequence has a unique solution, then there is no need to solve the rest LPs in the sequence. Thus, the method for examining the uniqueness of LPs’ optimal solutions described in Proposition 3 is also useful here. Embedded with this solutions’ uniqueness examination method, the complete algorithm for computing $\phi'(\mathbf{y}; \mathbf{M})$ in Assumption 3 is presented below.

If the original LP in (19) for some $\bar{\mathbf{y}} \in Y$ has a unique solution, then [5, Proposition 1.13 and Theorem 2.1] imply that ϕ is differentiable at $\bar{\mathbf{y}}$. In this case, Algorithm 2 only requires solving one LP to furnish the desired LD-derivative, namely the solutions’ uniqueness examination LP of this original LP. Similarly to the discussion after Algorithm 1, we expect this to happen frequently in practice, since ϕ is differentiable almost everywhere. Algorithm 2 requires solving at most $2p$ LPs.

Now, we compare Algorithm 2 to an established method [5] which can be used to compute a generalized Jacobian element of ϕ in Assumption 3. According to [5, Theorem 2.1], an element $\mathbf{s} \in \partial\phi(\bar{\mathbf{y}})$ is given by:

$$\mathbf{s} \equiv [\mathbf{Jc}(\bar{\mathbf{y}})]^T \mathbf{x}, \quad \text{for any } \mathbf{x} \in M(\bar{\mathbf{y}}), \tag{26}$$

where $M(\bar{\mathbf{y}})$ is the optimal solution set of the LP in (19) for $\mathbf{y} := \bar{\mathbf{y}}$. Both this method and Algorithm 2 can compute a valid element of the generalized Jacobian. Recall that an L-derivative can be computed from an LD-derivative via (2), and an L-derivative in this scalar-valued function case is an element of the generalized Jacobian. It appears that (26) is

Algorithm 2 Computes an LD-derivative $\phi'(\bar{y}; \mathbf{M})$ of ϕ for some $\bar{y} \in Y$ in Assumption 3

Require: $\phi(\bar{y})$ for some $\bar{y} \in Y$ in Assumption 3, a primal optimal solution \mathbf{x}^* and a dual optimal solution λ^* of the LP in (19) for $\mathbf{y} := \bar{y}$, and $\mathbf{M} := [\mathbf{m}_{(1)} \cdots \mathbf{m}_{(p)}] \in \mathbb{R}^{n_y \times p}$.
 Use $(\mathbf{x}^*, \lambda^*)$ to construct and solve the solutions' uniqueness examination LP (as in Proposition 3) of the LP in (19), and set $t \leftarrow$ the optimal objective value.
if $t \equiv 0$ **then**
 Compute $\phi'(\bar{y}; \mathbf{M})$ by (25) with $\mathbf{x} := \mathbf{x}^*$
 return $\phi'(\bar{y}; \mathbf{M})$
else
 continue
end if
 Set $j \leftarrow 0$.
while $j < p$ **do**
 Compute $\phi_{\mathbf{y}, \mathbf{M}}^{(j)}(\mathbf{m}_{(j+1)})$ by solving the LP in (22), and set $\mathbf{x}^* \leftarrow$ a primal optimal solution, $\lambda^* \leftarrow$ a dual optimal solution
 if $j < p - 1$ **then**
 Use $(\mathbf{x}^*, \lambda^*)$ to construct and solve the solutions' uniqueness examination LP of the LP defined for $\phi_{\mathbf{y}, \mathbf{M}}^{(j)}(\mathbf{m}_{(j+1)})$, and set $t \leftarrow$ the optimal objective value
 if $t \equiv 0$ **then**
 break
 else
 continue
 end if
 else
 continue
 end if
 Set $j \leftarrow j + 1$
end while
 Compute $\phi'(\mathbf{y}; \mathbf{M})$ by (25) with $\mathbf{x} := \mathbf{x}^*$
return $\phi'(\mathbf{y}; \mathbf{M})$

more efficient since no auxiliary LP needs to be solved. However, Algorithm 2 offers two advantages. Firstly, unlike the generalized Jacobian, the LD-derivatives obtained by Algorithm 3 satisfy a sharp chain rule [27], and thus allow dealing with more complex problems with ϕ embedded. Secondly, since ϕ is piecewise differentiable as shown in Theorem 4, in this case, an L-derivative is also guaranteed to be an element of the B-subdifferential in Definition 1 ([27, Corollary 3.6]). Such element enables solving nonsmooth equation systems with ϕ embedded using the B-subdifferential-based Newton method [30]. It has been shown in [27] that for piecewise differentiable functions, this generalized Newton method exhibits local Q-quadratic convergence under less stringent conditions than the semismooth Newton method [40] based on elements of the generalized Jacobian.

4 Conclusions and future work

This article has proposed new methods for computing LD-derivatives of optimal-value functions with embedded parameterized convex programs, with potential applications in nonsmooth equation solving and optimization. We have considered three cases of parameterized convex programs, namely the partial convexity case, the joint convexity case, and the case where the parameters appear in the objective function of LPs. In each case, our new method computes an LD-derivative by solving a sequence of LPs. In the partial convexity case, the key assumptions of the new method are that the strong Slater conditions hold for the embedded convex NLP's decision space, and this convex NLP has a unique optimal

solution for the parameters' values of interest. It has been shown that these assumptions are essentially less stringent than the conditions required by the only established methods [48, 49] that may be used for our purpose. Moreover, under these less stringent conditions, our new Algorithm 1 is also evidently computationally preferable over these methods. We then showed in Theorem 3 that in the joint convexity case, the proposed method does not require a unique optimal solution of the embedded convex programs, and any optimal solution can be used for furnishing the desired LD-derivatives. This feature is particularly beneficial for implementation. Prior to this article, there is no established method for furnishing computationally relevant generalized derivatives in this case. In the case of parameterized LP's objective function, our new method does not require Slater conditions for the decision space, does not require a unique optimal solution, and yields valid B-subdifferential elements for facilitating piecewise differentiable equation-solving. Lastly, Algorithms 1 and 2 employ an LP's solutions' uniqueness examination scheme as described in Proposition 3, which may significantly reduce computational effort for evaluating LD-derivatives in practice.

An interesting application of the proposed results is to design convergent algorithms for solving large-scale decomposable programs of the form

$$\begin{aligned} \min_{\mathbf{x} \in X, \mathbf{y} \in Y} \quad & f(\mathbf{x}, \mathbf{y}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0}, \\ & \mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{0}, \end{aligned} \tag{27}$$

where \mathbf{y} are the complicating variables, and the functions f , \mathbf{g} , and \mathbf{h} are convex with respect to \mathbf{x} for each fixed \mathbf{y} . This problem is equivalent to

$$\min_{\mathbf{y} \in Y^*} \phi(\mathbf{y}), \tag{28}$$

where Y^* is the projection of the original decision space onto \mathbf{y} -space [17], and $\phi(\mathbf{y}) := \min\{f(\mathbf{x}, \mathbf{y}) : \mathbf{g}(\mathbf{x}, \mathbf{y}) \leq \mathbf{0}, \mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{0}, \mathbf{x} \in X\}$. The generalized Benders decomposition [17] is a prevalent approach for solving such problems. This approach progressively refines upper and lower bounds of the globally optimal objective value, by solving certain auxiliary subproblems. The results of this article may open a new avenue for solving such NLPs with partial convexity. When \mathbf{y} are continuous variables, the proposed results may be applied to evaluate a generalized derivative of $\phi(\mathbf{y})$, and thus enables solving (28) directly using nonsmooth optimization methods, while still exploiting any decomposable structure when evaluating $\phi(\mathbf{y})$. Such methods would not require the functions in (27) to have *Property P* (c.f. [17]), which is crucial to apply the generalized Benders decomposition.

Funding 'Open Access funding provided by the MIT Libraries'

Data availability statements Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is

not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Barton, P.I., Khan, K.A., Stechlin, P., Watson, H.A.: Computationally relevant generalized derivatives: theory, evaluation, and applications. *Optim. Methods Softw.* **33**(4–6), 1030–1072 (2018)
2. Bertsekas, D.P.: *Nonlinear Programming*. Athena Scientific, Belmont (1999)
3. Bertsimas, D.P., Tsitsiklis, J.N.: *Introduction to Linear Optimization*. Athena Scientific, Belmont (1997)
4. Bonnans, J.F., Shapiro, A.: *Perturbation Analysis of Optimization Problems*. Springer, Berlin (2000)
5. Clarke, F.H.: Generalized gradients and applications. *Trans. Am. Math. Soc.* **205**, 247–262 (1975)
6. Clarke, F.H.: *Optimization and Nonsmooth Analysis*. SIAM, Philadelphia (1990)
7. Craven, B., Janin, R.: Regularity properties of the optimal value function in non linear programming. *Optimization* **28**(1), 1–7 (1993)
8. Danskin, J.M.: The theory of max–min, with applications. *SIAM J. Appl. Math.* **14**(4), 641–664 (1966)
9. De Wolf, D., Smeers, Y.: Generalized derivatives of the optimal value of a linear program with respect to matrix coefficients. *Eur. J. Oper. Res.* **291**(2), 491–496 (2021)
10. Facchinei, F., Fischer, A., Herrich, M.: An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions. *Math. Program.* **146**(1), 1–36 (2014)
11. Facchinei, F., Pang, J.S.: *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, New York (2003)
12. Fiacco, A.V.: *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, vol. 165. Academic Press, New York (1983)
13. Fiacco, A.V., Kyparisis, J.: Convexity and concavity properties of the optimal value function in parametric nonlinear programming. *J. Optim. Theory Appl.* **48**(1), 95–126 (1986)
14. Fiacco, A.V., McCormick, G.P.: *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. Wiley, New York (1968)
15. Fischer, A., Herrich, M., Izmailov, A.F., Solodov, M.V.: A globally convergent LP-Newton method. *SIAM J. Optim.* **26**(4), 2012–2033 (2016)
16. Gauvin, J.: The generalized gradient of a marginal function in mathematical programming. *Math. Oper. Res.* **4**(4), 458–463 (1979)
17. Geoffrion, A.M.: Generalized Benders decomposition. *J. Optim. Theory Appl.* **10**(4), 237–260 (1972)
18. Gomez, J.A., Höffner, K., Khan, K.A., Barton, P.I.: Generalized derivatives of lexicographic linear programs. *J. Optim. Theory Appl.* **178**(2), 477–501 (2018)
19. Heinonen, J.: *Lectures on Lipschitz Analysis*. University of Jyväskylä, Jyväskylä (2005)
20. Hiriart-Urruty, J.B., Lemaréchal, C.: *Convex Analysis and Minimization Algorithms I: Fundamentals*. Springer, Berlin (2013)
21. Höffner, K., Khan, K.A., Barton, P.I.: Generalized derivatives of dynamic systems with a linear program embedded. *Automatica* **63**, 198–208 (2016)
22. Hogan, W.W.: Directional derivatives for extremal-value functions with applications to the completely convex case. *Oper. Res.* **21**(1), 188–209 (1973)
23. Hogan, W.W.: Point-to-set maps in mathematical programming. *SIAM Rev.* **15**(3), 591–603 (1973)
24. IBM ILOG: V20.1.0: User's manual for CPLEX. International Business Machines Corporation (2020)
25. Khan, K.A.: Branch-locking AD techniques for nonsmooth composite functions and nonsmooth implicit functions. *Optim. Methods Softw.* **33**(4–6), 1127–1155 (2018)
26. Khan, K.A., Barton, P.I.: Generalized derivatives for solutions of parametric ordinary differential equations with non-differentiable right-hand sides. *J. Optim. Theory Appl.* **163**(2), 355–386 (2014)
27. Khan, K.A., Barton, P.I.: A vector forward mode of automatic differentiation for generalized derivative evaluation. *Optim. Methods Softw.* **30**(6), 1185–1212 (2015)
28. Klatte, D., Kummer, B.: Stability properties of infima and optimal solutions of parametric optimization problems. In: Demyanov, V.F., Pallaschke, D. (eds.) *Nondifferentiable Optimization: Motivations and Applications*, pp. 215–229. Springer, Berlin (1985)
29. Klatte, D., Kummer, B.: *Nonsmooth Equations in Optimization: Regularity, Calculus, Methods and Applications*. Kluwer, Dordrecht (2002)
30. Kojima, M., Shindo, S.: Extension of Newton and quasi-Newton methods to systems of PC¹ equations. *J. Oper. Res. Soc. Jpn.* **29**(4), 352–375 (1986)

31. Lemaréchal, C., Strodiot, J.J., Bihain, A.: On a bundle algorithm for nonsmooth optimization. In: Mangasarian, O.L., Meyer, R.R., Robinson, S.M. (eds.) *Nonlinear Programming 4*, pp. 245–282. Academic Press, New York (1981)
32. Lukšan, L., Vlček, J.: A bundle-Newton method for nonsmooth unconstrained minimization. *Math. Program.* **83**(1), 373–391 (1998)
33. Mangasarian, O.: Uniqueness of solution in linear programming. *Linear Algebra Appl.* **25**, 151–162 (1979)
34. Mordukhovich, B.S., Nam, N.M., Yen, N.D.: Subgradients of marginal functions in parametric mathematical programming. *Math. Program.* **116**(1), 369–396 (2009)
35. Nesterov, Y.: Lexicographic differentiation of nonsmooth functions. *Math. Program.* **104**(2–3), 669–700 (2005)
36. Nielsen, C.J., Barton, P.I.: 110th anniversary: a generalized nonsmooth operator for process integration. *Ind. Eng. Chem. Res.* **59**(1), 253–264 (2019)
37. Pang, J.S.: Newton’s method for B-differentiable equations. *Math. Oper. Res.* **15**(2), 311–341 (1990)
38. Penot, J.P.: Differentiability properties of optimal value functions. *Can. J. Math.* **56**(4), 825–842 (2004)
39. Qi, L.: Convergence analysis of some algorithms for solving nonsmooth equations. *Math. Oper. Res.* **18**(1), 227–244 (1993)
40. Qi, L., Sun, J.: A nonsmooth version of Newton’s method. *Math. Program.* **58**(1), 353–367 (1993)
41. Ralph, D., Dempe, S.: Directional derivatives of the solution of a parametric nonlinear program. *Math. Program.* **70**(1), 159–172 (1995)
42. Rockafellar, R.T.: *Convex Analysis*. Princeton University Press, Princeton (1970)
43. Rockafellar, R.T.: Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization. *Math. Oper. Res.* **6**(3), 424–436 (1981)
44. Rockafellar, R.T.: Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming. *Math. Program. Study* **17**, 28–66 (1982)
45. Rockafellar, R.T.: Directional differentiability of the optimal value function in a nonlinear programming problem. In: *Sensitivity, Stability and Parametric Analysis*, pp. 213–226. Springer (1984)
46. Rockafellar, R.T., Wets, R.J.B.: *Variational Analysis*, vol. 317. Springer, Berlin (2009)
47. Scholtes, S.: *Introduction to Piecewise Differentiable Equations*. Springer, Berlin (2012)
48. Stechliniski, P., Jäschke, J., Barton, P.I.: Generalized sensitivity analysis of nonlinear programs using a sequence of quadratic programs. *Optimization* **68**(2–3), 485–508 (2019)
49. Stechliniski, P., Khan, K.A., Barton, P.I.: Generalized sensitivity analysis of nonlinear programs. *SIAM J. Optim.* **28**(1), 272–301 (2018)
50. Sweetser, T.: A minimal set-valued strong derivative for vector-valued Lipschitz functions. *J. Optim. Theory Appl.* **23**(4), 549–562 (1977)
51. Thibault, L.: On subdifferentials of optimal value functions. *SIAM J. Control. Optim.* **29**(5), 1019–1036 (1991)
52. Ulbrich, M.: Semismooth Newton methods for operator equations in function spaces. *SIAM J. Optim.* **13**(3), 805–841 (2002)
53. Vikse, M., Watson, H.A., Kim, D., Barton, P.I., Gundersen, T.: Optimization of a dual mixed refrigerant process using a nonsmooth approach. *Energy* **196**, 116999 (2020)
54. Wächter, A., Biegler, L.T.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.* **106**(1), 25–57 (2006)
55. Watson, H.A., Vikse, M., Gundersen, T., Barton, P.I.: Optimization of single mixed-refrigerant natural gas liquefaction processes described by nondifferentiable models. *Energy* **150**, 860–876 (2018)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.