Maximal monotone inclusions and Fitzpatrick functions

Dedicated to the memory of Simon Fitzpatrick (1952–2004)

Jonathan M. Borwein (CARMA, Newcastle)

Joydeep Dutta (IIT, Kanpur)

Revised 04-02-15

Paper at: www.carma.newcastle.edu.au/jon/inclusion.pdf

February 4, 2015

CARMA

Prepared for ANZMC, December 8-12, 2014
Revised for SPCOM 2015, February 8–12, 2015

Dutta and Borwein (2013)

SPCOM 2015

South Pacific Continuous Optimization Meeting

8 — 12 February, 2015
Adelaide, South Australia

Sponsored by

University of South Australia | Centre for Industrial and Applied Mathematics

CARMA

AMSI Australian Mathematical Sciences Institute

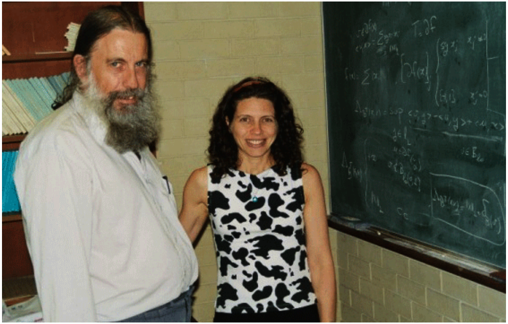
UNC Université de la Nouvelle-Calédonie

AustMS Australian Mathematical Society Inc.

In Memoriam

In his '23' "*Mathematische Probleme*" lecture to the Paris ICM in 1900*, David Hilbert wrote

"Besides it is an error to believe that rigor in the proof is the enemy of simplicity."



Simon Fitzpatrick† (1953–2004).

*See Ben Yandell's fine account of the *Hilbert Problems* and their solvers in *The Honors Class*, AK Peters, 2002. (He also died young in 2004.)

†At his blackboard with Regina Burachik

Abstract. We study maximal monotone inclusions from the perspective of (convex) *gap functions*.

We propose a very natural gap function and will demonstrate how this function arises from the *Fitzpatrick function* — a convex function used effectively to represent maximal monotone operators.

- This approach 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 generalised variational inequality problem.
    - The associated notion of a *scalar gap* is also considered.
    - Corresponding local and global error bounds are developed for the maximal monotone inclusion.

1 Introduction and Motivation

1.1 Monotone inclusions

We consider a set-valued map $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ which is maximal monotone. Recall that a set-valued map $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be *monotone* if 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

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

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

In this talk (article) we focus on the following well-studied problem [5]:

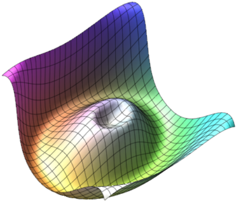
Given a set-valued map $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ which is maximal monotone the *monotone inclusion problem* requests a point $x \in \mathbb{R}^n$ such that

$$0 \in T(x). \tag{1}$$

It is clear that

$$T^{-1}(0) := \{x \in \mathbb{R}^n : 0 \in T(x)\}$$

is the solution set, which may be empty, of our inclusion problem (1).



1.2 Variational inequalities

We are interested in two special cases. First, we consider the case that

$$T(x) := S(x) + N_C(x)$$

for each $x \in \mathbb{R}^n$, where $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone and N_C the normal cone map associated with the closed convex set C; we recall that the *normal cone map* $N_C : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is given by

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

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

- Thus $\operatorname{dom} T = C \cap \operatorname{dom} S$. We assume (without much loss) that C has a non-empty interior, and that $\operatorname{dom} S \cap \operatorname{int} C \neq \emptyset$ so that $S + N_C$ is maximal monotone [5].

Since $T = S + N_C$ 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 is often referred to as the *generalized variational inequality problem* determined by S and C, denoted by $GVI(S, C)$.

- When $S := \partial f$ for $f : \mathbb{R}^n \to \mathbb{R}$ a 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 $0 \in T(x)$ as $GVI(T, \mathbb{R}^n)$.

The second problem consists of the further specialisation

$$T(x) := F(x) + N_C(x)$$

for all $x \in \mathbb{R}^n$, where $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and monotone (hence maximal) and C as before is a closed convex set. Since $\text{dom}T = C$ and $\text{dom}F = \mathbb{R}^n$ it is clear that $T = F + N_C$ is maximal [5].

Thus, this inclusion problem 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 the *variational inequality problem* determined by F and C denoted by $VI(F, C)$.

- If $C = \mathbb{R}^n_+$ then the problem reduces to the *non-linear complementarity problem* (NCP(F)) where one wishes to find $x \in \mathbb{R}^n$ such that

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

If $C = \mathbb{R}^n$ then the variational inequality problem reduces to the problem of solving equations i.e. finding an $x \in \mathbb{R}^n$ such that $F(x) = 0$.

- For more details on variational inequalities see, for example, the two volumes of Facchinei and Pang [10] ; and for monotone operators we refer to [6, Chapter 9].

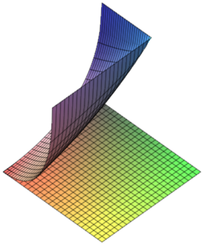
1.3 Our goals

Gap functions have played a fundamental role in the study of variational inequalities (see for example Fukushima [11] and Facchinei and Pang [10]). They allow us to:

- reformulate a (VI) as an optimization problem;
- and design error bounds for certain classes of (VI) .

Though there is a large literature regarding the monotone inclusion problem, most of is geared towards developing algorithms. One of the earliest such papers is due to Rockafellar [14].

- To our knowledge there has been no broad qualitative study of monotone inclusions from the perspective of gap functions.



Gap for -1/y

- More surprisingly, we show that the appropriate gap function for a monotone inclusion is derived from the Fitzpatrick function.
- We will also see the pivotal role played by the strong Fitzpatrick inequality [6]) in understanding aspects of the inclusion problem.
- We provide limiting examples to illustrate our results (most of which extend to reflexive Banach space).

2 Gap Functions

A *(convex) gap function* associated with the maximal monotone inclusion (1) is a function $\varphi := \varphi_T$ is a (convex) function such that

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

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

- We will show that a convex gap function can be constructed from the celebrated Fitzpatrick function, [5, Ch. 8] and [6, Ch. 8].

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

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

An immediate property is that for any maximal monotone T we have

$$F_T(x, x^*) \geq \langle x^*, x \rangle,$$

with equality holding if and only if $(x, x^*) \in \text{ graph } T$.

- In particular, $F_T(x, 0) \geq 0$ while $F_T(x, 0) = 0$ iff $0 \in T(x)$. Thus, $x \mapsto F_T(x, 0)$ is indeed a gap function for our monotone inclusion.

Let us set $G_T(x) := F_T(x, 0)$. Then explicitly

$$G_T(x) = \sup_{y \in \operatorname{dom} T} \sup_{y^* \in T(y)} \langle y^*, x - y \rangle. \tag{2}$$

Moreover, G_T is clearly a closed convex function.

- $F_T - \langle \cdot, \cdot \rangle$ is always separately convex but not often jointly convex and produces the smallest translation invariant convex gap function.

Remark 2.1 (Finitization of G_T)**.** Without much loss we can assume G_T is finite-valued. This is achieved as follows. Following Crouzeix [7] we consider $G_{\widehat{T}}$ instead of G_T where we define $\widehat{T}$ as follows

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

- $\widehat{T}(y)$ is bounded for all $y \in \mathbb{R}^n$ as $||z^*|| \le 1$ for any $z^* \in \widehat{T}(y)$. Further $G_{\widehat{T}}$ is a gap function for the pseudo-monotone inclusion $0 \in \widehat{T}(x)$.

The solution set coincides with that of the original inclusion $0 \in T(x)$. Thus $G_{\widehat{T}}$ is a gap function for the monotone inclusion $0 \in T(x)$. $\Diamond$

It is natural to ask when G_T is finite-valued without recourse to Remark 2.1. The most natural assumption on T is its coercivity: Given $x \in \mathbb{R}^n$ the operator T is *(strongly) coercive* at x if

$$\liminf_{||y|| \to \infty, y^* \in T(y)} \frac{\langle y^*, y - x \rangle}{||y||^2} > 0. \tag{3}$$

If $T(x)$ is single-valued and continuous on its domain then we have

$$\inf_{y^* \in T(y)} \langle y^*, y - x \rangle \geq q_x(y) \tag{4}$$

for some convex quadratic term $q_x(y) := c_x \|y\|^2 - b_x$ with $b_x, c_x > 0$; we call T *lower quadratic* at x.

- Clearly if T is lower quadratic at x, $G_T(x)$ is finite.

We obtain the following proposition.

Proposition 2.1 (Finiteness of G_T)**.** 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) T is coercive for all $x \in \mathbb{R}^n$, and is bounded on bounded sets.

iii) T is coercive for all $x \in \mathbb{R}^n$, and is locally bounded on its domain.

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 $\operatorname{dom} T$.

Corollary 2.1. Consider the monotone inclusion problem $0 \in T(x)$ where T is maximal monotone with $\operatorname{dom} T = \mathbb{R}^n$ and suppose T is everywhere coercive. Then G_T is finite-valued, convex and continuous.

Proof: It is well known that a maximal monotone operator is locally bounded on the interior of its domain [5, §8.2, Exercise 16]. In this case it is locally bounded over $\mathbb{R}^n$. Thus, if it is coercive, we conclude G_T is finite and the rest follows. □

Example 2.1 (Gap functions for $VI(F,C)$ or $GVI(S,C)$)**.** In particular, under our hypotheses, for $VI(F,C)$ or $GVI(S,C)$ the *variational 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, for $VI(F,C)$ the following is a convex gap function:

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

If we set $T := F + N_C$ then $VI(F,C)$ is the monotone inclusion $0 \in T(x)$. Without loss we assume $\operatorname{int} C \neq \emptyset$ — otherwise we may use relative interior. Then T is a maximal monotone operator. $\Diamond$

Proposition 2.2. If $T = F + N_C$, then for each $x \in C$, $G_T(x) = G(x)$.

We next examine the gap function, g, for $GVI(S,C)$ 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 $g(x) = G_S(x)$.

Proposition 2.3 (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 convex-valued, locally bounded, graph closed, monotone map on C.

Remark 2.2. This result was already proved in Crouzeix [7] under similar assumptions. Our proof (via Minty's (VI)) is completely different and relies essentially on the use of gap functions.

If $C = \mathbb{R}^n$ and S is maximal then g is a gap function without any additional assumptions on S. In this case $g = G_S$. $\Diamond$

The scalar gap. Define

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

This scalar value $\gamma = \gamma_T$ is called the *gap associated with the gap function* G_T. We have the following existence theorem.

Theorem 2.1. If the 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

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

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 $\gamma = 0$ then G_T is proper and lower-semicontinuous and since G_T is weakly coercive, 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 inclusion. $\square$

2.1 Complementarity problems

A *cone complementarity problem* is the special case of $VI(F, C)$ where $C = K$ is a closed convex cone and consists of 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*

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

We begin with $K^* = K := \mathbb{R}^n_+$ and with $F(x) := Mx + q$, where M is a positive semidefinite $n \times n$ matrix — but need not be symmetric. This yields the *linear complementarity problem* denoted by $LCP(M, q)$. Following Borwein [4] define the gap associated with $LCP(M, q)$ by

$$\gamma(q) := \inf\{\langle Mx + q, x\rangle : Mx + q \geq_K 0, x \geq_K 0\}.$$

Proposition 2.4. 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: **1.** The optimization problem which defines the gap is a convex quadratic problem with linear 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 and using the Frank-Wolfe Theorem we conclude there exists a minimizer. In other words

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

2. We show that $\gamma(q) = 0$. The *Lagrangian* is given by

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

Since $\gamma(q)$ is the infimal value, by separation or subgradient arguments, there exists $\bar{\lambda} \in \mathbb{R}^n_+$ such that

$$L(x, \bar{\lambda}) \geq \gamma(q), \forall x \in \mathbb{R}^n. \tag{7}$$

Since $\bar{\lambda} \in \mathbb{R}^n_+ = \mathbb{R}^{n\,*}_+$, we may set $x := \bar{\lambda}$ in (7) and see $\gamma(q) \leq 0$. This shows $\gamma(q) = 0$. Having established that $\gamma(q) = 0$ it is simple to show

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

This establishes the result. □

Remark 2.3 (Asymmetry)**.** We emphasize we have not assumed M to be symmetric. We can write $M = S + A$ where S is the symmetric part of M and A is the skew-symmetric part. If M is semidefinite then we have $\langle x, Sx\rangle \geq 0$ for all x since $\langle x, Ax\rangle = 0$ for all x. In important cases $F(x) = Mx + q$ is be monotone without M being symmetric. $\Diamond$

Such is the case of *abstract (conic) linear programming*. Consider the following pair of primal-dual linear programming problems:

$$\min\langle c, x\rangle \quad \text{subject to} \quad Ax \geq b, \quad x \geq 0, \tag{8}$$

and

$$\max\langle b, y\rangle \quad \text{subject to} \quad A^T y \leq c, \quad y \geq 0. \tag{9}$$

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 inequalities are component-wise.

From [5, Ch. 8] it follows that primal and dual solvability of the above primal-dual pair of linear programming problems is equivalent to the variational inequality $VI(F(x,y), \mathbb{R}^n_+ \times \mathbb{R}^m_+)$, where

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

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

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

and $q := (c, -b)^T$.

Note M is 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 monotone but $M \neq 0$ is never symmetric.

The variational inequality and primal-dual pair are equivalent in that the solution set of $VI(F(x,y), \mathbb{R}^n_+ \times \mathbb{R}^m_+)$ coincides with the combined primal-dual solution set. For a general version for conic programming, see [5, Thm 8.3.13].

- Even though the matrix above is skew symmetric the operator is monotone. This tempts us to consider the nature of G_T for the equation $0 = Sx + q$ where S is a skew-symmetric matrix.

Proposition 2.5 (Gap functions in the skew case)**.** Consider the problem $Sx + q = 0$ where S is a $n \times n$ skew-symmetric matrix and $q \in \mathbb{R}^n$. Then the following hold:

i) If $\bar{x}$ is a solution of $Sx + q = 0$, then $G_T(\bar{x}) = \langle q, \bar{x} \rangle = 0$.

ii) If x is not a solution of $Sx + q = 0$, then $G_T(x) = +\infty$.

Remark 2.4. Consider the consequences for the variational inequality $VI(F, C)$, where $F(x) = Sx + q$ and S is skew-symmetric. We have

$$G(x) = \langle q, x \rangle + \sup_{y \in C} \langle -(Sx + q), y \rangle. \tag{10}$$

If x is a solution of the $VI(F, C)$ we have $G(x) = 0$. If x is not a solution of $VI(Sx + q, C)$ the value $G(x)$ depends on the set C. $\Diamond$

Proposition 2.6. Consider the variational inequality associated the pair of primal-dual linear programs as above. Then we have

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

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

To conclude this section, consider the cone complementarity problem where $F(x) := Mx + q$ but K is any closed convex cone. This is the *generalized linear complementarity problem* (GLCP) [4]. Thus, we have the problem: find

$$x \in K, Mx + q \in K^*, \langle x, Mx + q \rangle = 0 \tag{11}$$

The associated gap problem as given in [4] is as follows,

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

Proposition 2.7 ((GLCP) [4])**.** Consider the complementarity problem of (11) Assume that K is a closed and convex pointed cone so K^* has nonempty interior. Suppose the Slater condition holds, in that there exists $x \in K$ such that $Mx + q \in \operatorname{int} K^*$. Then $\gamma(q) = 0$.

3 Strong Fitzpatrick Inequality and Existence of Solutions

We focus on existence of solutions for the maximal monotone inclusion. We also define and study approximate solutions.

- It is useful to compare *Celina's approximate selection* theorem for cuscos [5].

3.1 Exact solutions to inclusions

The main vehicle is two deep and recent results from the theory of maximal monotone operators (see Thm 9.7.2 and Cor. 9.7.3 in [6][1]). They are a subtle consequence of Fenchel duality.

We combine these results, which hold for *all* maximal monotone operators in reflexive Banach space, in the following theorem.

[1]As discussed in [6], the constant $1/4$ is not best possible; $1/2$ is.

Theorem 3.1 (Strong Fitzpatrick inequality)**.** Let $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a maximal monotone operator then

$$F_T(x, x^*) - \langle x, x^* \rangle \geq \frac{1}{4} d^2_{\mathrm{gph}\,(T)}(x, x^*). \tag{12}$$

Moreover

$$d^2_{\mathrm{gph}\,T}(x, x^*) \geq \max\{d^2_{\mathrm{dom}\,(T)}(x), d^2_{\mathrm{ran}\,(T)}(x^*)\}. \tag{13}$$

- The first inequality above is the strong Fitzpatrick inequality.
- With no additional hypothesis imposed on T, we always have

  $$\sqrt{2G_T(x)} \geq d_{\mathrm{gph}\,(T)}(x, 0), \tag{14}$$

  when T is maximal monotone.
  (Translating T yields the gap function for $q \in T(x)$.)

An almost immediate application of the above result is:

Theorem 3.2. Suppose T is maximal monotone, is coercive in the sense of (3), and is locally bounded in its domain. Then there is $q \in \mathbb{R}^n$ such that $\|q\| \leq 2\sqrt{G_T(0)}$ so that inclusion $0 \in T(x) - q$ has a solution.

3.2 Approximate solutions to inclusions

It is rarely easy to get the exact solutions to inclusions.

Given $\varepsilon > 0$ we say x is an *ε-approximate solution* of the maximal monotone inclusion if there exists $y^* \in T(x)$ with $\|y^*\| < \varepsilon$. The *associated gap problem* seeks a minimizer for the problem

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

Thus it is reasonable to call $x \in \mathbb{R}^n$ an *ε-approximate solution to the gap problem* if $G_T(x) < \varepsilon$. The following result connects approximate solutions of the inclusion and its associated gap problem.

Theorem 3.3 (Approximate solutions)**.** Let $\varepsilon > 0$ be given. If z is an $\dfrac{\varepsilon}{2}$-approximate solution of the gap problem, there exists y with $\|y - z\| < \sqrt{\varepsilon}$ such that y is an $\sqrt{\varepsilon}$-approximate solution of the inclusion problem $0 \in T(x)$.

Proof: Using Theorem 3.1 for the given $\varepsilon > 0$ we obtain existence of $(y, y^*) \in \operatorname{gph}(T)$ such that

$$\|y^*\|^2 + \|z - y\|^2 \leq G_T(z) + \frac{\varepsilon}{2}$$

As z is a $\dfrac{\varepsilon}{2}$-approximate minimizer of the gap function, $G_T(z) < \dfrac{\varepsilon}{2}$.

From the above inequality 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. $\square$.

- Theorem 3.3 is a *variational principle for maximal monotone inclusions*: if one as an approximate minimizer of the gap problem then there is a nearby approximate solution of the inclusion.

The following is an obvious corollary.

Corollary 3.1. If the gap problem has $\gamma = 0$, then for any $\varepsilon > 0$ there is a $\sqrt{\varepsilon}$-approximate solution to the inclusion problem $0 \in T(x)$.

4 Error Bounds for a Monotone Inclusion

4.1 Metric regularity and local error bounds

We begin by assuming the maximal monotone map is T to be metrically regular in an appropriate sense.

We say the maximal monotone mapping T is *metrically regular at* $(\bar{x}, \bar{y}) \in \operatorname{gph} T$ if there exist numbers $k > 0$, $\delta > 0$, and $\gamma > 0$ such that

$$d_{T^{-1}(y)}(x) \leq k d_{T(x)}(y) \quad \forall x \in B_\delta(\bar{x}) \quad \text{and} \quad y \in B_\gamma(\bar{y}), \tag{15}$$

and T is metrically regular over the graph if T is metrically-regular for every $(\bar{x}, \bar{y}) \in \operatorname{gph} T$.

- Metric regularity is by itself a kind of error bound which can be tuned in our setting to develop a local error bound.
- Sadly, even subdifferentials of simple convex functions can fail to be metrically regular.

Nonetheless, (16) implies that

$$d_{T^{-1}(0)}(x) \leq k d_{T(x)}(0) \qquad \forall x \in B_\delta(\bar{x}) \quad \text{and} \quad y \in B_\gamma(0). \tag{16}$$

- That said, as described in [8], for a monotone operator, metric reqularity will force the mapping to be single valued and indeed strongly monotone as discussed in Section 4.3.

Without any regularity assumption, (12) implies for all x in $B_\delta(\bar{x})$ that

$$\sqrt{G_T(x)} \geq \frac{1}{2} d_{\text{gph}\,(T)}(x, 0). \tag{17}$$

Putting (16) and (17) together we deduce that when T is metrically regular

$$4k\sqrt{G_T(x)} \geq d_{T^{-1}(0)}(x), \tag{18}$$

since we have $y \in T(z)$ with $\max\{\|x-z\|, \|y\|\} \leq \sqrt{G_T(x)}$. The inequality (18) follows by noting that $x \in \operatorname{dom} T$, so we can set $z = x$.

4.2 The convex case

The previous discussion motivates the need to exploit weaker notions such as metric subregularity even for the subdifferential of a convex function.

We say the subdifferential map ∂f of a convex function $f : \mathbb{R}^n \to \mathbb{R}$ is *metric-subregularity at a point* $(\bar{x}, \bar{y}) \in \operatorname{gph}(\partial f)$ if there exist neighbourhoods U and V of $\bar{x}$ and $\bar{y}$ respectively and $k > 0$ such that

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

Then ∂f *metrically subregular* if it is metrically subregular at each $(\bar{x}, \bar{y}) \in \operatorname{gph} \partial f$.

- Note that $f(x) := |x|$, $x \in \mathbb{R}$ is indeed metrically subregular. This leads to the following simple consequence of [1, Thm 3.3].

Proposition 4.1. Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and let S the denote the set of all global minimizers of f. Assume S is non-empty and that ∂f is metrically sub-regular.

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 $c_x > 0$ such that

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

- By contrast, we can exploit Theorem 3.1 for any proper lower semicontinuous convex function f, as soon as $\mu := \inf f$ is finite.

We begin by checking that for all x, x^*, in terms of the Fenchel conjugate

$$F_{\partial f}(x, x^*) \leq f(x) + f^*(x^*),$$

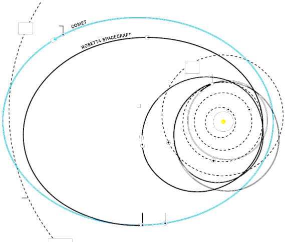
Thus, when $\mu = -f^*(0)$ is finite, we derive

$$G_{\partial f}(x) \leq f(x) - \mu. \tag{19}$$

Hence

$$\sqrt{G_{\partial f}(x)} \leq \sqrt{f(x) - \mu} \tag{20}$$

and also $\gamma = 0$.



Rosetta orbit

4.3 Error bounds in the strongly monotone case

Finally, we present a new gap function for the maximal monotone inclusion when $T:\mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *strongly monotone*. That is, 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*. This gap function is based on regularization of the gap function G_T and extends an approach of Nesterov and Scrimali [13].

We define the function $\widehat{G}_T$ as follows

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

We begin with the following result.

Proposition 4.2. If T is strongly monotone with modulus ρ then the function $\widehat{G}_T$ is a finite-valued, strongly convex and continuous function. Under natural conditions $\widehat{G}_T$ is a gap for the associated inclusion.

Theorem 4.1. Let T be strongly monotone with non-empty compact-values throughout its domain. Suppose T is locally bounded and graph closed. Then $\widehat{G}_T$ is a gap function for the monotone inclusion $0 \in T(x)$.

This leads to:

Theorem 4.2. Let T be strongly monotone and let $\bar{x}$ be the unique solution of the monotone inclusion $0 \in T(x)$. Further assume that T is nonempty compact-valued, locally bounded and graph closed. Then for any $x \in \mathbb{R}^n$ we have

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

5 Related Examples

We provide some examples associated with G_T and the gap γ.

Example 5.1 (Non-coercivity)**.** Consider the convex function

$$f(x) = -\log x, \qquad x > 0.$$

Then for $x > 0$ let us set

$$T(x) := \partial f(x) = \left\{ -\frac{1}{x} \right\}.$$

Now $G_T(x) = 1$ for all $x \geq 0$ and $G_T(x) = +\infty$, otherwise. Hence $\gamma = 1$. Since $\gamma \neq 0$ the inclusion problem has no solution. That is, $\inf_{\mathbb{R}} f = +\infty$, which indeed is true. Since G_T is not finite we conclude that $T = \nabla f$ is not coercive in the sense of (3). $\Diamond$

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 5.2 (Finiteness of G_T)**.** Consider $T(x) := e^x$. Simple calculations show that $G_T(x) = e^{x-1}$. Thus G_T is finite even though T is not coercive in the sense needed in this work. Also note that the gap $\gamma_T = 0$ but is not attained. $\Diamond$

Example 5.3 (Affine variational inequalities)**.** Consider the problem $VI(F, C)$ for $F(x) = Mx+q$ where F is a monotone map. As observed, this is equivalent to the monotone inclusion problem where

$$T(x) := F(x) + N_C(x).$$

We shall try to compute the gap function $G(x)$ under various assumptions on M and C.

First consider the case where M is skew-symmetric and thus monotone, so we have shown in Section 2 that

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

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

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

When M is symmetric and positive definite it was shown in [**?**] that

$$G(x) = \langle My(x) + q, x - y(x)\rangle,$$

where

$$y(x) := p_{M,C}\left(\frac{1}{2}(x - M^{-1}q)\right).$$

In the above expression $p_{M,C}$ is the oblique projection (see [10]) on C with respect to the matrix M. When M is only positive semi- definite it becomes difficult to provide an explicit expression for G. In these cases we can always define G_T to be equal to G on C and $+\infty$ otherwise. $\Diamond$

Let us show G_T can be weakly coercive without being lower quadratic.

Example 5.4. Let us again consider $f(x) := |x|$, $x \in \mathbb{R}$ and the inclusion $0 \in T(x) := \partial f(x)$. The unique solution is $x = 0$. Thus

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

Now $G_T(0) = 0$, $G_T(x) = x$, if $x > 0$ and $G_T(x) = -x$ if $x < 0$. Thus we see that G_T is a coercive function in the sense that $\liminf_{|x| \to +\infty} G_T(x)/|x| \to 1$. $\Diamond$

Example 5.5 (Computation of $\widehat{G}_T$). It is hard to compute the regularised gap function $\widehat{G}_T(x)$. We can simplify the computation when we consider a strongly convex function given as $f(x) := g(x) + \frac{\rho}{2}\|x\|^2$, where $\rho > 0$. Now ∇f is strongly monotone with modulus $\frac{\rho}{2}$ since $\partial f(x) = \partial g(x) + x$. Hence we have

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

This reduces:

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

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

$$\hat{G}_T(0) = \sup_{y \in \mathbb{R}^n} \left\{ -|y| - \frac{\rho}{2}|y|^2 \right\} = 0.$$

and so on. ◊

- It is also possible to explicitly compute the gap function G_T in various other interesting cases. For example, if we start again with $f(y) = -\log y, y > 0$ and look for solutions of $0 \in T(x) := \partial f(x) - z$ we arrive, for $z < 0$, at the gap function

$$G_T(x) = F_T(x, z) - \langle x, z \rangle = (1 - \sqrt{-zx})^2.$$

- Finally, we note that the construction in [12] can be used to show that the gap γ in Proposition 2.7 may be finite and positive.

References

[1] F. J. Aragon Atacho and M. H. Geoffroy, Characterization of metric regularity of subdifferentials, Preprint 2014.

[2] D. Aussel and J. Dutta, On gap functions for multivalued Stampacchia variational inequalities, *Journal of Optimization Theory and Applications*, Vol 149, (2011), pp 513–527.

[3] C. Berge, *Topological Spaces: Including a Treatment of Multi-valued Functions, Vector Spaces and Convexity*, Translated from the French original by E. M. Paterson, Reprint of 1963 translation, Dover Publications, NY, 1997.

[4] J. M. Borwein, Generalized linear complementarily problems treated without fixed-point theory, *Journal of Optimization Theory and Applications*, Vol 43, (1984), pp 343–356.

[5] J. M. Borwein and A. Lewis, *Convex Analysis and Nonlinear Optimization : Theory and Examples*, Springer, 2000. (2nd Edition, 2006).

[6] J. M. Borwein and J. D. Vanderwerff, *Convex Functions : Constructions, Characterizations and Counterexamples*, Cambridge University Press, 2010.

[7] J.-P. Crouzeix, Pseudomonotone variational inequality problems: Existence of solutions, *Mathematical Programming*, Vol 78 (1997), pp 305–314.

[8] A. L. Dontchev and R. T. Rockafellar, *Implicit Functions and Solution Mappings: A View from Variational Analysis*, Springer Monographs in Mathematics, 2009.

[9] J. Dutta, Gap functions and error bounds for variational and generalized variational inequalities, *Vietnam Journal of Mathematics*, Vol 40, (2012), pp 231–253.

[10] F. Facchinei and J.-S. Pang, *Finite Dimensional Variational Inequalities and Complementarity Problems*, Vol 1 and Vol 2, Springer 2003.

[11] M. Fukushima, Equivalent differentiable optimization and descent methods for asymmetric variational inequalities, *Mathematical Programming*, Ser. A, Vol 53, (1992), pp 99–110.

[12] A. Shapiro and A. Nemirowskii, Duality of linear conic problems, Preprint 2003.

[13] Y. Nesterov and L. Scrimali. Solving strongly monotone variational and quasi-variational inequalities. *Discrete Contin. Dyn. Syst.* Vol 31 (2011), pp 1383–1396.

[14] R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM Journal of Control Optimization* Vol 14 (1976), pp 877–898.