Incremental constraint projection methods for stochastic variational inequalities

Philip Thompson

Workshop on *Analysis and Applications of Stochastic Systems*

Center for Mathematical Modeling (CMM), Chile

28th, March, 2016, Rio de Janeiro

Contents

Contents

Contents

Contents

Stochastic variational inequality: generalization of stochastic optimization

Stochastic variational inequality: generalization of stochastic optimization

Assume

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

Stochastic variational inequality: generalization of stochastic optimization

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 $(\Omega, \mathcal{F}, \mathbb{P})$,

Stochastic variational inequality: generalization of stochastic optimization

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 $(\Omega, \mathcal{F}, \mathbb{P})$,
- **Mean operator**: $T : \mathbb{R}^n \to \mathbb{R}^n$ satisfies

$$T(x) = \mathbb{E}[F(\xi, x)], \quad \forall x \in \mathbb{R}^n.$$

Stochastic variational inequality: generalization of stochastic optimization

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 $(\Omega, \mathcal{F}, \mathbb{P})$,
- **Mean operator**: $T : \mathbb{R}^n \to \mathbb{R}^n$ satisfies

$$T(x) = \mathbb{E}[F(\xi, x)], \quad \forall x \in \mathbb{R}^n.$$

- **Feasible set**: $X \subset \mathbb{R}^n$ closed and convex.

Stochastic variational inequality: generalization of stochastic optimization

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 $(\Omega, \mathcal{F}, \mathbb{P})$,
- **Mean operator**: $T : \mathbb{R}^n \to \mathbb{R}^n$ satisfies

$$T(x) = \mathbb{E}[F(\xi, x)], \quad \forall x \in \mathbb{R}^n.$$

- **Feasible set**: $X \subset \mathbb{R}^n$ closed and convex.

Definition (SVI)

Supposing $T := \mathbb{E}[F(\xi, \cdot)]$, almost surely find $x^* \in X$ s.t. $\forall x \in X$,

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

Stochastic Equilibrium

Stochastic Equilibrium

Multi-agent stochastic optimization:

$$\begin{array}{ll} \min & \sum_{i=1}^{m} \mathbb{E}[f_i(\xi, x)] \\ \text{s.t.} & x \in X^1 \times \ldots \times X^m. \end{array}$$

SVI with

- $X := X^1 \times \ldots \times X^m$,
- $F := (\nabla f_1, \ldots, \nabla f_m)$.

Stochastic Equilibrium

Multi-agent stochastic optimization:

$$\begin{aligned} \min \quad & \sum_{i=1}^{m} \mathbb{E}[f_i(\xi, x)] \\ \text{s.t.} \quad & x \in X^1 \times \ldots \times X^m. \end{aligned}$$

SVI with

- $X := X^1 \times \ldots \times X^m$,
- $F := (\nabla f_1, \ldots, \nabla f_m)$.

Stochastic Nash-Equilibria: find $x^* \in \prod_{i=1}^m X^i$ s.t. for all $i \in [m]$,

$$x_i^* \in \text{argmin}_{x_i \in X_i} \mathbb{E}[f_i(\xi, x_i, x_{-i}^*)].$$

SVI with

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

Stochastic Approximation (SA) methods for SVI

Stochastic Approximation (SA) methods for SVI

Framework:

- **Unavailability** of T.

Stochastic Approximation (SA) methods for SVI

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) methods for SVI

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): use a deterministic method with $F(\xi^k, x^k)$ instead of $T(x^k)$.

Stochastic Approximation (SA) methods for SVI

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): use a deterministic method with $F(\xi^k, x^k)$ instead of $T(x^k)$.
- **Oracle error**:

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

Stochastic Approximation (SA) methods for SVI

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): use a deterministic method with $F(\xi^k, x^k)$ instead of $T(x^k)$.
- **Oracle error**:

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

- **Monotonicity**-type.

Stochastic Approximation (SA) methods for SVI

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): use a deterministic method with $F(\xi^k, x^k)$ instead of $T(x^k)$.
- **Oracle error**:

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

- **Monotonicity**-type.

SA related to Stochastic Gradient methods: **stochastic optimization** & **statistical inference** (Robbins-Monro 1950's), **online optimization**, **optimization for machine learning**.

Random incremental constraint projection

Framework:

- **Difficult access** to $X = \cap_{i \in \mathcal{I}} X_i$:
  - Difficult projections,
  - Large number of constraints,
  - Online learning of constraints.

Random incremental constraint projection

Framework:

- **Difficult access** to $X = \cap_{i \in \mathcal{I}} X_i$:
  - Difficult projections,
  - Large number of constraints,
  - Online learning of constraints.
- Alternative: **random sampling** of constraint component X_{ω_k},

Random incremental constraint projection

Framework:

- **Difficult access** to $X = \cap_{i \in \mathcal{I}} X_i$:
  - Difficult projections,
  - Large number of constraints,
  - Online learning of constraints.
- Alternative: **random sampling** of constraint component X_{ω_k},
- **Feasibility error:**

$$\mathrm{d}(x^k) := \mathrm{d}(x^k, X).$$

Random incremental constraint projection

Framework:

- **Difficult access** to $X = \cap_{i \in \mathcal{I}} X_i$:
  - Difficult projections,
  - Large number of constraints,
  - Online learning of constraints.
- Alternative: **random sampling** of constraint component X_{ω_k},
- **Feasibility error:**

$$\mathrm{d}(x^k) := \mathrm{d}(x^k, X).$$

- **Regularity of set**: Slater-type conditions. Mild condition in practice.

Random incremental constraint projection

Framework:

- **Difficult access** to $X = \cap_{i \in \mathcal{I}} X_i$:
  - Difficult projections,
  - Large number of constraints,
  - Online learning of constraints.
- Alternative: **random sampling** of constraint component X_{ω_k},
- **Feasibility error:**

$$\mathrm{d}(x^k) := \mathrm{d}(x^k, X).$$

- **Regularity of set**: Slater-type conditions. Mild condition in practice.

Applications: large-data set problems, online optimization and equilibrium, distributed learning (eg, distributed regression).

An incremental constraint one-projection method

An incremental constraint one-projection method

Set-up:

$$X = X_0 \cap \left(\cap_{i \in \mathcal{I}} X_i\right).$$

An incremental constraint one-projection method

Set-up:

$$X = X_0 \cap \left(\cap_{i \in \mathcal{I}} X_i \right).$$

Assume:

- projection onto X_0 is computationally easy (HARD CONSTRAINT),

An incremental constraint one-projection method

Set-up:

$$X = X_0 \cap \left(\cap_{i \in \mathcal{I}} X_i\right).$$

Assume:

- projection onto X_0 is computationally easy (HARD CONSTRAINT),
- $\forall i \in \mathcal{I}$, $X_i = \{x \in \mathbb{R}^n : g_i(x) \leq 0\}$, (SOFT CONSTRAINTS)
  - subgradients of $g_i^+(x)$ are easily computable,

An incremental constraint one-projection method

Set-up:

$$X = X_0 \cap \left(\cap_{i\in\mathcal{I}} X_i\right).$$

Assume:

- projection onto X_0 is computationally easy (HARD CONSTRAINT),
- $\forall i \in \mathcal{I}$, $X_i = \{x \in \mathbb{R}^n : g_i(x) \leq 0\}$, (SOFT CONSTRAINTS)
  - subgradients of $g_i^+(x)$ are easily computable,
  - $\{\partial g_i^+ : i \in \mathcal{I}\}$ is uniformly bounded over X_0.

An incremental constraint one-projection method

Set-up:

$$X = X_0 \cap \left(\cap_{i\in\mathcal{I}} X_i\right).$$

Assume:

- projection onto X_0 is computationally easy (HARD CONSTRAINT),
- $\forall i \in \mathcal{I}$, $X_i = \{x \in \mathbb{R}^n : g_i(x) \leq 0\}$, (SOFT CONSTRAINTS)
  - subgradients of $g_i^+(x)$ are easily computable,
  - $\{\partial g_i^+ : i \in \mathcal{I}\}$ is uniformly bounded over X_0.

Typical example: if X_i has easy projections set $g_i := \mathrm{d}(\cdot, X_i)$:

- $\sup_{x\in\mathbb{R}^n} \|\partial g_i(x)\| \leq 1$,
- $\frac{x-\Pi_{X_i}(x)}{\|x-\Pi_{X_i}(x)\|} \in \partial g_i(x)$.

An incremental constraint projection method

Algorithm

$$y^k = \Pi_{X_0}\left[x^k - \alpha_k\left(F(\xi^k, x^k) + \epsilon_k x^k\right)\right],$$

$$x^{k+1} = \Pi_{X_0}\left[y^k - \beta_k \frac{g_{\omega_k}^+(y^k)}{\|d^k\|^2} d^k\right],$$

where $d^k \in \partial g_{\omega_k}^+(y^k) - \{0\}$ *if* $g_{\omega_k}(y^k) > 0$.

An incremental constraint projection method

Typical example: soft constraints with easy **projection**:

Algorithm

$$y^k = x^k - \alpha_k \left(F(\xi^k, x^k) + \epsilon_k x^k \right),$$
$$x^{k+1} = \Pi_{X_0} \left[y^k - \beta_k \left(y^k - \Pi_{\omega_k}(y^k) \right) \right].$$

Case 1: Monotone weak-sharp SVI & $\epsilon^k \equiv 0$

Case 1: Monotone weak-sharp SVI & $\epsilon^k \equiv 0$

Assumptions:

- T is monotone and weak-sharp:

$$\langle T(x^*), x - x^* \rangle \geq \rho\, \mathrm{d}(x, X^*), \forall x \in X, \forall x^* \in X^*.$$

Case 1: Monotone weak-sharp SVI & $\epsilon^k \equiv 0$

Assumptions:

- T is monotone and weak-sharp:

$$\langle T(x^*), x - x^* \rangle \geq \rho\, \mathrm{d}(x, X^*), \forall x \in X, \forall x^* \in X^*.$$

- Bounded operator or Lipschitz operator.

Case 1: Monotone weak-sharp SVI & $\epsilon^k \equiv 0$

Assumptions:

- T is monotone and weak-sharp:

$$\langle T(x^*), x - x^* \rangle \geq \rho\, \mathrm{d}(x, X^*), \forall x \in X, \forall x^* \in X^*.$$

- Bounded operator or Lipschitz operator.
- Unbiased oracle with *finite* variance (non-uniform variance).

Case 1: Monotone weak-sharp SVI & $\epsilon^k \equiv 0$

Assumptions:

- T is monotone and weak-sharp:

$$\langle T(x^*), x - x^* \rangle \geq \rho \, \mathrm{d}(x, X^*), \forall x \in X, \forall x^* \in X^*.$$

- Bounded operator or Lipschitz operator.
- Unbiased oracle with *finite* variance (non-uniform variance).
- **Constraint sampling and regularity**:

$$\mathrm{d}(x, X)^2 \leq c\mathbb{E}\left[\left(g^+_{\omega_k}(x)\right)^2 \middle| \mathcal{F}_k\right], \forall x \in X_0.$$

Case 1: Monotone weak-sharp SVI & $\epsilon^k \equiv 0$

Assumptions:

- T is monotone and weak-sharp:

$$\langle T(x^*), x - x^* \rangle \geq \rho \, \mathrm{d}(x, X^*), \forall x \in X, \forall x^* \in X^*.$$

- Bounded operator or Lipschitz operator.
- Unbiased oracle with *finite* variance (non-uniform variance).
- **Constraint sampling and regularity**:

$$\mathrm{d}(x, X)^2 \leq c\mathbb{E}\left[\left(g_{\omega_k}^+(x)\right)^2 \middle| \mathcal{F}_k\right], \forall x \in X_0.$$

- **Small stepsizes**: $\alpha_k > 0$, $\beta_k \in (0,2)$ **without knowledge of problem parameters** and

$$\sum_{k=0}^{\infty} \alpha_k = \infty, \quad \sum_{k=0}^{\infty} \alpha_k^2 < \infty, \quad \sum_{k=0}^{\infty} \frac{\alpha_k^2}{\beta_k(2-\beta_k)} < \infty.$$

Constraint sampling and regularity

Typical case: $\{X_i : i \in \mathcal{I}\}$ with easy projection and $1 << |I| < \infty$:

Constraint sampling and regularity

Typical case: $\{X_i : i \in \mathcal{I}\}$ with easy projection and $1 << |I| < \infty$:

- **Linear regularity:** for all $x \in X_0$

$$\mathrm{d}(x, X)^2 \leq \eta \max_{i \in \mathcal{I}} \mathrm{d}(x, X_i)^2,$$

Constraint sampling and regularity

Typical case: $\{X_i : i \in \mathcal{I}\}$ with easy projection and $1 << |I| < \infty$:

- **Linear regularity:** for all $x \in X_0$

$$\mathrm{d}(x, X)^2 \leq \eta \max_{i \in \mathcal{I}} \mathrm{d}(x, X_i)^2,$$

- **Uniform independent sampling:** $i \in \mathcal{I}$,

$$\mathbb{P}(\omega_k = i | \mathcal{F}_k) = \frac{1}{|\mathcal{I}|}.$$

Constraint sampling and regularity

Typical case: $\{X_i : i \in \mathcal{I}\}$ with easy projection and $1 << |I| < \infty$:

- **Linear regularity:** for all $x \in X_0$

$$\mathrm{d}(x, X)^2 \leq \eta \max_{i \in \mathcal{I}} \mathrm{d}(x, X_i)^2,$$

- **Uniform independent sampling:** $i \in \mathcal{I}$,

$$\mathbb{P}(\omega_k = i | \mathcal{F}_k) = \frac{1}{|\mathcal{I}|}.$$

Then condition holds with $c = \mathcal{O}(|\mathcal{I}|/\eta)$.

Results

Theorem (Asymptotic convergence)

A.s. the sequence $\{x^k\}$ *is bounded and*

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

Results

Theorem (Asymptotic convergence)

A.s. the sequence $\{x^k\}$ *is bounded and*

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

Proposition (Boundedness in L^2)

The generated sequence $\{x^k\}$ *is bounded in* L^2 *with explicit constant estimates for* $\mathbb{E}[\|x^k - x^*\|^2]$.

Results

Theorem (Rate of convergence: unbounded case)

Given $\theta > 0$ *and* $\lambda > 0$ *take*

$$\alpha_k := \frac{\theta}{\sqrt{k\,(\ln k)^{1+\lambda}}}, \quad \beta_k \equiv \beta \in (0,2).$$

Remark: ROBUST STEPSIZES.

Results

Theorem (Rate of convergence: unbounded case)

Given $\theta > 0$ *and* $\lambda > 0$ *take*

$$\alpha_k := \frac{\theta}{\sqrt{k\,(\ln k)^{1+\lambda}}}, \quad \beta_k \equiv \beta \in (0,2).$$

Then **a.s.-asymptotic convergence** *holds*

Remark: ROBUST STEPSIZES.

Results

Theorem (Rate of convergence: unbounded case)

Given $\theta > 0$ *and* $\lambda > 0$ *take*

$$\alpha_k := \frac{\theta}{\sqrt{k\,(\ln k)^{1+\lambda}}}, \quad \beta_k \equiv \beta \in (0,2).$$

Then **a.s.-asymptotic convergence** *holds and*

$$\mathbb{E}\left[\mathrm{d}(\widehat{x}^k, X^*)\right] \lesssim \mathcal{O}(1)\max\{\theta,\theta^{-1}\}C \cdot \frac{(\ln k)^{\frac{1+\lambda}{2}}}{\sqrt{k}},$$

$$C := \inf_{x^* \in X^*}\left\{B(x^*)^2 \cdot \max_{0 \leq k \leq k_0} \mathbb{E}\left[\|x^k - x^*\|^2\right]\right\}.$$

Remark: ROBUST STEPSIZES.

Results

Theorem (Rate of convergence: bounded case)

Suppose **bounded operator** *or* **compact** X_0 *with same stepsize as before. Then a.s.-***asymptotic-convergence** *holds with*

$$\mathbb{E}\left[\mathrm{d}(\widehat{x}^k, X^*)\right] \lesssim \mathcal{O}(1)\max\{\theta, \theta^{-1}\}\,\mathrm{d}(x^0, X^*)^2 \cdot \frac{(\ln k)^{\frac{1+\lambda}{2}}}{\sqrt{k}},$$

or

$$\mathbb{E}\left[\mathrm{d}(\widehat{x}^k_{\lceil rk\rceil}, X^*)\right] \lesssim \mathcal{O}(1)\max\{\theta, \theta^{-1}\}\,\mathrm{diam}(X_0)^2 \cdot \frac{(\ln k)^{\frac{1+\lambda}{2}}}{\sqrt{k}},$$

respectively.

Results

Theorem (Rate of convergence: bounded case)

Suppose **bounded operator** *or* **compact** X_0 *with same stepsize as before. Then a.s.-***asymptotic-convergence** *holds with*

$$\mathbb{E}\left[\mathrm{d}(\widehat{x}^k, X^*)\right] \lesssim \mathcal{O}(1)\max\{\theta, \theta^{-1}\}\,\mathrm{d}(x^0, X^*)^2 \cdot \frac{(\ln k)^{\frac{1+\lambda}{2}}}{\sqrt{k}},$$

or

$$\mathbb{E}\left[\mathrm{d}(\widehat{x}^k_{\lceil rk \rceil}, X^*)\right] \lesssim \mathcal{O}(1)\max\{\theta, \theta^{-1}\}\,\mathrm{diam}(X_0)^2 \cdot \frac{(\ln k)^{\frac{1+\lambda}{2}}}{\sqrt{k}},$$

respectively.

- ROBUST STEPSIZES.

Results: asymptotic-convergence not required

Corollary (Convergence rates for larger stepsizes: bounded case)

Suppose compact case. Then

- **Constant stepsize**: *if* $\alpha_k \equiv \theta\alpha$,

$$\mathbb{E}\left[\mathsf{d}(\widehat{x}^k, X^*)\right] \lesssim \frac{1}{k} + O(\alpha).$$

Results: asymptotic-convergence not required

Corollary (Convergence rates for larger stepsizes: bounded case)

Suppose compact case. Then

- **Constant stepsize**: *if* $\alpha_k \equiv \theta\alpha$,

$$\mathbb{E}\left[\mathrm{d}(\widehat{x}^k, X^*)\right] \lesssim \frac{1}{k} + O(\alpha).$$

- *if* $\alpha_k := \frac{\theta}{\sqrt{k}}$, *then*

$$\mathbb{E}\left[\mathrm{d}(\widehat{x}^k_{\lceil rk \rceil}, X^*)\right] \lesssim \frac{1}{\sqrt{k}}.$$

Results

Corollary (An auxiliary simpler optimization problem)

Suppose that T is (L,δ)-Hölder continuous and

1. *T is unbounded and $\delta = 1$ or,*
2. *T is bounded or X_0 is compact.*

Then, there exists $\mathsf{V} > 0$, such that for all $k \geq 2$ with

$$k \sim \left(\frac{\mathsf{V}L^{\delta}}{\rho^{1+\delta}}\right)^2,$$

we have

$$argmin_{x\in X}\left\langle \mathbb{E}\left[F(\xi,\widehat{x}^k)\right], x\right\rangle \subset X^*.$$

Case 2: **Plain monotone** SVI & $\epsilon^k > 0$

Case 2: **Plain monotone** SVI & $\epsilon^k > 0$

Cartesian structure: m agents, (DISTRIBUTED SOLUTION)

- $\mathbb{R}^n = \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m}$,
- $n = n_1 + \ldots + n_m$
- $\langle x, y \rangle = \sum_{j=1}^m \langle x_j, y_j \rangle$,
- $X = X^1 \times \cdots \times X^m$,
- $F = (F_1, \ldots, F_m)$,

Case 2: **Plain monotone** SVI & $\epsilon^k > 0$

Cartesian structure: m agents, (DISTRIBUTED SOLUTION)

- $\mathbb{R}^n = \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m}$,
- $n = n_1 + \ldots + n_m$
- $\langle x, y\rangle = \sum_{j=1}^m \langle x_j, y_j\rangle$,
- $X = X^1 \times \cdots \times X^m$,
- $F = (F_1, \ldots, F_m)$,

Constraint structure: given $j \in [m]$, (INCREMENTAL PROJ.)

- $X^j = X_0^j \cap \left(\cap_{i \in \mathcal{I}_j} X_i^j\right)$,
- for all $i \in \mathcal{I}_j$, $X_i^j = \{x \in \mathbb{R}^n : g_i(j|x) \leq 0\}$,
- X_0^j has easy projections,
- for every $i \in \mathcal{I}_j$, subgradients of $g_i^+(j|\cdot)$ are easily computable,
- $\{\partial g_i^+(j|\cdot) : i \in \mathcal{I}_j\}$ is uniformly bounded over X_0^j.

Case 2: **Plain monotone** SVI & $\epsilon^k > 0$

Algorithm (Incremental constraint projection method: distributed case)

$$
\begin{aligned}
y_j^k &= \Pi_{X_0^j}\left[x_j^k - \alpha_{k,j}\left(F_j(v^k, x^k) + \epsilon_{k,j} x_j^k\right)\right], \\
x_j^{k+1} &= \Pi_{X_0^j}\left[y_j^k - \beta_{k,j}\frac{g^+_{\omega_{k,j}}(j|y_j^k)}{\|d_j^k\|^2} d_j^k\right],
\end{aligned}
$$

where $d_j^k \in \partial g^+_{\omega_{k,j}}(j|y_j^k) - \{0\}$ *if* $g_{\omega_{k,j}}(j|y_j^k) > 0$.

OBS: Includes the case of **agents** with **different stepsizes** and **regularization parameters**.

Case 2: **Plain monotone** SVI & $\epsilon^k > 0$

Typical example: soft constraints with easy **projection**:

Algorithm

$$y_j^k = x_j^k - \alpha_{k,j}\left(F_j(\xi^k, x^k) + \epsilon_{k,j} x_j^k\right),$$
$$x_j^{k+1} = \Pi_{X_0^j}\left[y_j^k - \beta_{k,j}\left(y_j^k - \Pi_{\omega_{k,j}}(y_j^k)\right)\right].$$

Case 2: Plain monotone SVI & $\epsilon^k > 0$

Same assumptions as before, but

- T is **monotone** and Lipschitz.

Case 2: Plain monotone SVI & $\epsilon^k > 0$

Same assumptions as before, but

- T is **monotone** and Lipschitz.
- **Regularization** parameters: $\lim_{k\to\infty} \epsilon_{k,j} = 0$.

Case 2: Plain monotone SVI & $\epsilon^k > 0$

Same assumptions as before, but

- T is **monotone** and Lipschitz.
- **Regularization** parameters: $\lim_{k\to\infty} \epsilon_{k,j} = 0$.
- **Partial Coordination** between stepsize and regularization, including:

$$\sum_{k=0}^{\infty} \frac{(\alpha_{k,\max} - \alpha_{k,\min})^2}{\alpha_{k,\min}\epsilon_{k,\min}} < \infty.$$

Case 2: Plain monotone SVI & $\epsilon^k > 0$

Same assumptions as before, but

- T is **monotone** and Lipschitz.
- **Regularization** parameters: $\lim_{k\to\infty} \epsilon_{k,j} = 0$.
- **Partial Coordination** between stepsize and regularization, including:

$$\sum_{k=0}^{\infty} \frac{(\alpha_{k,\max} - \alpha_{k,\min})^2}{\alpha_{k,\min}\epsilon_{k,\min}} < \infty.$$

Typical: $\alpha_{k,j} = (k + C_j)^{-c}$ and $\epsilon_{k,j} = (k + D_j)^{-d}$ with $0 < c + d < 1$.

Results

Theorem (Asymptotic convergence)

(i) *If* $\limsup_{k\to\infty} \frac{\epsilon_{k,\max}}{\epsilon_{k,\min}} < \infty$, *then a.s.* $\{x^k\}$ *is bounded and all cluster points of* $\{x^k\}$ *belong to* X^*,

(ii) *If* $\limsup_{k\to\infty} \frac{\epsilon_{k,\max}}{\epsilon_{k,\min}} \leq 1$, *then a.s.* $\{x^k\}$ *converges to the least-norm solution in* X^*.

Incremental constraint SA-extragradient method

Objective: remove **regularization**:

- better rate of convergence,
- less coordination between agents' parameters (important in distributed solutions).

Incremental constraint SA-extragradient method

Objective: remove **regularization**:

- better rate of convergence,
- less coordination between agents' parameters (important in distributed solutions).

Algorithm

$$
\begin{aligned}
y_1^k &:= x^k - \alpha_k F(\xi^k, x^k), \\
z^k &:= \Pi_{X_0}\left[y_1^k - \beta_k\left(y_1^k - \Pi_{\omega_k}(y_1^k)\right)\right], \\
y_2^k &:= x^k - \alpha_k F(\eta^k, z^k), \\
x^{k+1} &:= \Pi_{X_0}\left[y_2^k - \beta_k\left(y_2^k - \Pi_{\omega_k}(y_2^k)\right)\right].
\end{aligned}
$$

Results

Assume:

- Hard constraint X_0 compact,
- same set of assumptions as before (no Lipschitz-continuity required).

Results

Assume:

- Hard constraint X_0 compact,
- same set of assumptions as before (no Lipschitz-continuity required).

Theorem (Rate of convergence)

For $\alpha_k := \frac{\theta}{\sqrt{k}}$,

- *For the* **ergodic average** $\bar{z}^k := \frac{\sum_{i=0}^k \alpha_i z^i}{\sum_{i=0}^k \alpha_i}$,

$$\mathbb{E}[G(\bar{z}^k)] \leq \mathcal{O}(1) \max\{\theta, \theta^{-1}\}(M^2 + c)\frac{\ln k}{\sqrt{k}}.$$

- *For* **Nesterov-type weights** $\hat{z}^k := (1 - \theta_k)\hat{z}^k + \theta_k z^k$,

$$\mathbb{E}[G(\hat{z}^k)] \leq \mathcal{O}(1) \max\{\theta, \theta^{-1}\}(M^2 + c)\frac{1}{\sqrt{k}}.$$

Remark: ROBUST STEPSIZES and no Lipschitz-continuity requirement.

Results

Corollary (Constant stepsize)

If $\alpha_k \equiv \theta\alpha$ *then*

$$\mathbb{E}[G(\bar{z}^k)] \leq \frac{1}{k} + \mathcal{O}(\alpha).$$

THANK YOU!