Solving Monotone Stochastic Variational Inequalities and Complementarity Problems by Progressive Hedging

Jie Sun
National University of Singapore and Curtin University, Australia

In Collaboration with

Terry Rockafellar
University of Washington and University of Florida, USA

CMCAA 2018
BIT, June 30-July 1, 2018
Deterministic Variational Inequalities

Expression of conditions for optimality or equilibrium posed in a Hilbert space, finite- or infinite-dimensional

For $C \subset \mathcal{H}$ nonempty closed convex, $F : \mathcal{H} \to \mathcal{H}$ continuous,
determine $x \in C$ such that $-F(x) \in N_C(x)$
i.e., $\langle F(x), x' - x \rangle \geq 0 \ \forall x' \in C$

Monotone case: $F$ monotone, hence maximal monotone

$\langle F(x') - F(x), x' - x \rangle \geq 0 \ \forall x', x$
Elementary optimization: minimizing $g(x)$ over $x \in C$

$-\nabla g(x) \in N_C(x) \quad \rightarrow \quad \text{first-order optimality, take } F = \nabla g$

Lagrangian V.I.: for $L(y; z)$ on $Y \times Z$ closed convex-concave

$-\nabla_y L(y; z) \in N_Y(y), \quad -\nabla_z L(y; z) \in N_Z(z)$ corresponding to $x = (y, z), C = Y \times Z, F(x) = (\nabla_y L(y, z), -\nabla_z L(y, z))$

This encompasses KKT conditions in NLP and much more!

Complementarity problems as a special case: $C = \mathbb{R}^n_+$

$-F(x) \in N_C(x) \iff 0 \leq x \perp F(x) \geq 0$
Stochastic Variational Inequalities (SVI)

Multi-stage stochastic optimization and equilibrium

Pattern of “decisions” and “observations” in $N$ stages:

$$x_1, \xi_1, x_2(\xi_1), \xi_2, x_3(\xi_1, \xi_2), \xi_3, \cdots$$

Generally, consider two spaces and a map between them:

$$x = (x_1, ..., x_N) \in \mathbb{R}^n = \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_N}$$
$$\xi = (\xi_1, ..., \xi_N) \in \Xi := \Xi_1 \times \cdots \times \Xi_N$$

$$x(\cdot) : \xi \rightarrow x(\xi) = (x_1(\xi), ..., x_N(\xi))$$

the response function

Interpretation: Each $\xi \in \Xi$ (a probability space) is an information scenario. It is a path on the scenario tree.

A solution to a sto-opt problem should be a response function of form (1).
**Finite-dimensional function space** to work in:

\[ \mathcal{L} = \text{all functions from scenario space } \Xi \text{ to decision space } \mathbb{R}^n \]

**Response functions (solutions to SVI)** as elements of this space:

\[ x(\cdot) : \Xi \rightarrow \mathcal{L} \]

Expectation inner product giving **Hilbert structure**:

\[
\langle x(\cdot), w(\cdot) \rangle := \mathbb{E}_p(x(\xi), w(\xi)) = \sum_{\xi \in \Xi} p(\xi)x(\xi)^T w(\xi).
\]

**Note:** We solve for a **function**, not a **vector**!

That is, suppose \( \Xi = \{\xi^1, \xi^2, ..., \xi^K\} \) we need to find

\[
x^*(\cdot) = \begin{pmatrix} x^*(\xi^1) \\ \vdots \\ x^*(\xi^K) \end{pmatrix} \in \mathbb{R}^{Kn}.
\]
Multistage SVI (MSVI) in this $\mathcal{L}$ Setting

Nonanticipativity of decisions

$x_k$ can respond to $\xi_1, \ldots, \xi_{k-1}$ but not to $\xi_k, \ldots, \xi_N$

$x(\xi) = (x_1, x_2(\xi_1), x_3(\xi_1, \xi_2), \ldots, x_N(\xi_1; \xi_2, \ldots, \xi_{N-1}))$

Subspace of Nonanticipative Response Functions and its Complement Space

$\mathcal{N} := \{x(\cdot) : x(\cdot) \text{ satisfies the nonanticipativity constraint}\}$

$\mathcal{M} := \mathcal{N}^\perp = \{w(\cdot) \in \mathcal{L}_n : \mathbb{E}_{\xi_k, \ldots, \xi_N|\xi_1, \ldots, \xi_{k-1}} w_k(\xi) = 0, \ \forall k\}$. 
The MSVI Problem

Given a stochastic feasible set $C(\xi)$, and a stochastic map $F(\cdot, \xi) : \mathbb{R}^n \to \mathbb{R}^n$, find a stochastic $x(\xi)$, satisfying nonanticipativity, such that

$$-F(x(\xi), \xi) \in N_{C(\xi)}(x(\xi)) \quad \forall \xi \in \Xi.$$ 

Equivalently, we define

**The Feasible Set of** $x(\cdot)$:

$$C = \{x(\cdot) : x(\xi) \in C(\xi) \quad \forall \xi \in \Xi\}$$

The mapping $F(x(\cdot)) : \mathcal{L} \to \mathcal{L}$ take $\xi$ to $F(x(\xi), \xi)$. 
The **Basic Form** of MSVI:

\[
\text{Given } F : \mathcal{L} \to \mathcal{L} \text{ and probability space } \Xi, \text{ find } x(\cdot) \in \mathcal{L} \\
\text{such that } -F(x(\cdot)) \in N_{\text{GrW}}(x(\cdot))
\]

**Monotone case:** \( \mathcal{F} \) is monotone if \( F(\cdot, \xi) \) is monotone \( \forall \xi \in \Xi \)
**Calculus of Normals:** If $\exists \tilde{x}(\cdot) \in \mathcal{N}$ with $\tilde{x}(\xi) \in \text{ri } C(\xi)$ $\forall \xi$ (or without “ri” if $C(\xi)$ polyhedral), then

$$N_{C \cap \mathcal{N}}(x(\cdot)) = N_C(x(\cdot)) + N_N(x(\cdot))$$  
where $N_N(x(\cdot)) = \mathcal{N}^\perp = \mathcal{M}$

and moreover $v(\cdot) \in N_C(x(\cdot)) \iff v(\xi) \in N_{C(\xi)}(x(\xi))$ $\forall \xi$

**The Extensive Form of MSVI:**

Given $F : \mathcal{L} \to \mathcal{L}$ and probability space $\Xi$,

find $x(\cdot) \in \mathcal{N}, w(\cdot) \in \mathcal{M}$

such that $-F(x(\cdot)) - w(\cdot) \in N_C(x(\cdot))$

$$-F(x(\xi), \xi) - w(\xi) \in N_{C(\xi)}(x(\xi)) \quad \forall \xi$$
The Progressive Hedging Algorithm (PHA)

\[ \hat{x}^\nu(\xi) = \text{The unique } x(\xi) \text{ such that} \]

\[ -F(x(\xi)) - w^\nu(\xi) - r[x(\xi) - x^\nu(\xi)] \in N_C(\xi)(x(\xi)), \]

and then

\[ x^\nu+1(\cdot) = P_N(\hat{x}^\nu(\cdot)) \]

\[ w^\nu+1(\cdot) = w^\nu(\cdot) - rP_M(\hat{x}^\nu(\cdot)). \]

The \( P_N \) is done via computing a certain conditional expectation – quick and easy – and \( P_M = \mathcal{I} - P_N \).

**Convergence under Monotonicity.** As long as the MSVI has a solution,

\[ \{x^\nu(\cdot), w^\nu(\cdot)\} \rightarrow (x^*(\cdot), w^*(\cdot)) \] (a solution of the MSVI).

With additional regularity, at linear rate w.r.t. the \( r \)-norm

\[ \|(x, w)\|_r := \left(\|x\|^2 + r^{-2}\|w\|^2\right)^{\frac{1}{2}}. \]
Example: Two-Stage Linear Stochastic Complementarity

Corresponding SVI problem in extensive form:

Find $x(\cdot) \in \mathcal{N}$ along with $w(\cdot) \in \mathcal{M}$ such that, $\forall \xi \in \Xi$,

$$0 \leq x(\xi) \perp F(x(\xi), \xi) + w(\xi) \geq 0$$

PHA in this setting

$$F(x(\xi), \xi) = M(\xi)x(\xi) + b(\xi) = \begin{bmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{21}(\xi) & M_{22}(\xi) \end{bmatrix} \begin{bmatrix} x_1(\xi) \\ x_2(\xi) \end{bmatrix} + \begin{bmatrix} b_1(\xi) \\ b_2(\xi) \end{bmatrix},$$

with $M(\xi) + M(\xi)^T$ positive semidefinite.

$$x(\xi) = \begin{bmatrix} x_1(\xi) \\ x_2(\xi) \end{bmatrix}, \quad w(\xi) = \begin{bmatrix} w_1(\xi) \\ 0 \end{bmatrix}.$$
At iteration $\nu$, we solve $\forall \xi \in \Xi$,

\[
0 \leq \bar{x}_1^\nu(\xi) \perp [M_{11}(\xi) + rI]\bar{x}_1^\nu(\xi) + M_{12}(\xi)\bar{x}_2^\nu(\xi) + b_1^\nu(\xi) \geq 0,
\]

\[
0 \leq \bar{x}_2^\nu(\xi) \perp M_{21}(\xi)\bar{x}_1^\nu(\xi) + [M_{22}(\xi) + rI]\bar{x}_2^\nu(\xi) + b_2^\nu(\xi) \geq 0,
\]

where

\[
b_1^\nu(\xi)(\xi) = b_1(\xi) - rx_1^\nu + w_1^\nu(\xi), \quad b_2^\nu(\xi) = b_2(\xi) - rx_2^\nu(\xi).
\]

Then find $x^{\nu+1}(\cdot)$ and $w^{\nu+1}(\cdot)$ by projection.
Numerical results

The dimensions of $x_1(\xi)$ and $x_2(\xi)$ range from 15 to 500 while the number of scenarios ("sn" in the tables) ranges from 5 to 200, which we think would cover the range of quite many problems arising in practice. For each fixed dimension and number of scenarios, we generate 10 random problems and count the average number of iterations ("avg iter" in the tables) and the average CPU time (in seconds, "avg time (s)" in the tables).
### Table 1 Numerical results while sn increases (dim=[15,15])

<table>
<thead>
<tr>
<th>sn</th>
<th>avg-iter (r=1)</th>
<th>avg-time(s) (r=1)</th>
<th>avg-iter (r=√30)</th>
<th>avg-time(s) (r=√30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>60.9</td>
<td>0.1</td>
<td>65.1</td>
<td>0.1</td>
</tr>
<tr>
<td>10</td>
<td>77.1</td>
<td>0.2</td>
<td>67.7</td>
<td>0.2</td>
</tr>
<tr>
<td>20</td>
<td>108.9</td>
<td>0.5</td>
<td>70.9</td>
<td>0.3</td>
</tr>
<tr>
<td>50</td>
<td>148.1</td>
<td>1.5</td>
<td>94.3</td>
<td>1.0</td>
</tr>
<tr>
<td>100</td>
<td>179.9</td>
<td>3.2</td>
<td>94.0</td>
<td>1.8</td>
</tr>
<tr>
<td>150</td>
<td>143.0</td>
<td>3.8</td>
<td>104.9</td>
<td>2.8</td>
</tr>
<tr>
<td>200</td>
<td>193.7</td>
<td>6.9</td>
<td>122.1</td>
<td>4.4</td>
</tr>
</tbody>
</table>
Convergence time when scenario number increases
### Table 2 Numerical results while dimension increases (sn=100)

<table>
<thead>
<tr>
<th>dimension</th>
<th>( r=1 )</th>
<th></th>
<th>( r=\sqrt{\text{dim}} )</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>avg-iter</td>
<td>avg-time(s)</td>
<td>avg-iter</td>
<td>avg-time(s)</td>
</tr>
<tr>
<td>[15,15]</td>
<td>179.9</td>
<td>3.2</td>
<td>94.0</td>
<td>1.7</td>
</tr>
<tr>
<td>[30,30]</td>
<td>208.0</td>
<td>4.3</td>
<td>65.6</td>
<td>1.4</td>
</tr>
<tr>
<td>[50,50]</td>
<td>148.5</td>
<td>4.3</td>
<td>44.3</td>
<td>1.5</td>
</tr>
<tr>
<td>[100,100]</td>
<td>289.8</td>
<td>12.3</td>
<td>29.4</td>
<td>1.6</td>
</tr>
<tr>
<td>[200,200]</td>
<td>535.3</td>
<td>49.8</td>
<td>23.3</td>
<td>3.4</td>
</tr>
<tr>
<td>[300,300]</td>
<td>509.3</td>
<td>118.9</td>
<td>22.2</td>
<td>8.1</td>
</tr>
<tr>
<td>[500,500]</td>
<td>876.6</td>
<td>709.8</td>
<td>22.7</td>
<td>33.1</td>
</tr>
</tbody>
</table>
Convergence time when dimension increases
An Application: A Two-Stage Game

Consider a $K$-person game, where player $i$ solves a two-stage stochastic optimization problem with recourse,

$$
\min_{x_i \geq 0} f_i(x_i; x_{-i}) + E_{\xi} \left[ \min_{y_i \geq 0} \phi_i(x_i, y_i; x_{-i}, \xi) \right],
$$

Let

$$z_i(\xi) := \begin{pmatrix} x_i \\ y_i(\xi) \end{pmatrix}, \quad F_i(z(\xi), \xi) = \begin{pmatrix} \nabla_x f_i(x(\xi), \xi) + \phi_i(x(\xi), y(\xi), \xi) \\ \nabla_y \phi_i(x(\xi), y(\xi), \xi) \end{pmatrix},$$

and let

$$F(z(\xi), \xi) = [F_1(z(\xi), \xi), ..., F_K(z(\xi), \xi)].$$
Then as discussed above, the Nash equilibrium condition of this game is a stochastic complementarity problem:

\[
\text{Find } z(\cdot) \in \mathcal{N}, \ w(\cdot) \in \mathcal{M}, \text{ such that } 0 \leq \mathcal{F}(z(\cdot)) + w(\cdot) \perp z(\cdot) \geq 0.
\]
A Special Case: A Two-Stage Quadratic Game under Uncertainty

Player $i$ is solving the following problem:

$$\min_{x_i \geq 0} f_i(x_i; x_{-i}) + E_{\xi}[\min_{y_i \geq 0} \phi_i(x_i, y_i; x_{-i}, \xi)],$$

where

$$f_i(x_i, x_{-i}) = \frac{1}{2} x_i^\top Q_i x_i + q_i^\top x_i + \sum_{j \neq i} x_i^\top R_{ij} x_j,$$

and

$$\phi_i(y_i; x_i, x_{-i}, \xi) := \frac{1}{2} y_i^\top T_i(\xi) y_i + d_i(\xi)^\top y_i + x_i^\top S_i(\xi) y_i + \sum_{j \neq i} y_i^\top P_{ij}(\xi) x_j,$$
We can show in this case, the Nash equilibrium problem becomes a linear stochastic complementarity problem with the form

Find \( z(\cdot) \in \mathcal{N}, \ w(\cdot) \in \mathcal{M} \), such that

\[
0 \leq M(\xi)z(\xi) + b(\xi) + w(\xi) \perp z(\xi) \geq 0, \ \forall \xi \in \Xi.
\]
\[
M(\xi) = \begin{pmatrix}
Q_1 & S_1(\xi) & R_{12} & 0 & \cdots & R_{1N} & 0 \\
S_1(\xi)^\top & T_1(\xi) & P_{12}(\xi) & 0 & \cdots & P_{1N}(\xi) & 0 \\
R_{12}^\top & 0 & Q_2 & S_2(\xi) & \cdots & R_{2N} & 0 \\
P_{21}(\xi) & 0 & S_2(\xi)^\top & T_2(\xi) & \cdots & P_{2N}(\xi) & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
R_{1N}^\top & 0 & R_{2N}^\top & 0 & \cdots & Q_N & S_N(\xi) \\
P_{N1}(\xi) & 0 & P_{N2}(\xi) & 0 & \cdots & S_N(\xi)^\top & T_N(\xi)
\end{pmatrix}
\]

and

\[
b(\xi) = (q_1^\top, d_1(\xi)^\top, \cdots, q_N^\top, d_N(\xi)^\top)^\top.
\]
**Theorem (Quadratic Dominance)** Matrix $M(\xi)$ is semi-positive definite for all $\xi \in \Xi$ if there exist some nonzero constants $\alpha, \beta, \gamma$ such that the following conditions hold:

1. $Q_j - [\alpha^2 + \beta^2(K - j) + \frac{\gamma^2}{4}(K - 1)]I - \frac{1}{\beta^2} \sum_{i=1}^{j-1} R_{ij}^T R_{ij} \succeq 0$, for $j = 1, \cdots, K$;

2. $T_i(\xi) - \frac{1}{\alpha^2} S_i(\xi)^T S_i(\xi) - \frac{1}{\gamma^2} \sum_{j \neq i} P_{ij}(\xi) P_{ij}(\xi)^T \succeq 0$ for all $\xi \in \Xi$;

where $A \succeq 0$ means $A$ is symmetric semi-positive definite.
References


