Simulation-based Schemes for Hierarchical Optimization and Games under Uncertainty

Uday V. Shanbhag[+]

Jointly with Shisheng Cui[+] and Farzad Yousefian[%]

East Coast Optimization Meeting, George Mason University
April 14, 2023

+: Penn State; %: Rutgers University

Mathematical Programs with Equilibrium Constraints (MPEC) I

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & f(\mathbf{x},\mathbf{y}) \\ \text{subject to} \quad & \mathbf{y} \in \text{SOL}(\mathcal{Y}, F(\mathbf{x},\bullet)), \\ & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{MPEC}$$

where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is a real-valued function, $F : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^m$, $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^m$ denote closed and convex sets, and $\text{SOL}(\mathcal{Y}, F(\mathbf{x},\bullet))$ denotes the solution set of the parametrized variational inequality problem $\text{VI}(\mathcal{Y}, F(\mathbf{x},\bullet))$, given an upper-level decision $\mathbf{x}$.

Background on VIs

1. Recall that the variational inequality problem $\mathrm{VI}(\mathcal{Y}, F(\mathbf{x}, \bullet))$ requires a vector $\mathbf{y}$ in the set $\mathcal{Y}$ such that

$$(\tilde{\mathbf{y}} - \mathbf{y})^T F(\mathbf{y}) \ \geq \ 0, \qquad \forall\, \tilde{\mathbf{y}} \ \in \ \mathcal{Y}. \tag{VI($\mathcal{Y}, F$)}$$

2. VIs [Facchinei and Pang, 2003] subsume
   1. Smooth convex optimization problems
   2. A range of equilibrium problems (including Nash, traffic, and economic equilibrium problems)

Example: Convex optimization with quadratic objectives

Suppose $Q \succeq 0$.

$$\min_{\mathbf{y} \in \mathcal{Y}} \ \frac{1}{2}\mathbf{y}^T Q\mathbf{y} + c^T\mathbf{y} \tag{1}$$

y is a minimizer if and only if y is a solution of VI$(\mathcal{Y}, F)$ where $F(y) = Qy + c$, i.e. y solves

$$(\tilde{y} - \mathbf{y})^T(Q\mathbf{y} + c) \ \geq \ 0, \quad \forall \tilde{\mathbf{y}} \in \mathcal{Y}.$$

Example: Convex noncooperative games with quadratic objectives

Suppose $Q_{ii} \succeq 0$ for $i = 1, \cdots, N$.

$$\min_{\mathbf{y}_i \in \mathcal{Y}_i} \ \frac{1}{2}\mathbf{y}_i^T Q_{ii}\mathbf{y}_i + \sum_{j \neq i} \mathbf{y}_j^T Q_{ij}\mathbf{y}_j + c_i^T\mathbf{y}_i \tag{2}$$

Then $\{\mathbf{y}_1, \cdots, \mathbf{y}_N\}$ is a Nash equilibrium if and only if **y** is a solution of VI$(\mathcal{Y}, F)$ where $F(y) = \mathbf{Q}y + c$ and $\mathcal{Y} \triangleq \prod_{i=1}^N \mathcal{Y}_i$.

Stochastic variational inequality problem I

1. Recall that the stochastic variational inequality problem $\mathrm{VI}(\mathcal{Y}, F)$ where $F_i(\mathbf{y}) \triangleq \mathbb{E}[G_i(\mathbf{y}, \omega)]$ for $i = 1, \cdots, n$.
2. Capture stochastic convex optimization problems and noncooperative games

Example: Convex stochastic noncooperative games

Suppose the ith player solves

$$\min_{\mathbf{y}_i \in \mathcal{Y}_i} \ \mathbb{E}[f_i(\mathbf{y}, \omega)], \qquad (\mathrm{Player}_i(\mathbf{y}^{-i}))$$

given $\mathbf{y}^{-i}$ for $i = 1, \cdots, N$. Then $\{\mathbf{y}_1, \cdots, \mathbf{y}_N\}$ is a Nash equilibrium if and only if $\mathbf{y}$ is a solution of $\mathrm{VI}(\mathcal{Y}, F)$ where

$$F(y) = \begin{pmatrix} \mathbb{E}[\nabla_{\mathbf{y}_1} f_1(\mathbf{y}, \omega)] \\ \vdots \\ \mathbb{E}[\nabla_{\mathbf{y}_N} f_N(\mathbf{y}, \omega)] \end{pmatrix} \text{ and } \mathcal{Y} \triangleq \prod_{i=1}^{N} \mathcal{Y}_i.$$

Stochastic variational inequality problem II

Much research on SVIs: Existence/uniqueness [Ravat and S, 2012,2015], SAA schemes [Shapiro, Xu, etc.], SA schemes [Xu, Lan, S, Yousefian, Jofre & Thompson, etc]

SVI($\mathcal{Y}, F$) requires a vector $\mathbf{y}$ in the set $\mathcal{Y}$ such that

$$(\tilde{\mathbf{y}} - \mathbf{y})^T F(\mathbf{x}, \mathbf{y}) \geq 0, \qquad \forall\, \tilde{\mathbf{y}} \in \mathcal{Y}. \qquad \text{(VI}(\mathcal{Y}, F))$$

MPECs

A hierarchical framework where **y** is a solution to a lower-level variational inequality problem and **x** is an upper-level decision. Captures...

- A subclass of bilevel optimization problems
- Stackelberg equilibrium problems
- Frictional contact problems
- Market design problems in power systems
- Design in traffic equilibrium problems

See [Luo, Pang and Ralph, 1996]

Challenges

In the space of x, y, the MPEC is a challenging generalization of nonconvex and nonlinear programs

- Such problems lack and interior and standard regularity conditions fail (LICQ, MFCQ) at at any feasible point.
- Standard convergence theory for NLP does not apply

Algorithms

- Interior schemes [Pang, Ralph, Wright, Nocedal, Leyffer etc., 2000–2010]
- SQP schemes [Anitescu, Ralph, Wright, etc., 2000–2010]
- Implicit schemes [Require $y(\bullet)$ to be single-valued] [Pang, Outrata, etc., 1995–2010]
- Limited extensions to stochastic MPECs [Shapiro, Xu, Ralph, 2010-2015]

Gap: No available non-asymptotic guarantees even for deterministic regimes

The stochastic MPEC of interest

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}} \quad & f(\mathbf{x},\mathbf{y}) \\ \text{subject to} \quad & \mathbf{y} \in \mathrm{SOL}(\mathcal{Y}, \mathbb{E}[G(\mathbf{x},\bullet,\omega)]), \\ & \mathbf{x} \in \mathcal{X}. \end{aligned} \qquad (\mathrm{SMPEC}^{\mathrm{exp}})$$

$$\begin{aligned} \min_{\mathbf{x},\mathbf{y}(\omega)} \quad & \mathbb{E}[f(\mathbf{x},\mathbf{y}(\omega))] \\ \text{subject to} \quad & \mathbf{y}(\omega) \in \mathrm{SOL}(\mathcal{Y}(\mathbf{x},\omega), G(\mathbf{x},\bullet,\omega)), \ \text{for a.e. } \omega \\ & \mathbf{x} \in \mathcal{X}. \end{aligned} \qquad (\mathrm{SMPEC}^{\mathrm{as}})$$

Assumptions I

Assumption (Properties of $f, F, \mathcal{X}, \mathcal{Y}$)

(a.i) $f(\bullet, \mathbf{y}(\bullet))$ is L_0-Lipschitz continuous on $\mathcal{X} + \eta_0\mathbb{B}$ for some $\eta_0 > 0$. $f(\mathbf{x}, \bullet)$ is Lipschitz with the parameter $\tilde{L}_0 > 0$ for all $\mathbf{x} \in \mathcal{X} + \eta_0\mathbb{B}$ for some $\eta_0 > 0$.
(a.ii) $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^m$ are nonempty, closed, and convex sets.
(a.iii) $F(\mathbf{x}, \bullet)$ is a μ_F-strongly monotone and L_F-Lipschitz continuous map on $\mathcal{Y}$ uniformly in $\mathbf{x} \in \mathcal{X}$.

Consider the problem (SMPEC$^{\mathrm{as}}$).
(b.i) $\tilde{f}(\bullet, \mathbf{y}(\bullet, \omega), \omega)$ is L_0-Lipschitz continuous on $\mathcal{X} + \eta_0\mathbb{B}$ for every $\omega \in \Omega$ and for some $\eta_0 > 0$. $f(\mathbf{x}, \bullet)$ be Lipschitz with the parameter $\tilde{L}_0 > 0$ for all $\mathbf{x} \in \mathcal{X} + \eta_0\mathbb{B}$ for some $\eta_0 > 0$.
(b.ii) $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^m$ are closed and convex sets.
(b.iii) $G(\mathbf{x}, \bullet, \omega)$ is a μ_F-strongly monotone and L_F-Lipschitz continuous map on $\mathcal{Y}$ uniformly in $\mathbf{x} \in \mathcal{X}$ for every $\omega \in \Omega$. □

Recall that F is said to be μ-strongly monotone if $(F(x) - F(y))^T(x - y) \geq \eta\|x - y\|^2$ for any $x, y \in \mathcal{X}$.

Assumptions II

Proposition (Patrikkson and Wynter, 1999)

Consider the problem (SMPEC$^{\rm as}$). Suppose Assumption **??** (a.ii, a.iii) hold. Suppose $\tilde{f}(\bullet, \bullet, \omega)$ is continuously differentiable on $\mathcal{C} \times \mathbb{R}^m$ where $\mathcal{C}$ is an open set containing $\mathcal{X}$, and $\mathcal{X}$ is bounded. Then the function f, defined as $f(\mathbf{x}) \triangleq \mathbb{E}[\tilde{f}(\mathbf{x}, \mathbf{y}(\mathbf{x}, \omega), \omega)]$, is Lipschitz and directionally differentiable on $\mathcal{X}$.

Algorithmic framework

Consider the problem (SMPEC$^{\rm exp}$). When the lower-level problem is single-valued, this problem reduces to the following.

$$\min_{\mathbf{x}\in\mathcal{X}} \; f(\mathbf{x},\mathbf{y}(\mathbf{x})), \tag{3}$$

where $\tilde{\mathbf{y}} \equiv \mathbf{y}(\mathbf{x})$ is the unique solution to the strongly monotone variational inequality problem:

$$(\hat{\mathbf{y}}-\tilde{\mathbf{y}})^\top F(\mathbf{x},\tilde{\mathbf{y}}) \geq 0, \qquad \forall \hat{\mathbf{y}} \in \mathcal{Y}. \tag{4}$$

Consider the smoothing of f given by f_η, defined as

$$f_\eta(\mathbf{x},\mathbf{y}(\mathbf{x})) \triangleq \mathbb{E}_{u\in\mathbb{B}}[f(\mathbf{x}+\eta u,\mathbf{y}(\mathbf{x}+\eta u))], \tag{G-Smooth}$$

Background on smoothing

Randomized or convolution-based smoothing goes back to Steklov (1907) and Sobolev.

- Nemirovski and Yudin [1983]
- Norkin [1995]
- .....
- De Farias and Lakshmanan [2005]
- Fixed smoothing for stochastic convex optimization: Yousefian, Nedić, and S [2010,2012], Duchi [2012]
- Extensions to diminishing smoothing with application to games [Yousefian, Nedić and S, 2016...]
- Fixed smoothing: Zeroth-order deterministic methods [Nesterov and Spokoiny [2017] (This has some interesting relationships to work on FD/SP methods (see work by J. Spall on SPSA).

Algorithm definition I

A naive projected gradient framework

$$\mathbf{x}_{k+1} := \Pi_{\mathcal{X}} \left[\mathbf{x}_k - \gamma_k \nabla_{\mathbf{x}} f(\mathbf{x}_k, \mathbf{y}(\mathbf{x}_k)) \right]. \quad (5)$$

Problem. $f(\bullet, y(\bullet))$ is not smooth. In fact, we do not even have access to a subgradient even if $f(\bullet, y(\bullet))$ is convex over $\mathcal{X}$.

Resolution. Employ convolution-based smoothing.

Lemma (Properties of spherical smoothing)

Suppose $h : \mathbb{R}^n \to \mathbb{R}$ is a continuous function and $\eta > 0$ is a given scalar. Let h_η be defined as

$$h_\eta(\mathbf{x}) \triangleq \mathbb{E}_{u \in \mathbb{B}}[h(\mathbf{x} + \eta u)].$$

(i) The smoothed function h_η is C^1 over $\mathcal{X}$ and for any $\mathbf{x} \in \mathcal{X}$, where

$$\nabla_{\mathbf{x}} h_\eta(\mathbf{x}) = \left(\frac{n}{\eta}\right) \mathbb{E}_{v \in \eta\mathbb{S}} \left[\frac{v((h(\mathbf{x}+v) - h(\mathbf{x}))}{\|v\|}\right]. \tag{6}$$

(ii) Suppose h is L_0-Lipschitz, i.e. $h \in C^{0,0}(\mathcal{X}_\eta)$. For any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$,

$$|h_\eta(\mathbf{x}) - h_\eta(\mathbf{y})| \leq L_0 \|\mathbf{x} - \mathbf{y}\|, \quad (h_\eta \text{ is } L_0\text{-Lipschitz}) \tag{7}$$

$$|h_\eta(\mathbf{x}) - h(\mathbf{x})| \leq L_0 \eta, \quad (\text{Bound on diff. between } h \text{ and } h_\eta) \tag{8}$$

$$\|\nabla_{\mathbf{x}} h_\eta(\mathbf{x}) - \nabla_{\mathbf{x}} h_\eta(\mathbf{y})\| \leq \frac{L_0 n}{\eta} \|\mathbf{x} - \mathbf{y}\|. \quad \left(\nabla_{\mathbf{x}} h_\eta \text{ is } \frac{L_0}{\eta}\text{-smooth}\right) \tag{9}$$

(iii) If h is convex and $h \in C^{0,0}(\mathcal{X}_\eta)$, then h_η is convex and $h(\mathbf{x}) \leq h_\eta(\mathbf{x}) \leq h(\mathbf{x}) + \eta L_0$ for any $\mathbf{x} \in \mathcal{X}$. Further $\nabla_x h_\eta(\mathbf{x}) \in \partial_\epsilon h(\mathbf{x})$ where $\epsilon \triangleq \eta L_0$.

(v) Suppose $h \in C^{0,0}(\mathcal{X}_\eta)$ with parameter L_0. For any $v \in \eta\mathbb{S}$, $g_\eta(\mathbf{x}, v) \triangleq \left(\frac{n}{\eta}\right) \frac{(h(\mathbf{x}+v) - h(\mathbf{x}))v}{\|v\|}$. Then, $\forall \mathbf{x} \in \mathcal{X}$, $\mathbb{E}_{v \in \eta\mathbb{S}}[\|g_\eta(\mathbf{x}, v)\|^2] \leq L_0^2 n^2$.

Comments

- Spherical smoothing studied by Nemirovski and Yudin (1983);
- Part (i) of our Lemma inspired by Flaxman et al. (2005) who considered Gaussian smoothing.
- Other parts either follow in a fashion similar to Gaussian smoothing (Nesterov and Spokoiny, 2017) or are directly proven.

Smoothing of implicit function f I

Consider the smoothing of f given by f_η, defined as

$$f_\eta(\mathbf{x}, \mathbf{y}(\mathbf{x})) \triangleq \mathbb{E}_{u \in \mathbb{B}}[f(\mathbf{x} + \eta u, \mathbf{y}(\mathbf{x} + \eta u))], \qquad \text{(G-Smooth)}$$

where u is uniformly distributed in the unit ball $\mathbb{B}$.

A gradient-based framework

$$\mathbf{x}_{k+1} := \Pi_{\mathcal{X}}\left[\mathbf{x}_k - \gamma_k g_\eta(\mathbf{x}_k, \mathbf{y}(\mathbf{x}_k))\right], \qquad (10)$$

where

$$g_\eta(\mathbf{x}, \mathbf{y}(\mathbf{x})) \triangleq \left(\frac{n}{\eta}\right) \mathbb{E}_{v \in \eta\mathbb{S}}\left[\frac{v((f(\mathbf{x}+v, \mathbf{y}(x+v)) - f(\mathbf{x}, \mathbf{y}(\mathbf{x})))}{\|v\|}\right].$$

Asymptotic guarantees

Lack of asymptotics. Unforunately, a fixed η leads to approximate solutions.

Introduce iterative smoothing with $\eta_k \downarrow 0$ at suitable rate.

A iteratively smoothed gradient-based framework

$$\mathbf{x}_{k+1} := \Pi_{\mathcal{X}} \left[\mathbf{x}_k - \gamma_k g_{\eta_k}(\mathbf{x}_k, \mathbf{y}(\mathbf{x}_k)) \right]. \tag{11}$$

Exact gradients unavailable

Lack of access to $g_\eta(\mathbf{x}, \mathbf{y}(\mathbf{x}))$. Utilize an unbiased estimate given by $g_\eta(\mathbf{x}, \mathbf{y}(\mathbf{x}), v)$, defined as

$$g_\eta(\mathbf{x}, \mathbf{y}(\mathbf{x}), v) \triangleq \left(\frac{n}{\eta}\right) \left[\frac{(f(\mathbf{x}+v, \mathbf{y}(\mathbf{x}+v)) - f(\mathbf{x}, \mathbf{y}(\mathbf{x})))\, v}{\|v\|}\right]. \tag{12}$$

Introduce iterative smoothing + sampling.

A iteratively smoothed sampled gradient framework

$$\mathbf{x}_{k+1} := \Pi_{\mathcal{X}}\left[\mathbf{x}_k - \gamma_k g_{\eta_k}(\mathbf{x}_k, \mathbf{y}(\mathbf{x}_k), v_k)\right]. \tag{13}$$

Exact evaluation of $\mathbf{y}(\mathbf{x})$ unavailable

Problem. Unfortunately $\mathbf{y}(\mathbf{x})$ is unavailable in closed-form.

Introduce iterative smoothing + sampling + inexact solutions of $\mathbf{y}(\mathbf{x})$.

An inexact iteratively smoothed sampled gradient framework

$$\mathbf{x}_{k+1} := \Pi_{\mathcal{X}} \left[\mathbf{x}_k - \gamma_k \left(\nabla_{\mathbf{x}} f_{\eta_k}(\mathbf{x}_k, \mathbf{y}_{\epsilon_k}(\mathbf{x}_k), v_k)\right)\right]. \tag{14}$$

where $\mathbb{E}\left[\|\mathbf{y}_{\epsilon_k}(\mathbf{x}_k) - \mathbf{y}(\mathbf{x}_k)\|^2 \mid \mathbf{x}_k\right] \leq \epsilon_k$ and $\mathbf{y}_{\epsilon_k}(\mathbf{x}_k) \triangleq \mathbf{y}(\mathbf{x}_k) + \tilde{\epsilon}_k$.

Challenge. Unfortunately, $\tilde{\epsilon}_k$ is not necessarily conditionally zero mean

Computing approximate values of $\mathbf{y}(\mathbf{x})$

1. If $F(\mathbf{x}, \mathbf{y}(\mathbf{x})) \triangleq \left(\mathbb{E}[G_i(\mathbf{x}, \mathbf{y}(\mathbf{x}), \omega)]\right)_{i=1}^n$, $\mathbf{y}(\mathbf{x})$ is a solution to the parametrized stochastic variational inequality problem $\mathrm{VI}(\mathcal{Y}, F(\mathbf{x}, \bullet))$.

$$(\tilde{\mathbf{y}} - \mathbf{y}(\mathbf{x}))^T F(\mathbf{x}, \mathbf{y}(\mathbf{x})) \geq 0, \qquad \forall \tilde{y} \in \mathcal{Y} \qquad (\mathrm{VI}(\mathcal{Y}, F(\mathbf{x}, \bullet)))$$

2. Generate random realizations of the stochastic mapping $G(\hat{\mathbf{x}}_k, \mathbf{y}_t, \omega_{\ell,t})$ for $\ell = 1, \ldots, M_t$

$$\mathbf{y}_{t+1} := \Pi_{\mathcal{Y}} \left[\mathbf{y}_t - \alpha \frac{\sum_{\ell=1}^{M_t} G(\hat{\mathbf{x}}_k, \mathbf{y}_t, \omega_{\ell,t})}{M_t} \right] \tag{15}$$

3. Under some conditions, we may recover linear rates of convergence, i.e. after t_k steps of (**??**), we obtain that

$$\mathbb{E}[\|\mathbf{y}_\epsilon(\mathbf{x}_k) - \mathbf{y}(\mathbf{x}_k)\|^2] \leq C q^{t_k} \triangleq \epsilon_k. \tag{16}$$

An inexact zeroth-order framework

Given an $\mathbf{x}_0 \in \mathcal{X}$, for $k = 0, 1, 2, \cdots$

$$\mathbf{x}_{k+1} := \Pi_{\mathcal{X}} \left[\mathbf{x}_k - \gamma_k g_{\eta_k}(\mathbf{x}_k, \mathbf{y}_{\epsilon_k}(\mathbf{x}_k), v_k) \right], \tag{17}$$

where $\{\mathbf{y}_{\epsilon_k}(\mathbf{x}_k)\}$ represents an increasingly accurate approximation of $\{y(x_k)\}$.

Properties of inexact zeroth-order gradient

Lemma (**Properties of the inexact zeroth-order gradient**)

Suppose $F(\mathbf{x},\bullet)$ is a μ_F-strongly monotone and Lipschitz continuous map on $\mathcal{Y}$ uniformly on $\mathcal{X}$. In addition, $f(\mathbf{x},\bullet)$ is $\tilde{L}_0$-Lipschitz for all $\mathbf{x} \in \mathcal{X} + \eta_0\mathbb{B}$.

Suppose $\mathbb{E}[\|\mathbf{y}_\epsilon(\mathbf{x}) - \mathbf{y}(\mathbf{x})\|^2 \mid \mathbf{x}] \leq \epsilon$ almost surely for any $\mathbf{x} \in \mathcal{X}$.

(a) $\mathbb{E}[\|g_{\eta,\epsilon}(\mathbf{x},v)\|^2 \mid \mathbf{x}] \leq 3n^2\left(\frac{2\tilde{L}_0^2\epsilon}{\eta^2} + L_0^2\right)$, a.s. .

(b) $\mathbb{E}\left[\|g_{\eta,\epsilon}(\mathbf{x},v) - g_\eta(\mathbf{x},v)\|^2 \mid \mathbf{x}\right] \leq \frac{4\tilde{L}_0^2 n^2\epsilon}{\eta^2}$, a.s. .

Setting: $f(\bullet, \mathbf{y}(\bullet))$ is convex on $\mathcal{X}$

Definition (Parameters)

1. $\gamma_k := \frac{\gamma_0}{\sqrt{k+1}}$ *and* $\eta_k := \frac{\eta_0}{\sqrt{k+1}}$*, respectively for all* $k \geq 0$*.*
2. *Suppose* $\alpha \leq \frac{\mu_F}{2L_F^2}$, $M_t := \lceil M_0 \rho^{-t} \rceil$ *for* $t \geq 0$ *for* $\rho \in (0,1)$ *and* $M_0 \geq \frac{2\nu_{\mathbf{y}}^2}{L_F^2}$*.*
3. *Let* $t_k := \lceil \tau \ln(k+1) \rceil$ *where* $\tau \geq \frac{-2}{\ln(\max\{1-\mu_F\alpha, \rho\})}$*.*

Theorem (**Rate statements and complexity results**)

Consider the sequence $\{\bar{\mathbf{x}}_k\}$ generated by applying inexact zeroth-order scheme.

(a) For all K, we have $\mathbb{E}\left[f(\bar{\mathbf{x}}_K, \mathbf{y}(\bar{\mathbf{x}}_K))\right] - f^* \leq \mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$.

(b) Suppose $\gamma_0 = \mathcal{O}(\frac{1}{nL_0})$ and $r = 0$. Let $\epsilon > 0$ be an arbitrary scalar and K_ϵ be such that $\mathbb{E}\left[f(\bar{\mathbf{x}}_{K_\epsilon}, \mathbf{y}(\bar{\mathbf{x}}_{K_\epsilon}))\right] - f^* \leq \epsilon$. Then the sample complexity of

(b-1) upper-level evals. of $\mathbf{y}(\bullet)$ is $\mathcal{O}\left(n^2 L_0^2 \epsilon^{-2}\right)$.

(b-2) lower-level evals. is $\mathcal{O}\left(n^{2\bar{\tau}} L_0^{2\bar{\tau}} \epsilon^{-2\bar{\tau}}\right)$, where $\bar{\tau} \geq 1 - \tau \ln(\rho)$.

An inexact accelerated zeroth-order scheme

In this part, we consider the following (SMPEC).

$$\min_{\mathbf{x}\in\mathcal{X}} \ \mathbb{E}[\tilde{f}(\mathbf{x},\mathbf{y}(\mathbf{x},\omega))] \tag{SMPEC$^{\rm as}$}$$

This problem is equivent to this possibly semi-infinite MPEC.

$$\min_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[\tilde{f}(\mathbf{x},\mathbf{y}(\omega))] \tag{SMPEC$^{\rm as}$}$$

$$\text{subject to } \mathbf{y}(\omega) \text{ solves } \mathrm{VI}(\mathcal{Y}, G(\mathbf{x},\bullet,\omega)) \text{ for every } \omega \in \Omega. \tag{18}$$

Given an $\mathbf{x}_0 \in \mathcal{X}$, for $k = 0, 1, 2, \cdots$

$$\begin{aligned} \mathbf{z}_{k+1} &:= \Pi_{\mathcal{X}}\left[\mathbf{x}_k - \gamma_k g_{\eta_k,N_k}(\mathbf{x}_k, v_k)\right] \\ \mathbf{x}_{k+1} &= \mathbf{z}_{k+1} + \beta_k\left(\mathbf{z}_{k+1} - \mathbf{z}_k\right), \end{aligned} \tag{19}$$

where $g_{\eta_k,N_k}(\mathbf{x}_k, v_k)$ represents a mini-batch zeroth-order gradient estimator.

Inexact accelerated zeroth-order method

Proposition (Rate and complexity statement)

Suppose $\eta_k = \frac{1}{k}, \gamma_k = \frac{1}{2k}$, and $N_k = \lfloor k^a \rfloor$ for $a = 1 + \delta$.

(a) For any k, $\mathbb{E}[f(\mathbf{z}_k) - f(\mathbf{x}^*)] \leq \mathcal{O}\left(\frac{1}{k}\right)$.

(b) Suppose K^ϵ is such that $\mathbb{E}[f(\mathbf{z}_k) - f(\mathbf{x}^*)] \leq \epsilon$. Then $\sum_{k=1}^{K_\epsilon} N_k \leq \mathcal{O}(1/\epsilon^{2+\delta})$.

Setting: $f(\bullet, \mathbf{y}(\bullet))$ is not necessarily convex

Problem of interest where $\mathbf{y}(\mathbf{x})$ is a solution to a stochastic VI.

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{X}} \ & f(\mathbf{x}, \mathbf{y}(\mathbf{x})) \\ \text{subject to } & \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{20}$$

Challenge: Nonsmoothness, nonconvexity, and stochasticity.

Smoothing. Consider the smoothed implicit problem.

$$\begin{aligned} \min_{\mathbf{x} \in \mathcal{X}} \ & f_{\eta}(\mathbf{x}, \mathbf{y}(\mathbf{x})) \\ \text{subject to } & \mathbf{x} \in \mathcal{X}. \end{aligned} \tag{21}$$

Stationarity for nonsmooth nonconvex problems

We begin by defining the directional derivative, a key object necessary in addressing nonsmooth and possibly nonconvex optimization problems.

Definition (Clarke (1998))

The directional derivative of h at $\mathbf{x}$ in a direction v is defined as

$$h^\circ(\mathbf{x}, v) \triangleq \limsup_{\mathbf{y}\to\mathbf{x}, t\downarrow 0} \left(\frac{h(\mathbf{y} + tv) - h(\mathbf{y})}{t} \right). \tag{22}$$

The Clarke generalized gradient at $\mathbf{x}$ can then be defined as

$$\partial h(\mathbf{x}) \triangleq \{\xi \in \mathbb{R}^n \mid h^\circ(\mathbf{x}, v) \geq \langle \xi, v\rangle, \quad \forall v \in \mathbb{R}^n\}. \tag{23}$$

1. If h is differentiable, $\partial_{\mathbf{x}} h(\mathbf{x}) = \nabla_{\mathbf{x}} h(\mathbf{x})$.
2. If h is convex, this object reduces to the standard subdifferential.

ϵ-Clarke generalized gradient

Proposition (Properties of Clarke generalized gradients (Clarke, 1998))

Suppose h is Lipschitz continuous on $\mathbb{R}^n$. Then the following hold.

(i) $\partial h(\mathbf{x})$ is a nonempty, convex, and compact set and $\|g\| \leq L$ for any $g \in \partial h(\mathbf{x})$.

(ii) h is differentiable almost everywhere.

(iii) $\partial h(\mathbf{x})$ is an upper semicontinuous map defined as

$$\partial h(\mathbf{x}) = \text{conv}\left\{ g \mid g = \lim_{k \to \infty} \nabla_{\mathbf{x}} h(\mathbf{x}_k), \mathcal{C}_h \not\ni \mathbf{x}_k \to \mathbf{x} \right\}.$$

We may also define the ϵ-generalized gradient (Goldstein (1977)) as

$$\partial_\epsilon h(\mathbf{x}) \triangleq \text{conv}\left\{ \xi : \xi \in \partial h(\mathbf{y}), \|\mathbf{x} - \mathbf{y}\| \leq \epsilon \right\}. \tag{24}$$

Stationarity of smoothed problem vs $\epsilon-$Clarke stationarity

Proposition

Suppose h is locally Lipschitz and $\mathcal{X} \subseteq \mathbb{R}^n$ is closed, convex, and bounded.

(i) (Unconstrained, Goldstein 1977) For any $\eta > 0$ and any $\mathbf{x} \in \mathbb{R}^n$, $\nabla h_\eta(\mathbf{x}) \in \partial_{2\eta} h(\mathbf{x})$. If $0 \notin \partial h(\mathbf{x})$, $\exists \eta$ such that $\nabla_{\mathbf{x}} h_{\tilde{\eta}}(\mathbf{x}) \neq 0$ for $\tilde{\eta} \in (0, \eta]$.

(ii) (Constrained) For any $\eta > 0$ and any $\mathbf{x} \in \mathcal{X}$,

$$[0 \in \nabla_{\mathbf{x}} h_\eta(\mathbf{x}) + \mathcal{N}_{\mathcal{X}}(\mathbf{x})] \implies [0 \in \partial_{2\eta} h(\mathbf{x}) + \mathcal{N}_{\mathcal{X}}(\mathbf{x})] \,. \tag{25}$$

In short: **Stationarity of $\eta-$smoothed problem $\implies$ 2η-Clarke stationarity.**

Stationarity and relation to original problem

Definition (The residual mappings)

Given $\beta > 0$ and $\eta > 0$, for any $\mathbf{x} \in \mathbb{R}^n$. Let the residual mappings $G_{\eta,\beta}(\mathbf{x})$ and $\tilde{G}_{\eta,\beta}(\mathbf{x})$ be defined as follows where $\tilde{e} \in \mathbb{R}^n$ is an arbitrary given vector.

$$G_{\eta,\beta}(\mathbf{x}) \triangleq \beta\left(\mathbf{x} - \Pi_{\mathcal{X}}\left[\mathbf{x} - \tfrac{1}{\beta}\nabla_x f_\eta(\mathbf{x},\mathbf{y}(\mathbf{x}))\right]\right), \quad (26)$$

$$\tilde{G}_{\eta,\beta}(\mathbf{x}) \triangleq \beta\left(\mathbf{x} - \Pi_{\mathcal{X}}\left[\mathbf{x} - \tfrac{1}{\beta}(\nabla_x f_\eta(\mathbf{x},\mathbf{y}(\mathbf{x})) + \tilde{e})\right]\right), \quad (27)$$

Lemma

For any $\eta, \beta > 0$,

$$[G_{\eta,\beta}(\mathbf{x}) = 0] \iff [0 \in \nabla_{\mathbf{x}} f_\eta(\mathbf{x},\mathbf{y}(\mathbf{x})) + \mathcal{N}_{\mathcal{X}}(\mathbf{x})] \implies [0 \in \partial_{2\eta} f_\eta(\mathbf{x},\mathbf{y}(\mathbf{x})) + \mathcal{N}_{\mathcal{X}}(\mathbf{x})]$$

Rate and complexity statements

Theorem (Rate statements and complexity results)

For any $\gamma < \frac{\eta}{nL_0}$, $\ell \triangleq \lceil \lambda K \rceil$, and all $K > \frac{2}{1-\lambda}$, $\mathbb{E}\left[\|G_{\eta,1/\gamma}(\mathbf{x}_R)\|^2\right] \leq \mathcal{O}\left(\frac{1}{K}\right)$.

Suppose $\gamma = \frac{\eta}{2nL_0}$ and $\eta = \frac{1}{L_0}$. Let $\epsilon > 0$ be an arbitrary scalar and K_ϵ be such that $\mathbb{E}\left[\|G_{\eta,1/\gamma}(\mathbf{x}_R)\|^2\right] \leq \epsilon$. Then the sample complexity

1. of upper-level projs. is $\mathcal{O}\left(n^4 L_0^4 \epsilon^{-2}\right)$.
2. of lower-level projs is $\mathcal{O}\left(n^6 L_0^6 \epsilon^{-3}\right)$.

Numerics: A stochastic Stackelberg-Nash-Cournot game

For a given $x \geq 0$, let $(q_1(x,\omega), \ldots, q_N(x,\omega))$ be a set of quantities for every $\omega \in \Omega$ and each $q_i(x,\omega)$ solve the following, assuming that $q_j(x,\omega)$, $j \neq i$ are fixed:

$$\max_{q_i \geq 0} \quad q_i p\left(q_i + x + \textstyle\sum_{j=1, j\neq i}^{N} q_j(x,\omega), \omega\right) - f_i(q_i). \qquad \text{(Stack-follower}_i\text{)}$$

Accordingly, let $Q(x,\omega) \triangleq \sum_{i=1}^{N} q_i(x,\omega)$. Then $(x^*, q(x^*))$ is said to be a Stackelberg-Nash-Cournot equilibrium solution if x^* solves

$$\max_{0 \leq x \leq x^u} \mathbb{E}[xp(x + Q(x,\omega),\omega)] - f(x), \qquad \text{(Stack-leader)}$$

where $p(x,\omega) = a(\omega) - bx$ and let $f_i(q) = \frac{1}{2}cq^2$ for $i = 1, \cdots, N$, and $f(x) = \frac{1}{2}dx^2$.

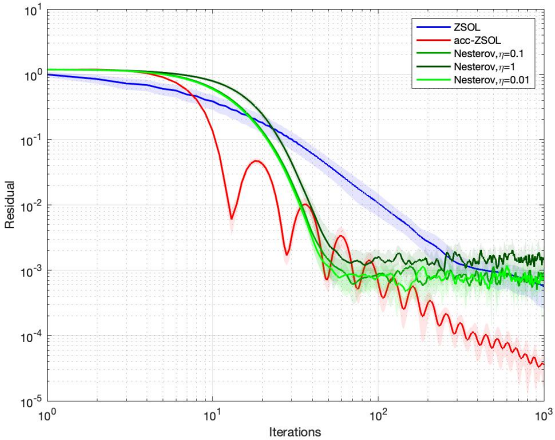


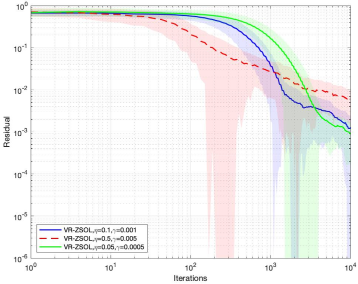
Figure: Trajectories for (ZSOL), (ac-ZSOL) and Nesterov on the convex SMPEC$^{\text{as}}$

Table: Errors and CPU time comparison of the three schemes with different parameters

			(ZSOL)		(acc-ZSOL)		(SAA)	
			$\mathbb{E}[f^* - f(\bar{x})]$	CPU	$\mathbb{E}[f^* - f(\bar{x})]$	CPU	$\mathbb{E}[f^* - f(x)]$	CPU
$N = 10$	$b = 1$	$c = 0.05$	1.3e-3	0.8	5.2e-5	5.1	2.5e-4	16.8
		$c = 0.1$	5.8e-4	0.9	2.2e-5	5.4	3.5e-4	20.4
	$b = 0.5$	$c = 0.05$	1.0e-3	0.8	4.8e-5	5.5	1.9e-4	19.1
		$c = 0.1$	1.1e-3	0.8	5.8e-5	5.3	4.2e-4	18.5
$N = 15$	$b = 1$	$c = 0.05$	7.8e-4	1.0	2.9e-5	5.3	1.4e-4	58.3
		$c = 0.1$	5.5e-4	1.0	2.1e-5	5.5	2.3e-4	54.3
	$b = 0.5$	$c = 0.05$	9.9e-4	1.0	2.5e-5	5.0	1.5e-4	49.9
		$c = 0.1$	8.2e-4	1.0	2.0e-5	5.1	8.4e-4	49.2
$N = 20$	$b = 1$	$c = 0.05$	4.3e-4	1.3	2.5e-5	5.3	1.3e-4	152.7
		$c = 0.1$	3.9e-4	1.3	1.1e-5	5.1	3.1e-4	123.1
	$b = 0.5$	$c = 0.05$	6.8e-4	1.2	2.0e-5	5.2	2.8e-4	129.7
		$c = 0.1$	4.0e-4	1.3	3.1e-5	5.1	7.2e-4	101.3

The error and CPU time are the average results of 20 runs

$$\begin{aligned}
\underset{x}{\text{minimize}} \quad & -x_1^2 - 3x_2 - 4y_1 + y_2^2 \\
\text{subject to} \quad & x_1^2 + 2x_2 \leq 4 \\
& 0 \leq x_1 \leq 1 \\
& 0 \leq x_2 \leq 2 \\
& \underset{y}{\text{minimize}} \quad \mathbb{E}[2x_1^2 + y_1^2 + y_2^2 - \xi(\omega)y_2] \\
& \text{subject to} \quad x_1^2 - 2x_1 + x_2^2 - 2y_1 + y_2 \geq -3 \\
& \qquad\qquad\quad x_2 + 3y_1 - y_2 \geq 4 \\
& \qquad\qquad\quad y_1 \geq 0, y_2 \geq 0,
\end{aligned}$$



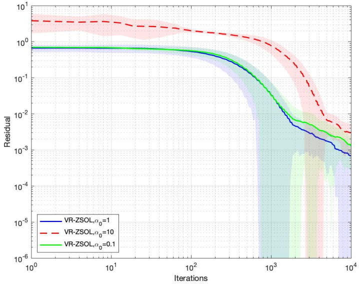


Figure: Trajectories for (VR-ZSOL) on the non-convex $\text{SMPEC}^{\text{exp}}$

Table: Errors comparison of the three schemes with different parameters

		ZSOL	NLPEC	BARON
		$\mathbb{E}[f(\bar{x})]$	local optimum	global optimum
$(a,b)=(1,0)$	$(c,d)=(1,1)$	-7.50	-7.20	-7.50
	$(c,d)=(2,2)$	-9.23	-9.04	-9.23
	$(c,d)=(3,3)$	-9.25	-9.10	-9.25
$(a,b)=(5,0)$	$(c,d)=(1,1)$	-11.50	-7.20	-11.50
	$(c,d)=(2,2)$	-13.23	-9.04	-13.23
	$(c,d)=(3,3)$	-13.25	-9.10	-13.25
$(a,b)=(10,0)$	$(c,d)=(1,1)$	-16.48	-7.20	-16.50
	$(c,d)=(2,2)$	-18.20	-9.04	-18.23
	$(c,d)=(3,3)$	-18.23	-9.10	-18.25

The error of (ZSOL) in the table is the average results of 20 runs

Definition of game

We consider an N-player noncooperative game where the ith player's problem is defined as the following hierarchical optimization problem.

$$\min_{\mathbf{x}^i \in \mathcal{X}^i} \; f_i(\mathbf{x}^i, \mathbf{x}^{-i}) \triangleq \mathbb{E}[\tilde{g}_i(\mathbf{x}^i, \mathbf{x}^{-i}, \omega)] + \mathbb{E}[\tilde{h}_i(\mathbf{x}^i, \mathbf{y}^i(\mathbf{x}, \omega), \omega)] \quad (\text{Player}_i(\mathbf{x}^{-i}))$$

- Player problems are stochastic and nonconvex in full space

Smoothed games I

Definition (An η-smoothed noncooperative game $\mathcal{G}_\eta$)

Consider a game $\mathcal{G} \in \mathcal{G}^{\rm cbl}_{\rm pot}$ in which the ith player solves (Player$_i(\mathbf{x}^{-i})$). Suppose $\mathcal{G}_\eta$ denotes a related game in which for $i = 1, \cdots, N$, the ith player's smoothed problem is defined as .

$$\min_{\mathbf{x}^i \in \mathcal{X}^i} \ f_{i,\eta}(\mathbf{x}^i, \mathbf{x}^{-i}), \qquad (\text{Player}_{i,\eta}(\mathbf{x}^{-i}))$$

where $f_{i,\eta}(\mathbf{x}^i, \mathbf{x}^{-i}) \triangleq \mathbb{E}[f_i(\mathbf{x}^i + u_i, \mathbf{x}^{-i})]$ where $\mathbb{B}_i \subseteq \mathbb{R}^{n_i}$ is a sphere centered at the origin.

Smoothed games II

$$B_i(\mathbf{x}) \triangleq \underset{\mathbf{v}^i \in \mathcal{X}_i}{\operatorname{argmin}} \ \left[f_i(\mathbf{v}^i, \mathbf{x}^{-i}) + \tfrac{c}{2}\|\mathbf{v}^i - \mathbf{x}^i\|^2 \right]. \qquad (\mathrm{PBR}_i(\mathbf{x}))$$

Similarly, we may define the η-*smoothed* proximal best-response of player i as follows.

$$B_{i,\eta}(\mathbf{x}) \triangleq \underset{\mathbf{v}^i \in \mathcal{X}_i}{\operatorname{argmin}} \ \left[f_{i,\eta}(\mathbf{v}^i, \mathbf{x}^{-i}) + \tfrac{c}{2}\|\mathbf{v}^i - \mathbf{x}^i\|^2 \right]. \qquad (\mathrm{SPBR}_{i,\eta}(\mathbf{x}))$$

Smoothed games III

Proposition (Fixed-point of (SPBR_η) is NE of $\mathcal{G}_\eta$)

Consider an N-player noncooperative game $\mathcal{G}$ where the ith player solves ($\text{Player}_i(\mathbf{x}^{-i})$), given rival decisions $\mathbf{x}^{-i}$. For $i = 1, \cdots, N$, suppose $f_i(\bullet, \mathbf{x}^{-i})$ is convex on $\mathcal{X}_i$ for any $\mathbf{x}^{-i} \in \mathcal{X}_{-i}$. Suppose $\mathbf{x}^\eta \triangleq \{\mathbf{x}^{1,\eta}, \cdots, \mathbf{x}^{N,\eta}\}$ is a fixed point of the η-smoothed best-response map. Then the following hold.

(a) $\mathbf{x}^\eta$ is a fixed point of ($\text{SPBR}_\eta(\bullet)$), i.e. $\mathbf{x}^{i,\eta}$ is a minimizer of $\text{SPBR}_{i,\eta}(\mathbf{x}^{-i,\eta})$ for $i = 1, \cdots, N$ if and only if $\mathbf{x}^\eta$ is a Nash equilibrium of $\mathscr{G}_\eta$.

(b) If $\mathbf{x}^\eta$ is a fixed point of $\text{SPBR}_\eta(\bullet)$, then $\mathbf{x}^\eta$ is an $\eta\bar{\beta}$-Nash equilibrium of $\mathscr{G}$ where $\bar{\beta} \triangleq \max_{i \in \{1,\cdots,N\}} \beta_i$.

Research gaps I

- No available schemes for computing equilibria for relatively general settings
- Approximations where subproblems are solved as NLPs and joint necessary conditions are resolved [Leyffer and Munson [2005]]; SAA schemes [De Miguel and Xu [2009]; again SAA problems are MPECs for which stationary points are available.

Are there asymptotics for an efficient algorithm for computing an ϵ-equilibrium?

An asynchronous inexact proximal best-response scheme

Asynchronous relaxed smoothed proximal best-response (ARSPBR) scheme

(0) Let $k = 0$, $\mathbf{z}^{i,0} = \mathbf{x}^{i,0} \in \mathcal{X}_i$ for $i = 1, \cdots, N$, and $p_i \in (0,1)$ for $i = 1, \cdots, N$ with $\sum_{i=1}^N p_i = 1$. Given $\eta > 0$ and relax. seq. $\{\gamma_k\}$.

(1) Select a player $i_k = i \in \{1, \cdots, N\}$ with probability $p_i > 0$.

(2) Update $\mathbf{z}^{k+1}$ and $\mathbf{x}^{k+1}$ as follows.

$$\mathbf{z}^{i,k+1} := \begin{cases} (1-\gamma_k)\mathbf{x}^{i,k} + \gamma_k \left(B_{i,\eta}(\mathbf{x}^k)\right); & i = i_k \\ \mathbf{x}^{i,k}; & i \neq i_k \end{cases}$$
$$\mathbf{x}^{i,k+1} := \begin{cases} \mathbf{z}^{i,k+1} + \epsilon^{i,k+1}; & i = i_k \\ \mathbf{z}^{i,k+1}; & i \neq i_k. \end{cases} \quad \text{(ARSPBR)}$$

(3) Stop if $k > K$, Stop; else return to Step 1, $k := k+1$.

Almost sure convergence guarantees available.

Proposition (**Almost-sure convergence for asynchronous relaxed inexact best-response scheme**)

Consider a game $\mathscr{G} \in \mathcal{G}^{\rm chl}$. For any $i \in \{1, \cdots, N\}$, suppose $f_i(\bullet, \mathbf{x}^{-i})$ is a convex function for any $\mathbf{x}^{-i} \in \mathcal{X}_{-i}$ and $\mathcal{X}_i \subseteq \mathbb{R}^{n_i}$ is a closed and convex set. Consider the smoothed counterpart of $\mathscr{G}$, denoted by $\mathscr{G}_\eta$ where $\mathscr{G}_\eta \in \mathcal{G}^{\rm chl}_{\rm pot}$; for any $i \in \{1, \cdots, N\}$, suppose $f_{i,\eta}(\bullet, \mathbf{x}^{-i})$ denotes the η-smoothing of $f_i(\bullet, \mathbf{x}^{-i})$. Suppose P_η denotes the potential function of $\mathscr{G}_\eta$ where $P_\eta(\mathbf{x}) \geq \tilde{P}$ for any $\mathbf{x} \in \mathcal{X} + \eta\mathbb{B}$, where $\tilde{P}$ denotes a lower bound on P_η. Suppose B_i and $\mathrm{B}_{i,\eta}$ denote the proximal best-response and smoothed proximal-response for $i \in \{1, \cdots, N\}$. Then the following hold for any $i \in \{1, \cdots, N\}$. Consider a sequence $\{\mathbf{x}^k\}$ generated by (ASRPBR) scheme. Then the following hold.

(a) For $k \geq 0$, the following holds almost surely.

$$\mathbb{E}[P_\eta(\mathbf{x}^{k+1}) - \tilde{P}_\eta \mid \mathcal{F}_k] \leq (P_\eta(\mathbf{x}^k) - \tilde{P}_\eta) - \gamma_k \left(c - \frac{L\gamma_k}{2}\right) \|B_{i,\eta}(\mathbf{x}^k) - \mathbf{x}^{i,k}\|^2 + \sum_{i=1}^N M_i \mathbb{E}[\|\epsilon^{i,k+1}\| \mid \mathcal{F}_k]. \quad (28)$$

(b) Suppose one of the following hold. (i) $\{\gamma_k\}$ is a decreasing non-summable but square-summable sequence where $\gamma_k < \frac{2c}{L}$ for every k; (ii) $\gamma_k = \gamma = 1$ and $c > \frac{L}{2}$. Furthermore, suppose $\sum_{k=0}^\infty \sum_{i=1}^N M_i \mathbb{E}[\|\epsilon^{i,k+1}\| \mid \mathcal{F}_k] < \infty$. Then

$$\lim_{k\to\infty} \sum_{i=1}^N \|\mathbf{x}^{i,k} - B_{i,\eta}(\mathbf{x}^k)\|^2 = 0 \text{ almost surely}. \quad (29)$$

(c) Suppose (??) holds. Then $\{\mathbf{x}^k\}$ converges to the set of Nash equilibria of $\mathscr{G}_\eta$ in an a.s. sense.

Numerics

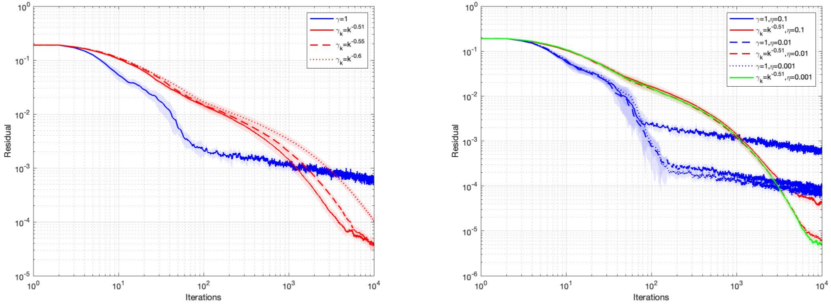


Figure: Trajectories for (**ARSPBR**) with different relaxation and smoothing parameters

Table: Errors and CPU time of (**ARSPBR**) with ($\gamma_k = 1$) and ($\gamma_k = k^{-0.51}$)

N	$\gamma_k = 1$		$\gamma_k = k^{-0.51}$	
	$\mathtt{res}(\mathbf{x}^k)$	CPU	$\mathtt{res}(\mathbf{x}^k)$	CPU
13	6.1e-4	7.6	4.0e-5	7.6
23	6.2e-4	9.3	1.3e-4	9.2
33	6.4e-4	11.4	4.8e-4	11.4
43	6.0e-4	15.4	6.3e-4	15.3

ζ_t	$\gamma_k = 1$		$\gamma_k = k^{-0.51}$	
	$\mathtt{res}(\mathbf{x}^k)$	CPU	$\mathtt{res}(\mathbf{x}^k)$	CPU
1e-2	6.1e-4	7.6	4.0e-5	7.6
5e-3	6.8e-4	7.4	4.8e-5	7.3
2e-3	7.2e-4	7.5	5.2e-5	7.4
1e-3	7.6e-4	7.7	5.6e-5	7.6

a	$\gamma_k = 1$		$\gamma_k = k^{-0.51}$	
	$\mathtt{res}(\mathbf{x}^k)$	CPU	$\mathtt{res}(\mathbf{x}^k)$	CPU
[33, 37]	6.1e-4	7.6	4.0e-5	7.6
[30, 40]	6.8e-4	7.7	4.3e-5	7.4
[25, 45]	7.2e-4	7.5	5.0e-5	7.6
[20, 50]	8.0e-4	7.7	5.1e-5	7.8

The error and CPU time in the table is the average results of 20 runs

Concluding remarks

Many questions/directions remain.

- Suppose lower-level problem loses strong monotonicity or even monotonicity.
- Schemes should apply for computing ϵ-Clarke stationary points of nonsmooth nonconvex stochastic optimization.
- Can we extend these zeroth-order schemes to other types of hierarchical simulation optimization problems?

- S. Cui*, UVS, and F. Yousefian, Complexity guarantees for an implicit smoothing-enabled method for stochastic MPECs, Mathematical Programming, 1–73, (2022).
- S. Cui* and UVS, On the computation of equilibria in monotone and potential stochastic hierarchical games, Mathematical Programming, 1–59, (2022).