KURDYKA-ŁOJASIEWICZ INEQUALITY AND ERROR BOUNDS OF D-GAP FUNCTIONS FOR NONSMOOTH AND NONMONOTONE VARIATIONAL INEQUALITY PROBLEMS

MINGHUA LI*, KAIWEN MENG†, AND XIAOQI YANG‡

Abstract. In this paper, we study the D-gap function associated with a nonsmooth and nonmonotone variational inequality problem. We present some exact formulas for the subderivative, the regular subdifferential set, and the limiting subdifferential set of the D-gap function. By virtue of these formulas, we provide some sufficient and necessary conditions for the Kurdyka-Łojasiewicz inequality property and the error bound property for the D-gap functions. As an application of our Kurdyka-Łojasiewicz inequality result and the abstract convergence result in [Attouch, et al., Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, Math. Program., 137(2013)91-129], we show that the sequence generated by a derivative free descent algorithm with an inexact line search converges linearly to some solution of the variational inequality problem.

Key words. variational inequality problem, D-gap function, Kurdyka-Łojasiewicz inequality, error bound, inexact line search, linear convergence rate

AMS subject classifications. Primary, 65K10, 65K15; Secondary, 90C26, 49M37

1. Introduction.

In this paper, we consider a variational inequality problem (VIP) of finding $x \in K$ such that

$$\langle F(x), y - x \rangle \geq 0 \quad \forall y \in K,$$

where K is a closed and convex subset of $\mathbb{R}^n$ and the mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ is locally Lipschitz continuous and not necessarily monotone. (VIP) has many applications in various fields such as mathematical programming, traffic network equilibrium problems and economics. We refer the reader to the very informative book [10] by Facchinei and Pang for the background information and motivations of (VIP).

One popular approach to study (VI) is based on reformulating (VIP) as equivalent constrained/unconstrained optimization problems through the consideration of appropriate gap (merit) functions; see [1, 2, 7, 10, 11, 13, 15, 16, 17, 19, 22, 25, 26, 27, 28, 29, 30, 33, 34, 36, 38, 39]. Among various reformulations in the literature, we recall that $\bar{x}$ solves (VIP) if and only if $\bar{x}$ solves the following unconstrained optimization problem with 0 as its optimal value:

$$\min_{x \in \mathbb{R}^n} \quad f_{ab}(x) := f_a(x) - f_b(x),$$

where $b > a > 0$, and for each $c > 0$,

$$f_c(x) := \max_{y \in K} \{\langle F(x), x - y \rangle - \frac{c}{2} ||y - x||^2\}.$$

While f_c is known as the regularized gap function [1, 11] with c being the regularized parameter, f_{ab} is often known as the D-gap function [28] with 'D' standing for the

*School of Mathematics and Big Data, Chongqing University of Arts and Sciences, Yongchuan, Chongqing, 402160, China (minghuali20021848@163.com)

†School of Mathematics, Southwestern University of Finance and Economics, Chengdu 611130, China (mengkw@swufe.edu.cn)

‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Kowloon, Hong Kong (mayangxq@polyu.edu.hk)

'difference' of two parameterized regularized gap functions. By replacing the quadratic term in defining f_c with some general term having very similar properties as those of the quadratic term, the corresponding *generalized* regularized gap and *generalized* D-gap functions have also been extensively studied in the literature; see [18, 19, 36, 39].

The (generalized) differentiability properties of these regularized gap and D-gap functions have been extensively investigated, and have been utilized to study the property of error bounds [10] and the property of the Kurdyka-Łojasiewicz (KL, for short) inequality [9]. The latter properties have played very important roles in convergence analysis for algorithms designed based upon gap functions.

We review a few of typical results related to the (generalized) D-gap function as follows. Peng [28] showed that if F is continuously differentiable and strongly monotone, the D-gap function is also continuously differentiable and its square root provides a global error bound for (VIP). Yamashita et al. [39] introduced the generalized D-gap function and obtained its continuous differentiability by assuming that F is continuously differentiable. Moreover, by assuming that F is strong monotone and that either F is Lipschitz continuous or K is compact, they showed that the square root of the generalized D-gap function provides a global error bound for (VIP), and that the sequence generated by a descent algorithm with an inexact line search converges to the unique solution of (VIP). Based on the D-gap function and by assuming that F is continuously differentiable and monotone, Solodov and Tseng [32] developed two unconstrained methods that are similar to the feasible method in Zhu and Marcotte [40] which is based on the regularized gap function. By assuming that F is locally Lipschitz continuous, Xu [37] obtained a formula for the Clarke subdifferential set of the D-gap function, and a global convergence result for a descent algorithm with an inexact line search under the circumstance that F is strongly monotone and Lipschitz continuous. By the same assumption that F is locally Lipschitz continuous, Ng and Tan [23] obtained some formulas for the Clarke directional derivative and the Clarke subdifferential set of the D-gap function. By assuming that F is coercive and locally Lipschitz continuous, and by introducing a condition expressed in terms of the Clarke generalized Jacobian of F, Li and Ng [18] showed that the square root of the generalized D-gap function provides a local error bound for (VIP), and by virtue of which, they proved that any cluster point of the sequence generated by a descent algorithm with an inexact line search is a solution of (VIP), and that the convergence rate is linear when F is smooth, strongly monotone and ∇F is locally Lipschitz continuous. Note that Li and Ng [18] also provided some formulas for the Clarke directional derivative and the Clarke subdifferential set of the generalized D-gap function, which were very crucial for their arguments. Later Li et al. [19] established some error bound results for the generalized D-gap function by assuming that F is (Lipschitz) continuous, locally monotone and coercive.

From the literature review above, it is clear to see that most of the existing results for error bounds and the convergence of a descent algorithm were obtained by assuming that F is strongly monotone, with an exception being that, the error bound result in Li and Ng [18], though having difficulty in verification, was applied to some cases when F is nonmonotone. As for the property of the KL inequality, there is almost no result, to the best of our knowledge, presented in a straightforward way for the case when F is locally Lipschitz continuous. By examining the definition for the KL inequality (see Definition 2.4 below) and the theory of error bounds in [5, 21], it is reasonable that the notion of the subderivative, the regular/Fréchet subdifferential set, and the general/limiting subdifferential set (see Definition 2.2) should have played a role in studying the generalized differentiability properties of the regularized gap and D-gap functions. But it is quite surprising that there is no such a related result in the literature for the case when F is locally Lipschitz continuous and not necessarily monotone.

To fill this gap, we will investigate the KL inequality and error bounds of the D-gap function for nonsmooth and nonmonotone (VIP) by providing formulas for the subderivative and the (limiting) subdifferential sets of the D-gap functions, and as an application of our result for the KL inequality and the abstract convergence result in [4] for inexact descent methods, we will establish the linear convergence rate for a descent algorithm with an inexact line search.

The main contributions of the paper are as follows.

(i) We obtain a number of exact formulas for the subderivatives, the regular/Fréchet subdifferential sets, and the general/limiting subdifferential sets of the regularized gap function f_c and the D-gap function f_{ab}, respectively. See Propositions 3.2-3.4 below. Taking the limiting subdifferential set $\partial f_{ab}(\bar{x})$ of f_{ab} at a point $\bar{x}$ for instance, we obtain

$$\partial f_{ab}(\bar{x}) = D^*F(\bar{x})\,(\pi_b(\bar{x}) - \pi_a(\bar{x})) - b\,(\pi_b(\bar{x}) - \pi_a(\bar{x})) + (b-a)(\bar{x} - \pi_a(\bar{x})),$$

where $D^*F(\bar{x})$ denotes the coderivative of F at $\bar{x}$ (cf. Definition 2.6), and $\pi_\xi(x) := P_K\left(x - \frac{F(x)}{\xi}\right)$ for any given $\xi > 0$ with $P_K(\cdot)$ being the projection operator onto K. To the best of our knowledge, these formulas have not been seen from the literature, although, as mentioned above, exact formulas have been obtained for the Clarke directional derivatives and the Clarke subdifferential sets of f_c and f_{ab}, respectively.

(ii) By virtue of the formula obtained for the general/limiting subdifferential set of the D-gap function f_{ab}, we present a few sharp results on the properties of the KL inequality and the error bounds for f_{ab}. In particular, by assuming that the following inequality holds for some $\mu > 0$ and for all $x \in \mathbb{R}^n$ where F is differentiable:

$$(1.1) \qquad \langle \nabla F(x)(\pi_a(x) - \pi_b(x)),\ \pi_a(x) - \pi_b(x)\rangle \geq \mu||\pi_a(x) - \pi_b(x)||^2,$$

which can be considered as a restricted (weaker) notion of strong monotonicity, we show that

$$d(0, \partial f_{ab}(x)) \geq \mu\|\pi_b(x) - \pi_a(x)\| \quad \forall x \in \mathbb{R}^n,$$

and that f_{ab} is a KL function with an exponent of $\frac{1}{2}$, and moreover that some local/global error bound results holds. See Theorem 4.7 below.

(iii) By assuming (1.1) and applying our result on the KL property for f_{ab}, we obtained the linear convergence rate for a derivative free descent algorithm, which is essentially the same algorithm as those studied in [15, 18, 29, 30, 37, 39]. See Theorem 5.6 below. Starting from any initial point x_0, the algorithm generates a sequence $\{x_n\}$ in the manner of $x_{n+1} = x_n + t_n d_n$, where d_n is the search direction, either being $\pi_a(x_n) - x_n$ or $\pi_a(x_n) - \pi_b(x_n)$, and t_n is the stepsize determined by an Armijo line search. Under some other mild assumptions, except for (1.1), we show that the stepsize sequence $\{t_n\}$ has a positive lower bound $t^* > 0$ (cf. Proposition 5.4 below), and moreover the following hold (cf. Proposition 5.5 below):

$$f_{ab}(x_{n+1}) - f_{ab}(x_n) \leq -M_1||x_{n+1} - x_n||^2$$

and

$$d(0, \partial f_{ab}(x_n)) \leq \frac{M_2}{t^*}||x_{n+1} - x_n||,$$

where M_1 and M_2 are two positive constants. That is, the sequence $\{x_n\}$ satisfies the assumptions **(H1)**, a variant of **(H2)**, and **(H3)** proposed in [4], and our convergence analysis falls into the framework of the abstract convergence for inexact descent methods studied in [4].

The outline of the paper is as follows. Section 2 is about notation and terminology, and some mathematical preliminaries. In section 3, we present some exact formulas for the subderivatives, the regular/Fréchet subdifferential sets, and the general/limiting subdifferential sets of the regularized gap function f_c and the D-gap function f_{ab}, respectively. By virtue of these formulas for the D-gap function, we present in Section 4 some sufficient and necessary conditions for the error bound property and the KL inequality property. As an application of our KL inequality result and the abstract convergence result in [4] for inexact descent methods, we show in section 5 that the sequence generated by a descent algorithm (based upon the D-gap function) with an inexact line search converges linearly to some solution of (VIP).

2. Notation and Mathematical Preliminaries. Throughout the paper we use the standard notations of variational analysis; see the seminal book [31] by Rockafellar and Wets. The Euclidean norm of a vector x is denoted by $||x||$, and the inner product of vectors x and y is denoted by $\langle x, y\rangle$. Let $A \subset \mathbb{R}^n$ be a nonempty set. We denote by $\operatorname{conv} A$ the convex hull of A. The polar cone of A is defined by $A^* := \{v \in \mathbb{R}^n \mid \langle v, x\rangle \leq 0 \ \forall x \in A\}$. The distance from x to A is defined by $d(x, A) := \inf_{y\in A} ||y - x||$. The projection mapping P_A is defined by $P_A(x) := \{y \in A \mid \|y - x\| = d(x, A)\}$.

DEFINITION 2.1. *Let $C \subset \mathbb{R}^n$ and let $x \in C$.*

(i) *The tangent cone to C at x is denoted by $T_C(x)$, i.e., $w \in T_C(x)$ if there exist sequences $t_k \downarrow 0$ and $\{w_k\} \subset \mathbb{R}^n$ with $w_k \to w$ and $x + t_k w_k \in C \ \forall k$.*

(ii) *The regular normal cone to C at x is denoted by $\widehat{N}_C(x)$, i.e., $v \in \widehat{N}_C(x)$ if*

$$\langle v, x - \bar{x}\rangle \leq o(\|x - \bar{x}\|) \quad \textit{for all } x \in C.$$

Another way of defining the regular normal cone is via the equality $\widehat{N}_C(x) = T_C(x)^*$.

(iii) *The normal cone to C at x is denoted by $N_C(x)$, i.e., $v \in N_C(x)$ if there exist sequences $x_k \to x$ and $v_k \to v$ with $x_k \in C$ and $v_k \in \widehat{N}_C(x_k)$ for all k.*

(iv) *C is said to be regular at x in the sense of Clarke if it is locally closed at x (i.e., $C \cap U$ is closed for some closed neighborhood U of x) and $\widehat{N}_C(x) = N_C(x)$.*

Let $f : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ be an extended real-valued function. We denote the epigraph of f by $\operatorname{epi} f := \{(x, \alpha) \mid f(x) \leq \alpha\}$. The lower level set with a level of α is defined and denoted by $[f \leq \alpha] := \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$. In a similar way, we define $[f < \alpha] := \{x \in \mathbb{R}^n \mid f(x) < \alpha\}$ and $[\alpha < f < \beta] := \{x \in \mathbb{R}^n \mid \alpha < f(x) < \beta\}$.

DEFINITION 2.2. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be an extended real-valued function and let $\bar{x}$ be a point with $f(\bar{x})$ finite.*

(i) *The vector $v \in \mathbb{R}^n$ is a regular/Fréchet subgradient of f at $\bar{x}$, written $v \in \widehat{\partial} f(\bar{x})$, if*

$$f(x) \geq f(\bar{x}) + \langle v, x - \bar{x}\rangle + o(||x - \bar{x}||).$$

(ii) *The vector* $v \in \mathbb{R}^n$ *is a general/limiting subgradient of* f *at* $\bar{x}$*, written* $v \in \partial f(\bar{x})$*, if there exist sequences* $x_k \to \bar{x}$ *and* $v_k \to v$ *with* $f(x_k) \to f(\bar{x})$ *and* $v_k \in \widehat{\partial} f(x_k)$*.*

(iii) *The function* f *is said to be (subdifferentially) regular at* $\bar{x}$ *if* epi f *is regular in the sense of Clarke at* $(\bar{x}, f(\bar{x}))$ *as a subset of* $\mathbb{R}^n \times \mathbb{R}$*.*

(iv) *The subderivative* $df(\bar{x}) : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is defined by*

$$df(\bar{x})(w) := \liminf_{t \downarrow 0, w' \to w} \frac{f(\bar{x} + tw') - f(\bar{x})}{t}.$$

Remark 2.3. The regular subgradients can be derived from the subderivative as follows [31, Exercise 8.4]:

$$\widehat{\partial} f(\bar{x}) = \{v \in \mathbb{R}^n | \langle v, w \rangle \leq df(\bar{x})(w) \ \forall w \in \mathbb{R}^n\}.$$

Following [3, 6, 20], we introduce the notion of the Kurdyka-Łojasiewicz (KL, for short) inequality.

Definition 2.4. *For a proper lower semicontinuous function* $f : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$*, a point* $\bar{x} \in \mathbb{R}^n$ *with* $\partial f(\bar{x}) \neq \emptyset$*, and some* $\alpha \in [0, 1)$*, we say that* f *satisfies the KL inequality at* $\bar{x}$ *with an exponent of* α*, if there exist* $\mu, \epsilon > 0$ *and* $\nu \in (0, +\infty]$ *so that*

$$d(0, \partial f(x)) \geq \mu(f(x) - f(\bar{x}))^\alpha$$

whenever $\|x - \bar{x}\| \leq \epsilon$ *and* $f(\bar{x}) < f(x) < f(\bar{x}) + \nu$*. If* f *satisfies the KL inequality at every* $x \in \mathbb{R}^n$ *with* $\partial f(x) \neq \emptyset$ *and with the same exponent* α*, we say that* f *is a KL function with an exponent of* α*.*

Following [10], we introduce the notion of local and global error bounds as follows.

Definition 2.5. *For a proper function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *and a set* $C \subset \mathbb{R}^n$*, we say that* f *has a local error bound on* C *if there exist two positive constants* τ *and* ϵ *such that for all* $x \in [f \leq \epsilon] \cap C$

$$d(x, [f \leq 0] \cap C) \leq \tau \max\{f(x), 0\}.$$

Furthermore, we say that f *has a global error bound on* C *if there exists a constant* $\tau > 0$ *such that the above inequality holds for all* $x \in C$*.*

Definition 2.6. *Let* $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *be a set-valued mapping and* $(\bar{x}, \bar{u}) \in \operatorname{gph} S := \{(x, u) \mid u \in S(x)\}$*.*

(i) *The graphical derivative of* S *at* $\bar{x}$ *for* $\bar{u}$ *is the mapping* $DS(\bar{x} \mid \bar{u}) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *defined by*

$$z \in DS(\bar{x} \mid \bar{u})(w) \Longleftrightarrow (w, z) \in T_{\operatorname{gph} S}(\bar{x}, \bar{u}).$$

(ii) *The regular coderivative of* S *at* $\bar{x}$ *for* $\bar{u}$ *is the mapping* $\widehat{D}^* S(\bar{x} \mid \bar{u}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ *defined by*

$$x^* \in \widehat{D}^* S(\bar{x} \mid \bar{u})(u^*) \Longleftrightarrow (x^*, -u^*) \in \widehat{N}_{\operatorname{gph} S}(\bar{x}, \bar{u}).$$

(iii) *The coderivative of* S *at* $\bar{x}$ *for* $\bar{u}$ *is the mapping* $D^* S(\bar{x} \mid \bar{u}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ *defined by*

$$x^* \in D^* S(\bar{x} \mid \bar{u})(u^*) \Longleftrightarrow (x^*, -u^*) \in N_{\operatorname{gph} S}(\bar{x}, \bar{u}).$$

Here the notation $DS(\bar{x} \mid \bar{u})$*,* $D^* S(\bar{x} \mid \bar{u})$ *and* $\widehat{D}^* S(\bar{x} \mid \bar{u})$ *is simplified to* $DS(\bar{x})$*,* $D^* S(\bar{x})$ *and* $\widehat{D}^* S(\bar{x})$ *when* S *is single-valued at* $\bar{x}$*, i.e.,* $S(\bar{x}) = \{\bar{u}\}$*.*

Definition 2.7. *Let F be a single-valued mapping defined on $\mathbb{R}^n$, with values in $\mathbb{R}^m$.*

(i) *F is globally Lipschitz continuous if there exists $\kappa \in \mathbb{R}_+ := [0, \infty)$ with*

$$\|F(x') - F(x)\| \le \kappa \|x' - x\| \quad \forall x, x' \in \mathbb{R}^n.$$

Then κ is called a Lipschitz constant for F.

(ii) *F is locally Lipschitz continuous at a point $\bar{x} \in \mathbb{R}^n$ if the value*

$$\operatorname{lip} F(\bar{x}) := \limsup_{x, x' \to \bar{x},\, x \neq x'} \frac{\|F(x') - F(x)\|}{\|x' - x\|}$$

is finite. Here $\operatorname{lip} F(\bar{x})$ is the Lipschitz modulus of F at $\bar{x}$.

(iii) *F is locally Lipschitz continuous if F is locally Lipschitz continuous at every $\bar{x} \in \mathbb{R}^n$.*

Lemma 2.8. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be an extended real-valued function and let $\bar{x}$ be a point with $f(\bar{x})$ finite. Assume that f is locally Lipschitz continuous at $\bar{x}$. The following properties hold:*

(a) *$\partial f(\bar{x})$ is nonempty and compact.*

(b) *$df(\bar{x})(w) = \liminf_{t \downarrow 0} \dfrac{f(\bar{x} + tw) - f(\bar{x})}{t}$.*

(c) *$\overline{\partial} f(\bar{x}) = \operatorname{conv}(\partial f(\bar{x}))$, where $\overline{\partial} f(\bar{x})$ denotes the Clarke subdifferential set of f at $\bar{x}$.*

Proof. **(a-c)** can be found in [31, Theorem 9.13, Exercise 9.15, Theorem 9.61], respectively. □

Lemma 2.9. *Assume that $F : \mathbb{R}^n \to \mathbb{R}^m$ is locally Lipschitz continuous at a point $\bar{x} \in \mathbb{R}^n$. The following properties hold:*

(a) *$D^*F(\bar{x})(0) = \{0\}$, which is also sufficient for F being locally Lipschitz continuous at $\bar{x}$.*

(b) *The mappings $DF(\bar{x})$ and $D^*F(\bar{x})$ are nonempty-valued and locally bounded.*

(c) *$||z|| \le (\operatorname{lip} F(\bar{x}))\, ||w||$ holds for all $(w, z) \in \operatorname{gph}(DF(\bar{x}))$.*

(d) *$||x^*|| \le (\operatorname{lip} F(\bar{x}))\, ||u^*||$ holds for all $(u^*, x^*) \in \operatorname{gph}(D^*F(\bar{x}))$.*

(e) *$z \in DF(\bar{x})(w)$ if and only if there is some $\tau^\nu \downarrow 0$ such that $\frac{F(\bar{x} + \tau^\nu w) - F(\bar{x})}{\tau^\nu} \to z$.*

Proof. **(a)** follows directly from the Mordukhovich criterion [31, Theorem 9.40]. **(b-d)** follow from [31, Proposition 9.24]. **(e)** follows from the definitions of the graphical derivative and the local Lipschitzian continuity. □

Assume now that $F : \mathbb{R}^n \to \mathbb{R}^m$ is a locally Lipschitz continuous function and let D be the subset of $\mathbb{R}^n$ consisting of the points where F is differentiable. By the Rademacher Theorem [31, Theorem 9.60], F is differentiable almost everywhere with $\mathbb{R}^n \backslash D$ being negligible. For each $\bar{x} \in \mathbb{R}^n$, define

$$\overline{\nabla} F(\bar{x}) := \{A \in \mathbb{R}^{m \times n} \mid \exists x^\nu \to \bar{x} \text{ with } x^\nu \in D,\ \nabla F(x^\nu) \to A\}, \tag{2.1}$$

in terms of which, the generalized Jacobian $\overline{\partial} F(x)$ [8, Definition 2.6.1] of F at $\bar{x}$ can be written as

$$\overline{\partial} F(\bar{x}) := \operatorname{conv} \overline{\nabla} F(\bar{x}). \tag{2.2}$$

According to [31, Theorem 9.62], $\overline{\nabla} F(\bar{x})$ is a nonempty, compact set of matrices, and for every $w \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ one has

$$\operatorname{conv} D^*F(\bar{x})(y) = \operatorname{conv}\{A^T y \mid A \in \overline{\nabla} F(\bar{x})\} = \{A^T y \mid A \in \operatorname{conv} \overline{\nabla} F(\bar{x})\} \tag{2.3}$$

and

$$\operatorname{conv} D_*F(\bar{x})(w) = \operatorname{conv}\{Aw \mid A \in \overline{\nabla}F(\bar{x})\} = \{Aw \mid A \in \operatorname{conv} \overline{\nabla}F(\bar{x})\}, \tag{2.4}$$

where $D_*F(\bar{x})$ stands for the strict derivative mapping of F at $\bar{x}$ [31, Definition 9.53], and has the following definition by taking into account that F is locally Lipschitz continuous:

$$D_*F(\bar{x})(w) := \{z \mid \exists \tau^\nu \downarrow 0, x^\nu \to \bar{x} \text{ with } (F(x^\nu + \tau^\nu w) - F(x^\nu))/\tau^\nu \to z\}. \tag{2.5}$$

Note that $D_*F(\bar{x})$ is also known as the Thibault's strict derivative (cf. [35]), and that by definition

$$\operatorname{gph} DF(\bar{x}) \subset \operatorname{gph} D_*F(\bar{x}). \tag{2.6}$$

Definition 2.10. *[10] Let C be a subset of $\mathbb{R}^n$, and let F be a single-valued mapping defined on $\mathbb{R}^n$, with values in $\mathbb{R}^n$. F is said to be coercive on C if*

$$\lim_{x \in C,\, \|x\| \to \infty} \frac{\langle F(x), x - y\rangle}{\|x\|} = +\infty$$

holds for all $y \in C$ (if C is bounded, then F is by convention coercive on C); and F is said to be strongly monotone on C (with modulus $\mu > 0$) if $\langle F(x) - F(y),\ x - y\rangle \geq \mu\|x - y\|^2$ holds for all $x, y \in C$.

3. Subderivatives and subgradients of gap functions. In the remainder of the paper, we make the following blanket assumptions on problem data and some constants, and for the sake of simplicity, we will not mention them in stating a result.

- $K \subset \mathbb{R}^n$ is a nonempty closed and convex set.
- $F : \mathbb{R}^n \to \mathbb{R}^n$ is a locally Lipschitz continuous function.
- a, b, c are fixed positive numbers with $a < b$.

The aim of this section is to study subderivatives and subgradients of f_{ab} and f_c at some $\bar{x}$ by virtue of the graphical derivative $DF(\bar{x})$ and the coderivatives, $D^*F(\bar{x})$ and $\widehat{D}^*F(\bar{x})$, and frequently, the following projection operator associated with F, K and some $\xi > 0$:

$$\pi_\xi(x) := P_K\left(x - \frac{F(x)}{\xi}\right).$$

To begin with, we summarize below some basic properties of the regularized gap function f_c and the D-gap function f_{ab}, most of which can be found in the literature and are useful for further development in the sequel.

Lemma 3.1. *The following properties hold:*

(a) $\frac{b-a}{2}||x - \pi_b(x)||^2 + \frac{a}{2}||\pi_b(x) - \pi_a(x)||^2 \leq f_{ab}(x) \leq \frac{b-a}{2}||x - \pi_a(x)||^2 - \frac{b}{2}||\pi_b(x) - \pi_a(x)||^2$.

(b) $||\pi_b(x) - \pi_a(x)|| \leq \frac{b-a}{a}||x - \pi_a(x)||$ *and* $||x - \pi_b(x)|| \leq ||x - \pi_a(x)|| \leq \frac{b}{a}||x - \pi_b(x)||$.

(c) $x \in \mathbb{R}^n$ *solves (VIP)* $\Leftrightarrow x = \pi_\xi(x)$ *for any* $\xi > 0 \Leftrightarrow f_{ab}(y) \geq f_{ab}(x) = 0$ *for all* $y \in \mathbb{R}^n \Leftrightarrow x \in K$ *and* $f_c(y) \geq f_c(x) = 0$ *for all* $y \in K$.

(d) $\langle a(x - \pi_a(x)) - b(x - \pi_b(x)), \pi_a(x) - \pi_b(x)\rangle \geq 0$.

(e) $\pi_a(x) - \pi_b(x) \in T_{ab}(x, F, K) := T_K(\pi_b(x)) \cap (-T_K(\pi_a(x))) \cap (F(x))^*$.

(f) π_a, π_b, π_c, f_c *and* f_{ab} *are locally Lipschitz continuous. If F is globally Lipschitz continuous, then* π_a, π_b, π_c, f_c *and* f_{ab} *are also globally Lipschitz continuous.*

(g) *The following hold:*

$$\begin{aligned}
&\arg\max_{y\in K}\left\{\langle F(x), x-y\rangle - \tfrac{\xi}{2}||y-x||^2\right\} = \{\pi_\xi(x)\} \quad \forall \xi>0,\\
&f_c(x) = \langle F(x), x-\pi_c(x)\rangle - \tfrac{c}{2}||x-\pi_c(x)||^2,\\
&f_{ab}(x) = \langle F(x), \pi_b(x)-\pi_a(x)\rangle - \tfrac{a}{2}||x-\pi_a(x)||^2 + \tfrac{b}{2}||x-\pi_b(x)||^2.
\end{aligned}$$

Proof. **(a)** and **(b)** can be found in [32, Lemma 1] and [23], respectively. **(c)** can be found in [11] and [36]. **(d)** and **(e)** can be found in [18, Lemma 4.4] or in [10, Theorem 10.3.4]. **(f)** can be found in [19, Lemma 3.1]. **(g)** can be found in [36] or deduced from standard optimality condition for convex programs. This completes the proof. ∎

3.1. Subderivatives and subgradients of f_c. We first present the formulas for the subderivative, the regular subdifferential set and the limiting subdifferential set of f_c at a point $\bar{x}$.

PROPOSITION 3.2. *Let $\bar{x}\in\mathbb{R}^n$ and let $w\in\mathbb{R}^n$. We have the following formulas:*

$$\begin{aligned}
&df_c(\bar{x})(w) = \langle F(\bar{x}),\ w\rangle + \min\langle (DF(\bar{x})-cI)\, w,\ \bar{x}-\pi_c(\bar{x})\rangle,\\
&\widehat{\partial} f_c(\bar{x}) = \left(\widehat{D}^*F(\bar{x})-cI\right)(\bar{x}-\pi_c(\bar{x})) + F(\bar{x}),\\
&\partial f_c(\bar{x}) = \left(D^*F(\bar{x})-cI\right)(\bar{x}-\pi_c(\bar{x})) + F(\bar{x}).
\end{aligned}$$

Proof. Let $w\in\mathbb{R}^n$ be fixed. Since F is locally Lipschitz continuous, it follows from Lemma 2.9 **(b)** and **(e)** that for any continuous function $M:\mathbb{R}\to\mathbb{R}^n$,

$$\liminf_{t\downarrow 0}\langle \frac{F(\bar{x}+tw)-F(\bar{x})}{t},\ M(t)\rangle = \min_{v\in DF(\bar{x})(w)}\langle v,\ M(0)\rangle. \tag{3.1}$$

By Lemma 3.1 **(f)**, f_c is a locally Lipschitz continuous function, which implies by Lemma 2.8 **(b)** that $df_c(\bar{x})(w) = \liminf_{t\downarrow 0}\frac{f_c(\bar{x}+tw)-f_c(\bar{x})}{t}$. In view of Lemma 3.1 **(g)**, we have for all t, $f_c(\bar{x}) \geq \langle F(\bar{x}), \bar{x}-\pi_c(\bar{x}+tw)\rangle - \frac{c}{2}||\bar{x}-\pi_c(\bar{x}+tw)||^2$, and $f_c(\bar{x}+tw) = \langle F(\bar{x}+tw), \bar{x}+tw-\pi_c(\bar{x}+tw)\rangle - \frac{c}{2}||\bar{x}+tw-\pi_c(\bar{x}+tw)||^2$. This, together with (3.1) and the fact that π_c is locally Lipschitz continuous (cf. Lemma 3.1 **(f)**), implies that

$$\begin{aligned}
df_c(\bar{x})(w) \leq\ & \liminf_{t\downarrow 0}\langle \frac{F(\bar{x}+tw)-F(\bar{x})}{t},\ \bar{x}-\pi_c(\bar{x}+tw)\rangle + \lim_{t\downarrow 0}\langle F(\bar{x}+tw),\ w\rangle\\
& + \lim_{t\downarrow 0}\frac{c}{2}\langle 2(\bar{x}-\pi_c(\bar{x}+tw))+tw,\ -w\rangle\\
=\ & \min_{v\in DF(\bar{x})(w)}\langle v, \bar{x}-\pi_c(\bar{x})\rangle + \langle F(\bar{x}),\ w\rangle - c\langle \bar{x}-\pi_c(\bar{x}),\ w\rangle\\
=:\ & \langle F(\bar{x}),\ w\rangle + \min\langle (DF(\bar{x})-cI)\, w,\ \bar{x}-\pi_c(\bar{x})\rangle.
\end{aligned}$$

To prove the inequality in the other direction, we simply follow a similar way by observing from Lemma 3.1 **(g)** that for all t, $f_c(\bar{x}) = \langle F(\bar{x}), \bar{x}-\pi_c(\bar{x})\rangle - \frac{c}{2}||\bar{x}-\pi_c(\bar{x})||^2$, and $f_c(\bar{x}+tw) \geq \langle F(\bar{x}+tw), \bar{x}+tw-\pi_c(\bar{x})\rangle - \frac{c}{2}||\bar{x}+tw-\pi_c(\bar{x})||^2$.

To get the formula for $\widehat{\partial} f_c(\bar{x})$, we resort to the formula for $df_c(\bar{x})$ and the equality in Remark 2.3. Specifically, in terms of $\bar{v} := F(\bar{x}) - c(\bar{x} - \pi_c(\bar{x}))$, we have

$$\begin{aligned} & v \in \widehat{\partial} f_c(\bar{x}) \\ \iff & \langle v, w\rangle \le \langle \bar{v}, w\rangle + \min\langle DF(\bar{x})(w), \bar{x} - \pi_c(\bar{x})\rangle \quad \forall w \in \mathbb{R}^n, \\ \iff & \langle v - \bar{v}, w\rangle \le \langle z, \bar{x} - \pi_c(\bar{x})\rangle \quad \forall (w, z) \in \operatorname{gph}(DF(\bar{x})) = T_{\operatorname{gph} F}(\bar{x}, F(\bar{x})), \\ \iff & (v - \bar{v}, -\bar{x} + \pi_c(\bar{x})) \in (T_{\operatorname{gph} F}(\bar{x}, F(\bar{x})))^* = \widehat{N}_{\operatorname{gph} F}(\bar{x}, F(\bar{x})), \\ \iff & v - \bar{v} \in \widehat{D}^* F(\bar{x})(\bar{x} - \pi_c(\bar{x})). \end{aligned}$$

This gives us the formula for $\widehat{\partial} f_c(\bar{x})$.

To show $\partial f_c(\bar{x}) \subset U := (D^*F(\bar{x}) - cI)\,(\bar{x} - \pi_c(\bar{x})) + F(\bar{x})$, let $v \in \partial f_c(\bar{x})$. Then by the formula for $\widehat{\partial} f_c(x_k)$, there are some $x_k \to \bar{x}$ and $v_k \to v$ such that

$$(v_k - \bar{v}_k, \pi_c(x_k) - x_k) \in \widehat{N}_{\operatorname{gph} F}(x_k, F(x_k)) \quad \forall k,$$

where $\bar{v}_k := F(x_k) - c(x_k - \pi_c(x_k))$. In view of the fact that F and π_c are locally Lipschitz continuous functions (cf. Lemma 3.1 **(f)**), we have $\bar{v}_k \to F(\bar{x}) - c(\bar{x} - \pi_c(\bar{x}))$, $x_k - \pi_c(x_k) \to \bar{x} - \pi_c(\bar{x})$, and hence $(v - F(\bar{x}) + c(\bar{x} - \pi_c(\bar{x})), \pi_c(\bar{x}) - \bar{x}) \in N_{\operatorname{gph} F}(\bar{x}, F(\bar{x}))$, or in other words, $v - F(\bar{x}) + c(\bar{x} - \pi_c(\bar{x})) \in D^*F(\bar{x})(\bar{x} - \pi_c(\bar{x}))$. This verifies that $v \in U$ and hence that $\partial f_c(\bar{x}) \subset U$.

To show $U \subset \partial f_c(\bar{x})$, let $v \in (D^*F(\bar{x}) - cI)\,(\bar{x} - \pi_c(\bar{x})) + F(\bar{x})$. Then we have

$$z \in D^*F(\bar{x})(\bar{x} - \pi_c(\bar{x})) \iff (z, -\bar{x} + \pi_c(\bar{x})) \in N_{\operatorname{gph} F}(\bar{x}, F(\bar{x})),$$

where $z := v + c(\bar{x} - \pi_c(\bar{x})) - F(\bar{x})$. According to the definition of normal cone (cf. Definition 2.1) and the definition of regular coderivative (cf. Definition 2.6), there exist $x_k \to \bar{x}$, $z_k \to z$ and $w_k \to \bar{x} - \pi_c(\bar{x})$ such that for all k,

$$(z_k, -w_k) \in \widehat{N}_{\operatorname{gph} F}(x_k, F(x_k)) \iff (z_k, -w_k) \in (\operatorname{gph} DF(x_k))^*,$$

or explicitly,

$$\langle z_k, w\rangle - \langle x_k - \pi_c(x_k), z\rangle \le \langle w_k - x_k + \pi_c(x_k), z\rangle \quad \forall z \in DF(x_k)(w). \tag{3.2}$$

By the Cauchy-Schwarz inequality and Lemma 2.9 **(c)**, we have for all k,

$$\langle w_k - x_k + \pi_c(x_k), z\rangle \le \epsilon_k \|w\| \quad \forall z \in DF(x_k)(w),$$

where $\epsilon_k := \operatorname{lip} F(x_k)\|w_k - x_k + \pi_c(x_k)\|$. It then follows from (3.2) that for all k,

$$\langle z_k, w\rangle \le \min\langle DF(x_k)(w), x_k - \pi_c(x_k)\rangle + \epsilon_k\|w\| \quad \forall w \in \mathbb{R}^n.$$

By the formula for the subderivative $df_c(x_k)(w)$, we have for all k,

$$\langle z_k - c(x_k - \pi_c(x_k)) + F(x_k),\ w\rangle \le df_c(x_k)(w) + \epsilon_k\|w\| \quad \forall w \in \mathbb{R}^n. \tag{3.3}$$

In view of the fact that F and π_c are locally Lipschitz continuous functions (cf. Lemma 3.1 **(f)**) and by letting $k \to +\infty$, we have $z_k - c(x_k - \pi_c(x_k)) + F(x_k) \to z - c(\bar{x} - \pi_c(\bar{x})) + F(\bar{x}) = v$, and $\epsilon_k \to 0$ (due to $\operatorname{lip} F(\cdot)$ being upper semicontinuous ([31, Theorem 9.2]) and $w_k - x_k + \pi_c(x_k) \to 0$). Then by [31, Proposition 10.46] and (3.3), we have $v \in \partial f_c(\bar{x})$. This completes the proof. ∎

By virtue of the formula for the limiting subdifferential set $\partial f_c(\bar{x})$ in Proposition 3.2, we can easily get the formula for the Clarke subdifferential set $\overline{\partial} f_c(\bar{x})$, which has been obtained first in [37, Lemma 3.2].

COROLLARY 3.3. *Let $\bar{x} \in \mathbb{R}^n$. We have*

$$\overline{\partial} f_c(\bar{x}) = \left(\overline{\partial} F(\bar{x})^T - cI\right)(\bar{x} - \pi_c(\bar{x})) + F(\bar{x}),$$

where $\overline{\partial} F(\bar{x})$ denotes the generalized Jacobian of F at $\bar{x}$ (cf. (2.2)).

Proof. By Lemma 3.1 **(f)** and Lemma 2.8 **(c)**, f_c is locally Lipschitz continuous and hence $\overline{\partial} f_c(\bar{x}) = \mathrm{conv}(\partial f_c(\bar{x}))$. The formula for $\overline{\partial} f_c(\bar{x})$ then follows directly from Proposition 3.2 and the coderivative duality (2.3). This completes the proof. ∎

3.2. Subderivatives and subgradients of f_{ab}. In parallel fashion as we have done in subsection 3.1, we present in this subsection some differential properties of the D-gap function f_{ab}. Most of the proofs are omitted because they are very similar with the corresponding ones in subsection 3.1.

PROPOSITION 3.4. *Let $\bar{x} \in \mathbb{R}^n$ and $w \in \mathbb{R}^n$. We have the following formulas:*

$$\begin{aligned} df_{ab}(\bar{x})(w) &= (b-a)\langle \bar{x} - \pi_a(\bar{x}),\ w\rangle + \min\langle (DF(\bar{x}) - bI)\,w,\ \pi_b(\bar{x}) - \pi_a(\bar{x})\rangle, \\ \widehat{\partial} f_{ab}(\bar{x}) &= \left(\widehat{D}^* F(\bar{x}) - bI\right)(\pi_b(\bar{x}) - \pi_a(\bar{x})) + (b-a)(\bar{x} - \pi_a(\bar{x})), \\ \partial f_{ab}(\bar{x}) &= (D^* F(\bar{x}) - bI)(\pi_b(\bar{x}) - \pi_a(\bar{x})) + (b-a)(\bar{x} - \pi_a(\bar{x})). \end{aligned}$$

Proof. In view of the fact that $f_{ab} = f_a - f_b$ is a locally Lipschitz continuous function, we have

$$df_{ab}(\bar{x})(w) = \liminf_{t \downarrow 0} \left[\frac{f_a(\bar{x} + tw) - f_a(\bar{x})}{t} - \frac{f_b(\bar{x} + tw) - f_b(\bar{x})}{t} \right].$$

According to Lemma 3.1 **(g)**, we have for all t, $f_a(\bar{x}) \geq \langle F(\bar{x}), \bar{x} - \pi_a(\bar{x} + tw)\rangle - \frac{a}{2}||\bar{x} - \pi_a(\bar{x} + tw)||^2$ and $f_b(\bar{x} + tw) \geq \langle F(\bar{x} + tw), \bar{x} + tw - \pi_b(\bar{x})\rangle - \frac{b}{2}||\bar{x} + tw - \pi_b(\bar{x})||^2$. This, together with (3.1) and the fact that π_a and π_b are locally Lipschitz continuous functions (see Lemma 3.1 **(f)**), implies that

$$\begin{aligned} df_{ab}(\bar{x})(w) \leq{} & \liminf_{t \downarrow 0} \langle \frac{F(\bar{x} + tw) - F(\bar{x})}{t}, \pi_b(\bar{x}) - \pi_a(\bar{x} + tw)\rangle \\ & - \lim_{t \downarrow 0} \frac{a}{2} \frac{||\bar{x} + tw - \pi_a(\bar{x} + tw)||^2 - ||\bar{x} - \pi_a(\bar{x} + tw)||^2}{t} \\ & + \lim_{t \downarrow 0} \frac{b}{2} \frac{||\bar{x} + tw - \pi_b(\bar{x})||^2 - ||\bar{x} - \pi_b(\bar{x})||^2}{t} \\ ={} & \langle b(\bar{x} - \pi_b(\bar{x})) - a(\bar{x} - \pi_a(\bar{x})), w\rangle + \min_{v \in DF(\bar{x})(w)} \langle v, \pi_b(\bar{x}) - \pi_a(\bar{x})\rangle. \end{aligned}$$

To prove the inequality in the other direction, we simply follow a similar way by observing from Lemma 3.1 **(g)** that for all t, $f_a(\bar{x} + tv) \geq \langle F(\bar{x} + tv), \bar{x} + tv - \pi_a(\bar{x})\rangle - \frac{a}{2}||\bar{x} + tv - \pi_a(\bar{x})||^2$ and $f_b(\bar{x}) \geq \langle F(\bar{x}), \bar{x} - \pi_b(\bar{x} + tv)\rangle - \frac{b}{2}||\bar{x} - \pi_b(\bar{x} + tv)||^2$. This completes the proof of the formula for $df_{ab}(\bar{x})(w)$. The other two formulas can be obtained in a similar way as we have done in Proposition 3.2. ∎

COROLLARY 3.5. *Let $\bar{x} \in \mathbb{R}^n$. The following properties hold:*

(a) *We have the formula for the Clarke subdifferential set of* f_{ab} *at* $\bar{x}$ *as follows:*

$$\overline{\partial} f_{ab}(\bar{x}) = \left(\overline{\partial} F(\bar{x})^T - bI\right)(\pi_b(\bar{x}) - \pi_a(\bar{x})) + (b-a)(\bar{x} - \pi_a(\bar{x})).$$

(b) $\bar{x}$ *solves (VIP) if and only if* $0 \in \partial f_{ab}(\bar{x})$ *and* $\pi_a(\bar{x}) = \pi_b(\bar{x})$.

Remark 3.6. The formula for $\overline{\partial} f_{ab}(\bar{x})$ was first obtained in [37, Lemma 3.3], and then in [23, Theorem 4.1] and [18, Theorem 3.1] for some generalized D-gap functions. According to the generalized Fermat's rule [31, Theorem 10.1], the condition

$$0 \in \partial f_{ab}(\bar{x}) \tag{3.4}$$

is necessary for $\bar{x}$ to be locally optimal for the optimization problem

$$\min f_{ab}(x) \quad \text{s.t.} \quad x \in \mathbb{R}^n,$$

and hence necessary for $\bar{x}$ to be a solution of (VIP) (cf. Lemma 3.1 **(c)**). Another necessary condition for $\bar{x}$ to be a solution of (VIP) is, by Lemma 3.1 **(c)**, the equality

$$\pi_a(\bar{x}) = \pi_b(\bar{x}). \tag{3.5}$$

Although these two necessary conditions together become sufficient for $\bar{x}$ to be a solution of (VIP), it is interesting to note that either one alone is not sufficient.

To see that (3.4) alone is not enough to guarantee that $\bar{x}$ solves (VIP), we simply consider the case that $K = \mathbb{R}^n$ and F is smooth with $\nabla F(\bar{x})^T F(\bar{x}) = 0$ but $F(\bar{x}) \neq 0$, for which case, (3.4) holds as f_{ab} is smooth with $\nabla f_{ab}(\bar{x}) = \frac{b-a}{ab} \nabla F(\bar{x})^T F(\bar{x}) = 0$, but $\bar{x}$ does not solve (VIP) as $F(\bar{x}) \neq 0$. In this case, (3.5) does not hold as it amount to $F(\bar{x}) = 0$.

To see that (3.5) alone is not enough to guarantee that $\bar{x}$ solves (VIP), we simply consider the case that $K = \mathbb{R}^n_+$ and $\bar{x} \in \mathbb{R}^n$ with $F_i(\bar{x}) \geq 0$ and $\bar{x}_i < 0$ for all i, for which case, (3.5) holds as $\pi_a(\bar{x}) = \pi_b(\bar{x}) = 0$, but $\bar{x}$ does not solve (VIP) as $\bar{x} \notin K$. In this case, (3.4) does not hold as $0 \notin \partial f_{ab}(\bar{x}) = \{(b-a)\bar{x}\}$.

It was shown in [18, Theorem 4.3] that $\bar{x}$ solves (VIP) if and only if $0 \in \overline{\partial} f_{ab}(\bar{x})$ and

$$\left.\begin{array}{c} w \in T_{ab}(x, F, K), \quad Z \in \overline{\partial} F(x) \\ Z^T w \in T_{ab}(x, F, K)^* \end{array}\right\} \Rightarrow F(x)^T w = 0, \tag{3.6}$$

where $T_{ab}(x, F, K)$ is a cone defined as in Lemma 3.1 **(e)**. However, by resorting to Corollary 3.5 **(b)** and noting that $\overline{\partial} f_{ab}(\bar{x}) = \partial f_{ab}(\bar{x})$ in the presence of (3.5), we can refine [18, Theorem 4.3] as follows: $\bar{x}$ solves (VIP) if and only if $0 \in \overline{\partial} f_{ab}(\bar{x})$ and (3.5) holds. Note that $\pi_a(\bar{x})$ and $\pi_b(\bar{x})$ are involved in the definition of $T_{ab}(x, F, K)$. So in contrast to the verification of (3.6), it is much easier to verify (3.5). It is also noteworthy that (3.5) is implied by (3.4) whenever the inequality

$$d(0, \partial f_{ab}(\bar{x})) \geq \mu \|\pi_b(\bar{x}) - \pi_a(\bar{x})\| \tag{3.7}$$

holds for some $\mu > 0$. Inequalities in the form of (3.7) will play a crucial role in the next section.

4. The Kurdyka-Łojasiewicz inequality and error bounds of f_{ab}. In this section, we study the KL inequality and error bounds for the D-gap function f_{ab} by virtue of the formula for the limiting subdifferential sets $\partial f_{ab}(x)$ presented in last section. Before summarizing our main results in Theorem 4.7, we present in Lemmas 4.1-4.4 several results on necessary and sufficient conditions for the following inequalities:

$$d(0, \partial f_{ab}(x)) \geq \mu \|\pi_b(x) - \pi_a(x)\| \quad \forall x \in V,$$

where V is some open set in $\mathbb{R}^n$.

LEMMA 4.1. *Let $x \in \mathbb{R}^n$ and let $\mu > 0$. If $d(0, \partial f_{ab}(x)) \geq \mu \|\pi_b(x) - \pi_a(x)\|$, then*

$$d(0, \partial f_{ab}(x)) \geq \frac{\mu(b-a)}{\mu + b + \operatorname{lip} F(x)} \|x - \pi_a(x)\|. \tag{4.1}$$

Proof. Let $w := \pi_b(x) - \pi_a(x)$ and let $u := x - \pi_a(x)$. By invoking the formula for $\partial f_{ab}(x)$ in Proposition 3.4, we can find some $z^* \in D^*F(x)(w)$ such that $d(0, \partial f_{ab}(x)) = \|z^* - bw + (b-a)u\|$. Then we get (4.1), as we have

$$\begin{aligned} d(0, \partial f_{ab}(x)) &\geq -\|z^*\| - b\|w\| + (b-a)\|u\| \\ &\geq -(b + \operatorname{lip} F(x))\|w\| + (b-a)\|u\| \\ &\geq -\tfrac{b + \operatorname{lip} F(x)}{\mu} d(0, \partial f_{ab}(x)) + (b-a)\|u\|, \end{aligned}$$

where the first inequality follows from the triangle inequality, the second one from Lemma 2.9 **(d)**, and the last one from the assumption that $d(0, \partial f_{ab}(x)) \geq \mu\|w\|$. This completes the proof. ∎

LEMMA 4.2. *Assume that $\operatorname{lip} F(x)$ is bounded from above on a nonempty subset V of $\mathbb{R}^n$, as is true in particular when V is bounded. Then the following properties are equivalent:*
(a) *There is some $\mu > 0$ such that $d(0, \partial f_{ab}(x)) \geq \mu\sqrt{f_{ab}(x)} \quad \forall x \in V$.*
(b) *There is some $\mu > 0$ such that $d(0, \partial f_{ab}(x)) \geq \mu\|x - \pi_a(x)\| \quad \forall x \in V$.*
(c) *There is some $\mu > 0$ such that $d(0, \partial f_{ab}(x)) \geq \mu\|\pi_b(x) - \pi_a(x)\| \quad \forall x \in V$.*
Therefore, f_{ab} satisfies the KL inequality at any solution $\bar{x}$ of (VIP) with an exponent of $\frac{1}{2}$ if and only if any of **(a)**, **(b)** *and* **(c)** *holds with V being some neighborhood of $\bar{x}$.*

Proof. The relations **(a)** $\iff$ **(b)** $\implies$ **(c)** follow directly from Lemma 3.1 **(a)**. As $\operatorname{lip} F(x)$ is upper semicontinuous ([31, Theorem 9.2]), it follows from [31, Corollary 1.10] that $\operatorname{lip} F(x)$ is bounded from above on each bounded subset of $\mathbb{R}^n$. We now show **(c)** $\implies$ **(b)** by assuming that **(c)** holds with some $\mu > 0$ and that there is some $L > 0$ such that $\operatorname{lip} F(x) \leq L \; \forall x \in V$. By Lemma 4.1, we get **(b)** as we have

$$d(0, \partial f_{ab}(x)) \geq \frac{\mu(b-a)}{\mu + b + \operatorname{lip} F(x)}\|x - \pi_a(x)\| \geq \frac{\mu(b-a)}{\mu + b + L}\|x - \pi_a(x)\| \quad \forall x \in V.$$

Let $\bar{x}$ be a solution of (VIP). We first note that f_{ab} is locally Lipschitz continuous with $f_{ab} \geq 0$ and $f_{ab}(\bar{x}) = 0$ (cf. Lemma 3.1 **(c)**). Then f_{ab} satisfies the KL inequality at $\bar{x}$ with an exponent of $\frac{1}{2}$ if, according to Definition 2.4, **(a)** holds with V being some bounded neighborhood of $\bar{x}$. By the previous argument, **(a)**, **(b)** and **(c)** are equivalent whenever V is bounded, and therefore the last assertion is true. This completes the proof. ∎

LEMMA 4.3. *Assume that the solution set of (VIP) is nonempty. If there are some $\mu \in (0, +\infty)$ and $\varepsilon \in (0, +\infty]$ such that*

$$d\left(0, \partial f_{ab}(x)\right) \geq \mu\|\pi_b(x) - \pi_a(x)\| \quad \forall x \in [f_{ab} < \varepsilon], \tag{4.2}$$

and

$$L := \sup_{x\in[0<f_{ab}<\varepsilon]} \operatorname{lip} F(x) < +\infty, \tag{4.3}$$

then

$$\sqrt{\frac{b-a}{2}}\frac{\mu}{\mu+b+L}\, d\left(x, [f_{ab}\le\theta]\right) \le \left(\sqrt{f_{ab}(x)}-\sqrt{\theta}\right)_+ \quad \forall\theta\in[0,\varepsilon),\ \forall x\in[f_{ab}<\varepsilon], \tag{4.4}$$

which, in particular, implies the following error bound property:

$$\sqrt{\frac{b-a}{2}}\frac{\mu}{\mu+b+L}\, d\left(x, [f_{ab}\le 0]\right) \le \sqrt{f_{ab}(x)} \quad \forall x\in[f_{ab}\le\varepsilon].$$

Proof. It suffices to show (4.4) by assuming (4.2) and (4.3) for some given $\mu\in(0,+\infty)$ and $\varepsilon\in(0,+\infty]$. As the solution set of (VIP) is nonempty, we deduce from Lemma 3.1 **(c)** that $[f_{ab}\le 0]\neq\emptyset$. In what follows, we assume that $[0<f_{ab}<\varepsilon]$ is nonempty, for otherwise (4.4) holds trivially. Fix any $x\in[0<f_{ab}<\varepsilon]$. In view of (4.2) and (4.3), we get from Lemma 4.1 that $d(0,\partial f_{ab}(x))\ge\frac{\mu(b-a)}{\mu+b+L}\|x-\pi_a(x)\|$. Then by Lemma 3.1 **(a)**, we have $d(0,\partial f_{ab}(x))\ge\frac{\mu\sqrt{2(b-a)}}{\mu+b+L}\sqrt{f_{ab}(x)}$. By some direct calculation, we have $\partial\sqrt{f_{ab}}(x)=\frac{\partial f_{ab}(x)}{2\sqrt{f_{ab}}(x)}$ and hence $d\left(0,\partial\sqrt{f_{ab}}(x)\right)\ge\sqrt{\frac{b-a}{2}}\frac{\mu}{\mu+b+L}$. Then by [21, Lemma 2.1 **(ii')**], we have

$$|\nabla\sqrt{f_{ab}}|(x)\ge\sqrt{\frac{b-a}{2}}\frac{\mu}{\mu+b+L},$$

where for a function $f:\mathbb{R}^n\to\mathbb{R}$ and a point $\bar{y}\in\mathbb{R}^n$,

$$|\nabla f|(\bar y) := \limsup_{y\to\bar y,\ y\neq\bar y}\frac{(f(\bar y)-f(y))_+}{\|y-\bar y\|}$$

denotes the the strong slope of f at $\bar y$, introduced by De Giorgi et al. [12]. As $x\in[0<f_{ab}<\varepsilon]$ is chosen arbitrarily, we can apply [5, Theorem 2.1] to deduce that

$$\begin{aligned}\inf_{0\le\sqrt{\theta}<\sqrt{\varepsilon}}\ \inf_{x\in[\sqrt{\theta}<\sqrt{f_{ab}}<\sqrt{\varepsilon}]}\frac{\sqrt{f_{ab}(x)}-\sqrt{\theta}}{d\left(x,\left[\sqrt{f_{ab}}\le\sqrt{\theta}\right]\right)} &= \inf_{x\in\left[0<\sqrt{f_{ab}}<\sqrt{\varepsilon}\right]}|\nabla\sqrt{f_{ab}}|(x)\\ &\ge\sqrt{\frac{b-a}{2}}\frac{\mu}{\mu+b+L},\end{aligned}$$

from which, (4.4) follows readily. This completes the proof. ∎

Many existing conditions in the literature are sufficient for Lemma 4.2 **(c)** or (4.2), as can be seen from the following lemma, where we also provide a new sufficient condition which can be considered as some restricted strong monotonicity.

LEMMA 4.4. *Let $\mu>0$ and let $V\subset\mathbb{R}^n$ be open. Consider the following properties:*

(a) *F is strongly monotone on V with modulus μ, which holds in the case of V being convex if and only if the following inequality holds for all $x \in V$ where F is differentiable:*

$$\langle \nabla F(x)w,\ w\rangle \geq \mu\|w\|^2 \quad \forall w \in \mathbb{R}^n. \tag{4.5}$$

(b) *The following holds for all $x \in V$ where F is differentiable and $f_{ab}(x) > 0$:*

$$\langle \nabla F(x)w,\ w\rangle \geq \mu\|w\|^2 \quad \forall w \in T_{ab}(x, F, K).$$

(c) *The following holds for all $x \in V$ where F is differentiable:*

$$\langle \nabla F(x)(\pi_a(x) - \pi_b(x)),\ \pi_a(x) - \pi_b(x)\rangle \geq \mu\|\pi_a(x) - \pi_b(x)\|^2.$$

(d) $d(0, \partial f_{ab}(x)) \geq \mu\|\pi_b(x) - \pi_a(x)\| \quad \forall x \in V$.

We have **(a)** $\Longrightarrow$ **(b)** $\Longrightarrow$ **(c)** $\Longrightarrow$ **(d)**.

Proof. According to [14, Proposition 2.3 **(b)**], the following holds for all $x \in V$:

$$\langle Zw,\ w\rangle \geq \mu\|w\|^2 \quad \forall Z \in \overline{\nabla} F(x),\ \forall w \in \mathbb{R}^n, \tag{4.6}$$

if F is strongly monotone on V with modulus μ, and the converse is true whenever V is convex. As $\nabla F(x) \in \overline{\nabla} F(x)$ when F is differentiable at x, (4.5) is implied by (4.6). To show that (4.6) is implied by (4.5), let $x \in V$ and let $Z \in \overline{\nabla} F(x)$. By the definition of $\overline{\nabla} F(x)$ (cf. (2.1)), there is $x_k \to x$ such that F is differentiable at x_k for all k and $\nabla F(x_k) \to Z$. Then by (4.5), we have for all sufficiently large k: $\langle \nabla F(x_k)w,\ w\rangle \geq \mu\|w\|^2 \quad \forall w \in \mathbb{R}^n$, which implies (4.6) by letting $k \to \infty$.

By the previous argument, we get **(b)** from **(a)** in a straightforward way. To get **(c)** from **(b)**, it suffices to note the following facts: (1) $\pi_a(x) - \pi_b(x) \in T_{ab}(x, F, K)$ (cf. Lemma 3.1 **(e)**); (2) $\pi_a(x) = \pi_b(x)$ whenever $f_{ab}(x) = 0$ (cf. Lemma 3.1 **(c)**).

We now show (**c**) $\Longrightarrow$ (**d**). Let $x \in V$. Set $w := \pi_b(x) - \pi_a(x)$ and $u := x - \pi_a(x)$. We first claim that the following holds for all $z^* \in \operatorname{conv} D^*F(x)(w)$:

$$\langle z^*, w\rangle \geqslant \mu\|w\|^2. \tag{4.7}$$

By the coderivative duality (2.3) for a locally Lipschitz continuous mapping, we have $z^* \in \{A^T w \mid A \in \operatorname{conv} \overline{\nabla} F(x)\}$. Then there exist a positive integer r and some $A^i \in \overline{\nabla} F(x)$ such that

$$z^* = \left(\sum_{i=1}^r \lambda^i A^i\right)^T w = \sum_{i=1}^r \lambda^i \left(A^i\right)^T w, \tag{4.8}$$

where $\lambda^i \geq 0$ for all i and $\sum_{i=1}^r \lambda^i = 1$. For each $A^i \in \overline{\nabla} F(x)$, there exists by definition some sequence $\{x_k^i\}$ such that F is differentiable at x_k^i for all k, $x_k^i \to x$ and $\nabla F(x_k^i) \to A^i$ as $k \to \infty$. Then by **(c)**, we have for all k large enough,

$$\langle \nabla F(x_k^i)(\pi_a(x_k^i) - \pi_b(x_k^i)), \pi_a(x_k^i) - \pi_b(x_k^i)\rangle \geqslant \mu\|\pi_b(x_k^i) - \pi_a(x_k^i)\|^2.$$

Thus, by noting that π_a and π_b are locally Lipschitz continuous and letting $k \to \infty$, we get $\langle A^i(\pi_a(x) - \pi_b(x)), \pi_a(x) - \pi_b(x)\rangle \geqslant \mu\|\pi_b(x) - \pi_a(x)\|^2$, or in terms of w, $\langle (A^i)^T w, w\rangle \geq \mu\|w\|^2$. This, together with (4.8), yields (4.7).

By invoking the formula for $\partial f_{ab}(x)$ in Proposition 3.4, we can find some $\bar{z}^* \in D^*F(x)(w) \subset \operatorname{conv} D^*F(x)(w)$ such that $d(0, \partial f_{ab}(x)) = \|\bar{z}^* - bw + (b-a)u\|$. Then we get **(d)**, as we have $d(0, \partial f_{ab}(x))\,\|w\| \geq \langle \bar{z}^* - bw + (b-a)u,\ w\rangle \geq \langle \bar{z}^*, w\rangle \geq \mu\|w\|^2$, where the first inequality follows from the Cauchy-Schwarz inequality, the second one from Lemma 3.1 **(d)**, and the last one from (4.7). This completes the proof. ∎

Remark 4.5. As $\nabla F(x) \in \overline{\nabla} F(x) \subset \overline{\partial} F(x)$ when F is differentiable at x, Lemma 4.4 **(b)** holds if the following holds for all $x \in V$ with $f_{ab}(x) > 0$:

$$\langle Z^T w, w\rangle \geq \mu ||w||^2 \quad \forall Z \in \overline{\partial} F(x),\ \forall w \in T_{ab}(x, F, K). \tag{4.9}$$

When $V = \mathbb{R}^n$, the supremum of all possible positive μ satisfying (4.9) can be reformulated as

$$\mu_{ab} := \inf\{w^T Z w \mid Z \in \overline{\partial} F(x),\ w \in T_{ab}(x, F, K),\ \|w\| = 1,\ f_{ab}(x) > 0\}. \tag{4.10}$$

The quantity μ_{ab} was first introduced for a general case in [18, Theorem 4.2], where the condition $\mu_{ab} > 0$ was utilized to study the local error bounds for f_{ab}.

Remark 4.6. Lemma 4.4 **(c)** can be reformulated as

$$\langle z^*, \pi_b(x) - \pi_a(x)\rangle \geq \mu||\pi_a(x) - \pi_b(x)||^2 \quad \forall x \in V,\ z^* \in \operatorname{conv} D^*F(x)(\pi_b(x) - \pi_a(x)), \tag{4.11}$$

or

$$\langle z, \pi_a(x) - \pi_b(x)\rangle \geq \mu||\pi_a(x) - \pi_b(x)||^2 \quad \forall x \in V,\ z \in \operatorname{conv} D_*F(x)(\pi_a(x) - \pi_b(x)), \tag{4.12}$$

where $D_*F(x)$ stands for the strict derivative mapping of F at x (cf. (2.5)). As

$$\nabla F(x)^T(\pi_b(x) - \pi_a(x)) \in \operatorname{conv} D^*F(x)(\pi_b(x) - \pi_a(x))$$

and

$$\nabla F(x)(\pi_a(x) - \pi_b(x)) \in \operatorname{conv} D_*F(x)(\pi_a(x) - \pi_b(x))$$

whenever F is differentiable at x (cf. (2.3) and (2.4)), Lemma 4.4 **(c)** is clearly implied by (4.11) or (4.12). In the proof of (**c**) $\Longrightarrow$ (**d**) in Lemma 4.4, we have already shown that (4.11) is implied by Lemma 4.4 **(c)**. By the coderivative duality (2.4) for a locally Lipschitz continuous mapping, we can show in a similar way that (4.12) is also implied by Lemma 4.4 **(c)**.

Example 1. *Let $A \in \mathbb{R}^{n\times n}$ and $q \in \mathbb{R}^n$ be such that $q + \operatorname{rge} A \neq \{0\}$, where $\operatorname{rge} A$ denotes the range space of A. Consider a (VIP) instance with $K = \mathbb{R}^n$ and $F(x) = Ax + q$. In this case, to find a solution of (VIP) is to find a solution to the linear equation $Ax + q = 0$, which exists if and only if $q \in \operatorname{rge} A$. Clearly, F is continuously differentiable on $\mathbb{R}^n$ with $\nabla F(\cdot) = A$, implying that f_{ab} is continuously differentiable on $\mathbb{R}^n$. By some direct computation we have*

$$\pi_b(x) - \pi_a(x) = \frac{b-a}{ab}(Ax + q), \quad f_{ab}(x) = \frac{b-a}{2ab}\|Ax + q\|^2,$$

and

$$\nabla f_{ab}(x) = \frac{b-a}{ab}A^T(Ax + q), \quad T_{ab}(x, F, K) = \{w \mid \langle Ax + q, w\rangle \leq 0\}.$$

Then in the case of $V := \mathbb{R}^n$, Lemma 4.4 **(a)-(d)** *can be reduced respectively to the following:*

(a) *$A - \mu I$ is positive-semidefinite on $\mathbb{R}^n$.*

(b) *$A - \mu I$ is positive-semidefinite on at least one closed-half space containing the origin and hence on the whole space $\mathbb{R}^n$.*
(Therefore, **(a)** *and* **(b)** *coincide, both of which implies that A is positive-definite on $\mathbb{R}^n$ and that the linear equation $Ax+q = 0$ has a unique solution.)*

(c) *$A-\mu I$ is positive-semidefinite on the linear subspace $\mathbb{R}\{q\}+\operatorname{rge} A$, which entails positive-semidefiniteness of $A^TAA-\mu A^TA$ on $\mathbb{R}^n$ and is equivalent to it when $q \in \operatorname{rge} A$. (The latter property can be fulfilled for a symmetric matrix A if and only if A is positive-semidefinite and $0<\mu<\lambda_i$ with λ_i being any positive eigenvalue of A.)*

(d) *$AA^T-\mu^2 I$ is positive-semidefinite on the linear subspace $\mathbb{R}\{q\}+\operatorname{rge} A$, which entails positive-semidefiniteness of $(A^TA)^2-\mu^2A^TA$ on $\mathbb{R}^n$ and is equivalent to it when $q\in\operatorname{rge} A$. (The latter property can be fulfilled as long as $0<\mu\le\sqrt{\lambda_i}$ with λ_i being any positive eigenvalue of A^TA.)*

Therefore, in the case of $q\in\operatorname{rge} A$ with A being symmetric and positive-semidefinite (but not positive-definite), Lemma 4.4 **(a)-(b)** *cannot hold, but Lemma 4.4* **(c)** *can as long as $0<\mu<\lambda_i$ with λ_i being any positive eigenvalue of A. This demonstrates that Lemma 4.4* **(c)** *can be strictly weaker than Lemma 4.4* **(a)-(b)**. *While in the case of $q\in\operatorname{rge} A$ with A being symmetric but not positive-semidefinite, Lemma 4.4* **(c)** *cannot hold, but Lemma 4.4* **(d)** *can as long as μ is less than or equal to the square root of the smallest positive eigenvalue of A^TA. This demonstrates that Lemma 4.4* **(d)** *can be strictly weaker than Lemma 4.4* **(c)**.

THEOREM 4.7. *Assume that any of* **(a)-(d)** *in Lemma 4.4 holds with some $\mu>0$ and $V=\mathbb{R}^n$. Then the following properties hold:*

(a) *f_{ab} is a KL function with an exponent of $\frac{1}{2}$.*

(b) *If F is coercive on $\mathbb{R}^n$, then the solution set of (VIP) is nonempty and compact, and $\sqrt{f_{ab}}$ has a local error bound on $\mathbb{R}^n$, i.e., the following holds for any given $\varepsilon>0$:*

$$\sqrt{\frac{b-a}{2}}\frac{\mu}{\mu+b+L}\, d\,(x,\,[f_{ab}\le 0]\,)\le\sqrt{f_{ab}(x)}\quad\forall x\in[f_{ab}\le\varepsilon].$$

where L is any number such that $L\ge\operatorname{lip}F(x)$ for all $x\in[0<f_{ab}<\varepsilon]$.

(c) *If the solution set of (VIP) is nonempty and F is globally Lipschitz continuous with a constant $L>0$, then $\sqrt{f_{ab}}$ has a global error bound on $\mathbb{R}^n$, i.e., the following holds:*

$$\sqrt{\frac{b-a}{2}}\frac{\mu}{\mu+b+L}\, d\,(x,[f_{ab}\le 0])\le\sqrt{f_{ab}(x)}\quad\forall x\in\mathbb{R}^n.$$

Proof. For each x that is a solution of (VIP), it follows from Lemma 4.2 that f_{ab} is a KL function at x with an exponent of $\frac{1}{2}$. For each x that is not a solution of (VIP), we claim that $0\notin\partial f_{ab}(x)$ and hence f_{ab} is a KL function at x with an exponent of 0, for otherwise the inclusion $0\in\partial f_{ab}(x)$, together with the equality $\pi_a(x)=\pi_b(x)$ as can be guaranteed by Lemma 4.4 **(d)**, would imply that x is a solution of (VIP) (cf. Corollary 3.5 **(b)**). As a whole f_{ab} is indeed a KL function with an exponent of $\frac{1}{2}$. This verifies **(a)**.

To show **(b)**, fix any $\varepsilon>0$ and let $\bar{L}:=\sup_{x\in[0<f_{ab}<\varepsilon]}\operatorname{lip}F(x)$. By the coerciveness of F on $\mathbb{R}^n$ (hence on K), the solution set of (VIP) is nonempty and compact (cf. [10, Proposition 2.2.7]), and the level set $[f_{ab}\le\varepsilon]$ is bounded (cf. [18, Lemma 4.1]). As $\operatorname{lip}F(x)$ is upper semicontinuous (cf. [31, Theorem 9.2]), it follows from [31, Corollary 1.10] that $\operatorname{lip}F(x)$ is bounded from above on each bounded subset of $\mathbb{R}^n$. So we have $\bar{L}<+\infty$. Then by Lemma 4.3, we get **(b)** in a straightforward way.

To show **(c)**, we apply Lemma 4.3 again by noting that

$$\sup_{x\in[0<f_{ab}<+\infty]}\operatorname{lip}F(x)\le L.$$

This completes the proof. ∎

Remark 4.8. In the presence of Lemma 4.4 **(a)** with some $\mu > 0$ and $V = \mathbb{R}^n$ (i.e., F is strongly monotone on $\mathbb{R}^n$ with modulus μ), it was pointed out by [18, Remark 2.1 (ii)] that F is coercive on $\mathbb{R}^n$. In this case, Theorem 4.7 **(b)** holds without explicitly assuming coerciveness. While in the presence of Lemma 4.4 **(b)** with $V = \mathbb{R}^n$ and some $\mu > 0$, Theorem 4.7 **(b)** can be deduced from [18, Theorem 4.2](cf. Remark 4.5). To the best of our knowledge, all the results in Theorem 4.7, except for the mentioned ones, are new.

EXAMPLE 2 ([18], Example 4.4). *Consider a (VIP) instance with $K = \mathbb{R}^2_+$ and $F : \mathbb{R}^2 \to \mathbb{R}^2$ being given by $F(x) = \left(x_1 + (x_1)_+(x_2)_+, \quad x_2 + \frac{3}{2}(x_1)_+\right)^T$. Clearly, F is differentiable at $x \in \mathbb{R}^2$ if and only if $x_1x_2 \neq 0$, and moreover,*

$$\nabla F(x) = \begin{cases} \begin{pmatrix} 1+x_2 & x_1 \\ \frac{3}{2} & 1 \end{pmatrix} & \text{if } x_1 > 0, x_2 > 0, \\ \begin{pmatrix} 1 & 0 \\ \frac{3}{2} & 1 \end{pmatrix} & \text{if } x_1 > 0, x_2 < 0, \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & \text{if } x_1 < 0, x_2 \neq 0. \end{cases}$$

Let $a \in (0,1)$ and $b = 1$. According to [18, Example 4.4], F is coercive on $\mathbb{R}^2$, $\sqrt{f_{ab}}$ has a local error bound on $\mathbb{R}^2$ (with some error bound modulus expressed in an abstract way), and $\mu_{ab} \geq 1$, where μ_{ab} is defined by (4.10).

In what follows, by virtue of Lemma 4.4 **(c)**, *we can show that $\mu_{ab} = 1$ and that some error bound modulus expressed in an explicit way can be provided. First, by some direct calculation, we have $\pi_b(x) = (0,0)^T$ for all $x \in \mathbb{R}^2$ and*

$$\pi_a(x) - \pi_b(x) = \begin{cases} \left(\dfrac{a-1}{a}x_1,\ 0\right)^T & \text{if } x_1 \leq 0, x_2 \geq 0, \\ \left(\dfrac{a-1}{a}x_1,\ \dfrac{a-1}{a}x_2\right)^T & \text{if } x_1 \leq 0, x_2 \leq 0, \\ \left(0,\ \dfrac{a-1}{a}x_2 - \dfrac{3}{2a}x_1\right)^T & \text{if } 0 \leq x_1 \leq \frac{2(a-1)}{3}x_2, \\ (0,\ 0)^T & \text{otherwise.} \end{cases}$$

Then it is straightforward to verify that the inequality

$$\langle \nabla F(x)(\pi_a(x) - \pi_b(x)),\ \pi_a(x) - \pi_b(x)\rangle \geq \mu||\pi_a(x) - \pi_b(x)||^2$$

holds for all $x \in \mathbb{R}^2$ with $x_1x_2 \neq 0$ if and only if $0 < \mu \leq 1$. That is, Lemma 4.4 **(c)** *holds with $V = \mathbb{R}^2$ if and only if $0 < \mu \leq 1$. As Lemma 4.4* **(c)** *is implied by Lemma 4.4* **(b)**, *we deduce that Lemma 4.4* **(b)** *cannot hold with $V = \mathbb{R}^2$ and $\mu > 1$, which implies that μ_{ab} cannot be greater than 1 (cf. Remark 4.5). Therefore, we confirm that $\mu_{ab} = 1$. Furthermore, we can apply Theorem 4.7 to get the following: (i) f_{ab} is a KL function with an exponent of $\frac{1}{2}$; (ii) $\sqrt{f_{ab}}$ has a local error bound on $\mathbb{R}^2$, i.e., for any given $\varepsilon > 0$,*

$$\sqrt{\frac{b-a}{2}}\frac{1}{1+b+L}\, d\left(x, [f_{ab} \leq 0]\right) \leq \sqrt{f_{ab}(x)} \quad \forall x \in [f_{ab} \leq \varepsilon],$$

where L is any number such that $L \geq \sup_{x \in [0 < f_{ab} < \varepsilon]} \operatorname{lip} F(x)$.

5. A derivative free descent method for (VIP). In this section, we analyze the convergence behavior of the following descent algorithm with an Armijo line search, which is essentially the same as those studied in [15, 18, 29, 30, 37, 39], especially the same in the way how descent directions are chosen.

Algorithm

Step 1. Set $0 < a < b$ and $0 < \rho < 1$. Choose three positive constants α, β, τ such that β and τ are small and that α is close to $b - a$. Select a start point $x_0 \in \mathbb{R}^n$, and set $n = 0$.

Step 2. If $f_{ab}(x_n) = 0$, stop. Otherwise, go to Step 3.

Step 3. Let $u_n = \pi_a(x_n) - x_n$ and $w_n = \pi_a(x_n) - \pi_b(x_n)$. If $\beta||u_n|| < ||w_n||$, set $d_n = w_n$ and select m_n as the smallest nonnegative integer m such that

$$f_{ab}(x_n + \rho^m d_n) - f_{ab}(x_n) \leq -\tau\rho^m ||d_n||^2. \tag{5.1}$$

Otherwise, set $d_n = u_n$ and select m_n as the smallest nonnegative integer m such that

$$f_{ab}(x_n + \rho^m d_n) - f_{ab}(x_n) \leq -(b - a - \alpha)\,\rho^m ||d_n||^2. \tag{5.2}$$

Step 4. Set $t_n = \rho^{m_n}$, $x_{n+1} = x_n + t_n d_n$ and $n = n + 1$, and go to Step 2.

In what follows, we make the following assumptions.

Assumption (i) The level set $[f_{ab} \leq f_{ab}(x_0)]$ is bounded, which can be guaranteed by the coerciveness of F on $\mathbb{R}^n$ as pointed out by [18, Lemma 4.1].

Assumption (ii) F is globally Lipschitz continuous with a constant $L > 0$ (implying that f_{ab}, π_a and π_b are all globally Lipschitz continuous).

Assumption (iii) There exists some $\mu^* > 0$ such that the inequality

$$\langle \nabla F(x)(\pi_a(x) - \pi_b(x)),\ \pi_a(x) - \pi_b(x)\rangle \geq \mu^* ||\pi_a(x) - \pi_b(x)||^2$$

holds for all $x \in \mathbb{R}^n$ where F is differentiable. This implies by Theorem 4.7 that f is a KL function with an exponent of $\frac{1}{2}$, and by Remark 4.6 and (2.6) that

$$\min_{z \in DF(x)(\pi_a(x) - \pi_b(x))} \langle z, \pi_a(x) - \pi_b(x)\rangle \geq \mu^* ||\pi_a(x) - \pi_b(x)||^2 \quad \forall x \in \mathbb{R}^n.$$

Assumption (iv) The parameters α, β, τ in the Algorithm are chosen such that

$$0 < \beta < \frac{b - a}{b + L}, \quad (b + L)\beta < \alpha < b - a, \quad 0 < \tau < \mu^*.$$

To begin with, we give two technical lemmas, which are helpful for our further analysis.

Lemma 5.1. *Under* **Assumption (ii)**, *we have*

$$\|v\| \leq (b + L)\|\pi_b(x) - \pi_a(x)\| + (b - a)\|x - \pi_a(x)\| \quad \forall x \in \mathbb{R}^n,\ \forall v \in \partial f_{ab}(x).$$

Proof. In view of Lemma 2.9 **(d)** and **Assumption (ii)**, we get this result directly from the formula for $\partial f_{ab}(x)$ presented in Proposition 3.4. The proof is completed. ∎

Lemma 5.2. *Consider a locally Lipschitz continuous function $g : \mathbb{R}^n \to \mathbb{R}$. For some $x \in \mathbb{R}^n$ and $w \in \mathbb{R}^n \backslash \{0\}$, assume that there are some $\sigma > 0$ and $0 < t_0 < t_1$ such that*

$$g(x + t_0 w) - g(x) \leq -\sigma t_0 ||w||^2 \ \text{ and } \ g(x + t_1 w) - g(x) > -\sigma t_1 ||w||^2.$$

Then there exist some $\theta^* \in (0, 1)$ and $v^* \in \partial g(x + \theta^* t_1 w)$ such that

$$g(x + t_1 w) - g(x) = t_1 \langle v^*, w\rangle.$$

Proof. Define $\varphi : \mathbb{R} \to \mathbb{R}$ by $\varphi(\theta) := g(x + \theta t_1 w) - g(x) + \theta[g(x) - g(x + t_1 w)]$. Clearly, φ is locally Lipschitz continuous, and $\varphi(0) = \varphi(1) = 0$. Moreover, it follows from the assumption that $\varphi(t_0/t_1) = g(x+t_0 w) - g(x) + (t_0/t_1)[g(x) - g(x+t_1 w)] < 0$. This entails the existence of at least one $\theta^* \in (0, 1)$ such that φ attains its minimum over $[0, 1]$ at θ^*, implying by the Fermat's rule that $0 \in \partial\varphi(\theta^*)$. In view of the local Lipschitzian continuity of g, we get from the calculus rules [31, Exercise 8.8 and Theorem 10.6] that $\partial\varphi(\theta^*) \subset g(x) - g(x+t_1 w) + \{t_1\langle v, w\rangle \mid v \in \partial g(x + \theta^* t_1 w)\}$. This completes the proof. ∎

PROPOSITION 5.3. *Under* **Assumptions (ii)-(iv)**, *Step 3 of the Algorithm is well defined.*

Proof. To show that Step 3 in the Algorithm is well defined, it suffices to show that if $\beta||u_n|| < ||w_n||$, $-d(-f_{ab})(x_n)(w_n) < -\tau\|w_n\|^2$, and if $\beta||u_n|| \geq ||w_n||$, $-d(-f_{ab})(x_n)(u_n) < -(b - a - \alpha)\|u_n\|^2$. Following from the proof of the formula for $df_{ab}(\bar{x})(w)$ in Proposition 3.4, we get the formula for the subderivative of $-f_{ab}$ at a point $\bar{x} \in \mathbb{R}^n$ as follows:

$$-d(-f_{ab})(\bar{x})(w) = (b - a)\langle \bar{x} - \pi_a(\bar{x}),\ w\rangle - \min\langle (DF(\bar{x}) - bI)\, w,\ -\pi_b(\bar{x}) + \pi_a(\bar{x})\rangle.$$

In the case of $\beta||u_n|| < ||w_n||$, we have

$$\begin{aligned} & -d(-f_{ab})(x_n)(w_n) \\ = \;& \langle b(x_n - \pi_b(x_n)) - a(x_n - \pi_a(x_n)), w_n\rangle - \min\nolimits_{z \in DF(x_n)(w_n)}\langle z, w_n\rangle \\ \leq \;& -\min\nolimits_{z \in DF(x_n)(w_n)}\langle z, w_n\rangle \\ \leq \;& -\mu^*||w_n||^2 \\ < \;& -\tau||w_n||^2, \end{aligned}$$

where the first inequality follows from Lemma 3.1 **(d)**, the second inequality follows from **Assumption (iii)**, and the third inequality follows from **Assumption (iv)**. In the case of $\beta||u_n|| \geq ||w_n||$, we have

$$\begin{aligned} & -d(-f_{ab})(x_n)(u_n) \\ = \;& \langle b(x_n - \pi_b(x_n)) - a(x_n - \pi_a(x_n)), u_n\rangle - \min\nolimits_{z \in DF(x_n)(u_n)}\langle z, w_n\rangle \\ = \;& -(b - a)||u_n||^2 + b\langle \pi_a(x_n) - \pi_b(x_n), u_n\rangle + \max\nolimits_{z \in DF(x_n)(u_n)}\langle z, -w_n\rangle \\ \leq \;& -[(b - a) - b\beta]||u_n||^2 + \max\nolimits_{z \in DF(x_n)(u_n)}\langle z, -w_n\rangle \\ \leq \;& -[(b - a) - b\beta]||u_n||^2 + L||u_n|| \cdot ||w_n|| \\ \leq \;& -[(b - a) - (b + L)\beta]||u_n||^2 \\ < \;& -[(b - a) - \alpha]||u_n||^2, \end{aligned}$$

where the first inequality follows by using the Cauchy-Schwarz inequality and the inequality $\beta||u_n|| \geq ||w_n||$, the second inequality follows from Lemma 2.9 **(c)** and **Assumption (ii)**, the third inequality follows from the inequality $\beta||u_n|| \geq ||w_n||$, and the last inequality follows from **Assumption (iv)**. This completes the proof. ∎

PROPOSITION 5.4. *Assume that the sequence $\{x_n\}$ generated by the Algorithm satisfies $f_{ab}(x_n) > 0$ for all n. Under* **Assumptions (ii)-(iv)**, *there is some $t^* > 0$ such that $t_n \geq t^*$ for all n, i.e., the step length sequence $\{t_n\}$ generated by the Algorithm has a lower bound.*

Proof. Recall that in Step 3 of the Algorithm, we set $u_n := \pi_a(x_n) - x_n$, $w_n := \pi_a(x_n) - \pi_b(x_n)$, and $d_n := u_n$ if $\beta\|u_n\| \geq \|w_n\|$, and $d_n := w_n$ if $\beta\|u_n\| < \|w_n\|$. In view of the setting for d_n and our assumption that $f_{ab}(x_n) > 0$ for all n, we get from Lemma 3.1 **(c)** that $d_n \neq 0$ for all n.

Suppose by contradiction that the step length sequence $\{t_n\}$ does not have a positive lower bound, i.e., by taking a subsequence if necessary we assume that $t_n \to 0+$ as $n \to +\infty$. Due to $t_n = \rho^{m_n}$, we have $m_n \to +\infty$ as $n \to +\infty$. Without loss of generality, we may assume that $m_n \geq 1$ for all n. In view of the line search strategy in Step 3 of the Algorithm, we apply Lemma 5.2 to get

$$f_{ab}(x_n + \rho^{m_n-1} d_n) - f_{ab}(x_n) = \rho^{m_n-1}\langle v_n, d_n\rangle \quad \forall n, \tag{5.3}$$

where $v_n \in \partial f_{ab}(y_n)$ with $y_n := x_n + \theta_n^* \rho^{m_n-1} d_n$ and $\theta_n^* \in (0,1)$. By the formula for $\partial f_{ab}(y_n)$ in Proposition 3.4, there exists some $z_n^* \in D^*F(\pi_b(y_n) - \pi_a(y_n))$ such that

$$v_n = z_n^* + b(y_n - \pi_b(y_n)) - a(y_n - \pi_a(y_n)). \tag{5.4}$$

In view of Lemma 2.9 **(d)** and **Assumption (ii)**, we have

$$||z_n^*|| \leq L||\pi_b(y_n) - \pi_a(y_n)||. \tag{5.5}$$

First, we consider the case that $\beta\|u_n\| \geq \|w_n\|$ in Step 3. In this case, we have $d_n = u_n = \pi_a(x_n) - x_n$ and $y_n := x_n + \theta_n^* \rho^{m_n-1} u_n$. Due to the line search strategy proposed in the Algorithm, we have $f_{ab}(x_n + \rho^{m_n-1}u_n) - f_{ab}(x_n) > -(b - a - \alpha)\rho^{m_n-1}\|u_n\|^2$. This, together with (5.3), (5.4) and (5.5), implies that

$$\begin{aligned}
-(b-a-\alpha)||u_n||^2 &< \langle v_n, u_n\rangle \\
&= \langle z_n^*, u_n\rangle + \langle b(y_n - \pi_b(y_n)) - a(y_n - \pi_a(y_n)), u_n\rangle \\
&= \langle z_n^*, u_n\rangle + b\langle \pi_a(y_n) - \pi_b(y_n), u_n\rangle + (b-a)\langle y_n - \pi_a(y_n), u_n\rangle \\
&\leq (L+b)||\pi_b(y_n) - \pi_a(y_n)|| \cdot ||u_n|| - (b-a)\langle \pi_a(y_n) - y_n, u_n\rangle. \qquad (5.6)
\end{aligned}$$

Moreover, by **Assumption (ii)**, we have

$$\begin{aligned}
||\pi_a(y_n) - \pi_b(y_n)|| &\leq ||w_n|| + ||\pi_a(y_n) - \pi_b(y_n) - w_n|| \\
&\leq \|w_n\| + ||\pi_a(y_n) - \pi_a(x_n)\| + \|\pi_b(y_n) - \pi_b(x_n)|| \\
&\leq \beta\|u_n\| + (1 + \frac{L}{a})\|y_n - x_n\| + (1 + \frac{L}{b})\|y_n - x_n\| \\
&= [\beta + (2 + \frac{L}{a} + \frac{L}{b})\theta_n^*\rho^{m_n-1}]||u_n||, \qquad (5.7)
\end{aligned}$$

and

$$\begin{aligned}
||\pi_a(y_n) - y_n - u_n|| &= ||\pi_a(y_n) - y_n - \pi_a(x_n) + x_n|| \\
&\leq \|\pi_a(y_n) - \pi_a(x_n)\| + \|y_n - x_n\| \\
&\leq (2 + \tfrac{L}{a})||y_n - x_n|| = (2 + \tfrac{L}{a})\theta_n^*\rho^{m_n-1}||u_n||.
\end{aligned}$$

The latter condition entails that

$$\langle \pi_a(y_n) - y_n, u_n\rangle = ||u_n||^2 + (2 + \frac{L}{a})\theta_n^*\rho^{m_n-1}||u_n||^2\langle c_n, \frac{u_n}{\|u_n\|}\rangle, \tag{5.8}$$

where $c_n := \frac{\pi_a(y_n) - y_n - u_n}{(2+\frac{L}{a})\theta_n^*\rho^{m_n-1}||u_n||}$ having the property that $\|c_n\| \leq 1$. Combining (5.6-5.8), we have

$$\begin{aligned}
-(b-a-\alpha) \quad < \quad & (L+b)[\beta + (2 + \tfrac{L}{a} + \tfrac{L}{b})\theta_n^*\rho^{m_n-1}] \\
& -(b-a)[1 + (2 + \tfrac{L}{a})\theta_n^*\rho^{m_n-1}\langle c_n, \tfrac{u_n}{||u_n||}\rangle].
\end{aligned} \tag{5.9}$$

Next, we consider the case that $\beta\|u_n\| < \|w_n\|$ in Step 3. In this case, we have $d_n = w_n = \pi_a(x_n) - \pi_b(x_n)$ and $y_n := x_n + \theta_n^*\rho^{m_n-1}w_n$. Due to the line search strategy proposed in the Algorithm, we have $f_{ab}(x_n + \rho^{m_n-1}w_n) - f_{ab}(x_n) > -\tau\rho^{m_n-1}\|w_n\|^2$, which, together with (5.3), (5.4) and (5.5), implies that

(5.10)
$$\begin{aligned}
& -\tau\|w_n\|^2 \\
< \;& \langle v_n,\ w_n\rangle \\
= \;& \langle z_n^* + b(y_n - \pi_b(y_n)) - a(y_n - \pi_a(y_n)), w_n\rangle \\
\leq \;& \langle z_n^*,\ \pi_a(y_n) - \pi_b(y_n)\rangle + \langle z_n^*, w_n - (\pi_a(y_n) - \pi_b(y_n))\rangle \\
& + \langle b(y_n - \pi_b(y_n)) - a(y_n - \pi_a(y_n)), w_n - (\pi_a(y_n) - \pi_b(y_n))\rangle \\
\leq \;& -\mu^*||\pi_a(y_n) - \pi_b(y_n)||^2 + \langle z_n^*, w_n - (\pi_a(y_n) - \pi_b(y_n))\rangle \\
& + \langle b(y_n - \pi_b(y_n)) - a(y_n - \pi_a(y_n)), w_n - (\pi_a(y_n) - \pi_b(y_n))\rangle \\
\leq \;& -\mu^*||\pi_a(y_n) - \pi_b(y_n)||^2 + L||\pi_a(y_n) - \pi_b(y_n)|| \cdot ||w_n - (\pi_a(y_n) - \pi_b(y_n))|| \\
& + [(b-a)||\pi_a(y_n) - y_n|| + b||\pi_a(y_n) - \pi_b(y_n)||]||w_n - (\pi_a(y_n) - \pi_b(y_n))||,
\end{aligned}$$

where the second inequality follows from Lemma 3.1 **(d)**, the third one from **Assumption (iii)**, the last one from Cauchy-Schwarz inequality. Moreover, by **Assumption (ii)**, we have

$$||\pi_a(y_n) - \pi_b(y_n) - w_n|| \leq (2 + \frac{L}{a} + \frac{L}{b})||y_n - x_n|| = (2 + \frac{L}{a} + \frac{L}{b})\theta_n^*\rho^{m_n-1}||w_n||, \tag{5.11}$$

$$||\pi_a(y_n) - \pi_b(y_n)|| \leq [1 + (2 + \frac{L}{a} + \frac{L}{b})\theta_n^*\rho^{m_n-1}]||w_n||, \tag{5.12}$$

$$\begin{aligned}
||\pi_a(y_n) - y_n|| &\leq \|u_n\| + ||\pi_a(y_n) - y_n - u_n|| \\
&\leq ||u_n|| + (2 + \frac{L}{a})\theta_n^*\rho^{m_n-1}||w_n|| \\
&\leq [\frac{1}{\beta} + (2 + \frac{L}{a})\theta_n^*\rho^{m_n-1}]||w_n|| \qquad (5.13)
\end{aligned}$$

and then there exists b_n with $||b_n|| \leq 1$ such that

$$\pi_a(y_n) - \pi_b(y_n) = w_n + (2 + \frac{L}{a} + \frac{L}{b})\theta_n^*\rho^{m_n-1}||w_n||b_n. \tag{5.14}$$

Combining (5.10-5.14), we have

$$\begin{aligned}
-\tau < & -\mu^*[1 + 2\langle\frac{w_n}{||w_n||}, (2 + \frac{L}{a} + \frac{L}{b})\theta_n^*\rho^{m_n-1}b_n\rangle + (2 + \frac{L}{a} + \frac{L}{b})^2(\theta_n^*\rho^{m_n-1})^2||b_n||^2] \\
& + L[1 + (2 + \frac{L}{a} + \frac{L}{b})\theta_n^*\rho^{m_n-1}](2 + \frac{L}{a} + \frac{L}{b})\theta_n^*\rho^{m_n-1} \\
& + (b-a)[\frac{1}{\beta} + (2 + \frac{L}{a})\theta_n^*\rho^{m_n-1}](2 + \frac{L}{a} + \frac{L}{b})\theta_n^*\rho^{m_n-1} \\
& + b[1 + (2 + \frac{L}{a} + \frac{L}{b})\theta_n^*\rho^{m_n-1}](2 + \frac{L}{a} + \frac{L}{b})\theta_n^*\rho^{m_n-1}. \qquad (5.15)
\end{aligned}$$

Our assumption that $f_{ab}(x_n) > 0$ for all n suggests that there are infinitely many positive integers n such that either $\beta\|u_n\| \geq \|w_n\|$ or $\beta\|u_n\| < \|w_n\|$, implying that there are infinitely many positive integers n such that either the inequality (5.9) or (5.15) holds. In view of $\rho^{m_n-1} \to 0+$, we have correspondingly either $-(b-a-\alpha) \le (L+b)\beta - (b-a)$ or $-\tau \le -\mu^*$, both contradicting to **Assumption (iv)**. This contradiction indicates that the step length sequence $\{t_n\}$ generated by the Algorithm has a positive lower bound. This completes the proof. ∎

PROPOSITION 5.5. *Assume that the sequence $\{x_n\}$ generated by the Algorithm satisfies $f_{ab}(x_n) > 0$ for all n. Under* **Assumptions (ii)-(iv)**, *the following inequalities hold for all n:*

$$f_{ab}(x_{n+1}) - f_{ab}(x_n) \le -M_1 ||x_{n+1} - x_n||^2 \tag{5.16}$$

and

$$d(0, \partial f_{ab}(x_n)) \le \frac{M_2}{t^*} ||x_{n+1} - x_n||, \tag{5.17}$$

where $M_1 := \min\{b-a-\alpha, \tau\}$, $M_2 := L + b + \frac{b-a}{\beta}$ and t^* is a positive lower bound of $\{t_n\}$.

Proof. By Steps 3 and 4 of the Algorithm, we have $0 < t_n \le 1$, $x_{n+1} = x_n + t_n d_n$ and $f_{ab}(x_{n+1}) - f_{ab}(x_n) \le -M_1 t_n \|d_n\|^2$ for all n, from which we get (5.16) immediately. By Lemma 5.1, we have $d(0, \partial f_{ab}(x_n)) \le (L+b)||w_n|| + (b-a)||u_n||$, where L is given as in **Assumption (ii)**, and $w_n = \pi_a(x_n) - \pi_b(x_n)$ and $u_n = \pi_a(x_n) - x_n$ are set as in Step 3. If $\beta||u_n|| < ||w_n||$, we get from Steps 3 and 4 of the Algorithm that $\|x_{n+1} - x_n\| = t_n \|w_n\|$ and hence that

$$(L+b)||w_n|| + (b-a)||u_n|| < (L + b + \frac{b-a}{\beta})||w_n|| = \frac{M_2}{t_n}||x_{n+1} - x_n||.$$

Alternatively if $\beta||u_n|| \ge ||w_n||$, we get from Steps 3 and 4 of the Algorithm that $\|x_{n+1} - x_n\| = t_n \|u_n\|$ and hence that

$$(L+b)||w_n|| + (b-a)||u_n|| \le \beta(L + b + \frac{b-a}{\beta})||u_n|| \le \frac{M_2}{t_n}||x_{n+1} - x_n||,$$

where the second inequality follows from the fact that $0 < \beta < \frac{b-a}{b+L} < 1$ according to **Assumption (iv)**. In both cases, we get (5.17) by noting that the existence of a positive lower bound t^* of $\{t_n\}$ is guaranteed by Proposition 5.4. This completes the proof. ∎

THEOREM 5.6. *Assume that the sequence $\{x_n\}$ generated by the Algorithm satisfies $f_{ab}(x_n) > 0$ for all n. Under* **Assumptions (i)-(iv)**, *the following assertions hold:*

(a) *The sequence x_n has a finite length, i.e., $\sum_{n=0}^{+\infty} ||x_{n+1} - x_n|| < +\infty$.*
(b) *The sequence $f_{ab}(x_n)$ converges* Q*-linearly to 0.*
(c) *The sequence x_n converges* R*-linearly to a solution $\bar{x}$ of (VIP).*

Proof. From Proposition 5.5, it follows that (5.16) and (5.17) holds with $M_1 := \min\{\tau, b-a-\alpha\}$, $M_2 := L + b + \frac{b-a}{\beta}$ and t^* being a positive lower bound of $\{t_n\}$. By **Assumption (i)**, the level set $[f_{ab} \le f_{ab}(x_0)]$ is bounded, which, together with (5.16), implies that the sequence $\{x_n\}$ is also bounded. Denote by $\bar{x}$ any cluster point of the sequence $\{x_n\}$. By **Assumption (iii)**, f satisfies the KL inequality at $\bar{x}$ with an exponent of $\frac{1}{2}$. In view of these facts and the continuity of f_{ab}, we confirm that the sequence $\{x_n\}$ satisfies the assumptions **(H1)** and **(H3)** and a variant of the assumption **(H2)** in [4]. Note that the assumption **(H2)** in [4] requires that $d(0, \partial f_{ab}(x_{n+1}))$, instead of $d(0, \partial f_{ab}(x_n))$, has an upper estimate as in the form of (5.17). In this case, [4, Theorem 2.9] cannot be applied directly, but we can still follow the proof of [4, Theorem 2.9] to deduce the following: (i) **(a)** holds; (ii) $x_n \to \bar{x}$ and $f_{ab}(x_n) \to f_{ab}(\bar{x})$ as n goes to ∞; and (iii) $0 \in \partial f_{ab}(\bar{x})$. In view of **Assumption (iii)** and Lemma 4.4, we have $\pi_a(\bar{x}) = \pi_b(\bar{x})$. Then by Corollary 3.5 **(b)**, $\bar{x}$ is a solution of (VIP) or equivalently $f_{ab}(\bar{x}) = 0$ (cf. Lemma 3.1 **(c)**).

It remains to show the convergence rate. By the line search strategy in Step 3 of the Algorithm, the following hold for all n:

$$\|d_n\| \geq \beta \|x_n - \pi_a(x_n)\|, \tag{5.18}$$

and

$$\begin{aligned} f_{ab}(x_{n+1}) - f_{ab}(x_n) &\leq -\min\{\tau, b-a-\alpha\} t_n \|d_n\|^2 \\ &\leq -\min\{\tau, b-a-\alpha\} t^* \|d_n\|^2 \\ &< 0. \end{aligned} \tag{5.19}$$

In view of (5.18), we get from Lemma 3.1 **(a)** that $\|d_n\|^2 \geq \frac{2\beta^2}{b-a} f_{ab}(x_n)$, which, together with (5.19) and the definition of M_1, implies that

$$f_{ab}(x_{n+1}) \leq -M_1 t^* ||d_n||^2 + f_{ab}(x_n) \leq (1 - \frac{2\beta^2 M_1 t^*}{b-a}) f_{ab}(x_n),$$

and hence that,

$$\frac{f_{ab}(x_{n+1})}{f_{ab}(x_n)} \leq 1 - \frac{2\beta^2 M_1 t^*}{b-a} =: \eta. \tag{5.20}$$

Clearly, we have $0 < \eta < 1$. Then by definition [24, pp.619-620], the sequence $f_{ab}(x_n)$ converges Q-linearly to 0. That is, **(b)** follows.

By the triangle inequality, the following holds for all positive integers n and m with $m > n$: $\|x_n - \bar{x}\| \leq \sum_{k=n}^{m} \|x_{k+1} - x_k\| + \|x_{m+1} - \bar{x}\|$. In view of **(a)** and the fact that $\|x_{m+1} - \bar{x}\| \to 0$ as $m \to \infty$, we have $\sum_{k=n}^{m} \|x_{k+1} - x_k\| + \|x_{m+1} - \bar{x}\| \to \sum_{k=n}^{\infty} \|x_{k+1} - x_k\|$ as $m \to \infty$, and hence $\|x_n - \bar{x}\| \leq \sum_{k=n}^{\infty} \|x_{k+1} - x_k\|$. In view of (5.16) and (5.20), we further have

$$\|x_n - \bar{x}\| \leq \sum_{k=n}^{\infty} \sqrt{\frac{f_{ab}(x_k)}{M_1}} \leq \sqrt{\frac{f_{ab}(x_n)}{M_1}} \sum_{k=0}^{\infty} \sqrt{\eta^k} = \sqrt{\frac{f_{ab}(x_n)}{M_1}} \frac{1}{1-\sqrt{\eta}} =: \zeta_n,$$

and $\frac{\zeta_{n+1}}{\zeta_n} = \sqrt{\frac{f_{ab}(x_{n+1})}{f_{ab}(x_n)}} \leq \sqrt{\eta}$. As $0 < \eta < 1$, we have $0 < \sqrt{\eta} < 1$. Then by definition [24, pp.619-620], ζ_n converges Q-linearly to 0, and x_n converges R-linearly to $\bar{x}$. This completes the proof. ∎

REFERENCES

[1] G. Auchmuty, Variational principles for variational inequalities, Numer. Funct. Anal. Optim., 10(1989)863-874.

[2] A. Auslender, Optimization, Méthodes numériques, Masson, Paris, France, 1976.

[3] H. Attouch, J. Bolte, P. Redont, A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Łojasiewicz inequality, Math. Oper. Res., 35(2010)438-457.
[4] H. Attouch, J. Bolte, B.F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, Math. Program., 137(2013)91-129.
[5] D. Azé and J.-N. Corvellec. Characterizations of error bounds for lower semicontinuous functions on metric spaces. ESAIM: Control, Optimisation and Calculus of Variations, 10(2004)409-425.
[6] J. Bolte, A. Daniilidis, A. Lewis, The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems, SIAM J. Optim., 17(2007)1205-1223.
[7] G.Y. Chen, C.J. Goh, X.Q. Yang, On gap functions and duality of variational inequality problems, J. Math. Anal. Appl., 214(1997)658-673. A
[8] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.
[9] B. V. Dinh, T.S. Pham, Error bounds of regularized gap functions for polynomial variational inequalities, J. Optim. Theory Appl., 192(2022)226-247.
[10] F. Facchinei, J.S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems, Volumes I-II, Springer, Berlin Helidelberg, New York, 2003.
[11] M. Fukushima, Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems, Math. Program., 53(1992)99-110.
[12] E. De Giorgi, A. Marino, and M. Tosques. Problemi di evoluzione in spazi metrici e curve di massima pendenza (evolution problems in metric spaces and curves of maximal slope). Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur., 68(1980)180-187.
[13] P.T. Harker, J.S. Pang, Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms, and applications, Math. Program., 48(1990)161-220.
[14] H.Y. Jiang, L.Q. Qi, Local uniqueness and convergence of iterative methods for nonsmooth variational inequalities, J. Math. Anal. Appl., 196(1995)314-331.
[15] C. Kanzow, M. Fukushima, Theoretical and numerical investigation of the D-gap function for box constrained variational inequalities, Math. Program., 83(1998)55-87.
[16] I. Konnov, Descent methods for nonsmooth variational inequalities, Comput. Math. Math. Phys., 46(2006)1186-1192.
[17] T. Larsson, M. Patriksson, A class of gap functions for variational inequalities, Math. Program., 64(1994)53-79.
[18] G.Y. Li, K.F. Ng, Error bounds of generalized D-gap functions for nonsmooth and nonmonotone variational inequality problems, SIAM J. Optim., 20(2009)667-690.
[19] G.Y. Li, C. Tang, Z. Wei, Error bound results for generalized D-gap functions of nonsmooth variational inequality problems, J. Comput. Appl. Math., 233(2010)2795-2806.
[20] G.Y. Li, T.K. Pong, Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods, Found. Comput. Math., 18(2018)1199-1232.
[21] K.W. Meng, X.Q. Yang, Equivalent conditions for local error bounds. Set-Valued Var. Anal. 20(2012)617-636.
[22] K.F. Ng, L.L. Tan, Error bounds of regularized gap functions for nonsmooth variational inequality problems, Math. Program., 110(2007)405-429.
[23] K.F. Ng, L.L. Tan, D-gap functions for nonsmooth variational inequality problems, J. Optim. Theory Appl., 133(2007)77-97.
[24] J. Nocedal, S. Wright, Numerical Optimization (2nd ed.), Berlin, New York: Springer-Verlag, 2006.
[25] J.-S. Pang, A posteriori error bounds for the linearly-constrained variational inequality problem, Math. Oper. Res., 12(1987)474-484.
[26] B. Panicucci, M. Pappalardo, M. Passacantando, A globally convergent descent method for nonsmooth variational inequalities, Comput. Optim. Appl., 43(2009)197-211.
[27] M. Pappalardo, G. Mastroeni, M. Passacantando, Merit functions: a bridge between optimization and equilibria, 4OR, 12(2014)1-33.
[28] J.M. Peng, Equivalence of variational inequality problems to unconstrained optimization, Math. Program., 78(1997)347-355.
[29] J.M. Peng, M. Fukushima, A hybrid Newton method for solving the variational inequality problem via the D-gap function, Math. Program., 86(1999)367-386.
[30] B. Qu, C.Y. Wang, J.Z. Zhang, Convergence and error bound of a method for solving variational inequality problems via the generalized D-gap function, J. Optim. Theory Appl.,

119(2003)535-552.
[31] R.T. Rockafellar, R. J.-B. Wets, Variational Analysis, Springer, Berlin, 1998.
[32] M.V. Solodov, P. Tseng, Some methods based on the D-gap function for solving monotone variational inequalities, Comput. Optim. Appl., 17(2000)255-277.
[33] M.V. Solodov, Merit functions and error bounds for generalized variational inequalities, J. Math. Anal. Appl., 287(2003)405-414.
[34] L.L. Tan, Regularized gap functions for nonsmooth variational inequality problems, J. Math. Anal. Appl., 334(2007)1022-1038.
[35] L. Thibault, Tangent cones and quasi-interiorly tangent cones to multifunctions, Trans. Amer. Math. Soc., 277(1983)601-621.
[36] J.H. Wu, M. Florian, P. Marcotte, A general descent framework for the monotone variational inequality problem, Math. Program., 61(1993)281-300.
[37] H.F. Xu, Regularized gap functions and D-gap functions for nonsmooth variational inequalities, Optimization and Related Topics, A. Rubinov and B. Glover eds., Kluwer Academic Publishers, 2001.
[38] N. Yamashita, M. Fukushima, Equivalent unconstrained minimization and global error bounds for variational inequality problems, SIAM J. Control Optim., 35(1997)273-284.
[39] N. Yamashita, K. Taji, M. Fukushima, Unconstrained optimization reformulations of variational inequality problems, J. Optim. Theory Appl., 92(1997)439-456.
[40] D.L. Zhu, P. Marcotte, Modified descent methods for solving the monotone variational inequality problem, Oper. Res. Lett., 14(1993)111-120.
[41] D.L. Zhu, P. Marcotte, An extended descent framework for variational inequalities, J. Optim. Theory Appl., 80(1994)349-366.