# When is a Gap Function Good for Error Bounds?

Joydeep Dutta [*]

**Abstract**

It is well known that gap functions play a major role in developing error bounds for strongly monotone and affine monotone variational inequalities. However the error bound would be important from the point of view of computation if the gap function values diminishes on sequence of points which converge to a solution of the variational inequality. A gap function exhibiting such behavior is termed as a well-behaved gap function. In this article we would like to consider some important gap functions for both variational inequalities with single-valued maps and set-valued maps and see what additional conditions one must impose on them so that they are well behaved.

## 1 Introduction and Motivation

Variational inequalities in finite dimensions generalize the necessary and sufficient optimality condition for minimizing a differentiable convex function over a convex set. More formally if $F : \mathbb{R}^n \to \mathbb{R}^n$ and let $C$ be a convex set then the variational inequality problem $VI(F, C)$ consists of finding $\bar{x}$ such that

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

It is simple to observe that if $F(x) = \nabla f(x)$ for all $x \in \mathbb{R}^n$, where $f$ is a differentiable convex function then the expression in the variational inequality is nothing but the necessary and sufficient condition for minimizing the convex function $f$ over the convex set $C$. Note that we had considered $f$ to be differentiable. More generally if we consider $f$ to be non-differentiable then the the generalization of the necessary and sufficient optimality condition leads to the so-called *generalized variational inequality* or GVI for short. Given the a set-valued map $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and convex set $C$ the generalized variational inequality $GVI(T, C)$ consists of finding $\bar{x} \in C$ and $\xi \in T(\bar{x})$ such that

$$\langle \xi, x - \bar{x} \rangle \geq 0 \quad \forall x \in C.$$

It is simple to observe that if $T = \partial f$ where $\partial f$ represents the subdifferential multifunction of the convex function $f$ then the above GVI is nothing but the non-smooth necessary and sufficient optimality condition for minimizing a the convex function over the convex set $C$. However in the set-valued setting we have a flexibility and we can define what we call a weak GVI denoted as $WGVI(T, C)$, where in one needs top find $\bar{x} \in C$ such that for each $x \in C$ there exists $\xi_x \in T(\bar{x})$ ( depending on $x$) such that

$$\langle \xi_x, x - \bar{x} \rangle \geq 0 \quad \forall x \in C.$$

The above type of variational inequalities has been referred to in the literature as the *Stampacchia type variational inequalities.* For more details on the problem $VI(F, C)$, see for example Facchinei and Pang [9] and for recent results on $GVI(T, C)$ and $WGVI(T, C)$ see Aussel and Dutta [4] and the references there in. However there is another class of variational inequalities known as the *Minty type variational equalities.* Given a function

[*]Dept. of Mathematics Statistics, Indian Institute of Technology, Kanpur-208016, India

$F : \mathbb{R}^n \to \mathbb{R}^n$ and a convex set $C$ then Minty variational inequality problem $MVI(F, C)$ is the problem of finding $\bar{x} \in C$ such that

$$\langle F(x), x - \bar{x} \rangle \geq 0 \quad \forall x \in C$$

An interesting feature is that the solution set of $MVI(F, C)$ is always convex while that of $VI(F, C)$ need not be so. Further when $F$ is monotone (or pseudomonotone) the solution sets of these two types of variational inequalities coincide. In our study here we will be essentially interested in the class of variational inequalities where $F$ is monotone or $T$ is monotone as set-valued map.
There are various ways to approach a variational inequality. One of the principal ways is to devise an optimization problem whose solution would lead to the solution of the original variational inequality. This is done by introducing the notion of a *merit function* or *gap function*. Given any of the above variational inequalities a gap function is a function $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ which satisfies the following two properties

i) $\psi(x) \geq 0$ for all $x \in C$

ii)$\psi(\bar{x}) = 0$, $x \in$ if and only if $\bar{x}$ solves the variational inequality.

Thus it is clear that if we minimize the function $\psi$ over the convex set $C$ and if the $\bar{x}$ is the minimum with $\psi(\bar{x}) = 0$ then $\bar{x}$ is a solution of the variational inequality. However as one might observe that it is not so simple to minimize the function $\psi$ since it could be not only be non-convex but every value of the function $\psi$ is evaluated by solving an optimization problem. Thus it becomes quiet cumbersome to minimize the function $\psi$. However it has been observed that gap functions play a fundamental role for developing error bounds for certain classes of monotone variational inequalities. An *error bound* is an expression which provides an upper estimate of the distance of an arbitrary feasible point to the solution set of the variational inequality. For more details see for example Pang [14] and Facchinei and Pang [9]. More precisely let $S$ denote the solution set of a variational inequality and $x \in C$ then we an error bound is given by the following expression

$$d(x, S) \leq k\beta(\psi(x)),$$

where $\beta$ is a real-valued function on $\mathbb{R}$ such that $\beta(0) = 0$ and $k > 0$ is a constant. In fact all the error bounds appearing in this article will be of the form

$$d(x, S) \leq k\sqrt{\psi(x)}.$$

In order for the above error bounds to be meaningful it is important to note that first of all gap function must be finite valued and then when $x$ is indeed very near the solution set then the gap function value at $x$ must also be very near zero. More precisely if there is a sequence $\{x^k\}$ with $x^k \in C$ such that $x^k \to \bar{x}$ where $\bar{x}$ is a solution then one should expect $\psi(x^k) \to 0$ as $k \to \infty$ or may in the more weaker sense $\liminf_{k\to\infty} \psi(x^k) \to 0$. Thus in some sense we are expecting the function $\psi$ to be continuous or at least lower semicontinuous. In this article we try to focus on classes of gap functions both for VI and GVI for which the above properties hold. Among these gap functions we focus on the *primal generalized gap function* due to Auchmuty [1]. Though Auchmuty's primal regularized gap function is not so well studied in the literature we will show that it displays very interesting properties and is truly an useful gap function. We will show that the primal generalized gap function can be finite under some mild conditions without any boundedness assumption on the set $C$. Further we will show that some important class of gap function like the Fukushima's regularized gap function [10] are just special cases of the primal generalized gap functions. Thus the primal gap function is a unifying one and thus can play a vital role in the study of gap functions and error bounds.

The article is arranged as follows. In section 2 we discuss the Auslender gap function which is probably one of the most popular gap function in the literature on variational inequalities. In this section we will also discuss hemivariational inequalities and the dual gap function.. In section 3 we will discuss the primal generalized gap function due Auchmuty [1] and we will show that very well behaved gap functions like the Fukushima's regularized gap function [10] are a special case of Auchmuty's construction. In section 4 we will concentrate on the gap function and its Fukushima type regularization for GVI and WGVI. In fact using a simple observation and the gap function developed in [4] for WGVI to develop a gap function for GVI and also regularize it along the lines of Fukushima [10] and show that it is indeed a useful gap function to devise an error for GVI and we shall also present the error bound. We will also show that a gap function developed in [4] specifically for GVI also has the desirable properties needed to devise an error bound using such a gap function. As we will see that these will be much more complicated that the usual VI.

We would also like to stress that this article is essentially pedagogical in nature. To put it more directly in this paper we intend to survey some important class of gap functions and show under what conditions they posses the propeties as discussed above in order to create error bounds. For example we will discuss the gap functions of Auslender [3], Auchmuty [1] and Fukushima [10]. Further we will also present some error bounds with proofs using the regularized gap function of Fukushima [10] which has not being explicitly mentioned in the literature though error bounds with more generalized regularized gap functions have appeared in the literature.
We shall end this section by mentioning the definitions of monotonicity and strong monotonicity which will be used through out the sequel.
A function $F : \mathbb{R}^n \to \mathbb{R}^n$ is said to be monotone if for any $x, y \in \mathbb{R}^n$,

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

The function $F$ is called strongly monotone with strong monotonicity parameter $\mu > 0$ if for any $x, y \in \mathbb{R}^n$

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

An important example of a strongly monotone operator is $F(x) = Mx + q$ where $M$ is a $n \times n$ is a symmetric positive definite matrix with $\mu = \lambda_{\min}(M)$ where $\lambda_{\min}(M)$ denotes the minimum eigenvalue of $A$. We shall usually refer a strongly monotone map with strong monotonicity parameter $\mu > 0$ as a $\mu$-strongly monotone function.

Let $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be monotone if for any $x, y \in \mathbb{R}^n$ and for any $x^* \in T(x)$ and $y^* \in T(y)$ we have

$$\langle y^* - x^*, y - x \rangle \geq 0.$$

and $T$ is said to be $\mu$-strongly monotone if for any $x, y \in \mathbb{R}^n$ and any $x^* \in T(x)$ and $y^* \in T(y)$ we have

$$\langle y^* - x^*, y - x\rangle \geq \mu||y - x||^2.$$

## 2 Auslender Gap Function

In the literature on variational inequalities the gap function it is usually mentioned that the gap function referred to Auslender appeared in his book on numerical optimization [2]. However to the best of our knowledge it seems that it first appeared in [3]
The Auslender gap functions denoted by $\theta$ is defined as

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

First of all let us note that in order to calculate the value of $\theta$ at a given value $x$ we need to minimize an affine function in $y$ over the convex set $C$. First of all it is important to see that $\theta$ is an extended valued function. Of course we need additional properties on $C$ so that$\theta$ becomes finite. One such property is that $C$ is compact. If $F$ is a continuous function then it is quiet simple to see that $\theta$ is a lower-semicontinuous function. Thus at the very least we need to have $F$ to be continuous. The following result due to Marcotte [12], whose proof we provide for completeness, tells under what condition the function $\theta$ can be Lipschitz over the set $C$.

**Theorem 2.1.** *Let $C$ be a compact convex set and $F$ be a continuously differentiable map. Then $\theta$ is Lipschitz continuous on $C$, i.e. there exits a constant $l \geq 0$ such that*

$$|\theta(x) - \theta(z)| \leq l||x - z||.$$

*Proof:* Let us set $f(x, y) = \langle F(x), x - y\rangle$. It is clear that $f$ is continuous as a function in the variables $x$ and $y$. Further since $F$ is continuously differentiable it is Lipschitz on $C$. Thus $f(., y)$ is Lipschitz on $C$ for each $y \in C$. To see this observe that we have

$$\nabla_x f(x, y) = F(x) + \nabla F(x)^T x - \nabla F(x)^T y,$$

where $\nabla F(x)$ is the Jacobian of the map $F$ at $x$ and $T$ denotes the transpose of the matrix. Thus for each fixed $y \in C$ the function $x \mapsto \nabla_x f(x, y)$ is continuous and thus $f(., y)$ is Lipschitz over $C$. Thus for any $x_1, x_2 \in$ , $x_1 \neq x_2$ and given $y \in C$ there exists $l(y) > 0$ such that

$$|f(x_1, y) - f(x_2, y)| < l(y)||x_1 - x_2||.$$

Now using the continuity of $f$ in the variables $(x, y)$, we have that there exists $\delta_y > 0$ such that for all $z \in B_{\delta_y}(y)$ we have

$$|f(x_1, z) - f(x_2, z)| < l(y)||x_1 - x_2||.$$

Further $\bigcup_{y\in C} B_{\delta_y}(y)$ forms a cover of $C$. Since $C$ is a compact set we have that there exists a finite subcover of $C$ i.e. there exists $y_i$, $i = 1, \ldots, m$, such that $\bigcup_{i=1}^m B_{\delta_{y_i}}(y_i)$ is a cover of $C$. Now for any $y \in C$ there exists $r \in \{1, \ldots, m\}$, such that $y \in B_{\delta_{y_r}}(y_r)$. Thus we have

$$|f(x_1, y_i) - f(x_2, y_i)| < l(y_r)||x_1 - x_2|$$

Put $l = \max\{l(y_1), \ldots, l(y_m)\}$. This shows that for all $y \in C$

$$|f(x_1, y) - f(x_2, y)| < l||x_1 - x_2||.$$

Thus we have

$$f(x_1, y) \leq f(x_2, y) + l||x_1 - x_2||.$$

This leads the fact

$$\theta(x_1) \leq \theta(x_2) + l||x_1 - x_2||.$$

The result is established by swapping the roles of $x_1$ and $x_2$. □

An important example for such an $F$ is the affine function $F(x) = Mx + q$ where $M$ is a $ntimesn$ symmetric matrix and $q \in \mathbb{R}^n$. Further when $F = Mx + q$ and $M$ is symmetric and positive semidefinite then it is easy to see that $\theta$ is a convex function and when $C$ is compact the subdifferential of $\theta$ can be computed very easily using the Danskin's formula. When $C$ is compact and $F$ is $\mu$-strongly monotone then it is simple to show that if $\bar{x}$ be the unique solution of $VI(F, C)$ and if $x \in C$ then

$$||x - \bar{x}|| \leq \sqrt{\frac{\theta(x)}{\mu}}$$

Unfortunately even if $C$ is polyhedral there is no way to guarantee that *theta* will be finite. The most well known case is the one where $C = \mathbb{R}^n_+$. Then have the famous *nonlinear complementarity problem* in which we have to find $x \in \mathbb{R}^n_+$ such that $F(x) \in \mathbb{R}^n_+$ and $\langle F(x), x \rangle = 0$. In this case the function $\theta(x)$ is given as follows. If $F(x) \in \mathbb{R}^n_+$ then $\theta(x) = \langle x, F(x) \rangle$ and if $F(x) \notin \mathbb{R}^n_+$ then $\theta(x) = +\infty$.
There is another class of variational inequalities which is called is usually called *Hemivariational Inequalities*(HVI for short) is also studied in the literature. Given $F : \mathbb{R}^n \to \mathbb{R}^n$ and a convex function $f : \mathbb{R}^n \to \mathbb{R}$ the problem $HVI(F, f, C)$ consists in finding $\bar{x}$ such that

$$\langle F(\bar{x}), y - \bar{x} \rangle + f(y) - f(\bar{x}) \geq 0, \quad \forall y \in C.$$

Note that if $f$ is a differentiable convex function then $HVI(F, f, C)$ is equivalent to $VI(F + \nabla f, C)$. If $f$ is not differentiable then $H(F, f, C)$ would correspond to a $GVI(F + \partial f, C)$. This means that if $\bar{x}$ solves $HVI(F, f, C)$ then there exists $\xi \in \partial f(\bar{x})$ such that

$$\langle F(\bar{x}) + \xi, y - \bar{x} \rangle \geq 0, \quad \forall y \in C.$$

Of course if $\bar{x}$ solves $GVI(F(\bar{x}) + \partial f, C)$ then it of course solves $HVI(F, f, C)$. The above facts will be crucial in finding the error bounds. In Chen, Goh and Yang [8] the function $f$ is considered to be an extended valued proper lower-semicontinuous convex function. They call such an HVI an *extended variational inequality.* Thus if we just know that $f$ is convex in our present set up we need to define its gap function in a different way. As proposed by Chen, Goh and Yang [8] the gap function for $HVI(F, f, C)$ is defined as follows

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

Note that that just like the Auslender gap function the in general it is not possible to say whether the above function $\phi$ is finite or not unless $C$ has some additional properties like compactness. However if $f$ is a strongly convex then $\phi$ is finite without any boundedness assumption on the set $C$. In fact observe that one can write

$$\phi(x) = -\inf\{\langle F(x), y - x\rangle + f(y) - f(x)\}.$$

Note that for a fixed $x$ the function

$$y \to \langle F(x), y - x\rangle + f(y) - f(x)$$

is a strongly convex function and thus attains a unique minimum on the closed convex $C$. However unless we have a particular form of $f$ in the strongly convex case it is not possible to explicitly compute this minimum. Of course it is simple enough to show that $\phi$ is a gap function for $HVI(F, f, C)$. Note that if $\bar{x}$ is a solution of $HVI(F, f, C)$ then we have

$$\langle F(\bar{x}), x - \bar{x}\rangle \geq f(\bar{x}) - f(x), \quad \forall x \in C.$$

This will show us two things. If $x_0$ is a solution of $HVI(F, f, C)$ and is also the maximum of the convex function $f$ over $C$ then $x_0$ also solves $VI(F, C)$. Further if $x_0$ is a solution of $VI(F, C)$ and $x_0$ is a minimum of the convex function $f$ over the convex set $C$ then $x_0$ is also a solution of the problem $HVI(F, f, C)$. Let us now present a result on error bound for $HVI(F, f, c)$. Let us note that if $F$ is strongly monotone or $f$ is strongly convex $HVI(F, f, C)$ has a unique solution. This can be seen easily by viewed by considering the equivalence of $HVI(F, f, C)$ with $GVI(F + \partial f, C)$.

**Theorem 2.2.** *Let us consider $HVI(F, f, C)$ and let $\bar{x}$ be the unique solution of $HVI(F, f, C)$, then we have the following*

*i) Let $C$ be compact and $F$ be $\mu$-strongly monotone then for any $x \in C$*

$$||x - \bar{x}|| \leq \sqrt{\frac{\phi(x)}{\mu}}$$

*ii) Let $F$ be monotone and $f$ be $\rho$-strongly convex ( $\rho > 0$ ) then*

$$||x - \bar{x}|| \leq \sqrt{\frac{\phi(x)}{\rho}}$$

**Proof :** Let us consider the case i) where we assume that $F$ is $\mu$-strongly monotone. Now from the definition of the gap function $\phi$ we have for all $y \in C$

$$\phi(x) \geq \langle F(x), x - y\rangle + f(x) - f(y).$$

Now in particular for $y = \bar{x}$ we have

$$\phi(x) \geq \langle F(x), x - x^*\rangle + f(x) - f(x^*).$$

Now using the strong monotonicity of $F$ and the convexity of $f$ we have for any $\xi \in \partial f(\bar{x})$ we have

$$\phi(x) \geq \langle F(\bar{x}), x - \bar{x}\rangle + \mu||x - \bar{x}||^2 + \langle \xi, x - \bar{x}\rangle.$$

Since $\bar{x}$ is the unique solution of $HVI(F, f, C)$ it is also a solution $GVI(F + \partial f, C)$ there exists $\bar{\xi} \in \partial f(\bar{x})$ such that

$$\langle F(\bar{x}) + \bar{\xi}, x - \bar{x} \rangle \geq 0.$$

Hence we have that

$$\phi(x) \geq \mu ||x - \bar{x}||^2$$

. This establishes the error bound in the first case. In the second case we have assumed that $f$ is strongly convex. Hence we know from Vial [17] that for all $\xi \in \partial f(\bar{x})$

$$f(x) - f(\bar{x}) \geq \langle \xi, x - \bar{x} \rangle + \rho ||x - \bar{x}||^2.$$

Using this fact and the monotonicity of $F$ we conclude that for all $\xi \in \partial f(\bar{x})$ we have

$$\phi(x) \geq \langle F(\bar{x}), x - \bar{x} \rangle + \langle \xi, x - \bar{x} \rangle + \rho ||x - \bar{x}||^2$$

The rest of the proof follows as the first case by noting that $\bar{x}$ also solves $GVI(F + \partial f, C)$. This completes the proof. □

It is important to note that we have considered $f$ to be a finite valued convex function on $\mathbb{R}^n$ while Chen, Goh and Yang [8] and Patriksson [15] considers $f$ to be an extended-valued proper lower-semicontinous convex function. However it is important to note that when $f$ is proper and lower-semicontinuous and if $\bar{x}$ is the solution then necessarily $f(\bar{x})$ has to be finite. Thus in without loss of generality we can always consider $f$ to be finite. This also makes sense from the computational point of view. Further when $f$ is extended-valued, proper and lower-semicontinuous convex function and $\text{int}(dom f) \cap C \neq \emptyset$ it has been shown in Patriksson [15] $HVI(F, f, C)$ is equivalent to $GVI(F + \partial f, C)$. Further it is important to note that even if $\bar{x} \in dom f$ there is no guarantee that $\partial f(\bar{x}) \neq \emptyset$ unless $\bar{x} \in \text{int} dom f$. Note that if $\partial f(\bar{x})$ is empty then the inequality associated with $GVI(F + \partial f, C)$ has no meaning. Further in order to develop an error bound for $GVI(F + \partial f, C)$ the approach taken by Patriksson [15] closely follows the line of treatment of gap functions by Auchmuty [1]. On the other hand in Chen, Goh and Yang [8] under the assumption that $F$ $\mu$-strongly pseudomonotone with respect to $f$ and some further additional assumptions an error bound is developed for $HVI(F, f, C)$. See [8] for details. However as we have seen our proof of the error bound for $HVI(F, f, C)$ is based on very simple and natural assumptions.

It is important to note that unless $F(x) = Mx + q$ the function $\theta$ need not be convex. The natural question is whether we can have a gap function for $VI(F, C)$ which is always convex under natural assumptions one of which is say $F$ being monotone. The answer surprisingly turns out to be "yes". We shall denote such a gap function by $\theta_D$ and define it as

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

It is interesting to note that $\theta_D$ is always convex and is a gap function for $VI(F, C)$ when $F$ is monotone. Note that even when $F(x) = Mx + q$, with $M$ symmetric and positive definite it is not possible to guarantee that the Auslender gap is finite unless $C$ is compact.

One just to take the case $C = \mathbb{R}^n_+$ to see this. While if $F(x) = Mx + q$ with $M$ symmetric and positive definite then we will show that $\theta_D$ is always finite irrespective of whether $C$ is compact or not. Note that when $F(x) = Mx + q$ we have

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

Note that the function $y \mapsto \langle y, My\rangle + \langle q, y\rangle - \langle Mx, y\rangle - \langle q, x\rangle$ is strongly convex and is we assume that $C$ is closed it has a unique minimum over $C$. Let us call this as $y(x)$. Thus in this case

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

Thus $\theta_D$ is finite. It will be shown in the next section that when $F$ is strongly monotone then $\theta_D$ is finite. However it is important to note that when $C$ is compact then $\theta_D$ can be used to compute the solution of $VI(F, C)$ with $F$ monotone. Note that in this case we have

$$\partial\theta_D(x) = co\{F(y(x)) : y(x) \in J(x)\},$$

where $J(x) = \text{argmax}_{y \in C}\langle F(y), x - y\rangle$. Note that as always we assume that $F$ is continuous. If $F(x) = Mx + q$ and $M$ is symmetric and positive definite and $C$ is compact then $\theta_D$ is differentiable since $J(x)$ is singleton and we have $\nabla\theta_D(x) = F(y(x))$ where $J(x) = \{y(x)\}$. Thus it seems that that for monotone or strongly monotone problems the dual gap function is a useful device to compute the solution of $VI(F, C)$. The natural question is that whether this gap function is also amenable in devising error bounds when $F$ is strongly monotone. It appears that it is quiet a difficult matter to devise error bounds with the dual gap function. A detailed study of the dual gap function was carried out in the general case of a GVI in Aussel and Dutta [5] where by modifying the dual gap function suitably they developed error bounds in combination with other gap functions but with sharper constraints. Further very recent paper which is under preparation Aussel and Dutta show that the dual gap function can provide an error bound in the case when $F(x) = Mx + q$, with $M$ is symmetric and positive definite without any additional assumption on $C$. In the same paper some additional modifications have been carried out on the dual gap function and the role of the modified dual gap function in devising error bounds is also explored.

# 3 The Primal Generalized Gap Function

At this point we take a detour and have a look at Auchmuty's approach to gap functions. Though Auchmuty had and elegant approach using convex analysis his approach is not so well documented in the literature. The gap function introduced by Auchmuty [1] is denoted as $\theta_A$ as is given as

$$\theta_A(x) = \langle F(x), x\rangle + \sigma_C(-F(x),$$

where $\sigma_C$ denotes the support function of the set $C$. Thus note that

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

This shows that $\theta_A(x) = \theta(x)$. Of course Auchmuty did not stop at this and went onto develop a more general approach to the notion of a gap function. Auchmuty [1] the notion of a primal generalized gap function which is given as

$$G(x) = \sup_{y \in C} \hat{L}(x, y),$$

where $L(x, y)$ is given as follows

$$L(x, y) = f(x) - f(y) - \langle x - y, \nabla f(x) - F(x) \rangle,$$

where $f$ is a proper lower-semicontinuous convex function which is smooth on an open set containing $C$. We can thus also assume that $f$ is finite and smooth on $\mathbb{R}^n$.It is not much difficult to prove that

$$G(x) = f(x) + g^*(\nabla f(x) - F(x)) + \langle x, F(x) - \nabla f(x) \rangle.$$

Note that in the above expression $g^*$ denotes the Fenchel conjugate of $g$ where $g = f + \delta_C$ where $\delta_C$ is the indicator function of the convex set $C$. This can be seen by observing that

$$G(x) = \sup_{y \in \mathbb{R}^n} \{ \langle \nabla f(x) - F(x), y \rangle - (f(y) + \delta_C(y)) \} + f(x) + \langle F(x) - \nabla f(x), x \rangle.$$

Let us now list down some important properties of the primal generalized gap function $G$ which is stated [11] and [1]. We provide the proof for completeness

**Theorem 3.1.** *The function $G$ is lower-semicontinuous and $G$ is a gap function for $VI(F, C)$. Further if $f$ be finite valued, differentiable and strongly convex on $\mathbb{R}^n$ then $G$ is finite and continuous.*

Proof : Since $f$ is smooth on an open set containing $C$ the function $L(x, y)$ is a continuous function of $y$ on $C$ and that shows that $G$ is lower-semicontinuous being the supremum of a family of continuous functions. We will now show that $G$ is finite and continuous when $f$ is finite valued and differentiable strongly convex function. Note that we can write

$$G(x) = -\inf_{y \in C}(-L(x, y)) = -\inf_{y \in C}(f(y) - f(x) + \langle F(x) - \nabla f(x), y - x \rangle).$$

Note that for each fixed $x$ the function $y \mapsto f(y) - f(x) + \langle F(x) - \nabla f(x), y - x \rangle$ is a strongly convex function. Since the set $C$ is closed it has a unique minimum $y(x)$ ( depending on $x$ ). Thus $G$ is finite valued. Further it is well known that $y(x)$ is continuous. So we have

$$G(x) = f(x) - f(y(x)) + \langle F(x) - \nabla f(x), x - y(x) \rangle.$$

Since $y(x)$ is continuous function of $x$ and $f$ is smooth this shows that $G$ is finite valued and continuous.
We will show that $G$ is a gap function. Putting $x = y$ we see that $L(x, y) \geq 0$ and thus $G(x) \geq 0$ for all $x \in C$. Now let us assume that $G(\bar{x}) = 0$ and $\bar{x} \in C$. Hence we have for all $y \in C$,

$$f(\bar{x}) - f(y) + \langle F(\bar{x}) - \nabla f(\bar{x}), \bar{x} - y \rangle \leq 0.$$

Consider any arbitatry $x \in C$ fixed for the time being. Then $\bar{x} + \lambda(x - \bar{x}) \in C$ for any $\lambda \in (0, 1)$. Hence we have for any $\lambda \in (0, 1)$,

$$f(\bar{x}) - f(\bar{x} + \lambda(x - \bar{x}) + \langle F(\bar{x}) - \nabla f(\bar{x}), \bar{x} - \bar{x} + \lambda(x - \bar{x}) \rangle \leq 0.$$

Using the fact that $f$ is smooth on an open set containing $C$ by Taylor's expansion we have

$$\langle F(\bar{x}) - \nabla f(\bar{x}), -\lambda(x - \bar{x})\rangle + \langle \nabla f(\bar{x} + \lambda(x - \bar{x}), -\lambda(x - \bar{x})\rangle + o(\lambda) \leq 0$$

Dividing by $\lambda$ and passing to the limit as $\lambda \downarrow 0$ and noting that $f$ is smooth on an open set containing $C$ we have

$$\langle F(\bar{x}), \bar{x} - x\rangle \leq 0.$$

Since $x \in C$ was arbitrary we see that $\bar{x}$ is a solution to $VI(F, C)$.
Conversely let us assume that $\bar{x}$ solves $VI(F, C)$ This shows that

$$\langle F(\bar{x}), \bar{x} - y\rangle \leq 0, \quad \forall y \in C.$$

Since $f$ is convex and smooth on an open set containing $C$ we have that

$$f(\bar{x}) - f(y) - \langle f(\bar{x}), \bar{x} - y\rangle \leq 0 \quad \forall y \in C,$$

This clearly shows that $L(\bar{x}, y) \leq 0$ for all $y \in C$ thus proving that $G(\bar{x}) \leq 0$. Since $G(\bar{x}) \geq 0$ it is clear that $G(\bar{x}) = 0$. This completes the proof that $G$ is a gap function for $VI(F, C)$. □

It is clear that if $\theta$ is a constant function then $G(x) = \theta(x)$ for all $x \in \mathbb{R}^n$. However as we have seen that when $f$ is strongly convex then $G$ is finite and continuous irrespective of the fact whether the set $C$ is bounded or not. This is one of the fundamental reason for regularizing the Auslender gap function with a strongly convex function. The notion of a regularized gap function due to Fukushima [10] in fact popularized the concept among optimization researchers. However as we will now see that Fukushima's regularized gap function [10] is a special case of the primal generalized gap function of Auchmuty [1]. Let us consider the case when $f(x) = \frac{1}{2}||x||^2$, $x \in \mathbb{R}^n$. Then we have

$$L(x, y) = \frac{1}{2}||x||^2 - \frac{1}{2}||y||^2 + \langle F(x) - x, x - y\rangle.$$

Some simple calculations will show that

$$L(x, y) = \langle F(x), x - y\rangle - \frac{1}{2}||y - x||^2.$$

This shows that $G(x) = g(x)$ which is a simpler version of the regularized gap function of Fukushima [10]. In fact we write

$$g(x) = \sup_{y \in C}\{\langle F(x), x - y\rangle - \frac{1}{2}||y - x||^2\}.$$

Of course since our chosen $f(x) = \frac{1}{2}||x||^2$ is strongly convex we conclude from Thoerem 3.1 that $g$ is finite-valued function which is continuous. In fact it can be shown that

$$g(x) = \langle F(x), x - \text{proj}_C(x - F(x))\rangle - \frac{1}{2}||\text{proj}_C(x - F(x)) - x||^2.$$

Let us note that in order to get a meaningful error bound when $F$ is strongly monotone it is useful to slightly modify the definition of $g$ by using a tuning parameter $\alpha > 0$ and define

$$g_\alpha(x) = \sup_{y \in C}\{\langle F(x), x - y\rangle - \frac{\alpha}{2}||y - x||^2\}$$

Further it can be easily shown that when $f(x) = \frac{\alpha}{2}$ then $G(x) = g_\alpha(x)$. In fact it is not much difficult to show that

$$g_\alpha(x) = \langle F(x), x - \text{proj}_C(x - \frac{1}{\alpha}F(x))\rangle - \frac{1}{2}||\text{proj}_C(x - \frac{1}{\alpha}F(x)) - x||^2.$$

However keeping in view of the pedagogical let us give a brief explanation of how the above expression comes about. Let $y_\alpha(x)$ be the unique minimum of the strongly convex function

$$y \mapsto \langle F(x), y - x\rangle + \frac{\alpha}{2}||y - x||^2$$

over the closed convex set $C$. Then the standard optimality conditions for convex programming tells us that

$$-F(x) - \alpha(y_\alpha(x) - x) \in N_C(y_\alpha(x)).$$

Noting that $N_C(y_\alpha(x))$ is a cone we have

$$(x - \frac{1}{\alpha}) - y_\alpha(x) \in N_C(y_\alpha(x)).$$

This shows that

$$y_\alpha(x) = \text{proj}_C(x - \frac{1}{\alpha}(x)).$$

Note that in order to compute $y_\alpha(x)$ one needs to compute the projection which is equivalent to solving a strongly convex optimization problem. If $C = \mathbb{R}^n_+$ then the projection is quiet simple to compute and it is given by

$$y_\alpha(x) = \max\{x - \frac{1}{\alpha}F(x), 0\},$$

where the maximum is taken in the componentwise sense.
When $C$ is described in terms of convex inequalities and affine equalities then one really had to compute the projection problem using the methods laid down for solving convex optimization problems. However for example in some special cases one can give an explicit formula for $y_\alpha(x)$. For example consider

$$C = \{x \in \mathbb{R}^n : Ax = b\},$$

where $A$ is a $m \times n$ with matrix of full row rank. In this case it is not difficult to show that

$$y_\alpha(x) = (I - A^T(AA^T)^{-1}A)(x - \frac{1}{\alpha}F(x)) + A^T(AA^T)^{-1}b.$$

where $A^T$ is the transpose of the matrix $A$.
Note that if $F$ is $\mu$-strongly monotone we have and if we choose $\alpha > 0$ such that $2\mu > \alpha$ and let $\bar{x}$ be the unique solution of $VI(F, C)$, then for any $x \in C$ we have

$$||x - \bar{x}|| \leq \sqrt{\frac{2}{2\mu - \alpha}}\sqrt{g_\alpha(x)}.$$

Let us now consider the case when $f(x) = \frac{\alpha}{2}\langle x, Px\rangle$, where $P$ is a $n \times n$ symmetric and positive definite matrix. Then we have

$$G(x) = \sup_{y \in C}\{\langle F(x), x - y\rangle - \frac{\alpha}{2}\langle y - x, P(y - x)\rangle\}$$

When $\alpha = 1$ this coincides with the regularized gap function introduced by Fukushima [10]. In this special case we shall denote $G(x)$ as $g_\alpha^F(x)$. This function as before is also finite by arguments as similar as above. However it is slightly more complicated to find out the point which minimizes the function

$$y \mapsto \langle F(x), y - x\rangle + \frac{\alpha}{2}\langle y - xP(y - x)\rangle.$$

This is obtained through the operation of oblique projection. Observe that if $y_\alpha^F(x)$ is the unique solution of the above function then the optimality condition states that

$$-(F(x) + \alpha P(y_\alpha^F(x) - x)) \in N_C(y_\alpha^F(x))$$

Thus we have that there exists $v \in N_C(y_\alpha^F(x))$ such that

$$-\frac{1}{\alpha}F(x) + P(y_\alpha^F(x) - x) = v$$

Thus

$$-\frac{1}{\alpha}P^{-1}F(x) - y_\alpha^F(x) + x = P^{-1}v.$$

This shows that

$$P((x - \frac{1}{\alpha}P^{-1}F(x)) - y_\alpha^F(x)) \in N_C(y_\alpha^F(x)).$$

The above expression is the necessary and sufficient optimality condition for the following strongly convex optimization problem whose solution is $y_\alpha^F(x)$,

$$\min ||y - (x - \frac{1}{\alpha}P^{-1}F(x))||_P^2, \quad \text{subject to} \quad x \in C,$$

where $||x||_P$ is the norm with respect to the positive definite matrix $P$ , i.e. $||x||_P^2 = \langle x, Px\rangle$. The above strongly convex problem is the problem of oblique projection and we write

$$y_\alpha^F(x) = proj_{C,P}(x - \frac{1}{\alpha}G^{-1}F(x)).$$

It is amply clear that even if the oblique projection has nice properties like Lipschitz continuity it is not so simple to compute and give an explicit expression for $y_\alpha^F(x)$ even for the case when $C = \mathbb{R}_+^n$. However we will now show that for the set $C$ given as

$$C = \{x \in \mathbb{R}^n : Ax = b\},$$

where $A$ as before is a $m \times n$ matrix with full row rank and $b \in \mathbb{R}^m$ once can provide an explicit expression for $y_\alpha^F(x)$. Noting that for the above $C$ we have $N_C(y) = \mathrm{Im}A^T$ for any $y \in C$ and thus we have that there exits $\lambda \in \mathbb{R}^m$ such that

$$-\frac{1}{\alpha}P^{-1}F(x) - (y_\alpha^F(x) - x) = P^{-1}A^T\lambda.$$

This shows that

$$y_\alpha^F(x) = x - \frac{1}{\alpha}P^{-1}F(x) - P^{-1}A^T\lambda.$$

Hence we have

$$b = Ax - \frac{1}{\alpha} AP^{-1}F(x) - AP^{-1}A^T\lambda$$

Note that since $P$ is positive definite so is $P^{-1}$ and since $A$ is of full row rank it is simple to show that $AP^{-1}A^T$ is also positive definite and thus invertible. Hence we have

$$\lambda = (AP^{-1}A^T)^{-1}Ax - (AP^{-1}A^T)^{-1}b - \frac{1}{\alpha}(AP^{-1}A^T)^{-1}AP^{-1}F(x)$$

Further some simple calculations shows us that

$$y_\alpha^F(x) = (I - P^{-1}A^T(AP^{-1}A^T)^{-1}A)x + \frac{1}{\alpha}(P^{-1}A^T(AP^{-1}A^T)^{-1}AP^{-1} - P^{-1})F(x) \\ +P^{-1}A^T(AP^{-1}A^T)^{-1}b.$$

Let us now write down an error bound in terms of the gap function $g_\alpha^F(x)$ where $F$ is $\mu$-monotone and and let $\mu > \frac{\alpha}{2}||P||$. If $x^*$ is the unique solution of $VI(F,C)$ then

$$||x - x^*|| \leq \sqrt{\frac{2}{2\mu - \alpha||P||}}\sqrt{g_\alpha^F(x)}.$$

In the above result for error bounds one observes that $\mu$ and $\alpha$ are related. One can ask the question whether an error bound with $g_\alpha^F$ can be developed with out any relation between $\mu$ and $\alpha$. In this case one would have to assume that $F$ is Lipschitz over $C$. To the best of our knowledge we did not see the theorem below and its corollary in the literature though they can be derived from more general results. We provide the proofs for completeness.

**Theorem 3.2.** *Let us consider the problem $VI(F,C)$ where $F$ is $\mu$-strongly monotone and Lipschitz on $C$ with Lipschitz constant $L \geq 0$. Let $x^* \in C$ be the unique solution of $VI(F,C)$. Then for any $x \in C$ we have*

$$||x - x^*|| \leq \frac{(\alpha||P|| + L)}{\mu}\sqrt{\frac{2}{\alpha\lambda_{\min}(P)}}\sqrt{g_\alpha^F(x)}.$$

**Proof :** Let $y_\alpha^F(x)$ be the unique minimizer of the strongly convex function

$$y \mapsto \langle F(x), y - x\rangle + \frac{\alpha}{2}\langle y - x, P(y - x)\rangle.$$

Then using the standard optimality conditions for convex optimization we have

$$\langle F(x) + \alpha P(y_\alpha^F(x) - x), x^* - y_\alpha^F(x) \geq 0$$

Further as $x^*$ solves $VI(F,C)$ we have

$$\langle F(x^*), y_\alpha^F(x) - x^*\rangle \geq 0.$$

This shows that

$$\langle F(x) - F(x^*) + \alpha P(y_\alpha^F(x) - x), y_\alpha^F(x) - x^*\rangle \leq 0.$$

A little manipulation of the above inequality will lead to the following inequality

$$\langle F(x) - F(x^*), x - x^* \rangle + \langle y_\alpha^F(x) - x, P(y_\alpha^F(x) - x) \rangle \leq -\alpha \langle P(y_\alpha^F(x) - x), x - x^* \rangle$$
$$- \langle F(x) - F(x^*), y_\alpha^F(x) - x \rangle.$$

Now using the fact that $P$ is positive definite, $F$ is Lipschitz and strongly monotoem we have that

$$\mu ||x - x^*||^2 \leq \alpha ||P|| ||y_\alpha^F(x) - x|| ||x - x^*|| + L ||x - x^*|| ||y_\alpha^F(x) - x||.$$

This leads to the fact

$$||x - x^*|| \leq \frac{\alpha ||P|| + L}{\mu} ||y_\alpha^F(x) - x||. \tag{3. 1}$$

Further observe that from the optimality conditions we have

$$\langle F(x) + \alpha P(y_\alpha^F(x) - x), x - y_\alpha^F(x) \rangle \geq 0 \tag{3. 2}$$

Thus we have

$$\alpha \langle P(y_\alpha^F(x) - x), y_\alpha^F(x) - x \rangle \leq - \langle F(x), y_\alpha^F(x) - x \rangle.$$

Since $P$ is positive definite we have

$$\frac{\alpha}{2} \langle y_\alpha^F(x) - x, P(y_\alpha^F(x) - x) \rangle \geq \frac{\alpha}{2} \lambda_{\min}(P) ||y_\alpha^F(x) - x||^2.$$

This shows that

$$\alpha \langle y_\alpha^F(x) - x, P(y_\alpha^F(x) - x) \rangle - \frac{\alpha}{2} \langle y_\alpha^F(x) - x, P(y_\alpha^F(x) - x) \rangle \geq \frac{\alpha}{2} \lambda_{\min}(P) ||y_\alpha^F(x) - x||^2.$$

Now from further manipulation using (3. 2) along with the above inequality we obtain that

$$g_\alpha^F(x) \geq \frac{\alpha}{2} \lambda_{\min}(P) ||y_\alpha^F(x) - x||^2.$$

Now the result is established by using (3. 1). □

**Corollary 3.1.** *Let us consider the problem $VI(F, C)$ with $F(x) = Mx + q$, where $M$ is a $n \times n$ symmetric and positive definite matrix. Let $x^*$ be the unique solution of $VI(F, C)$. Then for any $x \in C$ we have*

$$||x - x^*|| \leq \frac{(\alpha \sqrt{trace P^2} + \sqrt{trace M^2})}{\lambda_{\min}(M)} \sqrt{\frac{2}{\alpha \lambda_{\min}(P)}} \sqrt{g_\alpha^F(x)}.$$

**Proof :** The result can be established by noting that in this case $L = ||M||$ and $\mu = \lambda_{min} M$ and the applying Theorem 3.2. Note that we use the Frobenious norm for matrix $M$ and $P$ and thus $||M||^2 = \text{trace}(M^2)$ and $||P||^2 = \text{trace} P^2$. □

It is important to note that in the above formula any matrix norm can be used since all matrix norms are equivalent and possibly the norm that we have used might be easy to compute in most case. For example if $P$ is a diagonal matrix with positive entries then the above choice of the norm possibly will turn out to be quite efficient.

Let us now consider the case where $f(x) = \alpha h(x)$ where $g : \mathbb{R}^n \to \mathbb{R}$ is a differentiable strongly convex function with $\rho > 0$ as the modulus of strong monotonicity. In this case we denote $G(x)$ as $g_\alpha^h(x)$ and we have

$$g_\alpha^h(x) = \sup_{y \in C}\{\langle F(x), x - y\rangle - \alpha D_h(y, x)\}$$

where $D_h(y, x) = h(y) - h(x) - \langle \nabla h(x), y - x\rangle$. The function $D_h(y, x)$ is called the Bregman distance induced by $h$ which is a non-Euclidean distance and which plays a pivotal role in the generalized version of the proximal point algorithm for solving convex optimization problem and variational inequalities with maximal monotone map. For more details see for example the recent monograph by Burachik and Iusem [7]. By the arguments similar to the previous cases it is clear that $g_\alpha^h$ is well defined in the sense that it is finite and continuous. Thus the above gap functions generalizes the regularized gap function due to Fukushima [10]. However the function $g_\alpha^h$ can also be defined by considering that $h$ is just convex and not strongly convex and in such a case we can just conclude that $g_\alpha^h(x)$ is an extended-valued lower-semicontinuous function. In fact when $h$ is just convex an error bound can be obtained in terms of $g_\alpha^h$ by slightly tweaking the proof of theorem of Patriksson [11] however it might be difficult to verify the condition under which it holds and also the question about its finiteness remains. Thus a better approach would be use the case when $h$ is strongly convex which will guarantee the finiteness and continuity of $g_\alpha^h$ and in the proof of the error bound one just need to use the fact that $h$ is convex rather than using the more stronger property of strong convexity. Thus we will have the following result.

**Theorem 3.3.** *Let us consider the problem $VI(F, C)$ where $F$ is $\mu$-strongly monotone and the function $h$ is strongly convex whose derivative is Lipschitz continuous of rank $L > 0$ over $C$. Let us choose $\alpha > 0$ such that $\frac{\mu}{L} > \alpha$. If $x^* \in C$ is the unique solution of $VI(F, C)$ and $x \in C$, then we have*

$$||x - x^*|| \leq \sqrt{\frac{1}{\mu - \alpha L}}\sqrt{g_\alpha^h(x)}.$$

*Further if $F(x) = Mx + q$ where $M$ is a $n \times n$ symmetric and positive definite matrix and $q \in \mathbb{R}^n$ we have*

$$||x - x^*|| \leq \sqrt{\frac{1}{\lambda_{\min}(M) - \alpha L}}\sqrt{g_\alpha^h(x)}.$$

The class of functions $g_\alpha^h$ is naturally a subclass of the following family of gap functions given as

$$\hat{g}_\alpha(x) = \sup_{y \in C}\{\langle F(x), x - y\rangle - \alpha\phi(x, y)\}.$$

This class of gap functions were studied by Wu, Florian and Marcotte [18]. In fact the above function can be shown to be a gap function to $VI(F, C)$ is the function $\phi$ satisfies the four propeties

i) $\phi(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^n$.

ii) The function $\phi$ is continuously differentiable on $\mathbb{R}^n \to \mathbb{R}^n$

iii) For each fixed $x$ the function $y \mapsto \phi(x, y)$ is strongly convex in $y$ in a uniform way in the sense that the modulus of strong convexity is independent of the choice of $x$.

iv)$\phi(x, y) = 0$ if and only $y = x$.

In fact using the above four properties of $\phi$ one can establish that $\hat{g}_\alpha$ is a gap function in a straight forward way and in fact in a more simpler fashion than that of [18] whose proof is motivated by the approach in [10]. In fact the approach that we will just outline depends on the crucial fact that $\nabla_y \phi(x, y) = 0$ if and only if $x = y$. This can be proved very simply based on the property iv). It is in fact trivial to see that $\hat{g}_\alpha(x) \geq 0$ for all $x \in C$. Let $\bar{x} \in C$ is such that $\hat{g}_\alpha(\bar{x}) = 0$. For $k \in \mathbb{N}$ set

$$y^k = \bar{x} + \frac{1}{k}(y - \bar{x}),$$

for any arbitrary ( but fixed ) $y \in C$. From the the definition of $\hat{g}_\alpha$ we have

$$\langle F(\bar{x}), \bar{x} - y \rangle \leq k\phi(\bar{x}, y^k). \tag{3. 3}$$

Now by the strong convexity of $\phi$ in the second variable, we have using that fact that $\phi(\bar{x}, \bar{x}) = 0$

$$-k\phi(\bar{x}, y^k) \geq \langle \nabla_y \phi(\bar{x}, y^k).\bar{x} - y \rangle + \frac{\rho}{k^2}|y - \bar{x}, ||^2.$$

where $\rho > 0$ is the modulus of strong convexity. Note that since $\phi(\bar{x}, y)$ is strongly convex in $y$ and also differentiable, it is also continuously differentiable and further noting that $\nabla_y \phi(\bar{x}, \bar{x}) = 0$ we have as $k \to \infty$ from the inequality

$$\lim_{k \to \infty} k\phi(\bar{x}, y^k) \leq 0.$$

Now using (3. 3) we conclude that $\bar{x}$ solves $VI(F, C)$.
Now the converse that if $\bar{x}$ solves $VI(F, C)$ then that fact $\hat{g}_\alpha(\bar{x}) = 0$ follows very simply and we avoid the proof here.
Note that if we set $\phi(x, y) = D_h(y, x)$ where $h$ is a differentiable and strongly convex then it is simple to see that $D_h(y, x)$ satisfies all the properties of $\phi(x, y)$ listed above. In fact we know by choosing specific forms of $h$ we can get different types of gap functions including the regularized version of Fukushima [10]. An important question is that is the above representation of $\phi(x, y)$ and $D_h(y, x)$ is the only possible one for which all the four properties of $\phi$ holds true.
In fact in order to derive an error bound we have to impose Lipschitz condition on $\nabla_y \phi(x, .)$ for every $x$ and we will also have to assume that the Lipschitz constant does not depend on $x$. For more details see [18].
It is important to note that in this article we do not discuss the notion of $D$-gap functions or difference gap functions which is obtained by taking the difference of the same type of regularized gap function but with two different tuning parameters. For more details see Peng [16] and Yamashita, Taji and Fukushima [19]

# 4 Gap Functions for GVI

As compared to the usual $VI(F, C)$ there has been only a few attempts to develop gap functions and error bounds for variational inequality with multivalued maps which we term here as $GVI(T, C)$ and $WGVI(T, C)$. One of the important work in this direction is due to Patriksson [15] where he had considered $T(x) = F(x) + \partial f(x)$ for all $x \in \mathbb{R}^n$ where $F$ is a single-valued function and $f$ is a convex function. In this article Patriksson had modified the primal generalized gap function and developed a gap function for $GVI(F + \partial f, C)$. Very recently Aussel and Dutta [4] developed gap functions and regularized gap function for GVI and WGVI. However in their work they did not discuss whether the gap functions that they have introduced posses the properties which makes them useful for error bounds. In their main definitions Aussel and Dutta [4] they did not consider the map $T$ is compact-valued. Instead they considered an auxiliary set-valued map which was then used to define the gap functions. However we will only analyze the situation when $T$ is compact-valued. This is primarily motivated by the fact that most important subdifferentials are compact-valued. However it is important to note for example the basic subdifferential of Mordukhovich [?] which is also known as the Mordukhovich subdifferential is compact-valued when the underlying function is locally Lipschitz but is not convex-valued. So we do not assume that $T$ is convex-valued but state this assumption whenever needed. Our aim is to use the definition of the gap function devised in [4] for $WGVI(T, C)$ to devise a gap function for $GVI(T, C)$ with out assuming that the map $T$ is convex-valued. Let us without loss of generality assume that T is non-empty-valued and let us also assume that $T$ is compact-valued. Then Aussel and Dutta [4] defined the following function

$$H(x) = \sup_{y \in C} \inf_{x^* \in T(x)} \langle x^*, x - y \rangle.$$

It is proved in [4] that $H$ is a gap function for $WGVI(T, C)$. Further it is mentioned the statement of Proposition 4.1 in [4] that if $T$ is also convex-valued then $H$ is also a gap function for $GVI(T, C)$. It was mentioned in the proof of Proposition 4.1 that by applying the Sion's minimax theorem ( see for example Berge [6]) one can switch the infimum and the supremum and that will lead to the proof that $H$ is a gap function for $GVI(T, C)$. However no details of the proof were provided in [4]. However as we will see that working out the details would turn out to be instructive and would lead to a different gap function for $GVI(T, C)$ where we would no longer use the fact that $T$ is convex-valued.

Let us now under the assumption that $T$ is also convex-valued prove the fact that it is also a gap function for $GVI(T, C)$. For a given $x$ since $T(x)$ is convex and compact valued by using the Sion's minimax theorem it is immediate that

$$H(x) = \inf_{x^* \in T(x)} \sup_{y \in C} \langle x^*, x - y \rangle.$$

It is simple to see that $H(x) \geq 0$ for all $x \in C$. Further let $\bar{x} \in C$ such that $H(\bar{x}) = 0$. Then it is clear that

$$0 = \inf_{x^* \in T(\bar{x})} \sup_{y \in C} \langle x^*, \bar{x} - y \rangle.$$

Let us set

$$\beta(x^*, \bar{x}) = \sup_{y \in C} \langle x^*, \bar{x} - y \rangle.$$

Note that as $\bar{x}$ it is clear that $\beta(x^*, \bar{x})$ is a lower-semicontinuous convex function and it is simple to see that it is also proper. Since $T(\bar{x})$ is compact there exists $\hat{x}^*$ such that $\beta(\hat{x}^*, \bar{x}) = 0$ This shows that for all $y \in C$

$$\langle \hat{x}^*, \bar{x} - y \rangle \leq 0.$$

This shows that $\bar{x}$ solves $GVI(T, C)$. Conversely if $\bar{x} \in C$ is such that $H(\bar{x}) = 0$ then it is clear that there exists $x^* \in T(\bar{x})$ such that

$$\langle x^*, \bar{x} - y \rangle \leq 0.$$

This allows us to conclude that $H(\bar{x}) \leq 0$ and thus $H(\bar{x}) = 0$.
Note that in the above proof we do not require the convexity of $T(x)$ once we have switched the infimum with the supremum. This shows that even when $T$ is just compact-valued one can devise a gap function for $GVI(T, C)$. This gap function is given as

$$\hat{H}(x) = \inf_{x^* \in T(x)} \sup_{y \in C} \langle x^*, x - y \rangle.$$

It is however important to note that the compactness of $C$ will be pivotal in guaranteeing the finiteness of $H$ or $\hat{H}$. Thus it is important to note that in order to have a gap function which remains finite even when $C$ is unbounded. This can be done by regularizing $H$ and $\hat{H}$ in the sense of Fukushima [10]. The regularization for $H$ was done in Aussel and Dutta [4] and for any $\alpha > 0$ the regularized version of $H$ was give as

$$H_\alpha(x) = \sup_{y \in C} \left\{ \inf_{x^* \in T(x)} \langle x^*, x - y \rangle - \frac{\alpha}{2} ||y - x||^2 \right\}.$$

It was shown in [4] that $H_\alpha$ was finite and is a gap function for $WGVI(T, C)$ and an error bound can also be devised when $T$ is strongly monotone. We refer the reader to [4] for details. However it can be showed that $H$ is also a lower-semicontinuous function. We will now demonstrate that a regularization of $\hat{H}$ will lead to a finite-valued gap function for $GVI(T, C)$ which is also has nice properties under mild assumptions which makes it amenable to devise error bounds. The regularized version of $\hat{H}$ is given as

$$\hat{H}_\alpha(x) = \inf_{x^* \in T(x)} \left\{ \sup_{y \in C} \left\{ \langle x^*, x - y \rangle - \frac{\alpha}{2} ||y - x||^2 \right\} \right\}.$$

To begin with we will first prove that $\hat{H}_\alpha$is a gap function.

**Theorem 4.1.** *Let us consider the problem $GVI(T, C)$ where $T$ is compact-valued. Then for any $\alpha > 0$ the function $\hat{H}_\alpha$ is finite-valued and is a gap function for $GVI(T, C)$.*

Proof : We will first show that $\hat{H}_\alpha$ is finite for any $\alpha > 0$. Observe that we can rewrite $\hat{H}_\alpha$ as $\hat{H}_\alpha(x) = \inf_{y^* \in T(x)} q_\alpha(x^*, x)$ where $q_\alpha$ given below is defined for all pairs $(x^*, x) \in \mathbb{R}^n \times \mathbb{R}^n$.

$$q_\alpha(x^*, x) = \left\{ \sup_{y \in C} \left\{ \langle x^*, x - y \rangle - \frac{\alpha}{2} ||y - x||^2 \right\} \right\}.$$

Note that $q_\alpha$ can be rewritten as

$$q_\alpha(x^*, x) = - \inf_{y \in C} \left\{ \langle x^*, y - x \rangle + \frac{\alpha}{2} ||y - x||^2 \right\}.$$

Further it is clear that as $(x^*, x)$ is fixed

$$y \to \langle x^*, y - x \rangle + \frac{\alpha}{2} ||y - x||^2$$

is strongly convex and since $C$ is a closed and convex set there exists a unique minimizer $y_\alpha(x^*, x)$ which is given as

$$y_\alpha(x^*, x) = \text{proj}_C(x - \frac{1}{\alpha}x^*)$$

Hence $q_\alpha(x^*, x)$ is finite for any pair $(x^*, x) \in \mathbb{R}^n \times \mathbb{R}^n$. Further $q_\alpha$ is continuous $(x^*, x)$ since the projection map is Lipschitz on $\mathbb{R}^n$. Since $T$ is compact-valued this clearly shows that $\hat{H}_\alpha$ is finite-valued.
Our next aim is to show that $\hat{H}_\alpha$ is a gap function. It is clear that $\hat{H}_\alpha(x) \geq 0$ for all $x \in C$. Now let $\bar{x} \in C$ be such that $\hat{H}_\alpha(\bar{x}) = 0$. Now consider an arbitrary point $x \in C$ and consider all points $y(\lambda) = \bar{x} + \lambda(x - \bar{x})$ where $\lambda \in (0, 1)$. It is clear that since $C$ is a convex set $y(\lambda) \in C$ for all $\lambda \in (0, 1)$. Now we have

$$0 = \inf_{x^* \in T(\bar{x})} q_\alpha(x^*, \bar{x}).$$

Since $T$ is compact-valued and we have shown above that $q_\alpha$ is finite and continuous it clear that there exists $w^* \in T(\bar{x})$ such that

$$0 = q_\alpha(w^*, \bar{x})$$

Thus we have that for all $y \in C$

$$\langle w^*, \bar{x} - y\rangle \leq \frac{\alpha}{2}||y - x||^2$$

In particular for $y = y(\lambda)$ with $\lambda \in (0, 1)$ we have

$$\langle w^*, \bar{x} - x\rangle \leq \frac{\lambda\alpha}{2}||x - \bar{x}||^2$$

As $\lambda \to 0$ we see that

$$\langle w^*, \bar{x} - x\rangle \leq 0.$$

Since $x$ was arbitrary we see that $\bar{x}$ is a solution of $GVI(T, C)$.
Conversely assume that $\bar{x}$ solves $GVI(T, C)$. Then there exists $p^* \in T(\bar{x})$ such that

$$\langle p^*, \bar{x} - y\rangle \leq 0 \quad \forall y \in C.$$

This shows that $q_\alpha(p^*, \bar{x}) \leq 0$ and thus proving that $\hat{H}_\alpha(\bar{x}) \leq 0$. This shows that $\hat{H}_\alpha(\bar{x}) = 0$. This completes the proof $\square$.
Our next aim is to see if an error bound can be devised for $GVI(T, C)$ in terms of $\hat{H}_\alpha$ when $T$ is $\mu$-strongly monotone. Once that is done then we see whether $\hat{H}_\alpha$ has the desired properties which would make it a suitable choice for devising error bounds.

**Theorem 4.2.** *Let us consider the problem $GVI(T, C)$ where $T$ is compact-valued and $\mu$-strongly monotone. Let $\alpha > 0$ be such that $2\mu > \alpha$ and let $\bar{x}$ be the unique solution of $GVI(T, C)$. Then for any $x \in C$ we have*

$$||x - \bar{x}|| \leq \sqrt{\frac{2}{2\mu - \alpha}}\sqrt{\hat{H}_\alpha(x)}.$$

**Proof :** Observe that for any given $x \in C$ we can find $w^* \in T(x)$ such that

$$q_\alpha(w^*, x) = \hat{H}_\alpha(x),$$

This shows that

$$\hat{H}_\alpha(x) \geq \langle w^*, x - \bar{x} \rangle - \frac{\alpha}{2}||x - \bar{x}||^2 \tag{4. 4}$$

Since $\bar{x}$ solves $GVI(T, C)$ there exists $x^* \in T(\bar{x})$ such that

$$\langle x^*, x - \bar{x} \rangle \geq 0. \tag{4. 5}$$

Since $T$ is $\mu$-strongly monotone we have

$$\langle w^*, x - \bar{x} \rangle \geq \langle x^*, x - \bar{x} \rangle + \mu||x - \bar{x}||^2.$$

Now using (4. 5) we see that

$$\langle w^*, x - \bar{x} \rangle \geq \mu||x - \bar{x}||^2$$

Using this fact in (4. 4) we have

$$\hat{H}_\alpha(x) \geq (\mu - \frac{\alpha}{2})||x - \bar{x}||^2.$$

The result follows from the assumption that $2\mu > \alpha$. □

Now comes the most important question from the computational point of view. Is the gap function $\hat{H}_\alpha$ well behaved ?. If we recall our discussion in the beginning of the article then the function $\hat{H}_\alpha$ is well-behaved if for any sequence $x_k$ in $C$ such that $x_k \to \bar{x}$ we must have $\hat{H}_\alpha(x_k) \to 0$. This fact might be not so easy to establish since if we observe carefully then it is clear that one can write $\hat{H}_\alpha$ as follows

$$\hat{H}_\alpha(x) = \inf_{x^* \in T(x)} q_\alpha(x, x^*),$$

where $q_\alpha(x, x^*)$ is as given in Theorem 4.1. Note that this is a very general form of a parametric optimization problem. Thus it might be quiet difficult to guarantee that continuity or at the best the well-behaved-ness of $\hat{H}_\alpha$ with minimal assumptions. However under slightly stringent conditions we can actually show that $\hat{H}_\alpha$ is well behaved. To begin with let us consider the set-valued map $\Lambda : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ which is given as

$$\Lambda(x) = \{x^* \in T(x) : \hat{H}_\alpha(x) = q_\alpha(x, x^*)\}.$$

Note that the assumption that $T$ is non-empty and compact-valued guarantees that $\Lambda$ is also non-empty and compact-valued. We shall now show that if the set-valued map $\Lambda$ is lower-semicontinuous in the sense of Berge [6] then one can guarantee that $\hat{H}_\alpha$ is well behaved. Note that by lower-semicontinuity of the mapping $\Lambda$ we mean that for $x_k \to \bar{x}$ and $\bar{x}^* \in \Lambda(\bar{x})$ there exists $x_k^* in \Lambda(x_k)$ such that $x_k^* \to \bar{x}^*$. Note that by definition of $\hat{H}_\alpha$ for any $\bar{x}^* \in \Lambda(\bar{x})$ we have $q_\alpha(\bar{x}, \bar{x}^*) = 0$. Further since if $x_k$ is a sequence in $C$ converging to the solution $\bar{x}$ then by lower-semicontinuity of the set-valued map $\Lambda$ there exists $x_k^* \in \Lambda(x_k)$ with $x_k^* \to \bar{x}^*$. Thus we have

$$\hat{H}_\alpha(x_k) = q_\alpha(x^k, x^*).$$

Now using the continuity of the function $q_\alpha$ we see that

$$\lim_{k\to\infty} \hat{H}_\alpha(x_k) = q_\alpha(\bar{x}, \bar{x}^*) = 0.$$

This shows that $\hat{H}_\alpha$ is well behaved. Thus we can now summarize our discussion as follows

**Theorem 4.3.** *Let us consider the problem $GVI(T, C)$, where $T$ is compact-valued. Let $\bar{x}$ be a solution to $GVI(T, C)$ and let $\{x^k\}$ be a sequence in $C$ such that $x^k \to \bar{x}$. Further assume that set-valued map $\Lambda$ is lower-semicontinuous in the sense of Berge. Then*

$$\lim_{k\to\infty} \hat{H}_\alpha(x^k) = 0.$$

*Hence $\hat{H}_\alpha$ is well behaved.*

Though the assumption of lower-semicontinuity of the set-valued map $\Lambda$ may appear to be stringent it seems for the present it is difficult to remove this assumption. Thus it remains open to see if there are much simpler conditions on the problem data which will guarantee the function $\hat{H}_\alpha$ is well behaved.

## References


[1] Auchmuty, G., Variational principles for variational inequalities. Numer. Funct. Anal. Optim. 10 (1989), no. 9-10, 863-874.

[2] Auslender, A. Optimisation. Mthodes numriques. Matrise de Mathmatiques et Applications Fondamentales. Masson, Paris, New York, Barcelona, 1976.

[3] Auslender, Alfred Rsolution numrique d'ingalits variationnelles. (French) C. R. Acad. Sci. Paris Sr. A-B 276 (1973), 1063-1066.

[4] Aussel, D.; Dutta, J. On gap functions for multivalued Stampacchia variational inequalities. J. Optim. Theory Appl. 149 (2011), no. 3, 513-527.

[5] Aussel, D. ; Dutta, J., Revisiting a gap function for set-valued variational inequalities, preprint 2012.

[6] Berge, C., Topological spaces. Including a treatment of multi-valued functions, vector spaces and convexity. Translated from the French original by E. M. Patterson. Reprint of the 1963 translation. Dover Publications, Inc., Mineola, NY, 1997.

[7] Burachik, R. S.; Iusem, A. N., Set-valued mappings and enlargements of monotone operators. Springer Optimization and Its Applications, 8. Springer, New York,

[8] Chen, G. Y.; Goh, C. J.; Yang, X. Q. On gap functions and duality of variational inequality problems. J. Math. Anal. Appl. 214 (1997), no. 2, 658-673.

[9] Facchinei, F; Pang, J. S., Finite-dimensional variational inequalities and complementarity problems. Vol. I. Springer Series in Operations Research. Springer-Verlag, New York, 2003.

[10] Fukushima, Masao Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. Math. Programming 53 (1992), no. 1, Ser. A, 99-110

[11] Larsson, T.; Patriksson, M. A class of gap functions for variational inequalities. Math. Programming 64 (1994), no. 1, Ser. A, 53-79.

[12] Marcotte, Patrice A new algorithm for solving variational inequalities with application to the traffic assignment problem. Math. Programming 33 (1985), no. 3, 339-351,

[13] Mordukhovich, Boris S. Variational analysis and generalized differentiation. 330. Springer-Verlag, Berlin, 2006

[14] Pang, J. S., Error bounds in mathematical programming. Math. Programming 79 (1997), no. 1-3, Ser. B, 299-332.

[15] Patriksson, M. Merit functions and descent algorithms for a class of variational inequality problems. Optimization 41 (1997), no. 1, 37-55,

[16] Peng, J., Equivalence of variational inequality problems to unconstrained minimization. Math. Programming 78 (1997), no. 3, Ser. A, 347-355.

[17] Vial, J. P., Strong and weak convexity of sets and functions. Mathematics of Operations Research, 8,(1983), 231-259.

[18] Wu, J. H.; Florian, M; Marcotte, P., A general descent framework for the monotone variational inequality problem. Math. Programming 61 (1993), no. 3,Ser. A, 281-300.

[19] Yamashita, N.; Taji, K.; Fukushima, M. Unconstrained optimization reformulations of variational inequality problems. J. Optim. Theory Appl. 92 (1997), no. 3, 439-456.