BEYOND MONOTONE VARIATIONAL INEQUALITIES: SOLUTION METHODS AND ITERATION COMPLEXITIES

Kevin Huang and Shuzhong Zhang*

This paper is dedicated to the 75th birthday of Prof. Masao Fukushima, who has made great contributions to our community.

Abstract: In this paper, we discuss variational inequality (VI) problems without monotonicity from the perspective of convergence of projection-type algorithms. In particular, we identify existing conditions as well as present new conditions that are sufficient to guarantee convergence. The first half of the paper focuses on the case where a Minty solution exists (also known as *Minty condition*), which is a common assumption in the recent developments for non-monotone VI. The second half explores alternative sufficient conditions that are different from the existing ones such as monotonicity or Minty condition, using an *algorithm-based* approach. Through examples and convergence analysis, we show that these conditions are capable of characterizing different classes of VI problems where the algorithms are guaranteed to converge.

Key words: *variational inequality, Minty solution, non-monotone VI, projection-type method*

Mathematics Subject Classification: *49J40, 65K15, 65Y20*

1 Introduction

Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set, and $F(x) : \mathcal{X} \mapsto \mathbb{R}^n$ be a vector mapping. The following problem is known as the *variational inequality problem* (VI):

$$\text{Find } x^* \in \mathcal{X} \text{ such that } F(x^*)^\top (x - x^*) \geq 0 \text{ for all } x \in \mathcal{X}.$$

The study of finite-dimensional VI problems dates back to 1960's where the complementarity problem was developed to solve for various equilibria, such as economic equilibrium, traffic equilibrium, and in general Nash equilibrium. For a comprehensive study of the applications, theories and algorithms of VI, readers are referred to the celebrated monograph by Facchinei and Pang [5].

We say a problem is a *monotone* VI if the operator F is monotone:

$$\langle F(x) - F(y), x - y \rangle \geq \mu \|x - y\|^2, \quad \forall x, y \in \mathcal{X} \tag{1.1}$$

for some $\mu \geq 0$. If there exists some $\mu > 0$ such that (1.1) holds, then F is referred to as *strongly monotone*. Most contemporary researches on VI have focused on designing algorithms for monotone problems, such as modified forward-backward method [33], mirror-prox

---

*Corresponding author

method [22], dual-extrapolation method [24, 25], hybrid proximal extra-gradient method [20], OGDA [18, 19], and extra-point method [9]. These are also known as "projection-type" methods, while the earliest projection-type methods date back to (gradient) projection method by Sibony [30], proximal method by Martinet [16], and extra-gradient method by Korpelevich [13]. The convergence of these methods for monotone VI is studied by Tseng [32]. The aforementioned methods are all *first-order* projection methods. Recently there are also research on developing *high-order* projection methods and establishing their global convergence iteration complexities. They include [8, 21, 2, 26, 23, 11, 1, 15, 10].

The convergence of the above first-order and high-order (projection-type) methods and the corresponding iteration complexities are analyzed in the framework of monotone VI (or more generally, pseudo-monotone VI, which will be formally defined in the next section). For non-monotone VI, earlier research has developed non-projection-type methods such as the KKT condition based methods and merit function based methods (see [4, 29, 28, 5] and the references therein). However, it is in general difficult to establish iteration complexity for non-projection-type methods for non-monotone VI. In recent years, research on developing algorithms for non-monotone VI has focused on the VI problems where the so-called *Minty solutions* exist. A Minty solution to VI is a solution x^* where the following inequality is satisfied:

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

for all $x \in \mathcal{X}$. When the constraint set is a close convex set and the operator F is continuous and monotone, all solutions to the VI (if any) are Minty solutions. The existence of Minty solutions turns out to be critical in establishing convergence for the projection-type methods for non-monotone VI, and there have been recent developments of such results; see e.g. [15, 31, 3, 34, 14]. Recently, Zhao *et al.* [35] proposed a primal-dual approach to convex-constrained non-monotone VI problems based on augmented Lagrangian and the Minty condition.

In this paper, we follow this line of research on the convergence of projection-type methods for non-monotone VI. We start from the common assumption made in the literature, that is, a Minty solution exists. We show that a general extra-gradient-type method, the Approximation-Based Regularized Extra-Gradient Method (ARE) update proposed in [10], converges in a guaranteed rate with a similar convergence behavior as Perseus in [15]. In addition, we are interested in the concept of Minty solution itself, especially in the non-monotone setting where a VI solution is not necessarily a Minty solution. Therefore, we investigate implications given by the Minty solutions in different problem classes such as optimization and Nash games. Finally, we explore the possibilities of alternative sufficient conditions for convergence of projection-type methods through an *algorithm-based* approach. Conventionally, algorithms are devised to ensure convergence under a given problem framework, such as monotone VI or VI with Minty solutions, and the convergence behavior is analyzed within the framework. In this paper, we follow an opposite direction by deriving sufficient conditions for convergence based on the algorithms we are interested in. In other words, for a given algorithm, we aim to identify VI with specific structures where the algorithm is guaranteed to converge to a solution. It turns out that this approach makes it possible to characterize structures of VI models that are different from commonly encountered conditions such as the monotonicity or the Minty condition. We present several conditions of this kind and demonstrate examples as well as proving convergence of gradient projection methods and extra-gradient methods under these conditions.

The rest of the paper is organized as follows. Section 2 starts the discussion with non-monotone VI with Minty solution. We first provide formal definitions of the solution concepts and merit functions that are relevant to the discussion in this paper. Then the convergence

of ARE is established, followed by discussions on implications of Minty solutions in optimization and Nash games. Section 3 explores algorithm-based sufficient conditions, starting with formal definitions and examples. Convergence of gradient-projection method and extragradient method are established under these conditions. Finally, Section 4 concludes the paper with some further remarks.

2 Non-monotone VI with Minty Solution

2.1 Definitions and solution concepts

2.1.1 VI solutions and Minty solutions

In order to set the background for the discussion in this paper, let us first formally define the variational inequality problem and its solution set. For a given set $\mathcal{X} \subseteq \mathbb{R}^n$ and a continuous mapping $F : \mathcal{X} \mapsto \mathbb{R}^n$, consider the following VI model, to be denoted by $\mathrm{VI}(F; \mathcal{X})$:

$$
\begin{aligned}
&\text{Find} \qquad x^* \in \mathcal{X} \\
&\text{such that} \quad \langle F(x^*), x - x^* \rangle \geq 0, \ \forall x \in \mathcal{X}.
\end{aligned}
$$

Let the solution set of the above model be $\mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$. It is also referred to as the set of *strong solutions*, or simply the *solutions*, to the VI model. The non-emptiness of $\mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$ can be guaranteed by imposing some assumptions on the basic problem structure.

Assumption 2.1. F is a continuous mapping. $\mathcal{X}$ is non-empty, convex and compact.

Assumption 2.1 ensures that $\mathrm{Sol}(\mathrm{VI}(F; \mathcal{X})) \neq \emptyset$ [5, 6], and we shall make this assumption throughout the paper. In addition to the (strong) solutions to the VI, there is another important solution concept, which is the so-called *Minty solutions* or *weak solutions*, defined as the set of x^* such that

$$
\langle F(x), x - x^* \rangle \geq 0, \ \forall x \in \mathcal{X}.
$$

Let the set of Minty solutions to $\mathrm{VI}(F; \mathcal{X})$ be denoted by $\mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X}))$. The well-known Minty's Lemma states the follwing:

Lemma 2.1 (Minty's Lemma [17])**.** *If F is continuous, $\mathcal{X}$ is non-empty, closed and convex, then $\mathrm{Sol}_m(VI(F; \mathcal{X})) \subseteq \mathrm{Sol}(VI(F; \mathcal{X}))$.*

Proof. See [12], Lemma 1.5. $\qquad\square$

If additionally F is monotone, then $\mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X})) = \mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$. Indeed, for every $x^* \in \mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$, we have:

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

thus $x^* \in \mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X}))$. In this paper, while we always assume the non-emptiness of $\mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$ (by Assumption 2.1), we extend the discussion to the broader class of VI where F is not necessarily monotone. Alternatively, we focus on the Minty solutions and assume $\mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X})) \neq \emptyset$ in this section, and we discuss other conditions in Section 3 where no Minty solutions exist.

2.1.2 Merit functions

In the context of strongly monotone VI where the solution x^* uniquely exists, it is common to use (squared) distance to the solution $\|x - x^*\|^2$ in the iteration complexity analysis. For VI problems that are merely monotone, there are two other merit functions that are widely used, known as the *gap function* and the *dual gap function*. While monotonicity is not assumed in this paper, we may still use these two merit functions as the measurement of convergence. In this section we re-introduce them in a fashion that relates them with the two solution concepts, VI solutions $\mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$ and Minty solutions $\mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X}))$, introduced earlier.

Proposition 2.2. *Suppose that $\mathcal{X}$ is compact. It holds that*

$$\mathrm{Sol}(\mathit{VI}(F; \mathcal{X})) \neq \emptyset \Longleftrightarrow \min_{y \in \mathcal{X}} \max_{x \in \mathcal{X}} \langle F(y), y - x \rangle = 0.$$

Proof. Observe that for any $y \in \mathcal{X}$, we always have

$$\max_{x \in \mathcal{X}} \langle F(y), y - x \rangle \geq \langle F(y), y - y \rangle = 0.$$

Hence, $\min_{y \in \mathcal{X}} \max_{x \in \mathcal{X}} \langle F(y), y - x \rangle \geq 0$, or equivalently $\max_{y \in \mathcal{X}} \min_{x \in \mathcal{X}} \langle F(y), x - y \rangle \leq 0$. The equivalence is due to the following relation:

$$
\begin{aligned}
&\min_{y \in \mathcal{X}} \max_{x \in \mathcal{X}} \langle F(y), y - x \rangle \\
={} & \min_{y \in \mathcal{X}} \left(-\min_{x \in \mathcal{X}} \langle F(y), x - y \rangle \right) \\
={} & -\max_{y \in \mathcal{X}} \min_{x \in \mathcal{X}} \langle F(y), x - y \rangle.
\end{aligned}
$$

$\Longrightarrow$: Choose any $x^* \in \mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$. We have $\min_{x \in \mathcal{X}} \langle F(x^*), x - x^* \rangle = 0$, implying

$$\max_{y \in \mathcal{X}} \min_{x \in \mathcal{X}} \langle F(y), x - y \rangle = -\min_{y \in \mathcal{X}} \max_{x \in \mathcal{X}} \langle F(y), y - x \rangle = 0.$$

$\Longleftarrow$: Let

$$y^* \in \arg\max_{y \in \mathcal{X}} \left[\min_{x \in \mathcal{X}} \langle F(y), x - y \rangle \right].$$

It follows that $\min_{x \in \mathcal{X}} \langle F(y^*), x - y^* \rangle = 0$, or equivalently put

$$\langle F(y^*), x - y^* \rangle \geq 0, \text{ for any } x \in \mathcal{X}.$$

Hence, $y^* \in \mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$. $\qquad \square$

In a similar vein, we have:

Proposition 2.3. *Suppose that $\mathcal{X}$ is compact. It holds that*

$$\mathrm{Sol}_m(\mathit{VI}(F; \mathcal{X})) \neq \emptyset \Longleftrightarrow \max_{x \in \mathcal{X}} \min_{y \in \mathcal{X}} \langle F(y), y - x \rangle = 0.$$

Proof. First, observe that in general, we have

$$\max_{x \in \mathcal{X}} \min_{y \in \mathcal{X}} \langle F(y), y - x \rangle \leq 0$$

because for any given $x \in \mathcal{X}$, it follows that $\min_{y \in \mathcal{X}} \langle F(y), y - x \rangle \leq \langle F(x), x - x \rangle = 0$.

$\Longrightarrow$: Choose any $x^* \in \mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X}))$. Since $\langle F(y), y - x^* \rangle \geq 0$ for all $y \in \mathcal{X}$, we have

$$\min_{y \in \mathcal{X}} \langle F(y), y - x^* \rangle = 0,$$

implying $\max_{x \in \mathcal{X}} \min_{y \in \mathcal{X}} \langle F(y), y - x \rangle = 0$.

$\Longleftarrow$: Let

$$x^* \in \arg\max_{x \in \mathcal{X}} \left[\min_{y \in \mathcal{X}} \langle F(y), y - x \rangle \right].$$

We have $\min_{y \in \mathcal{X}} \langle F(y), y - x^* \rangle = 0$, or equivalently

$$\langle F(y), y - x^* \rangle \geq 0, \text{ for all } y \in \mathcal{X},$$

implying $x^* \in \mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X}))$. $\qquad\square$

The above analysis naturally leads to the following notions of merit functions:

$$G(x) := \max_{y \in \mathcal{X}} \langle F(x), x - y \rangle,$$

also known as the *gap function*, and

$$H(x) := \max_{y \in \mathcal{X}} \langle F(y), x - y \rangle,$$

also known as the *dual gap function*.

Based on our analysis, we have:

Proposition 2.4.

$G(x) \geq 0$ *for all* $x \in \mathcal{X}$, *and* $G(x) = 0$ *if and only if* $x \in \mathrm{Sol}(\mathit{VI}(F; \mathcal{X}))$.

$H(x) \geq 0$ *for all* $x \in \mathcal{X}$, *and* $H(x) = 0$ *if and only if* $x \in \mathrm{Sol}_m(\mathit{VI}(F; \mathcal{X}))$.

Therefore, we may introduce the following notion of ϵ-solutions.

Definition 2.5. For $\epsilon > 0$, we call x to be an ϵ-VI solution if $G(x) \leq \epsilon$; we call x to be an ϵ-Minty solution if $H(x) \leq \epsilon$.

2.1.3 Relaxation of monotonicity

We remark that there are several conditions can be made on the structure of F, under which the (pure) monotonicity is relaxed but the connections between $\mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$ and $\mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X}))$ still exist. While we do not assume most of these conditions, we summarize them below for the benefit of easy referencing. In particular, we only consider the *Minty condition* among others, which simply states $\mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X})) \neq \emptyset$.

Weak Sharpness:

$$\langle F(x^*), x - x^* \rangle \geq \mu \|x - x^*\|^2, \quad \forall x \in \mathcal{X}, \quad x^* \in \mathrm{Sol}(\mathrm{VI}(F, \mathcal{X})),$$

for some $\mu \geq 0$.

Pseudo-monotonicity:

$$\langle F(y), x - y\rangle \geq 0 \implies \langle F(x), x - y\rangle \geq 0, \quad \forall x, y \in \mathcal{X}.$$

Strongly pseudo-monotonicity:

$$\langle F(y), x - y\rangle \geq 0 \implies \langle F(x), x - y\rangle \geq \mu\|x - y\|^2, \quad \forall x, y \in \mathcal{X},$$

for some $\mu > 0$.

Quasi-monotonicity:

$$\langle F(y), x - y\rangle > 0 \implies \langle F(x), x - y\rangle \geq 0, \quad \forall x, y \in \mathcal{X}.$$

Minty's condition: $\mathrm{Sol}_m(\mathrm{VI}(F;\mathcal{X})) \neq \emptyset$, i.e. there exists $x^* \in \mathcal{X}$ such that

$$\langle F(x), x - x^*\rangle \geq 0, \quad \forall x \in \mathcal{X}.$$

Strong Minty's condition (Generalized monotonicity): there exists $x^* \in \mathcal{X}$ such that

$$\langle F(x), x - x^*\rangle \geq \mu\|x - x^*\|^2, \quad \forall x \in \mathcal{X},$$

for some $\mu \geq 0$.

A few remarks are in place to specify some implications given by the above conditions.

Remark 2.6.

- If F is pseudo-monotone, then $\mathrm{Sol}(\mathrm{VI}(F;\mathcal{X})) \subset \mathrm{Sol}_m(\mathrm{VI}(F;\mathcal{X}))$. If further F is continuous, $\mathcal{X}$ is nonempty, closed and convex, then $\mathrm{Sol}(\mathrm{VI}(F;\mathcal{X})) = \mathrm{Sol}_m(\mathrm{VI}(F;\mathcal{X}))$.

- If $\mathcal{X}$ is closed and bounded, then $\mathrm{Sol}_m(\mathrm{VI}(F;\mathcal{X})) \neq \emptyset$ if and only if F is quasi-monotone [7].

- Assume $\mathrm{Sol}(\mathrm{VI}(F;\mathcal{X})) \neq \emptyset$, then the following relations hold:

$$\text{monotone} \implies \text{pseudo-monotone} \implies \text{Minty's condition}$$

and

$$\text{strongly monotone} \implies \text{strongly pseudo-monotone} \implies \text{strong Minty's condition}$$

2.2 Convergence of projection-type methods

In this section, we present a solution method of projection type that can be shown to converge to Minty solutions by simply assuming $\mathrm{Sol}_m(\mathrm{VI}(F;\mathcal{X})) \neq \emptyset$. This method, known as *Approximation-based Regularized Extra-gradient method*, or simply ARE, was a general p^{th}-order method of extra-gradient type, proposed in [10] originally for solving monotone VI with convergence rate $\mathcal{O}(N^{-\frac{p+1}{2}})$. It turns out that ARE not only solves monotone VI, but also solves non-monotone VI as long as $\mathrm{Sol}_m(\mathrm{VI}(F;\mathcal{X})) \neq \emptyset$. A similar technique in the analysis has been used in [15]. It is shown that a different projection-type method—Perseus—can converge to Minty solutions for non-monotone VI, and the rate is the same as the one developed in this paper.

The ARE update proceeds as follows:

$$
\begin{cases}
x^{k+0.5} & := & \mathrm{VI}_{\mathcal{X}}\left(\tilde{F}(x; x^k) + L_p\|x - x^k\|^{p-1}(x - x^k)\right), \\
x^{k+1} & := & \arg\min_{x \in \mathcal{X}} \langle F(x^{k+0.5}), x - x^k \rangle + \frac{L_p\|x^{k+0.5} - x^k\|^{p-1}}{2}\|x - x^k\|^2,
\end{cases}
\tag{2.1}
$$

for $k = 1, 2, ...$, where L_p is the Lipschitz constant for $\nabla^{p-1}F(x)$ satisfying the condition

$$
\|\nabla^{p-1}F(x) - \nabla^{p-1}F(y)\| \leq L_p\|x - y\|,
\tag{2.2}
$$

$\tilde{F}(\cdot; y) : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a general approximation mapping estimated at y satisfying the bound:

$$
\|\tilde{F}(x; y) - F(x)\| \leq \tau L_p\|x - y\|^p,
\tag{2.3}
$$

and we use the notation $\mathrm{VI}_{\mathcal{X}}(F)$ to denote solving for a solution in $\mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$ as a subroutine in the update. In the ARE update (2.1), the subroutine at iteration k specifically solve the VI model associated with the regularized approximation operator $\tilde{F}(x; x^k) + L_p\|x - x^k\|^{p-1}(x - x^k)$.

The next theorem summarizes the convergence rate of the ARE update (2.1) under the assumption $\mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X})) \neq \emptyset$.

Theorem 2.7. *Consider the ARE update (2.1). Suppose that conditions (2.2) and (2.3) are satisfied, and $\mathrm{Sol}_m(VI(F; \mathcal{X})) \neq \emptyset$. Then, the sequence produced by ARE converges at the following rate:*

$$
\|x^{k_N+0.5} - x^{k_N}\|^2 = \mathcal{O}(1/N), \quad G(x^{k_N+0.5}) = \mathcal{O}(1/N^{\frac{p}{2}}).
$$

Proof. The initial steps in the analysis will follow the exact same logic as the proof in [10] for Theorem 3.1. By the definition of $x^{k+0.5}$, we have

$$
\langle \tilde{F}(x^{k+0.5}; x^k) + L_p\|x^{k+0.5} - x^k\|^{p-1}(x^{k+0.5} - x^k), x - x^{k+0.5} \rangle \geq 0, \ \forall x \in \mathcal{X}.
\tag{2.4}
$$

Denote $\gamma_k = L_p\|x^{k+0.5} - x^k\|^{p-1}$. Substituting $x = x^{k+1}$ in (2.4) we have

$$
\begin{aligned}
& \langle \tilde{F}(x^{k+0.5}; x^k), x^{k+1} - x^{k+0.5} \rangle \\
\geq \ & \gamma_k \langle x^{k+0.5} - x^k, x^{k+0.5} - x^{k+1} \rangle \\
= \ & \frac{\gamma_k}{2}\left(\|x^{k+0.5} - x^k\|^2 + \|x^{k+1} - x^{k+0.5}\|^2 - \|x^{k+1} - x^k\|^2\right).
\end{aligned}
\tag{2.5}
$$

On the other hand, by the optimality condition at x^{k+1} we have

$$
\langle F(x^{k+0.5}) + \gamma_k(x^{k+1} - x^k), x - x^{k+1} \rangle \geq 0, \ \text{for all } x \in \mathcal{X}.
$$

Hence,

$$
\begin{aligned}
& \langle F(x^{k+0.5}), x - x^{k+1} \rangle \\
\geq \ & \gamma_k \langle x^{k+1} - x^k, x^{k+1} - x \rangle \\
= \ & \frac{\gamma_k}{2}\left(\|x^{k+1} - x\|^2 + \|x^{k+1} - x^k\|^2 - \|x^k - x\|^2\right), \ \text{for all } x \in \mathcal{X}.
\end{aligned}
\tag{2.6}
$$

Continue with the above inequality, for any given $x \in \mathcal{X}$ we have

$$\frac{\gamma_k}{2} \left(\|x^{k+1} - x\|^2 + \|x^{k+1} - x^k\|^2 - \|x^k - x\|^2 \right)$$

$$\overset{(2.6)}{\leq} \langle F(x^{k+0.5}), x - x^{k+1} \rangle$$

$$= \langle F(x^{k+0.5}), x - x^{k+0.5} \rangle + \langle F(x^{k+0.5}), x^{k+0.5} - x^{k+1} \rangle$$

$$= \langle F(x^{k+0.5}), x - x^{k+0.5} \rangle + \langle F(x^{k+0.5}) - \tilde{F}(x^{k+0.5}; x^k), x^{k+0.5} - x^{k+1} \rangle$$
$$+ \langle \tilde{F}(x^{k+0.5}; x^k), x^{k+0.5} - x^{k+1} \rangle$$

$$\leq \langle F(x^{k+0.5}), x - x^{k+0.5} \rangle + \|F(x^{k+0.5}) - \tilde{F}(x^{k+0.5}; x^k)\| \cdot \|x^{k+0.5} - x^{k+1}\|$$
$$+ \langle \tilde{F}(x^{k+0.5}; x^k), x^{k+0.5} - x^{k+1} \rangle$$

$$\leq \langle F(x^{k+0.5}), x - x^{k+0.5} \rangle + \frac{\|F(x^{k+0.5}) - \tilde{F}(x^{k+0.5}; x^k)\|^2}{2\gamma_k} + \frac{\gamma_k \|x^{k+0.5} - x^{k+1}\|^2}{2}$$

$$+ \langle \tilde{F}(x^{k+0.5}; x^k), x^{k+0.5} - x^{k+1} \rangle$$

$$\overset{(2.3)}{\leq} \langle F(x^{k+0.5}), x - x^{k+0.5} \rangle + \frac{\tau^2 L_p^2 \|x^{k+0.5} - x^k\|^{2p}}{2\gamma_k} + \frac{\gamma_k \|x^{k+0.5} - x^{k+1}\|^2}{2}$$

$$+ \langle \tilde{F}(x^{k+0.5}; x^k), x^{k+0.5} - x^{k+1} \rangle.$$

Noticing that $\frac{\tau^2 L_p^2 \|x^{k+0.5} - x^k\|^{2p}}{2\gamma_k} = \frac{\tau^2 \gamma_k \|x^{k+0.5} - x^k\|^2}{2}$, and further using (2.5) we derive from the above that

$$\frac{\gamma_k}{2} \left(\|x^{k+1} - x\|^2 + \|x^{k+1} - x^k\|^2 - \|x^k - x\|^2 \right)$$

$$\leq \langle F(x^{k+0.5}), x - x^{k+0.5} \rangle + \frac{\tau^2 \gamma_k \|x^{k+0.5} - x^k\|^2}{2} + \frac{\gamma_k \|x^{k+0.5} - x^{k+1}\|^2}{2}$$

$$+ \frac{\gamma_k}{2} \left[-\|x^{k+0.5} - x^k\|^2 - \|x^{k+1} - x^{k+0.5}\|^2 + \|x^{k+1} - x^k\|^2 \right].$$

Canceling out terms, we simplify the above inequality into

$$\langle F(x^{k+0.5}), x^{k+0.5} - x \rangle + \frac{\gamma_k}{2} \left(1 - \tau^2 \right) \|x^{k+0.5} - x^k\|^2 \leq \frac{\gamma_k}{2} \left[\|x^k - x\|^2 - \|x^{k+1} - x\|^2 \right]. \quad (2.7)$$

In the original analysis in [10], the rest of the proof continues with the monotonicity of F. In this analysis, we assume the Minty condition (i.e. $\mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X})) \neq \emptyset$) holds instead of monotonicity of F. Taking any fixed $x = x^* \in \mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X}))$ in the above inequality, we have:

$$\left(1 - \tau^2 \right) \|x^{k+0.5} - x^k\|^2 \leq \left[\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 \right], \quad (2.8)$$

since $\langle F(x^{k+0.5}), x^{k+0.5} - x^* \rangle \geq 0$. Summing up this inequality for $k = 1, ..., N$ gives us:

$$\min_{1 \leq k \leq N} \|x^{k+0.5} - x^k\|^2 \leq \frac{1}{N} \sum_{k=1}^{N} \|x^{k+0.5} - x^k\|^2 \leq \frac{1}{N(1 - \tau^2)} \|x^1 - x^*\|^2, \quad (2.9)$$

which proves the first convergence result.

Condition (2.4) for updating $x^{k+0.5}$ implies for all $x \in \mathcal{X}$, we have:

$$\langle \tilde{F}(x^{k+0.5}; x^k), x^{k+0.5} - x \rangle \leq -L_p \|x^{k+0.5} - x^k\|^{p-1} (x^{k+0.5} - x^k)^\top (x^{k+0.5} - x)$$
$$\leq L_p D \|x^{k+0.5} - x^k\|^p,$$

where $D := \max\limits_{x,x' \in \mathcal{X}} \|x - x'\|$.

Denote $k_N := \arg\min\limits_{1 \le k \le N} \|x^{k+0.5} - x^k\|^2$, we have:

$$
\begin{aligned}
&\langle F(x^{k_N+0.5}), x^{k_N+0.5} - x \rangle \\
=\ &\langle F(x^{k_N+0.5}) - \tilde{F}(x^{k_N+0.5}; x^{k_N}), x^{k_N+0.5} - x \rangle + \langle \tilde{F}(x^{k_N+0.5}; x^{k_N}), x^{k_N+0.5} - x \rangle \\
\le\ &\left\| F(x^{k_N+0.5}) - \tilde{F}(x^{k_N+0.5}; x^{k_N}) \right\| \cdot \|x^{k_N+0.5} - x\| + L_p D \|x^{k_N+0.5} - x^{k_N}\|^p \\
\le\ &(1+\tau)L_p D \|x^{k_N+0.5} - x^{k_N}\|^p \\
\le\ &(1+\tau)L_p D \frac{1}{N^{\frac{p}{2}}(1-\tau^2)^{\frac{p}{2}}} \|x^1 - x^*\|^p \le \frac{(1+\tau)L_p D^{p+1}}{N^{\frac{p}{2}}(1-\tau^2)^{\frac{p}{2}}},
\end{aligned}
$$

which holds for all $x \in \mathcal{X}$. Therefore,

$$
G(x^{k_N+0.5}) = \max_{x \in \mathcal{X}} \langle F(x^{k_N+0.5}), x^{k_N+0.5} - x \rangle \le \frac{(1+\tau)L_p D^{p+1}}{N^{\frac{p}{2}}(1-\tau^2)^{\frac{p}{2}}}, \tag{2.10}
$$

which proves the second convergence result. □

Remark 2.8. In the above analysis, by using the same $x^* \in \text{Sol}_m(\text{VI}(F; \mathcal{X}))$ in (2.8) repetitively, it can be seen that the sequence $\{x^k\}$ converges to the specific Minty solution in terms of squared distance. On the other hand, while $\|x^{k_N+0.5} - x^{k_N}\|^2$ also converges at a rate $1/N$ by (2.9), the final result guarantees a rate of convergence $1/N^{\frac{p}{2}}$ in terms of the merit function $G(x^{k_N+0.5})$, which gives an ϵ-VI solution (but not necessarily an ϵ-Minty solution) based on Definition 2.5.

Finally, we note that if F is monotone (in which case $\text{Sol}(\text{VI}(F; \mathcal{X})) = \text{Sol}_m(\text{VI}(F; \mathcal{X}))$), the convergence rate is $1/N^{\frac{p+1}{2}}$ in terms of the merit function $H(\bar{x}_N)$, where $\bar{x}_N$ is the weighted average of $x^{k+0.5}$ (see [10], Theorem 3.1). It is an improved rate compared to the above result where we only assume $\text{Sol}_m(\text{VI}(F; \mathcal{X})) \neq \emptyset$.

2.3 Minty solutions beyond general VI

In the previous section, we see that the existence of Minty solutions is indeed an important property to have for solving non-monotone VI. Without assuming any other conditions, it allows projection-type methods such as extra-gradient method (a first-order specialized method of ARE) to converge (to a Minty solution) with guaranteed rate. It is then natural to ask whether the same solution concept presents with similar significance in other problem classes related to VI and what are the implications of the Minty solution therein. In this section, we first discuss the role of the Minty solution in optimization. The discussion proceeds in the context of Nash games, where we present the implications of the Minty solution and its connections to the VI model.

2.3.1 Minty solutions in optimization

Consider the optimization problem:

$$
\min_{x \in \mathcal{X}} f(x), \tag{2.11}
$$

where $f(x)$ is continuously differentiable, $\mathcal{X}$ is convex and closed. The local first-order optimality condition is given by:

$$
\langle \nabla f(x^*), x - x^* \rangle \ge 0, \quad \forall x \in \mathcal{X}, \tag{2.12}
$$

which is equivalent to the VI model $\text{VI}(\nabla f; \mathcal{X})$ with solution set $\text{Sol}(\text{VI}(\nabla f; \mathcal{X}))$. Now suppose a Minty solution exists for this VI model, that is, $\text{Sol}_m(\text{VI}(\nabla f; \mathcal{X})) \neq \emptyset$ and any element $x^* \in \text{Sol}_m(\text{VI}(\nabla f; \mathcal{X}))$ satisfies

$$\langle \nabla f(x), x - x^* \rangle \geq 0, \quad \forall x \in \mathcal{X}. \tag{2.13}$$

Note that Minty's Lemma (Lemma 2.1) applies here due to our assumption. Therefore, we have $\text{Sol}_m(\text{VI}(\nabla f; \mathcal{X})) \subseteq \text{Sol}(\text{VI}(\nabla f; \mathcal{X}))$, and any x^* satisfying (2.13) is a local first-order stationary point ((2.12) holds). The next theorem states that a Minty solution in optimization, if exists, is in fact a global solution to the problem.

Theorem 2.9 (Optimality of Minty solution). *For the optimization problem* (2.11) *where* f *is continuously differentiable,* $\mathcal{X}$ *is convex and closed. The following holds:*

$$\langle \nabla f(x), x - x^* \rangle \geq 0, \quad \forall x \in \mathcal{X}$$
$$\implies \quad f(x^*) \leq f(x), \quad \forall x \in \mathcal{X}.$$

In other words, if a Minty solution exists, it is a global solution to the problem.

Proof. Using the following identity:

$$f(x^*) - f(x) = \int_{t=0}^{1} \nabla f(x + t(x^* - x))^\top (x^* - x) dt,$$

we have for $t = 1$, $\nabla f(x^*)^\top (x^* - x) \leq 0$ since (2.12) holds. For $t = 0$, we have $\nabla f(x)^\top (x^* - x) \leq 0$ due to (2.13). For $0 < t < 1$, let $\hat{x} = x + t(x^* - x) \in \mathcal{X}$, then

$$\nabla f(x + t(x^* - x))^\top (x^* - x) = \frac{1}{1-t} \nabla f(\hat{x})^\top (x^* - \hat{x}) \leq 0, \quad 0 < t < 1,$$

where the last inequality is again due to (2.13). Therefore, we can conclude that

$$f(x^*) - f(x) = \int_{t=0}^{1} \nabla f(x + t(x^* - x))^\top (x^* - x) dt \leq 0$$
$$\implies \quad f(x^*) \leq f(x).$$

$\square$

Remark 2.10. The Minty solution is always a global solution in optimization, provided that $f(x)$ is continuously differentiable and $\mathcal{X}$ is closed convex set. However, a global solution needs not be a Minty solution. For a meaningful optimization problem, i.e. a global solution is well defined and exists, a Minty solution may not exist.

Consider a one-dimensional optimization problem:

$$\min_{-1 \leq x \leq 1} \; -x^2,$$

the global solutions are $x^* = -1, 1$. For $x^* = -1$:

$$\langle \nabla f(x), x - x^* \rangle = \langle -2x, x + 1 \rangle < 0, \; 0 < x \leq 1;$$

for $x^* = 1$,

$$\langle \nabla f(x), x - x^* \rangle = \langle -2x, x - 1 \rangle < 0, \; -1 \leq x < 0.$$

Therefore, neither of the global solutions is a Minty solution. Same as the VI model, when the objective function is convex, the set of local solutions $(\text{Sol}(\text{VI}(\nabla f; \mathcal{X})))$ coincides with the set of Minty solutions $(\text{Sol}_m(\text{VI}(\nabla f; \mathcal{X})))$, thus every global solution is a Minty solution.

$\boxed{\textbf{2.3.2}}$ **Minty solutions in Games**

Consider a two-player game:

$$\begin{cases} x: & \min_{x \in \mathcal{X}} \; \theta_x(x,y) \\ y: & \min_{y \in \mathcal{Y}} \; \theta_y(x,y), \end{cases} \tag{2.14}$$

where we use x, y to denote both the players and their corresponding strategies. Assume θ_x, θ_y are both continuously differentiable for fixed y, x, and $\mathcal{X}, \mathcal{Y}$ are closed convex sets. Let us first define three different notions of equilibria in this game, starting from the well-known Nash equilibrium.

Definition 2.11 (Nash equilibrium (NE)). A solution pair $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$ is said to be in Nash equilibrium if and only if

$$\theta_x(x^*, y^*) \leq \theta_x(x, y^*), \, \forall x \in \mathcal{X}, \quad \theta_y(x^*, y^*) \leq \theta_y(x^*, y), \, \forall y \in \mathcal{Y}.$$

In other words, for player x it is not possible to be better off by deviating from the Nash equilibrium strategy x^* if the opponent continues to play $y = y^*$ and vice versa. For a Nash equilibrium pair (x^*, y^*), x^* (resp. y^*) is the global minimizer of the objective function $\theta_x(\cdot, y^*)$ (resp. $\theta_y(x^*, \cdot)$) for fixed y^* (resp. x^*).

Definition 2.12 (Quasi-Nash equilibrium [27] (QNE)). A solution pair $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$ is said to be in quasi-Nash equilibrium if and only if

$$\langle \nabla_x \theta_x(x^*, y^*), x - x^* \rangle \geq 0, \, \forall x \in \mathcal{X}, \quad \langle \nabla_y \theta_y(x^*, y^*), y - y^* \rangle \geq 0, \, \forall y \in \mathcal{Y}.$$

Unlike Nash equilibrium where x^* and y^* have to be global minimizers of their respective objective functions when the opponent plays the equilibrium strategy, a pair of quasi-Nash equilibrium only requires the first-order stationarity condition to be satisfied in their respective optimization problem. Hence quasi-Nash equilibrium can be viewed as a relaxation of Nash equilibrium.

Definition 2.13 (Minty Nash equilibrium (MNE)). A solution pair $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$ is said to be in Minty Nash equilibrium if and only if

$$\langle \nabla_x \theta_x(x, y^*), x - x^* \rangle \geq 0, \, \forall x \in \mathcal{X}, \quad \langle \nabla_y \theta_y(x^*, y), y - y^* \rangle \geq 0, \, \forall y \in \mathcal{Y}.$$

The third definition given above pertains to the notion of Minty solution discussed thus far. It requires x^* (resp. y^*) to be a Minty solution of $\theta_x(\cdot, y^*)$ (resp. $\theta_y(x^*, \cdot)$) for fixed y^* (resp. x^*). By the discussion in the previous section, the set of Minty solutions is only a subset of the global solutions, therefore the Minty Nash equilibrium defines a stronger concept of equilibrium than the usual notion of Nash equilibrium.

It is straightforward to conclude the following relation among these three different notions of equilibria:

$$\text{MNE} \implies \text{NE} \implies \text{QNE}.$$

If the objective functions possess an additional property known as *block multiconvex*, i.e. $\theta_x(\cdot, y)$ is convex for fixed $y \in \mathcal{Y}$ and $\theta_y(x, \cdot)$ is convex for fixed $x \in \mathcal{X}$, then the above relation becomes:

$$\text{MNE} = \text{NE} = \text{QNE}.$$

Let us now consider the connections among the above notions of equilibria to the solutions in the VI formulation of the two-player game (2.14):

$$F(z) := \begin{pmatrix} \nabla_x \theta_x(x,y) \\ \nabla_y \theta_y(x,y) \end{pmatrix}, \quad z := (x,y)^\top, \quad \mathcal{Z} := \mathcal{X} \times \mathcal{Y}, \tag{2.15}$$

which can be expressed as the VI model $\mathrm{VI}(F;\mathcal{Z})$.

If $z^* = (x^*, y^*) \in \mathrm{Sol}(\mathrm{VI}(F;\mathcal{Z}))$, i.e.

$$\langle F(z^*), z - z^* \rangle = \langle \nabla_x \theta_x(x^*, y^*), x - x^* \rangle + \langle \nabla_y \theta_y(x^*, y^*), y - y^* \rangle \geq 0, \quad \forall z \in \mathcal{Z},$$

it is obvious that (x^*, y^*) is a pair of quasi-Nash equilibrium of the original two-player game by taking $x = x^*$ and $y = y^*$ in the above inequality. On the other hand, if (x^*, y^*) is a pair of quasi-Nash equilibrium, then the above inequality holds trivially and $z^* = (x^*, y^*) \in \mathrm{Sol}(\mathrm{VI}(F;\mathcal{Z}))$. Therefore, quasi-Nash equilibrium of the game is equivalent to the (strong) solution to the VI formulation.

Now if $z^* = (x^*, y^*) \in \mathrm{Sol}_m(\mathrm{VI}(F;\mathcal{Z}))$, i.e.

$$\langle F(z), z - z^* \rangle = \langle \nabla_x \theta_x(x,y), x - x^* \rangle + \langle \nabla_y \theta_y(x,y), y - y^* \rangle \geq 0, \quad \forall z \in \mathcal{Z}, \tag{2.16}$$

by taking any arbitrary $x \in \mathcal{X}$ and $y = y^*$ in the above inequality, we have

$$\langle \nabla_x \theta_x(x, y^*), x - x^* \rangle \geq 0, \quad \forall x \in \mathcal{X}.$$

Similarly we have

$$\langle \nabla_y \theta_y(x^*, y), y - y^* \rangle \geq 0, \quad \forall y \in \mathcal{Y}.$$

The above two inequalities combined indicates that (x^*, y^*) is a pair of Minty Nash equilibrium of the original two-player game. However, we note that the opposite direction is in general not true, since a Minty solution z^* to $\mathrm{VI}(F;\mathcal{X})$ requires the inequality (2.16) to be satisfied with $\nabla_x \theta_x(x,y)$ (resp. $\nabla_y \theta_y(x,y)$) while the Minty Nash equilibrium only defines on $\nabla_x \theta_x(x, y^*)$ (resp. $\nabla_y \theta_y(x^*, y)$), in which the opponent's strategy is fixed to be the equilibrium strategy.

We can include the two solution concepts in the VI formulation, $\mathrm{Sol}(\mathrm{VI}(F;\mathcal{Z}))$ and $\mathrm{Sol}_m(\mathrm{VI}(F;\mathcal{Z}))$, in the previous relation and obtain

$$\mathrm{Sol}_m(\mathrm{VI}(F;\mathcal{Z})) \implies \mathrm{MNE} \implies \mathrm{NE} \implies \mathrm{QNE} = \mathrm{Sol}(\mathrm{VI}(F;\mathcal{Z})).$$

In the case where both $\theta_x(x,y)$ and $\theta_y(x,y)$ are bock multiconvex, then

$$\mathrm{Sol}_m(\mathrm{VI}(F;\mathcal{Z})) \implies \mathrm{MNE} = \mathrm{NE} = \mathrm{QNE} = \mathrm{Sol}(\mathrm{VI}(F;\mathcal{Z})).$$

Note that it is not sufficient to state the equivalence between the Minty solution in VI and others even if the objective functions are block multiconvex. However, the above relation does offer a quick argument for the existence of Nash equilibrium for non-cooperative games where the payoff functions are block multiconvex and continuously differentiable, and the constraints are convex compact sets. Indeed, the latter two conditions are exactly given in Assumption 2.1, which guarantees that $\mathrm{Sol}(\mathrm{VI}(F;\mathcal{Z})) \neq \emptyset$. The functions being block multiconvex indicates that $\mathrm{NE} = \mathrm{Sol}(\mathrm{VI}(F;\mathcal{Z}))$, which proves the existence of Nash equilibrium. This conclusion is summarized in the next proposition.

Proposition 2.14. *For the two-player game* (2.14)*, if both* $\theta_x(x, y)$ *and* $\theta_y(x, y)$ *are block multiconvex and continuously differentiable, and* $\mathcal{X}, \mathcal{Y}$ *are convex compact sets, then a Nash equilibrium exists.*

Remark 2.15. Indeed, the celebrated Nash Theorem has shown that a mixed strategy Nash equilibrium exists in n-player multilinear games using Brouwer's fixed-point theorem. The above discussion only points out an alternative route for showing the same conclusion with a potentially lighter algebraic derivation. The equivalence between NE and QNE as well as between QNE and $\mathrm{Sol}(\mathrm{VI}(F; \mathcal{Z}))$ is straightforward. Proving $\mathrm{Sol}(\mathrm{VI}(F; \mathcal{Z})) \neq \emptyset$ can be accomplished via the same Brouwer's fixed-point theorem or degree theory [5], where the details are omitted here.

We note that restricting the number of players to be two in this section is only for the purpose of clear illustrations of the ideas. All the discussions can be easily extended to general n-player games.

3 | Algorithm-Based Conditions on VI

In the discussion thus far, we have focused on non-monotone VI with the condition $\mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X})) \neq \emptyset$, and we show that it is sufficient for a projection-type method such as ARE to converge globally with a guaranteed rate. In this section, we continue to ask the question: Are there other sufficient conditions different from the existing ones that are able to guarantee the convergence of algorithms of certain class? It turns out that through an *algorithm-based* approach, it is possible to characterize structures of the VI model by deriving sufficient conditions for which the algorithms converge. These VI problems can be of special interest since they are not necessarily monotone or satisfy the Minty condition, nonetheless the algorithms converge regardless. In particular, we present conditions on VI models based on projection-type methods, analyze their convergence behavior, and provide examples of problems satisfying these conditions.

3.1 | Conditions for projection-type methods

In order to present the conditions to be introduced later with more precise expressions, let us first define two projection-type mappings, which play a central role in these conditions since the purpose is to characterize VI problems with guaranteed convergence for projection-type methods.

Definition 3.1 (Gradient projection mapping). For a given $t > 0$, assuming $\mathcal{X}$ is a closed convex set, define the "gradient projection mapping" as

$$M(x; t) := \mathrm{Proj}_{\mathcal{X}}(x - tF(x)). \tag{3.1}$$

Note that the term "gradient" follows the convention in optimization, while in general F can be any vector mapping that is not necessarily a gradient mapping.

It is a well-known fact that for a fixed $t > 0$, $x^* \in \mathcal{X}$ is a fixed point of the gradient projection mapping $M(\cdot; t)$ if and only if $x^* \in \mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$. It is then natural to use a third merit function other than the (dual) gap function introduced in Section 2.1.2:

$$P(x) := \|M(x; t) - x\|^2.$$

We summarize the above observations in the next proposition and provide a proof for completeness.

Proposition 3.2. $x^* = M(x^*; t)$ *if and only if* $x^* \in \mathrm{Sol}(VI(F; \mathcal{X}))$. *Therefore,* $P(x) = 0$ *if and only if* $x \in \mathrm{Sol}(VI(F; \mathcal{X}))$.

Proof. By the optimality condition of the projection operation, we have

$$(y - M(x; t))^\top (M(x; t) - x + tF(x)) \geq 0, \quad \forall y \in \mathcal{X}. \tag{3.2}$$

If $M(x; t) \neq x$, then by setting $y = x$ in (3.2) we observe

$$(M(x; t) - x)^\top F(x) \leq -\frac{1}{t} \|M(x; t) - x\|^2 < 0$$

implying that $x \notin \mathrm{Sol}(VI(F; \mathcal{X}))$. On the other hand, if $M(x; t) = x$, then (3.2) yields

$$t(y - x)^\top F(x) \geq 0, \quad \forall y \in \mathcal{X},$$

and so $x \in \mathrm{Sol}(VI(F; \mathcal{X}))$. $\qquad\square$

Remark 3.3. In the literature, the mapping $M(x; 1) - x$ is also referred to as the "natural map" and solving $M(x; 1) - x = 0$ can be used as an equation reformulation of the VI model $VI(F; \mathcal{X})$. In our discussion, we do not explicitly adopt the equation reformulation approach, but only use $P(x)$ as one of the measurements of convergence.

In view of the gradient projection mapping defined earlier, let us also define the following "extra-gradient projection mapping", which expresses the extra-gradient-type methods and the sufficient condition for convergence more succinctly.

Definition 3.4 (Extra-gradient projection mapping). For a given $t > 0$, assuming $\mathcal{X}$ is a closed convex set, define the "extra-gradient mapping" as

$$M^+(x; t) := \mathrm{Proj}_{\mathcal{X}}(x - tF(M(x; t))),$$

where $M(x; t)$ is the gradient projection mapping defined in (3.1).

We are now ready to introduce conditions that can guarantee the convergence for different projection-type methods. These conditions provide additional characterizations of the structure of a VI problem when in general we do not assume monotonicity nor Minty condition.

Definition 3.5. Condition (Local Minty).
For some fixed $t > 0$ and any $x \in \mathcal{X}$ there is $x^* \in \mathrm{Sol}(VI(F; \mathcal{X}))$ such that

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

and for the same x^* the above inequality also holds if we replace x by $M(x; t)$.

Definition 3.6. Condition (Local Minty+).
For some fixed $t > 0$ and any $x \in \mathcal{X}$ there is $x^* \in \mathrm{Sol}(VI(F; \mathcal{X}))$ such that

$$\langle F(M(x; t)), M(x; t) - x^* \rangle \geq 0,$$

and for the same x^* the above inequality also holds if we replace x by $M^+(x; t)$.

Definition 3.7. Condition (Local Minty*).

For some fixed $t > 0$ and any $x \in \mathcal{X}$ there is $x^* \in \mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$ such that

$$\langle F(x), M(x; t) - x^* \rangle \geq 0,$$

and for the same x^* the above inequality also holds if we replace x by $M(x; t)$.

Definition 3.8. Condition (GP).

For some fixed $\delta > 0$ and $t > 0$, and any $x \in \mathcal{X}$ there is $x^* \in \mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$ such that

$$4(1 + \delta)t\langle F(M(x; t)), M(x; t) - x^* \rangle + \|M(x; t) - x\|^2 \geq 0,$$

and for the same x^* the above inequality also holds if we replace x by $M(x; t)$.

Definition 3.9. Condition (GP+).

For some fixed $\delta > 0$ and $t > 0$, and any $x \in \mathcal{X}$ there is $x^* \in \mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$ such that

$$4(1 + \delta)t\langle F(M(x; t)), M(x; t) - x^* \rangle + \|M(x; t) - x\|^2 \geq 0,$$

and for the same x^* the above inequality also holds if we replace x by $M^+(x; t)$.

Definition 3.10. Condition (GP*).

For some fixed $\delta > 0$ and $t > 0$, and any $x \in \mathcal{X}$ there is $x^* \in \mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$ such that

$$2(1 + \delta)t\langle F(x), M(x; t) - x^* \rangle + \|M(x; t) - x\|^2 \geq 0,$$

and for the same x^* the above inequality also holds if we replace x by $M(x; t)$.

We make the following remarks on the conditions introduced above.

Remark 3.11.

- Condition (Local Minty) (Local Minty*): Unlike other conditions that are seen before, both conditions are defined on sequences in $\mathcal{X}$ rather than arbitrary points. In particular these sequences are generated by the gradient projection mapping $M(x; t)$. For any specific sequence, there is a "local Minty solution" for which the inequality defined in the respective condition continues to hold.

- Condition (Local Minty+): When generating a sequence from the extra-gradient projection mapping $M^+(x; t)$, we immediately obtain another sequence that maps the previous sequence to their gradient projection mapping $M(x; t)$. (Local Minty+) is defined on the latter sequences.

- Condition (GP), (GP+), (GP*): These three conditions can be viewed as relaxations of Condition (Local Minty), (Local Minty+), and (Local Minty*), by allowing a positive term $P(x)$ in the defining inequality.

The conditions introduced above have one property in common: they all require the defining inequality to hold for the whole sequence generated from some pre-determined mapping. In other words, a solution $x^* \in \mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$ can "attract" some $x \in \mathcal{X}$ following the particular sequence. Condition (Local Minty)/(Local Minty+)/(Local Minty*) (Definition 3.5, 3.6, and 3.7) guarantee the existence of such "local Minty solution" x^* with respect to arbitrary $x \in \mathcal{X}$, while Condition (GP)/(GP+)/(GP*) (Definition 3.8, 3.9, and 3.10)

relax the previous three conditions. The term "local" is in contrast to the normal Minty solution, which is "global" since any $x^* \in \mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X}))$ is able to attract *every* point in $\mathcal{X}$ in terms of the Minty inequality $\langle F(x), x - x^* \rangle \geq 0$ in the definition. This property turns out to be critical to derive these algorithm-based conditions, which helps establish convergence of projection-type methods for those problems with more general structures than the common monotonicity or Minty condition.

Remark 3.12.

- If F is monotone, then Condition (Local Minty) and Condition (Local Minty+) hold trivially (therefore so do Conditions (GP) and (GP+)).

- If $\mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X})) \neq \emptyset$, then Condition (Local Minty) and Condition (Local Minty+) hold trivially (therefore so do Condition (GP) and (GP+)).

- It is possible that F is monotone but Conditions (Local Minty*) and (GP*) do not hold. Conversely, it is also possible that Conditions (Local Minty*) and (GP*) hold but F is not monotone or $\mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X})) = \emptyset$.

The next two examples demonstrate problem instances where $\mathrm{Sol}_m(\mathrm{VI}(F)) = \emptyset$ while Condition (GP)/(GP+)/(GP*) still hold.

Example 3.13. Consider $\mathcal{X} = [-1, 1]$, $F(x) = -x$. In that case, $\mathrm{Sol}(\mathrm{VI}(F; \mathcal{X})) = \{-1, 0, +1\}$ and $\mathrm{Sol}_m(\mathrm{VI}(F; \mathcal{X})) = \emptyset$. Note that

$$M(x; t) = \begin{cases} \min\{1, (1+t)x\}, & \text{if } x > 0 ; \\ 0, & \text{if } x = 0 ; \\ \max\{-1, (1+t)x\}, & \text{if } x < 0 . \end{cases}$$

In particular, for $x > 0$ we choose $x^* = 1$; for $x < 0$ we choose $x^* = -1$; for $x = 0$ we choose $x^* = 0$. It is now easy to verify that Condition (GP)/(GP+)/(GP*) hold in this case. In fact, Condition (Local Minty)/(Local Minty+)/(Local Minty*) all hold in this example.

Example 3.14. Consider $\mathcal{X} = \|x\| \leq 1$, $F(x) = Qx$, where $Q = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$. In this case, $\mathrm{Sol}(\mathrm{VI}(F; \mathcal{X})) = \{(1, 0)^\top, (0, 0)^\top, (-1, 0)^\top\}$. None of them is a Minty solution, but we can show that $\mathrm{VI}(F; \mathcal{X})$ satisfies (Local Minty)/(Local Minty+)/(Local Minty+) for any fixed $t \in (0, 1]$. For the first two conditions, it suffices to provide the following two observations as the proof. Also note that we can focus on $x_1 \geq 0$, since the behavior is symmetric for the case $x_1 \leq 0$.

1. For any $x = (x_1, x_2)^\top \in \mathcal{X}$ such that $x_1 \geq 0$, the whole sequence generated by $M(x; t)$ and $M^+(x; t)$ will remain $x_1 \geq 0$. This can be easily verified.

2. For any $x = (x_1, x_2)^\top \in \mathcal{X}$ such that $x_1 \geq 0$, we have

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

for $x^* = (1, 0)^\top$. Since the above inequality results in $x_2^2 \geq x_1(x_1 - 1)$, which always holds for any $0 \leq x_1 \leq 1$.

It remains to show that Condition (Local Minty*) also holds. Similarly, let us focus on $x_1 \geq 0$ and use $x^* = (1,0)^\top$. Let us denote $x^+ = x - tF(x) = ((1+t)x_1, (1-t)x_2)^\top$, then we have

$$M(x;t) = \begin{cases} x^+, & \text{if } \|x^+\| \leq 1 ; \\ x^+/\|x^+\|, & \text{if } \|x^+\| > 1 . \end{cases}$$

For the case $\|x^+\| \leq 1$, the condition

$$\langle F(x), M(x;t) - x^* \rangle \geq 0 \tag{3.3}$$

reduces to

$$(1-t)x_2^2 \geq (1+t)x_1^2 - x_1,$$

where the RHS is always non-positive for $x_1 \leq \frac{1}{1+t}$, which is the case for $\|x^+\| \leq 1$. Therefore inequality (3.3) holds. For $\|x^+\| > 1$, condition (3.3) can be reduced to

$$(1-t)x_2^2 \geq (1+t)x_1^2 - x_1 \cdot \|x^+\|.$$

Since $\|x^+\| \geq (1+t)x_1$, the RHS is always non-positive and condition (3.3) holds.

The next example shows that, even if F is monotone, Condition (GP*) does not necessarily hold. Otherwise, since Condition (GP*) is sufficient for the gradient projection method to converge (as will be shown in the next section), monotonicity would have been sufficient for the convergence as well (which is not the case for the gradient projection method).

Example 3.15. Consider $\mathcal{X} = \|x\| \leq 1$, $F(x) = Qx$, where $Q = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. In this case, $\text{Sol}(\text{VI}(F;\mathcal{X})) = \{(0,0)^\top\}$. This problem is originated from the saddle point problem $\min\max_{\|x\|^2+\|y\|^2\leq 1} xy$ and is monotone. For a small $\epsilon > 0$, consider $x = (\epsilon, 0)^\top$, where $F(x) = (0, -\epsilon)^\top$.

Consider first $t \leq \frac{\sqrt{1-\epsilon^2}}{\epsilon}$. In this case, $M(x;t) = x - tF(x) = \epsilon \cdot (1,t)^\top$. Therefore,

$$2(1+\delta)t\langle F(x), M(x;t) - x^* \rangle + \|M(x;t) - x\|^2 = -2(1+\delta)t^2\epsilon^2 + t^2\epsilon^2 < 0.$$

On the other hand, if $t > \frac{\sqrt{1-\epsilon^2}}{\epsilon}$, then $M(x;t) = (1,t)^\top \cdot (1+t^2)^{-\frac{1}{2}}$, and

$$\|M(x;t) - x\|^2 = \left\| \frac{(1 - \epsilon\sqrt{1+t^2}, t)^\top}{\sqrt{1+t^2}} \right\|^2 = 1 + \epsilon^2 - \frac{2\epsilon}{\sqrt{1+t^2}} \leq 1 + \epsilon^2,$$

whereas

$$2(1+\delta)t\langle F(x), M(x;t) - x^* \rangle = -2(1+\delta)\epsilon \cdot t^2 \cdot (1+t^2)^{-\frac{1}{2}} < -2(1+\delta)\cdot(1-\epsilon^2),$$

where the last inequality we take $t = \frac{\sqrt{1-\epsilon^2}}{\epsilon}$. It is then clear that for small enough ϵ, Condition (GP*) will not hold, even if F is monotone.

3.2 Convergence of projection-type methods

In the previous section, we present several algorithm-based conditions that are defined for sequences generated from either the gradient projection mapping or the extra-gradient projection mapping. In this section we show how these conditions are applied in the convergence analysis for two projection-type methods, the vanilla gradient projection method and the extra-gradient method.

$\boxed{\textbf{3.2.1}}$ **The gradient projection method**

Consider the gradient projection method:

$$x^{k+1} \quad := \quad \arg\min_{x \in \mathcal{X}} \langle F(x^k), x - x^k \rangle + \frac{1}{2t}\|x - x^k\|^2,$$

or equivalently written as

$$x^{k+1} \quad := \quad M(x^k; t).$$

It is now clear why the conditions (Local Minty)/(Local Minty*) and their relaxations (GP)/(GP*) are defined for sequences generated from the gradient projection mapping. They assume that for each sequence generated by the gradient projection method there exists at least one solution $x^* \in \mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$ such that their respective defining inequalities continue to hold. We first derive a key intermediate inequality from the gradient projection update itself, which makes it clearer the exact condition to be used in the following convergence analysis.

Lemma 3.16. *For the gradient projection method, we have*

$$\frac{1}{2}\|x^k - x^*\|^2 \geq \frac{1}{2}\|x^{k+1} - x^*\|^2 + t\langle F(x^k), x^{k+1} - x^* \rangle + \frac{1}{2}\|x^{k+1} - x^k\|^2. \qquad (3.4)$$

for any $x^* \in \mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$.

Proof. Since

$$\langle tF(x^k) + x^{k+1} - x^k, x - x^{k+1} \rangle \geq 0 \quad \forall x \in \mathcal{X},$$

with $x = x^*$, we have

$$\langle tF(x^k), x^* - x^{k+1} \rangle \geq \frac{1}{2}\left(\|x^{k+1} - x^k\|^2 + \|x^{k+1} - x^*\|^2 - \|x^k - x^*\|^2\right).$$

Rearranging terms gives the result. $\square$

In view of the inequality (3.4), it is straightforward that Conditions (Local Minty*) and (GP*) can provide a bound on $\langle F(x^k), x^{k+1} - x^* \rangle$ for the whole sequence with respect to some $x^* \in \mathrm{Sol}(\mathrm{VI}(F; \mathcal{X}))$, thus the convergence follows. The results are summarized in the next theorem.

Theorem 3.17. *Under Condition (GP*), and assume F is Lipschitz continuous with constant L, the gradient projection algorithm is convergent for $\mathrm{VI}(F; \mathcal{X})$. Moreover,*

$$\min_{1 \leq k \leq N} P(x^k) = O(1/N), \quad \min_{1 \leq k \leq N} G(x^k) = O(1/N^{\frac{1}{2}}).$$

Proof. In view of Lemma 3.16 and Condition (GP*), we have:

$$
\begin{aligned}
\frac{1}{2}\left(\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2\right) \quad &\geq \quad t\langle F(x^k), x^{k+1} - x^* \rangle + \frac{1}{2}\|x^{k+1} - x^k\|^2 \\
&\overset{\text{(GP*)}}{\geq} \quad -\frac{1}{2(1+\delta)}\|x^{k+1} - x^k\|^2 + \frac{1}{2}\|x^{k+1} - x^k\|^2 \\
&= \quad \frac{\delta}{2(1+\delta)}\|x^{k+1} - x^k\|^2.
\end{aligned}
$$

By Condition (GP), there exists an x^* such that the above inequality holds for $k = 1, ..., N$. Therefore, summing up the inequality for $k = 1, ..., N$, we have:

$$\sum_{k=1}^{N} \|x^{k+1} - x^k\|^2 \leq \left(1 + \frac{1}{\delta}\right) \|x^1 - x^*\|^2,$$

which implies

$$\min_{1 \leq k \leq N} P(x^k) \leq \frac{1}{N} \sum_{k=1}^{N} \|x^{k+1} - x^k\|^2 \leq \frac{1}{N} \cdot \left(1 + \frac{1}{\delta}\right) \|x^1 - x^*\|^2 = \mathcal{O}(1/N).$$

Moreover, we can transform the measurement in $\min\limits_{1 \leq k \leq N} P(x^k)$ into $\min\limits_{1 \leq k \leq N} G(x^k)$. Since

$$\langle tF(x^k) + x^{k+1} - x^k, x - x^{k+1} \rangle \geq 0 \quad \forall x \in \mathcal{X},$$

we have

$$\langle F(x^k), x^{k+1} - x \rangle \leq -\frac{1}{t}(x^{k+1} - x)^\top (x^{k+1} - x^k) \leq \frac{D}{t} \|x^{k+1} - x^k\|.$$

Therefore,

$$\begin{aligned}
\langle F(x^{k+1}), x^{k+1} - x \rangle &= \langle F(x^k), x^{k+1} - x \rangle + \langle F(x^{k+1}) - F(x^k), x^{k+1} - x \rangle \\
&\leq \frac{D}{t} \|x^{k+1} - x^k\| + \|F(x^{k+1}) - F(x^k)\| \cdot \|x^{k+1} - x\| \\
&\leq D\left(\frac{1}{t} + L\right) \|x^{k+1} - x^k\|.
\end{aligned}$$

Define $k_N := \arg\min\limits_{1 \leq k \leq N} P(x^k)$, then

$$\begin{aligned}
\langle F(x^{k_N+1}), x^{k_N+1} - x \rangle &\leq D\left(\frac{1}{t} + L\right) \|x^{k_N+1} - x^{k_N}\| \\
&\leq D\left(\frac{1}{t} + L\right) \frac{1}{N^{\frac{1}{2}}} \left(1 + \frac{1}{\delta}\right)^{\frac{1}{2}} \|x^1 - x^*\| \\
&\leq D^2\left(\frac{1}{t} + L\right) \frac{1}{N^{\frac{1}{2}}} \left(1 + \frac{1}{\delta}\right)^{\frac{1}{2}},
\end{aligned}$$

which holds for all $x \in \mathcal{X}$. This implies

$$\min_{1 \leq k \leq N} G(x^k) = \max_{x \in \mathcal{X}} \langle F(x^{k_N}), x^{k_N} - x \rangle \leq D^2 \left(\frac{1}{t} + L\right) \frac{1}{N^{\frac{1}{2}}} \left(1 + \frac{1}{\delta}\right)^{\frac{1}{2}} = \mathcal{O}(1/N^{\frac{1}{2}}).$$

$\square$

It is well-known that if F is strongly monotone, gradient projection method can be guaranteed to converge with iteration complexity $\mathcal{O}\left(\frac{L^2}{\mu^2} \log \frac{1}{\epsilon}\right)$, while F being merely monotone is insufficient for the convergence. On the other hand, Theorem 3.17 provides a different sufficient condition (GP*) such that gradient projection method converges globally in terms of $\min_{1 \leq k \leq N} G(x^k) = G(x^{k_N})$. As shown in the previous examples, there exist problems which are not monotone but Condition (GP*) is satisfied.

3.3 The extra-gradient method

Consider

$$
\begin{cases}
x^{k+0.5} & := \quad \arg\min_{x \in \mathcal{X}} \langle F(x^k), x - x^k \rangle + \frac{1}{2t} \|x - x^k\|^2 \\
x^{k+1} & := \quad \arg\min_{x \in \mathcal{X}} \langle F(x^{k+0.5}), x - x^k \rangle + \frac{1}{2t} \|x - x^k\|^2,
\end{cases}
$$

or it can be equivalently written as

$$
\begin{cases}
x^{k+0.5} & := \quad M(x^k; t) \\
x^{k+1} & := \quad M^+(x^k; t),
\end{cases}
$$

Note that the extra-gradient method can be viewed as a special case of the ARE update discussed in Section 2.2 with $p = 1$. Below we introduce a key inequality derived from the update, which also appears with a similar form in the analysis for ARE (cf. (2.7)). Both the notation and the parameter constraint are slightly adjusted in the following lemma, therefore a proof is provided for completeness.

Lemma 3.18. *For the extra-gradient method, assume that F is Lipschitz continuous with constant L, and $t \leq \frac{1}{\sqrt{2}L}$, the following inequality holds:*

$$
\langle F(x^{k+0.5}), x^{k+0.5} - x \rangle + \frac{1}{4t} \|x^{k+0.5} - x^k\|^2 \leq \frac{1}{2t} \left[\|x^k - x\|^2 - \|x^{k+1} - x\|^2 \right]. \tag{3.5}
$$

Proof. Optimality of $x^{k+0.5}$:

$$
\langle F(x^k) + \frac{1}{t}(x^{k+0.5} - x^k), x - x^{k+0.5} \rangle \geq 0, \ \forall x \in \mathcal{X}. \tag{3.6}
$$

Substituting $x = x^{k+1}$ in (3.6) we have

$$
\begin{aligned}
& \langle F(x^k), x^{k+1} - x^{k+0.5} \rangle \\
\geq \ & \frac{1}{t} \langle x^{k+0.5} - x^k, x^{k+0.5} - x^{k+1} \rangle \\
= \ & \frac{1}{2t} \left(\|x^{k+0.5} - x^k\|^2 + \|x^{k+1} - x^{k+0.5}\|^2 - \|x^{k+1} - x^k\|^2 \right). \tag{3.7}
\end{aligned}
$$

On the other hand, by the optimality condition at x^{k+1} we have

$$
\langle F(x^{k+0.5}) + \frac{1}{t}(x^{k+1} - x^k), x - x^{k+1} \rangle \geq 0, \ \text{for all } x \in \mathcal{X}.
$$

Hence,

$$
\begin{aligned}
& \langle F(x^{k+0.5}), x - x^{k+1} \rangle \\
\geq \ & \frac{1}{t} \langle x^{k+1} - x^k, x^{k+1} - x \rangle \\
= \ & \frac{1}{2t} \left(\|x^{k+1} - x\|^2 + \|x^{k+1} - x^k\|^2 - \|x^k - x\|^2 \right), \ \text{for all } x \in \mathcal{X}. \tag{3.8}
\end{aligned}
$$

Continue with the above inequality, for any given $x \in \mathcal{X}$ we have

$$\frac{1}{2t}\left(\|x^{k+1} - x\|^2 + \|x^{k+1} - x^k\|^2 - \|x^k - x\|^2\right)$$

$$\overset{(3.8)}{\leq} \langle F(x^{k+0.5}), x - x^{k+1}\rangle$$

$$= \langle F(x^{k+0.5}), x - x^{k+0.5}\rangle + \langle F(x^{k+0.5}), x^{k+0.5} - x^{k+1}\rangle$$

$$= \langle F(x^{k+0.5}), x - x^{k+0.5}\rangle + \langle F(x^{k+0.5}) - F(x^k), x^{k+0.5} - x^{k+1}\rangle$$
$$\quad + \langle F(x^k), x^{k+0.5} - x^{k+1}\rangle$$

$$\leq \langle F(x^{k+0.5}), x - x^{k+0.5}\rangle + \|F(x^{k+0.5}) - F(x^k)\| \cdot \|x^{k+0.5} - x^{k+1}\|$$
$$\quad + \langle F(x^k), x^{k+0.5} - x^{k+1}\rangle$$

$$\leq \langle F(x^{k+0.5}), x - x^{k+0.5}\rangle + \frac{t\|F(x^{k+0.5}) - F(x^k)\|^2}{2} + \frac{\|x^{k+0.5} - x^{k+1}\|^2}{2t}$$
$$\quad + \langle F(x^k), x^{k+0.5} - x^{k+1}\rangle$$

$$\leq \langle F(x^{k+0.5}), x - x^{k+0.5}\rangle + \frac{tL^2\|x^{k+0.5} - x^k\|^2}{2} + \frac{\|x^{k+0.5} - x^{k+1}\|^2}{2t}$$
$$\quad + \langle F(x^k), x^{k+0.5} - x^{k+1}\rangle.$$

Since $t \leq \frac{1}{\sqrt{2}L}$, and with (3.7) we have

$$\frac{1}{2t}\left(\|x^{k+1} - x\|^2 + \|x^{k+1} - x^k\|^2 - \|x^k - x\|^2\right)$$

$$\leq \langle F(x^{k+0.5}), x - x^{k+0.5}\rangle + \frac{\|x^{k+0.5} - x^k\|^2}{4t} + \frac{\|x^{k+0.5} - x^{k+1}\|^2}{2t}$$
$$\quad + \frac{1}{2t}\left[-\|x^{k+0.5} - x^k\|^2 - \|x^{k+1} - x^{k+0.5}\|^2 + \|x^{k+1} - x^k\|^2\right].$$

Canceling out terms, we simplify the above inequality into

$$\langle F(x^{k+0.5}), x^{k+0.5} - x\rangle + \frac{1}{4t}\|x^{k+0.5} - x^k\|^2 \leq \frac{1}{2t}\left[\|x^k - x\|^2 - \|x^{k+1} - x\|^2\right].$$

$$\square$$

One can immediately identify the connections between inequality (3.5) and Condition (GP+). In fact, inequality (3.5) plays the central role in the convergence of the extra-gradient method (or in general, the extra-gradient-type method such as ARE), and iteration complexities of different orders can be established following this inequality based on the conditions imposed on the VI model. The most conventional assumption will be the (strong) monotonicity of F, while in Section 2.2 it is relaxed to be Minty condition $\text{Sol}_m(\text{VI}(F; \mathcal{X})) \neq \emptyset$. Here, Condition (GP+) provides a more direct way to guide the convergence analysis of the extra-gradient method based on inequality (3.5), as summarized in the next theorem.

Theorem 3.19. *Under Condition (GP+), and assume F is Lipschitz continuous with constant L, the extra-gradient method with $t \leq \frac{1}{\sqrt{2}L}$ is convergent for $\text{VI}(F; \mathcal{X})$. Moreover,*

$$\min_{1 \leq k \leq N} P(x^k) = O(1/N), \quad \min_{1 \leq k \leq N} G(x^{k+0.5}) = O(1/N^{\frac{1}{2}}).$$

Proof. In view of Lemma 3.18 and Condition (GP+), we have:

$$\frac{1}{2t}\left[\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2\right] \geq \langle F(x^{k+0.5}), x^{k+0.5} - x^*\rangle + \frac{1}{4t}\|x^{k+0.5} - x^k\|^2$$

$$\overset{\text{(GP+)}}{\geq} -\frac{1}{4t(1+\delta)}\|x^{k+0.5} - x^k\|^2 + \frac{1}{4t}\|x^{k+0.5} - x^k\|^2$$

$$= \frac{\delta}{1+\delta}\frac{1}{4t}\|x^{k+0.5} - x^k\|^2.$$

Since Condition (GP+) also asserts that there exists an x^* such that the above inequality holds for $k = 1, ..., N$, summing up the inequality for $k = 1, ..., N$ gives us:

$$\sum_{k=1}^{N}\|x^{k+0.5} - x^k\|^2 \leq 2\left(1 + \frac{1}{\delta}\right)\|x^1 - x^*\|^2,$$

which implies

$$\min_{1 \leq k \leq N} P(x^k) \leq \frac{1}{N}\sum_{k=1}^{N}\|x^{k+0.5} - x^k\|^2 \leq \frac{2}{N} \cdot \left(1 + \frac{1}{\delta}\right)\|x^1 - x^*\|^2 = \mathcal{O}(1/N).$$

Moreover, we can transform the measurement in $\min_{1 \leq k \leq N} P(x^k)$ into $\min_{1 \leq k \leq N} G(x^{k+0.5})$. Since

$$\langle tF(x^k) + x^{k+0.5} - x^k, x - x^{k+0.5}\rangle \geq 0 \quad \forall x \in \mathcal{X},$$

we have

$$\langle F(x^k), x^{k+0.5} - x\rangle \leq -\frac{1}{t}(x^{k+0.5} - x)^\top(x^{k+0.5} - x^k) \leq \frac{D}{t}\|x^{k+0.5} - x^k\|.$$

Therefore,

$$\langle F(x^{k+0.5}), x^{k+0.5} - x\rangle = \langle F(x^k), x^{k+0.5} - x\rangle + \langle F(x^{k+0.5}) - F(x^k), x^{k+0.5} - x\rangle$$

$$\leq \frac{D}{t}\|x^{k+0.5} - x^k\| + \|F(x^{k+0.5}) - F(x^k)\| \cdot \|x^{k+0.5} - x\|$$

$$\leq D\left(\frac{1}{t} + L\right)\|x^{k+0.5} - x^k\|.$$

Define $k_N := \arg\min_{1 \leq k \leq N} P(x^k)$, then

$$\langle F(x^{k_N+0.5}), x^{k_N+0.5} - x\rangle \leq D\left(\frac{1}{t} + L\right)\|x^{k_N+0.5} - x^{k_N}\|$$

$$\leq D\left(\frac{1}{t} + L\right)\frac{\sqrt{2}}{N^{\frac{1}{2}}}\left(1 + \frac{1}{\delta}\right)^{\frac{1}{2}}\|x^1 - x^*\|$$

$$\leq D^2\left(\frac{1}{t} + L\right)\frac{\sqrt{2}}{N^{\frac{1}{2}}}\left(1 + \frac{1}{\delta}\right)^{\frac{1}{2}},$$

which holds for all $x \in \mathcal{X}$. This implies

$$\min_{1 \leq k \leq N} G(x^{k+0.5}) = \max_{x \in \mathcal{X}}\langle F(x^{k_N+0.5}), x^{k_N+0.5} - x\rangle \leq D^2\left(\frac{1}{t} + L\right)\frac{\sqrt{2}}{N^{\frac{1}{2}}}\left(1 + \frac{1}{\delta}\right)^{\frac{1}{2}} = \mathcal{O}(1/N^{\frac{1}{2}}).$$

$\square$

The convergence rate in Theorem 3.19 turns out to be the same as the rate in Theorem 3.17 for gradient projection method when Condition (GP*) is satisfied instead, as well as the rate in Theorem 2.7 for ARE ($p = 1$) when the Minty condition is satisfied. As discussed in the earlier examples, there exists problems where Conditions (GP+) or (GP*) are satisfied but no Minty solution exists. On the other hand, while these assumptions are able to provide alternative sufficient conditions for the convergence of certain class of algorithms, in general they can be difficult to verify *a priori* due to the requirement to hold for the whole sequence.

$\boxed{4}$ Conclusion

In this paper, we discuss sufficient conditions for projection-type methods to converge in VI problems that are not necessarily monotone. We first focus on the problem where a Minty solution exists, which is a relaxation of the monotonicity assumption. We derive the guaranteed global convergence rate for a general extra-gradient type method ARE under the Minty condition, and then we extend the discussion to properties and implications of Minty solutions in more specific problem classes such as optimization and Nash games. Finally, we present conditions on VI problems that are algorithm-based, in the sense that they are closely connected to the algorithms we are interested in (in particular, projection-type methods) and can suitably serve as sufficient conditions to guarantee the convergence. Conventionally, the algorithms are designed for problems where assumptions on the structure are made *a priori*, and the convergence is only guaranteed under these assumptions. In this paper, we provide an alternative aspect, by "desinging" conditions on the VI model such that they are sufficient to guarantee convergence of certain class of algorithms. We show that this approach is indeed capable of characterizing different classes of VI problems (potentially broader) from the existing ones such as monotone VI or VI with Minty solutions. We analyze the convergence of gradient projection method and extra-gradient method under the proposed conditions. There are still questions remaining, such as: if there are other algorithm-based conditions that can be derived for different projection-type methods or non-projection-type methods; if there exist different characterizations of these conditions such that they can be more easily verified. Answering these questions require some efforts in the future research.

References

[1] D. Adil, B. Bullins, A. Jambulapati and S. Sachdeva, Optimal methods for higher-order smooth monotone variational inequalities, *arXiv preprint arXiv:2205.06167*, 2022.

[2] B. Bullins and K. A. Lai, Higher-order methods for convex-concave min-max optimization and monotone variational inequalities, *SIAM Journal on Optimization* 32 (2022) 2208–2229.

[3] R.S. Burachik and R.D. Millan, A projection algorithm for non-monotone variational inequalities, *Set-Valued and Variational Analysis* 28 (2020) 149–166.

[4] F. Facchinei, A. Fischer and C. Kanzow, Regularity properties of a semismooth reformulation of variational inequalities, *SIAM Journal on Optimization* 8 (1998) 850–869.

[5] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer Science & Business Media, 2007.

[6] P.T. Harker and J.S. Pang, Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications, *Mathematical programming* 48 (1990) 161–220.

[7] Y. He, Solvability of the Minty variational inequality, *Journal of Optimization Theory and Applications*, 174 (2017), 686–692.

[8] K. Huang, J. Zhang and S. Zhang, Cubic regularized Newton method for the saddle point models: A global and local convergence analysis, *Journal of Scientific Computing* 91 (2022): 60.

[9] K. Huang and S. Zhang, A unifying framework of accelerated first-order approach to strongly monotone variational inequalities, *arXiv preprint arXiv:2103.15270*, 2021.

[10] K. Huang and S. Zhang, An approximation-based regularized extra-gradient method for monotone variational inequalities, *arXiv preprint arXiv:2210.04440*, 2022.

[11] R. Jiang and A. Mokhtari, Generalized optimistic methods for convex-concave saddle point problems, *arXiv preprint arXiv:2202.09674*, 2022.

[12] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications.* SIAM, 2000.

[13] G.M. Korpelevich, The extragradient method for finding saddle points and other problems, *Matecon* 12 (1976) 747–756.

[14] M. Lei and Y. He, An extragradient method for solving variational inequalities without monotonicity, *Journal of Optimization Theory and Applications* 188 (2021) 432–446.

[15] T. Lin and M. Jordan, Perseus: A simple high-order regularization method for variational inequalities, *arXiv preprint arXiv:2205.03202*, 2022.

[16] B. Martinet, Brève communication. régularisation d'inéquations variationnelles par approximations successives, *Revue française d'informatique et de recherche opérationnelle. Série rouge* 4 (1970) 154–158.

[17] G.J. Minty, Monotone (nonlinear) operators in hibert space, *Duke Math. J.* 29 (1973) 341–346.

[18] A. Mokhtari, A. Ozdaglar and S. Pattathil, A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach, *arXiv preprint arXiv:1901.08511*, 2019.

[19] A. Mokhtari, A. Ozdaglar and S. Pattathil, Convergence rate of $O(1/k)$ for optimistic gradient and extragradient methods in smooth convex-concave saddle point problems, *SIAM Journal on Optimization* 30 (2020) 3230–3251.

[20] R.D.C. Monteiro and B. F. Svaiter, On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean, *SIAM Journal on Optimization* 20 (2010) 2755–2787.

[21] R.D.C. Monteiro and B.F. Svaiter, Iteration-complexity of a Newton proximal extragradient method for monotone variational inequalities and inclusion problems, *SIAM Journal on Optimization* 22 (2012) 914–935.

[22] A. Nemirovski, Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems, *SIAM Journal on Optimization* 15 (2004) 229–251.

[23] Y. Nesterov, Cubic regularization of Newton's method for convex problems with constraints, 2006.

[24] Y. Nesterov, Dual extrapolation and its applications to solving variational inequalities and related problems, *Mathematical Programming* 109 (2007) 319–344.

[25] Y. Nesterov and L. Scrimali, Solving strongly monotone variational and quasi-variational inequalities. *SSRN 970903*, 2006.

[26] P. Ostroukhov, R. Kamalov, P. Dvurechensky and A. Gasnikov, Tensor methods for strongly convex strongly concave saddle point problems and strongly monotone variational inequalities, *arXiv preprint arXiv:2012.15595*, 2020.

[27] J.-S. Pang and G. Scutari, Nonconvex games with side constraints, *SIAM Journal on Optimization* 21 (2011) 1491–1522.

[28] J.-M. Peng and M. Fukushima, A hybrid Newton method for solving the variational inequality problem via the D-gap function, *Mathematical Programming* 86 (1999) 367–386.

[29] L. Qi and H. Jiang, Semismooth Karush-Kuhn-Tucker equations and convergence analysis of Newton and quasi-Newton methods for solving these equations, *Mathematics of Operations Research* 22 (1997) 301–325.

[30] M. Sibony, Méthodes itératives pour les équations et inéquations aux dérivées partielles non linéaires de type monotone, *Calcolo* 7 (1970) 65–183.

[31] C. Song, Z. Zhou, Y. Zhou, Y. Jiang and Y. Ma, Optimistic dual extrapolation for coherent non-monotone variational inequalities, *Advances in Neural Information Processing Systems* 33 (2020) 14303–14314.

[32] P. Tseng, On linear convergence of iterative methods for the variational inequality problem, *Journal of Computational and Applied Mathematics* 60 (1995) 237–252.

[33] P. Tseng, A modified forward-backward splitting method for maximal monotone mappings, *SIAM Journal on Control and Optimization* 38 (2000) 431–446.

[34] M. Ye, An infeasible projection type algorithm for nonmonotone variational inequalities, *Numerical Algorithms* 89 (2022) 1723–1742.

[35] L. Zhao, D. Zhu and S. Zhang, An augmented Lagrangian approach to conically constrained non-monotone variational inequality problems, *arXiv preprint arXiv:2306.01214*, 2023.

Kevin Huang
Institute of Industrial Engineering
National Taiwan University
E-mail address: kdhuang@ntu.edu.tw

Shuzhong Zhang
Department of Industrial and Systems Engineering
University of Minnesota
E-mail address: zhangs@umn.edu