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

Lagrange Multipliers, (Exact) Regularization and Error Bounds for Monotone Variational Inequalities

C. Charitha, Joydeep Dutta, and
D. Russell Luke

the date of receipt and acceptance should be inserted later

Abstract We examine two central regularization strategies for monotone variational inequalities, the first a direct regularization of the operative monotone mapping, and the second via regularization of the associated dual gap function. A key link in the relationship between the solution sets to these various regularized problems is the idea of exact regularization, which, in turn, is fundamentally associated with the existence of Lagrange multipliers for the regularized variational inequality. A regularization is said to be exact if a solution to the regularized problem is a solution to the unregularized problem for all parameters beyond a certain value. The Lagrange multipliers corresponding to a particular regularization of a variational inequality, on the other hand, are defined via the dual gap function. Our analysis suggests various conceptual, iteratively regularized numerical schemes, for which we provide error bounds, and hence stopping criteria, under the additional assumption that the solution set to the unregularized problem is what we call *weakly sharp* of order greater than one.

2010 Mathematics Subject Classification: Primary 49J40; 47J20 Secondary 47H04, 49M20, 49M37, 65K05, 90C30.

Keywords: Variational inequality, exact regularization, error bound, gap functional, dual gap functional, D-gap function, weak sharp solutions.

Institut für Numerische und Angewandte Mathematik, Universität Göttingen, 37083 Göttingen, Germany. CC was supported by DFG Grant SFB755-A4. E-mail: `c.cherugondi@math.uni-goettingen.de`

Department of Humanities and Social Sciences, Indian Institute of Technology Kanpur 208016 Kanpur, India. Email: `jdutta@iitk.ac.in`

Institut für Numerische und Angewandte Mathematik, Universität Göttingen, 37083 Göttingen, Germany. DRL was supported by DFG Grant SFB755-C2. Email: `r.luke@math.uni-goettingen.de`

Address(es) of author(s) should be given

1 Introduction.

Given a mapping $F : \mathbb{R}^n \to \mathbb{R}^n$, a closed set $\Omega \in \mathbb{R}^n$, we consider the variational inequality problem VI(F, Ω):

find a vector $\overline{x} \in \Omega$ such that

$$\langle F(\overline{x}), y - \overline{x} \rangle \geq 0, \quad \forall y \in \Omega \tag{VI}$$

and the (strong) generalized variational inequality GVI(T, Ω) for a multivalued mapping $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$:

find $\overline{x} \in \Omega$ such that

$$\exists\, v \in T(\overline{x}) \text{ with } \langle v,\ x - \overline{x} \rangle \geq 0 \quad \forall x \in \Omega. \tag{GVI}$$

We will denote the sets of solutions to these problems by SOL(F, Ω) and SOL(T, Ω) respectively where the corresponding problem, VI(F, Ω) or GVI(T, Ω), is clear from context. Though the subject of variational inequalities is well-established (see [10] for the basic theory and algorithms), we recall some basic definitions.

Definition 1 *[(pseudo) monotone mappings] A mapping $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to be* pseudomonotone *on Ω if for all $x, y \in \Omega$*

$$\langle x - y, F(y) \rangle \geq 0 \quad \Longrightarrow \quad \langle x - y, F(x) \rangle \geq 0. \tag{1}$$

F is said to be pseudomonotone$^+$ on Ω if F is pseudomonotone and for all x, y in Ω,

$$\langle F(x), y - x \rangle \geq 0 \quad \text{and} \quad \langle F(y), y - x \rangle = 0 \Rightarrow F(y) = F(x). \tag{2}$$

F is called monotone *on Ω if*

$$\langle F(x) - F(y), x - y \rangle \geq 0 \quad \forall\, x, y \in \Omega. \tag{3}$$

F is called strongly monotone *on Ω if there exists a $\mu > 0$ such that*

$$\langle F(x) - F(y), x - y \rangle \geq \mu \|x - y\|^2 \quad \forall\, x, y \in \Omega. \tag{4}$$

We recall that for monotone functions, the solution set, if exists, is convex. Throughout this work we assume the following.

Assumption 2

(i) $\Omega \subset \mathbb{R}^n$ is nonempty, closed and convex.
(ii) $F : \Omega \to \mathbb{R}^n$ is continuous and monotone.

Assumption 2(ii) above is understood in the context of Definition 1 by the obvious extension of F to a mapping defined on $\mathbb{R}^n$ by the mapping whose effective domain is Ω, that is $F(x) = \emptyset$ for all $x \notin \Omega$. We recall a standard result on existence and boundedness of the set of solutions to $\mathrm{VI}(F, \Omega)$. Define the *recession* or *asymptotic cone* Ω^∞ by

$$\Omega^\infty \equiv \{w \in \mathbb{R}^n \mid \text{ for any } x \in \Omega,\ x + w\tau \in \Omega \ \forall\, \tau \geq 0\}. \tag{5}$$

Lemma 1 (Exercise 12.52, [19]) *Under assumption 2,* $\mathrm{SOL}(F, \Omega)$ *is nonempty and bounded if and only if*

$$w \in \Omega^\infty \setminus \{0\} \implies \exists x \in \Omega \text{ with } \langle F(x),\ w\rangle > 0. \tag{6}$$

There is a vast literature on how to solve a variational inequality under various assumptions (see [10] and references therein). Of particular interest for us are *ill-posed variational inequalities*. There are many definitions of ill-posedness. Here we will consider ill-posed any variational inequality $\mathrm{VI}(F, \Omega)$ for which F is not strongly monotone. The conventional approach to such problems is to *regularize*, or otherwise modify the problem so that the regularized problem is well-posed and has one or more solutions that are reasonable approximations to solutions to the original problem. A solution to the desired ill-posed problem, if exists, is then achieved as a limit of solutions to well-posed approximate problems.

A central motivation of this paper is the concept of *exact regularization* for a variational inequality, that is, a regularization for which the regularized solution corresponds to a solution to the unregularized problem for all regularization parameters below a certain threshold. Exact *penalization* is a well understood concept in constrained optimization, and the relation to the existence of Lagrange multipliers has been extensively studied. This has recently been extended to penalized variational inequalities where the connection to Lagrange multipliers also appears [8, Lemma 4]. We take our inspiration from the concept of exact *regularization* developed in the context of convex programming by Friedlander and Tseng [11]. Given a convex mapping $f : \mathbb{R}^n \to \mathbb{R}$, a nonempty closed convex set $\Omega \in \mathbb{R}^n$ and a continuous convex map $\varphi : \mathbb{R}^n \to \mathbb{R}$, $\varepsilon > 0$, consider the following regularization scheme:

$$(\mathcal{P}_0) \qquad \underset{x\in\Omega}{\text{minimize}}\ f(x) \quad \to \quad (\mathcal{P}_\varepsilon) \qquad \underset{x\in\Omega}{\text{minimize}}\ f(x) + \varepsilon\varphi(x).$$

When $\varphi = \|.\|^2$, it is the well known Tikhonov regularization and when $\varphi = \|.\|_1$, an l_1 regularization. The regularization is said to be exact if solutions to $(\mathcal{P}_\varepsilon)$ are solutions to $(\mathcal{P}_0)$ for ε below some threshold value.

Generalizing this to variational inequalities, for any continuous convex mapping $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ with $\mathrm{dom}\ \varphi = \Omega$, denote $T_\varepsilon \equiv F + \varepsilon\partial\varphi$, with $\varepsilon > 0$ fixed, and F extended by $\emptyset$ to a mapping $\mathbb{R}^n \to \mathbb{R}^n$. We consider the following regularization strategy for $\mathrm{VI}(F, \Omega)$ which turns out to be a specialization of (GVI):

find $\overline{x} \in \Omega$ such that

$$\exists\, v \in T_\varepsilon(\overline{x}) \equiv F(\overline{x}) + \varepsilon\partial\varphi(\overline{x}) \text{ with } \langle v,\ x - \overline{x}\rangle \geq 0,\ \forall x \in \Omega. \qquad (\mathrm{GVI}_{T_\varepsilon})$$

In our extension of the notion of exact regularization to variational inequalities we introduce *Lagrange multipliers* for variational inequalities, the existence of which are closely tied to the existence of exact regularization strategies. The central tool for our analysis is the *gap function*.

For a given variational inequality $\mathrm{VI}(F,\Omega)$, a *gap function* is a function $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ with $\Omega \subseteq \mathrm{dom}\ \psi$ and

1. $\psi(x) \geq 0$ for all $x \in \Omega$;
2. $\psi(\overline{x}) = 0$, $\overline{x} \in \Omega$ if and only if $\overline{x}$ solves $\mathrm{VI}(F,\Omega)$.

It is clear that any minimizer $\bar{x}$ of the gap function ψ over Ω with $\psi(\bar{x}) = 0$ is a solution to $\mathrm{VI}(F,\Omega)$. The first occurrence of the gap function for $\mathrm{VI}(F,\Omega)$ is Auslender's gap function [1]:

$$\theta(x) = \sup_{y\in\Omega}\langle F(x), x - y\rangle, \tag{7}$$

where we use the convention that the value of a function on the emptyset is $+\infty$ so that $\theta(x) = +\infty$ at points $x \notin \Omega$. The dual gap function for $\mathrm{VI}(F,\Omega)$ is given by

$$G(x) = \sup_{y\in\Omega}\langle F(y), x - y\rangle. \tag{8}$$

For each fixed y, the function $x \mapsto \langle F(y), x - y\rangle$ is affine. Thus the dual gap function G is closed and convex on Ω since it is the pointwise supremum over affine functions. The dual gap function is not necessarily a gap function for $\mathrm{VI}(F,\Omega)$, however, with additional assumptions on F, it is indeed a gap function. In particular, if the mapping F is pseudomonotone and continuous, then G is in fact a gap function for $\mathrm{VI}(F,\Omega)$ [10, Theorem 2.3.5].

Note that neither θ nor G is finite valued in general. If Ω is assumed to be compact, then both are finite-valued, but we will avoid such restrictions in what follows. A regularized gap function for $\mathrm{VI}(F,\Omega)$ with regularization parameter $\alpha > 0$, is given by

$$\theta_\alpha(x) = \sup_{y\in\Omega}\left\{\langle F(x), x - y\rangle - \frac{\alpha}{2}\|y - x\|^2\right\}. \tag{9}$$

This was introduced in [12] and is finite valued for any closed convex set Ω. Note that when F is strongly monotone then G is finite valued even without Ω being compact. For, the strong monotonicity of F on Ω with constant μ implies that

$$\langle F(y), x - y\rangle \leq \langle F(x), x - y\rangle - \mu\|x - y\|^2. \tag{10}$$

Hence

$$G(x) = \sup_{y\in\Omega}\langle F(y), x-y\rangle \leq \sup_{y\in\Omega}\{\langle F(x), x-y\rangle - \mu\|x-y\|^2\} = \theta_{2\mu}(x) < \infty. \tag{11}$$

One can reformulate $\mathrm{VI}(F,\Omega)$ as a constrained optimization problem using θ_α. Since the objective function in (9) is strongly concave, for every x there exists a unique solution $y_\alpha(x)$ which is explicitly given by

$$y_\alpha(x) = P_\Omega\left(x - \frac{1}{\alpha}F(x)\right) \tag{12}$$

where $P_\Omega(z) \equiv \operatorname{argmin}_{y\in\Omega}\|y-z\|$ is the *projection* onto the set Ω. Hence $\theta_\alpha(x)$ can be explicitly written as

$$\theta_\alpha(x) = \left\{\langle F(x), x - y_\alpha(x)\rangle - \frac{\alpha}{2}\|y_\alpha(x) - x\|^2\right\}. \tag{13}$$

When F is continuously differentiable, θ_α is continuously differentiable [12, Theorem 3.2] and hence we can reformulate $\mathrm{VI}(F,\Omega)$ as a constrained optimization problem with the differentiable objective function θ_α.

We show in Section 2 that, although the solution set of the regularized problem ($\mathrm{GVI}_{T_\varepsilon}$) has some relation to the solution set of (VI), we will achieve a more precise correspondence via the dual gap function G defined by (8) and the equivalence between solutions to the problem (VI) and the convex optimization problem

$$\underset{x\in\Omega}{\text{minimize}}\quad G(x). \tag{$\mathcal{P}_G$}$$

If $\mathrm{SOL}(F,\Omega)\neq\emptyset$, then solving $\mathrm{VI}(F,\Omega)$ is equivalent to solving ($\mathcal{P}_G$). The corresponding regularization of the the above convex optimization problem in the spirit of [11] gives us the problem

$$\underset{x\in\Omega}{\text{minimize}}\ G_{\varepsilon\varphi}(x) \equiv G(x) + \varepsilon\varphi(x). \tag{$\mathcal{P}_{G_{\varepsilon\varphi}}$}$$

Definition 3 (exact regularization of variational inequalities) *A regularization of the variational inequality* (VI) *is said to be* exact *if solutions to the convex optimization problem* ($\mathcal{P}_{G_{\varepsilon\varphi}}$) *are also solutions to* (VI) *for all values of ε below some threshold value $\overline{\varepsilon} > 0$.*

Another advantage of gap functions is the availability of computable error bounds for strongly monotone variational inequalities. Error bounds, in turn, are essential for principled stopping criteria for algorithms. These are discussed in Section 3 where we derive an upper bound on the error under the assumption that the solution set $\mathrm{VI}(F,\Omega)$ is *weakly-sharp of order gamma* (46). Error bounds can also be achieved for the special case of monotone mappings where $F(x) = Mx + q$ and $\Omega = \mathbb{R}^n_+$ or a polyhedron with a positive semidefinite matrix M. For general monotone variational inequalities, however, we are unaware of any results on error bounds using the gap function.

Through the study of unconstrained reformulations for variational inequalities the closely related D-gap function $\theta_{\alpha\beta}$ for $\mathrm{VI}(F,\Omega)$ was introduced [15]. It is defined as the difference of two regularized gap functions θ_α and θ_β with $\beta > \alpha$ and is given by

$$\theta_{\alpha\beta}(x) = \theta_\alpha(x) - \theta_\beta(x);\quad \beta > \alpha > 0. \tag{14}$$

The D-gap function satisfies the following properties [20, Theorem 3.2]:

1. $\theta_{\alpha\beta}(x) \geq 0$ for all $x \in \mathbb{R}^n$;
2. $\theta_{\alpha\beta}(\overline{x}) = 0$, $x \in \mathbb{R}^n$ if and only if $\overline{x}$ solves $\mathrm{VI}(F, \Omega)$.

The D-gap function provides an unconstrained reformulation of the variational inequality [20, Theorem 3.2]. As with the gap function, when F is continuously differentiable, the D-gap function is smooth and the resulting unconstrained optimization problem of minimizing $\theta_{\alpha\beta}$ is smooth [20, Theorem 3.1].

As the theory for gap and D-gap functions for generalized variational inequalities is underdeveloped, particularly with regard to numerical algorithms, we will, when necessary, restrict our attention to *differentiable* strongly convex regularizers φ. Our numerical approach for solving the regularized problems $\mathrm{VI}(T_\varepsilon, \Omega)$ with $T_\varepsilon = F + \varepsilon\nabla\varphi$ is via D-gap functions for which there is ample choice of appropriate methods. We use the attendant error bounds developed in [7] and [20] for iterative methods for solving $\mathrm{VI}(T_\varepsilon, \Omega)$ with ε fixed and $T_\varepsilon = F + \varepsilon\nabla\varphi$ strongly monotone. In the limit as $\varepsilon \to 0$ we approach the solution set to $\mathrm{VI}(F, \Omega)$. If our regularization φ is *exact*, then, for some ε below a threshold value, the procedure for solving the regularized problem will converge to a point in $\mathrm{SOL}(F, \Omega)$ with computable error bounds. When the regularization is not exact, error bounds on the distance from the regularized solution to the original solution set is provided in [11] for a general convex minimization. We derive in Section 4.3 conditions for a similar error bound for the generalization to variational inequalities. Unlike the case of convex minimization our generalization demands more than the existence of the "weak sharp minima" in order to achieve exact regularization. Our characterization (74) appears to be new.

In section 2, we study the properties of the solution sets of generalized variational inequalities with the purpose of understanding the solution sets of the regularized problem $\mathrm{GVI}(T_\varepsilon, \Omega)$ with $T_\varepsilon = F + \varepsilon\partial\varphi$. Here the essential role of the dual gap function for characterizing exact regularization becomes apparent. In Section 3 we focus on the solution methods for monotone variational inequalities via iterative regularization of the dual gap function. The analysis in Section 4 is refined to the special case when φ is differentiable , where we study direct regularization of the variational inequality via $(\mathrm{GVI}_{T_\varepsilon})$. In the same section we present some numerical results illustrating the theory.

2 Solution Sets

We begin with a study of the relationship between regularized generalized variational inequalities and their limit as the regularization parameter goes to zero.

2.1 Basic Facts, Notation and Assumptions

Our focus in this section is on the solution sets of $\mathrm{GVI}(T, \Omega)$ where T is a *maximal monotone map* and Ω is a non-empty closed and convex set.

Definition 4 (normal cone) *A Normal cone to a closed convex set Ω at a point $\overline{x} \in \Omega$ is defined as*

$$N_\Omega(\overline{x}) = \{v \in \mathbb{R}^n : \langle v, x - \overline{x}\rangle \leq 0 \quad \forall x \in \Omega\}. \tag{15}$$

Definition 5 ((maximal) monotone mappings) *A set-valued map $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is* ξ-monotone *for some $\xi > 1$ if there exists $\mu > 0$ such that*

$$\langle v - w, x - y\rangle \geq \mu\|x - y\|^\xi \qquad \forall\ (x, v) \in \operatorname{gph} T \quad \text{and } \forall\ (y, w) \in \operatorname{gph} T. \tag{16}$$

It is simply said to be monotone *if it is $\xi = 1$ and $\mu = 0$ in the above equation. F is* maximally monotone *if there is no monotone operator $\overline{T} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ such that the graph of $\overline{T}$ properly contains the graph of T. T is* strongly monotone *if there exists $\mu > 0$ such that*

$$\langle v - w, x - y\rangle \geq \mu\|x - y\|^2 \quad \forall\ (x, v) \in \operatorname{gph} T \quad \text{and } \forall\ (y, w) \in \operatorname{gph} T. \tag{17}$$

Using this notion we can alternatively write (GVI) as a maximal monotone inclusion:

$$0 \in T(x) + N_\Omega(x).$$

Note that N_Ω is maximal monotone [4, Example 20. 41]. If dom $T = \Omega$ then $T + N_\Omega$ is also maximal monotone. More generally, if $\operatorname{ri}(\operatorname{dom} T) \cap \operatorname{ri}(\operatorname{dom} N_\Omega) = \operatorname{ri}(\operatorname{dom} T) \cap \operatorname{ri}(\Omega) \neq \emptyset$ then $T + N_\Omega$ is also maximal monotone.

Another central property of set-valued mappings that we will make use of concerns the notion of continuity.

Definition 6 *A map $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is* outer semi-continuous *at a point $\overline{x} \in \mathbb{R}^n$ if*

$$\bigcup_{x^k \to \overline{x}} \limsup_{k \to \infty} T(x^k) \subset T(\overline{x})$$

In another useful characterization, a set-valued map T is outer semicontinuous everywhere if and only if its graph is closed ([19, Theorem 5.7]).

We begin by studying some of the fundamental properties of the solution set $\mathrm{SOL}(T, \Omega)$, such as convexity and boundedness. We begin with convexity. For this, we introduce the notion of a *Minty GVI*, denoted by $\mathrm{MGVI}(T, \Omega)$, wherein we seek $\overline{x} \in \Omega$, such that for each $y \in \Omega$ and any $v \in T(y)$

$$\langle v, y - \overline{x}\rangle \geq 0. \tag{18}$$

Compare this to the *weak GVI*, denoted by $\mathrm{WGVI}(T, \Omega)$ [2], wherein we seek to find $\overline{x}$, such that for each $x \in \Omega$, there exists $w_x \in T(\overline{x})$ such that

$$\langle w_x, x - \overline{x}\rangle \geq 0. \tag{19}$$

Let us denote the solution sets of $\mathrm{MGVI}(T, \Omega)$ and $\mathrm{WGVI}(T, \Omega)$ by $\mathrm{SOL}^M(T, \Omega)$ and $\mathrm{SOL}^W(T, \Omega)$. We will first show that $\mathrm{SOL}^M(T, \Omega)$ is a convex set.

Lemma 2 $\mathrm{SOL}^M(T,\Omega)$ *is a convex set.*

(i) If T is monotone, then $\mathrm{SOL}(T,\Omega) \subseteq \mathrm{SOL}^M(T,\Omega)$.
(ii) If T is locally bounded and graph closed, then $\mathrm{SOL}^M(T,\Omega) \subseteq \mathrm{SOL}^W(T,\Omega)$.

Proof. Let $\overline{x}_1, \overline{x}_2 \in \mathrm{SOL}^M(T,\Omega)$ and let $y \in \Omega$. Then for any $v \in T(y)$,

$$\langle v, y - \overline{x}_1 \rangle \geq 0 \tag{20}$$
$$\langle v, y - \overline{x}_2 \rangle \geq 0. \tag{21}$$

Now multiplying (20) with λ and (21) with $(1-\lambda)$ where $0 \leq \lambda \leq 1$, we have

$$\langle v, y - (\lambda \overline{x}_1 + (1-\lambda)\overline{x}_2) \rangle \geq 0. \tag{22}$$

Since $y \in \Omega$ was chosen arbitrarily, we have

$$\lambda \overline{x}_1 + (1-\lambda)\overline{x}_2 \in \mathrm{SOL}^M(T,\Omega). \tag{23}$$

Hence $\mathrm{SOL}^M(T,\Omega)$ forms a convex set.

Part (i). By invoking the monotonicity of T, it is simple to show that

$$\mathrm{SOL}(T,\Omega) \subseteq \mathrm{SOL}^M(T,\Omega). \tag{24}$$

Part (ii). Let $\overline{x} \in \mathrm{SOL}^M(T,\Omega))$. Then for any $y \in \Omega$ and $v \in T(y)$

$$\langle v, y - \overline{x} \rangle \geq 0. \tag{25}$$

Let us construct the sequence

$$y_n = \overline{x} + \frac{1}{n}(x - \overline{x}),$$

where $x \in \Omega$ is a fixed but arbitrary point. Of course we have $y_n \in \Omega$, since Ω is a closed convex set. Hence for any $v_n \in T(y_n)$ we have

$$\langle v_n, y_n - \overline{x} \rangle \geq 0, \quad \forall n \in \mathbb{N} \tag{26}$$

and hence

$$\langle v_n, x - \overline{x} \rangle \geq 0, \quad \forall n \in \mathbb{N}. \tag{27}$$

As T is locally bounded, by noting that $y_n \to \overline{x}$, we can conclude that v_n is a bounded sequence. Without loss of generality let us assume that $v_n \to v_x$. Note that the limit must depend on the chosen x. Hence we have

$$\langle v_x, x - \overline{x} \rangle \geq 0. \tag{28}$$

Further, as T is graph closed, $v_x \in T(\overline{x})$. Now, this limit v_x will change with x. Since $x \in \Omega$ is arbitrary, the above argument can be repeated for each $x \in \Omega$. This shows that $\overline{x} \in \mathrm{SOL}^W(T,\Omega)$. □

The next result determines appropriate conditions on T that guarantee that

$$\mathrm{SOL}^M(T,\Omega) \subseteq \mathrm{SOL}(T,\Omega),$$

from which it follows by Lemma 2 that $\mathrm{SOL}(T,\Omega)$ is a convex set.

Theorem 7 *Let $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a non-empty, convex and compact valued map. Further, assume that T is monotone, locally bounded and graph closed. Then $sol(GVI(T, \Omega))$ is a convex set.*

Proof. Using the fact that T is compact-valued, Aussel and Dutta [2] had constructed the following gap function for $WGVI(T, \Omega)$. This given as

$$\widehat{g}(x) = \sup_{y \in \Omega} \inf_{v \in T(x)} \langle v, x - y \rangle.$$

Let $\overline{x} \in \mathrm{SOL}^M(T, \Omega)$. Note that any $\overline{x} \in \mathrm{SOL}^W(T, \Omega)$ satisfies $\widehat{g}(\overline{x}) = 0$. Since $\mathrm{SOL}^M(T, \Omega) \subseteq \mathrm{SOL}^W(T, \Omega)$ from Lemma 2, we can now write

$$\widehat{g}(\overline{x}) = 0 = \sup_{y \in \Omega} \inf_{v \in T(\overline{x})} \langle v, \overline{x} - y \rangle.$$

Now since $\overline{x}$ is fixed, the function $v \mapsto \langle v, \overline{x} - y \rangle$ is linear for each fixed y and the function $y \mapsto \langle v, \overline{x} - y \rangle$ is affine (and hence concave) for each fixed v. Hence, as $T(\overline{x})$ is convex and compact valued we can invoke the famous Sion's minimax theorem to conclude that

$$0 = \inf_{v \in T(\overline{x})} \sup_{y \in \Omega} \langle v, \overline{x} - y \rangle$$

Let

$$\varsigma(v, \overline{x}) = \sup_{y \in \Omega} \langle v, \overline{x} - y \rangle$$

Note that for each $y \in \Omega$, as $\langle v, \overline{x} - y \rangle$ is linear we conclude that $\varsigma(v, \overline{x})$ is in convex in v and lower semicontinous. Moreover, $\varsigma(v, \overline{x})$ is a proper function since

$$0 = \inf_{v \in T(\overline{x})} \varsigma(v, \overline{x})$$

Thus as $T(\overline{x})$ is convex and compact we conclude the existence of $v^* \in T(\overline{x})$ such that

$$0 = \varsigma(v^*, \overline{x})$$

Hence

$$\sup_{y \in \Omega} \langle v^*, \overline{x} - y \rangle = 0$$

Thus for all $y \in \Omega$ we have

$$\langle v^*, \overline{x} - y \rangle \leq 0$$

or

$$\langle v^*, y - \overline{x} \rangle \geq 0$$

This shows that $\overline{x} \in \mathrm{SOL}(T, \Omega)$ and hence $\mathrm{SOL}^M(T, \Omega) \subseteq \mathrm{SOL}(T, \Omega)$. Using Lemma 2 we conclude that

$$\mathrm{SOL}^M(T, \Omega) = \mathrm{SOL}^W(T, \Omega)$$

Therefore, again from Lemma 2 $\mathrm{SOL}(T, \Omega)$ is a convex set. □

Next, we determine conditions that guarantee boundedness of $\mathrm{SOL}(T, \Omega)$.

Proposition 8 (existence and boundedness of $\mathrm{SOL}(T,\Omega)$**)** *Let* $\Omega \subset \mathbb{R}^n$ *be closed convex and nonempty, let* $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *be maximal monotone with* $\mathrm{dom}\, T = \Omega$*. The set of solutions to* $\mathrm{GVI}(T,\Omega)$ *is nonempty and bounded if and only if*

$$w \in \Omega^\infty \setminus \{0\} \implies \exists x \in \Omega \text{ with } \langle v,\ w\rangle > 0 \text{ for some } v \in T(x). \tag{29}$$

If Ω *is bounded, then* $\Omega^\infty = \{0\}$ *and the implication holds trivially.*

Proof. The proof follows from [19, Theorem 12.51] in a minor extension of [19, Exercise 12.52]. We show that (29) is equivalent to the existence of $v \in \mathrm{range}(T + N_\Omega)$ with $\langle v,\ w\rangle > 0$ for each nonzero $w \in \Omega^\infty$. Existence and boundedness of the solution set to $\mathrm{GVI}(T,\Omega)$ then follows directly from [19, Theorem 12.51], since the solution set of $\mathrm{GVI}(T,\Omega)$ coincides with the set $(T + N_\Omega)^{-1}(0)$.

Indeed, if $\Omega^\infty \setminus \{0\}$ is empty then Ω is bounded and there is nothing to prove. Suppose, then, that $w \in \Omega^\infty \setminus \{0\}$. For each $x \in \mathrm{ri}\ \Omega$ and for all $\tau > 0$ we can write $w = \frac{(x_\tau - x)}{\tau}$ for some $x_\tau \in \Omega$, hence $w \in T_\Omega(x)$, the *tangent cone* to Ω for all $x \in \Omega$ [19, Definition 6.25 and Corollary 6.29]. Hence

$$(\forall x \in \Omega) \quad \langle w,\ z\rangle \le 0 \quad \text{for all } z \in N_\Omega(x). \tag{30}$$

Now, by [19, Theorem 12.51] $(T + N_\Omega)^{-1}(0)$ – that is the solution set to $\mathrm{GVI}(T,\Omega)$ – is nonempty and bounded if and only if for each nonzero $w \in (\mathrm{dom}\,(T + N_\Omega))^\infty = \Omega^\infty$ there exists $\hat{v} \in \mathrm{range}(T + N_\Omega)$ with $\langle \hat{v},\ w\rangle > 0$. This means that there exists $x \in \mathrm{dom}\, T \cap \Omega$, $v \in T(x)$ and $z \in N_\Omega(x)$ such that $\hat{v} = v + z$ and

$$\langle v + z, w\rangle > 0.$$

Since $\langle z, w\rangle \le 0$ it follows that $\langle v, w\rangle > 0$. This is exactly the statement in (29). □

To guarantee maximal monotonicity of the related set-valued mapping, which is central to the application of Proposition 8, we will restrict our attention to regularizing functions φ satisfying the following assumption.

Assumption 9

(i) $\varphi : \Omega \to \mathbb{R}$ *is continuous and convex.*
(ii) $0 \in \mathrm{ri}\,(\mathrm{dom}\ \partial\varphi - \Omega)$.

An understanding of convergence of solutions to ($\mathrm{GVI}_{T_\varepsilon}$) to the unregularized monotone problem (VI) is achieved through the solution set to the following generalized variational inequality.

Find $\overline{x} \in \mathrm{SOL}(F,\Omega)$ such that

$$\exists\ v \in \partial\varphi(\overline{x}) \text{ with } \langle v,\ x - \overline{x}\rangle \ge 0 \quad \forall\ x \in \mathrm{SOL}(F,\Omega). \tag{GVIφ}$$

To achieve compactness of the problem (GVIφ) we will require the following assumption.

Assumption 10

(i) $\mathrm{SOL}(F, \Omega)$ *is nonempty and closed.*
(ii) $0 \in \mathrm{ri}\,(\mathrm{dom}\ \partial\varphi - \mathrm{SOL}(F, \Omega))$.

Definition 11 *The indicator function* ι_C *of a set* C *is defined by*

$$\iota_C(x) = \begin{cases} 0, & if\ x \in C, \\ +\infty, & x \notin C. \end{cases}$$

Note that for a closed convex set C, the subdifferential of the indicator function is the normal cone

$$N_C(x) = \partial(\iota_C)(x) = \begin{cases} \{v \in \mathbb{R}^n : \langle v, y - x\rangle \le 0 \quad \forall y \in C\}, & if\ x \in C, \\ \emptyset, & x \notin C. \end{cases}$$

Corollary 12 *Let* S_0 *denote the solution set to* $\mathrm{VI}(F, \Omega)$*. Under Assumptions 2, 9 and 10, the solution set* $\mathrm{SOL}(\partial\varphi, S_0)$ *is nonempty, bounded and convex if and only if for each* $w \in S_0^\infty \setminus \{0\}$*, if any, there is an* $x \in S_0$ *with*

$$\langle v,\ w\rangle > 0 \text{ for some } v \in \partial\varphi(x). \tag{31}$$

Proof. By Assumption 10(i) the solution set S_0 is closed and nonempty. Furthermore, S_0 is convex for Ω convex by the monotonicity and continuity of F (Assumption 2(ii)). The solution set $\mathrm{SOL}(\partial\varphi, S_0)$ can be characterized as $T^{-1}(0) \equiv \{x \in \mathbb{R}^n \mid 0 \in T(x)\}$ where $T \equiv \partial\varphi + N_{S_0}$. For S_0 closed convex, the normal cone mapping $N_{S_0} = \partial\iota_{S_0}$ is maximal monotone, and for φ continuous and convex on Ω with

$$0 \in \mathrm{ri}\ (\mathrm{dom}\ \partial\varphi - S_0) \subset \mathrm{ri}\ (\mathrm{dom}\ \partial\varphi - \Omega)$$

(Assumptions 9(i) and 10(ii)) the operator T is maximal monotone [4, Corollary 24.4]. That $\mathrm{SOL}(\partial\varphi, S_0)$ is nonempty and bounded then follows from Proposition 8. To see that $\mathrm{SOL}(\partial\varphi, S_0)$ is convex, note that $\mathrm{SOL}(\partial\varphi, S_0) = \mathrm{argmin}_{S_0}\,\varphi$, the solution set to a convex optimization problem, and thus $\mathrm{SOL}(\partial\varphi, S_0)$ is convex. This completes the proof. □

Condition (29) holds in particular for *coercive* mappings with respect to Ω.

Definition 13 (coercive mappings) *A mapping* $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *is said to be* coercive with respect to Ω *if, for any* $x_0 \in \Omega$ *and for some* $\gamma > 0$,

$$\liminf_{\|x\|\to\infty} \frac{\langle v,\ x - x_0\rangle}{\|x\|^\gamma} > 0 \quad \forall\ v \in T(x). \tag{32}$$

The above definition uses the convention that the infimum over an empty set is $+\infty$ in the case that $T(x) = \emptyset$ (and in particular, if Ω is bounded and $\mathrm{dom}\ T = \Omega$).

Proposition 14 (existence and boundedness with coercivity) *If T satisfies* (32)*, then* (29) *holds. Moreover, if T is maximally monotone with* $\operatorname{dom} T = \Omega$ *then* (32) *is sufficient for* $\mathrm{SOL}(T, \Omega)$ *to be nonempty and bounded.*

Proof. Let us set $\hat{T} = T + N_\Omega$. If $\Omega^\infty \setminus \{0\}$ is empty, then there is nothing to prove. So let $w \in \Omega^\infty \setminus \{0\}$ and define $x = x_0 + w\tau \in \Omega$ for $x_0 \in \operatorname{ri} \Omega$ and $\tau > 0$. The inequality (32) is equivalent to the existence of a constant $c > 0$ such that for all x large enough

$$\langle v,\ x - x_0\rangle \geq c\|x\|^\gamma \quad \forall\, v \in \hat{T}(x). \tag{33}$$

But this is equivalent to

$$\langle v,\ w\rangle \geq \frac{c\|x_0 + w\tau\|^\gamma}{\tau} > 0 \quad \forall\, v \in \hat{T}(x). \tag{34}$$

The rest follows from Proposition 8. □

2.2 Lagrange Multipliers for Variational Inequalities

Our main result shows the relationship between exact regularization and Lagrange multipliers. What is meant by the latter is developed next. To reduce clutter we will use the notation

$$S_0 \equiv \mathrm{SOL}(F, \Omega), \quad S_\varepsilon \equiv \mathrm{SOL}(T_\varepsilon, \Omega) \text{ for } T_\varepsilon \equiv F + \varepsilon\partial\varphi$$
$$S_\varphi \equiv \mathrm{SOL}(\partial\varphi, S_0) \quad \text{and} \quad S_{G_{\varepsilon\varphi}} \equiv \operatorname{argmin}_{\Omega}\{G + \varepsilon\varphi\}.$$

As noted in the proof of Corollary 12, S_φ is convex for φ convex and S_0 convex. Moreover, for G defined by (8) (the dual gap function associated with $\mathrm{VI}(F, \Omega)$), we have

$$S_0 = \{x \in \Omega \mid G(x) = 0\}.$$

Thus $\operatorname{argmin}_{S_0} \varphi$ is equivalent to the solution of the following convex programming problem

$$\underset{x\in\Omega}{\text{minimize}}\ \varphi(x) \quad \text{subject to} \quad G(x) \leq 0. \tag{$\mathcal{P}_{\varphi,G}$}$$

Problem $(\mathcal{P}_{\varphi,G})$ is then a convex program whose solution set coincides with S_φ. as can be seen by the equivalence of $GVI(\partial\varphi, S_0)$ to the problem of minimizing the convex function φ over the convex set S_0.

The Lagrangian associated with $(\mathcal{P}_{\varphi,G})$ is

$$L(x, \lambda) \equiv \varphi(x) + \lambda G(x) - \sigma_{\mathbb{R}_-}(\lambda) + \iota_\Omega(x) \tag{35}$$

where ι_Ω is the indicator function of Ω and $\sigma_{\mathbb{R}_-}$ is the *support function* – equivalently, the Fenchel conjugate of the indicator function– of the negative orthant. The optimality condition for $(\mathcal{P}_{\varphi,G})$ in Lagrangian form is then (see, for example [19, Chapter 11, Section I])

$$0 \in \partial\varphi(\overline{x}) + \overline{\lambda}\partial G(\overline{x}) + N_\Omega(\overline{x}) \quad \text{ for some } \overline{\lambda} \in N_{\mathbb{R}_-}(G(\overline{x})). \tag{36}$$

Implicitly, we are assuming that $\partial G(\overline{x}) \neq \emptyset$. This leads naturally to the following definition.

Definition 15 (Lagrange multiplier for variational inequalities) *Let S_0 be the solution set to* $\mathrm{VI}(F, \Omega)$ *and $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Let G be the dual gap function associated with* $\mathrm{VI}(F, \Omega)$ *defined by* (8)*. A Lagrange multiplier of the generalized variational inequality* $\mathrm{GVI}(\partial\varphi, S_0)$ *is a constant $\lambda \geq 0$ that is also a Lagrange multiplier of the convex programming problem $(\mathcal{P}_{\varphi,G})$, when it exists.*

Regarding existence, if $\mathrm{argmax}_{y\in\Omega}\langle F(y), \overline{x} - y\rangle \neq \emptyset$ then $F(\overline{y}) \in \partial G(\overline{x})$ where $\overline{y} \in \mathrm{argmax}_{y\in\Omega}\langle F(y), \overline{x} - y\rangle$. The argmax always exists if, for instance, Ω is compact. We will attain existence, instead, under less restrictive conditions.

Proposition 16 *Suppose Ω and F satisfy Assumption 2 and let F be coercive on Ω. Then*

(i) S_0 is nonempty and bounded and
(ii) the dual gap function $G(x) = \sup_{y\in\Omega}\langle F(y), x - y\rangle$ is finite valued for all $x \in \Omega$. Consequently, the supremum is attained and $\partial G(x) \neq \emptyset$ for all $x \in \Omega$.

Proof. (i). Since F is continuous and monotone with $\mathrm{dom}\, F = \Omega$ it is, in fact, maximally monotone. The statement then follows from Proposition 14.

(ii) For the second statement, let us assume, on contrary, that $G(\overline{x}) = +\infty$ for some $\overline{x} \in \Omega$. Then, there exists a sequence $y^k \in \Omega$ such that $\lim_{k\to\infty}\langle F(y^k), \overline{x} - y^k\rangle = \infty$. Since Ω is closed and F is continuous on Ω, it must be that $\|y^k\| \to \infty$ as $k \to \infty$. Thus, for $R_k \equiv \|y^k\|$,

$$-\infty = -\lim_{k\to\infty}\langle F(y^k), \overline{x} - y^k\rangle \geq \liminf_{y\in\mathbb{R}^n\setminus\mathbb{B}_{R_k},\ R_k\to\infty}\langle F(y), y - \overline{x}\rangle.$$

Hence, for any fixed $\gamma > 0$,

$$\liminf_{y\in\mathbb{R}^n\setminus\mathbb{B}_{R_k},\ R_k\to\infty}\frac{\langle F(y), y - \overline{x}\rangle}{\|y\|^\gamma} \leq -\lim_{k\to\infty}\frac{\langle F(y^k), \overline{x} - y^k\rangle}{\|y^k\|^\gamma} \leq 0$$

which is a contradiction to the coercivity of F.

For a given $\overline{x} \in \Omega$, since Ω is closed, either the supremum in $G(\overline{x})$ is achieved at some point $\overline{y} \in \Omega$, or it is achieved in the limit at some point in the asymptotic cone of Ω. In the former case there is nothing to prove. Assume, therefore that there exists a sequence $(y^k)_{k\in\mathbb{N}}$ on Ω with $\|y^k\| \to \infty$ as $k \to \infty$ and $\lim_{k\to\infty}\langle F(y^k), \overline{x} - y^k\rangle = G(\overline{x}) < \infty$. This, however, contradicts the assumption that F satisfies (32), so the supremum must be attained on Ω. □

The next theorem is a transposition of [11, Theorem 2.1] to the setting of generalized variational inequalities and illuminates the connection between exact regularization, the dual gap function and the existence of Lagrange multipliers for GVI$(\partial\varphi, S_0)$.

Theorem 17 *Let Ω, F satisfy Assumption 2.*

(i) If there exists $\overline{\varepsilon} > 0$ such that $S_0 \cap S_{G_{\overline{\varepsilon}\varphi}} \neq \emptyset$, then $S_{G_{\varepsilon\varphi}} \subset S_0$ for all $\varepsilon \in (0, \overline{\varepsilon})$. If, moreover, F is coercive with respect to Ω, then $S_{G_{\varepsilon\varphi}}$ is bounded for all $\varepsilon \in (0, \overline{\varepsilon})$.

(ii) Let φ satisfy Assumption 9. If $x \in \Omega$ then $x \in S_0$ and $x \notin S_\varphi$ implies that $x \notin S_\varepsilon$ for all $\varepsilon > 0$.

(iii) Let φ satisfy Assumption 9. For any $\varepsilon > 0$, $S_0 \cap S_\varepsilon \subset S_\varphi$.

(iv) Let F, Ω and φ together satisfy Assumptions 9 and 10 and let F be coercive with respect to Ω. Then for all $\varepsilon > 0$, $S_0 \cap S_{G_{\varepsilon\varphi}} \subset S_\varphi$.

(v) Let φ satisfy Assumption 9 and let $\overline{\lambda} \geq 0$ be a Lagrange multiplier of GVI$(\partial\varphi, S_0)$. *If $\overline{\lambda} = 0$ then $S_0 \cap S_\varepsilon = S_\varphi$. If $\overline{\lambda} > 0$ and, in addition, Assumption 10 holds for F coercive with respect to Ω, then $S_\varphi = S_0 \cap S_{G_{\varepsilon\varphi}}$ for all $\varepsilon \in (0, \frac{1}{\overline{\lambda}}]$.*

(vi) Let F, Ω and φ together satisfy Assumptions 9 and 10 and let F be coercive with respect to Ω. Let the regularization parameter $\overline{\varepsilon} > 0$ be such that $S_0 \cap S_{G_{\overline{\varepsilon}\varphi}} \neq \emptyset$. Then $\dfrac{1}{\overline{\varepsilon}}$ is a Lagrange multiplier of GVI$(\partial\varphi, S_0)$ *and $S_0 \cap S_{G_{\varepsilon\varphi}} = S_\varphi$ for all $\varepsilon \in (0, \overline{\varepsilon}]$ with $S_{G_{\varepsilon\varphi}} = S_\varphi$ for all $\varepsilon \in (0, \overline{\varepsilon})$.*

Proof. (i). Let $\overline{x} \in S_0 \cap S_{G_{\overline{\varepsilon}\varphi}}$. Since $\overline{x} \in S_0$ we have $G(\overline{x}) = 0$ and thus $\overline{x}$ minimizes the convex function G over the convex set Ω. In fact the set of all minimizers of G over Ω is exactly S_0. Now choose any $x \in \Omega \setminus S_0$. At such points we have $G_{\bar{\varepsilon}}(\overline{x}) \leq G_{\bar{\varepsilon}}(x)$ and $G(\overline{x}) < G(x)$, where $G_{\bar{\varepsilon}}(x) = G(x) + \bar{\varepsilon}\varphi(x)$. Let $\varepsilon \in (0, \bar{\varepsilon})$ and note that

$$\frac{\varepsilon}{\bar{\varepsilon}} G_{\bar{\varepsilon}}(\overline{x}) = \frac{\varepsilon}{\bar{\varepsilon}} G_{\bar{\varepsilon}}(\overline{x}) + (1 - \frac{\varepsilon}{\bar{\varepsilon}}) G(\overline{x}),$$

and, for any $y \in \Omega$,

$$\frac{\varepsilon}{\bar{\varepsilon}} G_{\bar{\varepsilon}}(y) + (1 - \frac{\varepsilon}{\bar{\varepsilon}}) G(y) = G_\varepsilon(y).$$

Since $0 < \dfrac{\varepsilon}{\bar{\varepsilon}} < 1$, this yields, for $x \in \Omega \setminus S_0$,

$$G_\varepsilon(\overline{x}) < \frac{\varepsilon}{\bar{\varepsilon}} G_{\bar{\varepsilon}}(x) + (1 - \frac{\varepsilon}{\bar{\varepsilon}}) G(x) = G_\varepsilon(x),$$

thus $x \notin S_{G_{\varepsilon\varphi}}$. By contraposition we have $x \in S_{G_{\varepsilon\varphi}}$ for $\varepsilon \in (0, \bar{\varepsilon})$ implies $x \in S_0$. This yields the first statement. If, in addition F is coercive, by Proposition 16(i), S_0 is bounded, hence $S_{G_{\varepsilon\varphi}}$ is bounded for all $\varepsilon \in (0, \bar{\varepsilon})$. △

(ii). Let $x \in S_0 \setminus S_\varphi$. For each $v \in \partial\varphi(x)$, there exists $y \in S_0$ (depending on v) such that

$$\langle v, y - x \rangle < 0.$$

On the other hand, for the same pair y and x, since $y \in S_0$, we have

$$\langle F(y), y - x \rangle \leq 0.$$

Since F is monotone this implies that

$$\langle F(x), y - x \rangle \leq 0.$$

Then for any $\varepsilon > 0$ we have

$$\langle F(x) + \varepsilon v, y - x \rangle < 0.$$

Hence $x \notin S_\varepsilon$ as claimed. △

(iii). Let $\overline{x} \in S_0 \cap S_\varepsilon$. Then for some $v \in \partial\varphi(\overline{x})$, we have

$$\langle F(\overline{x}) + \varepsilon v, x - \overline{x} \rangle \geq 0, \quad \forall x \in \Omega \tag{37}$$

and

$$\langle F(\overline{x}), x - \overline{x} \rangle \geq 0 \quad \forall x \in \Omega.$$

On other hand for any $x \in S_0$ we have

$$\langle F(x), \overline{x} - x \rangle \geq 0.$$

Now by the monotonicity of F we have

$$\langle F(\overline{x}), \overline{x} - x \rangle \geq \langle F(x), \overline{x} - x \rangle \geq 0.$$

Hence $\langle F(\overline{x}), x - \overline{x} \rangle = 0$. Thus using (37) we conclude that there exists $v \in \partial\varphi(\overline{x})$ such that for all $x \in S_0$ we have $\langle v, x - \overline{x} \rangle \geq 0$. In other words, $\overline{x} \in S_\varphi$, as claimed. △

(iv). Let $x_\varepsilon \in S_0 \cap S_{G_{\varepsilon\varphi}}$. For all $x \in \Omega$,

$$G(x_\varepsilon) + \varepsilon\varphi(x_\varepsilon) \leq G(x) + \varepsilon\varphi(x).$$

Since $\varepsilon > 0$ we have, for all $x \in \Omega$,

$$\frac{1}{\varepsilon}G(x_\varepsilon) + \varphi(x_\varepsilon) \leq \frac{1}{\varepsilon}G(x) + \varphi(x).$$

This shows that x_ε solves

$$\underset{x \in \Omega}{\text{minimize}} \ \varphi(x) + \frac{1}{\varepsilon}G(x). \tag{38}$$

By Proposition 16(ii), G is finite-valued on Ω since F is coercive on an open set that contains Ω, hence, in particular, dom $G \supset \Omega$. The first-order optimality conditions for (38) are

$$0 \in \partial(\varphi + \frac{1}{\varepsilon} G)(x_\varepsilon) + N_\Omega(x_\varepsilon).$$

By Assumption 10, $0 \in \mathrm{ri}\ (\mathrm{dom}\ \partial\varphi - \Omega) \subset \mathrm{ri}\ (\mathrm{dom}\ \partial\varphi - \mathrm{dom}\ G)$ so we may apply the sum rule for subdifferentials (see, for example, [16, Theorem 3.39]) for the equivalent inclusion

$$0 \in \partial\varphi(x_\varepsilon) + \frac{1}{\varepsilon}\partial G(x_\varepsilon) + N_\Omega(x_\varepsilon).$$

As $\frac{1}{\varepsilon} \in N_{\mathbb{R}_-}(G(x_\varepsilon))$, the above inclusion is just (36), hence $\dfrac{1}{\varepsilon}$ is a Lagrange multiplier of $(\mathcal{P}_{\varphi,G})$ paired with the solution x_ε. Since the solution set to $(\mathcal{P}_{\varphi,G})$ coincides with S_φ, this completes the proof of part (iv). △

(v). Suppose that $\overline{x} \in S_\varphi$ is a solution, paired with the Lagrange multiplier $\overline{\lambda} \geq 0$, to GVI$(\partial\varphi, S_0)$. That is, by (36) the pair $(\overline{x}, \overline{\lambda})$ satisfies

$$0 \in \partial\varphi(\overline{x}) + \overline{\lambda}\partial G(\overline{x}) + N_\Omega(\overline{x}). \tag{39}$$

We consider first the case $\overline{\lambda} = 0$. The optimality condition (39) then simplifies to

$$0 \in \partial\varphi(\overline{x}) + N_\Omega(\overline{x}),$$

hence there exists $v \in \partial\varphi(\overline{x})$ such that

$$\langle v, x - \overline{x}\rangle \geq 0, \quad \forall x \in \Omega. \tag{40}$$

Moreover, since $\overline{x} \in S_\varphi$, we know that $\overline{x} \in S_0$ and hence

$$\langle F(\overline{x}), x - \overline{x}\rangle \geq 0 \quad \forall x \in \Omega. \tag{41}$$

Thus multiplying (40) by $\varepsilon > 0$ and adding to (41) yields

$$\langle F(\overline{x}) + \varepsilon v, x - \overline{x}\rangle \geq 0 \quad \forall x \in \Omega,$$

that is, $\overline{x} \in S_\varepsilon$ and hence $S_\varphi \subseteq S_\varepsilon \cap S_0$. Now by Part (iii) we conclude that, for $\overline{\lambda} = 0$, we have $S_\varphi = S_\varepsilon \cap S_0$.

Consider next the case $\overline{\lambda} > 0$. Note that dom $\varphi \cap$ dom $G \neq \emptyset$ since $S_0 \subset$ dom $\varphi \cap$ dom G. Further φ is continuous on int dom φ and thus continuous on S_0. By Proposition 16 and Assumption 10 we can again apply the sum rule to yield

$$0 \in \partial\varphi(\overline{x}) + \overline{\lambda}\partial G(\overline{x}) + N_\Omega(\overline{x}) = \partial(\varphi + \overline{\lambda}G)(\overline{x}) + N_\Omega(\overline{x}).$$

We conclude that $\overline{x}$ is a minimizer of the convex optimization problem

$$\min_{x\in\Omega} \varphi(x) + \overline{\lambda}G(x)$$

and, hence,

$$\frac{1}{\lambda}\varphi(\overline{x}) + G(\overline{x}) \le \frac{1}{\lambda}\varphi(x) + G(x), \quad \forall x \in \Omega.$$

Now since, $\overline{x} \in S_0$, we have, in fact, $G(\overline{x}) = 0$, so the above inequality simplifies to

$$\frac{1}{\lambda}\varphi(\overline{x}) \le \frac{1}{\lambda}\varphi(x) + G(x), \quad \forall x \in \Omega. \tag{42}$$

Also note that for any $x \in \Omega$

$$0 \le G(x). \tag{43}$$

Multiplying (42) by η and (43) by $(1-\eta)$ with $\eta \in (0,1]$ and adding yields

$$\frac{\eta}{\lambda}\varphi(\overline{x}) \le \frac{\eta}{\lambda}\varphi(x) + G(x), \quad \forall x \in \Omega.$$

Again using the fact that $G(\overline{x}) = 0$, the above inequality can be written as

$$\frac{\eta}{\lambda}\varphi(\overline{x}) + G(\overline{x}) \le \frac{\eta}{\lambda}\varphi(x) + G(x), \quad \forall x \in \Omega. \tag{44}$$

For all $\varepsilon \in (0, \frac{1}{\lambda}]$ there is an $\eta \in (0,1]$ with $\varepsilon = \dfrac{\eta}{\lambda}$. Then, by (44), for all $\varepsilon \in (0, \frac{1}{\lambda}]$,

$$\varepsilon\varphi(\overline{x}) + G(\overline{x}) \le \varepsilon\varphi(x) + G(x) \quad \forall\ x \in \Omega.$$

Hence, for all $\varepsilon \in (0, \frac{1}{\lambda}]$, $\overline{x} \in S_{G_{\varepsilon\varphi}}$, and thus $\overline{x} \in S_0 \cap S_{G_{\varepsilon\varphi}}$. This establishes the inclusion $S_\varphi \subseteq S_0 \cap S_{G_{\varepsilon\varphi}}$. Now by part (iv) this implies that $S_\varphi = S_0 \cap S_{G_{\varepsilon\varphi}}$, as claimed. △

(vi). Suppose that there exists $\bar{\varepsilon} > 0$ such that $S_0 \cap S_{G_{\bar{\varepsilon}\varphi}} \neq \emptyset$. Choose $\overline{x} \in S_0 \cap S_{G_{\bar{\varepsilon}\varphi}}$. Since $\overline{x} \in S_{G_{\bar{\varepsilon}\varphi}}$ we have

$$G(\overline{x}) + \bar{\varepsilon}\varphi(\overline{x}) \le G(x) + \bar{\varepsilon}\varphi(x) \quad \forall x \in \Omega,$$

and hence

$$\frac{1}{\bar{\varepsilon}}G(\overline{x}) + \varphi(\overline{x}) \le \frac{1}{\bar{\varepsilon}}G(x) + \varphi(x) \quad \forall x \in \Omega.$$

Thus $\overline{x}$ solves the convex optimization problem.

$$\min_{x\in\Omega} \varphi(x) + \frac{1}{\bar{\varepsilon}}G(x).$$

Since F is coercive we may apply By Proposition 16 to conclude that G is a finite convex function and $\overline{x}$ satisfies

$$0 \in \partial(\varphi + \frac{1}{\bar{\varepsilon}}G)(\overline{x}) + N_\Omega(\overline{x}).$$

Thus using the sum rule we obtain that

$$0 \in \partial\varphi(\overline{x}) + \frac{1}{\overline{\varepsilon}}\partial G(\overline{x}) + N_{\Omega}(\overline{x}).$$

This shows that $\frac{1}{\overline{\varepsilon}} > 0$ is a Lagrange multiplier of the problem $(\mathcal{P}_{\varphi,G})$. Thus using (v) we conclude that $S_0 \cap S_{G_{\varepsilon\varphi}} = S_{\varphi}$ for all $\varepsilon \in (0, \bar{\varepsilon}]$. By part (i), $S_{G_{\varepsilon\varphi}} \subset S_0$ for all $\varepsilon \in (0, \overline{\varepsilon})$, which yields the second statement and completes the proof. △
□

Corollary 18 (boundedness of solutions to $(\mathcal{P}_{G_{\varepsilon\varphi}})$**)** *Let Assumption 2 hold and let F be coercive with respect to Ω. Assume further that there exists $\overline{\varepsilon} > 0$ such that $S_0 \cap S_{G_{\overline{\varepsilon}\varphi}} \neq \emptyset$. Denote by $\mathcal{U}_{\varepsilon'}$ the set*

$$\mathcal{U}_{\varepsilon'} \equiv \bigcup_{0<\varepsilon\leq\varepsilon'} S_{G_{\varepsilon\varphi}}.$$

For all $\varepsilon' < \overline{\varepsilon}$, the set $\mathcal{U}_{\varepsilon'}$ is nonempty and bounded.

Proof. This is a direct consequence of Theorem 17 (i). □

3 Convergence of regularized VI: regularizing the dual gap function G with φ

In this section we briefly discuss the solution strategies for the regularization approach given by $(\mathcal{P}_{G_{\varepsilon\varphi}})$; that is, we regularize the dual gap function G of $\mathrm{VI}(F, \Omega)$ by $\varepsilon_k\varphi$ and examine solutions x_{ε_k} to $(\mathcal{P}_{G_{\varepsilon\varphi}})$ with parameter ε_k. Abstractly, this simply concerns regularization of convex optimization problems, and therefore is well understood. Our primary interest here is what relation the sequence of solutions to the regularized optimization problems has to the solution set to the unregularized monotone variational inequality. If the condition for the exact regularization (Theorem 17(i)) holds, then the regularized solutions, x_{ε_k}, lie in the solution set S_0 for all k such that $\varepsilon_k < \overline{\varepsilon}$. Moreover, if F is coercive, then by Corollary 18 the sequence (x_{ε_k}) has cluster points, all of which are solutions to $\mathrm{VI}(F, \Omega)$. Therefore, for some k large enough, in order to solve $\mathrm{VI}(F, \Omega)$ for F monotone, it suffices to solve $(\mathcal{P}_{G_{\varepsilon\varphi}})$ for ε_k.

Proposition 19 *Suppose Ω, F satisfy Assumptions 2, and let F be coercive with respect to Ω. Let $(\varepsilon_k)_{k\in\mathbb{N}}$ be a decreasing sequence on $\mathbb{R}_+$ with $\varepsilon_k \searrow 0$ and let x_{ε_k} solve $(\mathcal{P}_{G_{\varepsilon\varphi}})$ with parameter ε_k for each $k \in \mathbb{N}$. If there exists $\varepsilon > 0$ such that $S_0 \cap S_{G_{\varepsilon\varphi}} \neq \emptyset$, then the sequence $(x_{\varepsilon_k})_{k\in\mathbb{N}}$ is bounded and, for all k large enough, $x_{\varepsilon_k} \in S_0$.*

Proof. Boundedness of the sequence $(x_{\varepsilon_k})_{k\in\mathbb{N}}$ follows from Corollary 18. Indeed, since S_0 is bounded (Lemma 1) with $S_{G_{\varepsilon_k\varphi}} \subset S_0$ for all $\varepsilon_k \in (0, \overline{\varepsilon})$ (Theorem 17(i)), then the result follows immediately. □

Motivated by the study of error bounds in [11], we now derive an error bound for $d(S_0, S_{G_{\varepsilon\varphi}})$ in a analogous framework to [11, Theorem 5.1]. For this we introduce the concept of *weak sharpness of order* $\gamma > 1$ for the solution sets of variational inequalities.

The notion of weak sharp minimum for a convex minimization problem has been introduced by Burke and Ferris [5]. We recall that the solution set $S_f \equiv \operatorname{argmin}_{x\in\Omega}\{f(x)\}$ is weakly sharp if there exists a positive number α (sharpness constant) such that

$$f(x) \geq f(\overline{x}) + \alpha\ d(x, S_f) \quad \forall \overline{x} \in S_f.$$

Similarly, the solution set S_f is weakly sharp of order γ if there exists a positive number α (sharpness constant) such that, for each $x \in \Omega$,

$$f(x) \geq f(\overline{x}) + \alpha\ d(x, S_f)^\gamma \quad \forall \overline{x} \in S_f.$$

For any $A \subset \mathbb{R}^n$, it's polar cone is defined as $A^\circ = \{y \in \mathbb{R}^n : \langle y, x\rangle \leq 0 \quad \forall x \in A\}$. The relationship between a cone and its polar cone is, similar to that between a linear subspace and its orthogonal complement. From the characterization of a weak sharp solution for a convex minimization problem with a closed proper objective function f, Marcotte and Zhu [14] extended the concept of weak sharp minima for the variational inequality problem. The solution set S_0 of $\mathrm{VI}(F, \Omega)$ is weakly sharp if, for any $\overline{x} \in S_0$,

$$-F(\overline{x}) \in \operatorname{int}\left(\bigcap_{x\in S_0} [T_\Omega(x) \cap N_{S_0}(x)]^\circ\right). \tag{45}$$

However, it is not obvious how to extend (45) for orders $\gamma > 1$. Since the dual gap function $G(x)$ casts $\mathrm{VI}(F, \Omega)$ as a convex minimization problem, an alternative notion of weak sharp minima of a variational inequality of order 1 based on $G(x)$ has been proposed in [14]. We extend this to orders $\gamma > 1$ and propose a generalization. That is, the set S_0 is weakly sharp of order $\gamma > 1$ if there exists a positive number α (the sharpness constant) such that

$$G(x) \geq \alpha\ d(x, S_0)^\gamma \quad \forall x \in \Omega. \tag{46}$$

Theorem 20 *Let F, Ω and φ together satisfy Assumptions 2, 9 and 10, and let F be coercive with respect to Ω. Suppose that the solution set S_0 is weakly sharp of order $\gamma > 1$ with sharpness constant $\alpha > 0$. Then there exists $\tau > 0$ such that, for all $\varepsilon > 0$,*

$$d(x_\varepsilon, S_0)^{\gamma-1} \leq \tau\varepsilon \quad \forall \ \ x_\varepsilon \in S_{G_{\varepsilon\varphi}}. \tag{47}$$

In particular, $S_{G_{\varepsilon\varphi}}$ is bounded for each $\varepsilon > 0$.

Proof. Let $x_\varepsilon \in S_{G_{\varepsilon\varphi}}$ for some $\varepsilon > 0$ and let $\overline{x}_\varepsilon = P_{S_0}(x_\varepsilon)$, the projection being nonempty by Assumption 10. Then, from the definition of weak sharp minima

$$G(\overline{x}_\varepsilon) + \varepsilon\varphi(\overline{x}_\varepsilon) \geq G(x_\varepsilon) + \varepsilon\varphi(x_\varepsilon) \geq \alpha d(x_\varepsilon, S_0)^\gamma + \varepsilon\varphi(x_\varepsilon).$$

Note that, $G(\overline{x}_\varepsilon) = 0$, hence

$$\alpha d(x_\varepsilon, S_0)^\gamma = \alpha\|x_\varepsilon - \overline{x}_\varepsilon\|^\gamma \leq \varepsilon(\varphi(\overline{x}_\varepsilon) - \varphi(x_\varepsilon)). \tag{48}$$

From the definition of the subdifferential of a convex, real-valued map φ, we have

$$\varphi(\overline{x}_\varepsilon) - \varphi(x_\varepsilon) \leq \langle v_\varepsilon, \overline{x}_\varepsilon - x_\varepsilon\rangle \leq \|v_\varepsilon\|\|x_\varepsilon - \overline{x}_\varepsilon\| \quad \text{for all} \quad v_\varepsilon \in \partial\varphi(\overline{x}_\varepsilon),$$

thus, it follows from (48) that

$$\alpha\|x_\varepsilon - \overline{x}_\varepsilon\|^{\gamma-1} \leq \varepsilon\|v_\varepsilon\|. \tag{49}$$

Now, for F and Ω satisfying Assumption 2 with F coercive on Ω, the solution set S_0 is bounded (Proposition 16(i)). Moreover, by Assumption 9, φ is convex and continuous on Ω, and hence convex and continuous on S_0. Consequently, $\partial\varphi$, that is $\|v_\varepsilon\|$, is bounded, uniformly, on the compact set S_0. Hence the statement follows with $\tau' = \alpha^{-1}M$, where M is the uniform bound for $\|v\|$ with $v \in \partial\varphi(S_0)$. □

Note that this error bound is independent of the existence of Lagrange multipliers or the coincidence of the solution sets S_0 and $S_{G_{\varepsilon\varphi}}$ for some ε (Theorem 17(vi)).

4 Convergence of regularized VI: regularizing F with $\nabla\varphi$

In this section we study the other case of the regularization, where we solve VI(F, Ω) through a sequence of regularized problems VI(T_ε, Ω), where $T_\varepsilon = F + \varepsilon\nabla\varphi$. We restrict ourselves to a differentiable regularization to make the computation easier. We are interested in the approximate solutions to VI(T_ε, Ω) in view of the algorithm that we present later in this section which, in principle, involves generating a sequence of solutions to the regularized problems $VI(T_\varepsilon, \Omega)$ as $\varepsilon \to 0$. Since, it is not possible to compute the exact solution to $VI(T_\varepsilon, \Omega)$ in practice, we seek an approximate solution to $VI(T_\varepsilon, \Omega)$ for every $\varepsilon > 0$ with some error tolerance. Knowing that we are within a given error tolerance is the chief concern of *error bounds*, which we determine in Proposition 23. Error bounds between points in S_ε and S_0 are discussed briefly in section 4.3.

4.1 Convergence of regularized solutions

We begin with a study of the behavior of the path $\{x_\varepsilon : \varepsilon > 0\}$ where x_ε is the unique solution to $VI(T_\varepsilon, \Omega)$ and proceed to show that all the cluster points of the sequences of solutions (exact or approximate) to $\mathrm{VI}(T_\varepsilon, \Omega)$ are the solutions to $\mathrm{VI}(F, \Omega)$ as $\varepsilon \to 0$.

Theorem 21 *When φ is strongly convex and Fréchet differentiable and F is coercive, then the map $\varepsilon \mapsto x_\varepsilon$ is continuous.*

Proof. Let x_ε solve $\mathrm{VI}(T_\varepsilon, \Omega)$. Since φ is strongly convex, $\nabla\varphi$ is strongly monotone and hence $T_\varepsilon = F + \varepsilon\nabla\varphi$ is strongly monotone. Hence, there exists a $\mu_\varepsilon > 0$ such that for any $x, y \in \Omega$,

$$\langle T_\varepsilon(x) - T_\varepsilon(y), x - y\rangle \geq \mu_\varepsilon \|x - y\|^2.$$

This implies

$$\sup_{y\in\Omega} \left[\langle T_\varepsilon(x), x - y\rangle - \mu_\varepsilon \|x - y\|^2\right] \geq \sup_{y\in\Omega} \langle T_\varepsilon(y), x - y\rangle. \tag{50}$$

The expression on the left side of (50) is the regularized gap function $\theta(\cdot, \varepsilon\varphi)$, which is zero at $x = x_\varepsilon$. The right hand side is the dual gap function for $\mathrm{VI}(T_\varepsilon, \Omega)$ (see (8)), which we denote by $G(x, \varepsilon\varphi)$. Since $G(x; \varepsilon\varphi) \geq 0$ for all $x \in \Omega$, (50) implies that $G(x_\varepsilon; \varepsilon\varphi) = 0$ and hence $x_\varepsilon \in S_{G(.;\varepsilon\varphi)}$, where $S_{G(.;\varepsilon\varphi)}$ is the solution set of the convex minimization problem $\min_{x\in\Omega} G(.; \varepsilon\varphi)$. Now by Proposition 16(ii), G is finite valued, and hence G continuous (since it is convex [18, Theorem 10.1]) on the relative interior of Ω, which is nonempty as Ω is nonempty [18, Theorem 6.2]. Therefore, the map $\varepsilon \mapsto S_{G(.;\varepsilon\varphi)}$ is upper-semicontinuous as a set-valued map in the sense of [3, Theorem 4.3.3]. However, the strong monotonicity of T_ε implies that $S_\varepsilon (= S_{G(.;\varepsilon\varphi)})$ is singleton and hence the map $\varepsilon \mapsto x_\varepsilon$ is continuous. □

The type of continuity used in the above proof is not the same as outer semincontinuity defined in 6, which is the same as graph closedness.

Our next results are on the convergence of the sequences of solutions (exact or approximate) to $\mathrm{VI}(T_\varepsilon, \Omega)$. Before we define the concept of an approximate solution to $\mathrm{VI}(T_\varepsilon, \Omega)$, let us introduce some notation. For a given $\varepsilon > 0$ we denote the regularized gap function for $VI(T_\varepsilon, \Omega)$ by $\theta_\alpha(.; \varepsilon\varphi)$ and the D-gap function for $VI(T_\varepsilon, \Omega)$ by $\theta_{\alpha\beta}(.; \varepsilon\varphi)$. These are given by (similar to (9) and (14))

$$\theta_\alpha(x; \varepsilon\varphi) = \sup_{y\in\Omega} \left\{\langle T_\varepsilon(x), x - y\rangle - \frac{\alpha}{2}\|y - x\|^2\right\} \tag{51}$$

$$\theta_{\alpha\beta}(x; \varepsilon\varphi) = \theta_\alpha(x; \varepsilon\varphi) - \theta_\beta(x; \varepsilon\varphi), \quad (\alpha < \beta). \tag{52}$$

We write this more succinctly using the projection.

$$\theta_\alpha(x; \varepsilon\varphi) = \langle T_\varepsilon(x), x - y^\varepsilon_\alpha(x)\rangle - \frac{\alpha}{2}\|y^{\phi,\varepsilon}_\alpha(x) - x_\varepsilon\|^2, \tag{53}$$

with
$$y^{\varepsilon}_{\alpha}(x) = P_{\Omega}[x - \frac{1}{\alpha}T_{\varepsilon}(x)]. \tag{54}$$
The regularized gap function $\theta_\beta(.;\varepsilon\varphi)$ is defined analogously with, instead, the projection $y^{\varepsilon}_{\beta}(x)$.

Recall that for any solution $\overline{x}_\varepsilon$ of $VI(T_\varepsilon, \Omega)$, $\theta_{\alpha\beta}(\overline{x}_\varepsilon;\varepsilon\varphi) = 0$. We define a point x to be an *approximate solution* to $VI(T_\varepsilon, \Omega)$ with an error $\zeta > 0$ if
$$\theta_{\alpha\beta}(x;\varepsilon\varphi) \le \zeta.$$

Theorem 22 *Let $(\varepsilon_k)_{k\in\mathbb{N}}$ be a sequence of nonnegative scalars with $\varepsilon_k \searrow 0$, and let $(x^k)_{k\in\mathbb{N}} \in \mathbb{R}^n$ be a sequence of approximate solutions of* $\mathrm{VI}(T_{\varepsilon_k}, \Omega)$ *with errors $\zeta_k \ge 0$. Assume that φ is continuously differentiable. If $x^k \to \bar{x}$ as $k \to \infty$ and if $\zeta_k \searrow 0$, then $\bar{x}$ solves* $\mathrm{VI}(F, \Omega)$.

Proof. Choose $0 < \alpha < \beta$ and assume that $x^k \to \bar{x}$. Since the projection map onto a closed convex set is continuous, we have $y^{\varepsilon_k}_{\alpha}(x^k) \to y_\alpha(\overline{x})$ and $y^{\varepsilon_k}_{\beta}(x^k) \to y_\beta(\overline{x})$ as $k \to \infty$, where $y^{\varepsilon_k}_{\alpha}$ is defined by (54). Using (53),
$$\lim_{k\to\infty} \theta_\alpha(x^k;\varepsilon_k\varphi) = \langle F(\bar{x}), \bar{x} - y_\alpha(\bar{x})\rangle - \frac{\alpha}{2}\|y_\alpha(\bar{x}) - \bar{x}\|^2 = \theta_\alpha(\bar{x})$$
and
$$\lim_{k\to\infty} \theta_\beta(x^k;\varepsilon_k\varphi) = \langle F(\bar{x}), \bar{x} - y_\beta(\bar{x})\rangle - \frac{\beta}{2}\|y_\beta(\bar{x}) - \bar{x}\|^2 = \theta_\beta(\bar{x}). \tag{55}$$
Now, form (52)
$$0 \le \lim_{k\to\infty} \theta_{\alpha\beta}(x^k;\varepsilon_k\varphi) = \lim_{k\to\infty} \theta_\alpha(x^k;\varepsilon_k\varphi) - \lim_{k\to\infty} \theta_\beta(x^k;\varepsilon_k\varphi) \tag{56}$$
$$= \theta_\alpha(\bar{x}) - \theta_\beta(\bar{x}) = \theta_{\alpha\beta}(\bar{x}) \tag{57}$$
and since x^k is a sequence of approximate solutions
$$\lim_{k\to\infty} \theta_{\alpha\beta}(x^k;\varepsilon_k\varphi) \le \lim_{k\to\infty} \zeta_k = 0. \tag{58}$$
We conclude from (56) and (58) that $\theta_{\alpha\beta}(\bar{x}) = 0$ and therefore $\bar{x}$ solves $\mathrm{VI}(F, \Omega)$ [10, Theorem 10.3.3]. □

As noted in the introduction, solving $\mathrm{VI}(T_\varepsilon, \Omega)$ is equivalent to minimizing the gap function $\theta_\alpha(.;\varepsilon\varphi)$ over Ω or $\theta_{\alpha\beta}(.;\varepsilon\varphi)$ over $\mathbb{R}^n$. If we minimize $\theta_\alpha(.;\varepsilon\varphi)$ over Ω using standard optimization methods for the constrained case, we will in effect generate a sequence of solutions x_ε which are actually be considered as solutions to $\mathrm{VI}(T_\varepsilon, \Omega)$. Alternatively, if we minimize the D-gap function $\theta_{\alpha\beta}(.;\varepsilon\varphi)$ over $\mathbb{R}^n$, we also generate a sequence of solutions to $\mathrm{VI}(T_\varepsilon, \Omega)$.

Using the error bounds for a strongly monotone variational inequalities, we now deduce an error bound for the distance between any point and a true solution of $\mathrm{VI}(T_\varepsilon, \Omega)$ in terms of the corresponding D-gap function $\theta_{\alpha\beta}(.;\varepsilon\varphi)$, provided that F is Lipschitz continuous and φ and strong convex. This error bound can be used as an implementable stopping criterion for the algorithms aimed at approximately solving $\mathrm{VI}(T_\varepsilon, \Omega)$.

The proof of following lemma goes along the lines the proof of Theorem 3.2 in [7] adapted to $\mathrm{VI}(T_\varepsilon, \Omega)$.

Lemma 3 *Let φ be strongly convex with modulus ρ, F and $\nabla\varphi$ be Lipschitz on Ω with constants L and M respectively. If x_ε solves $\mathrm{VI}(T_\varepsilon, \Omega)$, then, for any $x \in \Omega$,*

$$\|x - x_\varepsilon\| \le \frac{\beta + L + \varepsilon M}{\varepsilon\rho}\|y^\varepsilon_\beta(x) - x\|, \tag{59}$$

where $y^\varepsilon_\beta(x)$ is the point where the supremum in $\theta_\beta(x_\varepsilon, \varepsilon\varphi)$ is attained and given by $y^\varepsilon_\beta(x) = P_\Omega[x - \frac{1}{\beta}T_\varepsilon(x)]$.

Proof. Since φ is strongly convex on Ω with modulus ρ, $\nabla\varphi$ is strongly monotone on Ω with modulus ρ. Since F is monotone, T_ε is strongly monotone with modulus of strong monotonicity $\varepsilon\rho$. Also, T_ε is Lipschitz with constant $L+\varepsilon M$. From (51), $y^\varepsilon_\beta(x)$ maximizes the function $y \to \langle T_\varepsilon(x), x - y\rangle - \frac{\beta}{2}\langle y - x, y - x\rangle$. Hence, $y^\varepsilon_\beta(x)$ is the unique minimizer of the strongly convex function $y \to \langle T_\varepsilon(x), y - x\rangle + \frac{\beta}{2}\langle y - x, y - x\rangle$. The optimality conditions yield

$$\langle T_\varepsilon(x) + \beta(y^\varepsilon_\beta(x) - x), x_\varepsilon - y^\varepsilon_\beta(x)\rangle \ge 0.$$

Moreover, since x_ε solves $\mathrm{VI}(T_\varepsilon, \Omega)$, we have

$$\langle T_\varepsilon(x_\varepsilon), y^\varepsilon_\beta(x) - x_\varepsilon\rangle \ge 0.$$

Hence the two inequalities above yield

$$\langle T_\varepsilon(x) - T_\varepsilon(x_\varepsilon) + \beta(y^\varepsilon_\beta(x) - x), y^\varepsilon_\beta(x) - x_\varepsilon\rangle \le 0.$$

Now,

$$\langle T_\varepsilon(x) - T_\varepsilon(x_\varepsilon), y^\varepsilon_\beta(x) - x_\varepsilon + x - x\rangle + \beta\langle y^\varepsilon_\beta(x) - x, y^\varepsilon_\beta(x) - x_\varepsilon + x - x\rangle \ge 0,$$

which implies

$$\begin{aligned}\langle T_\varepsilon(x) - T_\varepsilon(x_\varepsilon), x - x_\varepsilon\rangle + \beta\langle y^\varepsilon_\beta(x) - x, y^\varepsilon_\beta(x) - x\rangle &\le -\beta\langle y^\varepsilon_\beta(x) - x, x - x_\varepsilon\rangle \\ &\quad -\langle T_\varepsilon(x) - T_\varepsilon(x_\varepsilon), y^\varepsilon_\beta(x) - x\rangle.\end{aligned}$$

Since T_ε is Lipschitz with constant $L + \varepsilon M$ and strongly monotone with modulus $\varepsilon\rho$, we get

$$\varepsilon\rho\|x - x_\varepsilon\|^2 \le \beta\|y^\varepsilon_\beta(x) - x\|\|x - x_\varepsilon\| + (L + \varepsilon M)\|y^\varepsilon_\beta(x) - x\|\|x - x_\varepsilon\|.$$

Therefore,

$$\|x - x_\varepsilon\| \le \frac{\beta + L + \varepsilon M}{\varepsilon\rho}\|y^\varepsilon_\beta(x) - x\|$$

as claimed. ☐

Proposition 23 *Let φ be strongly convex with modulus ρ, and let F and $\nabla\varphi$ be Lipschitz on Ω with constants L and M respectively. If x_ε solves* $\mathrm{VI}(T_\varepsilon, \Omega)$, *then, for any $x \in \Omega$,*

$$\|x - x_\varepsilon\| \leq \frac{\beta + L + \varepsilon M}{\varepsilon\rho}\sqrt{\frac{2}{(\beta-\alpha)}\theta_{\alpha\beta}(x;\varepsilon\varphi)}. \tag{60}$$

Proof. Adapting [20, Lemma 4.2, Eq(19)] for $\mathrm{VI}(T_\varepsilon, \Omega)$,

$$\|x - y^\varepsilon_\beta(x)\|^2 \leq \frac{2}{(\beta-\alpha)}\theta_{\alpha\beta}(x;\varepsilon\varphi). \tag{61}$$

Now, from (59) and (61)

$$\|x - x_\varepsilon\| \leq \frac{\beta + L + \varepsilon M}{\varepsilon\rho}\sqrt{\frac{2}{(\beta-\alpha)}\theta_{\alpha\beta}(x;\varepsilon\varphi)} \tag{62}$$

This completes the proof. □

4.2 Sequential inexact descent method

In this section we propose a sequential inexact descent method to solve the $VI(F, \Omega)$ through the regularized problems $VI(T_\varepsilon, \Omega)$ where $T_\varepsilon = F + \varepsilon\nabla\varphi$. It is natural to look for the exact solutions of $\mathrm{VI}(T_{\varepsilon_k}, \Omega)$, however, it is not practically possible to run the algorithm infinitely. We therefore must be satisfied with *approximation* of the solutions to $\mathrm{VI}(T_{\varepsilon_k}, \Omega)$ for each k with an error tolerance τ_k. Convergence behavior of the sequence of approximate solutions will then follow from Proposition 22.

Choose a starting point $x^{k,0} = \overline{x}^0$, ε_0, α_0 and β_0. We solve the unconstrained minimization problem with the objective function $\theta_{\alpha_k\beta_k}(., \varepsilon_k\varphi)$ for $k = 0, 1, 2, ...$ For each k, we collect the approximate solution x^k and initialize the inner iteration for solving $\mathrm{VI}(T_{\varepsilon_{k+1}}, \Omega)$ with the point $x^{k+1,0} = x^k$. The descent method in the inner iteration of Algorithm 1 can be chosen to be any descent method that achieves sufficient decrease in the direction of the descent so that the convergence is guaranteed. The regularization parameters ε_k are updated so that $\varepsilon_k \to 0$ as $k \to \infty$ and the parameters α_k and β_k are updated so that $\alpha_{k+1} \geq \alpha_k$ and $\beta_{k+1} \leq \beta_k$.

We note that many choices exist for the descent method that is used in the inner iteration of Algorithm 1. For example, it can be the descent method proposed in [20] which is free from calculating the derivative of $\theta_{\alpha_k\beta_k}(., \varepsilon_k\varphi)$. Another possibility is the descent method in [13].

Remark 1 The termination of the inner iteration requires the knowledge of the solution x_{ε_k}. It is clear that the Algorithm 1 is implementable as long as the error estimates for $\|x^{k,j} - x_{\varepsilon_k}\|$ are computable. The error bound for $\mathrm{VI}(T_{\varepsilon_k}, \Omega)$ in Proposition 23 is very useful to fill this gap. If $\nabla\varphi$ is ρ-strongly monotone,

Data: Fix sequences of error tolerances $(\tau_k)_{k\in\mathbb{N}}$ and regularization parameters $(\varepsilon_k)_{k\in\mathbb{N}}$ with $\varepsilon_k \to 0$ as $k \to \infty$. For $k = j = 0$, choose the point x^0, parameters: $\beta_0 > \alpha_0 > 0$.
for $k = 0, 1, 2, \dots$ **do**
  Inner iteration: *approximately solve* $\mathrm{VI}(T_{\varepsilon_k}, \Omega)$.
  $x^{k,0} = x^k$
  while $\|x^{k,j} - x_{\varepsilon_k}\| > \tau_k$ **do**
    Apply a descent method to the unconstrained minimization of $\theta_{\alpha_k\beta_k}(., \varepsilon_k\varphi)$ with $x^{k,0}$ as starting point while updating j.
  Update: Set $x^{k+1} = x^{k,j}$, choose $\alpha_{k+1}, \beta_{k+1}$, increment $k = k + 1$, and reset $j = 0$.

Algorithm 1: Sequential inexact descent algorithm

Lipschitz continuous over Ω with modulus M and if F is Lipschitz continuous over Ω with constant L, then T_{ε_k} is strongly monotone with modulus $\varepsilon_k\rho$ and Lipschitz with constant $L + \varepsilon_k M$. Then according to (60),

$$\|x - x_{\varepsilon_k}\| \le L_k\sqrt{\theta_{\alpha_k\beta_k}(x, \varepsilon_k\varphi)}, \text{ where } L_k = \frac{\beta_k + L + \varepsilon_k M}{\varepsilon_k\rho}\sqrt{\frac{2}{(\beta_k - \alpha_k)}}. \tag{63}$$

Hence the stopping criterion in the inner iteration of Algorithm 1 can now be replaced by the implementable rule

$$\text{while } \ \theta_{\alpha_k\beta_k}(x^{k,j}, \varepsilon_k\varphi) > \frac{\tau_k^2}{L_k^2} \tag{64}$$

since $x^{k,j}$ satisfying (64) also satisfies $\|x^{k,j} - x_{\varepsilon_k}\| > \tau_k$.

We now discuss the convergence of Algorithm 1 under appropriate assumptions on F, φ and Ω based on the assumption that for each k, the descent method chosen for the inner iteration converges.

Theorem 24 *Consider the Algorithm 1 with the stopping rule replaced by the alternative stopping rule* (64)*. Assume that F and $\nabla\varphi$ are Lipschitz and $\nabla\varphi$ is strongly monotone. Given a sequence of parameters $(\varepsilon_k)_{k\in\mathbb{N}}$ such that $\varepsilon_k \to 0$ as $k \to \infty$ and and a sequence of stable error tolerances $(\tau_k)_{k\in\mathbb{N}}$, $\tau_k = \tau > 0$, assume that for each k, the descent method in the inner iteration converges. Then all the cluster points of the sequence $(x^k)_{k\in\mathbb{N}}$ of inexact solutions generated by Algorithm 1 are solutions to* $\mathrm{VI}(F, \Omega)$.

Proof. Since our stopping rule $\theta_{\alpha_k\beta_k}(x^{k,j}, \varepsilon_k\varphi) > \frac{\tau_k}{L_k}$ terminates the iterations early, the inner iteration in Algorithm 1 is an early terminated variant of the descent method that is chosen. Hence for any fixed k, any accumulation point of the sequence $x^{k,j}$ delivers an approximate solution x^k to the problem $\mathrm{VI}(T_{\varepsilon_k}, \Omega)$. Since x^k violates the stopping rule, $\theta_{\alpha_k\beta_k}(x^k, \varepsilon_k\varphi) \le \frac{\tau_k}{L_k}$ where L_k

is given as in (63), and since we chose stable error tolerances $\tau_k = \tau$, we have $\frac{\tau_k}{L_k} \searrow 0$ since $L_k \to \infty$ as $k \to \infty$. Hence by Theorem 22 all the cluster points of $(x^k)_{k\in\mathbb{N}}$ are solutions to VI(F, Ω). □

Remark 2 The convergence of the sequence (a subsequence if necessary) of inexact solutions $(x^k)_{k\in\mathbb{N}}$ generated in the Algorithm 1 is guaranteed provided $(x^k)_{k\in\mathbb{N}}$ is bounded. We now establish the sufficient conditions for the boundedness of the sequence $(x^k)_{k\in\mathbb{N}}$ along the similar lines of [9], however, without using the Mountain Pass Theorem. We need the following Lemma.

Lemma 4 *Let $K \subset \mathbb{R}^n$ be compact set and let F and $\nabla\varphi$ be continuous functions on K. Then for any $\varepsilon' > 0$ the gap function $\theta_{\alpha\beta}(.,\varepsilon\varphi)$, $0 < \alpha < \beta$ is uniformly continuous as a function of (x, ε) on $K \times [0, \varepsilon']$. In particular, for every $\delta > 0$, there exists an $\overline{\varepsilon} > 0$ such that*

$$|\theta_{\alpha\beta}(x, \varepsilon\varphi) - \theta_{\alpha\beta}(x)| \le \delta \tag{65}$$

for all $(x, \varepsilon) \in K \times [0, \overline{\varepsilon}]$.

Proof. Recall the D-gap function $\theta_{\alpha\beta}(.,\varepsilon\varphi)$ for $VI(T_\varepsilon, \Omega)$ is given by

$$\theta_{\alpha\beta}(x, \varepsilon\varphi) = \theta_\alpha(x, \varepsilon\varphi) - \theta_\beta(x, \varepsilon\varphi)$$

where

$$\theta_\alpha(x, \varepsilon\varphi) = \langle F(x) + \varepsilon\nabla\varphi(x), x - y_\alpha^\varepsilon(x)\rangle - \frac{\alpha}{2}\|y_\alpha^\varepsilon(x) - x\|^2$$

with

$$y_\alpha^\varepsilon(x) = P_\Omega[x - \frac{1}{\alpha}(F(x) + \varepsilon\nabla\varphi(x)]. \tag{66}$$

Let $(x^n, \varepsilon_n)_{n\in\mathbb{N}}$ be a sequence in $K \times \mathbb{R}_+$ and let $(x^n, \varepsilon_n) \to (x, \varepsilon)$ as $n \to \infty$. Since F and $\nabla\varphi$ are continuous, and since the projection map on a closed convex set is continuous, we have from (66) that

$$\begin{aligned}\lim_{n\to\infty} y_\alpha^{\varepsilon_n}(x^n) &= \lim_{n\to\infty} P_\Omega[x^n - \frac{1}{\alpha}(F(x^n) + \varepsilon_n\nabla\varphi(x^n))] \\ &= P_\Omega[x - \frac{1}{\alpha}(F(x) + \varepsilon\nabla\varphi(x))] = y_\alpha^\varepsilon(x).\end{aligned}$$

Thus $y_\alpha^\varepsilon(x)$ viewed as a function of x and ε is continuous on $K \times \mathbb{R}_+$. This implies that the function $\theta_\alpha(.;\varepsilon\varphi)$ is continuous on $K \times \mathbb{R}_+$ as a function of (x, ε) and so is $\theta_{\alpha\beta}(.;\varepsilon\varphi)$. Since K is a compact set, for any $\varepsilon' \in \mathbb{R}_+$, $\theta_\alpha(x;\varepsilon\varphi)$ is uniformly continuous on $K \times [0, \varepsilon']$. In particular, for a fixed $x \in K$, it holds that for any $\delta > 0$, there exists a $0 < \overline{\varepsilon} < \varepsilon'$ such that for every $\varepsilon \in [0, \overline{\varepsilon}]$

$$|\theta_{\alpha\beta}(x;\varepsilon\varphi) - \theta_{\alpha\beta}(x)| \le \delta.$$

□

Theorem 25 *Consider Algorithm 1 with the stopping rule* (64)*. Assume that F and $\nabla\varphi$ are Lipschitz and $0 < \alpha_k < \beta_k$ for each k. Assume that the solution set S_0 is nonempty and bounded and that $\varepsilon_k \to 0$. Then the sequence $(x^k)_{k\in\mathbb{N}}$ generated by the Algorithm 1 is bounded.*

Proof. Assume that the sequence $(x^k)_{k\in\mathbb{N}}$ generated by the Algorithm 1 is not bounded. Then there exists a compact set $K \in \mathbb{R}^n$ such that $S_0 \subset \text{int}\, K$ and $x^k \notin K$ for sufficiently large k. Denote

$$\overline{m}_k := \min_{x\in\partial K} \theta_{\alpha_k\beta_k}(x), \tag{67}$$

where we use ∂K to denote the boundary of K (not to be confused with the subdifferential, though this should be clear from context). Since the gap function $\theta_{\alpha_k\beta_k}$ is non-negative on $\mathbb{R}^n$ and since $S_0 \subset \text{int}\, K$, it is clear that $\overline{m}_k > 0$. Since $\theta_{\alpha_k\beta_k}(x) \geq \overline{m}_k$ for any $x \in \partial K$, choosing $\delta = c\overline{m}_k, c \in (0,1)$ we have from Lemma 4 that

$$\theta_{\alpha_k\beta_k}(x;\varepsilon_k\varphi) \geq \theta_{\alpha_k\beta_k}(x) - c\overline{m}_k \geq \overline{m}_k - c\overline{m}_k = (1-c)\overline{m}_k \quad \forall x \in \partial K,$$

which implies that

$$m_k := \min_{x\in\partial K} \theta_{\alpha_k\beta_k}(x;\varepsilon_k\varphi) \geq (1-c)\overline{m}_k. \tag{68}$$

Let $\overline{x} \in S_0$. Then $\theta_{\alpha_k\beta_k}(\overline{x}) = 0$ and hence, again from Lemma 4,

$$\theta_{\alpha_k\beta_k}(\overline{x};\varepsilon_k\varphi) = \theta_{\alpha_k\beta_k}(\overline{x};\varepsilon_k\varphi) - \theta_{\alpha_k\beta_k}(\overline{x}) \leq c\overline{m}_k. \tag{69}$$

Since $\theta_{\alpha_k\beta_k}(x^k;\varepsilon_k\varphi) \leq \frac{\tau_k^2}{L_k^2}$ by the stopping rule (64), and since $\frac{\tau_k^2}{L_k^2} \to 0$, for sufficiently large k, we have

$$\theta_{\alpha_k\beta_k}(x^k;\varepsilon_k\varphi) \leq c\overline{m}_k. \tag{70}$$

Let k be sufficiently large such that $x^k \notin K$ and the inequalities (68)-(70) hold. Since $c \in (0,1)$, from (68) we have $c\overline{m}_k \leq \frac{c}{1-c} m_k < m_k$.

Without loss of generality assume that $\theta_{\alpha_k\beta_k}(x^k;\varepsilon_k\varphi) \leq \theta_{\alpha_k\beta_k}(\overline{x};\varepsilon_k\varphi)$. From Weierstrass' extremal value theorem, $\theta_{\alpha_k\beta_k}(.;\varepsilon_k\varphi)$ must attain a maximum at least once in $[x^k,\overline{x}]$. Let $\widehat{x}_k \in [x^k,\overline{x}]$ be the point where $\theta_{\alpha_k\beta_k}(.;\varepsilon_k\varphi)$ attains its maximum. Now, viewing $\widehat{x}_k$ as a local maximizer, it satisfies [6, Proposition 2.3.2]

$$0 = \nabla\theta_{\alpha_k\beta_k}(\widehat{x}_k;\varepsilon_k\varphi). \tag{71}$$

Since $\overline{x} \in \text{int}\, K$ and $x^k \notin K$, there exists a $\lambda \in (0,1)$ such that $x^k_\lambda = \lambda x^k + (1-\lambda)\overline{x} \in \partial K \cap [x^k,\overline{x}]$. Now $\theta_{\alpha_k\beta_k}(\widehat{x}_k;\varepsilon_k\varphi) \geq \theta_{\alpha_k\beta_k}(x^k_\lambda;\varepsilon_k\varphi) \geq m_k$. Hence $\theta_{\alpha_k\beta_k}(\widehat{x}_k;\varepsilon_k\varphi) > 0$. But, the stationary point $\widehat{x}_k$ must be a global minimizer of the D-gap function $\theta_{\alpha_k\beta_k}(.;\varepsilon_k\varphi)$ [13, Theorem 4.3], which is a contradiction. ∎

4.3 Error bounds

Our goal in this section is to develop error bounds for the distance between the solution sets S_ε and S_0. In [14, Theorem 4.1] it is shown that, if the solution set S_0 of $VI(F,\Omega)$ is weakly sharp, Ω is compact and F is pseudomonotone$^+$, then there exists a positive number α such that

$$G(x) \geq \alpha\ d(x, S_0) \quad \forall x \in \Omega. \tag{72}$$

Hence under these three assumptions, we can have one type of error bound in terms of the dual gap function G for the distance between S_ε and S_0. That is, for any $x_\varepsilon \in S_\varepsilon$

$$G(x_\varepsilon) \geq \alpha\ d(x_\varepsilon, S_0) \quad \forall x \in \Omega.$$

We show that, even in the absence of compactness on Ω and the pseudomonotone$^+$ property on F, we can derive an error bound for $d(x_\varepsilon, S_0)$.

In the proof of [14, Theorem 4.1], it is shown that when the solution set S_0 of $VI(F,\Omega)$ is weakly sharp, that is, if (45) holds for S_0, then there exists a positive number α such that for any $x \in \Omega$ and $\overline{x} = P_{S_0}(x)$

$$\langle F(\overline{x}), x - \overline{x}\rangle \geq \alpha\ d(x, S_0). \tag{73}$$

We use this fact to construct an error bound along the lines of Theorem 20. We need the following property, which is a stronger condition than (46) and is, to our knowledge, new.

There exist $\alpha > 0$ and $\gamma \geq 1$ such that

$$\langle F(y), x - y\rangle \geq \alpha\ d(x, S_0)^\gamma \quad \forall x \in \Omega, y = P_{S_0}(x). \tag{74}$$

Theorem 26 *Let Ω and F satisfy Assumption 2 and Assumption 10*(i)*, and let the function φ be convex and differentiable on Ω.*

(i) Assume that the solution set S_0 is weakly sharp (satisfies (45)*) with sharpness constant α. Then, for any $\varepsilon > 0$, and any $x_\varepsilon \in S_\varepsilon$,*

$$\alpha\ dist(x_\varepsilon, S_0) \leq \varepsilon(\varphi(\overline{x}_\varepsilon) - \varphi(x_\varepsilon)), \tag{75}$$

where $\overline{x}_\varepsilon = P_{S_0}(x_\varepsilon)$.

(ii) Let F be coercive with respect to Ω. Suppose that there exist $\gamma > 1$ and $\alpha > 0$ such that (74) *holds. Then there exists $\tau > 0$ such that, for all $\varepsilon > 0$,*

$$d(x_\varepsilon, S_0)^{\gamma-1} \leq \tau\varepsilon \quad \forall \ \ x_\varepsilon \in S_\varepsilon. \tag{76}$$

In particular, S_ε is bounded for each $\varepsilon > 0$.

Proof. We begin with some general observations. Let $\varepsilon > 0$. For any $x_\varepsilon \in S_\varepsilon$

$$\langle F(x_\varepsilon) + \varepsilon \nabla\varphi(x_\varepsilon),\ y - x_\varepsilon\rangle \geq 0 \quad \forall y \in \Omega.$$

Rearranging yields

$$\langle F(x_\varepsilon),\ x_\varepsilon - y\rangle \leq \varepsilon \langle \nabla\varphi(x_\varepsilon),\ y - x_\varepsilon\rangle \quad \forall y \in \Omega.$$

Since F is monotone and φ is convex, for all $y \in \Omega$, it holds that

$$\langle F(y),\ x_\varepsilon - y\rangle \leq \langle F(x_\varepsilon),\ x_\varepsilon - y\rangle \leq \varepsilon \langle \nabla\varphi(x_\varepsilon),\ y - x_\varepsilon\rangle \leq \varepsilon(\varphi(y) - \varphi(x_\varepsilon)).$$

In particular, for $\overline{x}_\varepsilon := P_{S_0}(x_\varepsilon) \in \Omega$ (the projection is nonempty by Assumption 10(i)), we have

$$\langle F(\overline{x}_\varepsilon),\ x_\varepsilon - \overline{x}_\varepsilon\rangle \leq \varepsilon(\varphi(\overline{x}_\varepsilon) - \varphi(x_\varepsilon)). \tag{77}$$

(i). The inequality, (73), together with (77) immediately yields

$$\alpha\ d(x_\varepsilon, S_0) \leq \varepsilon(\varphi(\overline{x}_\varepsilon) - \varphi(x_\varepsilon))$$

as claimed.

(ii). Inequalities (77) and (74) yield

$$\alpha\ d(x_\varepsilon, S_0)^\gamma \leq \varepsilon(\varphi(\overline{x}_\varepsilon) - \varphi(x_\varepsilon)). \tag{78}$$

Since φ is convex real valued,

$$\langle \nabla\varphi(\overline{x}_\varepsilon),\ x_\varepsilon - \overline{x}_\varepsilon\rangle \leq (\varphi(x_\varepsilon) - \varphi(\overline{x}_\varepsilon)).$$

Using Cauchy-Schwarz inequality

$$-\|\nabla\varphi(\overline{x}_\varepsilon)\|_2 \|x_\varepsilon - \overline{x}_\varepsilon\|_2 \leq \langle \nabla\varphi(\overline{x}_\varepsilon),\ x_\varepsilon - \overline{x}_\varepsilon\rangle \leq \varphi(x_\varepsilon) - \varphi(\overline{x}_\varepsilon).$$

Since $\overline{x}_\varepsilon := P_{S_0}(x_\varepsilon)$. this implies that

$$-\ d(x_\varepsilon, S_0)\|\nabla\varphi(\overline{x}_\varepsilon)\|_2 \leq \varphi(x_\varepsilon) - \varphi(\overline{x}_\varepsilon). \tag{79}$$

Combining (79) and (78) yields

$$\alpha d(x_\varepsilon, S_0)^{\gamma-1} \leq \varepsilon \|\nabla\varphi(\overline{x}_\varepsilon)\|_2.$$

Now, for F and Ω satisfying Assumption 2 with F coercive on Ω, the solution set S_0 is bounded (Proposition 16(i)). Moreover, by Assumption 9, φ is convex and, by assumption differentiable, on Ω, and hence convex and differentiable on S_0. Consequently, $\nabla\varphi$ is bounded uniformly on the compact set S_0. Hence the proof follows with $\tau = \alpha^{-1}M$ where M is the uniform bound for $\|\nabla\varphi(\cdot)\|$. ∎

5 Numerical Illustration and Conclusion

We illustrate the theory explored in the previous sections and indicate directions for future investigation with numerical experiments on the following simple example.

5.1 Best Approximation

Example 27 Let

$$\Omega \equiv \left\{ x = (x_1, x_2, x_3)^T \in \mathbb{R}^3 \;\middle|\; \langle n, x\rangle = -1,\ x_1 \leq 1, n = (0,1,1)^T \right\}$$

and define $F(x) \equiv x - P_C(x)$ where $C \equiv \mathbb{R}^3_+ + (0, -1/4, 1/4)$. We compare two regularizing functions, $\varphi_1(x) \equiv \|x\|_1$ and $\varphi_2(x) \equiv \frac{1}{2}\|x\|_2^2$ (shifted Tikhonov) for the approaches to solving (VI) explored separately in Section 3 and Section 4, namely by solving $(\mathcal{P}_{G_{\varepsilon\varphi}})$ and $(\mathrm{GVI}_{T_\varepsilon})$ respectively.

For this problem we know the following.

- $S_0 = \{(x, -\frac{3}{4}, -\frac{1}{4}) \mid x \in [0,1]\}$.
- $S_{\varphi_j} = \{(0, -\frac{3}{4}, -\frac{1}{4})\}$, $(j = 1, 2)$.
- For the regularizer φ_1, $S_{G_{\varepsilon\varphi}} = \{(0, -\frac{3}{4}, -\frac{1}{4})\}$ for all $\varepsilon > 0$.
  Proof sketch. The nearest points in Ω to the point $x_0 = (0,0,0)$ with respect to the ℓ^1 norm are points on the line segment $y(t) = t(0,-1,1) + (0,0,-1)$ for $t \in [0,1]$, and this line segment intersects S_0 at the point $(0, -\frac{3}{4}, -\frac{1}{4})$, where, we know, $G(x)$ attains its minimum. △
- For the regularizer φ_2, the shifted Tikhonov regularizer, $S_{G_{\varepsilon\varphi}} \cap S_0 = \emptyset$ for all $\varepsilon > 0$.
  Proof sketch. The global minimum of φ_2 on Ω, namely the point $(0, -1/2, -1/2)$, does not coincide with those of $G(x)$ (S_0). Moreover, φ_2 is strictly convex on Ω, so the global minimum of the sum cannot be on S_0 for any value of ε. △
- For the regularization φ_1, $S_0 \cap S_\varepsilon = \{(0, -\frac{3}{4}, -\frac{1}{4})\}$ and, in fact $S_\varepsilon = \{(0, -\frac{3}{4}, -\frac{1}{4})\}$ for all $\varepsilon > 0$.
  Proof sketch. Again, because the nearest point in Ω to the origin with respect to the ℓ^1 norm are all points $y(t) = t(0,-1,1) + (0,0,-1)$ for $t \in [0,1]$, and this line segment intersects S_0 at the point $(0, -\frac{3}{4}, -\frac{1}{4})$, by Theorem 17(iii) the claim follows. △
- The regularization φ_2 in $(\mathrm{GVI}_{T_\varepsilon})$ is not exact.
  Proof sketch. For this regularization, a short calculation shows that, for all $\varepsilon > 0$, S_ε is a unique point on the line segment $y(t) = t(0,-1,1)+(0,0,-1)$ for t in the open interval $(\frac{1}{4}, \frac{1}{2})$. This interval does not intersect S_0, that is, $S_0 \cap S_\varepsilon = \emptyset$ for all $\varepsilon > 0$. △
- $\lambda \in [0, +\infty)$ are Lagrange multipliers of $\mathrm{GVI}(\partial\varphi_1, S_0)$.

The numerical results reported in Table 1 were generated from the same initial point, $(1, -2, 1)$. □

Table 1 Comparison of optimization models ($\mathrm{GVI}_{T_\varepsilon}$) and ($\mathcal{P}_{G_{\varepsilon\varphi}}$) with different regularizations (ℓ^1 or ℓ^2) and different regularization parameters ε.

Problem	Iteration	CPU (sec)	Distance to Solution	Distance to S_0
($\mathcal{P}_{G_{\varepsilon\varphi}}$), $\varphi = \ell^1$, $\varepsilon = 0.5$	20	6.889	1.357×10^{-5}	1.357×10^{-5}
($\mathcal{P}_{G_{\varepsilon\varphi}}$), $\varphi = \ell^1$, $\varepsilon = 0.1$	22	6.440	2.264×10^{-9}	2.264×10^{-9}
($\mathcal{P}_{G_{\varepsilon\varphi}}$), $\varphi = \ell^1$, $\varepsilon = 0.01$	32	6.353	8.412×10^{-10}	8.412×10^{-10}
($\mathcal{P}_{G_{\varepsilon\varphi}}$), $\varphi = \ell^1$, $\varepsilon = 0.005$	37	8.552	2.660×10^{-9}	2.660×10^{-9}
($\mathcal{P}_{G_{\varepsilon\varphi}}$), $\varphi = \ell^1$, $\varepsilon = 0.0001$	29	6.903	3.285×10^{-9}	3.285×10^{-9}
($\mathcal{P}_{G_{\varepsilon\varphi}}$), $\varphi = \ell^2$, $\varepsilon = 0.5$	8	8.650	1.768×10^{-1}	1.768×10^{-1}
($\mathcal{P}_{G_{\varepsilon\varphi}}$), $\varphi = \ell^2$, $\varepsilon = 0.1$	10	8.670	5.893×10^{-2}	5.893×10^{-2}
($\mathcal{P}_{G_{\varepsilon\varphi}}$), $\varphi = \ell^2$, $\varepsilon = 0.01$	20	11.68	6.931×10^{-3}	6.931×10^{-3}
($\mathcal{P}_{G_{\varepsilon\varphi}}$), $\varphi = \ell^2$, $\varepsilon = 0.005$	19	12.85	3.500×10^{-3}	3.500×10^{-3}
($\mathcal{P}_{G_{\varepsilon\varphi}}$), $\varphi = \ell^2$, $\varepsilon = 0.0001$	29	23.35	6.078×10^{-5}	1.83×10^{-2}
($\mathrm{GVI}_{T_\varepsilon}$), $\varphi = \ell^1$, $\varepsilon = 0.5$	71	.1099	7.930×10^{-10}	$.7930 \times 10^{-9}$
($\mathrm{GVI}_{T_\varepsilon}$), $\varphi = \ell^1$, $\varepsilon = 0.1$	74	.0860	.0248	$.9277 \times 10^{-9}$
($\mathrm{GVI}_{T_\varepsilon}$), $\varphi = \ell^1$, $\varepsilon = 0.01$	75	.0868	.8125	$.9653 \times 10^{-9}$
($\mathrm{GVI}_{T_\varepsilon}$), $\varphi = \ell^1$, $\varepsilon = 0.005$	75	.0859	.9063	$.9876 \times 10^{-9}$
($\mathrm{GVI}_{T_\varepsilon}$), $\varphi = \ell^1$, $\varepsilon = 0.0001$	76	.0883	.9981	$.7570 \times 10^{-9}$
($\mathrm{GVI}_{T_\varepsilon}$), $\varphi = \ell^2$, $\varepsilon = 0.5$	66	.1018	1.571×10^{-9}	1.179×10^{-1}
($\mathrm{GVI}_{T_\varepsilon}$), $\varphi = \ell^2$, $\varepsilon = 0.1$	281	.3143	1.725×10^{-9}	3.21×10^{-2}
($\mathrm{GVI}_{T_\varepsilon}$), $\varphi = \ell^2$, $\varepsilon = 0.01$	1910	2.153	1.763×10^{-9}	3.5×10^{-3}
($\mathrm{GVI}_{T_\varepsilon}$), $\varphi = \ell^2$, $\varepsilon = 0.005$	3267	3.697	1.766×10^{-9}	1.8×10^{-3}
($\mathrm{GVI}_{T_\varepsilon}$), $\varphi = \ell^2$, $\varepsilon = 0.0001$	6932	7.621	1.768×10^{-9}	3.54×10^{-5}

Example (27) has been purposely designed for simplicity - there are clearly other ways to solve the variational inequality. Recognizing that the problem is one of finding nearest points on the half-plane Ω to the shifted orthant, simple alternating projections would converge to an exact solution finitely, without recourse to regularization. Our purpose, however, is not to explore efficient algorithms for solving this particular problem, but rather to illustrate the theory of (exact) regularization and to underscore the possible advantages of different modelling approaches.

The optimization problem ($\mathcal{P}_{G_{\varepsilon\varphi}}$) was solved using Matlab's *fmincon* with an interior point solver. Evaluation of the dual gap function G given by (8) also involves solving an optimization problem. For the problem in Example (27) this has an explicit representation, but in general this will not be the case. We therefore evaluate the dual gap function numerically so that the experimental results will accurately simulate a practical implementation.

There are a variety of ways to solve ($\mathrm{GVI}_{T_\varepsilon}$). We briefly describe an approach here where the error bounds derived in Section 4.3 are put to use. This is an Armijo descent type algorithm by Li and Ng applicable for a Lipschitz, coercive mapping [13]. Hence, the analysis of Li and Ng applies to the problem $VI(T_\varepsilon, \Omega)$, with appropriate assumptions on F and φ under which T_ε is Lipschitz and coercive on Ω. This method uses the descent direction

$$d^k := \begin{cases} y^{\varphi}_{\alpha_k,\varepsilon_k}(x) - y^{\varphi}_{\beta_k,\varepsilon_k}(x), & \text{if } c_k\|x - y^{\varphi}_{\alpha_k,\varepsilon_k}(x)\| \le \|y^{\varphi}_{\alpha_k,\varepsilon_k}(x) - y^{\varphi}_{\beta_k,\varepsilon_k}(x)\|, \\ y_{\alpha_k,\varepsilon_k}(x) - x, & \text{otherwise;} \end{cases} \tag{80}$$

where $y^{\varphi}_{\alpha_k,\varepsilon_k}(x) = P_\Omega[x - \frac{1}{\alpha_k}T_{\varepsilon_k}(x)]$, the point where the supremum in $\theta_{\alpha_k}(x, \varepsilon_k\varphi)$ is attained and c_k is chosen to satisfy

$$c_k \leq \min\left\{1, \frac{\beta_k - \alpha_k}{2(L^\theta_k + \beta_k)}\right\}.$$

Here L^θ_k is the Lipschitz constant of $\theta_{\alpha_k\beta_k}(., \varepsilon_k\varphi)$ on $Lev^\theta_k = \{x : \theta_{\alpha_k\beta_k}(x; \varepsilon_k\varphi) \leq \theta_{\alpha_k\beta_k}(x^{k,0}; \varepsilon_k\varphi)\}$, where $x^{k,0}$ is the chosen initial point for the inner iteration for each k. It has a step rule that finds the smallest non-negative integer m such that,

$$\sqrt{\theta_{\alpha_k\beta_k}(x^{k,j} + \gamma^m_k d^{k,j}, \varepsilon_k\varphi)} - \sqrt{\theta_{\alpha_k\beta_k}(x^{k,j}, \varepsilon_k\varphi)} \leq -\frac{\delta_k}{4}\gamma^m_k\|d^k\|,$$

and updates $x^{k,j}$ as

$$x^{k,j+1} = x^{k,j} + t_{k,j}d^{k,j} \quad \text{where} \quad t_{k,j} = \gamma^m_k, \gamma_k \in (0,1),$$

where the constant δ_k chosen for a strongly monotone map T_{ε_k} satisfies [13, Remark 4.3]

$$\delta_k \leq \min\left\{\frac{1}{2}\sqrt{\frac{\beta_k - \alpha_k}{2}}, \frac{\sqrt{2}c_k\mu^k_{\alpha_k\beta_k}}{\sqrt{\beta_k - \alpha_k}}\right\},$$

and $\mu^k_{\alpha_k\beta_k}$ is the modulus of strong monotonicity of T_{ε_k}. We use an estimate for L^θ_k and a step size $\gamma_k = .9$ for all k. We note that, since P_C for $C \equiv \mathbb{R}^3_+ + (0, -\frac{1}{4}, \frac{1}{4})$ is nonexpansive, the Lipschitz constant of F defined by $I - P_C$ where I is the identity mapping, is $L = 2$.

– *The regularizer* $\varphi_2 = \frac{1}{2}\|\cdot\|^2_2$: Since the modulus of strong monotonicity of $\nabla\varphi_2$ is $\rho = 1$, T_{ε_k} is strongly monotone with modulus of strong monotonicity ε_k and hence δ_k in this case is chosen to satisfy

$$\delta_k \leq \min\left\{\frac{1}{2}\sqrt{\frac{\beta_k - \alpha_k}{2}}, \frac{\sqrt{2}c_k\varepsilon_k}{\sqrt{\beta_k - \alpha_k}}\right\}.$$

  Note that T_{ε_k} is Lipschitz on Ω. Since T_{ε_k} is strongly monotone, it is coercive too on Ω [13, Remark 2.1]. So, we can apply the method also calculate the error bound p_k in (64). Noting that the Lipschitz constant of $\nabla\varphi_2$ is $M = 1$, choosing $\alpha_k = 1$ and $\beta_k = 2$ for all k, the constant L_k in (63) is calculated as $L_k = \frac{\beta_k + L + \varepsilon_k M}{\varepsilon_k\rho}\sqrt{\frac{2}{(\beta_k - \alpha_k)}} = \frac{4+\varepsilon_k}{\varepsilon_k}\sqrt{2}$. We choose $\tau_k = 10^{-8}/\varepsilon$ in the tolerance $p_k = \frac{\tau^2_k}{L^2_k}$ in (64).
– *The regularizer* $\varphi_1 = \|\cdot\|_1$*:* There is no available theory. We include this experiment to indicate the potential for this approach, and, hopefully, to inspire more research to explain these results.

Remark 3 A few trends from Table 1 are worth noting before we conclude. First, while exact regularization of ($\mathcal{P}_{G_{\varepsilon\varphi}}$) converges to a solution of the unregularized variational inequality, it requires more iterations than (inexact) regularization via the ℓ^2 norm. Nevertheless, the per iteration computational cost, as shown by the CPU times, indicates that the nonsmooth regularization is still more efficient. This could be due to our solution technique for the smooth regularization. If a more efficient method for smooth regularization were available, an iteratively regularized problem, along the lines of Algorithm 1, could be a reasonable strategy. Such a strategy is made possible by the error bound established in Theorem 20. For direct regularization following model (GVI$_{T_\varepsilon}$), we have implemented Algorithm 1 with stopping criteria given by the error bounds established in Theorem 23. This performs as expected for smooth regularization. The distance to the solution to the regularized problem is reported according to the upper bound established in Theorem 23. What is not covered by the theory developed here are the results of our solution to model (GVI$_{T_\varepsilon}$) with the nonsmooth regularization $\varphi = \|\cdot\|_1$. Since we know the answer, we monitored the distance of the iterates to the solution of the regularized and unregularized problems. The gap functions and stopping criteria developed for the case of smooth regularization was not useful or even remotely informative regarding the progress of the iterates. Nevertheless, the direction choice and backtracking procedures appear to function well for this example. The algorithm appears to move quickly to the set S_0, but then cannot make further progress to the solution to the regularized problem, which consists of a single element from S_0. Finally, we note that, as indicated by the tabulated CPU times, the iteration counts should only be used as an indication of the *relative* computational complexity. One iteration of the method of Li and Ng for solving (GVI$_{T_\varepsilon}$) is a tiny fraction of the computational cost of one iteration of our approach to solving ($\mathcal{P}_{G_{\varepsilon\varphi}}$).

5.2 Conclusion

Our inspiration for this study was the theory of exact regularization in optimization developed in [11]. As with optimization, exact regularization for variational inequalities is closely related to the existence of Lagrange multipliers for a related optimization problem, namely ($\mathcal{P}_{\varphi,G}$) (Definition 15. We have found that the dual gap function defined by (8) plays a central role here. The dual gap function is difficult to work with in practice since it is itself the supremum of a nonlinear objective. We determined that, even in the absence of exact regularization, it is possible to establish error bounds for both model approaches to the true solution set introducing the notion of *weak-sharp minimum of degree* γ defined by (46) and (74) respectively. Two avenues for further exploration present themselves. One direction is an investigation of efficient numerical strategies based approximations to the dual gap functional G. The second direction is an investigation of generalized variational inequalities to accommodate nonsmooth, set-valued regularization for regularized variational inequalities of the form (GVI_{T_ε}). Both of these topics are formidable challenges.

5.3 Acknowledgments

The research of CC and DRL was supported by the Deutsche Forschungsgemeinschaft/German Research Foundation grant SFB755-TPA4.

References

1. Auslender, A.: Resolution Numerique d'inegalities variationanelles. Acad. Sci. Paris. Ser. A-B. 276, 1063–1066 (1973).
2. Aussel, D. and Dutta, J.: On gap functions for multivalued Stampacchia variational Inequalities, J. Optim. Theory and Appl., 149(3), 513-527, 2011.
3. Bank, B., Guddat, J., Klatte, D., Kummer, B., Tammer, K.: Non-linear Parametric Optimization. Birkhuser Verlag, Basel (1983).
4. Bauschke, H. H., Combettes, P. L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011).
5. Burke, J. V., Ferris, M. C.: Weak sharp minima in mathematical programming. SIAM J. Control. Optim. 31, 1340–1359 (1993).
6. Clarke, F. H.: Optimization and Nonsmooth Analysis. John Wiley and Sons, New York (1983).
7. Dutta, J.: Gap functons and Errorbounds for Variational and Generalized Variational Inequalities. Vietnam J. Math. 40, 231–253 (2012).
8. Facchinei, F., Pang, J. S., Lampariello, L.: VI-constrained hemivariational inequalities: distributed algorithms and power control in ad-hoc networks. Math. Prog. 145 (1-2), 59-96 (2014).
9. Facchinei, F., Kanzow, C.: Beyond monotonicity in regularization methods for nonlinear complimentarity problems. SIAM J. Control. Optim. 37(4), 1150–1162 (1999).
10. Facchinei, F., Pang, J. S.: Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume I and II. Springer, New York (2003).
11. Friedlander, M. P., Tseng, P.: Exact regularization of convex programs. SIAM J. Optim. 18, 1326–1350 (2007).
12. Fukushima, M.: Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. Math. Prog. Ser. A. 53(1), 99–110 (1992).
13. Li, G., Ng, K. F.: Error bounds of generalized D-gap functions for nonsmooth and nonmonotone variational inequality problems. SIAM J. Optim. 20(No. 2), 667–690 (1998).
14. Marcotte, P., Zhu, D.: Weak sharp solutions of variational inequalities. SIAM J. Optim. 9 (No. 1), 179–189 (1998).
15. Peng, J. M.: Equivalence of variational inequality problems to unconstrained minimization. Math. Prog. 78, 347–355 (1997).
16. Penot, J-P.: Calculus Without Derivatives. Springer, New York (2013).
17. Polyak, B.: Introduction to Optimization. Optimization Software Inc., New York (1987).
18. Rockafellar, R. T.: Convex Analysis. Princeton University Press, Princeton, NJ (2005).
19. Rockafellar, R. T., Wets, R. J-B.: Variational Analysis, Springer, corrected 3rd printing (2009).
20. Yamashita, N., Taji, K., Fukushima, M.: Unconstrained optimization reformulations of variational inequality problems. J. Optim. Theory and Appl. 92, 439–456 (1997).