

Mathematical programs with geometric constraints in Banach spaces: enhanced optimality, exact penalty, and sensitivity

Lei Guo*, Jane J. Ye[†] and Jin Zhang[‡]

February 2013, Revised June 2013

Abstract. In this paper we study the mathematical program with geometric constraints (MPGC) such that the image of a mapping from a Banach space is included in a closed subset of a finite dimensional space. We obtain the nonsmooth enhanced Fritz John necessary optimality conditions in terms of the approximate subdifferential. In the case where the Banach space is a weakly compactly generated Asplund space, the optimality condition obtained can be expressed in terms of the limiting subdifferential while in the general case it can be expressed in terms of the Clarke subdifferential. One of the technical difficulties in obtaining such a result in an infinite dimensional space is that no compactness results can be used to show the existence of local minimizers of a perturbed problem. In this paper we employ the celebrated Ekeland's variational principle to obtain the results instead. The enhanced Fritz John condition allows us to obtain the enhanced Karush-Kuhn-Tucker condition under the pseudonormality and the quasinormality conditions which are weaker than the classical normality conditions. We then prove that the quasinormality is a sufficient condition for the existence of local error bounds of the constraint system. Finally we obtain a tighter upper estimate for the subdifferentials of the value function of the perturbed problem in terms of the enhanced multipliers.

Key Words. mathematical program with geometric constraints, variational analysis, approximate subdifferential, optimality, error bound, exact penalty, sensitivity.

2010 Mathematics Subject Classification. 90C26, 90C30, 90C31, 90C46.

*School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China. E-mail: guolayne@gmail.com.

[†]Department of Mathematics and Statistics, University of Victoria, Victoria, BC, V8W 3R4 Canada. E-mail: janeye@uvic.ca. The research of this author was partially supported by NSERC.

[‡]Department of Mathematics and Statistics, University of Victoria, Victoria, BC, V8W 3R4 Canada. E-mail: zhangjin@uvic.ca. The research of this author was partially supported by China Scholarship Council and partially by NSERC.

1 Introduction

In this paper unless otherwise stated, we denote by \mathbb{X} a Banach space and by \mathbb{X}^* its dual space equipped with the weak* topology and by \mathbb{Y} a m -dimensional Hilbert space over \mathbb{R} together with the inner product $\langle \cdot, \cdot \rangle$ equipped with the orthogonal basis $\mathcal{E} = \{e_1, \dots, e_m\}$. We study the following mathematical program with geometric constraints such that the image of a mapping from a Banach space is included in a closed subset of a finite dimensional space:

$$\begin{aligned} \text{(MPGC)} \quad & \min_{x \in \Omega} f(x) \\ & \text{s.t.} \quad F(x) \in \Lambda, \end{aligned} \tag{1.1}$$

where $f : \mathbb{X} \rightarrow \mathbb{R}$ and $F : \mathbb{X} \rightarrow \mathbb{Y}$ are Lipschitzian near the point of interest, Ω and Λ are nonempty closed subset of \mathbb{X} and \mathbb{Y} respectively. Problem (MPGC) includes as special cases the conventional nonlinear program, the cone constrained program, the mathematical program with equilibrium constraints [17, 25], the problems considered in [10, 27], the semidefinite program, and the mathematical program with semidefinite cone complementarity constraints [8].

The classical Fritz John (FJ) necessary optimality condition for (MPGC) with continuously differentiable functions $\{f, F\}$, $\Omega = \mathbb{X}$, and convex set constraint Λ takes the following form: There exists scalar $r \geq 0$ and $\mu \in \mathbb{Y}$ not all equal to zero such that

$$0 = r \nabla f(x^*) + \nabla F(x^*)^* \mu \quad \text{and} \quad \mu \in \mathcal{N}_\Lambda(F(x^*)), \tag{1.2}$$

where $\nabla \varphi(x)$ is the Fréchet derivative of mapping φ at x , \mathcal{A}^* denotes the adjoint of a linear operator \mathcal{A} , and $\mathcal{N}_\Lambda(y)$ denotes the normal cone of Λ at y in the sense of convex analysis [28]:

$$\mathcal{N}_\Lambda(y) = \begin{cases} \{d \in \mathbb{Y} \mid \langle d, z - y \rangle \leq 0 \ \forall z \in \Lambda\} & \text{if } y \in \Lambda, \\ \emptyset & \text{if } y \notin \Lambda. \end{cases}$$

From the FJ condition, it follows immediately that if x^* is a locally optimal solution of (MPGC) and the *no nonzero abnormal multiplier constraint qualification* (NNAMCQ) or *basic constraint qualification* (Basic CQ) [29] holds at x^* , i.e., there is no nonzero μ such that

$$0 = \nabla F(x^*)^* \mu \quad \text{and} \quad \mu \in \mathcal{N}_\Lambda(F(x^*)),$$

then there exist $r > 0$ (which can be taken as 1) and $\mu \in \mathbb{Y}$ such that the KKT condition holds

$$0 = \nabla f(x^*) + \nabla F(x^*)^* \mu \quad \text{and} \quad \mu \in \mathcal{N}_\Lambda(F(x^*)).$$

Since \mathbb{Y} is assumed to be a finite dimensional space and Λ is a closed convex set, by virtue of [4, Corollary 2.98], the NNAMCQ is equivalent to the Robinson's CQ

$$0 \in \text{int}\{F(x^*) + \nabla F(x^*)\mathbb{X} - \Lambda\},$$

where “int” denotes the topological interior of a given set.

When $\{f, F\}$ are nonsmooth but locally Lipschitzian and Λ is not a convex set, the FJ condition can be obtained by replacing the usual derivatives and the normal cone in the sense of convex analysis by the limiting subdifferential and the limiting normal cone respectively if the underlying space \mathbb{X} is an Asplund space (an Asplund space is a Banach space \mathbb{X} such that the dual of every closed separable subspace of \mathbb{X} is separable and a Banach space, see Mordukhovich [20]), and by the Clarke subdifferential and the Clarke normal cone respectively if \mathbb{X} is a general Banach space (see Clarke [5]).

Although the NNAMCQ or the Basic CQ provides an easy way to verify constraint qualification, it may be fairly strong for some applications and in particular for certain classes of optimization problems such as bilevel programs, mathematical programs with equilibrium constraints [8, 17, 25, 34–36], they are never satisfied. In the last two decades, tremendous progress has been made towards developing weaker constraint qualifications and stronger necessary optimality conditions for the classical nonlinear program

$$\begin{aligned} \text{(NLP)} \quad & \min_{x \in \Omega} f(x) \\ & \text{s.t.} \quad h(x) = 0, \\ & \quad \quad g(x) \leq 0, \end{aligned}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}, h : \mathbb{R}^n \rightarrow \mathbb{R}^p, g : \mathbb{R}^n \rightarrow \mathbb{R}^q$ are continuously differentiable and Ω is a nonempty closed subset in \mathbb{R}^n . For (NLP) with $\Omega = \mathbb{R}^n$, the corresponding FJ condition asserts that for a locally optimal solution x^* of problem (NLP), there exist scalars $r, \lambda_1, \dots, \lambda_p$ and μ_1, \dots, μ_q , not all zero, such that $r, \mu_j \geq 0, j = 1, \dots, q$, and

$$r \nabla f(x^*) + \sum_{i=1}^p \lambda_i \nabla h_i(x^*) + \sum_{j=1}^q \mu_j \nabla g_j(x^*) = 0, \quad (1.3)$$

$$\mu_j g_j(x^*) = 0 \quad j = 1, \dots, q. \quad (1.4)$$

It follows immediately that the KKT condition holds under the NNAMCQ: There is no nonzero vector $\{\lambda, \mu\}$ such that

$$\begin{aligned} \sum_{i=1}^p \lambda_i \nabla h_i(x^*) + \sum_{j=1}^q \mu_j \nabla g_j(x^*) &= 0, \\ \mu_j \geq 0, \quad \mu_j g_j(x^*) &= 0 \quad j = 1, \dots, q. \end{aligned}$$

Using the Motzkin's transposition theorem, the NNAMCQ for problem (NLP) with $\Omega = \mathbb{R}^n$ can be shown to be equivalent to the well-known Mangasarian-Fromovitz constraint qualification (MFCQ), i.e., the gradient vectors

$$\{\nabla h_i(x^*) \mid i = 1, \dots, p\}$$

are linearly independent and there exists a vector $d \in \mathbb{R}^n$ such that

$$\langle \nabla h_i(x^*), d \rangle = 0 \quad i = 1, \dots, p, \quad \langle \nabla g_j(x^*), d \rangle < 0 \quad j \in I(x^*),$$

where $I(x^*) := \{j \mid g_j(x^*) = 0\}$ is the set of active inequality constraints at x^* . An enhanced version of the FJ condition (1.3)–(1.4) was proposed by Bertsekas in [1], i.e., if x^* is a locally optimal solution of problem (NLP), then there exist scalars $\lambda_1, \dots, \lambda_p$, and $\mu_1 \geq 0, \dots, \mu_q \geq 0$ not all zero satisfying (1.3) and the following sequential property: If the index set $I \cup J$ is nonempty, where

$$I = \{i \mid \lambda_i \neq 0\}, \quad J = \{j \mid \mu_j > 0\},$$

then there exists a sequence $\{x^k\}$ converging to x^* such that for all k ,

$$f(x^k) < f(x^*),$$

$$\lambda_i h_i(x^k) > 0 \quad \forall i \in I, \tag{1.5}$$

$$g_j(x^k) > 0 \quad \forall j \in J. \tag{1.6}$$

While there is no sign condition for an equality constraint in the classical FJ condition, condition (1.5) imposes a sequential sign condition for an equality constraint in the enhanced version. Condition (1.6), which is called the complementarity violation (CV) condition by Bertsekas and Ozdaglar [3], is stronger than the complementary slackness condition (1.4) since taking limits as k goes to infinity in condition CV (1.6) results in $g_j(x^*) \geq 0$, which implies that $g_j(x^*) = 0$ since x^* is feasible. The classical FJ condition is equivalent to the KKT condition under the NNAMCQ. Since the enhanced FJ condition is stronger than the classical FJ condition, it results in constraint qualifications such as quasinormality that are weaker than the NNAMCQ. Very recently, the enhanced KKT conditions for problem (NLP) with locally Lipschitzian data based on the limiting subdifferential and limiting normal cone were derived in [33] and the sensitivity of the value function was established in terms of the set of the enhanced KKT multipliers which may be smaller than the set of the classical KKT multipliers and hence provides sharper results.

The main purpose of this paper is to study the enhanced optimality condition for (MPGC) when the space \mathbb{X} is a Banach space and \mathbb{Y} is a finite dimensional space. Such a result is new even for the classical smooth nonlinear program. There are two

technical difficulties involved when the space \mathbb{X} is not finite dimensional. First, unlike the finite dimensional case, the quadratic penalization approach in [33] cannot be employed anymore because the compactness of the closed unit ball is possibly invalid in a Banach space \mathbb{X} , which plays a key role in guaranteeing the existence of enhanced sequential approximating solution by using the Weierstrass theorem in [33]. Nevertheless, by virtue of the optimization process, for any $\epsilon > 0$, a problem in the form as (MPGC) always possesses an ϵ -optimal solution (see [4] for definition), provided that the optimal value of the problem is finite. Inspired by this fact, we employ the Ekeland's variational principle instead to construct a cluster of ϵ -optimal solutions, and each of them becomes the minimizer of a certain slightly perturbed function. We then employ generalized calculus to obtain necessary optimality conditions for the perturbed problem. The second difficulty lies in applying the basic calculus rules and passing to the limit as ϵ tends to zero. When the space \mathbb{X} is finite dimensional, the limiting subdifferential and the limiting normal cone have nice calculus rules and are known to be closed as set-valued maps. The nice calculus rules and the robust property allow one to obtain the desired result. However, when \mathbb{X} is an infinite dimensional Banach space, the limiting subdifferential for a locally Lipschitzian function may even be empty and hence the basic calculus rules may fail and the robust property may not hold in general. To cope with the second difficulty, we use the approximate subdifferential developed by Ioffe [13, 14] instead. The approximate subdifferential is the most natural analytic tool in our situation since it has fairly rich calculus rule for a locally Lipschitzian function and the approximate subdifferential and the approximate normal cone are known to be closed as set-valued maps. Moreover, the approximate subdifferential for a locally Lipschitzian function is minimal (as a set) among all subdifferentials that have desired properties, and is in general smaller than the Clarke subdifferential. When the underlying space \mathbb{X} is a weakly compactly generated (WCG) Asplund space, the approximate subdifferential coincides with the limiting subdifferential [20, Theorem 3.59] and hence in this case, we obtain the desired result in terms of limiting subdifferential. Recall that \mathbb{X} is WCG if there is a weakly compact set $\mathbb{K} \subset \mathbb{X}$ such that \mathbb{X} is equal to the closure of the span of \mathbb{K} . Canonical examples of WCG Asplund spaces are reflexive Banach spaces. See, e.g., [20] for further discussions.

In recent years, it has been shown that constraint qualifications have strong connections with the stability of feasible region under certain perturbation p :

$$\mathcal{X}(p) := \{x \in \Omega \mid F(x, p) \in \Lambda\},$$

where p is in a topological space \mathbb{P} . For the case of a smooth optimization problem with convex geometric constraint Λ and $\mathbb{X} = \Omega$, it is known that Robinson's CQ at $x^* \in \mathcal{X}(p^*)$ implies the stability for the constraint region (see [4, Theorem 2.87]), i.e.,

the existence of a neighborhood \mathcal{U} of (x^*, p^*) such that for all $(x, p) \in \mathcal{U} \cap (\mathbb{X} \times \mathbb{P})$,

$$\text{dist}_{\mathcal{X}(p)}(x) = O(\text{dist}_\Lambda(F(x, p))),$$

and hence the existence of local error bounds, i.e., the existence of positive constants $\{\kappa, \delta_0\}$ such that

$$\text{dist}_{\mathcal{X}(p^*)}(x) \leq \kappa \text{dist}_\Lambda(F(x, p^*)) \quad \forall x \in \mathcal{B}(x^*, \delta_0) \cap \mathbb{X}.$$

In fact, the above stability results still hold in an infinite dimensional space even if the set Λ is not convex and \mathbb{X} is replaced with a closed subset Ω under the NNAMCQ which can be easily derived by using the error bound result as in [32, Theorem 2.4].

Error bounds have important applications in sensitivity analysis of mathematical programming and in convergence analysis of some algorithms. In his seminal paper [11], Hoffman showed that a linear inequality system in a finite dimensional space has a global error bound. Such a result was generalized to an infinite dimensional Banach space by Ioffe [12]. For a general constraint system, the existence of error bounds usually requires some conditions. As we discussed above, the Robinson's CQ and the NNAMCQ imply a local error bound for (MPGC). Therefore, the error bound estimates can be obtained straightforwardly for smooth nonlinear programs and nonlinear semidefinite programs (NLSDP) with the constraint systems taking the geometric forms respectively (see [4, Example 2.92, Example 2.93]). Very recently, for the case of nonsmooth (NLP), it was shown in [33] that either the pseudonormality or the quasinormality with regularity on the constraints implies the existence of local error bounds, which extends the result in [19] where all constraints are assumed to be twice continuously differentiable. In this paper, we show that a local error bound for nonsmooth (MPGC) exists under the quasinormality, which generalizes and improves all earlier results since except the constraint qualification, neither additional regularity condition nor continuous differentiability assumption is required.

The organization of this paper is as follows. We first give some background material in section 2. In section 3, we derive the enhanced FJ condition for (MPGC) and specialize the result to the case of conventional (NLP) and (NLSDP). Section 4 introduces some new weaker constraint qualifications for (MPGC) and discusses the relations between them. As our applications, under the new constraint qualifications, we show the local error bound results in section 5 and investigate the sensitivity results in section 6.

2 Preliminaries

We first give notation that will be used throughout the paper. We denote by \mathcal{X} the feasible region of (MPGC) and denote by $\mathcal{B}_\delta(x) := \{y \in \mathbb{X} \mid \|y - x\| < \delta\}$ the open ball

centered at x with radius $\delta > 0$. As usual, $\mathbb{B}_{\mathbb{X}}$ and $\widehat{\mathbb{B}}_{\mathbb{X}^*}$ stand for the closed unit balls of the space \mathbb{X} and its dual \mathbb{X}^* respectively. For a point $x \in \mathbb{X}$ and a set $C \subseteq \mathbb{X}$, we denote by $\text{dist}_C(x)$ the distance from x to C .

We next summarize some preliminary material in variational analysis that will be needed in this paper. We refer the reader to [5, 13, 14, 20, 30] for more details and discussions.

For a set-valued map $S : \mathbb{X} \rightrightarrows \mathbb{X}^*$, unless specified, we denote by

$$\begin{aligned} \text{Lim sup}_{x \rightarrow x^*} S(x) := \{v \in \mathbb{X}^* \mid \exists \text{ sequences } x_k \rightarrow x^* \text{ and } v_k \xrightarrow{w^*} v \\ \text{with } v_k \in S(x_k) \text{ for all } k\}. \end{aligned}$$

the sequential Painlevé–Kuratowski upper limit with respect to the norm topology of \mathbb{X} and the weak* topology of \mathbb{X}^* .

Given $\Omega \subset \mathbb{X}$ and $\epsilon \geq 0$, define the collection of ϵ -normals to Ω at $x^* \in \Omega$ by

$$\widehat{\mathcal{N}}_\epsilon(x^*, \Omega) := \{v \in \mathbb{X}^* \mid \limsup_{x \xrightarrow{\Omega} x^*} \frac{\langle v, x - x^* \rangle}{\|x - x^*\|} \leq \epsilon\}. \quad (2.1)$$

where $x \xrightarrow{\Omega} x^*$ means that $x \rightarrow x^*$ with $x \in \Omega$. When $\epsilon = 0$, elements of (2.1) are called Fréchet normals and their collection, denoted by $\widehat{\mathcal{N}}_\Omega(x^*)$, is the prenormal cone to Ω at x^* . The basic/limiting normal cone $\mathcal{N}_\Omega^L(x^*)$ to Ω at x^* is defined as

$$\mathcal{N}_\Omega^L(x^*) := \text{Lim sup}_{x \xrightarrow{\Omega} x^*, \epsilon \downarrow 0} \widehat{\mathcal{N}}_\epsilon(x, \Omega).$$

If \mathbb{X} is a Asplund space, then the limiting normal cone has the following simpler expression (see [20, Theorem 2.35])

$$\mathcal{N}_\Omega^L(x^*) := \text{Lim sup}_{x \xrightarrow{\Omega} x^*} \widehat{\mathcal{N}}_\Omega(x).$$

For $\Omega \subset \mathbb{X}$ and $x^* \in \Omega$, the contingent cone $\mathcal{T}_\Omega(x^*)$ to Ω at x^* is the set defined by

$$\mathcal{T}_\Omega(x^*) := \text{Lim sup}_{\epsilon \downarrow 0} \frac{\Omega - x^*}{t}, \quad (2.2)$$

where the “Lim sup” is taken with respect to the norm topology of \mathbb{X} . If the “Lim sup” in (2.2) is taken with respect to the weak topology of \mathbb{X} , then the resulting construction, denoted by $\mathcal{T}_\Omega^w(x^*)$, is called the weak contingent cone to Ω at x^* .

The Clarke tangent cone to Ω at x^* is defined by

$$\mathcal{T}_\Omega^c(x^*) := \{v \mid \forall x^k \rightarrow x^*, \forall t_k \downarrow 0, \exists v^k \rightarrow v^* \text{ s.t. } x^k + t_k v^k \in \Omega \text{ for all } k\},$$

and the Clarke normal cone to Ω at x^* is the dual to the Clarke tangent cone to Ω at x^* , i.e.,

$$\mathcal{N}_\Omega^c(x^*) := \mathcal{T}_\Omega^c(x^*)^\circ,$$

where $C^\circ := \{x : \langle x, v \rangle \leq 0, \forall v \in C\}$ denotes the polar of set C . In the general Banach space setting, we have

$$\text{cl}^* \text{conv} \mathcal{N}_\Omega^L(x^*) \subseteq \mathcal{N}_\Omega^c(x^*),$$

where “cl*conv” denotes the weak* closure of the convex hull and the inclusion relationship above holds with equality when \mathbb{X} is an Asplund space.

Let $\varphi : \mathbb{X} \rightarrow \overline{\mathbb{R}}$ be an extended-real-valued function with $\varphi(x^*)$ finite. The set

$$\hat{\partial}_\epsilon \varphi(x^*) := \{v \in \mathbb{X}^* \mid \liminf_{x \rightarrow x^*} \frac{\varphi(x) - \varphi(x^*) - \langle v, x - x^* \rangle}{\|x - x^*\|} \geq -\epsilon\}$$

is called the (Fréchet-like) ϵ -subdifferential of φ at x^* . When $\epsilon = 0$, the Fréchet-like ϵ -subdifferential reduces to the Fréchet subdifferential, denoted by $\hat{\partial} \varphi(x^*)$. The basic/limiting subdifferential of φ at x^* is defined by

$$\partial^L \varphi(x^*) := \text{Lim sup}_{x \xrightarrow{\varphi} x^*, \epsilon \downarrow 0} \hat{\partial}_\epsilon \varphi(x),$$

where $x \xrightarrow{\varphi} x^*$ means that $x \rightarrow x^*$ and $\varphi(x) \rightarrow \varphi(x^*)$. The singular subdifferential of $\varphi : \mathbb{X} \rightarrow \overline{\mathbb{R}}$ at x^* is defined by

$$\partial^\infty \varphi(x^*) := \text{Lim sup}_{x \xrightarrow{\varphi} x^*, \epsilon \downarrow 0, t \downarrow 0} t \hat{\partial}_\epsilon \varphi(x).$$

If \mathbb{X} is a Asplund space, then we have the following simpler form ([20, Theorems 2.34 and 2.38])

$$\partial^L \varphi(x^*) := \text{Lim sup}_{x \xrightarrow{\varphi} x^*} \hat{\partial} \varphi(x) \quad \text{and} \quad \partial^\infty \varphi(x^*) := \text{Lim sup}_{x \xrightarrow{\varphi} x^*, t \downarrow 0} t \hat{\partial} \varphi(x).$$

Next we introduce the approximate subdifferential developed by Ioffe [13, 14]. The lower Dini directional derivative of φ at x^* along the direction d is given by

$$D^- \varphi(x^*, d) := \liminf_{d' \rightarrow d, t \downarrow 0} \frac{\varphi(x^* + td') - \varphi(x^*)}{t},$$

and the Dini ϵ -subdifferential of φ at x^* is defined by

$$\partial_\epsilon^- \varphi(x^*) := \{v \in \mathbb{X}^* \mid \langle v, d \rangle \leq D^- \varphi(x^*, d) + \epsilon \|d\|, \quad \forall d \in \mathbb{X}\}.$$

As usual we set $\partial_\epsilon^- \varphi(x^*) := \emptyset$ if $|\varphi(x^*)| = \infty$. The approximate subdifferential of φ at x^* is given by

$$\partial^a \varphi(x^*) := \bigcap_{L \in \mathcal{L}} \overline{\text{Lim sup}_{x \xrightarrow{\varphi} x^*} \partial_0^-(\varphi + \delta(\cdot, L))(x)} = \bigcap_{L \in \mathcal{L}, \epsilon > 0} \overline{\text{Lim sup}_{x \xrightarrow{\varphi} x^*} \partial_\epsilon^-(\varphi + \delta(\cdot, L))(x)},$$

where \mathcal{L} is the collection of all finite dimensional subspaces of \mathbb{X} , $\delta(\cdot, L)$ is the indicator function of L , and $\overline{\text{Lim sup}}$ stands for the topological counterpart of the Painlevé-Kuratowski upper limit with sequences replaced by nets. The G-normal cone \mathcal{N}^g and its nucleus $\tilde{\mathcal{N}}^g$ to Ω at x^* are defined by

$$\mathcal{N}_\Omega^g(x^*) = cl^* \tilde{\mathcal{N}}_\Omega^g(x^*) \quad \text{and} \quad \tilde{\mathcal{N}}_\Omega^g(x^*) := \bigcup_{\lambda > 0} \lambda \partial^a \text{dist}_\Omega(x^*).$$

The A-normal cone to Ω at x^* is defined by

$$\mathcal{N}_\Omega^a(x^*) := \partial^a \delta(x^*, \Omega).$$

It follows from [14, Proposition 3.4], [20, Section 2.5.2, Page 238], and [14, Proposition 3.3] that

$$\mathcal{N}_\Omega^c(x^*) = cl^* \text{conv} \mathcal{N}_\Omega^g(x^*), \quad \mathcal{N}_\Omega^L(x^*) \subseteq \tilde{\mathcal{N}}_\Omega^g(x^*), \quad \text{and} \quad \mathcal{N}_\Omega^g(x^*) \subseteq \mathcal{N}_\Omega^a(x^*).$$

Clearly,

$$\mathcal{N}_\Omega^L(x^*) \subseteq \tilde{\mathcal{N}}_\Omega^g(x^*) \subseteq \mathcal{N}_\Omega^g(x^*) \subseteq \mathcal{N}_\Omega^c(x^*).$$

If Ω is convex, then $\mathcal{N}^L(x^*) = \tilde{\mathcal{N}}_\Omega^g(x^*) = \mathcal{N}^g(x^*) = \mathcal{N}^c(x^*)$ is the normal cone of Ω at x^* in the sense of convex analysis.

Now we introduce the Clarke subdifferential of locally Lipschitzian functions. In this paragraph we assume that φ is Lipschitzian near x^* . Recall that the Clarke's generalized derivative of φ at x^* along the direction d is defined by

$$\varphi^\circ(x^*, d) := \limsup_{x \rightarrow x^*, t \downarrow 0} \frac{\varphi(x + td) - \varphi(x)}{t}.$$

The Clarke subdifferential of φ at x^* is defined by

$$\partial^c \varphi(x^*) := \{v \in \mathbb{X}^* \mid \langle v, d \rangle \leq \varphi^\circ(x^*, d) \quad \forall d \in \mathbb{X}\}.$$

In the general Banach space setting, we have

$$cl^* \text{conv} \partial \varphi^L(x^*) \subseteq \partial^c \varphi(x^*),$$

where the inclusion relationship above holds with equality when \mathbb{X} is an Asplund space. It follows from [20, Section 2.5.2, Page 238] and [13, Proposition 3.3] that

$$\partial^L \varphi(x^*) \subseteq \partial^a \varphi(x^*) \quad \text{and} \quad \partial^a \varphi(x^*) \subseteq \partial^c \varphi(x^*).$$

If, in addition, φ is convex, then $\partial^a \varphi(x) = \partial^L \varphi(x) = \partial^c \varphi(x)$ is the same as the subdifferential of φ at x^* in the sense of convex analysis.

The following propositions provide a summary of some of the important properties of the approximate subdifferential, see [9, 13–15, 20]. For a set-valued map $S : \mathbb{X} \rightrightarrows \mathbb{X}^*$, we say S is closed if its graph is closed in the appropriate topology.

Proposition 2.1 *Let $f : \mathbb{X} \rightarrow \mathbb{R}$ be Lipschitzian near x^* with positive modulus L_f . Then the following results hold:*

- (i) [13, Proposition 3.3] and [5, Proposition 2.1.2] $cl^*conv\partial^a f(x^*) = \partial^c f(x^*) \subset L_f \mathbb{B}_{\mathbb{X}^*}$.
- (ii) [9, Theorem 1.1] *If x^* is a local minimizer of f on \mathbb{X} , then $0 \in \partial^a f(x^*)$.*

Proposition 2.2 [9, Theorem 1.4] *Let $f : \mathbb{X} \rightarrow \mathbb{R}$ be Lipschitzian near x^* . Then the set-valued map*

$$(\lambda, x) \rightarrow \partial^a(\lambda f)(x)$$

is closed at (λ^, x^*) , i.e.,*

$$\partial^a(\lambda^* f)(x^*) = \text{Lim sup}_{\lambda \rightarrow \lambda^*, x \xrightarrow{f} x^*} \partial^a(\lambda f)(x) \quad \forall \lambda^* \in \mathbb{R}.$$

For a locally Lipschitzian function in WCG Asplund spaces, at each point the limiting subdifferential set coincides with the approximate subdifferential set [20, Theorem 3.59]. Thus, the limiting subdifferential enjoys the namely, robust property as in Proposition 2.2 in the WCG Asplund setting. Note that even for a locally Lipschitzian function, the limiting subdifferential might not enjoy the robustness property in a non-WCG Banach space (see [20, Example 3.61]).

Proposition 2.3 [13, Proposition 2.3] *The A -normal cone mapping $\mathcal{N}_\Omega^a(\cdot) = \partial^a \delta(\cdot, \Omega)$ is closed, i.e.,*

$$\mathcal{N}_\Omega^a(x^*) = \text{Lim sup}_{x \xrightarrow{\Omega} x^*} \mathcal{N}_\Omega^a(x) \quad \forall x^* \in \Omega.$$

Proposition 2.4 (Calculus rules) (i) [13, Corollary 4.1.1] *Let $f, g : \mathbb{X} \rightarrow \overline{\mathbb{R}}$ be lower semicontinuous near x^* , finite at x^* , and at least one of them is Lipschitzian near $x^* \in \mathbb{X}$. Let α, β be nonnegative scalars. Then*

$$\partial^a(\alpha f + \beta g)(x^*) \subset \alpha \partial^a f(x^*) + \beta \partial^a g(x^*).$$

- (ii) [15, Theorem 2.5 and Remark (2)] *Let $\varphi : \mathbb{X} \rightarrow \mathbb{Y}$ be Lipschitzian near x^* and $f : \mathbb{Y} \rightarrow \mathbb{R}$ be Lipschitzian near $\varphi(x^*)$. Then $f \circ \varphi$ is locally Lipschitzian at x^* and*

$$\partial^a(f \circ \varphi)(x^*) \subset \cup_{\xi \in \partial^a f(\varphi(x^*))} \partial^a \langle \xi, \varphi \rangle(x^*).$$

- (iii) [9, Corollary 1.2] *Let $f_i : \mathbb{X} \rightarrow \mathbb{R}$ ($i = 1, \dots, n$) be Lipschitzian near x^* and $f(x) := \max\{f_i(x) \mid i = 1, \dots, n\}$. Then $f(x)$ is Lipschitzian near $x^* \in \mathbb{X}$ and*

$$\partial^a f(x^*) \subset \text{conv}\{\partial^a f_i(x^*) \mid i \in \mathcal{I}(x^*)\},$$

where $\mathcal{I}(x^) = \{i \mid f_i(x^*) = f(x^*)\}$ is the set of active indices.*

3 Enhanced Fritz John condition

For nonsmooth problem (MPGC), the classical Fritz John necessary optimality condition is generalized to one where the classical gradient is replaced by limiting subdifferential (see Mordukhovich [20]) and the Clarke subdifferential (see Clarke [5]), respectively. The following theorem strengthens the Fritz John condition Theorem 3.1(i)-(ii) through a stronger condition Theorem 3.1(iii), and hence their effectiveness has been significantly enhanced. Taking limits in (3.1) it is easy to see that condition (ii) is included in condition (iii). In order to emphasize the enhanced properties, however, we keep the redundant condition (ii) in Theorem 3.1. Note that the following result depends on the chosen basis $\mathcal{E} = \{e_1, \dots, e_m\}$ and since \mathbb{Y} is assumed to be finite dimensional, the limiting normal cone of Λ coincides with the nucleus of the G-normal cone of Λ at any point [20, Theorem 3.59(ii)].

Theorem 3.1 *Let x^* be a local minimizer of problem (MPGC). Then there exist a scalar $r \geq 0$ and a vector $\eta^* \in \mathbb{Y}$ not all zero, such that the following conditions hold:*

- (i) $0 \in r\partial^a f(x^*) + \sum_{i=1}^m \partial^a \langle \eta^*, e_i \rangle \langle F, e_i \rangle (x^*) + \tilde{\mathcal{N}}_\Omega^g(x^*);$
- (ii) $\eta^* \in \mathcal{N}_\Lambda^L(F(x^*));$
- (iii) *There exists a sequence $\{(x^k, y^k, \eta^k)\} \subseteq \Omega \times \Lambda \times \mathbb{Y}$ converging to $(x^*, F(x^*), \eta^*)$ such that for all k ,*

$$\begin{aligned} f(x^k) &< f(x^*), \\ \eta^k &\in \mathcal{N}_\Lambda^L(y^k), \end{aligned} \tag{3.1}$$

$$\langle \eta^*, e_i \rangle \neq 0 \implies \langle \eta^*, e_i \rangle \langle F(x^k) - y^k, e_i \rangle > 0. \tag{3.2}$$

Proof. Without loss of generality we may assume that x^* is a global minimizer of problem (MPGC). First of all, we observe that if x^* is a local minimizer of the problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in \Omega, \end{aligned} \tag{3.3}$$

then by the Clarke exact penalty principle [5, Proposition 2.4.3], then there exists $\kappa > 0$ such that x^* is a local minimizer for

$$\min \quad f(x) + \kappa \text{dist}(x, \Omega).$$

Then by Proposition 2.1(ii) and Proposition 2.4(i), we have

$$0 \in \partial^a f(x^*) + \tilde{\mathcal{N}}_\Omega^g(x^*).$$

If $\eta^* = 0$, then the existence of the desired sequence is trivial and hence the proof is complete by letting $r = 1$ and $\eta^* = 0$. In the following, we assume that x^* is not a local minimizer of problem (3.3). By introducing a slack variable $y \in \mathbb{Y}$ for the geometric constraint $F(x) \in \Lambda$, we first reformulate problem (MPGC) as follows:

$$\begin{aligned} \text{(MPGC)'} \quad & \min && f(x) \\ & \text{s.t.} && F(x) - y = 0, \\ & && x \in \Omega, y \in \Lambda. \end{aligned}$$

Then (x^*, y^*) with $y^* = F(x^*)$ is a global minimizer for problem (MPGC)'. For each $k = 1, 2, \dots$, we consider the function $F^k : \mathbb{X} \times \mathbb{Y} \rightarrow \mathbb{R}$ defined by

$$F^k(x, y) := \max\left\{f(x) - f(x^*) + \frac{1}{2k}, |\langle F(x) - y, e_1 \rangle|, \dots, |\langle F(x) - y, e_m \rangle|\right\}.$$

Since (x^*, y^*) is a global minimizer of (MPGC)', we have

$$F^k(x, y) > 0, \quad \forall (x, y) \in \Omega \times \Lambda,$$

which, together with $F^k(x^*, y^*) = \frac{1}{2k}$, implies

$$F^k(x^*, y^*) < \inf_{(x, y) \in \Omega \times \Lambda} F^k(x, y) + \frac{1}{k}. \quad (3.4)$$

Since x^* is not a local minimizer of problem (3.3), then there exists a sequence $\{\tilde{x}^k\}$ with $\tilde{x}^k \in \Omega$ such that $\tilde{x}^k \rightarrow x^*$ and

$$f(\tilde{x}^k) < f(x^*) \quad \forall k.$$

Then $\|F(\tilde{x}^k) - y^*\| = \|F(\tilde{x}^k) - F(x^*)\| \rightarrow 0$. Thus we can choose the sequence $\{\tilde{x}^k\}$ such that

$$F(\tilde{x}^k, y^*) < \frac{1}{2k} = F^k(x^*, y^*). \quad (3.5)$$

This and (3.4) imply

$$F(\tilde{x}^k, y^*) < \inf_{(x, y) \in \Omega \times \Lambda} F^k(x, y) + \frac{1}{k}.$$

Clearly, F^k is Lipschitzian near (x^*, y^*) and hence, by the Ekeland's variational principle (see, e.g., [20, Theorem 2.26], [30, Corollary 8.2.6]), there exists $(x^k, y^k) \in \Omega \times \Lambda$ such that

$$\begin{cases} \|(x^k, y^k) - (\tilde{x}^k, y^*)\| \leq \frac{1}{\sqrt{k}}, \\ F^k(x^k, y^k) \leq F^k(\tilde{x}^k, y^*), \\ F^k(x^k, y^k) \leq F^k(x, y) + \frac{1}{\sqrt{k}}\|(x, y) - (x^k, y^k)\|, \quad \forall (x, y) \in \Omega \times \Lambda. \end{cases} \quad (3.6)$$

It follows that $(x^k, y^k) \rightarrow (x^*, y^*)$ as $k \rightarrow \infty$ and for each k , $(x, y) = (x^k, y^k)$ is a global minimizer of the problem

$$\begin{aligned} \min \quad & \tilde{F}^k(x, y) := F^k(x, y) + \frac{1}{\sqrt{k}} \|(x, y) - (x^k, y^k)\| \\ \text{s.t.} \quad & (x, y) \in \Omega \times \Lambda. \end{aligned}$$

Then by the Clarke exact penalty principle [5, Proposition 2.4.3], there exists $\kappa \geq 0$ such that $(x, y) = (x^k, y^k)$ is a global minimizer of the problem

$$\begin{aligned} \min \quad & \tilde{F}^k(x, y) + \kappa \text{dist}_{\Omega \times \Lambda}(x, y) \\ \text{s.t.} \quad & (x, y) \in \mathbb{X} \times \mathbb{Y}. \end{aligned}$$

Thus, we have by the necessary optimality condition (Proposition 2.1(ii)) and calculus rule (Proposition 2.4(i)) that

$$0 \in \partial^a \tilde{F}^k(x^k, y^k) + \kappa \partial^a \text{dist}_{\Omega}(x^k) \times \kappa \partial^a \text{dist}_{\Lambda}(y^k). \quad (3.7)$$

Applying the calculus rules (Proposition 2.4(iii)), there exist nonnegative scalars $\{r^k, \hat{\eta}_1^k, \dots, \hat{\eta}_m^k\}$ such that for each k ,

$$r^k + \sum_{i=1}^m \hat{\eta}_i^k = 1 \quad (3.8)$$

and

$$0 \in r^k \begin{pmatrix} \partial^a f(x^k) \\ 0 \end{pmatrix} + \sum_{i=1}^m \hat{\eta}_i^k \partial^a |\psi_i|(x^k, y^k) + \frac{1}{\sqrt{k}} \begin{pmatrix} \mathbb{B}_{\mathbb{X}^*} \\ \mathbb{B}_{\mathbb{Y}} \end{pmatrix} + \kappa \begin{pmatrix} \partial^a \text{dist}_{\Omega}(x^k) \\ \partial^a \text{dist}_{\Lambda}(y^k) \end{pmatrix}, \quad (3.9)$$

where $\psi_i(x, y) := \langle F(x) - y, e_i \rangle$. Since the active indices only count in the maximum rule and $F^k(x^k, y^k) > 0$, we have for each k , $\psi_i(x^k, y^k) = \langle F(x^k) - y^k, e_i \rangle = 0$ implies $\hat{\eta}_i^k = 0$. Define

$$\tilde{\eta}_i^k := (\text{sign} \langle F(x^k) - y^k, e_i \rangle) \hat{\eta}_i^k,$$

where $\text{sign} 0 = 0$. We then obtain by the chain rule (Proposition 2.4(ii)) that

$$\hat{\eta}_i^k \partial^a |\psi_i|(x^k, y^k) = \begin{pmatrix} \partial^a \langle F, \tilde{\eta}_i^k e_i \rangle(x^k) \\ -\tilde{\eta}_i^k e_i \end{pmatrix}.$$

This and (3.9) imply that

$$0 \in r^k \begin{pmatrix} \partial^a f(x^k) \\ 0 \end{pmatrix} + \sum_{i=1}^m \begin{pmatrix} \partial^a \langle F, \tilde{\eta}_i^k e_i \rangle(x^k) \\ -\tilde{\eta}_i^k e_i \end{pmatrix} + \frac{1}{\sqrt{k}} \begin{pmatrix} \mathbb{B}_{\mathbb{X}^*} \\ \mathbb{B}_{\mathbb{Y}} \end{pmatrix} + \kappa \begin{pmatrix} \partial^a \text{dist}_{\Omega}(x^k) \\ \partial^a \text{dist}_{\Lambda}(y^k) \end{pmatrix},$$

that is,

$$\begin{cases} 0 \in r^k \partial^a f(x^k) + \sum_{i=1}^m \partial^a \tilde{\eta}_i^k \langle F, e_i \rangle(x^k) + \frac{1}{\sqrt{k}} \mathbb{B}_{\mathbb{X}^*} + \kappa \partial^a \text{dist}_\Omega(x^k), \\ \tilde{\eta}^k \in \frac{1}{\sqrt{k}} \mathbb{B}_{\mathbb{Y}} + \kappa \partial^a \text{dist}_\Lambda(y^k), \end{cases} \quad (3.10)$$

where $\tilde{\eta}^k := \sum_{i=1}^m \tilde{\eta}_i^k e_i$.

Since by construction we have $r^k + \sum_{i=1}^m |\tilde{\eta}_i^k| = 1$, the sequence $\{(r^k, \tilde{\eta}_1^k, \dots, \tilde{\eta}_m^k)\}$ is bounded and must contain a subsequence that converges to some limit $(r, \bar{\eta}_1, \dots, \bar{\eta}_m)$, where $r \geq 0$ and $(r, \bar{\eta}_1, \dots, \bar{\eta}_m) \neq 0$. By virtue of the closedness of the subdifferential (Proposition 2.2), it follows from (3.10) that

$$\begin{cases} 0 \in r \partial^a f(x^*) + \sum_{i=1}^m \partial^a \langle \eta^*, e_i \rangle \langle F, e_i \rangle(x^*) + \kappa \partial^a \text{dist}_\Omega(x^*), \\ \eta^* \in \kappa \partial^a \text{dist}_\Lambda(y^*), \end{cases} \quad (3.11)$$

where $\eta^* = \sum_{i=1}^m \bar{\eta}_i e_i$. Thus,

$$\begin{cases} 0 \in r \partial^a f(x^*) + \sum_{i=1}^m \partial^a \langle \eta^*, e_i \rangle \langle F, e_i \rangle(x^*) + \tilde{\mathcal{N}}_\Omega^g(x^*), \\ \eta^* \in \tilde{\mathcal{N}}_\Lambda^g(F(x^*)). \end{cases}$$

To show that condition (iii) is satisfied, assume that $\eta^* \neq 0$ (otherwise the existence of the desired sequence is trivial). Since

$$\tilde{\eta}^k \in \frac{1}{\sqrt{k}} \mathbb{B}_{\mathbb{Y}} + \kappa \partial^g \text{dist}_\Lambda(y^k) \subseteq \frac{1}{\sqrt{k}} \mathbb{B}_{\mathbb{Y}} + \tilde{\mathcal{N}}_\Lambda^g(y^k),$$

there exists $\rho^k \in \mathbb{B}_{\mathbb{Y}}$ such that

$$\eta^k := \tilde{\eta}^k + \frac{1}{\sqrt{k}} \rho^k \in \tilde{\mathcal{N}}_\Lambda^g(y^k).$$

Since $\tilde{\eta}^k \rightarrow \eta^*$ as $k \rightarrow \infty$, it is easy to see that $\eta^k \rightarrow \eta^*$. Since $\tilde{\eta}_i^k \rightarrow \bar{\eta}_i = \langle \eta^*, e_i \rangle \neq 0$ for each $i \in \mathcal{I}$, $\tilde{\eta}_i^k$ has the same sign as $\langle \eta^*, e_i \rangle$ for sufficiently large k . Hence we must have $\langle \eta^*, e_i \rangle \tilde{\eta}_i^k > 0$ for all $i \in \mathcal{I}$ and sufficiently large k . By the definition, $\tilde{\eta}_i^k$ have the same sign as $\langle F(x^k) - y^k, e_i \rangle$, therefore we must have $\langle \eta^*, e_i \rangle \langle F(x^k) - y^k, e_i \rangle > 0$ for all $i \in \mathcal{I}$ and sufficiently large k . Moreover, it follows from the definition of F^k and (3.5)–(3.6) that

$$\begin{aligned} f(x^k) - f(x^*) + \frac{1}{2k} &\leq F^k(x^k, y^k) \\ &\leq F^k(\tilde{x}^k, y^*) \\ &< \frac{1}{2k} \end{aligned}$$

and hence $f(x^k) < f(x^*)$. The proof is complete by noting that the limiting normal cone of Λ coincides with the nucleus of the G-normal cone of Λ at any point in the finite dimensional setting [20, Theorem 3.59(ii)]. \blacksquare

Since for any function φ and set S , it must hold that (see, e.g., [14, Proposition 3.4])

$$\partial^g \varphi(x) \subseteq \partial^c \varphi(x), \quad \tilde{\mathcal{N}}_S^g(x) \subseteq \mathcal{N}_S^c(x),$$

the following result holds immediately.

Corollary 3.1 *Let x^* be a local minimizer of problem (MPGC). Then there exist a scalar $r \geq 0$ and a vector $\eta^* \in \mathbb{Y}$ not all zero, such that the following conditions hold:*

- (i) $0 \in r\partial^c f(x^*) + \sum_{i=1}^m \partial^c \langle \eta^*, e_i \rangle \langle F, e_i \rangle(x^*) + \mathcal{N}_\Omega^c(x^*);$
- (ii) $\eta^* \in \mathcal{N}_\Lambda^L(F(x^*));$
- (iii) *There exists a sequence $\{(x^k, y^k, \eta^k)\} \subseteq \Omega \times \Lambda \times \mathbb{Y}$ converging to $(x^*, F(x^*), \eta^*)$ such that for all k ,*

$$\begin{aligned} f(x^k) &< f(x^*), \\ \eta^k &\in \mathcal{N}_\Lambda^L(y^k), \\ \langle \eta^*, e_i \rangle \neq 0 &\implies \langle \eta^*, e_i \rangle \langle F(x^k) - y^k, e_i \rangle > 0. \end{aligned}$$

Since in the WCG Asplund space setting, the limiting subdifferential and limiting normal cone coincide with the approximate subdifferential and the nucleus of the G-normal cone respectively [20, Theorem 3.59], we have the following result immediately.

Corollary 3.2 *Assume that \mathbb{X} is a WCG Asplund space. Let x^* be a local minimizer of problem (MPGC). Then there exist a scalar $r \geq 0$ and a vector $\eta^* \in \mathbb{Y}$ not all zero, such that the following conditions hold:*

- (i) $0 \in r\partial^L f(x^*) + \sum_{i=1}^m \partial^L \langle \eta^*, e_i \rangle \langle F, e_i \rangle(x^*) + \mathcal{N}_\Omega^L(x^*);$
- (ii) $\eta^* \in \mathcal{N}_\Lambda^L(F(x^*));$
- (iii) *There exists a sequence $\{(x^k, y^k, \eta^k)\} \subseteq \Omega \times \Lambda \times \mathbb{Y}$ converging to $(x^*, F(x^*), \eta^*)$ such that for all k ,*

$$\begin{aligned} f(x^k) &< f(x^*), \\ \eta^k &\in \mathcal{N}_\Lambda^L(y^k), \\ \langle \eta^*, e_i \rangle \neq 0 &\implies \langle \eta^*, e_i \rangle \langle F(x^k) - y^k, e_i \rangle > 0. \end{aligned}$$

We now specialize Theorem 3.1 to problem (NLP) where $f : \mathbb{X} \rightarrow \mathbb{R}, h : \mathbb{X} \rightarrow \mathbb{R}^p, g : \mathbb{X} \rightarrow \mathbb{R}^q$ are Lipschitzian near the optimal solution and Ω is a nonempty closed subset of \mathbb{X} . Let

$$F(x) := (h(x), g(x)) \quad \text{and} \quad \Lambda := \{0\}^p \times \mathbb{R}_-^q, \quad (3.12)$$

By virtue of Theorem 3.1, we are now able to establish the enhanced Fritz John condition for the nonsmooth NLP in a Banach space, which improves [33, Theorem 1]. Note that the set Λ in (3.12) is a convex cone.

Corollary 3.3 *Let x^* be a local minimizer of problem (NLP). Then there exists $r \geq 0, \lambda^* \in \mathbb{R}^p, \mu^* \in \mathbb{R}^q$ not all zero, such that*

- (a) $0 \in r\partial^a f(x^*) + \sum_{i=1}^p \partial^a(\lambda_i^* h_i)(x^*) + \sum_{j=1}^q \mu_j^* \partial^a g_j(x^*) + \tilde{\mathcal{N}}_\Omega^g(x^*);$
- (b) $0 \leq -g(x^*) \perp \mu^* \geq 0;$
- (c) *There exists a sequence $\{x^k\} \subset \Omega$ converging to x^* such that for all $k, f(x^k) < f(x^*)$ and*

$$\lambda_i^* \neq 0 \implies \lambda_i^* h_i(x^k) > 0, \quad \mu_j^* > 0 \implies \mu_j^* g_j(x^k) > 0.$$

Proof. Letting F and Λ be defined as in (3.12), it is not hard from Theorem 3.1 and the explicit expression for the normal cone $\mathcal{N}_\Lambda^L(F(x^*))$ to see that there exists $r \geq 0, \lambda^* \in \mathbb{R}^p, \mu^* \in \mathbb{R}^q$ not all zero, such that conditions (a)–(b) hold and there exists a sequence $\{(x^k, \hat{y}^k, \tilde{y}^k, \lambda^k, \mu^k)\} \in \Omega \times \{0\}^p \times \mathbb{R}_-^q \times \mathbb{R}^p \times \mathbb{R}^q$ converging to $(x^*, h(x^*), g(x^*), \lambda^*, \eta^*)$ such that for all $k, f(x^k) < f(x^*),$

$$(\lambda^k, \mu^k) \in \mathcal{N}_{\{0\}^p \times \mathbb{R}_-^q}(\hat{y}^k, \tilde{y}^k) \tag{3.13}$$

and

$$\langle \eta^*, e_i \rangle \neq 0 \implies \langle \eta^*, e_i \rangle \langle F(x^k) - y^k, e_i \rangle > 0, \tag{3.14}$$

where $\eta^* := (\lambda^*, \mu^*),$ and $F(x^k) := (h(x^k), g(x^k)).$ Since $\hat{y}^k = 0,$ it is easy to see from (3.14) that

$$\lambda_i^* \neq 0 \implies \lambda_i^* h_i(x^k) > 0.$$

If $\mu_j^* > 0,$ then it follows from (3.14) that $\mu_j^* g_j(x^k) > \mu_j^* \tilde{y}_j^k.$ We next show that there exists a subsequence $\{\tilde{y}_j^{k_i}\}_{i \in \mathbb{N}}$ such that $\tilde{y}_j^{k_i} = 0, \forall i \in \mathbb{N}.$ Assume to the contrary that $\tilde{y}_j^k < 0$ for each sufficiently large k and then it follows from (3.13) that $\mu_j^k = 0,$ which implies that $\mu_j^* = 0$ by taking a limit as $k \rightarrow \infty.$ This contradicts assumption $\mu_j^* > 0$ and hence we have

$$\mu_j^* > 0 \implies \mu_j^* g_j(x^{k_i}) > 0, \quad \forall i \in \mathbb{N}.$$

Therefore, condition (c) also holds by choosing and resetting this subsequence. The proof is complete. \blacksquare

Our next task is to specialize our result to the nonlinear semidefinite program:

$$\begin{aligned} \text{(NLSDP)} \quad & \min_{x \in \mathbb{X}} && f(x) \\ & \text{s.t.} && H(x) \in \mathcal{S}_-^l, \end{aligned}$$

where $f : \mathbb{X} \rightarrow \mathbb{R}$ and $H : \mathbb{X} \rightarrow \mathcal{S}^l$, \mathcal{S}^l is the linear space of all $l \times l$ real symmetric matrices equipped with the usual Frobenius inner product $\langle \cdot, \cdot \rangle$, and \mathcal{S}_-^l is the cone of all $l \times l$ negative semidefinite matrices in \mathcal{S}^l . Note that for simplicity, we omit the usual equality and inequality constraints since they can be handled as in the usual nonlinear program. For $A \in \mathcal{S}^l$, we denote by $\lambda(A) \in \mathbb{R}^l$ the vector of its eigenvalues ordered in a decreasing order as follows:

$$\lambda_1(A) \geq \cdots \geq \lambda_l(A).$$

Clearly, (NLSDP) is equivalent to the problem

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & \lambda_1(H(x)) \leq 0. \end{aligned} \tag{3.15}$$

For $A \in \mathcal{S}^l$, the notation $\text{diag}(\lambda(A)) \in \mathcal{S}^l$ is used for the diagonal matrix with the vector $\lambda(A)$ on the main diagonal. It is known that any $A \in \mathcal{S}^l$ admits an eigenvalue decomposition as follows:

$$A = U \text{diag}(\lambda(A)) U^T$$

with a square orthogonal matrix $U = U(A)$ such that $U^T U = I$ whose columns are eigenvectors of A . Let $u_i(A)$ be the i th column of matrix $U(A)$. Note that since λ_1 is convex (see, e.g., [18, Proposition 1.1]), the approximate subdifferential coincides with its Clarke subdifferential.

Lemma 3.1 [16, 24] *The approximate subdifferential and Clarke subdifferential of $\lambda_1(A) : \mathcal{S}^l \rightarrow \mathbb{R}$ is given by*

$$\begin{aligned} \partial^a \lambda_1(A) = \partial^c \lambda_1(A) &= \text{conv}\{u_i(A)u_i(A)^T \mid i = 1, \dots, d(A)\} \\ &= \left\{ \sum_{i=1}^{d(A)} \tau_i u_i(A)u_i(A)^T \mid \sum_{i=1}^{d(A)} \tau_i = 1, \tau_i \geq 0 \ i = 1, \dots, d(A) \right\}, \end{aligned}$$

where $d(A)$ is the multiplicity of the largest eigenvalue of the matrix A .

We get the following results immediately by applying Corollary 3.3 to the (NLSDP) problem (3.15). Recall that $\mathcal{S}_+^l = -\mathcal{S}_-^l$.

Corollary 3.4 *Assume that x^* is a local minimizer of problem (NLSDP). Then there exist $r \geq 0$ and $\Gamma^* \in \mathcal{S}_+^l$, which are not all zero, such that*

- (a) $0 \in r \partial^a f(x^*) + \partial^a \langle \Gamma^*, H \rangle(x^*)$;
- (b) $\Gamma^* \in \mathcal{S}_+^l$, $\langle \Gamma^*, H(x^*) \rangle = 0$;

(c) *there exists a sequence $\{x^k\}$ converging to x^* such that for all k , $f(x^k) < f(x^*)$ and*

$$\lambda_1(\Gamma^*) > 0 \implies \lambda_1(H(x^k)) > 0.$$

Proof. Since x^* is a local minimizer of problem (3.15), it follows from Corollary 3.3 that there exist $\{r, \mu^*\}$, which are not all zero, such that

- (i) $0 \in r\partial^a f(x^*) + \mu^*\partial^a(\lambda_1 \circ H)(x^*)$;
- (ii) $r \geq 0, 0 \leq -\lambda_1(H(x^*)) \perp \mu^* \geq 0$;
- (iii) There exist a sequence $\{x^k\} \subseteq \mathbb{X}$ converging to x^* such that for all k , $f(x^k) < f(x^*)$ and

$$\mu^* > 0 \implies \mu^*(\lambda_1(H(x^k))) > 0.$$

It follows from Proposition 2.4(ii), Lemma 3.1, and (i) above that there exists

$$\Gamma^* = \mu^* \sum_{i=1}^{d(H(x^*))} \tau_i^* u_i(H(x^*)) u_i(H(x^*))^T \in \mu^* \text{conv}\{u_i(H(x^*)) u_i(H(x^*))^T \mid i = 1, \dots, d(H(x^*))\}$$

such that

$$0 \in r\partial^a f(x^*) + \partial^a \langle \Gamma^*, H \rangle(x^*), \quad (3.16)$$

where $d(H(x^*))$ is the multiplicity of the largest eigenvalue of the matrix $H(x^*)$.

It is easy to see that $\Gamma^* \in \mathcal{S}_+^l$ and, from the definition of $\partial^c \lambda_1(H(x^*))$ and (ii)–(iii) above that

$$\begin{aligned} \langle \Gamma^*, H(x^*) \rangle &= \left\langle \mu^* \sum_{i=1}^{d(H(x^*))} \tau_i^* u_i(H(x^*)) u_i(H(x^*))^T, H(x^*) \right\rangle \\ &= \mu^* \sum_{i=1}^{d(H(x^*))} \tau_i^* \left\langle u_i(H(x^*)) u_i(H(x^*))^T, H(x^*) \right\rangle \\ &= \mu^* \sum_{i=1}^{d(H(x^*))} \tau_i^* \left\langle 1, u_i(H(x^*))^T H(x^*) u_i(H(x^*)) \right\rangle \\ &= \mu^* \lambda_1(H(x^*)) \sum_{i=1}^{d(H(x^*))} \left(\tau_i^* u_i(H(x^*))^T u_i(H(x^*)) \right) \\ &= \mu^* \lambda_1(H(x^*)) \\ &= 0. \end{aligned}$$

Then, conditions (a) and (b) in this corollary hold. We next show condition (c). From the definition of Γ^* , we have that $\mu^* > 0$ once $\lambda_1(\Gamma^*) > 0$ holds. Then we have from (iii) above the desired result. The proof is complete. \blacksquare

4 Enhanced KKT condition and weaker constraint qualification

Based on the enhanced Fritz John condition for problem (MPGC) in the previous section, we define the following enhanced KKT condition for problem (MPGC). We denote by $\mathcal{N}_\Lambda^e(F(x^*))$ the set of elements in the normal cone $\eta^* \in \mathcal{N}_\Lambda^L(F(x^*))$ such that there exists a sequence $\{(x^k, y^k, \eta^k)\} \subset \Omega \times \Lambda \times \mathbb{Y}$ converging to $(x^*, F(x^*), \eta^*)$ such that for all k ,

$$\begin{aligned} \eta^k &\in \mathcal{N}_\Lambda^L(y^k) \\ \langle \eta^*, e_i \rangle \neq 0 &\implies \langle \eta^*, e_i \rangle \langle F(x^k) - y^k, e_i \rangle > 0. \end{aligned}$$

Definition 4.1 (Enhanced KKT point) Let x^* be a feasible point of the problem (MPGC).

(a) We say that x^* is an *enhanced KKT point* if there exists $\eta^* \in \mathcal{N}_\Lambda^L(F(x^*))$ such that

- (i) $0 \in \partial^a f(x^*) + \sum_{i=1}^m \partial^a \langle \eta^*, e_i \rangle \langle F, e_i \rangle(x^*) + \tilde{\mathcal{N}}_\Omega^g(x^*)$,
- (ii) There exists a sequence $\{x^k, y^k, \eta^k\} \subseteq \Omega \times \Lambda \times \mathbb{Y}$ converging to $(x^*, F(x^*), \eta^*)$ such that for all k ,

$$\begin{aligned} f(x^k) &< f(x^*), \\ \eta^k &\in \mathcal{N}_\Lambda^L(y^k), \\ \langle \eta^*, e_i \rangle \neq 0 &\implies \langle \eta^*, e_i \rangle \langle F(x^k) - y^k, e_i \rangle > 0. \end{aligned}$$

(b) We say that x^* is a *weaker enhanced KKT point* if there exists $\eta^* \in \mathcal{N}_\Lambda^e(F(x^*))$ such that

$$0 \in \partial^a f(x^*) + \sum_{i=1}^m \partial^a \langle \eta^*, e_i \rangle \langle F, e_i \rangle(x^*) + \tilde{\mathcal{N}}_\Omega^g(x^*).$$

Definition 4.2 Let $x^* \in \mathcal{X}$.

(a) x^* is said to satisfy the *no nonzero abnormal multiplier constraint qualification* (NNAMCQ) if there is no nonzero vector $\eta^* \in \mathbb{Y}$ such that

$$0 \in \sum_{i=1}^m \partial^a \langle \eta^*, e_i \rangle \langle F, e_i \rangle(x^*) + \tilde{\mathcal{N}}_\Omega^g(x^*), \quad \eta^* \in \mathcal{N}_\Lambda^L(F(x^*)); \quad (4.1)$$

- (b) x^* is said to be *pseudonormal* for \mathcal{X} if there is no nonzero vector $\eta^* \in \mathbb{Y}$ such that (4.1) holds and there exists a sequence $\{(x^k, y^k, \eta^k)\} \subset \Omega \times \Lambda \times \mathbb{Y}$ converging to $(x^*, F(x^*), \eta^*)$ such that for each k ,

$$\eta^k \in \mathcal{N}_\Lambda^L(y^k) \quad \langle \eta^*, F(x^k) - y^k \rangle > 0;$$

- (c) x^* is said to be *quasinormal* for \mathcal{X} if there is no nonzero vector $\eta^* \in \mathbb{Y}$ such that

$$0 \in \sum_{i=1}^m \partial^a \langle \eta^*, e_i \rangle \langle F, e_i \rangle(x^*) + \tilde{\mathcal{N}}_\Omega^g(x^*), \quad \eta^* \in \mathcal{N}_\Lambda^e(F(x^*));$$

- (d) x^* is said to satisfy the *enhanced Guignard constraint qualification* (EGCQ) if F is Fréchet differentiable at x^* and

$$\widehat{\mathcal{N}}_{\mathcal{X}}(x^*) \subseteq \nabla F(x^*)^* \mathcal{N}_\Lambda^e(F(x^*)) + \tilde{\mathcal{N}}_\Omega^g(x^*).$$

The relationships of the first three constraint qualifications are obvious:

$$\text{NNAMCQ} \implies \text{pseudonormality} \implies \text{quasinormality}.$$

The enhanced KKT condition under the quasinormality follows immediately from Theorem 3.1 and the definition of the quasinormality.

Theorem 4.1 *Let x^* be a local minimizer of problem (MPGC). Suppose that x^* is quasinormal. Then x^* is an enhanced KKT point.*

We now make some comments on the EGCQ. It is well-known that $\mathcal{T}_{\mathcal{X}}(x^*)^o = \widehat{\mathcal{N}}_{\mathcal{X}}(x^*)$ in a finite dimensional space. We next consider the case of standard nonlinear constraints, i.e., $\mathcal{X} := \{x \in \Omega \mid F(x) \in \Lambda\}$ with $\Omega = \mathbb{R}^n$, $F(x)$ and Λ are defined as in (3.12). In this case,

$$\mathcal{L}_{\mathcal{X}}(x^*)^o = \nabla F(x^*)^* \mathcal{N}_\Lambda(F(x^*)),$$

where

$$\mathcal{L}_{\mathcal{X}}(x^*) := \{d \mid \nabla F(x^*)d \in \mathcal{T}_\Lambda(F(x^*))\}$$

is the linearized cone of \mathcal{X} at x^* . Since the inclusion $\mathcal{N}_\Lambda^e(F(x^*)) \subset \mathcal{N}_\Lambda(F(x^*))$ may hold strictly, in the case of standard nonlinear constraints, the EGCQ is stronger than the condition $\mathcal{T}_{\mathcal{X}}(x^*)^o \subseteq \mathcal{L}_{\mathcal{X}}(x^*)^o$, which is the so-called Guignard constraint qualification (GCQ).

Next we show that the quasinormality implies the EGCQ in the case where \mathbb{X} admits a Fréchet smooth renorm [20, Page 35]. To this end, we first show that the EGCQ is the weakest constraint qualification for weaker enhanced KKT point when the objective is Fréchet smooth in a Banach space.

Lemma 4.1 [20, Theorem 1.30] *Assume that \mathbb{X} admits a Fréchet smooth renorm. Then for every $d \in \widehat{\mathcal{N}}_S(x^*)$, there is a concave Fréchet smooth function $\varphi : \mathbb{X} \rightarrow \mathbb{R}$ that achieves its global maximum relative to S uniquely at x^* and such that $\nabla\varphi(x^*) = d$.*

Theorem 4.2 *Suppose that $x^* \in \mathcal{X}$ is a local minimizer for the optimization problem $\min_{x \in \mathcal{X}} \theta(x)$, where θ is Fréchet differentiable function, and*

$$\widehat{\mathcal{N}}_{\mathcal{X}}(x^*) \subseteq \nabla F(x^*)^* \mathcal{N}_{\Lambda}^e(F(x^*)) + \widetilde{\mathcal{N}}_{\Omega}^g(x^*). \quad (4.2)$$

Then, x^ must be a weaker enhanced KKT point of $\min_{x \in \mathcal{X}} \theta(x)$. Conversely, assume that \mathbb{X} admits a Fréchet smooth renorm and $x^* \in \mathcal{X}$ is a weaker enhanced KKT point of $\min_{x \in \mathcal{X}} \theta(x)$ for any convex Fréchet smooth function θ at x^* with x^* being a local minimizer, then (4.2) holds.*

Proof. Let x^* be locally optimal for problem $\min_{x \in \mathcal{X}} \theta(x)$. Then it follows from [20, Proposition 5.1] that $-\nabla\theta(x^*) \in \widehat{\mathcal{N}}_{\mathcal{X}}(x^*)$. Thus if $\widehat{\mathcal{N}}_{\mathcal{X}}(x^*) \subseteq \nabla F(x^*)^* \mathcal{N}_{\Lambda}^e(F(x^*)) + \widetilde{\mathcal{N}}_{\Omega}^g(x^*)$, then $-\nabla\theta(x^*) \in \nabla F(x^*)^* \mathcal{N}_{\Lambda}^e(F(x^*)) + \widetilde{\mathcal{N}}_{\Omega}^g(x^*)$ and hence x^* is an enhanced KKT point of $\min_{x \in \mathcal{X}} \theta(x)$.

Conversely suppose that if $x^* \in \mathcal{X}$ is a local minimizer for an optimization problem $\min_{x \in \mathcal{X}} \theta(x)$ with convex Fréchet smooth objective functions, then x^* must be a weaker enhanced KKT point of the problem. Let $d \in \widehat{\mathcal{N}}_{\mathcal{X}}(x^*)$, by Lemma 4.1, there exists a convex Fréchet smooth function φ such that $-\nabla\varphi(x^*) = d$ and $\operatorname{argmin}_{x \in \mathcal{X}} \varphi(x) = \{x^*\}$. It follows that x^* is an enhanced KKT point of $\min_{x \in \mathcal{X}} \varphi(x)$, i.e.,

$$-\nabla\varphi(x^*) \in \nabla F(x^*)^* \mathcal{N}_{\Lambda}^e(F(x^*)) + \mathcal{N}_{\Omega}(x^*),$$

that is, $d = -\nabla\varphi(x^*) \in \nabla F(x^*)^* \mathcal{N}_{\Lambda}^e(F(x^*)) + \widetilde{\mathcal{N}}_{\Omega}^g(x^*)$. Therefore, by the arbitrariness of $d \in \widehat{\mathcal{N}}_{\mathcal{X}}(x^*)$, we have $\widehat{\mathcal{N}}_{\mathcal{X}}(x^*) \subseteq \nabla F(x^*)^* \mathcal{N}_{\Lambda}^e(F(x^*)) + \widetilde{\mathcal{N}}_{\Omega}^g(x^*)$. \blacksquare

The following result follows from Theorem 4.2.

Corollary 4.1 *Assume that $\{f, F\}$ are Fréchet differentiable at x^* . If x^* is a local minimizer of problem (MPGC) and the EGCQ holds at x^* , then x^* is a weaker enhanced KKT point.*

Corollary 4.2 *Assume that \mathbb{X} admits a Fréchet smooth renorm and F is Fréchet differentiable at x^* . Then the quasinormality implies the EGCQ.*

Proof. It follows from Theorem 4.1 that for any locally Lipschitzian objective function f , if a local minimizer satisfies the quasinormality, then it is an enhanced KKT point. Since a convex Fréchet smooth function is locally Lipschitzian ([4, Proposition 2.107]), it follows from Theorem 4.2 that the EGCQ holds at this point. \blacksquare

5 Error bound and exact penalty

In this section, we prove that a local error bound exists under the quasinormality in the general Banach space. Our results are new even for the finite dimensional space.

For nonsmooth finite dimensional (NLP) problem, the existence of a local error bound has been proved under the pseudonormality and under the quasinormality with extra regularity conditions on the constraint functions in [33], where [31, Theorem 3.1] plays a significant role. In this section, we show that the quasinormality alone implies the existence of a local error bound without imposing any regularity conditions. We first establish the following estimate, which will lead to the possibility of applying [31, Theorem 3.1].

Lemma 5.1 *Let x^* be feasible for problem (MPGC) and*

$$\Phi(x, y) := \max_{1 \leq i \leq m} \{|\psi_i(x, y)|\} \quad \text{with} \quad \psi_i(x, y) := \langle F(x) - y, e_i \rangle.$$

If x^ is quasinormal, then there exist $\delta > 0$ and $c > 0$ such that for all $(\xi, v) \in \partial^a(\Phi + \text{dist}_{\Omega \times \Lambda})(x, y)$ with $(x, y) \in \mathcal{B}_\delta(x^*, F(x^*)) \cap (\Omega \times \Lambda)$ and $x \notin \mathcal{X}$,*

$$\|(\xi, v)\| \geq \frac{1}{c}.$$

Proof. Suppose to the contrary that there exists a sequence $\{(x^k, y^k)\} \subseteq \Omega \times \Lambda$ converging to $(x^*, F(x^*))$ with $x^k \notin \mathcal{X}$ and $(\xi^k, v^k) \in \partial^a(\Phi + \text{dist}_{\Omega \times \Lambda})(x^k, y^k)$ such that $\|(\xi^k, v^k)\| \rightarrow 0$. Since $F(x^k) \notin \Lambda$ and $y^k \in \Lambda$ for all k , we have $\|F(x^k) - y^k\| > 0$ and hence $\Phi(x^k, y^k) > 0$. By the sum rule Proposition 2.4(i), we have

$$(\xi^k, v^k) \in \partial^a \Phi(x^k, y^k) + \partial^a \text{dist}_\Omega(x^k) \times \partial^a \text{dist}_\Lambda(y^k). \quad (5.1)$$

Since F is assumed to be locally Lipschitzian, applying the maximum rule (Proposition 2.4(iii)) in calculating the subdifferential of $\Phi(x, y) := \max_{1 \leq i \leq m} \{|\psi_i(x, y)|\}$ at (x^k, y^k) yields the existence of nonnegative scalars $\{\hat{\mu}_1^k, \dots, \hat{\mu}_m^k\}$ such that

$$\sum_{i=1}^m \hat{\mu}_i^k = 1 \quad \text{and} \quad \partial^a \Phi(x^k, y^k) \subset \sum_{i=1}^m \hat{\mu}_i^k \partial^a |\psi_i|(x^k, y^k), \quad (5.2)$$

where $\hat{\mu}_i^k = 0$ if i is not an active index. Since $\Phi(x^k, y^k) > 0$, any $i \in \{1, \dots, m\}$ such that $\psi_i(x^k, y^k) = 0$ is not an active index. Hence, for all $i = 1, \dots, m$, $\psi_i(x^k, y^k) = \langle F(x^k) - y^k, e_i \rangle = 0$ implies $\hat{\mu}_i^k = 0$. Define

$$\tilde{\mu}_i^k := (\text{sign} \langle F(x^k) - y^k, e_i \rangle) \hat{\mu}_i^k.$$

We then obtain by the chain rule that

$$\tilde{\mu}_i^k \partial^a |\psi_i|(x^k, y^k) = \begin{pmatrix} \partial^a \tilde{\mu}_i^k \langle F, e_i \rangle(x^k) \\ -\tilde{\mu}_i^k e_i \end{pmatrix} \quad (5.3)$$

From (5.1)–(5.3), we obtain

$$\begin{pmatrix} \xi^k \\ v^k \end{pmatrix} \in \sum_{i=1}^m \begin{pmatrix} \partial^a \tilde{\mu}_i^k \langle F, e_i \rangle(x^k) \\ -\tilde{\mu}_i^k e_i \end{pmatrix} + \begin{pmatrix} \partial^a \text{dist}_\Omega(x^k) \\ \partial^a \text{dist}_\Lambda(y^k) \end{pmatrix},$$

that is,

$$\begin{cases} \xi^k \in \sum_{i=1}^m \partial^a \tilde{\mu}_i^k \langle F, e_i \rangle(x^k) + \partial^a \text{dist}_\Omega(x^k), \\ v^k \in \sum_{i=1}^m \tilde{\mu}_i^k (-e_i) + \partial^a \text{dist}_\Lambda(y^k). \end{cases} \quad (5.4)$$

Since by the construction $\sum_{i=1}^m |\tilde{\mu}_i^k| = 1$, the sequence $\{(\tilde{\mu}_1^k, \dots, \tilde{\mu}_m^k)\}$ is bounded and must contain a subsequence that converges to some limit $(\bar{\mu}_1, \dots, \bar{\mu}_m) \neq 0$. Taking limits as $k \rightarrow \infty$, by virtue of the closedness of the subdifferentials (Proposition 2.2), it follows that from (5.4) that

$$\begin{cases} 0 \in \sum_{i=1}^m \partial^a \langle \mu^*, e_i \rangle \langle F, e_i \rangle(x^*) + \partial^a \text{dist}_\Omega(x^*), \\ \mu^* \in \partial^a \text{dist}_\Lambda(y^*), \end{cases}$$

where $\mu^* = \sum_{i=1}^m \bar{\mu}_i e_i$. Then we have

$$\begin{cases} 0 \in \sum_{i=1}^m \partial^a \langle \mu^*, e_i \rangle \langle F, e_i \rangle(x^*) + \tilde{\mathcal{N}}_\Omega^g(x^*), \\ \mu^* \in \tilde{\mathcal{N}}_\Lambda^g(F(x^*)). \end{cases}$$

Since \mathbb{Y} is finite dimensional, by [20, Theorem 3.59], we have

$$\mu^* \in \tilde{\mathcal{N}}_\Lambda^g(F(x^*)) = \mathcal{N}_\Lambda^L(F(x^*)).$$

Since $\tilde{\mu}_i^k \rightarrow \bar{\mu}_i = \langle \mu^*, e_i \rangle \neq 0$ as $k \rightarrow \infty$ for $i \in \mathcal{I}$ where $\mathcal{I} = \{i \mid \langle \mu^*, e_i \rangle \neq 0\}$, μ_i^k has the same sign as $\langle \mu^*, e_i \rangle$ for sufficiently large k . Hence we must have $\langle \mu^*, e_i \rangle \mu_i^k > 0$ for all $i \in \mathcal{I}$ and sufficiently large k . By the definition, μ_i^k has the same sign as $\langle F(x^k) - y^k, e_i \rangle$, thus we must have $\langle \mu^*, e_i \rangle \langle F(x^k) - y^k, e_i \rangle > 0$ for all $i \in \mathcal{I}$ and sufficiently large k . Since $v^k \rightarrow 0$, $\mu^k := \sum_{i=1}^m \tilde{\mu}_i^k e_i + v^k \rightarrow \mu^*$ and then it follows from (5.4) and the fact that \mathbb{Y} is finite dimensional that

$$\mu^k \in \tilde{\mathcal{N}}_\Lambda^g(y^k) = \mathcal{N}_\Lambda^L(y^k).$$

However, these facts above and $\mu^* \neq 0$ imply that the quasinormality in the general Banach space is violated at x^* and hence a contradiction. \blacksquare

Now we are ready to give the main result of this section about the existence of local error bounds.

Theorem 5.1 *Let x^* be feasible for problem (MPGC). Suppose that x^* is quasinormal. Then the local error bound holds, i.e., there exists δ_0 such that*

$$\text{dist}_{\mathcal{X}}(x) \leq c \text{dist}_{\Lambda}(F(x)) \quad \forall x \in \mathcal{B}_{\delta_0}(x^*) \cap \Omega,$$

where c is defined as in Lemma 5.1.

Proof. According to Lemma 5.1, there exist constants $\delta > 0$ and $c > 0$ such that for all $(\xi, v) \in \partial^a(\Phi + \text{dist}_{\Omega \times \Lambda})(x, y)$ with $(x, y) \in (\mathcal{B}_{\delta}(x^*) \times \mathcal{B}_{\delta}(F(x^*))) \cap (\Omega \times \Lambda)$ and $x \notin \mathcal{X}$,

$$\|(\xi, v)\| \geq \frac{1}{c},$$

where $\Phi(x, y) = \max_{1 \leq i \leq m} \{|\langle F(x) - y, e_i \rangle|\}$. It follows from [31, Theorem 3.1] that for all $x \in \mathcal{B}_{\frac{\delta}{2}}(x^*) \cap \Omega$ and $y \in \mathcal{B}_{\frac{\delta}{2}}(F(x^*)) \cap \Lambda$,

$$\text{dist}_{\mathcal{S}}(x, y) \leq c \|F(x) - y\|, \tag{5.5}$$

where $\mathcal{S} := \{(x, y) \in \Omega \times \Lambda \mid F(x) - y = 0\}$. Let $\text{dist}_{\Lambda}(F(x)) = \|F(x) - y_x\|$ with $y_x \in \Lambda$. It follows from the continuity that there exists $\delta_0 \in (0, \frac{\delta}{2})$ such that if $x \in \mathcal{B}_{\delta_0}(x^*) \cap \Omega$, then $y_x \in \mathcal{B}_{\frac{\delta}{2}}(F(x^*))$. Thus, it follows from (5.5) that for each $x \in \mathcal{B}_{\delta_0}(x^*) \cap \Omega$,

$$\text{dist}_{\mathcal{S}}(x, y_x) \leq c \|F(x) - y_x\| = c \text{dist}_{\Lambda}(F(x)). \tag{5.6}$$

It is clear that for each x ,

$$\text{dist}_{\mathcal{X}}(x) \leq \text{dist}_{\mathcal{S}}(x, y_x).$$

This and (5.6) imply that

$$\text{dist}_{\mathcal{X}}(x) \leq c \text{dist}_{\Lambda}(F(x))$$

holds for each $x \in \mathcal{B}_{\delta_0}(x^*) \cap \Omega$. ■

As one of the main results, [19] has proven that the quasinormality implies the existence of a local error bound for smooth nonlinear programs in \mathbb{R}^n . Still in \mathbb{R}^n , [33, Theorem 5] extends [19, Theorem 2.1] to nonlinear programs with nonsmooth objective and constraints, and shows that the quasinormality implies local error bounds under some regularity conditions. Taking into account the previous results for problem (MPGC), we can now eliminate the extra regularity conditions and hence complete the investigation for local error bounds under the quasinormality for nonsmooth (NLP) problems in an infinite dimensional space. The improvement of our result owes much to the new approach constructing the enhanced sequential structure.

Corollary 5.1 *Let x^* be feasible for problem (NLP). Suppose that x^* is quasinormal. Then the local error bound holds, i.e., there are $\delta > 0$ and $c > 0$ such that*

$$\text{dist}_{\mathcal{F}_1}(x) \leq c(\|h(x)\| + \|g^+(x)\|) \quad \forall x \in \mathcal{B}_\delta(x^*) \cap \Omega,$$

where \mathcal{F}_1 is the feasible region of (NLP).

Corollary 5.2 *Let x^* be feasible for problem (NLSDP). Suppose that x^* is quasinormal. Then the local error bound holds, i.e., there are $\delta > 0$ and $c > 0$ such that*

$$\text{dist}_{\mathcal{F}_2}(x) \leq c\lambda_1(H(x))_+ \quad \forall x \in \mathcal{B}_\delta(x^*),$$

where \mathcal{F}_2 is the feasible region of (NLSDP).

Taking Theorem 5.1 into account, we can now follow the Clarke's exact penalty principle [5, Proposition 2.4.3] and then get an exact penalty result for (MPGC) immediately.

Corollary 5.3 *Let x^* be a local minimizer of problem (MPGC). If the quasinormality holds at x^* , then x^* is a local minimizer of the following penalized problem:*

$$\min_{x \in \Omega} f(x) + cL_f \text{dist}_\Lambda(F(x))$$

where L_f is the Lipschitzian constant of f near x^* and c is the error bound constant.

6 Sensitivity analysis

Mordukhovich and Nam [21], Mordukhovich et al [23], and Mordukhovich et al [22] studied the limiting subdifferential and singular subdifferential of value functions (or marginal functions) of a class of general optimization problems with abstract set-valued mapping constraints in Banach space and, Dempe et al [6] and [7] investigated the sensitivity of two-level value functions of pessimistic bilevel program and optimistic bilevel program in \mathbb{R}^n respectively in terms of classical KKT conditions by making use of the advanced tools of variational analysis [20]. In this section, we will study the sensitivity of value functions of (MPGC) and give much tighter upper estimate in terms of enhanced KKT conditions. Consider the following parametric mathematical program with geometric constraints

$$\begin{aligned} (\text{MPGC}_p) \quad & \min_{x \in \Omega} && f(x, p) \\ & \text{s.t.} && F(x, p) \in \Lambda, \end{aligned}$$

where $f : \mathbb{X} \times \mathbb{P} \rightarrow \mathbb{Y}$ and $F : \mathbb{X} \times \mathbb{P} \rightarrow \mathbb{Y}$ are locally Lipschitzian, where \mathbb{P} is a Banach space. Denote by $\mathcal{X}(p)$ the feasible region of problem (MPGC_p) . We focus on the value function

$$\mathcal{V}(p) := \inf\{f(x, p) \mid x \in \mathcal{X}(p)\}$$

and the solution mapping

$$\mathcal{O}(p) := \{x \mid f(x, p) = \mathcal{V}(p)\}.$$

To derive the sensitivity result in this section, we need to use the closedness of the approximal subdifferential and approximate normal cone. Since the nucleus G-normal cone $\tilde{\mathcal{N}}_{\Omega}^g(x^*)$ as a set-valued mapping is not necessarily closed in Banach spaces, we consider the following slightly stronger quasinormality throughout this section by noting that the A-normal cone includes the nucleus G-normal cone as a subset.

Definition 6.1 (x^*, p^*) is said to be *strongly quasinormal* for $\{(x, p) \in \Omega \times \mathbb{P} \mid F(x, p) \in \Lambda\}$ if there is no nonzero vector $\eta^* \in \mathbb{Y}$ such that

- (1) $0 \in \sum_{i=1}^m \partial^a \langle \eta^*, e_i \rangle \langle F, e_i \rangle(x^*, p^*) + \mathcal{N}_{\Omega}^a(x^*) \times \{0\}$, $\eta^* \in \mathcal{N}_{\Lambda}^L(F(x^*, p^*))$;
- (2) There exists a sequence $\{x^k, p^k, y^k, \eta^k\} \subseteq \Omega \times \mathbb{P} \times \Lambda \times \mathbb{Y}$ converging to $(x^*, p^*, F(x^*, p^*), \eta^*)$ such that for all k ,

$$\eta^k \in \mathcal{N}_{\Lambda}^L(y^k), \quad \langle \eta^*, e_i \rangle \neq 0 \implies \langle \eta^*, e_i \rangle \langle F(x^k, p^k) - y^k, e_i \rangle > 0.$$

The set of multipliers $\eta^* \in \mathcal{N}_{\Lambda}^L(F(x^*, p^*))$ satisfying (2) above is also denoted by $\mathcal{N}_{\Lambda}^e(F(x^*, p^*))$.

The following result shows that the strong quasinormality is robust. Since the proof is similar to [33, Lemma 1] and [26, Lemma 2], we omit it here.

Proposition 6.1 *If the strong quasinormality holds at $(x^*, p^*) \in \{(x, p) \in \Omega \times \mathbb{P} \mid F(x, p) \in \Lambda\}$, then it holds at all feasible points near (x^*, p^*) .*

It is well known that the MFCQ implies that the multiplier mapping is locally bounded (i.e., uniformly compact). The following result shows that the strong quasinormality implies that the ϵ -quasinormality multiplier mapping

$$\begin{aligned} (\epsilon, x, p) \rightarrow \mathcal{M}^Q(\epsilon, x, p) := \\ \{\eta \in \mathcal{N}_{\Lambda}^e(F(x, p)) \mid \epsilon \in \partial^a f(x, p) + \sum_{i=1}^m \partial^a \langle \eta, e_i \rangle \langle F, e_i \rangle(x, p) + \mathcal{N}_{\Omega}^a(x) \times \{0\}\} \end{aligned}$$

is locally bounded. Since its proof is similar to [33, Theorem 3], we also omit it here.

Proposition 6.2 *If the strong quasinormality holds at*

$$(x^*, p^*) \in \{(x, p) \in \Omega \times \mathbb{P} \mid F(x, p) \in \Lambda\},$$

then the ϵ -quasinormality multiplier mapping \mathcal{M}^Q is locally bounded at (ϵ^, x^*, p^*) , where ϵ^* is an arbitrary given element in \mathbb{X}^* .*

For sake of simplicity, given $\epsilon \geq 0$ and $r \geq 0$, we denote by $\mathcal{Q}_\epsilon^r(x^*, p^*)$ the set of vectors (η^*, ζ) satisfying

- (i) $0 \in r\partial^a f(x^*, p^*) + \sum_{i=1}^m \partial^a \langle \eta^*, e_i \rangle \langle F, e_i \rangle (x^*, p^*) - (0, \zeta) + \mathcal{N}_\Omega^a(x^*) \times \epsilon \mathbb{B}_{\mathbb{P}^*}$ with $r \geq 0$ and $\eta^* \in \mathcal{N}_\Lambda^L(F(x^*, p^*))$,
- (ii) There exists a sequence $\{(x^k, p^k, y^k, \eta^k)\} \subset \Omega \times \mathbb{P} \times \Lambda \times \mathbb{Y}$ converging to $(x^*, p^*, F(x^*, p^*), \eta^*)$ such that for all k ,

$$\begin{aligned} f(x^k) &< f(x^*), \\ \eta^k &\in \mathcal{N}_\Lambda^L(y^k), \\ \langle \eta^*, e_i \rangle &\neq 0 \implies \langle \eta^*, e_i \rangle \langle F(x^k, p^k) - y^k, e_i \rangle > 0. \end{aligned}$$

Theorem 6.1 *Let $x^* \in \mathcal{O}(p^*)$. Assume that (x^*, p^*) is strongly quasinormal for the region $\{(x, p) \in \Omega \times \mathbb{Y} \mid F(x, p) \in \Lambda\}$. Then for any $\epsilon > 0$, we have*

$$\hat{\partial}_\epsilon \mathcal{V}(p^*) \subseteq \{\zeta \mid (\eta^*, \zeta) \in \mathcal{Q}_\epsilon^1(x^*, p^*)\}.$$

Proof. Let $\zeta \in \hat{\partial}_\epsilon \mathcal{V}(p^*)$. Then by the definition of ϵ -subdifferential, for each $\bar{\epsilon} > 0$, there exists $\delta_{\bar{\epsilon}} > 0$ such that

$$\mathcal{V}(p) - \mathcal{V}(p^*) \geq \langle \zeta, p - p^* \rangle - (\epsilon + \bar{\epsilon}) \|p - p^*\|, \quad \forall p \in \mathcal{B}_{\delta_{\bar{\epsilon}}}(p^*).$$

By the definition of value functions, we have $f(x, p) \geq \mathcal{V}(p)$ for every $x \in \mathcal{X}(p)$ and hence

$$f(x, p) - \langle \zeta, p - p^* \rangle + (\epsilon + \bar{\epsilon}) \|p - p^*\| \geq f(x^*, p^*), \quad \forall x \in \mathcal{X}(p), \quad \forall p \in \mathcal{B}_{\delta_{\bar{\epsilon}}}(p^*).$$

Thus, (x^*, p^*) is a locally optimal solution to the optimization problem

$$\begin{aligned} \min_{x \in \Omega, p \in \mathbb{P}} \quad & f(x, p) - \langle \zeta, p - p^* \rangle + (\epsilon + \bar{\epsilon}) \|p - p^*\| \\ \text{s.t.} \quad & F(x, p) \in \Lambda. \end{aligned}$$

Since (x^*, p^*) is strongly quasinormal for the above problem, it follows from Theorem 4.1 and (3.11) that there exist $\eta^* \in \mathcal{N}_\Lambda^L(F(x^*, p^*))$ and $\kappa \geq 0$ such that

- (i) $0 \in \partial^a f(x^*, p^*) + \sum_{i=1}^m \partial^a \langle \eta^*, e_i \rangle \langle F, e_i \rangle (x^*, p^*) - (0, \zeta) + \mathcal{N}_\Omega^a(x^*) \times (\epsilon + \tilde{\epsilon}) \mathbb{B}_{\mathbb{P}^*}$.
- (ii) There exists a sequence $\{(x^k, p^k, y^k, \eta^k)\} \subset \Omega \times \mathbb{P} \times \Lambda \times \mathbb{Y}$ converging to $(x^*, p^*, F(x^*, p^*), \eta^*)$ such that for all k ,

$$\begin{aligned} f(x^k) &< f(x^*), \\ \eta^k &\in \mathcal{N}_\Lambda^L(y^k), \\ \langle \eta^*, e_i \rangle \neq 0 &\implies \langle \eta^*, e_i \rangle \langle F(x^k, p^k) - y^k, e_i \rangle > 0. \end{aligned}$$

The desired result is obtained since $\tilde{\epsilon}$ is arbitrary. ■

Definition 6.2 We say that the *inf-compactness* holds for (MPGC_p) with $p = p^*$ if there exist a number α and a compact set S such that for each p in some neighborhood of p^* , the level set

$$\{x \in \mathcal{X}(p) \mid f(x, p) \leq \alpha\}$$

is nonempty and contained in S .

Theorem 6.2 *Assume that the inf-compactness holds for problem (MPGC). Suppose that for each $x^* \in \mathcal{O}(p^*)$, (x^*, p^*) is strongly quasinormal for the constraint region $\{(x, p) \in \Omega \times \mathbb{Y} \mid F(x, p) \in \Lambda\}$. Then the value function $\mathcal{V}(p)$ is lower semicontinuous around p^* and*

$$\begin{aligned} \partial^L \mathcal{V}(p^*) &\subset \bigcup_{x^* \in \mathcal{O}(p^*)} \{\zeta \mid (\eta^*, \zeta) \in \mathcal{Q}_0^1(x^*, p^*)\}, \\ \partial^\infty \mathcal{V}(p^*) &\subset \bigcup_{x^* \in \mathcal{O}(p^*)} \{\zeta \mid (\eta^*, \zeta) \in \mathcal{Q}_0^0(x^*, p^*)\}. \end{aligned}$$

Proof. The lower semicontinuity follows from the proof of [4, Proposition 4.4] immediately. We complete the proof by considering the following two cases:

- (a) Let $\zeta \in \partial^L \mathcal{V}(p^*)$. By the definition, there exist sequences $p^l \xrightarrow{\mathcal{V}} p^*$, $\epsilon_l \downarrow 0$, and $\zeta^l \xrightarrow{w^*} \zeta$ with $\zeta^l \in \hat{\partial}_{\epsilon_l} \mathcal{V}(p^l)$. Since the inf-compactness holds, the set $\{x \in \mathcal{X}(p^l) \mid f(x, p^l) \leq \alpha\}$ is nonempty when l is sufficiently large. By the inf-compactness again, there exists $x^l \in \mathcal{O}(p^l)$ and, without loss of generality, we may assume that $x^l \rightarrow x^*$. Since $\mathcal{V}(p^l) \rightarrow \mathcal{V}(p^*)$ and

$$\mathcal{V}(p^l) = f(x^l, p^l) \rightarrow f(x^*, p^*),$$

we have $f(x^*, p^*) = \mathcal{V}(p^*)$. Thus, $x^* \in \mathcal{O}(p^*)$. Since the strong quasinormality holds at (x^*, p^*) and $(p^l, x^l) \rightarrow (p^*, x^*)$, by Proposition 6.1, the strong quasinormality holds at (x^l, p^l) for each sufficiently large l . Thus, we have from Theorem 6.1 that, for each sufficiently large l , there exist η^l and $\kappa \geq 0$ such that

- (1) $(0, \zeta^l) \in \partial^a f(x^l, p^l) + \sum_{i=1}^m \partial^a \langle \eta^l, e_i \rangle \langle F, e_i \rangle (x^l, p^l) + \mathcal{N}_\Omega^a(x^l) \times \epsilon_l \mathbb{B}_{\mathbb{P}^*}$ with $\eta^l \in \mathcal{N}_\Lambda^L(F(x^l, p^l))$.
- (2) There exists a sequence $\{(x^{l,k}, p^{l,k}, y^{l,k}, \eta^{l,k})\} \subset \mathbb{X} \times \mathbb{P} \times \Lambda \times \mathbb{Y}$ converging to $(x^l, p^l, F(x^l, p^l), \eta^l)$ such that for all k ,

$$\begin{aligned} f(x^{l,k}) &< f(x^l), \\ \eta^{l,k} &\in \mathcal{N}_\Lambda^L(y^{l,k}), \\ \langle \eta^l, e_i \rangle \neq 0 &\implies \langle \eta^l, e_i \rangle \langle F(x^{l,k}, p^{l,k}) - y^{l,k}, e_i \rangle > 0. \end{aligned}$$

By the quasinormality assumption and Proposition 6.2, the sequence $\{\eta^l\}$ is bounded. Thus, without loss of generality, we may assume that $\{\eta^l\}$ converges to η^* . Taking a limit in (1) above, it is not hard to see from the weak* closedness of basic subdifferential and normal cone that

$$\begin{cases} (0, \zeta) \in \partial^a f(x^*, p^*) + \sum_{i=1}^m \partial^a \langle \eta^*, e_i \rangle \langle F, e_i \rangle (x^*, p^*) + \mathcal{N}_\Omega^a(x^*) \times \{0\}, \\ \eta^* \in \mathcal{N}_\Lambda^L(F(x^*, p^*)). \end{cases}$$

Also by the diagonal rule, we can find a sequence $\{(x^{l,k_l}, p^{l,k_l}, y^{l,k_l}, \eta^{l,k_l})\}$ converging to (x^*, p^*, y^*, η^*) as $l \rightarrow \infty$ is such that for all l ,

$$\begin{aligned} f(x^{l,k_l}) &< f(x^*) \\ \eta^{l,k_l} &\in \mathcal{N}_\Lambda^L(y^{l,k_l}), \\ \langle \eta^*, e_i \rangle \neq 0 &\implies \langle \eta^*, e_i \rangle \langle F(x^{l,k_l}, p^{l,k_l}) - y^{l,k_l}, e_i \rangle > 0. \end{aligned}$$

Therefore, it follows that $(\eta^*, \zeta) \in \mathcal{Q}_0^1(x^*, p^*)$.

- (b) Let $\zeta \in \partial^\infty \mathcal{V}(p^*)$. By the definition, there exist sequence $p^l \xrightarrow{\mathcal{V}} p^*$, $\epsilon_l \downarrow 0$, $\zeta^l \in \hat{\partial}_{\epsilon_l} \mathcal{V}(p^l)$, and $t_l \downarrow 0$ such that $t_l \zeta^l \rightarrow \zeta$. Similar as (a) in this proof, for each l sufficiently large l , there exist η^l and $\kappa \geq 0$ such that (1)–(2) hold. It is easy to get from (1) that

$$\begin{cases} (0, t_l \zeta^l) \in t_l \partial^a f(x^l, p^l) + \sum_{i=1}^m \partial^a \langle t_l \eta^l, e_i \rangle \langle F, e_i \rangle (x^l, p^l) + \mathcal{N}_\Omega^a(x^l) \times t_l \epsilon_l \mathbb{B}_{\mathbb{P}^*}, \\ t_l \eta^l \in \mathcal{N}_\Lambda^L(F(x^l, p^l)). \end{cases} \quad (6.1)$$

By the quasinormality assumption and Proposition 6.2, the sequence $\{t_l \eta^l\}$ is bounded. Without loss of generality, assume that $\{t_l \eta^l\}$ converges to η^* . Taking a limit as $k \rightarrow \infty$ in (6.1), we have from the weak* closedness of the basic subdifferential and the normal cone that

$$\begin{cases} (0, \zeta) \in \sum_{i=1}^m \partial^a \langle \eta^*, e_i \rangle \langle F, e_i \rangle (x^*, p^*) + \mathcal{N}_\Omega^a(x^*) \times \{0\}, \\ \eta^* \in \mathcal{N}_\Lambda^L(F(x^*, p^*)). \end{cases}$$

The rest of the proof is similar to (a).

■

Acknowledgements. The authors are grateful to the two anonymous referees for their extremely helpful comments and suggestions.

References

- [1] D.P. Bertsekas, *Nonlinear Programming*. Athena Scientific Publishers, 1999.
- [2] D.P. Bertsekas, A. Nedic and A.E. Ozdaglar, *Convex Analysis and Optimization*, Athena Scientific, Belmont, Massachusetts, 2003.
- [3] D.P. Bertsekas and A.E. Ozdaglar, *Pseudonormality and a Lagrange multiplier theory for constrained optimization*. *Journal of Optimization Theory and Applications*, 114 (2002), 287–343.
- [4] J.F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.
- [5] F.H. Clarke, *Optimization and Nonsmooth Analysis*. Wiley Interscience, New York, 1983; reprinted as vol. 5 of *Classics Appl. Math.* 5, SIAM, Philadelphia, PA, 1990.
- [6] S. Dempe, B.S. Mordukhovich and A.B. Zemkoho, *Necessary optimality conditions in pessimistic bilevel programming*, *Optimization*, doi: 10.1080/02331934.2012.696641.
- [7] S. Dempe, B.S. Mordukhovich and A.B. Zemkoho, *Sensitivity analysis for two-level value functions with applications to bilevel programming*, *SIAM Journal on Optimization*, 22 (2012), 1309–1343.
- [8] C. Ding, D. Sun and J.J. Ye, *First order optimality conditions for mathematical programs with semidefinite cone complementarity constraints*, revised for *Mathematical Programming*, series A.
- [9] B.M. Glover and B.D. Craven, *A Fritz John optimality condition using the approximate subdifferential*, *Journal of Optimization Theory and Applications*, 82 (1994), 253–265
- [10] L. Guo, G.H. Lin and J.J. Ye, *Stability analysis for parametric mathematical programs with geometric constraints and its applications*, *SIAM Journal on Optimization*, 22 (2012), 1151–1176.

- [11] A.J. Hoffman, *On approximate solutions of systems of linear inequalities*, Journal of Research of the National Bureau of Standards, 49 (1952), 263–265.
- [12] A.D. Ioffe, *Regular points of Lipschitz functions*, Transaction of American Mathematical Society, 251 (1979), 61–69.
- [13] A.D. Ioffe, *Approximate subdifferentials and applications, II: Functions on locally convex spaces*, Mathematika, 33 (1986), 111–128.
- [14] A.D. Ioffe, *Approximate subdifferentials and applications, III: The metric theory*, Mathematika, 36 (1989), 1–38.
- [15] A. Jourani and L. Thibault, *The approximate subdifferential of composite functions*, Bulletin of the Australian Mathematical Society, 47 (1993), 443–456.
- [16] A.S. Lewis, *Nonsmooth analysis of eigenvalues*, Mathematical Programming, 84 (1999), 1–24.
- [17] Z.Q. Luo, J.S. Pang and D. Ralph, *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, UK, 1996.
- [18] P. Maréchal and J.J. Ye *Optimizing condition numbers*, SIAM Journal on Optimization, 20 (2009), 935–947.
- [19] L. Minchenko and A. Tarakanov, *On error bounds for quasinormal programs*, Journal of Optimization Theory and Applications, 148 (2011), 571–579.
- [20] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation I: Basic Theory, II: Application*, Grundlehren der Mathematischen Wissenschaften 330, Springer, Berlin, 2006.
- [21] B.S. Mordukhovich and N.M. Nam, *Variational stability and marginal functions via generalized differentiation*, Mathematics of Operations Research, 30 (2005), 800–816.
- [22] B.S. Mordukhovich, N.M. Nam and H.M. Phan, *Variational analysis of marginal functions with applications to bilevel programming*, Journal of Optimization Theory and Applications, 152 (2012), 557–586.
- [23] B.S. Mordukhovich, N.M. Nam and N.D. Yen, *Subgradients of marginal functions in parametric mathematical programming*, Mathematical Programming, 116 (2009), 369–396.

- [24] M.L. Overton and R.S. Womersley, *On the sum of the largest eigenvalues of a symmetric matrix*, SIAM Journal on Matrix Analysis and Applications, 13 (1992), 41–45.
- [25] J.V. Outrata, M. Kocvara and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results*, Kluwer Academic Publishers, Boston, 1998.
- [26] A.E. Ozdaglar and D.P. Bertsekas, *The relation between pseudonormality and quasiregularity in constrained optimization*, Optimization Methods and Software, 19 (2004), 493–506.
- [27] S.M. Robinson, *Generalized equations and their solution, part II: Applications to nonlinear programming*, Mathematical Programming Study, 19 (1982), 200–221.
- [28] R.T. Rockafellar, *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1970.
- [29] R.T. Rockafellar and R.J.-B. Wets, *Variational Analysis*. Springer, Berlin, 1998.
- [30] W. Schirotzek, *Nonsmooth Analysis*. Springer, Berlin, 2007.
- [31] Z.L. Wu and J.J. Ye, *Sufficient conditions for error bounds*, SIAM Journal on Optimizaiton, 12 (2001), 421–435.
- [32] Z.L. Wu and J.J. Ye, *First and second order conditions for error bounds*, SIAM Journal on Optimization, 14 (2003), 621–645.
- [33] J.J. Ye and J. Zhang, *Enhanced Karush-Kuhn-Tucker condition and weaker constraint qualifications*, Mathematical Programming, 139 (2013), 353–381, series B, doi 10.1007/s10107-013-0667-7.
- [34] J.J. Ye and J. Zhang, *Enhanced Karush-Kuhn-Tucker condition for mathematical programs with equilibrium constraints*, Journal of Optimization Theory and Applications, to appear.
- [35] J.J. Ye and D.L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization, 33 (1995), 9–27.
- [36] J.J. Ye, D.L. Zhu and Q.J. Zhu, *Exact penalization and necessary optimality conditions for generalized bilevel programming problems*, SIAM Journal on Optimization, 7 (1997), 481–507.