Stochastic methods for stochastic variational inequalities

Philip Thompson
Santiago, 2017

Center for Mathematical Modeling (CMM) - Universidad de Chile

Workshop on Variational and Stochastic Analysis,
Santiago, March 16th, 2017

Joint work with: A. Iusem (IMPA), A. Jofré (CMM-DIM) and R. Oliveira (IMPA).

Expected value formulation of SVI

Expected value formulation of SVI

Assume

- **Random operator**: $F : \Xi \times \mathbb{R}^n \to \mathbb{R}^n$ is a Carathéodory map on a measurable space $(\Xi, \mathcal{G})$,

Expected value formulation of SVI

Assume

- **Random operator**: $F : \Xi \times \mathbb{R}^n \to \mathbb{R}^n$ is a Carathéodory map on a measurable space $(\Xi, \mathcal{G})$,
- **Randomness**: $\xi : \Omega \to \Xi$ is a r.v. on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$,

Expected value formulation of SVI

Assume

- **Random operator**: $F : \Xi \times \mathbb{R}^n \to \mathbb{R}^n$ is a Carathéodory map on a measurable space $(\Xi, \mathcal{G})$,
- **Randomness**: $\xi : \Omega \to \Xi$ is a r.v. on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$,
- **Feasible set**: $X \subset \mathbb{R}^n$ closed and convex.

Expected value formulation of SVI

Assume

- **Random operator**: $F : \Xi \times \mathbb{R}^n \to \mathbb{R}^n$ is a Carathéodory map on a measurable space $(\Xi, \mathcal{G})$,
- **Randomness**: $\xi : \Omega \to \Xi$ is a r.v. on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$,
- **Feasible set**: $X \subset \mathbb{R}^n$ closed and convex.

Definition (SVI)

Assuming $T : \mathbb{R}^n \to \mathbb{R}^n$ is given by $T = \mathbb{E}[F(\xi, \cdot)]$, find $x^* \in X$ s.t. $\forall x \in X$,

$$\langle T(x^*), x - x^* \rangle \geq 0.$$

(Solution set: X^*).

Expected value formulation of SVI

Assume

- **Random operator**: $F : \Xi \times \mathbb{R}^n \to \mathbb{R}^n$ is a Carathéodory map on a measurable space $(\Xi, \mathcal{G})$,
- **Randomness**: $\xi : \Omega \to \Xi$ is a r.v. on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$,
- **Feasible set**: $X \subset \mathbb{R}^n$ closed and convex.

Definition (SVI)

Assuming $T : \mathbb{R}^n \to \mathbb{R}^n$ is given by $T = \mathbb{E}[F(\xi, \cdot)]$, find $x^* \in X$ s.t. $\forall x \in X$,

$$\langle T(x^*), x - x^* \rangle \geq 0.$$

(Solution set: X^*).

Under some conditions, variational equilibria can be solved in stochastic generalized Nash games with **expected value constraints** and reduced to the above setting with $X = K \times \mathbb{R}^m_+$ (unbounded).

Expected valued formulation of SVI

Definition (SVI)

Assuming $T : \mathbb{R}^n \to \mathbb{R}^n$ is given by $T = \mathbb{E}[F(\xi, \cdot)]$, solve

$$0 \in T(x) + \mathcal{N}_X(x).$$

- Dates back to King and Rockafellar (1993) and Gürkan, Özge and Robinson (1999),

Expected valued formulation of SVI

Definition (SVI)

Assuming $T : \mathbb{R}^n \to \mathbb{R}^n$ is given by $T = \mathbb{E}[F(\xi, \cdot)]$, solve

$$0 \in T(x) + \mathcal{N}_X(x).$$

- Dates back to King and Rockafellar (1993) and Gürkan, Özge and Robinson (1999),
- **Unavailability** of T: only samples of $F(\xi, \cdot)$ available,

Expected valued formulation of SVI

Definition (SVI)

Assuming $T : \mathbb{R}^n \to \mathbb{R}^n$ is given by $T = \mathbb{E}[F(\xi, \cdot)]$, solve

$$0 \in T(x) + \mathcal{N}_X(x).$$

- Dates back to King and Rockafellar (1993) and Gürkan, Özge and Robinson (1999),
- **Unavailability** of T: only samples of $F(\xi, \cdot)$ available,
- Stochastic Optimization and SVIs need to be studied in terms *perturbation theory* of the mean objective function or operator with respect to the probability measure,

Expected valued formulation of SVI

Definition (SVI)

Assuming $T : \mathbb{R}^n \to \mathbb{R}^n$ is given by $T = \mathbb{E}[F(\xi, \cdot)]$, solve

$$0 \in T(x) + \mathcal{N}_X(x).$$

- Dates back to King and Rockafellar (1993) and Gürkan, Özge and Robinson (1999),
- **Unavailability** of T: only samples of $F(\xi, \cdot)$ available,
- Stochastic Optimization and SVIs need to be studied in terms *perturbation theory* of the mean objective function or operator with respect to the probability measure,
- **Two** complexity metrics: *optimization error* and *sample size*.

Sample Average Approximation (SAA)

Sample Average Approximation (SAA)

Framework:

- **Unavailability** of T.

Sample Average Approximation (SAA)

Framework:

- **Unavailability** of T.
- **SAA problem**: *choose a sample* $\{\xi_j\}_{j=1}^N$ and solve

$$0 \in \widehat{F}_N(x) + \mathcal{N}_X(x),$$

where

$$\widehat{F}_N(\cdot) := \frac{1}{N}\sum_{j=1}^{N} F(\xi_j, \cdot)$$

is the *empirical sample mean* operator.

Sample Average Approximation (SAA)

Framework:

- **Unavailability** of T.
- **SAA problem**: *choose a sample* $\{\xi_j\}_{j=1}^N$ and solve

$$0 \in \widehat{F}_N(x) + \mathcal{N}_X(x),$$

where

$$\widehat{F}_N(\cdot) := \frac{1}{N}\sum_{j=1}^{N} F(\xi_j, \cdot)$$

is the *empirical sample mean* operator.

- Choose a deterministic algorithm to solve the SAA problem.

Stochastic Approximation (SA)

Stochastic Approximation (SA)

Framework:

- **Unavailability** of T.

Stochastic Approximation (SA)

Framework:

- **Unavailability** of T.
- **Stochastic oracle**: given $x^k \in \mathbb{R}^n$ and sample ξ^k of ξ, $F(\xi^k, x^k)$ is available.

Stochastic Approximation (SA)

Framework:

- **Unavailability** of T.
- **Stochastic oracle**: given $x^k \in \mathbb{R}^n$ and sample ξ^k of ξ, $F(\xi^k, x^k)$ is available.
- **Stochastic approximation** (SA): construct an *explicit method* using online samples with $T(x^k)$ replaced by $F(\xi^k, x^k)$.

Stochastic Approximation (SA)

Framework:

- **Unavailability** of T.
- **Stochastic oracle**: given $x^k \in \mathbb{R}^n$ and sample ξ^k of ξ, $F(\xi^k, x^k)$ is available.
- **Stochastic approximation** (SA): construct an *explicit method* using online samples with $T(x^k)$ replaced by $F(\xi^k, x^k)$.
- **Oracle error**:

$$\epsilon(\xi^k, x^k) := F(\xi^k, x^k) - T(x^k).$$

Stochastic Approximation (SA)

Framework:

- **Unavailability** of T.
- **Stochastic oracle**: given $x^k \in \mathbb{R}^n$ and sample ξ^k of ξ, $F(\xi^k, x^k)$ is available.
- **Stochastic approximation** (SA): construct an *explicit method* using online samples with $T(x^k)$ replaced by $F(\xi^k, x^k)$.
- **Oracle error**:

$$\epsilon(\xi^k, x^k) := F(\xi^k, x^k) - T(x^k).$$

- **Monotonicity**: explore this property in the *stochastic setting* pushing forward significantly previous known convergence properties.

Examples: Stochastic Nash Equilibria and Simulation Optimization

Stochastic Nash Equilibria: find $x^* \in \Pi_{i=1}^m X^i$ s.t.

$$\forall i = 1, \ldots, m, \quad x_i^* \in \text{argmin}_{x_i \in X^i} \mathbb{E}\left[f_i(\xi, x_i, x_{-i}^*)\right].$$

Examples: Stochastic Nash Equilibria and Simulation Optimization

Stochastic Nash Equilibria: find $x^* \in \Pi_{i=1}^m X^i$ s.t.

$$\forall i = 1, \ldots, m, \quad x_i^* \in \text{argmin}_{x_i \in X^i} \mathbb{E}\left[f_i(\xi, x_i, x_{-i}^*)\right].$$

Assuming *agent-wise* smooth convex pay-offs, equivalent to a SVI with

- $X := X^1 \times \ldots \times X^m$,
- $F := (\nabla_{x_1} f_1, \ldots, \nabla_{x_m} f_m)$.

Examples: Stochastic Nash Equilibria and Simulation Optimization

Stochastic Nash Equilibria: find $x^* \in \Pi_{i=1}^m X^i$ s.t.

$$\forall i = 1, \ldots, m, \quad x_i^* \in \mathrm{argmin}_{x_i \in X^i} \mathbb{E}\left[f_i(\xi, x_i, x_{-i}^*)\right].$$

Assuming *agent-wise* smooth convex pay-offs, equivalent to a SVI with

- $X := X^1 \times \ldots \times X^m$,
- $F := (\nabla_{x_1} f_1, \ldots, \nabla_{x_m} f_m)$.

Simulation optimization: *Handbook of Simulation Optimization* (2014).

Examples: Statistical Machine Learning

(Linear) Empirical risk minimization (ERM): given *sample data* $\{(x_j, y_j)\}_{j=1}^N$ and *loss* $\ell(\cdot)$ the proposed estimator is

$$\beta_N \in \text{argmin}_{\beta\in\Theta}\frac{1}{N}\sum_{j=1}^{N}\ell\left(Y_j - \beta^T X_j\right).$$

Examples: Statistical Machine Learning

(Linear) Empirical risk minimization (ERM): given *sample data* $\{(x_j, y_j)\}_{j=1}^N$ and *loss* $\ell(\cdot)$ the proposed estimator is

$$\beta_N \in \text{argmin}_{\beta \in \Theta} \frac{1}{N} \sum_{j=1}^{N} \ell\left(Y_j - \beta^T X_j\right).$$

Or the SAA problem associated to $F(X, Y; \beta) := \ell\left(Y - \beta^T X\right)$ where

$$N \gg 1.$$

Examples: Statistical Machine Learning

(Linear) Empirical risk minimization (ERM): given *sample data* $\{(x_j, y_j)\}_{j=1}^N$ and *loss* $\ell(\cdot)$ the proposed estimator is

$$\beta_N \in \text{argmin}_{\beta\in\Theta}\frac{1}{N}\sum_{j=1}^{N}\ell\left(Y_j - \beta^T X_j\right).$$

Or the SAA problem associated to $F(X,Y;\beta) := \ell\left(Y - \beta^T X\right)$ where

$$N \gg 1.$$

Stochastic gradient descent methods (SGD) are SA-type methods for the ERM problem (in a discrete distribution) also known as random *incremental methods*.

Examples: Statistical Machine Learning

(Linear) Empirical risk minimization (ERM): given *sample data* $\{(x_j, y_j)\}_{j=1}^N$ and *loss* $\ell(\cdot)$ the proposed estimator is

$$\beta_N \in \text{argmin}_{\beta \in \Theta} \frac{1}{N} \sum_{j=1}^{N} \ell\left(Y_j - \beta^T X_j\right).$$

Or the SAA problem associated to $F(X, Y; \beta) := \ell\left(Y - \beta^T X\right)$ where

$$N \gg 1.$$

Stochastic gradient descent methods (SGD) are SA-type methods for the ERM problem (in a discrete distribution) also known as random *incremental methods*. Other incremental variations with respect to other discrete *high-dimension* parameters (coordinates, constraints, number of agents, etc).

Applications

Applications

Monotonicity in:

- mechanical problems,

Applications

Monotonicity in:

- mechanical problems,
- partial economic equilibrium,

Applications

Monotonicity in:

- mechanical problems,
- partial economic equilibrium,
- transportation: Wardrop equilibrium,

Applications

Monotonicity in:

- mechanical problems,
- partial economic equilibrium,
- transportation: Wardrop equilibrium,
- wireline and wireless communication networks (Koshal, Nedić and Shanbhag),

Applications

Monotonicity in:

- mechanical problems,
- partial economic equilibrium,
- transportation: Wardrop equilibrium,
- wireline and wireless communication networks (Koshal, Nedić and Shanbhag),
- bandwith allocation and design of cognitive radio systems (Scutari, Facchinei and Pang),

Applications

Monotonicity in:

- mechanical problems,
- partial economic equilibrium,
- transportation: Wardrop equilibrium,
- wireline and wireless communication networks (Koshal, Nedić and Shanbhag),
- bandwith allocation and design of cognitive radio systems (Scutari, Facchinei and Pang),
- optical networks (Pan and Pavel),

Applications

Monotonicity in:

- mechanical problems,
- partial economic equilibrium,
- transportation: Wardrop equilibrium,
- wireline and wireless communication networks (Koshal, Nedić and Shanbhag),
- bandwith allocation and design of cognitive radio systems (Scutari, Facchinei and Pang),
- optical networks (Pan and Pavel),
- zero-sum Nash games, potential games

Applications

Monotonicity in:

- mechanical problems,
- partial economic equilibrium,
- transportation: Wardrop equilibrium,
- wireline and wireless communication networks (Koshal, Nedić and Shanbhag),
- bandwith allocation and design of cognitive radio systems (Scutari, Facchinei and Pang),
- optical networks (Pan and Pavel),
- zero-sum Nash games, potential games
- Some structured Nash-equilibria (e.g. a class of Nash-Cournot problems with nonlinear prices: Kannan and Shanbhag).

Applications

Monotonicity in:

- mechanical problems,
- partial economic equilibrium,
- transportation: Wardrop equilibrium,
- wireline and wireless communication networks (Koshal, Nedić and Shanbhag),
- bandwith allocation and design of cognitive radio systems (Scutari, Facchinei and Pang),
- optical networks (Pan and Pavel),
- zero-sum Nash games, potential games
- Some structured Nash-equilibria (e.g. a class of Nash-Cournot problems with nonlinear prices: Kannan and Shanbhag).

NOTE: *Local monotonicity* satisfied by a large class of non-monotone equilibrium problems.

A variance-based stochastic extragradient method

Notation: Given sample $\xi^N := \{\xi_1, \ldots, \xi_N\}$, $\widehat{F}(\xi^N, x) = \frac{1}{N}\sum_{j=1}^{N} F(\xi_j, x)$.

Algorithm (A variance-based stochastic extragradient method)

$$\begin{aligned} z^k &= \Pi_X\left[x^k - \alpha_k \widehat{F}(\xi^k, x^k)\right], \\ x^{k+1} &= \Pi_X\left[x^k - \alpha_k \widehat{F}(\eta^k, z^k)\right], \end{aligned}$$

where $\xi^k := \{\xi_j^k : j = 1, \ldots, N_k\}$ *and* $\eta^k := \{\eta_j^k : j = 1, \ldots, N_k\}$.

A variance-based stochastic extragradient method

Assumptions

- *T is* **pseudo-monotone***:*
  $\langle T(y), x - y\rangle \geq 0 \Longrightarrow \langle T(x), x - y\rangle \geq 0, \forall x, y \in \mathbb{R}^n.$

A variance-based stochastic extragradient method

Assumptions

- *T is* **pseudo-monotone***:*
  $\langle T(y), x-y\rangle \geq 0 \Longrightarrow \langle T(x), x-y\rangle \geq 0, \forall x,y \in \mathbb{R}^n.$
- **Lipschitz-continuity***: There exists measurable $L:\Xi \to \mathbb{R}_+$ with finite p-moment, $p \geq 2$,*
  $\|F(\xi,x)-F(\xi,y)\| \leq L(\xi)\|x-y\|, \forall x,y \in \mathbb{R}^n.$

A variance-based stochastic extragradient method

Assumptions

- *T is* **pseudo-monotone***:* $\langle T(y), x-y\rangle \geq 0 \Longrightarrow \langle T(x), x-y\rangle \geq 0, \forall x, y \in \mathbb{R}^n$.
- **Lipschitz-continuity***: There exists measurable $L : \Xi \to \mathbb{R}_+$ with finite p-moment, $p \geq 2$,* $\|F(\xi, x) - F(\xi, y)\| \leq L(\xi)\|x-y\|, \forall x, y \in \mathbb{R}^n$.
- **Sampling rate***:* $\sum \frac{1}{N_k} < \infty$.

A variance-based stochastic extragradient method

Assumptions

- *T is* **pseudo-monotone***:*
  $\langle T(y), x-y\rangle \geq 0 \Longrightarrow \langle T(x), x-y\rangle \geq 0, \forall x, y \in \mathbb{R}^n$.
- **Lipschitz-continuity***: There exists measurable* $L:\Xi \to \mathbb{R}_+$ *with finite p-moment,* $p \geq 2$*,*
  $\|F(\xi,x)-F(\xi,y)\| \leq L(\xi)\|x-y\|, \forall x, y \in \mathbb{R}^n$.
- **Sampling rate***:* $\sum \frac{1}{N_k} < \infty$.
- **Stepsize bounded away from zero***:* $0 < \inf_k \alpha_k \leq \sup_k \alpha_k < \frac{1}{\sqrt{6}L}$.

A variance-based stochastic extragradient method

Assumptions

- *T is* **pseudo-monotone***:*
  $\langle T(y), x-y\rangle \geq 0 \Longrightarrow \langle T(x), x-y\rangle \geq 0, \forall x, y \in \mathbb{R}^n$.
- **Lipschitz-continuity***: There exists measurable $L:\Xi\to\mathbb{R}_+$ with finite p-moment, $p\geq 2$,*
  $\|F(\xi,x)-F(\xi,y)\| \leq L(\xi)\|x-y\|, \forall x, y \in \mathbb{R}^n$.
- **Sampling rate***:* $\sum \frac{1}{N_k} < \infty$.
- **Stepsize bounded away from zero***:* $0 < \inf_k \alpha_k \leq \sup_k \alpha_k < \frac{1}{\sqrt{6}L}$.

Previous assumptions **not** required:

- bounded T or X
- regularization
- Uniformly bounded variance of oracle:
  $\sup_{x\in X}\mathbb{E}\left[\|F(\xi,x)-T(x)\|^2\right] \leq \sigma^2$.

Stochastic extragradient method with linear search

OBJECTIVE: Absence of **knowledge of** L.

Stochastic extragradient method with linear search

OBJECTIVE: Absence of **knowledge of** L.
Notation: Given sample $\xi^N := \{\xi_1, \ldots, \xi_N\}$, $\widehat{F}(\xi^N, x) = \frac{1}{N}\sum_{j=1}^N F(\xi_j, x)$.

Algorithm (A variance-based stochastic extragradient method with linear search)

Choose any $\hat{\alpha} > 0$. *If* $x^k = \Pi\left[x^k - \hat{\alpha}\widehat{F}(\xi^k, x^k)\right]$ *stop. Otherwise:*

Linear search rule: *define* α_k *as the maximum* $\alpha \in \{\Theta^j \hat{\alpha} : j \in \mathbb{N}_0\}$ *such that*

$$\alpha \left\| \widehat{F}\left(\xi^k, z^k(\alpha)\right) - \widehat{F}\left(\xi^k, x^k\right) \right\| \leq \lambda \|z^k(\alpha) - x^k\|,$$

where $z^k(\alpha) := \Pi_X\left[x^k - \alpha\widehat{F}(\xi^k, x^k)\right]$ *for all* $\alpha > 0$. *Set*

$$\begin{aligned} z^k &= \Pi_X\left[x^k - \alpha_k \widehat{F}(\xi^k, x^k)\right], \\ x^{k+1} &= \Pi_X\left[x^k - \alpha_k \widehat{F}(\eta^k, z^k)\right]. \end{aligned}$$

Results

Theorem (Asymptotic convergence)

For both methods, *a.s. the sequence* $\{x^k\}$ *is bounded,*

$$\lim_{k\to\infty} \mathrm{d}(x^k, X^*) = 0,$$

and

$$r_{\alpha_k}(x^k) \xrightarrow[k\to\infty]{a.s., L^2} 0.$$

Natural residual:

$$r_\alpha(x) := \|x - \Pi_X[x - \alpha T(x)]\|.$$

Proposition (Uniform boundedness of p-moment)

For both methods, *given* $x^* \in X^*$, *there exists* $\mathsf{c}_p(x^*) \geq 1$ *and* $k_0 := k_0(x^*) \in \mathbb{N}$ *s.t.*

$$\sup_{k \geq k_0} \left| \|x^k - x^*\| \right|_p^2 \leq \mathsf{c}_p(x^*) \left[1 + \left| \|x^{k_0} - x^*\| \right|_p^2 \right].$$

NOTE: $\mathsf{c}_p(x^*)$ and $k_0(x^*)$ are explicitly estimated.
NOTE: boundedness not assumed a priori!

Results

Theorem (Convergence rate and oracle complexity: known L)

Take $\alpha_k \equiv \alpha \in (0, 1/\sqrt{6}L)$ *and* N_k *as*

$$N_k = \left\lceil \theta(k+\mu)(\ln(k+\mu))^{1+b} \right\rceil.$$

Then **a.s.-convergence** *holds and for all* $x^* \in X^*$, *there are non-negative constants* $\overline{\mathsf{Q}}(x^*)$, $\mathsf{P}(x^*)$ *and* $\mathsf{I}(x^*)$ *such that for all* $\epsilon > 0$, *there exists* $K := K_\epsilon \in \mathbb{N}$ *such that*

$$\mathbb{E}[r_\alpha(x^K)^2] \leq \epsilon \leq \frac{\max\{1, \theta^{-2}\}\overline{\mathsf{Q}}(x^*)}{K},$$

$$\sum_{k=1}^{K} 2N_k \leq \frac{\max\{1, \theta^{-4}\}\max\{1, \theta\}\mathsf{I}(x^*)\left\{\left[\ln\left(\mathsf{P}(x^*)\epsilon^{-1}\right)\right]^{1+b} + \frac{1}{\mu}\right\}}{\epsilon^2}.$$

Results

Theorem (Convergence rate and oracle complexity: unknown L)

Take any $\hat{\alpha} > 0$, N_k *as*

$$N_k = \left\lceil \theta(k+\mu)(\ln(k+\mu))^{1+b} \right\rceil.$$

Then **a.s.-convergence** *holds and for all* $x^* \in X^*$ *and all* $\epsilon > 0$*, there exists* $K := K_\epsilon \in \mathbb{N}$ *such that*

$$\mathbb{E}[r_{\hat{\alpha}}(x^K)^2] \leq \epsilon \lesssim_{x^*} \frac{\max\{1, \theta^{-2}\}}{K},$$

$$\sum_{k=1}^{K} j_k \cdot 2N_k \lesssim_{x^*} \ln_{\frac{1}{\Theta}} \left(\widetilde{L}_k\right) \frac{\max\{1, \theta^{-4}\} \max\{1, \theta\} \left\{ \left[\ln\left(\epsilon^{-1}\right)\right]^{1+b} + \frac{1}{\mu} \right\}}{\epsilon^2},$$

where $\widetilde{L}_k = \frac{1}{N_k} \sum_{j=1}^{N_k} L(\xi_j^k)$ *and* j_k *is the number of oracle calls at the* k*-th linear search.*

Contributions: stochastic extragradient methods

Contributions: stochastic extragradient methods

- **Unbounded** T or X,

Before:

- Boundedness

Contributions: stochastic extragradient methods

- **Unbounded** T or X,
- **Non-uniform variance**: $\forall x \in X, \mathbb{E}\left[\|F(\xi,x) - T(x)\|^2\right] < \infty$,

Before:

- Boundedness
- Uniform variance: $\sup_{x\in X} \mathbb{E}\left[\|F(\xi,x) - T(x)\|^2\right] \leq \sigma^2$,

Contributions: stochastic extragradient methods

- **Unbounded** T or X,
- **Non-uniform variance**: $\forall x \in X, \mathbb{E}\left[\|F(\xi,x) - T(x)\|^2\right] < \infty$,
- **No regularization**,

Before:

- Boundedness
- Uniform variance: $\sup_{x \in X} \mathbb{E}\left[\|F(\xi,x) - T(x)\|^2\right] \leq \sigma^2$,
- Regularization
  - additional coordination
  - sub-optimality

Contributions: stochastic extragradient methods

- **Unbounded** T or X,
- **Non-uniform variance**: $\forall x \in X, \mathbb{E}\left[\|F(\xi,x) - T(x)\|^2\right] < \infty$,
- **No regularization**,
- **Faster rate** $O(1/K)$ & **constant stepsize**,

Before:

- Boundedness
- Uniform variance: $\sup_{x \in X} \mathbb{E}\left[\|F(\xi,x) - T(x)\|^2\right] \leq \sigma^2$,
- Regularization
  - additional coordination
  - sub-optimality
- Rate $O(1/\sqrt{K})$ & small stepsizes,

Contributions: stochastic extragradient methods

- **Unbounded** T or X,
- **Non-uniform variance**: $\forall x \in X, \mathbb{E}\left[\|F(\xi, x) - T(x)\|^2\right] < \infty$,
- **No regularization**,
- **Faster rate** $O(1/K)$ & **constant stepsize**,
- **Oracle complexity** $O(\epsilon^{-2}) \cdot \left[\ln(\epsilon^{-1})\right]^{1+b}$

Before:

- Boundedness
- Uniform variance: $\sup_{x \in X} \mathbb{E}\left[\|F(\xi, x) - T(x)\|^2\right] \leq \sigma^2$,
- Regularization
  - additional coordination
  - sub-optimality
- Rate $O(1/\sqrt{K})$ & small stepsizes,
- Oracle complexity $O(\epsilon^{-2})$

Contributions: stochastic extragradient methods

- **Unbounded** T or X,
- **Non-uniform variance**: $\forall x \in X, \mathbb{E}\left[\|F(\xi,x) - T(x)\|^2\right] < \infty$,
- **No regularization**,
- **Faster rate** $O(1/K)$ & **constant stepsize**,
- **Oracle complexity** $O(\epsilon^{-2}) \cdot \left[\ln(\epsilon^{-1})\right]^{1+b}$ (efficient)

Before:

- Boundedness
- Uniform variance: $\sup_{x \in X} \mathbb{E}\left[\|F(\xi,x) - T(x)\|^2\right] \leq \sigma^2$,
- Regularization
  - additional coordination
  - sub-optimality
- Rate $O(1/\sqrt{K})$ & small stepsizes,
- Oracle complexity $O(\epsilon^{-2})$

Contributions: stochastic extragradient methods

- **Unbounded** T or X,
- **Non-uniform variance**: $\forall x \in X, \mathbb{E}\left[\|F(\xi,x) - T(x)\|^2\right] < \infty$,
- **No regularization**,
- **Faster rate** $O(1/K)$ & **constant stepsize**,
- **Oracle complexity** $O(\epsilon^{-2}) \cdot \left[\ln(\epsilon^{-1})\right]^{1+b}$ (efficient)
- **Uniform boundedness of p-moments.**

Before:

- Boundedness
- Uniform variance: $\sup_{x \in X} \mathbb{E}\left[\|F(\xi,x) - T(x)\|^2\right] \leq \sigma^2$,
- Regularization
  - additional coordination
  - sub-optimality
- Rate $O(1/\sqrt{K})$ & small stepsizes,
- Oracle complexity $O(\epsilon^{-2})$

Contributions: stochastic extragradient methods

Contributions: stochastic extragradient methods

- **Absence of** L:
  - First stochastic approximation method with linear search for SVI,
  - Essentially same performance as with knowledge of L up to a factor of

    $$O\left(\ln_{\frac{1}{\Theta}} L\right)$$

    in the number of projections per iteration and oracle complexity.

Contributions: stochastic extragradient methods

- **Absence of L:**
  - First stochastic approximation method with linear search for SVI,
  - Essentially same performance as with knowledge of L up to a factor of

    $$O\left(\ln_{\frac{1}{\Theta}} L\right)$$

    in the number of projections per iteration and oracle complexity.
- To the best of our knowledge, we obtain the best error bound known for SA of monotone SVIs improving works of Juditsky-Nemirovski-Tauvel, Lan et al. and others.

Contributions: stochastic extragradient methods

- **Absence of L**:
  - First stochastic approximation method with linear search for SVI,
  - Essentially same performance as with knowledge of L up to a factor of

$$O\left(\ln_{\frac{1}{\Theta}} L\right)$$

  in the number of projections per iteration and oracle complexity.
- To the best of our knowledge, we obtain the best error bound known for SA of monotone SVIs improving works of Juditsky-Nemirovski-Tauvel, Lan et al. and others.
- Previous results not shared by the SAA method or classical SA method:
  - Possibility of using **local** and **distributed** empirical averages,
  - Avoid *uniform* Central Limit Theorem,
  - **Robust** sampling.

Contributions: stochastic extragradient methods

- **Absence of L**:
  - First stochastic approximation method with linear search for SVI,
  - Essentially same performance as with knowledge of L up to a factor of

$$O\left(\ln_{\frac{1}{\Theta}} L\right)$$

    in the number of projections per iteration and oracle complexity.
- To the best of our knowledge, we obtain the best error bound known for SA of monotone SVIs improving works of Juditsky-Nemirovski-Tauvel, Lan et al. and others.
- Previous results not shared by the SAA method or classical SA method:
  - Possibility of using **local** and **distributed** empirical averages,
  - Avoid *uniform* Central Limit Theorem,
  - **Robust** sampling.
- Empirical evidence in Nemirovski-Juditsky-Lan-Shapiro (2009) that SA can outperform SAA in a large class of convex-structured problems.

Final comments

Final comments

Non-uniform variance of oracle:

- Affine variational inequalities, LCP or systems of equations:

$$\sigma(x)^2 := \mathbb{E}\left[\|F(\xi,x) - T(x)\|^2\right] \sim \|x\|^2,$$

Final comments

Non-uniform variance of oracle:

- Affine variational inequalities, LCP or systems of equations:

$$\sigma(x)^2 := \mathbb{E}\left[\|F(\xi,x) - T(x)\|^2\right] \sim \|x\|^2,$$

- The performance of our methods depend on the variance at points of the trajectory and in X^* (but not on the whole X):

$$\widehat{Q} := \inf_{x^* \in X^*} \left\{ \|x^0 - x^*\|^2 + \sigma(x^*)^4 \cdot \max_{0 \leq k \leq k_0(x^*)} \mathbb{E}\left[\|x^k - x^*\|^2\right] \right\}$$

Final comments

Non-uniform variance of oracle:

- Affine variational inequalities, LCP or systems of equations:

$$\sigma(x)^2 := \mathbb{E}\left[\|F(\xi,x) - T(x)\|^2\right] \sim \|x\|^2,$$

- The performance of our methods depend on the variance at points of the trajectory and in X^* (but not on the whole X):

$$\widehat{Q} := \inf_{x^* \in X^*} \left\{ \|x^0 - x^*\|^2 + \sigma(x^*)^4 \cdot \max_{0 \le k \le k_0(x^*)} \mathbb{E}\left[\|x^k - x^*\|^2\right] \right\}$$

- Better performance also for the case of **compact** X or **uniform variance** if

$$\sigma(x^*)^2 \ll \sigma^2,$$

(e.g. **affine variational inequalities** over compact X).

Final comments: some new proof techniques

Final comments: some new proof techniques

Known L:

- Martingale methods,

Final comments: some new proof techniques

Known L:

- Martingale methods,
- Oracle error control: Burkholder-Davis-Gundy inequality in Hilbert spaces.

Final comments: some new proof techniques

Known L:

- Martingale methods,
- Oracle error control: Burkholder-Davis-Gundy inequality in Hilbert spaces.

Unknown L:

- Endogenous random stepsize: must control an oracle error which is **not a martingale difference**,

Final comments: some new proof techniques

Known L:

- Martingale methods,
- Oracle error control: Burkholder-Davis-Gundy inequality in Hilbert spaces.

Unknown L:

- Endogenous random stepsize: must control an oracle error which is **not a martingale difference**,
- Nevertheless, we can condition on a ball centred at the previous k-th iterate and control the **empirical process**

$$\sup_{x \in B(x^k)} \frac{1}{N_k} \left\| \sum_{j=1}^{N_k} \epsilon(\xi_j^k, x) - \epsilon(\xi_j^k, x^*) \right\|,$$

where $\epsilon(\xi, x) := F(\xi, x) - T(x)$,

Final comments: some new proof techniques

Known L:

- Martingale methods,
- Oracle error control: Burkholder-Davis-Gundy inequality in Hilbert spaces.

Unknown L:

- Endogenous random stepsize: must control an oracle error which is **not a martingale difference**,
- Nevertheless, we can condition on a ball centred at the previous k-th iterate and control the **empirical process**

$$\sup_{x \in B(x^k)} \frac{1}{N_k} \left\| \sum_{j=1}^{N_k} \epsilon(\xi_j^k, x) - \epsilon(\xi_j^k, x^*) \right\|,$$

  where $\epsilon(\xi, x) := F(\xi, x) - T(x)$,
- Use **locally** moment and concentration inequalities of empirical process theory.

References

[1] A. Iusem, A. Jofré, R.I. Oliveira and P. Thompson, *Extragradient method with variance reduction for stochastic variational inequalities*. To appear in *SIAM Journal on Optimization*.

[2] A. Iusem, A. Jofré, R.I. Oliveira and P. Thompson, *Variance-based extragradient method with linear search for stochastic variational inequalities*. Submitted.

[3] A. Iusem, A. Jofré and P. Thompson, *Incremental constraint projection methods for monotone stochastic variational inequalities*. Under second round revision.

THANK YOU VERY MUCH!