Generalized Sensitivity Analysis of Nonlinear Programs using a Sequence of Quadratic Programs

Peter Stechlinski[a], Johannes Jäschke[b], Paul I. Barton[c,*]

[a]*Department of Mathematics and Statistics, University of Maine, Orono, ME 04469, United States*
[b]*Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), N-7491 Trondheim, Norway*
[c]*Process Systems Engineering Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, United States*

Abstract

Local sensitivity information is obtained for KKT points of parametric NLPs that may exhibit active set changes under parametric perturbations; under appropriate regularity conditions, computationally relevant generalized derivatives of primal and dual variable solutions of parametric NLPs are calculated. Ralph and Dempe obtained directional derivatives of solutions of parametric NLPs exhibiting active set changes from the unique solution of an auxiliary quadratic program. This article uses lexicographic directional derivatives, a newly developed tool in nonsmooth analysis, to generalize the classical NLP sensitivity analysis theory of Ralph and Dempe. By viewing said auxiliary quadratic program as a parametric NLP, the results of Ralph and Dempe are applied to furnish a sequence of coupled QPs, whose unique solutions yield generalized derivative information for the NLP. A practically implementable algorithm is provided. The theory developed here is motivated by widespread applications of nonlinear programming sensitivity analysis, such as in dynamic control and optimization problems.

Keywords: Sensitivity analysis, Nonsmooth analysis, Generalized derivatives, B-subdifferential, Parametric optimization, NLP KKT systems.
2010 MSC: 49J52, 90C31 90C33

1. Introduction

Consider the following parametric NLP:

$$\begin{aligned}
\min_{\mathbf{x}\in D_x} \quad & f(\mathbf{p},\mathbf{x}),\\
\text{s.t.} \quad & g_i(\mathbf{p},\mathbf{x}) \le 0, \quad \forall i \in \mathcal{G} := \{1,\dots,n_g\},\\
& h_i(\mathbf{p},\mathbf{x}) = 0, \quad \forall i \in \mathcal{H} := \{1,\dots,n_h\},
\end{aligned} \tag{1}$$

where $\mathbf{p} \in D_p$ is a problem parameter; $f : D_p \times D_x \to \mathbb{R}$, $\mathbf{g} : D_p \times D_x \to \mathbb{R}^{n_g}$ and $\mathbf{h} : D_p \times D_x \to \mathbb{R}^{n_h}$ are C^2 on their respective domains; and the sets $D_p \subset \mathbb{R}^{n_p}$ and $D_x \subset \mathbb{R}^{n_x}$ are open. Given a reference parameter value $\mathbf{p}^0 \in D_p$, the focus of this article is obtaining (generalized) derivative information of the primal and dual variable solutions of (1), under parametric perturbations which may cause active set changes.

Since Fiacco and McCormick [1] established classical sensitivity analysis of parametric NLPs under regularity assumptions including the linear independence constraint qualification (LICQ) and strict complementarity (i.e. an absence of active set changes), a number of authors [2–5] have investigated sensitivity analysis for parametric NLPs with active set changes; a broad and comprehensive sensitivity analysis theory for mathematical programs is found in [6, 7]. This article focuses on generalizing the practical method by Ralph and Dempe [8] for calculating directional derivatives of primal variable solution mappings using a quadratic program with an auxiliary linear program embedded.

*Corresponding author. Tel.: +1 617 2536526.
Email addresses: `peter.stechlinski@maine.edu` (Peter Stechlinski), `johannes.jaschke@ntnu.no` (Johannes Jäschke), `pib@mit.edu` (Paul I. Barton)

The theory of Ralph and Dempe [8] assumes the Magnasarian-Fromovitz constraint qualification (MFCQ), constant rank constraint qualification (CRCQ) and general strong second-order sufficient condition (GSSOSC). Said theory was specialized by Scholtes [9] to the LICQ and strong second-order sufficient condition (SSOSC) setting, using Kojima's nonsmooth reformulation of the parametric NLP KKT system: Given a reference parameter value $\mathbf{p}^0 \in D_p$, let $(\mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0) \in D_x \times \mathbb{R}_+^{n_g} \times \mathbb{R}^{n_h}$ be a KKT point of (1). LICQ holds at $(\mathbf{p}^0, \mathbf{x}^0)$ if the set of vectors

$$\{(\mathbf{J_x} g_i(\mathbf{p}^0, \mathbf{x}^0))^{\mathrm{T}} : i \in \mathcal{A}\} \cup \{(\mathbf{J_x} h_i(\mathbf{p}^0, \mathbf{x}^0))^{\mathrm{T}} : i \in \mathcal{H}\}$$

are linearly independent, where $\mathbf{J_x} g_i(\mathbf{p}^0, \mathbf{x}^0) \in \mathbb{R}^{1 \times n_x}$ denotes the partial Jacobian matrix of g_i with respect to $\mathbf{x}$ evaluated at $(\mathbf{p}^0, \mathbf{x}^0)$ (with the other partial Jacobians defined similarly), and where the *active set* of (1) at $(\mathbf{p}^0, \mathbf{x}^0)$ is denoted by

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

The *strong second-order sufficient condition (SSOSC)* holds at $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0)$ if $\mathbf{v}^{\mathrm{T}} \nabla^2_{\mathbf{xx}} L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0)\mathbf{v} > 0$ for all $\mathbf{v} \in \mathbb{R}^{n_x} \setminus \{\mathbf{0}_{n_x}\}$ satisfying

$$\begin{aligned}
\mathbf{J_x} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{v} &= 0, \quad \forall i \in \mathcal{A}^+, \\
\mathbf{J_x} h_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{v} &= 0, \quad \forall i \in \mathcal{H}.
\end{aligned}$$

where L is the usual *Lagrangian* function associated with (1), $\nabla^2_{\mathbf{xx}} L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0)$ is the Hessian matrix of L with respect to $\mathbf{x}$ evaluated at $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0)$ and the *strongly active*, *weakly active*, and *inactive* sets of $\mathbf{g}$ in (1) at $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$ are defined as, respectively,

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

Given $\mathbf{d} \in \mathbb{R}^{n_p}$, let $QP_{(1)}(\mathbf{d})$ denote the following quadratic program:

$$\begin{aligned}
QP_{(1)}(\mathbf{d}): \quad \min_{\mathbf{z} \in \mathbb{R}^{n_x}} \quad & 0.5\mathbf{z}^{\mathrm{T}} \nabla^2_{\mathbf{xx}} L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0)\mathbf{z} + \mathbf{z}^{\mathrm{T}} \nabla^2_{\mathbf{xp}} L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0)\mathbf{d}, \\
\text{s.t.} \quad & \mathbf{J_x} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{z} + \mathbf{J_p} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{d} \leq 0, \quad \forall i \in \mathcal{A}^0, \qquad \leftarrow \text{multipliers } \boldsymbol{\eta} \in \mathbb{R}_+^{|\mathcal{A}^0|} \\
& \mathbf{J_x} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{z} + \mathbf{J_p} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{d} = 0, \quad \forall i \in \mathcal{A}^+, \qquad \leftarrow \text{multipliers } \boldsymbol{\gamma} \in \mathbb{R}^{|\mathcal{A}^+|} \\
& \mathbf{J_x} h_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{z} + \mathbf{J_p} h_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{d} = 0, \quad \forall i \in \mathcal{H}, \qquad \leftarrow \text{multipliers } \boldsymbol{\rho} \in \mathbb{R}^{n_h}
\end{aligned} \tag{2}$$

whose feasible set is the critical cone of (1) at $(\mathbf{p}^0, \mathbf{x}^0)$ with respect to $(\boldsymbol{\mu}^0, \boldsymbol{\lambda}^0)$ in the direction $\mathbf{d}$. To improve readability, the notation chosen for the multipliers associated with the quadratic program's constraints is outlined above. Directional differentiability of (1) is obtained via $QP_{(1)}(\mathbf{d})$, under the regularity assumptions outlined above, in the following adaptation of Theorem 5.2.1 and Proposition 5.2.1 in [9].

Theorem 1.1. Let $(\mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0) \in D_x \times \mathbb{R}_+^{n_g} \times \mathbb{R}^{n_h}$ be a KKT point of (1) satisfying SSOSC and let LICQ hold at $(\mathbf{p}^0, \mathbf{x}^0)$. Then there exist a neighborhood $N_{\mathbf{p}^0} \subset D_p$ of $\mathbf{p}^0$ and PC^1 mappings $(\widetilde{\mathbf{x}}, \widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\lambda}}) : N_{\mathbf{p}^0} \to D_x \times \mathbb{R}_+^{n_g} \times \mathbb{R}^{n_h}$ such that, for each $\mathbf{p} \in N_{\mathbf{p}^0}$, $\widetilde{\mathbf{x}}(\mathbf{p})$ is an isolated strict local minimizer of (1) and $(\widetilde{\mathbf{x}}(\mathbf{p}), \widetilde{\boldsymbol{\mu}}(\mathbf{p}), \widetilde{\boldsymbol{\lambda}}(\mathbf{p}))$ is an isolated KKT point of (1) in a neighborhood of $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0)$. Moreover, for any $\mathbf{d} \in \mathbb{R}^{n_p}$, the directional derivatives of $(\widetilde{\mathbf{x}}, \widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\lambda}})$ at $\mathbf{p}^0$ in the direction $\mathbf{d}$ satisfy

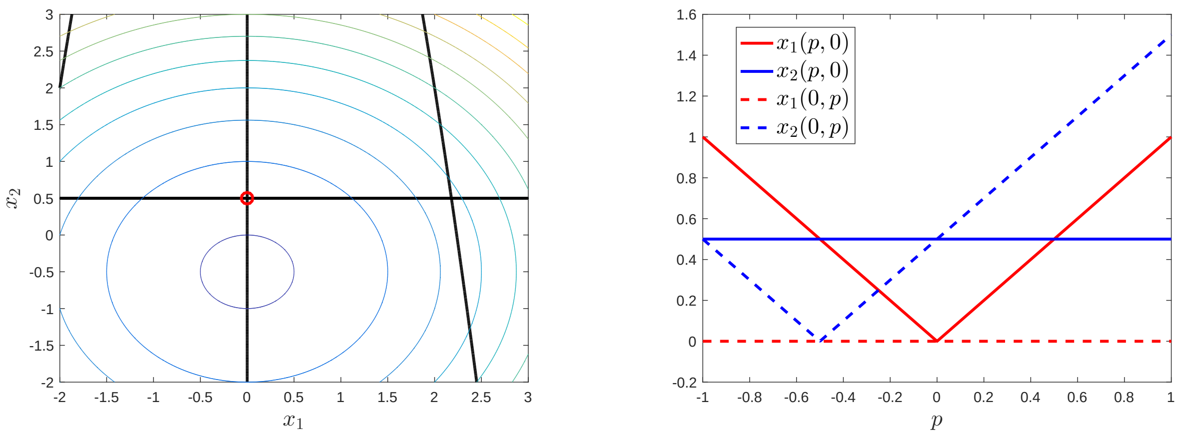
$$\begin{aligned}
\widetilde{\mathbf{x}}'(\mathbf{p}^0; \mathbf{d}) &= \mathbf{z}_{(1)}(\mathbf{d}), \\
\widetilde{\boldsymbol{\lambda}}'(\mathbf{p}^0; \mathbf{d}) &= \boldsymbol{\rho}_{(1)}(\mathbf{d}), \\
\widetilde{\boldsymbol{\mu}}_i'(\mathbf{p}^0; \mathbf{d}) &= \begin{cases} \eta_{(1),i}(\mathbf{d}), & \text{if } i \in \mathcal{A}^0, \\ \gamma_{(1),i}(\mathbf{d}), & \text{if } i \in \mathcal{A}^+, \\ 0, & \text{if } i \in \mathcal{A}^-, \end{cases}
\end{aligned} \tag{3}$$

where $\mathbf{z}_{(1)}(\mathbf{d})$ and $(\boldsymbol{\eta}_{(1)}(\mathbf{d}), \boldsymbol{\gamma}_{(1)}(\mathbf{d}), \boldsymbol{\rho}_{(1)}(\mathbf{d}))$ are the unique primal and dual solutions of $QP_{(1)}(\mathbf{d})$, respectively, evaluated at $\mathbf{d}$.

Example 1.2. Consider the following parametric NLP, inspired by the example studied in [10]:

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathbb{R}^2} \quad & x_1^2 + x_2^2 + 2(p_1x_1 + p_2x_2) + x_2, \\
\text{s.t.} \quad & -x_1 + p_1 \le 0, \\
& 2x_1^2 + x_2 - 10 \le 0, \\
& -x_2 + 0.5 + p_2 \le 0.
\end{aligned} \tag{4}
$$

With reference parameter value $\mathbf{p}^0 = (0,0)$, $(\mathbf{x}^0, \boldsymbol{\mu}^0)$ is a KKT point of (4) where $\mathbf{x}^0 = (0, 0.5)$ and $\boldsymbol{\mu}^0 = (0,0,2)$; see Figure 1 for an illustration.



(a) Constraints (black lines), computed solution (red dot), and objective function contours.

(b) Primal variable solutions.

Figure 1: Illustration of (4).

The quadratic program (2) associated with (4) at $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$ for any $\mathbf{d} \in \mathbb{R}^2$ is given by

$$
\begin{aligned}
QP_{(1)}(\mathbf{d}): \quad \min_{\mathbf{z}\in\mathbb{R}^2} \quad & z_1^2 + z_2^2 + 2(d_1z_1 + d_2z_2), \\
\text{s.t.} \quad & -z_1 + d_1 \le 0, \\
& -z_2 + d_2 = 0,
\end{aligned} \tag{5}
$$

since $\mathcal{A}^+ = \{3\}$, $\mathcal{A}^0 = \{1\}$ and $\mathcal{A}^- = \{2\}$. As a function of $\mathbf{d} \in \mathbb{R}^2$, primal and dual variable solutions of (5) are given by

$$\mathbf{z}_{(1)}(\mathbf{d}) \equiv (|d_1|, d_2), \quad \eta_{(1)}(\mathbf{d}) \equiv \max(4d_1, 0), \quad \gamma_{(1)}(\mathbf{d}) \equiv 4d_2.$$

As a function of $\mathbf{p} \in N_{\mathbf{p}^0} = (-1,1)$, the isolated strict local minimum of (4) is given by

$$\widetilde{\mathbf{x}}(\mathbf{p}) \equiv (|p_1|, |p_2 + 0.5|).$$

Moreover, for each $\mathbf{p} \in N_{\mathbf{p}^0}$, $(\widetilde{\mathbf{x}}(\mathbf{p}), \widetilde{\boldsymbol{\mu}}(\mathbf{p}))$ is an isolated KKT point of (4) in a neighborhood of $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0)$, where

$$\widetilde{\boldsymbol{\mu}}(\mathbf{p}) \equiv (\max(4p_1, 0), 0, \max(4p_2 + 2, 0)).$$

As expected, $\widetilde{\mathbf{x}}'(\mathbf{p}^0; \mathbf{d}) = \mathbf{z}_{(1)}(\mathbf{d})$ and $\widetilde{\boldsymbol{\mu}}'(\mathbf{p}^0; \mathbf{d}) = (\eta_{(1)}(\mathbf{d}), 0, \gamma_{(1)}(\mathbf{d}))$.

Theorem 1.1 provides a method to calculate directional derivatives and allows for active index set changes (i.e., strict complementarity is not required), but does not yield B-subdifferential elements. This article approaches this parametric NLP sensitivity analysis problem by applying the theory of Scholtes [9] in combination with the recently developed theory of lexicographical directional (LD-)derivatives [11]. Based on lexicographic differentiation [12], LD-derivatives extend the classical directional derivative and always satisfy sharp calculus rules, unlike Clarke's generalized derivative [13]. Tractable numerical methods [11] have been developed to compute LD-derivatives in an automatable way. LD-derivatives can be used to furnish lexicographic (L-)derivatives [12], by solving a linear equation system, which are elements of the Bouligand (B-)subdifferential (and therefore Clarke's generalized derivative), as well the Mordukhovich (M-)subdifferential [14], when the participating functions are piecewise differentiable (PC^1) in the sense of Scholtes [9]. Consequently, L-derivatives are computationally relevant generalized derivative elements that can be supplied to dedicated nonsmooth algorithms (e.g. nonsmooth Newton methods [15–18] and optimization methods [19–23]).

This new approach yields a sequence of quadratic programs whose unique solutions are used to furnish LD-derivatives of primal and dual variable solution mappings. Since said solution mappings are piecewise differentiable on a neighborhood of a reference parameter, the L-derivatives obtained from this procedure are computationally relevant generalized derivative elements. A recent work [24] provides the first computationally relevant theory to obtain generalized derivatives of parametric NLPs with active set changes, and is therefore the only competing theory to the contributions made here. In [24], L-derivatives are obtained by applying a nonsmooth implicit function theorem to a nonsmooth reformulation of the NLP KKT system, generalizing Fiacco and McCormick's [1] approach to allow for active set changes. Since the computational costs of the aforementioned approach in [24] are currently unclear, the methods detailed in this article may prove to be superior for evaluating generalized derivative elements. Moreover, since the present approach only relies on the ability to furnish directional derivatives, it is hopeful that it can be applied to other types of mathematical programs, such as complementarity problems, variational inequalities, mathematical programs with equilibrium constraints, etc.

Establishing sensitivity analysis for NLPs and other types of mathematical programs has widespread application. For example, a wide variety of process operation problems require dynamic optimization, often posed as open loop optimal control problems and solved via sequential methods (e.g., multiple shooting) or simultaneous methods (e.g., collocation on finite elements). Such an approach necessitates accurate and efficient computation of NLP solutions and sensitivity information, which motivates the contributions made in this article. Consequently, we are hopeful the theory in this article leads to improvements in optimal control methods (e.g., nonlinear model predictive control [25–27]) and solving dynamic control applications (e.g., temperature control in batch reactors, optimal catalyst mixing problems, bioreactor control, distillation column problems see , including temperature control in batch reactors, catalyst mixing, stirred tank reactors, bioreactor control, and distillation column problems [28, 29]).

2. Background: Generalized Derivatives

Generalized derivatives theory is reviewed before presenting lexicographic differentiation and the lexicographic directional derivative. (For a broader view of nonsmooth analysis, the reader is referred to [7, 13, 30, 31].) Given a locally Lipschitz continuous function $\mathbf{f} : Z \subset \mathbb{R}^n \to \mathbb{R}^m$, Z open, it follows by Rademacher's Theorem that $\mathbf{f}$ is differentiable on $Z \setminus S_{\mathbf{f}}$, where $S_{\mathbf{f}} \subset Z$ is a zero (Lebesgue) measure subset. The *B-subdifferential* of $\mathbf{f}$ at $\mathbf{z}^0 \in Z$ is

$$\partial_{\mathrm{B}}\mathbf{f}(\mathbf{z}^0) := \left\{ \lim_{j\to\infty} \mathbf{Jf}(\mathbf{z}_{(j)}) : \mathbf{z}_{(j)} \to \mathbf{z}^0,\ \mathbf{z}_{(j)} \in Z \setminus S_{\mathbf{f}} \right\},$$

which is nonempty and compact. If $\mathbf{f}$ is C^1 at $\mathbf{z}$, then $\partial_{\mathrm{B}}\mathbf{f}(\mathbf{z}) = \{\mathbf{Jf}(\mathbf{z})\}$. Assume that $\mathbf{f}$ is PC^1 at $\mathbf{z}$ for the remainder of this section. Then

$$\partial_{\mathrm{B}}\mathbf{f}(\mathbf{z}^0) = \left\{ \mathbf{Jf}_{(i)}(\mathbf{z}^0) : i \in \{1, \ldots, n_{\mathrm{ess}}\} \right\},$$

by [32], where $\{\mathbf{f}_{(1)}, \ldots, \mathbf{f}_{(n_{\mathrm{ess}})}\}$ is a set of $n_{\mathrm{ess}} \in \mathbb{N}$ essentially active C^1 selection functions of $\mathbf{f}$ at $\mathbf{z}^0$.

Dedicated numerical nonsmooth algorithms nominally require an element of the Clarke (generalized) *Jacobian* of $\mathbf{f}$ at $\mathbf{z}^0$ [13], which is the convex hull of the B-subdifferential; $\partial\mathbf{f}(\mathbf{z}^0) := \mathrm{conv}\,\partial_{\mathrm{B}}\mathbf{f}(\mathbf{z}^0)$. It is difficult in general to evaluate said elements in an automatable way for the following reasons, among others [13, 31]:

1. Clarke's Jacobian satisfies calculus rules with inclusions; given $\mathbf{g} : \mathbb{R}^m \to \mathbb{R}^q$ that is Lipschitz continuous on a neighborhood of $\mathbf{f}(\mathbf{z}^0)$, then $\mathbf{G} \in \partial\mathbf{g}(\mathbf{f}(\mathbf{z}^0))$ and $\mathbf{F} \in \partial\mathbf{f}(\mathbf{z}^0)$ may satisfy $\mathbf{GF} \notin \partial[\mathbf{g} \circ \mathbf{f}](\mathbf{z}^0)$. For

example, consider the functions $f(x) \equiv \min(0,x)$ and $g(x) \equiv \max(0,x)$. Then $1 \in \partial f(0) = [0,1]$ and $1 \in \partial g(f(0)) = [0,1]$, but, noting that $[g \circ f](x) \equiv 0$, $1 \notin \partial[g \circ f](0) = \{0\}$.

2. Taking directional derivatives in the unit coordinate directions $\mathbf{e}_{(i)}$ does not necessarily yield an element of the B-subdifferential; supposing $\mathbf{f}$ is PC^1 at $\mathbf{z}^0$ and $i_1, \ldots, i_n \in \{1, \ldots, n_{\mathrm{ess}}\}$,

$$\begin{aligned}\mathbf{F} &= \begin{bmatrix}\mathbf{f}'(\mathbf{z}^0;\mathbf{e}_{(1)}) & \mathbf{f}'(\mathbf{z}^0;\mathbf{e}_{(2)}) & \ldots \mathbf{f}'(\mathbf{z}^0;\mathbf{e}_{(n)})\end{bmatrix} \\ &= \begin{bmatrix}\mathbf{J}\mathbf{f}_{(i_1)}(\mathbf{z}^0)\mathbf{e}_{(1)} & \mathbf{J}\mathbf{f}_{(i_2)}(\mathbf{z}^0)\mathbf{e}_{(2)} & \ldots & \mathbf{J}\mathbf{f}_{(i_n)}(\mathbf{z}^0)\mathbf{e}_{(n)}\end{bmatrix},\end{aligned}$$

may satisfy $\mathbf{F} \notin \partial_{\mathrm{B}}\mathbf{f}(\mathbf{z}^0)$. For example, consider the function $f(\mathbf{x}) \equiv |x_1 - x_2|$. Observe that

$$\partial_{\mathrm{B}} f(\mathbf{0}_2) = \{[1 \quad -1], [-1 \quad 1]\}, \quad \partial f(\mathbf{0}_2) = \{[1-\lambda \quad -1+\lambda] : 0 \leq \lambda \leq 2\},$$

but

$$\begin{bmatrix}f'(\mathbf{0}_2;(1,0)) & f'(\mathbf{0}_2;(0,1))\end{bmatrix} = [1 \quad 1] \notin \partial_{\mathrm{B}} f(\mathbf{0}_2).$$

3. The Cartesian product of componentwise Clarke gradients may be a strict superset of the Clarke Jacobian; $\mathbf{F} = \prod_{i=1}^m \partial f_i(\mathbf{z}^0)$ may satisfy $\mathbf{F} \notin \partial\mathbf{f}(\mathbf{z}^0)$. For example, the function $\mathbf{f}(\mathbf{x}) \equiv (x_1 + |x_2|, x_1 - |x_2|)$ satisfies

$$\partial\mathbf{f}(\mathbf{0}_2) = \left\{\begin{bmatrix}1 & 2\lambda - 1 \\ 1 & 1 - 2\lambda\end{bmatrix} : 0 \leq \lambda \leq 1\right\} \subset \partial f_1(\mathbf{0}_2) \times \partial f_2(\mathbf{0}_2) = \left\{\begin{bmatrix}1 & 2\lambda_1 - 1 \\ 1 & 2\lambda_2 - 1\end{bmatrix} : (\lambda_1, \lambda_2) \in [0,1]^2\right\},$$

where the inclusion is strict.

The lexicographic derivative [12] is an element of the B-subdifferential in the PC^1 setting [11], and can be computed more easily thanks to its strict calculus rules; for any $k \in \mathbb{N}$ and full row rank $\mathbf{M} = [\mathbf{m}_{(1)} \ \cdots \ \mathbf{m}_{(k)}] \in \mathbb{R}^{n \times k}$, the *lexicographic (L-)derivative* of $\mathbf{f}$ at $\mathbf{z}^0$ in the directions $\mathbf{M}$ is given as

$$\mathbf{J}_{\mathrm{L}}\mathbf{f}(\mathbf{z}^0;\mathbf{M}) := \mathbf{J}\mathbf{f}^{(k)}_{\mathbf{z}^0,\mathbf{M}}(\mathbf{0}_n) \in \mathbb{R}^{m \times n},$$

where the directional derivative mappings are

$$\begin{aligned}&\mathbf{f}^{(0)}_{\mathbf{z}^0,\mathbf{M}} : \mathbb{R}^n \to \mathbb{R}^m : \mathbf{d} \mapsto \mathbf{f}'(\mathbf{z}^0;\mathbf{d}), \\ &\mathbf{f}^{(j)}_{\mathbf{z}^0,\mathbf{M}} : \mathbb{R}^n \to \mathbb{R}^m : \mathbf{d} \mapsto [\mathbf{f}^{(j-1)}_{\mathbf{z}^0,\mathbf{M}}]'(\mathbf{m}_{(j)};\mathbf{d}), \quad \forall j \in \{1,\ldots,k\}.\end{aligned} \tag{6}$$

($\mathbf{M}$ being full row rank guarantees linearity of the mapping $\mathbf{f}^{(k)}_{\mathbf{z}^0,\mathbf{M}}$.) The *lexicographic (L-)subdifferential* of $\mathbf{f}$ at $\mathbf{z}^0$ is the set of all L-derivatives;

$$\partial_{\mathrm{L}}\mathbf{f}(\mathbf{z}^0) := \{\mathbf{J}_{\mathrm{L}}\mathbf{f}(\mathbf{z}^0;\mathbf{M}) : k \in \mathbb{N},\ \mathbf{M} \in \mathbb{R}^{n \times k} \text{ is full row rank}\}.$$

The L-subdifferential is defined for the class of *lexicographically (L-)smooth* functions, which are functions that are locally Lipschitz continuous and have well-defined directional derivative mappings in (6) for any matrix $\mathbf{M}$. All C^1, PC^1 and convex functions, as well as compositions of L-smooth functions are L-smooth. The computational relevancy of L-derivatives of PC^1 functions is captured in the following relation [11, 12, 33]:

$$\partial_{\mathrm{L}}\mathbf{f}(\mathbf{z}^0) = \partial_{\mathrm{B}}[\mathbf{f}'(\mathbf{z}^0;\cdot)](\mathbf{0}_n) \subset \partial_{\mathrm{B}}\mathbf{f}(\mathbf{z}^0) \subset \partial\mathbf{f}(\mathbf{z}^0). \tag{7}$$

The *lexicographic directional (LD-)derivative* [11] of $\mathbf{f}$ at $\mathbf{z}^0$ in the directions $\mathbf{M}$ (not necessarily full row rank) is defined as

$$\mathbf{f}'(\mathbf{z}^0;\mathbf{M}) := \begin{bmatrix}\mathbf{f}^{(0)}_{\mathbf{z}^0,\mathbf{M}}(\mathbf{m}_{(1)}) & \mathbf{f}^{(1)}_{\mathbf{z}^0,\mathbf{M}}(\mathbf{m}_{(2)}) & \cdots & \mathbf{f}^{(k-1)}_{\mathbf{z}^0,\mathbf{M}}(\mathbf{m}_{(k)})\end{bmatrix}.$$

Mirroring the relationship between the Jacobian matrix and directional derivatives in the smooth case,

$$\mathbf{f}'(\mathbf{z}^0;\mathbf{M}) = \mathbf{J}_{\mathrm{L}}\mathbf{f}(\mathbf{z}^0;\mathbf{M})\mathbf{M}, \tag{8}$$

if $\mathbf{M}$ is square and nonsingular ($\mathbf{f}'(\mathbf{z}^0;\mathbf{M}) = \mathbf{J}\mathbf{f}(\mathbf{z}^0)\mathbf{M}$ if $\mathbf{f}$ is differentiable at $\mathbf{z}^0$). Moreover, the LD-derivative obeys sharp calculus rules; the composition $\mathbf{q}\circ\mathbf{f}$ is L-smooth at $\mathbf{z}^0$ and satisfies

$$[\mathbf{q}\circ\mathbf{f}]'(\mathbf{z}^0;\mathbf{M}) = \mathbf{q}'(\mathbf{f}(\mathbf{z}^0);\mathbf{f}'(\mathbf{z}^0;\mathbf{M})), \tag{9}$$

given that $\mathbf{f}$ is L-smooth at $\mathbf{z}^0$, and $\mathbf{q}: Y \subset \mathbb{R}^m \to \mathbb{R}^q$, Y open, is L-smooth at $\mathbf{f}(\mathbf{z}^0)$ [11]. These observations motivate a general procedure for evaluating computationally relevant generalized derivative elements of an L-smooth function as follows:

1. Choose a nonsingular $\mathbf{M}$.
2. Compute an LD-derivative $\mathbf{f}'(\mathbf{z}^0;\mathbf{M})$ by taking advantage of its sharp calculus rules.
3. Solve the linear equation system (8) for the L-derivative $\mathbf{J}_\mathrm{L}\mathbf{f}(\mathbf{z}^0;\mathbf{M})$.

Example 2.1. The mapping $\widetilde{\mathbf{x}}$ in Example 1.2 is PC^1 at $\mathbf{p}^0$; the essentially active selection functions at $\mathbf{p}^0$ are $\{\mathbf{x}_{(1)},\mathbf{x}_{(2)}\}$, where

$$\mathbf{x}_{(1)}(\mathbf{p}) \equiv (p_1, p_2+0.5), \quad \mathbf{x}_{(2)}(\mathbf{p}) \equiv (-p_1, p_2+0.5).$$

Hence, the B-subdifferential of $\widetilde{\mathbf{x}}$ at $\mathbf{p}^0$ is given by

$$\partial_\mathrm{B}\widetilde{\mathbf{x}}(\mathbf{p}^0) = \{\mathbf{J}\mathbf{x}_{(i)}(\mathbf{p}^0) : i=1,2\} = \left\{\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix}\right\}.$$

Given

$$\mathbf{P} = \begin{bmatrix}\mathbf{p}_{(1)} & \mathbf{p}_{(2)}\end{bmatrix} = \begin{bmatrix}P_{11} & P_{12}\\ P_{21} & P_{22}\end{bmatrix} \in \mathbb{R}^{2\times 2},$$

the LD-derivative of $\widetilde{\mathbf{x}}$ at $\mathbf{p}^0$ in the directions $\mathbf{P}$ is calculated as follows: for any $\mathbf{d}\in\mathbb{R}^2$,

$$\begin{aligned}\widetilde{\mathbf{x}}^{(0)}_{\mathbf{p}^0,\mathbf{P}}(\mathbf{d}) &= \widetilde{\mathbf{x}}'(\mathbf{p}^0;\mathbf{d}) = (|d_1|, d_2),\\ \widetilde{\mathbf{x}}^{(1)}_{\mathbf{p}^0,\mathbf{P}}(\mathbf{d}) &= \begin{cases}(d_1,d_2), & \text{if } P_{11}>0 \text{ or } P_{11}=0, d_1\geq 0,\\ (-d_1,d_2), & \text{if } P_{11}<0 \text{ or } P_{11}=0, d_1<0.\end{cases}\end{aligned}$$

Thus,

$$\widetilde{\mathbf{x}}'(\mathbf{p}^0;\mathbf{P}) = \begin{bmatrix}\widetilde{\mathbf{x}}^{(0)}_{\mathbf{p}^0,\mathbf{P}}(\mathbf{p}_{(1)}) & \widetilde{\mathbf{x}}^{(1)}_{\mathbf{p}^0,\mathbf{P}}(\mathbf{p}_{(2)})\end{bmatrix} = \begin{bmatrix}\mathrm{fsign}(P_{11},P_{12})P_{11} & \mathrm{fsign}(P_{11},P_{12})P_{12}\\ P_{21} & P_{22}\end{bmatrix},$$

where the *first-sign function* fsign [34] returns the sign of the first nonzero element in its argument (or zero if the argument is the zero vector). If $\mathbf{P}$ is nonsingular, then

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

and $\mathbf{J}_\mathrm{L}\widetilde{\mathbf{x}}(\mathbf{p}^0;\mathbf{P}) \in \partial_\mathrm{L}\widetilde{\mathbf{x}}(\mathbf{p}^0) \subset \partial_\mathrm{B}\widetilde{\mathbf{x}}(\mathbf{p}^0)$ since $\mathrm{fsign}(P_{11},P_{12}) \neq 0$ by nonsingularity of $\mathbf{P}$.

3. Main Results

Viewing $QP_{(1)}(\mathbf{d})$ in (2) as a parametric NLP (with problem parameter $\mathbf{d}$), a repeated application of Theorem 1.1 yields LD-derivatives of the primal and dual variable solutions of the NLP (1), and thus L-derivatives for a square and nonsingular directions matrix. First, some notational conventions are introduced: for each $j\in\{1,\ldots,k\}$, where $k\in\mathbb{N}$, and any $\mathbf{d}\in\mathbb{R}^{n_p}$, let $QP_{(j)}(\mathbf{d})$ denote the following quadratic program:

$$\begin{aligned}QP_{(j)}(\mathbf{d}):\quad \min_{\mathbf{z}\in\mathbb{R}^{n_x}}\quad & 0.5\mathbf{z}^\mathrm{T}\nabla^2_{\mathbf{xx}}L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0,\boldsymbol{\lambda}^0)\mathbf{z} + \mathbf{z}^\mathrm{T}\nabla^2_{\mathbf{xp}}L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0,\boldsymbol{\lambda}^0)\mathbf{d},\\ \text{s.t.}\quad & \mathbf{J_x}g_i(\mathbf{p}^0,\mathbf{x}^0)\mathbf{z} + \mathbf{J_p}g_i(\mathbf{p}^0,\mathbf{x}^0)\mathbf{d} \leq 0, \quad \forall i\in\mathcal{A}^0_{(j-1)}, && \leftarrow \text{multipliers } \boldsymbol{\eta}_{(j)} \in \mathbb{R}_+^{|\mathcal{A}^0_{(j-1)}|}\\ & \mathbf{J_x}g_i(\mathbf{p}^0,\mathbf{x}^0)\mathbf{z} + \mathbf{J_p}g_i(\mathbf{p}^0,\mathbf{x}^0)\mathbf{d} = 0, \quad \forall i\in\mathcal{A}^+_{(j-1)}, && \leftarrow \text{multipliers } \boldsymbol{\gamma}_{(j)} \in \mathbb{R}^{|\mathcal{A}^+_{(j-1)}|}\\ & \mathbf{J_x}h_i(\mathbf{p}^0,\mathbf{x}^0)\mathbf{z} + \mathbf{J_p}h_i(\mathbf{p}^0,\mathbf{x}^0)\mathbf{d} = 0, \quad \forall i\in\mathcal{H}, && \leftarrow \text{multipliers } \boldsymbol{\rho}_{(j)} \in \mathbb{R}^{n_h}\end{aligned} \tag{10}$$

Optimization Problem	Primal and Dual Solutions	Number of Constraints
NLP	$\mathbf{x}^0 = \widetilde{\mathbf{x}}(\mathbf{p}^0) \quad \& \quad (\boldsymbol{\mu}^0, \boldsymbol{\lambda}^0) = (\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0), \widetilde{\boldsymbol{\lambda}}(\mathbf{p}^0))$	$n_g + n_h$
$\hookrightarrow QP_{(1)}(\mathbf{d})$	$\mathbf{z}_{(1)}(\mathbf{d}) \quad \& \quad (\boldsymbol{\eta}_{(1)}(\mathbf{d}), \boldsymbol{\gamma}_{(1)}(\mathbf{d}), \boldsymbol{\rho}_{(1)}(\mathbf{d}))$	$\searrow$ $\lvert\mathcal{A}^0\rvert + \lvert\mathcal{A}^+\rvert + n_h$
$\hookrightarrow QP_{(2)}(\mathbf{d})$	$\mathbf{z}_{(2)}(\mathbf{d}) \quad \& \quad (\boldsymbol{\eta}_{(2)}(\mathbf{d}), \boldsymbol{\gamma}_{(2)}(\mathbf{d}), \boldsymbol{\rho}_{(2)}(\mathbf{d}))$	$\searrow$ $\lvert\mathcal{A}^0_{(1)}\rvert + \lvert\mathcal{A}^+_{(1)}\rvert + n_h$
$\vdots$	$\vdots$	$\vdots$
$\hookrightarrow QP_{(k)}(\mathbf{d})$	$\mathbf{z}_{(k)}(\mathbf{d}) \quad \& \quad (\boldsymbol{\eta}_{(k)}(\mathbf{d}), \boldsymbol{\gamma}_{(k)}(\mathbf{d}), \boldsymbol{\rho}_{(k)}(\mathbf{d}))$	$\searrow$ $\lvert\mathcal{A}^0_{(k-1)}\rvert + \lvert\mathcal{A}^+_{(k-1)}\rvert + n_h$

Table 1: Summary of relations between optimization problems. Each quadratic program is obtained by linearizing the constraints of its predecessor and quadratically approximating the Lagrangian associated with its predecessor. The number of constraints of each optimization problem is derived from the active sets associated with the predecessor optimization problem in the hierarchy.

where $(\boldsymbol{\eta}_{(j)}, \boldsymbol{\gamma}_{(j)}, \boldsymbol{\rho}_{(j)})$ are dual variables associated with the j^{th} weakly active set, j^{th} strongly active set, and index set $\mathcal{H}$, respectively, where, assuming that $(\mathbf{z}^0_{(j-1)}, \boldsymbol{\eta}^0_{(j-1)}, \boldsymbol{\gamma}^0_{(j-1)}, \boldsymbol{\rho}^0_{(j-1)})$ is a KKT point of $QP_{(j-1)}(\mathbf{d}^0_{(j-1)})$,

$$\begin{aligned}
\mathcal{A}^0_{(j)} &:= \{i \in \mathcal{A}^0_{(j-1)} : \mathbf{J_x} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{z}^0_{(j)} + \mathbf{J_p} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{d}^0_{(j)} = 0 = \eta^0_{(j),i}\}, \\
\mathcal{A}^+_{(j)} &:= \mathcal{A}^+_{(j-1)} \cup \{i \in \mathcal{A}^0_{(j-1)} : \mathbf{J_x} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{z}^0_{(j)} + \mathbf{J_p} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{d}^0_{(j)} = 0 < \eta^0_{(j),i}\}, \\
\mathcal{A}^-_{(j)} &:= \mathcal{A}^-_{(j-1)} \cup \{i \in \mathcal{A}^0_{(j-1)} : \mathbf{J_x} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{z}^0_{(j)} + \mathbf{J_p} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{d}^0_{(j)} < 0 = \eta^0_{(j),i}\},
\end{aligned} \tag{11}$$

and $\mathcal{A}^+_{(0)} := \mathcal{A}^+$, $\mathcal{A}^0_{(0)} := \mathcal{A}^0$, $\mathcal{A}^-_{(0)} := \mathcal{A}^-$. The hierarchy of the optimization problems is illustrated in Table 1. Before giving the main result, the following notation convention is adopted: given a matrix $\mathbf{H} \in \mathbb{R}^{m \times n}$, $\mathbf{H}_{\mathcal{I},j}$ denotes the components of the j^{th} column of $\mathbf{H}$, indexed by the set $\mathcal{I} \equiv \{i_1, \dots, i_s\} \subset \{1, \dots, m\}$:

$$\mathbf{H}_{\mathcal{I},j} \equiv \begin{bmatrix} h_{i_1,j} \\ h_{i_2,j} \\ \vdots \\ h_{i_s,j} \end{bmatrix} \in \mathbb{R}^{s \times 1}.$$

Theorem 3.1. Let $(\mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0) \in D_x \times \mathbb{R}^{n_g}_+ \times \mathbb{R}^{n_h}$ be a KKT point of (1) satisfying SSOSC and let LICQ hold at $(\mathbf{p}^0, \mathbf{x}^0)$. Let $(\widetilde{\mathbf{x}}, \widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\lambda}})$ satisfy the conclusions of Theorem 1.1. Then, for any $k \in \mathbb{N}$ and $\mathbf{P} = [\mathbf{p}_{(1)} \cdots \mathbf{p}_{(k)}] \in \mathbb{R}^{n_p \times k}$, the LD-derivatives of $(\widetilde{\mathbf{x}}, \widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\lambda}})$ at $\mathbf{p}^0$ in the directions $\mathbf{P}$, denoted $(\mathbf{X}, \mathbf{U}, \mathbf{W})$, are given by

$$\begin{aligned}
\mathbf{X} &= \begin{bmatrix} \mathbf{z}_{(1)}(\mathbf{p}_{(1)}) & \mathbf{z}_{(2)}(\mathbf{p}_{(2)}) & \cdots & \mathbf{z}_{(k)}(\mathbf{p}_{(k)}) \end{bmatrix}, \\
\mathbf{W} &= \begin{bmatrix} \boldsymbol{\rho}_{(1)}(\mathbf{p}_{(1)}) & \boldsymbol{\rho}_{(2)}(\mathbf{p}_{(2)}) & \cdots & \boldsymbol{\rho}_{(k)}(\mathbf{p}_{(k)}) \end{bmatrix},
\end{aligned}$$

and for each $j \in \{1, \dots, k\}$,

$$\begin{aligned}
\mathbf{U}_{\mathcal{A}^0_{(j)},j} &= \boldsymbol{\eta}_{(j)}(\mathbf{p}_{(j)}), \\
\mathbf{U}_{\mathcal{A}^+_{(j)},j} &= \boldsymbol{\gamma}_{(j)}(\mathbf{p}_{(j)}), \\
\mathbf{U}_{\mathcal{A}^-_{(j)},j} &= \mathbf{0}_{\lvert\mathcal{A}^-_{(j)}\rvert},
\end{aligned} \tag{12}$$

where $\mathbf{z}_{(j)}(\mathbf{p}_{(j)})$ and $(\boldsymbol{\eta}_{(j)}(\mathbf{p}_{(j)}), \boldsymbol{\gamma}_{(j)}(\mathbf{p}_{(j)}), \boldsymbol{\rho}_{(j)}(\mathbf{p}_{(j)}))$ are the unique primal and dual solutions of $QP_{(j)}(\mathbf{p}_{(j)})$, respectively, evaluated at $\mathbf{p}_{(j)}$.

Proof. Suppose, without loss of generality, that $\mathcal{A}^0, \mathcal{A}^+, \mathcal{A}^- \neq \emptyset$. It will be shown by induction that, for each $j \in \{1, \ldots, k-1\}$, there exist PC^1 mappings

$$\mathbf{z}_{(j)} : \mathbb{R}^{n_p} \to \mathbb{R}^{n_x}$$

and

$$(\boldsymbol{\eta}_{(j)}, \boldsymbol{\gamma}_{(j)}, \boldsymbol{\rho}_{(j)}) : \mathbb{R}^{n_p} \to \mathbb{R}_+^{|\mathcal{A}^0_{(j-1)}|} \times \mathbb{R}^{|\mathcal{A}^+_{(j-1)}|} \times \mathbb{R}^{n_h}$$

such that

$$(\mathbf{z}_{(j)}(\mathbf{p}_{(j)}), \boldsymbol{\eta}_{(j)}(\mathbf{p}_{(j)}), \boldsymbol{\gamma}_{(j)}(\mathbf{p}_{(j)}), \boldsymbol{\rho}_{(j)}(\mathbf{p}_{(j)}))$$

is a KKT point of $QP_{(j)}(\mathbf{p}_{(j)})$ satisfying SSOSC, with LICQ holding at $(\mathbf{p}_{(j)}, \mathbf{z}_{(j)}(\mathbf{p}_{(j)}))$. Moreover, for any $\mathbf{d} \in \mathbb{R}^{n_p}$, the directional derivatives of $(\mathbf{z}_{(j)}, \boldsymbol{\eta}_{(j)}, \boldsymbol{\gamma}_{(j)}, \boldsymbol{\rho}_{(j)})$ at $\mathbf{p}_{(j)}$ in the direction $\mathbf{d}$ satisfy

$$\begin{aligned}
\mathbf{z}'_{(j)}(\mathbf{p}_{(j)}; \mathbf{d}) &= \mathbf{z}_{(j+1)}(\mathbf{d}), \\
\boldsymbol{\gamma}'_{(j)}(\mathbf{p}_{(j)}; \mathbf{d}) &= \boldsymbol{\gamma}_{(j+1)}(\mathbf{d}), \\
\boldsymbol{\rho}'_{(j)}(\mathbf{p}_{(j)}; \mathbf{d}) &= \boldsymbol{\rho}_{(j+1)}(\mathbf{d}), \\
\eta'_{(j),i}(\mathbf{p}_{(j)}; \mathbf{d}) &= \begin{cases} \eta_{(j+1),i}(\mathbf{d}), & \text{if } i \in \mathcal{A}^0_{(j)}, \\ \gamma_{(j+1),i}(\mathbf{d}), & \text{if } i \in \mathcal{A}^+_{(j)}, \\ 0, & \text{if } i \in \mathcal{A}^-_{(j)}, \end{cases}
\end{aligned} \tag{13}$$

where $\mathbf{z}_{(j+1)}(\mathbf{d})$ and $(\boldsymbol{\eta}_{(j+1)}(\mathbf{d}), \boldsymbol{\gamma}_{(j+1)}(\mathbf{d}), \boldsymbol{\rho}_{(j+1)}(\mathbf{d}))$ are the unique primal and dual solutions of $QP_{(j+1)}(\mathbf{d})$, respectively.

Theorem 1.1 implies that, for any $\mathbf{d} \in \mathbb{R}^{n_p}$, $\tilde{\mathbf{x}}'(\mathbf{p}^0; \mathbf{d}) = \mathbf{z}_{(1)}(\mathbf{d})$. Since $\tilde{\mathbf{x}}$ is a PC^1 mapping at $\mathbf{p}^0$, the mapping $\mathbf{z}_{(1)}$ is piecewise linear (and therefore PC^1) on $\mathbb{R}^{n_p}$ [9]. From Equation (3), similar arguments can be made to conclude that $(\boldsymbol{\eta}_{(1)}, \boldsymbol{\gamma}_{(1)}, \boldsymbol{\rho}_{(1)})$ are PC^1 mappings on $\mathbb{R}^{n_p}$. Rewrite the quadratic program $QP_{(1)}(\mathbf{d})$ as follows:

$$\begin{aligned}
\min_{\mathbf{z} \in \mathbb{R}^{n_x}} \quad & f_{(1)}(\mathbf{d}, \mathbf{z}), \\
\text{s.t.} \quad & \mathbf{g}_{(1)}(\mathbf{d}, \mathbf{z}) \leq \mathbf{0}_{|\mathcal{A}^0|}, \\
& \mathbf{h}_{(1)}(\mathbf{d}, \mathbf{z}) = \mathbf{0}_{|\mathcal{A}^+| + n_h},
\end{aligned} \tag{14}$$

where

$$\begin{aligned}
f_{(1)}(\mathbf{d}, \mathbf{z}) &\equiv 0.5\mathbf{z}^{\mathrm{T}} \nabla^2_{\mathbf{xx}} L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0)\mathbf{z} + \mathbf{z}^{\mathrm{T}} \nabla^2_{\mathbf{xp}} L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0)\mathbf{d}, \\
\mathbf{g}_{(1)}(\mathbf{d}, \mathbf{z}) &\equiv \mathbf{J_x g}_{\mathcal{A}^0}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{z} + \mathbf{J_p g}_{\mathcal{A}^0}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{d}, \\
\mathbf{h}_{(1)}(\mathbf{d}, \mathbf{z}) &\equiv \begin{bmatrix} \mathbf{J_x g}_{\mathcal{A}^+}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{z} + \mathbf{J_p g}_{\mathcal{A}^+}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{d} \\ \mathbf{J_x h}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{z} + \mathbf{J_p h}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{d} \end{bmatrix}.
\end{aligned}$$

Let the active index set of $QP_{(1)}(\mathbf{p}_{(1)})$ at $(\mathbf{p}_{(1)}, \mathbf{z}_{(1)}(\mathbf{p}_{(1)}))$ be denoted by

$$\mathcal{A}_{(1)} := \{i \in \mathcal{A}^0 : \mathbf{J_x} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{z}_{(1)}(\mathbf{p}_{(1)}) + \mathbf{J_p} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{p}_{(1)} = 0\}.$$

Noting that, for any $(\mathbf{d}, \mathbf{z})$,

$$\begin{aligned}
\mathbf{J_z g}_{(1)}(\mathbf{d}, \mathbf{z}) &= \mathbf{J_x g}_{\mathcal{A}^0}(\mathbf{p}^0, \mathbf{x}^0), \\
\mathbf{J_z h}_{(1)}(\mathbf{d}, \mathbf{z}) &= \begin{bmatrix} \mathbf{J_x g}_{\mathcal{A}^+}(\mathbf{p}^0, \mathbf{x}^0) \\ \mathbf{J_x h}(\mathbf{p}^0, \mathbf{x}^0) \end{bmatrix},
\end{aligned} \tag{15}$$

it follows that LICQ holds at $(\mathbf{p}_{(1)}, \mathbf{z}_{(1)}(\mathbf{p}_{(1)}))$ with respect to $QP_{(1)}(\mathbf{p}_{(1)})$ since

$$\begin{aligned}
&\{(\mathbf{J_x} g_i(\mathbf{p}^0, \mathbf{x}^0))^{\mathrm{T}} : i \in \mathcal{A}_{(1)} \cup \mathcal{A}^+\} \cup \{(\mathbf{J_x} h_i(\mathbf{p}^0, \mathbf{x}^0))^{\mathrm{T}} : i \in \mathcal{H}\} \\
&\subset \{(\mathbf{J_x} g_i(\mathbf{p}^0, \mathbf{x}^0))^{\mathrm{T}} : i \in \mathcal{A}\} \cup \{(\mathbf{J_x} h_i(\mathbf{p}^0, \mathbf{x}^0))^{\mathrm{T}} : i \in \mathcal{H}\},
\end{aligned}$$

and LICQ holds at $(\mathbf{p}^0, \mathbf{x}^0)$ with respect to the NLP (1).

Letting $L_{(1)}$ denote the Lagrangian associated with (14), it follows that

$$\begin{aligned}\nabla^2_{\mathbf{zz}} L_{(1)}(\mathbf{d},\mathbf{z},\boldsymbol{\eta},\boldsymbol{\gamma},\boldsymbol{\rho}) &= \nabla^2_{\mathbf{xx}} L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0,\boldsymbol{\lambda}^0),\\ \nabla^2_{\mathbf{zd}} L_{(1)}(\mathbf{d},\mathbf{z},\boldsymbol{\eta},\boldsymbol{\gamma},\boldsymbol{\rho}) &= \nabla^2_{\mathbf{xp}} L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0,\boldsymbol{\lambda}^0),\end{aligned} \tag{16}$$

for any $(\mathbf{d},\mathbf{z},\boldsymbol{\eta},\boldsymbol{\gamma},\boldsymbol{\rho})$. Since SSOSC holds at $(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0,\boldsymbol{\lambda}^0)$ with respect to the NLP (1),

$$\mathbf{v}\nabla^2_{\mathbf{xx}} L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0,\boldsymbol{\lambda}^0)\mathbf{v} > 0, \quad \forall \mathbf{v} \in \mathcal{K} := \{\mathbf{v} \in \mathbb{R}^{n_x} \setminus \{\mathbf{0}_{n_x}\} : \mathbf{J_x g}_{\mathcal{A}^+}(\mathbf{p}^0,\mathbf{x}^0)\mathbf{v} = \mathbf{0}_{|\mathcal{A}^+|},\ \mathbf{J_x h}(\mathbf{p}^0,\mathbf{x}^0)\mathbf{v} = \mathbf{0}_{n_h}\}.$$

Letting $\mathcal{A}^+_{(1)}$ be defined as in Equation (11), observe that

$$\begin{aligned}\mathcal{K}_{(1)} &:= \{\mathbf{v} \in \mathbb{R}^{n_x} \setminus \{\mathbf{0}_{n_x}\} : \mathbf{J_z g}_{(1),\mathcal{A}^+_{(1)}}(\mathbf{d},\mathbf{z}_{(1)}(\mathbf{d}))\mathbf{v} = \mathbf{0}_{|\mathcal{A}^+_{(1)}|},\ \mathbf{J_z h}_{(1)}(\mathbf{d},\mathbf{z}_{(1)}(\mathbf{d}))\mathbf{v} = \mathbf{0}_{|\mathcal{A}^+|+n_h}\},\\ &= \{\mathbf{v} \in \mathbb{R}^{n_x} \setminus \{\mathbf{0}_{n_x}\} : \mathbf{J_x g}_{\mathcal{A}^+_{(1)}\cup\mathcal{A}^+}(\mathbf{p}^0,\mathbf{x}^0)\mathbf{v} = \mathbf{0}_{|\mathcal{A}^+_{(1)}|+|\mathcal{A}^+|},\ \mathbf{J_x h}(\mathbf{p}^0,\mathbf{x}^0)\mathbf{v} = \mathbf{0}_{n_h}\},\\ &\subset \{\mathbf{v} \in \mathbb{R}^{n_x} \setminus \{\mathbf{0}_{n_x}\} : \mathbf{J_x g}_{\mathcal{A}^+}(\mathbf{p}^0,\mathbf{x}^0)\mathbf{v} = \mathbf{0}_{|\mathcal{A}^+|},\ \mathbf{J_x h}(\mathbf{p}^0,\mathbf{x}^0)\mathbf{v} = \mathbf{0}_{n_h}\},\\ &= \mathcal{K}.\end{aligned}$$

Hence,

$$\mathbf{v}^{\mathrm{T}}\nabla^2_{\mathbf{zz}} L_{(1)}(\mathbf{p}_{(1)},\mathbf{z}_{(1)}(\mathbf{p}_{(1)}),\boldsymbol{\eta}_{(1)}(\mathbf{p}_{(1)}),\boldsymbol{\gamma}_{(1)}(\mathbf{p}_{(1)}),\boldsymbol{\rho}_{(1)}(\mathbf{p}_{(1)}))\mathbf{v} = \mathbf{v}^{\mathrm{T}}\nabla^2_{\mathbf{xx}} L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0,\boldsymbol{\lambda}^0)\mathbf{v} > 0$$

for all $\mathbf{v} \in \mathcal{K}_{(1)} \subset \mathcal{K}$, from which it follows that $(\mathbf{p}_{(1)},\mathbf{z}_{(1)}(\mathbf{p}_{(1)}),\boldsymbol{\eta}_{(1)}(\mathbf{p}_{(1)}),\boldsymbol{\gamma}_{(1)}(\mathbf{p}_{(1)}),\boldsymbol{\rho}_{(1)}(\mathbf{p}_{(1)}))$ satisfies SSOSC with respect to $QP_{(1)}(\mathbf{p}_{(1)})$.

The conditions of Theorem 1.1 are satisfied by (14) at $(\mathbf{p}_{(1)},\mathbf{z}_{(1)}(\mathbf{p}_{(1)}),\boldsymbol{\eta}_{(1)}(\mathbf{p}_{(1)}),\boldsymbol{\gamma}_{(1)}(\mathbf{p}_{(1)}),\boldsymbol{\rho}_{(1)}(\mathbf{p}_{(1)}))$. Moreover, it follows from the observations above (namely, (15) and (16)) that the auxiliary quadratic program (i.e., (2)) associated with (14) at $(\mathbf{p}_{(1)},\mathbf{z}_{(1)}(\mathbf{p}_{(1)}),\boldsymbol{\eta}_{(1)}(\mathbf{p}_{(1)}),\boldsymbol{\gamma}_{(1)}(\mathbf{p}_{(1)}),\boldsymbol{\rho}_{(1)}(\mathbf{p}_{(1)}))$ is given by $QP_{(2)}(\mathbf{d})$. Thus, for any $\mathbf{d} \in \mathbb{R}^{n_p}$, the directional derivatives of $(\mathbf{z}_{(1)},\boldsymbol{\eta}_{(1)},\boldsymbol{\gamma}_{(1)},\boldsymbol{\rho}_{(1)})$ at $\mathbf{p}_{(1)}$ in the direction $\mathbf{d}$ satisfy

$$\begin{aligned}\mathbf{z}'_{(1)}(\mathbf{p}_{(1)};\mathbf{d}) &= \mathbf{z}_{(2)}(\mathbf{d}),\\ \boldsymbol{\gamma}'_{(1)}(\mathbf{p}_{(1)};\mathbf{d}) &= \boldsymbol{\gamma}_{(2)}(\mathbf{d}),\\ \boldsymbol{\rho}'_{(1)}(\mathbf{p}_{(1)};\mathbf{d}) &= \boldsymbol{\rho}_{(2)}(\mathbf{d}),\\ \eta'_{(1),i}(\mathbf{p}_{(1)};\mathbf{d}) &= \begin{cases}\eta_{(2),i}(\mathbf{d}), & \text{if } i \in \mathcal{A}^0_{(1)},\\ \gamma_{(2),i}(\mathbf{d}), & \text{if } i \in \mathcal{A}^+_{(1)},\\ 0, & \text{if } i \in \mathcal{A}^-_{(1)},\end{cases}\end{aligned} \tag{17}$$

where $\mathbf{z}_{(2)}(\mathbf{d})$ and $(\boldsymbol{\eta}_{(2)}(\mathbf{d}),\boldsymbol{\gamma}_{(2)}(\mathbf{d}),\boldsymbol{\rho}_{(2)}(\mathbf{d}))$ are the unique primal and dual solutions of $QP_{(2)}(\mathbf{d})$, respectively. Thus, the base case is proved.

Assume the claim holds for $j^* \in \{2,\ldots,k-2\}$. That is, there exist PC^1 mappings

$$\mathbf{z}_{(j^*)} : \mathbb{R}^{n_p} \to \mathbb{R}^{n_x}$$

and

$$(\boldsymbol{\eta}_{(j^*)},\boldsymbol{\gamma}_{(j^*)},\boldsymbol{\rho}_{(j^*)}) : \mathbb{R}^{n_p} \to \mathbb{R}_+^{|\mathcal{A}^0_{(j^*-1)}|} \times \mathbb{R}^{|\mathcal{A}^+_{(j^*-1)}|} \times \mathbb{R}^{n_h}$$

such that

$$(\mathbf{z}_{(j^*)}(\mathbf{p}_{(j^*)}),\boldsymbol{\eta}_{(j^*)}(\mathbf{p}_{(j^*)}),\boldsymbol{\gamma}_{(j^*)}(\mathbf{p}_{(j^*)}),\boldsymbol{\rho}_{(j^*)}(\mathbf{p}_{(j^*)}))$$

is a KKT point of $QP_{(j^*)}(\mathbf{p}_{(j^*)})$ satisfying SSOSC, with LICQ holding at $(\mathbf{p}_{(j^*)},\mathbf{z}_{(j^*)}(\mathbf{p}_{(j^*)}))$. Moreover, for any $\mathbf{d} \in \mathbb{R}^{n_p}$, the directional derivatives of $(\mathbf{z}_{(j^*)},\boldsymbol{\eta}_{(j^*)},\boldsymbol{\gamma}_{(j^*)},\boldsymbol{\rho}_{(j^*)})$ at $\mathbf{p}_{(j^*)}$ in the direction $\mathbf{d}$ satisfy

$$\begin{aligned}\mathbf{z}'_{(j^*)}(\mathbf{p}_{(j^*)};\mathbf{d}) &= \mathbf{z}_{(j^*+1)}(\mathbf{d}),\\ \boldsymbol{\gamma}'_{(j^*)}(\mathbf{p}_{(j^*)};\mathbf{d}) &= \boldsymbol{\gamma}_{(j^*+1)}(\mathbf{d}),\\ \boldsymbol{\rho}'_{(j^*)}(\mathbf{p}_{(j^*)};\mathbf{d}) &= \boldsymbol{\rho}_{(j^*+1)}(\mathbf{d}),\\ \eta'_{(j^*),i}(\mathbf{p}_{(j^*)};\mathbf{d}) &= \begin{cases}\eta_{(j^*+1),i}(\mathbf{d}), & \text{if } i \in \mathcal{A}^0_{(j^*)},\\ \gamma_{(j^*+1),i}(\mathbf{d}), & \text{if } i \in \mathcal{A}^+_{(j^*)},\\ 0, & \text{if } i \in \mathcal{A}^-_{(j^*)},\end{cases}\end{aligned} \tag{18}$$

where $\mathbf{z}_{(j^*+1)}(\mathbf{d})$ and $(\boldsymbol{\eta}_{(j^*+1)}(\mathbf{d}), \boldsymbol{\gamma}_{(j^*+1)}(\mathbf{d}), \boldsymbol{\rho}_{(j^*+1)}(\mathbf{d}))$ are the unique primal and dual solutions of $QP_{(j^*+1)}(\mathbf{d})$, respectively, evaluated at $\mathbf{d}$.

Rewrite $QP_{(j^*+1)}(\mathbf{d})$ as

$$\begin{aligned} \min_{\mathbf{z}\in\mathbb{R}^{n_x}} \quad & f_{(j^*+1)}(\mathbf{d}, \mathbf{z}), \\ \text{s.t.} \quad & \mathbf{g}_{(j^*+1)}(\mathbf{d}, \mathbf{z}) \leq \mathbf{0}_{|\mathcal{A}^0_{(j^*)}|}, \\ & \mathbf{h}_{(j^*+1)}(\mathbf{d}, \mathbf{z}) = \mathbf{0}_{|\mathcal{A}^+_{(j^*)}|+n_h}, \end{aligned} \tag{19}$$

where

$$\begin{aligned} f_{(j^*+1)}(\mathbf{d}, \mathbf{z}) &\equiv 0.5\mathbf{z}^{\mathrm{T}}\nabla^2_{\mathbf{xx}}L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0)\mathbf{z} + \mathbf{z}^{\mathrm{T}}\nabla^2_{\mathbf{xp}}L(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0)\mathbf{d}, \\ \mathbf{g}_{(j^*+1)}(\mathbf{d}, \mathbf{z}) &\equiv \mathbf{J_x g}_{\mathcal{A}^0_{(j^*)}}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{z} + \mathbf{J_p g}_{\mathcal{A}^0_{(j^*)}}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{d}, \\ \mathbf{h}_{(j^*+1)}(\mathbf{d}, \mathbf{z}) &\equiv \begin{bmatrix} \mathbf{J_x g}_{\mathcal{A}^+_{(j^*)}}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{z} + \mathbf{J_p g}_{\mathcal{A}^+_{(j^*)}}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{d} \\ \mathbf{J_x h}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{z} + \mathbf{J_p h}(\mathbf{p}^0, \mathbf{x}^0)\mathbf{d} \end{bmatrix}. \end{aligned}$$

Again, by the same arguments as in the base case, $\mathbf{z}_{(j^*+1)}$ and $(\boldsymbol{\eta}_{(j^*)}, \boldsymbol{\gamma}_{(j^*)}, \boldsymbol{\rho}_{(j^*)})$ are piecewise linear (and thus PC^1) mappings on $\mathbb{R}^{n_p}$, with LICQ holding at $(\mathbf{p}_{(j^*+1)}, \mathbf{z}_{(j^*+1)}(\mathbf{p}_{(j^*+1)}))$ and SSOSC holding at the KKT point

$$(\mathbf{z}_{(j^*+1)}(\mathbf{p}_{(j^*+1)}), \boldsymbol{\eta}_{(j^*+1)}(\mathbf{p}_{(j^*+1)}), \boldsymbol{\gamma}_{(j^*+1)}(\mathbf{p}_{(j^*+1)}), \boldsymbol{\rho}_{(j^*+1)}(\mathbf{p}_{(j^*+1)}))$$

with respect to $QP_{(j^*+1)}(\mathbf{p}_{(j^*+1)})$. Moreover, it follows similarly as in the base case that the auxiliary quadratic program (i.e., (2)) associated with (19) at

$$(\mathbf{p}_{(j^*+1)}, \mathbf{z}_{(j^*+1)}(\mathbf{p}_{(j^*+1)}), \boldsymbol{\eta}_{(j^*+1)}(\mathbf{p}_{(j^*+1)}), \boldsymbol{\gamma}_{(j^*+1)}(\mathbf{p}_{(j^*+1)}), \boldsymbol{\rho}_{(j^*+1)}(\mathbf{p}_{(j^*+1)}))$$

is given by $QP_{(j^*+2)}(\mathbf{d})$. Thus, Theorem 1.1 may be applied to yield

$$\begin{aligned} \mathbf{z}'_{(j^*+1)}(\mathbf{p}_{(j^*+1)}; \mathbf{d}) &= \mathbf{z}_{(j^*+2)}(\mathbf{d}), \\ \boldsymbol{\gamma}'_{(j^*+1)}(\mathbf{p}_{(j^*+1)}; \mathbf{d}) &= \boldsymbol{\gamma}_{(j^*+2)}(\mathbf{d}), \\ \boldsymbol{\rho}'_{(j^*+1)}(\mathbf{p}_{(j^*+1)}; \mathbf{d}) &= \boldsymbol{\rho}_{(j^*+2)}(\mathbf{d}), \\ \eta'_{(j^*+1),i}(\mathbf{p}_{(j^*+1)}; \mathbf{d}) &= \begin{cases} \eta_{(j^*+2),i}(\mathbf{d}), & \text{if } i \in \mathcal{A}^0_{(j^*+1)}, \\ \gamma_{(j^*+2),i}(\mathbf{d}), & \text{if } i \in \mathcal{A}^+_{(j^*+1)}, \\ 0, & \text{if } i \in \mathcal{A}^-_{(j^*+1)}, \end{cases} \end{aligned} \tag{20}$$

where $\mathbf{z}_{(j^*+2)}(\mathbf{d})$ and $(\boldsymbol{\eta}_{(j^*+2)}(\mathbf{d}), \boldsymbol{\gamma}_{(j^*+2)}(\mathbf{d}), \boldsymbol{\rho}_{(j^*+2)}(\mathbf{d}))$ are the unique primal and dual solutions of $QP_{(j^*+2)}(\mathbf{d})$, respectively, evaluated at $\mathbf{d}$, and the claim is proved.

Recalling that $\widetilde{\mathbf{x}}^{(0)}_{\mathbf{p}^0,\mathbf{P}}(\mathbf{p}_{(1)}) = \widetilde{\mathbf{x}}'(\mathbf{p}^0; \mathbf{p}_{(1)}) = \mathbf{z}_{(1)}(\mathbf{p}_{(1)})$ from Theorem 1.1, it has been shown that

$$\widetilde{\mathbf{x}}^{(j)}_{\mathbf{p}^0,\mathbf{P}}(\mathbf{p}_{(j+1)}) = \mathbf{z}_{(j+1)}(\mathbf{p}_{(j+1)}), \quad \forall j \in \{0, \ldots, k-1\}.$$

The result then follows by definition of LD-derivative;

$$\begin{aligned} \widetilde{\mathbf{x}}'(\mathbf{p}^0; \mathbf{P}) &= \begin{bmatrix} \widetilde{\mathbf{x}}^{(0)}_{\mathbf{p}^0,\mathbf{P}}(\mathbf{p}_{(1)}) & \widetilde{\mathbf{x}}^{(1)}_{\mathbf{p}^0,\mathbf{P}}(\mathbf{p}_{(2)}) & \cdots & \widetilde{\mathbf{x}}^{(k-1)}_{\mathbf{p}^0,\mathbf{P}}(\mathbf{p}_{(k)}) \end{bmatrix}, \\ &= \begin{bmatrix} \mathbf{z}_{(1)}(\mathbf{p}_{(1)}) & \mathbf{z}_{(2)}(\mathbf{p}_{(2)}) & \cdots & \mathbf{z}_{(k)}(\mathbf{p}_{(k)}) \end{bmatrix}, \end{aligned}$$

with $\widetilde{\boldsymbol{\lambda}}'(\mathbf{p}^0; \mathbf{P})$ and $\widetilde{\boldsymbol{\mu}}'(\mathbf{p}^0; \mathbf{P})$ following similarly. □

Example 3.2. Consider again Example 1.2 and let

$$\mathbf{P} = [\mathbf{p}_{(1)} \quad \mathbf{p}_{(2)}] = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \in \mathbb{R}^{2\times 2}.$$

Viewing the quadratic program $QP_{(1)}$ in (5) as a parametric nonlinear program, its auxiliary quadratic program $QP_{(2)}$ depends on the choice of $\mathbf{p}_{(1)}$ since

$$\mathcal{A}^0_{(1)} = \begin{cases} \{1\}, & \text{if } P_{11} = 0, \\ \emptyset, & \text{if } P_{11} \neq 0, \end{cases}$$

$$\mathcal{A}^+_{(1)} = \begin{cases} \{1,3\}, & \text{if } P_{11} > 0, \\ \{3\}, & \text{if } P_{11} \leq 0, \end{cases}$$

$$\mathcal{A}^-_{(1)} = \begin{cases} \{1,2\}, & \text{if } P_{11} < 0, \\ \{2\}, & \text{if } P_{11} \geq 0. \end{cases}$$

If $P_{11} = 0$ then the quadratic program $QP_{(2)}$ in (10) associated with (5) is given by

$$\begin{aligned} Q_{(2)}(\mathbf{d}): \quad \min_{\mathbf{z}\in\mathbb{R}^2} \quad & z_1^2 + z_2^2 + 2(d_1 z_1 + d_2 z_2), \\ \text{s.t.} \quad & -z_1 + d_1 \leq 0, \\ & -z_2 + d_2 = 0, \end{aligned}$$

(i.e., the same quadratic program as (5)), which admits primal and dual variable solutions, for any $\mathbf{d} \in \mathbb{R}^2$,

$$\mathbf{z}_{(2)}(\mathbf{d}) \equiv (|d_1|, d_2), \quad \eta_{(2)}(\mathbf{d}) \equiv \max(4d_1, 0), \quad \gamma_{(2)}(\mathbf{d}) \equiv 4d_2.$$

If $P_{11} > 0$ then the quadratic program $QP_{(2)}$ in (10) associated with (5) is given by

$$\begin{aligned} Q_{(2)}(\mathbf{d}): \quad \min_{\mathbf{z}\in\mathbb{R}^2} \quad & z_1^2 + z_2^2 + 2(d_1 z_1 + d_2 z_2), \\ \text{s.t.} \quad & -z_1 + d_1 = 0, \\ & -z_2 + d_2 = 0, \end{aligned}$$

which admits primal and dual variable solutions, for any $\mathbf{d} \in \mathbb{R}^2$,

$$\mathbf{z}_{(2)}(\mathbf{d}) \equiv (d_1, d_2), \quad \boldsymbol{\gamma}_{(2)}(\mathbf{d}) \equiv (4d_1, 4d_2).$$

If $P_{11} < 0$ then the quadratic program $QP_{(2)}$ in (10) associated with (5) is given by

$$\begin{aligned} Q_{(2)}(\mathbf{d}): \quad \min_{\mathbf{z}\in\mathbb{R}^2} \quad & z_1^2 + z_2^2 + 2(d_1 z_1 + d_2 z_2), \\ \text{s.t.} \quad & -z_2 + d_2 = 0, \end{aligned}$$

which admits primal and dual variable solutions, for any $\mathbf{d} \in \mathbb{R}^2$,

$$\mathbf{z}_{(2)}(\mathbf{d}) \equiv (-d_1, d_2), \quad \gamma_{(2)}(\mathbf{d}) \equiv 4d_2.$$

According to Theorem 3.1, the parametric sensitivities of the primal and dual variables of the original nonlinear program (i.e., (4)) are constructed as follows:

$$\widetilde{\mathbf{x}}'(\mathbf{p}^0; \mathbf{P}) = \begin{bmatrix} \mathbf{z}_{(1)}(\mathbf{p}_{(1)}) & \mathbf{z}_{(2)}(\mathbf{p}_{(2)}) \end{bmatrix} = \begin{cases} \begin{bmatrix} 0 & |P_{12}| \\ P_{21} & P_{22} \end{bmatrix}, & \text{if } P_{11} = 0, \\ \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}, & \text{if } P_{11} > 0, \\ \begin{bmatrix} -P_{11} & -P_{12} \\ P_{21} & P_{22} \end{bmatrix}, & \text{if } P_{11} < 0, \end{cases}$$

and

$$\widetilde{\boldsymbol{\mu}}'(\mathbf{p}^0;\mathbf{P}) = \begin{cases} \begin{bmatrix} \eta_{(1)}(\mathbf{p}_{(1)}) & \eta_{(2)}(\mathbf{p}_{(2)}) \\ 0 & 0 \\ \gamma_{(1)}(\mathbf{p}_{(1)}) & \gamma_{(2)}(\mathbf{p}_{(2)}) \end{bmatrix} = \begin{bmatrix} 0 & \max(4P_{12},0) \\ 0 & 0 \\ 4P_{21} & 4P_{22} \end{bmatrix}, & \text{if } P_{11} = 0, \\ \begin{bmatrix} \eta_{(1)}(\mathbf{p}_{(1)}) & \gamma_{(2),1}(\mathbf{p}_{(2)}) \\ 0 & 0 \\ \gamma_{(1)}(\mathbf{p}_{(1)}) & \gamma_{(2),2}(\mathbf{p}_{(2)}) \end{bmatrix} = \begin{bmatrix} 4P_{11} & 4P_{12} \\ 0 & 0 \\ 4P_{21} & 4P_{22} \end{bmatrix}, & \text{if } P_{11} > 0, \\ = \begin{bmatrix} \eta_{(1)}(\mathbf{p}_{(1)}) & 0 \\ 0 & 0 \\ \gamma_{(1)}(\mathbf{p}_{(1)}) & \gamma_{(2)}(\mathbf{p}_{(2)}) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 4P_{21} & 4P_{22} \end{bmatrix}, & \text{if } P_{11} < 0. \end{cases}$$

The LD-derivatives furnished in Theorem 3.1 can be used to compute L-derivatives using a nonsingular directions matrix and solving the linear equation system (8). See Algorithm 1 for a practically implementable method for computing such L-derivatives: the sequence of QPs are solved in Line 3, which requires, for example, an interior-point method or active-set method (see [35, 36]). The active sets are updated in the loop beginning on Line 5, possibly resulting in inequality constraints being removed or becoming equalities in the next QP solve. Line 19 furnishes an L-derivative, from the linear equation system (8).

Algorithm 1 Evaluate L-Derivatives of Primal and Dual Variable Solutions

Require: KKT point $(\mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0)$; index sets $\mathcal{A}^0$, $\mathcal{A}^+$, $\mathcal{A}^-$; nonsingular $\mathbf{P} = [\mathbf{p}_{(1)} \quad \cdots \quad \mathbf{p}_{(n_p)}] \in \mathbb{R}^{n_p \times n_p}$

1: **procedure** CALCULATE $(\mathbf{J}_\mathrm{L}\widetilde{\mathbf{x}}(\mathbf{p}^0;\mathbf{P}), \mathbf{J}_\mathrm{L}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0;\mathbf{P}), \mathbf{J}_\mathrm{L}\widetilde{\boldsymbol{\lambda}}(\mathbf{p}^0;\mathbf{P}))$
2:   **for** $j = 1, \ldots, p$ **do**
3:     Solve $QP_{(j)}(\mathbf{p}_{(j)})$ for unique primal and dual solutions $\mathbf{z}_{(j)}$ and $(\boldsymbol{\eta}_{(j)}, \boldsymbol{\gamma}_{(j)}, \boldsymbol{\rho}_{(j)})$, respectively.
4:     Set $\mathcal{A}^0_{(j)} \leftarrow \mathcal{A}^0$, $\mathcal{A}^+_{(j)} \leftarrow \mathcal{A}^+$, $\mathcal{A}^-_{(j)} \leftarrow \mathcal{A}^-$.
5:     **for all** $i \in \mathcal{A}^0$ **do**
6:       **if** $\mathbf{J}_\mathbf{x} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{z}_{(j)} + \mathbf{J}_\mathbf{p} g_i(\mathbf{p}^0, \mathbf{x}^0)\mathbf{p}_{(j)} < 0$ **then**
7:         Set $\mathcal{A}^0_{(j)} \leftarrow \mathcal{A}^0_{(j)} \setminus \{i\}$, $\mathcal{A}^-_{(j)} \leftarrow \mathcal{A}^-_{(j)} \cup \{i\}$
8:       **else if** $\eta_{(j),i} > 0$ **then**
9:         Set $\mathcal{A}^0_{(j)} \leftarrow \mathcal{A}^0_{(j)} \setminus \{i\}$, $\mathcal{A}^+_{(j)} \leftarrow \mathcal{A}^+_{(j)} \cup \{i\}$
10:       **end if**
11:     **end for**
12:     Set $\mathcal{A}^0 \leftarrow \mathcal{A}^0_{(j)}$, $\mathcal{A}^+ \leftarrow \mathcal{A}^+_{(j)}$, $\mathcal{A}^- \leftarrow \mathcal{A}^-_{(j)}$.
13:     Set $\mathbf{X} \leftarrow \begin{bmatrix} \mathbf{X} & \mathbf{z}_{(j)} \end{bmatrix}$
14:     Set $\mathbf{W} \leftarrow \begin{bmatrix} \mathbf{W} & \boldsymbol{\rho}_{(j)} \end{bmatrix}$
15:     Set $\mathbf{U}_{\mathcal{A}^0,j} \leftarrow \begin{bmatrix} \mathbf{U}_{\mathcal{A}^0,j} & \boldsymbol{\eta}_{(j)} \end{bmatrix}$
16:     Set $\mathbf{U}_{\mathcal{A}^+,j} \leftarrow \begin{bmatrix} \mathbf{U}_{\mathcal{A}^+,j} & \boldsymbol{\rho}_{(j)} \end{bmatrix}$
17:     Set $\mathbf{U}_{\mathcal{A}^-,j} \leftarrow \begin{bmatrix} \mathbf{U}_{\mathcal{A}^-,j} & \mathbf{0}_{|\mathcal{A}^0|} \end{bmatrix}$
18:   **end for**
19:   Solve the equation system

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{U} \\ \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_\mathrm{L} \\ \mathbf{U}_\mathrm{L} \\ \mathbf{W}_\mathrm{L} \end{bmatrix} \mathbf{P}$$

20:   for $(\mathbf{X}_\mathrm{L}, \mathbf{U}_\mathrm{L}, \mathbf{W}_\mathrm{L}) \in \mathbb{R}^{(n_x+n_g+n_h)\times n_p}$.
21:   **return** $(\mathbf{X}_\mathrm{L}, \mathbf{U}_\mathrm{L}, \mathbf{W}_\mathrm{L})$.
22: **end procedure**

Example 3.3. Returning to the LD-derivatives found in Example 3.2, choosing $\mathbf{P} = \mathbf{I}_2$ in Algorithm 1 yields

$$\mathbf{X}_{\mathrm{L}} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}^{-1} \in \partial_{\mathrm{B}}\widetilde{\mathbf{x}}(\mathbf{p}^0) = \left\{\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}, \begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix}\right\},$$

and

$$\mathbf{U}_{\mathrm{L}} = \begin{bmatrix}4 & 0\\ 0 & 0\\ 0 & 4\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}^{-1} = \begin{bmatrix}4 & 0\\ 0 & 0\\ 0 & 4\end{bmatrix} \in \partial_{\mathrm{B}}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0) = \left\{\begin{bmatrix}4 & 0\\ 0 & 0\\ 0 & 4\end{bmatrix}, \begin{bmatrix}0 & 0\\ 0 & 0\\ 0 & 4\end{bmatrix}\right\}.$$

If instead $\mathbf{P} = -\mathbf{I}_2$ is chosen, then

$$\mathbf{X}_{\mathrm{L}} = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}\begin{bmatrix}-1 & 0\\ 0 & -1\end{bmatrix}^{-1} = \begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix} \in \partial_{\mathrm{B}}\widetilde{\mathbf{x}}(\mathbf{p}^0),$$

and

$$\mathbf{U}_{\mathrm{L}} = \begin{bmatrix}0 & 0\\ 0 & 0\\ 0 & -4\end{bmatrix}\begin{bmatrix}-1 & 0\\ 0 & -1\end{bmatrix}^{-1} = \begin{bmatrix}0 & 0\\ 0 & 0\\ 0 & 4\end{bmatrix} \in \partial_{\mathrm{B}}\widetilde{\boldsymbol{\mu}}(\mathbf{p}^0).$$

(Recall that $\mathcal{H} = \emptyset$.)

Remark 3.4. Theorem 3.1 can be placed in the context of the findings in [24], where sensitivities of parametric NLPs with active index set changes are calculated by instead applying a nonsmooth implicit function theorem to a nonsmooth NLP KKT system reformulation. For these purposes, functions based on lexicographical ordering are introduced: define the generalized inequalities $\prec$ and $\preceq$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ as

$$\begin{aligned}&\mathbf{a} \prec \mathbf{b} \text{ if and only if } \exists j \in \{1, \dots, n\} \text{ s.t. } a_i = b_i\ \forall i < j \text{ and } a_j < b_j,\\ &\mathbf{a} \preceq \mathbf{b} \text{ if and only if } \mathbf{a} = \mathbf{b} \text{ or } \mathbf{a} \prec \mathbf{b},\end{aligned}$$

with $\succ$ and $\succeq$ similarly defined. Define the *lexicographic-minimum function* by

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

and the *lexicographic-matrix-minimum* by

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

where $\mathbf{A}_i$ and $\mathbf{B}_i$ denote the i^{th} rows of $\mathbf{A}$ and $\mathbf{B}$, respectively. $\mathbf{LMmin}$ compares two matrices lexicographically (by rows). Then, assuming the setting of Theorem 3.1, the LD-derivatives of $(\widetilde{\mathbf{x}}, \widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\lambda}})$ at $\mathbf{p}^0$ in the directions $\mathbf{P}$, denoted $(\mathbf{X}, \mathbf{U}, \mathbf{W})$, are the unique solution of the following nonsmooth equation system:

$$\begin{aligned}\left[\begin{array}{c|c|c}\nabla^2_{\mathbf{xx}}L & (\mathbf{J_x g}_{\mathcal{A}^+\cup\mathcal{A}^0})^{\mathrm{T}} & (\mathbf{J_x h})^{\mathrm{T}}\\ \hline -\mathbf{J_x g}_{\mathcal{A}^+} & \mathbf{0}_{|\mathcal{A}^+|\times(|\mathcal{A}^+|+|\mathcal{A}^0|)} & \mathbf{0}_{|\mathcal{A}^+|\times n_h}\\ \hline -\mathbf{J_x h} & \mathbf{0}_{n_h\times(|\mathcal{A}^+|+|\mathcal{A}^0|)} & \mathbf{0}_{n_h\times n_h}\end{array}\right]\begin{bmatrix}\mathbf{X}\\ \mathbf{U}_{\mathcal{A}^+\cup\mathcal{A}^0,\bullet}\\ \mathbf{W}\end{bmatrix} &= \begin{bmatrix}-\nabla^2_{\mathbf{xp}}L\\ \mathbf{J_p g}_{\mathcal{A}^+}\\ \mathbf{J_p h}\end{bmatrix}\mathbf{P},\\ \mathbf{U}_{\mathcal{A}^-,\bullet} &= \mathbf{0}_{|\mathcal{A}^-|\times k},\\ \mathbf{LMmin}\left(-\mathbf{J_p g}_{\mathcal{A}^0}\mathbf{P} - \mathbf{J_x g}_{\mathcal{A}^0}\mathbf{X}, \mathbf{U}_{\mathcal{A}^0,\bullet}\right) &= \mathbf{0}_{|\mathcal{A}^0|\times k},\end{aligned} \tag{21}$$

where the arguments of the Hessians associated with L are $(\mathbf{p}^0, \mathbf{x}^0, \boldsymbol{\mu}^0, \boldsymbol{\lambda}^0)$ and Jacobians associated with $\mathbf{g}$ and $\mathbf{h}$ are $(\mathbf{p}^0, \mathbf{x}^0)$, and $\mathbf{U}_{\mathcal{J},\bullet}$ denotes the rows of $\mathbf{U}$ indexed by $\mathcal{J} \subset \mathcal{G}$.

Theorem 1.1 can be connected to Equation (21) as follows: the KKT system associated with the quadratic program $QP_{(j)}$ in (10) is given by

$$\begin{aligned}
&\begin{bmatrix}\nabla^2_{\mathbf{xp}}L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0,\boldsymbol{\lambda}^0) & \nabla^2_{\mathbf{xx}}L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0,\boldsymbol{\lambda}^0)\end{bmatrix}\begin{bmatrix}\mathbf{d}\\ \mathbf{z}\end{bmatrix}\\
&+\begin{bmatrix}(\mathbf{J_x g}_{\mathcal{A}^0_{(j-1)}}(\mathbf{p}^0,\mathbf{x}^0))^{\mathrm{T}} & (\mathbf{J_x g}_{\mathcal{A}^+_{(j-1)}}(\mathbf{p}^0,\mathbf{x}^0))^{\mathrm{T}} & (\mathbf{J_x h}(\mathbf{p}^0,\mathbf{x}^0))^{\mathrm{T}}\end{bmatrix}\begin{bmatrix}\boldsymbol{\eta}\\ \boldsymbol{\gamma}\\ \boldsymbol{\rho}\end{bmatrix} = \mathbf{0}_{n_x},\\
&\mathbf{J_x g}_{\mathcal{A}^+_{(j-1)}}(\mathbf{p}^0,\mathbf{x}^0)\mathbf{z} + \mathbf{J_p g}_{\mathcal{A}^+_{(j-1)}}(\mathbf{p}^0,\mathbf{x}^0)\mathbf{d} = \mathbf{0}_{|\mathcal{A}^+_{(j-1)}|},\\
&\mathbf{J_x h}(\mathbf{p}^0,\mathbf{x}^0)\mathbf{z} + \mathbf{J_p h}(\mathbf{p}^0,\mathbf{x}^0)\mathbf{d} = \mathbf{0}_{n_h},\\
&\mathbf{min}(-\mathbf{J_x g}_{\mathcal{A}^0_{(j-1)}}(\mathbf{p}^0,\mathbf{x}^0)\mathbf{z} - \mathbf{J_p g}_{\mathcal{A}^0_{(j-1)}}(\mathbf{p}^0,\mathbf{x}^0)\mathbf{d}, \boldsymbol{\eta}) = \mathbf{0}_{|\mathcal{A}^0_{(j-1)}|},
\end{aligned} \tag{22}$$

where **min** is the componentwise minimum function. The nonsmooth and nonlinear equation system (22) is identical to solving the j^{th} column of the nonsmooth and nonlinear sensitivity system (21) by noting that $\mathbf{U}_{\mathcal{A}^-,\bullet} = \mathbf{0}_{|\mathcal{A}^-|\times k}$ in (21) is not enforced in $QP_{(j)}$ but is instead enforced after the fact in (12). The optimality conditions associated with the sequence of quadratic programs (10) are the columnwise nonsmooth and nonlinear equation systems in (21).

Remark 3.5. Given a reference parameter value $\mathbf{p}^0 \in D_p$ and denoting the set of all multipliers satisfying the KKT conditions at $(\mathbf{p}^0,\mathbf{x}^0) \in D_p \times D_x$ by

$$M(\mathbf{p}^0,\mathbf{x}^0) := \{(\boldsymbol{\mu},\boldsymbol{\lambda}) \in \mathbb{R}^{n_g+n_h} : (\mathbf{x}^0,\boldsymbol{\mu},\boldsymbol{\lambda}) \text{ is a KKT point of (1)}\},$$

suppose that MFCQ holds at $(\mathbf{p}^0,\mathbf{x}^0)$; the vectors in the set $\{(\mathbf{J_x}h_i(\mathbf{p}^0,\mathbf{x}^0))^{\mathrm{T}} : i \in \mathcal{H}\}$ are linearly independent and there exists $\mathbf{v} \in \mathbb{R}^{n_x}$ such that

$$\begin{aligned}
\mathbf{J_x}g_i(\mathbf{p}^0,\mathbf{x}^0)\mathbf{v} &< 0, \quad \forall i \in \mathcal{A},\\
\mathbf{J_x}h_i(\mathbf{p}^0,\mathbf{x}^0)\mathbf{v} &= 0, \quad \forall i \in \mathcal{H}.
\end{aligned}$$

(If MFCQ holds at $(\mathbf{p}^0,\mathbf{x}^0)$ then $M(\mathbf{p}^0,\mathbf{x}^0)$ is a closed convex polytope [37].) Suppose that CRCQ holds at $(\mathbf{p}^0,\mathbf{x}^0)$ (see [2, 8] for details). (Note that if LICQ holds at $(\mathbf{p}^0,\mathbf{x}^0)$ then $M(\mathbf{p}^0,\mathbf{x}^0)$ is a singleton and MFCQ and CRCQ hold at $(\mathbf{p}^0,\mathbf{x}^0)$.) Lastly, suppose that GSSOSC holds at $(\mathbf{p}^0,\mathbf{x}^0)$; SSOSC holds at $(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu},\boldsymbol{\lambda})$ for all multipliers $(\boldsymbol{\mu},\boldsymbol{\lambda}) \in M(\mathbf{p}^0,\mathbf{x}^0)$. Then directional derivatives of the primal variable solution of (1) are obtained as follows [8]: for any $\mathbf{d} \in \mathbb{R}^{n_p}$, $\widetilde{\mathbf{x}}'(\mathbf{p}^0;\mathbf{d})$ is the unique solution $\mathbf{z}_{(1)}(\mathbf{d})$ of $QP_{(1)}(\boldsymbol{\mu},\boldsymbol{\lambda};\mathbf{d})$ if $(\boldsymbol{\mu},\boldsymbol{\lambda})$ is an element of the solution set of $LP_{(1)}(\mathbf{d})$, where $QP_{(1)}(\boldsymbol{\mu},\boldsymbol{\lambda};\mathbf{d})$ denotes the following quadratic program:

$$\begin{aligned}
QP_{(1)}(\boldsymbol{\mu},\boldsymbol{\lambda};\mathbf{d}): \quad \min_{\mathbf{z}\in\mathbb{R}^{n_x}} \quad & 0.5\mathbf{z}^{\mathrm{T}}\nabla^2_{\mathbf{xx}}L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu},\boldsymbol{\lambda})\mathbf{z} + \mathbf{z}^{\mathrm{T}}\nabla^2_{\mathbf{xp}}L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu},\boldsymbol{\lambda})\mathbf{d},\\
\text{s.t.} \quad & \mathbf{J_x}g_i(\mathbf{p}^0,\mathbf{x}^0)\mathbf{z} + \mathbf{J_p}g_i(\mathbf{p}^0,\mathbf{x}^0)\mathbf{d} \le 0, \quad \forall i \in \mathcal{A}^0,\\
& \mathbf{J_x}g_i(\mathbf{p}^0,\mathbf{x}^0)\mathbf{z} + \mathbf{J_p}g_i(\mathbf{p}^0,\mathbf{x}^0)\mathbf{d} = 0, \quad \forall i \in \mathcal{A}^+,\\
& \mathbf{J_x}h_i(\mathbf{p}^0,\mathbf{x}^0)\mathbf{z} + \mathbf{J_p}h_i(\mathbf{p}^0,\mathbf{x}^0)\mathbf{d} = 0, \quad \forall i \in \mathcal{H},
\end{aligned} \tag{23}$$

and where $LP_{(1)}(\mathbf{d})$ denotes the following linear program:

$$\begin{aligned}
LP_{(1)}(\mathbf{d}): \quad \max_{(\boldsymbol{\mu},\boldsymbol{\lambda})\in\mathbb{R}^{n_g}\times\mathbb{R}^{n_h}} \quad & \boldsymbol{\mu}^{\mathrm{T}}\mathbf{J_p g}(\mathbf{p}^0,\mathbf{x}^0)\mathbf{d} + \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{J_p h}(\mathbf{p}^0,\mathbf{x}^0)\mathbf{d},\\
\text{s.t.} \quad & (\boldsymbol{\mu},\boldsymbol{\lambda}) \in M(\mathbf{p}^0,\mathbf{x}^0).
\end{aligned} \tag{24}$$

In the same vein as the approach in Theorem 3.1 (which applies Theorem 1.1 iteratively), it seems fruitful to apply the results of [8], as outlined above, iteratively to obtain generalized derivative information in the setting of non-unique multipliers. That is, furnishing a sequence of quadratic programs $QP_{(j)}(\boldsymbol{\mu},\boldsymbol{\lambda};\mathbf{d})$ and a sequence of embedded linear programs $LP_{(j)}(\mathbf{d})$. However, the parametric quadratic program $QP_{(j)}$ need not inherit MFCQ at its solution from $QP_{(j-1)}$. For example, consider the NLP

$$
\begin{aligned}
\min_{\mathbf{x}\in\mathbb{R}^2} \quad & 0.5(x_1-p_1)^2+0.5(x_2-p_2)^2+x_1+x_2,\\
\text{s.t.} \quad & -x_1 \le 0,\\
& -x_2 \le 0,\\
& -x_1-x_2+p_1+p_2 \le 0,
\end{aligned}
\tag{25}
$$

which is similar to the one studied in [4]. Let $\mathbf{p}^0 = (0,0)$ and $\mathbf{x}^0 = (0,0)$. Then $\mathcal{A} = \{1,2,3\}$ and $M(\mathbf{p}^0,\mathbf{x}^0) = \{(u,u,1-u) : 0 \le u \le 1\}$ and, for any such multiplier, $(\mathbf{x}^0,\boldsymbol{\mu})$ is a KKT point satisfying MFCQ, GSSOSC and CRCQ.

Given $\mathbf{d} \in \mathbb{R}^2$, the linear program in (24) takes the form

$$
\begin{aligned}
\max_{\boldsymbol{\mu}\in\mathbb{R}^3} \quad & \mu_3(d_1+d_2),\\
\text{s.t.} \quad & \boldsymbol{\mu} \in \{(u,u,1-u) : u \in [0,1]\}.
\end{aligned}
$$

from which the following observations can be made:

1. If $d_1+d_2>0$, then the solution set of $LP_{(1)}(\mathbf{d})$ is $\{(0,0,1)\}$;
2. If $d_1+d_2<0$, then the solution set of $LP_{(1)}(\mathbf{d})$ is $\{(1,1,0)\}$;
3. If $d_1+d_2=0$, then the solution set of $LP_{(1)}(\mathbf{d})$ is $M(\mathbf{p}^0,\mathbf{x}^0)$.

In attempting to construct the successors in the optimization problem heirarchy (i.e., $LP_{(2)}$ and $QP_{(2)}$) in the same manner as outlined above in the LICQ setting, suppose that $(1,1,0) \in M(\mathbf{p}^0,\mathbf{x}^0)$ and $\mathbf{d} = (1,-1)$ are chosen. This results in the quadratic program

$$
\begin{aligned}
QP_{(1)}(1,1,0;(1,-1)) : \quad \min_{\mathbf{z}\in\mathbb{R}^2} \quad & 0.5(z_1^2+z_2^2)-z_1+z_2,\\
\text{s.t.} \quad & -z_1 = 0,\\
& -z_2 = 0,\\
& -z_1-z_2 \le 0,
\end{aligned}
$$

whose primal variable solution is $\mathbf{z}_{(1)} = (0,0)$ and dual variable solution set is $\{(-1-s,1-s,s) : s \ge 0\}$ (i.e., an unbounded set). Consequently, for any $\mathbf{d}$ such that $d_1+d_2>0$, $LP_{(2)}(\mathbf{d})$ has no solution. A repeated application of Ralph and Dempe's result is not possible since the regularity assumptions of the original NLP (1) are not inherited by the auxiliary quadratic program. In particular, MFCQ is not inherited by $QP_{(1)}$ in this example.

4. Conclusions

A new theory is provided in this article for computing generalized derivatives of parametric NLPs. The results in this article require LICQ to hold at a KKT point of interest, which is more restrictive than the results of Ralph and Dempe [8] (where MFCQ is assumed), but allows for computation of generalized derivative elements of parametric NLPs with active index set changes. Moreover, as detailed in [24], it is straightforward to use the theory in this article to evaluate a generalized gradient element of the parametric NLP (1) objective-value function.

As discussed in Section 1, the computational costs associated with the competing theory (i.e., solving the nonsmooth equation system (21)) are currently unclear, and there may be cases where the current approach is more tractable. If LICQ and SSOSC hold in $QP_{(1)}(\mathbf{d}^0)$ in (2), given some $\mathbf{d}^0 \in \mathbb{R}^{n_p}$, then

$$
\mathbf{z}^{\mathrm{T}}\nabla^2_{\mathbf{xx}}L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0,\boldsymbol{\lambda}^0)\mathbf{z} \ge 0,
$$

for all $\mathbf{z}$ in its feasible set [8]. This property therefore holds when solving $QP_{(j)}(\mathbf{p}_{(j)})$ in Line 3 of Algorithm 1. This does not necessarily match the typical definition of a convex QP (i.e., $\nabla^2_{\mathbf{xx}}L(\mathbf{p}^0,\mathbf{x}^0,\boldsymbol{\mu}^0,\boldsymbol{\lambda}^0)$ being positive semidefinite) [35, 36, 38], for which polynomial-time algorithms exist (e.g., interior-point algorithms [39]). Moreover, since the sequence of QPs in the heirarchy are closely related to each other (the objective-value functions are identical and the feasible set only differs in inequality constraints being removed or becoming equalities), numerical methods which warm-start the QPs may be possible for improving computational time.

As mentioned in Section 1, the authors are hopeful that the theory presented here finds extension to parametric mathematical programs which currently admit a sensitivity theory in the form of directional derivatives (e.g., variational inequalities); if, under certain regularity assumptions, directional derivatives can be computed from an auxiliary mathematical program, which itself inherits appropriate regularity assumptions, LD-derivatives can be furnished by an approach similar to the theory in this article. Such extensions would provide methods for use in practical dynamic optimization problems, such as optimizing the startup of a binary batch distillation problem, involving mixed complementarity systems [40] and classes of hybrid dynamic models using mathematical programs with equilibrium constraints formulations [41].

Lastly, an extension of the theory here to Ralph and Dempe's [8] results in the MFCQ and CRCQ setting is a current limitation to this approach and direction for future work; the NLP (25) is a counterexample that shows the convex quadratic program need not inherit the MFCQ assumption. Possible remedies include choosing a directions matrix to avoid this issue or a multiparametric programming approach. Advancing the theory to other types of mathematical programs under analogous regularity conditions (i.e., those implying non-unique multipliers) is also desirable.

Acknowledgments

This material was based in part on work supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[1] A. V. Fiacco, G. P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, Wiley, New York, 1968.

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

[3] A. Shapiro, Sensitivity analysis of nonlinear programs and differentiability properties of metric projections, SIAM Journal on Control and Optimization 26 (3) (1988) 628–645.

[4] J. Kyparisis, Sensitivity analysis for nonlinear programs and variational inequalities with nonunique multipliers, Mathematics of Operations Research 15 (2) (1990) 286–298.

[5] S. Dempe, Directional differentiability of optimal solutions under Slater's condition, Mathematical Programming 59 (1) (1993) 49–69.

[6] J. F. Bonnans, A. Shapiro, Perturbation Analysis of Optimization Problems, Springer-Verlag, New York, 2000.

[7] F. Facchinei, J.-S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems, Springer, New York, 2003.

[8] D. Ralph, S. Dempe, Directional derivatives of the solution of a parametric nonlinear program, Mathematical programming 70 (1-3) (1995) 159–172.

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

[10] I. J. Wolf, W. Marquardt, Fast NMPC schemes for regulatory and economic NMPC–A review, Journal of Process Control 44 (2016) 162–183.

[11] K. A. Khan, P. I. Barton, A vector forward mode of automatic differentiation for generalized derivative evaluation, Optimization Methods and Software 30 (6) (2015) 1185–1212.

[12] Y. Nesterov, Lexicographic differentiation of nonsmooth functions, Mathematical Programming 104 (2005) 669–700.

[13] F. H. Clarke, Optimization and Nonsmooth Analysis, SIAM, Philadelphia, 1990.

[14] B. S. Mordukhovich, Generalized differential calculus for nonsmooth and set-valued mappings, Journal of Mathematical Analysis and Applications 183 (1) (1994) 250–288.

[15] M. Kojima, S. Shindoh, Extensions of Newton and quasi-Newton methods to systems of PC^1 equations, Journal of Operations Research Society of Japan 29 (1986) 352–374.

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

[17] M. Ulbrich, Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces, MOS-SIAM Series on Optimizaiton, SIAM, Philadelphia, 2011.

[18] F. Facchinei, A. Fischer, M. Herrich, An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions, Mathematical Programming 146 (2014) 1–36.

[19] K. C. Kiwiel, Methods of Descent for Nondifferentiable Optimization. Lecture Notes in Mathematics, Springer, Berlin, 1985.

[20] C. Lemaréchal, J. J. Strodiot, A. Bihain, On a bundle algorithm for nonsmooth optimization, in: O. L. Mangasarian, R. R. Meyer, S. M. Robinson (Eds.), Nonlinear Programming 4, Academic Press, New York, 1981.

[21] L. Lukšan, J. Vlček, A bundle-Newton method for nonsmooth unconstrained minimization, Mathematical Programming 83 (1998) 373–391.

[22] L. Lukšan, J. Vlček, Algorithm 811: NDA: Algorithms for nondifferentiable optimization, ACM Transactions on Mathematical Software 27 (2) (2001) 193–213.

[23] J.-B. Hiriart-Urruty, C. Lemaréchal, Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods, A Series of Comprehensive Studies in Mathematics, Springer, Berlin, 1996.

[24] P. Stechlinski, K. A. Khan, P. I. Barton, Generalized sensitivity analysis of nonlinear programs, SIAM Journal on Optimization 28 (1) (2018) 272–301.

[25] D. Q. Mayne, J. B. Rawlings, C. V. Rao, P. O. Scokaert, Constrained model predictive control: Stability and optimality, Automatica 36 (6) (2000) 789–814.

[26] Y. Wang, S. Boyd, Fast model predictive control using online optimization, IEEE Transactions on Control Systems Technology 18 (2) (2010) 267–278.

[27] J. H. Lee, Model predictive control: Review of the three decades of development, International Journal of Control, Automation and Systems 9 (3) (2011) 415–424.

[28] W. Chen, Z. Shao, L. T. Biegler, A bilevel NLP sensitivity-based decomposition for dynamic optimization with moving finite elements, AIChE Journal 60 (3) (2014) 966–979.

[29] W. Chen, L. T. Biegler, Nested direct transcription optimization for singular optimal control problems, AIChE Journal 62 (10) (2016) 3611–3627.

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

[31] P. I. Barton, K. A. Khan, P. Stechlinski, H. A. J. Watson, Computationally relevant generalized derivatives: theory, evaluation and applications, Optimization Methods and Software, In Press.

[32] J.-S. Pang, D. Ralph, Piecewise smoothness, local invertibility, and parametric analysis of normal maps, Mathematics of Operations Research 21 (2) (1996) 401–426.

[33] K. A. Khan, P. I. Barton, Generalized derivatives for solutions of parametric ordinary differential equations with non-differentiable right-hand sides, Journal of Optimization Theory and Applications 163 (2014) 355–386.

[34] A. Griewank, On stable piecewise linearization and generalized algorithmic differentiation, Optimization Methods and Software 28 (6) (2013) 1139–1178.

[35] S. Wright, J. Nocedal, Numerical Optimization, 2nd Edition, Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006.

[36] D. den Hertog, Interior Point Approach to Linear, Quadratic and Convex Programming, Kluwer Academic Publishers, Dordrecht, 1994.

[37] J. Gauvin, A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming, Mathematical Programming 12 (1) (1977) 136–138.

[38] S. Boyd, L. Vandenberghe, Convex optimization, Cambridge University Press, Cambridge, 2004.

[39] Y. Ye, E. Tse, An extension of Karmarkar's projective algorithm for convex quadratic programming, Mathematical Programming 44 (1-3) (1989) 157–179.

[40] A. U. Raghunathan, M. S. Diaz, L. T. Biegler, An MPEC formulation for dynamic optimization of distillation operations, Computers & chemical engineering 28 (10) (2004) 2037–2052.

[41] B. T. Baumrucker, L. T. Biegler, MPEC strategies for optimization of a class of hybrid dynamic systems, Journal of Process Control 19 (8) (2009) 1248–1256.