STOCHASTIC VARIATIONAL INEQUALITIES AND EQUILIBRIUM

Terry Rockafellar
University of Washington, Seattle
University of Florida, Gainesville

"Smart Energy and Stochastic Optimization"
ENSTA ParisTech
May 30 to June 3, 2016

Modeling with Variational Inequalities

Targeted applications: here with stochastic ingredients

expression of conditions for optimality or equilibrium

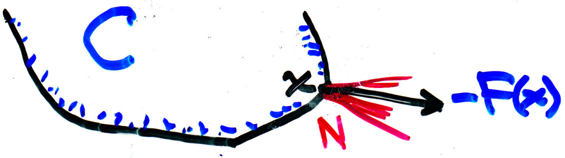
posed in a Hilbert space, finite- or infinite-dimensional

Variational inequality problem in a space $\mathcal{H}$

For $C \subset \mathcal{H}$ nonempty closed convex, $F : \mathcal{H} \to \mathcal{H}$ continuous,

$$\text{determine } x \in C \text{ such that } -F(x) \in N_C(x)$$

$$\text{i.e., } \langle F(x), x' - x \rangle \geq 0 \;\; \forall x' \in C$$



Monotone case: F monotone, hence actually maximal monotone

$$\langle F(x') - F(x), x' - x \rangle \geq 0 \text{ for all } x, x'$$

Motivations in Optimization — Without Stochastics

general V.I.: $-F(x) \in N_C(x)$

Elementary optimization: minimizing $g(x)$ over $x \in C$

$-\nabla g(x) \in N_C(x)$ $\longrightarrow$ first-order optimality

To formulate this as a V.I., take $F = \nabla g$

Lagrangian V.I.: for $l(y,z)$ on $Y \times Z$ closed convex

$$-\nabla_y l(y,z) \in N_Y(y), \quad \nabla_z l(y,z) \in N_Z(z),$$

To formulate this as a V.I., take

$$x = (x,y), \quad C = Y \times Z, \quad F(x) = (\nabla_y l(y,z), -\nabla_z l(y,z))$$

$\longrightarrow$ this encompasses KKT conditions in NLP and much more!

(also, it can model a saddle-point in a two-person game)

Motivations in Equilibrium — Without Stochastics

Game-type equilibrium: for agents $i = 1, \dots, m$

- agent i chooses $x_i \in X_i$ closed convex $\subset \boldsymbol{R}^{n_i}$
- minimization of $f_i(x_i, x_{-i})$ over $x_i \in X_i$ is desired

  where x_{-i} stands for the choices of all the other agents
- equilibrium represented by $-\nabla_{x_i} f_i(x_i, x_{-i}) \in N_{X_i}(x_i)$ for all i

  agent global optimality is replaced here by "stationarity"!

To formulate this as a V.I., take

$$x = (x_1, \dots, x_m), \qquad C = X_1 \times \cdots \times X_m$$

$$F(x) = (\nabla_{x_1} f_1(x_1, x_{-1}), \dots, \nabla_{x_m} f_m(x_m, x_{-m}))$$

Stochastic Structure with Emerging Information

capturing dynamics in stochastic optimization and equilibrium

Pattern of "decisions" and "observations" in N stages:

$$x_1,\ \xi_1,\ x_2,\ \xi_2,\dots,x_N,\ \xi_N \quad \text{with} \quad x_k \in \boldsymbol{R}^{n_k},\ \xi_k \in \Xi_k$$

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

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

Interpretation: each $\xi \in \Xi$ is an information **scenario**

Nonanticipativity of decisions

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

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

Simplifying assumptions here:

- the scenario space Ξ has only finitely many elements ξ
- each scenario $\xi \in \Xi$ has known probability $p(\xi) > 0$

$\longrightarrow$ Ξ is a probability space

Function Space Framework for Responses

$\mathcal{L}$ = **all** functions from **scenario** space Ξ to **decision** space R^n

$$\Xi \subset \Xi_1 \times \cdots \Xi_N, \qquad R^n = R^{n_1} \times \cdots \times R^{n_N}$$

Response functions as elements of this space:

$$x(\cdot) : \xi = (\xi_1, \ldots, \xi_N) \mapsto x(\xi) = (x_1(\xi), \ldots, x_N(\xi))$$

Expectation inner product giving Hilbert structure:

$$\langle x(\cdot), w(\cdot) \rangle = E_\xi[x(\xi), w(\xi)] = \sum_{\xi \in \Xi} p(\xi) \sum_{k=1}^N x_k(\xi) \cdot w_k(\xi)$$

Nonanticipativity subspace:

in terms of scenarios $\xi = (\xi_1, \ldots, \xi_{k-1}, \xi_k, \ldots, \xi_N)$,

$$\mathcal{N} = \{ x(\cdot) \in \mathcal{L} \mid x_k(\xi) \text{ depends only on } \xi_1, \ldots, \xi_{k-1} \}$$

$\longrightarrow$ $x(\cdot)$ is nonanticipative $\iff$ $x(\cdot) \in \mathcal{N}$

Complementary subspace: $\mathcal{M} = \mathcal{N}^\perp$ (source of "multipliers")

$$\mathcal{M} = \{ w(\cdot) \in \mathcal{L} \mid E_{\xi_k, \ldots, \xi_N}[w_k(\xi_1, \ldots, \xi_{k-1}, \xi_k \ldots, \xi_N)] = 0 \}$$

Multistage Stochastic Optimization in this Setting

Basic constraints beyond nonanticipativity: $x(\cdot) \in \mathcal{C}$

$$\mathcal{C} = \{x(\cdot) \in \mathcal{L} \mid x(\xi) \in C(\xi) \subset \boldsymbol{R}^n \text{ for all } \xi \in \Xi\}$$

Convexity assumption: making $\mathcal{C}$ be closed convex $\neq \emptyset$ in $\mathcal{L}$

$C(\xi)$ closed convex $\neq \emptyset$ in $\boldsymbol{R}^n$ for every scenario $\xi \in \Xi$

Stochastic programming problem in a classical format

minimize $\mathcal{G}(x(\cdot)) = E_\xi[g(x(\xi), \xi)]$ over all functions $x(\cdot) \in \mathcal{C} \cap \mathcal{N}$

Smoothness assumption: making $\mathcal{G}$ be a $\mathcal{C}^1$ function on $\mathcal{L}$

$g(\cdot, \xi)$ is a $\mathcal{C}^1$ function on $\boldsymbol{R}^n$ for every scenario $\xi \in \Xi$

Directional derivatives: $d\mathcal{G}(x(\cdot))(u(\cdot)) = \langle \nabla_x \mathcal{G}(x(\cdot)), u(\cdot) \rangle$

the gradient $\nabla \mathcal{G}(x(\cdot)) \in \mathcal{L}$ takes ξ to $\nabla g(x(\xi), \xi)$

Convex minimization case: $\mathcal{G}$ is convex if each $g(\cdot, \xi)$ is convex

Variational Inequalities in the Response Function Space $\mathcal{L}$

First-order optimality condition in stochastic programming

$$-\nabla\mathcal{G}(x(\cdot)) \in N_{\mathcal{C}\cap\mathcal{N}}(x(\cdot))$$

the V. I. for the gradient mapping $\nabla\mathcal{G}$ and the convex set $\mathcal{C}\cap\mathcal{N}$

this condition is **sufficient** in the case of **convex** minimization

Generalization: replace $\nabla\mathcal{G}$ by any mapping $\mathcal{F}:\mathcal{L}\to\mathcal{L}$ where

$\mathcal{F}(x(\cdot))$ takes ξ to $F(x(\xi),\xi)$ for functions $F(\cdot,\xi):\boldsymbol{R}^n\to\boldsymbol{R}^n$

stochastic programming had $F(\cdot,\xi)=\nabla_x g(\cdot,\xi)$

Continuity assumption: $F(\cdot,\xi)$ continuous making $\mathcal{F}$ continuous

Definition of a stochastic variational inequality, basic form

$$-\mathcal{F}(x(\cdot)) \in N_{\mathcal{C}\cap\mathcal{N}}(x(\cdot))$$

Monotone case: $\mathcal{F}$ is monotone if $F(\cdot,\xi)$ is monotone $\forall\,\xi\in\Xi$

Stochastic Decomposition

Basic S.V.I. to understand further: $-\mathcal{F}(x(\cdot)) \in N_{\mathcal{C}\cap\mathcal{N}}(x(\cdot))$

$\mathcal{C} = \{x(\cdot) \mid x(\xi) \in C(\xi),\ \forall\xi\}$, $\mathcal{N}$ = nonanticipativity subspace

Calculus of normals: if $\exists\,\tilde{x}(\cdot) \in \mathcal{N}$ with $\tilde{x}(\xi) \in \mathrm{ri}\, C(\xi)\ \forall\xi$, then

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

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

Definition of a stochastic variational inequality, extensive form

$$x(\cdot)\in\mathcal{N} \text{ and } \exists w(\cdot)\in\mathcal{M}: \quad -F(x(\xi),\xi)-w(\xi) \in N_{C(\xi)}(x(\xi)),\ \forall\xi$$

the basic and extensive forms of are "essentially equivalent"

Meaning in stochastic programming: when $F(\cdot,\xi) = \nabla_x g(\cdot,\xi)$

$x(\xi)$ minimizes $g(\cdot,\xi) + \langle\cdot, w(\xi)\rangle$ over $C(\xi)$ for each scenario ξ

the multiplier vectors $w(\xi)$ give the **price of information**

having $w(\cdot) \in \mathcal{M}$ is a "martingale-like" condition

Projection Tool for Aggregating Responses

Recalling the structure of the complementary subspaces:

$$\mathcal{N} = \{x(\cdot) \in \mathcal{L} \mid x_k(\xi) \text{ depends only on } \xi_1, \ldots, \xi_{k-1}\}$$
$$\mathcal{M} = \{w(\cdot) \in \mathcal{L} \mid E_{\xi_k,\ldots,\xi_N}[w_k(\xi_1, \ldots, \xi_{k-1}, \xi_k \ldots, \xi_N)] = 0\}$$

Aggregation: let $\mathcal{P}$ = projection onto $\mathcal{N}$

then $\mathcal{I} - \mathcal{P}$ = projection onto $\mathcal{M}$, since $\mathcal{M} = \mathcal{N}^\perp$

Execution relative to the information structure:

- Scenarios $\xi = (\xi_1, \ldots, \xi_N)$ and $\xi' = (\xi'_1, \ldots, \xi'_N)$ are at stage k information-equivalent if $(\xi_1, \ldots, \xi_{k-1}) = (\xi'_1, \ldots, \xi'_{k-1})$
- Let $A_k(\xi)$ = kth-stage equivalence class containing ξ
- Then $x(\cdot) = \mathcal{P}(\bar{x}(\cdot))$ has its kth-stage component given by

$$x_k(\xi) = \sum_{\xi' \in A_k(\xi)} p(\xi')\bar{x}_k(\xi') / \sum_{\xi' \in A_k(\xi)} p(\xi')$$

thus $x_k(\xi)$ is the **conditional expectation** of $\bar{x}_k(\xi)$ relative to the kth-stage information-equivalence class containing ξ

Progressive Hedging in Stochastic Programming

Rock. & Wets, 1991: stochastic decomposition realized iteratively

Algorithm statement in the convex case with parameter $r > 0$

Having $x^\nu(\cdot) \in \mathcal{N}$ and $w^\nu(\cdot) \in \mathcal{M}$, get $\bar{x}^\nu(\cdot) \in \mathcal{L}$ by

$$\bar{x}^\nu(\xi) = \operatorname{argmin}_{x \in C(\xi)} \left\{ g(x,\xi) + x \cdot w^\nu(\xi) + \frac{r}{2} ||x - x^\nu(\xi)||^2 \right\}$$

Then get $x^{\nu+1}(\cdot) \in \mathcal{N}$ and $w^{\nu+1}(\cdot) \in \mathcal{M}$ by aggregation:

$$x^{\nu+1}(\cdot) = \mathcal{P}(\bar{x}^\nu(\cdot)), \qquad w^{\nu+1}(\cdot) = w^\nu(\cdot) + r[\bar{x}^\nu(\cdot) - x^{\nu+1}(\cdot)]$$

Convergence theorem — when a solution pair $x(\cdot)$, $w(\cdot)$, exists

The sequence $\{(x^\nu(\cdot), w^\nu(\cdot))\}_{\nu=1}^{\infty}$ generated by the algorithm will always converge to a particular solution pair $(x^*(\cdot), w^*(\cdot))$, with

$$\begin{aligned} &||x^{\nu+1}(\cdot) - x^*(\cdot)||^2 + r^{-2}||w^{\nu+1}(\cdot) - w^*(\cdot)||^2 \\ &\qquad \le ||x^\nu(\cdot) - x^*(\cdot)||^2 + r^{-2}||w^\nu(\cdot) - w^*(\cdot)||^2 \end{aligned}$$

Progressive Hedging for Stochastic Variational Inequalities

Recall extensive form of S.V.I: $x(\cdot) \in \mathcal{N}$, $w(\cdot) \in \mathcal{M}$, and
$-F(x(\xi), \xi) - w(\xi) \in N_{C(\xi)}(x(\xi))$ for every scenario $\xi \in \Xi$

Algorithm statement in the monotone case with parameter $r > 0$

Having $x^\nu(\cdot) \in \mathcal{N}$ and $w^\nu(\cdot) \in \mathcal{M}$, get $\bar{x}^\nu(\cdot) \in \mathcal{L}$ by solving for **each individual scenario** $\xi \in \Xi$ the V.I.

$$-F^\nu(x, \xi) \in N_{C(\xi)}(x)$$

with respect to $x \in \boldsymbol{R}^n$ to get $x^\nu(\xi)$, where

$$F^\nu(x, \xi) = F(x, \xi) + w^\nu(\xi) + r[x - x^\nu(\xi)].$$

Then get $x^{\nu+1}(\cdot) \in \mathcal{N}$ and $w^{\nu+1}(\cdot) \in \mathcal{M}$ by aggregation:

$$x^{\nu+1}(\cdot) = \mathcal{P}(\bar{x}^\nu(\cdot)), \qquad w^{\nu+1}(\cdot) = w^\nu(\cdot) + r[\bar{x}^\nu(\cdot) - x^{\nu+1}(\cdot)]$$

Convergence theorem — when a solution pair $x(\cdot)$, $w(\cdot)$, exists

same result again!

The Role of Monotonicity and its Prospects

Technical basis for the algorithm:

proximal point algorithm applied to a certain $T : \mathcal{L} \rightrightarrows \mathcal{L}$

T is derived from a "twisting" of the ingredients of the S.V.I.

Global convergence: T must be a maximal monotone mapping

Local convergence: T maximal monotone just around solution

Prospects for extended stochastic programming:

- convexity of the cost functions can be weakened to a second-order optimality condition right at a solution
- expectation may be replaced by risk measure in objective

Prospects for game models and equilibrium

- game versions of multistage stochastic programming
- local convergence in circumstances of "moderate interaction"
- an augmented Lagrangian technique may actually elicit that!

Some References

[1] R.T. Rockafellar, R.J -B Wets (2016) "Stochastic variational inequalities: single-stage to multistage," *Mathematical Programming B,* forthcoming.

[2] R.T. Rockafellar and R.J-B Wets (1991) "Scenarios and policy aggregation in optimization under uncertainty," *Mathematics of Operations Research* 16, 119–147.

[3] R.T. Rockafellar and R.J-B Wets (1976) "Nonanticipativity and L^1-martingales in stochastic optimization problems," *Mathematical Programming Study* 6, 170-187.

website: **www.math.washington.edu/∼rtr/mypage.html**