Stochastic Variational Inequalities

Presenter: Peijun Xiao

Department of Industrial and Enterprise Systems Engineering (ISE)
University of Illinois at Urbana-Champaign (UIUC)

IE598 Fall 2020 Course Project

April 2, 2020
Overview

Introduction to Variational Inequalities

Weakly-Convex-Weakly-Concave Saddle-Point Problems

Stochastic Variational Inequalities for Multi-stages

Stochastic Mirror Prox
Outline

Introduction to Variational Inequalities

Weakly-Convex-Weakly-Concave Saddle-Point Problems

Stochastic Variational Inequalities for Multi-stages

Stochastic Mirror Prox
Start with Optimality Conditions

- Consider minimizing a smooth convex function \( f(x) \) over a compact set \( C \),

\[
\min_{x \in C} f(x).
\]

- The feasible set at \( x^* \) is \( S_{\text{feasible}}(x^*) = \{ s \in \mathbb{R}^n \mid s := x' - x^*, x' \in C \} \)

- The descent set at \( x^* \) is \( S_{\text{descent}}(x^*) = \{ s \in \mathbb{R}^n \mid \langle s, \nabla f(x^*) \rangle < 0 \} \)

- A point \( x^* \) is optimal if \( x^* \in C \) and \( S_{\text{feasible}}(x^*) \cap S_{\text{descent}}(x^*) = \emptyset \), i.e.

\[
\langle \nabla f(x^*), x' - x^* \rangle \geq 0, \quad \forall x' \in C
\]
Start with Optimality Conditions

Consider minimizing a smooth convex function $f(x)$ over a compact set $C$,

$$\min_{x \in C} f(x).$$

- The feasible set at $x^*$ is $S_{feasible}(x^*) = \{ s \in \mathbb{R}^n \mid s := x' - x^*, x' \in C \}$
- The descent set at $x^*$ is $S_{descent}(x^*) = \{ s \in \mathbb{R}^n \mid \langle s, \nabla f(x^*) \rangle < 0 \}$
- A point $x^*$ is optimal if $x^* \in C$ and $S_{feasible}(x^*) \cap S_{descent}(x^*) = \emptyset$, i.e.

$$\langle \nabla f(x^*), x' - x^* \rangle \geq 0, \ \forall x' \in C$$

- This is equivalent to say

$$\nabla f(x^*) \in N_C(x^*) = \{ h \in \mathbb{R}^n \mid h^T (x' - x^*) \geq 0, \forall x' \in C \}. \quad (1)$$
Definition of Variational Inequalities (VI)

- Consider a **general** case: we replace $\nabla f(x^*)$ by a mapping $F(x^*)$, where $F(x) : \mathbb{R}^d \to \mathbb{R}^d$ is a set-valued mapping, and $C \subset \mathbb{R}^d$ is a compact set.

- A **variational inequality (VI)** condition at $x^*$ with respect to a mapping $F(x^*)$ over a compact set $C$ is

  $$F(x^*) \in N_C(x^*) = \left\{ h \in \mathbb{R}^n : h^T(x' - x^*) \geq 0, \forall x' \in C \right\}$$  

  \hspace{1cm}(2)

- A VI Problem is to find an $x^*$ which satisfies eq. (2)

- Finding the optimal solution $x^*$ is a **special case** of a VI problem, by letting $F(x^*) = \nabla f(x^*)$. 

KKT Condition is Also a VI

▶ Suppose $C = \{ x \in X : G(x) := (g_1(x), \cdots, g_m(x)) \in K \}$

▶ $X$ - closed convex set, $K$ - closed convex cone, $g_i$ - continuously differentiable.

▶ The Lagrangian function

$$L(x, y) = f(x) + \sum_{i=1}^{m} y_ig_i(x) \quad (3)$$

where $y = (y_1, \cdots, y_m) \in Y = K^*$ ($K^*$ is the dual cone of $K$) with $G(x) \in N_Y(y)$.

▶ One can show the KKT conditions can also be represented as a VI

$$\nabla_x L(x, y) \in N_X(x), \quad -\nabla_y L(x, y) \in N_C(x) \quad (4)$$

▶ Or $H(z) \in N_Z(z)$ for $z = (x, y)$, $Z = X \times Y$, $H(z) = (\nabla_x L(x, y), -\nabla_y L(x, y))$. 

Introduction to Variational Inequalities
History of Variational Inequalities

- In early 1960s, Stampacchia and his collaborators used **VI as an analytic tool** for studying free boundary problems.
- At the same time, Cottle (student of Dantzig) introduced **nonlinear complementarity problem (NCP)**. Lemke and Howson solved a **bimatrix game** formulated as a linear complementarity problem.
History of Variational Inequalities

▶ In early 1960s, Stampacchia and his collaborators used \textit{VI as an analytic tool} for studying free boundary problems.

▶ At the same time, Cottle (student of Dantzig) introduced \textit{nonlinear complementarity problem} (NCP). Lemke and Howson solved a \textit{bimatrix game} formulated as a linear complementarity problem.

▶ In 1967, Scarf developed the first constructive iterative method for \textit{approximating a fixed point} of a continuous mapping. The field of \textit{equilibrium programming} was thus born.

▶ At this infant stage of linear programming, the \textit{primal-dual relation} provided clear evidence of the interplay between complementarity and equilibrium.
History of Variational Inequalities

- In 1970s, Karamardian provided extensive **existence theory for the NCP and its cone generalization** and made clear the connection between NCP and the VI. This is the beginning of contemporary chapter of finite-dimensional VI.

- **Problems from fields like physics and economics were posted in VI forms.** A few algorithms like Newton’s method were proposed at that time.
History of Variational Inequalities

▶ In 1970s, Karamardian provided extensive existence theory for the NCP and its cone generalization and made clear the connection between NCP and the VI. This is the beginning of contemporary chapter of finite-dimensional VI.

▶ Problems from fields like physics and economics were posted in VI forms. A few algorithms like Newton’s method were proposed at that time.

▶ The convergence analysis of Newton’s method influenced the modern development of sensitivity analysis of mathematical programs.

▶ In 1990s, many developments such as piece-wise smooth functions, error bounds, interior point methods, smoothing methods, methods of the projection family, and regularization were also closely related to VI.

▶ Reference: (Facchinei and Pang, 2007)
This Talk: Focus on Stochastic Programming

We can use VI as a tool to

- define optimality conditions;
- design efficient algorithms;
- show convergence of algorithms.

We focus on three settings in Stochastic Programming:

1. Nonconvex nonconcave saddle-point problems (Lin et al., 2018)
2. Multi-stage stochastic optimization (Rockafellar and Wets, 2017)
3. Stochastic Mirror-Prox (Juditsky, Nemirovski, and Tauvel, 2011)
Outline

Introduction to Variational Inequalities

Weakly-Convex-Weakly-Concave Saddle-Point Problems

Stochastic Variation Inequalities for Multi-stages

Stochastic Mirror Prox
First work that establishes the non-asymptotic convergence to a nearly stationary point of a non-convex non-concave min-max problem

<table>
<thead>
<tr>
<th>Setting</th>
<th>Algorithms for sub-problems</th>
<th>Lipchitz</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stochastic</td>
<td>Stochastic Subgradient Method</td>
<td>No</td>
<td>$O \left( \frac{1}{\epsilon^6} \right)$</td>
</tr>
<tr>
<td>Deterministic</td>
<td>Subgradient Method</td>
<td>No</td>
<td>$O \left( \frac{1}{\epsilon^6} \right)$</td>
</tr>
<tr>
<td></td>
<td>Gradient Descent Method</td>
<td>Yes</td>
<td>$\tilde{O} \left( \frac{L^2}{\epsilon^2} \right)$</td>
</tr>
<tr>
<td></td>
<td>Nesterov’s Accelerated Method</td>
<td>Yes</td>
<td>$\tilde{O} \left( \frac{L\rho}{\epsilon^2} \right)$</td>
</tr>
<tr>
<td>Finite-sum</td>
<td>Variance Reduction</td>
<td>Yes</td>
<td>$\tilde{O} \left( \frac{n\rho^2}{\epsilon^2} + \frac{L^2}{\epsilon^2} \right)$</td>
</tr>
<tr>
<td>(with $n$ components)</td>
<td>Accelerated Variance Reduction</td>
<td>Yes</td>
<td>$\tilde{O} \left( \frac{n\rho^2}{\epsilon^2} + L\rho/\epsilon^2 \right)$</td>
</tr>
</tbody>
</table>
Things That We Need to Address

▶ What is a general min-max problem? Optimality condition?
▶ What kinds of non-convex non-concave functions?
▶ How to relate non-convex non-concave min-max problem problem to VI?
▶ How to design algorithm for non-convex non-concave problems?
▶ Finally, how to use VI to give convergence analysis of the algorithm?
Saddle Point Problem

\[
\min_{x \in X} \max_{y \in Y} f(x, y)
\]  
$(5)$

- $X, Y$ are closed convex sets
- $f(x, y)$ is non-convex w.r.t $x$, and non-concave w.r.t $y$
- Saddle point $(x^*_*, y^*_*)$

\[
f(x^*_*, y^*_*) \leq f(x^*, y^*_*) \leq f(x, y^*_*) , \quad \forall x \in X, y \in Y.
\]  
$(6)$

- First-order stationary solution $F^*_*$

\[
F^*_* = \{(x, y) \mid 0 \in \partial_x [f(x, y) + 1_X(x)], 0 \in \partial_y [f(x, y) + 1_Y(y)]\}
\]  
$(7)$

- Applications: GAN, robust training, game theory, ...

Weakly-Convex-Weakly-Concave Saddle-Point Problems
Weakly-Convex-Weakly-Concave Saddle-point Problem

- $f(x)$ is strongly convex if $\alpha$-strongly convex if
  - $f(x) - \frac{\alpha}{2} \|x\|^2$ is convex

- $f(x, y)$ is $\rho$-weakly-convex-weakly-concave, if
  - $f(x, y) + \frac{\rho}{2} \|x\|^2$ is convex in terms of $x$
  - $f(x, y) - \frac{\rho}{2} \|y\|^2$ is concave in terms of $y$

- We say a saddle point problem is $\rho$-weakly-convex-weakly-concave if the $f(x, y)$ is $\rho$-weakly-convex-weakly-concave.
Strongly or Weakly Monotone

A set-valued mapping $F(z) : \mathbb{R}^d \to \mathbb{R}^d$ is

- monotone
  \[ \langle \xi - \xi', z - z' \rangle \geq 0; \quad (8) \]
  \[ \forall z, z' \in Z, \forall \xi \in F(z), \forall \xi' \in F(z'). \quad (9) \]
  
  - The sub-differential of a convex function $f$, $F(x) := \partial f(x)$ is monotone.

Weakly-Convex-Weakly-Concave Saddle-Point Problems
Strongly or Weakly Monotone

A set-valued mapping \( F(z) : \mathbb{R}^d \to \mathbb{R}^d \) is

- monotone
  \[ \langle \xi - \xi', z - z' \rangle \geq 0; \]  
  \( \forall z, z' \in Z, \forall \xi \in F(z), \forall \xi' \in F(z') \). \( \tag{8} \)

  - The sub-differential of a convex function \( f \), \( F(x) := \partial f(x) \) is monotone.

- \( \mu \)-strongly monotone
  \[ \langle \xi - \xi', z - z' \rangle \geq \mu \| z - z' \|^2; \]  
  \( \tag{10} \)

- \( \rho \)-weakly monotone
  \[ \langle \xi - \xi', z - z' \rangle \geq -\rho \| z - z' \|^2. \]  
  \( \tag{11} \)
Recall, a VI is defined on a set $Z$ and a set-valued mapping $F(z) : \mathbb{R}^d \rightarrow \mathbb{R}^d$.

A VI problem is to find a $z^* \in Z$, such that

$$
\exists \ \xi^* \in F(z^*) \ \text{s.t.} \ \langle \xi^*, z - z^* \rangle \geq 0, \ \forall z \in Z
$$

In this paper, we let $Z = X \times Y$ is compact (bounded by $D$).

$F(z) = (\partial_x f(x,y), \partial_y [-f(x,y)])$.

We assume $f(x,y)$ is $\rho$-weakly-convex-weakly-concave.

Lemma $f(x,y)$ is $\rho$-weakly-convex-weakly-concave if and only if $F(z)$ is $\rho$-weakly monotone.
Recall, a VI is defined on a set $Z$ and a set-valued mapping $F(z) : \mathbb{R}^d \to \mathbb{R}^d$.

A VI problem is to find a $z^* \in Z$, such that

$$\exists \, \xi^* \in F(z^*) \text{ s.t. } \langle \xi^*, z - z^* \rangle \geq 0, \quad \forall z \in Z$$

In this paper, we let $Z = X \times Y$ is compact (bounded by $D$)

$F(z) = (\partial_x f(x, y), \partial_y [-f(x, y)])$

We assume $f(x, y)$ is $\rho$-weakly-convex-weakly-concave.
Recall, a VI is defined on a set $Z$ and a set-valued mapping $F(z) : \mathbb{R}^d \to \mathbb{R}^d$.

A VI problem is to find a $z^* \in Z$, such that

$$\exists \, \xi^* \in F(z^*) \text{ s.t. } \langle \xi^*, z - z^* \rangle \geq 0, \quad \forall z \in Z$$

In this paper, we let $Z = X \times Y$ is compact (bounded by $D$)

$$F(z) = (\partial_x f(x, y), \partial_y [-f(x, y)])$$

We assume $f(x, y)$ is $\rho$-weakly-convex-weakly-concave

\textbf{Lemma}

$f(x, y)$ is $\rho$-weakly-convex-weakly-concave if and only if $F(z)$ is $\rho$-weakly monotone.
Title: “Solving Weakly-Convex-Weakly-Concave Saddle-Point Problems as Successive Strongly Monotone Variational Inequalities”

- Recall $F(z) = (\partial_x f(x, y), \partial_y [-f(x, y)])$
- Use proximal-point mapping to get a sequence of strongly monotone stochastic variational inequalities (will explain soon)
- Solve the strongly monotone stochastic variational inequalities approximately in each iteration
- As long as we can solve the strongly monotone stochastic variational inequalities to nearly optimal, then the algorithm should converge
Lemma for Proximal-point Mapping

Lemma

If $F(z) : \mathbb{R}^d \to \mathbb{R}^d$ is $\rho$-weakly monotone, then the proximal-point mapping $F_w^\gamma(z) := F(z) + \frac{1}{\gamma}(z - w)$ is $(\frac{1}{\gamma} - \rho)$ strongly monotone for any $\gamma < \rho^{-1}$ and any $w \in Z$.

Proof.

For any $z, z' \in Z$, any $\xi \in F(z)$, and any $\xi' \in F(z')$,

$$\langle \xi + \frac{1}{\gamma}(z - w) - \xi' - \frac{1}{\gamma}(z' - w), z - z' \rangle \geq \langle \xi - \xi', z - z' \rangle + \frac{1}{\gamma} \|z - z'\|^2 \tag{12}$$

$$\geq (\frac{1}{\gamma} - \rho) \|z - z'\|^2 \tag{13}$$

$$\geq (\frac{1}{\gamma} - \rho) \|z - z'\|^2 \tag{14}$$
**Algorithm 1** Inexact Proximal Point (IPP) Method for Weakly-Monotone SVI

1. **Input:** step size $\eta_k$, integers $T_k$ and non-decreasing weights $\theta_k$, $z_0 \in \mathcal{Z}$, $0 < \gamma < \rho^{-1}$
2. **for** $k = 0, \ldots, K - 1$ **do**
   3. Let $F_k \equiv F_{z_k} = F(z) + \gamma^{-1}(z - z_k)$
   4. $z_{k+1} = \text{ApproxSVI}(F_k, \mathcal{Z}, z_k, \eta_k, T_k)$
5. **end for**
6. Sample $\tau$ randomly from $\{0, 1, \ldots, K - 1\}$ with $\text{Prob}(\tau = k) = \frac{\theta_k}{\sum_{k=0}^{K-1} \theta_k}$, $k = 0, 1 \ldots, K - 1$.
7. **Output:** $z^{(\tau)}$.

- Line 4, ApproxSVI can be Gradient Descent, Nesterov’s accelerated method, or Stochastic Subgradient Method.
Focus on Stochastic Subgradient Method

Assumptions on stochastic oracle

- For any $z \in Z$, the oracle returns a random variable $\xi(z)$ s.t.

\[
\mathbb{E}[\xi(z)] \in F(z), \quad \mathbb{E}[\|\xi(z)\|^2] \leq G^2
\]  

(15)

Algorithm 2 Stochastic Subgradient Method for SVI($F, \mathcal{Z}$): SG($F, \mathcal{Z}, z(0), \eta, T$)

1: **Input**: Monotone Mapping $F$, set $\mathcal{Z}$, $z(0) \in \mathcal{Z}$, $\eta > 0$ and an integer $T \geq 1$.
2: **for** $t = 0, \ldots, T - 1$ **do**
3: \[ z(t+1) = \text{Proj}_{\mathcal{Z}} (z(t) - \eta \xi(z(t))) \] given a random vector $\xi(z(t))$ that satisfies $\mathbb{E}[\xi(z(t))] \in F(z(t))$
4: **end for**
5: Sample $\tau$ uniformly randomly from $\{0, 1, \ldots, T - 1\}$.
6: **Output**: $z(\tau)$
Strong Assumption: Minty VI

- Stampacchia variational inequality ($SVI(F, Z)$) concerns finding $z^* \in Z$ such that

\[ \exists \, \xi^* \in F(z^*) \quad s.t. \quad \langle \xi^*, z - z^* \rangle \geq 0, \; \forall z \in Z \]  \hfill (16)

- Minty variational inequality ($MVI(F, Z)$) concerns finding $z^* \in Z$ such that

\[ \langle \xi, z - z^* \rangle \geq 0, \; \forall z \in Z, \; \forall \xi \in F(z) \]  \hfill (17)

- This paper assumes that the solution of Minty variational inequality exists when

\[ F(z) = (\partial_x f(x, y), \partial_y [-f(x, y)]) \]

- This assumption is “strong and difficult to check, ...., there is no practical problem for which the Minty variational inequality condition has been proven” (cited from (Nouiehed et al., 2019))
Convergence

Under the above assumptions, running Algorithm 1 with Stochastic Subgradient Method as the \text{ApproxSVI}(\eta_k = 1/k, T_k = (k + 1)^2) with total of steps $K = \mathcal{O}(1/\epsilon^2)$, we have

\[
\mathbb{E}[\|w_\tau - w'_\tau\|^2] \leq \mathcal{O}(\epsilon^2) \quad (18)
\]

\[
\mathbb{E}[dist(0, \partial [f(w'_\tau) + 1_Z(w'_\tau)])] \leq \mathcal{O}(\epsilon) \quad (19)
\]

where $w'_\tau$ is the solution to SVI($F^\gamma_w, Z$).

The total iteration complexity is $\mathcal{O}(1/\epsilon^6)$, derived as $\sum_{k=1}^{K} k^2 = \mathcal{O}(1/\epsilon^6)$. 

Weakly-Convex-Weakly-Concave Saddle-Point Problems 22 / 47
Proof Outline

- $f(x, y)$ is $\rho$-weakly-convex-weakly-concave if and only if $F(z)$ is $\rho$-weakly monotone.
- Given $w'_\tau$ is the solution to $\text{SVI}(F^\gamma_{w_\tau}, Z)$,

$$ \text{dist}(0, \partial [f(w'_\tau) + 1_Z(w'_\tau)]) \leq \frac{1}{\gamma} \|w_\tau - w'_\tau\| $$

- The inner loop algorithm converges in $O(1/T)$ by using strongly-monotone of $F^\gamma_{w_\tau}$.
- Finally, we use the assumption of the solution of the Minty VI exists and the strongly-monotone $F^\gamma_{w_\tau}$ to get results for outer loop algorithm.
Outline

Introduction to Variational Inequalities

Weakly-Convex-Weakly-Concave Saddle-Point Problems

Stochastic Variational Inequalities for Multi-stages

Stochastic Mirror Prox
Recall Stochastic Multi-stage Programming

Assumptions: the probability space $\Xi$ is discrete, i.e. there are only finitely many scenarios

$$x = (x_1, \cdots, x_N) \in \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_N}$$

$$\xi = (\xi_1, \cdots, \xi_N) \in \Xi_1 \times \cdots \times \Xi_N$$

Response function: For every $\xi \in \Xi$, we define the mapping

$$x(\cdot) : \xi \rightarrow x(\xi)$$

Space of functions $\mathcal{L}_n$:

$$\mathcal{L}_n(\Xi, p) = \text{the collection of all functions } x(\cdot) : \Xi \rightarrow \mathbb{R}^n.$$
Nonanticipativity

- **Nonanticipativity**: \( x_t \) depends only on \( \xi_t := (\xi_1, \cdots, \xi_t) \)
- \( N \)-stage pattern

\[
x(\xi) = (x_1, x_2(\xi_1), x_3(\xi_1, \xi_2), \cdots, x_N(\xi_1, \cdots, \xi_{N-1}))
\]

- "You cannot see the future"
- For example, in the stock market, the time-indexed decision variables that are associated with making an investment at time \( t \) **cannot utilize any information** that is revealed after time \( t \).
Nonanticipativity Subspace

- Recall $N$-stage pattern

$$x(\xi) = (x_1, x_2(\xi_1), x_3(\xi_1, \xi_2), \cdots, x_N(\xi_1, \cdots, \xi_{N-1}))$$  \hspace{1cm} (25)

- The nonanticipativity subspace

$$\mathcal{N} = \{x(\cdot) = (x_1(\cdot), \cdots, x_N(\cdot)) \mid x_k(\xi) \text{ does not depend on } \xi_k, \cdots, \xi_N \}$$  \hspace{1cm} (26)

- Nonanticipativity constraint is

$$x(\cdot) \in \mathcal{N}$$
Due to dependence on certain scenarios $\xi$, the constraint set $C$ has dependence on $\xi$

$$C = \{x(\cdot) \in \mathcal{L}_n \mid x(\xi) \in C(\xi) \text{ for all } \xi \in \Xi\}$$  \hspace{1cm} (27)$$

The *expectational* inner product is defined as

$$\langle x(\cdot), w(\cdot) \rangle = \sum_{\xi \in \Xi} p(\xi) \sum_{k=1}^{N} \langle x_k(\xi), w_k(\xi) \rangle.$$  \hspace{1cm} (28)$$
Define a concatenated vector $F(x, \xi)$

$$F(x, \xi) = (F_1(x, \xi), \cdots, F_N(x, \xi))$$  \hfill (29)

where each $F_i(x, \xi)$ is continuous in $x \in \mathbb{R}^n$.

Define the mapping $\mathcal{F}(x(\cdot))$

$$\mathcal{F}(x(\cdot)) : \xi \rightarrow F(\xi, x(\xi)) = (F_1(x(\xi), \xi), \cdots, F_N(x(\xi), \xi))$$  \hfill (30)
Definition of Stochastic Variational Inequalities for Multi-stages

- Recall an variational inequality is defined as

\[ F(x) \in N_C(x) = \{ h \in \mathbb{R}^n : h^T(x' - x) \geq 0, \forall x' \in C \} \]  

(31)

- We have defined the the mapping \( F(x) \), the set \( C \), and the inner product for multi-stages stochastic programming, which is \( \mathcal{F}(x(\cdot)) \) and \( C \cap \mathcal{N} \), and \( \langle u(\cdot), v(\cdot) \rangle \) on the space of functions \( \mathcal{L}_n \).

- **Stochastic Variational Inequalities for Multi-stages** (basic form):

\[ \mathcal{F}(x(\cdot)) \in N_{C \cap \mathcal{N}}(x(\cdot)) \]  

(32)
A Closer Look at the Basic Form

\[ \mathcal{F}(x(\cdot)) \in N_{\mathcal{C} \cap \mathcal{N}}(x(\cdot)) \] (33)

- \( \mathcal{C} \cap \mathcal{N} \) makes sure the solution satisfies the constraints and nonanticipativity.
- This stochastic variational inequality is difficult to solve. We have to find \( x(\cdot) \) such that

\[
x(\cdot) \in \mathcal{C} \cap \mathcal{N}
\]
\[
\langle (F_1(x(\xi), \xi), \cdots, F_N(x(\xi), \xi)), x'(\cdot) - x(\cdot) \rangle \geq 0 \quad \forall x'(\cdot) \in \mathcal{C} \cap \mathcal{N}
\] (35)

where \( \mathcal{N} = \{ x(\cdot) = (x_1(\cdot), \cdots, x_N(\cdot)) \mid x_k(\xi) \text{ does not depend on } \xi_k, \cdots, \xi_N \} \)
Solution: Dualization

- Dualize the nonanticipativity constraint \( x(\cdot) \in \mathcal{N} \) by nonanticipativity multipliers

- Nonanticipativity multipliers

  \[
  \mathcal{M} = \{ w(\cdot) \mid \mathbb{E}_{\xi_k, \cdots, \xi_N} [w_k(\xi_1, \cdots, \xi_k-1, \xi_k, \cdots, \xi_N)] = 0 \} \quad (36)
  \]

- \( \mathcal{M} \) is orthogonal complement of \( \mathcal{N} \) w.r.t expectational inner product

  \[
  w(\cdot) \in N_{\mathcal{N}}(x(\cdot)) \quad (37)
  \]

  \[
  \iff x(\cdot) \in N_{\mathcal{M}}(x(\cdot)) \quad (38)
  \]

  \[
  \iff x(\cdot) \in \mathcal{N}, \text{ and } w(\cdot) \in \mathcal{M} \quad (39)
  \]
Recall the basic form is
\[ \mathcal{F}(x(\cdot)) \in N_{C \cap \mathcal{N}}(x(\cdot)) \]

The extensive form in multistage programming is
\[
x(\cdot) \in \mathcal{N} \quad \text{there exists} \quad w(\cdot) \in \mathcal{M} \quad \text{s.t.}
F(x(\xi), \xi) + w(\xi) \in N_{C(\xi)}(x(\xi)) \quad \forall \xi \in \Xi.
\]

Note: these two forms are equivalent.
- Extensive form implies basic form.
- To obtain extensive form from basic form, we need \( x(\xi) \in \text{ri}(C(\xi)) \) for all \( \xi \).
The idea of decomposition is used in Progressive Hedging Algorithm. Consider two-stage problem

\[
\begin{align*}
\min_{(x_1,y_1),\ldots,(x_K,y_K),x_0} & \sum_k p_k f(x_k,y_k) \\
\text{s.t.} & \quad (x_k,y_k) \in X_k \quad \forall k = 1,\ldots,K \\
& \quad p_k x_k = p_k x_0 \quad \forall k = 1,\ldots,K
\end{align*}
\]  

The augmented Lagrangian is

\[
L_\rho(x_0, x, y, \lambda) = \sum_k p_k \left( f(x_k, y_k) + \lambda_k^T (x_k - x_0) + \frac{\rho}{2} \|x_k - x_0\|^2 \right)
\]
Decomposed Large-scale Problem Into Small Sub-problems

Instead of solving the original problem directly, we solve $K$ sub-problems in each step $i$

$$(x_k^{i+1}, y_k^{i+1}) \in \arg \min_{(x_k, y_k) \in X_k} f(x_k, y_k) + \lambda_k^i (x_k - x_i^0) + \frac{\rho}{2} \|x_k^i - x_{i0}\|_2^2, \quad \forall k = 1, \ldots, K$$

(44)

Then we update $x_0$ with the updated $x_i$ values where $i$ is from 1 to $K$

$$x_0^{i+1} = \sum_k p_k x_k^{i+1}$$

(45)

Finally, we update the dual variables

$$\lambda_k^{i+1} = \lambda_k^i + \rho (x_k^{i+1} - x_{i0}^{i+1}), \quad \forall k = 1, \ldots, K$$

(46)
Proof of Basic-extensive Forms Equivalence

▶ (Extensive form to basic form)

\[ x(\cdot) \in \mathcal{N} \quad \exists w(\cdot) \in \mathcal{M} \quad s.t. F(x(\xi), \xi) + w(\xi) \in N_C(\xi)(x(\xi)) \quad \forall \xi \quad (47) \]

\[ \implies \mathcal{F}(x(\cdot)) \in N_{\mathcal{C}\cap\mathcal{N}}(x(\cdot)) \quad (48) \]

▶ Start with \( \mathcal{F}(x(\cdot)) \in N_{\mathcal{C}\cap\mathcal{N}}(x(\cdot)) \), apply intersection rule on VI,

\[ N_{\mathcal{C}\cap\mathcal{N}}(x(\cdot)) \supseteq \mathcal{N}_C(x(\cdot)) + \mathcal{N}(x(\cdot)) \quad (49) \]

\[ := \{ v(\cdot) + w(\cdot) \mid v(\cdot) \in \mathcal{N}_C(x(\cdot)), w(\cdot) \in \mathcal{N}(x(\cdot)) \} \quad (50) \]

▶ By orthogonal complement between \( \mathcal{M} \) and \( \mathcal{N} \),

\[ N_{\mathcal{N}}(x(\cdot)) = \mathcal{M} \quad \text{for all} \quad x(\cdot) \in \mathcal{N} \quad (51) \]
Monotonicity and the Existence of Solutions

- Recall $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is monotone relative to $C$

  \[ \langle F(x') - F(x), x' - x \rangle \geq 0 \quad \forall x, x' \in C \quad (52) \]

- $\mathcal{F} : \mathcal{L}_n \rightarrow \mathcal{L}_n$ is monotone relative to $C$

  \[ \langle \mathcal{F}(x'(\cdot)) - \mathcal{F}(x(\cdot)), x'(\cdot) - x(\cdot) \rangle \geq 0 \quad \forall x(\cdot), x'(\cdot) \in C \quad (53) \]

Theorem

- Given $\mathcal{F}(x(\cdot))$ in eq. (30) is monotone relative to $C$ where $F(\cdot, \xi)$ is monotone relative to $C'(\xi)$, then the set of solutions to the SVI in basic form is convex. Note that the conditions on $\mathcal{N}$ and $\mathcal{C}$ are hidden in statement that $\mathcal{F}(x(\cdot)) \in \mathcal{N}_{C \cap \mathcal{N}}(x(\cdot))$.

- Under strict monotonicity, the solution is unique if it exits.
Summary of SVI in Multi-stage

Define the basic form of SVI in multi-stage programming
- Space of functions, expectational inner product
- Nonanticipativity, nonanticipativity subspace
- Constraint subspace
- Mapping $\mathcal{F}(x(\cdot))$
Summary of SVI in Multi-stage

- Define the basic form of SVI in multi-stage programming
  - Space of functions, expectational inner product
  - Nonanticipativity, nonanticipativity subspace
  - Constraint subspace
  - Mapping $F(x(\cdot))$

- Extensive from
  - Motivation: basic form is difficult to solve
  - Solution: dualize nonanticipativity constraints
  - Equivalent of two forms
  - Related to Progressive Hedging Algorithm, decompose large-scale problem

- Monotonicity and the existence of solutions
Outline

Introduction to Variational Inequalities

Weakly-Convex-Weakly-Concave Saddle-Point Problems

Stochastic Variational Inequalities for Multi-stages

Stochastic Mirror Prox
Preliminaries

▶ Given a norm $\| \cdot \|$, its conjugate norm is

$$\| \xi \|_* = \max_{z : \|z\| \leq 1} \langle \xi, z \rangle$$ (54)

▶ $\omega(z) : X \times Y \to \mathbb{R}$ be a distance generating function where $\omega$ is 1-strongly convex function w.r.t. some norm $\| \cdot \|$ on the underlying space and is continuously differentiable

▶ The prox-function

$$V(z, u) = \omega(u) - \omega(z) - \langle \omega'(z), u - z \rangle$$ (55)

▶ Prox-mapping: given an input $z$ and a vector $\xi$,

$$P_z(\xi) = \arg\min_{u \in Z} V_\omega(u, x) + \langle \xi, u \rangle$$ (56)
Deterministic Mirror-Prox Algorithm

1. **Construct a VI condition** with the set $Z$ and the mapping $F(z)$ based on the problem.
   - In saddle point problem, 
     \[ Z = X \times Y, \quad F(z) = (\partial_x f(x,y), \partial_y [-f(x,y)]) \]

2. **Initialization**: \( r_0 \in Z \), stepsize \( \gamma_t > 0 \).

3. **Update at each iteration** \( t \),
   \[
   w_t = P_{r_{t-1}} (\gamma_t F(r_{t-1})) \\
   r_t = P_{r_{t-1}} (\gamma_t F(w_t))
   \]

4. **At step T**, output
   \[
   \hat{z}_T = \left[ \sum_{t=1}^{T} \gamma_t \right]^{-1} \sum_{t=1}^{T} \gamma_t w_t.
   \]
Stochastic Mirror-Prox Algorithm

\[ w_r = P_{r_{t-1}} \left( \gamma_t \hat{F}(r_{t-1}) \right) \]  
(60)

\[ r_r = P_{r_{t-1}} \left( \gamma_t \hat{F}(w_r) \right) \]  
(61)

- Stochastic Oracle: $\hat{F}(z)$ is the output of the oracle w.r.t. the input $z$, and is the approximation of $F(\cdot)$

- Assumptions: for any $z \in Z$, the oracle returns $\hat{F}(z)$ s.t.

\[ \|\mathbb{E}[\hat{F}(z) - F(z)]\|_* \leq \mu, \quad \mathbb{E}[\|\hat{F}(z) - F(z)\|_*^2] \leq \sigma^2 \]  
(62)
Main Result

Assume $\|F(z) - F(z')\|_* \leq L\|z - z'\| + M$ for all $z, z' \in Z$, then running the Stochastic Mirror Prox algorithm with a constant stepsize $0 \leq \gamma \leq \frac{1}{\sqrt{3L}}$, then

$$\mathbb{E}[Err_{vi}(\hat{z}_T)] \leq K_0(T) := \frac{\Omega^2}{\gamma^T} + 3.5\gamma(M^2 + 2\sigma^2) + 2\mu\Omega$$  \hspace{1cm} (63)

where $\Omega^2 = 2\max_{u \in Z} V(z, u)$.

- Recall $z$ is the optimal solution, $\langle F(u), z - u \rangle \geq 0$

- Define error bound of VI

$$Err_{vi}(z) := \max_{u \in Z}\langle F(u), z - u \rangle$$
Similar to the proof of non-stochastic Mirror Prox, we use the optimality condition during the updates and \textit{Bregman three-point identity}.

\begin{equation}
V(x, z) = V_\omega(x, y) + V_\omega(y, z) - \langle \nabla \omega(z) - \nabla \omega(z), x, y \rangle \tag{64}
\end{equation}

As for the error $\epsilon_z = \|\hat{F}(z) - F(z)\|_*$ caused by the randomness, we use the assumptions of the oracle to bound it.
Application: Composite Minimization Problem

Problem: \( \min_{x \in X} g(f_1(x), \ldots, f_m(x)) \)

- \( X \) and \( Y \) are convex compact
- Each \( f_i \) are Lipschitz continuous
- \( g(x) \) is a convex function given by the Fenchel-type representation with a given Lipschitz continuous convex function \( h_*(y) \):

\[
g(u_1, \ldots, u_m) = \max_{y \in Y} \sum_{i=1}^{m} \langle u_i, A_i y + b_i \rangle - h_*(y) \tag{65}
\]
We rewrite the composite minimization problem as a saddle point problem which is convex-concave problem (due to convexity of $h_*(y)$)

$$\min_{x \in X} \max_{y \in Y} g(x, y) = \sum_{i=1}^{m} \langle f_i(x), A_i y + b_i \rangle - h_*(y)$$  \hspace{1cm} (66)

Now we have a monotone mapping

$$F(z) = \left( \sum_{i=1}^{m} \langle f'_i(x), A_i^* y + b_i \rangle - h_*(y)' \right)$$  \hspace{1cm} (67)

Note that $F$ is linear in $f_i(\cdot)$, so the stochastic oracle is unbiased, and the convergence rate becomes $\mathcal{O}(1/\sqrt{T})$. 

Thank You!

Questions?