SUTD 40.616—Special Topics in Games, Learning, and Optimisation

Lecture 17—Learning with Variational Inequalities

Iosif Sakos

November 20, 2025

Abstract

In these notes, we study the effectiveness of iterative methods for solving variational inequality (VI) problems defined over compact, convex sets. We begin by introducing the Minty variational inequality (MVI) problem, a dual formulation of the classical Stampacchia variational inequality (SVI) problem. We then present the Proximal Point Algorithm (PPA), a conceptual method for solving VIs defined by strongly monotone operators. We show how Projected Gradient Descent (PGD) achieves comparable performance when applied to such problems. Finally, we explain how the Extra-Gradient Method (EGM) can be used to solve VIs problems with monotone, rather than strongly monotone, operators

Disclaimer. These lecture notes are a working draft and will be revised and expanded over time. They do not aim to cover the subject exhaustively; the goal is to highlight key ideas and develop some central proofs in detail. The topic is an active research area, so both the notes and our understanding of the material may evolve.

Contents

List of Abbreviations

List of Symbols

1 Variational inequalities

Recall that, given a nonempty set $\mathcal{X} \subseteq \mathbb{R}^d$ and an operator $F\colon \mathcal{X} \to \mathbb{R}^d$, the Stampacchia variational inequality (SVI) problem $P_{\mathrm{SVI}}(F, \mathcal{X})$ seeks $x^\star \in \mathcal{X}$ such that

$$\left\langle F(x^\star),\, x - x^\star \right\rangle \geq 0, \qquad \forall x \in \mathcal{X}. \tag{1}$$

In a previous lecture, we saw that whenever $\mathcal{X}$ is convex and compact and F is continuous, solutions to the SVI problem exist [1]. Moreover, if F is strongly monotone, the solution is unique [2].

Furthermore, we saw that in the context of continuous games, SVIs provide a powerful framework to characterize the Nash equilibria (NE) of the game. In particular, suppose $\mathscr{G} = \left\{ [\![n]\!], (\mathcal{S}_i)_{i\in[\![n]\!]}, (u_i)_{i\in[\![n]\!]} \right\}$ is an n-player continuous game with convex strategy sets $\mathcal{S}_1, \ldots, \mathcal{S}_n$, and differentiable payoff functions $u_i\colon \mathcal{S} \to \mathbb{R}$, for all $i \in [\![n]\!]$, such that each $u_i(\cdot, s_{-i})$ is concave for every fixed $s_{-i} \in \mathcal{S}_{-i}$, where $\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_n$ denotes the joint strategy space. Then, the NE of $\mathscr{G}$ coincide with solutions of the SVI problem $P_{\mathrm{SVI}}(F, \mathcal{S})$ [3], where $F\colon \mathcal{S} \to \mathbb{R}^d$ is the pseudo-gradient of $\mathscr{G}$ defined as

$$F(s) = \begin{bmatrix} -\nabla_{s_1} u_1(s) \\ \vdots \\ -\nabla_{s_n} u_n(s) \end{bmatrix}, \qquad \forall s \in \mathcal{S}. \tag{pseudo-gradient}$$

1.1 The Minty variational inequalities problem

We now introduce another canonical formulation of variational inequality (VI) problems due to Minty [4], which is defined as follows:

Definition 1 (Minty VI.)**.** Given a nonempty set $\mathcal{X} \subseteq \mathbb{R}^d$ and an operator $F\colon \mathcal{X} \to \mathbb{R}^d$, the Minty variational inequality (MVI) problem $P_{\mathrm{MVI}}(F, \mathcal{X})$ is to

$$\text{find } x^\star \in \mathcal{X} \text{ such that } \left\langle F(x),\, x - x^\star \right\rangle \geq 0, \text{ for all } x \in \mathcal{X}. \tag{2}$$

Here, $\langle \cdot, \cdot \rangle$ denotes the standard inner product in $\mathbb{R}^d$. Any $x^\star \in \mathcal{X}$ satisfying (2) is called a solution of the MVI problem $P_{\mathrm{MVI}}(F, \mathcal{X})$.

Observe that the evaluation point of the operator F is inverted in (1) and (2). In particular, in (1) F is evaluated at *the unknown solution* $x^\star$, whereas in (2) it is evaluated at *the test point* x. This has important algorithmic implications, as we will see in subsequent sections. In particular, the merit functions associated with these two problems are different, which leads to different convergence analyses for algorithms designed to solve them. However, before proceeding, we introduce a couple of basic properties of the solutions to the MVI problem and those of the SVI problem. The proofs of these results are provided for completeness; a mre extensive treatment can be found in Facchinei et al. [5].

Relationship between SVI and MVI. Let us first observe that under mild assumptions, any solution of the MVI problem is also a solution of the SVI problem. Formally, we have the following theorem.

Theorem 2. *Suppose $\mathcal{X} \subseteq \mathbb{R}^d$ is a nonempty convex set and $F\colon \mathcal{X} \to \mathbb{R}^d$ is a continuous operator. Then any solution of the MVI problem $P_{MVI}(F, \mathcal{X})$ is also a solution of the SVI problem $P_{SVI}(F, \mathcal{X})$.*

Proof. Suppose $x^\star \in \mathcal{X}$ is a solution of $P_{\mathrm{MVI}}(F, \mathcal{X})$, i.e., it satisfies (2).

Fix any arbitrary $x \in \mathcal{X}$. For each $t \in [0, 1]$, define the point

$$x_t = x^\star + t(x - x^\star) \in \mathcal{X}. \tag{3}$$

Since $\mathcal{X}$ is convex, it follows that $x_t \in \mathcal{X}$ for every $t \in [0, 1]$. Thus, by (2), we have that

$$\langle F(x_t),\, x_t - x^\star \rangle \geq 0 \implies \langle F(x_t),\, t(x - x^\star) \rangle \geq 0 \implies \langle F(x_t),\, x - x^\star \rangle \geq 0, \qquad \forall t \in [0, 1]. \tag{4}$$

Furthermore, since F is continuous, we have that

$$\lim_{t \to 0^+} \langle F(x_t),\, x - x^\star \rangle = \langle F(\lim_{t\to 0^+} x_t),\, x - x^\star \rangle = \langle F(x^\star),\, x - x^\star \rangle. \tag{5}$$

Thus, by taking the limit as $t \to 0^+$ on both sides of the previous inequality, we obtain

$$\langle F(x^\star),\, x - x^\star \rangle \geq 0. \tag{6}$$

Finally, since $x \in \mathcal{X}$ was chosen arbitrarily, we conclude that $x^\star$ satisfies (1), and therefore, is a solution of $P_{\mathrm{SVI}}(F, \mathcal{X})$. □

The converse of Theorem 2 does not hold in general. However, under the assumption that F is monotone, we have the following result.

Theorem 3. *Suppose $F \colon \mathcal{X} \to \mathbb{R}^d$ is a monotone operator on a nonempty set $\mathcal{X} \subseteq \mathbb{R}^d$. Then any solution of the SVI problem $P_{SVI}(F, \mathcal{X})$ is also a solution of the MVI problem $P_{MVI}(F, \mathcal{X})$.*

Proof. Suppose $x^\star \in \mathcal{X}$ is a solution of $P_{\mathrm{SVI}}(F, \mathcal{X})$, i.e., it satisfies (1).

Recall that a monotone operator $F \colon \mathcal{X} \to \mathbb{R}^d$ saitisfies

$$\langle F(x) - F(y),\, x - y \rangle \geq 0, \qquad \forall x, y \in \mathcal{X}. \tag{7}$$

Thus, by setting $y = x^\star$, it follows that

$$\langle F(x) - F(x^\star),\, x - x^\star \rangle \geq 0 \implies \langle F(x),\, x - x^\star \rangle \geq \langle F(x^\star),\, x - x^\star \rangle \geq 0, \qquad \forall x \in \mathcal{X}. \tag{8}$$

where the last inequality follows from (1). Hence, $x^\star$ also satisfies (2), and is therefore a solution of the MVI problem $P_{\mathrm{MVI}}(F, \mathcal{X})$. □

In summary, we have established the following important relationship between SVIs and MVIs.

Corollary 4. *Suppose $\mathcal{X} \subseteq \mathbb{R}^d$ is a nonempty convex set and $F \colon \mathcal{X} \to \mathbb{R}^d$ is a continuous monotone operator. Then the solutions of the SVI problem $P_{SVI}(F, \mathcal{X})$ and the MVI problem $P_{MVI}(F, \mathcal{X})$ are identical.*

1.2 Merit functions for VI problems

To analyze the convergence of algorithms for solving VI problems, it is useful to introduce appropriate merit functions that quantify how close a given point is to being a solution to the corresponding variational inequality problems. In particular, for the SVI problem $P_{\mathrm{SVI}}(F, \mathcal{X})$, we define the Stampacchia gap function,

$$G_S(y) = \sup_{x \in \mathcal{X}} \langle F(y),\, y - x \rangle, \qquad \forall y \in \mathcal{X}. \tag{G_S}$$

Similarly, for the MVI problem $P_{\mathrm{MVI}}(F, \mathcal{X})$, we define the Minty gap function,

$$G_M(y) = \sup_{x \in \mathcal{X}} \langle F(x),\, y - x \rangle, \qquad \forall y \in \mathcal{X}. \tag{G_M}$$

Note that if $x = y$, we have that $\langle F(y),\, y - x \rangle = \langle F(x),\, y - x \rangle = 0$. Thus, $G_S(y), G_M(y) \geq 0$ for all $y \in \mathcal{X}$. Moreover, by (1), $y^\star \in \mathcal{X}$ is a solution of the SVI problem $P_{\mathrm{SVI}}(F, \mathcal{X})$ if and only if $G_S(y^\star) \leq 0$; thus, $G_S(y^\star) = 0$. Similarly, by (2), $y^\star$ is a solution of the MVI problem $P_{\mathrm{MVI}}(F, \mathcal{X})$ if and only if $G_M(y^\star) = 0$.

Now observe that if F is monotone, then, by definition, we have that

$$\langle F(x) - F(y),\, x - y \rangle \geq 0 \implies \langle F(x),\, y - x \rangle \leq \langle F(y),\, y - x \rangle \implies G_M(y) \leq G_S(y), \qquad \forall x, y \in \mathcal{X}. \tag{9}$$

Thus, for monotone operators, the MVI merit function lower bounds the SVI merit function. In addition, if $\mathcal{X}$ is convex, then by Corollary 4, the solutions of the two problems coincide. Therefore, for monotone operators on convex sets, both merit functions can be used to quantify the quality of approximate solutions to either VI problem, with G_M *providing a potentially tighter measure* than G_S.

2 Proximal methods

Having established the relationship between Stampacchia VI and Minty VI problems, we now study algorithms for solving them. We begin our discussion with the Proximal Point Algorithm (PPA), a conceptual benchmark against which we are going to compare more practical algorithms in later sections.

2.1 The resolvent operator

Consider a convex set $\mathcal{X} \subseteq \mathbb{R}^d$ and *a strongly monotone operator* $F\colon \mathcal{X} \to \mathbb{R}^d$. Recall that, since $\mathcal{X}$ is convex, $x^\star \in \mathcal{X}$ is a solution of the Stampacchia VI problem $P_{\mathrm{SVI}}(F, \mathcal{X})$ if and only if

$$-F(x^\star) \in \mathcal{N}_{\mathcal{X}}(x^\star) \iff 0 \in (F + \mathcal{N}_{\mathcal{X}})(x^\star), \tag{10}$$

where $\mathcal{N}_{\mathcal{X}}(x^\star)$ denotes the normal cone of $\mathcal{X}$ at the point $x^\star$; i.e.,

$$\mathcal{N}_{\mathcal{X}}(x^\star) = \big\{y \in \mathbb{R}^d \;\big|\; \langle y,\, x - x^\star\rangle \le 0,\, \forall x \in \mathcal{X}\big\}. \tag{normal cone}$$

Observe that, by the $\mathcal{N}$'s definition, we have that $0 \in \mathcal{N}_{\mathcal{X}}(x)$ for all $x \in \mathcal{X}$.

Define the operator $T\colon \mathcal{X} \to 2^{\mathbb{R}^d}$ as the set-valued operator given by

$$T(x) = F(x) + \mathcal{N}_{\mathcal{X}}(x), \qquad \forall x \in \mathcal{X}. \tag{11}$$

Then, by (10), solving the Stampacchia VI problem $P_{\mathrm{SVI}}(F, \mathcal{X})$ is equivalent to finding a point $x^\star \in \mathcal{X}$ such that $0 \in T(x^\star)$. Observe that, since F is a strongly monotone operator, it follows that the operator T defined in (11) is also strongly monotone. Indeed, if F is μ-strongly monotone for some $\mu > 0$, then for all $x, y \in \mathcal{X}$, and for all $u \in T(x)$ and $v \in T(y)$, we have that exist $z_x \in \mathcal{N}_{\mathcal{X}}(x)$ and $z_y \in \mathcal{N}_{\mathcal{X}}(y)$ such that $u = F(x) + z_x$ and $v = F(y) + z_y$. Thus, since F is μ-strongly monotone, we have that

$$\langle u - v,\, x - y\rangle = \big\langle F(x) - F(y),\, x - y\big\rangle + \underbrace{\langle z_x - z_y,\, x - y\rangle}_{=0} \ge \mu \cdot \|x - y\|_2^2, \tag{12}$$

and therefore, T is also μ-strongly monotone.

2.2 The Proximal Point Algorithm

Given $\lambda > 0$, define the resolvent $J_\lambda\colon \mathbb{R}^d \to \mathcal{X}$ as

$$J_{\lambda T}(y) = (\mathbf{I} + \lambda T)^{-1}(y), \qquad \forall y \in \mathbb{R}^d. \tag{13}$$

Note that, for general operators T, the resolvent $J_{\lambda T}$ is a set-valued mapping. However, since T is strongly monotone, $J_{\lambda T}(y)$ is single-valued for all $y \in \mathbb{R}^d$ [6]; thus, it is well defined.

The PPA generates a sequence $\{x_k\}_{k\in\mathbb{N}}$ according to the following recursion:

$$x_{k+1} = J_{\lambda T}(x_k), \qquad \forall k \in \mathbb{N}. \tag{14}$$

However, since $J_{\lambda T}(x_k)$ involves computing the preimage of x_k under the mapping $\mathbf{I} + \lambda T$, each iteration of the PPA *may be computationally expensive*. Nevertheless, the PPA enjoys strong convergence guarantees when F is a strongly monotone.

Theorem 5 (Rockafellar [6]). *Suppose $\mathcal{X} \subseteq \mathbb{R}^d$ is a nonempty, compact, and convex set and $F\colon \mathcal{X} \to \mathbb{R}^d$ is a continuous and strongly monotone operator. Then for any initial point $x_0 \in \mathbb{R}^d$, the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by the PPA converges to the unique solution of the SVI problem $P_{SVI}(F, \mathcal{X})$ at a linear rate.*

Proof. Suppose F is μ-strongly monotone for some $\mu > 0$. Then, for all $y \in \mathbb{R}^d$, the resolvent $J_{\lambda T}(y)$ is single-valued. Thus, by (13), we have that

$$J_{\lambda T}(y) = x \implies y \in x + \lambda T(x) \forall x \in \mathcal{X},\ y \in \mathbb{R}^d. \tag{15}$$

Furthermore, since $\mathcal{X}$ is nonempty, compact, and convex, and F is continuous and strongly monotone, the Stampacchia VI problem $P_{\mathrm{SVI}}(F, \mathcal{X})$ has a unique solution $x^\star \in \mathcal{X}$. Then, by (10), we have that

$$0 \in T(x^\star) \implies x^\star \in x^\star + \lambda T(x^\star). \implies x^\star = J_{\lambda T}(x^\star). \tag{16}$$

Fix $k \in \mathbb{N}$ to be arbitrary. Then, by (15), there exists $y \in T(x_k)$ such that

$$x_k = x_{k+1} + \lambda y, \overset{\lambda > 0}{\implies} y = \frac{1}{\lambda}(x_k - x_{k+1}). \tag{17}$$

Thus, since T is μ-strongly monotone, we have that

$$\begin{aligned} \mu \cdot \|x_{k+1} - x^\star\|_2^2 &\overset{0 \in T(x^\star)}{\leq} \langle y,\, x_{k+1} - x^\star \rangle && \text{(18a)} \\ &= \frac{1}{\lambda}\langle x_k - x_{k+1},\, x_{k+1} - x^\star \rangle && \text{(18b)} \\ &= \frac{1}{\lambda}\langle x_k - x^\star,\, x_{k+1} - x^\star \rangle - \frac{1}{\lambda}\|x_{k+1} - x^\star\|_2^2, && \text{(18c)} \end{aligned}$$

which implies that

$$(1 + \lambda\mu) \cdot \|x_{k+1} - x^\star\|_2^2 \leq \langle x_k - x^\star,\, x_{k+1} - x^\star \rangle \leq \|x_k - x^\star\|_2 \cdot \|x_{k+1} - x^\star\|_2. \tag{19}$$

where the last inequality follows from the Cauchy–Schwarz inequality.

If $x_{k+1} = x^\star$, then the sequence $\{x_k\}_{k\in\mathbb{N}}$ has converged to the solution of the Stampacchia VI problem $P_{\mathrm{SVI}}(F, \mathcal{X})$. If $x_{k+1} \neq x^\star$, we can divide both sides by $(1 + \lambda\mu) \cdot \|x_{k+1} - x^\star\|_2$ to obtain

$$\|x_{k+1} - x^\star\|_2 \geq \frac{1}{1 + \lambda\mu} \cdot \|x_k - x^\star\|_2 \geq \cdots \geq \frac{1}{(1 + \lambda\mu)^{k+1}} \cdot \|x_0 - x^\star\|_2. \tag{20}$$

□

Although the PPA is not a practical algorithm due to the computational cost of each iteration, it serves as a useful conceptual benchmark for more practical algorithms that we are going to discuss in the next sections. In particular, we are going to see that Projected Gradient Descent (PGD) can match (asymptotically) the linear convergence rate of the PPA when F is strongly monotone and Lipschitz continuous. However, it fails to converge when F is only monotone, whereas the PPA still converges in this case (albeit at a sublinear rate). Finally, we are going to see that the Extra-Gradient Method (EGM) converges at a sublinear rate when F is only monotone, asymptotically matching the convergence rate of the PPA in this case.

3 The Projected Gradient Descent

We have already discussed the PPA, which is a conceptual algorithm for solving SVI problems. Let us now discuss a more practical algorithm for solving SVI problems, namely PGD. Recall that PGD is an iterative method that, given an initial point $x_0 \in \mathcal{X}$ and a step-size $\alpha > 0$, generates a sequence $\{x_k\}_{k\in\mathbb{N}}$ according to the recursion

$$x_{k+1} = \Pi_{\mathcal{X}}\big(x_k - \alpha F(x_k)\big) \tag{21}$$

In contrast to the PPA, to prove the convergence of PGD we need to impose stronger assumptions on the operator F. In particular, we need to assume that F is Lipschitz continuous. Its rate of convergence depends on both the Lipschitz constant and the strong monotonicity parameter of F. Formally, we have the following theorem.

Theorem 6 (Zarantonello [7]). *Suppose $\mathcal{X} \subseteq \mathbb{R}^d$ is a nonempty, compact, and convex set, and $F\colon \mathcal{X} \to \mathbb{R}^d$ is an L-Lipschitz continuous and μ-strongly monotone operator for some $L, \mu > 0$. Then for any initial point $x_0 \in \mathcal{X}$, the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by the PGD with step-size $\alpha \in [0,\, 2\mu/L^2]$ converges to the unique solution of the SVI problem $P_{SVI}(F, \mathcal{X})$ at a linear rate.*

Proof. Fix $y \in \mathbb{R}^d$ to be arbitrary. Then, since $\mathcal{X}$ is a compact and convex set, and the function $f(x) = \frac{1}{2}\|x-y\|_2^2$ is strongly convex, the projection $\Pi_{\mathcal{X}}(y)$ is a Karush–Kuhn–Tucker (KKT) point for $\min_{x\in\mathcal{X}} f(x)$; that is, a point $z \in \mathcal{X}$ such that

$$\langle z-y,\, x-z\rangle \geq 0, \quad \forall x \in \mathcal{X}. \tag{22}$$

Since F is strongly monotone, the SVI problem $P_{\mathrm{SVI}}(F, \mathcal{X})$ has a unique solution $x^\star \in \mathcal{X}$ that satisfies (1). We show that $x^\star = \Pi_{\mathcal{X}}(x^\star - \alpha F(x^\star))$, i.e., (22) holds for $z = x^\star$ and $y = x^\star - \alpha F(x^\star)$.

$$\langle z-y,\, x-z\rangle = \big\langle x^\star - \big(x^\star - \alpha F(x^\star)\big),\, x - x^\star\big\rangle = \alpha\langle F(x^\star),\, x - x^\star\rangle \geq 0, \tag{23}$$

where the last inequality follows from (1). Thus, $x^\star$ is a fixed point of the PGD update.

Next, we show that the sequence $\{x_k\}_{k\in\mathbb{N}}$ converges to $x^\star$ at a linear rate. Since $\mathcal{X}$ is a compact, convex set, the projection operator $\Pi_{\mathcal{X}}$ is 1-Lipschitz continuous. This is intuitive, but it can also be formally show this using (22). Thus, for any $k \in \mathbb{N}$, we have that

$$\|x_{k+1} - x^\star\|_2 = \big\|\Pi_{\mathcal{X}}\big(x_k - \alpha F(x_k)\big) - \Pi_{\mathcal{X}}\big(x^\star - \alpha F(x^\star)\big)\big\|_2 \leq \big\|\big(x_k - \alpha F(x_k)\big) - \big(x^\star - \alpha F(x^\star)\big)\big\|_2. \tag{24}$$

Then by squaring both sides, we obtain

$$\|x_{k+1} - x^\star\|_2^2 \leq \|x_k - x^\star\|_2^2 - 2\alpha\langle F(x_k) - F(x^\star),\, x_k - x^\star\rangle + \alpha^2\|F(x_k) - F(x^\star)\|_2^2. \tag{25}$$

Suppose that F is L-Lipschitz continuous and μ-strongly monotone for some $L, \mu > 0$. Then, by the definitions of Lipschitz continuity and strong monotonicity, we have that

$$\|F(x_k) - F(x^\star)\|_2^2 \leq L^2 \cdot \|x_k - x^\star\|_2^2, \quad \text{and} \quad \langle F(x_k) - F(x^\star),\, x_k - x^\star\rangle \geq \mu \cdot \|x_k - x^\star\|_2^2. \tag{26}$$

Thus, substituting these inequalities into the previous one, we obtain

$$\|x_{k+1}-x^\star\|_2^2 \leq \big(1-2\alpha\mu+\alpha^2L^2\big)\cdot\|x_k-x^\star\|_2^2 \leq \cdots \leq \big(1-2\alpha\mu+\alpha^2L^2\big)^{k+1}\cdot\|x_0-x^\star\|_2^2. = q^{k+1}\cdot\|x_0-x^\star\|_2^2, \tag{27}$$

where we defined $q = 1 - 2\alpha\mu + \alpha^2L^2$. Finally, observe that for any step-size $\alpha \in [0,\, 2\mu/L^2]$, we have that $q \in [0,\, 1]$, which establishes the desired linear convergence rate. □

So we have established that PGD converges linearly to the unique solution of the SVI problem $P_{\mathrm{SVI}}(F, \mathcal{X})$ when F is Lipschitz continuous and strongly monotone. However, unlike the PPA, PGD is a practical algorithm since each iteration only requires evaluating the operator F once and computing a projection onto the set $\mathcal{X}$.

4 The Extra-Gradient Method

Let us now focus on the same setting as in Theorem 6, i.e., suppose $\mathcal{X} \subseteq \mathbb{R}^d$ is a nonempty, compact, and convex set, and let $F\colon \mathcal{X} \to \mathbb{R}^d$ be an L-Lipschitz continuous operator for some $L > 0$. However, instead of assuming that F is strongly monotone, we are only going to assume that it is monotone.

In this case, the PGD method may fail to converge to a solution of the SVI problem $P_{\mathrm{SVI}}(F, \mathcal{X})$. In particular, it is possible to construct simple examples of monotone and Lipschitz continuous operators F for which the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by the PGD method oscillates and does not converge. To address this issue, we can use EGM, which is an iterative method that, given an initial point $x_0 \in \mathcal{X}$ and a step-size $\alpha > 0$, generates sequences $\{()x_k)_{k\in\mathbb{N}}$ and $\{()y_k)_{k\in\mathbb{N}}$ according to the recursion

$$y_k = \Pi_{\mathcal{X}}\big(x_k - \alpha F(x_k)\big), \tag{28a}$$
$$x_{k+1} = \Pi_{\mathcal{X}}\big(x_k - \alpha F(y_k)\big). \tag{28b}$$

It is possible to show the following intermediate bound for the sequences generated by the EGM method.

Lemma 7. *Suppose $\mathcal{X} \subseteq \mathbb{R}^d$ is a nonempty, compact, and convex set, and $F\colon \mathcal{X} \to \mathbb{R}^d$ is an L-Lipschitz continuous and monotone operator for some $L > 0$. Then for any initial point $x_0 \in \mathcal{X}$, the sequences $(x_k)_{k\in\mathbb{N}}$ and $(y_k)_{k\in\mathbb{N}}$ generated by the EGM with step-size $\alpha = 1/L$ satisfy*

$$\langle F(y_k),\, y_k - x\rangle \le \frac{1}{2\alpha}\cdot\left(\|x_k - x\|_2^2 - \|x_{k+1} - x\|_2^2\right), \qquad \forall x \in \mathcal{X},\ \forall k \in \mathbb{N}. \tag{29}$$

The proof of this result can be found in Nemirovski [8]. Using Lemma 7, we can now establish the convergence of the EGM in the monotone case.

Theorem 8 (Nemirovski [8])**.** *Suppose $\mathcal{X} \subseteq \mathbb{R}^d$ is a nonempty, compact, and convex set, and $F\colon \mathcal{X} \to \mathbb{R}^d$ is a L-Lipschitz continuous and monotone operator for some $L > 0$. Then for any initial point $x_0 \in \mathcal{X}$, the sequence $\{x_k\}_{k\in\mathbb{N}}$ generated by the EGM with step-size $\alpha = 1/L$ converges on average to a solution of the SVI problem $P_{SVI}(F, \mathcal{X})$ at a rate of $\mathcal{O}(1/k)$.*

Proof. Fix $K \in \mathbb{N}$ to be arbitrary. Then, by summing both sides of (29) from $k = 0$ to $k = K - 1$, we have that

$$\sum_{k=0}^{K-1}\langle F(y_k),\, y_k - x\rangle \le \frac{1}{2\alpha}\cdot\left(\|x_0 - x\|_2^2 - \|x_K - x\|_2^2\right) \le \frac{1}{2\alpha}\cdot\|x_0 - x\|_2^2, \qquad \forall x \in \mathcal{X}. \tag{30}$$

Since $\mathcal{X}$ is compact, there exists $R > 0$ such that $\|x_0 - x\|_2 \le R$ for all $x \in \mathcal{X}$. Then, by dividing both sides by K, we have that

$$\frac{1}{K}\sum_{k=0}^{K-1}\langle F(y_k),\, y_k - x\rangle \le \frac{R^2}{2\alpha K}. \tag{31}$$

By the monotonicity of F, we have that

$$\langle F(x) - F(y_k),\, x - y_k\rangle \ge 0, \implies \langle F(y_k),\, y_k - x\rangle \ge \langle F(x),\, y_k - x\rangle \qquad \forall x \in \mathcal{X},\ \forall k \in \mathbb{N}. \tag{32}$$

Define the average iterate

$$\bar{y}_K = \frac{1}{K}\sum_{k=0}^{K-1} y_k. \tag{33}$$

Then, by the convexity of the set $\mathcal{X}$, we have that $\bar{y}_K \in \mathcal{X}$, and therefore, by the linearity of the inner product, we have that

$$\langle F(x),\, \bar{y}_K - x\rangle = \frac{1}{K}\sum_{k=0}^{K-1}\langle F(x),\, y_k - x\rangle \le \frac{1}{K}\sum_{k=0}^{K-1}\langle F(y_k),\, y_k - x\rangle \le \frac{R^2}{2\alpha K}, \qquad \forall x \in \mathcal{X}. \tag{34}$$

Thus, we have established that the Minty gap function of the average iterate $\bar{y}_K$ satisfies

$$G_{M\mathcal{X}}(\bar{y}_K) = \operatorname{supp}_{x\in\mathcal{X}}\langle F(x),\, \bar{y}_K - x\rangle \le \frac{R^2}{2\alpha K} = \mathcal{O}\left(\frac{1}{K}\right). \tag{35}$$

□

References

[1] Gérard Debreu. "A Social Equilibrium Existence Theorem." In: *Proceedings of the National Academy of Sciences* 38.10 (1952), pp. 886–893. ISSN: 1557-7317. DOI: 10.1073/pnas.38.10.886.

[2] Judah Ben Rosen. "Existence and Uniqueness of Equilibrium Points for Concave N-Person Games." In: *Econometrica* 33.3 (1965), pp. 520–534. ISSN: 1468-0262. DOI: 10.2307/1911749.

[3] Alain Bensoussan. "Points de Nash dans le cas de fonctionnelles quadratiques et jeux différentiels linéaires à N personnes." In: *SIAM Journal on Control and Optimization* 12 (1974), pp. 460–499. ISSN: 1095-7138. DOI: 10.1137/0312037.

[4] George J. Minty. "Monotone (Nonlinear) Operators in Hilbert Space." In: *Duke Mathematical Journal* 29.3 (1962), pp. 341–346.

[5] Francisco Facchinei and Jong-Shi Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, 2003.

[6] R. Tyrrell Rockafellar. "Monotone Operators and the Proximal Point Algorithm." In: *SIAM Journal on Control and Optimization* 14.5 (1976), pp. 877–898.

[7] E. H. Zarantonello. *Solving Functional Equations by Contractive Averaging*. MRC Technical Summary Report 160. Madison, Wisconsin: Mathematics Research Center, United States Army, The University of Wisconsin, June 1960.

[8] Arkadi 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." In: *SIAM Journal on Optimization* 15.1 (2004), pp. 229–251.

Index