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

Maximal Monotone Inclusions and Fitzpatrick Functions

J. M. Borwein · J. Dutta

Communicated by Michel Thera.

Abstract In this paper, we study maximal monotone inclusions from the perspective of *gap functions*. We propose a very natural gap function for an arbitrary maximal monotone inclusion, and will demonstrate how naturally this gap function arises from the Fitzpatrick function, which is a convex function, used to represent maximal monotone operators. This allows us to use the powerful *strong Fitzpatrick inequality* to analyse solutions of the inclusion. We also study the special cases of a variational inequality and of a generalized

J. M. Borwein
CARMA, University of Newcastle
Callaghan, NSW, Australia
jon.borwein@gmail.com

J. Dutta, Corresponding author
Economics Group, Dept. of Humanities and Social Science
Indian Institute of Technology Kanpur,
Kanpur, India jdutta@iitk.ac.in

variational inequality problem. The associated notion of a scalar gap is also considered in some detail. Corresponding local and global error bounds are also developed for the maximal monotone inclusion.

Keywords Maximal monotone operator · monotone inclusions · variational inequality · Fitzpatrick function · gap functions · error bounds

Mathematics Subject Classification (2000) 90C30 · 49J52

1 Introduction

Our main focus in this article is the study of the monotone and the maximal monotone inclusion problem through the use of a gap function. Given a monotone or a maximal monotone map, the monotone inclusion problem seeks to find a vector, such that the zero vector lies in the image of that vector under the monotone map. We shall also consider particular forms of the monotone operator, which will lead us to the variational and generalized variational inequality problems. We shall also study these particular problems in terms of their gap function. The new feature of our approach here is that the gap function for the maximal monotone inclusion evolves naturally from the celebrated Fitzpatrick function, which is used to represent a maximal monotone operator. We will describe our goals and the organization of the paper below.

In this article, we study various aspects of the monotone inclusion problem through the lens of a gap function. Gap functions have played a fundamental role in the study of variational inequalities (see for example, Fukushima [1] and Facchinei and Pang [2]). Gap functions allow us to reformulate a variational inequality problem as an optimization problem. They also play a key role in devising error bounds for certain classes of variational inequality problems.

Though there has been a large volume of literature studying the monotone inclusion problem, most of it is geared towards developing algorithms. One of the earliest such papers is due to Rockafellar [3]. To the best of our knowledge, however, there has been no qualitative study of monotone inclusions from the perspective of gap functions. More interestingly, we demonstrate here that an appropriate gap function for a monotone inclusion is derived from the so-called Fitzpatrick function, which is a convex function used to represent maximal monotone operators. We will also see the pivotal role played by the strong Fitzpatrick inequality (see Borwein and Vanderwerff [4]) in understanding various aspects of the monotone inclusion problem. We shall also provide limiting examples to illustrate our results. While we work entirely in finite dimensions, almost all of our results have counterparts for reflexive Banach space.

The paper is organized as follows. In Section 2, we formally introduce the notion of a monotone inclusion and also discuss the related variational inequality problems. In Section 3, we introduce a convex gap function associated with the monotone inclusion problem. We subsequently relate this gap function to obtain convex gap functions associated with variational inequality, and the generalized variational inequality, respectively. We then discuss the issue of finiteness of the gap function wherein coercivity of the set-valued map plays a fundamental role. We additionally introduce the notion of the scalar gap associated with the monotone inclusion. Moreover, we study this gap for special problems such as the complementarity problem, and the variational inequality associated with a primal-dual pair of linear programming problems. In Section 4, we first discuss existence of solutions of the monotone inclusion, and then examine approximate solutions. The strong Fitzpatrick inequality will play a pivotal role here. In Section 5, we study both local and global error bounds for the monotone inclusion. Metric regularity is used as the main vehicle for expressing local error bounds in terms of the gap function. Global error bounds are developed for the monotone inclusion, when T is strongly monotone, by using a regularization of the gap function. In Section 6, we finish by presenting some examples to illustrate the results obtained in this article. At the end there is an Appendix, which consists of a proof of a result in the paper and also the statements of two results used in the paper.

2 Monotone Inclusions and Variational Inequalities

In this section, we formally introduce the notion of a monotone inclusion problem and the variational inequalities, that arise when we consider particular forms of the set-valued map which describes the monotone inclusion.

2.1 Monotone Inclusions

We shall consider a set-valued map $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, which is maximal monotone in the following sense. Recall that a set-valued map $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be *monotone* iff for any x and y in $\mathbb{R}^n$ we have, for all $u \in T(y)$ and $v \in T(x)$,

$$\langle u - v, y - x \rangle \geq 0.$$

The *graph* of a set-valued map T is given as

$$\operatorname{gph} T := \{(x, y) : y \in T(x)\}.$$

The domain of the set-valued map T is given as

$$\operatorname{dom} T := \{x \in \mathbb{R}^n : T(x) \neq \emptyset\},$$

while the range of the set-valued map T is given as

$$\operatorname{ran} T := \bigcup_{x \in \operatorname{dom} T} T(x).$$

A monotone map T is said to be *maximal monotone* iff there is no monotone map whose graph properly contains the graph of T.

In this work, we focus on the following well-studied problem [5] : Given a set-valued map $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, which is maximal monotone, the *monotone inclusion problem* requests a point $x \in \mathbb{R}^n$ such that $0 \in T(x)$.

The set $T^{-1}(0) := \{x \in \mathbb{R}^n : 0 \in T(x)\}$ is the solution set (which may be empty) of our monotone inclusion problem.

2.2 Variational Inequalities

We shall also be interested in two special cases of the monotone inclusion problem.

(a) First, we consider the case where $T(x) := S(x) + N_C(x)$ for each $x \in \mathbb{R}^n$, where $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a maximal monotone map and N_C is the normal cone map associated with the closed and convex set C; we recall that given a closed and convex set C the *normal cone map* $N_C : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is given as follows,

$$N_C(x) := \{v \in \mathbb{R}^n : \langle v, y - x\rangle \leq 0 \quad \forall y \in C\},$$

when $x \in C$ and $N_C(x) := \emptyset$ if $x \notin C$.

Thus, we have $\operatorname{dom} T = C \cap \operatorname{dom} S$. We assume (without any loss of generality) that C has a non-empty interior. We assume that $\operatorname{dom} S \cap \operatorname{int} C \neq \emptyset$, so that $S + N_C$ is maximal monotone [5]. In this case, the monotone inclusion problem requires finding $x \in C$ and $\xi \in S(x)$ such that $\langle \xi, y - x\rangle \geq 0, \quad \forall y \in C$.

This problem is often referred to as the *generalized variational inequality problem* determined by the set-valued map S and the convex set C, and is denoted by $GVI(S, C)$. When $S := \partial f$ for $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, a proper and lower-semicontinuous convex function, then the generalized variational inequality problem reduces to the well known Rockafellar-Pschenychni condition [5] in convex optimization. We note that $GVI(S, C)$ itself reduces to the inclusion problem if $C = \mathbb{R}^n$. Indeed, we can also view the inclusion $0 \in T(x)$ as $GVI(T, \mathbb{R}^n)$.

(b) In the second case we consider $T(x) := F(x) + N_C(x)$ for all $x \in \mathbb{R}^n$, where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a single-valued everywhere continuous and monotone map (hence maximal monotone) and C, as before is a closed and convex set. Since $\operatorname{dom} T = C$ and $\operatorname{dom} F = \mathbb{R}^n$, then $T = F + N_C$ is maximal monotone [5]. Thus, this monotone inclusion problem now reduces to finding $x \in C$ such that $\langle F(x), y - x\rangle \geq 0, \forall y \in C$.

This is traditionally known as *variational inequality* (VI for short), determined by F and C and is denoted by $VI(F, C)$. It is well known that, if $C = \mathbb{R}^n_+$, then the VI reduces to the so called *non-linear complementarity problem* where one wishes to find $x \in \mathbb{R}^n$ such that

$$x \in \mathbb{R}^n_+, F(x) \in \mathbb{R}^n_+, \langle x, F(x)\rangle = 0.$$

The non-linear complementary problem is denoted by $NCP(F)$. If $C = \mathbb{R}^n$, then the VI reduces to the problem of solving equations, that is, of finding an $x \in \mathbb{R}^n$ such that $F(x) = 0$. For more details on variational inequalities see, for example, the two volume monograph of Facchinei and Pang [2], and for monotone operators we refer to [4, Chapter 9].

3 Gap Functions

We begin by describing the notion of a *(convex) gap function* associated with the maximal monotone inclusion (refmonotone-inclusion). This is a (convex) function φ which satisfies,

i) $\varphi(x) \geq 0$ *for all* $x \in \mathbb{R}^n$*.*

ii) $\varphi(x) = 0$ *if and only if* $x \in T^{-1}(0)$*.*

It is important to note that the gap function φ is, in general, an extended-valued function. We will now show that such a function can be constructed out of the celebrated Fitzpatrick function from the theory of monotone operators. See for example [5, Chapter 8] and [4, Chapter 9]. The Fitzpatrick function is a convex function used to represent maximal monotone operators. The original idea of the function was given in Fitzpatrick [6].

The *Fitzpatrick function*, representing a maximal monotone operator T, is the convex function on $\mathbb{R}^n \times \mathbb{R}^n$, given as follows

$$F_T(x, x^*) := \sup_{(y,y^*)\in \operatorname{gph} T} \{\langle y^*, x - y\rangle + \langle x^*, y\rangle\}.$$

An immediate property is that for any maximal monotone operator T we have $F_T(x, x^*) \geq \langle x^*, x\rangle$, with equality holding if and only if $(x, x^*) \in$ graph T. In particular, $F_T(x, 0) \geq 0$, while $F_T(x, 0) = 0$ if and only if $0 \in T(x)$.

Thus, $x \mapsto F_T(x, 0)$ is indeed a gap function for our monotone inclusion. For convenience let us set $G_T(x) := F_T(x, 0)$. Then explicitly

$$G_T(x) = \sup_{y\in \operatorname{dom} T} \sup_{y^*\in T(y)} \langle y^*, x - y\rangle. \tag{1}$$

Moreover, G_T is clearly a convex function. We also have the following less obvious fact.

Theorem 3.1 (Minimality of G_T) *The function G_T is the smallest translation invariant gap function associated with the maximal monotone inclusion.*

Proof: If we apply the definition of G_T to the mapping $x \mapsto T(x) - x^*$, then we determine that $G_{T-x^*}(x) = F_T(x, x^*) - \langle x^*, x\rangle$. Since F_T is known to be minimal among closed *representative functions* [4, Chapter 8], we may deduce that G_T is the smallest translation invariant convex gap function. $\square$

Remark 3.1 (Finitization of G_T) If we are willing to relinquish monotonicity of T, then we may assume that G_T is finite-valued. This can be achieved as follows. Following Crouzeix [7] we consider the function $G_{\widehat{T}}$ instead of G_T, where we define $\widehat{T}$ as follows,

$$\widehat{T}(y) := \left\{ z^*\colon z^* = \frac{y^*}{\max(||y^*||, 1)\max(||y||, 1)}, y^* \in T(y) \right\}.$$

Now $\widehat{T}(y)$ is bounded for all $y \in \mathbb{R}^n$ since examination of the above definition reveals that $||z^*|| \leq 1$ for any $z^* \in \widehat{T}(y)$. Following the arguments above we immediately see that $G_{\widehat{T}}$ is a gap function for the probably non-monotone inclusion $0 \in \widehat{T}(x)$. It is easy to see that $\widehat{T}$ is still *pseudomonotone* in the sense that

$$x^* \in \hat{T}(x); \langle x^*, y - x\rangle \geq 0 \quad \text{implies that} \quad \langle y^*, y - x\rangle \geq 0, \forall y^* \in \hat{T}(y).$$

Though the solution sets of the inclusions $0 \in T(x)$ and $0 \in \widehat{T}(x)$ coincide, the fact that $\widehat{T}$ is usually non-monotone makes it unsuitable for further analysis based on maximal monotonicity. $\diamond$

It is thus natural to ask how to tell if G_T is finite-valued using only conditions on the operator T itself. Perhaps the most natural assumption on T is its coercivity as described next. Given $x \in \mathbb{R}^n$ the monotone operator T is said to be *coercive* at x iff

$$\liminf_{||y|| \to \infty, y^* \in T(y)} \frac{\langle y^*, y - x\rangle}{||y||^2} > 0. \tag{2}$$

We now examine how coercivity of T leads to the finiteness of G_T. To prove finiteness of G_T when T is maximal monotone, we use a notion of *lower-quadraticity* — a condition which is motivated by the core case when T is single-valued and coercive. We note that, if $T(x)$ is single-valued and continuous on its domain, then whenever (2) holds we have

$$\inf_{y^* \in T(y)} \langle y^*, y - x \rangle \geq q_x(y) \tag{3}$$

for some convex quadratic term $q_x(y) := c_x \|y\|^2 - b_x$ with $b_x, c_x > 0$, and we call T *lower quadratic* at x. Now clearly, whenever T is lower quadratic at x, $G_T(x)$ is finite.

More generally, (3) holds when T is coercive and is bounded on bounded sets (but (3) can hold for unbounded mappings). Note, if T is maximally monotone then it is locally bounded on the interior of its domain. Now, in Euclidean space, when T is everywhere defined, then is bounded on bounded sets, and so coercivity again implies lower quadraticity.

We arrive at the following proposition.

Proposition 3.1 (Finiteness of G_T for T monotone) *Consider the maximal monotone inclusion problem $0 \in T(x)$. Then, G_T is everywhere finite and convex, hence continuous, if any one of the following conditions holds.*

i) T is lower-quadratic for all $x \in \mathbb{R}^n$.

ii) Let $\operatorname{dom} T = \mathbb{R}^n$ and let T be coercive for all $x \in \operatorname{dom} T$.

Of course we can deduce the corresponding result that G_T is finite on $\operatorname{dom} T$ by requiring the conditions to hold only on $\operatorname{dom} T$. Indeed, if T is coercive at x in its domain, then monotonicity alone shows T is lower quadratic at such an x.

The following result, showing that G_T is finite-valued function when T is locally bounded on its domain without any assumption of monotonicity on T, is of some independent interest.

Proposition 3.2 (Finiteness of G_T for T non-monotone) *Let T be a set-valued map which is coercive on $\mathbb{R}^n$ and locally bounded on its domain. Then G_T is everywhere finite and convex, hence continuous.*

Proof: Assume that G_T is not finite-valued. Then, there exists $x \in \mathbb{R}^n$ such that $G_T(x) = +\infty$. Thence

$$\sup_{(y.y^*)\in \operatorname{gph} T} \langle y^*, x - y\rangle = +\infty.$$

This shows there exist sequences $y_n \in \mathbb{R}^n$ and $y_n^* \in T(y_n)$, such that

$$\lim_{n\to\infty} \langle y_n^*, x - y_n\rangle = +\infty.$$

We claim that $\{y_n\}$ is unbounded. To the contrary assume $\{y_n\}$ is bounded and, without any loss of generality, assume $y_n \to y$. Since T is locally bounded we can assume $\{y_n^*\}$ is bounded. Thus, we have positive numbers K and L such that $\|y_n\| \le K$ and $\|y_n^*\| \le L$ for all $n \in \mathbb{N}$. Hence,

$$\lim_{n\to\infty} \langle y_n^*, x - y_n\rangle \le L\|x\| + LK < +\infty.$$

This is clearly a contradiction. Hence, for $M > 0$ we have $n_0 \in \mathbb{N}$ such that, for all $n \ge n_0$, $\langle y_n^*, x - y_n\rangle > M$. Hence we have for n sufficiently large that $\dfrac{\langle y_n^*, y_n - x\rangle}{\|y_n\|^2} < -\dfrac{M}{\|y_n\|^2} < 0$. Recall that $\{y_n\}$ is an unbounded sequence. Hence we can assume, without any loss of generality, $\|y_n\| \to \infty$, and so

$$\liminf_{\|y\|\to\infty, (y,y^*)\in \operatorname{gph} T} \frac{\langle y^*, y - x\rangle}{\|y\|^2} \le \liminf_{\|y_n\|\to\infty, (y_n,y_n^*)\in \operatorname{gph} T} \frac{\langle y_n^*, y_n - x\rangle}{\|y_n\|^2} \le 0.$$

This is a contradiction to the fact that T is coercive and hence concludes the result. □

We have discussed above the conditions which ensure that G_T is finite. The finiteness of G_T is one of the key features we shall require to develop useful error bounds for monotone inclusions in terms of the gap function G_T.

We now turn to the notion of a scalar *gap*, which is nothing but the infimal value of the gap function over the feasible set — or of some other set obtained from the problem data. Define

$$\gamma := \gamma_T = \inf_{x \in \mathbb{R}^n} G_T(x).$$

The scalar value $\gamma = \gamma_T$ is called the gap associated with the gap function G_T. We have the following.

Theorem 3.2 *Suppose the maximal monotone inclusion $0 \in T(x)$ has a solution. Then $\gamma = 0$. Conversely assume that $\gamma = 0$ and that G_T is weakly coercive in the following sense that*

$$\lim_{||x|| \to \infty} G_T(x) = +\infty. \tag{4}$$

Then the corresponding maximal monotone inclusion has a solution.

Proof: Let $\bar{x}$ be a solution of the maximal monotone inclusion. Then, $G_T(\bar{x}) = 0$ and thus $\gamma = 0$. Conversely if $\gamma = 0$, by noting that G_T is proper and lower-semicontinuous and weakly coercive we conclude that the infimum of G_T is attained. Thus, there exists $\bar{x} \in \mathbb{R}^n$ such that $0 = G(\bar{x})$, and hence $\bar{x}$ is a solution of the maximal monotone inclusion. □

Remark 3.2 If, in the above result, we had assumed that T is coercive and $\operatorname{dom} T = \mathbb{R}^n$, then we could have easily concluded, using Proposition 3.1, that G_T is finite-valued and hence continuous. That along with the weak coercivity of G_T , gives us our desired result that , if the gap is zero ,then, there exists a solution for the maximal monotone inclusion. $\diamond$

It is important to note from a theoretical point of view, one may consider solving the maximal monotone inclusion problem by unconstrained minimization of the convex function G_T. In such a case, the notion of the gap will play a fundamental role, as illustrated by Theorem 3.2.

Example 3.1 (Gap functions for $VI(F,C)$ or $GVI(S,C)$) In particular, under our hypotheses, for the $VI(F,C)$ or $GVI(S,C)$ problem, the gap function is an extended-valued function ψ such that

i) $\psi(x) \geq 0$ for all $x \in C$ (or for all $x \in \mathbb{R}^n$)

ii) $\psi(x) = 0$, $x \in C$ if and only if x solves $VI(F,C)$ or $GVI(S,C)$.

Hence, $VI(F,C)$ the following is a convex gap function:

$$G(x) := \sup_{y \in C} \langle F(y), x - y \rangle. \tag{5}$$

See the Appendix (Section 7) for a proof that G is a gap function. Note that , if we set $T := F + N_C$, then $VI(F,C)$ becomes the monotone inclusion $0 \in T(x)$. These two problems are equivalent in the sense that the solution sets of these two problems coincide. Further without any loss of generality we can consider $\operatorname{int} C \neq \emptyset$ — as otherwise we may use the relative interior. Then, the sum T is a maximal monotone operator. $\diamond$

The following proposition relates G_T of (1) to G of (5) when $T = F + N_C$.

Proposition 3.3 *If $T = F + N_C$, then for each $x \in C$, $G_T(x) = G(x)$.*

Proof: We have $T = F + N_C$. In this case we have $\operatorname{dom} T = C$. Thence

$$\begin{aligned} G_T(x) = F_T(x,0) &= \sup_{(y,y^*)\in \operatorname{gph}(F+N_C)} \langle y^*, x-y\rangle \\ &= \sup_{y\in\mathbb{R}^n} \sup_{y^*\in F(y)+N_C(y)} \langle y^*, x-y\rangle. \\ &= \sup_{y\in C} \sup_{z^*\in N_C(y)} \langle F(y)+z^*, x-y\rangle. \end{aligned}$$

Given $x \in C$, consider a fixed $y \in C$. Thus we have

$$\sup_{z^*\in N_C(y)} \{\langle F(y), x-y\rangle + \langle z^*, x-y\rangle\} = \langle F(y), x-y\rangle + \sup_{z^*\in N_C(y)} \langle z^*, x-y\rangle.$$

Now using the fact that $z^* \in N_C(x)$ we have

$$\sup_{z^*\in N_C(y)} \{\langle F(y), x-y\rangle + \langle z^*, x-y\rangle\} \le \langle F(y), x-y\rangle.$$

This shows that $G_T(x) \le G(x)$. It remains to prove the reverse inequality. Since $0 \in N_C(y)$ we have

$$\langle F(y), x-y\rangle \le \sup_{z^*\in N_C(y)} \langle F(y)+z^*, x-y\rangle.$$

This shows that $G(x) \le G_T(x)$, and completes the proof. □

Thus when $\operatorname{int} C \neq \emptyset$ and $T(x) = F(x) + N_C(x)$ for all $x \in \mathbb{R}^n$, then we can define $G_T(x) := G(x)$, when $x \in C$ and $G_T(x) := +\infty$ otherwise.

We are now going to examine a gap function designed specifically for $GVI(S,C)$. This will be denoted as g, and is given by

$$g(x) := \sup_{y\in C} \sup_{y^*\in S(y)} \langle y^*, x-y\rangle. \tag{6}$$

When $C = \mathbb{R}^n$, then we have $g(x) = G_S(x)$. Our next task is to show that g is indeed a gap function when S is monotone with some additional properties:

Proposition 3.4 (Gap function for $GVI(S,C)$) *The function g of (6) is a gap function for $GVI(S,C)$, provided S is a non-empty, compact-valued, locally bounded and graph closed, monotone map on C.*

Proof: Let us first show that g is a gap function for the *Minty variational inequality* which seeks an $x \in C$ such that, for all $\xi \in S(y)$, one has

$$\langle \xi, y - x \rangle \geq 0, \quad \forall y \in C.$$

To establish this we first observe that $g(x) \geq 0$ for all $x \in C$, as is seen by setting $y = x$. Note that $g(x) = 0$ leads to the fact that for each $y \in C$, we have $\langle \xi, x - y \rangle \leq 0, \quad \forall \xi \in S(y)$, and hence x is a solution of the Minty variational inequality. Conversely, if x is a solution of the Minty variational inequality, then, we reverse the arguments above to reach the conclusion that $g(x) = 0$.
We shall now show that, if S is locally bounded and graph closed, then every solution of a Minty variational inequality is a solution of the *weak variational inequality*; we say that x is a solution of the weak variational inequality iff for each $y \in C$ there exists $\xi_y \in S(x)$ such that $\langle \xi_y, y - x \rangle \geq 0$.
Suppose x is a solution of the Minty variational inequality, and fix any $y \in C$. Consider the sequence $\{y_n := x + \frac{1}{n}(y - x)\}$. We deduce that for $\xi_n \in S(y_n)$,

$$\langle \xi_n, y_n - x \rangle \geq 0.$$

Now, as $n \to +\infty$ we have $y_n \to x$, and the local boundedness of S guarantees that $\{\xi_n\}$ forms a bounded sequence, which we assume, without any loss of generality to converge to ξ_y. As T is graph closed, we have $\xi_y \in S(x)$. On passing to the limit from the previous inequality, as $n \to \infty$ we have $\langle \xi_y, y-x\rangle \geq 0$. Since $y \in C$ was arbitrary, we have shown that x solves the weak variational inequality problem. In [8] Aussel and Dutta introduced a gap function for the weak variational inequality. It was given as

$$\widehat{g}(x) := \sup_{y\in C} \inf_{\xi\in S(x)} \langle \xi, x-y\rangle.$$

It is proven in [8, Prop. 4.1] (see also Appendix (Section 7)) that $\widehat{g}$ is a gap function for the weak variational inequality problem. Now we observe the following: for each fixed x we have that $\langle \xi, x-y\rangle$ is a convex function in ξ, when y is fixed and is a concave function in y for fixed ξ. By Sion's minimax theorem (see [19] or Appendix (Section 7)) we conclude that

$$\widehat{g}(x) = \inf_{\xi\in S(x)} \sup_{y\in C} \langle \xi, x-y\rangle.$$

Let us consider $x \in C$ such that $\hat{g}(x) = 0$. Then we have

$$0 = \inf_{\xi\in S(x)} \sup_{y\in C} \langle \xi, x-y\rangle.$$

Let us set $\psi(x,\xi) := \sup_{y\in C}\langle \xi, x-y\rangle$. Since x is fixed, $\psi(x,\cdot)$ is a convex function of ξ and is lower-semicontinuous and proper. Thus, appealing to the compactness of $S(x)$ we conclude that there exists $\xi^* \in S(x)$ with $0 = \psi(x,\xi^*)$. This shows that for all $y \in C$, we have $\langle \xi^*, x-y\rangle \leq 0$, and hence x solves $GVI(S,C)$.

Conversely, if $x \in C$ is a solution of $GVI(S,C)$, then, it is straight forward to show that $\hat{g}(x) = 0$. Hence, we deduce that $\widehat{g}$ is a gap function for $GVI(S,C)$.

Let $x \in C$ be such that $g(x) = 0$. Then, x is a solution of the Minty variational inequality. Under our hypothesis, we can conclude that x is a solution of the weak variational inequality. This shows that $\hat{g}(x) = 0$. Thus from the proof above we conclude that $x \in C$ is a solution of $GVI(S, C)$.

Further, when x is a solution of $GVI(S, C)$, by monotonicity of S, it is a solution of the Minty variational inequality, and $g(x) = 0$, as needed. □

Remark 3.3 We note that the above result was already proved in Crouzeix [7] under similar assumptions. The proof we give here is completely different, since it relies essentially on ideas of gap functions. Further note that in general if $T = S + N_C$ where S is a monotone map then the approach in the proof of Proposition 3.3 shows that $G_T(x) = g(x)$ for all $x \in C$. ◇

3.1 Complementarity Problems

In this subsection, we consider the case when $T(x) := F(x) + N_K(x)$ for all x, where K is a closed and convex cone. Then, the monotone inclusion problem reduces to what is known as the *cone complementarity problem*, which requires finding $x \in \mathbb{R}^n$ such that

$$x \in K, F(x) \in K^*, \langle F(x), x \rangle = 0,$$

where K^* is the *dual cone* of K, which is defined by

$$K^* := \{w \in \mathbb{R}^n : \langle w, v \rangle \geq 0, \quad \forall v \in K\}.$$

We will see that in this particular setting, the notion of a gap will provide us much more information about the nature of the cone complementarity problem. The notion of a gap for the nonlinear cone complementarity problem was first introduced by Borwein [9].

A cone complementarity problem can also be viewed as a special case of $VI(F, C)$, where $C = K$ is a closed and convex cone. Let us begin with the case that $K = \mathbb{R}^n_+$ and $F(x) = Mx + q$, where M is a $n \times n$, matrix which is positive semidefinite, but may well not be symmetric. Then, the cone complementarity problem becomes the *linear complementarity problem* and is denoted by $LCP(M, q)$. Following [9] let us define the gap associated with $LCP(M, q)$ as follows:

$$\gamma(q) := \inf\{\langle Mx + q, x\rangle : Mx + q \geq 0, x \geq 0\}, \tag{7}$$

where the ordering in the above expression is taken component-wise. We shall show below that the gap value is always zero and the set of minimizers of the above problem is exactly the solution set of $LCP(M, q)$.

Before we show that the gap value is zero we need to motivate how we arrived at the above definition of the gap for $LCP(M, q)$. At a first glance it is not clear that the gap defined in (7) is related to G_T. The gap in (7) arises from concept of the *Auslender gap function* for $VI(F, C)$ (see [2]) which is given as

$$\theta(x) = \sup_{x \in C} \langle F(x), x - y\rangle.$$

Now monotonicity of F shows that for all x, we have $G(x) \leq \theta(x)$. Thus, from Proposition 3.3 we have for $x \in C$, $G_T(x) \leq \theta(x)$. Now consider the case when $C = \mathbb{R}^n_+$ and $F(x) = Mx + q$, where M is a positive semidefinite matrix.

Since we have $G_T(x) = +\infty$ when $x \notin \mathbb{R}^n_+$, we can write

$$\gamma = \gamma_T = \inf_{x \in \mathbb{R}^n_+} G_T(x) \leq \inf_{x \in \mathbb{R}^n_+} \theta(x).$$

We now claim that $\gamma(q) = \inf_{x \in \mathbb{R}^n_+} \theta(x)$. In fact, if we can establish this we may conclude that for $T(x) = Mx + q + N_{\mathbb{R}^n_+}(x)$, $\gamma_T \leq \gamma(q)$. Thus, when $\gamma(q) = 0$ we will have $\gamma_T = 0$. Further, $\theta(x) = \langle Mx + q, x\rangle - \inf_{y \in \mathbb{R}^n_+} \langle Mx + q, y\rangle$.
A simple calculation shows us that $\theta(x) = Mx + q$ when $Mx + q \geq 0$ and $\theta(x) = +\infty$ otherwise. Hence,

$$\inf_{x \in \mathbb{R}^n_+} \theta(x) = \inf_{x \in \mathbb{R}^n_+} \{\langle Mx + q, x\rangle : Mx + q \geq 0\} = \gamma(q).$$

Proposition 3.5 *Consider the problem $LCP(M, q)$ where M is positive semi-definite. Then, $\gamma(q) = 0$ and*

$$\operatorname{argmin} \{\langle Mx + q, x\rangle : Mx + q \geq 0, x \geq 0\} = \operatorname{sol}(LCP(M, q)),$$

where $\operatorname{sol}(LCP(M, q))$ *denotes the solution set of $LCP(M, q)$.*

Proof: The proof is based on ideas and techniques from [9]. Note that the optimization problem which defines the gap is a convex quadratic problem under linear (or rather affine) constraints; indeed, the objective is $\langle Qx + q, x\rangle$, where $Q = \frac{M+M^*}{2}$ is symmetric. For any x which is feasible for the above problem we have $\langle Mx + q, x\rangle \geq 0$. Thus, the problem is bounded below on the feasible set, and using the Frank-Wolfe Theorem (see [10]) we may conclude that there exists a minimizer for the problem. In other words

$$\operatorname{argmin} \{\langle Mx + q, x\rangle : Mx + q \geq 0, x \geq 0\} \neq \emptyset.$$

We claim that $\gamma(q) = 0$. Let $\bar{x} \in \operatorname{argmin}\{\langle Mx + q, x\rangle : Mx + q \geq 0, x \geq 0\}$. Since the constraints in the convex quadratic optimization problem which defines the gap are affine, then without any constraint qualification, the saddle point conditions hold. Hence, we write down the *Lagrangian*, which is the function $L : \mathbb{R}^n \times \mathbb{R}^n_+ \to \mathbb{R}$, given by

$$L(x, \lambda) := \langle Mx + q, x\rangle - \langle \lambda, Mx + q\rangle.$$

Since $\gamma(q)$ is the infimal value, by the saddle point theorem there exists $\bar{\lambda} \in \mathbb{R}^n_+$ such that

$$L(x, \bar{\lambda}) \geq L(\bar{x}, \bar{\lambda}) = \gamma(q), \forall x \in \mathbb{R}^n_+. \tag{8}$$

Since $\bar{\lambda} \in \mathbb{R}^n_+$, on substituting $x := \bar{\lambda}$ in (8) we have $\gamma(q) \leq 0$. This shows that $\gamma(q) = 0$.

Now that we have established that $\gamma(q) = 0$, it is simple to show that

$$\operatorname{argmin}\{\langle Mx + q, x\rangle : Mx + q \geq 0, x \geq 0\} = \operatorname{sol}(LCP(M, q)).$$

This establishes the result. $\square$

Remark 3.4 (Asymmetry) We emphasize that above we have not assumed M to be symmetric. In general, we can write $M = S + A$, where S is the symmetric part of M and A is the skew-symmetric part. If M is positive semidefinite then we have $\langle x, Sx\rangle \geq 0$ for all x since $\langle x, Ax\rangle = 0$ for all x. In some important cases, $F(x) = Mx + q$ will be monotone without M being symmetric. $\diamond$

One such case comes from linear programming. Consider the following pair of primal-dual linear programming problems:

$$\min\langle c, x\rangle \quad \text{subject to} \quad Ax \geq b, \quad x \geq 0, \tag{9}$$

and

$$\max\langle b, y\rangle \quad \text{subject to} \quad A^T y \leq c, \quad y \geq 0, \tag{10}$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, A is a $m \times n$ matrix, $c \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$. Here the vector inequalities are again in the component-wise sense. From Borwein and Lewis [5, Chapter 8] it follows that primal and dual solvability of the above primal-dual pair of linear programming problem is equivalent to solvability of the following variational inequality $VI(F(x,y), \mathbb{R}^n_+ \times \mathbb{R}^m_+)$, where

$$F(x, y) := Mz + q$$

for $z := (x, y)^T$, and

$$M := \begin{bmatrix} 0 & -A^T \\ A & 0 \end{bmatrix}$$

and $q := (c, -b)^T$. Note that M is positive semi-definite since it is skew:

$$\langle (x, y), M(x, y)\rangle = 0$$

for all $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$. Thus F is a monotone map, but M is never a symmetric matrix if non-trivial.

Equivalence between the variational inequality and the primal-dual pair of linear programs is in the sense that the solution set of $VI(F(x,y), \mathbb{R}^n_+ \times \mathbb{R}^m_+)$ coincides with the combined primal-dual solution set. This is indeed easy to show. We briefly outline only the proof of the fact that any solution of $VI(F(x,y), \mathbb{R}^n_+ \times \mathbb{R}^m_+)$, is a solution of the primal-dual pair of variational inequalities.

Consider $(\bar{x}, \bar{y})$ a solution of $VI(F(x,y), \mathbb{R}^n_+ \times \mathbb{R}^m_+)$. Then we have

$$\langle c - A^T\bar{y}, x - \bar{x}\rangle + \langle A\bar{x} - b, y - \bar{y}\rangle \geq 0, \forall (x,y) \in \mathbb{R}^n_+ \times \mathbb{R}^m_+.$$

Putting $y = \bar{y}$ we conclude that $A^T\bar{y} \leq c$; and putting $x = \bar{x}$ shows that $A\bar{x} \geq b$. Setting $x = 0$ and $y = 0$ leads to the expression

$$-\langle c, \bar{x}\rangle + \langle A^T\bar{y}, \bar{x}\rangle - \langle A\bar{x}, \bar{y}\rangle + \langle b, \bar{y}\rangle \geq 0.$$

This shows that $\langle b, \bar{y}\rangle \geq \langle c, \bar{x}\rangle$. Then using weak duality we conclude that $\langle c, \bar{x}\rangle = \langle b, \bar{y}\rangle$. Hence $\bar{x}$ solves the primal and $\bar{y}$ solves the dual. The converse can be proved very easily. For a complete and more general version of the above discussion, for conic programming, see Borwein and Lewis [5, Theorem 8.3.13].

Thus, it is interesting to consider the consequences for the variational inequality $VI(F, C)$, where $F(x) = Sx + q$ with S being skew-symmetric. In this case using the definition of the gap function G as in (5) we have,

$$G(x) = \langle q, x\rangle + \sup_{y \in C}\langle -(Sx + q), y\rangle. \tag{11}$$

Of course, if x is a solution of the $VI(F, C)$, then, $G(x) = 0$. If x is not a solution of $VI(Sx + q, C)$ then the value $G(x)$ depends on the set C. Of course the variational inequality, associated with the pair of primal dual linear programming problem, is of this type and it is interesting to compute the gap function for it.

Proposition 3.6 *Consider the variational inequality associated to the pair of primal-dual linear programs, as above. Using the definition of the gap function G as in (5), we obtain*

$$G(x, y) = \langle c, x\rangle - \langle b, y\rangle,$$

if (x, y) is feasible to the primal-dual pair, of linear programming problems. If (x, y) is not feasible for the primal-dual pair then we have $G(x, y) = +\infty$.

Proof: For the variational inequality associated with the primal-dual pair of linear programming problems we have $C = \mathbb{R}^n_+ \times \mathbb{R}^m_+$. Then, using (11) we have

$$G(x, y) = \langle (c, -b), (x, y)\rangle + \sup_{(x'y')\in\mathbb{R}^n_+\times\mathbb{R}^m_+} \langle (A^T y - c, b - Ax), (x', y')\rangle. \quad (12)$$

Suppose (x, y) is feasible, then $A^T y \leq c$ and $Ax - b \geq 0$, showing that for all $(x', y') \in \mathbb{R}^n_+ \times \mathbb{R}^m_+$,

$$\langle (A^T y - c, b - Ax), (x', y')\rangle \leq 0.$$

Thus it is simple to see that

$$\sup_{(x'y')\in\mathbb{R}^n_+\times\mathbb{R}^m_+} \langle (A^T y - c, b - Ax), (x', y')\rangle = 0.$$

This shows that

$$G(x, y) = \langle c, x\rangle - \langle b, y\rangle.$$

When (x, y) is not feasible to the primal-dual pair of linear programs it is very simple to see that $G(x, y) = +\infty$. Hence the result. □

To conclude this section, let us consider the cone complementarity problem, where $F(x) := Mx+q$, but now K is merely a closed and convex cone and not necessarily $\mathbb{R}^n_+$. This is referred to as the *generalized linear complementarity problem* (GLCP) [9]. Thus, we have problem

$$x \in K, Mx + q \in K^*, \langle x, Mx + q\rangle = 0. \tag{13}$$

Then, the associated gap problem, as given in Borwein [9], is as follows,

$$\gamma(q) := \inf\{\langle Mx + q, x\rangle : Mx + q \in K^*, x \in K\}.$$

We have the following result.

Proposition 3.7 ((GLCP) [9]) *Consider the complementarity problem of (13). Assume that K is a closed and convex pointed cone with nonempty interior. Suppose the Slater condition holds, that is, there exists $x \in K$ such that $Mx + q \in \operatorname{int} K^*$. Then, $\gamma(q) = 0$.*

Remark 3.5 Proposition 3.5 has a much stronger conclusion than Proposition 3.7 since polyhedrality of the feasible set is, in general, lost in the (GLCP) gap problem. ◇

4 Strong Fitzpatrick Inequality and Existence of Solutions

In this section, we focus on the existence of solutions for the maximal monotone inclusion. We also define and study the nature of approximate solutions for this inclusion.

4.1 Exact Solutions to Inclusions

The main vehicle of our investigation will be two deep and recent results from the theory of maximal monotone operators. These are given as Theorem 9.7.2 and Corollary 9.7.3 in Borwein and Vanderwerff [4].[1] We combine these results, which hold for *all* maximal monotone operators in reflexive Banach space, in the following theorem. They are a consequence of a subtle application of the Fenchel duality theorem.

Theorem 4.1 (Strong Fitzpatrick inequality) *Let $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a maximal monotone operator, then*

$$F_T(x, x^*) - \langle x, x^* \rangle \geq \frac{1}{4} d^2_{\mathrm{gph}\,(T)}(x, x^*), \tag{14}$$

and

$$d^2_{\mathrm{gph}\,T}(x, x^*) \geq \max\{d^2_{\mathrm{dom}\,(T)}(x), d^2_{\mathrm{ran}\,(T)}(x^*)\}. \tag{15}$$

Proof: We shall prove only the second part of the lemma. For the proof of the first part of the result see [4]. Note that

$$d^2_{\mathrm{gph}\,(T)}(x, x^*) = \inf_{(y,y^*) \in \mathrm{gph}\,(T), y \in \mathrm{dom}\,T} \{||y - x||^2 + ||y^* - x^*||^2\}.$$

This shows that

$$d^2_{\mathrm{gph}\,(T)}(x, x^*) \geq \inf_{y^* \in T(y), y \in \mathrm{dom}\,T} ||y - x||^2 = d^2_{\mathrm{dom}(T)}(x).$$

We also have

$$d^2_{\mathrm{gph}\,(T)}(x, x^*) \geq \inf_{y^* \in T(y)} ||y^* - x^*||^2 = d^2_{\mathrm{ran}\,(T)}(x^*).$$

[1] As discussed in [4], the constant 1/4 is not best possible; 1/2 is.

This completes the proof. □

The first inequality in Theorem 4.1 is known as the strong Fitzpatrick inequality. We emphasize that, with no additional hypothesis imposed on T, we always have

$$2\sqrt{G_T(x)} \geq d_{\mathrm{gph}\,(T)}(x,0), \tag{16}$$

when T is maximal monotone.

Let us now establish an almost immediate application of the above result.

Theorem 4.2 *Suppose T is maximal monotone, is coercive in the sense of (2), and let $\operatorname{dom} T = \mathbb{R}^n$. Then, there exists $q \in \mathbb{R}^n$ such that $\|q\| \leq 2\sqrt{G_T(0)}$ such that the inclusion $0 \in T(x) - q$ has a solution.*

Proof: Since T is coercive and locally bounded on its domain by Proposition 3.1 ii) we see that G_T is finite on $\mathbb{R}^n$, and since G_T is convex, it is continuous on $\mathbb{R}^n$. Hence G_T is continuous at $x = 0$. Now consider a decreasing sequence $\{\varepsilon_k\}$ with $\varepsilon_k > 0$ for all $k \in \mathbb{N}$ and $\varepsilon_k \to 0$ as $k \to \infty$. Using the continuity of G_T at $x = 0$, we conclude that for each ε_k there exist $\delta_k \leq \varepsilon_k$ and x_k such that $\|x_k\| \leq \delta_k$ and $G_T(x_k) \leq G_T(0) + \dfrac{\varepsilon_k}{8}$. Now, using Theorem 4.1 and (16) we deduce that for each $\varepsilon_k > 0$ there exists $(y_k, y_k^*) \in \mathrm{gph}\,(T)$ such that

$$\|y_k^*\|^2 + \|x_k - y_k\|^2 \leq 4G_T(x_k) + \frac{\varepsilon_k}{2}.$$

Hence we have $\|y_k^*\|^2 + \|x_k - y_k\|^2 \leq 4G_T(0) + \varepsilon_k$. This shows that

$$\|y_k^*\|^2 \leq 4G_T(0) + \varepsilon_k \tag{17}$$

and

$$\|x_k - y_k\|^2 \leq 4G_T(0) + \varepsilon_k. \tag{18}$$

From (17) we see that $\{y_k^*\}$ is a bounded sequence, and (18) gives us

$$\|y_k\| \leq \|x_k\| + \sqrt{4G_T(0) + \varepsilon_k}.$$

By construction $\{x_k\}$ is bounded, and so this shows that $\{y_k\}$ is also bounded. Thus, without any loss of generality, we assume that $y_k \to \bar{y}$, and since $\{y_k^*\}$ is bounded, let us assume that $y_k^* \to q$. Since T is maximal monotone, it is graph closed and thus we have $q \in T(\bar{y}) \quad \text{i.e.,} \quad 0 \in T(\bar{y}) - q$. Now, passing to the limit as $k \to \infty$ in (17) we see, $\|q\| \leq 2\sqrt{G_T(0)}$. Hence the result. □

Remark 4.1 Note, in the proof of the above result we have considered only the point $x = 0$. In fact, any $x_0 \in \mathbb{R}^n$ would be sufficient for our purpose. Then, we simply deduce $\|q\| \leq 2\sqrt{G_T(x_0)}$. ◇

4.2 Approximate Solutions to Inclusions

In practice, however, it is rarely easy to get the exact solution of a maximal monotone inclusion. This calls for the study of approximate solutions of the maximal monotone inclusion problem. Given $\varepsilon > 0$ we say that x is an ε-*approximate solution* of the maximal monotone inclusion iff there exists $y^* \in T(x)$ with $\|y^*\| < \varepsilon$. Let us also note that associated with the inclusion problem is the gap problem, which seeks a minimizer for the problem

$$\gamma = \inf_{x \in \mathbb{R}^n} G_T(x).$$

We say that $x \in \mathbb{R}^n$ is an *ε-approximate solution to the gap problem* iff $G_T(x) < \varepsilon$. The following result will connect the approximate solutions of the inclusion problem and of its associated gap problem.

Theorem 4.3 (Approximate solutions) *Let $\varepsilon > 0$ be given. Let z be an $\dfrac{\varepsilon}{8}$-approximate solution of the gap problem. Then, there exists y with $\|y-z\| < \sqrt{\varepsilon}$ such that y is an $\sqrt{\varepsilon}$-approximate solution of the maximal monotone inclusion problem $0 \in T(x)$.*

Proof: Using Theorem 4.1 for the given $\varepsilon > 0$ we obtain existence of $(y, y^*) \in \operatorname{gph}(T)$ such that, $\|y^*\|^2 + \|z-y\|^2 \leq 4G_T(z) + \frac{\varepsilon}{2}$. Now noting that z is a $\dfrac{\varepsilon}{8}$-approximate minimizer of the gap function we have $G_T(z) < \dfrac{\varepsilon}{8}$. Hence we conclude that, $\|y^*\|^2 + \|z-y\|^2 < \varepsilon$.This certainly shows that $\|y^*\| < \sqrt{\varepsilon}$ and that$\|y-z\| < \sqrt{\varepsilon}$, and hence establishes the result. $\square$.

Remark 4.2 Theorem 4.3 can be viewed as a variational principle for maximal monotone inclusions. What it says is that, if one can obtain an approximate minimizer of the gap problem, then one can obtain a nearby approximate minimizer of the inclusion problem. $\diamond$

Our intention above was to highlight the logic, not to provide the best possible estimate. It was brought to our attention by a referee that, if we apply the Bronsted-Rockafellar theorem for Fitzpatrick functions, then we obtain a better estimate for the approximate solution with which we begin in Theorem 4.3. This is achieved using Theorem 3.4 in Alves and Svaiter [11] . If x is a an

ε-approximate solution to the gap problem, then we have

$$G_T(x) = F_T(x, 0) - \langle x, 0 \rangle < \varepsilon.$$

Now applying Theorem 3.4 in [11] we deduce the existence of $(y, y^*) \in \operatorname{gph} T$ such that $\|y-x\| < \sqrt{\varepsilon}$ and $\|y^*\| < \sqrt{\varepsilon}$. This shows that y is a $\sqrt{\varepsilon}$-approximate minimum to the inclusion problem.

It is interesting to note further that Monterio and Svaiter [12] had developed certain notions of approximate solutions of the monotone inclusion in terms of the notion of the enlargement of a monotone operator. It would be interesting to see if their ideas can be used in our framework. The following is an obvious corollary.

Corollary 4.1 *Suppose the gap problem has $\gamma = 0$. Then, for any $\varepsilon > 0$ there exists a $\sqrt{\varepsilon}$-approximate solution to the inclusion problem $0 \in T(x)$.*

5 Error Bounds for the Maximal Monotone Inclusion

Before we begin our study of local error bounds let us take another look at inequality (16), which provides a global error bound for the maximal monotone inclusion but viewed from a higher dimension. The gap function G_T then provides an error bound without strong monotonicity assumptions. Traditionally, error bounds are understood as providing estimates for the distance of a point from the actual solution set.

5.1 Metric Regularity and Local Error Bounds

In this section we will assume that T will satisfy the property of metric regularity. We say the maximal monotone mapping T is *metrically regular at* $(\bar{x}, \bar{y}) \in \operatorname{gph} T$ iff there exist real numbers $k > 0$, $\delta > 0$, and $\gamma > 0$ such that

$$d_{T^{-1}(y)}(x) \leq k d_{T(x)}(y) \quad \forall x \in B_\delta(\bar{x}) \quad \text{and} \quad y \in B_\gamma(\bar{y}), \tag{19}$$

and we say that T is metrically regular over the graph if T is metrically-regular for every $(\bar{x}, \bar{y}) \in \operatorname{gph} T$. For more details on metric regularity see for example Dontchev and Rockafellar [13]. In fact, metric regularity is itself a kind of error bound, which can be fine tuned in our setting. Sadly, even subdifferentials of simple convex functions can fail to be metrically regular. Nonetheless, by setting $\bar{y} = 0$, (19) implies that

$$d_{T^{-1}(0)}(x) \leq k d_{T(x)}(0) \quad \forall x \in B_\delta(\bar{x}). \tag{20}$$

To set $\bar{y} = 0$ we have tacitly assumed $(\bar{x}, 0)$ is in the graph of T and so that $\bar{x}$ is a solution. Using (16), we have $\sqrt{G_T(x)} \geq \frac{1}{2} d_{\operatorname{gph}(T)}(x, 0)$, for all $x \in B_\delta(\bar{x})$. Further $d_{\operatorname{gph}(T)}(x, 0) \leq d_{T(x)}(0)$. For developing bounds, however, it seems to be crucial that we have $d_{T(x)}(0) \leq k\, d_{\operatorname{gph}(T)}(x, 0)$, for some $k > 0$. This does not hold in general and so we consider the following set:

$$U_k(T) = \{x \in \mathbb{R}^n : d_{T(x)}(0) \leq k\, d_{\operatorname{gph}(T)}(x, 0)\}.$$

Let us impose the qualification condition that $U_k(T) \neq \emptyset$. Then for any $x \in U_k(T) \cap B_\delta(\bar{x})$, we have, $d_{T^{-1}(0)}(x) \leq 2k\sqrt{G_T(x)}$.

All this said, as is well known, for a maximal monotone operator, metric regularity will force the inverse mapping to be LSC [14] and so single-valued and indeed strongly monotone as discussed in Section 5.3. For completeness we add the fundamental underlying reason for this result (see Borwein [15]).

Proposition 5.1 (Lower-semicontinuity (LSC) and singlevaluedness) *Suppose T is monotone and is lower semicontinuous at x in the domain of T. Then $T(x)$ is singleton.*

Proof: Assume not let y and z lie in $T(x)$. Select h such that $\langle h, y\rangle < \langle h, z\rangle$ and consider the neighbourhood $N = \{w : \langle w, h\rangle < \langle z, h\rangle\}$. Then y lies in $T(x)$ and in N hence $T(x + th)$ meets N for t small and positive since T is LSC at x. But by monotonicity $\langle T(x + th) - z, h\rangle \geq 0$ which is a contradiction. □

5.2 The Convex Case

The previous discussion motivates the need to exploit weaker regularity notions even for the subdifferential of a convex function. The subdifferential map ∂f of a convex function $f : \mathbb{R}^n \to \mathbb{R}$ is *metrically subregular* at $(\bar{x}, \bar{y}) \in \operatorname{gph}(\partial f)$ iff there exist neighbourhoods U and V of $\bar{x}$ and $\bar{y}$, respectively, and $k > 0$ such that

$$d_{(\partial f)^{-1}(\bar{y})}(x) \leq k d_{\partial f(x)\cap V}(\bar{y}) \quad \forall x \in U.$$

We shall call ∂f metrically subregular iff it is metrically subregular at each $(\bar{x}, \bar{y}) \in \operatorname{gph} \partial f$. Note that the function $f(x) := |x|$, $x \in \mathbb{R}$ is indeed metrically subregular. This leads us to the following result which is a simple consequence of [16, Theorem 3.3].

Proposition 5.2 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let S denote the set of all global minimizers of f. Assume that S is non-empty and that ∂f is metrically subregular. Let $\alpha := \inf_{x \in \mathbb{R}^n} f$. Then, for any $\bar{x}$ in the boundary of S there exists a neighbourhood $U_{\bar{x}}$ and a positive number $c_{\bar{x}} > 0$ such that*

$$d_S(x) \leq \sqrt{\frac{f(x) - \alpha}{c_{\bar{x}}}} \quad \forall x \in U_{\bar{x}}.$$

Proof: Note that $S = (\partial f)^{-1}(0)$. Let us consider $\bar{x}$ in the boundary of S. Since ∂f is metrically subregular, we now apply Theorem 3.3 in [16] with $\bar{v} = 0$ to conclude that there exists $c_{\bar{x}} > 0$ and a neighbourhood of $U_{\bar{x}}$ of $\bar{x}$ such that

$$f(x) \geq f(\bar{x}) + c_{\bar{x}} d^2_{(\partial f)^{-1}(0)}(x), \quad \forall x \in U_{\bar{x}}.$$

The final form of the error bound is easily derived by noting that $f(\bar{x}) = \alpha$.
$\square$

Remark 5.1 We ask how we can interpret the above result so that it becomes a useful tool in practice. In fact, we need not always consider the infimum α, but rather can take any available lower bound of f. In many situations we can find such a lower-bound. When running an algorithm for solving a convex optimization problem, if for some iterates the above error bound holds, then, we may argue that such points are near the boundary of the solution set. Sadly, in practice it is often not possible to figure out whether the subdifferential satisfies the notion of metric subregularity. $\diamond$

By contrast, we can exploit Theorem 4.1 for a proper and lower semicontinuous convex function f, as soon as $\mu := \inf f$ is finite. We begin by checking that from the definition of the subgradient and the Young-Fenchel inequality, we have $F_{\partial f}(x, x^*) \leq f(x) + f^*(x^*)$, for all x, x^*. See [17] for details. Thus, when $\mu = -f^*(0)$ is finite, we derive $G_{\partial f}(x) \leq f(x) - \mu$. Hence $\sqrt{G_{\partial f}(x)} \leq \sqrt{f(x) - \mu}$. Thus using (16) we can conclude that

$$d_{\mathrm{gph}\,\partial f}(x, 0) \leq 2\sqrt{f(x) - \mu}.$$

This tells us that, if we know the lower bound of a convex function over $\mathbb{R}^n$, then, just using the functional values and the lower-bound one can develop an error bound though viewed from an higher dimension.

5.3 Error Bounds in the Strongly Monotone Case

In this final subsection we present a new gap function for the maximal monotone inclusion when T is *strongly monotone*. We say that, T is strongly monotone or ρ-strongly monotone, if there exists $\rho > 0$ for all $\xi \in T(y)$ and $\eta \in T(x)$ we have

$$\langle \xi - \eta, y - x \rangle \geq \rho \|y - x\|^2, \quad \forall x, y \in \mathbb{R}^n.$$

The scalar $\rho > 0$ is the *modulus of strong monotonicity*. Our gap fuction is based on regularization of G_T and generalizes an approach of Nesterov and Scrimali [18].

Let T be strongly monotone with modulus of strong monotonicity ρ. We define the function $\widehat{G}_T$ as follows

$$\widehat{G}_T(x) := \sup_{y\in\mathbb{R}^n}\left\{\sup_{y^*\in T(y)}\langle y^*, x-y\rangle + \frac{\rho}{2}\|y-x\|^2\right\}. \tag{21}$$

It is simple to note that the function $\widehat{G}_T$ can be written as

$$\widehat{G}_T(x) := \sup_{y\in\operatorname{dom} T}\left\{\sup_{y^*\in T(y)}\langle y^*, x-y\rangle + \frac{\rho}{2}\|y-x\|^2\right\}.$$

We begin by the following result.

Proposition 5.3 *If T is strongly monotone with modulus ρ, then the function $\widehat{G}_T$ is a finite-valued, strongly convex and continuous function.*

Proof: We first show that $\widehat{G}_T$ is finite-valued. The definition of strong monotonicity shows that for any fixed x and y, for all $x^* \in T(x)$ and $y^* \in T(y)$ we have

$$\langle y^*, x-y\rangle + \frac{\rho}{2}\|y-x\|^2 \leq \langle x^*, x-y\rangle - \frac{\rho}{2}\|y-x\|^2.$$

The above inequality leads easily to the following,

$$\sup_{y\in\mathbb{R}^n}\left\{\sup_{y^*\in T(y)}\langle y^*, x-y\rangle + \frac{\rho}{2}\|y-x\|^2\right\} \leq \sup_{y\in\mathbb{R}^n}\left\{\inf_{x^*\in T(x)}\langle x^*, x-y\rangle - \frac{\rho}{2}\|y-x\|^2\right\}. \tag{22}$$

Let us set $\varphi(x,y) := \inf_{x^*\in T(x)}\langle x^*, x-y\rangle$. Note that when x is fixed for each x^*, the function $y \mapsto \langle x^*, x-y\rangle$ is affine in y. Thus, for each fixed x we see that $\varphi(x,.)$ is a concave function. Thus the function $\varphi(x,y) - \frac{\rho}{2}\|y-x\|^2$ is a strongly concave function in y for each x. Hence, there is a unique maximizer of the problem,

$$\sup_{y\in\mathbb{R}^n}\left\{\varphi(x,y) - \frac{\rho}{2}\|y-x\|^2\right\}.$$

Consequently, we have

$$\widehat{G}_T(x) = \sup_{y\in\mathbb{R}^n} \left\{ \sup_{y^*\in T(y)} \langle y^*, x-y\rangle + \frac{\rho}{2}\|y-x\|^2 \right\} < +\infty.$$

By setting $y = x$ we conclude that $\widehat{G}_T(x) \geq 0$ for all $x \in \mathbb{R}^n$. This shows that $\widehat{G}_T$ is finite. We now show that $\widehat{G}_T$ is (strongly) convex. Let us set

$$\psi(x, y) := \sup_{y^*\in T(y)} \langle y^*, x-y\rangle.$$

Note that for every fixed y, the function $\langle y^*, x-y\rangle$ is affine in x for each $y^* \in T(y)$. Thus for each fixed y, the function $\psi(., y)$ is convex in x. Thus, $\widehat{G}_T$ is a strongly convex function in x as the supremum of a family of strongly convex functions in x. Hence the result is established. □

We now demonstrate that under natural conditions, $\widehat{G}_T$ is a gap function for the associated monotone inclusion.

Theorem 5.1 *Let T be maximal monotone operator with non-empty convex and compact-values throughout $\mathbb{R}^n$. Suppose that T is strongly monotone with $\rho > 0$ as the modulus of strong convexity. Then, $\widehat{G}_T$ is a gap function for the maximal monotone inclusion $0 \in T(x)$.*

Proof: Let us begin by considering $\bar{x}$ to be a solution of the monotone inclusion. This implies that for any $y \in \mathbb{R}^n$, we have

$$\inf_{x^*\in T(\bar{x})} \langle x^*, \bar{x}-y\rangle \leq 0 \leq \frac{\rho}{2}\|\bar{x}-y\|^2.$$

This shows that

$$\sup_{y\in\mathbb{R}^m} \left\{ \inf_{x^*\in T(\bar{x})} \langle x^*, \bar{x}-y\rangle - \frac{\rho}{2}\|\bar{x}-y\|^2 \right\} \leq 0.$$

Now, using (22) we conclude that $\widehat{G}_T(\bar{x}) \leq 0$. It is simple to see that $\widehat{G}_T(x) \geq 0$ for all $x \in \mathbb{R}^n$. This shows that $\widehat{G}_T(\bar{x}) = 0$.

Now suppose only that $\bar{x}$ satisfies $\widehat{G}_T(\bar{x}) = 0$. Then, for all $y \in \mathbb{R}^n$ we have

$$\sup_{y^* \in T(y)} \langle y^*, \bar{x} - y \rangle \leq -\frac{\rho}{2} \|y - \bar{x}\|^2. \tag{23}$$

Consider a fixed but arbitrary $y \in \mathbb{R}^n$, and set $y_k := \bar{x} + \dfrac{1}{k}(y - \bar{x})$. Thus $y_k \to \bar{x}$ as $k \to \infty$. Now, using (23) we have

$$\sup_{y^* \in T(y_k)} \langle y^*, \bar{x} - y_k \rangle \leq -\frac{\rho}{2} \|y_k - \bar{x}\|^2.$$

As T has compact images, we conclude that for each k there exists $y_k^* \in T(y_k)$ such that $\langle y_k^*, \bar{x} - y_k \rangle \leq -\frac{\rho}{2} \|y_k - \bar{x}\|^2$. A simple calculation shows

$$\langle y_k^*, \bar{x} - y \rangle \leq -\frac{\rho}{2k} \|y - \bar{x}\|^2. \tag{24}$$

Since T is locally bounded it follows that $\{y_k^*\}$ is a bounded sequence. Since T is graph closed, we can assume without any loss of generality that $y_k^* \to x_y^* \in T(\bar{x})$. Note here that x_y^* is dependent on the choice of y. Hence, as $k \to \infty$ from (24), we have $\langle x_y^*, \bar{x} - y \rangle \leq 0$.

Noting that y is chosen arbitrarily we conclude that for each $y \in \mathbb{R}^n$, we have

$$\inf_{x^* \in T(\bar{x})} \langle x^*, \bar{x} - y \rangle - \frac{\rho}{2} \|y - \bar{x}\|^2 \leq 0.$$

Hence we have

$$\sup_{y \in \mathbb{R}^n} \inf_{x^* \in T(\bar{x})} \left\{ \langle x^*, \bar{x} - y \rangle - \frac{\rho}{2} \|y - \bar{x}\|^2 \right\} \leq 0.$$

Observe that the function, $\langle x^*, \bar{x} - y \rangle - \frac{\rho}{2} \|y - \bar{x}\|^2$, is convex in x^* for a fixed y, and is concave in y for fixed x^*. Since $T(\bar{x})$ is compact we can again use

Sion's minimax Theorem [19] to conclude that

$$\inf_{x^* \in T(\bar{x})} \sup_{y \in \mathbb{R}^n} \left\{ \langle x^*, \bar{x} - y \rangle - \frac{\rho}{2} \|y - \bar{x}\|^2 \right\} \leq 0. \tag{25}$$

Now, by using the well known fact that the function $\frac{1}{2}\|.\|^2$ is self-conjugate we have

$$\sup_{y \in \mathbb{R}^n} \left\{ \langle x^*, \bar{x} - y \rangle - \frac{\rho}{2} \|y - \bar{x}\|^2 \right\} = \frac{1}{2\rho} \|x^*\|^2.$$

Hence from (25) we see that $\inf_{x^* \in T(\bar{x})} \frac{1}{2\rho} \|x^*\|^2 \leq 0$. Now using the compactness of $T(\bar{x})$ we conclude the existence of $\hat{x}^* \in T(\bar{x})$ such that $\frac{1}{2\rho} \|\hat{x}^*\|^2 \leq 0$. This allows us to conclude that $\hat{x^*} = 0$, and thus proving that $\bar{x}$ is a solution of the monotone inclusion. $\square$

While in the above theorem we assumed that $\operatorname{dom} T = \mathbb{R}^n$, the proof works as soon as $\operatorname{dom} T \subset \mathbb{R}^n$ is open.

Theorem 5.2 *Let T be a maximal monotone operator with nonempty values on all of $\mathbb{R}^n$. Let T be strongly monotone with $\rho > 0$, as the modulus of strong monotonicity. Let $\bar{x}$ be the unique solution of the maximal monotone inclusion $0 \in T(x)$. Then, for any $x \in \mathbb{R}^n$ we have*

$$\|x - \bar{x}\| \leq \sqrt{\frac{2}{\rho} \widehat{G}_T(x)}.$$

Proof: By the definition of $\widehat{G}_T$ we see that for each $y \in \mathbb{R}^n$ we have

$$\widehat{G}_T(x) \geq \sup_{y^* \in T(y)} \langle y^*, x - y \rangle + \frac{\rho}{2} \|x - y\|^2.$$

Now, setting $y = \bar{x}$ we have

$$\widehat{G}_T(x) \geq \sup_{y^* \in T(\bar{x})} \langle y^*, x - \bar{x} \rangle + \frac{\rho}{2} \|x - \bar{x}\|^2. \tag{26}$$

Now, as $\bar{x}$ is a solution of the monotone inclusion, we have $0 \in T(\bar{x})$. This shows that $\sup_{y^* \in T(\bar{x})} \langle y^*, x - \bar{x} \rangle \geq 0$. Hence from (26) we have $\hat{G}_T(x) \geq \frac{\rho}{2} \|x - \bar{x}\|^2$. This completes the proof. $\square$

In fact, we can modify the above theorem for the case when $\operatorname{dom} T \subset \mathbb{R}^n$. We have only to assume that T is locally bounded on $\operatorname{dom} T$ (necessarily open).

6 Related Examples

Let us now provide some examples associated with the gap function G_T and the gap γ.

Example 6.1 (Non-coercivity) Consider the convex function $f(x) = -\log x$, $x > 0$ and $f(x) = +\infty$ for $x \leq 0$. This means that $\inf_{\mathbb{R}} = -\infty$. Then, for any $x > 0$ let us set $T(x) = \partial f(x) = \{-\frac{1}{x}\}$. A simple calculation will show that $G_T(x) = 1$ for all $x \geq 0$ and $G_T(x) = +\infty$, otherwise. Hence $\gamma = 1$. Since $\gamma \neq 0$, it clearly shows that, the monotone inclusion problem has no solution, which is same as saying that $\inf_{\mathbb{R}} f = -\infty$, which indeed is true in this case. Since G_T is not finite we can conclude from here that $T = \nabla f$ is not coercive in the sense in Theorem 3.2. $\diamond$

We have shown that coercivity of T leads to finiteness of G_T. The following example shows that coercivity of T is only sufficient and not necessary.

Example 6.2 (Finiteness of G_T) Consider $T(x) = e^x$. Then, from very simple calculations we can conclude that $G_T(x) = e^{x-1}$. Thus G_T is finite, though T is not coercive in the sense used in this article. Note that the gap $\gamma_T = 0$ but is not attained. $\diamond$

Example 6.3 (Affine variational inequalities) In this example we shall consider the problem $VI(F,C)$ for $F(x) = Mx+q$, where F is a monotone map, which is equivalent to saying that M is positive semi-definite. As observed, this problem is equivalent to the monotone inclusion problem where $T(x) = F(x) + N_C(x)$.

Let us consider the case where M is skew-symmetric and thus monotone. In that case we have already shown in Section 2 that

$$G(x) = \langle q, x\rangle + \sup_{y\in C}\langle -(Mx+q), y\rangle.$$

If for example, we choose $C = \overline{\mathbb{B}}$, the unit ball in $\mathbb{R}^n$, then we have

$$G(x) = \langle q, x\rangle + \|Mx+q\|.$$

$\diamond$

Let us now provide a very simple example to illustrate that G_T can indeed be weakly coercive in the sense of (4) but not strongly so.

Example 6.4 Let us consider the function $f(x) = |x|$, $x \in \mathbb{R}$. Now consider the inclusion $0 \in \partial f(x)$. It is well known that the unique solution is $x = 0$. Thus in this case we have

$$G_T(x) = \sup_{y\in\mathbb{R}} \sup_{y^*\in\partial f(y)} y^*(x-y).$$

A simple calculation will show that $G_T(0) = 0$, $G_T(x) = x$ if $x > 0$, and $G_T(x) = -x$ if $x < 0$. It is simple to observe that G_T is only weakly coercive in the sense of (4). $\diamond$

Let us note that the gap functions computed in Examples 6.1 and Example 6.4 can also be obtained easily from Example 3.1 and 3.3 in [17]. We shall now show that we can obtain the following interesting example using Example 3.6 from [17]

Example 6.5 (Negative Entropy)

Let us the consider the convex function $f : \mathbb{R} \to \overline{\mathbb{R}}$ which is given as follows. We set $f(x) := x \ln x - x$ if $x > 0$, $f(x) = 0$ if $x = 0$, and $f(x) = +\infty$ if $x < 0$. This function has a unique minimizer at $x = 1$. Let us consider $T = \partial f$. Now, using Example 3.6 in [17] we have $G_T(x) = +\infty$ if $x < 0$, $G_T(x) = e^{-1}$ if $x = 0$, and when $x > 0$ we have

$$G_T(x) = x(W(xe) + \frac{1}{W(xe)} - 2), \tag{27}$$

where W is the inverse of the real function $x \mapsto xe^x$ and is known as the *Lambert W function*. Thence, we have $W(x)e^{W(x)} = x$. This shows that $W(e) = 1$, and so from (27) we see that $G_T(1) = 0$, showing that $x = 1$ is indeed the minimizer. Note also that for any $x > 0$ we have $G_T(x) > 0$, thus establishing that $x = 1$ is indeed the unique minimizer. $\diamond$

Example 6.6 (Computation of $\widehat{G}_T$) It is not easy to compute the regularised gap function $\widehat{G}_T(x)$. However, let us see to what extent we can simplify the computation when we consider a strongly convex function given as $f(x) := g(x) + \frac{\rho}{2}\|x\|^2$, where $g : \mathbb{R}^n \to \mathbb{R}$ is a convex function and $\rho > 0$. Now ∂f is ρ-strongly monotone and $\partial f(x) = \partial g(x) + \rho x$. Hence we have

$$\widehat{G}_{\partial f}(x) = \sup_{y \in \mathbb{R}^n} \left\{ \sup_{y^* \in \partial g(y)} \langle y^* + \rho y, x - y \rangle + \frac{\rho}{2}\|y - x\|^2 \right\}.$$

This reduces to the following formula:

$$\widehat{G}_{\partial f}(x) = \sup_{y \in \mathbb{R}^n} \left\{ g'(y, x-y) + \rho\langle x, y\rangle - \rho\|y\|^2 + \frac{\rho}{2}\|y-x\|^2 \right\}.$$

Consider the case where $g(x) := |x|$, $x \in \mathbb{R}$. Then, $x = 0$ is the minimizer of f over $\mathbb{R}$. Thence

$$\hat{G}_{\partial f}(0) = \sup_{y \in \mathbb{R}} \left\{ g'(y, -y) - \frac{\rho}{2}|y|^2 \right\} = \sup_{y \in \mathbb{R}} \left\{ -|y| - \frac{\rho}{2}|y|^2 \right\} = 0.$$

◇

Example 6.7 In this example we show that, even if we have a very simple looking structure for T which can be easily solved it may not be possible to get an explicit form of G_T, and one will have to use modern mathematical packages to get an idea of the nature of G_T. In fact, even if we know the solutions of the inclusions, it may be even difficult to exactly compute the value of G_T at the solution points and show it to be zero. However mathematical packages can be used to get a good insight into the computation of the gap function. For example let us $f(x) = -1/x$ if $x < 0$, and $f(x) = +\infty$ if $x \geq 0$, and look for solutions of $0 \in T(x) := \partial f(x) - z$, where $z > 0$. It is simple to note that, for any $z > 0$, we have $0 \in T(x)$, if and only if $x = -\sqrt{\frac{1}{z}}$. Given $z > 0$, the function G_T is given as

$$G_T(x) = \sup_{y<0} \left(\frac{1}{y^2} - z \right)(x - y).$$

Now set

$$\varphi_{x,z}(y) = \left(\frac{1}{y^2} - z \right)(x - y).$$

In fact, the natural instinct is to compute the derivative of $\varphi_{x,z}$ in terms of y, and equate it to zero. However that would result in a cubic equation which only complicates the matter. So we considered particular values of z i.e. $z = 2$ and $z = 3$ at the solution points, and tried to compute approximately $G_T\left(-\sqrt{\frac{1}{2}}\right)$ and $G_T\left(-\sqrt{\frac{1}{3}}\right)$ and also draw the graph of $\varphi_{x,z}$ with $x = -\sqrt{\frac{1}{2}}, -\sqrt{\frac{1}{3}}$ with $z = 2, 3$. We observed that $(\frac{1}{y^2}-2)(-\sqrt{\frac{1}{2}}-y) < 0$ for all $y < 0$, and the function value approaches zero asymptotically. This shows that $G_T\left(-\sqrt{\frac{1}{2}}\right) = 0$. This is same when $z = 3$ and $x = -\sqrt{\frac{1}{3}}$. However using MATLAB (version R2014) we have computed approximate values of $G_T\left(-\sqrt{\frac{1}{2}}\right)$, and $G_T\left(-\sqrt{\frac{1}{3}}\right)$, which we present below. First we present for the case $z = 2$, and then $z = 3$.

For $z = 2$ we have,

$$G_T\left(-\sqrt{\frac{1}{2}}\right) = -5.8609 \times 10^{-11} (\text{obtained in} \quad 10^5 \quad \text{iterates}),$$

and

$$G_T\left(-\sqrt{\frac{1}{2}}\right) = -2.0022 \times 10^{-15} (\text{obtained in} \quad 10^7 \quad \text{iterates}).$$

Now for $z = 3$ we have,

$$G_T\left(-\sqrt{\frac{1}{3}}\right) = -7.5306 \times 10^{-13} (\text{obtained in} \quad 10^5 \quad \text{iterates}),$$

and

$$G_T\left(-\sqrt{\frac{1}{3}}\right) = -9.8652 \times 10^{-15} (\text{obtained in} \quad 10^7 \quad \text{iterates}).$$

So we see that as we increase the number of iterations the approximate values that we obtain for the gap function at the solution points continue to increase towards zero. $\diamond$

Finally, we note that the construction in [20] can be used to show that the gap γ in Proposition 3.7 may be finite and positive.

Appendix

In this appendix we first prove that the function

$$G(x) = \sup_{y \in C} \langle F(y), x - y \rangle,$$

is a gap function when F is monotone.

We establish this directly. The function G is convex, proper and lower-semicontinuous. When C is compact, G is continuous since it is finite. It is easy to see that $G(x) \geq 0$. If $G(x) = 0$, then we have, for all $y \in C$, $\langle F(y), x - y \rangle \leq 0$. Consider a fixed $y \in C$, and construct the sequence $y_n = x + \frac{1}{n}(y - x)$. Since C is convex, $y_n \in C$. Hence we have $\langle F(y_n), x - y_n \rangle \leq 0$. As $n \to \infty$, using the continuity of F, we have $\langle F(x), x - y \rangle \leq 0$. Thus x is a solution of $VI(F, C)$ since $y \in C$ was chosen arbitrarily.

Now assume that x is a solution of $VI(F, C)$. Hence $\langle F(x), x - y \rangle \leq 0$, for all $y \in C$. Then, using monotonicity of F we have $\langle F(y), x - y \rangle \leq 0$, for all $y \in C$. This shows that $G(x) \leq 0$. Hence $G(x) = 0$, and proves that G is a gap function when F is monotone.

We shall now state the Proposition 4.1 of Aussel and Dutta [8] adapted to our setting.

Proposition A.1 *Assume that the set-valued map T is compact-valued. The the function $\hat{g}$ is a gap function for the weak variational inequality.*

We end the Appendix by stating the Sion's minimax theorem as given in Komiya [19] but will present it only in our finite dimensional setting. We recall that a *quasi-convex function* is a function, whose lower level sets are always convex. If h is quasi-convex, then $-h$ is *quasi-concave*. Thus, every convex function is quasi-convex, and every concave function is quasi-concave. For more details on quasi-convex functions see, for example, [21].

Theorem A.1 (Sion's minimax theorem)
Let X be a compact convex set of $\mathbb{R}^n$ and Y be a convex set of $\mathbb{R}^m$ and let f be a real-valued function on $X \times Y$ such that

i) $f(x,.)$ is quasi-concave and upper-semicontinuous on Y for each $x \in X$;

ii) $f(.,y)$ is quasi-convex and lower-semicontinuous on X for each $y \in Y$.

Then, we have

$$\inf_{x \in X} \sup_{y \in Y} f(x,y) = \sup_{y \in Y} \inf_{x \in X} f(x,y).$$

Acknowledgements We are thankful to the anonymous referees for their constructive suggestions which has improved the presentation of the paper and also bringing to our notice the references [12] and [17]. We would also like to thank Poonam Kesarwani for her help with the MATLAB.

References

1. Fukushima, M. : Equivalent differentiable optimization and descent methods for asymmetric variational inequalities, Mathematical Programming. 53, 99-110 (1992)

2. Facchinei . F ; Pang, J. -S. : Finite Dimensional Variational Inequalities and Complementarity Problems. Vol 1 and Vol 2, Springer (2003)
3. Rockafellar, R. T. : Monotone operators and the proximal point algorithm, SIAM Journal of Control Optimization. 14, 877-898 (1976)
4. Borwein, J. M. ; Vanderwerff, J. D. : Convex Functions : Constructions, Characterizations and Counterexamples. Cambridge University Press, (2010)
5. Borwein J. M. ; Lewis, A. S. Convex Analysis and Nonlinear Optimization : Theory and Examples. Springer, (2000) (2nd Edition, 2006)
6. Fitzpatrick, S. : Representing monotone operators by convex functions, *Workshop/Miniconference on Functional Analysis and Optimization (Canberra 1988)*, Proc. Center Math. Anal. Austral. Nat. Univ 20, Austral. Nat. Univ. Canberra, 1988, 59-65.
7. Crouzeix, J. -P. : Pseudomonotone variational inequality problems: Existence of solutions, Mathematical Programming. 78, 305-314 (1997)
8. Aussel, D. and Dutta, J. : On gap functions for multivalued Stampacchia variational inequalities. Journal of Optimization Theory and Applications. 149, 513-527 (2011)
9. Borwein, J. M. : Generalized linear complementarily problems treated without fixed-point theory. Journal of Optimization Theory and Applications. 43, 343-356 (1984)
10. Frank, M.; Wolfe, P. : An algorithm for quadratic programming. Naval Research Logistics Quarterly. 3, 95-110 (1956).
11. Alves, M. M. ; Svaiter,B. F. : Bronsted-Rockafellar property and the maximality of monotone operators representable by convex functions in non-relexive Banach spaces. Journal of Convex Analysis, 15, 693-706 (2008)
12. Monteiro, R. D. C. ; Svaiter, B. F. : On the complexity of the hybrid extragradient method for the iterates and ergodic mean, SIAM Journal on Optimization. 20, 2755-2787 (2010)
13. Dontchev, A. L. ; Rockafellar, R. T. : Implicit Functions and Solution Mappings: A View from Variational Analysis. Springer Monographs in Mathematics. (2009)
14. Borwein, J. M. ; Zhuang, D. M. : Verifiable necessary and sufficient conditions for openness and regularity of set-valued and single-valued maps, J. Math. Analysis and Applications. 134, 441-459 (1992)

15. Borwein, J. M. : ‘Minimal cuscos and subgradients of Lipschitz functions. In: J-B. Baillon and M. Thera (eds.) : Fixed Point Theory and its Applications, Pitman Lecture Notes in Mathematics. Longman, Essex (1991)

16. Aragon Atacho , F. J. ; Geoffroy, M. H. : Characterization of metric regularity of subdifferentials. Preprint (2014).

17. Bauschke, H. H.; McLaren, D. A. : Sendov, H. S. : Fitzpatrick functions, inequalities, examples and remarks on a problem by S. Fitzpatrick. Journal of Convex Analysis. 13, 499-523 (2006)

18. Nesterov, Y. ; Scrimali, L. : Solving strongly monotone variational and quasi-variational inequalities. Discrete Contin. Dyn. Syst. 31, 1383-1396 (2011)

19. Komiya, H. : Elementary proof of Sion’s minimax theorem. Kodai Mathematics Journal, 11, 5-7(1988)

20. Shapiro, A. ; Nemirovski, A. : Duality of linear conic problems, preprint, Optimization-online (2003)

21. Bazaara, M. S. ; Sherali, H. D. ; Shetty, C. M. : Nonlinear Programming: Theory and Algorithms, 3rd Edition, Wiley, (2006)