A Note on Sensitivity of Value Functions of Mathematical Programs with Complementarity Constraints

Third revision February, 2002
The final version of this paper is published in:
Mathematical Programmming 93 (2002), 265–279

Xinmin Hu*
Danny Ralph†

Abstract

Using standard nonlinear programming (NLP) theory, we establish formulas for first and second order directional derivatives for optimal value functions of parametric mathematical programs with complementarity constraints (MPCCs). The main point is that under a linear independence condition on the active constraint gradients, optimal value sensitivity of MPCCs is essentially the same as for NLPs, in spite of the combinatorial nature of the MPCC feasible set. Unlike NLP however, second order directional derivatives of the MPCC optimal value function show combinatorial structure.

1 Introduction

In this note, derived largely from the doctoral thesis of the first author [8], we will study the rate of change of the optimal value function and optimal solutions of the following parametric mathematical program with complementarity constraints (MPCC):

$$\begin{array}{ll} \underset{z}{\text{Minimize}} & f(z,p) \\ \text{subject to} & g(z,p) \leq 0, \quad h(z,p) = 0 \\ & G(z,p) \geq 0, \quad H(z,p) \geq 0 \\ & G(z,p)^T H(z,p) = 0 \end{array} \qquad (P(p))$$

where $f : \mathbb{R}^{n_1+n_2} \to \mathbb{R}, g : \mathbb{R}^{n_1+n_2} \to \mathbb{R}^{m_1}, h : \mathbb{R}^{n_1+n_2} \to \mathbb{R}^{m_2}, G : \mathbb{R}^{n_1+n_2} \to \mathbb{R}^m, H : \mathbb{R}^{n_1+n_2} \to \mathbb{R}^m$ are smooth and $p \in \mathbb{R}^{n_2}$ is the parameter vector. Problem $(P(p))$ is the parametric version of the problem studied in [5, 20, 21]. In fact, the literature on this subject often concerns unperturbed MPCCs in the form $z = (x, y) \in \mathbb{R}^{(n_1-m)} \times \mathbb{R}^m$ and $H(x, y) = y$ [12, 13, 14]. The motivation and mathematical techniques are identical, however.

The area of mathematical programs with complementarity constraints has received much attention within recent years (see, [1, 13, 16] and references therein). Unfortunately, the presence of the complementarity constraint $G(z,p)^T H(z,p) = 0$ in $(P(p))$ implies a lack of nonlinear programming

*Department of Mathematics and Statistics, The University of Melbourne
†Judge Institute of Management Studies, The University of Cambridge

(NLP) regularity of the constraints [3] (see also [26]) and leads to theoretical and numerical difficulties; see discussion in [13, Chapter 3] or [9, Section 5]. An alternative piecewise or disjunctive approach recognizes the nonconvex, combinatorial nature of the problem by explicitly decomposing the feasible region into a possibly huge number of "branches", each of which has the format of a standard, often regular, nonlinear programming feasible set. This means that checking MPCC optimality conditions is equivalent to checking NLP optimality conditions on many branches, in fact 2^c branches where c is the number of indices i for which $G_i(z,p) = 0 = H_i(z,p)$. However the combinatorial aspect of optimality conditions is completely relieved if a linear independence condition on active gradients, MPCC-LICQ, holds [14, 20]; see [25] for generalizations. Since the MPCC-LICQ is a generic property of MPCCs [22], this means first order sensitivity analysis of the optimal value function of an MPCC is generically noncombinatorial, that is independent of branches.

Our main point is that under the MPCC-LICQ, standard NLP sensitivity analysis yields sensitivity of the optimal value function of the MPCC as a natural consequence, despite the combinatorial properties of the MPCC feasible set. We summarize the results of Sections 2 and 3. Given $p \in \mathbb{R}^{m_2}$, the MPCC Lagrangean [13, 20] of $(P(p))$ is the function

$$\begin{aligned} &\mathbb{L}(z,p;\lambda,\mu,\xi,\eta) \\ &= \ f(z,p) + \lambda^T g(z,p) + \mu^T h(z,p) - \xi^T G(z,p) - \eta^T H(z,p) \end{aligned} \tag{1}$$

where $z \in \mathbb{R}^{n_1}$, and (λ,μ,ξ,η) in $\mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \mathbb{R}^{m} \times \mathbb{R}^{m}$ is the vector of MPCC multipliers. This is the usual Lagrangean for the *globally relaxed NLP* which is derived from $(P(p))$ by dropping the complementarity condition. Fix $\bar{p} \in \mathbb{R}^{m_2}$ and let $\bar{z}$ be a local minimizer of $(P(\bar{p}))$ at which the usual LICQ for the globally relaxed NLP holds. It is well known [14, 20] that there exists a unique MPCC multiplier vector $(\bar{\lambda},\bar{\mu},\bar{\xi},\bar{\eta})$ such that the partial gradient $\nabla_z \mathbb{L}(\bar{z},\bar{p};\bar{\lambda},\bar{\mu},\bar{\xi},\bar{\eta})$ is zero. We suppose a second-order sufficient condition or inf-compactness condition, that is related to but weaker than an associated condition for the globally relaxed NLP, also holds at $\bar{z}$. Then optimal value function, $\mathbb{W}(p)$, for $(P(p))$ restricted to a neighborhood of $\bar{z}$, is differentiable at $\bar{p}$ with

$$\nabla \mathbb{W}(\bar{p}) \ = \ \nabla_p \mathbb{L}(\bar{z},\bar{p};\bar{\lambda},\bar{\mu},\bar{\xi},\bar{\eta}).$$

That is, MPCC sensitivity is precisely classical sensitivity for the relaxed NLP. Likewise, the MPCC multipliers $\bar{\lambda}$, $\bar{\mu}$, $\bar{\xi}$, $\bar{\eta}$ are the usual shadow costs associated with changes to their respective constraint functions g, h, G, H.

Even under the MPCC-LICQ however, second order analysis of the optimal value function of MPCC cannot generally avoid a combinatorial curse. Recall that the second order directional differentiability of the optimal value function of a parametric NLP is linked with some kind of second order sufficient conditions which, in turn, imply some kind of first order sensitivity of optimal solutions [2, 23]. In Section 4 we will show that for parametric MPCCs, a similar story is true. However an important difference between optimal solution sensitivity for MPCC and NLP is that even a strong second order sufficient condition cannot guarantee local uniqueness of the MPCC solution as the parameter varies. The difficulty is that each NLP branch has its own locally unique solution depending on the same parameter; some of these solutions will be local or global MPCC solutions.

Lucet and Ye [12] address the sensitivity problem for the optimal value function of mathematical programs with variational inequality constraints, which include MPCCs, in a nonsmooth setting. They establish an upper estimate of the set of the generalized gradients of optimal value function using the Mordukhovich calculus [15]. The piecewise programming approach is used in [12] to compare various types of multipliers. In this note, we focus on the classical directional differentiability of optimal value functions of MPCCs, by using their straightforward and standard NLP counterparts under the MPCC-LICQ. Analysis of the MPCC optimal value function does not necessarily require the MPCC-LICQ if an appropriate generalized directional derivative is used, as in [12], but this constraint qualification is attractive from the point of view of clarity of hypothesis, results and analysis.

Scheel and Scholtes [20] also study sensitivity issues for parametric MPCCs. [20, Theorem 11] gives conditions, including an "upper level strict complementarity" condition and a strong second order sufficient condition, such that all relevant NLP branches share the same perturbed optimal solution as the perturbation parameter changes. That is, under these conditions, the perturbed MPCC solution is a locally unique function of the parameter. Unfortunately this pleasant outcome is not as natural for MPCC as it appears to be for NLP; see Section 4. Another result, [20, Theorem 12], shows the existence of a continuous selection of stationary points under a constant sign condition on the determinants of certain matrices relating to the Lagrangean on each branch. Example 12 in Section 4, however, shows that the global optimal solution set of a parametric MPCC need not have a continuous selection.

2 Strict differentiability of local optimal value function

We first define the following active index sets corresponding to a feasible point $\bar{z}$ of $(P(\bar{p}))$:

$$\begin{aligned} I_G(\bar{z},\bar{p}) &= \{i \in \{1,\ldots,m\} : G_i(\bar{z},\bar{p}) = 0\} \\ I_H(\bar{z},\bar{p}) &= \{i \in \{1,\ldots,m\} : H_i(\bar{z},\bar{p}) = 0\} \end{aligned}$$

We further define a nonempty family of index sets $J \subseteq \{1,\ldots,m\}$,

$$\mathcal{J}(\bar{z},\bar{p}) = \{J : J \subset I_G(\bar{z},\bar{p}),\ J^C \subseteq I_H(\bar{z},\bar{p})\},$$

where $J^C = \{1,\ldots,m\} \setminus J$. Next we have the following nonlinear program corresponding to $J \subset \{1,\ldots,m\}$:

$$\begin{array}{lllll} \underset{z}{\text{Minimize}} & f(z,p) & & & \\ \text{subject to} & g(z,p) \le 0, & h(z,p) = 0 & & \\ & G_i(z,p) = 0, & H_i(z,p) \ge 0, & i \in J & \\ & G_i(z,p) \ge 0, & H_i(z,p) = 0, & i \in J^C. & \end{array} \qquad (NLP_J(p))$$

Each of these nonlinear programs is called an NLP branch of $(P(p))$ and its feasible set is called a branch [9] of the MPCC. An important but trivial fact is that the the branches over $J \in \mathcal{J}(\bar{z},\bar{p})$ form a neighborhood of $\bar{z}$ in the feasible set of $(P(\bar{p}))$. As a result, first and second order conditions can be specified in terms of NLP branches, e.g. Definition 1 and 4 below. Unfortunately the cardinality of $\mathcal{J}(\bar{z},\bar{p})$ may be rather large, namely 2^c where c is the number of indices i such that $G_i(\bar{z},\bar{p}) = 0 = H_i(\bar{z},\bar{p})$.

Recall the MPCC Lagrangian function $\mathbb{L}(z,p;\lambda,\mu,\xi,\eta)$ defined in (1). Observe that it is independent of J, and coincides with the usual Lagrangian function for each $(NLP_J(p))$.

Definition 1 *A feasible point $\bar{z}$ of $(P(\bar{p}))$ is said to be a piecewise stationary [13] or B-stationary [20] point of $(P(\bar{p}))$ if $\bar{z}$ is a stationary point of $NLP_J(\bar{p})$ for each $J \in \mathcal{J}(\bar{z},\bar{p})$, that is, for each $J \in \mathcal{J}(\bar{z},\bar{p})$, there exist KKT multipliers $\bar{\lambda}^J, \bar{\mu}^J, \bar{\xi}^J$ and $\bar{\eta}^J$ such that*

$$\begin{aligned}
&\nabla_z \mathbb{L}(\bar{z},\bar{p};\bar{\lambda}^J,\bar{\mu}^J,\bar{\xi}^J,\bar{\eta}^J) = 0, \\
&(\bar{\lambda}^J)^T g(\bar{z},\bar{p}) = 0, \bar{\lambda}^J \geq 0 \\
&\bar{\xi}_i^J G_i(\bar{z},\bar{p}) = 0, \bar{\xi}_i^J \geq 0, \text{ for } i \in J^C \\
&\bar{\eta}_i^J H_i(\bar{z},\bar{p}) = 0, \bar{\eta}_i^J \geq 0, \text{ for } i \in J
\end{aligned} \tag{2}$$

The definition of B-stationarity above is actually slightly different from, but equivalent to, the original one given by Scheel and Scholtes [20].

Definition 2 *The following nonlinear program is called the global NLP relaxation of $(P(p))$:*

$$\begin{array}{ll}
\underset{z}{\text{Minimize}} & f(z,p) \\
\text{subject to} & g(z,p) \leq 0, h(z,p) = 0 \\
& G(z,p) \geq 0, H(z,p) \geq 0
\end{array} \tag{3}$$

and the following nonlinear program is called the (local) NLP relaxation [13] at $(\bar{z},\bar{p})$ of $(P(\bar{p}))$:

$$\begin{array}{lll}
\underset{z}{\text{Minimize}} & f(z,\bar{p}) & \\
\text{subject to} & g(z,\bar{p}) \leq 0,\ h(z,\bar{p}) = 0 & \\
& G_i(z,\bar{p}) = 0 & i:\ H_i(\bar{z},\bar{p}) > 0, \\
& H_i(z,\bar{p}) = 0 & i:\ G_i(\bar{z},\bar{p}) > 0, \\
& G_i(z,\bar{p}) \geq 0 \text{ and } H_i(z,\bar{p}) \geq 0 & \text{all other } i.
\end{array} \tag{4}$$

The following condition has been widely used in MPCC literature [5, 13, 14, 21]. It is just the usual linear independence constraint qualification at $\bar{z}$ for the global NLP relaxation of $(P(\bar{p}))$. Let $I_G(z,p)$ and $I_H(z,p)$ be as above and, similarly, $I_g(z,p)$ be the set of active indices of g at (z,p).

Definition 3 *The linear independence constraint qualification for MPCC $(P(\bar{p}))$ (MPCC-LICQ) is said to hold at a feasible point $\bar{z}$ of this problem if the following gradients,*

$$\left\{\nabla_z g_i(\bar{z},\bar{p})\right\}_{i\in I_g(\bar{z},\bar{p})}, \quad \left\{\nabla_z h_i(\bar{z},\bar{p})\right\}_{i=1}^{m_2}, \quad \left\{\nabla_z G_i(\bar{z},\bar{p})\right\}_{i\in I_G(\bar{z},\bar{p})}, \quad \left\{\nabla_z H_j(\bar{z},\bar{p})\right\}_{i\in I_H(\bar{z},\bar{p})}$$

are linearly independent.

MPCC-LICQ is a nontrivial assumption but one that holds in some sense generically [22]. Under MPCC-LICQ, local minima of $(P(\bar{p}))$ are B-stationary points [14, 20]. Moreover, if $\bar{z}$ is a B-stationary point at which MPCC-LICQ holds, then

i) for each $J \in \mathcal{J}(\bar{z},\bar{p})$, there exists a unique KKT multiplier vector $(\bar{\lambda}^J,\bar{\mu}^J,\bar{\xi}^J,\bar{\eta}^J)$ corresponding to $\bar{z}$ for $(NLP_J(\bar{p}))$; and

ii) since the Lagrangian gradient equation (2) specifies the KKT multiplier uniquely and is independent of J, the associated multiplier vector is also independent of J.

In short, there is a unique multiplier vector $(\bar\lambda, \bar\mu, \bar\xi, \bar\eta)$, called the MPCC multiplier vector, that satisfies (2), and this MPCC multiplier vector is the unique KKT multiplier at $\bar z$ for each $(NLP_J(\bar p))$.

Under the MPCC-LICQ one can easily deduce [14, 20] that the MPCC multiplier is also the unique KKT multiplier $\bar z$ of the local NLP relaxation (4). If the vectors $\bar\xi$ and $\bar\eta$, which correspond to complementarity functions, are nonnegative, then the MPCC multiplier is also the unique KKT multiplier at $\bar z$ of the global NLP relaxation (3).

We give a strong second order sufficient condition, denoted MPCC-SSOSC, for the problem $(P(\bar p))$ based on the branching structure of the feasible set about a given feasible point $\bar z$. It is closely related to the second-order sufficient conditions for mathematical programs with nonlinear complementarity constraints discussed in [13]. The standard SSOSC [18] for the global NLP relaxation at $\bar z$ is sufficient but not necessary for MPCC-SSOSC, though it is a useful concept in some MPCC algorithms [21].

Definition 4 *Let $\bar z$ be a B-stationary point of $(P(\bar p))$ with a unique MPCC multiplier vector $(\bar\lambda, \bar\mu, \bar\xi, \bar\eta)$. We say that the strong second order sufficient condition (MPCC-SSOSC) for $(P(\bar p))$ holds at $\bar z$ if*

$$d^T \nabla^2_{zz} \mathbb{L}(\bar z, \bar p; \bar\lambda, \bar\mu, \bar\xi, \bar\eta) d > 0$$

for each $J \in \mathcal{J}(\bar z, \bar p)$ and every nonzero d satisfying

$$\begin{array}{ll} \nabla_z g_i(\bar z, \bar p)^T d = 0, & i : \bar\lambda_i > 0 \\ \nabla_z h(\bar z, \bar p) d = 0, & \\ \nabla_z G_i(\bar z, \bar p)^T d = 0, & i : \textit{either } i \in J \textit{ or } (i \in J^C \textit{ and } \bar\xi_i > 0) \\ \nabla_z H_i(\bar z, \bar p)^T d = 0, & i : \textit{either } i \in J^c \textit{ or } (i \in J \textit{ and } \bar\eta_i > 0) \end{array} \tag{5}$$

Remark 5 The above second order condition is piecewise, i.e. posed with respect to active branches. A related condition is used in [20, Theorem 12], that is a constant sign condition on the determinants of certain matrices of the Lagrangean functions corresponding to $(NLP_J(\bar p))$, for some J.

A sufficient condition for MPCC-SSOSC that does not use the branching structure of the feasible set, is positive definiteness of the Hessian matrix $\nabla^2_{zz} \mathbb{L}(\bar z, \bar p; \bar\lambda, \bar\mu, \bar\xi, \bar\eta)$ on the directions d satisfying the first two conditions of (5) as well as

$$\begin{array}{ll} \nabla_z G_i(\bar z, \bar p)^T d = 0, & i : \text{either } H_i(\bar z, \bar p) > 0 \text{ or } \bar\xi_i > 0 \\ \nabla_z H_i(\bar z, \bar p)^T d = 0, & i : \text{either } G_i(\bar z, \bar p) > 0 \text{ or } \bar\eta_i > 0. \end{array}$$

This condition is implied by the conditions of [20, Theorem 11]. Moreover, it is precisely the standard SSOSC for the local NLP relaxation of $(P(\bar p))$ at $(\bar z, \bar p)$.

Our main result in this section is next. We need some notation for localized global optimal value functions. Fix $\bar z$ as a feasible point of $(P(\bar p))$ and let $\mathcal{N}(\bar z)$ denote a sufficiently small neighborhood of $\bar z$. For p near $\bar p$ and $J \in \mathcal{J}(\bar z, \bar p)$,

let $\mathbb{W}(p)$ be the global optimal value function of $(P(p))$ with the extra constraint $z \in \mathcal{N}(\bar z)$; and

let $W_J(p)$ be the global optimal value function of $(NLP_J(p))$ with the extra constraint $z \in \mathcal{N}(\bar z)$.

The role of $\mathcal{N}(\bar{z})$ is indirect; under the conditions of the next theorem, for p near $\bar{p}$, the values of $\mathbb{W}(p)$ and each $W_J(p)$ are attained in the interior of $\mathcal{N}(\bar{z})$ at local minimizers of $(P(p))$ and $(NLP_J(p))$ respectively.

Recall strict differentiability at $\bar{p}$ of a Lipschitz function $W : \mathbb{R}^n \to \mathbb{R}^m$ is equivalent [4] to the derivatives $\nabla W(p)$, where they exist, converging to $\nabla W(\bar{p})$ as $p \to \bar{p}$.

Theorem 6 *Let $\bar{z}$ be a B-stationary point of $(P(\bar{p}))$. If MPCC-LICQ and MPCC-SSOSC hold for $(P(\bar{p}))$ at $\bar{z}$ then*

1) *$\bar{z}$ is an isolated optimal solution of $(P(\bar{p}))$;*

2) *the optimal value function $\mathbb{W}(p)$ is piecewise smooth (hence Lipschitz) near $\bar{p}$; and*

3) *the optimal value function $\mathbb{W}(p)$ is strictly differentiable at $\bar{p}$, with gradient $\nabla\mathbb{W}(\bar{p}) = \nabla_p \mathbb{L}(\bar{z}, \bar{p}; \bar{\lambda}, \bar{\mu}, \bar{\xi}, \bar{\eta})$.*

Proof First of all, MPCC-LICQ and MPCC-SSOSC imply the usual LICQ and SSOSC at $\bar{z}$ for each $(NLP_J(\bar{p}))$, $J \in \mathcal{J}(\bar{z}, \bar{p})$. Standard NLP theory [10, 18] says $\bar{z}$ is an isolated optimal solution of each $(NLP_J(\bar{p}))$, hence of the MPCC since these branches form a neighborhood of $\bar{z}$ in the feasible set of $(P(\bar{p}))$. In fact, for each $J \in \mathcal{J}(\bar{z}, \bar{p})$, by the well known implicit function theorem for NLP [11, Theorem 7.2], there is a neighborhood $O^J(\bar{p})$ of $\bar{p}$ and a neighborhood $\mathcal{N}^J(\bar{z})$ of $\bar{z}$ such that the optimization problem $(NLP_J(p))$ with the additional constraint $z \in \mathcal{N}^J(\bar{z})$ added has a unique optimal solution $z^J(p)$ for $p \in O^J(\bar{p})$ and $z^J(\cdot)$ is continuous in $p \in O^J(\bar{p})$. Now, let $\mathcal{N}(\bar{z}) = \bigcap_{J \in \mathcal{J}(\bar{z}, \bar{p})} \mathcal{N}^J(\bar{z})$ in the definition of $\mathbb{W}(\cdot)$ and $W_J(\cdot)$.

Second, for $J \in \mathcal{J}(\bar{z}, \bar{p})$, the same NLP theory says $W_J(p)$ is C^1 near $\bar{p}$ and its derivative $\nabla W_J(\bar{p})$ equals the partial derivative with respect to p of its Lagrangean function at $(\bar{z}, \bar{p})$. Since the Lagrangian function and KKT multipliers at $\bar{z}$ for each $(NLP_J(\bar{p}))$ are independent of J, and coincide with the MPCC Lagrangean and MPCC multipliers at $\bar{z}$, respectively, we have $\nabla W_J(\bar{p}) = \nabla_p \mathbb{L}(\bar{z}, \bar{p}; \bar{\lambda}, \bar{\mu}, \bar{\xi}, \bar{\eta})$ for each such J.

Finally, for p near $\bar{p}$, $\mathbb{W}(p)$ is the minimum value of the C^1 functions $W_J(p)$ for $J \in \mathcal{J}(\bar{z}, \bar{p})$, hence it is piecewise smooth (PC^1). It is, furthermore, strictly differentiable at $\bar{p}$ since the gradients $\nabla W_J(\bar{p})$ are independent of J; this conclusion relies on standard arguments for min (or max) functions or PC^1 functions. ■

Part 1) has appeared in a number of forms, see [13, 14, 20], and Part 3) is related to Lucet and Ye's recent work [12].

Remark 7 The theorem implies that, under MPCC-LICQ and MPCC-SSOSC, the rate of change of the MPCC optimal value function with respect to the parameter coincides with the rate of change of the optimal value function of any one of $(NLP_J(\bar{p}))$ with $J \in \mathcal{J}(\bar{z}, \bar{p})$. That is, $\nabla\mathbb{W}(\bar{p}) = \nabla W_J(\bar{p})$ for each $J \in \mathcal{J}(\bar{z}, \bar{p})$.

Remark 8 The MPCC multipliers are, as usual in NLP, shadow prices corresponding to perturbations of the constraint functions (excluding the complementarity function $G(z,p)^T H(z,p)$). Consider a special case of $(P(p))$ with only right-hand side perturbations, that is

$$\begin{array}{ll} \underset{z}{\text{Minimize}} & f(z) \\ \text{subject to} & g(z) - p^g \leq 0, h(z) - p^h = 0 \\ & G(z) - p^G \geq 0, H(z) - p^H \geq 0 \\ & (G(z) - p^G)^T (H(z) - p^H) = 0 \end{array}$$

with the parameter vector $p = (p^g, p^h, p^G, p^H)$. Theorem 6 gives conditions under which the change rate of the optimal value function under any of the constraints is exactly the corresponding MPCC multiplier. Lucet and Ye [12, Corollary 3.5] obtained essentially the same result relating optimal value sensitivity to shadow prices, with a slightly different problem format, as a corollary of more general results in nonsmooth analysis.

3 Directional differentiability of global optimal value functions

Here we establish the directional differentiability of the global optimal value function of $(P(p))$ under MPCC-LICQ and the *inf-compactness* assumption. First, we introduce notation for global optimal value functions.

Let $\mathbb{V}(p)$ be the global (infimal) optimal value function of $(P(p))$ and $\mathbb{S}(p)$ denote the corresponding set of global optimal solutions (possible empty).

Let $V_J(p)$ be the global (infimal) optimal value function of $(NLP_J(p))$, and $S_J(p)$ be the set of global optimal solutions.

Inf-Compactness Assumption [23, Assumption 1, p.217]: There exist a number α and a compact set $S \subset \mathbb{R}^{n_1}$ such that $\alpha > \mathbb{V}(\bar{p})$ and the set

$$\{z : f(z,p) \leq \alpha \text{ and } z \text{ is feasible for } (P(p))\} \subset S \tag{6}$$

for all p in a neighborhood of $\bar{p}$.

It is easy to see that MPCC-LICQ at a feasible point z^0 of $(P(\bar{p}))$ allows, following the standard implicit function theorem (see [7] for the NLP case), construction of feasible solutions of $(P(p))$ and $(NLP_J(p))$ for $J \in \mathcal{J}(z^0, \bar{p})$ near z^0 for p in some neighborhood of $\bar{p}$. Moreover, it is easy to see that the Inf-Compactness Assumption implies the inf-compactness of $(NLP_J(p))$ for each $J \in \bar{\mathcal{J}} := \{J : J \in \mathcal{J}(\bar{z}, \bar{p}), \bar{z} \in \mathbb{S}(\bar{p})\}$ since for the same α, S, and the same neighborhood of $\bar{p}$ in the Inf-Compactness Assumption the set $\{z : f(z,p) \leq \alpha, z \text{ is feasible for } (NLP_J(p))\}$ is a subset of the level set in (6) and $\mathbb{V}(\bar{p}) = V_J(\bar{p})$ for $J \in \bar{\mathcal{J}}$.

We will apply the standard sensitivity results [6, 7, 23] for the optimal value function of a NLP to that of $(P(p))$ under the inf-compactness, instead of the stronger *uniform compactness* [6] or *inf-boundedness* [19], together with the MPCC-LICQ. The result for a NLP under the inf-compactness and the NLP-LICQ can be established by repeating the processes in [7, 23, 6], but we follow another way here to apply the explicitly-stated results which hold under the uniform compactness assumption in [6]. First we state a result for a parametric NLP. Its proof is straightforward and is given in Appendix.

Consider the parametric NLP with smooth functions f, g, h:

$$(NLP(p)) \qquad \begin{array}{ll} \underset{z}{\text{Minimize}} & f(z,p) \\ \text{subject to} & g(z,p) \leq 0, h(z,p) = 0. \end{array}$$

Let $V(p)$ and $S(p)$ denote the global optimal value function and the set of global optimal solutions of $(NLP(p))$, respectively and let $L(z, p; \lambda, \mu) = f(z,p) + \lambda^T g(z,p) + \mu^T h(z,p)$ be the Lagrangean of $(NLP(p))$ at (z, p).

Proposition 9 Suppose the Inf-Compactness Assumption holds with $V(p)$ replacing $\mathbb{V}(p)$ and $(NLP(p))$ replacing $(P(p))$. Suppose further that the usual LICQ holds for $(NLP(\bar{p}))$ at each $z \in S(\bar{p})$. Then the optimal value function $V(p)$ is Lipschitz and directionally differentialbe near $\bar{p}$, and its directional derivative at $\bar{p}$ in the direction q is given by

$$V'(\bar{p}; q) = \min_{\bar{z} \in S(\bar{p})} \{\nabla_p L(\bar{z}, \bar{p}; \bar{\lambda}, \bar{\mu})^T q\}$$

where $(\bar{\lambda}, \bar{\mu})$ denotes the Lagrangian multiplier vector of $(NLP(\bar{p}))$ at the optimal solution $\bar{z} \in S(\bar{p})$.

Under MPCC-LICQ, the first order sensitivity result for the optimal value function of $(P(p))$ is exactly the same as that in standard nonlinear programming theory:

Theorem 10 *Under the Inf-Compactness Assumption, if MPCC-LICQ holds at each $\bar{z} \in \mathbb{S}(\bar{p})$ then the optimal value function $\mathbb{V}(\cdot)$ is Lipschitz and directionally differentiable near $\bar{p}$. Moreover, the directional derivative of $\mathbb{V}(\cdot)$ at $\bar{p}$ in any direction q is given by*

$$\mathbb{V}'(\bar{p}; q) = \min_{\bar{z} \in \mathbb{S}(\bar{p})} \{\left(\nabla_p \mathbb{L}(\bar{z}, \bar{p}; \bar{\lambda}, \bar{\mu}, \bar{\xi}, \bar{\eta})\right)^T q\} \tag{7}$$

where $(\bar{\lambda}, \bar{\mu}, \bar{\xi}, \bar{\eta})$ denotes the vector of MPCC multipliers that is uniquely defined by each $\bar{z} \in \mathbb{S}(\bar{p})$. In particular, when $\mathbb{S}(\bar{p})$ is a singleton, the optimal value function $\mathbb{V}(\cdot)$ is strictly differentiable at $\bar{p}$.

Proof Without any assumptions we know $\mathbb{V}(p)$ is the minimum of $V_J(p)$ over all branches $J \subset \{1, \ldots, m\}$, where V_J takes the value $+\infty$ or $-\infty$ if $(NLP_J(p))$ is infeasible or unbounded, respectively. First, we claim for p near $\bar{p}$ that

$$\mathbb{V}(p) = \min_{J \in \bar{\mathcal{J}}} V_J(p). \tag{8}$$

Nonemptiness of $\bar{\mathcal{J}}$ is due to nonemptiness of $\mathbb{S}(\bar{p})$, which is a consequence of the Inf-Compactness Assumption. Of course $\mathbb{V}(p) \leq \min_{J \in \bar{\mathcal{J}}} V_J(p)$. The reverse inequality follows from finite cardinality of the set of all branch indices J and the Inf-Compactness Assumption, for if $p^k \to \bar{p}$ then there is a branch index J such that $\mathbb{V}(p^k) = V_J(p^k)$ for infinitely many k, and, using inf compactness in the limit, we get $\mathbb{V}(\bar{p}) = V_J(\bar{p})$ and $\mathbb{S}(\bar{p}) \cap S_J(\bar{p}) \neq \emptyset$. To avoid a contradiction, (8) follows.

Second, for any $J \in \bar{\mathcal{J}}$, $\mathbb{S}(\bar{p}) \cap S_J(\bar{p})$ is nonempty and thus $\emptyset \neq S_J(\bar{p}) \subset \mathbb{S}(\bar{p})$. The MPCC-LICQ assumption therefore gives the LICQ for $(NLP_J(\bar{p}))$ at each member of $S_J(\bar{p})$, hence also nonemptiness of the NLP feasible set for p near $\bar{p}$. The condition i) in Proposition 9 for $(NLP_J(p))$ follows from the Inf-Compactness Assumption. By Proposition 9, the optimal value function $V_J(p)$ is Lipschitz near $\bar{p}$ and directionally differentiable near $\bar{p}$ with

$$V_J'(\bar{p}; q) = \min_{z^* \in S_J(\bar{p})} \nabla_p \mathbb{L}(z^*, \bar{p}; \lambda^*, \mu^*, \xi^*, \eta^*)^T q \tag{9}$$

where, from the MPCC-LICQ, $(\lambda^*, \mu^*, \xi^*, \eta^*)$ is the unique KKT multiplier corresponding to z^* for $(NLP_J(\bar{p}))$ and we are using the MPCC Lagrangean since it coincides with the Lagrangean for the NLP branch.

Finally, we apply the standard formula for directional derivatives of min functions to (8), as described in the Appendix, see (12); details (for max functions) appear in [17, Section 5.4]. We have

$$\mathbb{V}'(\bar{p}; q) = \min_{J \in \bar{\mathcal{J}}(\bar{p})} V_J'(\bar{p}; q)$$

where $\bar{\mathcal{J}}(\bar p) = \{J \in \mathcal{J} : \mathbb{V}_J(\bar p) = \mathbb{V}(\bar p)\}$. Using (9) gives

$$\mathbb{V}'(\bar p; q) = \min_{J \in \bar{\mathcal{J}}(\bar p)} \min_{z^* \in S_J(\bar p)} \{ \left(\nabla_p \mathbb{L}(z^*, \bar p; \lambda^*, \mu^*, \xi^*, \eta^*)\right)^T q \}.$$

It can be checked that the set $\{z^* : z^* \in S_J(\bar p), J \in \bar{\mathcal{J}}(\bar p)\}$ is exactly $\mathbb{S}(\bar p)$ and (7) follows. ■

Lucet and Ye [12] establish some inclusion relations between the set of the generalized gradients of the optimal value function of an MPCC in a slightly different format to $(P(p))$. Their result [12, Theorem 4.8] is related to Theorem 10, but describes the generalized gradient of the optimal value function by using a generalized equation approach involving nonsmooth functions, where the Mordukhovich calculus is applied to obtain CD (coderivative) multipliers. It may be worth noting that the set of generalized gradients does not necessarily describe the directional derivative or vice versa.

Corollary 11 is a local version of Theorem 10. It provides most of the conclusions of Theorem 10, except piecewise smoothness of $\mathbb{W}(\cdot)$, while weakening the MPCC-SSOSC to the requirement that $\bar z$ is an isolated local minimizer of $(P(\bar p))$. The latter requirement implies that the Inf-Compactness Assumption will hold if feasibility is restricted to a small enough neighborhood of $\bar z$.

Corollary 11 *Let $\bar z$ be an isolated local minimum of $(P(\bar p))$ at which the MPCC-LICQ holds. Then the optimal value function $\mathbb{W}(p)$ is Lipschitz near $\bar p$ and strictly differentiable at $\bar p$ with $\nabla \mathbb{W}(\bar p) = \nabla_p \mathbb{L}(\bar z, \bar p; \bar\lambda, \bar\mu, \bar\xi, \bar\eta)$.*

4 Second order sensitivity analysis and some elementary observations

Unlike in the NLP case, the following example shows that the MPCC-LICQ plus MPCC-SSOSC do not ensure local uniqueness of the optimal solution, or existence of a continuous selection from the set of global optimal solutions, or smoothness of the optimal value function near $\bar p$. Nevertheless it is easy to see that the set of all local optimal solutions of $(P(p))$ near $\bar z$ is a continuous set-mapping at $\bar p$ under MPCC-LICQ and MPCC-SSOSC.

Example 12 Consider

$$\begin{array}{ll} \underset{z}{\text{Minimize}} & (x-p)^2 + (y - q(p))^2 \\ \text{subject to} & x \ge 0, y \ge 0 \\ & xy = 0 \end{array}$$

where $q : \mathbb{R}^1 \to \mathbb{R}^1$ is the twice continuously differentiable function given by

$$q(p) = \begin{cases} p + p^5 \sin(\frac{\pi}{p}), & \text{if } p \neq 0; \\ 0, & \text{if } p = 0. \end{cases}$$

So, the set $\mathbb{S}(p)$ of global optimal solutions for $p > 0$ is

$$\mathbb{S}(p) = \begin{cases} \{(0, q(p))\}, & \text{if } p \in \cup_{k\in\mathbb{N}}(\frac{1}{2k+1}, \frac{1}{2k}); \\ \{(p, 0)\}, & \text{if } p \in \cup_{k\in\mathbb{N}}(\frac{1}{2k}, \frac{1}{2k-1}); \\ \{(p,0), (0, q(p))\}, & \text{if } p \in \{\frac{1}{k} : k \in \mathbb{N}\}; \\ \{(0,0)\}, & \text{otherwise.} \end{cases}$$

where $\mathbb{N}$ is the set of positive integers. Observe that every selection function of $\mathbb{S}(p)$ is discontinuous at $p = 1/k, k \in \mathbb{N}$, hence $\mathbb{S}(p)$ has no continuous selection function in a neighborhood of $p = 0$.

The optimal value function of the problem for $p > 0$ is

$$\mathbb{V}(p) = \begin{cases} p^2, & \text{for } p \in \cup_{k\in\mathbb{N}}(\frac{1}{2k+1}, \frac{1}{2k}); \\ q(p)^2, & \text{for } p \in \cup_{k\in\mathbb{N}}(\frac{1}{2k}, \frac{1}{2k-1}); \\ p^2 + q(p)^2, & \text{otherwise} . \end{cases}$$

which is not differentiable at $p = \frac{1}{2k}$ or $\frac{1}{2k-1}$ for all $k \in \mathbb{N}$.

Now we establish that second order sensitivity of the optimal value function of $(P(p))$, though similar to that in NLP case, is no longer independent of the branches.

We use the notation of [23] for second order directional derivatives: if, for $\varphi : \mathbb{R}^{n_2} \to \mathbb{R}$, the following limit

$$\lim_{t\to 0^+} \frac{\varphi(p(t)) - \varphi(\bar{p}) - t\varphi'(\bar{p}; d)}{t^2}$$

exists for all $d, w \in \mathbb{R}^{n_2}$ and any $p(t) = \bar{p} + td + t^2 w + o(t^2)$, then we say $\varphi(\cdot)$ is second order directionally differentiable at $\bar{p}$; the limit is denoted by $\varphi''(\bar{p}; d, w)$.

The next example is taken from [20], which also shows the lack of local uniqueness of parametric solutions.

Example 13 [20] Let $z = (x, y) \in \mathbb{R}^2$ and consider the following simple program:

$$\begin{array}{ll} \underset{z}{\text{Minimize}} & f(z,p) = (x - p_1)^2 + (y - p_2)^2 \\ \text{subject to} & G_1(z,p) = x \geq 0, H_1(z,p) = y \geq 0 \\ & G_1(z,p)H_1(z,p) = xy = 0 \end{array}$$

with parameter vector $p = (p_1, p_2)$ at $\bar{p} = (\bar{p_1}, \bar{p_2}) = (0, 0)$, where $\bar{z} = (\bar{x}, \bar{y}) = (0, 0)$ is the unique local and global optimal solution of the problem at $\bar{p}$. Note that SSOSC mentioned in Remark 5 holds for this MPCC at $(\bar{z}, \bar{p})$.

The family of branches $\mathcal{J}(\bar{z}, \bar{p})$ consists of the index sets $J = \emptyset$ and $J = \{1\}$. It is easy to see that

$$V_\emptyset(p_1, p_2) = \begin{cases} p_2^2, & \text{if } p_1 \geq 0, \\ p_1^2 + p_2^2, & \text{otherwise} \end{cases} \quad \text{and} \quad V_{\{1\}}(p_1, p_2) = \begin{cases} p_1^2, & \text{if } p_2 \geq 0, \\ p_1^2 + p_2^2, & \text{otherwise} \end{cases}$$

Each of the derivatives of $V_\emptyset$ and $V_{\{1\}}$ is zero at $\bar{p}$.

The global optimal value function of the MPCC is

$$\mathbb{V}(p_1, p_2) = \begin{cases} \min\{p_1^2, p_2^2\}, & \text{if } p_1 \geq 0, p_2 \geq 0; \\ \big((p_1)_-\big)^2 + \big((p_2)_-\big)^2, & \text{otherwise} \end{cases}$$

which is also differentiable at $(0, 0)$, where $(\alpha)_- = \min\{\alpha, 0\}$.

For $d = (d_1, d_2)$ with $d_1 > 0, d_2 > 0$ and $w = (w_1, w_2) \in \mathbb{R}^2$, it can be seen that

$$\mathbb{V}''(\bar{p}; d, w) = \min\{d_1^2, d_2^2\} = \min\{V''_{\{1\}}(\bar{p}; d, w), V''_\emptyset(\bar{p}; d, w)\}$$

A second order directional derivative formula for the optimal value function of $(P(p))$ under MPCC-SSOSC in addition to the Inf-Compactness Assumption and MPCC-LICQ is given below. Note the explicit combinatorial construction needed there. Before stating this result, we need a technical result on the second order directional derivatives of min (or max) function, whose proof uses standard ideas and appears in Appendix.

Recall that a function $g : \mathbb{R}^{n_2} \to \mathbb{R}$ is said to be directionally differentiable in the Hadamard sense at $\bar{p} \in \mathbb{R}^{n_2}$ if it is directionally differentiable at $\bar{p}$ such that for each $d \in \mathbb{R}^{n_2}$, there exists the limit

$$\lim_{t\to 0^+, d'\to d} \frac{g(\bar{p}+td') - g(\bar{p})}{t}$$

which must therefore coincide with $g'(\bar{p}; d)$.

It is easy to see that if $g(\cdot)$ is locally Lipschitz near $\bar{p}$ and it is directionally differentiable at $\bar{p}$ then $g(\cdot)$ is directionally differentiable at $\bar{p}$ in the Hadamard sense. In particular, under the assumptions of Theorem 10, the global optimal value function $\mathbb{V}(p)$ is directionally differentiable at $\bar{p}$ in the Hadamard sense.

Proposition 14 *Consider*

$$\phi(\cdot) = \min_{w\in W} \phi_w(\cdot)$$

where W is a finite index set and $\phi_w : \mathbb{R}^{n_1} \to \mathbb{R}^1$. Let $\phi_w(\cdot)$ be directionally differentiable at $\bar{z}$ in the Hadamard sense for each $w \in W$. If $\phi''_w(\bar{z}; d, q)$ exists for all $d, q \in \mathbb{R}^{n_1}$, then $\phi(\cdot)$ is also twice directionally differentiable at $\bar{z}$ and its second order directional derivative is

$$\phi''(\bar{z}; d, q) = \min_{w\in I_{\phi'}(\bar{z};d)} \phi''_w(\bar{z}; d, q) \tag{10}$$

where $I_\phi(\bar{z}) = \{w : \phi_w(\bar{z}) = \phi(\bar{z})\}$ and $I_{\phi'}(\bar{z}; d) = \{w \in I_\phi(\bar{z}) : \phi'(\bar{z}; d) = \phi'_w(\bar{z}; d)\}$.

Theorem 15 *Under the Inf-Compactness Assumption, if MPCC-LICQ and MPCC-SSOSC hold at each $\bar{z} \in \mathbb{S}(\bar{p})$, then $\mathbb{V}(\cdot)$ is twice directionally differentiable at $\bar{p}$, and the second order directional derivative is*

$$\mathbb{V}''(\bar{p}; d, w) = \min_{\bar{z}\in M(\bar{p})} \min_{J\in\mathcal{J}(\bar{z},\bar{p})} V''_J(\bar{p}; d, w) \tag{11}$$

where $M(\bar{p}) := \{\bar{z} \in \mathbb{S}(\bar{p}) : \left(\nabla_p \mathbb{L}(\bar{z}, \bar{p}; \bar{\lambda}, \bar{\mu}, \bar{\xi}, \bar{\eta})\right)^T d = \mathbb{V}'(\bar{p}; d)\}$.

Proof First, note from the proof of Theorem 10 that $\mathbb{V}(p)$ is the min function $\min_{J\in\bar{\mathcal{J}}} V_J(p)$ for p near $\bar{p}$, where each $V_J(\cdot)$ is Lipschitz near $\bar{p}$ and directionally differentiable at $\bar{p}$. Second, the MPCC-SSOSC implies the usual SSOSC at $\bar{z}$ for each $(NLP_J(\bar{p}))$, $J \in \bar{\mathcal{J}}$. Standard results [23] give second order directional differentiability of the associated global optimal value functions $V_J(\cdot)$ at $\bar{p}$. Third, min functions of locally Lipschitz, twice directionally differentiable mappings are seen to be twice directionally differentiable by standard techniques (see Appendix), giving

$$\mathbb{V}''(\bar{p}; d, w) = \min_{J\in\bar{\mathcal{J}}'(\bar{z},\bar{p})} V''_J(\bar{p}; d, w)$$

where $\bar{\mathcal{J}}'(\bar{z}, \bar{p}) = \{J : V_J(\bar{p}) = \mathbb{V}(\bar{p}), V'_J(\bar{p}; d) = \mathbb{V}'(\bar{p}; d)\}$. Finally, (11) follows after showing that $\bar{\mathcal{J}}'(\bar{z}, \bar{p}) = \{J : J \in \mathcal{J}(\bar{z}, \bar{p})$ for some $\bar{z} \in M(\bar{p})\}$. ■

Some investigations, e.g. [23], relax the constraint qualification and second order conditions for NLPs and still obtain second order directional derivatives of the optimal value functions. Theorem 15 can be similarly relaxed.

Also, there are several different notions of second order directional derivatives used to analyse the second order variations of NLP optimal value functions, see the survey paper [2]. We can replace $\mathbb{V}''(\bar{p}; d, w)$ in (11) by any of them under the appropriate conditions.

Theorem 15 can be applied to the case of locally unique MPCC solution $\bar{z} \in \mathbb{S}(\bar{p})$. Here $\mathbb{V}(\cdot)$ becomes the global minimum of MPCC near $\bar{z}$.

Acknowledgements

The authors would like to thank Stefan Scholtes and Jane Ye for very helpful discussions, and also to anonymous referees for their comments including drawing our attention to reference [19]. The authors are grateful for partial support of this work by the Australian Research Council.

References

[1] Bard, J. F (1998): Practical Bilevel Optimization: Algorithms and Applications, Kluwer Academic Publishers.

[2] Bonnans, J. F., Shapiro, F. (1998): Optimization Problems with Perturbations, A Guided tour, SIAM Rev **40**, 228–264.

[3] Chen, Y., Florian, F. (1995): The nonlinear bilevel programming problem: formulations, regularity and optimality conditions, Optimization **32**, 193–209.

[4] Clarke, F. H. (1983): Optimization and Nonsmooth Analysis, A Wiley-Interscience Publication, New York.

[5] Fukushima, M., Pang, J.-S. (2000): Convergence of a smoothing continuation method for mathematical programs with complementarity constraints. In: Théra, M., Tichatschke, R., eds., Ill-posed Variational Problems and Regularization Techniques, Springer, 99–110.

[6] Gauvin, J., Dubeau, F. (1982): Differential properties of the marginal function in mathematical programming, Math. Prog. Study **19**, 101–119.

[7] Gauvin, J., Janin, R. (1990): Directional derivative of the value function in parametric optimization, Ann. Oper. Res. **27**, 237–252.

[8] Hu, X. (2000): Mathematical Programs with Complementarity Constraints, PhD dissertation, The University of Melbourne, Department of Mathematics and Statistics, in progress.

[9] Jiang, J., Ralph, D. (1999): QPECgen, a MATLAB generator for mathematical programs with quadratic objectives and affine variational inequality constraints, Comput. Optim. Appl. **13**, 25–59.

[10] Jittorntrum, K. (1984): Solution point differentiability without strict complementarity in nonlinear programming, Math. Prog. Study Ser. **21**, 127–138.

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

[12] Lucet, Y., Ye, J. J. (2001): Sensitivity analysis for optimization problems with variational inequality constraints, SIAM J. Control Optim. **40**, 699–723.

[13] Luo, Z.-Q., Pang, J.-S., Ralph, D. (1996): Mathematical programs with equilibrium constraints, Cambridge University Press, Cambridge.

[14] Luo, Z.-Q., Pang, J.-S., Ralph, D. (1998): Piecewise sequential quadratic programming for mathematical programs with nonlinear complementarity constraints. In: Migdalas, A. et al., ed., Multilevel Optimization: Algorithms, Complexity and Applications, Kluwer Academic Publishers.

[15] Mordukhovich, B. M. (1980): Metric approximations and necessary optimality conditions for general classes of nonsmooth extremal problems, Soviet Math. Dokl. **22**, 526–530.

[16] Outrara, J. V., Kocvara, M., Zowe, J. (1998): Nonsmooth approach to optimization problems with equilibrium constraints: Theory, applications and numerical results, Kluwer Academic Publishers.

[17] Polak, E. (1997): *Optimization : algorithms and consistent approximations*, Springer-Verlag, New York.

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

[19] Rockafellar, R. T. (1984): Directional differentiability of the optimal value function in a nonlinear programming problem, Math. Prog. Study **21**, 213–226.

[20] Scheel, M., Scholtes, S. (2000): Mathematical Programs with Complementarity Constraints: Stationarity, Optimality, and Sensitivity, Math. Oper. Res. **25**, 1–22.

[21] Scholtes, S. (2001): Convergence properties of a regularization scheme for mathematical programs with complementarity constraints, SIAM J. Optim. **11**, 918–936.

[22] Scholtes, S., Stöhr, M. (2001): How stringent is the linear independence assumption for mathematical programs with stationarity constraints? Math. Oper. Res. **21**, 851–863.

[23] Shapiro, A. (1988): Perturbation theory of nonlinear programs when the set of optimal solutions is not singleton, Appl. Math. Optim. **18**, 215–229.

[24] Wets, R. J. B. (1973): On inf-compact mathematical programs, Fifth Conference on Optimization Techniques, Part 1, Lect. Notes in Comput. Sci. **3**.

[25] Ye, J. J. (1998): Optimality conditions for optimization problems with complementarity constraints, SIAM J. Optim. **9**, 374–387.

[26] Ye, J.J., Zhu, D. L., Zhu, Q. J. (1997): Exact penalization and necessary optimality conditions for generalized bilevel programming problems, SIAM J. Optim. **7**, 481–507.

Appendix

Proof of Proposition 9

First, it is easy to see that there exists a neighborhood $N_1(\bar{p}) \subset N(\bar{p})$ of $\bar{p}$ such that $V(p) < \alpha$ for $p \in N_1(\bar{p})$ under the assumptions of the proposition since $(NLP(p))$ is feasible for all p near $\bar{p}$ due to the LICQ at all solutions of $(NLP(\bar{p}))$, whose solution set is not empty under the assumptions. Then we consider the following NLP:

$$(NLP^\alpha(p)) \qquad \begin{array}{ll} \underset{z}{\text{Minimize}} & f(z,p) \\ \text{subject to} & g(z,p) \le 0,\ h(z,p) = 0 \\ & f(z,p) \le \alpha \end{array}$$

where the constraint $f(z,p) \le \alpha$ is inactive at any optimal solution of $(NLP^\alpha(p))$ when $p \in N_1(\bar{p})$. In particular, the LICQ holds for $(NLP^\alpha(\bar{p}))$ and the uniform compactness [6, Definition 3.3] holds for $(NLP^\alpha(p))$ with p near $\bar{p}$. Let $V^\alpha(p)$ denote the optimal value function of $(NLP^\alpha(p))$. Then $V^\alpha(p)$ is locally Lipschitz [6, Theorem 5.1] near $\bar{p}$ and $V^\alpha(p)$ is directionally differentiable [6, Corollary 4.4] at $\bar{p}$. Moreover, its directional derivative [6, Corollary 4.4] is given by the formula specified in the proposition. Note that the conditions of the proposition remain valid under small perturbations of $\bar{p}$, hence V^α is directionally differentiable near $\bar{p}$.

Now, for $p \in N_1(\bar{p})$, we have $V(p) = V^\alpha(p)$. So, the proposition follows. ■

Consider the min function $\phi(z) = \min_w \phi_w(z)$ defined in Proposition 14. If $\phi_w(\cdot)$ is Lipschitz near $\bar{z}$ and directionally differentiable at $\bar{z}$ for each $w \in W$, then by standard results, see [17, Section 4.5], $\phi(\cdot)$ is directionally differentiable in the Hadamard sense at $\bar{z}$ and its directional derivative is given by

$$\phi'(\bar{z};d) = \min_{w \in I(\bar{z})} \phi'_w(\bar{z};d) \tag{12}$$

where $I(\bar{z}) := \{w \in W : \phi_w(\bar{z}) = \phi(\bar{z})\}$.

Proof of Proposition 14

It is easy to see that $I_{\phi'}(\bar{z};d) \neq \emptyset$ and that for any $z(t) = \bar{z} + td + t^2 q + o(t^2)$,

$$\limsup_{t \to 0^+} \frac{\phi(z(t)) - \phi(\bar{z}) - t\phi'(\bar{z};d)}{t^2} \le \phi''_w(\bar{z};d,q) \text{ for any } w \in I_{\phi'}(\bar{z};d).$$

Now, for those indices $w' \in I_\phi(\bar{z}) \setminus I_{\phi'}(\bar{z};d)$, one has

$$\phi_{w'}(z(t)) > \phi_w(z(t)), \text{ for any } w \in I_{\phi'}(\bar{z};d)$$

because $\phi_{w'}(\cdot) - \phi_w(\cdot)$ is directionally differentiable in the Hadamard sense at $\bar{z}$ for any $w \in I_{\phi'}(\bar{z};d)$ and $\phi'_{w'}(\bar{z};d) > \phi'_w(\bar{z};d)$. Therefore, one has

$$\liminf_{t \to 0^+} \frac{\phi_{w'}(z(t)) - \phi(\bar{z}) - t\phi'(\bar{z};d)}{t^2} \ge \lim_{t \to 0^+} \frac{\phi_w(z(t)) - \phi(\bar{z}) - t\phi'(\bar{z};d)}{t^2} = \phi''_w(\bar{z};d,q) \tag{13}$$

for any $w \in I_{\phi'}(\bar{z};d)$.

Let $\{t_k\} \to \{0^+\}$ such that

$$\lim_{t_k \to 0^+} \frac{\phi(z(t_k)) - \phi(\bar{z}) - t_k\phi'(\bar{z};d)}{t_k^2} = \liminf_{t \to 0^+} \frac{\phi(z(t)) - \phi(\bar{z}) - t\phi'(\bar{z};d)}{t^2}.$$

Since W is a finite set, there is at least one index v such that

$$\liminf_{t\to 0^+} \frac{\phi(z(t)) - \phi(\bar{z}) - t\phi'(\bar{z};d)}{t^2} = \lim_{t_k\to 0^+} \frac{\phi_v(z(t_k)) - \phi_v(\bar{z}) - t_k\phi_v'(\bar{z};d)}{t_k^2}$$

where if necessary, we can choose some convergent subsequence of $\{t_k\}$. By (13), without loss of generality, we can assume that $v \in I_{\phi'}(\bar{z};d)$. Hence,

$$\liminf_{t\to 0^+} \frac{\phi(z(t)) - \phi(\bar{z}) - t\phi'(\bar{z};d)}{t^2} = \phi_v''(\bar{z};d,q) \geq \min_{w\in I_{\phi'}(\bar{z};d)} \phi_w''(\bar{z};d,q)$$

which implies the formula (10). ■