Directional constraint qualifications and optimality conditions with application to bilevel programs

by

Kuang Bai
B.Sc., Lanzhou University, 2013
M.Sc., Dalian University of Technology, 2016

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Mathematics and Statistics

Directional constraint qualifications and optimality conditions with application to bilevel programs

by

Kuang Bai
B.Sc., Lanzhou University, 2013
M.Sc., Dalian University of Technology, 2016

Supervisory Committee

Dr. Jane Ye, Supervisor
(Department of Mathematics and Statistics)

Dr. David Goluskin, Departmental Member
(Department of Mathematics and Statistics)

Dr. Yang Shi, Outside Member
(Department of Mechanical Engineering)

Supervisory Committee

Dr. Jane Ye, Supervisor
(Department of Mathematics and Statistics)

Dr. David Goluskin, Departmental Member
(Department of Mathematics and Statistics)

Dr. Yang Shi, Outside Member
(Department of Mechanical Engineering)

ABSTRACT

The main purpose of this dissertation is to investigate directional constraint qualifications and necessary optimality conditions for nonsmooth set-constrained mathematical programs.

First, we study sufficient conditions for metric subregularity of the set-constrained system. We introduce the directional version of the quasi-/pseudo-normality as a sufficient condition for metric subregularity, which is weaker than the classical quasi-/pseudo-normality, respectively. Then we apply our results to complementarity and Karush-Kuhn-Tucker systems.

Secondly, we study directional optimality conditions of bilevel programs. It is well-known that the value function reformulation of bilevel programs provides equivalent single-level optimization problems which are nonsmooth and never satisfy the usual constraint qualifications such as the Mangasarian-Fromovitz constraint qualification (MFCQ). We show that even the first-order sufficient condition for metric subregularity (which is generally weaker than MFCQ) fails at each feasible point of bilevel programs. We introduce the directional Clarke calmness condition and show that under the directional Clarke calmness condition, the directional necessary optimality condition holds. We perform directional sensitivity analysis of the value function and propose the directional quasi-normality as a sufficient condition for the directional Clarke calmness.

Contents

List of Symbols

Spaces and Orthants

$\bar{\mathbb{R}}$ the extended real numbers

$\mathbb{R}$ the real numbers

$\mathbb{R}^n$ the n-dimension real vector space

$\mathbb{R}^n_+$ the nonnegative orthant in $\mathbb{R}^n$

$\mathbb{R}^n_-$ the nonpositive orthant in $\mathbb{R}^n$

$\mathbb{R}^{n\times m}$ the set of all real n by m matrices

$\mathcal{X}$ a finite dimensional Euclidean space

Cones

$\widehat{N}_\Omega(x)$ the regular/Fréchet normal cone to set Ω at x

$N_\Omega(x)$ the limiting normal cone to set Ω at x

$N_\Omega(x;d)$ the limiting normal cone to set Ω at x in direction d

$T_\Omega(x)$ the tangent cone to set Ω at x

$N^c_\Omega(x)$ the Clarke normal cone to set Ω at x

Functions

$\delta_\Omega(x)$ the indicator function of set Ω

$\nabla f(x)$ the gradient of the function $f:\mathbb{R}^n\to\mathbb{R}$ at x

$\nabla\phi(x)$ the Jacobian matrix in $\mathbb{R}^{n\times m}$ of single-valued map $\phi:\mathbb{R}^n\to\mathbb{R}^m$ at x

$\partial f(x)$ the limiting subdifferential of a function f at x

$\partial f(x;(d,\xi))$ the directional limiting subdifferential of a function f at x in direction (d,ξ)

$\partial^c f(x)$ the Clarke subdifferential of a function f at x

$\partial^c f(x;d)$ the directional Clarke subdifferential of a function f at x in direction d

$\partial_a f(x;d)$ the analytic directional limiting subdifferential of a function f at x in direction d

$\widehat{\partial} f(x)$ the Fréchet subdifferential of a function f at x

$\widehat{D}^*\Psi(x)$ the regular/Fréchet coderivative of multifunction Ψ at x

$D\Psi(x)$ the graphical derivative of multifunction Ψ at x

$D^*\Psi(x)$ the coderivative of multifunction Ψ at x

$D^2\Psi(x)(d)$ the second-order graphical derivative of multifunction Ψ at x in direction d

$d_\Omega(x)/\mathrm{dist}(x,\Omega)$ the distance between point x and set Ω

$f'(x;d)$ the directional derivative of a function f at x in direction d

$o(h)$ an infinitesimal of higher order than h

Abrreviations

CS complementarity system

FOSCMS first-order sufficient condition for metric subregularity

KKT Karush-Kuhn-Tucker

LICQ linearly independent constraint qualification

MFCQ Mangasarian-Fromovitz constraint qualification

MPEC mathematical program with equilibrium constraints

NNAMCQ no nonzero abnormal multiplier constraint qualification

PLICQ positive linearly independent constraint qualification

RCPLD the relaxed constant positive linear dependence constraint qualification

RCR relaxed constant rank

RS the Robinson stability property

SOSCMS second-order sufficient condition for metric subregularity

WSCMS weak sufficient condition for metric subregularity

Sequences

$\liminf_{x\to\bar{x}} \Psi(x)$ the lower/inner limit for a multifunction Ψ at x

$\limsup_{x\to\bar{x}} \Psi(x)$ the upper/outer limit for a multifunction Ψ at x

$x^k \xrightarrow{\Omega} \bar{x}$ a sequence $\{x^k\}$ in set Ω converging to $\bar{x} \in \Omega$

$x^k \xrightarrow{d} \bar{x}$ a sequence $\{x^k\}$ converging to $\bar{x}$ along direction d

Sets

$\mathbb{B}$ the unit open ball centered at $0 \in \mathbb{R}^n$ in $\mathbb{R}^n$

$\mathbb{B}_\delta(x)$ the open ball centered at $x \in \mathbb{R}^n$ with radius δ in $\mathbb{R}^n$

$\mathcal{V}_{\epsilon,\delta}(d)$ a neighborhood of 0 in direction d with radius (ϵ, δ)

Ω^n the n-Cartesian product of set Ω

Ω^o the polar of set Ω

Ω_{EC} the union of $\{0\} \times \mathbb{R}_+$ and $\mathbb{R}_+ \times \{0\}$

$\{x\}$ the set consisting of the vector x

$cl\Omega$ the closure of set Ω

$co\Omega$ the convex hull of set Ω

$domf$ the domain of the function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$

$int\Omega$ the interior of set Ω

epi$f$ the epigraph of function f

gph$f$ the graph of function f

ACKNOWLEDGEMENTS

I would like to thank:

my supervisor, Dr. Jane J. Ye for her countless patience, dedicated support and guidance throughout my PhD studies. Without her encouragement and persistent help this dissertation would not have been possible.

Dr. J. Zhang for his dedicated help and valuable suggestions on my research.

Dr. H. Gfrerer for an inspiring discussion on my research.

UVic and China Scholarship Council for the fellowships and scholarships.

Last but not least, I would like to thank my family for their unconditional love, support and encougement.

Chapter 1

Introduction

This chapter introduces the main objects of our research. Comprehensive literature reviews on necessary optimality conditions, constraint qualifications and metric subregularity are presented. The contribution is summarized. Finally, necessary background materials are presented.

1.1 Research Objects

Our research mainly studies the following set-constrained mathematical program

$$(\text{P}) \qquad \begin{array}{ll} \min & f(x) \\ s.t. & P(x) \in \Lambda, \end{array}$$

where $f : \mathcal{X} \to \mathbb{R}$ and $P : \mathcal{X} \to \mathcal{Y}$ are continuous, $\Lambda \subseteq \mathcal{Y}$ is closed and $\mathcal{X}, \mathcal{Y}$ are finite dimensional Euclidean spaces.

(P) is very general. It includes many classical optimization problems as special cases. For example, when $\mathcal{X} := \mathbb{R}^n, \mathcal{Y} := \mathbb{R}^{p+q}, P(x) := (h(x), g(x))$ with $h : \mathbb{R}^n \to \mathbb{R}^p, g : \mathbb{R}^n \to \mathbb{R}^q$ and $\Lambda := \{0\}^p \times \mathbb{R}^q_-$, (P) becomes the mathematical program with equality and inequality constraints, which is also known as the nonlinear program

$$(\text{NLP}) \qquad \begin{array}{ll} \min & f(x) \\ s.t. & h(x) = 0, \\ & g(x) \leq 0. \end{array}$$

When $\mathcal{X}$ is a finite dimensional Euclidean space, $\mathcal{Y}$ is the linear space of $p \times p$ symmetric matrices and Λ consists of positive semidefinite matrices, (P) is called the semidefinite program

$$\text{(SDP)} \quad \begin{array}{ll} \min & f(x) \\ s.t. & P(x) \succeq 0, \end{array}$$

where $A \succeq 0$ means A is positive semidefinite.

When $\mathcal{X} := \mathbb{R}^n$, $\mathcal{Y} := \mathbb{R}^{m+1}$ and $\Lambda := \{(y,z)|z \geq \|y\|_2\}$ is the so-called second order cone/Lorentz cone/icecream cone, (P) is called the second-order cone program, where $\|\cdot\|_2$ denotes the Euclidean norm.

Our research can be applied to cases where Λ is nonconvex. Particularly, in **Chapter 2**, we will study the Karush-Kuhn-Tucker (KKT) system (2.25) of NLP and the cone complementarity system (2.21).

1.2 Literature Review on Metric Subregularity

When solving optimization problems numerically, round-off errors and noisy data are often unavoidable. To study the influence of errors, it is meaningful to study the following parameterized set-constrained system

$$P(x) + \alpha \in \Lambda,$$

where α denotes the perturbation parameter. A critical issue is the stability analysis of the perturbed set-valued map $\mathcal{F}(\alpha) := \{x \in \mathcal{X} | P(x) + \alpha \in \Lambda\}$. We define $\mathcal{F} := \mathcal{F}(0)$.

To make this chapter more readable, in the sequel, we present most of the classical results in the case of NLP. Consider the constraint system of parametric NLP in the form

$$\begin{aligned} h(x) + u &= 0, \\ g(x) + v &\leq 0, \end{aligned} \tag{1.1}$$

where $(u,v) \in \mathbb{R}^{p+q}$ is the purturbation parameter.

The error bound property is an important tool to study the stability of the constraint system $P(x) \in \Lambda$.

Definition 1.2.1 (Error bound property)**.** *We say that the local error bound property* *holds for (1.1) at $\bar{x}$ if there exist $\kappa > 0,\ \epsilon > 0$ such that*

$$\mathrm{dist}(x, \mathcal{F}) \leq \kappa(\|h(x)\| + \|g_+(x)\|) \tag{1.2}$$

holds for $\forall x \in \bar{x} + \epsilon\mathbb{B}$, where $\mathcal{F} := \{x | h(x) = 0, g(x) \leq 0\}$ and $(g_j)_+(x) := \max\{0, g_j(x)\}$ for $j = 1, \ldots, q$.

We say the global error bound property for (1.1) holds, if (1.2) holds for any x.

Generally, it is hard to calculate the distance between a point x and the solution set $\mathcal{F}$, but if error bound holds, we can majorize $\mathrm{dist}(x, \mathcal{F})$ by the error $\kappa(\|h(x)\| + \|g_+(x)\|)$, which is easier to compute.

In 1952, Hoffman [37] proved that the global error bound holds automatically for linear constraint systems. We say (1.1) is a linear system, if $h(x) := A_1 x + b_1$ and $g(x) := A_2 x + b_2$, with A_1, A_2 being matrices and b_1, b_2 being vectors.

For general nonlinear programs, error bound property does not necessarily hold automatically and extra conditions are needed. For smooth NLP, there are two classical sufficient conditions for the error bound property. One is the so-called linear independent constraint qualification (LICQ). Define the active index set $\bar{\mathcal{I}}_g := \{j = 1, \ldots, q | g_j(\bar{x}) = 0\}$.

Definition 1.2.2 (LICQ). *Let $\bar{x}$ be a feasible point of NLP. Suppose that h, g are smooth. We say that LICQ holds at $\bar{x}$ if the set of vectors $\{\nabla h_i(\bar{x}), \nabla g_j(\bar{x}) | i = 1, \ldots, p,\ j \in \bar{\mathcal{I}}_g\}$ are linearly independent.*

The other is the Magasarian-Fromovitz constraint qualification (MFCQ).

Definition 1.2.3 (MFCQ). *Let $\bar{x}$ be a feasible point of NLP. Suppose that h, g are smooth. We say that MFCQ holds at $\bar{x}$ if the set of vectors $\{\nabla h_i(\bar{x}) | i = 1, \ldots, p\}$ are linearly independent and there exists a vector d such that $\nabla h_i(\bar{x})d = 0$ for all $i = 1, \ldots, p$ and $\nabla g_j(\bar{x})d < 0$ for any $j \in \bar{\mathcal{I}}_g$.*

One can easily find that LICQ is stronger than MFCQ. And MFCQ is equivalent to the positive linear independence constraint qualification (PLICQ).

Definition 1.2.4 (PLICQ). *Let $\bar{x}$ be a feasible point of NLP. Suppose that h, g are smooth. We say that PLICQ holds at $\bar{x}$ if there exists no nonzero (λ, μ) satisfying*

$$\begin{aligned} \nabla h(\bar{x})^T \lambda + \nabla g(\bar{x})^T \mu = 0, \\ 0 \leq \mu \perp g(\bar{x}). \end{aligned} \tag{1.3}$$

We give a brief proof for the equivalence between MFCQ and PLICQ in the sequel. First, we need the following theorem. We define $\mathbb{R}^{n\times m}$ as the set of all n by m matrices.

Theorem 1.2.1 (Motzkin's Transposition Theorem)**.** *(see e.g. [56]) Let $A \in \mathbb{R}^{n_1\times m}, B \in \mathbb{R}^{n_2\times m}$ and $C \in \mathbb{R}^{n_3\times m}$ be real constrant matrices with A nonempty. Then either the system*

$$Ay < 0,\ By \leq 0,\ Cy = 0$$

has a solution $\bar{y} \in \mathbb{R}^m$, or the system

$$\lambda_1^T A + \lambda_2^T B + \lambda_3^T C = 0,\ \lambda_1 \geq 0,\ \lambda_1 \neq 0,\ \lambda_2 \geq 0$$

has a solution $(\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^{n_1+n_2+n_3}$, but never both.

Proof. Assume PLICQ holds while MFCQ fails at $\bar{x}$. This means either (i) the set of vectors $\{\nabla h_i(\bar{x}) | i = 1, \ldots, p\}$ are linearly dependent, or (ii) the system

$$\nabla h_i(\bar{x})d = 0,\ \nabla g_j(\bar{x})d < 0,\ i = 1, \ldots, p,\ j \in \bar{\mathcal{I}}_g$$

has no solution. For case (i), we obtain there exists $0 \neq \lambda \in \mathbb{R}^p$ such that $(\lambda, 0)$ satisfy (1.3), contradicting PLICQ. For case (ii), letting A be the matrix whose columns consists of $\{\nabla g_j(\bar{x}) | j \in \bar{\mathcal{I}}_g\}$, by Motzkin's transposition theorem, we can find nonzero (λ, μ) satisfying (1.3), which implies a contradiction. In summary, we obtain PLICQ can imply MFCQ.

Conversely, we assume MFCQ holds while PLICQ fails. Then we can find nonzero (λ, μ) satisfying (1.3). Multiplying both sides of the equation in (1.3) by the vector d defined in MFCQ, we obtain

$$0 > \sum_{i=1}^{p} \lambda_i \nabla h(\bar{x})^T d + \sum_{j \in \bar{\mathcal{I}}_g} \mu_j \nabla g_j(\bar{x})^T d = 0,$$

which is a contradiction. Hence, we obtain MFCQ implies PLICQ. Thus, MFCQ is equivalent to PLICQ. ■

Robinson [67] proved that when the solution set $\mathcal{F}$ is bounded convex closed, under Slater condition, the global error bound for NLP holds.

Definition 1.2.5 (Slater's condition)**.** *There exists x^0 such that $h(x^0) = 0$ and*

$g(x^0) < 0$.

For nonsmooth NLP, Ye and Zhang [90] proved that under the quasi-/pseudo-normality condition, the local error bound property holds. Following is the definition of quasi-/pseudo-normality for smooth NLP.

Definition 1.2.6 (Quasi-/pseudo-normality)**.** *[9] Let $\bar{x}$ be a feasible point of NLP. Suppose that h, g are smooth. We say that quasi-normality holds at $\bar{x}$ if there exists no nonzero $(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}^q_+$ satisfying condions in (1.3) and such that there exists a sequence $\{x^k\}$ converging to $\bar{x}$ satisfying*

$$\begin{aligned} &\lambda_i h_i(x^k) > 0, \text{ for } \forall i \in I := \{i = 1, \ldots, p | \lambda_i \neq 0\} \\ &\mu_j g_j(x^k) > 0, \text{ for } \forall j \in J := \{j = 1, \ldots, q | \mu_j > 0\}. \end{aligned}$$

And we say that pseudo-normality holds at $\bar{x}$ if there exists no nonzero $(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}^q_+$ satisfying condions in (1.3) and such that there exists a sequence $\{x^k\}$ converging to $\bar{x}$ satisfying

$$\lambda^T h(x^k) + \mu^T g(x^k) > 0.$$

By definition, the above sufficient conditions for error bound have the following relationship.

$$\text{LICQ} \implies \text{PLICQ} \Leftrightarrow \text{MFCQ} \implies \text{pseudo-normality} \implies \text{quasi-normality}. \quad (1.4)$$

And when $g(x)$ is smooth convex and $h(x) := Ax - b$ with $A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p$. We have that MFCQ is equivalent to Slater's condition.

The error bound property plays an important role in optimization. It serves as a sufficient condition for exact penalty/Clarke calmness.

Lemma 1.2.1. *[12, Proposition 2.4.3] Let $f(x)$ be a Lipschitz function of rank K over a set S. Let $\bar{x} \in C \subseteq S$. Suppose that $f(x)$ attains a local minimum over C at $\bar{x}$. Then for any $\hat{K} > K$, the function $f(x) + \hat{K}\mathrm{dist}(x, C)$ attains a local minimum over S at $\bar{x}$.*

In Lemma 1.2.1 let $S := \mathbb{R}^n$, $C := \mathcal{F}$. Then the error bound property means $\mathrm{dist}(x, \mathcal{F}) \leq \kappa(\|h(x)\| + \|g_+(x)\|)$. By Lemma 1.2.1, one can easily obtain the following theorem.

Theorem 1.2.2 (Clarke calmness)**.** *(see e.g. [12, Proposition 6.4.3]) Suppose that $f(x)$ is Lipschitz continuous and $\bar{x}$ is a local solution to NLP. If the local error bound property holds at $\bar{x}$, then the Clarke calmness holds at $\bar{x}$, i.e., there exists $\bar{\rho} > 0$ such that $\bar{x}$ is a local solution to the following program*

$$\min_x \; f(x) + \rho(\|h(x)\| + \|g_+(x)\|)$$

for any $\rho \geq \bar{\rho}$.

Besides, error bound property can serve as a sufficient condition for linear convergence of certain first-order numerical algorithms, see e.g. [76]. Suppose that $\bar{x}$ is a solution to NLP and KKT condition (1.8) holds at $\bar{x}$, equivalently,

$$\begin{aligned} &\nabla f(\bar{x}) + \nabla h(\bar{x})^T \lambda + \nabla g(\bar{x})^T \mu = 0, \\ &-\mu \leq 0, \;\; g(\bar{x}) \leq 0, \\ &\mu^T g(\bar{x}) = 0. \end{aligned} \tag{1.5}$$

Denote by $\mathcal{O}$ the set of (x, λ, μ) satisfying (1.5). Generally, it is difficult to calculate the distance of a point (x, λ, μ) to the set $\mathcal{O}$. But if error bound holds at $(\bar{x}, \bar{\lambda}, \bar{\mu})$ for (1.5), there exists positive scalars κ, δ such that

$$\operatorname{dist}((x, \lambda, \mu), \mathcal{O}) \leq \kappa \left(\|\nabla f(x) + \nabla h(x)^T \lambda + \nabla g(x)^T \mu\| + \|(-\mu)_+\| + \|g(x)_+\| + \|\mu^T g(x)\| \right),$$

whenever $\|(x, \lambda, \mu) - (\bar{x}, \bar{\lambda}, \bar{\mu})\| < \epsilon$. Then, one can obtain a sequence $(x, \lambda, \mu) \to \mathcal{O}$ by making $\|\nabla f(x) + \nabla h(x)^T \lambda + \nabla g(x)^T \mu\| + \|(-\mu)_+\| + \|g(x)_+\| + \|\mu^T g(x)\| \downarrow 0$.

When it comes to general set-constrained system

$$P(x) \in \Lambda, \tag{1.6}$$

the metric subregularity was introduced to play a similar role as error bound property of NLP. In **Chapter 2**, we will focus on the study of the metric subregularity of the set-valued map $-P(x) + \Lambda$ induced by (1.6).

1.3 Literature review on optimality conditions

To solve an optimization problem, an important issue is to study its optimality conditions. In this dissertation, we focus on studying first order necessary optimality conditions of (P). And this section is dedicated to review classical results on this topic for NLP.

For unconstrained optimization problems, it is well-known that if a function $f(x)$ attains a local extremum at a point $\bar{x}$ and $f(x)$ is differentiable at $\bar{x}$, the Fermat's extremum theorem (Fermat's rule) provides a necessary optimality condition $\nabla f(\bar{x}) = 0$.

Then consider constrained optimization programs with only equality constraints

$$\begin{aligned} &\min\ f(x) \\ &\ s.t.\ h(x) = 0, \end{aligned}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^p$ are continuously differentiable. Let $\bar{x}$ be a locally optimal solution. By the classical Lagrange multiplier theorem, if the Jacobian $\nabla h(\bar{x})$ has full row rank, then there exists a unique $\lambda \in \mathbb{R}^p$ such that $\nabla f(\bar{x}) = \nabla h(\bar{x})^T \lambda$.

For optimization problems with inequality constraints, the first result on describing the local minimizer of an optimization program is presented by Fritz John [45] in 1948, where he studied NLP with inequality constraints only. The result is now known as the Fritz-John necessary optimality condition and has been extended to many other optimization programs. Following is the Fritz-John condition for NLP given by Mangasarian and Fromovitz [56] in 1967.

Definition 1.3.1 (Fritz-John condition)**.** *Let $\bar{x}$ be a local optimal solution to NLP. Suppose f, h, g are smooth. Then there exists a nonzero vector $(\lambda_0, \lambda, \mu) \in \mathbb{R}^{1+p+q}$ such that*

$$\begin{aligned} \lambda_0 \nabla f(\bar{x}) + \nabla h(\bar{x})^T \lambda + \nabla g(\bar{x})^T \mu = 0, \\ \lambda_0 \geq 0,\ 0 \leq \mu \perp g(\bar{x}). \end{aligned} \tag{1.7}$$

In (1.7), $(\lambda_0, \lambda, \mu)$ is called the generalized Lagrange multiplier. We call (λ, μ) as the normal Lagrange multiplier if $\lambda_0 \neq 0$ and abnormal Lagrange multiplier if $\lambda_0 = 0$.

Notice that if $\lambda_0 = 0$, $\nabla f(\bar{x})$ disappears and the Fritz-John condition provides no information on $f(x)$, so it fails to serve as an optimaliy condition. Hence, it is meaningful to study optimality conditions where $\lambda_0 \neq 0$. If $\lambda_0 \neq 0$, by normalization we can obtain that there exists a Lagrange multiplier of the form $(1, \lambda', \mu')$. Then we say that $\bar{x}$ satisfies the KKT condition.

Definition 1.3.2 (KKT condition)**.** *Let $\bar{x}$ be a feasible point of NLP. Suppose that f, h, g are smooth. We say that KKT condition holds at $\bar{x}$ if there exists a nonzero vector $(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}^q_+$ such that*

$$\begin{aligned} &\nabla f(\bar{x}) + \nabla h(\bar{x})^T \lambda + \nabla g(\bar{x})^T \mu = 0, \\ &\mu \perp g(\bar{x}). \end{aligned} \tag{1.8}$$

The second line in (1.8) is called the complementary slackness condition. KKT condition does not always hold at a local minimizer. Following is an example where KKT condition fails at an optimal solution.

Example 1.3.1.

$$\begin{aligned} &\min\ x \\ &s.t.\ x^2 \leq 0. \end{aligned}$$

Obviously, $\bar{x} = 0$ is the unique optimal solution to the above program. Define $f(x) := x,\ g(x) := x^2$. We have $f'(\bar{x}) = 1,\ g'(\bar{x}) = 0$. There exists no $\mu \neq 0$ such that $f'(\bar{x}) + \mu g'(\bar{x}) = 0$. And the KKT condition fails at $\bar{x}$.

Remark 1.3.1. *In Example 1.3.1, one can easily find that the feasible region can be equivalently defined as $x = 0$. In this case, the KKT condition holds with $\lambda = -1$. This means that the fulfillment of the KKT conditions also depends on how the feasible set is expressed in terms of constraint functions.*

Consequently, how to guarantee that the KKT condition holds at all the local minimizers of a program becomes a critical problem. This introduces the study on constraint qualifications.

A constraint qualification is a condition imposed on the constraint region of (P), under which the KKT condition holds at any local optimal solution. Based on Fritz-John condition, if there is no nonzero abnormal Lagrange multiplier, the KKT condition holds. This is actually PLICQ or Basic Constraint Qualification (BCQ) for NLP.

By (1.4), LICQ and MFCQ can both imply PLICQ. Hence, they can serve as constraint qualifications.

Theorem 1.3.1. *(see e.g., [8]) Let $\bar{x}$ be a local optimal solution to NLP. Suppose that h, g are smooth. If LICQ/MFCQ holds at $\bar{x}$, then KKT condition holds at $\bar{x}$.*

For convex programs, we have Slater's condition as a classical constraint qualification. The proof can be obtained by taking into account the relation between Slater's condition and PLICQ.

Theorem 1.3.2. *(see e.g., [8]) Let $\bar{x}$ be a local optimal solution to NLP. Suppose that g is convex and $h := Ax - b$ with $A \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^p$. If Slater's condition holds, then the KKT condition holds at $\bar{x}$.*

Actually, the error bound property can also be used as a constraint qualification.

Theorem 1.3.3. *Suppose that $f(x)$ is Lipschitz continuous and $\bar{x}$ is a local optimal solution to NLP. If the local error bound property holds at $\bar{x}$, then KKT condition holds at $\bar{x}$.*

Proof. By Theorem 1.2.2, the Clarke calmness holds at $\bar{x}$, i.e., there exists $\bar{\rho} > 0$ such that $\bar{x}$ is a local optimal solution to the following program

$$\min_x \ f(x) + \rho(\|h(x)\| + \|g_+(x)\|)$$

for any $\rho \geq \bar{\rho}$. Then by the Fermat's rule and the calculus rules of limiting subdifferential (see e.g. Proposition 1.5.1), there exist $(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}^q_+$ with $\mu_j = 0$ for $\forall j \notin \bar{I}(\bar{x})$, satisfying that

$$\nabla f(\bar{x}) + \sum_{i=1}^{p} (\bar{\rho}\lambda_i) \nabla h(\bar{x}) + \sum_{j=1}^{q} \bar{\rho}\mu_j \nabla g_j(\bar{x}).$$

One can easily find that KKT condition is satisfied. ■

From the above theorem, all sufficient conditions for error bound property such as the quasi-/pseudo-normality can be used as constraint qualifications.

Theorem 1.3.4. *[9] Let $\bar{x}$ be a local optimal solution to NLP. Suppose that h, g are smooth. If quasi-/pseudo-normality holds at $\bar{x}$, then KKT condition holds at $\bar{x}$.*

The results presented above are for smooth NLP. However, general (P) may have complicated structures and include nonsmooth data. To deal with nonsmooth (P), there are two main challenges. First, we need to introduce generalized gradients such as Clarke subdifferential and limiting subdifferential, for nonsmooth mappings. The readers can refer to **Chap 1.5** for definitions of various generalized gradients.

Secondly, those constraint qualifications derived for smooth NLP are no longer useful. As we have discussed in **Chap 1.1**, (P) includes many classes of optimization problems as special cases. Constraint qualifications for NLP have been extended to many of these problems to deduce new constraint qualifications. For example, Ye and Zhang [90, 86], Guo, Ye and Zhang [30] extended quasi-/pseudo-normality to nonsmooth (P) and introduced new constraint qualifications and optimality conditions for (P).

And one can also study constraint qualifications for bilevel programs

$$\begin{aligned} (\mathrm{BP}) \quad \min_{x,y} \quad & F(x,y) \\ s.t. \quad & y \in S(x),\ G(x,y) \leq 0, \end{aligned}$$

where for any given x, $S(x)$ denotes the solution set of the lower level program

$$(P_x) \quad \min_y f(x,y) \quad \text{s.t. } g(x,y) \leq 0,$$

and $F, f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R},\ G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$, $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ are continuously differentiable.

According to [90], one can equivalently reformulate (BP) using the value function of the lower level program

$$\begin{aligned} (\mathrm{VP}) \qquad \min_{x,y} \quad & F(x,y) \\ \text{s.t.} \quad & f(x,y) - V(x) \leq 0,\ g(x,y) \leq 0,\ G(x,y) \leq 0. \end{aligned}$$

where $V(x) := \min_y \{f(x,y) | g(x,y) \leq 0\}$ denotes the value function of the lower level program. It has been shown that $V(x)$ is usually not differentiable even if f, g are smooth. Hence, (VP) is usually a nonsmooth program. And one can easily find that if we let $P(x) := (f(x,y) - V(x), G(x,y), g(x,y))$ and $\Lambda := \mathbb{R}_- \times \mathbb{R}_-^{p_1} \times \mathbb{R}_-^{p_2}$, (VP) is a special case of (P).

The following example shows that even if $g(x,y)$ is linear, $V(x)$ can be nondifferentiable.

Example 1.3.2.

$$\begin{aligned} \min_y \; & x - y, \\ s.t. \; & -x + y - 1 \leq 0, \\ & x + y - 1 \leq 0. \end{aligned}$$

One can easily find that

$$V(x) = \begin{cases} 2x - 1, & x \geq 0, \\ -1, & x \leq 0, \end{cases}$$

which is not differentiable at $x = 0$*.*

There are two main challenges of solving (VP). One is that, many existing constraint qualifications are developed for smooth programs and one needs to extend them to nonsmooth cases. The other challenge is, even when $V(x)$ is differentiable, classical constraint qualifications such as MFCQ fail at optimal solutions to (VP). Thus, it is necessary to study nonsmooth (P) and develop new applicable constraint qualifications. And in **Chapter 3**, we will study differentiability of $V(x)$ and derive applicable constraint qualification for (VP).

1.4 Main contributions

We summarize our contributions as follows.
In **Chapter 2**, we study the nonsmooth set-constrained system and derive a sufficient condition for metric subregularity which is weaker than the first-order sufficient condition for metric subregularity (FOSCMS) by adding an extra sequential condition. Then we introduce directional versions of quasi-/pseudo-normality which are stronger than the new weak sufficient condition for metric subregularity but weaker than classical quasi-normality and pseudo-normality respectively. Moreover we introduce a nonsmooth version of the second-order sufficient condition for metric subregularity and show that it is a sufficient condition for the new sufficient condition for metric subregularity to hold. For the class of set-valued maps where the single-valued mapping is affine and the abstract set is the union of finitely many convex polyhedral sets, we show that pseudo-normality and hence directional pseudo-normality holds automatically at each point of the graph. Finally, we apply our results to complementarity and Karush-Kuhn-Tucker systems.

In **Chapter 3**, we study directional optimality conditions of value function reformulation of bilevel programs. We show that even the FOSCMS fails at each feasible point of the bilevel program. We introduce the concept of directional calmness condition and show that under the directional calmness condition, the directional necessary optimality condition holds. While the directional optimality condition is in general sharper than the non-directional one, the directional calmness condition is in general weaker than the classical calmness condition and hence is more likely to hold. We perform the directional sensitivity analysis of the value function and propose the directional quasi-normality as a sufficient condition for the directional calmness. An example is given to show that the directional quasi-normality condition may hold for the bilevel program.

1.5 Backgrounds on variational analysis

In this section, we review the definition of some subdifferentials and some useful calculus rules. For more detailed information on the subject, the reader is referred to Mordukhovich [59], and Rockafellar and Wets [72].

Definition 1.5.1. *(subdifferentials; see, e.g., [59]). Let* $f : \mathcal{X} \to [-\infty, +\infty]$ *and* $\bar{x}$ *is a point where* f *is finite. Then*

- *the Fréchet (regular) subdifferential of f at $\bar{x}$ is the set*

$$\widehat{\partial} f(\bar{x}) := \left\{ \xi \in \mathcal{X} \,\middle|\, \liminf_{h \to 0} \frac{f(\bar{x}+h) - f(\bar{x}) - \langle \xi, h \rangle}{\|h\|} \geq 0 \right\};$$

- *the limiting (Mordukhovich or basic) subdifferential of f at $\bar{x}$ is the set*

$$\partial f(\bar{x}) := \left\{ \xi \in \mathcal{X} \,\middle|\, \exists x_k \to \bar{x}, f(x_k) \to f(\bar{x}), \textit{and } \xi_k \to \xi \textit{ with } \xi_k \in \widehat{\partial} f(x_k) \right\};$$

When $f(x)$ is Lipschitz continuous, the Clarke subdifferential can be equivalently defined as

$$\partial^c f(\bar{x}) := co\partial f(\bar{x}),$$

where $co\Omega := \{\alpha x + (1-\alpha)y | x, y \in \Omega, \alpha \in [0,1]\}$ denotes the convex hull of a set Ω. We give some useful calculus rules for the limiting subdifferential in the following proposition.

If f is convex, then

$$\widehat{\partial} f(\bar{x}) = \partial f(\bar{x}) = \partial^c f(\bar{x}) = \{\xi \in \mathcal{X} | f(\bar{x}) + \langle \xi, x - \bar{x} \rangle \leq f(x), \forall x\}.$$

Furthermore, if f is continuously differentiable,

$$\widehat{\partial} f(\bar{x}) = \partial f(\bar{x}) = \partial^c f(\bar{x}) = \{\nabla f(\bar{x})\}.$$

We present the limiting subdifferential of two frequently used functions.

$$\partial \max\{0, x\} = \begin{cases} 0, & \text{if } x < 0, \\ \text{[0,1]}, & \text{if } x = 0, \\ 1, & \text{if } x > 0. \end{cases} \quad \partial \|x\|_2 = \begin{cases} \frac{x}{\|x\|_2}, & \text{if } x \neq 0, \\ \mathbb{B}, & \text{if } x = 0. \end{cases}$$

In the following, we review some useful calculus rules of limiting/Clarke subdifferentials.

Proposition 1.5.1. *(Calculus Rules of limiting subdifferential)(see e.g. [72])*

(1) Let $f : \mathcal{X} \to \mathbb{R}$ be Lipschitz near $\bar{x}$ and $g : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ be lower semi-continuous (l.s.c.) and finite at $\bar{x}$. Let α, β be nonnegative scalars. Then $\partial(\alpha f + \beta g)(\bar{x}) \subset \alpha \partial f + \beta \partial g(\bar{x})$.

(2) Let $\phi : \mathcal{X} \to \mathcal{Y}$ *be Lipschitz near* $\bar{x}$ *and* $f : \mathcal{Y} \to \mathbb{R}$ *be Lipschitz near* $\phi(\bar{x})$*. Then*

$$\partial(f \circ \phi)(\bar{x}) \subset \bigcup_{\xi \in \partial f(\phi(\bar{x}))} \partial \langle \xi, \phi \rangle (\bar{x}).$$

(3) Let $f : \mathcal{X} \to \mathbb{R}$ *be Lipschitz near* $\bar{x}$ *and* C *be a closed subset of* $\mathcal{X}$*. If* $\bar{x}$ *is a local minimizer of* f *on* C*, then* $0 \in \partial f(\bar{x}) + N_C(\bar{x})$*.*

We denote by $co(\Omega)$ the convex hull of set Ω, i.e., $co(\Omega) := \{\lambda x + (1-\lambda)y | x, y \in \Omega,\ 0 \leq \lambda \leq 1\}$. And $\overline{co}(\Omega)$ denotes the closure of $co(\Omega)$.

Proposition 1.5.2. *(Calculus Rules of Clarke subdifferential)(see e.g. [12])*

(1) Let $f, g : \mathcal{X} \to \mathbb{R}$ *be Lipschitz near* $\bar{x}$*. Let* α, β *be scalars. Then*

$$\partial^c(\alpha f)(\bar{x}) = \alpha \partial^c f(\bar{x}),\ \partial^c(\alpha f + \beta g)(\bar{x}) \subset \alpha \partial^c f + \beta \partial^c g(\bar{x}).$$

(2) Let $\phi : \mathcal{X} \to \mathcal{Y}$ *be Lipschitz near* $\bar{x}$ *and* $f : \mathcal{Y} \to \mathbb{R}$ *be Lipschitz near* $\phi(\bar{x})$*. Then*

$$\partial^c(f \circ \phi)(\bar{x}) \subset \overline{co} \left(\bigcup_{\xi \in \partial^c f(\phi(\bar{x}))} \partial^c \langle \xi, \phi \rangle (\bar{x}) \right).$$

Chapter 2

Directional quasi-/pseudo-normality as sufficient condition for metric subregularity

In **Chapter 1.2**, we reviewed error bound properties for NLP and its sufficient conditions. When it comes to the stability analysis of the set-constrained system (1.6), error bound property is generalized as metric subregularity.

This chapter is mainly dedicated to the study on sufficient conditions for metric subregularity of (1.6). All work in this chapter has been published as a journal paper, see [4]. Recently, Benko et al. [6] also studied the directional quasi-/pseudo-normality condition with application to a more general class of disjunctive constraint systems, which includes the complementarity system as a special case. In particular, for disjunctive constraint systems, they reformulated the directional pseudo-normality condition into a much simplified form and derived second-order and higher-order sufficient conditions for it. Similarly, they studied the so-called ortho-disjunctive constraint system, for which they simplified and derived a second-order sufficient condition for the directional quasi-normality condition.

2.1 Introduction

In this chapter, we study the stability analysis of the system (1.6). Throughout the chapter, unless otherwise specified, we assume that $\mathcal{Y}$ is an m-dimensional Euclidean space with inner product $\langle \cdot, \cdot \rangle$ equipped with the orthonormal basis $\mathscr{E} = \{e_1, \ldots, e_m\}$. Without loss of generality, throughout this chapter for any $y \in \mathcal{Y}$ we denote $\langle y, e_i \rangle$ by y_i, $i = 1, \ldots, m$.

Since the set Λ is not required to be convex, the system represented by $P(x) \in \Lambda$ is very general and many systems can be formulated in this form. In particular, various variational inequalities/complementarity systems can be reformulated in this form. For example, consider the cone complementarity system defined as

$$\mathcal{K} \ni \Phi(x) \perp \Psi(x) \in \mathcal{K},$$

where $\mathcal{K}$ is a convex cone in $\mathcal{Y}$, $\Phi, \Psi : \mathcal{X} \to \mathcal{Y}$, and $y \perp z$ means that $\langle y, z \rangle = 0$. Then the cone complementarity system can be reformulated in the form (1.6) by defining $P(x) := (\Phi(x), \Psi(x))$ and the complementarity set

$$\Lambda := \{(y, z) \in \mathcal{Y} \times \mathcal{Y} | \mathcal{K} \ni y \perp z \in \mathcal{K}\}.$$

Note that although $\mathcal{K}$ is convex, the complementarity set is not convex.

Denote by $G(x) := P(x) - \Lambda$, a set-valued map induced by the system $P(x) \in \Lambda$. An important stability issue to study is the metric subregularity. We say that the set-valued map G is metrically subregular at $(\bar{x}, 0) \in gphG$, where

$$gphG := \{(x, y) | y \in G(x)\}$$

is the graph of G, if there exist $\kappa \geq 0$ and a neighborhood U of $\bar{x}$ such that

$$d(x, G^{-1}(0)) \leq \kappa d(P(x), \Lambda) \quad \forall x \in U,$$

where $d(x, C)$ denotes the distance between a point x and a set C and $G^{-1}(y) := \{x | y \in G(x)\}$ denotes the inverse of G at y. By convention, we have $d(x, \emptyset) = +\infty$.

The concept of metric subregularity was introduced by Ioffe [39] using the terminology "regularity at a point." The terminology "metric subregularity" was suggested by Dontchev and Rockafellar in [15, Definition 3.1]. This property is also referred to as an error bound property since it enables us to estimate the distance from a point x near $\bar{x}$ to the set of solutions to the system (1.6) by its residue $d(P(x), \Lambda)$, which is much easier to deal with; see, e.g., [18, 77, 78, 79, 16, 64] and the references therein for related results and applications. Metric subregularity is a weaker condition than the more familiar property of metric regularity, which requires the existence of $\kappa \geq 0$ and $U,\ V$, neighborhoods of $\bar{x}, 0$, respectively, such that

$$d(x, G^{-1}(y)) \leq \kappa d(P(x) + y, \Lambda) \quad \forall x \in U, y \in V,$$

and strong metric subregularity (see, e.g., [14]) which requires the existence of $\kappa \geq 0$ and U, a neighborhood of $\bar{x}$ such that

$$\|x - \bar{x}\| \leq \kappa d(P(x), \Lambda) \quad \forall x \in U.$$

It is well known (see, e.g., [15, Theorem 3.2]) that the metric subregularity of a set-valued map is equivalent to the calmness of its inverse map, which means that there exist $\kappa \geq 0$ and neighborhoods U of $\bar{x}$ and V of 0 such that

$$G^{-1}(y) \cap U \subseteq G^{-1}(0) + \kappa\|y\|\mathbb{B} \quad \forall y \in V,$$

where $\|\cdot\|$ and $\mathbb{B}$ denote the norm and the closed unit ball in $\mathcal{Y}$, respectively. The concept of calmness was first introduced by J. J. Ye and X. Y. Ye in [84, Definition 2.8] under a different name, "pseudo-upper-Lipschitz continuity," and the terminology of "calmness" was coined by Rockafellar and Wets in [72]. Note that the calmness property is part of the property required in the notion of pseudo-Lipschitz continuity introduced by Klatte [48]. As suggested by the name "pseudo-upper-Lipschitz continuity," the concept of calmness is weaker than both the pseudo-Lipschitz continuity (or Aubin continuity) introduced by Aubin [2] and the upper-Lipschitz continuity introduced by Robinson [68, 69, 70]. Analogous to the fact that a set-valued map is metrically subregular if and only if its inverse map is calm, it is well known that the metric regularity of a set-valued map is equivalent to the pseudo-Lipschitz continuity of its inverse map (see [59, Theorem 1.49]).

Metric subregularity/calmness plays an important role in optimization. It serves as a constraint qualification and a sufficient condition for exact penalty; see e.g., [11, 39, 38, 40, 49, 75, 81, 84]. As pointed out in [41], metric subregularity/calmness is also an important tool in the subdifferential calculus of nonsmooth analysis. More recently, it has been discovered that it serves as a sufficient condition for linear convergence of certain numerical algorithms [50, 76] and quadratic convergence of the Newton-type method [17].

Although the metric subregularity/calmness/error bound condition is very important, it is by no means easy to verify. For a long time, there have been only two major *verifiable* sufficient conditions: one is derived by Robinson's multifunction theory and the other is by Mordukhovich's criteria. By Robinson's multifunction theory [70], if the linear constraint qualification (linear CQ) holds, i.e., $P(x)$ is affine and Λ is the union of finitely many polyhedral convex sets, then the set-valued map $G(x) = P(x) - \Lambda$ must be a polyhedral multivalued function and so is its inverse map G^{-1}. Hence the set-valued map G^{-1} must be upper Lipschitz and hence calm. Recall that in optimization we call a multiplier abnormal if it is a multiplier associated with an optimality system where its component corresponding to the objective function vanishes. Assuming P is continuously differentiable (C^1), if the no nonzero abnormal multiplier constraint qualification (NNAMCQ) holds, i.e., there is no nonzero abnormal multiplier ζ such that

$$0 = \nabla P(\bar{x})^* \zeta, \quad \zeta \in N_\Lambda(P(\bar{x})), \tag{2.1}$$

where $N_\Lambda(\cdot)$ is the limiting normal cone, ∇P denotes the Fréchet derivative of P, and * denotes the adjoint, then Mordukhovich's criterion for metric regularity (see, e.g., [72, Theorem 9.40]) holds and so does metric subregularity. These two criteria are relatively strong since they are actually sufficient conditions for stronger stability concepts. And therefore there are many situations where these sufficient conditions do not hold but the systems are still metrically subregular. In general metric subregularity is weaker than NNAMCQ but for the case of a differentiable convex inequality system, which is (1.6) with each component function of P convex and differentiable and Λ a nonnegative orthant, Li [53] has shown that all the following conditions are equivalent: metric subreguality, Abadie's constraint qualification [1], the Slater condition and the MFCQ (which is equivalent to NNAMCQ in this case).

Over the last fifteen years or so, some results for characterizing metric subregularity/calmness for general set-valued maps have been obtained; see, e.g., [33, 34, 35, 36, 92]. Recently the concept of a directional limiting normal cone which is in general a smaller set than the limiting normal cone was introduced [29, 21]. Based on the result for general set-valued maps in [21], Gfrerer and Klatte [25, Corollary 1] showed that metric subregularity holds for system (1.6) at $\bar{x}$ under the FOSCMS: assuming $P(x)$ is C^1, if for each nonzero direction u satisfying $\nabla P(\bar{x})u \in T_\Lambda(P(\bar{x}))$, there is no nonzero ζ such that

$$0 = \nabla P(\bar{x})^* \zeta, \quad \zeta \in N_\Lambda(P(\bar{x}); \nabla P(\bar{x})u),$$

where $T_\Lambda(\cdot)$ and $N_\Lambda(y; d)$ are the tangent cone and the limiting normal cone at y in direction d (see Definition 2.2.2). Moreover, if $P(x)$ is continuously differentiable and twice directionally differentiable and Λ is the union of finitely many polyhedral convex sets, it was shown in [22, Theorem 4.3] that metric subregularity holds at $(\bar{x}, 0)$ under the following second-order sufficient condition for metric subregularity (SOSCMS): for each nonzero direction u satisfying $\nabla P(\bar{x})u \in T_\Lambda(P(\bar{x}))$, there exists no $\zeta \neq 0$ such that

$$\begin{cases} 0 = \nabla P(\bar{x})^* \zeta, \quad \zeta \in N_\Lambda(P(\bar{x}); \nabla P(\bar{x})u), \\ \langle \zeta, P''(\bar{x}; u) \rangle \geq 0 \end{cases}$$

where $P''(\bar{x}; u)$ denotes the second-order derivative of $P(x)$ at $\bar{x}$ in the direction u. Some sufficient conditions for the metric subregularity/calmness/error bound condition for special complementarity systems based on the FOSCMS have been obtained in [27, 87].

Another direction in the effort of weakening the NNAMCQ is to add some extra conditions to (2.1). In the case where P is continuously differentiable at $\bar{x}$, we say that quasi-normality and pseudo-normality hold at $\bar{x}$ if there exists no $\zeta \neq 0$ such that (2.1) holds and

$$\exists (x^k, s^k, \zeta^k) \to (\bar{x}, P(\bar{x}), \zeta) \ \text{ s.t. } \zeta^k \in N_\Lambda(s^k) \text{ and } \zeta_i(P_i(x^k) - s_i^k) > 0 \text{ if } \zeta_i \neq 0,$$

$$\exists (x^k, s^k, \zeta^k) \to (\bar{x}, P(\bar{x}), \zeta) \ \text{ s.t. } \zeta^k \in N_\Lambda(s^k) \text{ and } \langle \zeta, P(x^k) - s^k \rangle > 0,$$

respectively. It is obvious that pseudo-normality implies quasi-normality. For a system with equality and inequality constraints where all constraint functions are C^{1+} which means that the gradients are locally Lipschitz, Minchenko and Tarakanov [58, Theorem 2.1] showed that quasi-normality implies the existence of a local error bound or equivalently metric subregularity/calmness at $\bar{x}$. In [90, Theorem 5], this result is extended to systems with continuously differentiable equality constraint functions and subdifferentially regular inequality constraint functions and a regular constraint set. Quasi-normality/pseudo-normality for the general system in the form (1.6) was introduced by Guo, Ye and Zhang [30, Definition 4.2] and proved to be a sufficient condition for error bound/metric subregularity/calmness in [30, Theorem 5.2] under the Lipschitz continuity of P and the closeness of the set Λ only.

The main purpose of this chapter is to combine the two approaches of weakening the NNAMCQ , i.e., to replace the limiting normal cone by the directional normal cone as in FOSCMS and SOSCMS and to add extra conditions as in quasi-/pseudo-normality and prove that our weaker sufficient conditions are still sufficient for verifying the metric subregularity/calmness.

Our assumptions are very general. We only assume the continuity of the mapping $P(x)$. Indeed, it is natural to study the case where $P(x)$ is only continuous, since it will widen the range of applications of formation (1.6). For example, consider the recovery of an unknown vector $x \in \mathbb{R}^n$ (such as a signal or an image) from noisy data $b \in \mathbb{R}^m$ by minimizing with respect to x a regularized cost function

$$F(x, b) = f(x, b) + \mu g(x), \tag{2.2}$$

where typically $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a data-fidelity term and $g : \mathbb{R}^n \to \mathbb{R}$ is a nonsmooth regularization term, with $\mu > 0$ a parameter. One usual choice for the data-fidelity term is

$$f(x, b) = \sum_{i=1}^{m} |a_i^T x - b_i|^\rho$$

with $a_i \in \mathbb{R}^n$ and ρ in the range $(0, \infty]$; see, e.g., [62, 65, 66]. Apparently when ρ takes a value in the interval $(1, 2)$, the optimality condition of minimizing function (2.2) with respect to x can be described by $0 \in \nabla_x f(x, b) + \partial g(x)$, where $\partial g(x)$ denotes a certain subdifferential of g at x, which can be reformulated as $P(x) \in \Lambda$ where $P(x) := (x, \nabla_x f(x, b))$ is continuous and $\Lambda := gph\partial g$ is a locally closed set. Therefore, thanks to the equivalence between calmness of different reformulations established in [27, Proposition 3], our results can be used to study the calmness of the optimality condition system of minimizing (2.2) without imposing an unnecessarily stronger condition.

We organize this chapter as follows. First, we summarize some preliminary results in **Section 2.2**. In **Section 2.3** we propose the weak sufficient condition for metric subregularity and show that it is sufficient for metric subregularity. In **Section 2.4**, we propose the concepts of directional quasi-/pseudo-normality and show that they are stronger than the new sufficient condition for metric subregularity. Moreover, in this section it is shown that the SOSCMS implies pseudo-normality. In **Section 2.5**, we apply our results to complementarity systems and Karush-Kuhn-Tucker (KKT) systems.

2.2 Preliminaries and preliminary results

In this section, we gather some preliminaries on variational analysis and nonsmooth analysis that will be used in the following sections. We only give concise definitions and results that will be needed in this chapter. For more detailed information on the subject, the reader is referred to Mordukhovich [59], and Rockafellar and Wets [72].

First, we give the definition of tangent cones and normal cones.

Definition 2.2.1. *(tangent cones and normal cones; see, e.g., [72, Definition 6.1]). Given a set $\Omega \subseteq \mathcal{Y}$ and a point $\bar{y} \in \Omega$, the tangent cone to Ω at $\bar{y}$ is defined as*

$$T_\Omega(\bar{y}) := \{d \in \mathscr{Y} | \exists t_k \downarrow 0, d_k \to d \ \ s.t. \ \bar{y} + t_k d_k \in \Omega \ \forall k\}.$$

The derivable cone to Ω at $\bar{y}$ is defined as

$$T^i_\Omega(\bar{y}) := \{d \in \mathscr{Y} | \forall t_k \downarrow 0, \exists d_k \to d \ \ s.t. \ \bar{y} + t_k d_k \in \Omega \ \forall k\}.$$

A set Ω is said to be geometrically derivable if the tangent cone coincides with the derivable cone at each point of Ω, or equivalently if $\lim_{t\downarrow 0} t^{-1} d(\bar{y} + tu, \Omega) = 0$.

The regular normal cone and the limiting normal cone to Ω at $\bar{y}$ are defined as

$$\widehat{N}_\Omega(\bar{y}) := \left\{ \zeta \in \mathcal{Y} \,\middle|\, \limsup_{y \xrightarrow{\Omega} \bar{y}} \frac{\langle \zeta, y - \bar{y} \rangle}{\|y - \bar{y}\|} \le 0 \right\},$$

and

$$N_\Omega(\bar{y}) := \left\{ \zeta \in \mathcal{Y} \,\middle|\, \exists \ y_k \xrightarrow{\Omega} \bar{y}, \ \zeta_k \to \zeta \ such \ that \ \zeta_k \in \widehat{N}_\Omega(y_k) \ \forall k \right\}$$

respectively, where $y_k \xrightarrow{\Omega} \bar{y}$ means $y_k \to \bar{y}$ and for each k, $y_k \in \Omega$.

Recently a directional version of limiting normal cones was introduced in [29, Definition 2.3] and extended to general Banach spaces in [21].

Definition 2.2.2 (Directional Normal Cone)**.** *([29, Definition 2.3] or [21, Definition 2]). Given a set $\Omega \subseteq \mathbb{R}^n$, a point $\bar{x} \in \Omega$ and a direction $d \in \mathbb{R}^n$, the limiting normal cone to Ω at $\bar{x}$ in direction d is defined by*

$$N_\Omega(\bar{x}; d) := \left\{ \zeta \in \mathbb{R}^n \middle| \exists\ t_k \downarrow 0, d_k \to d, \zeta_k \to \zeta \ s.t.\ \zeta_k \in \widehat{N}_\Omega(\bar{x} + t_k d_k)\ \forall k \right\}.$$

It is obvious that $N_\Omega(\bar{x}; 0) = N_\Omega(\bar{x})$, $N_\Omega(\bar{x}; d) = \emptyset$ if $d \notin T_\Omega(\bar{x})$ and $N_\Omega(\bar{x}; d) \subseteq N_\Omega(\bar{x})$. It is also obvious that for all $d \in T_\Omega(\bar{x}) \setminus T_{bd\Omega}(\bar{x})$, one has $N_\Omega(\bar{x}; d) = \{0\}$. Moreover when Ω is convex, by [24, Lemma 2.1] the directional and the classical normal cone have the following relationship

$$N_\Omega(\bar{x}; d) = N_\Omega(\bar{x}) \cap \{d\}^\perp \qquad \forall d \in T_\Omega(\bar{x}). \tag{2.3}$$

Proposition 2.2.1. *[87, Proposition 3.3] Let $\Omega := \Omega_1 \times \cdots \times \Omega_l$, where $\Omega_i \subseteq \mathbb{R}^{n_i}$ are closed for $i = 1, \ldots, l$ and $n = n_1 + \cdots + n_l$. Consider a point $\bar{y} = (\bar{y}_1, \ldots, \bar{y}_l) \in \Omega$ and a direction $d = (d_1, \ldots, d_l) \in \mathbb{R}^n$. Then*

$$\begin{aligned} T_\Omega(\bar{y}) &\subseteq T_{\Omega_1}(\bar{y}_1) \times \cdots \times T_{\Omega_l}(\bar{y}_l), \\ N_\Omega(\bar{y}; d) &\subseteq N_{\Omega_1}(\bar{y}_1; d_1) \times \cdots \times N_{\Omega_l}(\bar{y}_l; d_l). \end{aligned}$$

The equality holds if all except at most one of Ω_i for $i = 1, \ldots, l$ are directionally regular at y_i in the sense of [87, Definition 3.3].

Recently based on the concept of the directional limiting normal cone, the following directional version of the limiting subdifferential was introduced in [7].

Definition 2.2.3. *(directional subdifferentials; see [7]) Let $f : \mathcal{X} \to [-\infty, +\infty]$ and $\bar{x}$ be a point where f is finite. Then the limiting subdifferential of f at $\bar{x}$ in direction $(u, \zeta) \in \mathcal{X} \times \mathbb{R}$ is defined as*

$$\begin{aligned} \partial f(\bar{x}; (u, \zeta)) := \Big\{ \xi \in \mathcal{X} \Big| &\exists t_k \downarrow 0,\ u^k \to u,\ \zeta^k \to \zeta,\ \xi^k \to \xi,\ f(\bar{x}) + t_k \zeta^k = f(\bar{x} + t_k u^k), \\ &\xi^k \in \widehat{\partial} f(\bar{x} + t_k u^k) \Big\}. \end{aligned}$$

Remark 2.2.1. *Let f be continuously differentiable at $\bar{x}$. Then $\partial f(\bar{x}; (u, \zeta)) \neq \emptyset$ if* *and only if* $\zeta = \nabla f(\bar{x})u$*, in which case*

$$\partial f(\bar{x};(u,\zeta)) = \partial f(\bar{x}) = \{\nabla f(\bar{x})\}.$$

Definition 2.2.4. *(graphical derivatives; see, e.g., [14]) For a set-valued map* $G : \mathcal{X} \rightrightarrows \mathcal{Y}$ *and a pair* (x,y) *with* $y \in G(x)$*, the graphical derivative of* G *at* x *for* y *is the set-valued map* $DG(x|y) : \mathcal{X} \rightrightarrows \mathcal{Y}$ *whose graph is the tangent cone to gphG at* (x,y)*:*

$$v \in DG(x|y)(u) \Leftrightarrow (u,v) \in T_{gphG}(x,y).$$

Thus, $v \in DG(x|y)(u)$ *if and only if there exist sequences* $u_k \to u,\ v_k \to v$ *and* $\tau_k \downarrow 0$ *such that* $y + \tau_k v_k \in G(x + \tau_k u_k)$ *for all* k*.*

For a single-valued mapping $P : \mathcal{X} \to \mathcal{Y}$, its graphical derivative at x for $y = P(x)$ is

$$DP(x)(u) := \left\{ \xi \middle| \exists t_k \downarrow 0, u_k \to u \text{ s.t.} \lim_{k\to+\infty} \frac{P(x+t_k u_k) - P(x)}{t_k} = \xi \right\}. \tag{2.4}$$

Moreover if $P(x)$ is Hadamard directionally differentiable at x, then its graphical derivative is equal to the directional derivative: for any $u \in \mathcal{X}$,

$$DP(x)(u) = P'(x;u) := \lim_{t\downarrow 0, u'\to u} \frac{P(x+tu') - P(x)}{t}.$$

The following sum rule extends the sum rule in [14, Proposition 4A.2] by allowing $P(x)$ to be only continuous.

Proposition 2.2.2. *Let* $G(x) := -P(x) + \Lambda$ *and* $P(\bar{x}) \in \Lambda$*, where* $P(x) : \mathcal{X} \to \mathcal{Y}$ *is a continuous singled-valued map.*

Then either

$$DG(\bar{x}|0)(u) \subseteq -DP(\bar{x})(u) + T_\Lambda(P(\bar{x})) \tag{2.5}$$

or there exists $\zeta \neq 0$ *such that*

$$\zeta \in DP(\bar{x})(0) \cap T_\Lambda(P(\bar{x})).$$

If either $P(x)$ *is Hadamard directionally differentiable at* $\bar{x}$ *or* Λ *is geometrically derivable, then (2.5) holds as an equality.*

Proof. By definition, $v \in DG(\bar{x}|0)(u)$ if and only if $(u, v) \in T_{gphG}(\bar{x}, 0)$. It follows from the definition of the tangent cone that there exist sequences $(u_k, v_k) \to (u, v)$ and $\tau_k \downarrow 0$ such that $(\bar{x}, 0) + \tau_k(u_k, v_k) \in gphG$, which means that there exists $s_k \in \Lambda$ such that $\tau_k v_k = -P(\bar{x} + \tau_k u_k) + s_k$.

Case (i) ($\{\frac{P(\bar{x}+\tau_k u_k)-P(\bar{x})}{\tau_k}\}$ is bounded). Then without loss of generality we may assume that $\lim_{k\to+\infty} \frac{P(\bar{x}+\tau_k u_k)-P(\bar{x})}{\tau_k} = \xi$. Therefore we have

$$v = \lim_{k\to+\infty} v_k = -\lim_{k\to+\infty} \frac{P(\bar{x} + \tau_k u_k) - P(\bar{x})}{\tau_k} + \lim_{k\to+\infty} \frac{s_k - P(\bar{x})}{\tau_k}.$$

Since $s_k \in \Lambda$, we have

$$\lim_{k\to+\infty} \frac{s_k - P(\bar{x})}{\tau_k} \in T_\Lambda(P(\bar{x})).$$

Hence $v \in -DP(\bar{x})(u) + T_\Lambda(P(\bar{x}))$.

Case (ii) ($\{\frac{P(\bar{x}+\tau_k u_k)-P(\bar{x})}{\tau_k}\}$ is unbounded). Without loss of generality, assume that

$$\lim_{k\to+\infty} \frac{\|P(\bar{x} + \tau_k u_k) - P(\bar{x})\|}{\tau_k} = \infty.$$

Define $t_k := \|P(\bar{x} + \tau_k u_k) - P(\bar{x})\|$.

Since

$$\left\{\frac{P(\bar{x} + \tau_k u_k) - P(\bar{x})}{t_k}\right\} = \left\{\frac{P(\bar{x} + \tau_k u_k) - P(\bar{x})}{\|P(\bar{x} + \tau_k u_k) - P(\bar{x})\|}\right\}$$

is bounded, we may without loss of generality assume $\lim_{k\to+\infty}\{\frac{P(\bar{x}+\tau_k u_k)-P(\bar{x})}{t_k}\} = \zeta$. By definition of $DP(\bar{x})(0)$ and the fact that $\lim_{k\to\infty} \frac{\tau_k}{t_k} = 0$, we have

$$0 \neq \zeta = \lim_{k\to+\infty} \frac{P(\bar{x} + \tau_k u_k) - P(\bar{x})}{t_k} = \lim_{k\to+\infty} \frac{P(\bar{x} + t_k(\frac{\tau_k}{t_k} u_k)) - P(\bar{x})}{t_k} \in DP(\bar{x})(0).$$

Since $v_k \to v$ and $\lim_{k\to\infty} \frac{\tau_k}{t_k} = 0$, we have

$$\begin{aligned} 0 = \lim_{k\to\infty} \frac{\tau_k}{t_k} v_k &= -\lim_{k\to+\infty} \frac{P(\bar{x} + \tau_k u_k) - P(\bar{x})}{t_k} + \lim_{k\to+\infty} \frac{s_k - P(\bar{x})}{t_k} \\ &= -\zeta + \lim_{k\to+\infty} \frac{s_k - P(\bar{x})}{t_k}. \end{aligned}$$

Therefore $\zeta = \lim_{k\to+\infty} \frac{s_k - P(\bar{x})}{t_k}$ which implies that $\zeta \in T_\Lambda(P(\bar{x}))$.

Conversely, let $v \in -DP(\bar{x})(u) + T_\Lambda(P(\bar{x}))$. Then there exist $\xi \in DP(\bar{x})(u)$ and $\zeta \in T_\Lambda(P(\bar{x}))$ such that $v = -\xi + \zeta$.

If $P(x)$ is Hadamard directionally differentiable at $\bar{x}$, then the limit

$$\xi = \lim_{t\downarrow 0, u'\to u} \frac{P(\bar{x}+tu') - P(\bar{x})}{t}$$

exists and there exist sequences $\tau_k \downarrow 0,\ s_k \xrightarrow{\Lambda} P(\bar{x})$ such that

$$\zeta = \lim_{k\to+\infty} \frac{s_k - P(\bar{x})}{\tau_k}.$$

Define

$$v_k = -\frac{P(\bar{x}+\tau_k u_k) - P(\bar{x})}{\tau_k} + \frac{s_k - P(\bar{x})}{\tau_k} = \frac{-P(\bar{x}+\tau_k u_k) + s_k}{\tau_k}.$$

Then $\lim_{k\to\infty} v_k = v$ and $\tau_k v_k \in -P(\bar{x}+\tau_k u_k) + \Lambda$ for all k. Hence $v \in DG(\bar{x}|0)(u)$.

Now suppose that Λ is geometrically derivable. Let $\tau_k \downarrow 0,\ u_k \to u$ be sequences such that

$$\xi = \lim_{k\downarrow\infty} \frac{P(\bar{x}+\tau_k u_k) - P(\bar{x})}{\tau_k}.$$

Since Λ is geometrically derivable , there exists $s_k \in \Lambda$ such that

$$\zeta = \lim_{k\to+\infty} \frac{s_k - P(\bar{x})}{\tau_k}.$$

Define

$$v_k = -\frac{P(\bar{x}+\tau_k u_k) - P(\bar{x})}{\tau_k} + \frac{s_k - P(\bar{x})}{\tau_k} = \frac{-P(\bar{x}+\tau_k u_k) + s_k}{\tau_k}.$$

Then $\lim_{k\to\infty} v_k = v$ and $\tau_k v_k \in G(\bar{x}+\tau_k u_k)$ for all k. Hence $v \in DG(\bar{x}|0)(u)$. ■

Definition 2.2.5. *(coderivatives and directional coderivatives; see [7] and [59, Definition 1.32]) For a set-valued map $G : \mathcal{X} \rightrightarrows \mathcal{Y}$ and a point $(\bar{x},\bar{y}) \in gphG := \{(x,y) \in \mathcal{X}\times\mathcal{Y} | y \in G(x)\}$, the Fréchet coderivative (precoderivative) of G at $(\bar{x},\bar{y})$ is a multifunction $\widehat{D}^*G(\bar{x},\bar{y}) : \mathcal{Y} \rightrightarrows \mathcal{X}$ defined as*

$$\widehat{D}^*G(\bar{x},\bar{y})(\zeta) := \left\{ \eta \in \mathcal{X} \middle| (\eta, -\zeta) \in \widehat{N}_{gphG}(\bar{x},\bar{y}) \right\};$$

the limiting (Mordukhovich) coderivative of G at $(\bar{x},\bar{y})$ is a multifunction $D^*G(\bar{x},\bar{y}) :$

$\mathcal{Y} \rightrightarrows \mathcal{X}$ defined as

$$D^*G(\bar{x},\bar{y})(\zeta) := \{\eta \in \mathcal{X} | (\eta, -\zeta) \in N_{gphG}(\bar{x},\bar{y})\}.$$

The symbol $D^*G(\bar{x})$ is used when G is single valued. The limiting coderivative of G at $(\bar{x},\bar{y})$ in direction $(u,\xi) \in \mathcal{X} \times \mathcal{Y}$ is defined as

$$D^*G(\bar{x},\bar{y};(u,\xi))(\zeta) := \{\eta \in \mathcal{X} | (\eta, -\zeta) \in N_{gphG}(\bar{x},\bar{y};(u,\xi))\}.$$

Similarly the symbol $D^*G(\bar{x};(u,\xi))$ is used when G is single valued.

Remark 2.2.2. *In the special case when $P : \mathcal{X} \to \mathcal{Y}$ is a single-valued map which is Lipschitz continuous at $\bar{x}$, by [59, Theorem 3.28], the coderivative is related to the limiting subdifferential in the following way:*

$$D^*P(\bar{x})(\zeta) = \partial \langle P, \zeta \rangle (\bar{x}) \quad \textit{for all } \zeta \in \mathcal{Y}.$$

By [7, Proposition 5.1], if P is Lipschitz near $\bar{x}$ in direction u, then $D^*P(\bar{x};(u,\xi))(\zeta) \neq \emptyset$ if and only if $\xi \in DP(\bar{x})(u)$, in which case

$$D^*P(\bar{x};(u,\xi))(\zeta) = \partial \langle P, \zeta \rangle (\bar{x};(u, \langle \xi, \zeta \rangle)).$$

Let $P : \mathcal{X} \to \mathcal{Y}$ be C^1. By [7, Remark 2.1], one has $DP(\bar{x})(u) = \nabla P(\bar{x})u$ and thus $D^*P(\bar{x};(u,\xi))(\zeta) \neq \emptyset$ if and only if $\xi = \nabla P(\bar{x})u$, in which case

$$D^*P(\bar{x};(u,\xi))(\zeta) = D^*P(\bar{x})(\zeta) = \nabla P(\bar{x})^*\zeta.$$

To state our main results, given $P : \mathcal{X} \to \mathcal{Y}$ and $\Lambda \subseteq \mathcal{Y}$, we define the extended linearized cone as

$$\widetilde{\mathcal{L}}(x) := \{(u,\xi) \in \mathcal{X} \times \mathcal{Y} | \xi \in DP(x)(u) \cap T_\Lambda(P(x))\}. \tag{2.6}$$

It is easy to see that the projection of $\widetilde{\mathcal{L}}(x)$ onto the space $\mathcal{X}$ is the linearized cone defined by $\mathcal{L}(x) := \{u \in \mathcal{X} | \exists \xi \text{ s.t. } \xi \in DP(x)(u) \cap T_\Lambda(P(x))\}$. When P is differentiable at x, $DP(x)(u) = \nabla P(x)u$ and hence in this case

$$\widetilde{\mathcal{L}}(x) = \{(u, \nabla P(x)u) : 0 \in -\nabla P(x)u + T_\Lambda(P(x))\}.$$

Proposition 2.2.3. *Let $P : \mathcal{X} \to \mathcal{Y}$ be continuous and $\Lambda \subseteq \mathcal{Y}$. Then*

$$\widetilde{\mathcal{L}}(\bar{x}) = \{(0,0)\} \implies DG(\bar{x}|0)^{-1}(0) = \{0\}. \tag{2.7}$$

Proof. By virtue of Proposition 2.2.2, when $\widetilde{\mathcal{L}}(\bar{x}) = \{(0,0)\}$, one must have

$$DG(\bar{x}|0)(u) \subseteq -DP(\bar{x})(u) + T_\Lambda(P(\bar{x})).$$

Suppose that $u \in DG(\bar{x}|0)^{-1}(0)$. Then equivalently, $0 \in DG(\bar{x}|0)(u)$. Hence $0 \in -DP(\bar{x})(u) + T_\Lambda(P(\bar{x}))$ or equivalently $DP(x)(u) \cap T_\Lambda(P(x)) \neq \emptyset$. Since $\widetilde{\mathcal{L}}(\bar{x}) = \{(0,0)\}$, it means that $\forall u \neq 0,\ DP(x)(u) \cap T_\Lambda(P(x)) = \emptyset$. Hence we must have $u = 0$. ■

Proposition 2.2.4. *Let $P : \mathcal{X} \to \mathcal{Y}$ be continuous and $\Lambda \subseteq \mathcal{Y}$ be closed near $\bar{x} \in \mathcal{X}$. If $\widetilde{\mathcal{L}}(\bar{x}) = \{(0,0)\}$, then $G(x) = P(x) - \Lambda$ is strongly metrically subregular at $(\bar{x}, 0)$.*

Proof. By [14, Theorem 4C.1], G is strongly metrically subregular at $(\bar{x}, 0)$ if and only if $DG(\bar{x}|0)^{-1}(0) = \{0\}$. The result then follows from applying Proposition 2.2.3. ■

2.3 Weak sufficient condition for metric subregularity

In this section we will derive a sufficient condition for metric subregularity of the system $P(x) \in \Lambda$ where $P(x) : \mathcal{X} \to \mathcal{Y}$ is a continuous single-valued map and $\Lambda \subseteq \mathcal{Y}$ is locally closed. Recall that no ζ satisfying condition (2.10) alone is the FOSCMS, as established by Gfrerer and Klatte in [25, Corollary 1] for the case where P is smooth and extended to the nonsmooth but calmness case in [7, Proposition 2.2]. Our sufficient condition in Theorem 2.3.1 improves the FOSCMS in [7, Proposition 2.2] in two aspects. First, we allow $P(x)$ to be only continuous instead of being calm. Secondly, even in the case where $P(x)$ is calm, our condition is weaker in that the extra condition of the existence of sequences $(u_k, v_k, \zeta_k) \to (u, 0, \zeta)$ and $t_k \downarrow 0$ satisfying (2.11) and (2.12) is required.

We will derive our result based on the following sufficient conditions for metric subregularity for general set-valued maps by Gfrerer in [23].

Lemma 2.3.1. *(see [23, Corollary 1 and Remarks 1 and 2]) Let $G : \mathcal{X} \rightrightarrows \mathcal{Y}$ be a closed set-valued map, and take a point $(\bar{x}, \bar{y}) \in gphG$. Assume that for any direction $u \in \mathcal{X}$, there do not exist sequences $t_k \downarrow 0$, $\|(u_k, v_k)\| = 1$, $\|y_k^*\| = 1$ with $\|u_k\| \to 1$, $\|u\|u_k \to u$, $v_k \to 0$, $x_k^* \to 0$ satisfying*

$$(x_k^*, -y_k^*) \in \widehat{N}_{gphG}(x_k', y_k'),\ x_k' \notin G^{-1}(\bar{y})$$

and

$$\lim_{k\to\infty} \frac{\langle y_k^*, y_k' - \bar{y}\rangle}{\|y_k' - \bar{y}\|} = 1,$$

where $x_k' := \bar{x} + t_k u_k \neq \bar{x}$, $y_k' := \bar{y} + t_k v_k \neq \bar{y}$. Then G is metrically subregular at $(\bar{x}, \bar{y})$.

Note that as commented in [23, Remark 2], if the condition $x_k' \notin G^{-1}(\bar{y})$ is omitted then the resulting sufficient condition is stronger but may be easier to verify. However, in [61, Example 1], it was shown that sometimes these kinds of conditions can not be omitted in order to show the metric subregularity.

Lemma 2.3.2. *Let P be a single-valued map from $\mathcal{X}$ to $\mathcal{Y}$ and Λ be a subset of $\mathcal{Y}$. Define $G(x) := P(x) - \Lambda$, $y = P(x) - s$ for some $s \in \Lambda$. Then $(x^*, -y^*) \in \widehat{N}_{gphG}(x, y)$ implies that*

$$x^* \in \widehat{D}^* P(x)(y^*), \quad y^* \in \widehat{N}_\Lambda(P(x) - y).$$

Proof. Since $(x^*, -y^*) \in \widehat{N}_{gphG}(x, y)$, by definition for any $\epsilon > 0$,

$$\langle x^*, x' - x\rangle + \langle -y^*, y' - y\rangle \leq \epsilon\|(x' - x, y' - y)\| \tag{2.8}$$

for any $(x', y') \in gphG$ which is sufficiently close to (x, y). Let $y' := P(x) - s'$, $s' \in \Lambda$. Then when s' is close to s, $y' = P(x) - s'$ is close to $y = P(x) - s$. Hence, fixing $x' = x$ in (2.8) we obtain that for any $\epsilon > 0$ and any $s' \in \Lambda$ sufficiently close to s,

$$\langle -y^*, s - s'\rangle \leq \epsilon\|s - s'\| \Leftrightarrow \langle y^*, s' - s\rangle \leq \epsilon\|s - s'\|.$$

This means that $y^* \in \widehat{N}_\Lambda(s) = \widehat{N}_\Lambda(P(x) - y)$.

On the other hand, let $x' \in \mathcal{X}$ and $y' := P(x') - s$. Then $y' \in G(x')$ and when $(x', P(x'))$ is close to $(x, P(x))$, (x', y') is close to (x, y). Hence, by (2.8) we have

$$\langle x^*, x' - x\rangle + \langle -y^*, P(x') - P(x)\rangle \leq \epsilon\|(x' - x, P(x') - P(x))\|,$$

for any $(x', P(x'))$ which is close to $(x, P(x))$. This means that

$$(x^*, -y^*) \in \widehat{N}_{gphP}(x, P(x))$$

or equivalently $x^* \in \widehat{D}^* P(x)(y^*)$. The proof of the lemma is therefore complete. ■

Applying Lemmas 2.3.1 and 2.3.2, we obtain the following sufficient condition for metric subregularity.

Proposition 2.3.1. *Let $P : \mathcal{X} \to \mathcal{Y}$ be a single-valued map and $\Lambda \subseteq \mathcal{Y}$ be closed. Let $G(x) := P(x) - \Lambda$ and $P(\bar{x}) \in \Lambda$. Assume that $G(x)$ is a set-valued map which is closed around $\bar{x}$ and suppose that for any direction $u \in \mathcal{X}$, there do not exist sequences $t_k \downarrow 0$, $\|(u_k, v_k)\| = 1$, $\|y_k^*\| = 1$ with $\|u_k\| \to 1$, $\|u\| u_k \to u$, $v_k \to 0$, $x_k^* \to 0$ satisfying*

$$x_k^* \in \widehat{D}^* P(\bar{x} + t_k u_k)(y_k^*), \quad y_k^* \in \widehat{N}_\Lambda(P(\bar{x} + t_k u_k) - t_k v_k), \quad P(\bar{x} + t_k u_k) \notin \Lambda$$

and

$$\lim_{k \to \infty} \frac{\langle y_k^*, v_k \rangle}{\|v_k\|} = 1. \tag{2.9}$$

Then G is metrically subregular at $(\bar{x}, 0)$.

Note that by [59, Theorem 1.38], when P is Fréchet differentiable but not necessarily Lipschitz continuous, we have $\widehat{D}^* P(x)(y^*) = \{\nabla P(x)^* y^*\}$.

Theorem 2.3.1. *Let $P : \mathcal{X} \to \mathcal{Y}$ be continuous and $\Lambda \subseteq \mathcal{Y}$ be closed at $\bar{x} \in \mathcal{X}$. Suppose that the weak sufficient condition for metric subregularity (WSCMS) holds at $\bar{x}$, i.e., for all $(0, 0) \neq (u, \xi) \in \widetilde{\mathcal{L}}(\bar{x})$, there exists no unit vector ζ, sequences $(u_k, v_k, \zeta_k) \to (u, 0, \zeta)$ and $t_k \downarrow 0$ satisfying*

$$0 \in D^* P(\bar{x}; (u, \xi))(\zeta), \quad \zeta \in N_\Lambda(P(\bar{x}); \xi), \tag{2.10}$$

$$\zeta_k \in \widehat{N}_\Lambda(s_k), \ s_k = P(\bar{x} + t_k u_k) - t_k v_k, \ P(\bar{x} + t_k u_k) \notin \Lambda, \tag{2.11}$$

$$\lim_{k \to \infty} \langle \zeta_k, \frac{v_k}{\|v_k\|} \rangle = 1. \tag{2.12}$$

Then $G(x) = P(x) - \Lambda$ is metrically subregular at $(\bar{x}, 0)$.

Proof. If $\widetilde{\mathcal{L}}(\bar{x}) = \{(0, 0)\}$, then by Proposition 2.2.4, G is strongly metrically subregular and hence metrically subregular at $(\bar{x}, 0)$. We now prove the result for the

$\widetilde{\mathcal{L}}(\bar{x}) \neq \{(0,0)\}$ case by contradiction. To the contrary, suppose that $P(x) - \Lambda$ is not metrically subregular at $(\bar{x}, 0)$. By Proposition 2.3.1, there exist $u \in \mathcal{X}$ and sequences $t_k \downarrow 0,\ \|(u_k, v_k)\| = 1,\ \|y_k^*\| = 1$ with $\|u_k\| \to 1,\ \|u\|u_k \to u, v_k \to 0,\ x_k^* \to 0$, such that

$$(x_k^*, -y_k^*) \in \widehat{N}_{gphP}(\bar{x} + t_k u_k, P(\bar{x} + t_k u_k)), y_k^* \in \widehat{N}_\Lambda(P(\bar{x} + t_k u_k) - t_k v_k) \tag{2.13}$$

and (3.45) holds.

Since we have $\|y_k^*\| = 1,\ \|(u_k, v_k)\| = 1$, and $v_k \to 0$, passing to a subsequence if necessary, we assume that $\lim_{k\to\infty} y_k^* = \zeta$, $\lim_{k\to\infty} u_k = u$ for certain $\|u\| = 1$. It follows that $\|\zeta\| = 1$.

Case (1) ($\{\frac{P(\bar{x}+t_k u_k) - P(\bar{x})}{t_k}\}$ is bounded). Then without loss of generality we may assume that $\lim_{k\to+\infty} \frac{P(\bar{x}+t_k u_k)-P(\bar{x})}{t_k} = \xi$. Thus letting $\xi_k := \frac{P(\bar{x}+t_k u_k)-P(\bar{x})}{t_k}$, we have $P(\bar{x} + t_k u_k) = P(\bar{x}) + t_k \xi_k$. Combining with (2.13) we get

$$(x_k^*, -y_k^*) \in \widehat{N}_{gphP}((\bar{x}, P(\bar{x})) + t_k(u_k, \xi_k)), y_k^* \in \widehat{N}_\Lambda(P(\bar{x} + t_k u_k) - t_k v_k).$$

Since $(u_k, \xi_k) \to (u, \xi)$ as $k \to \infty$, we have

$$(0, -\zeta) \in N_{gphP}((\bar{x}, P(\bar{x})); (u, \xi)), \zeta \in N_\Lambda(P(\bar{x}); \xi).$$

Also from the proof of Proposition 2.2.2, we see that $\xi \in DP(\bar{x})(u) \cap T_\Lambda(P(\bar{x}))$ and hence $(u, \xi) \in \tilde{\mathcal{L}}(\bar{x})$.

In summary for *Case* (1), we have obtained a nonzero vector ζ, a nonzero vector $(u, \xi) \in \widetilde{L}(\bar{x})$, and sequences $(u_k, v_k, y_k^*) \to (u, 0, \zeta)$ and $t_k \downarrow 0$ such that

$$\begin{aligned}
&0 \in D^*P(\bar{x}; (u, \xi))(\zeta), \quad \zeta \in N_\Lambda(P(\bar{x}); \xi)\\
&y_k^* \in \widehat{N}_\Lambda(s_k), \quad s_k = P(\bar{x} + t_k u_k) - t_k v_k,\\
&\lim_{k\to\infty} \langle y_k^*, \frac{v_k}{\|v_k\|}\rangle = 1,
\end{aligned}$$

which contradicts the assumption in (WSCMS). Thus $P(x) - \Lambda$ is metrically subregular at $(\bar{x}, 0)$.

Case (2) ($\{\frac{P(\bar{x}+t_k u_k)-P(\bar{x})}{t_k}\}$ is unbounded). Without loss of generality, assume that

$\lim_{k\to+\infty}\frac{\|P(\bar{x}+t_ku_k)-P(\bar{x})\|}{t_k}=\infty$. Define

$$\tau_k := \|t_ku_k\| + \|P(\bar{x}+t_ku_k)-P(\bar{x})\|, \quad u'_k := \frac{t_ku_k}{\tau_k},$$
$$\xi_k := \tfrac{P(\bar{x}+\tau_ku'_k)-P(\bar{x})}{\tau_k}, \qquad v'_k := \frac{t_kv_k}{\tau_k}.$$

Since $t_k/\tau_k \le t_k/\|P(\bar{x}+t_ku_k)-P(\bar{x})\|$, we have $t_k/\tau_k \to 0$ and hence $v'_k \to 0$ and $u'_k \to 0$. Since $\{\xi_k\}$ is bounded, taking a subsequence if necessary , we have

$$\xi := \lim_{k\to\infty}\xi_k.$$

Then with $t_ku_k = \tau_ku'_k$ and $P(\bar{x}+t_ku_k) = P(\bar{x}) + \tau_k\xi_k$, by (2.13) we get

$$(x_k^*, -y_k^*) \in \widehat{N}_{gphP}((\bar{x}, P(\bar{x})) + \tau_k(u'_k, \xi_k)), y_k^* \in \widehat{N}_\Lambda(P(\bar{x}+\tau_ku'_k) - \tau_kv'_k).$$

Since $s_k = P(\bar{x}+\tau_ku'_k) - \tau_kv'_k$, we know that

$$\begin{aligned}\lim_{k\to\infty}\frac{s_k - P(\bar{x})}{\tau_k} &= \lim_{k\to\infty}\frac{P(\bar{x}+\tau_ku'_k) - \tau_kv'_k - P(\bar{x})}{\tau_k}\\ &= \lim_{k\to\infty}\frac{P(\bar{x}+\tau_ku'_k) - P(\bar{x})}{\tau_k} = \xi.\end{aligned}$$

Thus, $\xi \in DP(\bar{x})(0) \cap T_\Lambda(P(\bar{x}))$, which means $(0,\xi) \in \tilde{\mathcal{L}}(\bar{x})$. With $x_k^* \to 0$, we have

$$0 \in D^*P(\bar{x};(0,\xi))(\zeta), \quad \zeta \in N_\Lambda(P(\bar{x});\xi).$$

By (3.45), we can easily obtain that

$$\lim_{k\to\infty}\langle y_k^*, \frac{v'_k}{\|v'_k\|}\rangle = \lim_{k\to\infty}\langle y_k^*, \frac{t_kv_k}{\|t_kv_k\|}\rangle = \lim_{k\to\infty}\langle y_k^*, \frac{v_k}{\|v_k\|}\rangle = 1.$$

In summary for *Case* (2), we obtain a nonzero vector ζ, a nonzero vector $(0,\xi) \in \tilde{\mathcal{L}}(\bar{x})$, and sequences $(u'_k, v'_k, y_k^*) \to (0,0,\zeta)$ and $\tau_k \downarrow 0$ such that

$$\begin{aligned}&0 \in D^*P(\bar{x};(0,\xi))(\zeta), \quad \zeta \in N_\Lambda(P(\bar{x});\xi)\\ &y_k^* \in \hat{N}_\Lambda(s_k), \quad s_k = P(\bar{x}+\tau_ku'_k) - \tau_kv'_k,\\ &\lim_{k\to\infty}\langle y_k^*, \frac{v'_k}{\|v'_k\|}\rangle = 1,\end{aligned}$$

which contradicts the assumption in (WSCMS). Thus $P(x) - \Lambda$ is metrically subregular at $(\bar{x}, 0)$. $\blacksquare$

As an immediate consequence, if we discard the sequential conditions (2.11) and (2.12) in WSCMS, we derive from Theorem 2.3.1 the following sufficient condition for metric subregularity in the form of FOSCMS. The result improves [7, Proposition 2.2] in that P is only assumed to be continuous instead of being calm.

Corollary 2.3.1. *Let $P : \mathcal{X} \to \mathcal{Y}$ be continuous and $\Lambda \subseteq \mathcal{Y}$ be closed at $\bar{x} \in \mathcal{X}$. Suppose that FOSCMS holds at $\bar{x}$, i.e., for all (u, ξ) such that $\xi \in DP(\bar{x})(u) \cap T_\Lambda(P(\bar{x}))$,*

$$0 \in D^*P(\bar{x}; (u, \xi))(\zeta), \quad \zeta \in N_\Lambda(P(\bar{x}); \xi) \quad \Longrightarrow \quad \zeta = 0.$$

Then $G(x) = P(x) - \Lambda$ is metrically subregular at $(\bar{x}, 0)$.

2.4 Directional quasi-/pseudo- normality

As we mentioned in the introduction, quasi-/pseudo-normality are also sufficient for metric subregularity. In this section we propose directional versions of the quasi-/pseudo-normality and show that they are slightly stronger than the WSCMS. Moreover we show that the SOSCMS implies pseudo-normality. Our results are based on the following observations.

Proposition 2.4.1. *Let $P : \mathcal{X} \to \mathcal{Y}$, $(u^k, v^k, \zeta^k) \to (u, 0, \zeta)$, $\ t_k \downarrow 0$ with $\|\zeta\| = 1$, and $s^k = P(\bar{x} + t_k u^k) - t_k v^k$. Then the condition*

$$\lim_{k\to\infty} \langle \zeta^k, \frac{v^k}{\|v^k\|} \rangle = 1 \tag{2.14}$$

implies

$$\zeta_i (P_i(\bar{x} + t_k u^k) - s_i^k) > 0, \forall i \in I := \{i : \zeta_i \neq 0\} \tag{2.15}$$

which implies

$$\langle \zeta, P(\bar{x} + t_k u^k) - s^k \rangle > 0. \tag{2.16}$$

Proof. Suppose that (2.14) holds. Since

$$\begin{aligned} &\left\| \frac{\zeta^k}{\|\zeta^k\|} - \frac{v^k}{\|v^k\|} \right\|^2 \\ &= \langle \frac{\zeta^k}{\|\zeta^k\|} - \frac{v^k}{\|v^k\|}, \frac{\zeta^k}{\|\zeta^k\|} - \frac{v^k}{\|v^k\|} \rangle \\ &= \frac{\|\zeta^k\|^2}{\|\zeta^k\|^2} - 2\langle \frac{\zeta^k}{\|\zeta^k\|}, \frac{v^k}{\|v^k\|} \rangle + \frac{\|v^k\|^2}{\|v^k\|^2} \\ &= 2 - \frac{2}{\|\zeta^k\|} \langle \zeta^k, \frac{v^k}{\|v^k\|} \rangle, \end{aligned}$$

$\lim_{k\to\infty} \langle \zeta^k, \frac{v^k}{\|v^k\|} \rangle = 1$ and $\lim_{k\to\infty} \|\zeta^k\| = \|\zeta\| = 1$, we have

$$\lim_{k\to\infty} \left\| \frac{\zeta^k}{\|\zeta^k\|} - \frac{v^k}{\|v^k\|} \right\| = 0.$$

Consequently, $\lim_{k\to\infty} \frac{v^k}{\|v^k\|} = \frac{\zeta}{\|\zeta\|}$. Thus when k is large enough, for each $i = 1, \ldots, m$ with $\zeta_i \neq 0$, v_i^k has the same sign as ζ_i. This means

$$\zeta_i v_i^k > 0 \quad \forall i \in I := \{i : \zeta_i \neq 0\},$$

which implies (2.15). Since $\zeta \neq 0$, (2.15) obviously implies (2.16). ■

We are now in a position to define the concept of directional quasi/pseudo-normality.

Definition 2.4.1 (Directional quasi-/pseudo-normality)**.** *Let $P : \mathcal{X} \to \mathcal{Y}$ with $P(\bar{x}) \in \Lambda$.*

(a) We say that directional quasi-normality holds at $\bar{x}$ if for all

$$(0, 0) \neq (u, \xi) \in \widetilde{\mathcal{L}}(\bar{x}) := \{(u, \xi) \in \mathcal{X} \times \mathcal{Y} | \xi \in DP(x)(u) \cap T_\Lambda(P(x))\},$$

there exists no $\zeta \neq 0$ such that

$$0 \in D^*P(\bar{x}; (u, \xi))(\zeta), \quad \zeta \in N_\Lambda(P(\bar{x}); \xi) \tag{2.17}$$

and

$$\begin{cases} \exists (u^k, s^k, \zeta^k) \to (u, P(\bar{x}), \zeta) and\ t_k \downarrow 0, \\ s.t.\ \zeta^k \in \widehat{N}_\Lambda(s^k)\ and\ \zeta_i(P_i(\bar{x}+t_k u^k) - s_i^k) > 0\ if\ \zeta_i \neq 0. \end{cases}$$

(b) *We say that directional pseudo-normality holds at $\bar{x}$ if for all $(0,0) \neq (u, \xi) \in \widetilde{\mathcal{L}}(\bar{x})$, there exists no $\zeta \neq 0$ such that* (2.17) *holds and*

$$\begin{cases} \exists (u^k, s^k, \zeta^k) \to (u, P(\bar{x}), \zeta) \quad and \quad t^k \downarrow 0, \\ s.t.\ \zeta^k \in \widehat{N}_\Lambda(s^k)\ and\ \langle \zeta, P(\bar{x}+t_k u^k) - s^k \rangle > 0. \end{cases}$$

By virtue of Proposition 2.4.1, directional pseudo-normality is stronger than directional quasi-normality. And consequently from Theorem 2.3.1, they can provide sufficient conditions for metric subregularity.

Corollary 2.4.1. *Let $P : \mathcal{X} \to \mathcal{Y}$, $P(\bar{x}) \in \Lambda$, where $P(x)$ is continuous at $\bar{x}$ and Λ is closed near $\bar{x}$. If either directional pseudo-normality or directional quasi-normality holds at $\bar{x}$, then the set-valued map $G(x) = P(x) - \Lambda$ is metrically subregular at $(\bar{x}, 0)$.*

By definition, directional quasi-/pseudo-normality is weaker than quasi-/pseudo-normality, the following example shows that it is weaker than both quasi-normality and FOSCMS.

Example 2.4.1 (FOSCMS fails but directional pseudo-normality holds)**.** *Consider the constraint system defined by $P(x) = (x, -x^2) \in \Lambda$, where*

$$\Lambda := \{(x,y) | y \leq 0\ or\ y \leq x\}.$$

The point $\bar{x} = 0$ is feasible since $(0,0) \in \Lambda$. We have

$$P(\bar{x}) = (0,0), \quad \nabla P(\bar{x}) = \begin{pmatrix} 1 \\ -2\bar{x} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, T_\Lambda(P(\bar{x})) = \Lambda$$

and the linearized cone $\mathcal{L}(\bar{x}) = \{u \in \mathbb{R} | 0 \in -\nabla P(\bar{x})u + T_\Lambda(P(\bar{x}))\} = \mathbb{R}$. Let $\bar{u} := -1 \in \mathcal{L}(\bar{x})$, $\zeta := (0,1)$ and $(x^k, y^k) = P(\bar{x}) + \frac{1}{k}\nabla P(\bar{x})\bar{u} = (-\frac{1}{k}, 0)$. Then $\nabla P(\bar{x})^T \zeta = 0$ and for each k, $\zeta \in N_\Lambda(x^k, y^k)$. Thus $\nabla P(\bar{x})^T \zeta = 0$ and $\zeta \in N_\Lambda(P(\bar{x}); \nabla P(\bar{x})\bar{u})$. Hence FOSCMS fails at $\bar{x}$.

However, we can prove that directional pseudo-normality holds at $\bar{x}$. We prove it by contradiction. Assume that directional pseudo-normality fails at $\bar{x}$. Then there ex-

ist $0 \neq u \in \mathcal{L}(\bar{x})$, $0 \neq \zeta \in N_\Lambda(P(\bar{x}); \nabla P(\bar{x})u)$, *and a sequence* $\{u^k, s^k, \zeta^k\}$ *converging to* $(u, P(\bar{x}), \zeta)$ *and* $t_k \downarrow 0$ *such that*

$$\nabla P(\bar{x})^T \zeta = 0, \ \sum_{i=1}^{2} \zeta_i (P_i(\bar{x} + t_k u^k) - s_i^k) > 0, \ \zeta^k \in \widehat{N}_\Lambda(s^k). \tag{2.18}$$

Solving $\nabla P(\bar{x})^T \zeta = 0$, *we obtain* $\zeta_1 = 0$. *Moreover since* $N_\Lambda(P(\bar{x})) = \{0\} \times \mathbb{R}_+ \cup \{(-r, r) | r \geq 0\}$, *we have* $\zeta \in \{0\} \times \mathbb{R}_{++}$. *Since* $\zeta^k \to \zeta$ *and* $\zeta^k \in \widehat{N}_\Lambda(s^k)$, *we must have* $\zeta^k \in \{0\} \times \mathbb{R}_{++}$ *and* $s^k \in \{0\} \times \mathbb{R}_+$. *Thus we obtain*

$$\sum_{i=1}^{2} \zeta_i (P_i(z^k) - s_i^k) = \zeta_2 (P_2(z^k) - s_2^k) = \lambda(-(z^k)^2 - s_2^k) \leq 0,$$

where $z^k := \bar{x} + t_k u^k$. *But this contradicts* (2.18). *Hence directional pseudo-normality holds at* $\bar{x}$.

We now consider the case where Λ is the union of finitely many convex polyhedral sets in $\mathcal{Y}$, i.e. $\Lambda := \bigcup_{i=1}^p \Lambda_i$, where

$$\Lambda_i := \{y \in \mathcal{Y} | \langle \lambda_{ij}, y \rangle \leq b_{ij}, \quad j = 1, \ldots, m_i\}, \quad i = 1, \ldots, p,$$

with $\lambda_{ij} \in \mathcal{Y}$, $b_{ij} \in \mathbb{R}$ for $j = 1, \ldots, m_i$, are convex polyhedral sets. As noted in the introduction, by Robinson's multifunction theory [70], we know that when P is affine and Λ is the union of finitely many convex polyhedral sets, the set-valued map G^{-1} is upper Lipschitz continuous and hence calm at each point of the graph. What is more, we now show that the pseudo-normality always holds. To our knowledge, this result has never been shown in the literature before.

The following results will be needed in the proof. For every $s \in \Lambda$, we denote by $p(s) := \{i = 1, \ldots, p | s \in \Lambda_i\}$ the index set of the convex polyhedral sets containing s. Then we have from [22] that

$$T_\Lambda(s) = \bigcup_{i \in p(s)} T_{\Lambda_i}(s), \qquad \widehat{N}_\Lambda(s) = \bigcap_{i \in p(s)} \widehat{N}_{\Lambda_i}(s). \tag{2.19}$$

Proposition 2.4.2. *Let* $P : \mathcal{X} \to \mathcal{Y}$. *Suppose that* $P(x)$ *is affine and* Λ *is the union of finitely many convex polyhedral sets defined as above. Then for any feasible point* $\bar{x}$ *satisfying* $P(\bar{x}) \in \Lambda$, *pseudo-normality holds.*

Proof. We prove the proposition by contradiction. Assume that pseudo-normality does not hold at $\bar{x}$. Then there exists $\zeta \neq 0$ such that

$$\begin{cases} 0 = \nabla P(\bar{x})^* \zeta, \quad \zeta \in N_\Lambda(P(\bar{x})), \\ \exists (x^k, s^k, \zeta^k) \to (\bar{x}, P(\bar{x}), \zeta) \\ \text{ s.t. } \zeta^k \in \widehat{N}_\Lambda(s^k), \ \langle \zeta, P(x^k) - s^k \rangle > 0. \end{cases}$$

As $s^k \to P(\bar{x})$ when $k \to \infty$ and $s^k \in \Lambda = \bigcup_{i=1}^p \Lambda_i$, by virtue of (2.19), taking a subsequence if necessary, there exists $i \in \{1, \dots, p\}$ such that for k sufficiently large, $s^k \in \Lambda_i$, $P(\bar{x}) \in \Lambda_i$, $\zeta^k \in N_{\Lambda_i}(s^k)$. Define $J(s^k) := \{j = 1, \dots, m_i | \langle \lambda_{ij}, s^k \rangle = b_{ij}\}$ and $J(P(\bar{x})) := \{j = 1, \dots, m_i | \langle \lambda_{ij}, P(\bar{x}) \rangle = b_{ij}\}$. Since $\zeta^k \neq 0$, s^k is not an interior point of Λ_i and hence the index set $J(s^k)$ is not empty. Since $s^k \to P(\bar{x})$, we have $J(s^k) \subseteq J(P(\bar{x}))$ when k is sufficiently large. Hence without loss of generality, we can find a nonempty set $J \subseteq J(P(\bar{x}))$ such that $J(s^k) \equiv J$ for all k large enough. Define $C := \{\lambda_{ij} | j \in J\}$. Then we have $\zeta^k \in cone(C)$, where

$$cone(C) := \left\{ \sum_{j \in J} c_j \lambda_{ij} | c_j \geq 0 \ \forall j \in J \right\}$$

denotes the conic hull of C. It follows that $\zeta \in cone(C)$. Since when k large enough, for each $j \in J$, $\langle \lambda_{ij}, P(\bar{x}) - s^k \rangle = b_{ij} - b_{ij} = 0$, we obtain $\langle \zeta, P(\bar{x}) - s^k \rangle = 0$. Thus for sufficiently large k, we have

$$\begin{aligned} 0 &> \langle \zeta, s^k - P(x^k) \rangle + \langle \zeta, P(\bar{x}) - s^k \rangle \\ &= \langle \zeta, P(\bar{x}) - P(x^k) \rangle \\ &= \langle \zeta, \nabla P(\bar{x})(\bar{x} - x^k) \rangle, \end{aligned}$$

which contradicts the condition that $0 = \nabla P(\bar{x})^* \zeta$. Thus pseudo-normality holds at $\bar{x}$. ■

For a single-valued mapping $P : \mathcal{X} \to \mathcal{Y}$ which is C^1 at $\bar{x}$ and $u \in \mathcal{Y}$, we define its second-order graphical derivative of $P(x)$ at $\bar{x}$ in direction u as

$$\begin{aligned} & D^2 P(\bar{x})(u) \\ := & \left\{ l \in \mathcal{Y} | \exists t_k \downarrow 0, \ u^k \to u \ s.t. \ l = \lim_{k \to \infty} \frac{P(\bar{x} + t_k u^k) - P(\bar{x}) - t_k \nabla P(\bar{x}) u^k}{\frac{1}{2} t_k^2} \right\}. \end{aligned}$$

In [22, Theorem 4.3], a second-order sufficient condition for metric subregularity (SOSCMS) is presented for a split system in Banach spaces where one of the system is metrically subregular. Specializing the result in [22, Theorem 4.3] to our system (1.6), we may conclude that if $P(x)$ is C^1 and directionally second-order differentiable, Λ is the union of finitely many convex polyhedral sets and SOSCMS as stated in Theorem 2.4.1 holds, then the system is directionally pseudo-normal. In Theorem 2.4.1, we extend this result to the case where $P(x)$ is C^1 and $\nabla P(x)$ is directionally calm at $\bar{x}$ in each nonzero direction u lying in the linearization cone which means that there exist positive numbers $\epsilon,\ \delta,\ L_u$ such that

$$\|\nabla P(\bar{x}+tu') - \nabla P(\bar{x})\| \leq L_u \|tu'\| \quad \text{for all } 0 < t < \epsilon, \|u'-u\| < \delta.$$

Moreover we show that SOSCMS implies directional pseudo-normality.

Theorem 2.4.1. *Let $P(\bar{x}) \in \Lambda$, where $P(x)$ is C^1, Λ is the union of finitely many convex polyhedral sets in $\mathcal{Y}$ and $\nabla P(x)$ is directionally calm at $\bar{x}$ in each direction $0 \neq u$ such that $\nabla P(\bar{x})u \in T_\Lambda(P(\bar{x}))$. Suppose SOSCMS holds at $\bar{x}$, i.e., for all $0 \neq u$ such that $\nabla P(\bar{x})u \in T_\Lambda(P(\bar{x}))$, there exists no $\zeta \neq 0$ such that*

$$\begin{cases} \nabla P(\bar{x})^*\zeta = 0,\ \ \zeta \in N_\Lambda(P(\bar{x}); \nabla P(\bar{x})u), \\ \exists l \in D^2 P(\bar{x})(u)\ \ s.t.\ \langle \zeta, l\rangle > 0. \end{cases}$$

Then $\bar{x}$ is directionally pseudo-normal at $\bar{x}$.

Proof. We prove that SOSCMS is stronger than directional pseudo-normality by contradiction. Assume there exist $0 \neq u$ such that $\nabla P(\bar{x})u \in T_\Lambda(P(\bar{x}))$ and $\zeta \neq 0$ such that

$$\begin{cases} \nabla P(\bar{x})^*\zeta = 0, \quad \zeta \in N_\Lambda(P(\bar{x}); \nabla P(\bar{x})u) \\ \exists (u^k, s^k, \zeta^k) \to (u, P(\bar{x}), \zeta) \text{ and } t_k \downarrow 0 \\ \text{ s.t. } \zeta^k \in \widehat{N}_\Lambda(s^k),\ \sum_{i=1}^m \zeta_i (P_i(\bar{x}+t_k u^k) - s_i^k) > 0. \end{cases}$$

Notice that $\langle P(x), e_j\rangle$, where e_j is in the orthogonal basis $\mathscr{E}$, is a function on $\mathcal{X}$. By the mean value theorem, for each j and k, there exist $0 < c_j^k < t_k$ such that

$$\langle P(\bar{x}+t_k u^k) - P(\bar{x}), e_j\rangle = \langle \nabla P(\bar{x}+c_j^k u^k)(\bar{x}+t_k u^k - \bar{x}), e_j\rangle = \langle \nabla P(\bar{x}+c_j^k u^k) t_k u^k, e_j\rangle.$$

Thus

$$
\begin{aligned}
&\left\langle \frac{P(\bar{x}+t_k u^k)-P(\bar{x})-t_k\nabla P(\bar{x})u^k}{\frac{1}{2}t_k^2}, e_j \right\rangle \\
&\quad = \frac{1}{2t_k}\left(\frac{\langle P(\bar{x}+t_k u^k)-P(\bar{x}), e_j\rangle}{t_k} - \langle \nabla P(\bar{x})u^k, e_j\rangle \right) \\
&\quad = \frac{2}{t_k}\left(\frac{\langle \nabla P(\bar{x}+c_j^k u^k)u^k, e_j\rangle t_k}{t_k} - \langle \nabla P(\bar{x})u^k, e_j\rangle \right) \\
&\quad = \frac{2}{t_k}(\langle \nabla P(\bar{x}+c_j^k u^k)u^k, e_j\rangle - \langle \nabla P(\bar{x})u^k, e_j\rangle).
\end{aligned}
$$

Since $\nabla P(x)$ is directionally calm at $\bar{x}$ in direction u, there exists $L_u > 0$ such that for each j and sufficiently large k,

$$
\begin{aligned}
&\left\| \frac{2}{t_k}(\langle \nabla P(\bar{x}+c_j^k u^k)u^k, e_j\rangle - \langle \nabla P(\bar{x})u^k, e_j\rangle) \right\| \\
&\quad \leq \frac{2L_u\|\bar{x}+c_j^k u^k-\bar{x}\|\|u^k\|}{t_k} \\
&\quad \leq \frac{2L_u t_k\|u^k\|^2}{t_k} = 2L_u\|u^k\|^2.
\end{aligned}
$$

This implies that the sequence $\{\langle \frac{P(\bar{x}+t_k u^k)-P(\bar{x})-t_k\nabla P(\bar{x})u^k}{\frac{1}{2}t_k^2}, e_j\rangle\}$ is bounded. Consequently, the sequence $\{\frac{P(\bar{x}+t_k u^k)-P(\bar{x})-t_k\nabla P(\bar{x})u^k}{\frac{1}{2}t_k^2}\}$ is bounded. Taking a subsequence if necessary, there exists l such that

$$
l := \lim_{k\to\infty} \frac{P(\bar{x}+t_k u^k)-P(\bar{x})-t_k\nabla P(\bar{x})u^k}{\frac{1}{2}t_k^2} \in D^2P(\bar{x})(u).
$$

It follows that

$$
\begin{aligned}
0 \quad &< \langle \zeta, P(\bar{x}+t_k u^k)-s^k\rangle \\
&= \langle \zeta, P(\bar{x}+t_k u^k)-P(\bar{x})+P(\bar{x})-s^k\rangle \\
&= \langle \zeta, t_k\nabla P(\bar{x})u^k + \frac{t_k^2}{2}l + o(t_k^2)\rangle + \langle \zeta, P(\bar{x})-s^k\rangle. \qquad (2.20)
\end{aligned}
$$

By assumption, $\nabla P(\bar{x})^*\zeta = 0$, which means $\langle \zeta, t_k\nabla P(\bar{x})u^k\rangle = 0$. And since $s^k \to P(\bar{x})$ as $k\to\infty$, taking a subsequence if necessary, there exists $j\in\{1,\ldots,p\}$ such that for k sufficiently large, $s^k\in\Lambda_j$, $P(\bar{x})\in\Lambda_j$, $\zeta^k\in N_{\Lambda_j}(s^k)$. Since Λ_j is convex polyhedral, similar to the discussion in the proof of Proposition 2.4.2, we have

$\langle \zeta, P(\bar{x}) - s^k \rangle = 0$. Thus for k large enough, by (2.20) we have

$$\begin{aligned} 0 &< \langle \zeta, t_k \nabla P(\bar{x}) u^k + \frac{t_k^2}{2} l + o(t_k^2) \rangle + \langle \zeta, P(\bar{x}) - s^k \rangle \\ &\leq \frac{t_k^2}{2} \langle \zeta, l + o(1) \rangle. \end{aligned}$$

Then we obtain that $\exists l \in D^2 P(\bar{x})(u)$ such that $\langle \zeta, l \rangle \geq 0$. But this contradicts the assumption of the SOSCMS. The contradiction proves that the SOSCMS implies directional pseudo-normality. ■

Since directional calmness is obviously weaker than calmness, the following corollary follows immediately from Theorem 2.4.1. We say that $P(x)$ is $C^{1,c}$ at $\bar{x}$ if $P(x)$ is C^1 at $\bar{x}$ and $\nabla P(x)$ is calm at $\bar{x}$, i.e., there exist $\kappa > 0$ and a neighborhood U of $\bar{x}$ such that $\|\nabla P(x) - \nabla P(\bar{x})\| \leq \kappa \|x - \bar{x}\|$ for all $x \in U$.

Corollary 2.4.2. *Let $P(\bar{x}) \in \Lambda$, where P is $C^{1,c}$ and Λ is the union of finitely many convex polyhedral sets in $\mathcal{Y}$. Suppose SOSCMS holds at $\bar{x}$. Then $\bar{x}$ is directionally pseudo-normal.*

In summary, we have shown the following implications:

$$\begin{aligned} &SOSCMS \implies \text{directional pseudo-normality} \implies \text{directional quasi-normality} \\ &\implies \text{WSCMS} \implies \text{metric subregularity.} \end{aligned}$$

2.5 Applications to complementarity and KKT systems

In this section we apply our results to complementarity and KKT systems. When directional quasi-/pseudo-normality are applied to these systems we derive expressions that are much simpler and moreover can be directly compared with classical quasi/pseudo-normality.

First we consider the complementarity system formulated as follows:

$$\text{(CS)} \qquad H(x) = 0, \quad 0 \leq \Phi(x) \perp \Psi(x) \geq 0, \tag{2.21}$$

where $H(x) : \mathbb{R}^n \to \mathbb{R}^d$, Φ, $\Psi : \mathbb{R}^n \to \mathbb{R}^m$. For simplicity of explanation, we omit possible inequality and abstract constraints and moreover we assume that all functions are continuously differentiable. The results can be extended to the general case in a straightforward manner.

Define $\Omega_{EC} := \{(a,b) \in \mathbb{R}_+ \times \mathbb{R}_+ | ab = 0\}$. For any set C and any positive integer m we denote by C^m the m-Cartesian product of C. (CS) can be rewritten as

$$(H(x), (\Phi_1(x), \Psi_1(x)), \ldots, (\Phi_m(x), \Psi_m(x))) \in \{0\}^d \times \Omega_{EC}^m.$$

To derive the precise form of the directional quasi-/pseudo-normality, we review the formulas for the regular normal cone, the limiting normal cone, the tangent cone and the directional limiting normal cone of the set Ω_{EC}.

Lemma 2.5.1. *[24, Lemma 4.1] The Fréchet normal cone to Ω_{EC} is*

$$\widehat{N}_{\Omega_{EC}}(a,b) = \left\{ -(\gamma,\nu) \middle| \begin{array}{ll} \nu = 0, & \text{if } 0 = a < b \\ \gamma \geq 0, \nu \geq 0, & \text{if } a = b = 0 \\ \gamma = 0, & \text{if } a > b = 0 \end{array} \right\},$$

the limiting normal cone is

$$N_{\Omega_{EC}}(a,b) = \left\{ \begin{array}{ll} \widehat{N}_{\Omega_{EC}}(a,b) & \text{if } (a,b) \neq (0,0) \\ \{-(\gamma,\nu) | \text{either } \gamma > 0, \nu > 0 \text{ or } \ \gamma\nu = 0\}, & \text{if } (a,b) = (0,0) \end{array} \right\},$$

and the tangent cone is

$$T_{\Omega_{EC}}(a,b) = \left\{ (d_1,d_2) \middle| \begin{array}{ll} d_1 = 0, & \text{if } 0 = a < b \\ (d_1,d_2) \in \Omega_{EC}, & \text{if } a = b = 0 \\ d_2 = 0, & \text{if } a > b = 0 \end{array} \right\}.$$

For all $d = (d_1,d_2) \in T_{\Omega_{EC}}(a,b)$, the directional limiting normal cone to Ω_{EC} in direction d is

$$N_{\Omega_{EC}}((a,b);d) = \left\{ \begin{array}{ll} N_{\Omega_{EC}}(a,b) & \text{if } (a,b) \neq (0,0) \\ N_{\Omega_{EC}}(d_1,d_2), & \text{if } (a,b) = (0,0) \end{array} \right\}.$$

Let $\bar{x}$ be a feasible point of the system (CS). We define the index sets

$$\begin{aligned} I_{00} &:= I_{00}(\bar{x}) := \{i | \Phi_i(\bar{x}) = 0, \Psi_i(\bar{x}) = 0\}, \\ I_{0+} &:= I_{0+}(\bar{x}) := \{i | \Phi_i(\bar{x}) = 0, \Psi_i(\bar{x}) > 0\}, \\ I_{+0} &:= I_{+0}(\bar{x}) := \{i | \Phi_i(\bar{x}) > 0, \Psi_i(\bar{x}) = 0\}, \end{aligned}$$

and define the linearized cone as

$$\mathcal{L}(\bar{x}) := \left\{ u \in \mathbb{R}^n | \begin{array}{ll} 0 = \nabla H_i(\bar{x})u & i = 1, \ldots, d, \\ 0 = \nabla \Phi_i(\bar{x})u & i \in I_{0+}, \\ 0 = \nabla \Psi_i(\bar{x})u & i \in I_{+0}, \\ (\nabla \Phi_i(\bar{x})u, \nabla \Psi_i(\bar{x})u) \in \Omega_{EC}, & i \in I_{00} \end{array} \right\}.$$

Given $u \in \mathcal{L}(\bar{x})$ we define

$$\begin{aligned} I_{+0}(u) &:= \{i \in I_{00} | \nabla \Phi_i(\bar{x})u > 0 = \nabla \Psi_i(\bar{x})u\}, \\ I_{0+}(u) &:= \{i \in I_{00} | \nabla \Phi_i(\bar{x})u = 0 < \nabla \Psi_i(\bar{x})u\}, \\ I_{00}(u) &:= \{i \in I_{00} | \nabla \Phi_i(\bar{x})u = 0 = \nabla \Psi_i(\bar{x})u\}. \end{aligned}$$

Let $\bar{x}$ be a feasible point of (CS). By Definition 2.4.1 and Proposition 2.2.1, since the complementarity set Ω_{EC} is directionally regular, (CS) is directionally quasi- or pseudo-normal if and only if for all directions $0 \neq u \in \mathcal{L}(\bar{x})$ there exists no $(\eta, \gamma, \nu) \neq 0$ such that

$$0 = \nabla H(\bar{x})^T \eta - \nabla \Phi(\bar{x})^T \gamma - \nabla \Psi(\bar{x})^T \nu, \tag{2.22}$$

$$-(\gamma_i, \nu_i) \in N_{\Omega_{EC}}(\Phi_i(\bar{x}), \Psi_i(\bar{x}); \nabla \Phi_i(\bar{x})u, \nabla \Psi_i(\bar{x})u) \quad i = 1, \ldots, m, \tag{2.23}$$

$$\begin{aligned} &\exists (u^k, h^k, \phi^k, \psi^k, \eta^k, \gamma^k, \nu^k) \to (u, H(\bar{x}), \Phi(\bar{x}), \Psi(\bar{x}), \eta, \gamma, \nu), t_k \downarrow 0 \\ &\text{such that} \begin{cases} \eta^k \in N_{\{0\}^d}(h^k), -(\gamma_i^k, \nu_i^k) \in \widehat{N}_{\Omega_{EC}}(\phi_i^k, \psi_i^k) \quad i = 1, \ldots, m, \\ \text{if } \eta_i \neq 0, \ \eta_i(H_i(\bar{x} + t_k u^k) - h_i^k) > 0, \\ \text{if } \gamma_i \neq 0, \ \gamma_i(\Phi_i(\bar{x} + t_k u^k) - \phi_i^k) < 0, \\ \text{if } \nu_i \neq 0, \ \nu_i(\Psi_i(\bar{x} + t_k u^k) - \psi_i^k) < 0. \end{cases} \end{aligned} \tag{2.24}$$

or

$$\eta^T(H(\bar{x} + t_k u^k) - h^k) - \gamma^T(\Phi(\bar{x} + t_k u^k) - \phi^k) - \nu^T(\Psi(\bar{x} + t_k u^k) - \psi^k) > 0$$

respectively.

By the formula for the directional limiting normal cone in Lemma 2.5.1, (2.23) is equivalent to (ii) in the following definition. Since $\eta^k \in N_{\{0\}^d}(h^k)$, we have $h^k = 0$. Suppose $\gamma_i \neq 0$. Then for sufficiently large k, $\gamma_i^k \neq 0$. Since $-(\gamma_i^k,\ \nu_i^k) \in \widehat{N}_{\Omega_{EC}}(\phi_i^k, \psi_i^k)$ we must have $\phi_i^k = 0$. Similarly if $\nu_i \neq 0$, we must have $\psi_i^k = 0$. Based on these discussions, directional quasi-normality for (CS) can be written in the following form which is much more concise.

Definition 2.5.1. *Let $\bar{x}$ be a feasible solution of (CS). $\bar{x}$ is said to be directionally quasi- or pseudo-normal if for all directions $0 \neq u \in \mathcal{L}(\bar{x})$ there exists no $(\eta, \gamma, \nu) \neq 0$ such that*

(i) $0 = \nabla H(\bar{x})^T \eta - \nabla \Phi(\bar{x})^T \gamma - \nabla \Psi(\bar{x})^T \nu$;

(ii) $\gamma_i = 0,\ i \in I_{+0} \cup I_{+0}(u)$; $\nu_i = 0,\ i \in I_{0+} \cup I_{0+}(u)$; *eihter* $\gamma_i > 0,\ \nu_i > 0$ *or* $\gamma_i \nu_i = 0,\ i \in I_{00}(u)$;

(iii) $\exists u^k \to u$ *and* $t_k \downarrow 0$ *such that*

$$\begin{cases} \text{if } \eta_i \neq 0, & \eta_i H_i(\bar{x} + t_k u^k) > 0 \\ \text{if } \gamma_i \neq 0, & \gamma_i \Phi_i(\bar{x} + t_k u^k) < 0, \\ \text{if } \nu_i \neq 0, & \nu_i \Psi_i(\bar{x} + t_k u^k) < 0. \end{cases}$$

or

$$\eta^T H(\bar{x} + t_k u^k) - \gamma^T \Phi(\bar{x} + t_k u^k) - \nu^T \Psi(\bar{x} + t_k u^k) > 0$$

respectively.

Remark 2.5.1. *In Definition 2.5.1, if we only require that there exists no $(\eta, \gamma, \nu) \neq 0$ satisfying condition (i), then it reduces to the linearly independent constraint qualification (MPEC-LICQ) (see [73]). If we only require that there exists no $(\eta, \gamma, \nu) \neq 0$ satisfying condition (i) and change (ii) to*

$$\gamma_i = 0,\ i \in I_{+0};\ \nu_i = 0,\ i \in I_{0+},\quad \textit{eihter } \gamma_i \geq 0,\ \nu_i \geq 0 \quad \textit{or} \quad \gamma_i \nu_i = 0,\ i \in I_{00}$$

then it reduces to MPEC-NNAMCQ [82, Definition 2.10]. If we omit (iii), then it reduces to FOSCMS. If we take u to be any direction, then it reduces to the

MPEC quasi-/pseudo-normality first given in [47, Definition 3.2] and extended to the Lipschitz continuous case in [86, Definition 5]. Since for the set Ω_{EC} and any $0 \neq d \in T_{\Omega_{EC}}(0,0)$, the directional normal cone $N_{\Omega_{EC}}((0,0);d)$ is strictly smaller than the limiting normal cone $N_{\Omega_{EC}}(0,0)$, if there exists some $u \in \mathcal{L}(\bar{x})$ such that $(\nabla G(\bar{x})u, \nabla H(\bar{x})u) \neq (0,0)$, then directional quasi-/pseudo-normality will be strictly weaker than standard quasi-/pseudo-normality.

We now consider the following KKT system of an optimization problem with equality and inequality constraints:

$$\begin{aligned} &\nabla_x L(x,\mu,\lambda) = 0, \\ &\mu \geq 0, \quad g(x) \leq 0, \quad \langle g(x), \mu \rangle = 0, \\ &h(x) = 0, \end{aligned} \tag{2.25}$$

where $f : \mathbb{R}^p \to \mathbb{R},\ g : \mathbb{R}^p \to \mathbb{R}^m,\ h : \mathbb{R}^p \to \mathbb{R}^n$ are twice continuously differentiable, $\mu \in \mathbb{R}^m,\ \lambda \in \mathbb{R}^n$, and $L(x,\mu,\lambda) := f(x) + \mu^T g(x) + \lambda^T h(x)$ is the Lagrange function. Denote the feasible set of the KKT system by $\mathcal{F}_{KKT}$. We say that the error bound property holds at $(x^*,\mu^*,\lambda^*) \in \mathcal{F}_{KKT}$ if there exist $\alpha > 0$ and U, a neighborhood of (x^*,μ^*,λ^*), such that

$$d_{\mathcal{F}_{KKT}}(x,\mu,\lambda) \leq \alpha \max\{\|\nabla_x L(x,\mu,\lambda)\|, \|h(x)\|, \|\min\{\mu, -g(x)\}\|\}, \quad \forall (x,\mu,\lambda) \in U. \tag{2.26}$$

It is easy to see that this error bound property can be derived from the metric subregularity/calmness of KKT system and hence directional quasi-/pseudo-normality is a sufficient condition. Such an error bound property is crucial to the quadratic convergence of the Newton-type method (see [17]). The classical sufficient conditions for the error bound property are either MFCQ combined with the second-order sufficient condition (SOSC) or requiring $g,\ h$ to be affine and f to be quadratic (see e.g., [71]). These sufficient conditions were weakened in [19, 32] but still require SOSC. Recently, weaker sufficient conditions have been proposed including the existence of noncritical multipliers, a concept introduced by Izmailov for pure equality constraint in [42], extended by Izmailov and Solodov [43, Defnition 2] to problems with inequalities and further extended to a general variational system by Mordukhovich and Sarabi [60, Definition 3.1]. Note that as shown in [43, Proposition 3], the existence of noncritical multipliers is equivalent to a stronger type of error bound property: existence of

$\alpha > 0$ and U, a neighborhood of (x^*, μ^*, λ^*), such that

$$\|x - \bar{x}\| + d_{\mathcal{M}(\bar{x})}(\mu, \lambda) \leq \alpha \max\{\|\nabla_x L(x, \mu, \lambda)\|, \|h(x)\|, \|\min\{\mu, -g(x)\}\|\} \quad \forall (x, \mu, \lambda) \in U,$$

where $\mathcal{M}(\bar{x}) := \{(\mu, \lambda) : 0 = \nabla_x L(\bar{x}, \mu, \lambda),\ \mu \geq 0,\ \langle g(\bar{x}),\ \mu\rangle = 0\}$ denotes the set of multipliers. Obviously this is a stronger error bound property than the error bound property (2.26).

The KKT system is a special case of (CS) with

$$H(x, \mu, \lambda) := (\nabla_x L(x, \mu, \lambda), h(x)), \quad \Phi(x, \mu, \lambda) := -g(x), \quad \Psi(x, \mu, \lambda) := \mu.$$

Let $(\bar{x}, \bar{\mu}, \bar{\lambda})$ be a feasible point of the KKT system. We define the following index sets:

$$\begin{aligned}
I_{00} &:= I_{00}(\bar{x}, \bar{\mu}, \bar{\lambda}) := \{i | g_i(\bar{x}) = 0,\ \bar{\mu}_i = 0\},\\
I_{0+} &:= I_{0+}(\bar{x}, \bar{\mu}, \bar{\lambda}) := \{i | g_i(\bar{x}) = 0,\ \bar{\mu}_i > 0\},\\
I_{+0} &:= I_{+0}(\bar{x}, \bar{\mu}, \bar{\lambda}) := \{i | - g_i(\bar{x}) > 0,\ \bar{\mu}_i = 0\}.
\end{aligned}$$

The linearized cone for the KKT system is

$$\mathcal{L}(\bar{x}, \bar{\mu}, \bar{\lambda}) := \left\{ u = (u^x, u^\mu, u^\lambda) \left| \begin{array}{ll} 0 = \nabla^2_{xx} L(\bar{x}, \bar{u}, \bar{v}) u^x + \nabla g(\bar{x})^T u^\mu + \nabla h(\bar{x})^T u^\lambda, & \\ 0 = \nabla h(\bar{x}) u^x, & \\ 0 = \nabla g_i(\bar{x}) u^x, & i \in I_{0+}, \\ 0 = u_i^\mu, & i \in I_{+0}, \\ u_i^\mu \geq 0, \nabla g_i(\bar{x}) u^x \leq 0 \text{ and } u_i^\mu \nabla g_i(\bar{x}) u^x = 0, & i \in I_{00} \end{array} \right. \right\}.$$

Given $u \in \mathcal{L}(\bar{x}, \bar{\mu}, \bar{\lambda})$ we define the index sets

$$\begin{aligned}
I_{+0}(u) &:= \{i \in I_{00} | - \nabla g_i(\bar{x}) u^x > 0 = u_i^\mu\},\\
I_{0+}(u) &:= \{i \in I_{00} | \nabla g_i(\bar{x}) u^x = 0 < u_i^\mu\},\\
I_{00}(u) &:= \{i \in I_{00} | \nabla g_i(\bar{x}) u^x = 0 = u_i^\mu\}.
\end{aligned}$$

Then by Definition 2.5.1, we propose the following definition of directional quasi-normality for the KKT system.

Definition 2.5.2. *Let $(\bar{x}, \bar{\mu}, \bar{\lambda})$ be a feasible point of the KKT system. $(\bar{x}, \bar{\mu}, \bar{\lambda})$ is* *said to be directionally quasi-/pseudo-normal if for all directions*

$$0 \neq \bar{u} := (\bar{u}^x, \bar{u}^\mu, \bar{u}^\lambda) \in \mathcal{L}(\bar{x}, \bar{\mu}, \bar{\lambda})$$

there exists no $(\xi, \zeta, \eta) \neq 0$ *such that*

(i) $0 = \nabla^2_{xx} L(\bar{x}, \bar{\mu}, \bar{\lambda})\xi + \nabla h(\bar{x})^T \eta + \nabla g(\bar{x})^T \zeta$;

(ii) $\nabla h(\bar{x})\xi = 0$;

(iii) $\zeta_i = 0,\ i \in I_{+0} \cup I_{+0}(\bar{u})$; $\nabla g_i(\bar{x})\xi = 0,\ i \in I_{0+} \cup I_{0+}(\bar{u})$; *either* $\zeta_i > 0,\ \nabla g_i(\bar{x})\xi > 0$ *or* $\zeta_i \nabla g_i(\bar{x})\xi = 0,\ i \in I_{00}(\bar{u})$;

(iv) $\exists u_k := (u_k^x, u_k^\mu, u_k^\lambda) \to \bar{u}$ *and* $t_k \downarrow 0$ *such that*

$$\begin{cases} \text{if } \xi_i \neq 0,\ \ \xi_i \nabla_x L_i((\bar{x}, \bar{\mu}, \bar{\lambda}) + t_k u_k) > 0, \\ \text{if } \eta_i \neq 0,\ \ \eta_i h_i(\bar{x} + t_k u_k^x) > 0, \\ \text{if } \zeta_i \neq 0,\ \ \zeta_i g_i(\bar{x} + t_k u_k^x) > 0, \\ \text{if } (\nabla g(\bar{x})\xi)_i \neq 0,\ \ (\nabla g(\bar{x})\xi)_i(\bar{u}_i^\mu + t_k(u_k^\mu)_i) < 0, \end{cases}$$

or

$$\xi^T \nabla_x L((\bar{x}, \bar{\mu}, \bar{\lambda}) + t_k u_k) + \eta^T h(\bar{x} + t_k u_k^x) - (\nabla g(\bar{x})\xi)^T(\bar{u} + t_k u_k^\mu) > 0$$

respectively.

Remark 2.5.2. *Let* $(\bar{x}, \bar{\mu}, \bar{\lambda})$ *be a feasible point to the KKT system. By [43, Definition 2],* $(\bar{\mu}, \bar{\lambda}) \in \mathcal{M}(\bar{x})$ *is a critical multiplier associated with* $\bar{x}$ *if there exists* (ξ, ζ, η) *with* $\xi \neq 0$ *satisfying that*

$$\begin{cases} 0 = \nabla^2_{xx} L(\bar{x}, \bar{\mu}, \bar{\lambda})\xi + \nabla h(\bar{x})^T \eta + \nabla g(\bar{x})^T \zeta, & \\ 0 = \nabla h(\bar{x})\xi, & \\ 0 = \nabla g_i(\bar{x})\xi, & i \in I_{0+} \\ 0 = \zeta_i, & i \in I_{+0} \\ \zeta_i \geq 0, \nabla g_i(\bar{x})\xi \leq 0 \text{ and } \zeta_i \nabla g_i(\bar{x})\xi = 0, & i \in I_{00}. \end{cases}$$

Note that from Definition 2.5.2, we can see that even if $(\bar{\mu}, \bar{\lambda})$ *is a critical multiplier with* $\bar{x}$*, it is still possible for directional quasi-normality to hold. In particular let*

(ξ, ζ, η) satisfy Definition 2.5.2 with $\xi \neq 0$. Suppose that for $i \in I_{00}(\bar{u})$, it is not possible to have $\zeta_i > 0$, $\nabla g_i(\bar{x})\xi > 0$. Then $(\bar{\mu}, \bar{\lambda}) \in \mathcal{M}(\bar{x})$ is a critical multiplier associated with $\bar{x}$.

Chapter 3

Directional necessary optimality conditions for bilevel programs

In **Chapter 1.3**, in the proof of Theorem 1.3.3 we proved that, given a local optimal solution $\bar{x}$ to NLP with Lipschitz continuous objective function, under the Clarke calmness at $\bar{x}$, KKT condition holds at $\bar{x}$. Hence, sufficient conditions for Clarke calmness can imply KKT conditions. In **Chapter 2**, we introduced the directional quasi-normality condition as a sufficient condition for metric subregualrity weaker than FOSCMS. As an application, in this chapter, we apply the directional quasi-normality to bilevel programs and derive sharp optimality conditions. In particular, first, we derive directional optimality conditions for nonsmooth NLPs under the directional quasi-normality. Secondly, we present upper estimates for the directional subdifferential and directional derivative of the lower level value function $V(x)$.

All the content of this chapter has been submitted as a journal paper, see [3].

3.1 Introduction

The motivation for studying bilevel optimization originated in economics under the name of Stackelberg games [74] since 1934. In economics, it is used to model interactions between a leader and its follower of a two level hierarchical system and hence is referred to as leader and follower games or principal-agent problems. In recent years, bilevel programs find wider range of applications (see e.g. [13, 46, 55, 63] and references within). In particular, bilevel programs have been used to model hyper-parameter selection in machine learning (see e.g. [51, 52]) in recent years.

In this chapter, we consider the following bilevel program

$$\begin{aligned} \text{(BP)} \quad \min_{x,y} \quad & F(x,y) \\ s.t. \quad & y \in S(x),\ G(x,y) \leq 0, \end{aligned}$$

where for any given x, $S(x)$ denotes the solution set of the lower level program

$$(P_x) \quad \min_y f(x,y) \quad \text{s.t. } g(x,y) \leq 0,$$

and $F, f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R},\ G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q,\ g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ are continuously differentiable. Throughout this chapter, for simplicity we assume $S(x) \neq \emptyset$ for all x. To obtain an optimality condition for (BP), one may reformulate it as a single-level optimization problem and apply optimality conditions to the single-level problem. There are three approaches for reformulating (BP) as a single-level optimization problem in the literature. The earliest approach is the so-called first order approach or the Karush-Kuhn-Tucker (KKT) approach by which one replaces the constraint $y \in S(x)$ by its first order optimality conditions and minimizing over the original variables as well as the multipliers. The resulting single-level optimization problem is the so-called mathematical program with equilibrium constraints (MPEC), which was popularly studied over the last three decades; see e.g. [55, 63] for the general theory and [81, 84, 28] for the optimality conditions derived by using this approach. The value function approach (see e.g. [88]) replaces the constraint $y \in S(x)$ by $f(x,y) - V(x) \leq 0$, where $V(x) := \inf_y\{f(x,y)|g(x,y) \leq 0\}$ is the value function of the lower level program (P_x). And the combined approach ([90]) not only replaces the constraint $y \in S(x)$ by $f(x,y) - V(x) \leq 0$ but also adds the first order optimality conditions. The first order approach is obviously only applicable if the first order optimality condition is necessary and sufficient for optimality; e.g. when the lower level program is convex and certain constraint qualification holds. Both the KKT approach and the combined approach suffer from the drawback that the resulting MPEC may not be equivalent to the original (BP) if the local optimality is considered; see e.g. [83] and the reference within for further discussions on this issue.

In this chapter by the value function approach, we reformulate (BP) as (VP). Under fairly reasonable assumptions, the value function $V(x)$ is Lipschitz continuous and hence a nonsmooth Fritz John type necessary optimality condition holds at a local optimal solution. For a KKT type necessary optimality condition to hold, in general one needs to assume certain constraint qualifications. Unfortunately, it is known ([88, Proposition 3.1]) that the nonsmooth MFCQ or equivalently the no nonzero abnormal multiplier constraint qualification (NNAMCQ), a standard constraint qualification for nonsmooth mathematical programs, fails to hold at any feasible point of (VP). For an optimization problem with Lipschitz continuous problem data, it is known that the necessary optimality condition holds provided that the problem is calm in the sense of Clarke [12, Definition 6.41]. Ye and Zhu [88] introduced the partial calmness condition for problem (VP) which means that a local solution of problem (VP) is also a local solution of the partially penalized problem for certain $\rho > 0$

$$\begin{array}{lll} (\mathrm{VP})_\rho & \min\limits_{x,y} & F(x,y) + \rho(f(x,y) - V(x)) \\ & s.t. & g(x,y) \leq 0, G(x,y) \leq 0. \end{array}$$

Since the most difficult constraint $f(x,y) - V(x)$ is moved to the objective, the KKT condition would hold under some constraint qualifications for the partially penalized problem $(\mathrm{VP})_\rho$. It is easy to show that the full calmness implies the partial calmness and the partial calmness plus the full calmness of the partially penalized problem $(\mathrm{VP})_\rho$ implies the full calmness condition for problem (VP). Some sufficient conditions for partial calmness and its relationship with exact penalization were further discussed in [88, 89, 91]. Unfortunately for problem (VP), the partial calmness or the full calmness condition is still a fairly strong condition. And there are very few constraint qualifications or sufficient conditions for partial calmness for (VP) in the literature. Recently, [80] has extended the relaxed constant positive linear dependence constraint qualification (RCPLD) to bilevel programs and has shown that it is a constraint qualification.

Recently Gfrerer [21, Theorem 7] derived a directional version of the KKT type necessary optimality condition for set-constrained mathematical programs under the directional metric subregularity constraint qualification. The directional KKT condition is in general sharper than the nondirectional KKT condition and the directional metric subregularity is weaker than the nondirectional one. Inspired by this approach, in this chapter we aim at developing a directional KKT condition for problem (VP). First we review the following concept of directional neighborhood recently introduced by Gfrerer in [21]. Given a direction $d \in \mathbb{R}^n$, and positive numbers $\epsilon, \delta > 0$, the directional neighborhood of direction d is a set defined by

$$\mathcal{V}_{\epsilon,\delta}(d) := \{z \in \epsilon\mathbb{B} | \big\| \|d\|z - \|z\|d \big\| \leq \delta \|z\| \|d\| \},$$

where $\mathbb{B}$ denotes the open unit ball and $\|\cdot\|$ denotes the Euclidean norm. It is easy to see that the directional neighborhood of direction $d = 0$ is just the open ball $\epsilon\mathbb{B}$ and the directional neighborhood of a nonzero direction $d \neq 0$ is a smaller subset of $\epsilon\mathbb{B}$. Hence many regularity conditions can be extended to a directional version which is weaker than the original nondirectional one. We say that (VP) is calm at a feasible solution $(\bar{x}, \bar{y})$ in direction $d \in \mathbb{R}^{n+m}$ if there exist positive scalars ϵ, δ, ρ, such that for any $\alpha \in \epsilon\mathbb{B}$ and any $(x, y) \in (\bar{x}, \bar{y}) + \mathcal{V}_{\epsilon,\delta}(d)$ satisfying $\varphi(x, y) + \alpha \in \mathbb{R}^{1+p+q}_{-}$ with $\varphi(x, y) := (f(x, y) - V(x), g(x, y), G(x, y))$ one has,

$$F(x, y) - F(\bar{x}, \bar{y}) + \rho\|\alpha\| \geq 0.$$

It is obvious that when the direction $d = 0$, the directional calmness is reduced to the classical calmness condition [12, Definition 6.41]. When $d \neq 0$, since the directional neighborhood is in general smaller than the usual neighborhood, the directional calmness condition is in general weaker than the nondirectional calmness condition. It is obvious that if $(\bar{x}, \bar{y})$ solves (VP), then under the calmness condition in direction d, $(\bar{x}, \bar{y})$ is also a solution of the following penalized problem

$$\begin{array}{llll} \text{(DP)} & \min\limits_{x,y} & F(x, y) + \rho \text{dist}(\varphi(x, y), \mathbb{R}^{1+p+q}_{-}) \\ & s.t. & (x, y) \in (\bar{x}, \bar{y}) + \mathcal{V}_{\epsilon,\delta}(d). \end{array}$$

The directionally penalized problem (DP) is much easier to deal with than (VP) since all the inequality constraints are moved to the objective function. By using the nonsmooth calculus, one can then show that $(\bar{x}, \bar{y})$ satisfies a KKT condition provided the value function is Lipschitz continuous. In fact we can achieve more. When d is a critical direction, we can show that $(\bar{x}, \bar{y})$ satisfies a directional KKT condition in which a directional Clarke subdifferential (see Definition 3.2.4 and (3.1)) of the value function $V(x)$ at $\bar{x}$ in direction d is used instead of the Clarke subdifferential. Since the directional Clarke subdifferential is a subset of the Clarke subdifferential, the directional KKT condition is sharper than the nondirectional one.

To make the directional calmness condition and the directional KKT condition useful, we have two issues to consider. First, under what conditions, the value function is directionally Lipschitz continuous and directionally differentiable and how to calculate the directional limiting subdifferential and the directional derivative of the value function which will be needed in the directional KKT condition for problem (VP). In this chapter, we have derived some formulas for the directional derivative of the value function and an upper estimate for the Clarke directional subdifferential of the value function $V(x)$. Secondly, how to derive a verifiable constraint qualification which ensures the directional calmness condition of (VP)? It is known that the FOSCMS (introduced in Gfrerer and Klatte [25, Corollary] for the smooth case and [4, Proposition 2.2] for the nonsmooth case) is a sufficient condition for the metric subregularity of the set-valued map $\Phi(x, y) := \varphi(x, y) - \mathbb{R}_-^{p+q+1}$ which in turn implies the calmness of the problem (VP). FOSCMS is in general weaker than NNAMCQ and hence it is natural to ask if FOSCMS would hold for (VP). Unfortunately in Proposition 3.5.1, we show that FOSCMS also fails for problem (VP) in any critical direction. We propose the directional quasi-normality as a sufficient condition for the directional calmness condition and give an example to show that the directional quasi-normality is possible to hold for (VP).

Other than deriving a weaker constraint qualification and a sharper necessary optimality condition for bilevel programs, we have also made contributions that are of independent interest as summarized below.

- We introduce the concept of directional Clarke subdifferentials and derive some useful calculus rules for directional subdifferentials; see Propositions 3.2.4 and 3.2.6.

- For an optimization problem with directionally Lipschitz continuous objective function and directionally Lipschitz and directionally differentiable inequality constraints, we derive a directional KKT condition under the directional calmness condition; see Theorem 3.3.1. An example of a bilevel program is given to show that the directional calmness is weaker than the classical calmness; see Example 3.3.1.

- The classical results for the directional derivative of the value function are improved with weaker assumptions: see Propositions 3.4.3 and 3.4.4 and Corollary 3.4.1. Sufficient conditions for directional Lipschitz continuity of the value function is given in Theorem 3.4.1 and the upper estimate of the directional subdifferential of the value function is given in Theorems 3.4.2 and 3.4.3.

We organize the chapter as follows. In **Section 3.2**, we provide the notations, preliminaries and preliminary results. In **Section 3.3** we derive the directional KKT condition under the directional calmness condition for a general optimization problem with directionally Lipschitz inequality constraints. In **Section 3.4**, we study directional sensitivity analysis of the value function. Finally in **Section 3.5**, we apply the previous results to (VP) and derive a verifiable constraint qualification and a necessary optimality condition.

3.2 Preliminaries

We now review some basic concepts and results in directional variational analysis, which will be used later on. For more details see e.g. [7, 10, 12, 14, 54, 59, 72]. Moreover we derive some preliminary results that will be needed. For any $y \in \mathbb{R}^p_-$, define the active index set $I_y := \{i = 1, \ldots, p | y_i = 0\}$. One can easily obtain that $N_{\mathbb{R}^p_-}(y) = \{\zeta \in \mathbb{R}^p_+ | \zeta_i = 0,\ i \notin I_y\}$. The property stated in the following proposition will be useful.

Proposition 3.2.1. *Let $y, z \in \mathbb{R}^p_-$ be such that $I_y \subseteq I_z$. Then $N_{\mathbb{R}^p_-}(y) = N_{\mathbb{R}^p_-}(z) \cap [y - z]^\perp$.*

Proof. The proof follows immediately from [72, Example 6.10]. ∎

When $d = 0$ the following definition coincides with the Painlevé-Kuratowski inner/lower and outer/upper limit of Φ as $x \to \bar{x}$ respectively; see e.g., [59].

Definition 3.2.1. *Given a set-valued map $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and a direction $d \in \mathbb{R}^n$, the inner/lower and outer/upper limit of Φ as $x \xrightarrow{d} \bar{x}$ respectively is defined by*

$$\liminf_{x \xrightarrow{d} \bar{x}} \Phi(x) := \{y \in \mathbb{R}^m | \forall t_k \downarrow 0, d^k \to d, \exists y^k \to y \ s.t.\ y^k \in \Phi(\bar{x} + t_k d^k)\}$$
$$\limsup_{x \xrightarrow{d} \bar{x}} \Phi(x) := \{y \in \mathbb{R}^m | \exists t_k \downarrow 0, d^k \to d, y^k \to y \ s.t.\ y^k \in \Phi(\bar{x} + t_k d^k)\},$$

respectively.

Definition 3.2.2 (Directional Lipschitz continuity)**.** *([7, Page 719]) We say that a single-valued map $\phi(x) : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous at $\bar{x}$ in direction d if there* *exists a scalar $L > 0$ and a directional neighborhood $\mathcal{V}_{\epsilon,\delta}(d)$ of d such that*

$$\|\phi(x) - \phi(x')\| \leq L\|x - x'\| \quad \forall x, x' \in \bar{x} + \mathcal{V}_{\epsilon,\delta}(d).$$

Obviously, the directional Lipschitz continuity in direction $d = 0$ coincides with the classical Lipschitz continuity. Note that [72, Page 386] also defines the directional Lipschitz continuity, which however is different from the one presented in this paper.

Definition 3.2.3 (Directional derivatives)**.** *Let $\phi : \mathbb{R}^n \to \mathbb{R}^m$ and $x, u \in \mathbb{R}^n$. The usual directional derivative of ϕ at x in the direction u is*

$$\phi'(x; u) := \lim_{t \downarrow 0} \frac{\phi(x + tu) - \phi(x)}{t}$$

when this limit exists.

It is easy to see that if $\phi : \mathbb{R}^n \to \mathbb{R}^m$ is Lipschitz continuous and directionally differentiable at x in direction u, then for all sequence $\{u^k\}$ which converges to u, we have

$$\phi'(x; u) = \lim_{k \to \infty} \frac{\phi(x + t_k u^k) - \phi(x)}{t_k} = D\phi(x)(u).$$

We now recall the definition of some subdifferentials below.

Definition 3.2.4 (Analytic directional subdifferentials)**.** *[21, 29, 54, 7] Let $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $\bar{x} \in dom\varphi$. The analytic limiting subdifferential of φ at $\bar{x}$ in direction $u \in \mathbb{R}^n$ is defined as*

$$\partial_a \varphi(\bar{x}; u) := \{\xi \in \mathbb{R}^n | \exists t_k \downarrow 0, u^k \to u, \xi^k \to \xi \ s.t. \ \varphi(\bar{x} + t_k u^k) \to \varphi(\bar{x}), \ \xi^k \in \widehat{\partial}\varphi(\bar{x} + t_k u^k)\}.$$

It is easy to see that if $u \notin T_{dom\varphi}(\bar{x})$, then $\partial_a \varphi(\bar{x}; u) = \emptyset$.

Proposition 3.2.2. *[54, Theorem 5.4] Given $\varphi : \mathbb{R}^n \to \bar{\mathbb{R}}$, $\bar{x} \in dom\varphi$ and a direction $u \in \mathbb{R}^n$, one has*

$$\partial_a \varphi(\bar{x}; u) = \{\xi \in \mathbb{R}^n | \exists t_k \downarrow 0, u^k \to u, \xi^k \to \xi \ s.t. \ \varphi(\bar{x} + t_k u^k) \to \varphi(\bar{x}), \ \xi^k \in \partial\varphi(\bar{x} + t_k u^k)\}.$$

Remark 3.2.1. *When $\varphi(x)$ is calm at $\bar{x}$ in direction u and $D\varphi(\bar{x})(u) = \{\zeta\}$ is a singleton, we have $\partial\varphi(\bar{x}; (u, \zeta)) = \partial_a \varphi(\bar{x}; u)$. By [7, Proposition 5.1], if φ is Lipschitz continuous and directionally differentiable at $\bar{x}$ in direction u we have $D\varphi(\bar{x})(u) =$*

$\{\varphi'(\bar{x};u)\}$ *and*

$$D^*\varphi(\bar{x};(u,\xi))(\zeta) = \partial_a\langle\zeta,\varphi\rangle(\bar{x};u).$$

In this case $\partial\varphi(\bar{x};(u,\varphi'(\bar{x};u)) = \partial_a\varphi(\bar{x};u)$*; see [7, Corollary 4.1].*

Furthermore, we define the directional Clarke subdifferential of locally Lipschitz continuous functions, which is of its own interest.

Definition 3.2.5 (Directional Clarke subdifferential)**.** *Let* $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ *be locally Lipschitz continuous near* $\bar{x}$*. The directional Clarke subdifferential of* φ *at* $\bar{x}$ *in direction* u *is defined as*

$$\partial^c\varphi(\bar{x};u) := co(\partial_a\varphi(\bar{x};u)). \tag{3.1}$$

Proposition 3.2.3. *Let* $\varphi : \mathbb{R}^n \to R$ *be Lipschitz continuous at* $\bar{x}$ *in direction* u*. Then we have*

$$\partial^c\varphi(\bar{x};u) = co\limsup_{x\xrightarrow{u}\bar{x}}\partial^c\varphi(x),\ \ \partial^c(-\varphi)(\bar{x};u) = -\partial^c\varphi(\bar{x};u).$$

Proof. By Proposition 3.2.2, we have $\partial_a\varphi(\bar{x};u) = \limsup_{x\xrightarrow{u}\bar{x}}\partial\varphi(x)$. It follows that

$$\partial^c\varphi(\bar{x};u) := co(\partial_a\varphi(\bar{x};u)) = co(\limsup_{x\xrightarrow{u}\bar{x}}\partial\varphi(x)) = co(\limsup_{x\xrightarrow{u}\bar{x}}\partial^c\varphi(x)),$$

where the last equality follows from the fact that for any sequence of sets $\{A_k\} \subseteq \Omega$ where $\Omega \subseteq \mathbb{R}^n$ is compact, we have $\limsup_k co(A_k) \subseteq co(\limsup_k A_k)$. Indeed, for any sequence $x^k \xrightarrow{u} \bar{x}$, let $A_k = \partial\varphi(x^k)$ and by the Lipschitz continuity of $\varphi(x)$ and [72, Theorem 9.13], there exists $L > 0$ such that $\{A_k\} \subseteq L\bar{\mathbb{B}}$. And the second equality follows directly from the first equality and the scalar multiplication rule of Clarke subdifferential [12, Proposition 2.3.1]. ■

Proposition 3.2.4 (Sum Rules for analytic directional differentials)**.** *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be Lipschitz at* $\bar{x}$ *in direction* u *and* $g : \mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ *be l.s.c., proper and finite at* $\bar{x}$*. Let* α, β *be nonnegative scalars and* $u \in \mathbb{R}^n$*. Then* $\partial_a(\alpha f+\beta g)(\bar{x};u) \subseteq \alpha\partial_a f(\bar{x};u)+\beta\partial_a g(\bar{x};u)$*.*

Proof. Let $\xi \in \partial_a(\alpha f + \beta g)(\bar{x};u)$, by Proposition 3.2.2, there exist sequences $t_k \downarrow 0$, $u^k \to u$, $\xi^k \to \xi$ such that $(\alpha f + \beta g)(\bar{x} + t_k u^k) \to (\alpha f + \beta g)(\bar{x})$ and $\xi^k \in \partial(\alpha f + \beta g)(\bar{x} + t_k u^k)$. Since f is Lipschitz continuous at $\bar{x}$ in direction u, for all sufficiently large k, $f(x)$ is Lipschitz continuous near $\bar{x}+t_k u^k$ and hence $\beta g(\bar{x}+t_k u^k)$ is finite. It follows by the sum rule of limiting subdifferentials (see e.g., [72, Exercise 10.10]) that we have

$$\partial(\alpha f+\beta g)(\bar{x}+t_k u^k) \subseteq \alpha\partial f(\bar{x}+t_k u^k)+\beta\partial g(\bar{x}+t_k u^k).$$

That is, there exist $\xi_f^k \in \partial f(\bar{x}+t_k u^k)$ and $\xi_g^k \in \partial g(\bar{x}+t_k u^k)$ such that $\xi^k = \alpha\xi_f^k+\beta\xi_g^k$. By the continuity of f, $f(\bar{x}+t_k u^k) \to f(\bar{x})$. Hence, $\beta g(\bar{x}+t_k u^k) \to \beta g(\bar{x})$. Since $f(x)$ is Lipschitz continuous near $\bar{x}+t_k u^k$ with a Lipschitz constant K for all k large enough, $\|\xi_f^k\| \leq K$ (see e.g., [72, Theorem 9.13]). Hence both $\{\xi_f^k\}$ and $\{\beta\xi_g^k\}$ are bounded. Without loss of generality, we assume $\xi_f := \lim_k \xi_f^k$, $\beta\xi_g := \lim_k \beta\xi_g^k$. we have $\xi_f \in \partial f(\bar{x};u)$. If $\beta=0$, we have $\xi = \alpha\xi_f + 0\cdot\xi_g$. Otherwise if $\beta>0$, we have $g(\bar{x}+t_k u^k) \to g(\bar{x})$. By Proposition 3.2.2, we have $\xi_g \in \partial_a g(\bar{x};u)$ and $\xi = \alpha\xi_f+\beta\xi_g$. For both of these two cases, we can obtain $\xi \in \alpha\partial_a f(\bar{x};u)+\beta\partial_a g(\bar{x};u)$. By the choice of ξ, the desired inclusion is proved. $\blacksquare$

We now give the definition of directional metric subregularity.

Definition 3.2.6 (Directional Metric Subregularity)**.** *[21, Definition 2.1] Let $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ be a set-valued map and $(\bar{x},\bar{y}) \in gph\Phi$. Given a direction $u \in \mathbb{R}^n$, Φ is said to be metrically subregular in direction u at $(\bar{x},\bar{y})$, if there are positive reals $\epsilon>0, \delta>0$, and $\kappa>0$ such that*

$$\operatorname{dist}(x,\Phi^{-1}(\bar{y})) \leq \kappa \operatorname{dist}(\bar{y},\Phi(x)) \qquad \forall x \in \bar{x}+\mathcal{V}_{\epsilon,\delta}(u).$$

If $u=0$ in the above definition, then the set-valued map Φ is metrically subregular at $(\bar{x},\bar{y})$ in the classical sense.

Proposition 3.2.5. *Let $C \subseteq \mathbb{R}^p$ be closed and $\phi : \mathbb{R}^n \to \mathbb{R}^p$. The set-valued map $\Phi(x) := -\phi(x)+C$ is metrically subregular at $(\bar{x},0)$ in direction u if and only if the set-valued map $\Psi(x,\alpha) := epi\delta_C - (\phi(x),\alpha)$ is metrically subregular at $((\bar{x},0),(0,0))$ in direction (u,r) $\forall r \in \mathbb{R}$.*

Proof. It is easy to verify that $epi\delta_C = C\times\mathbb{R}_+$ and $\Psi^{-1}(0) = \Phi^{-1}(0)\times\mathbb{R}_+$. By the equivalence of norms in Euclidean space and the triangle inequality, we can find a positive scalar ρ such that for any $x,x' \in \mathbb{R}^p, \alpha,\alpha' \in \mathbb{R}$, it holds $\rho(\|x-x'\|+|\alpha-\alpha'|) \leq$

$\|(x,\alpha)-(x',\alpha')\| \le \|x-x'\|+|\alpha-\alpha'|$. Therefore there exists $\rho>0$ such that

$$\begin{aligned}\rho(\mathrm{dist}(x,\Phi^{-1}(0))+\mathrm{dist}(\alpha,\mathbb{R}_+)) &\le \mathrm{dist}((x,\alpha),\Psi^{-1}(0))\\ &\le \mathrm{dist}(x,\Phi^{-1}(0))+\mathrm{dist}(\alpha,\mathbb{R}_+).\end{aligned} \tag{3.2}$$

Similarly, there exists $\rho'>0$ such that

$$\begin{aligned}\rho'(\mathrm{dist}(0,\Phi(x))+\mathrm{dist}(\alpha,\mathbb{R}_+)) &\le \mathrm{dist}(0,\Psi(x,\alpha))\\ &\le \mathrm{dist}(0,\Phi(x))+\mathrm{dist}(\alpha,\mathbb{R}_+).\end{aligned} \tag{3.3}$$

Since $\Psi(x,\alpha)$ is metrically subregular at $((\bar{x},0),(0,0))$ in direction (u,r) if and only if there exist positive scalars κ,ϵ,δ such that

$$\mathrm{dist}((x,\alpha),\Psi^{-1}(0)) \le \kappa \mathrm{dist}(0,\Psi(x,\alpha)),\ \forall x\in \bar{x}+\mathcal{V}_{\epsilon,\delta}(u), \forall \alpha\in\delta\mathbb{B}, \tag{3.4}$$

and $\Phi(x)$ is metrically subregular at $(\bar{x},0)$ in direction u if there exist positive scalars $\kappa',\epsilon',\delta'$ such that

$$\mathrm{dist}(x,\Phi^{-1}(0)) \le \kappa' \mathrm{dist}(0,\Phi(x)),\ \forall x\in\bar{x}+\mathcal{V}_{\epsilon',\delta'}(u), \tag{3.5}$$

by (3.2) and (3.3), it follows that (3.4) holds if and only if (3.5) holds and the proof is complete. ■

We now derive a chain rule for the analytic directional subdifferential of the composition function of an indicator function and a smooth map.

Proposition 3.2.6. *Let $\phi:\mathbb{R}^n\to\mathbb{R}^p$ be continuously differentiable and $C\subseteq\mathbb{R}^p$ be closed. Suppose $\phi(\bar{x})\in C$ and the set-valued map $\Phi(x):=-\phi(x)+C$ is metrically subregular at $(\bar{x},0)$ in direction u. Then*

$$\partial_a(\delta_C\circ\phi)(\bar{x};u)\subseteq\nabla\phi(\bar{x})^T N_C(\phi(\bar{x});\nabla\phi(\bar{x})u).$$

Proof. Define $\varphi(x):=\delta_C\circ\phi(x)$. If $u\notin T_{dom\varphi}(\bar{x})$ with $\mathrm{dom}\varphi=\{x|\phi(x)\in C\}$, then $\partial_a\varphi(\bar{x};u)=\emptyset$. And the proposition holds trivially. Otherwise if $u\in T_{dom\varphi}(\bar{x})$, there exist sequences $t_k\downarrow 0, u^k\to u$ with $\bar{x}+t_k u^k\in dom\varphi$. Then it follows that for all such sequences we have $\varphi(\bar{x}+t_k u^k)\equiv 0$ for all k. Hence, $D\varphi(\bar{x})(u)=\{0\}$. By Remark 3.2.1, we have $\partial_a\varphi(\bar{x};u)=\partial\varphi(\bar{x};u,0)$. Since the set-valued mapping $\Phi(x)$

is metrically subregular at $(\bar{x}, 0)$ in direction u, by Proposition 3.2.5, the set-valued map given by $\Psi(x, \alpha) := epi\delta_C - (\phi(x), \alpha)$ is metrically subregular at $((\bar{x}, 0), (0, 0))$ in direction $(u, 0)$. Since ϕ is continuously differentiable by [7, Theorem 4.1] and Remark 3.2.1 we have

$$\partial_a(\delta_C \circ \phi)(\bar{x}; u) = \partial(\delta_C \circ \phi)(\bar{x}; u, 0) \subseteq \nabla\phi(\bar{x})^T \partial\delta_C(\phi(\bar{x}); \nabla\phi(\bar{x})u).$$

The desired result follows from the fact that $N_C(z; d) = \partial\delta_C(z; d)$ by virtue of [54, Theorem 5.5]. ■

3.3 Directional KKT conditions under directional calmness condition

In this section we derive directional KKT condition for the optimization problem

$$(\widetilde{\mathrm{P}}) \quad \min_{z} \quad \varphi(z) \quad \text{s.t. } \phi(z) \leq 0,$$

where $\varphi : \mathbb{R}^n \to \mathbb{R}$ and $\phi : \mathbb{R}^n \to \mathbb{R}^q$.

The concept of (Clarke) calmness for a mathematical program is first defined by Clarke [12, Definition 6.41]. We now introduce a directional version of the calmness condition for $(\widetilde{\mathrm{P}})$.

Definition 3.3.1 (Directional Clarke calmness)**.** *Suppose $\bar{z}$ solves $(\widetilde{\mathrm{P}})$. We say that $(\widetilde{\mathrm{P}})$ is (Clarke) calm at $\bar{z}$ in direction u if there exist positive scalars ϵ, δ, ρ, such that for any $\alpha \in \epsilon\mathbb{B}$ and any $z \in \bar{z} + \mathcal{V}_{\epsilon,\delta}(u)$ satisfying $\phi(z) + \alpha \leq 0$ one has,*

$$\varphi(z) - \varphi(\bar{z}) + \rho\|\alpha\| \geq 0.$$

We now prove that the directional metric subregularity implies the directional calmness of problem $(\widetilde{\mathrm{P}})$ provided the objective function is directional Lipschitz continuous.

Lemma 3.3.1. *Let $\bar{z}$ solve $(\widetilde{\mathrm{P}})$ and $\varphi(z)$ be Lipschitz continuous at $\bar{z}$ in direction u. Suppose that the set-valued map $\Phi(z) : -\phi(z) + \mathbb{R}^q_-$ is metrically subregular at $(\bar{z}, 0)$ in direction u. Then $(\widetilde{\mathrm{P}})$ is calm at $\bar{z}$ in direction u.*

Proof. Since $\Phi(z)$ is metrically subregular at $(\bar{z}, 0)$ in direction u, by Definition 3.2.6, there exist positive scalars ϵ, δ, κ such that

$$\mathrm{dist}(z, \Phi^{-1}(0)) \leq \kappa \mathrm{dist}(\phi(z), \mathbb{R}^q_-) \qquad \forall z \in \bar{z} + \mathcal{V}_{\epsilon,\delta}(u).$$

Let $\tilde{z}$ be the projection of z onto $\Phi^{-1}(0)$. Since $\varphi(z)$ is directionally Lipschitz continuous, without loss of generality, taking ϵ, δ small enough, there exists $L > 0$ such that $|\varphi(z) - \varphi(z')| \leq L\kappa\|z - z'\|$ for any $z, z' \in \bar{z} + \mathcal{V}_{\epsilon,\delta}(u)$. Then we have for any $\alpha \in \epsilon\mathbb{B}$ satisfying $\phi(z) + \alpha \leq 0$,

$$\begin{aligned}
\varphi(z) - \varphi(\bar{z}) + L\kappa\|\alpha\| &\geq \varphi(z) - \varphi(\bar{z}) + L\kappa \mathrm{dist}(\varphi(z), \mathbb{R}^q_-) \\
&\geq \varphi(z) - \varphi(\bar{z}) + L\kappa \mathrm{dist}(z, \Phi^{-1}(0)) \\
&= \varphi(z) - \varphi(\bar{z}) + L\kappa\|z - \tilde{z}\| \\
&\geq \varphi(z) - \varphi(\tilde{z}) + L\kappa\|z - \tilde{z}\| \\
&\geq 0,
\end{aligned}$$

where the third inequality follows from the optimality of $\varphi(z)$ at $\bar{z}$ and the last inequality follows from the directional Lipschitz continuity of $\varphi(z)$ at $\bar{z}$. Let $\rho := L\kappa$. The proof is complete. ■

Let $\bar{z}$ be a feasible solution to problem $(\widetilde{\mathrm{P}})$. We denote by $\bar{I}_\phi := I_\phi(\bar{z}) := \{j = 1, \ldots, q | \phi_j(\bar{z}) = 0\}$ the set of indexes of active constraints at $\bar{z}$. If φ is continuously differentiable and ϕ is Lipschitz and directionally differentiable, we define the linearized cone by $L(\bar{z}) := \{u \in \mathbb{R}^n | \phi'_j(\bar{z}; u) \leq 0, j \in I_\phi(\bar{z})\}$ and the critical cone by

$$C(z) := \{u \in L(\bar{z}) | \nabla\varphi(z)u \leq 0\} = \{u \in \mathbb{R}^n | \phi'_j(\bar{z}; u) \leq 0, j \in I_\phi(\bar{z}), \nabla\varphi(z)u \leq 0\}.$$

The following definition lists some sufficient conditions for the directional metric subregularity, hence are sufficient for directional calmness.

Definition 3.3.2. *Let $\phi(\bar{z}) \leq 0$ and $u \in \mathbb{R}^n$.*

- *Suppose that ϕ is Lipschitz at $\bar{z}$. We say that the NNAMCQ holds at $\bar{z}$ if*

$$0 \in \partial\langle\zeta, \phi\rangle(\bar{z}) \text{ and } 0 \leq \zeta \perp \phi(\bar{z}) \Longrightarrow \zeta = 0.$$

- *Suppose that ϕ is Lipschitz and directionally differentiable at $\bar{z}$ in direction u.*

We say that the FOSCMS holds at $(\bar{z}, 0)$ *in direction* u *if there exists no* $\zeta \neq 0$ *satisfying* $0 \leq \zeta \perp \phi(\bar{z}),\ \zeta \perp \phi'(\bar{z}; u)$ *and*

$$0 \in \partial_a \langle \zeta, \phi \rangle (\bar{z}; u). \tag{3.6}$$

- *Suppose that* ϕ *is Lipschitz and directionally differentiable at* $\bar{z}$ *in direction* u. *We say that the directional quasi-normality holds at* $\bar{z}$ *in direction* u *if there exists no* $\zeta \neq 0$ *satisfying* $0 \leq \zeta \perp \phi(\bar{z}),\ \zeta \perp \phi'(\bar{z}; u)$ *such that (3.6) holds and there exists sequences* $t_k \downarrow 0,\ u^k \to u$ *satisfying*

$$\phi_j(\bar{z} + t_k u_k) > 0, \quad \textit{if } j \in \bar{I}_\phi \textit{ and } \zeta_j > 0. \tag{3.7}$$

It is easy to see that for any given direction u, if ϕ is Lipschitz and directionally differentiable at $\bar{z}$ in direction u then the following implications hold:

$$\text{NNAMCQ} \Rightarrow \ \text{FOSCMS in direction } u \Rightarrow \ \text{quasi-normality in direction } u.$$

Proposition 3.3.1. *Let* $\phi(\bar{z}) \leq 0$ *and suppose that* ϕ *is Lipschitz and directionally differentiable at* $\bar{z}$ *in direction* $u \in L(\bar{z})$. *If the directional quasi-normality holds at* $\bar{z}$ *in direction* u *for the inequality system* $\phi(z) \leq 0$. *Then the set-valued map* $\Phi(z) := -\phi(z) + \mathbb{R}^q_-$ *is metrically subregular at* $(\bar{z}, 0)$ *in direction* u.

Proof. Since ϕ is Lipschitz and directionally differentiable at $\bar{z}$ in direction u, we have $D\phi(\bar{z})(u) = \{\phi'(\bar{z}; u)\}$. By Remark 3.2.1,

$$D^*\phi(\bar{z}; (u, \phi'(\bar{z}; u)))(\zeta) = \partial \langle \zeta, \phi \rangle (\bar{z}; (u, \phi'(\bar{z}; u))) = \partial_a \langle \zeta, \phi \rangle (\bar{z}; u).$$

By equality (2.3), we have $N_{\mathbb{R}^q_-}(\phi(\bar{z}); \phi'(\bar{z}; u)) = \{\mu \in \mathbb{R}^q | 0 \leq \mu \perp \phi(\bar{z}), \mu \perp \phi'(\bar{z}; u)\}$. For any sequences $\{\zeta^k\}$ and $\{s^k\}$, if $\zeta_j > 0$ and $\widehat{N}_{\mathbb{R}^q_-}(s^k_j) \ni \zeta^k_j \to \zeta_j$, then for large enough k, $\zeta^k_j > 0$ and hence $s^k_j = 0$. Hence the condition (3.7) is equivalent to the sequential condition in [4, Definition 4.1(a)]. Therefore the quasi-normality in direction $u \in L(\bar{z})$ means that there exists no $\zeta \neq 0$ such that

$$0 \in D^*\phi(\bar{z}; (u, \phi'(\bar{z}; u)))(\zeta), \qquad \zeta \in N_{\mathbb{R}^q_-}(\phi(\bar{z}); \phi'(\bar{z}; u))$$

and there exist sequences $t_k \downarrow 0,\ u^k \to u$ such that (3.7) holds.

From the proof of [4, Lemma 3.1 and Corollary 4.1] and [23, Corollary 1], one can easily obtain that the quasi-normality at $\bar{z}$ in direction u implies that $\Phi(z)$ is metrically subregular at $(\bar{z}, 0)$ in direction u. ■

In the following theorem, we derive the directional KKT condition under the directional calmness condition.

Theorem 3.3.1. *Let $\bar{z}$ be a local minimizer of $(\widetilde{\mathrm{P}})$. Suppose that $\varphi(z)$ is continuously differentiable at $\bar{z}$ and $\phi(z)$ is Lipschitz and directionally differentiable at $\bar{z}$ in direction $u \in C(\bar{z})$. Suppose that the $(\widetilde{\mathrm{P}})$ is calm at $\bar{z}$ in direction u. Then there exists a vector $\lambda_\phi \in \mathbb{R}^q$ such that $0 \le \lambda_\phi \perp \phi(\bar{z})$, $\lambda_\phi \perp \phi'(\bar{z}; u)$ and*

$$0 \in \nabla\varphi(\bar{z}) + \partial_a \langle \lambda_\phi, \phi \rangle(\bar{z}; u).$$

Proof. Since $(\widetilde{\mathrm{P}})$ is calm at $\bar{z}$ in direction u, there exist positive scalars ϵ, δ, ρ such that

$$\varphi(z) + \rho \mathrm{dist}(\phi(z), \mathbb{R}^q_-) \ge \varphi(\bar{z}) \qquad \forall z \in \bar{z} + \mathcal{V}_{2\epsilon,\delta}(u). \tag{3.8}$$

Since $u \in C(\bar{z})$, we have $\phi(\bar{z}) + t\phi'(\bar{z}; u) \in \mathbb{R}^q_-$ and hence

$$0 \le \frac{\mathrm{dist}(\phi(\bar{z} + tu), \mathbb{R}^q_-)}{t} \le \frac{\phi(\bar{z} + tu) - \phi(\bar{z}) - t\phi'(\bar{z}; u)}{t}.$$

Since $\nabla\varphi(\bar{z})u \le 0$, it follows that $\lim_{t\downarrow 0} \frac{\mathrm{dist}(\phi(\bar{z}+tu), \mathbb{R}^q_-)}{t} = 0$.

Since $\bar{z} + tu \in \bar{z} + cl(\mathcal{V}_{\epsilon,\delta}(u))$ for t sufficiently small, by (3.8),

$$\varphi(\bar{z} + tu) + \rho \mathrm{dist}(\phi(\bar{z} + tu), \mathbb{R}^q_-) \ge \varphi(\bar{z})$$

for all t small enough. Together with $\nabla\varphi(\bar{z})u \le 0$ we have

$$\lim_{t\downarrow 0} \frac{\varphi(\bar{z} + tu) + \rho \mathrm{dist}(\phi(\bar{z} + tu), \mathbb{R}^q_-) - \varphi(\bar{z})}{t} = 0. \tag{3.9}$$

For each $k = 0, 1, \ldots$, define $\sigma_k := 2(\varphi(\bar{z} + \frac{u}{k}) + \rho \mathrm{dist}(\phi(\bar{z} + \frac{u}{k}), \mathbb{R}^q_-) - \varphi(\bar{z}))$. If $\sigma_k \equiv 0$, then for each large enough k, by (3.8), $\bar{z} + \frac{u}{k}$ is a global minimizer of the function $\varphi(z) + \rho \mathrm{dist}(\phi(z), \mathbb{R}^q_-) + \delta_{\bar{z}+cl(\mathcal{V}_{\epsilon,\delta}(u))}(z)$. Since for each large enough k, $\bar{z} + \frac{u}{k}$ is an interior point of $\bar{z} + cl(\mathcal{V}_{\epsilon,\delta}(u))(u)$, by the well-known Fermat's rule and the calculus rule (see e.g., [72, Corollary 10.9]),

$$0 \in \nabla\varphi(\bar{z} + \frac{u}{k}) + \rho\partial(\mathrm{dist}_{\mathbb{R}^q_-} \circ \phi)(\bar{z} + \frac{u}{k}). \tag{3.10}$$

Otherwise, without loss of generality, we assume that for all k, $\sigma_k > 0$. Then by definition of σ_k we have for k sufficiently large,

$$\varphi(\bar{z}+\frac{u}{k})+\rho\mathrm{dist}(\phi(\bar{z}+\frac{u}{k}),\mathbb{R}^q_-)+\delta_{\bar{z}+cl(\mathcal{V}_{\epsilon,\delta}(u))}(\bar{z}+\frac{u}{k})<\varphi(\bar{z})+\sigma_k.$$

Define $\lambda_k := \frac{2\|u\|r}{k\epsilon}\sqrt{\frac{\sigma_k k\epsilon}{2\|u\|r}}$. By Ekeland's variation principle, there exists $\tilde{z}^k$ satisfying that $\|\tilde{z}^k-(\bar{z}+\frac{u}{k})\|\leq\lambda_k$, and $\varphi(z)+\rho\mathrm{dist}(\phi(z),\mathbb{R}^q_-)+\delta_{\bar{z}+cl(\mathcal{V}_{\epsilon,\delta}(u))}(z)+\frac{\sigma_k}{\lambda_k}\|z-(\bar{z}+\frac{u}{k})\|$ attains its global minimum at $\tilde{z}^k$. Since $\frac{\epsilon u}{2\|u\|}$ is an interior point of $cl(\mathcal{V}_{\epsilon,\delta}(u))$, there exists $r\in(0,\epsilon/2)$ such that $\frac{\epsilon u}{2\|u\|}+r\mathbb{B}\subset cl(\mathcal{V}_{\epsilon,\delta}(u))$. It is obvious that the following implication holds

$$z\in cl(\mathcal{V}_{\epsilon,\delta}(u)),0\leq\alpha\leq 1\Rightarrow \alpha z\in cl(\mathcal{V}_{\alpha\epsilon,\delta}(u))$$

Hence $(\frac{\epsilon u}{2\|u\|}+r\mathbb{B})\frac{2\|u\|}{\epsilon k}\subset cl(\mathcal{V}_{\epsilon,\delta}(u))$ and hence $\bar{z}+\frac{u}{k}+\frac{2\|u\|}{k\epsilon}r\mathbb{B}\subset(\bar{z}+cl(\mathcal{V}_{\epsilon,\delta}(u)))$ and since $\sigma_k=o(\frac{1}{k})$ by (3.9), $\tilde{z}^k$ is in the interior of $\bar{z}+cl(\mathcal{V}_{\epsilon,\delta}(u))$. Then by the well-known Fermat's rule, we obtain

$$0\in\nabla\varphi(\tilde{z}^k)+\rho\partial(\mathrm{dist}_{\mathbb{R}^q_-}\circ\phi)(\tilde{z}^k)+\frac{\sigma_k}{\lambda_k}\bar{\mathbb{B}}. \tag{3.11}$$

Since ϕ is Lipschitz continuous near $\bar{z}$ in direction u, it is Lipschitz continuous at $\tilde{z}^k$ for k large enough. So by the chain rule for limiting subdifferential [59, Corollary 3.43], we have

$$\partial(\mathrm{dist}_{\mathbb{R}^q_-}\circ\phi)(\tilde{z}^k)\subseteq\cup_{\zeta'\in\partial\mathrm{dist}_{\mathbb{R}^q_-}(\phi(\tilde{z}^k))}\partial\langle\zeta',\phi\rangle(\tilde{z}^k).$$

Therefore by (3.10) or (3.11), $\exists\zeta^k\in\partial\mathrm{dist}_{\mathbb{R}^q_-}(\phi(\bar{z}+\frac{u}{k}))$ or $\exists\zeta^k\in\partial\mathrm{dist}_{\mathbb{R}^q_-}(\phi(\tilde{z}^k))$ such that

$$0\in\nabla\varphi(\bar{z}+\frac{u}{k})+\rho\partial\langle\zeta^k,\phi\rangle(\bar{z}+\frac{u}{k}),\ \text{or}\ 0\in\nabla\varphi(\tilde{z}^k)+\rho\partial\langle\zeta^k,\phi\rangle(\tilde{z}^k)+\frac{\sigma_k}{\lambda_k}\bar{\mathbb{B}}. \tag{3.12}$$

Since distance functions are Lipschitz, by [72, Theorem 9.13], $\{\zeta^k\}$ is bounded. Without loss of generality, there exists $\zeta:=\lim_k\zeta^k$. By the way, one can easily obtain that $\lim_k(\bar{z}+u/k-\bar{z})/\frac{1}{k}=\lim_k(\tilde{z}^k-\bar{z})/\frac{1}{k}=u$. Since $\sigma_k=o(\frac{1}{k})$, $\lim_k\frac{\sigma_k}{\lambda_k}=0$. Taking the limit of (3.12) as $k\to\infty$, by Proposition 3.2.2 we have

$$0\in\nabla\varphi(\bar{z})+\rho\partial_a\langle\zeta,\phi\rangle(\bar{z};u).$$

Moreover by [7, Corollary 4.2], $\zeta \in \partial_a \mathrm{dist}_{\mathbb{R}^q_-}(\phi(\bar{z}); \phi'(\bar{z}; u)) \subseteq N_{\mathbb{R}^q_-}(\phi(\bar{z}); \phi'(\bar{z}; u))$. The desired result holds by taking $\lambda_\phi := \rho\zeta \in N_{\mathbb{R}^q_-}(\phi(\bar{z}); \phi'(\bar{z}; u)) = \{\xi \in \mathbb{R}^q | 0 \leq \xi \perp \phi(\bar{z}), \xi \perp \phi'(\bar{z}; u)\}$. ■

We now give an example of a bilevel program where the partial calmness and calmness fail but the calmness condition holds in a nonzero critical direction.

Example 3.3.1. *Consider the following bilevel program:*

$$\begin{array}{lll} \text{(BP)} & \min & F(x,y) := (x-y-1)^{\frac{5}{3}} + 4(x+y+1)^{\frac{5}{3}} \\ & s.t. & -1 \leq x \leq 1, y \in S(x), \end{array}$$

where for each x, $S(x)$ is the solution set for the lower level program:

$$\min_y \{f(x,y) := -(x+y)^2 + x^3(x+y-1), s.t. \ -y-x-1 \leq 0, y+x-1 \leq 0\}.$$

It is easy to see that the solution mapping $S(x)$ of the lower level problem is equal to

$$S(x) = \begin{cases} -x-1, & x > 0, \\ \{-1, 1\}, & x = 0, \\ -x+1, & x < 0 \end{cases} \tag{3.13}$$

And the global optimal solution of (BP) is $(\bar{x}, \bar{y}) = (0, -1)$. The constraints $y+x-1 \leq 0$ and $-1 \leq x \leq 1$ are inactive at $(0, -1)$. The value function

$$V(x) = \begin{cases} -1-2x^3 & x > 0 \\ -1 & x \leq 0 \end{cases}. \tag{3.14}$$

First, we prove that the partial calmness condition fails at $(\bar{x}, \bar{y})$. For any scalar $\rho > 0$, consider the partially penalized problem:

$$\begin{array}{lll} (VP)_\rho & \min & F(x,y) + \rho(f(x,y) - V(x)) \\ & s.t. & g_1(x,y) := -y-x-1 \leq 0, g_2(x,y) := y+x-1 \leq 0, \\ & & -1 \leq x \leq 1. \end{array}$$

Since $-1 < \bar{x} < 1, g_1(\bar{x},\bar{y}) = 0,\ g_2(\bar{x},\bar{y}) < 0$, *by (3.13)-(3.14), the critical cone is*

$$\begin{aligned} C(\bar{x},\bar{y}) &= \{(u,v) | \nabla F(\bar{x},\bar{y})(u,v) \leq 0, \nabla f(\bar{x},\bar{y})(u,v) - V'(\bar{x};u) = 0, \nabla g_1(\bar{x},\bar{y})(u,v) \leq 0\} \\ &= \{(u,v) | u + v = 0\}. \end{aligned}$$

Consider the sequence $(x^k, y^k) := (-\frac{1}{k}, \frac{1}{k} - 1)$ *which are feasible to* $(VP)_\rho$ *and converges to* $(\bar{x},\bar{y})$. *Since* $F(x^k,y^k) = -(\frac{2}{k})^{\frac{5}{3}},\ f(x^k,y^k) = -1 + \frac{2}{k^3}$ *and by* (3.14), $V(x^k) = -1$, *we have* $F(x^k,y^k) + \rho(f(x^k,y^k) - V(x^k)) = -(\frac{2}{k})^{\frac{5}{3}} + \frac{2\rho}{k^3}$. *Hence for* k *sufficiently large, we have*

$$F(x^k,y^k) + \rho(f(x^k,y^k) - V(x^k)) < 0 = F(\bar{x},\bar{y}) + \rho(f(\bar{x},\bar{y}) - V(\bar{x})).$$

This means that for any $\rho > 0$, $(\bar{x},\bar{y})$ *is not a local minimizer of* $(VP)_\rho$. *Hence, the partial calmness fails. Since the calmness condition is in general stronger than partial calmness, the calmness condition also fails. In fact for this example since the constraint functions for* $(VP)_\rho$ *are all affine, the partial calmness is equivalent to the fully calmness. Notice that* $(x^k,y^k) \to (\bar{x},\bar{y})$ *in direction* $(-1,1)$ *and so we have shown that problem (VP) is not calm in direction* $(-1,1)$. *Next, we prove that (VP) is calm at* $(\bar{x},\bar{y})$ *in direction* $(1,-1) \in C(\bar{x},\bar{y})$. *Since the constraints* $g_2(x,y) \leq 0$ *and* $-1 \leq x \leq 1$ *are inactive at* $(\bar{x},\bar{y}) = (0,-1)$, *it suffices to show that there exists a positive scalar* ρ *such that for any sequences* $t_k \downarrow 0,\ (u^k,v^k) \to (\bar{u},\bar{v}) := (1,-1)$, *for* k *sufficiently large,*

$$\begin{aligned} F(\bar{x}+t_k u^k, \bar{y}+t_k v^k) + \rho \text{dist}(f(\bar{x}+t_k u^k, \bar{y}+t_k v^k) - V(\bar{x}+t_k u^k), \mathbb{R}_-) \\ + \rho \text{dist}(g_1(\bar{x}+t_k u^k, \bar{y}+t_k v^k), \mathbb{R}_-) - F(\bar{x},\bar{y}) \geq 0. \end{aligned} \tag{3.15}$$

Suppose that $g_1(\bar{x}+t_k u^k, \bar{y}+t_k v^k) \leq 0$, *then for* k *sufficiently large,* $\bar{y}+t_k v^k$ *is a feasible solution for* $(P_{\bar{x}+t_k u^k})$ *and hence* $f(\bar{x}+t_k u^k, \bar{y}+t_k v^k) - V(\bar{x}+t_k u^k) \geq 0$ *by the definition of the value function. Moreover* $F(\bar{x}+t_k u^k, \bar{y}+t_k v^k) \geq F(\bar{x},\bar{y})$. *Hence (3.15) holds. Otherwise suppose that*

$$g_1(\bar{x}+t_k u^k, \bar{y}+t_k v^k) = -t_k(u^k+v^k) > 0.$$

Hence $t_k(u^k+v^k) < 0$. *Together with* $u^k > 0, t_k > 0$, *we can verify that* $f(\bar{x}+t_k u^k, \bar{y}+$

$t_k v^k) - V(\bar{x} + t_k u^k) < 0$. *Also since* $t_k \downarrow 0, -(u^k + v^k) \downarrow 0$, *we have*

$$\begin{aligned} & F(\bar{x} + t_k u^k, \bar{y} + t_k v^k) - F(\bar{x}, \bar{y}) + \rho \text{dist}(g_1(\bar{x} + t_k u^k, \bar{y} + t_k v^k), \mathbb{R}_-) \\ = & t_k^{\frac{5}{3}} (u^k + v^k)^{\frac{5}{3}} - t_k (u^k + v^k) \geq 0. \end{aligned}$$

Hence, we obtain (3.15). *Consequently* (VP) *is calm at* $(\bar{x}, \bar{y})$ *in direction* $(\bar{u}, \bar{v}) = (1, -1)$.

3.4 Directional sensitivity analysis of the value function

In this section we study the directional sensitivity analysis of the value function of the lower level program (P_x). Recall that $S(x)$ denotes the solution set of (P_x) and $V(x)$ is the value function. The results of this section could be of independent interest.

First we give some preliminary results that will be needed. We first introduce a directional version of the restricted inf-compactness condition which was first introduced in [12, Hypothesis 6.5.1] with the terminology introduced in [31, Definition 3.8].

Definition 3.4.1 (Directional Restricted Inf-compactness)**.** *We say that the restricted inf-compactness holds at* $\bar{x}$ *in direction* u *with compact set* $\Omega_u \subseteq \mathbb{R}^n$ *if* $V(\bar{x})$ *is finite and there exists positive numbers* $\epsilon > 0, \delta > 0$ *such that for all* $x \in \bar{x} + \mathcal{V}_{\epsilon,\delta}(u)$ *with* $V(x) < V(\bar{x}) + \epsilon$, *one always has* $S(x) \cap \Omega_u \neq \emptyset$. *When* $u = 0$ *in the above, we say the restricted inf-compactness holds at* $\bar{x}$.

Next we introduce a directional version of the inf-compactness condition (see e.g., [10, Page 272]). It is not difficult to verify that the directional inf-compactness implies the directional restricted inf-compactness.

Definition 3.4.2 (Directional Inf-compactness)**.** *We say that the inf-compactness holds at* $\bar{x}$ *in direction* u *if there exist a compact set* $\Lambda_u \subseteq \mathbb{R}^n$, *and positive numbers* $\alpha > V(\bar{x})$, ϵ, δ *such that for all* $x \in \bar{x} + \mathcal{V}_{\epsilon,\delta}(u)$, *one always has* $\{y | f(x, y) \leq \alpha,\ g(x, y) \leq 0\} \neq \emptyset$ *and contained in* Λ_u. *When* $u = 0$ *in the above, we say the inf-compactness holds at* $\bar{x}$.

When $f(x,y)$ satisfies the growth condition, i.e., there exists $\delta > 0$ such that the set

$$\{y \in \mathbb{R}^m | g(\bar{x},y) \leq \alpha,\ f(\bar{x},y) \leq M,\ \alpha \in \delta\mathbb{B}\}$$

is bounded for each $M \in \mathbb{R}$, the inf-compactness holds at $\bar{x}$. Similarly, if $f(x,y)$ is coercive or level bounded, the inf-compactness holds at $\bar{x}$.

The following definition gives a directional version of the classical inner semi-continuity (see e.g. [59, Definition 1.63]).

Definition 3.4.3 (Directional Inner Semi-continuity)**.** *Given $\bar{y} \in S(\bar{x})$, we say that the optimal solution map $S(x)$ is inner semi-continuous at $(\bar{x},\bar{y})$ in direction u, if for any sequences $t_k \downarrow 0, u^k \to u$, there exists a sequence $y^k \in S(\bar{x} + t_k u^k)$ converging to $\bar{y}$. When $u = 0$ in the above, we say that $S(x)$ is inner semi-continuous at $(\bar{x},\bar{y})$.*

Note that [54, Definition 4.4(i)] introduced a directional inner semicontinuity which requires $y^k \xrightarrow{v} \bar{y}$ for some v. Since $y^k \xrightarrow{v} \bar{y}$ implies that $y^k \to \bar{y}$, their directional inner semicontinuity is stronger than ours. Given a direction we define a subset of the solution $S(\bar{x})$ as below. It coincides with the solution set when $u = 0$ and may be strictly contained in the solution set if the direction u is nonzero.

Definition 3.4.4 (Directional Solution)**.** *The optimal solution in direction u is the set defined by*

$$S(\bar{x};u) = \{y \in S(\bar{x}) | \exists t_k \downarrow 0, u^k \to u, y^k \to y, y^k \in S(\bar{x} + t_k u^k)\}.$$

If $\bar{y} \in S(\bar{x};u)$, then $\bar{y}$ is upper stable in direction u in the sense of Janin (see [44, Definition 3.4]).

It is obvious that if the optimal solution map $S(x)$ is inner semi-continuous at $(\bar{x},\bar{y}) \in gphS$ in direction u, then $\bar{y} \in S(\bar{x};u)$.

Denote the feasible map of the problem (P_x) by

$$\mathcal{F}(x) := \{y \in \mathbb{R}^m | g(x,y) \leq 0\}$$

and the active index set $I_g(x,y) := \{i = 1,\ldots,p | g_i(x,y) = 0\}$.

Definition 3.4.5 (RCR regularity)**.** *([57, Definition 1]) We say that the the feasible map $\mathcal{F}(x)$ is relaxed constant rank (RCR) regular at $(\bar{x},\bar{y}) \in \mathrm{gph}\mathcal{F}$ if there exists $\delta > 0$ such that for any index subset $K \subseteq I_g(\bar{x},\bar{y})$, the family of gradient vectors $\nabla_y g_j(x,y), j \in K$, has the same rank at all points $(x,y) \in \mathbb{B}_\delta(\bar{x},\bar{y})$.*

The following lemma shows that the RCR regularity condition is slightly stronger than the calmness of the feasible map $\mathcal{F}$ at $(\bar{x}, \bar{y}) \in gph\mathcal{F}$ and is needed in Proposition 3.4.3 and Theorem 3.4.2.

Lemma 3.4.1. *(see [57, Lemma 5]) Suppose that the feasible map $\mathcal{F}(x)$ is RCR regular at $(\bar{x}, \bar{y}) \in \text{gph}\mathcal{F}$. Then there exist $\delta > 0, \kappa > 0$ such that for any $x \in \mathbb{B}_\delta(\bar{x})$ and $y \in \mathbb{B}_\delta(\bar{y}) \cap \mathcal{F}(x)$, there exists $\tilde{y} \in \mathcal{F}(\bar{x})$ such that $\|y - \tilde{y}\| \leq \kappa\|x - \bar{x}\|$ and $g_j(x, y) \leq g_j(\bar{x}, \tilde{y}) \leq 0$ for each $j \in I_g(\bar{x}, \bar{y})$.*

We now define a directional version of the Robinson Stability [26, Definition 1.1].

Definition 3.4.6 (Directional Robinson stability)**.** *We say that the feasible map $\mathcal{F}(x)$ satisfies Robinson stability* (RS) *property at $(\bar{x}, \bar{y}) \in \text{gph}\mathcal{F}$ in direction $u \in \mathbb{R}^n$ if there exist positive scalars κ, ϵ, δ such that*

$$\text{dist}(y, \mathcal{F}(x)) \leq \kappa \text{dist}(g(x, y), \mathbb{R}^p_-) \quad \forall x \in \bar{x} + \mathcal{V}_{\epsilon,\delta}(u), y \in \mathbb{B}_\epsilon(\bar{y}). \tag{3.16}$$

If RS holds at $(\bar{x}, \bar{y})$ in direction $u = 0$, we say that RS holds at $(\bar{x}, \bar{y})$ ([26, Definition 1.1]). Note that RS in direction u is equivalent to R-regularity with respect to set $\bar{x} + \mathcal{V}_{\epsilon,\delta}(u)$ as defined in [57, Definition 2].

Proposition 3.4.1 (Sufficient Conditions for RS)**.** *If the system $g(x, y) \leq 0$ satisfies one of the following conditions at $(\bar{x}, \bar{y})$, then RS holds at $(\bar{x}, \bar{y})$.*

- *$g(x, y)$ is the sum of a continuous function of x and an affine function of y, with $\mathcal{F}(x)$ nonempty near $\bar{x}$: $g(x, y) = p(x) + q(y)$ with $p(x)$ continuous and $q(y)$ affine.*

- *Partial linear independence constraint qualification: g is continuously differentiable at $(\bar{x}, \bar{y})$ and the set $\{\nabla_y g_i(\bar{x}, \bar{y}) | i \in I_g(\bar{x}, \bar{y})\}$ is linearly independent.*

- *Partial NNAMCQ: g is continuously differentiable at $(\bar{x}, \bar{y})$ and there exists no nonzero vector $\lambda \in \mathbb{R}^p_+$ such that $\lambda \perp g(\bar{x}, \bar{y})$ and $\nabla_y g(\bar{x}, \bar{y})^T \lambda = 0$.*

Note that the RS under the partial linear constraint qualification follows from Hoffman's lemma. One can refer to [57, 31, 26] and the references therein for more sufficient conditions for RS. The following proposition shows that the directional RS implies the directional metric subregularity.

Proposition 3.4.2. *Suppose that the feasible map $\mathcal{F}$ satisfies RS at $(\bar{x}, \bar{y}) \in \text{gph}\mathcal{F}$ in direction u. Then the metric subregularity of the system $g(x, y) \leq 0$ holds at $((\bar{x}, \bar{y}), 0)$ in direction (u, v) for any $v \in \mathbb{R}^m$.*

Proof. Since RS for $\mathcal{F}$ holds at $(\bar{x}, \bar{y})$ in direction u, i.e., there exist numbers $\kappa > 0, \epsilon > 0, \delta > 0$ such that

$$dist(y, \mathcal{F}(x)) \leq \kappa dist(g(x, y), \mathbb{R}^p_-),$$

for all $x \in \bar{x} + \mathcal{V}_{\epsilon,\delta}(u)$ and $y \in \mathbb{B}_\epsilon(\bar{y})$. Then we obtain

$$dist((x, y), g^{-1}(\mathbb{R}^p_-)) = dist((x, y), gph\mathcal{F}) \leq dist(y, \mathcal{F}(x)) \leq \kappa dist(g(x, y), \mathbb{R}^p_-),$$

for all $x \in \bar{x} + \mathcal{V}_{\epsilon,\delta}(u)$ and $y \in \mathbb{B}_\epsilon(\bar{y})$. This means that the metric subregularity of $g(x, y) \leq 0$ holds at $(\bar{x}, \bar{y}, 0)$ in direction (u, v) for any $v \in \mathbb{R}^m$. ■

Recall that the lower Dini directional derivative of the feasible map $\mathcal{F}(x)$ at a point $(\bar{x}, \bar{y}) \in \text{gph}\mathcal{F}$ in direction u is defined as

$$D_+\mathcal{F}(\bar{x}, \bar{y}; u) := \liminf_{t \downarrow 0} \frac{\mathcal{F}(\bar{x} + tu) - \bar{y}}{t} = \{v | \exists o(t) \text{ s.t. } \bar{y} + tv + o(t) \in \mathcal{F}(\bar{x} + tu)\}.$$

Define the y-projection of the linearization cone of gph$\mathcal{F}$ at $(\bar{x}, \bar{y})$ in direction u, i.e.,

$$\mathbb{L}(\bar{x}, \bar{y}; u) := \{v \in \mathbb{R}^m | \nabla g_i(\bar{x}, \bar{y})(u, v) \leq 0, i \in I_g(\bar{x}, \bar{y})\}.$$

By definition, one always has $D_+\mathcal{F}(\bar{x}, \bar{y}; u) \subseteq \mathbb{L}(\bar{x}, \bar{y}; u)$. Since the directional MPEC R-regularity introduced in [31, Lemma 3.3] is weaker than our directional RS and (P_x) is a special case of the problem studied in [31] when the equilibrium constraints are omitted, the following results follow from [31, Lemmas 3.3, 3.5].

Lemma 3.4.2. *[31, Lemmas 3.3, 3.5] Let $\bar{y} \in \mathcal{F}(\bar{x})$. Suppose either the feasible map $\mathcal{F}$ satisfies RS at $(\bar{x}, \bar{y})$ in direction u or $D_+\mathcal{F}(\bar{x}, \bar{y}; u) \neq \emptyset$ and $\mathcal{F}$ is RCR-regular at $(\bar{x}, \bar{y})$ in direction u. Then $D_+\mathcal{F}(\bar{x}, \bar{y}; u) = \mathbb{L}(\bar{x}, \bar{y}; u)$.*

The following results will be needed in Corollary 3.4.1 and Theorem 3.4.1.

Lemma 3.4.3. *Suppose that the restricted inf-compactness holds at $\bar{x}$ in direction u with compact set Ω_u and there exists $\bar{y} \in S(\bar{x})$ such that $\mathcal{F}$ satisfies RS at $(\bar{x}, \bar{y})$ in* *direction u. Then $D_+\mathcal{F}(\bar{x},\bar{y};u) = \mathbb{L}(\bar{x},\bar{y};u) \neq \emptyset$, $\bar{y} \in \liminf_{x \xrightarrow{u} \bar{x}} \mathcal{F}(x)$ and $S(\bar{x};u) \neq \emptyset$. And for any $l > 0, \exists \epsilon, \delta > 0$ such that for $\forall x \in \bar{x} + V_{\epsilon,\delta}(u)$, $\exists y \in S(x) \cap \Omega_u$ satisfying* $\mathrm{dist}(y, S(\bar{x};u) \cap \Omega_u) < l$.

Proof. Since RS holds at $(\bar{x},\bar{y})$ in direction u, by Lemma 3.4.2, $D_+\mathcal{F}(\bar{x},\bar{y};u) = \mathbb{L}(\bar{x},\bar{y};u)$. Moreover by (3.16) there exist positive scalars κ, ϵ, δ, such that for any $x \in \bar{x} + \mathcal{V}_{\epsilon,\delta}(u)$,

$$\mathrm{dist}(\bar{y}, \mathcal{F}(x)) \leq \kappa \mathrm{dist}(g(x,\bar{y}), \mathbb{R}^p_-) \leq \kappa \|g(x,\bar{y}) - g(\bar{x},\bar{y})\| \leq L_g\kappa\|x - \bar{x}\|, \tag{3.17}$$

where $L_g > 0$ is the Lipschitz modulus of $g(x,\bar{y})$ around $\bar{x}$. Then for any sequences $t_k \downarrow 0, u^k \to u$, by (3.17) we can find a sequence $y^k \in \mathcal{F}(\bar{x}+t_ku^k)$ such that $\|\bar{y}-y^k\| \leq L_g\kappa\|\bar{x} + t_ku^k - \bar{x}\|$, which implies that $y^k \to \bar{y}$. By Definition 3.2.1, this means that $\bar{y} \in \liminf_{x \xrightarrow{u} \bar{x}} \mathcal{F}(x)$. Since $\{\frac{y^k-\bar{y}}{t_k}\}$ is bounded, taking a subsequence if necessary, we can find $v \in \mathbb{R}^m$ such that $v^k := \frac{y^k-\bar{y}}{t_k}$ converges to v. Since for each $i \in I_g(\bar{x},\bar{y})$, $g_i(\bar{x}+t_ku^k, y^k) \leq 0$, it follows that $v \in \mathbb{L}(\bar{x},\bar{y};u)$. We also have $\limsup_k V(\bar{x}+t_ku^k) \leq \lim_k f(\bar{x}+t_ku^k, y^k) = V(\bar{x})$. It follows that since the restricted inf-compactness holds at $\bar{x}$ in direction u, for each k large enough, there exists $\tilde{y}^k \in S(\bar{x}+t_ku^k) \cap \Omega_u$. By the compactness of Ω_u, the sequence $\{\tilde{y}^k\}$ is bounded. Without loss of generality, assume $\tilde{y} := \lim_k \tilde{y}^k \in \Omega_u$. Since

$$f(\bar{x},\tilde{y}) = \lim_k f(\bar{x}+t_ku^k, \tilde{y}^k) = \lim_k V(\bar{x}+t_ku^k) \leq \lim_k f(\bar{x}+t_ku^k, \bar{y}+t_kv^k) = f(\bar{x},\bar{y}) = V(\bar{x})$$

and

$$g(\bar{x},\tilde{y}) = \lim_k g(\bar{x}+t_ku^k, \tilde{y}^k) \leq 0,$$

we can obtain $\tilde{y} \in S(\bar{x})$. Consequently, $\tilde{y} \in S(\bar{x};u) \cap \Omega_u$.

We prove the last statement by contradiction. Assume there exist $l > 0$ and for k large enough, $\bar{x} + t_ku^k \in \bar{x} + \mathcal{V}_{\frac{1}{k},\frac{1}{k}}(u)$ and $\tilde{y}^k \in S(\bar{x}+t_ku^k) \cap \Omega_u$ such that $\mathrm{dist}(\tilde{y}^k, S(\bar{x};u) \cap \Omega_u) \geq l$. Taking the limit as $k \to \infty$, $\mathrm{dist}(\tilde{y}, S(\bar{x};u) \cap \Omega_u) \geq l$, which contradicts $\tilde{y} \in S(\bar{x};u) \cap \Omega_u$. The proof is complete. ■

In general there may not exist relationship between RCR-regularity and RS condition. However under the inner semicontinuity of $S(x)$, we can show that RCR-regularity implies RS/R-regularity.

Lemma 3.4.4. *Let $\bar{y} \in S(\bar{x})$ and $S(x)$ is inner semi-continuous at $(\bar{x},\bar{y})$ in direction*

u. Suppose that $\mathcal{F}$ is RCR regular at $(\bar{x}, \bar{y})$ in direction u. Then $\mathcal{F}$ satisfies RS at $(\bar{x}, \bar{y})$ in direction u.

Proof. We approve the lemma by contradiction. Assume RS does not hold at $(\bar{x}, \bar{y})$ in direction u. Then there exist sequences $x^k \xrightarrow{u} \bar{x}$ and $y^k \to \bar{y}$ satisfying that

$$\text{dist}(y^k, \mathcal{F}(x^k)) > k\text{dist}(g(x^k, y^k), \mathbb{R}^p_-). \tag{3.18}$$

Since $S(x)$ is inner semi-continuous at $(\bar{x}, \bar{y})$ in direction u, we have

$$\bar{y} \in \liminf_{x \xrightarrow{u} \bar{x}} S(x) \subseteq \liminf_{x \xrightarrow{u} \bar{x}} \mathcal{F}(x).$$

Then for sufficiently large k there exists a sequence $\tilde{y}^k \in \mathcal{F}(x^k)$ such that $\tilde{y}^k \to \bar{y}$. Let $\bar{y}^k$ be the projection of y^k onto $\mathcal{F}(x^k)$. We obtain

$$\|y^k - \bar{y}^k\| \leq \|y^k - \tilde{y}^k\| \to 0 \text{ as } k \to \infty.$$

Then following the proof of [31, Lemma 3.5] for the case when the number of complementarity constraints is 0, we can find some scalar $M > 0$ such that

$$\text{dist}(y^k, \mathcal{F}(x^k)) \leq M\text{dist}(g(x^k, y^k), \mathbb{R}^p_-)$$

contradicting (3.18). Hence, the assumption is false and RS for $\mathcal{F}$ holds at $(\bar{x}, \bar{y})$ in direction u. ■

Define the Lagrange function of (P_x) by

$$\mathcal{L}(x, y; \lambda) := f(x, y) + g(x, y)^T \lambda.$$

From now on in this section we assume that the functions f, g are continuously differentiable. Then the set of Lagrange multiplier associated with $y \in \mathcal{F}(x)$ is

$$\Lambda(x, y) := \{\lambda \in \mathbb{R}^p | \nabla_y \mathcal{L}(x, y; \lambda) = 0,\ g(x, y)^T \lambda = 0,\ \lambda \geq 0\}.$$

3.4.1 Directional derivative of the value function

In this subsection, we study the directional differentiability of the value function. In the following proposition we derive the formula for the directional derivative of the value function. Our result improves the corresponding classical results in [57, Theorem 5] and [31, Theorem 3.9] in that weaker assumptions are required and in the formula the directional solution instead of the solution set is used.

Proposition 3.4.3. *Let u be a direction such that $S(\bar{x};u) \neq \emptyset$ and $D_+\mathcal{F}(\bar{x},y;u) \neq \emptyset$ $\forall y \in S(\bar{x};u)$. Suppose that the restricted inf-compactness holds at $\bar{x}$ in direction u. Moreover assume that $\mathcal{F}(x)$ is RCR-regular at each $y \in S(\bar{x};u)$. Then the value function $V(x)$ is directionally differentiable at $x = \bar{x}$ in direction u and*

$$V'(\bar{x};u) = \min_{y \in S(\bar{x};u)} \min_{v \in \mathbb{L}(\bar{x},y;u)} \nabla f(\bar{x},y)(u,v) = \min_{y \in S(\bar{x};u)} \max_{\lambda \in \Lambda(\bar{x},y)} \nabla_x \mathcal{L}(\bar{x},y;\lambda)u. \tag{3.19}$$

Proof. Since for any given $y \in S(\bar{x};u)$, $D_+\mathcal{F}(\bar{x},y;u) \neq \emptyset$, there is $v \in D_+\mathcal{F}(\bar{x},y;u)$. It follows that there exists $o(t)$ such that $y + tv + o(t) \in \mathcal{F}(\bar{x}+tu)$ for $t \geq 0$. Thus we have

$$\begin{aligned} V'_+(\bar{x};u) := \limsup_{t \downarrow 0} \frac{V(\bar{x}+tu) - V(\bar{x})}{t} \quad &\leq \quad \limsup_{t \downarrow 0} \frac{f(\bar{x}+tu, y+tv+o(t)) - f(\bar{x},y)}{t} \\ &= \quad \nabla f(\bar{x},y)(u,v). \end{aligned} \tag{3.20}$$

On the other hand, let $t_k \downarrow 0$ be the sequence satisfying

$$V'_-(\bar{x};u) := \liminf_{t \downarrow 0} \frac{V(\bar{x}+tu) - V(\bar{x})}{t} = \lim_{k \to \infty} \frac{V(\bar{x}+t_k u) - V(\bar{x})}{t_k}.$$

By (3.20), for any $\epsilon > 0$ and any sequence $t_k \downarrow 0$, $V(\bar{x}+t_k u) < V(\bar{x}) + \epsilon$ for k large enough. Since the restricted inf-compactness holds at $\bar{x}$ in direction u with a compact set Ω_u, there exists a sequence $y^k \in S(\bar{x}+t_k u) \cap \Omega_u$ for k large enough. Without loss of generality, define $\tilde{y} := \lim_k y^k$. Then

$$\begin{aligned} f(\bar{x},\tilde{y}) &= \lim_{k \to \infty} f(\bar{x}+t_k u, y^k) = \lim_{k \to \infty} V(\bar{x}+t_k u) \leq V(\bar{x}) \\ g(\bar{x},\tilde{y}) &= \lim_{k \to \infty} g(\bar{x}+t_k u, y^k) \leq 0. \end{aligned}$$

This means $\tilde{y} \in S(\bar{x}) \cap \Omega_u$. Moreover it is clear that $\tilde{y} \in S(\bar{x};u) \cap \Omega_u$.

Since $\mathcal{F}$ is RCR regular at each $y \in S(\bar{x};u)$ and $D_+\mathcal{F}(\bar{x},y;u) \neq \emptyset$, by Lemma 3.4.2 we have

$$D_+\mathcal{F}(\bar{x},y;u) = \mathbb{L}(\bar{x},y;u) \quad \forall y \in S(\bar{x};u). \tag{3.21}$$

Moreover by Lemma 3.4.1, for sufficiently large k, there exist $\kappa > 0$ independent of k

and a sequence $\bar{y}^k \in \mathcal{F}(\bar{x})$ such that

$$\|y^k - \bar{y}^k\| \leq \kappa\|\bar{x} + t_k u - \bar{x}\|,\ g_j(\bar{x} + t_k u, y^k) - g_j(\bar{x}, \bar{y}^k) \leq 0,\ j \in I_g(\bar{x}, \tilde{y}).$$

Consequently, $\{\frac{y^k - \bar{y}^k}{t_k}\}$ is bounded. Taking a subsequence if necessary, we assume that $\tilde{v} := \lim_{k\to\infty} \frac{y^k - \bar{y}^k}{t_k}$ and then $y^k = \bar{y}^k + t_k\tilde{v} + o(t_k)$. Thus, we obtain $\nabla g_i(\bar{x}, \tilde{y})(u, \tilde{v}) \leq 0, i \in I_g(\bar{x}, \tilde{y})$. This implies that $\tilde{v} \in \mathbb{L}(\bar{x}, \tilde{y}; u)$. Furthermore, since $\bar{y}^k \in \mathcal{F}(\bar{x})$, we have

$$\begin{aligned} V'_-(\bar{x}; u) \quad &= \lim_{k\to\infty} \frac{V(\bar{x} + t_k u) - V(\bar{x})}{t_k} \\ &\geq \lim_{k\to\infty} \frac{f(\bar{x} + t_k u, y^k) - f(\bar{x}, \bar{y}^k)}{t_k} \\ &= \lim_{k\to\infty} \frac{f(\bar{x} + t_k u, \bar{y}^k + t_k\tilde{v} + o(t_k)) - f(\bar{x}, \bar{y}^k)}{t_k} \\ &= \nabla f(\bar{x}, \tilde{y})(u, \tilde{v}). \end{aligned} \tag{3.22}$$

It follows that

$$V'_-(\bar{x}; u) \geq \nabla f(\bar{x}, \tilde{y})(u, \tilde{v}) \geq \min_{y \in S(\bar{x};u)\cap\Omega_u} \inf_{v \in \mathbb{L}(\bar{x},y;u)} \nabla f(\bar{x}, y)(u, v). \tag{3.23}$$

Since (3.20) holds for any $y \in S(\bar{x}; u) \subseteq S(\bar{x})$ and $v \in D_+\mathcal{F}(\bar{x}, y; u) = \mathbb{L}(\bar{x}, y; u)$, where the equality follows from (3.21), we have

$$V'_+(\bar{x}; u) \leq \inf_{y \in S(\bar{x};u)} \inf_{v \in \mathbb{L}(\bar{x},y;u)} \nabla f(\bar{x}, y)(u, v) \leq \min_{y \in S(\bar{x};u)\cap\Omega_u} \inf_{v \in \mathbb{L}(\bar{x},y;u)} \nabla f(\bar{x}, y)(u, v). \tag{3.24}$$

(3.23) and (3.24) imply that

$$V'_-(\bar{x}; u) \geq \inf_{y \in S(\bar{x};u)} \inf_{v \in \mathbb{L}(\bar{x},y;u)} \nabla f(\bar{x}, y)(u, v) \geq V'_+(\bar{x}; u).$$

Hence $V(x)$ is directionally differentiable at $\bar{x}$ in direction u with the first equality in (3.19) holds. And the minimum with respect to y in (3.19) can be attained on the set $S(\bar{x}; u) \cap \Omega_u$. By the linear programming duality theorem, the second equality in (3.19) holds and the minimum with respect to v can be attained. ■

In Proposition 3.4.3, the sets $S(\bar{x}; u)$ and $D_+\mathcal{F}(\bar{x}, y; u)$ are required to be both nonempty. However by Lemma 3.4.3 this condition can be guaranteed if in addition $\mathcal{F}$

satisfies RS in direction u. Consequently we have the following corollary. It improves the result of [31, Theorem 3.11] in that the NNAMCQ holding at each $y \in S(\bar{x})$ is replaced by the directional RS which is in general weaker.

Corollary 3.4.1. *Assume that $\mathcal{F}$ is RCR-regular at each $(\bar{x}, y) \in \mathrm{gph}S$. Suppose that the restricted inf-compactness holds at $\bar{x}$ in direction u and RS is satisfied at each $(\bar{x}, y) \in \mathrm{gph}S$ in direction u. Then the value function is directionally differentiable in direction u and (3.19) holds.*

In general, according to Corollary 3.4.1, one needs to ensure both RS and RCR regularity for the existence of the directional derivative. However thanks to Lemma 3.4.4, if the solution set $S(x)$ is inner semi-continuous, only RCR-regularity is needed.

Proposition 3.4.4. *Suppose that the solution set $S(x)$ is inner semi-continuous at $(\bar{x}, \bar{y}) \in \mathrm{gph}S$ in direction u. Moreover assume that $\mathcal{F}$ is RCR-regular at $(\bar{x}, \bar{y})$. Then the value function $V(x)$ is directionally differentiable at $\bar{x}$ in direction u and*

$$V'(\bar{x}; u) = \min_{v \in \mathbb{L}(\bar{x}, \bar{y}; u)} \nabla f(\bar{x}, \bar{y})(u, v) = \max_{\lambda \in \Lambda(\bar{x}, \bar{y})} \nabla_x \mathcal{L}(\bar{x}, \bar{y}; \lambda) u.$$

Proof. Since $S(x)$ is inner semi-continuous at $(\bar{x}, \bar{y}) \in \mathrm{gph}S$ in direction u we have that the restricted inf-compactness holds at $(\bar{x}, \bar{y})$ in direction u holds and by Lemma 3.4.4 both RCR and RS holds at $(\bar{x}, \bar{y})$ in direction u. By definition of the directional inner semicontinuity of $S(x)$, we can always choose $\tilde{y} = \bar{y}$ in the proof of Proposition 3.4.3. Hence the result follows from Corollary 3.4.1. ■

3.4.2 Directional Lipschitz continuity of the value function

In this subsection we study sufficient conditions for the directional Lipschitz continuity of $V(x)$.

The classical criterion for guaranteeing the Lipschitz continuity of the value function, is a combination of the uniform compactness condition and MFCQ holding at each $y \in S(\bar{x})$, see e.g. [20, Theorem 5.1]. The following theorem gives sufficient conditions for the directional Lipschitz continuity of the value function under weaker assumptions. When $u = 0$, it recovers the result in [5, Theorem 5.5].

Theorem 3.4.1. *(i) Suppose that $S(\bar{x}; u) \neq \emptyset$, the restricted inf-compactness holds at $\bar{x}$ in direction u with compact set Ω_u and the feasible map $\mathcal{F}(x) := \{y | g(x, y) \leq$*

0\} satisfies RS at $(\bar{x}, y)$ for each $y \in S(\bar{x}; u) \cap \Omega_u$. Then $V(x)$ is Lipschitz continuous at $\bar{x}$ in direction u.

(ii) Suppose there exists $\bar{y} \in S(\bar{x})$ such that $S(x)$ is inner semi-continuous at $(\bar{x}, \bar{y})$ in direction u, and the feasible map $\mathcal{F}(x) := \{y | g(x, y) \leq 0\}$ satisfies RS at $(\bar{x}, \bar{y})$ in direction u. Then $V(x)$ is Lipschitz continuous at $\bar{x}$ in direction u.

Furthermore, if (i) or (ii) holds in direction $u = 0$ then $V(x)$ is Lipschitz around $\bar{x}$.

Proof. Since RS is satisfied at each $(\bar{x}, y)$ for $y \in S(\bar{x}; u) \cap \Omega_u$ in direction u, by the compactness of Ω_u and Borel-Lebesgue covering theorem, there exist positive scalars ϵ, δ, κ such that

$$\operatorname{dist}(y, \mathcal{F}(x)) \leq \kappa \operatorname{dist}(g(x, y), \mathbb{R}^p_-) \quad \forall x \in \bar{x} + \mathcal{V}_{\epsilon,\delta}(u), y \in (S(\bar{x}; u) \cap \Omega_u) + \epsilon\mathbb{B}. \quad (3.25)$$

By Lemma 3.4.3, choosing ϵ, δ small enough, we have for any $x, x' \in \bar{x} + \mathcal{V}_{\epsilon,\delta}(u)$, there exist $y \in S(x) \cap \Omega_u$, $y' \in S(x') \cap \Omega_u$ close enough to $S(\bar{x}; u) \cap \Omega_u$. Without loss of generality assume $x, x' \in \bar{x} + \mathcal{V}_{\epsilon,\delta}(u)$ and $y, y' \in (S(\bar{x}; u) \cap \Omega_u) + \epsilon\mathbb{B}$. Then by (3.25) we can find $\bar{y} \in \mathcal{F}(x)$, $\bar{y}' \in \mathcal{F}(x')$ such that

$$\begin{aligned}
\|y - \bar{y}'\| &\leq \kappa \|g(x', y) - g(x, y)\| \leq 2\kappa \|\nabla_x g(x, y)\| \|x - x'\|, \\
\|y' - \bar{y}\| &\leq \kappa \|g(x, y') - g(x', y')\| \leq 2\kappa \|\nabla_x g(x', y')\| \|x - x'\|.
\end{aligned}$$

Since $\nabla_x g(x, y)$ is continuous and $\{\bar{x} + \mathcal{V}_{\epsilon,\delta}(u)\} \times (S(\bar{x}; u) \cap \Omega_u)$ is bounded, by Weirstrass extreme value theorem, there exists a positive scalar M such that $2\kappa \|\nabla_x g(x, y)\| \leq M$ for any $(x, y) \in \{\bar{x} + \mathcal{V}_{\epsilon,\delta}(u)\} \times (S(\bar{x}; u) \cap \Omega_u)$. Similarly, since $\nabla f(x, y)$ is continuous, hence, locally bounded. Choosing M' large enough, we have

$$\begin{aligned}
\|f(x, y) - f(x', \bar{y}')\| &\leq M' \|(x, y) - (x', \bar{y}')\| \leq M'(1 + M)\|x - x'\|, \\
\|f(x, \bar{y}) - f(x', y')\| &\leq M' \|(x, \bar{y}) - (x', y')\| \leq M'(1 + M)\|x - x'\|.
\end{aligned}$$

Then since $f(x, y) - f(x', \bar{y}') \leq V(x) - V(x') \leq f(x, \bar{y}) - f(x', y')$, we have

$$\|V(x) - V(x')\| \leq \max\{\|f(x, y) - f(x', \bar{y}')\|, \|f(x, \bar{y}) - f(x', y')\|\} \leq M'(1 + M)\|x - x'\|.$$

This means $V(x)$ is Lipschitz continuous at $\bar{x}$ in direction u and (i) is proved.

Next, we prove (ii). If there exists $\bar{y} \in S(\bar{x})$ such that $S(x)$ is inner semi-continuous at $(\bar{x}, \bar{y})$ in direction u, $\bar{y} \in S(\bar{x}; u) \neq \emptyset$ and the restricted inf-compactness holds at $\bar{x}$ in direction u. Then one can easily replace $S(\bar{x}; u) \cap \Omega_u$ by $\{\bar{y}\}$ in the proof above and obtain the Lipschitz continuity of $V(x)$ under RS at $(\bar{x}, \bar{y})$ in direction u. ■

3.4.3 Directional subdiffentials of the value function

In this subsection, we study the analytic directional subdifferential of the value function of (P_x). First, we derive an upper estimate for the analytic directional subdifferential of the value function in terms of the problem data. For any x, y, u, suppose $V'(x; u)$ exists. We denote by

$$\Sigma(x, y, u) := \{v \in \mathbb{L}(x, y; u) | V'(x; u) = \nabla f(x, y)(u, v)\}. \tag{3.26}$$

Theorem 3.4.2. *Let $u \in \mathbb{R}^n$.*

(i) Suppose that the restricted inf-compactness holds at $\bar{x}$ in direction u with compact set Ω_u. Suppose that $V(x)$ is directionally differentiable at $\bar{x}$ in direction u. Then $S(\bar{x}; u) \neq \emptyset$. Moreover suppose that the feasible map $\mathcal{F}(x) := \{y | g(x, y) \leq 0\}$ satisfies RS at $(\bar{x}, y)$ in direction u for each $y \in S(\bar{x}; u) \cap \Omega_u$. Then $V(x)$ is Lipschitz at $\bar{x}$ in direction u and

$$\emptyset \neq \partial_a V(\bar{x}; u) \subseteq \Theta(\bar{x}; u) \tag{3.27}$$

where

$$\begin{aligned}
&\Theta(\bar{x}; u) := \\
&\bigcup_{\tilde{y} \in S(\bar{x}; u) \cap \Omega_u} \Big(\bigcup_{v \in \Sigma(\bar{x}, \tilde{y}, u)} \left\{ \nabla_x f(\bar{x}, \tilde{y}) + \nabla_x g(\bar{x}, \tilde{y})^T \lambda_g | \ \ \lambda_g \in \Lambda(\bar{x}, \tilde{y}) \cap \{\nabla g(\bar{x}, \tilde{y})(u, v)\}^{\perp} \right\} \\
&\cup \bigcup_{v \in \Sigma(\bar{x}, \tilde{y}; 0) \cap \mathbb{S}} \left\{ \nabla_x f(\bar{x}, \tilde{y}) + \nabla_x g(\bar{x}, \tilde{y})^T \lambda_g | \ \lambda_g \in \Lambda(\bar{x}, \tilde{y}) \cap \{\nabla g(\bar{x}, \tilde{y})(0, v)\}^{\perp} \right\} \Big).
\end{aligned} \tag{3.28}$$

(ii) Suppose that there exists $\bar{y} \in S(\bar{x})$ such that $S(x)$ is inner semi-continuous at $(\bar{x}, \bar{y})$ in direction u and the feasible map $\mathcal{F}(x) := \{y | g(x, y) \leq 0\}$ satisfies RS at $(\bar{x}, \bar{y})$ in direction u. Suppose that $V(x)$ is directionally differentiable at $\bar{x}$ in direction u. Then $V(x)$ is Lipschitz at $\bar{x}$ in direction u. And (3.27) holds with the union over $\tilde{y} \in S(\bar{x}; u) \cap \Omega_u$ superfluous and $\tilde{y} = \bar{y}$.

(iii) Suppose that there exists $\bar{y} \in S(\bar{x})$ such that $S(x)$ is inner semi-continuous at $(\bar{x}, \bar{y})$ in direction u and $\mathcal{F}$ is RCR-regular at $(\bar{x}, \bar{y})$ in direction u, then $V(x)$ is

Lipschitz at $\bar{x}$ in direction u and

$$\emptyset \neq \partial_a V(\bar{x};u) \subseteq \bigcup_{v\in\Sigma(\bar{x},\bar{y},u)} \left\{ \nabla_x f(\bar{x},\bar{y}) + \nabla_x g(\bar{x},\bar{y})^T\lambda_g \middle| \ \lambda_g \in \Lambda(\bar{x},\bar{y}) \cap \{\nabla g(\bar{x},\bar{y})(u,v)\}^{\perp} \right\}.$$

Proof. (i)Since $V(x)$ is differetiable in direction u, for any sequence $\epsilon_k \downarrow 0$, we can find a sequence $t_k \downarrow 0$ such that for k large enough, we have $V(\bar{x}+t_k u) < V(\bar{x})+\epsilon_k$. Then by the assumption of the directional restricted inf-compactness, for k large enough, there exists $\hat{y}^k \in S(\bar{x}+t_k u)\cap\Omega_u$. Then $\{\hat{y}^k\}$ is bounded. Without loss of generality, there exists $\hat{y} = \lim_k \hat{y}^k$. And we know that $f(\bar{x},\hat{y}) = \lim_k V(\bar{x}+t_k u) \le V(\bar{x})$. Hence, $\hat{y} \in S(\bar{x};u) \neq \emptyset$.

Since the directional restricted inf-compactness holds at $\bar{x}$ in direction u and RS is satisfied at $(\bar{x},y)$ in direction u for each $y \in S(\bar{x},u)\cap\Omega_u$, by Theorem 3.4.1, $V(x)$ is Lipschitz continuous at $\bar{x}$ in direction u. Then by the well-known Rademacher's Theorem and [72, Theorem 9.13], $\partial_a V(\bar{x};u) \neq \emptyset$.

Let $\zeta \in \partial_a V(\bar{x};u)$. Then by definition, there exist sequences $t_k \downarrow 0$, $u^k \to u$, $\zeta^k \to \zeta$ such that $V(\bar{x}+t_k u^k) \to V(\bar{x})$ and $\zeta^k \in \widehat{\partial} V(\bar{x}+t_k u^k)$. It follows that $V(\bar{x}+t_k u^k) < V(\bar{x})+\epsilon$ for all k large enough and hence by the directional restricted inf-compactness, there exists $y^k \in S(\bar{x}+t_k u^k)\cap\Omega_u$. Passing to a subsequence if necessary, we may assume that $y^k \to \tilde{y}$. Hence, by the continuity of $f(x,y)$, $\tilde{y} \in S(\bar{x};u)\cap\Omega_u$.

For each k, since $\zeta^k \in \widehat{\partial} V(\bar{x}+t_k u^k)$, there exists a neighborhood $\mathcal{U}^k$ of $\bar{x}+t_k u^k$ satisfying

$$V(x) - V(\bar{x}+t_k u^k) - \langle \zeta^k, x-(\bar{x}+t_k u^k)\rangle + \frac{1}{k}\|x-(\bar{x}+t_k u^k)\| \ge 0 \ \forall x \in \mathcal{U}^k.$$

It follows from the fact $V(x) = \inf_y\{f(x,y)+\delta_{\mathbb{R}^p_-}(g(x,y))\}$ and $y^k \in S(\bar{x}+t_k u^k)$, that

$$f(x,y) - \langle \zeta^k, x-(\bar{x}+t_k u^k)\rangle + \frac{1}{k}\|x-(\bar{x}+t_k u^k)\| + \delta_{\mathbb{R}^p_-}(g(x,y)) \ge f(\bar{x}+t_k u^k, y^k),$$

for any $(x,y) \in \mathcal{U}^k \times \mathbb{R}^m$. Hence the function

$$\phi_k(x,y) := f(x,y) - \langle \zeta^k, x-(\bar{x}+t_k u^k)\rangle + \frac{1}{k}\|x-(\bar{x}+t_k u^k)\| + \delta_{\mathbb{R}^p_-}(g(x,y))$$

attains its local minimum at $(x,y) = (\bar{x}+t_k u^k, y^k)$. Thus, by the well known Fermat's rule and the sum rule ([72, Exercise 10.10]),

$$0 \in \nabla f(\bar{x}+t_k u^k, y^k) - (\zeta^k, 0) + \frac{1}{k}\bar{\mathbb{B}} \times \{0\} + \partial(\delta_{\mathbb{R}^p_-} \circ g)(\bar{x}+t_k u^k, y^k). \tag{3.29}$$

Now we consider two cases.
Case (a): $\{\frac{y^k-\tilde{y}}{t_k}\}$ is bounded. Define $v^k := \frac{y^k-\tilde{y}}{t_k}$. Passing to a subsequence if necessary, there exists $v \in \mathbb{R}^m$ such that $v^k \to v$. Since $y^k \in S(\bar{x}+t_k u^k)$, it follows that

$$\nabla f(\bar{x},\tilde{y})(u,v) = \lim_k \frac{f(\bar{x}+t_k u^k, y^k) - f(\bar{x},\tilde{y})}{t_k} = \lim_k \frac{V(\bar{x}+t_k u^k) - V(\bar{x})}{t_k} = V'(\bar{x};u)$$

and $\nabla g_i(\bar{x},\tilde{y})(u,v) \le 0,\ \forall i \in I_g(\bar{x},\tilde{y})$. This means $v \in \Sigma(\bar{x},\tilde{y};u)$. Since for each k, $\delta_{\mathbb{R}^p_-}(g(\bar{x}+t_k u^k, y^k)) = 0 = \delta_{\mathbb{R}^p_-}(g(\bar{x},\tilde{y}))$, taking limits as $k \to \infty$ in (3.29), by Proposition 3.2.2 we have

$$0 \in \nabla f(\bar{x},\tilde{y}) - (\zeta,0) + \partial_a(\delta_{\mathbb{R}^p_-} \circ g)(\bar{x},\tilde{y};u,v).$$

Since RS is satisfied at $(\bar{x},\tilde{y})$ in direction u, by Proposition 3.4.2, the metric subregularity for the system $g(x,y) \in \mathbb{R}^p_-$ holds at $(\bar{x},\tilde{y},0)$ in direction (u,v), then by Proposition 3.2.6, we have

$$\partial_a(\delta_{\mathbb{R}^p_-} \circ g)(\bar{x},\tilde{y};u,v) \subseteq \nabla g(\bar{x},\tilde{y})^T N_{\mathbb{R}^p_-}(g(\bar{x},\tilde{y}); \nabla g(\bar{x},\tilde{y})(u,v)).$$

With (3.29), we obtain there exists $\lambda_g \in N_{\mathbb{R}^p_-}(g(\bar{x},\tilde{y}); \nabla g(\bar{x},\tilde{y})(u,v))$ with $0 \in \nabla_y f(\bar{x},\tilde{y}) + \nabla_y g(\bar{x},\tilde{y})^T \lambda_g$ such that $\zeta = \nabla_x f(\bar{x},\tilde{y}) + \nabla_x g(\bar{x},\tilde{y})^T \lambda_g$. The proof follows from

$$N_{\mathbb{R}^p_-}(g(\bar{x},\tilde{y}); \nabla g(\bar{x},\tilde{y})(u,v)) = \{\lambda \in \mathbb{R}^p | 0 \le \lambda \perp g(\bar{x},\tilde{y}),\ \lambda \perp \nabla g(\bar{x},\tilde{y})(u,v)\}.$$

Case (b): $\{\frac{y^k-\tilde{y}}{t_k}\}$ is unbounded. Without loss of generality, assume $\lim_{k\to\infty} \frac{\|y^k-\tilde{y}\|}{t_k} = \infty$. Define $\tau_k := \|y^k - \tilde{y}\|$. Then $\frac{t_k}{\tau_k} \downarrow 0$. Since the sequence $\{\frac{y^k-\tilde{y}}{\tau_k}\}$ is bounded, passing to a subsequence if necessary, assume there exist $v \in \mathbb{S}$ and a sequence $v^k \to v$ such that $y^k = \tilde{y} + \tau_k v^k$. Define $\tilde{u}^k := \frac{t_k}{\tau_k} u^k$. Then $\bar{x} + t_k u^k = \bar{x} + \tau_k \tilde{u}^k$ and $\tilde{u}^k \to 0$. Since $y^k \in S(\bar{x}+t_k u^k)$, it follows that

$$0 = \lim_k \frac{V(\bar{x}+\tau_k \tilde{u}^k) - V(\bar{x})}{t_k}\frac{t_k}{\tau_k} = \lim_k \frac{f(\bar{x}+\tau_k \tilde{u}^k, \tilde{y}+\tau_k v^k) - f(\bar{x},\tilde{y})}{\tau_k} = \nabla f(\bar{x},\tilde{y})(0,v)$$

and $\nabla g_i(\bar{x},\tilde{y})(0,v) \leq 0,\ \forall i \in I_g(\bar{x},\tilde{y})$. Since $V'(\bar{x};0) = 0$, we obtain $V'(\bar{x};0) = \nabla f(\bar{x},\tilde{y})(0,v)$. Hence, $v \in \Sigma(\bar{x},\tilde{y};0) \cap \mathbb{S}$. Taking limits as $k \to \infty$ in (3.29), following a similar process as in Case (a) we have

$$\begin{aligned} 0 &\in \nabla f(\bar{x},\tilde{y}) - (\zeta,0) + \partial(\delta_{\mathbb{R}^p_-} \circ g)(\bar{x},\tilde{y};0,v) \\ &\subseteq \nabla f(\bar{x},\tilde{y}) - (\zeta,0) + \nabla g(\bar{x},\tilde{y})^T N_{\mathbb{R}^p_-}(g(\bar{x},\tilde{y});\nabla g(\bar{x},\tilde{y})(0,v)). \end{aligned}$$

So there exists $\lambda_g \in N_{\mathbb{R}^p_-}(g(\bar{x},\tilde{y});\nabla g(\bar{x},\tilde{y})(0,v))$ with $0 = \nabla_y f(\bar{x},\tilde{y}) + \nabla_y g(\bar{x},\tilde{y})^T\lambda_g$ such that $\zeta = \nabla_x f(\bar{x},\tilde{y}) + \nabla_x g(\bar{x},\tilde{y})^T\lambda_g$. This completes the proof.

(ii) When $S(x)$ is inner semi-continuous at some point $\bar{y} \in S(\bar{x})$ in direction u, one can choose $\tilde{y} = \bar{y}$. And the results follows similarly as the proof of (i).

(iii) Let $\zeta \in \partial_a V(\bar{x};u)$. As in the proof of (i) and taking into account the inner semicontinuity of $S(x)$ at $(\bar{x},\bar{y})$ in direction u, we obtain $t_k \downarrow 0,\ u^k \to u,\ \zeta^k \to \zeta$, $y^k \in S(\bar{x}+t_k u^k)$, $y^k \to \bar{y}$ satisfying (3.29). By Lemma 3.4.4 and Theorem 3.4.1, RS holds at $\bar{x}$ in direction u and $V(x)$ is Lipschitz continuous at $\bar{x}$ in direction u. Then by Proposition 3.4.2 we have metric subregularity for the system $g(x,y) \in \mathbb{R}^p_-$ holds at each k sufficiently large. Hence, by Proposition 3.2.6, for sufficiently large k, we have

$$\partial(\delta_{\mathbb{R}^p_-} \circ g)(\bar{x}+t_k u^k, y^k) \subseteq \nabla g(\bar{x}+t_k u^k, y^k)^T N_{\mathbb{R}^p_-}(g(\bar{x}+t_k u^k, y^k)).$$

Hence

$$0 \in \nabla f(\bar{x}+t_k u^k, y^k) - (\zeta^k,0) + \frac{1}{k}\mathbb{B}\times\{0\} + \nabla g(\bar{x}+t_k u^k, y^k)^T N_{\mathbb{R}^p_-}(g(\bar{x}+t_k u^k, y^k)). \tag{3.30}$$

Since RCR-regularity holds at $(\bar{x},\bar{y})$ and $y^k \in S(\bar{x}+t_k u^k)$, by Lemma 3.4.1, for sufficiently large k, there exist $\kappa > 0$ independent of k and a sequence $\bar{y}^k \in \mathcal{F}(\bar{x})$ such that

$$\|y^k - \bar{y}^k\| \leq \kappa\|\bar{x}+t_k u^k - \bar{x}\|,\ g_j(\bar{x}+t_k u^k, y^k) \leq g_j(\bar{x},\bar{y}^k),\ j \in I_g(\bar{x},\bar{y}). \tag{3.31}$$

Then $I_g(\bar{x}+t_k u^k, y^k) \subseteq I_g(\bar{x},\bar{y}^k)$ and by Proposition 3.2.1,

$$\begin{aligned} N_{\mathbb{R}^p_-}(g(\bar{x}+t_k u^k, y^k)) &= N_{\mathbb{R}^p_-}(g(\bar{x},\bar{y}^k)) \cap [g(\bar{x},\bar{y}^k) - g(\bar{x}+t_k u^k, y^k)]^\perp \\ &= N_{\mathbb{R}^p_-}(g(\bar{x},\bar{y}^k)) \cap [\frac{g(\bar{x},\bar{y}^k) - g(\bar{x}+t_k u^k, y^k)}{t_k}]^\perp. \end{aligned}$$

Define $v^k := \frac{y^k - \bar{y}^k}{t_k}$. $y^k = \bar{y}^k + t_k v^k$. By (3.31), $\{v^k\}$ is bounded. Without loss of generality, there exists $v = \lim_k v^k$. Then $\lim_k \frac{g(\bar{x}+t_k u^k, y^k) - g(\bar{x}, \bar{y}^k)}{t_k} = \nabla g(\bar{x}, \bar{y})(u, v)$. By (3.31), $\bar{y}^k \to \bar{y}$ and $v \in \mathbb{L}(\bar{x}, \bar{y}; u)$. Taking the limit in (3.30), we have

$$0 \in \nabla f(\bar{x}, \bar{y}) - (\zeta, 0) + \nabla g(\bar{x}, \bar{y})^T (N_{\mathbb{R}^p_-}(g(\bar{x}, \bar{y})) \cap [\nabla g(\bar{x}, \bar{y})(u, v)]^{\perp}).$$

We obtain the existence of $\lambda_g \in N_{\mathbb{R}^p_-}(g(\bar{x}, \bar{y})) \cap [\nabla g(\bar{x}, \bar{y})(u, v)]^{\perp}$ such that $\zeta = \nabla_x f(\bar{x}, \bar{y}) + \nabla_x g(\bar{x}, \bar{y})^T \lambda_g$. Furthermore, by Proposition 3.4.4, $V'(\bar{x}; u) = \min\{\nabla f(\bar{x}, \bar{y})(u, \nu) | \nu \in \mathbb{L}(\bar{x}, \bar{y}; u)\}$, it follows from a similar process as (3.22), we have $V'(\bar{x}; u) = \nabla f(\bar{x}, \bar{y})(u, v)$. The proof is complete. ■

[54, Theoerems 5.10 and 5.11] also gave an upper estimate of the value function of constrained programs in terms of the coderivatives of the constraint mapping $\mathcal{F}$ under a stronger version of directional inner semicontinuity [54, Definition 4.4(i)] of $S(x)$. Our result cannot be obtained from [54, Theoerems 5.10 and 5.11] and is in a more explicit form.

The following theorem provides an estimate of the directional Clarke subdifferential of the value function which will be used in the necessary optimality condition for bilevel programs. We give some notations. For any given (x, y, u, v) we define the set

$$W(x, y, u, v) \;:=\; \left\{ \nabla_x f(x, y) + \nabla_x g(x, y)^T \lambda_g \,\middle|\, \lambda_g \in \Lambda(x, y) \cap \{\nabla g(x, y)(u, v)\}^{\perp} \right\}.$$

Theorem 3.4.3. *Under the assumptions of Theorem 3.4.2(i), we have*

$$\partial^c V(\bar{x}; u) \subseteq co \bigcup_{\tilde{y} \in S(\bar{x}; u) \cap \Omega_u} (W(\bar{x}, \tilde{y}, u, v) \cup W(\bar{x}, \tilde{y}, 0, \nu)) \quad \forall v \in \Sigma(\bar{x}, \tilde{y}, u), \nu \in \Sigma(\bar{x}, \tilde{y}, 0) \cap \mathbb{S}.$$

Under the assumptions of Theorem 3.4.2(ii), we have

$$\partial^c V(\bar{x}; u) \subseteq \{\mu\zeta + (1-\mu)\xi | 0 \leq \mu \leq 1,\ \zeta \in W(\bar{x}, \bar{y}, u, v),\ \xi \in W(\bar{x}, \bar{y}, 0, \nu)\}$$

for $\forall v \in \Sigma(\bar{x}, \tilde{y}, u), \nu \in \Sigma(\bar{x}, \tilde{y}, 0) \cap \mathbb{S}$.
Under the assumptions of Theorem 3.4.2(iii), we have

$$\partial^c V(\bar{x}; u) \subseteq W(\bar{x}, \bar{y}, u, v), \quad \forall v \in \Sigma(\bar{x}, \bar{y}, u).$$

Proof. Since $\partial^c V(\bar{x}; u) = co\partial_a V(\bar{x}; u)$, by Theorem 3.4.2, we only need to show that

$W(x,y,u,v_1) = W(x,y,u,v_2)$ for any $v_1, v_2 \in \Sigma(x,y,u)$. Let

$$C(x,y,u,v) := \{\lambda_g | \nabla_y f(x,y) + \lambda_g \nabla_y g(x,y) = 0, 0 \le \lambda_g \perp g(x,y), \lambda_g \perp \nabla g(x,y)(u,v)\}.$$

It suffices to show that $C(x,y,u,v_1) = C(x,y,u,v_2)$ for any $v_1, v_2 \in \Sigma(x,y,u)$. By $\nabla_y f(\bar{x},\bar{y}) + \nabla_y g(\bar{x},\bar{y})^T \lambda_g = 0$, we have $\lambda_g^T \nabla_y g(x,y) v_i = -\nabla_y f(x,y) v_i$ for $i = 1,2$. And since $\nabla_y f(x,y) v_1 = \nabla_y f(x,y) v_2 = V'(x)(u) - \nabla_x f(x,y) u$, $\lambda_g^T \nabla_y g(x,y) v_1 = \lambda_g^T \nabla_y g(x,y) v_2$. Hence, $\lambda_g^T \nabla g(x,y)(u,v_1) = \lambda_g^T \nabla g(x,y)(u,v_2)$. This implies $C(x,y,u,v_1) = C(x,y,u,v_2)$. ■

3.5 Necessary optimality conditions for bilevel programs

The main purpose of this section is to apply Theorem 3.3.1 to problem (VP) and the result of the directional sensitivity analysis of the value functions in Section 4 to derive a sharp necessary optimality condition for (VP) under a weak and verifiable constraint qualification.

We first try to answer the question on whether it is possible for FOSCMS to hold at a feasible point of (VP). Given $(u,v) \in \mathbb{R}^{n+m}$, define the set-valued map $M_{(u,v)} : \mathbb{R}^n \times \mathbb{R}^m \rightrightarrows \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$ by

$$M_{(u,v)}(x,y) := (f(x,y) - V(x) + \langle u, x-\bar{x}\rangle^3 + \langle v, y-\bar{y}\rangle^3 - \mathbb{R}_-, g(x,y) - \mathbb{R}^p_-, G(x,y) - \mathbb{R}^q_-).$$

Assume that the value function is directionally differentiable at $\bar{x}$ in direction u. Define the linearization cone of (VP) at $(\bar{x},\bar{y})$ by

$$\mathbb{L}(\bar{x},\bar{y}) := \left\{ (u,v) | \begin{array}{l} \nabla f(\bar{x},\bar{y})(u,v) \le V'(\bar{x};u), \\ \nabla g_i(\bar{x},\bar{y})(u,v) \le 0 \ \forall i \in I_g(\bar{x},\bar{y}), \nabla G_i(\bar{x},\bar{y})(u,v) \le 0 \ \forall i \in I_G(\bar{x},\bar{y}) \end{array} \right\}.$$

Note that although $f(x,y) - V(x) \le 0$ is an inequality, it is in fact an equality constraint by the definition of the value function. Hence under the Abadie constraint qualification, one always have $\nabla f(\bar{x},\bar{y})(u,v) \ge V'(\bar{x};u)$ for all (u,v) satisfying $\nabla g_i(\bar{x},\bar{y})(u,v) \le 0$ $i \in I_g(\bar{x},\bar{y})$. Therefore if the Abadie constraint qualification holds, in the linearization cone the inequality $\nabla f(\bar{x},\bar{y})(u,v) \le V'(\bar{x};u)$ can be equivalently replaced by the equality.

Lemma 3.5.1. *Let* $(\bar{x},\bar{y})$ *be a feasible point of* (VP)*. Assume that the value function is directionally differentiable at* $\bar{x}$ *in any direction* u *and* $0 \neq (u,v) \in \mathbb{L}(\bar{x},\bar{y})$*. Then* $M_{(u,v)}(x,y)$ *is not metrically subregular at* $((\bar{x},\bar{y}),(0,0)) \in gphM_{(u,v)}$ *in direction* (u,v)*.*

Proof. To concentrate on the main idea we omit the upper level constraint $G(x,y) \leq 0$ in the proof. To the contrary, suppose that $M_{(u,v)}(x,y)$ is metrically subregular at $((\bar{x},\bar{y}),(0,0))$ in the nonzero direction $(u,v) \in \mathbb{L}(\bar{x},\bar{y})$. Then by definition of metric subregularity in direction (u,v), $\exists \kappa > 0$, for all sequences $t_k \downarrow 0$, $u^k \to u$, $v^k \to v$, we have for sufficiently large k

$$\begin{aligned}
&\operatorname{dist}\Big((\bar{x}+t_k u^k, \bar{y}+t_k v^k),\ M^{-1}_{(u,v)}(0,0)\Big)\\
&\quad\leq \kappa \operatorname{dist}(0, M_{(u,v)}(\bar{x}+t_k u^k, \bar{y}+t_k v^k))\\
&\quad\leq \kappa \left(\operatorname{dist}(f(\bar{x}+t_k u^k, \bar{y}+t_k v^k) - V(\bar{x}+t_k u^k) + t_k^3\langle u,u^k\rangle^3 + t_k^3\langle v,v^k\rangle^3, \mathbb{R}_-)\right.\\
&\quad\left.+\operatorname{dist}(g(\bar{x}+t_k u^k, \bar{y}+t_k v^k), \mathbb{R}^p_-)\right).
\end{aligned} \tag{3.32}$$

Since $(u,v) \in L(\bar{x},\bar{y})$, we have $g(\bar{x},\bar{y}) + t_k \nabla g(\bar{x},\bar{y})(u,v) \leq 0$. Hence,

$$\begin{aligned}
&\lim_{k\to\infty} \frac{\operatorname{dist}(g(\bar{x}+t_k u^k, \bar{y}+t_k v^k), \mathbb{R}^p_-)}{t_k}\\
&\leq \lim_{k\to\infty} \frac{\|g(\bar{x}+t_k u^k, \bar{y}+t_k v^k) - t_k \nabla g(\bar{x},\bar{y})(u,v) - g(\bar{x},\bar{y})\|}{t_k} = 0.
\end{aligned}$$

Similarly, since $f(\bar{x},\bar{y}) - V(\bar{x}) = 0$, we have $\nabla f(\bar{x},\bar{y})(u,v) - V'(\bar{x};u) \leq 0$,

$$\lim_{k\to\infty} \frac{\operatorname{dist}(f(\bar{x}+t_k u^k, \bar{y}+t_k v^k) - V(\bar{x}+t_k u^k) + t_k^3\langle u,u^k\rangle^3 + t_k^3\langle v,v^k\rangle^3, \mathbb{R}_-)}{t_k} = 0. \tag{3.33}$$

Since for every $t_k > 0$ sufficiently small, we can find a point $(x_{t_k}, y_{t_k}) \in M^{-1}_{(u,v)}(0,0)$ satisfying (3.32), then

$$f(x_{t_k}, y_{t_k}) - V(x_{t_k}) + \langle u, x_{t_k} - \bar{x}\rangle^3 + \langle v, y_{t_k} - \bar{y}\rangle^3 \leq 0. \tag{3.34}$$

And by (3.33), $\lim_{t_k\downarrow 0} t_k^{-1}\|(\bar{x}+t_k u^k, \bar{y}+t_k v^k) - (x_{t_k}, y_{t_k})\| = 0$.

Since $(u,v) \neq (0,0)$, by (3.34) we have for every k sufficiently large

$$\begin{aligned}
0 \quad & \geq f(x_{t_k}, y_{t_k}) - V(x_{t_k}) + \langle u, x_{t_k} - \bar{x}\rangle^3 + \langle v, y_{t_k} - \bar{y}\rangle^3 \\
& = f(x_{t_k}, y_{t_k}) - V(x_{t_k}) + \langle u, \bar{x} + t_k u - \bar{x}\rangle^3 + \langle v, \bar{y} + t_k v - \bar{y}\rangle^3 + o(t_k^3) \\
& \geq f(x_{t_k}, y_{t_k}) - V(x_{t_k}) + \frac{t_k^3}{2}(\|u\|^6 + \|v\|^6) \\
& > f(x_{t_k}, y_{t_k}) - V(x_{t_k}),
\end{aligned}$$

contradicting that $V(x_{t_k}) = \inf_{y \in \mathcal{F}(x_{t_k})} f(x_{t_k}, y) \leq f(x_{t_k}, y_{t_k})$. ■

We are now ready to give a negative answer on the question if the FOSCMS can be satisfied by a feasible solution of (VP). Let $(\bar{x}, \bar{y})$ be a feasible solution of (VP). Denote the critical cone of (VP) at $(\bar{x}, \bar{y})$ by

$$C(\bar{x}, \bar{y}) := \{(u,v) \in \mathbb{L}(\bar{x}, \bar{y}) | F(\bar{x}, \bar{y})(u,v) \leq 0\}.$$

Proposition 3.5.1. *Assume that the value function is directionally differentiable at $\bar{x}$ in any direction u and $(u,v) \in C(\bar{x}, \bar{y})$. Then there exists a nonzero vector $(\lambda, \mu, \nu) \in \mathbb{R}^{1+p+q}$ such that $\lambda \geq 0, 0 \leq \mu \perp g(\bar{x}, \bar{y}),\ \mu \perp \nabla g(\bar{x}, \bar{y})(u,v), 0 \leq \nu \perp G(\bar{x}, \bar{y}),\ \nu \perp \nabla G(\bar{x}, \bar{y})(u,v)$ and*

$$0 \in \lambda \partial_a (f - V)(\bar{x}, \bar{y}; (u,v)) + \nabla g(\bar{x}, \bar{y})^T \mu + \nabla G(\bar{x}, \bar{y})^T \nu. \tag{3.35}$$

Hence FOSCMS fails at any feasible solution of (VP) in any critical direction.

Proof. Since by Lemma 3.5.1, $M_{(u,v)}(x,y)$, hence $-M_{(u,v)}(x,y)$, is not metrically subregular at $(\bar{x}, \bar{y}, 0, 0)$ in direction (u,v) and metric subregularity is weaker than FOSCMS, FOSCMS for the inequality system

$$\psi(x,y) := (f(x,y) - V(x) + \langle u, x - \bar{x}\rangle^3 + \langle v, y - \bar{y}\rangle^3), g(x,y), G(x,y)) \leq 0$$

must fail at $(\bar{x}, \bar{y})$ in direction (u,v). By the sum rule [54, Theorem 5.6] of analytic directional subdifferential,

$$\partial_a (f(x,y) - V(x) + \langle u, x - \bar{x}\rangle^3 + \langle v, y - \bar{y}\rangle^3)(\bar{x}, \bar{y}; (u,v)) = \partial_a (f - V)(\bar{x}, \bar{y}; (u,v))).$$

Hence by Definition 3.3.2(2) the FOSCMS for the inequality system $\psi(x,y) \leq 0$ at $((\bar{x}, \bar{y}), (0,0))$ is the same as the (3.35) which means that FOSCMS for (VP) at $(\bar{x}, \bar{y})$

in direction (u, v) fails. ■

We now apply Lemma 3.3.1, Proposition 3.3.1 and Theorem 3.3.1 to (VP) and obtain the following necessary optimality condition for the bilevel program (BP).

Theorem 3.5.1. *Let $(\bar{x}, \bar{y})$ be a local minimizer of* (BP). *Suppose that the value function $V(x)$ is Lipschitz continuous and directionally differentiable near $\bar{x}$ in direction u and $(u, v) \in C(\bar{x}, \bar{y})$. Moreover suppose that the directional quasi-normality holds at $(\bar{x}, \bar{y})$ in direction (u, v), i.e., there exists no nonzero vector $(\alpha, \nu_g, \nu_G) \in \mathbb{R}_+^{1+p+q}$ and*

$$0 \in \alpha \nabla f(\bar{x}, \bar{y}) - \alpha \partial^c V(\bar{x}; u) \times \{0\} + \nabla g(\bar{x}, \bar{y})^T \nu_g + \nabla G(\bar{x}, \bar{y})^T \nu_G, \tag{3.36}$$

$$\nu_g \perp g(\bar{x}, \bar{y}),\ \nu_g \perp \nabla g(\bar{x}, \bar{y})(u, v),\ \nu_G \perp G(\bar{x}, \bar{y}),\ \nu_G \perp \nabla G(\bar{x}, \bar{y})(u, v), \tag{3.37}$$

and there exists sequences $t_k \downarrow 0,\ (u^k, v^k) \to (u, v)$ such that

$$\alpha(f(\bar{x} + t_k u^k, \bar{y} + t_k v^k) - V(\bar{x} + t_k u^k)) > 0, \text{ if } \alpha > 0, \tag{3.38}$$

$$g_i(\bar{x} + t_k u^k, \bar{y} + t_k v^k) > 0, \text{ if } (\nu_g)_i > 0, i \in I_g,$$

$$G_i(\bar{x} + t_k u^k, \bar{y} + t_k v^k) > 0, \text{ if } (\nu_G)_i > 0, i \in I_G. \tag{3.39}$$

Then the directional KKT condition holds. That is, there exists $(\lambda_V, \lambda_g, \lambda_G)$ such that

$$0 \in \nabla F(\bar{x}, \bar{y}) + \lambda_V \nabla f(\bar{x}, \bar{y}) - \lambda_V \partial^c V(\bar{x}; u) \times \{0\} + \nabla g(\bar{x}, \bar{y})^T \lambda_g + \nabla G(\bar{x}, \bar{y})^T \lambda_G,$$
$$\lambda_V \geq 0,\ 0 \leq \lambda_g \perp g(\bar{x}, \bar{y}),\ \lambda_g \perp \nabla g(\bar{x}, \bar{y})(u, v),\ 0 \leq \lambda_G \perp G(\bar{x}, \bar{y}),\ \lambda_G \perp \nabla G(\bar{x}, \bar{y})(u, v).$$

Proof. Define $\phi(x, y) := (f(x, y) - V(x), g(x, y), G(x, y))$ and $\lambda_\phi := (\alpha, \nu_g, \nu_G)$. Then by assumption, $\phi(x, y)$ is Lipschitz continuous and directionally differentiable at $(\bar{x}, \bar{y})$ in direction (u, v). Since $(u, v) \in C(\bar{x}, \bar{y})$, we have $\nabla\phi(\bar{x}, \bar{y})(u, v) \leq 0$. Then since $f(\bar{x}, \bar{y}) - V(\bar{x}) = 0, g(\bar{x}, \bar{y}) \leq 0, G(\bar{x}, \bar{y}) \leq 0$, (3.37) means $0 \leq \lambda_\phi \perp \phi(\bar{x}, \bar{y})$ and $\lambda_\phi \perp \nabla\phi(\bar{x}, \bar{y})(u, v)$. Since

$$\begin{aligned} \partial_a (f - V)(\bar{x}, \bar{y}; (u, v)) &= \nabla f(\bar{x}, \bar{y}) + \partial_a(-V)(\bar{x}; u) \times \{0\} \\ &\subseteq \nabla f(\bar{x}, \bar{y}) + \partial^c(-V)(\bar{x}; u) \times \{0\} \\ &\subseteq \nabla f(\bar{x}, \bar{y}) - \partial^c V(\bar{x}; u) \times \{0\}, \end{aligned}$$

where the first equation follows from [54, Theorem 5.6] and the second inclusion follows from Proposition 3.2.3, (3.36)-(3.39) imply that the directional quasi-normality defined in Definition 3.3.2(3) holds. Applying Theorem 3.3.1, the proof is complete. ■

When the conditions in Corollary 3.4.1 and Theorems 3.4.2(i) hold, one can apply the formulas of $V'(\bar{x};u)$ and the upper estimates for $\partial^c V(\bar{x};u)$ obtained in section 4 and derive the directional KKT condition in terms of the problem data under the directional quasi-normality as below.

Theorem 3.5.2. *Let $(\bar{x},\bar{y})$ be a local minimizer of* (BP) *and $u \in \mathbb{R}^n$. Suppose that the feasible map $\mathcal{F}(x) := \{y|g(x,y) \leq 0\}$ is RCR-regular at each $(\bar{x},y) \in gphS$ and satisfies RS at each $(\bar{x},y) \in gphS$ in direction u. Moreover assume that the restricted inf-compactness holds at $\bar{x}$ in direction u. Then the value function $V(x)$ is Lipschitz continuous and directionally differentiable at $\bar{x}$ in direction u with*

$$\begin{aligned} V'(\bar{x};u) &= \min_{y\in S(\bar{x};u)} \max_{\lambda\in\Lambda(\bar{x},y)} \nabla_x \mathcal{L}(\bar{x},y;\lambda)u, \\ \partial_a V(\bar{x};u) &\subseteq \Theta(\bar{x};u), \end{aligned}$$

where $\Theta(\bar{x};u)$ is defined as in (3.28). Suppose that the directional quasi-normality holds at $(\bar{x},\bar{y})$ in direction $(u,v) \in C(\bar{x},\bar{y})$ in Theorem 3.5.1, with $\partial^c V(\bar{x};u)$ replaced by $co(\Theta(\bar{x};u))$. Then the directional KKT condition and Theorem 3.5.1 holds with $\partial^c V(\bar{x};u)$ replaced by $co(\Theta(\bar{x};u))$.

When the the solution map $S(x)$ is directionally inner semi-continuous at the point of interest, we can obtain the directional quasi-normality condition and the KKT condition of (VP) in the following more verifiable forms.

Theorem 3.5.3. *Let $(\bar{x},\bar{y})$ be a local minimizer of* (BP). *Suppose that the feasible map $\mathcal{F}(x)$ is RCR-regular at $(\bar{x},\bar{y})$ and $S(x)$ is inner semi-continuous at $(\bar{x},\bar{y})$ in direction u. Then the value function is Lipschitz continuous and directional differentiable at $\bar{x}$ in direction u and $V'(\bar{x};u) = \max_{\lambda\in\Lambda(\bar{x},\bar{y})} \nabla_x \mathcal{L}(\bar{x},\bar{y};\lambda)u$. Suppose that there exists v such that $(u,v) \in C(\bar{x},\bar{y})$. Furthermore suppose that there exists no nonzero vector (α,ν_g,ν_G) satisfying*

$$\begin{aligned} &0 \in \alpha\nabla f(\bar{x},\bar{y}) - \alpha W(\bar{x},\bar{y},u,v) \times \{0\} + \nabla g(\bar{x},\bar{y})^T \nu_g + \nabla G(\bar{x},\bar{y})^T \nu_G, \\ &\alpha \geq 0,\ 0 \leq \nu_g \perp g(\bar{x},\bar{y}), \nu_g \perp \nabla g(\bar{x},\bar{y})(u,v),\ 0 \leq \nu_G \perp G(\bar{x},\bar{y}), \nu_G \perp \nabla G(\bar{x},\bar{y})(u,v), \end{aligned}$$

where $W(\bar{x}, \bar{y}, u, v) = \{\nabla_x f(\bar{x}, \bar{y}) + \nabla_x g(\bar{x}, \bar{y})^T \lambda | \lambda \in \Lambda(\bar{x}, \bar{y}) \cap \{\nabla g(\bar{x}, \bar{y})(u, v)\}^\perp\}$ *and there exists sequences* $t_k \downarrow 0$, $(u^k, v^k) \to (u, v)$ *such that (3.38)-(3.39) hold. Then there exists a vector* $(\lambda^V, \lambda_g, \lambda_G, \lambda) \in \mathbb{R}^{1+p+q+p}$ *satisfying*

$$
\begin{aligned}
&0 = \nabla_x F(\bar{x}, \bar{y}) - \lambda^V \nabla_x g(\bar{x}, \bar{y})^T \lambda + \nabla_x g(\bar{x}, \bar{y})^T \lambda_g + \nabla_x G(\bar{x}, \bar{y})^T \lambda_G, \\
&0 = \nabla_y F(\bar{x}, \bar{y}) + \lambda^V \nabla_y f(\bar{x}, \bar{y}) + \nabla_y g(\bar{x}, \bar{y})^T \lambda_g + \nabla_y G(\bar{x}, \bar{y})^T \lambda_G \\
&\lambda^V \geq 0,\ 0 \leq \lambda_g \perp g(\bar{x}, \bar{y}),\ \lambda_g \perp \nabla g(\bar{x}, \bar{y})(u, v)\ , 0 \leq \lambda_G \perp G(\bar{x}, \bar{y}),\ \lambda_G \perp \nabla G(\bar{x}, \bar{y})(u, v) \\
&0 = \nabla_y f(\bar{x}, \bar{y}) + \nabla_y g(\bar{x}, \bar{y})^T \lambda, 0 \leq \lambda \perp g(\bar{x}, \bar{y}), \lambda \perp \nabla g(\bar{x}, \bar{y})(u, v).
\end{aligned}
$$

Proof. By Proposition 3.4.4, since $\mathcal{F}(x)$ is RCR-regular at $(\bar{x}, \bar{y})$ and $S(x)$ is inner semi-continuous at $(\bar{x}, \bar{y})$ in direction u, $V(x)$ is directional differentiable at $\bar{x}$ in direction u and $V'(\bar{x}; u) = \min_{v \in \mathbb{L}(\bar{x}, \bar{y}; u)} \nabla f(\bar{x}, \bar{y})(u, v)$. Since $(u, v) \in C(\bar{x}, \bar{y})$, we have $\nabla f(\bar{x}, \bar{y})(u, v) - V'(\bar{x}; u) = 0$. Hence $v \in \Sigma(\bar{x}, \bar{y}, u) = \{v \in \mathbb{L}(x, y; u) | V'(x; u) = \nabla f(x, y)(u, v)\}$. Then by Theorem 3.4.3(iii), $V(x)$ is Lipschitz continuous at $\bar{x}$ in direction u and $\partial^c V(\bar{x}; u) \subseteq W(\bar{x}, \bar{y}; u, v)$. The rest of result follows from Theorem 3.5.1. ■

The following example verifies Theorem 3.5.3. For this example, $S(x)$ is not inner semi-continuous at $\bar{x}$ but it is directional inner semi-continuous, the classical quasi-normality fails but the directional quasi-normality holds.

Example 3.5.1. *Consider the following bilevel program*

$$
\begin{aligned}
\min_{x,y} \quad & F(x, y) := (\sqrt{3}x - y - \sqrt{3})^2 + x + \sqrt{3}y + 3 \\
\text{s.t.} \quad & y \in S(x) := \arg\min_y \{1 - (x - y)^2 : (x - 1)^2 + y^2 - 4 \leq 0, -\sqrt{3}x - y - \sqrt{3} \leq 0\}.
\end{aligned}
$$

It is easy to verify that

$$
S(x) = \begin{cases} \sqrt{4 - (x - 1)^2}, & -1 \leq x < 0, \\ \{-\sqrt{3}, \sqrt{3}\}, & x = 0, \\ -\sqrt{4 - (x - 1)^2}, & 0 < x \leq 3. \end{cases} \tag{3.40}
$$

$$
V(x) = \begin{cases} 1 - (x - \sqrt{4 - (x - 1)^2})^2, & -1 \leq x < 0, \\ -2, & x = 0, \\ 1 - (x + \sqrt{4 - (x - 1)^2})^2, & 0 < x \leq 3. \end{cases} \tag{3.41}
$$

Note that the value function is Lipschitz continuous at $\bar{x} = 0$ *but not smooth. The* *global optimal solution of the bilevel program is* $(\bar{x},\bar{y}) = (0,-\sqrt{3})$. *By* (3.40), $S(x)$ *is inner semi-continuous at* $\bar{y}$ *in any direction* $u > 0$. *Indeed, for any sequence* $x \to \bar{x}$ *in direction* $u > 0$, $S(x) \to \bar{y}$. *It follows that* $S(\bar{x}; u) = \{\bar{y}\}$. *Note that since for any sequence* $x \to \bar{x}$ *in direction* $u < 0$, $S(x) \not\to \bar{y}$, $S(x)$ *is not inner semi-continuous at* $\bar{x}$.

Denote by $f(x, y) := 1-(x-y)^2, g_1(x, y) := (x-1)^2+y^2-4, g_2(x, y) := -\sqrt{3}x - y - \sqrt{3}$. *Then*

$$\nabla F(\bar{x},\bar{y}) = \begin{bmatrix} 1 \\ \sqrt{3} \end{bmatrix}, \quad \nabla f(\bar{x},\bar{y}) = \begin{bmatrix} -2\sqrt{3} \\ 2\sqrt{3} \end{bmatrix} \quad \nabla g_1(\bar{x},\bar{y}) = \begin{bmatrix} -2 \\ -2\sqrt{3} \end{bmatrix}, \quad \nabla g_2(\bar{x},\bar{y}) = \begin{bmatrix} -\sqrt{3} \\ -1 \end{bmatrix}.$$

It is easy to see that the rank of the gradient vectors $\{\nabla_y g_1(x, y), \nabla_y g_2(x, y)\}$ *is always equal to* 1 *around* $(\bar{x},\bar{y})$ *and hence, RCR-regularity holds at* $(\bar{x},\bar{y})$. *Since* $g_1(\bar{x},\bar{y}) = 0, g_2(\bar{x},\bar{y}) = 0$,

$$\Lambda(\bar{x},\bar{y}) := \{(\lambda^1,\lambda^2) \in \mathbb{R}^2_+ | 2\sqrt{3} - 2\sqrt{3}\lambda^1 - \lambda^2 = 0\}.$$

Then by Theorem 3.5.3, $V(x)$ *is Lipschitz continuous and directionally differentiable in direction* $u > 0$ *and*

$$\begin{aligned} V'(\bar{x}; u) &= \max\{\nabla_x \mathcal{L}(\bar{x},\bar{y};\lambda^1,\lambda^2)u : (\lambda^1,\lambda^2) \in \Lambda(\bar{x},\bar{y})\} \\ &= \max\{(-2\sqrt{3} - 2\lambda^1 - \sqrt{3}\lambda^2)u | (\lambda^1,\lambda^2) \in \mathbb{R}^2_+, 2\sqrt{3} - 2\sqrt{3}\lambda^1 - \lambda^2 = 0\} \\ &= \max\{(-2\sqrt{3} + 4\lambda^1 - 6)u | 0 \le \lambda^1 \le 1\} \\ &= -(2\sqrt{3}+2)u. \end{aligned}$$

Moreover we can verify that this statement is correct by the expression (3.41). Now we prove that the directional quasi-normality holds at $(\bar{x},\bar{y})$. *The critical cone can be calculated as*

$$\begin{aligned} C(\bar{x},\bar{y}) := &\{(u, v) | \nabla F(\bar{x},\bar{y})(u, v) \le 0, \nabla f(\bar{x},\bar{y})(u, v) - V'(\bar{x}; u) = 0, \nabla g(\bar{x},\bar{y})(u, v) \le 0\} \\ = &\{(u, v) | u + \sqrt{3}v = 0, \sqrt{3}u + v \ge 0\}. \end{aligned}$$

Let $\bar{u} = \sqrt{3}$ *and* $\bar{v} = -1$, *we have* $(\bar{u},\bar{v}) \in C(\bar{x},\bar{y})$. *Since* $g_1(\bar{x},\bar{y}) = g_2(\bar{x},\bar{y}) =$

$0, \nabla g_1(\bar{x},\bar{y})(\bar{u},\bar{v}) = 0, \nabla g_2(\bar{x},\bar{y})(\bar{u},\bar{v}) = -\sqrt{3}\bar{u} - \bar{v} \neq 0,$ *we have*

$$\begin{aligned} W(\bar{x},\bar{y},\bar{u},\bar{v}) &:= \{\nabla_x f(\bar{x},\bar{y}) + \nabla_x g(\bar{x},\bar{y})^T \lambda_g | \lambda_g \in \Lambda(\bar{x},\bar{y}) \cap \{\nabla g(\bar{x},\bar{y})(\bar{u},\bar{v})\}^\perp\} \\ &= \{-2\sqrt{3} - 2\lambda_g^1 | \lambda_g^1 \geq 0, 2\sqrt{3} - 2\sqrt{3}\lambda_g^1 = 0\} \\ &= \{-2\sqrt{3} - 2\}. \end{aligned}$$

Since $(\bar{u},\bar{v}) \in C(\bar{x},\bar{y})$, *by (3.26) we have* $\bar{v} \in \Sigma(\bar{x},\bar{y},\bar{u})$. *Therefore by Theorem 3.4.3, we have* $\partial^c V(\bar{x};\bar{u}) \subseteq W(\bar{x},\bar{y},\bar{u},\bar{v})$. *Since* $V(x)$ *is a function of one variable, we can verify by the expression of the value function (3.41) that*

$$\partial^c V(\bar{x};\bar{u}) = W(\bar{x},\bar{y},\bar{u},\bar{v}) = \{-2\sqrt{3} - 2\}.$$

Let α, ν_1, ν_2 *be such that*

$$0 \in \alpha(\nabla_x f(\bar{x},\bar{y}) - W(\bar{x},\bar{y},\bar{u},\bar{v})) + \nu_1 \nabla_x g_1(\bar{x},\bar{y}) + \nu_2 \nabla_x g_2(\bar{x},\bar{y}), \tag{3.42}$$

$$0 = \alpha \nabla_y f(\bar{x},\bar{y}) + \nu_1 \nabla_y g_1(\bar{x},\bar{y}) + \nu_2 \nabla_y g_2(\bar{x},\bar{y}), \tag{3.43}$$

$$\nu_2 \nabla g_2(\bar{x},\bar{y})(\bar{u},\bar{v}) = 0, \alpha \geq 0,\ \nu_1 \geq 0,\ \nu_2 \geq 0 \tag{3.44}$$

and there exist sequences $t_k \downarrow 0,\ (u^k,v^k) \to (\bar{u},\bar{v})$, *such that*

$$f(\bar{x}+t_k u^k, \bar{y}+t_k v^k) - V(\bar{x}+t_k u^k) > 0 \text{ if } \alpha > 0, \tag{3.45}$$

$$g_1(\bar{x}+t_k u^k, \bar{y}+t_k v^k) > 0 \text{ if } \nu_1 > 0. \tag{3.46}$$

$$g_2(\bar{x}+t_k u^k, \bar{y}+t_k v^k) > 0 \text{ if } \nu_2 > 0. \tag{3.47}$$

(3.44) implies that $\nu_2 = 0$ *and (3.47) will not be needed. We now show the conditions (3.42)-(3.46) can only hold if* $\alpha = \nu_1 = \nu_2 = 0$. *By* (3.43), $2\sqrt{3}\alpha - 2\sqrt{3}\nu_1 = 0$. *Hence* $\alpha = \nu_1$. *To the contrary, assume* $\alpha > 0$. *Then* $\nu_1 = \alpha > 0$. *Let* $t_k \downarrow 0,\ (u^k,v^k) \to (\bar{u},\bar{v})$ *be arbitrary and suppose that (3.46) holds. Then* $g_1(x^k,y^k) > 0$ *for* $(x^k,y^k) := (\bar{x}+t_k u^k, \bar{y}+t_k v^k)$. *It follows that* $y^k < -\sqrt{4-(x^k-1)^2}$. *Since* $\nabla_y f(x^k, -\sqrt{4-(x^k-1)^2}) = 2(x^k + \sqrt{4-(x^k-1)^2}) > 0$ *and* $y^k < -\sqrt{4-(x^k-1)^2}$ *we have* $f(x^k,y^k) < f(x^k, -\sqrt{4-(x^k-1)^2}) = V(x^k)$, *where the last equality follows from* (3.40). *Hence (3.45) does not hold. The contradiction show that* $(\alpha,\nu_1,\nu_2) = (0,0,0)$ *and directional quasi-normality holds at* $(\bar{x},\bar{y})$ *in direction* $(\bar{u},\bar{v})$.

By now, the conditions in Theorem 3.5.3 are all verified and so the directional KKT condition should hold at $(\bar{x},\bar{y})$. *That is, there exists a nonzero vector* $(\lambda_V,\lambda,\lambda_g) \in$

$\mathbb{R}^{1+2+2}$ *such that*

$$\begin{aligned}
&0 = 1 - \lambda_V(-2\lambda^1 - \sqrt{3}\lambda^2) - 2\lambda_g^1 - \sqrt{3}\lambda_g^2,\\
&0 = \sqrt{3} + \lambda_V 2\sqrt{3} - 2\sqrt{3}\lambda_g^1 - \lambda_g^2,\\
&\lambda_g, \lambda \in \Lambda(\bar{x}, \bar{y}), \lambda_g \perp \nabla g(\bar{x}, \bar{y})(\bar{u}, \bar{v}), \lambda \perp \nabla g(\bar{x}, \bar{y})(\bar{u}, \bar{v}).
\end{aligned}$$

Obviously the vectors $(\lambda_V, \lambda, \lambda_g) := (\frac{1}{2}, (1,0), (1,0))$ *satisfies the above conditions.*

As we have mentioned before, NNAMCQ and FOSCMS always fail for (BP). In this example, the quasi-normality also fails at $(\bar{x}, \bar{y})$*. Indeed, let* $(\alpha, \nu_1, \nu_2) = (1, 1, 0)$*. We have* (α, ν_1, ν_2) *satisfies* (3.42) *and* (3.43)*. And choose* $(x^k, y^k) := (-1/k - \sqrt{4-(1/k+1)^2} - 1/k)$*, which converges to* $(\bar{x}, \bar{y})$*. By* (3.40)*, we have*

$$\begin{aligned}
f(x^k, y^k) &= 1 - \left(\sqrt{4-(1/k+1)^2}\right)^2 > 1 - \left(1/k + \sqrt{4-(1/k+1)^2}\right)^2 = V(x^k),\\
g_1(x^k, y^k) &= (1/k+1)^2 + \left(\sqrt{4-(1/k+1)^2} + 1/k\right)^2 - 4 > 0.
\end{aligned}$$

By the definition of the classical quasi-normality defined in [30, Definition 4.2] (one can refer to Definition 3.3.2 for the case $u = 0$*), this means that the quasi-normality fails at* $(\bar{x}, \bar{y})$*.*

Bibliography

[1] J. Abadie, *On the KuhnTucker theorem, in Nonlinear Programming*, J. Abadie, ed., North Holland, Amsterdam, 1967, pp. 1936.

[2] J. -P. Aubin, *Lipschitz behavior of solutions to convex minimization problems*, Math. Oper. Res., 9 (1984), pp. 87–111.

[3] K. Bai and J. J. Ye, *Directional necessary optimality conditions for bilevel programs*, submitted to Math. Oper. Res., 2020. arXiv: 2004.01783.

[4] K. Bai, J. J. Ye and J. Zhang, *Directional quasi-/pseudo-normality as sufficient conditions for metric subregularity*, SIAM J. Optim., 29 (2019), pp. 2625-2649.

[5] E. M. Bednarczuk, L. I. Minchenko and K. E. Rutkowski, *On Lipschitz-like continuity of a class of set-valued mappings*, Optimization, 2019, pp. 1-15.

[6] M. Benko, M. Červinka and T. Hoheisel, *Sufficient conditions for metric subregularity of constraint systems with applications to disjunctive and ortho-disjunctive programs*, submitted to Set-Valued Var. Anal., 2019. arXiv: 1906.08337.

[7] M. Benko, H. Gfrerer and J. V. Outrata, *Calculus for directional limiting normal cones and subdifferentials*, Set-Valued Var. Anal., (2018), pp. 713-745.

[8] D. P. Bertsekas, *Nonlinear Programming, 2nd ed.*, Athena Scientific, Belmont, MA, 1999.

[9] D. P. Bertsekas and A. E. Ozdaglar, *Pseudonormality and a Lagrange multiplier theory for constrained optimization*, J. Optim. Theory Appl., 114 (2002), pp. 287-343.

[10] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.

[11] J. V. Burke, *Calmness and exact penalization*, SIAM J. Control Optim., 29 (1991), pp. 493–497.

[12] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics Appl. Math. 5, SIAM, Philadelphia, PA, 1990.

[13] S. Dempe, *Foundations of Bilevel Programming*, Kluwer Academic, Dordrecht 2002.

[14] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings*, Springer, New York, 2009.

[15] A. L. Dontchev and R. T. Rockafellar, *Regularity and conditioning of solution mappings in variational analysis*, Set-Valued Anal., 12 (2004), pp. 79–109.

[16] M. J. Fabian, R. Henrion, A.Y. Kruger and J. V. Outrata, *Error bounds: necessary and sufficient conditions*, Set-Valued Var. Anal., 18 (2010), pp. 121–149.

[17] F. Facchinei, A. Fischer and M. Herrich, *An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions*, Math. Program., 146 (2014), pp. 1–36.

[18] F. Facchinei and J. -S Pang, *Finite Dimensional Variational Inequalities and Complementarity Problems*, Vol. I, Springer Ser. Oper. Res. Financ. Eng., Springer, New York, 2003.

[19] A. Fischer, *Local behavior of an iterative framework for generalized equations with nonisolated solutions*, Math. Program., 94 (2002), pp. 91–124.

[20] J. Gauvin and F. Dubeau, *Differential properties of the marginal function in mathematical programming*. Math. Program. Stud., 19 (1982), pp. 101-119.

[21] H. Gfrerer, *On directional metric regularity, subregularity and optimality conditions for nonsmooth mathematical programs*, Set-Valued Var. Anal., 21 (2013), pp. 151–176.

[22] H. Gfrerer, *On directional metric subregularity and second-order optimality conditions for a class of nonsmooth mathematical programs*, SIAM J. Optim., 23 (2013), pp. 632–665.

[23] H. Gfrerer, *On metric pseudo-(sub) regularity of multifunctions and optimality conditions for degenerated mathematical programs*, Set-Valued Var. Anal., 22 (2014), pp. 79–115.

[24] H. Gfrerer, *Optimality conditions for disjunctive programs based on generalized differentiation with application to mathematical programs with equilibrium constraints*, SIAM J. Optim., 24 (2014), pp. 898–931.

[25] H. Gfrerer and D. Klatte, *Lipschitz and Hölder stability of optimization problems and generalized equations*, Math. Program., 158 (2016), pp. 35–75.

[26] H. Gfrerer and B. Mordukhovich, *Robinson stability of parametric constraint systems via variational analysis*, SIAM J. Optim. 27 (2017), pp. 438–465.

[27] H. Gfrerer and J. J. Ye, *New constraint qualifications for mathematical programs with equilibrium constraints via variational analysis*, SIAM J. Optim., 27 (2017), pp. 842–865.

[28] H. Gfrerer and J. J. Ye, *New sharp necessary optimality conditions for mathematical programs with equilibrium constraints*, Set-Valued Var. Anal., 28 (2020), pp. 395-426.

[29] I. Ginchev and B. S. Mordukhovich, *On directionally dependent subdifferentials*, C. R. Bulg. Acad. Sci., 64 (2011), pp. 497–508.

[30] L. Guo, J. J. Ye and J. Zhang, *Mathematical programs with geometric constraints in Banach spaces: enhanced optimality, exact penalty, and sensitivity*, SIAM J. Optim., 23 (2013), pp. 2295-2319.

[31] L. Guo, G. H. Lin, J. J. Ye, and J. Zhang, *Sensitivity analysis of the value function for parametric mathematical programs with equlibrium constraints*, SIAM J. Optim. 24 (2014), pp. 1206-1237.

[32] W. Hager and M. Gowda, *Stability in the presence of degeneracy and error estimation*, Math. Program., 85 (1999), pp. 181–192.

[33] R. Henrion and A. Jourani, *Subdifferential conditions for calmness of convex constraints*, SIAM J. Optim., 13 (2002), pp. 520–534.

[34] R. Henrion, A. Jourani and J. V. Outrata, *On the calmness of a class of multifunctions*, SIAM J. Optim., 13 (2002), pp. 603–618.

[35] R. Henrion and J. V. Outrata, *A subdifferential condition for calmness of multifunctions*, J. Math. Anal. Appl., 258 (2001), pp. 110–130.

[36] R. Henrion and J. V. Outrata, *Calmness of constraint systems with applications*, Math. Program., 104 (2005), pp. 437–464.

[37] A. J. Hoffman, *On approximate solutions of systems of linear inequalities*, J. Res. Natl. Bur. Standards, 49 (1952), pp. 263- 265.

[38] A. D. Ioffe, *Necessary and sufficient conditions for a local minimum. 1: a reduction theorem and first order conditions*, SIAM J. Control Optim., 17 (1979), pp. 245–250.

[39] A. D. Ioffe, *Regular points of Lipschitz functions*, Trans. Amer. Math. Soc., 251 (1979), pp. 61–69.

[40] A. D. Ioffe, *Metric regularity and subdifferential calclulus*, Russian Math. Surveys, 55 (2000), pp. 501–558.

[41] A. D. Ioffe and J. V. Outrata, *On metric and calmness qualification conditions in subdifferential calculus*, Set-valued Anal., 16 (2008), pp. 199–227.

[42] A. F. Izmailov, *On the analytical and numerical stability of critical Lagrange multipliers*, Comput. Math. Math. Phys., 45 (2005), pp. 930–946.

[43] A. F. Izmailov and M. V. Solodov, *Stabilized SQP revisited*, Math. Program., 133 (2012), pp. 93–120.

[44] R. Janin, *Directional derivative of the marginal function in nonlinear programming*, Math. Program. Study, 21 (1984), pp. 110–126.

[45] F. John, *Extremum problems with inequalities as subsidiary conditions*, Traces and emergence of nonlinear programming (pp. 197-215). Springer Basel, 2014.

[46] V. Kalashnikov, S. Dempe, G. Prez-Valds, N. Kalashnykova and J. Camacho-Vallejo, *Bilevel programming and applications*. Math. Probl. Eng., 2015.

[47] C. Kanzow and A. Schwartz, *Mathematical programs with equilibrium constraints: enhanced Fritz-John conditions, new constraint qualifications, and improved exact penalty results*, SIAM J. Optim., 20 (2010), pp. 2730–2753.

[48] D. Klatte, *A note on quantitative stability results in nonlinear optimization*, in Proceedings of the 19. Jahrestagung Mathematische Optimierung, Seminarbericht 90, Humboldt-Universität Berlin, Berlin, Sektion Mathematik, 1987, pp. 77–86.

[49] D. Klatte and B. Kummer, *Constrained minima and Lipschitzian penalties in metric spaces*, SIAM J. Optim., 13 (2002), pp. 619–633.

[50] D. Klatte and B. Kummer, *Optimization methods and stability of inclusions in Banach spaces*, Math. program., 117 (2009), pp. 305–330.

[51] G. Kunapuli, K. P. Bennet, J. Hu and J. S. Pang, *Classification model selection via bilevel programming*. Optim. methods Softw., 23 (2008), pp. 475-489.

[52] K. Lachhwani and A. Dwivedi, *Bi-level and multi-Level programming problems: taxonomy of literature review and research issues*, Arch. Compt. Method E., 4 (2018), pp. 847-877.

[53] W. Li, *Abadie's constraint qualification, metric regularity, and error bounds for differentiable convex inequalities*, SIAM J. Optim., 7 (1997), pp. 966–978.

[54] P. Long, B. Wang and X. Yang, *Calculus of directional subdifferentials and coderivatives in Banach spaces*, Positivity, 21 (2017), pp. 223-254.

[55] Z. Q. Luo, J. S. Pang and D. Ralph, *Mathematical Programs with Equlibrium Constriants*, Cambridge University Press, 1996.

[56] O. L. Mangasarian and S. Fromovitz, *The Fritz-John necessary optimality conditions in the presence of equality and inequality constraints*, J. of Math. Anal. appl.. 17 (1967), pp. 37-47.

[57] L. Minchenko and S. Stakhovski, *Parametric nonlinear programming problems under the relaxed constant rank condition*, SIAM J. Optim. 21 (2011), pp. 314-332.

[58] L. Minchenko and A. Tarakanov, *On error bounds for quasinormal programs*, J. Optim. Theory Appl., 148 (2011), pp. 571–579.

[59] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation I. Basic Theory.* Ser. Comprehensive Stud. Math. 330, Springer, Berlin, 2006.

[60] B. S. Mordukhovich and M. E. Sarabi, *Critical multipliers in variational systems via second-order generalized differentiation*, Math. Program., 169 (2018), pp. 605–648.

[61] H. V. Ngai and P. N. Tinh, *Metric subregularity of multifunctions: first and second order infinitesimal characterizations*, Math. Oper. Res., 40 (2015), pp. 703–724.

[62] M. Nikolova, *Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers*, SIAM J. Numer. Anal., 40 (2002), pp. 965–994.

[63] J. V. Outrata, M. Koĉvara and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equlibrium Constraints: Theory, Applications and Numerical Results*, Kluwer, Dordrecht, The Netherlands, 1998.

[64] J. -P. Penot, *Error bounds, calmness and their applications in nonsmooth analysis*, in Nonlinear Analysis and Optimization II: Optimization, Comtemp. Math. 514, A. Leizarowitz, B.S. Mordukhovich, I. Shafrir and A.J. Zaslavski, eds., American Mathematical Society, Providence, RI, 2010, pp. 225–247.

[65] T. T. Pham, *Maximum likelihood estimation of a class of non-Gaussian densities with application to l_p deconvolution.* IEEE Trans. Acoust. Speech Signal Process., 37 (1989), pp. 73–82.

[66] J. R. Rice and J. S. White, *Norms for smoothing and estimation*, SIAM review, 6 (1964), pp. 243–256.

[67] S. M. Robinson, *An application of error bounds for convex programming in linear space*, SIAM J. Control, 13 (1975), pp. 271 -273.

[68] S. M. Robinson, *Stability theory for systems of inequalities. part* I: *linear systems*, SIAM J. Numer. Anal., 12 (1975), pp. 754–769.

[69] S. M. Robinson, *Stability theory for systems of inequalities. part* II: *differentiable nonlinear systems*, SIAM J. Numer. Anal., 13 (1976), pp. 497–513.

[70] S. M. Robinson, *Some continuity properties of polyhedral multifunctions*, Mathematical Programming at Oberwolfach, Math. Program. Stud., 14, Springer Berlin, 1981, pp. 206–214.

[71] S. M. Robinson, *Generalized equations and their solution. part II: applications to nonlinear programming*, in Optimality and Stability in Mathematical Programming, Math. Program. Stud., 19, Springer Berlin, 1982, pp. 200–221.

[72] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Grundlehren Math. Wiss., 317, Springer, Berlin, 1998.

[73] S. Scholtes and M. Stohr, *How stringent is the linear independence assumption for mathematical programs with complementarity constraints?*, Math. Oper. Res., 26 (2001), pp. 851–863.

[74] H. von Stackelberg, *Marktform and gleichgewicht*, Springer-Verlag, Berlin, 1934, engl. transl.: *The theory of the market economy*, Oxford University Press, Oxford, England, 1954.

[75] M. Studniarski and D. E. Ward, *Weak sharp minima: characterizations and sufficient conditions*, SIAM J. Control Optim., 38 (1999), pp. 219–236.

[76] X. Wang, J. J. Ye, X. Yuan, S. Zeng and J. Zhang, *Perturbation techniques for convergence analysis of proximal gradient method and other first-order algorithms via variational analysis*, to appear in Set-Valued Var. Anal., arXiv:1810.10051.

[77] Z. Wu and J. J. Ye, *Sufficient conditions for error bounds*, SIAM J. Optim., 12 (2002), pp. 421–435.

[78] Z. Wu and J. J. Ye, *On error bounds for lower semicontinuous functions*, Math. Program., 92 (2001), pp. 301–314.

[79] Z. Wu and J. J. Ye, *First-order and second-order conditions for error bounds*, SIAM J. Optim., 14 (2003), pp. 621–645.

[80] M. Xu and J. J. Ye, *Relaxed constant positive linear dependence constraint qualification and its application to bilevel programs*, J. Global Optim., Springer, 0 (2020), pp. 1-25.

[81] J. J. Ye, *Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints*, SIAM J. Optim.,10 (2000), pp. 943-962.

[82] J. J. Ye, *Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints*, J. Math. Anal. Appl., 307 (2005), pp. 350–369.

[83] J. J. Ye, *Constraint qualifications and necessary optimality conditions in bilevel optimization*, Preprint, arXiv (1910.04072).

[84] J. J. Ye and X. Y. Ye, *Necessary optimality conditions for optimization problems with variational inequality constraints*, Math. Oper. Res., 22 (1997), pp. 977–997.

[85] J. J. Ye and J. Zhang, *Enhanced Karush-Kuhn-Tucker condition and weaker constraint qualification*, Math. Program., 139 (2013), pp. 353–381.

[86] J. J. Ye and J. Zhang, *Enhanced Karush-Kuhn-Tucker condition for mathematical programs with equilibrium constraints*, J. Optim. Theory Appl., 163 (2014), pp. 777–794.

[87] J. J. Ye and J. C. Zhou, *Verifiable sufficient conditions for local error bound property of second-order cone complementarity problems*, Math. Program., 171 (2018), pp. 361–395.

[88] J. J. Ye and D. L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization, 33 (1995), pp. 9-27.

[89] J. J. Ye and D. L. Zhu, *A note on optimality conditions for bilevel programming problems*, Optimization, 39 (1997), pp. 361-366.

[90] J. J. Ye and D. L. Zhu, *New necessary optimality conditions for bilevel programs by combining MPEC and value function approaches*, SIAM J. Optim., 20 (2010), pp. 1885-1905.

[91] J. J. Ye, D. L. Zhu and Q. J. Zhu, *Exact penalization and necessary optimality conditions for generalized bilevel programming problems*, SIAM J. Optim., 2 (1997), pp. 481-507.

[92] X. Y. Zheng and K. F. Ng, *Metric subregularity and constraint qualifications for nonconvex generalized equations in Banach spaces*, SIAM J. Optim., 18 (2010), pp. 2119–2136.