Noname manuscript No.
(will be inserted by the editor)

Enhanced Karush-Kuhn-Tucker Condition and Weaker Constraint Qualifications

Jane J. Ye · Jin Zhang

Dedicated to Jon Borwein in honor of his 60th birthday.

Received: date / Accepted: date

Abstract In this paper we study necessary optimality conditions for nonsmooth optimization problems with equality, inequality and abstract set constraints. We derive the enhanced Fritz John condition which contains some new information even in the smooth case than the classical enhanced Fritz John condition. From this enhanced Fritz John condition we derive the enhanced Karush-Kuhn-Tucker (KKT) condition and introduce the associated pseudonormality and quasinormality condition. We prove that either pseudonormality or quasinormality with regularity on the constraint functions and the set constraint implies the existence of a local error bound. Finally we give a tighter upper estimate for the Fréchet subdifferential and the limiting subdifferential of the value function in terms of quasinormal multipliers which is usually a smaller set than the set of classical normal multipliers. In particular we show that the value function of a perturbed problem is Lipschitz continuous under the perturbed quasinormality condition which is much weaker than the classical normality condition.

Keywords enhanced Fritz John condition, enhanced KKT condition · constraint qualification · nonsmooth analysis · value function · local error bound · calmness

Mathematics Subject Classification (2010) 90C46 · 90C30 · 90C31

The first author's work was supported in part by NSERC. The second author's work was sponsored in part by China Scholarship Council and in part by NSERC.

Jane Ye
Department of Mathematics and Statistics, University of Victoria, Victoria, BC, V8W 3R4 Canada.
E-mail: janeye@uvic.ca
Jin Zhang
E-mail: zhangjin@uvic.ca

1 Introduction

In this paper we consider optimization problems in the form

$$(P) \qquad \min \ f(x) \qquad \text{s.t.} \ \ x \in \mathcal{C},$$

where the feasible region $\mathcal{C}$ is

$$\mathcal{C} = \{x \in \mathcal{X} : h_1(x) = 0, \ldots, h_p(x) = 0, g_1(x) \le 0, \ldots, g_q(x) \le 0\}. \quad (1.1)$$

Unless otherwise indicated we assume throughout this paper that $f, h_i (i = 1, \ldots, p), g_j (j = 1, \ldots, q) : \mathbb{R}^m \to \mathbb{R}$ are Lipschitz continuous around the point of interest and $\mathcal{X}$ is a nonempty closed subset of $\mathbb{R}^m$.

In 1948, Fritz John [14] proposed the now well-known Fritz John necessary optimality condition for smooth optimization problems with inequality constraints only. In 1967, Mangasarian and Fromovitz [20] extended the Fritz John condition to smooth optimization problems with equality and inequality constraints (i.e. $\mathcal{X} = \mathbb{R}^m$). For the smooth case, Fritz John condition asserts that if x^* is a local optimal solution of problem (P) with $\mathcal{X} = \mathbb{R}^m$, then there exist scalars $\lambda_1^*, \ldots, \lambda_p^*$ and $\mu_0^*, \cdots, \mu_q^*$ not all zero, satisfying $\mu_j^* \ge 0$ for all $j = 0, 1, \ldots, q$ and

$$0 = \mu_0^* \nabla f(x^*) + \sum_{i=1}^{p} \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^{q} \mu_j^* \nabla g_j(x^*), \quad (1.2)$$

$$0 = \mu_j^* g_j(x^*), \quad (1.3)$$

where $\nabla \varphi(x)$ denotes the gradient of the function φ at x. Condition (1.3) is referred to as the complementary slackness condition (CS for short). We call a multiplier $(\lambda_1^*, \cdots, \lambda_p^*, \mu_1^*, \cdots, \mu_q^*)$ satisfying the Fritz John condition (1.2)-(1.3) with $\mu_0^* = 1$ and $\mu_0^* = 0$ a normal multiplier and an abnormal multiplier respectively. It follows from the Fritz John condition that if there is no nonzero abnormal multiplier then there must exist a normal multiplier. This simple corollary from the Fritz John condition leads to the so-called No Nonzero Abnormal Multiplier Constraint Qualification (NNAMCQ for short) or the so-called Basic Constraint Qualification for the Karush-Kuhn-Tucker (KKT for short) condition to hold at a local minimum. It was Mangasarian and Fromovitz who first pointed out that the Fritz John condition can be used to derive the KKT condition under the condition that the gradient vectors

$$\nabla h_i(x^*), i = 1, \ldots, p$$

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

$$\nabla h_i(x^*)^T d = 0 \quad i = 1, \ldots, p, \quad \nabla g_j(x^*)^T d < 0 \quad j \in A(x^*),$$

where $A(x^*) := \{j : g_j(x^*) = 0\}$ is the set of active inequality constraints at x^*, using the fact that the above condition is equivalent to the NNAMCQ by the Motzkin's transposition theorem. The above condition is now well-known as the Mangasarian-Fromovitz Constraint Qualification (MFCQ).

The first but weaker versions of the enhanced Fritz John conditions were considered in a largely overlooked analysis by Hestenes [13] for the case of smooth optimization problem without an abstract set constraint. A version of the enhanced Fritz John condition first given by Bertsekas in [2] for a smooth problem with $\mathcal{X} = \mathbb{R}^m$ states that if x^* is a local optimal solution of problem (P) with $\mathcal{X} = \mathbb{R}^m$, then there exist scalars $\lambda_1^*, \dots, \lambda_p^*$ and $\mu_0^* \geq 0, \cdots, \mu_q^* \geq 0$ not all zero satisfying (1.2) and the following sequential property: If the index set $I \cup J$ is nonempty, where $I = \{i | \lambda_i^* \neq 0\}, \quad J = \{j \neq 0 | \mu_j^* > 0\}$, then there exists a sequence $\{x^k\} \subset \mathcal{X}$ converging to x^* such that for all k,

$$f(x^k) < f(x^*), \quad \lambda_i^* h_i(x^k) > 0, \ \forall i \in I, \quad \mu_j^* g_j(x^k) > 0, \ \forall j \in J. \tag{1.4}$$

Condition (1.4) is stronger than the complementary slackness condition (1.3) since if $\mu_j^* > 0$, then according to condition (1.4), the corresponding jth inequality constraint must be violated arbitrarily close to x^*, implying that $g_j(x^*) = 0$. For this reason, the condition (1.4) is called the complementarity violation condition (CV for short) by Bertsekas and Ozdaglar [4].

Since the enhanced Fritz John condition is stronger than the classical Fritz John condition, it results in a stronger KKT condition under a weaker constraint qualification than the MFCQ. The enhanced Fritz John condition has been further extended to the case of smooth problem data with a convex abstract set constraint in Bertsekas [2] and with nonconvex set in Bertsekas and Ozdaglar [4] and Bertsekas, Nedić and Ozdaglar [3].

The first result on the enhanced Fritz John condition for nonsmooth problems with no abstract set constraint can be found in Bector, Chandra and Dutta [1] where the classical gradient is replaced by the Clarke subdifferential. Duality results for convex problems in terms of the enhanced Fritz John condition have also been studied by Bertsekas, Ozdaglar and Tseng in [6]. One of the main results of this paper is an improved version of the enhanced Fritz John condition for problem (P) with Lipschitz problem data based on the limiting subdifferential and limiting normal cone. Even in the case of a smooth problem, our improved enhanced Fritz John condition provides some new information. In our improved CV, we have an extra condition that the sequence $\{x^k\}$ can be found such that the functions $f, h_i(i \in I), g_j(j \in J)$ are proximal subdifferentiable at x^k (see Definition 2). Note that our improved CV is stronger than the original CV for the smooth problem since a continuously differentiable function may not be proximal subdifferentiable (a sufficient condition for a function to be proximal subdifferentiable is C^{1+}, i.e. the gradient of the function is locally Lipschitz).

Based on the enhanced Fritz John condition, Bertsekas and Ozdaglar [4] introduced the so-called pseudonormality and quasinormality as constraint qualifications that are weaker than the MFCQ. Since our improved enhanced Fritz John condition is stronger than the original enhanced Fritz John condition even in the smooth case, our pseudonormality and quasinormality conditions are even weaker than the original pseudonormality and quasinormality respectively and are much weaker than the NNAMCQ (which is in general weaker than the MFCQ in the nonsmooth case).

In recent years, it has been shown that constraint qualifications have strong connections with certain Lipschitz-like property of the set-valued map $\mathcal{C}$: $\mathbb{R}^{p+q} \rightrightarrows \mathbb{R}^m$ defined by the perturbed feasible region

$$\mathcal{C}(\alpha,\beta) := \{x \in \mathcal{X} : h(x) = \alpha, g(x) \leq \beta\},$$

where $h := (h_1, \ldots, h_p), g := (g_1, \ldots, g_q)$. For the case of a smooth optimization problem with $\mathcal{X} = \mathbb{R}^m$, by Mordukhovich's criteria for pseudo-Lipschitz continuity ([24,25]), MFCQ (or equivalently NNAMCQ) at a feasible point x^* is equivalent to the pseudo-Lipschitz continuity (or so-called Aubin continuity) of the set-valued map $\mathcal{C}(\alpha,\beta)$ around $(0,0,x^*)$. Calmness of a set-valued map (introduced as the pseudo upper-Lipschitz continuity by Ye and Ye [33] and coined as calmness by Rockafellar and Wets [29]) is a much weaker condition than the pseudo-Lipschitz continuity. It is known that the calmness of the set-valued map $\mathcal{C}(\alpha,\beta)$ around $(0,0,x^*)$ is equivalent to the existence of local error bound for the constraint region, i.e., the existence of positive constants c, δ such that

$$d_{\mathcal{C}}(x) \leq c(\|h(x)\|_1 + \|g^+(x)\|_1) \quad \forall x \in \mathbb{B}(x^*,\delta) \cap \mathcal{X}, \tag{1.5}$$

where $d_{\mathcal{C}}(x)$ denotes the distance of x to set $\mathcal{C}$, $g^+(x) := \max\{0, g(x)\}$ where the maximization is taken componentwise, $\|\cdot\|_1$ denotes the one norm and $B(x^*,\delta)$ denotes the closed ball centered at x^* with radius δ. In this paper we show that either pseudonormality or quasinormality with regularity on the constraint functions and the set constraint implies that the set-valued map $\mathcal{C}(\alpha,\beta)$ is calm around the point $(0,0,x^*)$. Hence pseudonormality and quasinormality are much weaker than the NNAMCQ.

NNAMCQ plays an important role in the sensitivity analysis. In particular it is a sufficient condition for the value function of a perturbed problem to be Lispchitz continuous (see e.g. [17, 18]). In this paper we apply our improved enhanced KKT condition to derive an estimate for the Fréchet subdifferential and the limiting subdifferential of the value function. We provide a tighter upper estimate for the Fréchet subdifferential and the limiting subdifferentials of the value function in terms of the quasinormal multipliers. As a consequence we show that the value function is Lipschitz continuous under the perturbed quasinormality condition which is a much weaker condition than the NNAMCQ.

We organize our paper as follows. In the next section, we review the preliminaries for nonsmooth analysis that will be used in this paper. We derive the improved enhanced Fritz John condition in Section 3. New constraint qualifications, the enhanced KKT and the relationship between pseudonormality and quasinormality are given in Section 4. Section 5 is devoted to the sufficient condition for the existence of local error bounds. In Section 6, the results is applied to the sensitivity analysis to provide a tighter upper estimate for the subdifferential of the value function.

2 Preliminaries

This section contains some background material on nonsmooth analysis and preliminary results which will be used later. We give only concise definitions and results that will be needed in the paper. For more detailed information on the subject our references are Borwein and Lewis [7], Borwein and Zhu [8], Clarke [10], Clarke, Ledyaev, Stern and Wolenski [11], Loewen [19], Mordukhovich [25] and Rockafellar and Wets [29].

We first give the following notations that will be used throughout the paper. We denote by $\mathbb{B}(x^*, \epsilon)$ the closed ball centered at x^* with radius ϵ and $\mathbb{B}$ the closed unit ball centered at 0. For a set $\mathcal{C}$, we denote by int$\mathcal{C}$, cl$\mathcal{C}$, co$\mathcal{C}$ its interior, closure and convex hull respectively. For a function $g : \mathbb{R}^n \to \mathbb{R}$, we denote by $g^+(x) := \max\{0, g(x)\}$ and if it is vector-valued then the maximum is taken componentwise. For a cone $\mathcal{N}$, we denote by $\mathcal{N}^*$ its polar.

For a set-valued map $\Phi : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, we denote by $\limsup_{x \to x_0} \Phi(x)$ and $\liminf_{x \to x_0} \Phi(x)$ the Kuratowski-Painlevé upper (outer) and lower (inner) limit respectively.

Definition 1 (Subdifferentials) *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a lower semi-continuous (l.s.c.) function and $x_0 \in dom f := \{x \in \mathbb{R}^n : f(x) < +\infty\}$. The proximal subdifferential of f at x_0 is the set*

$$\partial^\pi f(x_0) := \left\{ \xi \in \mathbb{R}^n : \begin{array}{l} \exists \sigma > 0, \eta > 0 \ s.t. \\ f(x) \geq f(x_0) + \langle \xi, x - x_0 \rangle - \sigma \|x - x_0\|^2 \quad \forall x \in \mathbb{B}(x_0, \eta) \end{array} \right\}.$$

The Fréchet (regular) subdifferential of f at x_0 is the set

$$\partial^F f(x_0) := \left\{ \xi \in \mathbb{R}^n : \liminf_{h \to 0} \frac{f(x_0 + h) - f(x_0) - \langle \xi, h \rangle}{\| h \|} \geq 0 \right\}.$$

The limiting (Mordukhovich or basic) subdifferential of f at x_0 is the set

$$\begin{aligned} \partial f(x_0) &:= \left\{ \xi \in \mathbb{R}^n : \exists x_k \to x_0, \ and \ \xi_k \to \xi \ with \ \xi_k \in \partial^F f(x_k) \right\} \\ &= \left\{ \xi \in \mathbb{R}^n : \exists x_k \to x_0, \ and \ \xi_k \to \xi \ with \ \xi_k \in \partial^\pi f(x_k) \right\}. \end{aligned}$$

The singular limiting (Mordukhovich) subdifferential of f at x_0 is the set

$$\begin{aligned} \partial^\infty f(x_0) &:= \left\{ \xi \in \mathbb{R}^n : \exists x_k \to x_0, \ and \ t_k \xi_k \to \xi \ with \ \xi_k \in \partial^F f(x_k), t_k \downarrow 0 \right\} \\ &= \left\{ \xi \in \mathbb{R}^n : \exists x_k \to x_0, \ and \ t_k \xi_k \to \xi \ with \ \xi_k \in \partial^\pi f(x_k), t_k \downarrow 0 \right\}. \end{aligned}$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz near x_0. The Clarke subdifferential (generalized gradient) of f at x_0 is the set $\partial^c f(x_0) = clco \partial f(x_0)$.

When f is strictly differentiable (see the definition, e.g. in Clarke [10]), $\partial f(x_0) = \partial^c f(x_0) = \{\nabla f(x_0)\}$. A l.s.c. function f is said to be subdifferentially regular ([25, Definition 1.91]) at x_0 if $\partial f(x_0) = \partial^F f(x_0)$. It is known that for a locally Lipschitz continuous function, the subdifferential regularity is the same as the Clarke regularity (see [10, Definition 2.3.4] for the definition).

The following facts about the subdifferentials are well-known.

Proposition 1 *(i) A function $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz near x_0 and $\partial f(x_0) = \{\zeta\}$ if and only if f is strictly differentiable at x_0 and the gradient of f at x_0 is equal to ζ.*

(ii) If a function $f : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz near x_0 with positive constant L_f, then $\partial f(x_0) \subset L_f cl\mathbb{B}$.

(iii) A l.s.c. function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is Lipschitz near x_0 if and only if $\partial^\infty f(x_0) = \{0\}$.

(iv) Let $a \in \mathbb{R}$. Then

$$\partial \max\{0, a\} = \begin{cases} \{0\} & a < 0 \\ [0,1] & a = 0 \\ \{1\} & a > 0 \end{cases}, \qquad \partial |a| = \begin{cases} \{-1\} & a < 0 \\ [-1,1] & a = 0 \\ \{1\} & a > 0 \end{cases}.$$

Definition 2 (Proximal subdifferentiability) *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a l.s.c. function and $x_0 \in dom f$. We say that f is proximal subdifferentiable at x_0 if $\partial^\pi f(x_0) \neq \emptyset$.*

Definition 3 (Normal cones) *Let Ω be a nonempty subset of $\mathbb{R}^n$ and $x_0 \in cl\Omega$. The convex cone*

$$\mathcal{N}^\pi_\Omega(x_0) := \left\{\xi \in \mathbb{R}^n : \exists \sigma > 0 \ s.t. \ \langle \xi, x - x_0 \rangle \leq \sigma \|x - x_0\|^2 \quad \forall x \in \Omega \right\}$$

is called the proximal normal cone to Ω at x_0. The convex cone

$$\mathcal{N}^F_\Omega(x_0) := \left\{\xi \in \mathbb{R}^n : \limsup_{x \to x_0, x \in \Omega} \frac{\langle \xi, x - x_0 \rangle}{\|x - x_0\|} \leq 0 \right\}$$

is called the Fréchet (regular) normal cone to Ω at x_0. The nonempty cone

$$\mathcal{N}_\Omega(x_0) := \limsup_{x \to x_0} \mathcal{N}^F_\Omega(x_0) = \limsup_{x \to x_0} \mathcal{N}^\pi_\Omega(x_0)$$

is called the limiting (Mordukhovich or basic) normal cone to Ω at x_0. The Clarke normal cone is the closure of the convex hull of the limiting normal cone, i.e., $\mathcal{N}^c_\Omega(x_0) = clco\mathcal{N}_\Omega(x_0)$.

We say a set Ω is regular if $\mathcal{N}^F_\Omega(x) = \mathcal{N}_\Omega(x)$ for all $x \in \Omega$.

Proposition 2 (Calculus rules) *(i) Let $f : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz near x_0 and $g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be l.s.c. and finite at x_0. Let α, β be nonnegative scalars. Then $\partial(\alpha f + \beta g)(x_0) \subset \alpha \partial f(x_0) + \beta \partial g(x_0)$.*

(ii) [27, Corollary 3.4] Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be l.s.c. near x_0 and $g : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz near x_0. Assume that $\partial^F g(x_0) \neq \emptyset$ for all x near x_0. Then

$$\partial(f - g)(x_0) \subset \partial f(x_0) - \partial g(x_0).$$

(iii) Let $\varphi : \mathbb{R}^m \to \mathbb{R}^n$ be Lipschitz near x_0 and $f : \mathbb{R}^n \to \mathbb{R}$ be Lipschitz near $\varphi(x_0)$. Then

$$\partial(f \circ \varphi)(x_0) \subset \cup_{\xi \in \partial f(\varphi(x_0))} \partial \langle \xi, \varphi \rangle (x_0).$$

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

(v) Let $f : \mathbb{R}^n \to \mathbb{R}$ be Fréchet differentiable at x^* and $\mathcal{C}$ be a closed subset of $\mathbb{R}^n$. If x^* is a local minimizer of f on $\mathcal{C}$, then $0 \in \nabla f(x^*) + \mathcal{N}^F_\mathcal{C}(x^*)$.

3 Enhanced Fritz John necessary optimality condition

For nonsmooth problem (P), the classical Fritz John necessary optimality condition is generalized to one where the classical gradient is replaced by the generalized gradient by Clarke ([9], see also [10, Theorem 6.1.1]). The limiting subdifferential version of the Fritz John condition was first obtained by Mordukhovich in [23] (see also [30, Corollary 4.2] for more explicit expressions).

The following theorem strengthens the limiting subdifferential version of the Fritz John conditions by replacing the complementary slackness condition with a stronger condition [Theorem 1(iv)], and hence their effectiveness has been significantly enhanced. Although [Theorem 1(iv)] is slightly stronger than the complementarity violation condition of Bertsekas and Ozdaglar [4], for convenience we still refer to it as the complementarity violation condition (CV).

Theorem 1 *Let x^* be a local minimum of problem (P). Then there exist scalars $\mu_0^*, \lambda_1^*, \ldots, \lambda_p^*, \mu_1^*, \ldots, \mu_q^*$, satisfying the following conditions:*

(i) $0 \in \mu_0^* \partial f(x^*) + \sum_{i=1}^p \partial(\lambda_i^* h_i)(x^*) + \sum_{j=1}^q \mu_j^* \partial g_j(x^*) + \mathcal{N}_{\mathcal{X}}(x^*)$.
(ii) $\mu_j^* \geq 0, \qquad$ *for all* $j = 0, 1, \ldots, q$.
(iii) $\mu_0^*, \lambda_1^*, \ldots, \lambda_p^*, \mu_1^*, \ldots, \mu_q^*$ *are not all equal to 0.*
(iv) *The complementarity violation condition holds: If the index set $I \cup J$ is nonempty, where $I = \{i | \lambda_i^* \neq 0\}$, $\quad J = \{j \neq 0 | \mu_j^* > 0\}$, then there exists a sequence $\{x^k\} \subset \mathcal{X}$ converging to x^* such that for all k,*

$$f(x^k) < f(x^*), \qquad \lambda_i^* h_i(x^k) > 0,\ \forall i \in I, \qquad \mu_j^* g_j(x^k) > 0,\ \forall j \in J,$$

and $f, h_i (i \in I)$, $g_j (j \in J)$ are all proximal subdifferentiable at x^k.

Proof Similar to the differentiable case in Bertsekas and Ozdaglar [4], we use a quadratic penalty function approach originated with McShane [21] to prove the result. For each $k = 1, 2, \ldots,$ we consider the penalized problem

$$(P_k) \qquad \min\ F^k(x) = f(x) + \frac{k}{2}\sum_{i=1}^p (h_i(x))^2 + \frac{k}{2}\sum_{j=1}^q (g_j^+(x))^2 + \frac{1}{2}\|x - x^*\|^2$$
$$\text{s.t. } x \in \mathcal{X} \cap B(x^*, \epsilon),$$

where $\epsilon > 0$ is such that $f(x^*) \leq f(x)$ for all feasible x with $x \in B(x^*, \epsilon)$. Since $\mathcal{X} \cap B(x^*, \epsilon)$ is compact, by the Weierstrass theorem, an optimal solution $\mathbb{x}^k$ of the problem (P_k) exists. Consequently

$$f(\mathbb{x}^k) + \frac{k}{2}\sum_{i=1}^p (h_i(\mathbb{x}^k))^2 + \frac{k}{2}\sum_{j=1}^q (g_j^+(\mathbb{x}^k))^2 + \frac{1}{2}\|\mathbb{x}^k - x^*\|^2 = F^k(\mathbb{x}^k)$$
$$\leq F^k(x^*) = f(x^*). \tag{3.1}$$

Since $f(\mathbb{x}^k)$ is bounded over $x \in \mathcal{X} \cap B(x^*, \epsilon)$, we obtain from (3.1) that

$$\lim_{k\to\infty} |h_i(\mathbb{x}^k)| = 0, \qquad i = 1, \ldots, p, \qquad \lim_{k\to\infty} |g_j^+(\mathbb{x}^k)| = 0, \qquad j = 1, \ldots, q$$

and hence every limit point $\bar{x}$ of $\{\mathbb{x}^k\}$ is feasible; i.e., $\bar{x} \in \mathcal{C}$. Furthermore, (3.1) yields

$$f(\mathbb{x}^k) + \frac{1}{2}\|\mathbb{x}^k - x^*\|^2 \leq f(x^*), \qquad \forall k. \tag{3.2}$$

So by taking limit as $k \to \infty$, we obtain

$$f(\bar{x}) + \frac{1}{2}\|\bar{x} - x^*\|^2 \leq f(x^*).$$

Since $\bar{x} \in B(x^*, \epsilon)$ and $\bar{x}$ is feasible, we have $f(x^*) \leq f(\bar{x})$, which combined with the preceding inequality yields $\|\bar{x} - x^*\| = 0$ so that $\bar{x} = x^*$. Thus, the sequence $\{\mathbb{x}_k\}$ converges to x^*, and it follows that $\mathbb{x}^k$ is an interior point of the closed ball $B(x^*, \epsilon)$ for all k greater than some $\bar{k}$.

For $k > \bar{k}$, since $\mathbb{x}^k$ is an optimal solution of (P_k) and $\mathbb{x}^k$ is an interior point of the closed ball $B(x^*, \epsilon)$, we have by the necessary optimality condition in terms of limiting subdifferential in Proposition 2 (iv) that

$$0 \in \partial F^k(\mathbb{x}^k) + \mathcal{N}_{\mathcal{X}}(\mathbb{x}^k).$$

Applying the calculus rules in Proposition 2 (i),(iii) to $\partial F^k(\mathbb{x}^k)$ we have the existence of multipliers

$$\xi_i^k := k h_i(\mathbb{x}^k), \qquad \zeta_j^k := k g_j^+(\mathbb{x}^k) \tag{3.3}$$

such that

$$0 \in \partial f(\mathbb{x}^k) + \sum_{i=1}^{p} \partial(\xi_i^k h_i)(\mathbb{x}^k) + \sum_{j=1}^{q} \zeta_j^k \partial g_j(\mathbb{x}^k) + (\mathbb{x}^k - x^*) + \mathcal{N}_{\mathcal{X}}(\mathbb{x}^k). \tag{3.4}$$

Denote by

$$\delta^k := \sqrt{1 + \sum_{i=1}^{p}(\xi_i^k)^2 + \sum_{j=1}^{q}(\zeta_j^k)^2},$$
$$\mu_0^k := \frac{1}{\delta^k}, \quad \lambda_i^k := \frac{\xi_i^k}{\delta^k}, \quad i = 1, \ldots, p, \quad \mu_j^k := \frac{\zeta_j^k}{\delta^k}, \quad j = 1, \ldots, q. \tag{3.5}$$

Then since $\delta^k > 0$, dividing (3.4) by δ^k, we obtain for all $k > \bar{k}$,

$$\begin{aligned} 0 \in \mu_0^k \partial f(\mathbb{x}^k) &+ \sum_{i=1}^{p} \partial(\lambda_i^k h_i)(\mathbb{x}^k) + \sum_{j=1}^{q} \mu_j^k \partial g_j(\mathbb{x}^k) + \frac{1}{\delta^k}(\mathbb{x}^k - x^*) \\ &+ \mathcal{N}_{\mathcal{X}}(\mathbb{x}^k). \end{aligned} \tag{3.6}$$

Since by construction we have

$$(\mu_0^k)^2 + \sum_{i=1}^{p}(\lambda_i^k)^2 + \sum_{j=1}^{q}(\mu_j^k)^2 = 1 \tag{3.7}$$

the sequence $\{\mu_0^k, \lambda_1^k, \ldots, \lambda_p^k, \mu_1^k, \ldots, \mu_q^k\}$ is bounded and must contain a subsequence that converges to some limit $\{\mu_0^*, \lambda_1^*, \ldots, \lambda_p^*, \mu_1^*, \ldots, \mu_q^*\}$.

Since h_i is Lipschitz near x^*, we have

$$\begin{aligned}\partial(\lambda_i^k h_i)(\mathbb{x}^k) &\subset \partial[(\lambda_i^k - \lambda_i^*)h_i](\mathbb{x}^k) + \partial(\lambda_i^* h_i)(\mathbb{x}^k) \quad \text{by Proposition 2 (i)}\\ &\subset L_{h_i}|\lambda_i^k - \lambda_i^*| clB + \partial(\lambda_i^* h_i)(\mathbb{x}^k) \quad \text{by Proposition 1 (ii)},\end{aligned}$$

where L_{h_i} is the Lipschitz constant of h_i. Similarly,

$$\begin{aligned}\mu_0^k \partial f(\mathbb{x}^k) &\subset L_f|\mu_0^k - \mu_0^*| clB + \mu_0^* \partial f(\mathbb{x}^k),\\ \mu_j^k \partial g_j(\mathbb{x}^k) &\subset L_{g_j}|\mu_j^k - \mu_j^*| clB + \mu_j^* \partial g_j(\mathbb{x}^k),\end{aligned}$$

where L_f, L_{g_j} are the Lipschitz constants of f, g_j. Hence we have from (3.6) that

$$\begin{aligned}0 \in \mu_0^* \partial f(\mathbb{x}^k) + \sum_{i=1}^p \partial(\lambda_i^* h_i)(\mathbb{x}^k) + \sum_{j=1}^q \mu_j^* \partial g_j(\mathbb{x}^k) + \frac{1}{\delta^k}(\mathbb{x}^k - x^*)\\ +(L_f|\mu_0^k - \mu_0^*| + \sum_{i=1}^p L_{h_i}|\lambda_i^k - \lambda_i^*| + \sum_{j=1}^q L_{g_j}|\mu_j^k - \mu_j^*|) cl\mathbb{B} + \mathcal{N}_{\mathcal{X}}(\mathbb{x}^k).\end{aligned}$$

Taking limit as $k \to \infty$, by the definition of the limiting subdifferential and the limiting normal cone (or the fact ∂f is outer semicontinuous [29, Proposition 8.7]), we see that μ_0^*, λ_i^* and μ_j^* must satisfy condition (i). From (3.3) and (3.5), μ_0^* and μ_j^* must satisfy condition (ii) and from (3.7), μ_0^*, λ_i^* and μ_j^* must satisfy condition (iii).

Finally, to show that condition (iv) is satisfied, assume that $I \cup J$ is nonempty (otherwise there is nothing to prove). Since $\lambda_i^k \to \lambda_i^*$ as $k \to \infty$ and $\lambda_i^* \neq 0$ for $i \in I$, for sufficiently large k, λ_i^k have the same sign as λ_i^*. Hence we must have $\lambda_i^* \lambda_i^k > 0$ for all $i \in I$ and sufficiently large k. Similarly $\mu_j^* \mu_j^k > 0$ for all $j \in J$ and sufficiently large k. Therefore from (3.3) and (3.5) we must have $\lambda_i^* h_i(\mathbb{x}^k) > 0$ for all $i \in I$ and $\mu_j^* g_j(\mathbb{x}^k) > 0$ for all $j \in J$ and $k \geq K_0$ for some positive integer K_0. Consequently since $I \cup J$ is nonempty, it follows that there exists either $i \in I$ such that $h_i(\mathbb{x}^k) \neq 0$ or $j \in J$ such that $g_j(\mathbb{x}^k) \neq 0$ for all $k \geq K_0$ and hence from (3.1) we have $f(\mathbb{x}^k) < f(x^*)$ for all $k \geq K_0$. It remains to show the proximal subdifferentiability of the functions $f, h_i (i \in I), g_j (j \in J)$ at $\mathbb{x}^k$. By the density theorem in [11, Theorem 3.1], for each $\mathbb{x}^k$ with $k \geq K_0$, there exists a sequence $\{\mathbb{x}^{k,l}\} \subset \mathcal{X}$ with $\lim_{l \to \infty} \mathbb{x}^{k,l} = \mathbb{x}^k$ such that f, h_i, g_j are proximal subdifferentiable at $\mathbb{x}^{k,l}$. Since

$$f(\mathbb{x}^k) < f(x^*),\ \lambda_i^* h_i(\mathbb{x}^k) > 0, \forall i \in I, \quad \mu_j^* g_j(\mathbb{x}^k) > 0,\ \forall j \in J,$$

we have that and for all sufficiently large l,

$$f(\mathbb{x}^{k,l}) < f(x^*), \quad \lambda_i^* h_i(\mathbb{x}^{k,l}) > 0,\ \forall i \in I, \quad \mu_j^* g_j(\mathbb{x}^{k,l}) > 0,\ \forall j \in J.$$

For each $k \geq K_0$, choose an index l_k such that $l_1 < \ldots < l_{k-1} < l_k$ and

$$\lim_{k\to\infty} \mathbb{x}^{k,l_k} = x^*.$$

Consider the sequence $\{x^k\}$ defined by $x^k = \mathbb{x}^{(K_0+k),(l_{K_0+k})}$, $k = 1, 2, \ldots$. It follows from the preceding relations that $\{x^k\} \subset \mathcal{X}$,

$$\lim_{k\to\infty} x^k = x^*, \quad f(x^k) < f(x^*), \quad \lambda_i^* h_i(x^k) > 0, \forall i \in I, \quad \mu_j^* g_j(x^k) > 0, \forall j \in J,$$

and $f, h_i(i \in I)$, $g_j(j \in J)$ are all proximal subdifferentiable at x^k. ■

The condition (iv) is illustrated in Figure 1.

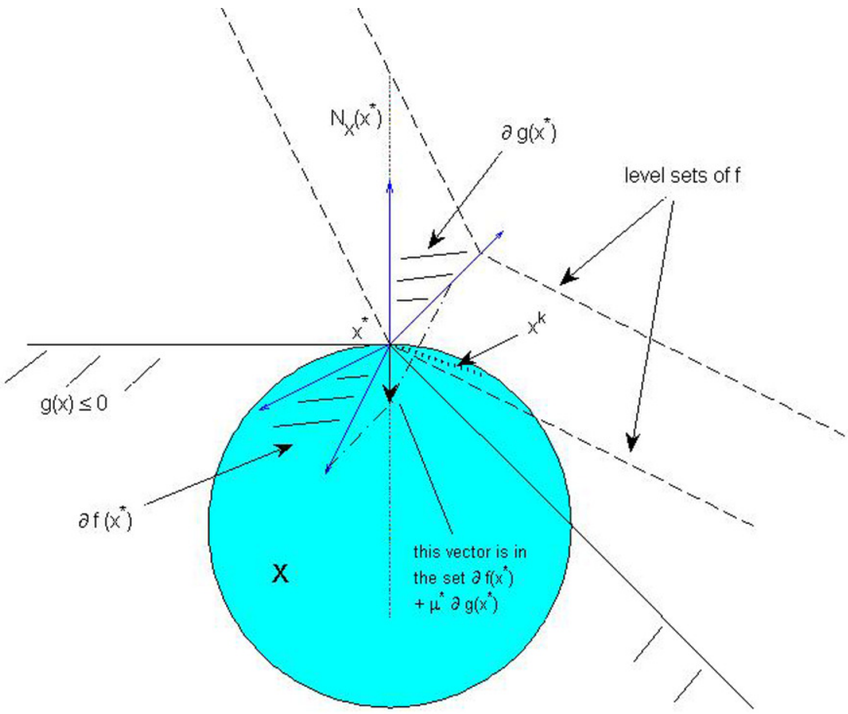

Fig. 1 Existence of μ^* and $\{x^k\}$

4 Enhanced KKT condition and weakened CQs

Based on the enhanced Fritz John condition, we define the following enhanced KKT condition.

Definition 4 (Enhanced KKT condition) *Let x^* be a feasible point of the problem (P). We say the enhanced KKT condition holds at x^* if the enhanced Fritz John condition holds with $\mu_0^* = 1$.*

Theorem 2 *Let x^* be a local minimum of problem (P). Suppose that there is no nonzero vector $(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}_+^q$ such that*

$$0 \in \sum_{i=1}^{p} \partial(\lambda_i h_i)(x^*) + \sum_{j=1}^{q} \mu_j \partial g_j(x^*) + \mathcal{N}_{\mathcal{X}}(x^*), \tag{4.1}$$

and the CV condition defined in [Theorem 1(iv)] hold. Then the enhanced KKT condition holds at x^*.

Proof Under the assumptions of the theorem, (i)-(iv) of Theorem 1 never hold if $\mu_0^* = 0$. Hence μ_0^* must be nonzero. The enhanced KKT condition then holds after a scaling. ■

Note that the condition in Theorem 4.1 is not a constraint qualification since it involves the objective function f. However Theorem 2 leads to the introduction of some constraint qualifications for a weaker version of the enhanced KKT condition to hold. In the smooth case, the pseudonormality and the quasinormality are slightly weaker than the original definitions introduced by Bertsekas and Ozdaglar [4].

Definition 5 *Let x^* be in the feasible region $\mathcal{C}$.*

(a) *x^* is said to satisfy NNAMCQ if there is no nonzero vector $(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}_+^q$ such that (4.1) and CS holds: $\mu_j g_j(x^*) = 0$ for all $j = 1, \ldots, q$.*

(b) *x^* is said to be pseudonormal (for the feasible region $\mathcal{C}$) if there is no vector $(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}_+^q$ and no infeasible sequence $\{x^k\} \subset \mathcal{X}$ converging to x^* such that (4.1) and the pseudo-complementary slackness condition (pseudo-CS for short) hold: if the index set $I \cup J$ is nonempty, where $I = \{i|\lambda_i \neq 0\}$, $J = \{j|\mu_j > 0\}$, then for each k*

$$\sum_{i=1}^{p} \lambda_i h_i(x^k) + \sum_{j=1}^{q} \mu_j g_j(x^k) > 0,$$

and $h_i(i \in I)$, $g_j(j \in J)$ are all proximal subdifferentiable at x^k for each k.

(c) *x^* is said to be quasinormal (for the feasible region $\mathcal{C}$) if there is no nonzero vector $(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}_+^q$ and no infeasible sequence $\{x^k\} \subset \mathcal{X}$ converging to x^* such that (4.1) and the quasi-complementary slackness condition (quasi-CS for short) hold: if the index set $I \cup J$ is nonempty, where $I = \{i|\lambda_i \neq 0\}$, $J = \{j|\mu_j > 0\}$, then for all $i \in I$, $j \in J$, $\lambda_i h_i(x^k) > 0$ and $\mu_j g_j(x^k) > 0$, and $h_i(i \in I)$, $g_j(j \in J)$ are all proximal subdifferentiable at x^k for each k.*

Since Quasi-CS $\Longrightarrow$ Pseudo-CS $\Longrightarrow$ CS, the following implications hold:

$$NNAMCQ \Longrightarrow Pseudonormality \Longrightarrow Quasinormality.$$

The first reverse implication is obviously not true. [4, Example 3] shows that the second reverse implication is not true either. We will show later that under the assumption that $\mathcal{N}_{\mathcal{X}}(x^*)$ is convex, quasinormality is in fact equivalent to a slightly weaker version of pseudonormality.

In [4, Proposition 3.1] Bertsekas and Ozadaglar showed that any feasible point of a constraint region where the equality functions are linear and inequality functions are concave and smooth and there is no abstract constraint must be pseudonormal. In what follows we extend it to the nonsmooth case.

Proposition 3 *Suppose that h_i are linear and g_j are concave and $\mathcal{X} = \mathbb{R}^m$. Then any feasible point of problem (P) is pseudonormal.*

Proof We prove it by contradiction. To the contrary, suppose that there is a feasible point x^* which is not pseudonormal. Then there exists nonzero vector $(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}^q_+$ and a sequence $\{x^k\} \subset \mathcal{X}$ converging to x^* such that (4.1) and the following condition hold: for each k

$$\sum_{i=1}^{p} \lambda_i h_i(x^k) + \sum_{j=1}^{q} \mu_j g_j(x^k) > 0. \tag{4.2}$$

By the linearity of h_i and concavity of g_j, we have that for all $x \in \mathbb{R}^m$,

$$\begin{aligned}
&h_i(x) = h_i(x^*) + \nabla h_i(x^*)^T(x - x^*) \quad i = 1, \ldots, p,\\
&g_j(x) \leq g_j(x^*) + \xi_j^T(x - x^*) \quad \forall \xi_j \in \partial g_j(x^*), j = 1, \ldots, q.
\end{aligned}$$

By multiplying these two relations with λ_i and μ_j and by adding over i and j, respectively, we obtain that for all $x \in \mathbb{R}^m$ and all $\xi_j \in \partial g_j(x^*), j = 1, \ldots, q$,

$$\begin{aligned}
&\sum_{i=1}^{p} \lambda_i h_i(x) + \sum_{j=1}^{q} \mu_j g_j(x)\\
&\leq \sum_{i=1}^{p} \lambda_i h_i(x^*) + \sum_{j=1}^{q} \mu_j g_j(x^*) + [\sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{q} \mu_j \xi_j]^T(x - x^*)\\
&= [\sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{q} \mu_j \xi_j]^T(x - x^*)
\end{aligned}$$

where the last equality holds because we have

$$\lambda_i h_i(x^*) = 0 \text{ for all } i \text{ and } \sum_{j=1}^{q} \mu_j g_j(x^*) = 0.$$

By (4.1), since $\mathcal{N}_{\mathbb{R}^m}(x^*) = \{0\}$ there exists $\xi_j^* \in \partial g_j(x^*), j = 1, \ldots, q$ such that

$$\sum_{i=1}^{p} \lambda_i \nabla h_i(x^*) + \sum_{j=1}^{q} \mu_j \xi_j^* = 0.$$

Hence it follows that for all $x \in \mathbb{R}^m$,

$$\sum_{i=1}^{p} \lambda_i h_i(x) + \sum_{j=1}^{q} \mu_j g_j(x) \le 0$$

which contradicts (4.2). Hence the proof is complete. ■

Definition 6 *Let x^* be a feasible point of problem (P). We call a vector $(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}^q_+$ satisfying the following weaker version of the enhanced KKT conditions a quasinormal multiplier:*

(i) $0 \in \partial f(x^*) + \sum_{i=1}^{p} \partial(\lambda_i^* h_i)(x^*) + \sum_{j=1}^{q} \mu_j^* \partial g_j(x^*) + \mathcal{N}_{\mathcal{X}}(x^*)$.

(ii) *There exists a sequence $\{x^k\} \subset \mathcal{X}$ converging to x^* such that the quasi-CS as defined in Definition 5 holds.*

Since the only difference of the quasinormality with the sufficient condition given in Theorem 2 is the condition $f(x^k) < f(x^*)$, it is obvious that the quasinormality is a constraint qualification for the weaker version of the enhanced KKT condition to hold and hence the following result follows immediately from Theorem 2 and the definitions of the three constraint qualifications.

Corollary 1 *Let x^* be a local minimizer of problem (P). Then if x^* satisfies NNAMCQ, or is pseudonormal, or is quasinormal, then the weaker version of the enhanced KKT condition holds at x^*.*

It is known that NNAMCQ implies the boundedness of the set of all normal multipliers (see e.g. [15]). In what follows, we show that the set of all quasinormal multipliers are bounded under the quasinormality condition.

Theorem 3 *Let x^* be a feasible point for problem (P). If quasinormality holds at x^*, then the set of all quasinormal multipliers $M_Q(x^*)$ is bounded.*

Proof To the contrary, suppose that $M_Q(x^*)$ is unbounded. Then there exists $(\lambda^n, \mu^n) \in M_Q(x^*)$ such that $\|(\lambda^n, \mu^n)\| \to \infty$ as n tends to infinity. By definition of a quasinormal multiplier, for each n, there exists a sequence $\{x_n^k\}_k \subset \mathcal{X}$ converging to x^* such that

$$0 \in \partial f(x^*) + \sum_{i=1}^{p} \partial(\lambda_i^n h_i)(x^*) + \sum_{j=1}^{q} \mu_j^n \partial g_j(x^*) + \mathcal{N}_{\mathcal{X}}(x^*), \tag{4.3}$$

$$\mu_j^n \ge 0, \quad \forall j = 1, \ldots, q, \tag{4.4}$$

$$\lambda_i^n h_i(x_n^k) > 0 \ \forall i \in I^n, \ \mu_j^n g_j(x_n^k) > 0 \ \forall j \in J^n, \tag{4.5}$$

$$h_i (i \in I^n), g_j (j \in J^n) \text{ are proximal subdifferential at } x_n^k, \tag{4.6}$$

where $I^n := \{i : \lambda_i^n \neq 0\}$ and $J^n := \{j : \mu_j^n > 0\}$.

Denote by $\xi^n := \frac{\lambda^n}{\|(\lambda^n, \mu^n)\|}$ and $\zeta^n := \frac{\mu^n}{\|(\lambda^n, \mu^n)\|}$. Assume without loss of generality that $(\xi^n, \mu^n) \to (\xi^*, \mu^*)$. Divide both sides of (4.3) by $\|(\lambda^n, \mu^n)\|$ and take the limit, we have

$$0 \in \sum_{i=1}^{p} \partial(\xi_i^* h_i)(x^*) + \sum_{j=1}^{q} \zeta_j^* \partial g_j(x^*) + \mathcal{N}_{\mathcal{X}}(x^*).$$

It follow from (4.4) that $\zeta_j^* \geq 0$, for all $j = 1, \dots, q$. Finally, let

$$I = \{i : \xi_i^* \neq 0\}; \qquad J = \{j : \zeta_j^* > 0\}.$$

Then $I \cup J$ is nonempty. By virtue of (4.5), there are some N_0 such that for $n > N_0$, we must have $\xi_i^* h_i(x_n^k) > 0$ for all $i \in I$ and $\zeta_j^* g_j(x_n^k) > 0$ for all $j \in J$. Moreover by (4.6), $h_i(i \in I^n), g_j(j \in J^n)$ are proximal subdifferential at x_n^k. Thus there exist scalars $\{\xi_1^*, \dots \xi_p^*, \zeta_1^*, \dots, \zeta_q^*\}$ not all zero and a sequence $\{x_n^k\} \subset \mathcal{X}$ that satisfy the preceding relation an so violate the quasinormality of x^*. Hence the proof is complete. ■

Combining the proof techniques of Theorem 1 and [5, Lemma 2] in the following proposition we can extend [5, Lemma 2] to our nonsmooth problem. We omit the proof here.

Lemma 1 *If a vector $x^* \in \mathcal{C}$ is quasinormal, then all feasible vectors in a neighborhood of x^* are quasinormal.*

In the following result we obtain a specific representation of the limiting normal cone to the constraint region in terms of the set of quasinormal multipliers. Note that our result is sharper than the result of Bertsekas and Ozdaglar [5, Proposition 1] which gives a representation of the Fréchet normal cone in terms of the set of quasinormal multipliers for the case of smooth problems with a closed abstract set constraint. The result is also sharper than the one given by Henrion, Jourani and Outrata [12, Theorem 4.1] in which the representation is given in terms of the usual normal multipliers.

Proposition 4 *If $\bar{x}$ is quasinormal for $\mathcal{C}$, then*

$$\mathcal{N}_{\mathcal{C}}(\bar{x}) \subset \left\{ \sum_{i=1}^{p} \partial(\lambda_i h_i)(\bar{x}) + \sum_{j=1}^{q} \mu_j \partial g_j(\bar{x}) + \mathcal{N}_{\mathcal{X}}(\bar{x}) : (\lambda, \mu) \in M_Q(\bar{x}) \right\}.$$

Proof Let v be a vector that belongs to $\mathcal{N}_{\mathcal{C}}(\bar{x})$. Then by definition, there are sequences $x^l \to \bar{x}$ and $v^l \to v$ with $v^l \in \mathcal{N}_{\mathcal{C}}^F(x^l)$ and $x^l \in \mathcal{C}$.

Step 1. By Lemma 1, for l sufficiently large, x^l is quasinormal for $\mathcal{C}$. By [29, Theorem 6.11], for each l there exists a smooth function φ^l that achieves a strict global minimum over $\mathcal{C}$ at x^l with $-\nabla\varphi^l(x^l) = v^l$. Since x^l is a quasinormal vector of $\mathcal{C}$, by Theorem 2, the weaker version of the enhanced KKT condition holds for problem

$$\min \varphi^l(x) \quad \text{ s.t. } x \in \mathcal{C}.$$

That is, there exists a vector $(\lambda^l, \mu^l) \in \mathbb{R}^p \times \mathbb{R}_+^q$ such that

$$v^l \in \sum_{i=1}^{p} \partial(\lambda_i^l h_i)(x^l) + \sum_{j=1}^{q} \mu_j^l \partial g_j(x^l) + \mathcal{N}_{\mathcal{X}}(x^l) \tag{4.7}$$

and a sequence $\{x^{l,k}\} \subset \mathcal{X}$ converging to x^l as $k \to \infty$ such that for all k, $\lambda_i^l h_i(x^{l,k}) > 0, \forall i \in I^l, \mu_j^l g_j(x^{l,k}) > 0, \forall j \in J^l$, and $h_i(i \in I^l)$, $g_j(j \in J^l)$ are proximal subdifferentiable at $x^{l,k}$, where $I^l = \{i : \lambda_i^l \neq 0\}, J^l = \{j : \mu_j^l > 0\}$.

Step 2. We show that the sequence $\{\lambda_1^l, \dots, \lambda_p^l, \mu_1^l, \dots, \mu_q^l\}$ is bounded. To the contrary suppose that the sequence $\{\lambda_1^l, \dots, \lambda_p^l, \mu_1^l, \dots, \mu_q^l\}$ is unbounded. For every l, denote

$$\delta^l = \sqrt{1 + \sum_{i=1}^p (\lambda_i^l)^2 + \sum_{j=1}^q (\mu_j^l)^2},\ \xi_i^l = \frac{\lambda_i^l}{\delta^l},\ i = 1, \dots, p,\ \zeta_j^l = \frac{\mu_j^l}{\delta^l}, \quad j = 1, \dots, q.$$

Then from (4.7) it follows that

$$\frac{v^l}{\delta^l} \in \sum_{i=1}^p \partial(\xi_i^l h_i)(x^l) + \sum_{j=1}^q \zeta_j^l \partial g_j(x^l) + \mathcal{N}_{\mathcal{X}}(x^l).$$

Since the sequence $\{\xi_1^l, \dots, \xi_p^l, \zeta_1^l, \dots, \zeta_q^l\}$ is bounded, for the sake of simplicity, we may assume that $\{\xi_1^l, \dots, \xi_p^l, \zeta_1^l, \dots, \zeta_q^l\} \to \{\xi_1^*, \dots, \xi_p^*, \zeta_1^*, \dots, \zeta_q^*\} \neq 0$ as $l \to \infty$. Taking limits in the above inclusion, similar to the proof of Theorem 1 we obtain

$$0 \in \sum_{i=1}^p \partial(\xi_i^* h_i)(\bar{x}) + \sum_{j=1}^q \zeta_j^* \partial g_j(\bar{x}) + \mathcal{N}_{\mathcal{X}}(\bar{x}),$$

where $\zeta_j^* \geq 0$ for all $j = 1, \dots, q$ and $\xi_1^*, \dots, \xi_p^*, \zeta_1^*, \dots, \zeta_q^*$ are not all zero. Let $i \in I^* := \{i : \xi_i^* \neq 0\}$. Since $\xi_i^l \to \xi_i^* \neq 0$ as $l \to \infty$, $\xi_i^l \neq 0$ and has the same sign as ξ_i^* for sufficiently large l. Consequently since $\xi_i^l h_i(x^{l,k}) > 0$ we have also $\xi_i^* h_i(x^{l,k}) > 0$ for all sufficiently large l and all k. Similarly let $j \in J^* := \{j : \zeta_j^* > 0\}$, we have $\zeta_j^* g_j(x^{l,k}) > 0$. Also similar to the proof of Theorem 1, by using the density theorem we can find a subsequence $\{x^{l,k_l}\} \subset \{x^{l,k}\} \subset \mathcal{X}$ converging to $\bar{x}$ as $l \to \infty$ such that for all sufficiently large l,

$$\xi_i^* h_i(x^{l,k_l}) > 0 \quad \forall i \in I^*, \qquad \zeta_j^* g_i(x^{l,k_l}) > 0 \ \forall j \in J^*$$

and $h_i(x^{l,k_l})(i \in I^*), g_j(x^{l,k_l})(j \in J^*)$ are proximal subdifferentiable at x^{l,k_l}. But this is impossible since $\bar{x}$ is assumed to be quasinormal and hence the sequence $\{\lambda_1^l, \dots, \lambda_p^l, \mu_1^l, \dots, \mu_q^l\}$ must be bounded.

Step 3. By virtue of Step 2, without loss of generality, we assume that $\{\lambda_1^l, \dots, \lambda_p^l, \mu_1^l, \dots, \mu_q^l\}$ converges to $\{\lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_q\}$ as $l \to \infty$. Taking the limit in (4.7) as $l \to \infty$, we have

$$v \in \sum_{i=1}^p \partial(\lambda_i h_i)(\bar{x}) + \sum_{j=1}^q \mu_j \partial g_j(\bar{x}) + \mathcal{N}_{\mathcal{X}}(\bar{x}).$$

Similar to Step 2, we can find a subsequence $\{x^{l,k_l}\} \subset \{x^{l,k}\} \subset \mathcal{X}$ converging to $\bar{x}$ as $l \to \infty$ such that for all sufficiently large l, $\lambda_i h_i(x^{l,k_l}) > 0, \forall i \in I,\ \mu_j g_j(x^{l,k_l}) > 0, \forall j \in J$, and $h_i(i \in I)$, $g_j(j \in J)$ are proximal subdifferentiable at x^{l,k_l}, where $I = \{i : \lambda_i \neq 0\}$ and $J = \{j : \mu_j > 0\}$. ∎

From Propositions 4 and 2 (v), the following enhanced KKT necessary optimality condition for the case where the objective function is Fréchet differentiable (but may not be Lipschitz) follows immediately. Note that for a Fréchet differentiable function which is not Lipschitz continuous, the limiting subdifferential may not coincide with the usual gradient and hence the following result provides a sharper result for this case.

Corollary 2 *Let x^* be a local minimizer of problem (P) where the objective function f is Fréchet differentiable at x^*. If x^* either satisfies NNAMCQ, is pseudonormal, or is quasinormal, then the weaker version of the enhanced KKT condition holds.*

We close this section with a result showing that quasinormality and a weaker version of pseudonormality coincide under the condition that the normal cone is convex and the constraint functions are strictly differentiable at the point x^*. This result is an extension of a similar result of Bertsekas and Ozdaglar [4, Proposition 3.2] in that we do not require the function to be continuously differentiable at x^*.

Proposition 5 *Let $x^* \in \mathcal{C}$. Assume that for each $i = 1, \ldots, p$, $j = 1, \ldots, q$, $h_i(x)$, $g_j(x)$ are strictly differentiable at x^*, and the limiting normal cone $\mathcal{N}_{\mathcal{X}}(x^*)$ is convex. Then x^* is quasinormal if and only if the following weaker version of pseudonormality holds: there are no vector $(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}^q_+$ and no sequence $\{x^k\} \subset \mathcal{X}$ converging to x^* such that*

(i) $0 \in \sum_{i=1}^p \lambda_i \nabla h_i(x^*) + \sum_{j=1}^q \mu_j \nabla g_j(x^*) + \mathcal{N}_{\mathcal{X}}(x^*)$.
(ii) $\lambda_i h_i(x^k) \geq 0$ *for all* i *and* $\mu_j g_j(x^k) \geq 0$ *for all* j, *and if the index sets* $I \cup J \neq \emptyset$ *where* $I = \{i|\lambda_i \neq 0\}$ $J = \{j|\mu_j > 0\}$ *then*

$$\sum_{i=1}^p \lambda_i h_i(x^k) + \sum_{j=1}^q \mu_j g_j(x^k) > 0, \qquad \forall k$$

and $h_i (i \in I)$, $g_j (j \in J)$ *are proximal subdifferentiable at* x^k.

Proof It is easy to see that the weaker version of pseudonormality implies the quasinormality. So what we have to do is to show the converse. To the contrary, suppose that the quasinormality holds but the weaker version of pseudonormality does not hold. Then there exist scalars $\lambda_1, \ldots, \lambda_p, \mu_1, \ldots, \mu_q$ and a sequence $\{x^k\} \subset \mathcal{X}$ converging to x^* such that (i)-(ii) hold. Condition (ii) implies that $\lambda_i h_i(x^k) > 0$ for some $\bar{i}$ such that $\lambda_{\bar{i}} \neq 0$ or $\mu_j g_j(x^k) > 0$ for some $\bar{j}$ such that $\mu_{\bar{j}} > 0$. We now suppose that such $\bar{j}$ exists (the case where $\bar{j}$ does not exist but $\bar{i}$ exists can be similarly proved and we omit it here). Without loss of generality, we can assume $\bar{j} = 1$ and $\mu_1 = 1$ (otherwise we can normalize it) such that (i) holds:

$$-\Big(\nabla g_1(x^*) + \sum_{i=1}^p \lambda_i \nabla h_i(x^*) + \sum_{j=2}^q \mu_j \nabla g_j(x^*)\Big) \in \mathcal{N}_{\mathcal{X}}(x^*). \tag{4.8}$$

Since $g_1(x^k) > 0$ for all k, $\mu_2, \ldots, \mu_q, \lambda_1, \ldots, \lambda_p$ are not all equal to 0, otherwise it would contradicts the quasinormality of x^*. Besides, because $\mathcal{N}_{\mathcal{X}}(x^*)$ is closed and convex, by [4, Lemma 2.2] there exists a vector $\bar{d} \in \mathcal{N}_{\mathcal{X}}(x^*)^*$ with $\langle \bar{d}, \nabla g_1(x^*)\rangle < 0$, $\langle \bar{d}, \nabla g_j(x^*)\rangle > 0$ for all $j = 2, \ldots, q$, such that $\mu_j > 0$ and $\langle \bar{d}, \lambda_i \nabla h_i(x^*)\rangle > 0$ for all $i = 1, \ldots, p$, such that $\lambda_i \neq 0$.

In the remaining part of the proof, we show that the scalars $\mu_1 = 1, \mu_2, \ldots, \mu_q, \lambda_1, \ldots, \lambda_p$ achieved above satisfy condition: $\lambda_i h_i(x^k) > 0 \quad \forall i \in I := \{i : \lambda_i \neq 0\}$, $\mu_j g_j(x^k) > 0 \quad \forall j \in J := \{j = 2, \ldots, q : \mu_j > 0\}$ which would contradicts the fact that x^* is quasinormal. Since g_j and h_i are strictly differentiable at x^*, the gradients coincide with the limiting subdifferentials, i.e.,

$$\nabla(\mu_j g_j)(x^*) = \lim_{k\to\infty} \rho_j^k \text{ for some } \rho_j^k \in \partial^\pi(\mu_j g_j)(x^k)$$
$$\nabla(\lambda_i h_i)(x^*) = \lim_{k\to\infty} \varrho_j^k \text{ for some } \varrho_i^k \in \partial^\pi(\lambda_i h_i)(x^k).$$

By [29, Theorem 6.26, Theorem 6.28], for vector $\bar{d} \in \mathcal{N}_{\mathcal{X}}(x^*)^*$ and the sequence x^k converging to x^* constructed above, there is a sequence $d^k \in \mathcal{T}_{\mathcal{X}}(x^k)$ such that $d^k \to \bar{d}$. By virtue of $x^k \to x^*$, $d^k \to \bar{d}$ and $\langle \bar{d}, \nabla(\mu_j g_j)(x^*)\rangle > 0$ for all $j = 2, \ldots, q$, with $\mu_j > 0$, $\langle \bar{d}, \nabla(\lambda_i h_i)(x^*)\rangle > 0$ for all $i = 1, \ldots, p$, with $\lambda_i \neq 0$, we have that, for all sufficiently large k, $\langle d^k, \rho_j^k\rangle > 0$ for all $j = 2, \ldots, q$, with $\mu_j > 0$, $\langle d^k, \varrho_i^k\rangle > 0$ for all $i = 1, \ldots, p$, with $\lambda_i \neq 0$. Since $d^k \in \mathcal{T}_{\mathcal{X}}(x^k)$, there exists a sequence $\{x^{k,l}\} \in \mathcal{X}$ such that, for each k, we have $x^{k,l} \neq x^k$ for all l and

$$x^{k,l} \to x^k, \qquad \frac{x^{k,l} - x^k}{\|x^{k,l} - x^k\|} \to \frac{d^k}{\|d^k\|}, \quad as\ l \to \infty,$$

h_i, g_j are proximal subdifferentiable at $x^{k,l}$. Since $\rho_j^k \in \partial^\pi(\mu_j g_j)(x^k) \subset \partial^F(\mu_j g_j)(x^k)$, by definition of the Fréchet subdifferential, for some vector sequence υ converging to 0, and for each $j = 2, \ldots, q$, with $\mu_j > 0$,

$$\begin{aligned}\mu_j g_j(x^{k,l}) &\geq \mu_j g_j(x^k) + \langle x^{k,l} - x^k, \rho_j^k\rangle + o(\|x^{k,l} - x^k\|)\\ &\geq \mu_j \langle \frac{d^k}{\|d^k\|} + \upsilon, \rho_j^k\rangle \|x^{k,l} - x^k\| + o(\|x^{k,l} - x^k\|)\end{aligned}$$

where the second inequality above follows from the assumption that $\mu_j g_j(x^k) \geq 0$, for all j and x^k. It follows that, for l and k sufficiently large, there exists $x^{k,l} \in \mathcal{X}$ arbitrary close to x^k such that $\mu_j g_j(x^{k,l}) > 0$ and, g_j are proximal subdifferentiable at $x^{k,l}$ for all $j = 2, \ldots, q$, with $\mu_j > 0$. Similarly, for l and k sufficiently large, there exists $x^{k,l} \in \mathcal{X}$ arbitrary close to x^k such that $\lambda_i h_i(x^{k,l}) > 0$ and, h_i are proximal subdifferentiable at $x^{k,l}$ for all $i = 1, \ldots, p$ with $\lambda_i \neq 0$. ■

5 Sufficient conditions for error bounds

In this section we show that either pseudonormality or quasinormality plus the subdifferential regularity condition on constraints implies the existence of local error bounds. Our results are new even for the smooth case.

In order to derive the desired error bound formula (1.5), let us first rewrite the constraint region (1.1) equivalently as follows:

$$\mathcal{C} = \{x \in \mathcal{X} : \|h(x)\|_1 + \|g^+(x)\|_1 = 0\}. \tag{5.1}$$

By [31, Theorem 3.3], to prove the desired error bound result we only need to derive the following estimation.

Lemma 2 *Let x^* be feasible for problem (P) such that pseudonormality holds. Then there are $\delta, c > 0$ such that*

$$\frac{1}{c} \leq \|\xi\|_1 \qquad \forall \xi \in \partial^\pi(\|h(x)\|_1 + \|g^+(x)\|_1 + \delta_{\mathcal{X}}(x)), x \in \mathbb{B}(x^*, \delta/2) \cap \mathcal{X}, x \notin \mathcal{C},$$

where $\delta_{\mathcal{C}}(x)$ denotes the indicator function of the set $\mathcal{C}$ at x.

Proof To the contrary, assume that there exists a sequence $\{x^k\} \to x^*$ with $x^k \in \mathcal{X} \setminus \mathcal{C}$ and $\xi^k \in \partial^\pi(\|h\|_1 + \|g^+\|_1 + \delta_{\mathcal{X}})(x^k)$ for all $k \in \mathbb{N}$ such that $\|\xi^k\|_1 \to 0$. By the calculus rule in Proposition 2 (i), (iii) and Proposition 1 (iv), we can find bounded multipliers (μ^k, λ^k) with $\mu^k \geq 0$ such that

$$\xi^k \in \sum_{i=1}^{q} \partial(\lambda_i^k h_i)(x^k) + \sum_{j=1}^{q} \mu_j^k \partial g_j(x^k) + \mathcal{N}_{\mathcal{X}}(x^k) \tag{5.2}$$

for all $k \in \mathbb{N}$. Hence, we may assume without loss of generality that it converges to a limit (λ, μ). Taking the limit as $k \to \infty$ in (5.2) yields

$$0 \in \sum_{i=1}^{p} \partial(\lambda_i h_i)(x^*) + \sum_{j=1}^{q} \mu_j \partial g_j(x^*) + \mathcal{N}_{\mathcal{X}}(x^*).$$

In addition, by the existence of λ_i^k, μ_j^k and Proposition 1 (iv), for k large enough, it is easy to see that

$$\lambda_i h_i(x^k) \geq 0 \ \forall i = 1, \ldots, p, \quad \mu_j g_j(x^k) \geq 0 \ \forall j = 1, \ldots, q.$$

Since $x^k \notin \mathcal{C}$ for all k, at least one functional constraint has to be violated infinitely many times. Using again Proposition 1 (iv), it is easy to see that there exists at least one multiplier λ_i or μ_j not equal to zero, and the corresponding product is strictly positive for all k such that the constraint is violated, i.e. if constraint $h_i(x^k) = 0$ is violated for infinitely many k, we may have $\lambda_i \neq 0$ and $\lambda_i h_i(x^k) > 0$ for all those k, if the constraint $g_j(x^k) \leq 0$ is violated for infinitely many k, we may have $\mu_j > 0$ and $\mu_j g_j(x^k) > 0$ for all those k. Therefore

$$\sum_{i=1}^{p} \lambda_i h_i(x^k) + \sum_{j=1}^{q} \mu_j g_j(x^k) > 0$$

at least on a subsequence. Moreover by the density theorem $h_i (i \in I), g_j (j \in J)$ can be selected to be proximal subdifferentiable at x^k. This, however, implies that pseudonormality is violated in x^* since x^k is chosen from $\mathcal{X}$, a contradiction. ∎

Using the local error bound result of [31, Theorem 3.3], we obtain the following error bound result.

Theorem 4 *Let x^* be feasible for problem (P) such that pseudonormality holds. Then the local error bound holds: there exist positive constants c and δ such that*

$$d_{\mathcal{C}}(x) \leq c(\|h(x)\|_1 + \|g^+(x)\|_1) \quad \forall x \in \mathbb{B}(x^*, \delta) \cap \mathcal{X}.$$

By Clarke's exact penalty principle [10, Proposition 2.4.3] we obtain the following exact penalty result immediately.

Corollary 3 *Let x^* be a local minimizer of problem (P). If pseudonormality holds at x^*, then x^* is a local minimizer of the penalized problem:*

$$\begin{aligned} \min \ & f(x) + \alpha(\|h(x)\|_1 + \|g^+(x)\|_1) \\ s.t. \ & x \in \mathcal{X}, \end{aligned}$$

where $\alpha \geq L_f c$, L_f is the Lipschitz constant of f and c is the error bound constant.

Notice that Corollary 3 even works for nonstrict local minima x^* in the nonsmooth case. However, we find that the exact penalty result in [4, Proposition 4.2], established in the smooth case, requires x^* to be a strict local minimum, and it is stated in [4, Example 7.7] that this assumption might be crucial. The example is the following:

$$\min \ f(x_1, x_2) := x_2 \ \text{ s.t. } h(x_1, x_2) := x_2/(x_1^2 + 1) = 0.$$

The feasible points are of the form $(x_1, 0)$ with $x_1 \in \mathbb{R}$. And each feasible point is a local minimum. Since the gradient $\nabla h(x_1, x_2)$ is nonzero, every feasible point is quasinormal. The authors claim that pseudonormality at a nonstrict local minimum may not imply the exact penalty since for any $c > 0$,

$$\inf_{(x_1,x_2)\in\mathbb{R}^2} \{x_2 + c|x_2|/|x_1^2 + 1|\} = -\infty,$$

which shows that a local optimal solution of the original problem is not a global optimal solution of the penalized problem. However this example is not a counter example since $(x_1, 0)$ is a local minimum of the function $x_2 + c|x_2|/|x_1^2 + 1|$ for large enough $c > 0$. Since the limiting subdifferential agrees with the classical gradient when a function is strictly differentiable, we stress, that not only did we extend the exact penalty result in [4] to a more general case, but we also improved their result in another way. We have now answered positively the open question raised in [16] in which the authors ask whether or not the proof technique based on error bound and the exact penalty principle of Clarke (which is completely different from the one used in [4]) can be used to prove the exact penalty result in [4] with a nonstrict local optimum.

The following example shows that the converse of Theorem 3 does not hold. Since when the objective function is Lipschitz continuous, the existence of an exact penalty function implies the exact penalty. It also shows that the existence of an exact penalty function does not imply pseudonormality.

Example 1 *Consider the locally Lipschitz optimization problem*

$$\begin{aligned} \min \quad & f(x) = |x_1| + |x_2| \\ s.t. \quad & g(x) = |x_1| - x_2 \leq 0 \\ & x \in \mathcal{X} := \{(x_1, x_2) : x_1^2 + (x_2 + 1)^2 \leq 1\}. \end{aligned}$$

At the only feasible point $x^* = (0,0)$, $\partial g(x^*) = \{(\zeta, -1) | -1 \leq \zeta \leq 1\}$ and $\mathcal{N}_{\mathcal{X}}(x^*) = \{t(0,1) | t \geq 0\}$. However, if we choose $\mu = 1$ and a sequence $\{x^k\}$ located in $\mathcal{X}$ where for each $k = 1, 2, \ldots$, $x^k = (\cos(\frac{\pi}{2} - \frac{\pi}{2k}), -1 + \sin(\frac{\pi}{2} - \frac{\pi}{2k}))$, we have $0 \in \mu \partial g(x^*) + \mathcal{N}_{\mathcal{X}}(x^*)$ and $\mu g(x^k) > 0$ for all $k > 1$. This implies x^* is not pseudonormal. However it is easy to see that the error bound holds:

$$d_{\mathcal{C}}(x) \leq |x_1| - x_2 = |x_1| + |x_2| \qquad \forall x \in \mathcal{X},$$

where $\mathcal{C} = \{(0,0)\}$.

Naturally, after showing that pseudonormality implies the existence of local error bound, we would like to explore the relation between quasinormality and the error bound property. In [22, Theorem 2.1], under the assumption that the constraint functions are C^{1+}, Minchenko and Tarakanov show that quasinormality implies the error bound for a smooth optimization problem with $\mathcal{X} = \mathbb{R}^m$. In what follows, we will show that quasinormality implies the error bound property for our nonsmooth optimization problem (P) under the condition that the constraint functions are subdifferential regular and the abstract constraint set is regular. Since a smooth function must be subdifferentially regular, our results show that the condition of C^{1+} for the constraint functions in Minchenko and Tarakanov [22, Theorem 2.1] can be removed.

Theorem 5 *Assume in the constraint system (1.1) that $\mathcal{X}$ is a nonempty closed regular set. Further let $x^* \in \mathcal{C}$, assume $h_i(x)$ are continuously differentiable, $g_j(x)$ are subdifferentially regular around x^* (automatically holds when g_j are convex or C^1 around x^*). If x^* is a quasinormal point of $\mathcal{C}$, then there exist positive numbers c and δ, such that*

$$d_{\mathcal{C}}(x) \leq c(\|h(x)\|_1 + \|g^+(x)\|_1) \quad \forall x \in \mathbb{B}(x^*, \delta) \cap \mathcal{X}. \tag{5.3}$$

Proof By assumption we can find $\delta_0 > 0$ such that $h_i(x), g_j(x)$ are subdifferentially regular for all $x \in \mathbb{B}(x^*, \delta_0)$. Since the required assertion is always true if $x^* \in int\mathcal{C}$, we only need to consider the case when $x^* \in \partial\mathcal{C}$. In this case, (5.3) can be violated only for $x \notin \mathcal{C}$. Let us take some sequences $\{\mathbb{x}^k\}$ and $\{x^k\}$, such that $\mathbb{x}^k \to x^*$,$\mathbb{x}^k \in \mathcal{X} \setminus \mathcal{C}$, and $x^k = \prod_{\mathcal{C}}(\mathbb{x}^k)$, the projection of $\mathbb{x}^k$ onto the set $\mathcal{C}$. Note that $x^k \to x^*$, since $\|x^k - \mathbb{x}^k\| \leq \|\mathbb{x}^k - x^*\|$. For simplicity we may assume both $\{\mathbb{x}^k\}$ and $\{x^k\}$ belong to $\mathbb{B}(x^*, \delta_0) \cap \mathcal{X}$.

Since $\mathbb{x}^k - x^k \in \mathcal{N}_{\mathcal{C}}^{\pi}(x^k) \subset \mathcal{N}_{\mathcal{C}}^F(x^k)$, we have

$$\eta^k = \frac{\mathbb{x}^k - x^k}{\|\mathbb{x}^k - x^k\|} \in \mathcal{N}_{\mathcal{C}}^F(x^k).$$

Since x^* is quasinormal, from Lemma 1 it follows that the point x^k is also quasinormal for all sufficiently large k and, without loss of generality, we may assume that all x^k are quasinormal. Then, by Proposition 4 there exists a sequence $\{\xi_1^k, \ldots, \xi_p^k, \zeta_1^k, \ldots, \zeta_q^k\}$ with $\zeta_j^k \geq 0$, such that

$$\eta^k \in \sum_{i=1}^p \xi_i^k \nabla h_i(x^k) + \sum_{j=1}^q \zeta_j^k \partial g_j(x^k) + \mathcal{N}_{\mathcal{X}}(x^k), \tag{5.4}$$

and there exists a sequence $\{x^{k,l}\} \subset \mathcal{X}$, such that $x^{k,l} \to x^k$ as $l \to \infty$ and for all $l = 1, 2, \ldots,$ $\xi_i^k h_i(x^{k,l}) > 0$ for $i \in I^k$; $\zeta_j^k g_j(x^{k,l}) > 0$ for $j \in J^k$, where $I^k = \{i : \xi_i^k \neq 0\}$ and $J^k = \{j : \zeta_j^k > 0\}$. As in the proof of Step 2 in Proposition 4, we can show that the quasinormality of x^* implies that the sequence $\{\xi_1^k, \ldots, \xi_p^k, \zeta_1^k, \ldots, \zeta_q^k\}$ is bounded. Therefore, without loss of generality, we may assume $\{\xi_1^k, \ldots, \xi_p^k, \zeta_1^k, \ldots, \zeta_q^k\}$ converges to some vector $\{\xi_1^*, \ldots, \xi_p^*, \zeta_1^*, \ldots, \zeta_q^*\}$. Then there exists a number $M_0 > 0$, such that for all k, $\|(\xi^k, \zeta^k)\| \leq M_0$.

Without loss of any generality, we may assume that $\mathbb{x}^k \in \mathbb{B}(x^*, \frac{\delta_0}{2}) \cap \mathcal{X} \setminus \mathcal{C}$ and $x^k \in \mathbb{B}(x^*, \delta_0) \cap \mathcal{X}$ for all k. Setting $(\bar{\xi}^k, \bar{\zeta}^k) = 2(\xi^k, \zeta^k)$, then from (5.4) for each k there exist $\rho_j^k \in \partial g_j(x^k), \forall j = 1, \ldots, q$, and $\omega^k \in \mathcal{N}_{\mathcal{X}}(x^k)$ such that

$$\frac{\mathbb{x}^k - x^k}{\|\mathbb{x}^k - x^k\|} = \frac{x^k - \mathbb{x}^k}{\|\mathbb{x}^k - x^k\|} + \sum_{i=1}^p \bar{\xi}_i^k \nabla h_i(x^k) + \sum_{j=1}^q \bar{\zeta}_j^k \rho_j^k + 2\omega^k.$$

We obtain from the discussion above that

$$\begin{aligned}
\|\mathbb{x}^k - x^k\| &= \frac{\langle \mathbb{x}^k - x^k, \mathbb{x}^k - x^k \rangle}{\|\mathbb{x}^k - x^k\|} \\
&= \Big\langle \frac{x^k - \mathbb{x}^k}{\|\mathbb{x}^k - x^k\|} + \sum_{i=1}^p \bar{\xi}_i^k \nabla h_i(x^k) + \sum_{j=1}^q \bar{\zeta}_j^k \rho_j^k + 2\omega^k, \mathbb{x}^k - x^k \Big\rangle \\
&\leq \Big\langle \frac{x^k - \mathbb{x}^k}{\|\mathbb{x}^k - x^k\|} + \sum_{i=1}^p \bar{\xi}_i^k \nabla h_i(x^k) + \sum_{j=1}^q \bar{\zeta}_j^k \rho_j^k, \mathbb{x}^k - x^k \Big\rangle + o(\|\mathbb{x}^k - x^k\|) \\
&\leq \sum_{i=1}^p \Big\langle \bar{\xi}_i^k \nabla h_i(x^k), \mathbb{x}^k - x^k \Big\rangle + \sum_{j=1}^q \Big\langle \bar{\zeta}_j^k \rho_j^k, \mathbb{x}^k - x^k \Big\rangle + o(\|\mathbb{x}^k - x^k\|) \\
&\leq \sum_{i=1}^p \bar{\xi}_i^k \Big(h_i(\mathbb{x}^k) - o(\|\mathbb{x}^k - x^k\|) \Big) \\
&\quad + \sum_{j=1}^q \bar{\zeta}_j^k \Big(g_j(\mathbb{x}^k) - o(\|\mathbb{x}^k - x^k\|) \Big) + o(\|\mathbb{x}^k - x^k\|) \\
&\leq 2 \Big| \sum_{i=1}^p \xi_i^k h_i(\mathbb{x}^k) + \sum_{j=1}^q \zeta_j^k g_j(\mathbb{x}^k) \Big| + 2 \Big| \sum_{i=1}^p \xi_i^k + \sum_{j=1}^q \zeta_j^k + 1 \Big| o(\|\mathbb{x}^k - x^k\|) \\
&\leq 2 \Big| \sum_{i=1}^p \xi_i^k h_i(\mathbb{x}^k) + \sum_{j=1}^q \zeta_j^k g_j(\mathbb{x}^k) \Big| + \frac{1}{2} \|\mathbb{x}^k - x^k\|
\end{aligned}$$

where the first inequality comes from the fact that $\mathcal{X}$ is regular, the third one arises from the subdifferential regularity assumption of $h_i(x)$ and $g_j(x)$ in $\mathbb{B}(x^*, \delta_0) \cap \mathcal{X}$, and the last one is valid because without loss of generality, we may assume for k sufficiently large,

$$o(\|\mathbb{x}^k - x^k\|) \leq \frac{1}{4(M_0+1)}\|\mathbb{x}^k - x^k\|$$

since $\mathbb{x}^k - x^k \to 0$ as k tends to infinity. This means

$$d_{\mathcal{C}}(\mathbb{x}^k) = \|\mathbb{x}^k - x^k\| \leq 4M_0\Big(\sum_{i=1}^{p}|h_i(\mathbb{x}^k)| + \sum_{i=1}^{q} g_j^+(\mathbb{x}^k)\Big).$$

Thus, for any sequence $\{\mathbb{x}^k\} \subset \mathcal{X}$ converging to x^* there exists a number $c > 0$ such that

$$d_{\mathcal{C}}(\mathbb{x}^k) \leq c(\|h(\mathbb{x}^k)\|_1 + \|g^+(\mathbb{x}^k)\|_1) \quad \forall k = 1, 2, \ldots.$$

This further implies the error bound property at x^*. Indeed, suppose the contrary. Then there exists a sequence $\tilde{\mathbb{x}}^k \to x^*$, such that $\tilde{\mathbb{x}}^k \in \mathcal{X} \setminus \mathcal{C}$ and $d_{\mathcal{C}}(\tilde{\mathbb{x}}^k) > c(\|h(\tilde{\mathbb{x}}^k)\|_1 + \|g^+(\tilde{\mathbb{x}}^k)\|_1)$ for all $k = 1, 2, \ldots$, which is a contradiction. ■

A natural question to ask is: Is the quasinormality strictly stronger than the error bound property. This question has been answered positively in [22, Example 2.1], with a smooth optimization problem without an abstract set constraint.

6 Sensitivity analysis of value functions

In this section we consider the following perturbed optimization problem:

$$P(a) \qquad \min\ \hat{f}(x,a) \ \text{ s.t. } \ x \in \mathcal{C}(a),$$

with

$$\mathcal{C}(a) = \{x \in \mathcal{X} : \hat{h}(x,a) = 0, \hat{g}(x,a) \leq 0\}, \tag{6.1}$$

where $\mathcal{X}$ is closed subset of $\mathbb{R}^m$, $\hat{f} : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}, \hat{h} : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^p, \hat{g} : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^q$ are Lipschitz continuous around $(\bar{x}, \bar{a})$.

In practice it is often important to know how well the model responds to the perturbation a. For this we need to consider, for instance, the value function $V(a)$ related to the parametric optimization problem:

$$V(a) := \inf_{x \in \mathcal{C}(a)} \hat{f}(x,a), \tag{6.2}$$

with the solution map $\mathcal{S}(\cdot)$ defined by

$$\mathcal{S}(a) := \{x \in \mathcal{C}(a) :\ V(a) = \hat{f}(x,a)\}. \tag{6.3}$$

In the recent paper [28], Mordukhovich, Nam and Yen obtain some new results for computing and estimating the Fréchet subgradient of the value function in parametric optimization (6.1)-(6.1) with smooth and nonsmooth data using normal multipliers. In the following result we estimate the Fréchet subdifferential of the value function by using the quasinormal multipliers instead. Since the set of quasinormal multipliers are smaller than the set of normal multipliers, our estimate provides a tighter bound for the Fréchet subdifferential of the value function.

Let $M_Q^r(\bar{x},\bar{a})$ denotes the set of vectors $(\lambda,\mu,\gamma)\in\mathbb{R}^p\times\mathbb{R}^q_+\times\mathbb{R}$ such that

$$0\in r\partial\hat{f}(\bar{x},\bar{a})+\sum_{i=1}^{p}\partial(\lambda_i\hat{h}_i)(\bar{x},\bar{a})+\sum_{j=1}^{q}\mu_i\partial\hat{g}_j(\bar{x},\bar{a})+(0,\gamma)+\mathcal{N}_{\mathcal{X}}(\bar{x})\times\{0\}$$

and there exists a corresponding sequence $\{(x^k,a^k)\}\subset\mathcal{X}\times\mathbb{R}^n$ converging to $(\bar{x},\bar{a})$ such that $\lambda_i\hat{h}_i(x^k,a^k)>0$ for all $i\in I:=\{i:\lambda_i\neq 0\}$, $\mu_j\hat{g}_j(x^k,a^k)>0$ for all $j\in J:=\{\mu_j>0\}$, and $\hat{h}_i(i\in I),\hat{g}_j(j\in J)$ are proximal subdifferentiable at (x^k,a^k) for each k.

Theorem 6 *Let $V(a)$ be the value function as defined in (6.2) and $\bar{x}\in\mathcal{S}(\bar{a})$. Assume also that $(\bar{x},\bar{a})$ is quasinormal for the constraint region*

$$\{(x,a)\in\mathcal{X}\times\mathbb{R}^n:\hat{h}(x,a)=0,\hat{g}(x,a)\leq 0\}.$$

Then one has the upper estimation:

$$\partial^F V(\bar{a})\subset\{-\gamma:(\lambda,\mu,\gamma)\in M_Q^1(\bar{x},\bar{a})\}. \tag{6.4}$$

Proof There is nothing to prove if $\partial^F V(\bar{a})=\emptyset$. Let $\gamma\in\partial^F V(\bar{a})\neq\emptyset$. Then by definition of the Fréchet subdifferential, for arbitrary $\kappa>0$, there exists $\delta_\kappa>0$ such that

$$V(a)-V(\bar{a})\geq\langle\gamma,a-\bar{a}\rangle-\kappa\|a-\bar{a}\|\qquad\forall a\in\mathbb{B}(\bar{a},\delta_\kappa).$$

By definition of the value function, for every $x\in\mathcal{C}(a)$, we have $\hat{f}(x,a)\geq V(a)$ and hence for all $x\in\mathcal{C}(a)$, $\hat{f}(x,a)-\langle\gamma,a-\bar{a}\rangle+\kappa\|a-\bar{a}\|\geq\hat{f}(\bar{x},\bar{a})$

Thus, $(\bar{x},\bar{a})$ is a local optimal solution to the optimization problem

$$\begin{aligned}\min\ &\hat{f}(x,a)-\langle\gamma,a-\bar{a}\rangle+\kappa\|a-\bar{a}\|\\ \text{s.t.}\ &\hat{h}_i(x,a)=0,\quad i=1,\ldots,p,\\ &\hat{g}_j(x,a)\leq 0,\quad j=1,\ldots,q,\\ &(x,a)\in\mathcal{X}\times\mathbb{R}^n.\end{aligned}$$

Since $(\bar{x},\bar{a})$ is quasinormal by assumption, by the enhanced KKT condition (Theorem 1), there exist a vector $(\lambda,\mu)\in\mathbb{R}^p\times\mathbb{R}^q_+$ and a sequence $\{(x^k,a^k)\}\subset$

$\mathcal{X} \times \mathbb{R}^n$ converging to $(\bar{x}, \bar{a})$ such that the following conditions hold:

$$
\begin{aligned}
&0 \in \partial \hat{f}(\bar{x}, \bar{a}) + \sum_{i=1}^{p} \partial(\lambda_i \hat{h}_i)(\bar{x}, \bar{a}) + \sum_{j=1}^{q} \mu_j \partial \hat{g}_j(\bar{x}, \bar{a}) + \\
&\qquad \mathcal{N}_{\mathcal{X}\times\mathbb{R}^m}(\bar{x}, \bar{a}) - \begin{pmatrix} 0 \\ \gamma \end{pmatrix} + \kappa \begin{pmatrix} 0 \\ \mathbb{B} \end{pmatrix}, \qquad (6.5)\\
&\lambda_i \hat{h}_i(x^k, a^k) > 0 \ \forall i \in I, \quad \mu_j \hat{g}_j(x^k, a^k) > 0 \ \forall j \in J, \\
&\hat{h}_i (i \in I), \hat{g}_j (j \in J) \text{ are proximal subdifferentiable at } (x^k, a^k).
\end{aligned}
$$

The desired upper estimation follows since κ is arbitrary. ■

We now give a tighter estimate for the limiting subdifferential of the value function in terms of the quasinormality.

Theorem 7 *Let $V(a)$ be the value function as defined in (6.2). Suppose that the growth hypothesis holds, i.e., there exists $\delta > 0$ such that the set*

$$\{x \in \mathcal{X} : \hat{h}(x, \bar{a}) = \alpha,\ \hat{g}(x, \bar{a}) \le \beta,\ \hat{f}(x, \bar{a}) \le M, (\alpha, \beta) \in \delta\mathbb{B}\}$$

is bounded for each $M \in \mathbb{R}$. Assume that for each $\bar{x} \in \mathcal{S}(\bar{a})$, $(\bar{x}, \bar{a})$ is quasinormal for the constraint region

$$\{(x, a) \in \mathcal{X} \times \mathbb{R}^n : \hat{h}(x, a) = 0,\ \hat{g}(x, a) \le 0\}. \qquad (6.6)$$

Then the value function $V(a)$ is l.s.c. near $\bar{a}$ and

$$
\begin{aligned}
\partial V(\bar{a}) &\subset \bigcup_{\bar{x} \in \mathcal{S}(\bar{a})} \{-\gamma : (\lambda, \mu, \gamma) \in M_Q^1(\bar{x}, \bar{a})\} \\
\partial^\infty V(\bar{a}) &\subset \bigcup_{\bar{x} \in \mathcal{S}(\bar{a})} \{-\gamma : (\lambda, \mu, \gamma) \in M_Q^0(\bar{x}, \bar{a})\}.
\end{aligned}
$$

Proof By [17, Theorem 3.6], the value function $V(a)$ is lower semicontinuous near $\bar{a}$ under our assumption.

Step 1. Let v be a vector that belongs to $\partial V(\bar{a})$, by definition there are sequences $a^l \to \bar{a}$ and $v^l \to v$ with $v^l \in \partial^F V(a^l)$. By the growth hypothesis, for l sufficiently large, we may find a solution $x^l \in \mathcal{S}(a^l)$. Following [10, Theorem 6.5.2], without loss of generality we may assume x^l converges to an element $\bar{x} \in \mathcal{S}(\bar{a})$. Since $(\bar{x}, \bar{a})$ is quasinormal and it is a limit point of the sequence $\{(x^l, a^l)\}$, by Lemma 1 we find that for sufficient large l, (x^l, a^l) is also quasinormal for the constraint region (6.6) and hence from Theorem 6 it follows that for each l there exist a vector $(\lambda^l, \mu^l) \in \mathbb{R}^p \times \mathbb{R}^q_+$ and a sequence $\{(x^{l,k}, a^{l,k})\}_k \subset \mathcal{X} \times \mathbb{R}^n$ converging to (x^l, a^l) as $k \to \infty$ such that

$$(0, v^l) \in \partial \hat{f}(x^l, a^l) + \sum_{i=1}^{p} \partial(\lambda_i^l \hat{h}_i)(x^l, a^l) + \sum_{j=1}^{q} \mu_j^l \partial \hat{g}_j(x^l, a^l) + \mathcal{N}_{\mathcal{X}}(x^l) \times \{0\}, \qquad (6.7)$$

$$\lambda_i^l \hat{h}_i(x^{l,k}, a^{l,k}) > 0 \ \forall i \in I^l, \quad \mu_j^l \hat{g}_j(x^{l,k}, a^{l,k}) > 0 \ \forall j \in J^l, \qquad (6.8)$$

$$\hat{h}_i (i \in I^l), \hat{g}_j (j \in J^l) \text{ are proximal subdifferentiable at } (x^{l,k}, a^{l,k}), \qquad (6.9)$$

where $I^l := \{i : \lambda_i^l \neq 0\}, J^l := \{j : \mu_j^l > 0\}$. Similar as in Step 2 of the proof of Proposition 4, we may obtain the boundedness of the multipliers sequence $\{\lambda_1^l, \dots, \lambda_p^l, \mu_1^l, \dots, \mu_q^l\}$. Therefore, without loss of generality, we may assume $\{\lambda_1^l, \dots, \lambda_p^l, \mu_1^l, \dots, \mu_q^l\}$ converges to $\{\lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_q\}$. Taking the limit on both sides of (6.7), similar to Theorem 1, we obtain

$$(0, v) \in \partial \hat{f}(\bar{x}, \bar{a}) + \sum_{i=1}^{p} \partial(\lambda_i \hat{h}_i)(\bar{x}, \bar{a}) + \sum_{j=1}^{q} \mu_j \partial \hat{g}_j(\bar{x}, \bar{a}) + \mathcal{N}_{\mathcal{X}}(\bar{x}) \times \{0\}.$$

Also we find a sequence $\{(x^{l,k_l}, a^{l,k_l})\} \subset \mathcal{X} \times \mathbb{R}^n$ converging to $\bar{x}$ as $l \to \infty$ and is such that for all l, $\lambda_i \hat{h}_i(x^{l,k_l}, a^{l,k_l}) > 0, \forall i \in I,\ \mu_j g_j(x^{l,k_l}, a^{l,k_l}) > 0, \forall j \in J$, and h_i, g_j are proximal subdifferentiable at x^{l,k_l}, where $I = \{i | \lambda_i \neq 0\}$ and $J = \{j | \mu_j > 0\}$.

Step 2. Let $v \in \partial^\infty V(\bar{a})$. By definition there are sequence $a^l \to \bar{a}$, $v^l \in \partial^F V(a^l)$ and $t^l \downarrow 0$ such that $t^l v^l \to v$. Similar as in Step 1, for each l there exist a vector $(\lambda^l, \mu^l) \in \mathbb{R}^p \times \mathbb{R}_+^q$ and a sequence $\{(x^{l,k}, a^{l,k})\} \subset \mathcal{X} \times \mathbb{R}^n$ converging to (x^l, a^l) such that (6.7)-(6.9) hold. Multiplying both sides of (6.7) by t^l we have

$$\begin{aligned}(0, t^l v^l) \in t^l \partial \hat{f}(x^l, a^l) &+ \sum_{i=1}^{p} \partial(t^l \lambda_i^l \hat{h}_i)(x^l, a^l) + \sum_{j=1}^{q} t^l \mu_j^l \partial \hat{g}_j(x^l, a^l) \\ &+ \mathcal{N}_{\mathcal{X}}(\bar{x}) \times \{0\}. \end{aligned} \tag{6.10}$$

Since $(\bar{x}, \bar{a})$ is quasinormal for the constraint region (6.6), similarly as in Step 2 of the proof of Proposition 4, the sequence $\{t^l \lambda_1^l, \dots, t^l \lambda_p^l, t^l \mu_1^l, \dots, t^l \mu_q^l\}$ must be bounded as $l \to \infty$. Without loss of generality assume that the limit is $\{\lambda_1, \dots, \lambda_p, \mu_1, \dots, \mu_q\}$. Talking limits in (6.10), we have

$$(0, v) \in \sum_{i=1}^{p} \lambda_i \partial \hat{h}_i(\bar{x}, \bar{a}) + \sum_{j=1}^{q} \mu_j \partial \hat{g}_j(\bar{x}, \bar{a}) + \mathcal{N}_{\mathcal{X}}(\bar{x}) \times \{0\}.$$

The rest of the proof is similar to Step 1. ■

From Theorem 7 we derive the following very interesting result which significantly improves the classical result in that our sufficient condition is the perturbed quasinormality which is much weaker than the classical condition of NNAMCQ (see e.g. [17, Corollary 3.7]).

Corollary 4 *Let $V(a)$ be the value function as defined in (6.2). Suppose that the growth hypothesis holds at each $\bar{x} \in \mathcal{S}(\bar{a})$.*

(i) Assume that $(\bar{x}, \bar{a})$ is quasinormal for the constraint region (6.6). If

$$\bigcup_{\bar{x} \in \mathcal{S}(\bar{a})} \{-\gamma : (\lambda, \mu, \gamma) \in M_Q^0(\bar{x}, \bar{a})\} = \{0\}, \tag{6.11}$$

then the value function $V(a)$ is Lipschitz continuous around $\bar{a}$ with

$$\emptyset \neq \partial V(\bar{a}) \subset \bigcup_{\bar{x} \in \mathcal{S}(\bar{a})} \{-\gamma : (\lambda, \mu, \gamma) \in M_Q^1(\bar{x}, \bar{a})\}.$$

In addtion to the above assumptions, if

$$\bigcup_{\bar{x} \in \mathcal{S}(\bar{a})} \{-\gamma : (\lambda, \mu, \gamma) \in M_Q^1(\bar{x}, \bar{a})\} = \{-\bar{\gamma}\}$$

for some $(\bar{\lambda}, \bar{\mu}, \bar{\gamma}) \in M_Q^1(\bar{x}, \bar{a})$, then V is strictly differentiable at $\bar{a}$ and $\nabla V(\bar{a}) = -\bar{\gamma}$.

(ii) For the functions $\phi = \hat{f}, \pm\hat{h}_i, \hat{g}_j$, suppose that the partial limiting subdifferential property holds at $(\bar{x}, \bar{a})$:

$$\partial\phi(\bar{x}, \bar{a}) = \partial_x \phi(\bar{x}, \bar{a}) \times \partial_a \phi(\bar{x}, \bar{a}).$$

Also assume that $(\bar{x}, \bar{a})$ is quasinormal for the constraint region (6.6). If

$$\bigcup_{\bar{x} \in \mathcal{S}(\bar{a})} \left\{ \sum_{i=1}^{p} \partial_a (\lambda_i \hat{h}_i)(\bar{x}, \bar{a}) + \sum_{j=1}^{q} \mu_j \partial_a \hat{g}_j(\bar{x}, \bar{a}) : (\mu, \lambda) \in \widetilde{M}_Q^0(\bar{x}, \bar{a}) \right\} = \{0\} \tag{6.12}$$

then the value function $V(a)$ is Lipschitz continuous around $\bar{a}$ and

$$\emptyset \neq \partial V(\bar{a}) \subset \bigcup_{\substack{\bar{x} \in \mathcal{S}(\bar{a}) \\ (\mu,\lambda) \in \widetilde{M}_Q^1(\bar{x}, \bar{a})}} \left\{ \partial_a \hat{f}(\bar{x}, \bar{a}) + \sum_{i=1}^{p} \partial_a (\lambda_i \hat{h}_i)(\bar{x}, \bar{a}) + \sum_{j=1}^{q} \mu_i \partial_a \hat{g}_j(\bar{x}, \bar{a}) \right\} \tag{6.13}$$

where $\widetilde{M}_Q^r(\bar{x}, \bar{a})$ denotes the set of perturbed quasinormal multipliers which are the set of vectors $(\lambda, \mu) \in \mathbb{R}^p \times \mathbb{R}_+^q$ such that

$$0 \in r\partial_x \hat{f}(\bar{x}, \bar{a}) + \sum_{i=1}^{p} \partial_x (\lambda_i \hat{h}_i)(\bar{x}, \bar{a}) + \sum_{j=1}^{q} \mu_j \partial_x \hat{g}_j(\bar{x}, \bar{a}) + \mathcal{N}_{\mathcal{X}}(\bar{x})$$

and there exists a corresponding sequence $\{(x^k, a^k)\} \subset \mathcal{X} \times \mathbb{R}^n$ converging to $(\bar{x}, \bar{a})$ such that $\lambda_i \hat{h}_i(x^k, a^k) > 0$ for all $i \in I := \{i : \lambda_i \neq 0\}$, $\mu_j \hat{g}_j(x^k, a^k) > 0$ for all $j \in J := \{\mu_j > 0\}$, and $\hat{h}_i (i \in I), \hat{g}_j (j \in J)$ are proximal subdifferentiable at (x^k, a^k) for each k.

(iii) Suppose that the partial limiting subdifferential property at $(\bar{x}, \bar{a})$ holds as in (ii) and $\widetilde{M}_Q^0(\bar{x}, \bar{a}) = \{0\}$ for each $\bar{x} \in \mathcal{S}(\bar{a})$. Then the value function $V(a)$ is Lipschitz continuous around $\bar{a}$ and (6.13) holds.

Proof (i) It follows from Theorem 7 that

$$\bigcup_{\bar{x}\in\mathcal{S}(\bar{a})}\{-\gamma : (\lambda,\mu,\gamma)\in M_Q^0(\bar{x},\bar{a})\} = \{0\}$$

implies that $\partial^\infty V(\bar{a}) = \{0\}$. We conclude that the value function is Lipschitz around $\bar{a}$ by virtue of Proposition 1 (iii). The assertion about the strict differentiability then follows from Proposition 1 (i).

(ii) It is clear that under the partial limiting subdifferential property, (6.11) is equivalent to (6.12). The conclusion then follows from applying Theorem 7 and Proposition 1 (iii).

(iii) follows immediately from (ii) and the fact that $\widetilde{M}_Q^0(\bar{x},\bar{a}) = \{0\}$ implies the quasinormality of $(\bar{x},\bar{a})$. ■

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

References

1. C.R. Bector, S. Chandra and J. Dutta, *Principle of Optimization Theory.* Narosa Publishers, India and Alpha Science Publishers, 2004.
2. D.P. Bertsekas, *Nonlinear Programming.* Athena Scientific Publishers, 1999.
3. D.P. Bertsekas, A. Nedić and A.E. Ozdaglar, *Convex Analysis and Optimization.* Belmont, MA, Athena Scientific, 2003.
4. D.P. Bertsekas and A.E. Ozdaglar, *Pseudonormality and a Lagrange multiplier theory for constrained optimization.* J. Optim. Theo. Appl., **114**(2002), 287-343.
5. D.P. Bertsekas and A.E. Ozdaglar, *The relation between pseudonormality and quasiregularity in constrained optimization.* Optim. Meth. Softw., **19** (2004), 493-506.
6. D.P. Bertsekas, A.E. Ozdaglar and P. Tseng, *Enhanced Fritz John conditions for convex programming*, SIAM J. Optim., **16**(2006), 766-797.
7. J.M. Borwein and A.S. Lewis, *Convex Analysis and Nonlinear Optimization.* Canadian Mathematical Society Books in Mathematics, Springer, New York, 2000.
8. J.M. Borwein and Q.J. Zhu, *Techniques of Variational Analysis.* Canadian Mathematical Society Books in Mathematics, Springer, New York, 2005.
9. F.H. Clarke, *A new approach to Lagrange multipliers*, Math. Oper. Res., **1**(1976), 165-174.
10. 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.
11. F.H. Clarke, Y.S. Ledyaev, R.J. Stern and P.R. Wolenski, *Nonsmooth Analysis and Control Theory.* Springer, New York, 1998.
12. R. Henrion, A. Jourani and J. Outrata, *On the calmness of a class of multifunctions*, SIAM J. Optim., **13** (2002), 603-618.
13. M.R. Hestenes, *Optimization Theory: The Finite Dimensional Case.* Wiley, Now York, 1975.
14. F. John, *Extremum problems with inequalities as side constraints.* In "Studies and Essays, Courant Anniversary Volume" (K.O. Friedrichs, O.E. Neugebauer and J. J. Stoker, eds), 187-204. Wiley (interscience), New York, 1948.
15. A. Jourani, *Constraint qualification and Lagrange multipliers in nondifferentiable programming problems.* J. Optim. Theo. Appl., **81**(1994), 533-548.
16. 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), 2730-2753.

17. Y. Lucet and J.J. Ye, *Sensitivity analysis of the value function for optimization problems with variational inequality constraints.* SIAM J. Contr. Optim., **40**(2001), 699-723.
18. Y. Lucet and J.J. Ye, *Erratum: sensitivity analysis of the value function for optimization problems with variational inequality constraints.* SIAM J. Contr. Optim., **41**(2001), 1315-1319.
19. P.D. Loewen, *Optimal Control via Nonsmooth Analysis.* American Mathematical Society, Providence, Rhode Island, 2003.
20. O.L. Mangasarian and S. Fromovitz, *The Fritz John necessary optimality conditions in the presence of equality and inequality constraints*, J. Math. Anal. Appl., **17**(1967), 37-47.
21. E.J. McShane, *The Lagrange multiplier rule*, Amer. Math. Monthly, **80** (1973), 922-925.
22. L. Minchenko and A. Turakanov, *On error bounds for quasinormal programs.* J. Optim. Theo. Appl., **148**(2011), 571-579.
23. B.S. Mordukhovich, *Metric approximation and necessary optimality conditions for general classes of nonsmooth extremal problems*, Soviet Math. Dokl. **22** (1980), 526-530.
24. B.S. Mordukhovich, *Sensitivity analysis in nonsmooth analysis*, in Theoretical Aspects of Industrial Design, edited by D.A. Field and V. Komkov, SIAM Proc. Appl. Math. **58** (1992), 32-46, Philadelphia, Pennsylvania.
25. B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation I. Basic Theory.* Ser. Comprehensive Stud. Math. 330, Springer, Berlin, 2006.
26. B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation II. Application.* Ser. Comprehensive Stud. Math. 331, Springer, Berlin, 2006.
27. B.S. Mordukhovich, N.M. Nam and N.D. Yen, *Fréchet subdifferential calculus and optimality conditions in nondifferentiable programming*, Optim., **55** (2006), 685-708.
28. B.S. Mordukhovich, N.M. Nam and N.D. Yen, *Subgradients of marginal functions in parametric mathematical programming*, Math. Program. Ser. B, **116**(2009), 369-396.
29. R.T. Rochafellar and R. J.-B. Wets, *Variational Analysis.* Springer-Verlag, Berlin, 1998.
30. J.S. Treiman, *Lagrange multipliers for nonconvex generalized gradients with equality, inequality, and set constraints.* SIAM J. Contr. Optim., **37**(1999), 1313-1329.
31. Z.L. Wu and J.J. Ye, *Sufficient conditions for error bounds.* SIAM J. Optim. **12**(2001), 421-435.
32. J.J. Ye, *Multiplier rules under mixed assumptions of differentiability and Lipschitz continuity.* SIAM J. Contr. Optim., **39**(2001), 1441-1460.
33. J.J. Ye and X.Y. Ye, *Necessary optimality conditions for optimization problems with variational inequality constraints*, Math. Oper. Res., **22**(1997), 977-997.