Generalized Sensitivity Analysis of Nonlinear Programs

As Published	10.1137/17M1120385
Publisher	Society for Industrial & Applied Mathematics (SIAM)
Version	Final published version
Citable link	https://hdl.handle.net/1721.1/133369
Terms of Use	Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.

SIAM J. OPTIM.
Vol. 28, No. 1, pp. 272–301

GENERALIZED SENSITIVITY ANALYSIS OF NONLINEAR PROGRAMS*

PETER STECHLINSKI†, KAMIL A. KHAN‡, AND PAUL I. BARTON§

Abstract. This paper extends classical sensitivity results for nonlinear programs to cases in which parametric perturbations cause changes in the active set. This is accomplished using lexicographic directional derivatives, a recently developed tool in nonsmooth analysis based on Nesterov's lexicographic differentiation. A nonsmooth implicit function theorem is augmented with generalized derivative information and applied to a standard nonsmooth reformulation of the parametric KKT system. It is shown that the sufficient conditions for this implicit function theorem variant are implied by a KKT point satisfying the linear independence constraint qualification and strong second-order sufficiency. Mirroring the classical theory, the resulting sensitivity system is a nonsmooth equation system which admits primal and dual sensitivities as its unique solution. Practically implementable algorithms are provided for calculating the nonsmooth sensitivity system's unique solution, which is then used to furnish B-subdifferential elements of the primal and dual variable solutions by solving a linear equation system. Consequently, the findings in this article are computationally relevant since dedicated nonsmooth equation-solving and optimization methods display attractive convergence properties when supplied with such generalized derivative elements. The results have potential applications in nonlinear model predictive control and problems involving dynamic systems with mathematical programs embedded. Extending the theoretical treatments here to sensitivity analysis theory of other mathematical programs is also anticipated.

Key words. parametric optimization, nonsmooth KKT systems, generalized derivatives, lexicographic differentiation, lexicographic directional derivatives, nonsmooth implicit functions

AMS subject classifications. 49J52, 90C31, 90C33

DOI. 10.1137/17M1120385

1. Introduction.

Fiacco and McCormick [11] characterized primal and dual variable first-order sensitivities of parametric nonlinear programs (NLPs) from linearized KKT conditions, furnished by application of the classical implicit function theorem under appropriate regularity conditions. However, the theory of Fiacco and McCormick yields no information in the presence of active index set changes under parametric perturbations. In this article, the aforementioned classical sensitivity results are extended to include active index set changes; parametric sensitivities of NLPs are characterized by evaluating lexicographic directional (LD-) derivatives of nonsmooth equation-based reformulations of KKT systems.

Built from the theory of lexicographic differentiation [30], the LD-derivative is a nonsmooth extension of the classical directional derivative and can be used in established methods for nonsmooth equation-solving and optimization. LD-derivatives

*Received by the editors March 10, 2017; accepted for publication (in revised form) August 22, 2017; published electronically February 1, 2018.
http://www.siam.org/journals/siopt/28-1/M112038.html
Funding: This material was based on work supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357.
†Process Systems Engineering Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 (pstechli@mit.edu, http://yoric.mit.edu). Current address: Department of Mathematics and Statistics, University of Maine, Orono, ME 04469 (peter.stechlinski@maine.edu).
‡Department of Chemical Engineering, McMaster University, Hamilton, ON L8S 4L7, Canada (kamilkhan@mcmaster.ca).
§Process Systems Engineering Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 (pib@mit.edu, http://yoric.mit.edu).

have been used successfully to furnish generalized derivative information for linear programs (for use in sensitivity analysis of ordinary-differential equations with linear programs embedded) [16], lexicographic linear programs [13], as well as optimization problems with nonsmooth dynamical systems embedded [20, 39]. As the LD-derivative satisfies sharp calculus rules, generalized derivatives of the problems outlined above can be evaluated using a tractable numerical implementation; a nonsmooth vector forward mode of automatic differentiation has been developed [21] for automatable and relatively cheap LD-derivative computation, extending established automatic differentiation methods [15]. For an overview of the theory of LD-derivatives, the reader is referred to [1].

This article shows that, when applied to the setting of parametric NLPs under appropriate regularity conditions, the LD-derivatives approach furnishes an auxiliary nonsmooth and nonlinear equation system that describes NLP primal and dual variable sensitivities. Said system admits a unique solution under complete coherent orientation [34] of nonsmooth equation-based reformulations of KKT systems. Complete coherent orientation is implied by the linear independence constraint qualification (LICQ) and strong second-order sufficiency and guarantees primal and dual variable NLP solution mappings that are piecewise differentiable in the sense of Scholtes [37] and unique in a neighborhood of a reference problem parameter. Moreover, the unique solution of said system can be used to obtain lexicographic derivatives of the primal and dual variable solution mappings, which are guaranteed to be elements of the B-subdifferential (and therefore Clarke's generalized Jacobian [3]) of the solution mappings and thus computationally relevant; dedicated nonsmooth equation-solving algorithms (e.g., semismooth Newton methods [32] and LP-Newton methods [7]) and nonsmooth optimization methods (e.g., bundle methods for local optimization [26]) can be applied with convergence properties similar to their smooth counterparts.

Tracing back to the influential works of Kojima [24] on NLPs and Robinson [35] on generalized equations, a series of results have been obtained in the literature to address parametric sensitivities of NLPs in the presence of active index set changes [17, 38, 25, 6], culminating in a practical method to calculate directional derivatives of the primal variable solution in the form of an auxiliary convex quadratic program (QP) with a linear program embedded [33]. It is straightforward to calculate B-subdifferential elements of nonsmooth equation-based reformulations of NLP KKT systems (see [9, Chapter 10] and, in particular, [9, Proposition 10.1.16]) for use in furnishing KKT points via nonsmooth equation-solver methods; various methods for finding KKT triples of variational inequalities (VIs) are discussed, with numerical advantages and drawbacks highlighted. Arguments advocating a nonsmooth equation system reformulation over smoothening approximations are given in [9, section 9.1]. However, there is currently no theory for computing B-subdifferential elements of primal and dual variable NLP solutions (i.e., generalized derivative first-order sensitivity information) until this article, extending the classical results of Fiacco and McCormick by removing their restrictive assumption of strict complementarity.

An overview of sensitivity analysis theory for mathematical programs is found in [10, 12, 2, 23, 8]; [8, section 5.7] presents a detailed account of current sensitivity analysis theory for parametric NLPs, as well as complementarity problems, VIs, and mathematical programs with equilibrium constraints (MPECs). This article focuses on a nonsmooth equation-based reformulation of NLP KKT systems using the minimum function, but the results can be generalized to any suitable nonlinear complementarity problem (NCP) function reformulation of NLP KKT systems and to mixed complementarity problems (MiCPs) and VIs under suitable regularity conditions. For example, an auxiliary variable formulation in the spirit of [7, Example 2], which has shown to have numerical benefits, can be treated with the present theory. Extensions to sensitivity analysis of other mathematical programs are expected using the theoretical machinery presented here.

The rest of this article is structured as follows. Necessary background material is presented in section 2. Section 3 details existence and computation of generalized derivatives of nonsmooth inverse and implicit functions, including a specialization to MiCPs useful for present purposes in section 4. Generalized derivatives of NLPs are given in section 5, including a connection to familiar regularity conditions and a scheme for evaluating the unique solution of the nonsmooth and nonlinear sensitivity system. Future work and conclusions are given in section 6.

2. Background material. Unless specified otherwise, the following notational conventions are used. $\mathbb{N}$ and $\mathbb{R}_+$ denote the positive integers and nonnegative real numbers, respectively. $\mathbb{R}^n$ is the Euclidean space of n-dimensions (equipped with the Euclidean norm $\|\cdot\|$) and the vector space $\mathbb{R}^{m\times n}$ is equipped with the corresponding induced norm. A set is denoted by an uppercase letter (e.g., H). The canonical projections of $H \subset \mathbb{R}^n \times \mathbb{R}^m$ onto $\mathbb{R}^n$ and $\mathbb{R}^m$ are denoted by $\pi_{\mathbf{x}}H$ and $\pi_{\mathbf{y}}H$, respectively. Vector-valued functions and vectors in $\mathbb{R}^n$ are denoted by lowercase boldface letters (e.g., $\mathbf{h}$) whose ith component is denoted by h_i. Given a function $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^m$ and a nonempty subset $\mathcal{J} \equiv \{j_1, \dots, j_s\} \subset \{1, \dots, m\}$ with $s \le m$ and $j_l < j_{l+1}$ $\forall l \in \{1, \dots, s-1\}$, let $\mathbf{h}_{\mathcal{J}} \equiv (h_{j_1}, \dots, h_{j_s})$ (i.e., the components of $\mathbf{h}$ indexed by $\mathcal{J}$). Similarly, given a vector $\mathbf{h} \in \mathbb{R}^m$, let $\mathbf{h}_{\mathcal{J}}$ denote its components indexed by $\mathcal{J}$.

Matrix-valued functions and matrices in $\mathbb{R}^{m\times n}$ are denoted by uppercase boldface letters (e.g., $\mathbf{H}$). Parenthetical subscripts are used to indicate the column vector of a matrix (e.g., the matrix $\mathbf{H} \in \mathbb{R}^{m\times n}$ has the kth column $\mathbf{h}_{(k)} \in \mathbb{R}^m$ whose ith component is $h_{(k),i}$), a leftmost submatrix of a matrix (e.g., $\mathbf{H}_{(k)} \equiv [\mathbf{h}_{(1)} \quad \cdots \quad \mathbf{h}_{(k)}] \in \mathbb{R}^{m\times k}$), or to indicate a sequence of vectors or vector-valued functions. The kth row of $\mathbf{H} \in \mathbb{R}^{m\times n}$ is denoted by $\mathbf{H}_k \in \mathbb{R}^{1\times n}$. Unless stated otherwise, parenthetical superscripts (e.g., $\mathbf{h}^{(k)}$) are used for lexicographic differentiation. $\mathbf{0}_n$ denotes the zero vector in $\mathbb{R}^n$, $\mathbf{0}_{m\times n}$ denotes the $m \times n$ zero matrix, and $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. The notation $(\mathbf{M}, \mathbf{N})$ is used for a well-defined vertical block matrix (or vector): $\left[\begin{smallmatrix}\mathbf{M}\\ \mathbf{N}\end{smallmatrix}\right]$. For convenience in inductive proofs, an empty matrix with m rows but no columns is denoted by $\emptyset_{m\times 0}$ and concatenated with $\mathbf{H} \in \mathbb{R}^{m\times n}$ as follows:

$$[\mathbf{H} \quad \emptyset_{m\times 0}] = [\emptyset_{m\times 0} \quad \mathbf{H}] = \mathbf{H}.$$

Let $\operatorname{diag}(a_1, \dots, a_m) \in \mathbb{R}^{m\times m}$ denote the diagonal matrix with (i,i)-entry $a_i \in \mathbb{R}$.

Given $\mathbf{H} \in \mathbb{R}^{m\times n}$, nonempty subset $\mathcal{J} \equiv \{j_1, \dots, j_s\} \subset \{1, \dots, m\}$ satisfying $s \le m$ and $j_l < j_{l+1}$ $\forall l \in \{1, \dots, s-1\}$, and nonempty subset $\mathcal{I} \equiv \{i_1, \dots, i_r\} \subset \{1, \dots, n\}$ satisfying $r \le n$ and $i_l < i_{l+1}$ $\forall l \in \{1, \dots, r-1\}$, let $\mathbf{H}_{\mathcal{J},\bullet}$ denote the rows of $\mathbf{H}$ indexed by $\mathcal{J}$ and $\mathbf{H}_{\bullet,\mathcal{I}}$ denote the columns of $\mathbf{H}$ indexed by $\mathcal{I}$. That is,

$$\mathbf{H}_{\mathcal{J},\bullet} \equiv \begin{bmatrix} h_{(1),j_1} & h_{(2),j_1} & \dots & h_{(n),j_1} \\ h_{(1),j_2} & h_{(2),j_2} & \dots & h_{(n),j_2} \\ \vdots & \vdots & \vdots & \vdots \\ h_{(1),j_s} & h_{(2),j_s} & \dots & h_{(n),j_s} \end{bmatrix} \in \mathbb{R}^{s\times n},$$

$$\mathbf{H}_{\bullet,\mathcal{I}} \equiv \begin{bmatrix} \mathbf{h}_{(i_1)} & \mathbf{h}_{(i_2)} & \dots & \mathbf{h}_{(i_r)} \end{bmatrix} \in \mathbb{R}^{m\times r}.$$

The following convention is adopted: given two distinct index sets, i.e., nonempty subsets $\mathcal{J}_1, \mathcal{J}_2 \subset \{1, \dots, m\}$ such that $\mathcal{J}_1 \cap \mathcal{J}_2 = \emptyset$, the matrix $\mathbf{H}_{\mathcal{J}_1\cup\mathcal{J}_2,\bullet}$ denotes the columns of $\mathbf{H}$ indexed by $\mathcal{J}^*$, where $\mathcal{J}^*$ is the index set formed by merging $\mathcal{J}_1$ and $\mathcal{J}_2$ with proper ordering. $\mathbf{H}_{\bullet,\mathcal{I}_1\cup\mathcal{I}_2}$ and $\mathbf{h}_{\mathcal{J}_1\cup\mathcal{J}_2}$ are defined in a similar spirit. Let $\mathbf{I}_{\mathcal{J},\bullet} \in \mathbb{R}^{s\times m}$ denote the rows of $\mathbf{I}_m$ indexed by $\mathcal{J}$ (i.e., the matrix whose kth row is equal to $\mathbf{e}_{(j_k)}^{\mathrm{T}}$, where $\mathbf{e}_{(i)}$ denotes the ith unit coordinate vector in $\mathbb{R}^m$).

2.1. Generalized derivatives. Let $X \subset \mathbb{R}^n$ be open and $\mathbf{f} : X \to \mathbb{R}^m$ be locally Lipschitz continuous on X. By Rademacher's theorem, $\mathbf{f}$ is differentiable at each point $\mathbf{x}^0 \in X\backslash Z_{\mathbf{f}}$, where the subset $Z_{\mathbf{f}} \subset X$ has zero (Lebesgue) measure. The *B-subdifferential* of $\mathbf{f}$ at $\mathbf{x}^0$ is equal to

$$\partial^{\mathrm{B}}\mathbf{f}(\mathbf{x}^0) \equiv \left\{\mathbf{H} \in \mathbb{R}^{m\times n} : \mathbf{H} = \lim_{i\to\infty} \mathbf{J}\mathbf{f}(\mathbf{x}_{(i)}),\ \lim_{i\to\infty} \mathbf{x}_{(i)} = \mathbf{x}^0,\ \mathbf{x}_{(i)} \in X\backslash Z_{\mathbf{f}}\ \forall i \in \mathbb{N}\right\}$$

and is nonempty and compact. The Clarke (generalized) *Jacobian* of $\mathbf{f}$ at $\mathbf{x}^0$ [3] is defined as

$$\partial\mathbf{f}(\mathbf{x}^0) \equiv \operatorname{conv} \partial^{\mathrm{B}}\mathbf{f}(\mathbf{x}^0).$$

If $\mathbf{f}$ is continuously differentiable (C^1) at $\mathbf{x}^0$, then $\partial\mathbf{f}(\mathbf{x}^0) = \partial^{\mathrm{B}}\mathbf{f}(\mathbf{x}^0) = \{\mathbf{J}\mathbf{f}(\mathbf{x}^0)\}$. If $\mathbf{f}$ is differentiable at $\mathbf{x}^0$, then $\mathbf{J}\mathbf{f}(\mathbf{x}^0) \in \partial\mathbf{f}^{\mathrm{B}}(\mathbf{x}^0)$. If $\mathbf{f}$ is *piecewise differentiable (PC^1)* [37] at $\mathbf{x}^0$ then, by [31, Lemma 2],

$$\partial^{\mathrm{B}}\mathbf{f}(\mathbf{x}^0) = \left\{\mathbf{J}\mathbf{f}_{(i)}(\mathbf{x}^0) : i \in I_{\mathbf{f}}^{\mathrm{ess}}(\mathbf{x}^0)\right\},$$

where

$$I_{\mathbf{f}}^{\mathrm{ess}}(\mathbf{x}^0) \equiv \{i \in \{1,\ldots,k\} : \mathbf{x}^0 \in \mathrm{cl}(\mathrm{int}\{\boldsymbol{\eta} \in N_{\mathbf{x}^0} : \mathbf{f}(\boldsymbol{\eta}) = \mathbf{f}_{(i)}(\boldsymbol{\eta})\})\}$$

is the set of *essentially active indices* of $\mathbf{f}$ at $\mathbf{x}^0$ with respect to a set of C^1 selection functions $\{\mathbf{f}_{(1)},\ldots,\mathbf{f}_{(k)}\}$ defined on a neighborhood $N_{\mathbf{x}^0} \subset X$ of $\mathbf{x}^0$. Corresponding to the essentially active indices are the set of *essentially active selection functions* of $\mathbf{f}$ at $\mathbf{x}^0$, defined as $\mathcal{E}_{\mathbf{f}}(\mathbf{x}^0) \equiv \{\mathbf{f}_{(i)} : i \in I_{\mathbf{f}}^{\mathrm{ess}}(\mathbf{x}^0)\}$. Note that the class of *piecewise differentiable functions of order-r (PC^r)* is defined analogously as the class of C^r functions.

Given $W \subset \mathbb{R}^n \times \mathbb{R}^m$ open and $\mathbf{g} : W \to \mathbb{R}^q$ Lipschitz continuous on a neighborhood of $(\mathbf{x}^0,\mathbf{y}^0) \in W$, let $\mathbf{g}_{\mathbf{x}^0} \equiv \mathbf{g}(\mathbf{x}^0,\cdot)$ and $\mathbf{g}_{\mathbf{y}^0} \equiv \mathbf{g}(\cdot,\mathbf{y}^0)$. Let $Z_{\mathbf{g}_{\mathbf{x}^0}} \subset \pi_{\mathbf{y}}(W;\mathbf{x}^0) \subset \mathbb{R}^m$ and $Z_{\mathbf{g}_{\mathbf{y}^0}} \subset \pi_{\mathbf{x}}(W;\mathbf{y}^0) \subset \mathbb{R}^n$ be the zero measure subsets for which $\mathbf{g}_{\mathbf{x}^0}$ and $\mathbf{g}_{\mathbf{y}^0}$ are not differentiable, respectively. The *partial* Clarke (generalized) *Jacobian* of $\mathbf{g}$ with respect to $\mathbf{y}$ at $(\mathbf{x}^0,\mathbf{y}^0)$ is the convex hull of the *partial B-subdifferential* of $\mathbf{g}$ with respect to $\mathbf{y}$ at $(\mathbf{x}^0,\mathbf{y}^0)$,

$$\partial_{\mathbf{y}}^{\mathrm{B}}\mathbf{g}(\mathbf{x}^0,\mathbf{y}^0) \equiv \partial^{\mathrm{B}}[\mathbf{g}(\mathbf{x}^0,\cdot)](\mathbf{y}^0).$$

The Clarke (generalized) *Jacobian projection* of $\mathbf{g}$ with respect to $\mathbf{y}$ at $(\mathbf{x}^0,\mathbf{y}^0)$ is defined as

$$\pi_{\mathbf{y}}\partial\mathbf{g}(\mathbf{x}^0,\mathbf{y}^0) \equiv \left\{\mathbf{N} \in \mathbb{R}^{q\times m} : \exists[\mathbf{M} \quad \mathbf{N}] \in \partial\mathbf{g}(\mathbf{x}^0,\mathbf{y}^0)\right\}.$$

If $\mathbf{g}$ is C^1 at $(\mathbf{x}^0,\mathbf{y}^0)$, then $\pi_{\mathbf{x}}\partial\mathbf{g}(\mathbf{x}^0,\mathbf{y}^0) = \partial_{\mathbf{x}}\mathbf{g}(\mathbf{x}^0,\mathbf{y}^0) = \left\{\mathbf{J}_{\mathbf{x}}\mathbf{g}(\mathbf{x}^0,\mathbf{y}^0)\right\}$. The generalized derivatives $\partial_{\mathbf{x}}^{\mathrm{B}}\mathbf{g}(\mathbf{x}^0,\mathbf{y}^0)$ and $\pi_{\mathbf{x}}\partial\mathbf{g}(\mathbf{x}^0,\mathbf{y}^0)$ are defined similarly.

Suppose that $\mathbf{g}$ is PC^1 at $(\mathbf{x}^0,\mathbf{y}^0) \in W$ with selection functions $\{\mathbf{g}_{(1)},\ldots,\mathbf{g}_{(k)}\}$ and essentially active indices $I_{\mathbf{g}}^{\mathrm{ess}}(\mathbf{x}^0,\mathbf{y}^0)$. For use in their analysis of piecewise smooth equations, Ralph and Scholtes [34, p. 607] made use of the Cartesian product of partial Jacobians of essentially active selection functions, called here the *combinatorial partial Jacobian* of $\mathbf{g}$ with respect to $\mathbf{y}$ at $(\mathbf{x}^0,\mathbf{y}^0)$,

$$\begin{aligned}
&\Lambda_{\mathbf{y}}\mathbf{g}(\mathbf{x}^0,\mathbf{y}^0)\\
&\quad\equiv \{\mathbf{M}\in\mathbb{R}^{q\times m} : \mathbf{M}_{\{i\},\bullet} = \mathbf{J}_{\mathbf{y}} g_{(\delta_i),i}(\mathbf{x}^0,\mathbf{y}^0), \forall i\in\{1,\ldots,q\},\ \boldsymbol{\delta}\in\Delta_{\mathbf{g}}(\mathbf{x}^0,\mathbf{y}^0)\}\\
&\quad= \prod_{j=1}^{q}\{\mathbf{J}_{\mathbf{y}} g_{(i),j} : i\in I^{\text{ess}}_{\mathbf{g}}(\mathbf{x}^0,\mathbf{y}^0)\},
\end{aligned}$$

where the *combinatorial vectorization of the essentially active indices* of $\mathbf{g}$ at $(\mathbf{x}^0,\mathbf{y}^0)$ is given by

$$\Delta_{\mathbf{g}}(\mathbf{x}^0,\mathbf{y}^0) \equiv \{\boldsymbol{\delta}\in\mathbb{N}^q : \delta_i\in I^{\text{ess}}_{\mathbf{g}}(\mathbf{x}^0,\mathbf{y}^0), i\in\{1,\ldots,q\}\}.$$

The *combinatorial partial Jacobian* of $\mathbf{g}$ with respect to $\mathbf{x}$ is defined similarly.

Example 2.1. Let $g:\mathbb{R}^2\to\mathbb{R}:(x,y)\mapsto\max(0,\min(x,y))$, which is PC^1 on $\mathbb{R}^2$. The subset of $\mathbb{R}^2$ at which g is not differentiable is given by

$$Z_g = \{(x,y): y=x, x\geq 0, y\geq 0\}\cup\{(x,y): y=0, x\geq 0\}\cup\{(x,y): x=0, y\geq 0\}.$$

Then $\{g_{(1)}, g_{(2)}, g_{(3)}\}$ is a set of essentially active C^1 selection functions of g at $(0,0)$, where

$$g_{(1)}:(x,y)\mapsto 0,\quad g_{(2)}:(x,y)\mapsto x,\quad g_{(3)}:(x,y)\mapsto y.$$

Hence,

$$\partial^{\mathrm{B}} g(0,0) = \{\mathbf{J}g_{(i)}(0,0) : i\in\{1,2,3\}\} = \{[1\quad 0],[0\quad 1],[0\quad 0]\},$$

from which it follows that

$$\partial g(0,0) = \{[\lambda_1\quad \lambda_2] : \lambda_1,\lambda_2\geq 0, \lambda_1+\lambda_2\leq 1\},$$

and thus

$$\pi_y\partial g(0,0) = \{\lambda : \lambda\in[0,1]\}.$$

To calculate the partial Clarke Jacobian, consider

$$\hat{g}\equiv g(0,\cdot):\mathbb{R}\to\mathbb{R}: y\mapsto\max\{0,\min\{0,y\}\},$$

and note that $\hat{g}(y)\equiv 0\ \forall y\in\mathbb{R}$. Hence,

$$\partial_y g(0,0) = \partial^{\mathrm{B}}_y g(0,0) = \{0\}.$$

The combinatorial partial Jacobian of g with respect to y at $(0,0)$ evaluates as

$$\Lambda_y g(0,0) = \{J_y g_{(i)}(0,0) : i\in\{1,2,3\}\} = \{0,1,0\}.$$

Therefore, $\partial^{\mathrm{B}}_y g(0,0)\subset\Lambda_y g(0,0)\subset\pi_y\partial g(0,0)$, where the inclusions are strict.

The Clarke Jacobian is the smallest convex-valued generalized derivative satisfying a number of useful properties: a mean-value, inverse, and implicit function theory; recovery of the subdifferential from convex analysis whenever the Lipschitzian function is scalar; and a necessary optimality condition, among others [3, 29]. Moreover, the Clarke Jacobian has practical application due to computational relevance; dedicated nonsmooth methods exhibit attractive convergence rates when supplied with Clarke Jacobian elements. However, it is challenging to compute Clarke Jacobian elements in general, for a number of reasons: the Clarke Jacobian satisfies inclusion-based calculus rules, componentwise computation may not yield a Clarke Jacobian element of a vector-valued function, etc.

Before proceeding, we briefly discuss other generalized derivatives that are prevalent in the literature. The Mordukhovich (M-) subdifferential [28] and coderivative [29, 27] satisfy a number of desirable properties, including ones the Clarke Jacobian lacks. For example, there are situations in which the M-subdifferential satisfies sharp calculus rules but the Clarke Jacobian does not. As the B-subdifferential is contained in the M-subdifferential, methods furnishing B-subdifferential elements yield M-subdifferential elements as well (which is the case in this article). The proximal subdifferential [3, 4] is applicable to functions that are not Lipschitzian and is used in stability analysis theory [5] but is not suitable for the aforementioned dedicated nonsmooth methods. Linear Newton approximations (LNAs; see [9], for example) are straightforward to compute but lack desirable properties of Clarke's constructions (e.g., an LNA element of a differentiable function is not necessarily the derivative [19, Example 5.2], and an LNA element of a convex scalar-valued function is not necessarily a subgradient).

2.2. Lexicographic differentiation and the lexicographic directional derivative. Let $X \subset \mathbb{R}^n$ be open and $\mathbf{f} : X \to \mathbb{R}^m$. The *directional derivative* of $\mathbf{f}$ at $\mathbf{x}^0 \in X$ in the direction $\mathbf{d} \in \mathbb{R}^n$ is denoted by $\mathbf{f}'(\mathbf{x}^0; \mathbf{d})$. The class of lexicographically smooth functions and the lexicographic (generalized) derivative were introduced by Nesterov [30]: given that $\mathbf{f}$ is locally Lipschitz continuous on X, $\mathbf{f}$ is said to be *lexicographically smooth (L-smooth)* at $\mathbf{x}^0 \in X$ if for any $k \in \mathbb{N}$ and any $\mathbf{M} = [\mathbf{m}_{(1)} \ \cdots \ \mathbf{m}_{(k)}] \in \mathbb{R}^{n\times k}$, the following higher-order directional derivatives are well-defined:

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

$\mathbf{f}$ is said to be *lexicographically smooth (L-smooth) on X* if it is L-smooth at each point $\mathbf{x} \in X$. The class of L-smooth functions is closed under composition and includes all C^1 functions, convex functions [30], and PC^1 functions [21]. L-smooth functions that are not PC^1 on their domain include the Euclidean norm (or any p-norm with $2 \le p < \infty$), solutions of parametric ODEs and DAEs with certain PC^1 right-hand-side functions [1], and the function

$$f : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto \inf_{k\in\mathbb{N}} \left| x - \frac{y}{k} \right|,$$

which was suggested by Jeffrey Pang and illustrated by Roshchina [36].

Given any nonsingular matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ and $\mathbf{f} : X \to \mathbb{R}^m$ L-smooth at $\mathbf{x}^0$, the mapping $\mathbf{f}^{(n)}_{\mathbf{x}^0,\mathbf{M}}$ is necessarily linear [30, Theorem 2] and the *lexicographic (L-) derivative* of $\mathbf{f}$ at $\mathbf{x}^0$ in the directions $\mathbf{M}$ is defined as

$$\mathbf{J}_{\mathrm{L}}\mathbf{f}(\mathbf{x}^0; \mathbf{M}) \equiv \mathbf{J}\mathbf{f}^{(n)}_{\mathbf{x}^0,\mathbf{M}}(\mathbf{0}_n) \in \mathbb{R}^{m\times n}.$$

The *lexicographic (L-) subdifferential* of $\mathbf{f}$ at $\mathbf{x}^0$ is defined as

$$\partial^{\mathrm{L}}\mathbf{f}(\mathbf{x}^0) \equiv \{\mathbf{J}_{\mathrm{L}}\mathbf{f}(\mathbf{x}^0; \mathbf{N}) : \mathbf{N} \in \mathbb{R}^{n\times n}, \det \mathbf{N} \neq 0\}.$$

Introduced by Khan and Barton [21], the LD-derivative is defined as follows: given any $k \in \mathbb{N}$, any $\mathbf{M} = [\mathbf{m}_{(1)} \ \cdots \ \mathbf{m}_{(k)}] \in \mathbb{R}^{n\times k}$, and $\mathbf{f} : X \to \mathbb{R}^m$ L-smooth at $\mathbf{x}^0 \in X$, the *LD-derivative* of $\mathbf{f}$ at $\mathbf{x}^0 \in X$ in the directions $\mathbf{M}$ is

$$\mathbf{f}'(\mathbf{x}^0; \mathbf{M}) \equiv \left[\mathbf{f}^{(0)}_{\mathbf{x}^0,\mathbf{M}}(\mathbf{m}_{(1)}) \quad \mathbf{f}^{(1)}_{\mathbf{x}^0,\mathbf{M}}(\mathbf{m}_{(2)}) \quad \cdots \quad \mathbf{f}^{(k-1)}_{\mathbf{x}^0,\mathbf{M}}(\mathbf{m}_{(k)})\right].$$

The LD-derivative is uniquely defined for any $\mathbf{M} \in \mathbb{R}^{n\times k}$ and $k \in \mathbb{N}$ and satisfies the linear equation system

$$\mathbf{f}'(\mathbf{x}^0; \mathbf{M}) = \mathbf{J}_{\mathrm{L}}\mathbf{f}(\mathbf{x}^0; \mathbf{M})\mathbf{M} \tag{1}$$

if, in addition, $\mathbf{M}$ is square and nonsingular. Equation (1) mirrors the relationship between the classical directional derivative and the Jacobian matrix. Indeed, if $\mathbf{f}$ is differentiable at $\mathbf{x}^0$, then $\mathbf{f}'(\mathbf{x}^0; \mathbf{M}) = \mathbf{J}\mathbf{f}(\mathbf{x}^0)\mathbf{M}$ and $\partial^{\mathrm{L}}\mathbf{f}(\mathbf{x}^0) = \{\mathbf{J}\mathbf{f}(\mathbf{x}^0)\}$. If $\mathbf{M}$ has one column, the LD-derivative is equivalent to the directional derivative.

A convincing case on the usefulness of L-derivatives has been made in [19, 21]: if $\mathbf{f}$ is PC^1 at $\mathbf{x}^0$ (which is pertinent to this article), then $\mathbf{f}$ is L-smooth at $\mathbf{x}^0$ and $\partial^{\mathrm{L}}\mathbf{f}(\mathbf{x}^0) \subset \partial^{\mathrm{B}}\mathbf{f}(\mathbf{x}^0)$; if f is a scalar-valued function that is L-smooth at $\mathbf{x}^0$ (i.e., objective functions), then $\partial^{\mathrm{L}} f(\mathbf{x}^0) \subset \partial f(\mathbf{x}^0)$; and if $\mathbf{f}$ is L-smooth at $\mathbf{x}^0$, then $\partial^{\mathrm{L}}\mathbf{f}(\mathbf{x}^0)$ is a subset of the plenary hull [42] of the Clarke Jacobian (whose elements are no less useful than Clarke Jacobian elements in nonsmooth methods using matrix-vector products, for example). The L-derivative $\mathbf{J}_{\mathrm{L}}\mathbf{f}(\mathbf{x}^0; \mathbf{M})$, which can be furnished via computing an LD-derivative for a square and nonsingular $\mathbf{M}$ and solving (1), is therefore computationally relevant in nonsmooth equation-solving methods (e.g., semismooth Newton methods and the LP-Newton methods) and nonsmooth optimization methods (e.g., bundle methods).

The L-subdifferential is nonempty and bounded, with the function's local Lipschitz constant providing a bound [30]. (The same properties hold true for the B-subdifferential, Clarke Jacobian, and M-subdifferential.) When viewed as a set-valued mapping, the B-subdifferential, Clarke Jacobian, and M-subdifferential are outer-semicontinuous (also called upper-semicontinuous), which is not true of the L-subdifferential. Outer-semicontinuity is desirable since it provides some robustness to numerical error. However, since the L-subdifferential is a subset of all three in the PC^1 setting, it benefits from the same robustness to numerical error. Of course, this moderate robustness is weaker than the usual sense, which would require the set-valued mappings to be continuous or even Lipschitz with respect to some appropriate metric. Notwithstanding, this moderate level of robustness can be sufficient to achieve attractive convergence behavior (e.g., in nonsmooth equation-solving methods). Other notions of robustness have been pursued (e.g., because of its normal-cone-based definition, the M-subdifferential benefits from other properties the Clarke Jacobian lacks).

Importantly, the LD-derivative obeys a sharp chain rule [21], unlike the Clarke Jacobian: let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ be open and $\mathbf{h} : X \to Y$ and $\mathbf{g} : Y \to \mathbb{R}^q$ be L-smooth at $\mathbf{x}^0 \in X$ and $\mathbf{h}(\mathbf{x}^0) \in Y$, respectively. Then, the composition $\mathbf{g} \circ \mathbf{h}$ is L-smooth at $\mathbf{x}^0$; for any $k \in \mathbb{N}$ and any $\mathbf{M} \in \mathbb{R}^{n\times k}$, the chain rule for LD-derivatives is given as

$$[\mathbf{g} \circ \mathbf{h}]'(\mathbf{x}^0; \mathbf{M}) = \mathbf{g}'(\mathbf{h}(\mathbf{x}^0); \mathbf{h}'(\mathbf{x}^0; \mathbf{M})), \tag{2}$$

which reduces to Nesterov's chain rule [30, Theorem 5] when the directions matrix is square and nonsingular and reduces further to the classical chain rule if the participating functions are differentiable. Thanks to said sharp chain rule, a nonsmooth vector forward mode of automatic differentiation to calculate LD-derivatives has been recently developed [21]. The property that LD-derivatives are well-defined for singular or nonsquare directions matrices is crucial for the LD-derivative chain rule (2); given a nonsingular directions matrix $\mathbf{M}$, the intermediate directions matrix $\mathbf{h}'(\mathbf{x}^0;\mathbf{M}) \in \mathbb{R}^{m\times k}$ is permitted to be singular or nonsquare. This is important in problems where the directions matrix is not chosen a priori (e.g., if the directions matrix is the output of an embedded problem).

Lexicographic differentiation and the LD-derivative are illustrated in the next example.

Example 2.2. Recall the function $g : \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto \max(0, \min(x, y))$ in Example 2.1. Let $\mathbf{M} = \mathbf{I}_2$ be the directions matrix. For any $\mathbf{d} \in \mathbb{R}^2$,

$$\begin{aligned} g^{(0)}_{\mathbf{0}_2,\mathbf{I}_2}(\mathbf{d}) &= \max(0, \min(d_1, d_2)),\\ g^{(1)}_{\mathbf{0}_2,\mathbf{I}_2}(\mathbf{d}) &= \max(0, d_2), \end{aligned}$$

from which it follows that

$$g'(0, 0; \mathbf{I}_2) = \begin{bmatrix} g^{(0)}_{\mathbf{0}_2,\mathbf{I}_2}(1, 0) & g^{(0)}_{\mathbf{0}_2,\mathbf{I}_2}(0, 1) \end{bmatrix} = \begin{bmatrix} 0 & 1 \end{bmatrix}.$$

Using (1),

$$\mathbf{J}_{\mathrm{L}}g(0, 0; \mathbf{I}_2) = \begin{bmatrix} 0 & 1 \end{bmatrix} \in \partial^{\mathrm{L}} g(0, 0) \subset \partial^{\mathrm{B}} g(0, 0).$$

(Observe that $g^{(2)}_{\mathbf{0}_2,\mathbf{I}_2}(\mathbf{d}) \equiv d_2$ is linear and $\mathbf{J}g^{(2)}_{\mathbf{0}_2,\mathbf{I}_2}(0, 0) = [0 \quad 1]$, as expected.) Choosing instead $\mathbf{M} = \left[\begin{smallmatrix} 0 & 1\\ 1 & 0 \end{smallmatrix}\right]$ yields

$$g'\left(0, 0; \left[\begin{smallmatrix} 0 & 1\\ 1 & 0 \end{smallmatrix}\right]\right) = \begin{bmatrix} 0 & 1 \end{bmatrix},$$

which furnishes the L-derivative

$$\mathbf{J}_{\mathrm{L}}g\left(0, 0; \left[\begin{smallmatrix} 0 & 1\\ 1 & 0 \end{smallmatrix}\right]\right) = g'\left(0, 0; \left[\begin{smallmatrix} 0 & 1\\ 1 & 0 \end{smallmatrix}\right]\right) \begin{bmatrix} 0 & 1\\ 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \end{bmatrix} \in \partial^{\mathrm{L}} g(0, 0) \subset \partial^{\mathrm{B}} g(0, 0).$$

Choosing $\mathbf{M} = -\mathbf{I}_2$ gives $g'(0, 0; -\mathbf{I}_2) = [0 \quad 0]$, which yields

$$\mathbf{J}_{\mathrm{L}}g\left(0, 0; -\mathbf{I}_2\right) = [0 \quad 0] \in \partial^{\mathrm{L}} g(0, 0) \subset \partial^{\mathrm{B}} g(0, 0).$$

Alternatively, define the functions

$$\begin{aligned} h &: \mathbb{R}^2 \to \mathbb{R} : (x, y) \mapsto \min(x, y),\\ f &: \mathbb{R} \to \mathbb{R} : z \mapsto \max(0, z). \end{aligned}$$

Then $g \equiv f \circ h$ and the LD-derivative chain rule (2) yields

$$\begin{aligned} [f \circ h]'(0, 0; \mathbf{I}_2) &= f'(h(0, 0); \mathbf{h}'(0, 0; \mathbf{I}_2)) = f'(0; [0 \quad 1]) = [0 \quad 1],\\ [f \circ h]'\left(0, 0; \left[\begin{smallmatrix} 0 & 1\\ 1 & 0 \end{smallmatrix}\right]\right) &= f'\left(h(0, 0); \mathbf{h}'\left(0, 0; \left[\begin{smallmatrix} 0 & 1\\ 1 & 0 \end{smallmatrix}\right]\right)\right) = f'(0; [0 \quad 1]) = [0 \quad 1],\\ [f \circ h]'(0, 0; -\mathbf{I}_2) &= f'(h(0, 0); \mathbf{h}'(0, 0; -\mathbf{I}_2)) = f'(0; [0 \quad 0]) = [0 \quad 0], \end{aligned}$$

as above.

3. Nonsmooth inverse and implicit functions: Existence and generalized derivative information. Clarke provided inverse and implicit function theorems for locally Lipschitz continuous functions (Theorem 7.1.1 and its corollary in [3]); a Lipschitzian function admits a local inverse (implicit) function near one of its domain points (zeroes) if Clarke Jacobian projections contain no singular matrices at said domain point (zero). However, these two results do not describe generalized derivative information of the inverse or implicit functions.

Khan and Barton [22] described generalized derivative information for inverse functions, in the form of an LD-derivative, assuming L-smoothness of the participating functions (see [22, Theorem 1]). A sufficient condition less restrictive than Clarke Jacobian projections containing no singular matrices exists for the piecewise differentiable case due to the work of Ralph and Scholtes [34]. Using the present terminology and the remark following [34, Definition 16], coherent orientation and complete coherent orientation of piecewise differentiable functions are defined.

Definition 3.1. *Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ be open and $\mathbf{g} : X \times Y \to \mathbb{R}^m$ PC^r at $(\mathbf{x}^0, \mathbf{y}^0) \in X \times Y$. The function $\mathbf{g}$ is called coherently oriented with respect to $\mathbf{y}$ at $(\mathbf{x}^0, \mathbf{y}^0)$ if all matrices in $\partial^{\mathrm{B}}_{\mathbf{y}}\mathbf{g}(\mathbf{x}^0, \mathbf{y}^0)$ have the same nonvanishing determinant sign. The function $\mathbf{g}$ is called completely coherently oriented with respect to $\mathbf{y}$ at $(\mathbf{x}^0, \mathbf{y}^0)$ if all matrices in $\Lambda_{\mathbf{y}}\mathbf{g}(\mathbf{x}^0, \mathbf{y}^0)$ have the same nonvanishing determinant sign. $\mathbf{g}$ is called (completely) coherently oriented with respect to $\mathbf{y}$ on $W \subset X \times Y$ if it is (completely) coherently oriented at each $(\mathbf{x}, \mathbf{y}) \in W$. $\mathbf{g}$ is called (completely) coherently oriented with respect to $\mathbf{y}$ if it is (completely) coherently oriented with respect to $\mathbf{y}$ on $X \times Y$.*

As noted by Ralph and Scholtes [34], complete coherent orientation of piecewise affine functions is a natural generalization of the P-matrix property (i.e., every principal minor has positive determinant sign) for linear complementarity problems. The authors provide the example

$$\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2 : \mathbf{x} \mapsto \begin{bmatrix} \min(x_1, 0.5x_1 + 0.5x_2) \\ \min(x_2, 0.5x_1 + 0.5x_2) \end{bmatrix},$$

which is coherently oriented on its domain, but not completely coherently oriented. (If the dimension of the preimage and image space of a function are equal, then the definition of (complete) coherent orientation extends as expected using the (combinatorial Jacobian) B-subdifferential instead of the (combinatorial partial Jacobian) partial B-subdifferential of said function [34].) In particular,

$$\partial^{\mathrm{B}}\mathbf{f}(0,0) = \left\{ \begin{bmatrix} 1 & 0 \\ 0.5 & 0.5 \end{bmatrix}, \begin{bmatrix} 0.5 & 0.5 \\ 0 & 1 \end{bmatrix} \right\},$$

but the matrix

$$\begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \in \Lambda\mathbf{f}(0,0)$$

has determinant equal to zero.

Theorem 1 in [22] is adapted here, including the piecewise differentiable case since a PC^r function is L-smooth.

Theorem 3.2. *Let $X \subset \mathbb{R}^n$ be open and $\mathbf{f} : X \to \mathbb{R}^n$ be L-smooth (PC^r) at $\mathbf{x}^0 \in X$. If $\mathbf{f}$ is a Lipschitz (PC^r-) homeomorphism at $\mathbf{x}^0$, then the corresponding local inverse function $\mathbf{f}^{-1}$ of $\mathbf{f}$ around $\mathbf{x}^0$ is L-smooth (PC^r) at $\mathbf{y}^0 \equiv \mathbf{f}(\mathbf{x}^0)$; for any*

$k \in \mathbb{N}$ and any $\mathbf{M} \in \mathbb{R}^{n\times k}$, $[\mathbf{f}^{-1}]'(\mathbf{y}^0;\mathbf{M})$ is the unique solution $\mathbf{N} \in \mathbb{R}^{n\times k}$ of the equation system

$$\mathbf{f}'(\mathbf{x}^0;\mathbf{N}) = \mathbf{M}. \tag{3}$$

Remark 3.3. As noted by the authors of [22], if $\mathbf{f}$ in Theorem 3.2 is L-smooth at $\mathbf{x}^0$, then $\mathbf{f}$ is a Lipschitz homeomorphism at $\mathbf{x}^0$ if $\partial\mathbf{f}(\mathbf{x}^0)$ contains no singular matrices by [3, Theorem 7.1.1]. Theorem 3.2 also augments [34, Theorem 5] with generalized derivative information; if $\mathbf{f}$ in Theorem 3.2 is PC^r at $\mathbf{x}^0$, then $\mathbf{f}$ is a PC^r-homemorphism at $\mathbf{x}^0$ if and only if $\mathbf{f}$ is coherently oriented at $\mathbf{x}^0$ and the directional derivative mapping $\mathbf{f}'(\mathbf{x}^0;\cdot)$ is invertible. (Other equivalencies are presented in [34, Theorem 5].)

Implicit functions can be built from homeomorphisms; the zero of a function determines an implicit function if and only if an auxiliary mapping admits a local inverse at said zero (see [37, Lemma 3.2.1]). Theorem 2 in [22] is adapted here and augmented with the PC^r case since complete coherent orientation of a PC^r function at one of its zeros gives a PC^r implicit function (see [34, Corollary 20]).

THEOREM 3.4. *Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ be open and $\mathbf{g} : X \times Y \to \mathbb{R}^m$ L-smooth (PC^r) at $(\mathbf{x}^0,\mathbf{y}^0) \in X \times Y$. Suppose that $\mathbf{g}(\mathbf{x}^0,\mathbf{y}^0) = \mathbf{0}_m$ and, in addition, the auxiliary mapping $\mathbf{f} : X \times Y \to \mathbb{R}^n \times \mathbb{R}^m : (\mathbf{x},\mathbf{y}) \mapsto (\mathbf{x},\mathbf{g}(\mathbf{x},\mathbf{y}))$ is a Lipschitz (PC^r-) homeomorphism at $(\mathbf{x}^0,\mathbf{y}^0)$. Then, there exist a neighborhood $N_{\mathbf{x}^0} \subset X$ of $\mathbf{x}^0$ and a Lipschitz continuous (PC^r-) function $\mathbf{r} : N_{\mathbf{x}^0} \to \mathbb{R}^m$ such that, for each $\mathbf{x} \in N_{\mathbf{x}^0}$, $(\mathbf{x},\mathbf{r}(\mathbf{x}))$ is the unique vector in a neighborhood of $(\mathbf{x}^0,\mathbf{y}^0)$ satisfying $\mathbf{g}(\mathbf{x},\mathbf{r}(\mathbf{x})) = \mathbf{0}_m$. Moreover, $\mathbf{r}$ is L-smooth (PC^r) at $\mathbf{x}^0$; for any $k \in \mathbb{N}$ and any $\mathbf{M} \in \mathbb{R}^{n\times k}$, the LD-derivative $\mathbf{r}'(\mathbf{x}^0;\mathbf{M})$ is the unique solution $\mathbf{N} \in \mathbb{R}^{m\times k}$ of the equation system*

$$\mathbf{g}'(\mathbf{x}^0,\mathbf{y}^0;(\mathbf{M},\mathbf{N})) = \mathbf{0}_{m\times k}. \tag{4}$$

In addition, the following statements hold:

(i) *If $\mathbf{g}$ is L-smooth at $(\mathbf{x}^0,\mathbf{y}^0)$ and $\pi_y\partial\mathbf{g}(\mathbf{x}^0,\mathbf{y}^0)$ contains no singular matrices, then $\mathbf{f}$ is a Lipschitz homeomorphism at $(\mathbf{x}^0,\mathbf{y}^0)$.*

(ii) *If $\mathbf{g}$ is PC^r at $(\mathbf{x}^0,\mathbf{y}^0)$ and completely coherently oriented with respect to $\mathbf{y}$ at $(\mathbf{x}^0,\mathbf{y}^0)$, then $\mathbf{f}$ is a PC^r-homeomorphism at $(\mathbf{x}^0,\mathbf{y}^0)$.*

Proof. The first part of the theorem is proved: the case in which $\mathbf{g}$ is L-smooth at $(\mathbf{x}^0,\mathbf{y}^0)$ and $\mathbf{f}$ is a Lipschitz homeomorphism at $(\mathbf{x}^0,\mathbf{y}^0)$ is a restatement of [22, Theorem 2]. Suppose that $\mathbf{g}$ is PC^r at $(\mathbf{x}^0,\mathbf{y}^0)$ and $\mathbf{f}$ is a PC^r-homeomorphism at $(\mathbf{x}^0,\mathbf{y}^0)$. According to [37, Lemma 3.2.1], the zero $(\mathbf{x}^0,\mathbf{y}^0)$ of $\mathbf{g}$ implies the existence of such a PC^r implicit function $\mathbf{r}$ since $\mathbf{r}(\mathbf{x}) = \mathbf{f}_{\mathbf{y}}^{-1}(\mathbf{x},\mathbf{0}_m)$ $\forall\mathbf{x} \in N_{\mathbf{x}^0}$, where $\mathbf{f}^{-1}(\mathbf{u},\mathbf{v}) \equiv (\mathbf{f}_{\mathbf{x}}^{-1}(\mathbf{u},\mathbf{v}),\mathbf{f}_{\mathbf{y}}^{-1}(\mathbf{u},\mathbf{v})) \in \mathbb{R}^n \times \mathbb{R}^m$ for $(\mathbf{u},\mathbf{v})$ in a neighborhood of $(\mathbf{x}^0,\mathbf{0}_m)$ is the corresponding PC^r local inverse of $\mathbf{f}$. Since $\mathbf{g}$ is PC^r (and therefore L-smooth) at $(\mathbf{x}^0,\mathbf{y}^0)$, [37, Proposition 4.2.1] implies that $\mathbf{f}$ is a Lipschitz homeomorphism at $\mathbf{x}^0$, from which (4) follows again by [22, Theorem 2].

Next, statement (i) follows from [40, Lemma 3.1] and [3, Theorem 7.1.1]. Statement (ii) holds by the following arguments: the PC^r function $\mathbf{r}$ is immediately furnished from [34, Corollary 20] and the auxiliary mapping $\mathbf{f}$ is therefore PC^r at $(\mathbf{x}^0,\mathbf{y}^0)$ by construction. As the zero $(\mathbf{x}^0,\mathbf{y}^0)$ of $\mathbf{g}$ implies the implicit function $\mathbf{r}$, $\mathbf{f}$ admits a local inverse $\mathbf{f}^{-1}$ at $(\mathbf{x}^0,\mathbf{y}^0)$ which satisfies $\mathbf{r}(\mathbf{x}) = \mathbf{f}_{\mathbf{y}}^{-1}(\mathbf{x},\mathbf{0}_m)$ $\forall\mathbf{x} \in N_{\mathbf{x}^0}$ by [37, Lemma 3.2.1]. Moreover, since $\mathbf{f}_{\mathbf{x}}^{-1}$ is the identity mapping, $\mathbf{f}^{-1}$ is PC^r at $\mathbf{x}^0$ since $\mathbf{r}$ is PC^r on $N_{\mathbf{x}^0}$. ∎

Numerical solution of (3) and (4) can be computed practically using the following lemmata in an approach that is described subsequently.

LEMMA 3.5. *Let $X \subset \mathbb{R}^n$ be open and $\mathbf{f}: X \to \mathbb{R}^m$ L-smooth at $\mathbf{x} \in X$. Given $\mathbf{M} \in \mathbb{R}^{n\times k}$ and some $j \in \{0, 1, \ldots, k-1\}$, define a function $\mathbf{h}: \mathbb{R}^n \to \mathbb{R}^m : \mathbf{d} \mapsto \mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}}(\mathbf{d})$. Then, $\mathbf{h}$ is L-smooth on $\mathbb{R}^n$; for any $\mathbf{d} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{n\times q}$,*

$$\mathbf{h}'(\mathbf{d}; \mathbf{A}) = \mathbf{f}'(\mathbf{x}; [\mathbf{M}_{(j)} \quad \mathbf{d} \quad \mathbf{A}]) \begin{bmatrix} \mathbf{0}_{(j+1)\times q} \\ \mathbf{I}_q \end{bmatrix},$$

where $\mathbf{M}_{(0)} \equiv \emptyset_{n\times 0}$.

Proof. Choose $\mathbf{d} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{n\times q}$, and set $\mathbf{B} \equiv [\mathbf{M}_{(j)} \quad \mathbf{d} \quad \mathbf{A}] \in \mathbb{R}^{n\times(j+1+q)}$. It follows from the definition of the intermediate directional derivatives $\mathbf{f}^{(i)}_{\mathbf{x},\mathbf{M}}$ that $\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{M}} \equiv \mathbf{f}^{(j)}_{\mathbf{x},\mathbf{B}}$, and so $\mathbf{h} \equiv \mathbf{f}^{(j)}_{\mathbf{x},\mathbf{B}}$. Thus, $\mathbf{h}$ is L-smooth, and

$$\mathbf{h}^{(0)}_{\mathbf{d},\mathbf{A}} \equiv \left[\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{B}}\right]'(\mathbf{d}; \cdot) \equiv \left[\mathbf{f}^{(j)}_{\mathbf{x},\mathbf{B}}\right]'(\mathbf{b}_{(j+1)}; \cdot) \equiv \mathbf{f}^{(j+1)}_{\mathbf{x},\mathbf{B}}.$$

Starting from this equivalence, the following inductive argument then shows that $\mathbf{h}^{(p)}_{\mathbf{d},\mathbf{A}} \equiv \mathbf{f}^{(j+1+p)}_{\mathbf{x},\mathbf{B}}$ for each $p \in \{0, 1, \ldots, q-1\}$. Assume that this statement is true for some $p \in \{0, 1 \ldots, q-2\}$; taking directional derivatives then yields

$$\mathbf{h}^{(p+1)}_{\mathbf{d},\mathbf{A}} \equiv \left[\mathbf{h}^{(p)}_{\mathbf{d},\mathbf{A}}\right]'(\mathbf{a}_{(p+1)}; \cdot) \equiv \left[\mathbf{f}^{(j+1+p)}_{\mathbf{x},\mathbf{B}}\right]'(\mathbf{a}_{(p+1)}; \cdot) \equiv \left[\mathbf{f}^{(j+1+p)}_{\mathbf{x},\mathbf{B}}\right]'(\mathbf{b}_{(j+2+p)}; \cdot),$$

from which it follows that $\mathbf{h}^{(p+1)}_{\mathbf{d},\mathbf{A}} \equiv \mathbf{f}^{(j+1+(p+1))}_{\mathbf{x},\mathbf{B}}$, completing the inductive step. The final claim of the lemma then follows immediately from the constructive definition of the LD-derivative. ∎

LEMMA 3.6. *Assume the setting of Theorem 3.2 with $\mathbf{f}$ an L-smooth, Lipschitz homeomorphism at $\mathbf{x}^0$. The jth column of $\mathbf{N} \equiv [\mathbf{f}^{-1}]'(\mathbf{y}^0; \mathbf{M})$ is the unique solution $\mathbf{n}$ of the equation system*

$$\mathbf{0}_n = \mathbf{f}^{(j-1)}_{\mathbf{x}^0,\mathbf{N}_{(j-1)}}(\mathbf{n}) - \mathbf{m}_{(j)}.$$

Denote the residual function for this equation system as $\mathbf{h}: \mathbb{R}^n \to \mathbb{R}^n : \mathbf{d} \mapsto \mathbf{f}^{(j-1)}_{\mathbf{x}^0,\mathbf{N}_{(j-1)}}(\mathbf{d}) - \mathbf{m}_{(j)}$. Then, $\mathbf{h}$ is L-smooth on $\mathbb{R}^n$; for any $\mathbf{d} \in \mathbb{R}^n$ and any $\mathbf{A} \in \mathbb{R}^{n\times q}$,

$$\mathbf{h}'(\mathbf{d}; \mathbf{A}) = \mathbf{f}'(\mathbf{x}^0; [\mathbf{N}_{(j-1)} \quad \mathbf{d} \quad \mathbf{A}]) \begin{bmatrix} \mathbf{0}_{j\times q} \\ \mathbf{I}_q \end{bmatrix}.$$

Proof. Consider any $\mathbf{d} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{n\times q}$, define an auxiliary mapping $\mathbf{r} \equiv \mathbf{f}^{(j-1)}_{\mathbf{x}^0,\mathbf{N}_{(j-1)}}$, and set $\mathbf{B} \equiv [\mathbf{N}_{(j-1)} \quad \mathbf{d} \quad \mathbf{A}] \in \mathbb{R}^{n\times(j+q)}$. According to Lemma 3.5, $\mathbf{r}$ is L-smooth on $\mathbb{R}^n$, and

$$\mathbf{r}'(\mathbf{d}; \mathbf{A}) = \mathbf{f}'(\mathbf{x}^0; \mathbf{B}) \begin{bmatrix} \mathbf{0}_{j\times q} \\ \mathbf{I}_q \end{bmatrix}.$$

Observe that $\mathbf{h}$ is the mapping $\mathbf{v} \mapsto \mathbf{r}(\mathbf{v}) - \mathbf{m}_{(j)}$; the chain rule for LD-derivatives implies that $\mathbf{h}$ is L-smooth on $\mathbb{R}^n$ and that $\mathbf{h}'(\mathbf{d}; \mathbf{A}) = \mathbf{r}'(\mathbf{d}; \mathbf{A})$, as required. ∎

LEMMA 3.7. *Assume the setting of Theorem 3.4 with $\mathbf{g}$ L-smooth at $(\mathbf{x}^0, \mathbf{y}^0)$ and $\mathbf{f}$ a Lipschitz homeomorphism at $(\mathbf{x}^0, \mathbf{y}^0)$. The jth column of $\mathbf{N} \equiv \mathbf{r}'(\mathbf{x}^0; \mathbf{M})$ is the unique solution $\mathbf{n}$ of the equation system*

$$\mathbf{0}_m = \mathbf{g}^{(j-1)}_{\begin{bmatrix}\mathbf{x}^0\\ \mathbf{y}^0\end{bmatrix}, \begin{bmatrix}\mathbf{M}_{(j-1)}\\ \mathbf{N}_{(j-1)}\end{bmatrix}}\left(\begin{bmatrix}\mathbf{m}_{(j)} \\ \mathbf{n}\end{bmatrix}\right). \tag{5}$$

Denote the residual function for this equation system as

$$\mathbf{h} : \mathbb{R}^m \to \mathbb{R}^m : \mathbf{d} \mapsto \mathbf{g}^{(j-1)}_{\begin{bmatrix}\mathbf{x}^0\\ \mathbf{y}^0\end{bmatrix}, \begin{bmatrix}\mathbf{M}_{(j-1)}\\ \mathbf{N}_{(j-1)}\end{bmatrix}}\left(\begin{bmatrix}\mathbf{m}_{(j)}\\ \mathbf{d}\end{bmatrix}\right).$$

Then, the function $\mathbf{h}$ *is L-smooth on* $\mathbb{R}^m$*; for any* $\mathbf{d} \in \mathbb{R}^m$ *and any* $\mathbf{A} \in \mathbb{R}^{m\times q}$*,*

$$\mathbf{h}'(\mathbf{d}; \mathbf{A}) = \mathbf{g}'\left(\begin{bmatrix}\mathbf{x}^0\\ \mathbf{y}^0\end{bmatrix}; \begin{bmatrix}\mathbf{M}_{(j-1)} & \mathbf{m}_{(j)} & \mathbf{0}_{n\times q}\\ \mathbf{N}_{(j-1)} & \mathbf{d} & \mathbf{A}\end{bmatrix}\right)\begin{bmatrix}\mathbf{0}_{j\times q}\\ \mathbf{I}_q\end{bmatrix}. \tag{6}$$

Proof. Choose any $\mathbf{d} \in \mathbb{R}^m$ and $\mathbf{A} \in \mathbb{R}^{m\times q}$, and define an auxiliary mapping

$$\mathbf{r} : \mathbb{R}^{n+m} \to \mathbb{R}^m : \mathbf{v} \mapsto \mathbf{g}^{(j-1)}_{\begin{bmatrix}\mathbf{x}^0\\ \mathbf{y}^0\end{bmatrix}, \begin{bmatrix}\mathbf{M}_{(j-1)}\\ \mathbf{N}_{(j-1)}\end{bmatrix}}(\mathbf{v}).$$

According to Lemma 3.5, $\mathbf{r}$ is L-smooth on $\mathbb{R}^{n+m}$, and, for any $\mathbf{v} \in \mathbb{R}^{n+m}$ and $\mathbf{C} \in \mathbb{R}^{(n+m)\times q}$,

$$\mathbf{r}'(\mathbf{v}; \mathbf{C}) = \mathbf{g}'\left(\begin{bmatrix}\mathbf{x}^0\\ \mathbf{y}^0\end{bmatrix}; \begin{bmatrix}\begin{bmatrix}\mathbf{M}_{(j-1)}\\ \mathbf{N}_{(j-1)}\end{bmatrix} & \mathbf{v} & \mathbf{C}\end{bmatrix}\right)\begin{bmatrix}\mathbf{0}_{j\times q}\\ \mathbf{I}_q\end{bmatrix}. \tag{7}$$

By construction of $\mathbf{r}$, for any $\mathbf{w} \in \mathbb{R}^m$,

$$\mathbf{h}(\mathbf{w}) = \mathbf{r}\left(\begin{bmatrix}\mathbf{m}_{(j)}\\ \mathbf{w}\end{bmatrix}\right);$$

the chain rule for LD-derivatives then implies that $\mathbf{h}$ is L-smooth on $\mathbb{R}^m$ and that

$$\mathbf{h}'(\mathbf{d}; \mathbf{A}) = \mathbf{r}'\left(\begin{bmatrix}\mathbf{m}_{(j)}\\ \mathbf{d}\end{bmatrix}; \begin{bmatrix}\mathbf{0}_{n\times q}\\ \mathbf{A}\end{bmatrix}\right).$$

The claimed result then follows from (7). ∎

Lemma 3.7 provides a way to solve the nonsmooth and nonlinear equation system (4). Assuming the settings of Theorem 3.4 and Lemma 3.7, the mapping $\mathbf{N} \mapsto \mathbf{g}'(\mathbf{x}^0, \mathbf{y}^0; (\mathbf{M}, \mathbf{N}))$ is not necessarily continuous, but the mappings

$$\mathbf{n} \mapsto \mathbf{g}^{(j-1)}_{\begin{bmatrix}\mathbf{x}^0\\ \mathbf{y}^0\end{bmatrix}, \begin{bmatrix}\mathbf{M}_{(j-1)}\\ \mathbf{N}_{(j-1)}\end{bmatrix}}\left(\begin{bmatrix}\mathbf{m}_{(j)}\\ \mathbf{n}\end{bmatrix}\right)$$

are continuous for each $j \in \{1, \ldots, k\}$ [22]. Consequently, (4) can be decomposed columnwise and solved from left to right, using a nonsmooth equation-solving method for each columnwise solve (i.e., solving (5) for $j = 1, 2, \ldots, k$). Invoking nonsmooth equation-solving methods requires a generalized derivative element at each iteration, which can be furnished from (6); for example, the $(l+1)$th iteration of a nonsmooth Newton method is obtained by solving the linear equation system

$$\mathbf{h}'(\mathbf{v}_{(l)}; \mathbf{I}_m)\,(\mathbf{v}_{(l+1)} - \mathbf{v}_{(l)}) = -\mathbf{h}(\mathbf{v}_{(l)}) \tag{8}$$

for $\mathbf{v}_{(l+1)}$, where $\mathbf{h}'(\mathbf{v}_{(l)}; \mathbf{I}_m)$ is given by (6) and can be computed by, for example, a nonsmooth vector forward mode of automatic differentiation [21]. The iteration scheme (8) produces $\mathbf{n}_{(j)}$ (i.e., the jth column of $\mathbf{N} \equiv \mathbf{r}'(\mathbf{x}^0; \mathbf{M})$).

If $\mathbf{g}$ is PC^1 at $(\mathbf{x}^0, \mathbf{y}^0)$, then Proposition 2 in [22] can also be applied to compute $\mathbf{N}$, which cycles through a set of essentially active selection functions and performs linear equation solves per cycle. As remarked in [22] after Proposition 2, said algorithm scales worst-case linearly with respect to the number of selection functions and according to the linear equation solves needed. (Similarly, Proposition 1 in [22] may be applied instead of Lemma 3.6 in the PC^1 case.) More recently, a branch-locking procedure has also been developed [18, section 4] to solve (4) more efficiently.

4. Specialization of results to the minimum function. To calculate parametric sensitivities of the motivating problem (i.e., NLP KKT nonsmooth equation systems), LD-derivatives of the mapping **min** are detailed in this section. Let the generalized inequalities $\prec$ and $\preceq$ denote lexicographic ordering; given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$\mathbf{x} \prec \mathbf{y} \text{ if and only if } \exists j \in \{1, \dots, n\} \text{ s.t. } x_i = y_i \; \forall i < j \text{ and } x_j < y_j,$$
$$\mathbf{x} \preceq \mathbf{y} \text{ if and only if } \mathbf{x} = \mathbf{y} \text{ or } \mathbf{x} \prec \mathbf{y}.$$

The generalized inequalities $\succ$ and $\succeq$ are defined similarly. Let the *lexicographic-minimum function* return the lexicographically ordered minimum of two vectors:

$$\mathbf{Lmin} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n : (\mathbf{x}, \mathbf{y}) \mapsto \begin{cases} \mathbf{x} & \text{if } \mathbf{x} \preceq \mathbf{y}, \\ \mathbf{y} & \text{if } \mathbf{x} \succ \mathbf{y}. \end{cases}$$

The *lexicographic-matrix-minimum*, which compares two matrices lexicographically (by rows), is defined as

$$\mathbf{LMmin} : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n} : (\mathbf{X}, \mathbf{Y}) \mapsto \begin{bmatrix} (\mathbf{Lmin}(\mathbf{X}_1^{\mathrm{T}}, \mathbf{Y}_1^{\mathrm{T}}))^{\mathrm{T}} \\ (\mathbf{Lmin}(\mathbf{X}_2^{\mathrm{T}}, \mathbf{Y}_2^{\mathrm{T}}))^{\mathrm{T}} \\ \vdots \\ (\mathbf{Lmin}(\mathbf{X}_m^{\mathrm{T}}, \mathbf{Y}_m^{\mathrm{T}}))^{\mathrm{T}} \end{bmatrix}.$$

Given $n \in \mathbb{N}$, $n \geq 2$, the *shifted-lexicographic-minimum* is defined as

$$\mathbf{SLmin} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{n-1} : (\mathbf{x}, \mathbf{y}) \mapsto \begin{cases} (x_2, \dots, x_n) & \text{if } \mathbf{x} \preceq \mathbf{y}, \\ (y_2, \dots, y_n) & \text{if } \mathbf{x} \succ \mathbf{y}. \end{cases}$$

The *shifted-lexicographic-matrix-minimum* $\mathbf{SLMmin} : \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times (n-1)}$ is defined similarly as **LMmin** with **Lmin** replaced by **SLmin** and returns the shifted-lexicographic-minimums of the rows of $\mathbf{X}$ and $\mathbf{Y}$.

LEMMA 4.1. *The LD-derivative of the componentwise minimum function* **min** : $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ *at* $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ *in the directions* $(\mathbf{M}, \mathbf{N}) \in \mathbb{R}^{2n \times k}$ *evaluates as*

$$\mathbf{min}'(\mathbf{x}, \mathbf{y}; (\mathbf{M}, \mathbf{N})) = \mathbf{SLMmin}\left([\mathbf{x} \quad \mathbf{M}], [\mathbf{y} \quad \mathbf{N}]\right) \in \mathbb{R}^{n \times k}.$$

Proof. Let $\mathbf{M} = [\mathbf{m}_{(1)} \quad \cdots \quad \mathbf{m}_{(k)}] \in \mathbb{R}^{n \times k}$ and $\mathbf{N} = [\mathbf{n}_{(1)} \quad \cdots \quad \mathbf{n}_{(k)}] \in \mathbb{R}^{n \times k}$. By virtue of [21, Example 4.3], the LD-derivative of min at (x_1, y_1) in the directions

$$\mathbf{D}_{(1)} \equiv \begin{bmatrix} m_{(1),1} & m_{(2),1} & \cdots & m_{(k),1} \\ n_{(1),1} & n_{(2),1} & \cdots & n_{(k),1} \end{bmatrix} = \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{N}_1 \end{bmatrix} \in \mathbb{R}^{2 \times k}$$

is evaluated as

$$\min{}'(x_1, y_1; \mathbf{D}_{(1)}) = \begin{cases} [1 \quad 0]\mathbf{D}_{(1)} & \text{if } \operatorname{fsign}\left((x_1, \mathbf{M}_1^{\mathrm{T}}) - (y_1, \mathbf{N}_1^{\mathrm{T}})\right) \leq 0, \\ [0 \quad 1]\mathbf{D}_{(1)} & \text{if } \operatorname{fsign}\left((x_1, \mathbf{M}_1^{\mathrm{T}}) - (y_1, \mathbf{N}_1^{\mathrm{T}})\right) > 0, \end{cases}$$

where the first-sign function [14] is given as

$$\operatorname{fsign} : \mathbb{R}^q \to \{-1, 0, 1\} : \boldsymbol{\eta} \mapsto \begin{cases} \operatorname{sign}(\eta_{i^*}), \text{ with } i^* \equiv \min\{i : \eta_i \neq 0\}, & \text{if } \boldsymbol{\eta} \neq \mathbf{0}_q, \\ 0 & \text{if } \boldsymbol{\eta} = \mathbf{0}_q, \end{cases}$$

and sign$(\cdot)$ denotes the signum function. Noting that $\mathbf{x} \preceq \mathbf{y}$ if and only if fsign$(\mathbf{x}-\mathbf{y}) \leq 0$ and $\mathbf{x} \succ \mathbf{y}$ if and only if fsign$(\mathbf{x}-\mathbf{y}) > 0$, it follows that

$$\begin{aligned}\min{}'(x_1, y_1; \mathbf{D}_{(1)}) &= \left(\mathbf{SLmin}\left((x_1, \mathbf{M}_1^{\mathrm{T}}) - (y_1, \mathbf{N}_1^{\mathrm{T}})\right)\right)^{\mathrm{T}} \\ &= (\mathbf{SLmin}([\mathbf{x} \quad \mathbf{M}]^{\mathrm{T}}\mathbf{e}_{(1)}, [\mathbf{y} \quad \mathbf{N}]^{\mathrm{T}}\mathbf{e}_{(1)}))^{\mathrm{T}}.\end{aligned}$$

The result follows by definition of the shifted-lexicographic-matrix-minimum and noting that $\mathbf{min}'(\mathbf{x}, \mathbf{y}; (\mathbf{M}, \mathbf{N})) = (\min'(x_1, y_1; \mathbf{D}_{(1)}), \ldots, \min'(x_n, y_n; \mathbf{D}_{(n)}))$, where

$$\mathbf{D}_{(i)} \equiv \begin{bmatrix} m_{(1),i} & m_{(2),i} & \cdots & m_{(k),i} \\ n_{(1),i} & n_{(2),i} & \cdots & n_{(k),i} \end{bmatrix} = \begin{bmatrix} \mathbf{M}_i \\ \mathbf{N}_i \end{bmatrix} \quad \forall i \in \{1, \ldots, n\}.$$ ∎

The LD-derivative chain rule yields LD-derivatives of compositions of minimum and C^1 functions.

Lemma 4.2. *Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ be open. Let $\mathbf{f} : X \times Y \to \mathbb{R}^v$ and $\mathbf{g} : X \times Y \to \mathbb{R}^v$ be C^1 at $(\mathbf{x}, \mathbf{y}) \in X \times Y$. The LD-derivative of the mapping $\mathbf{min} \circ (\mathbf{f}, \mathbf{g}) : X \times Y \to \mathbb{R}^v$ at $(\mathbf{x}, \mathbf{y})$ in the directions $(\mathbf{M}, \mathbf{N}) \in \mathbb{R}^{(n+m)\times k}$ is given by*

$$\begin{aligned}&[\mathbf{min} \circ (\mathbf{f}, \mathbf{g})]'(\mathbf{x}, \mathbf{y}; (\mathbf{M}, \mathbf{N})) \\ &\quad = \mathbf{SLMmin}([\mathbf{f}(\mathbf{x}, \mathbf{y}) \quad \mathbf{Jf}(\mathbf{x}, \mathbf{y})(\mathbf{M}, \mathbf{N})], [\mathbf{g}(\mathbf{x}, \mathbf{y}) \quad \mathbf{Jg}(\mathbf{x}, \mathbf{y})(\mathbf{M}, \mathbf{N})]) \in \mathbb{R}^{v\times k}.\end{aligned}$$

Proof. Define the mapping $\mathbf{q} : X \times Y \to \mathbb{R}^{2v} : (\boldsymbol{\eta}_x, \boldsymbol{\eta}_y) \mapsto (\mathbf{f}(\boldsymbol{\eta}_x, \boldsymbol{\eta}_y), \mathbf{g}(\boldsymbol{\eta}_x, \boldsymbol{\eta}_y))$. Then,

$$\mathbf{Jq}(\mathbf{x}, \mathbf{y}) = \begin{bmatrix} \mathbf{Jf}(\mathbf{x}, \mathbf{y}) \\ \mathbf{Jg}(\mathbf{x}, \mathbf{y}) \end{bmatrix} \in \mathbb{R}^{2v\times(n+m)},$$

and, by the LD-derivative chain rule (2) and Lemma 4.1,

$$\begin{aligned}&[\mathbf{min} \circ (\mathbf{f}, \mathbf{g})]'(\mathbf{x}, \mathbf{y}; (\mathbf{M}, \mathbf{N})) \\ &\quad = \mathbf{min}'(\mathbf{q}(\mathbf{x}, \mathbf{y}); \mathbf{q}'(\mathbf{x}, \mathbf{y}; (\mathbf{M}, \mathbf{N}))) \\ &\quad = \mathbf{min}'\left(\begin{bmatrix} \mathbf{f}(\mathbf{x}, \mathbf{y}) \\ \mathbf{g}(\mathbf{x}, \mathbf{y}) \end{bmatrix}; \mathbf{Jq}(\mathbf{x}, \mathbf{y}) \begin{bmatrix} \mathbf{M} \\ \mathbf{N} \end{bmatrix}\right) \\ &\quad = \mathbf{min}'\left(\begin{bmatrix} \mathbf{f}(\mathbf{x}, \mathbf{y}) \\ \mathbf{g}(\mathbf{x}, \mathbf{y}) \end{bmatrix}; \begin{bmatrix} \mathbf{Jf}(\mathbf{x}, \mathbf{y}) \\ \mathbf{Jg}(\mathbf{x}, \mathbf{y}) \end{bmatrix} \begin{bmatrix} \mathbf{M} \\ \mathbf{N} \end{bmatrix}\right) \\ &\quad = \mathbf{SLMmin}([\mathbf{f}(\mathbf{x}, \mathbf{y}) \quad \mathbf{Jf}(\mathbf{x}, \mathbf{y})(\mathbf{M}, \mathbf{N})], [\mathbf{g}(\mathbf{x}, \mathbf{y}) \quad \mathbf{Jg}(\mathbf{x}, \mathbf{y})(\mathbf{M}, \mathbf{N})]),\end{aligned}$$

as required. ∎

A specialization of Theorem 3.4 to a useful function form is given; generalized derivative information is obtained for an MiCP function.

Theorem 4.3. *Let $X \subset \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ be open. Let $\mathbf{f} : X \times Y \to \mathbb{R}^q$, $\mathbf{g} : X \times Y \to \mathbb{R}^q$, and $\mathbf{h} : X \times Y \to \mathbb{R}^{m-q}$ be C^1 at $(\mathbf{x}^0, \mathbf{y}^0) \in X \times Y$. Suppose that $\boldsymbol{\Theta}_{\min}(\mathbf{x}^0, \mathbf{y}^0) = \mathbf{0}_m$, where*

$$\boldsymbol{\Theta}_{\min} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m : (\mathbf{x}, \mathbf{y}) \mapsto \begin{bmatrix} \mathbf{h}(\mathbf{x}, \mathbf{y}) \\ \mathbf{min}(\mathbf{f}(\mathbf{x}, \mathbf{y}), \mathbf{g}(\mathbf{x}, \mathbf{y})) \end{bmatrix}.$$

Let

$$\begin{aligned}\alpha &\equiv \{i \in \{1, \ldots, q\} : f_i(\mathbf{x}^0, \mathbf{y}^0) = 0 < g_i(\mathbf{x}^0, \mathbf{y}^0)\}, \\ \beta &\equiv \{i \in \{1, \ldots, q\} : f_i(\mathbf{x}^0, \mathbf{y}^0) = 0 = g_i(\mathbf{x}^0, \mathbf{y}^0)\}, \\ \gamma &\equiv \{i \in \{1, \ldots, q\} : f_i(\mathbf{x}^0, \mathbf{y}^0) > 0 = g_i(\mathbf{x}^0, \mathbf{y}^0)\}.\end{aligned}$$

Suppose that $\mathbf{\Theta}_{\min}$ is completely coherently oriented with respect to $\mathbf{y}$ at $(\mathbf{x}^0, \mathbf{y}^0)$. Then, there exist a neighborhood $N_{\mathbf{x}^0} \subset X$ of $\mathbf{x}^0$ and a function $\mathbf{r} : N_{\mathbf{x}^0} \to \mathbb{R}^m$ that is PC^1 on $N_{\mathbf{x}^0}$ such that, for each $\mathbf{x} \in N_{\mathbf{x}^0}$, $(\mathbf{x}, \mathbf{r}(\mathbf{x}))$ is the unique vector in a neighborhood of $(\mathbf{x}^0, \mathbf{y}^0)$ satisfying $\mathbf{\Theta}_{\min}(\mathbf{x}, \mathbf{r}(\mathbf{x})) = \mathbf{0}_m$. Moreover, for any $k \in \mathbb{N}$ and any $\mathbf{M} \in \mathbb{R}^{n\times k}$, the LD-derivative $\mathbf{r}'(\mathbf{x}^0; \mathbf{M})$ is the unique solution $\mathbf{N} \in \mathbb{R}^{m\times k}$ of the equation system

$$\mathbf{0}_{(m-q)\times k} = \mathbf{J}_{\mathbf{x}}\mathbf{h}(\mathbf{x}^0, \mathbf{y}^0)\mathbf{M} + \mathbf{J}_{\mathbf{y}}\mathbf{h}(\mathbf{x}^0, \mathbf{y}^0)\mathbf{N}, \tag{9a}$$

$$\mathbf{0}_{|\alpha|\times k} = \mathbf{J}_{\mathbf{x}}\mathbf{f}_\alpha(\mathbf{x}^0, \mathbf{y}^0)\mathbf{M} + \mathbf{J}_{\mathbf{y}}\mathbf{f}_\alpha(\mathbf{x}^0, \mathbf{y}^0)\mathbf{N}, \tag{9b}$$

$$\mathbf{0}_{|\gamma|\times k} = \mathbf{J}_{\mathbf{x}}\mathbf{g}_\gamma(\mathbf{x}^0, \mathbf{y}^0)\mathbf{M} + \mathbf{J}_{\mathbf{y}}\mathbf{g}_\gamma(\mathbf{x}^0, \mathbf{y}^0)\mathbf{N}, \tag{9c}$$

$$\mathbf{0}_{|\beta|\times k} = \mathbf{LMmin}(\mathbf{J}\mathbf{f}_\beta(\mathbf{x}^0, \mathbf{y}^0)(\mathbf{M}, \mathbf{N}), \mathbf{J}\mathbf{g}_\beta(\mathbf{x}^0, \mathbf{y}^0)(\mathbf{M}, \mathbf{N})). \tag{9d}$$

Proof. Theorem 3.4 can immediately be applied to yield the following: there exist a neighborhood $N_{\mathbf{x}^0} \subset X$ of $\mathbf{x}^0$ and PC^1 mapping $\mathbf{r} : N_{\mathbf{x}^0} \to \mathbb{R}^m$ such that, for each $\mathbf{x} \in N_{\mathbf{x}^0}$, $(\mathbf{x}, \mathbf{r}(\mathbf{x}))$ is the unique vector in a neighborhood of $(\mathbf{x}^0, \mathbf{y}^0)$ satisfying $\mathbf{\Theta}_{\min}(\mathbf{x}, \mathbf{r}(\mathbf{x})) = \mathbf{0}_m$. Moreover, for any $k \in \mathbb{N}$ and any $\mathbf{M} \in \mathbb{R}^{n\times k}$, the LD-derivative $\mathbf{r}'(\mathbf{x}^0; \mathbf{M})$ is the unique solution $\mathbf{N} \in \mathbb{R}^{m\times k}$ of the equation system

$$\mathbf{0}_{m\times k} = \mathbf{\Theta}'_{\min}(\mathbf{x}^0, \mathbf{y}^0; \mathbf{W}),$$

where $\mathbf{W} \equiv (\mathbf{M}, \mathbf{N})$, which is equivalent to

$$\begin{aligned} \mathbf{0}_{(m-q)\times k} &= \mathbf{J}_{\mathbf{x}}\mathbf{h}(\mathbf{x}^0, \mathbf{y}^0)\mathbf{M} + \mathbf{J}_{\mathbf{y}}\mathbf{h}(\mathbf{x}^0, \mathbf{y}^0)\mathbf{N}, \\ \mathbf{0}_{q\times k} &= \mathbf{SLMmin}(\begin{bmatrix}\mathbf{f}(\mathbf{x}^0, \mathbf{y}^0) & \mathbf{J}\mathbf{f}(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}, \begin{bmatrix}\mathbf{g}(\mathbf{x}^0, \mathbf{y}^0) & \mathbf{J}\mathbf{g}(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}), \end{aligned} \tag{10}$$

by Lemma 4.2. By definition of the shifted-lexicographic-matrix-minimum, (10) is equivalent to the equation system

$$\mathbf{0}_{|\alpha|\times k} = \mathbf{SLMmin}\left(\begin{bmatrix}\mathbf{f}_\alpha(\mathbf{x}^0, \mathbf{y}^0) & \mathbf{J}\mathbf{f}_\alpha(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}, \begin{bmatrix}\mathbf{g}_\alpha(\mathbf{x}^0, \mathbf{y}^0) & \mathbf{J}\mathbf{g}_\alpha(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}\right), \tag{11}$$

$$\mathbf{0}_{|\gamma|\times k} = \mathbf{SLMmin}\left(\begin{bmatrix}\mathbf{f}_\gamma(\mathbf{x}^0, \mathbf{y}^0) & \mathbf{J}\mathbf{f}_\gamma(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}, \begin{bmatrix}\mathbf{g}_\gamma(\mathbf{x}^0, \mathbf{y}^0) & \mathbf{J}\mathbf{g}_\gamma(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}\right), \tag{12}$$

$$\mathbf{0}_{|\beta|\times k} = \mathbf{SLMmin}\left(\begin{bmatrix}\mathbf{f}_\beta(\mathbf{x}^0, \mathbf{y}^0) & \mathbf{J}\mathbf{f}_\beta(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}, \begin{bmatrix}\mathbf{g}_\beta(\mathbf{x}^0, \mathbf{y}^0) & \mathbf{J}\mathbf{g}_\beta(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}\right). \tag{13}$$

By definition of the set α, $\mathbf{f}_\alpha(\mathbf{x}^0, \mathbf{y}^0) = \mathbf{0}_{|\alpha|} < \mathbf{g}_\alpha(\mathbf{x}^0, \mathbf{y}^0)$ (i.e., componentwise), which gives that

$$\begin{aligned} &\mathbf{SLMmin}\left(\begin{bmatrix}\mathbf{f}_\alpha(\mathbf{x}^0, \mathbf{y}^0) & \mathbf{J}\mathbf{f}_\alpha(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}, \begin{bmatrix}\mathbf{g}_\alpha(\mathbf{x}^0, \mathbf{y}^0) & \mathbf{J}\mathbf{g}_\alpha(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}\right) \\ &\quad = \mathbf{SLMmin}\left(\begin{bmatrix}\mathbf{0}_{|\alpha|} & \mathbf{J}\mathbf{f}_\alpha(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}, \begin{bmatrix}\mathbf{g}_\alpha(\mathbf{x}^0, \mathbf{y}^0) & \mathbf{J}\mathbf{g}_\alpha(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}\right) \\ &\quad = \mathbf{J}\mathbf{f}_\alpha(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}. \end{aligned}$$

Similarly,

$$\begin{aligned} &\mathbf{SLMmin}\left(\begin{bmatrix}\mathbf{f}_\gamma(\mathbf{x}^0, \mathbf{y}^0) & \mathbf{J}\mathbf{f}_\gamma(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}, \begin{bmatrix}\mathbf{g}_\gamma(\mathbf{x}^0, \mathbf{y}^0) & \mathbf{J}\mathbf{g}_\gamma(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}\right) \\ &\quad = \mathbf{SLMmin}\left(\begin{bmatrix}\mathbf{f}_\gamma(\mathbf{x}^0, \mathbf{y}^0) & \mathbf{J}\mathbf{f}_\gamma(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}, \begin{bmatrix}\mathbf{0}_{|\gamma|} & \mathbf{J}\mathbf{g}_\gamma(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}\end{bmatrix}\right) \\ &\quad = \mathbf{J}\mathbf{g}_\gamma(\mathbf{x}^0, \mathbf{y}^0)\mathbf{W}. \end{aligned}$$

Also, $\mathbf{f}_\beta(\mathbf{x}^0,\mathbf{y}^0) = \mathbf{0}_{|\beta|} = \mathbf{g}_\beta(\mathbf{x}^0,\mathbf{y}^0)$ by definition of the set β:

$$\begin{aligned}
&\mathbf{SLMmin}\left(\begin{bmatrix}\mathbf{f}_\beta(\mathbf{x}^0,\mathbf{y}^0) & \mathbf{J}\mathbf{f}_\beta(\mathbf{x}^0,\mathbf{y}^0)\mathbf{W}\end{bmatrix}, \begin{bmatrix}\mathbf{g}_\beta(\mathbf{x}^0,\mathbf{y}^0) & \mathbf{J}\mathbf{g}_\beta(\mathbf{x}^0,\mathbf{y}^0)\mathbf{W}\end{bmatrix}\right)\\
&\quad = \mathbf{SLMmin}\left(\begin{bmatrix}\mathbf{0}_{|\beta|} & \mathbf{J}\mathbf{f}_\beta(\mathbf{x}^0,\mathbf{y}^0)\mathbf{W}\end{bmatrix}, \begin{bmatrix}\mathbf{0}_{|\beta|} & \mathbf{J}\mathbf{g}_\beta(\mathbf{x}^0,\mathbf{y}^0)\mathbf{W}\end{bmatrix}\right)\\
&\quad = \mathbf{LMmin}\left(\mathbf{J}\mathbf{f}_\beta(\mathbf{x}^0,\mathbf{y}^0)\mathbf{W}, \mathbf{J}\mathbf{g}_\beta(\mathbf{x}^0,\mathbf{y}^0)\mathbf{W}\right).
\end{aligned}$$

Therefore (11), (12), and (13) yield (9b), (9c), and (9d), respectively. ∎

A sufficient condition for complete coherent orientation of the mapping outlined in Theorem 4.3 is detailed next.

LEMMA 4.4. *Assume the setting of Theorem* 4.3. *For each $i \in \{1,\dots,m\}$ and $\mathcal{J} \subset \beta$, let*

$$v_i^{(\mathcal{J})} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} : (\mathbf{x},\mathbf{y}) \mapsto \begin{cases} h_i(\mathbf{x},\mathbf{y}) & \textit{if } i \in \{1,\dots,m-q\},\\ f_i(\mathbf{x},\mathbf{y}) & \textit{if } i-(m-q) \in \alpha \cup \mathcal{J},\\ g_i(\mathbf{x},\mathbf{y}) & \textit{if } i-(m-q) \in \gamma \cup (\beta \setminus \mathcal{J}).\end{cases}$$

If all matrices in the set $\{\mathbf{J}_\mathbf{y}\mathbf{v}^{(\mathcal{J})} \in \mathbb{R}^{m\times m} : \mathcal{J} \subset \beta\}$ have the same nonvanishing determinant sign, then the mapping $\boldsymbol{\Theta}_{\min}$ is completely coherently oriented with respect to $\mathbf{y}$ at $(\mathbf{x}^0,\mathbf{y}^0)$.

Proof. By construction of the sets α, β, and γ, $\mathcal{E} \equiv \{\mathbf{v}^{(\mathcal{J})} : \mathcal{J} \subset \beta\}$ is a set of selection functions of the mapping $\boldsymbol{\Theta}_{\min}$ at $(\mathbf{x}^0,\mathbf{y}^0)$. Assume for now that $\mathcal{E}$ is a set of essentially active selection functions. Noting that $|\mathcal{E}| = 2^{|\beta|}$, enumerate the power set of β by $\mathcal{P}(\beta) = \{\mathcal{J}_1,\mathcal{J}_2,\dots,\mathcal{J}_{|\mathcal{E}|}\}$. Let $\boldsymbol{\phi}_{(i)} \equiv \mathbf{v}^{(\mathcal{J}_i)}$ for each $i \in \{1,\dots,|\mathcal{E}|\}$. Then, $\{\boldsymbol{\phi}_{(1)},\dots,\boldsymbol{\phi}_{(|\mathcal{E}|)}\} = \mathcal{E}$ is a set of essentially active selection functions of the mapping $\boldsymbol{\Theta}_{\min}$ at $(\mathbf{x}^0,\mathbf{y}^0)$, with essentially active indices $I^{\text{ess}}_{\boldsymbol{\Theta}_{\min}}(\mathbf{x}^0,\mathbf{y}^0) = \{1,\dots,|\mathcal{E}|\}$. The combinatorial vectorization of the essentially active indices is thus given by

$$\Delta_{\boldsymbol{\Theta}_{\min}}(\mathbf{x}^0,\mathbf{y}^0) = \{\boldsymbol{\delta} \in \mathbb{R}^m : \delta_i \in \{1,\dots,|\mathcal{E}|\}, i \in \{1,\dots,m\}\} = \{1,\dots,|\mathcal{E}|\}^m.$$

For any $\mathcal{J}_1,\mathcal{J}_2 \in \mathcal{P}(\beta)$, $v_i^{(\mathcal{J}_1)} = v_i^{(\mathcal{J}_2)}$ for every $i \in \alpha\cup\gamma$ by definition of the mappings $\mathbf{v}^{(\mathcal{J})}$. Consequently, $\mathbf{J}_\mathbf{y} v_i^{(\mathcal{J}_1)}(\mathbf{x}^0,\mathbf{y}^0) = \mathbf{J}_\mathbf{y} v_i^{(\mathcal{J}_2)}(\mathbf{x}^0,\mathbf{y}^0)$ and, for any $i,l \in I^{\text{ess}}_{\boldsymbol{\Theta}_{\min}}$, $\mathbf{J}_\mathbf{y}\phi_{(i),j}(\mathbf{x}^0,\mathbf{y}^0) = \mathbf{J}_\mathbf{y}\phi_{(l),j}(\mathbf{x}^0,\mathbf{y}^0)$ $\forall j \in \alpha\cup\gamma$. For each $k \in \{1,\dots,|\mathcal{E}|\}$, let $\Delta_{\mathcal{J}_k} \equiv \{\boldsymbol{\delta} \in \Delta_{\boldsymbol{\Theta}_{\min}}(\mathbf{x}^0,\mathbf{y}^0) : \phi_{(k),j} = \phi_{(\delta_j),j}\ \forall j \in \beta\}$. Note that $|\Delta_{\boldsymbol{\Theta}_{\min}}(\mathbf{x}^0,\mathbf{y}^0)| = |\mathcal{E}|^m$,

$$\bigcup_{k\in\{1,\dots,|\mathcal{E}|\}} \Delta_{\mathcal{J}_k} = \Delta_{\boldsymbol{\Theta}_{\min}}(\mathbf{x}^0,\mathbf{y}^0),$$

and, by symmetry, $|\Delta_{\mathcal{J}_k}| = |\Delta_{\mathcal{J}_l}|$ for any $l \in \{1,\dots,|\mathcal{E}|\}$;

$$\sum_{i=1}^{|\mathcal{E}|} |\Delta_{\mathcal{J}_i}| = |\mathcal{E}||\Delta_{\mathcal{J}_k}| = |\Delta_{\boldsymbol{\Theta}_{\min}}(\mathbf{x}^0,\mathbf{y}^0)| = |\mathcal{E}|^m$$

implies that $|\Delta_{\mathcal{J}_k}| = |\mathcal{E}|^{m-1}$.

Given $\boldsymbol{\delta} \in \Delta_{\boldsymbol{\Theta}_{\min}}(\mathbf{x}^0,\mathbf{y}^0)$, let

$$\mathbf{M}(\boldsymbol{\delta}) \equiv \begin{bmatrix} \mathbf{J}_\mathbf{y}\phi_{(\delta_1),1}(\mathbf{x}^0,\mathbf{y}^0)\\ \mathbf{J}_\mathbf{y}\phi_{(\delta_2),2}(\mathbf{x}^0,\mathbf{y}^0)\\ \vdots \\ \mathbf{J}_\mathbf{y}\phi_{(\delta_m),m}(\mathbf{x}^0,\mathbf{y}^0)\end{bmatrix}$$

so that $\Lambda_{\mathbf{y}}\boldsymbol{\Theta}_{\min}(\mathbf{x}^0,\mathbf{y}^0) = \{\mathbf{M}(\boldsymbol{\delta}) \in \mathbb{R}^{m\times m} : \boldsymbol{\delta} \in \Delta_{\boldsymbol{\Theta}_{\min}}(\mathbf{x}^0,\mathbf{y}^0)\}$. Then, for any $k^* \in \{1,\ldots,|\mathcal{E}|\}$,

$$\mathbf{M}(\boldsymbol{\delta}^*) = \mathbf{M}\left(\begin{bmatrix} k^* \\ k^* \\ \vdots \\ k^* \end{bmatrix}\right) = \begin{bmatrix} \mathbf{J}_{\mathbf{y}}\phi_{(k^*),1}(\mathbf{x}^0,\mathbf{y}^0) \\ \mathbf{J}_{\mathbf{y}}\phi_{(k^*),2}(\mathbf{x}^0,\mathbf{y}^0) \\ \vdots \\ \mathbf{J}_{\mathbf{y}}\phi_{(k^*),m}(\mathbf{x}^0,\mathbf{y}^0) \end{bmatrix} \quad \forall \boldsymbol{\delta}^* \in \Delta_{\mathcal{J}_{k^*}}.$$

Moreover,

$$\begin{bmatrix} \mathbf{J}_{\mathbf{y}}\phi_{(k^*),1}(\mathbf{x}^0,\mathbf{y}^0) \\ \mathbf{J}_{\mathbf{y}}\phi_{(k^*),2}(\mathbf{x}^0,\mathbf{y}^0) \\ \vdots \\ \mathbf{J}_{\mathbf{y}}\phi_{(k^*),m}(\mathbf{x}^0,\mathbf{y}^0) \end{bmatrix} = \begin{bmatrix} \mathbf{J}_{\mathbf{y}}v_1^{(\mathcal{J}_{k^*})}(\mathbf{x}^0,\mathbf{y}^0) \\ \mathbf{J}_{\mathbf{y}}v_2^{(\mathcal{J}_{k^*})}(\mathbf{x}^0,\mathbf{y}^0) \\ \vdots \\ \mathbf{J}_{\mathbf{y}}v_m^{(\mathcal{J}_{k^*})}(\mathbf{x}^0,\mathbf{y}^0) \end{bmatrix}$$

for some $\mathcal{J}_{k^*} \in \mathcal{P}(\beta)$. That is, $\Lambda_{\mathbf{y}}\boldsymbol{\Theta}_{\min}(\mathbf{x}^0,\mathbf{y}^0)$ contains at most $|\mathcal{E}|$ distinct matrices, each of which corresponds to $\mathbf{J}_{\mathbf{y}}\mathbf{v}^{(\mathcal{J})}$ for some $\mathcal{J} \subset \beta$; $\boldsymbol{\Theta}_{\min}(\mathbf{x}^0,\mathbf{y}^0)$ is completely coherently oriented with respect to $\mathbf{y}$ at $(\mathbf{x}^0,\mathbf{y}^0)$ if and only if all matrices in the set $\{\mathbf{J}_{\mathbf{y}}\mathbf{v}^{(\mathcal{J})} \in \mathbb{R}^{m\times m} : \mathcal{J} \subset \beta\}$ have the same nonvanishing determinant sign.

Suppose that $\mathcal{E}$ is not a set of essentially active selection functions and let, without loss of generality, $\mathcal{Y} \subset \mathcal{E}$ be a set of essentially active selection functions of the mapping $\boldsymbol{\Theta}_{\min}$ at $(\mathbf{x}^0,\mathbf{y}^0)$. By the above arguments, there are $\chi \leq |\mathcal{E}|$ distinct matrices in $\Lambda_{\mathbf{y}}\boldsymbol{\Theta}_{\min}(\mathbf{x}^0,\mathbf{y}^0)$, denoted $\mathbf{Y}_1,\ldots,\mathbf{Y}_\chi$, which necessarily satisfy $\{\mathbf{Y}_1,\ldots,\mathbf{Y}_\chi\} \subset \{\mathbf{J}_{\mathbf{y}}\mathbf{v}^{(\mathcal{J})} \in \mathbb{R}^{m\times m} : \mathcal{J} \subset \beta\}$. It immediately follows that if all matrices in the set $\{\mathbf{J}_{\mathbf{y}}\mathbf{v}^{(\mathcal{J})} \in \mathbb{R}^{m\times m} : \mathcal{J} \subset \beta\}$ have the same nonvanishing determinant sign, then $\boldsymbol{\Theta}_{\min}$ is completely coherently oriented with respect to $\mathbf{y}$ at $(\mathbf{x}^0,\mathbf{y}^0)$. ∎

Remark 4.5. Given mappings $\mathbf{G} : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}_+ \to \mathbb{R}^{n_1}$, $\mathbf{H} : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}_+ \to \mathbb{R}^{n_2}$, the MiCP (see [8]) associated with $(\mathbf{G},\mathbf{H})$ is to find a pair $(\mathbf{u},\mathbf{v}) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ such that

$$\begin{aligned} \mathbf{G}(\mathbf{u},\mathbf{v}) &= \mathbf{0}_{n_1}, \\ \mathbf{0}_{n_2} \leq \mathbf{v} \perp \mathbf{H}(\mathbf{u},\mathbf{v}) &\geq \mathbf{0}_{n_2}. \end{aligned}$$

Theorem 4.3 therefore yields sensitivities for parametric MiCPs with C^1 mappings $(\mathbf{G},\mathbf{H})$ by setting $\mathbf{y} \equiv (\mathbf{u},\mathbf{v})$ and $\mathbf{f}(\mathbf{x},\mathbf{y}) \equiv \mathbf{v}$ (i.e., $\mathbf{x}$ is the problem parameter here). Since MiCPs generalize NCPs, sensitivities for parametric NCPs with smooth participating function are also obtained.

5. Lexicographic derivatives of solutions of nonlinear programs. Let $D_p \subset \mathbb{R}^p$ and $D_x \subset \mathbb{R}^n$ be open. Let $f : D_p \times D_x \to \mathbb{R}$, $\mathbf{g} : D_p \times D_x \to \mathbb{R}^m$ and consider the following parametric NLP:

$$\begin{aligned} \phi(\mathbf{p}) \equiv \min_{\mathbf{x}\in D_x} \quad & f(\mathbf{p},\mathbf{x}), \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{p},\mathbf{x}) \leq \mathbf{0}_m, \end{aligned} \tag{14}$$

in which $\mathbf{p}$ is a problem parameter and ϕ the objective-value function. Suppose that f and $\mathbf{g}$ are differentiable at $(\mathbf{p}^0,\mathbf{x}^0) \in D_p \times D_x$ and that $(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0) \in D_p \times D_x \times \mathbb{R}^m$ is a KKT point of (14); i.e., it satisfies the following MiCP:

$$\begin{aligned} \nabla_{\mathbf{x}} L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0) &= \mathbf{0}_n, \\ \mathbf{0}_m \leq \boldsymbol{\mu}^0 \perp -\mathbf{g}(\mathbf{p}^0,\mathbf{x}^0) &\geq \mathbf{0}_m, \end{aligned} \tag{15}$$

where L is the usual Lagrangian function associated with (14). Equation (15) can be written as a nonsmooth equation system using any suitable NCP function [41, 8]; for example, $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$ satisfies

$$\boldsymbol{\Phi}_{\min}(\mathbf{p}, \mathbf{x}, \boldsymbol{\mu}) = \mathbf{0}_{n+m}, \tag{16}$$

where

$$\boldsymbol{\Phi}_{\min} : D_p \times D_x \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m : (\mathbf{p}, \mathbf{x}, \boldsymbol{\mu}) \mapsto \begin{bmatrix} \nabla_{\mathbf{x}} L(\mathbf{p}, \mathbf{x}, \boldsymbol{\mu}) \\ \mathbf{min}(-\mathbf{g}(\mathbf{p}, \mathbf{x}), \boldsymbol{\mu}) \end{bmatrix}.$$

The *feasible set* of (14) with respect to $\mathbf{p}^0 \in D_p \subset \mathbb{R}^p$ is denoted by

$$K(\mathbf{p}^0) \equiv \{\mathbf{x} \in D_x : \mathbf{g}(\mathbf{p}^0, \mathbf{x}) \leq \mathbf{0}_m\}.$$

Let $\mathcal{I} \equiv \{1, \ldots, m\}$ and define the *active set* of (14) at $(\mathbf{p}^0, \mathbf{x}^0) \in D_p \times K(\mathbf{p}^0)$ by

$$A(\mathbf{p}^0, \mathbf{x}^0) \equiv \{i \in \mathcal{I} : g_i(\mathbf{p}^0, \mathbf{x}^0) = 0\}.$$

The set of all multipliers satisfying the KKT conditions at $(\mathbf{p}^0, \mathbf{x}^0)$ is denoted by

$$M(\mathbf{p}^0, \mathbf{x}^0) \equiv \{\boldsymbol{\mu} \in \mathbb{R}^m : (\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}) \text{ is a KKT point of (14)}\}.$$

For $\boldsymbol{\mu}^0 \in M(\mathbf{p}^0, \mathbf{x}^0)$, the *strongly active*, *degenerate* (or *weakly active*), and *inactive* sets of (14) at $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$ are defined as

$$\begin{aligned} A^+(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) &\equiv \{i \in \mathcal{I} : g_i(\mathbf{p}^0, \mathbf{x}^0) = 0 < \mu_i^0\}, \\ A^0(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) &\equiv \{i \in \mathcal{I} : g_i(\mathbf{p}^0, \mathbf{x}^0) = 0 = \mu_i^0\}, \\ A^-(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) &\equiv \{i \in \mathcal{I} : g_i(\mathbf{p}^0, \mathbf{x}^0) < 0 = \mu_i^0\}. \end{aligned}$$

The mapping **min** is PC^1 on its domain in the sense of Scholtes [37]. If f and $\mathbf{g}$ are C^2 on their respective domains, then $\nabla_{\mathbf{x}} L$ is C^1 on its domain and $\boldsymbol{\Phi}_{\min}$ is PC^1 (and thus L-smooth) on its domain. Complete coherent orientation allows for application of Theorem 4.3 to characterize parametric sensitivites of (14).

THEOREM 5.1. *Let f and $\mathbf{g}$ be C^2 at $(\mathbf{p}^0, \mathbf{x}^0) \in D_p \times D_x$ and $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) \in D_p \times D_x \times \mathbb{R}^m$ be a KKT point of (14). If $\boldsymbol{\Phi}_{\min}$ is completely coherently oriented with respect to $(\mathbf{x}, \boldsymbol{\mu})$ at $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$, then there exist a neighborhood $N_{\mathbf{p}^0} \subset D_p$ of $\mathbf{p}^0$ and a PC^1 mapping $(\widetilde{\mathbf{x}}, \widetilde{\boldsymbol{\mu}}) : N_{\mathbf{p}^0} \to \mathbb{R}^n \times \mathbb{R}^m$ such that, for each $\mathbf{p} \in N_{\mathbf{p}^0}$, $(\mathbf{p}, \widetilde{\mathbf{x}}(\mathbf{p}), \widetilde{\boldsymbol{\mu}}(\mathbf{p}))$ is the unique solution of (16) in a neighborhood of $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$. Moreover, for any $k \in \mathbb{N}$ and any $\mathbf{P} \in \mathbb{R}^{p \times k}$, the LD-derivatives $\widetilde{\mathbf{x}}'(\mathbf{p}^0; \mathbf{P})$ and $\widetilde{\boldsymbol{\mu}}'(\mathbf{p}^0; \mathbf{P})$ are the unique solutions $\mathbf{X} \in \mathbb{R}^{n \times k}$ and $\mathbf{U} \in \mathbb{R}^{m \times k}$, respectively, of the following nonlinear equation system:*

$$\left[\begin{array}{c|c} \nabla^2_{\mathbf{xx}} L & (\mathbf{J_x g}_{A^+ \cup A^0})^{\mathrm{T}} \\ \hline -\mathbf{J_x g}_{A^+} & \mathbf{0}_{|A^+| \times (|A^+| + |A^0|)} \end{array}\right] \begin{bmatrix} \mathbf{X} \\ \mathbf{U}_{A^+ \cup A^0, \bullet} \end{bmatrix} = \begin{bmatrix} -\nabla^2_{\mathbf{xp}} L \\ \mathbf{J_p g}_{A^+} \end{bmatrix} \mathbf{P}, \tag{17a}$$

$$\mathbf{U}_{A^-, \bullet} = \mathbf{0}_{|A^-| \times k}, \tag{17b}$$

$$\mathbf{LMmin}\left(-\mathbf{J_p g}_{A^0} \mathbf{P} - \mathbf{J_x g}_{A^0} \mathbf{X}, \mathbf{U}_{A^0, \bullet}\right) = \mathbf{0}_{|A^0| \times k}, \tag{17c}$$

where the arguments of the Hessians associated with L and Jacobians associated with $\mathbf{g}$ have been omitted for brevity. Moreover, the objective-value function ϕ is PC^1 on $N_{\mathbf{p}^0}$, and its LD-derivative is given by

$$\phi'(\mathbf{p}^0; \mathbf{P}) = \mathbf{J_p} f(\mathbf{p}^0, \mathbf{x}^0)\mathbf{P} + \mathbf{J_x} f(\mathbf{p}^0, \mathbf{x}^0)\mathbf{X}. \tag{18}$$

Proof. Let

$$\mathbf{u} : D_p \times D_x \times \mathbb{R}^m \to \mathbb{R}^m : (\mathbf{p}, \mathbf{x}, \boldsymbol{\mu}) \mapsto -\mathbf{g}(\mathbf{p}, \mathbf{x}),$$
$$\mathbf{v} : D_p \times D_x \times \mathbb{R}^m \to \mathbb{R}^m : (\mathbf{p}, \mathbf{x}, \boldsymbol{\mu}) \mapsto \boldsymbol{\mu}.$$

In the setting of Theorem 4.3 (with $\mathbf{h}$, $\mathbf{f}$, and $\mathbf{g}$ replaced by $\nabla_{\mathbf{x}} L$, $\mathbf{u}$, and $\mathbf{v}$, respectively), $\alpha = A^+(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$, $\beta = A^0(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$, and $\gamma = A^-(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$. The neighborhood $N_{\mathbf{p}^0}$ and PC^1 mapping $(\widetilde{\mathbf{x}}, \widetilde{\boldsymbol{\mu}}) : N_{\mathbf{p}^0} \to \mathbb{R}^n$ satisfying $\boldsymbol{\Phi}_{\min}(\mathbf{p}, \widetilde{\mathbf{x}}(\mathbf{p}), \widetilde{\boldsymbol{\mu}}(\mathbf{p})) = \mathbf{0}_{n+m}$ for each $\mathbf{p} \in N_{\mathbf{p}^0}$ exist by virtue of Theorem 4.3. Equation (9) implies that, for any $k \in \mathbb{N}$ and any $\mathbf{P} \in \mathbb{R}^{p \times k}$, the LD-derivatives $\widetilde{\mathbf{x}}'(\mathbf{p}^0; \mathbf{M})$ and $\widetilde{\boldsymbol{\mu}}'(\mathbf{p}^0; \mathbf{M})$ are the unique solutions $\mathbf{X} \in \mathbb{R}^{n \times k}$ and $\mathbf{U} \in \mathbb{R}^{m \times k}$ of the equation system

$$\mathbf{0}_{n \times k} = \mathbf{J}[\nabla_{\mathbf{x}} L](\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)(\mathbf{P}, \mathbf{X}, \mathbf{U}), \tag{19a}$$
$$\mathbf{0}_{|A^+| \times k} = \mathbf{J}\mathbf{u}_{A^+}(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)(\mathbf{P}, \mathbf{X}, \mathbf{U}), \tag{19b}$$
$$\mathbf{0}_{|A^-| \times k} = \mathbf{J}\mathbf{v}_{A^-}(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)(\mathbf{P}, \mathbf{X}, \mathbf{U}), \tag{19c}$$
$$\mathbf{0}_{|A^0| \times k} = \mathbf{LMmin}(\mathbf{J}\mathbf{u}_{A^0}(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)(\mathbf{P}, \mathbf{X}, \mathbf{U}), \mathbf{J}\mathbf{v}_{A^0}(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)(\mathbf{P}, \mathbf{X}, \mathbf{U})). \tag{19d}$$

Note that

$$\mathbf{J}\mathbf{v}_{A^-}(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)(\mathbf{P}, \mathbf{X}, \mathbf{U}) = \begin{bmatrix} \mathbf{0}_{|A^-| \times p} & \mathbf{0}_{|A^-| \times n} & \mathbf{I}_{A^-, \bullet} \end{bmatrix} \begin{bmatrix} \mathbf{P} \\ \mathbf{X} \\ \mathbf{U} \end{bmatrix} = \mathbf{U}_{A^-, \bullet}.$$

Hence, (19c) is equivalent to (17b). Furthermore,

$$\begin{aligned} \mathbf{J}\mathbf{u}_{A^+}(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)(\mathbf{P}, \mathbf{X}, \mathbf{U}) &= \begin{bmatrix} -\mathbf{J}_{\mathbf{p}}\mathbf{g}_{A^+}(\mathbf{p}^0, \mathbf{x}^0) & -\mathbf{J}_{\mathbf{x}}\mathbf{g}_{A^+}(\mathbf{p}^0, \mathbf{x}^0) & \mathbf{0}_{|A^+| \times m} \end{bmatrix} \begin{bmatrix} \mathbf{P} \\ \mathbf{X} \\ \mathbf{U} \end{bmatrix} \\ &= -\mathbf{J}_{\mathbf{p}}\mathbf{g}_{A^+}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{P} - \mathbf{J}_{\mathbf{x}}\mathbf{g}_{A^+}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{X}, \end{aligned}$$

and

$$\begin{aligned} &\mathbf{J}[\nabla_{\mathbf{x}} L](\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)(\mathbf{P}, \mathbf{X}, \mathbf{U}) \\ &\quad = \nabla^2_{\mathbf{xp}} L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)\mathbf{P} + \nabla^2_{\mathbf{xx}} L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)\mathbf{X} + (\mathbf{J}_{\mathbf{x}}\mathbf{g}(\mathbf{p}^0, \mathbf{x}^0))^{\mathrm{T}}\mathbf{U} \\ &\quad = \nabla^2_{\mathbf{xp}} L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)\mathbf{P} + \nabla^2_{\mathbf{xx}} L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)\mathbf{X} \\ &\qquad + \begin{bmatrix} (\mathbf{J}_{\mathbf{x}}\mathbf{g}_{A^+}(\mathbf{p}^0, \mathbf{x}^0))^{\mathrm{T}} & (\mathbf{J}_{\mathbf{x}}\mathbf{g}_{A^0}(\mathbf{p}^0, \mathbf{x}^0))^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{U}_{A^+, \bullet} \\ \mathbf{U}_{A^0, \bullet} \end{bmatrix}, \end{aligned}$$

since $\mathbf{U}_{A^-, \bullet} = \mathbf{0}_{|A^-| \times k}$. Thus, (17a) is furnished by rearranging (19a) and (19b). Last,

$$\begin{aligned} &\mathbf{LMmin}(\mathbf{J}\mathbf{u}_{A^0}(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)(\mathbf{P}, \mathbf{X}, \mathbf{U}), \mathbf{J}\mathbf{v}_{A^0}(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)(\mathbf{P}, \mathbf{X}, \mathbf{U})) \\ &\quad = \mathbf{LMmin}(-\mathbf{J}_{\mathbf{p}}\mathbf{g}_{A^0}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{P} - \mathbf{J}_{\mathbf{x}}\mathbf{g}_{A^0}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{X}, \mathbf{U}_{A^0, \bullet}); \end{aligned}$$

(17c) is recovered from (19d).

Hence, ϕ satisfies $\phi(\mathbf{p}) = f(\mathbf{p}, \widetilde{\mathbf{x}}(\mathbf{p}))$ for $\mathbf{p} \in N_{\mathbf{p}^0}$ and is therefore PC^1 on $N_{\mathbf{p}^0}$ as the composition of a C^2 and PC^1 function. Defining $\bar{\mathbf{x}}(\mathbf{p}) \equiv (\mathbf{p}, \widetilde{\mathbf{x}}(\mathbf{p}))$ for $\mathbf{p} \in N_{\mathbf{p}^0}$, it follows that $\phi(\mathbf{p}) = [f \circ \bar{\mathbf{x}}](\mathbf{p})$, from which the LD-derivative chain rule (2) yields

$$\phi'(\mathbf{p}^0; \mathbf{P}) = f'(\bar{\mathbf{x}}(\mathbf{p}^0); \bar{\mathbf{x}}'(\mathbf{p}^0; \mathbf{P})) = \begin{bmatrix} \mathbf{J}_{\mathbf{p}} f(\mathbf{p}^0, \mathbf{x}^0) & \mathbf{J}_{\mathbf{x}} f(\mathbf{p}^0, \mathbf{x}^0) \end{bmatrix} \begin{bmatrix} \mathbf{P} \\ \mathbf{X} \end{bmatrix}. \qquad \square$$

Familiar nonlinear programming regularity conditions can be shown to guarantee complete coherent orientation of the nonsmooth mapping $\mathbf{\Phi}_{\min}$. First, a sufficient condition for complete coherent orientation is given in terms of a set of matrices having the same nonvanishing determinant sign (compare to Lemma 4.4).

LEMMA 5.2. *Given the setting of Theorem* 5.1*, if all matrices in the set*

$$\{\mathbf{H}(\mathcal{J}) \in \mathbb{R}^{(n+|A^+|+|\mathcal{J}|)\times(n+|A^+|+|\mathcal{J}|)} : \mathcal{J} \subset A^0(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)\}$$

have the same nonvanishing determinant sign, where

$$\mathbf{H}(\mathcal{J}) \equiv \begin{bmatrix} \nabla^2_{\mathbf{xx}} L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) & (\mathbf{J_x g}_{A^+\cup\mathcal{J}}(\mathbf{p}^0, \mathbf{x}^0))^{\mathrm{T}} \\ -\mathbf{J_x g}_{A^+\cup\mathcal{J}}(\mathbf{p}^0, \mathbf{x}^0) & \mathbf{0}_{(|A^+|+|\mathcal{J}|)\times(|A^+|+|\mathcal{J}|)} \end{bmatrix}, \tag{20}$$

then $\mathbf{\Phi}_{\min}$ is completely coherently oriented with respect to $(\mathbf{x}, \boldsymbol{\mu})$ at $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$.

Proof. Define the mappings

$$\begin{aligned} &\mathbf{u} : D_p \times D_x \times \mathbb{R}^m \to \mathbb{R}^m : (\mathbf{p}, \mathbf{x}, \boldsymbol{\mu}) \mapsto -\mathbf{g}(\mathbf{p}, \mathbf{x}), \\ &\mathbf{v} : D_p \times D_x \times \mathbb{R}^m \to \mathbb{R}^m : (\mathbf{p}, \mathbf{x}, \boldsymbol{\mu}) \mapsto \boldsymbol{\mu}. \end{aligned}$$

In the same vein as Lemma 4.4, if all the matrices in the set $\{\mathbf{M}(\mathcal{J}) \in \mathbb{R}^{(n+m)\times(n+m)} : \mathcal{J} \subset A^0(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)\}$ have the same nonvanishing determinant sign, where

$$\mathbf{M}(\mathcal{J}) \equiv \left[\begin{array}{c|c} \nabla^2_{\mathbf{xx}} L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) & (\mathbf{J_x g}(\mathbf{p}^0, \mathbf{x}^0))^{\mathrm{T}} \\ \hline -\mathrm{diag}(a_1, \ldots, a_m)\mathbf{J_x g}(\mathbf{p}^0, \mathbf{x}^0) & \mathbf{I}_m - \mathrm{diag}(a_1, \ldots, a_m) \end{array}\right],$$

$$a_i(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) \equiv \begin{cases} 1 & \text{if } i \in A^+(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0), \\ 1 & \text{if } i \in \mathcal{J}, \\ 0 & \text{if } i \in \mathcal{J}', \\ 0 & \text{if } i \in A^-(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0), \end{cases}$$

for each $i \in \{1, \ldots, m\}$, and $\mathcal{J}' \equiv A^0(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) \setminus \mathcal{J}$, then $\mathbf{\Phi}_{\min}$ is completely coherently oriented with respect to $(\mathbf{x}, \boldsymbol{\mu})$ at $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$.

Without loss of generality, let the submatrix $(\mathbf{J_x g}(\mathbf{p}^0, \mathbf{x}^0))^{\mathrm{T}}$ be equal to

$$\begin{bmatrix} (\mathbf{Jg}_{A^+\cup\mathcal{J}}(\mathbf{p}^0, \mathbf{x}^0))^{\mathrm{T}} & (\mathbf{Jg}_{A^-\cup\mathcal{J}'}(\mathbf{p}^0, \mathbf{x}^0))^{\mathrm{T}} \end{bmatrix}$$

under $q \in \mathbb{N} \cup \{0\}$ column permutations. By symmetry, the submatrix

$$-\mathrm{diag}(a_1, \ldots, a_m)\mathbf{J_x g}(\mathbf{p}^0, \mathbf{x}^0)$$

is equal to the matrix

$$\begin{bmatrix} -\mathbf{Jg}_{A^+\cup\mathcal{J}}(\mathbf{p}^0, \mathbf{x}^0) \\ \mathbf{0}_{(|A^-|+|\mathcal{J}'|)\times n} \end{bmatrix}$$

under q row permutations. Hence,

$$\det(\mathbf{M}(\mathcal{J})) = \det\left(\left[\begin{array}{c|c} \nabla^2_{\mathbf{xx}}L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0) & (\mathbf{J_x g}(\mathbf{p}^0,\mathbf{x}^0))^{\mathrm{T}} \\ \hline -\mathrm{diag}(a_1,\ldots,a_m)\mathbf{J_x g}(\mathbf{p}^0,\mathbf{x}^0) & \mathbf{I}_m - \mathrm{diag}(a_1,\ldots,a_m) \end{array}\right]\right)$$

$$= (-1)^{2q}\det\left(\begin{array}{cc|c} \nabla^2_{\mathbf{xx}}L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0) & (\mathbf{Jg}_{A^+\cup\mathcal{J}}(\mathbf{p}^0,\mathbf{x}^0))^{\mathrm{T}} & \\ -\mathbf{Jg}_{A^+\cup\mathcal{J}}(\mathbf{p}^0,\mathbf{x}^0) & \mathbf{0}_{(|A^+|+|\mathcal{J}|)\times(|A^+|+|\mathcal{J}|)} & \mathbf{G}(\mathcal{J}) \\ \hline \multicolumn{2}{c|}{\mathbf{0}_{(|\mathcal{J}'|+|A^-|)\times(n+|A^+|+|\mathcal{J}|)}} & \mathbf{I}_{|\mathcal{J}'|+|A^-|} \end{array}\right),$$

where

$$\mathbf{G}(\mathcal{J}) \equiv \begin{bmatrix} (\mathbf{Jg}_{\mathcal{J}'}(\mathbf{p}^0,\mathbf{x}^0))^{\mathrm{T}} & (\mathbf{Jg}_{A^-}(\mathbf{p}^0,\mathbf{x}^0))^{\mathrm{T}} \\ \mathbf{0}_{(|A^+|+|\mathcal{J}|)\times|\mathcal{J}'|} & \mathbf{0}_{(|A^+|+|\mathcal{J}|)\times|A^-|} \end{bmatrix}.$$

It therefore follows that, for any $\mathcal{J} \subset A^0(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0)$,

$$\mathrm{sign}(\det(\mathbf{M}(\mathcal{J}))) = \mathrm{sign}\left(\det\left(\left[\begin{array}{c|c} \mathbf{H}(\mathcal{J}) & \mathbf{G}(\mathcal{J}) \\ \hline \mathbf{0}_{(|\mathcal{J}'|+|A^-|)\times(n+|A^+|+|\mathcal{J}|)} & \mathbf{I}_{|\mathcal{J}'|+|A^-|} \end{array}\right]\right)\right)$$
$$= \mathrm{sign}(\det\mathbf{H}(\mathcal{J})). \qquad \square$$

Assuming that $\mathbf{f}$ and $\mathbf{g}$ are C^2 on their respective domains, the strong second-order sufficient condition is recalled.

Definition 5.3 (strong second-order sufficient condition). *The strong second-order sufficient condition (SSOSC) is said to hold at* $(\mathbf{p},\mathbf{x},\boldsymbol{\mu}) \in D_p \times D_x \times \mathbb{R}^m$ *if* $\mathbf{d}^{\mathrm{T}}\nabla^2_{\mathbf{xx}}L(\mathbf{p},\mathbf{x},\boldsymbol{\mu})\mathbf{d} > 0$ $\forall\mathbf{d} \in \mathbb{R}^n \setminus \{\mathbf{0}_n\}$ *satisfying* $(\nabla_{\mathbf{x}} g_i(\mathbf{p},\mathbf{x}))^{\mathrm{T}}\mathbf{d} = 0$ $\forall i \in A^+(\mathbf{p},\mathbf{x},\boldsymbol{\mu})$.

If the LICQ, SSOSC, and strict complementarity hold at a KKT point $(\mathbf{p},\mathbf{x},\boldsymbol{\mu})$ (i.e., $\mu_i - g_i(\mathbf{p},\mathbf{x}) > 0$ $\forall i \in \mathcal{I}$), the active index set is unchanged under continuity of $\mathbf{g}$ and sufficiently small parameter perturbations, allowing for an application of the classical implicit function theorem to yield the sensitivities for primal and dual variable solutions of (14) [10]. To remove the strict complementarity condition, a lemma is first needed which shows that complete coherent orientation holds under LICQ and SSOSC.

Lemma 5.4. *Let* $(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0) \in D_p \times D_x \times \mathbb{R}^m$ *be a KKT point of* (14) *satisfying SSOSC. Assume that LICQ holds at* $(\mathbf{p}^0,\mathbf{x}^0)$*. Then,* $\boldsymbol{\Phi}_{\min}$ *is completely coherently oriented with respect to* $(\mathbf{x},\boldsymbol{\mu})$ *at* $(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0)$.

Proof. By Remark 5.2.3 and the proof of [37, Proposition 5.2.1] (SSOSC implies the second-order sufficiency condition used by the author),

$$\mathrm{sign}(\det(\mathbf{H}(\mathcal{J}))) = \mathrm{sign}(\det(\mathbf{V}(\mathcal{J})^{\mathrm{T}}\nabla^2_{\mathbf{xx}}L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0)\mathbf{V}(\mathcal{J}))) \tag{21}$$

$\forall \mathcal{J} \subset A^0(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$, where $\mathbf{H}(\mathcal{J})$ is defined in (20) and the columns of the matrix $\mathbf{V}(\mathcal{J})$ form a basis of the nullspace of the matrix $\mathbf{J_x g}_{A^+ \cup \mathcal{J}}(\mathbf{p}^0, \mathbf{x}^0)$. Moreover, all the matrices in the set

$$\{\mathbf{V}(\mathcal{J})^{\mathrm{T}} \nabla^2_{\mathbf{xx}} L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) \mathbf{V}(\mathcal{J}) : \mathcal{J} \subset A^0(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)\}$$

have the same positive determinant since SSOSC holds at $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$. Equation (21) therefore yields that all the matrices in the set

$$\{\mathbf{H}(\mathcal{J}) \in \mathbb{R}^{(n+|A^+|+|\mathcal{J}|)\times(n+|A^+|+|\mathcal{J}|)} : \mathcal{J} \subset A^0(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)\}$$

have the same positive determinant since LICQ holds at $(\mathbf{p}^0, \mathbf{x}^0)$. The result follows from Lemma 5.2. ∎

An extension of Fiacco and McCormick's classical result is thus given (without strict complementarity), which furnishes an equation system whose unique solution describes L-derivatives of the primal and dual variables with respect to parametric perturbations.

THEOREM 5.5. *Let f and $\mathbf{g}$ be C^2 at $(\mathbf{p}^0, \mathbf{x}^0) \in D_p \times D_x$ and $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) \in D_p \times D_x \times \mathbb{R}^m$ be a KKT point of* (14) *satisfying SSOSC. Assume that LICQ holds at $(\mathbf{p}^0, \mathbf{x}^0)$. Then, there exist a neighborhood $N_{\mathbf{p}^0} \subset D_p$ of $\mathbf{p}^0$ and a PC^1 mapping $(\widetilde{\mathbf{x}}, \widetilde{\boldsymbol{\mu}}) : N(\mathbf{p}^0) \to \mathbb{R}^n \times \mathbb{R}^m$ such that, for each $\mathbf{p} \in N_{\mathbf{p}^0}$, $\widetilde{\mathbf{x}}(\mathbf{p})$ is an isolated strict local minimum of* (14) *and $(\mathbf{p}, \widetilde{\mathbf{x}}(\mathbf{p}), \widetilde{\boldsymbol{\mu}}(\mathbf{p}))$ is the unique KKT point in a neighborhood of $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$. Moreover, for any nonsingular $\mathbf{P}_0 \in \mathbb{R}^{p\times p}$, the L-derivatives $\mathbf{J}_{\mathrm{L}}\widetilde{\mathbf{x}}(\mathbf{p}^0; \mathbf{P}_0)$ and $\mathbf{J}_{\mathrm{L}}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0; \mathbf{P}_0)$ are the unique solutions $\mathbf{X}_{\mathrm{L}} \in \mathbb{R}^{n\times p}$ and $\mathbf{U}_{\mathrm{L}} \in \mathbb{R}^{m\times p}$, respectively, of the following linear equation system:*

$$\begin{bmatrix}\mathbf{X}\\ \mathbf{U}\end{bmatrix} = \begin{bmatrix}\mathbf{X}_{\mathrm{L}}\\ \mathbf{U}_{\mathrm{L}}\end{bmatrix}\mathbf{P}_0, \tag{22}$$

where $(\mathbf{X}, \mathbf{U})$ is furnished as the LD-derivative solution of (17) *with $k = p$ and $\mathbf{P} = \mathbf{P}_0$. The L-derivative $\mathbf{J}_{\mathrm{L}}\phi(\mathbf{p}^0; \mathbf{P}_0)$ of the objective-value function ϕ, which is PC^1 on $N_{\mathbf{p}^0}$, is the unique vector $\mathbf{z} \in \mathbb{R}^{1\times n_p}$ that solves the following linear equation system: $\mathbf{z}\mathbf{P}_0 = \mathbf{J_p} f(\mathbf{p}^0, \mathbf{x}^0)\mathbf{P}_0 + \mathbf{J_x} f(\mathbf{p}^0, \mathbf{x}^0)\mathbf{X}$.*

Proof. The existence of $N_{\mathbf{p}^0}$ and PC^1 mapping $(\widetilde{\mathbf{x}}, \widetilde{\boldsymbol{\mu}})$ such that $(\mathbf{p}, \widetilde{\mathbf{x}}(\mathbf{p}), \widetilde{\boldsymbol{\mu}}(\mathbf{p}))$ is the unique KKT point in a neighborhood of $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$ follows immediately from Theorem 5.1 and Lemma 5.4. That $\widetilde{\mathbf{x}}(\mathbf{p})$ is an isolated strict local minimum of (14) for each $\mathbf{p} \in N_{\mathbf{p}^0}$ follows from the following observations: by reducing the neighborhood $N_{\mathbf{p}^0}$ as necessary, $(\mathbf{p}, \widetilde{\mathbf{x}}(\mathbf{p}), \widetilde{\boldsymbol{\mu}}(\mathbf{p}))$ satisfies SSOSC and $(\mathbf{p}, \widetilde{\mathbf{x}}(\mathbf{p}))$ satisfies LICQ for each $\mathbf{p} \in N_{\mathbf{p}^0}$ (by the classical sensitivities result in [10]). Moreover, a KKT point of (14) satisfying SSOSC implies $\widetilde{\mathbf{x}}(\mathbf{p})$ is a strict local minimum of (14) for each $\mathbf{p} \in N_{\mathbf{p}^0}$ by [10, Lemma 3.2.1]. Its isolation follows from the uniqueness of $(\mathbf{p}, \widetilde{\mathbf{x}}(\mathbf{p}), \widetilde{\boldsymbol{\mu}}(\mathbf{p}))$ in a neighborhood of $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$.

The PC^1 mapping ϕ satisfies $\phi(\mathbf{p}) = f(\mathbf{p}, \widetilde{\mathbf{x}}(\mathbf{p}))$ for $\mathbf{p} \in N_{\mathbf{p}^0}$ by Theorem 5.1. Since $\mathbf{P}_0$ is nonsingular, the linear equation system $\mathbf{z}\mathbf{P}_0 = \phi'(\mathbf{p}^0; \mathbf{P}_0)$ admits a unique solution $\mathbf{z}^* \in \partial^{\mathrm{L}}\phi(\mathbf{p}^0) \subset \partial^{\mathrm{B}}\phi(\mathbf{p}^0)$. The result follows from (18). ∎

Remark 5.6. If, in addition, strict complementarity holds at $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$ in Theorem 5.5, then the classical result of Fiacco and McCormick is recovered; in this case, $A^0(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) = \emptyset$ and the nonsmooth equation system (17) reduces to the classical sensitivity linear equation system by choosing $\mathbf{P} = \mathbf{I}_p$. Moreover, $\mathbf{X} = \mathbf{X}_{\mathrm{L}} = \mathbf{J}_{\mathrm{L}}\widetilde{\mathbf{x}}(\mathbf{p}^0; \mathbf{I}_p) = \mathbf{J}\widetilde{\mathbf{x}}(\mathbf{p}^0)$, $\mathbf{U} = \mathbf{U}_{\mathrm{L}} = \mathbf{J}_{\mathrm{L}}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0; \mathbf{I}_p) = \mathbf{J}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0)$, and $\mathbf{z} = \mathbf{J}_{\mathrm{L}}\phi(\mathbf{p}^0; \mathbf{I}_p) = \mathbf{J}\phi(\mathbf{p}^0)$.

Remark 5.7. It is straightforward to extend the results to generalized derivatives of parametric NLPs with equality constraints. In fact, a more general sensitivities result can be given, which is applicable to different KKT nonsmooth equation system reformulations and the case where the participating functions are not necessarily C^2: let the first part of the setting of Theorem 5.1 hold with $\mathbf{f}$ and $\mathbf{g}$ instead only C^1 with (Fréchet) derivatives that are L-smooth at $(\mathbf{p}^0, \mathbf{x}^0) \in D_p \times D_x$ and let $\mathbf{\Phi}_{\mathrm{NCP}}(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) = \mathbf{0}_{n+m}$, where

$$\mathbf{\Phi}_{\mathrm{NCP}} : D_p \times D_x \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m : (\mathbf{p}, \mathbf{x}, \boldsymbol{\mu}) \mapsto \begin{bmatrix} \nabla_{\mathbf{x}} L(\mathbf{p}, \mathbf{x}, \boldsymbol{\mu}) \\ \boldsymbol{\psi}_{\mathrm{NCP}}(-\mathbf{g}(\mathbf{p}, \mathbf{x}), \boldsymbol{\mu}) \end{bmatrix},$$

and $\boldsymbol{\psi}_{\mathrm{NCP}} : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ is any suitable NCP function that is L-smooth at $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$. If the auxiliary mapping $(\mathbf{p}, \mathbf{x}, \boldsymbol{\mu}) \mapsto (\mathbf{p}, \mathbf{\Phi}_{\mathrm{NCP}}(\mathbf{p}, \mathbf{x}, \boldsymbol{\mu}))$ is a Lipschitz homeomorphism at $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$, the conclusions of Theorem 5.1 hold with (17) replaced by the equation system

$$\begin{aligned} \mathbf{0}_{n \times k} &= [\nabla_{\mathbf{x}} L]'(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0; (\mathbf{P}, \mathbf{X}, \mathbf{U})), \\ \mathbf{0}_{m \times k} &= \boldsymbol{\psi}_{\mathrm{NCP}}{}' \left(\begin{bmatrix} -\mathbf{g}(\mathbf{p}^0, \mathbf{x}^0) \\ \boldsymbol{\mu}^0 \end{bmatrix}; \begin{bmatrix} -\mathbf{J_p g}(\mathbf{p}^0, \mathbf{x}^0) & -\mathbf{J_x g}(\mathbf{p}^0, \mathbf{x}^0) & \mathbf{0}_{m \times m} \\ \mathbf{0}_{m \times p} & \mathbf{0}_{m \times n} & \mathbf{I}_m \end{bmatrix} \begin{bmatrix} \mathbf{P} \\ \mathbf{X} \\ \mathbf{U} \end{bmatrix} \right). \end{aligned}$$

The theory is illustrated with an example, inspired by the one in [45].

Example 5.8. Consider the following parametric NLP:

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^2} \quad & x_1^2 + x_2^2 + 2(p_1 x_1 + p_2 x_2) + x_2 \\ \text{s.t.} \quad & -x_1 + p_1 \leq 0, \\ & 2x_1^2 + x_2 - 10 \leq 0, \\ & -x_2 + 0.5 + p_2 \leq 0. \end{aligned} \tag{23}$$

Let $\mathbf{p}^0 = (0, 0)$. Then, $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$ is the unique KKT point of (23) where $\mathbf{x}^0 = (0, 0.5)$ and $\boldsymbol{\mu}^0 = (0, 0, 2)$. As a function of the parameter value, the isolated strict local minimum of (23) in a neighborhood of $\mathbf{p}^0$ is given by

$$\widetilde{\mathbf{x}} : N_{\mathbf{p}^0} \to \mathbb{R}^2 : \mathbf{p} \mapsto (|p_1|, |p_2 + 0.5|),$$

where $N_{\mathbf{p}^0} = (-1, 1)$. Moreover, for each $\mathbf{p} \in N_{\mathbf{p}^0}$, $(\mathbf{p}, \widetilde{\mathbf{x}}(\mathbf{p}), \widetilde{\boldsymbol{\mu}}(\mathbf{p}))$ is the unique KKT point of (23) in a neighborhood of $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$, where

$$\widetilde{\boldsymbol{\mu}} : N_{\mathbf{p}^0} \to \mathbb{R}^3 : \mathbf{p} \mapsto (\max(4p_1, 0), 0, \max(4p_2 + 2, 0)).$$

The PC^1 mappings $\widetilde{\mathbf{x}}$ and $\widetilde{\boldsymbol{\mu}}$ are the primal and dual variable solutions of (23) for different parameter values and correspond to the implicit functions outlined in Theorem 5.5. The objective-value function satisfies

$$\phi : N_{\mathbf{p}^0} \to \mathbb{R}^2 : \mathbf{p} \mapsto p_1^2 + (p_2 + 0.5)^2 + 2(p_1|p_1| + p_2|p_2 + 0.5|) + |p_2 + 0.5|. \tag{24}$$

Note that $A(\mathbf{p}^0, \mathbf{x}^0) = \{1, 3\}$ with strongly active set $A^+(\mathbf{p}^0, \mathbf{x}^0) = \{3\}$ and weakly active set $A^0(\mathbf{p}^0, \mathbf{x}^0) = \{1\}$. Though strict complementarity does not hold $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$, Theorem 5.5 is applicable as LICQ holds at $(\mathbf{p}^0, \mathbf{x}^0)$ and SSOSC holds at $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$. Equation (17) yields the following nonsmooth and nonlinear equation system:

$$\begin{aligned} &\begin{bmatrix} 2 & 0 & -1 & 0 \\ 0 & 2 & 0 & -1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ U_{11} & U_{12} \\ U_{31} & U_{32} \end{bmatrix} = \begin{bmatrix} -2 & 0 \\ 0 & -2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}, \\ &\begin{bmatrix} U_{21} & U_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 \end{bmatrix}, \\ &\mathbf{LMmin}\left(\begin{bmatrix} X_{11} - P_{11} & X_{12} - P_{12} \end{bmatrix}, \begin{bmatrix} U_{11} & U_{12} \end{bmatrix}\right) = \begin{bmatrix} 0 & 0 \end{bmatrix}. \end{aligned} \tag{25}$$

The solution of (25) yields LD-derivatives as functions of $\mathbf{P}$:

$$\widetilde{\mathbf{x}}'(\mathbf{p}^0; \mathbf{P}) \equiv \begin{bmatrix} \mathrm{fsign}(P_{11}, P_{12}) & 0 \\ 0 & 1 \end{bmatrix} \mathbf{P}$$

and

$$\widetilde{\boldsymbol{\mu}}'(\mathbf{p}^0; \mathbf{P}) \equiv \mathbf{LMmax}\left(\begin{bmatrix} 4P_{11} & 4P_{12} \\ 0 & 0 \\ 4P_{21} & 4P_{22} \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 4P_{21} & 4P_{22} \end{bmatrix} \right),$$

where **LMmax** is the *lexicographic-matrix-maximum*, defined similarly as **LMmin**.

For $\mathbf{P}$ nonsingular,

$$\mathbf{J}_{\mathrm{L}}\widetilde{\mathbf{x}}(\mathbf{p}^0, \mathbf{P}) = \begin{bmatrix} \mathrm{fsign}(P_{11}, P_{12}) & 0 \\ 0 & 1 \end{bmatrix}$$

and

$$\mathbf{J}_{\mathrm{L}}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0, \mathbf{P}) = \begin{cases} \begin{bmatrix} 4 & 0 \\ 0 & 0 \\ 0 & 4 \end{bmatrix} & \text{if } \mathbf{LMmax}([4P_{11} \quad 4P_{12}], [0 \quad 0]) = [4P_{11} \quad 4P_{12}], \\ \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 4 \end{bmatrix} & \text{if } \mathbf{LMmax}([4P_{11} \quad 4P_{12}], [0 \quad 0]) = [0 \quad 0] \end{cases}$$

are elements of the L-subdifferentials of the primal and dual solutions, and

$$\mathbf{J}_{\mathrm{L}}\phi(\mathbf{p}^0, \mathbf{P}) = \begin{bmatrix} 0 & 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} \mathbf{I}_p \\ \mathbf{J}_{\mathrm{L}}\widetilde{\mathbf{x}}(\mathbf{p}^0, \mathbf{P}) \end{bmatrix}.$$

Calculate the B-subdifferentials of $\widetilde{\mathbf{x}}$ and $\widetilde{\boldsymbol{\mu}}$ at $\mathbf{p}^0$ from the closed-form solutions:

$$\partial^{\mathrm{B}}\widetilde{\mathbf{x}}(\mathbf{p}^0) = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \right\}, \quad \partial^{\mathrm{B}}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0) = \left\{ \begin{bmatrix} 4 & 0 \\ 0 & 0 \\ 0 & 4 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 4 \end{bmatrix} \right\}.$$

Noting that ϕ in (24) is C^1 at $\mathbf{p} = \mathbf{p}^0$, with $\mathbf{J}\phi(\mathbf{p}^0) = [0 \quad 3]$, observe that

$$\begin{aligned} &\mathbf{J}_{\mathrm{L}}\widetilde{\mathbf{x}}(\mathbf{p}^0; \mathbf{P}) \in \partial^{\mathrm{L}}\widetilde{\mathbf{x}}(\mathbf{p}^0) \subset \partial^{\mathrm{B}}\widetilde{\mathbf{x}}(\mathbf{p}^0), \\ &\mathbf{J}_{\mathrm{L}}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0; \mathbf{P}) \in \partial^{\mathrm{L}}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0) \subset \partial^{\mathrm{B}}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0), \\ &\mathbf{J}_{\mathrm{L}}\phi(\mathbf{p}^0; \mathbf{P}) \in \partial^{\mathrm{L}}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0) = \partial^{\mathrm{B}}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0) = \{\mathbf{J}\phi(\mathbf{p}^0)\} = \{[0 \quad 3]\}. \end{aligned}$$

Finally, complete coherent orientation is confirmed to hold (i.e., Lemma 5.4); the set $\Lambda_{\mathbf{x},\boldsymbol{\mu}}\boldsymbol{\Phi}_{\min}(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0)$ contains two distinct elements, which have determinant one and two, implying that complete coherent orientation holds at $(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0)$. The set of matrices outlined in Lemma 5.2 is equal to

$$\{\mathbf{H}(\mathcal{J}) : \mathcal{J} \in \{\{1\},\emptyset\}\} = \left\{ \begin{bmatrix} 2 & 0 & -1 & 0 \\ 0 & 2 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 0 \end{bmatrix} \right\};$$

both matrices have the same nonvanishing determinant sign (the determinants are equal to one and two, respectively), in agreement with Lemma 5.2.

Practically implementable methods to solve the nonsmooth and nonlinear sensitivity system (17) (and thereby obtain L-derivatives of primal and dual variable solutions via the linear equation system (22)) are outlined as follows. In the spirit of the discussion at the end of section 3, Proposition 2 in [22] can be applied to compute L-derivatives (and therefore B-subdifferential elements) of $\boldsymbol{\Phi}_{\min}$ by performing up to $2^{|A^0(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0)|}$ linear equation solves, instead of solving the nonsmooth and nonlinear equation system (17). In this approach, $|(A^0(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0)-1)p|$ comparisons of elements in the worst case are needed in the verification stage to terminate the for loop.

Alternatively, Lemma 3.7 provides a way to solve (17) directly using a nonsmooth equation-solving algorithm; assuming the setting of Theorem 5.5, the nonsmooth sensitivity system (17) can be solved columnwise (from left to right) using, for example, the nonsmooth Newton method (8) with $\mathbf{v}_{(l)} \equiv (\mathbf{x}^{(l)},\boldsymbol{\mu}^{(l)})$ to furnish the jth columns $\mathbf{x}_{(j)}$ and $\boldsymbol{\mu}_{(j)}$ of $\mathbf{X}$ and $\mathbf{U}$, respectively, and $\mathbf{h}'(\mathbf{v}_{(l)};\mathbf{I}_m)$ replaced by

$$(26)\qquad \boldsymbol{\Phi}'_{\min}\left(\begin{bmatrix}\mathbf{p}^0\\ \mathbf{x}^0\\ \boldsymbol{\mu}^0\end{bmatrix};\begin{bmatrix}\mathbf{P}_{(j-1)} & \mathbf{p}_{(j)} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times m}\\ \mathbf{X}_{(j-1)} & \mathbf{x}^{(l)} & \mathbf{I}_n & \mathbf{0}_{n\times m}\\ \mathbf{U}_{(j-1)} & \boldsymbol{\mu}^{(l)} & \mathbf{0}_{m\times n} & \mathbf{I}_m\end{bmatrix}\right)\begin{bmatrix}\mathbf{0}_{p\times n} & \mathbf{0}_{p\times m}\\ \mathbf{I}_n & \mathbf{0}_{n\times m}\\ \mathbf{0}_{m\times n} & \mathbf{I}_m\end{bmatrix}.$$

Motivated by this, a hybrid algorithm is proposed as Algorithm 1, with a user-chosen critical number β_{crit} dictating which of these two methods is used: either cycling through linear equation system solves, or nonsmooth equation-solving.

- $\beta_{\text{crit}} = m$ corresponds to application of Proposition 2 in [22] (i.e., solving the linear equation system (27) until a solution is verified in line 7).
- $\beta_{\text{crit}} = -1$ corresponds to application of nonsmooth equation-solving methods (i.e., solving (28) using (26)), regardless of the weakly active set size.
- $\beta_{\text{crit}} = 0$ corresponds to solving a linear equation system (i.e., Fiacco and McCormick's classical result), given an absence of weakly active indices and performing nonsmooth equation solves otherwise.

The verification test in line 7 for the cycling approach may mistakenly fail for a correct solution, because of numerical error (e.g., due to the coefficient matrix on the left-hand side of (27) having a high condition number). Such false negatives have been observed in practice [43] when applying the cycling approach (i.e., Proposition 2 in [22]) to solve PC^1 equation systems. The method of iterative refinement [44] has successfully alleviated this issue in the aforementioned work and can be optionally added to the cycling part of Algorithm 1.

The for-loop beginning on line 14 in Algorithm 1 updates the set of weakly active indices by definition of **LMmin**, in the spirit of [18, section 4]; if $-\mathbf{J}_{\mathbf{p}}g_l(\mathbf{p}^0,\mathbf{x}^0)\mathbf{p}_{(j)}-$

$\mathbf{J_x} g_l(\mathbf{p}^0, \mathbf{x}^0)\mathbf{a} < b_l$ for some $l \in A^0$ and $j = j^*$, then the corresponding vector (i.e., $-\mathbf{J_p} g_l(\mathbf{p}^0, \mathbf{x}^0)\mathbf{P} - \mathbf{J_x} g_l(\mathbf{p}^0, \mathbf{x}^0)\mathbf{A}$ instead of $\mathbf{B}_l$) is the lexicographically ordered minimum. The remaining $p - j^*$ comparisons (in the case of verifying the solution of (27)) or nonsmooth equation method solves (in the case of solving (28)) are not needed.

Algorithm 1. Evaluate L-derivatives of primal and dual variable solutions.

Require: $\beta_{\text{crit}} \in \{-1, 0, 1, \ldots, m\}$, nonsingular $\mathbf{P} \in \mathbb{R}^{p \times p}$

1: **procedure** CALCULATE $(\mathbf{J}_{\mathrm{L}}\widetilde{\mathbf{x}}(\mathbf{p}^0; \mathbf{P}), \mathbf{J}_{\mathrm{L}}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0; \mathbf{P}))$
2: Set $A^0 \leftarrow A^0(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$, $A^+ \leftarrow A^+(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$, $A^- \leftarrow A^-(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$.
3: **for** $j = 1, \ldots, p$ **do**
4: **if** $|A^0| \leq \beta_{\text{crit}}$ **then**
5: **for all** $\mathcal{J} \subset A^0$ **do**
6: Solve the following linear equation system for $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{(n+m)\times 1}$:

$$\begin{bmatrix} \nabla^2_{\mathbf{xx}} L & (\mathbf{J_x g}_{A^+ \cup \mathcal{J}})^{\mathrm{T}} \\ \hline -\mathbf{J_x g}_{A^+} & \mathbf{0}_{(|A^+|)\times(|A^+|+|\mathcal{J}|)} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \\ \mathbf{b}_{A^+ \cup \mathcal{J}} \end{bmatrix} = \begin{bmatrix} -\nabla^2_{\mathbf{xp}} L \\ \mathbf{J_p g}_{A^+} \end{bmatrix} \mathbf{p}_{(j)}, \qquad (27)$$
$$\mathbf{b}_{A^- \cup (A^0 \setminus \mathcal{J})} = \mathbf{0}_{|A^-|+|A^0 \setminus \mathcal{J}|}.$$

7: **if** $\mathbf{0}_{|A^0|} = \mathbf{min}\left(-\mathbf{J_p g}_{A^0}\mathbf{p}_{(j)} - \mathbf{J_x g}_{A^0}\mathbf{a}, \mathbf{b}_{A^0}\right)$ **then**
8: **go to** 21
9: **end if**
10: **end for**
11: **else**
12: Solve the following nonsmooth equation system for $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}^{(n+m)\times 1}$:

$$\begin{bmatrix} \nabla^2_{\mathbf{xx}} L & (\mathbf{J_x g}_{A^+ \cup A^0})^{\mathrm{T}} \\ \hline -\mathbf{J_x g}_{A^+} & \mathbf{0}_{(|A^+|)\times(|A^+|+|A^0|)} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \\ \mathbf{b}_{A^+ \cup A^0} \end{bmatrix} = \begin{bmatrix} -\nabla^2_{\mathbf{xp}} L \\ \mathbf{J_p g}_{A^+} \end{bmatrix} \mathbf{p}_{(j)}, \qquad (28)$$
$$\mathbf{b}_{A^-} = \mathbf{0}_{|A^-|},$$
$$\mathbf{min}\left(-\mathbf{J_p g}_{A^0}\mathbf{p}_{(j)} - \mathbf{J_x g}_{A^0}\mathbf{a}, \mathbf{b}_{A^0}\right) = \mathbf{0}_{|A^0|}.$$

13: **end if**
14: **for all** $l \in A^0$ **do**
15: **if** $-\mathbf{J_p} g_l(\mathbf{p}^0, \mathbf{x}^0)\mathbf{p}_{(j)} - \mathbf{J_x} g_l(\mathbf{p}^0, \mathbf{x}^0)\mathbf{a} < b_l$ **then**
16: Set $A^0 \leftarrow A^0 \setminus \{l\}$, $A^+ \leftarrow A^+ \cup \{l\}$
17: **else if** $-\mathbf{J_p} g_l(\mathbf{p}^0, \mathbf{x}^0)\mathbf{p}_{(j)} - \mathbf{J_x} g_l(\mathbf{p}^0, \mathbf{x}^0)\mathbf{a} > b_l$ **then**
18: Set $A^0 \leftarrow A^0 \setminus \{l\}$, $A^- \leftarrow A^- \cup \{l\}$
19: **end if**
20: **end for**
21: Set $\begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \leftarrow \begin{bmatrix} \mathbf{A} & \mathbf{a} \\ \mathbf{B} & \mathbf{b} \end{bmatrix}$
22: **end for**
23: Solve the equation system $(\mathbf{A}, \mathbf{B}) = (\mathbf{X}_{\mathrm{L}}, \mathbf{U}_{\mathrm{L}})\mathbf{P}$ for $(\mathbf{X}_{\mathrm{L}}, \mathbf{U}_{\mathrm{L}}) \in \mathbb{R}^{(n+m)\times p}$.
24: **return** $(\mathbf{X}_{\mathrm{L}}, \mathbf{U}_{\mathrm{L}})$.
25: **end procedure**

Example 5.9. Consider applying Algorithm 1 to the parametric NLP (23). Let $\mathbf{P} = \mathbf{I}_2$ and $\beta_{\text{crit}} = 1$. Then, $A^0 = \{1\}$, $A^- = \{2\}$, $A^+ = \{3\}$, and solving two equation systems (at worst) is guaranteed to furnish the sensitivities since $2^{|A^0|} = 2$. For $j = 1$, the linear equation system (27) associated with $\mathcal{J} = \emptyset$ is given by

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 0 \end{bmatrix}, \quad b_1 = 0, \quad b_2 = 0,$$

which is solved to yield $\mathbf{a} = (-1, 0)$ and $\mathbf{b} = (0, 0, 0)$. The LD-derivative test in line 7 fails since $\min(-2, 0) \neq 0$; the inner for loop does not terminate. The linear equation system (27) associated with $\mathcal{J} = \{1\}$ has solution $\mathbf{a} = (1, 0)$ and $\mathbf{b} = (4, 0, 0)$. In this case, $\min(0, 4) = 0$ and the LD-derivative test passes; the following assignments are made: $\mathbf{A} \leftarrow (1, 0)$, $\mathbf{B} \leftarrow (4, 0, 0)$, $A^0 \leftarrow \emptyset$, and $A^+ \leftarrow \{1, 3\}$. For $j = 2$, the linear equation system (27) associated with $\mathcal{J} = \emptyset$ is given by

$$\begin{bmatrix} 2 & 0 & 0 & -1 \\ 0 & 2 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_3 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \quad b_2 = 0, \tag{29}$$

since the weakly active set is empty (i.e., a solve of the classical sensitivity system). The solution is given by $\mathbf{a} = (0, 1)$ and $\mathbf{b} = (0, 0, 4)$,

$$\mathbf{A} \leftarrow \mathbf{I}_2, \quad \mathbf{B} \leftarrow \begin{bmatrix} 4 & 0 \\ 0 & 0 \\ 0 & 4 \end{bmatrix}, \tag{30}$$

and $(\mathbf{X}_\text{L}, \mathbf{U}_\text{L}) = (\mathbf{A}, \mathbf{B})$ is a correct L-derivative of primal and dual solutions.

Repeating the problem with $\mathbf{P} = \mathbf{I}_2$ and $\beta_{\text{crit}} = 0$ (i.e., an aversion to cycling through linear equation solves in the presence of weakly active sets) yields the following: for $j = 1$, the nonsmooth and nonlinear equation system (28) is given by

$$\begin{bmatrix} 2 & 0 & 0 & -1 \\ 0 & 2 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_3 \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad b_2 = 0,$$

$$\min\left(\begin{bmatrix} -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} -1 & 0 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, b_1 \right) = 0,$$

which admits unique solution $\mathbf{a} = (1, 0)$ and $\mathbf{b} = (4, 0, 0)$. The following assignments are made: $\mathbf{A} \leftarrow (1, 0)$, $\mathbf{B} \leftarrow (4, 0, 0)$, $A^0 \leftarrow \emptyset$, and $A^+ \leftarrow \{1, 3\}$. For $j = 2$, (29) is solved to yield $\mathbf{a} = (0, 1)$ and $\mathbf{b} = (0, 0, 4)$ and the assignments in (30) are made.

Application of the new method to a class of parametric QPs is straightforward.

Example 5.10. Consider the following parametric QP:

$$\begin{aligned} \phi(\mathbf{p}) \equiv \min_{\mathbf{x}\in D_x} \quad & 0.5\mathbf{x}^{\mathrm{T}}\mathbf{H}\mathbf{x} + \mathbf{x}^{\mathrm{T}}\mathbf{G}\mathbf{p} \\ \text{s.t.} \quad & \mathbf{C}\mathbf{x} \leq \mathbf{F}\mathbf{p}, \end{aligned} \tag{31}$$

where $\mathbf{C} \in \mathbb{R}^{m\times n}$, $\mathbf{F} \in \mathbb{R}^{m\times p}$, and $\mathbf{H} \in \mathbb{R}^{n\times n}$ are matrices with real-valued elements. Given a reference parameter value $\mathbf{p}^0 \in D_p \subset \mathbb{R}^p$, let $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) \in D_p \times D_x \times \mathbb{R}^m$ be a KKT point of (31); $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$ satisfies the system of equations

$$\begin{aligned} \mathbf{H}\mathbf{x} + \mathbf{G}\mathbf{p} + \boldsymbol{\mu}^{\mathrm{T}}\mathbf{C} &= \mathbf{0}_n, \\ \mathbf{min}(\mathbf{F}\mathbf{p} - \mathbf{C}\mathbf{x}, \boldsymbol{\mu}) &= \mathbf{0}_m. \end{aligned}$$

Given the index set $\mathcal{I} \equiv \{1, \ldots, m\}$, the active, strongly active, weakly active, and inactive index sets of (31) at $(\mathbf{p}^0, \mathbf{x}^0)$ are equal to, respectively,

$$\begin{aligned} A(\mathbf{p}^0, \mathbf{x}^0) &\equiv \{i \in \mathcal{I} : \mathbf{C}_i\mathbf{x}^0 = \mathbf{F}_i\mathbf{p}^0\}, \\ A^+(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) &\equiv \{i \in \mathcal{I} : \mathbf{C}_i\mathbf{x}^0 - \mathbf{F}_i\mathbf{p}^0 = 0 < \mu_i^0\}, \\ A^0(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) &\equiv \{i \in \mathcal{I} : \mathbf{C}_i\mathbf{x}^0 - \mathbf{F}_i\mathbf{p}^0 = 0 = \mu_i^0\}, \\ A^-(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0) &\equiv \{i \in \mathcal{I} : \mathbf{C}_i\mathbf{x}^0 - \mathbf{F}_i\mathbf{p}^0 < 0 = \mu_i^0\}. \end{aligned}$$

Let $\mathbf{H}$ be positive definite, which implies SSOSC holds at $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$. Let the active constraint matrix $\mathbf{C}_{A,\bullet}$ be full row rank, which implies LICQ holds at $(\mathbf{p}^0, \mathbf{x}^0)$. Then, according to Theorem 5.5, the nonsmooth equation system

$$\begin{aligned} \left[\begin{array}{c|c} \mathbf{H} & (\mathbf{C}_{A^+\cup A^0})^{\mathrm{T}} \\ \hline -\mathbf{C}_{A^+} & \mathbf{0}_{|A^+|\times(|A^+|+|A^0|)} \end{array}\right] \begin{bmatrix} \mathbf{X} \\ \mathbf{U}_{A^+\cup A^0,\bullet} \end{bmatrix} &= \begin{bmatrix} -\mathbf{G} \\ -\mathbf{F}_{A^+} \end{bmatrix}, \\ \mathbf{U}_{A^-,\bullet} &= \mathbf{0}_{|A^-|\times p}, \\ \mathbf{LMmin}\left(-\mathbf{F}_{A^0} - \mathbf{C}_{A^0}\mathbf{X}, \mathbf{U}_{A^0,\bullet}\right) &= \mathbf{0}_{|A^0|\times p}, \end{aligned} \tag{32}$$

admits the unique solution $(\mathbf{X}_{\mathrm{L}}, \mathbf{U}_{\mathrm{L}}) \in \mathbb{R}^{(n+m)\times p}$ which are B-subdifferential elements of primal and dual variable solutions of (31) at $\mathbf{p} = \mathbf{p}^0$:

$$\mathbf{X}_{\mathrm{L}} \in \partial^{\mathrm{L}}\widetilde{\mathbf{x}}(\mathbf{p}^0) \subset \partial^{\mathrm{B}}\widetilde{\mathbf{x}}(\mathbf{p}^0)$$

and

$$\mathbf{U}_{\mathrm{L}} \in \partial^{\mathrm{L}}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0) \subset \partial^{\mathrm{B}}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0).$$

(In this case, $\mathbf{P} = \mathbf{I}_p$ is chosen.) Moreover, a B-subdifferential element of the objective-value function ϕ at $\mathbf{p} = \mathbf{p}^0$ is calculated as $\mathbf{x}^{\mathrm{T}}\mathbf{H}\mathbf{X}_{\mathrm{L}} \in \partial^{\mathrm{L}}\phi(\mathbf{p}^0) \subset \partial^{\mathrm{B}}\phi(\mathbf{p}^0)$.

6. Conclusions. Parametric sensitivities for NLPs exhibiting active set changes have been obtained. The primal and dual variables sensitivities are characterized by L-derivatives, which are computationally relevant (as elements of the B-subdifferential, M-subdifferential, and Clarke Jacobian) and calculated as the unique solution of an auxiliary nonsmooth and nonlinear equation system. The classical sensitivity theory of Fiacco and McCormick is recovered in the absence of active index set changes (i.e., strict complementarity). The regularity conditions on the NLPs are shown to be implied by LICQ and SSOSC and there is no competing theory for furnishing computationally relevant generalized derivatives (i.e., B-subdifferential elements) of solutions of NLPs (and MiCPs). Often in practice a well-conditioned nonsingular (or even orthonormal) matrix is chosen for the directions matrix when calculating an LD-derivative. However, since LD-derivatives are well-defined for singular (or even nonsquare) directions matrices and satisfy a sharp chain rule, automatable computation of generalized derivative information in applications involving nonsmooth optimization problems with nonsmooth dynamical systems embedded (or vice versa) is possible. Though the minimum-function reformulation of the KKT system is used in this work, the theory developed here is without loss of generality in this regard as it can be extended to other NCP function reformulations.

As mentioned earlier, applying the theoretical tools used here to other types of mathematical programs (e.g., VIs and MPECs) is of interest going forward. The results in this article assume the computational costs associated with computing a KKT point and verifying its regularity are already incurred, as in the classical theory, since the focus here is the subsequent sensitivity analysis step. However, analyzing the computational complexity of computing sensitivities of solutions of NLPs using the theory in this article warrants investigation; choosing an appropriate value of β_{crit} in Algorithm 1 requires consideration of the computational complexity of the cycling approach (which scales exponentially with the number of weakly active sets in the worst case) and the nonsmooth equation-solving approach (which is presently unclear). Using the sensitivity results here in model predictive control and dynamic optimization problems is another direction for future work.

Acknowledgment. The authors would like to acknowledge the anonymous reviewers for their helpful comments.

REFERENCES

[1] P. I. Barton, K. A. Khan, P. Stechlinski, and H. A. J. Watson, *Computationally relevant generalized derivatives: Theory, evaluation and applications*, Optim. Methods Softw., to appear, doi:10.1080/10556788.2017.1374385.
[2] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.
[3] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics in Appl. Math., SIAM, Philadelphia, 1990.
[4] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, Grad. Texts in Math. 178, Springer, New York, 2008.
[5] J. Cortes, *Discontinuous dynamical systems*, IEEE Control Syst., 28 (2012), pp. 36–73.
[6] S. Dempe, *Directional differentiability of optimal solutions under Slater's condition*, Math. Program., 59 (1993), pp. 49–69.
[7] F. Facchinei, A. Fischer, and M. Herrich, *An LP-Newton method: Nonsmooth equations, KKT systems, and nonisolated solutions*, Math. Program., 146 (2014), pp. 1–36.
[8] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems: Volume* I, Springer, New York, 2003.
[9] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems: Volume* II, Springer, New York, 2003.
[10] A. V. Fiacco, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Academic Press, New York, 1983.
[11] A. V. Fiacco and G. P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York, 1968.
[12] T. Gal, *Postoptimal Analyses, Parametric Programming and Related Topics*, 2nd ed., Walter de Gruyter, Berlin, 1995.
[13] J. A. Gomez, K. Höffner, K. A. Khan, and P. I. Barton, *Generalized derivatives of lexicographic linear programs*, submitted.
[14] A. Griewank, *On stable piecewise linearization and generalized algorithmic differentiation*, Optim. Methods Softw., 28 (2013), pp. 1139–1178.
[15] A. Griewank and A. Walther, *Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation*, 2nd ed., SIAM, Philadelphia, 2008.

[16] K. Höffner, K. A. Khan, and P. I. Barton, *Generalized derivatives of dynamic systems with a linear program embedded*, Automatica, 63 (2016), pp. 198–208.

[17] R. Janin, *Directional derivative of the marginal function in nonlinear programming*, in Sensitivity, Stability and Parametric Analysis, Springer, New York, 1984, pp. 110–126.

[18] K. A. Khan, *Branch-locking AD techniques for nonsmooth composite functions and nonsmooth implicit functions*, Optim. Methods Softw., to appear, doi:10.1080/10556788.2017.1341506.

[19] K. A. Khan and P. I. Barton, *Generalized derivatives for solutions of parametric ordinary differential equations with non-differentiable right-hand sides*, J. Optim. Theory Appl., 163 (2014), pp. 355–386.

[20] K. A. Khan and P. I. Barton, *Generalized gradient elements for nonsmooth optimal control problems*, in 53rd IEEE Conference on Decision and Control, 2014, pp. 1887–1892.

[21] K. A. Khan and P. I. Barton, *A vector forward mode of automatic differentiation for generalized derivative evaluation*, Optim. Methods Softw., 30 (2015), pp. 1185–1212.

[22] K. A. Khan and P. I. Barton, *Generalized derivatives for hybrid systems*, IEEE Trans. Automat. Control, 62 (2017), pp. 3193–3208.

[23] D. Klatte and B. Kummer, *Nonsmooth Equations in Optimization: Regularity, Calculus, Methods and Applications*, Kluwer, Dordecht, the Netherlands, 2002.

[24] M. Kojima, *Strongly stable stationary solutions in nonlinear programs*, in Analysis and Computation of Fixed Points, S. M. Robinson, ed., Academic Press, New York, 1980, pp. 93–138.

[25] J. Kyparisis, *Sensitivity analysis for nonlinear programs and variational inequalities with nonunique multipliers*, Math. Oper. Res., 15 (1990), pp. 286–298.

[26] C. Lemaréchal, J. J. Strodiot, and A. Bihain, *On a bundle algorithm for nonsmooth optimization*, in Nonlinear Programming 4, O. L. Mangasarian, R. R. Meyer, and S. M. Robinson, eds., Academic Press, New York, 1981.

[27] A. B. Levy and B. S. Mordukhovich, *Coderivatives in parametric optimization*, Math. Program., 99 (2004), pp. 311–327.

[28] B. S. Mordukhovich, *Generalized differential calculus for nonsmooth and set-valued mappings*, J. Math. Anal. Appl., 183 (1994), pp. 250–288.

[29] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation* I*: Basic Theory*, Springer, Berlin, 2006.

[30] Y. Nesterov, *Lexicographic differentiation of nonsmooth functions*, Math. Program., 104 (2005), pp. 669–700.

[31] J.-S. Pang and D. Ralph, *Piecewise smoothness, local invertibility, and parametric analysis of normal maps*, Math. Oper. Res., 21 (1996), pp. 401–426.

[32] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Program., 58 (1993), pp. 353–367.

[33] D. Ralph and S. Dempe, *Directional derivatives of the solution of a parametric nonlinear program*, Math. Program., 70 (1995), pp. 159–172.

[34] D. Ralph and S. Scholtes, *Sensitivity analysis of composite piecewise smooth equations*, Math. Program., 76 (1997), pp. 593–612.

[35] S. M. Robinson, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), pp. 43–62.

[36] V. Roshchina, *Two applications of lexicographic differentiation*, presented at CIAO Showcase and Workshop, 2017, http://www.roshchina.com/wp-content/uploads/2017/04/lexicographic.pdf.

[37] S. Scholtes, *Introduction to Piecewise Differentiable Equations*, Springer, New York, 2012.

[38] A. Shapiro, *Sensitivity analysis of nonlinear programs and differentiability properties of metric projections*, SIAM J. Control Optim., 26 (1988), pp. 628–645.

[39] P. G. Stechlinski and P. I. Barton, *Generalized derivatives of optimal control problems with nonsmooth differential-algebraic equations embedded*, in Proceedings of the 55th IEEE Conference on Decision and Control, 2016, pp. 592–597.

[40] P. G. Stechlinski and P. I. Barton, *Dependence of solutions of nonsmooth differential–algebraic equations on parameters*, J. Differential Equations, 262 (2017), pp. 2254–2285.

[41] D. Sun and L. Qi, *On NCP-Functions*, Comput. Optim. Appl., 13 (1999), pp. 201–220.

[42] T. H. Sweetser, *A minimal set-valued strong derivative for vector-valued Lipschitz functions*, J. Optim. Theory Appl., 23 (1977), pp. 549–562.

[43] H. A. J. Watson, M. Vikse, T. Gundersen, and P. I. Barton, *Reliable flash calculations: Part* 2*. Process flowsheeting with nonsmooth models and generalized derivatives*, Ind. Eng. Chem. Res., 56 (2017), pp. 14848–14864.

[44] J. H. Wilkinson, *Rounding Errors in Algebraic Processes*, Prentice-Hall, Englewood Cliffs, NJ, 1963.

[45] I. J. Wolf and W. Marquardt, *Fast NMPC schemes for regulatory and economic NMPC—A review*, J. Process Control, 44 (2016), pp. 162–183.