Inexact and Distributed Best-Response Schemes for Stochastic Nash Equilibrium Problems

Uday V. Shanbhag

Jointly with Jinlong Lei[+], Jong-Shi Pang[%], and Suvrajeet Sen[%]

Industrial and Manufacturing Engg.
Pennsylvania State University
University Park, PA 16802

One World Optimization Seminar
November 2, 2020

+: Tongji University; %: USC

N-player Nash Equilibrium Problem

- players: $\mathcal{N} \triangleq \{1, \cdots, N\}$ is a set of N players, indexed by i
- strategy x_i and strategy set for player i: X_i

$$x \triangleq \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix} \in X \triangleq \prod_{i=1}^{N} X_i.$$

- objective of player i: $f_i(x_i, x_{-i}) : X \to \mathbb{R}$, where $x_{-i} \triangleq \{x_j\}_{j \neq i}$.
- objective: each player minimizes its objective given rivals' actions

$$\min_{x_i \in X_i} \quad f_i(x_i, x_{-i}).$$

- $x^* = \{x_i^*\}_{i=1}^{N}$ is a Nash Equilibrium if for any $i \in \mathcal{N}$

$$f_i(x_i^*, x_{-i}^*) \leq f_i(x_i, x_{-i}^*), \quad \forall x_i \in X_i.$$

- no player can improve her payoff by unilaterally deviating from strategy x_i^*

Introduction
○●○○○○○○○○○○○○○

BR for SNEPs
○○○○○○○○○○○○○○○○○○○○

Distributed BR and GR schemes
○○○○○○○○○○○○○○○○○○○

Background

Assumptions

- For $i \in \mathcal{N}$, the strategy set $X_i \subseteq \mathbb{R}^{n_i}$ is a closed and convex set

- For $i \in \mathcal{N}$, suppose $f_i(x_i, x_{-i})$ is convex and C^1 in x_i on an open set containing X_i.

- The resulting class of noncooperative games, referred to as $\mathcal{G}$, is a class of static convex games.

Background

Synchronous Gradient-Response (GR) schemes

Loosely speaking, in GR schemes, player i uses the GR, given current rival strategies.

Algorithm 1 Synchronous GR scheme

Set $k = 0$, $x_{i,0} \in X_i$ be a given sequence.

(1) For $i = 1, \ldots, N$, let $x_{i,k+1} \in X_i$ be $\widehat{x}_{i,k+1}$ defined as follows

$$\widehat{x}_{i,k+1} := \Pi_{X_i} \left[x_{i,k} - \gamma_{i,k} \nabla_{x_i} f_i(x_k) \right].$$

(2) $k := k + 1$; If $k < K$, return to (1); else STOP.

- Computing a GR requires player i to have access to $\nabla_{x_i} f_i(\cdot, x_{-i})$ and $x_{-i,k}$
- Lends itself to (partially) distributed implementations
- Under some conditions, synchronous GR leads to an NE

$\Pi_X(y) \triangleq \arg \min_{u \in X} \| u - y \|$

Introduction
○○○●○○○○○○○○○○○

BR for SNEPs
○○○○○○○○○○○○○○○○○○○○○○

Distributed BR and GR schemes
○○○○○○○○○○○○○○○○○○○○

Background

Convergence of SGR scheme

- Consider a game $\mathscr{G} \in \mathcal{G}$. Then x^* is an NE of $\mathscr{G}$ if and only if x^* is a solution of $VI(X, F)$ where

$$F(x) \triangleq \begin{pmatrix} \nabla_{x_1} f_1(x) \\ \vdots \\ \nabla_{x_N} f_N(x) \end{pmatrix}.$$

- Recall that the variational inequality problem $VI(X, F)$ requires an $x \in X$ such that

$$(y - x)^\top F(x) \geq 0, \qquad \forall y \in X.$$

- If F is $L-$Lipschitz and η-strongly monotone on X, then for $\gamma_{i,k} = \gamma < \frac{2\eta}{L^2}$ for all i and k, SGR converges to (unique) NE

Synchronous Iteratively Regularized Gradient-Response (GR) schemes

---

Algorithm 2 Synchronous RGR scheme

Set $k = 0$, $x_{i,0} \in X_i$ be a given sequence.

(1) For $i = 1, \ldots, N$, let $x_{i,k+1} \in X_i$ be $\widehat{x}_{i,k+1}$ defined as follows

$$\widehat{x}_{i,k+1} := \Pi_{X_i} \left[x_{i,k} - \gamma_{i,k} (\nabla_{x_i} f_i(x_k) + \epsilon_{i,k} x_{i,k}) \right].$$

(2) $k := k + 1$; If $k < K$, return to (1); else STOP.

- Suppose F is monotone on X.
- Under suitable conditions on $\gamma_{i,k}$ and $\epsilon_{i,k}$, synchronous RGR converges to (least-norm) NE

F is η-monotone if there exists $\eta > 0$, such that $(F(x) - F(y))^T (x - y) \geq \eta \|x - y\|^2$ for all $x, y \in X$.

Non-exhaustive summary of research

◆ Monograph on learning equilibria in games **[Fudenberg and Tirole, 1998]**
Single-timescale gradient-response schemes:

- Strongly monotone maps **[Alpcan and Başar (2003, 2007); Pavel (2006), Pan and Pavel (2009)]**
- Monotone maps via iterative regularization (single-projection) **[Yin, UVS and Mehta (2011); Kannan and UVS (2012)]**
- More recently, projected reflected gradient schemes (single projection) **[Malitksy (2015)]**
- Not "fully rational".

Synchronous Best-Response (BR) schemes

Loosely speaking, in BR schemes, player i uses the BR, given current rival strategies.

Algorithm 3 Synchronous BR scheme

Set $k = 0$, $x_{i,0} \in X_i$ be a given sequence.

(1) For $i = 1, \ldots, N$, let $x_{i,k+1} \in X_i$ be $\widehat{x}_{i,k+1}$ defined as follows:

$$\widehat{x}_{i,k+1} \in \operatorname*{argmin}_{x_i \in X_i} f_i(x_i, x_{-i,k}).$$

(2) $k := k + 1$; If $k < K$, return to (1); else STOP.

- Computing a BR requires player i to know f_i and x_{-i}^k but not $f_j, j \neq i$.
- Lends itself to (partially) distributed implementations
- BR schemes may converge to a NE or may cycle

Synchronous Proximal BR schemes

Algorithm 4 Synchronous proximal BR scheme

Set $k = 0$, $x_{i,0} \in X_i$ be a given sequence.

(1) For $i = 1, \ldots, N$, let $x_{i,k+1} \in X_i$ be $\widehat{x}_{i,k+1}(x_k)$ defined as follows:

$$\widehat{x}_{i,k+1}(x_k) = \operatorname*{argmin}_{x_i \in X_i} f_i(x_i, x_{-i,k}) + \frac{\mu}{2}\|x_i - x_{i,k}\|^2.$$

(2) $k := k + 1$; If $k < K$, return to (1); else STOP.

- Proximal BR adds a proximal term $\frac{\mu}{2}\|x_i - x_{i,k}\|^2$.
- If $f_i(\bullet, x_{-i})$ is convex in x_i, then $f_i(\bullet, x_{-i}) + \frac{\mu}{2}\| \bullet - x_{i,k}\|^2$ is μ-strongly convex (and BR is unique)
- Under some conditions, proximal BR converges to an NE

Introduction
0000000●0000000

BR for SNEPs
0000000000000000000

Distributed BR and GR schemes
0000000000000000000

Proximal BR [Facchinei and Pang, 2009]

◆ Fixed Point: $x^* = \widehat{x}(x^*)$

- x^* is an NE if and only if x^* is a fixed point of the proximal BR (PBR) map $\widehat{x}(\bullet)$
- $x_{k+1} = \widehat{x}(x_k)$ converges linearly to x^* when $\widehat{x}(\bullet)$ is contractive
- Define the $N \times N$ real matrix $\Gamma = [\gamma_{ij}]_{i,j=1}^N$:

$$\Gamma \triangleq \begin{pmatrix} \frac{\mu}{\mu+\zeta_{1,\min}} & \frac{\zeta_{12,\max}}{\mu+\zeta_{1,\min}} & \cdots & \frac{\zeta_{1N,\max}}{\mu+\zeta_{1,\min}} \\ \frac{\zeta_{21,\max}}{\mu+\zeta_{2,\min}} & \frac{\mu}{\mu+\zeta_{2,\min}} & \cdots & \frac{\zeta_{2N,\max}}{\mu+\zeta_{2,\min}} \\ \vdots & & \ddots & \\ \frac{\zeta_{N1,\max}}{\mu+\zeta_{N,\min}} & \frac{\zeta_{N2,\max}}{\mu+\zeta_{N,\min}} & \cdots & \frac{\mu}{\mu+\zeta_{N,\min}} \end{pmatrix}$$

  with $\zeta_{i,\min} \triangleq \inf_{x \in X} \lambda_{\min} \left(\nabla_{x_i}^2 f_i(x) \right)$, and $\zeta_{ij,\max} \triangleq \sup_{x \in X} \| \nabla_{x_i x_j}^2 f_i(x) \| \ \forall j \neq i$ measuring the coupling of players' subproblem.

- If the spectral radius $\rho(\Gamma) < 1$, then there exist a scalar $a \in (0,1)$ and monotonic norm $| \| \bullet \| |$ such that

$$\left| \left\| \begin{pmatrix} \|\widehat{x}_1(y') - \widehat{x}_1(y)\| \\ \vdots \\ \|\widehat{x}_N(y') - \widehat{x}_N(y)\| \end{pmatrix} \right\| \right| \leq a \left| \left\| \begin{pmatrix} \|y_1' - y_1\| \\ \vdots \\ \|y_N' - y_N\| \end{pmatrix} \right\| \right| .$$

Randomized proximal best-response scheme

For any $i \in \mathcal{N}$, let $\chi_{i,k} = 1$ (or 0) if player i updates at iteration k (or not).

◆ Assumption: For any $i \in \mathcal{N}$, $\mathbb{P}(\chi_{i,k} = 1) = p_i > 0$ and $\chi_{i,k}$ is independent of $\mathcal{F}_k$.

Algorithm 5 Randomized proximal best-response scheme

Let $k := 0$, $x_{i,0} \in X_i$ for $i = 1, \ldots, N$.

(1) If $\chi_{i,k} = 1$, then $x_{i,k+1} \in X_i$ is defined as follows:

$$\widehat{x}_{i,k+1}(x_k) = \underset{x_i \in X_i}{\operatorname{argmin}} \, f_i(x_i, x_{-i,k}) + \frac{\mu}{2} \| x_i - x_{i,k} \|^2.$$

Otherwise, $x_{i,k+1} = x_{i,k}$ when $\chi_{i,k} = 0$.

(2) $k := k + 1$; If $k < K$, return to (1); else STOP.

- A collection of players is randomly chosen to update via proximal BR

Introduction
○○○○○○○○○○○●○○○○○○

BR for SNEPs
○○○○○○○○○○○○○○○○○○○○○

Distributed BR and GR schemes
○○○○○○○○○○○○○○○○○○

Background

Non-exhaustive literature review on BR schemes

- Synchronous best-response schemes **[Facchinei and Pang (2009); Scutari, Facchinei, Palomar, Song, and Pang (2013)]**

- Customized schemes in signal processing **[Scutari, Palomar and Barbarossa (2008, 2009); Scutari and Palomar (2010)]**

N-player Stochastic Nash Equilibrium Problems

Key difference: Player objectives are expectation-valued; objective of player i:
$f_i : X \times \mathbb{R}^d \to \mathbb{R}$

$$f_i(x_i, x_{-i}) \triangleq \mathbb{E}\left[\psi_i(x_i, x_{-i}; \xi(\omega))\right]$$

where $\xi : \Omega \to \mathbb{R}^d$ denotes a random variable

- A direct extension of GR and BR is impossible since it requires access to either $\nabla_{x_i}\mathbb{E}[\psi_i(x_i, x_{-i}, \xi)]$ (GR) or exact solutions to (BR) in finite time.

◆ Existence of a stochastic first-order oracle (SFO):
For any $i \in \mathcal{N}$ and x, ξ, (SFO) returns a sampled gradient $\nabla_{x_i}\psi_i(x_i, x_{-i}; \xi)$ s.t.

- Unbiased: $\nabla_{x_i}f_i(x_i, x_{-i}) = \mathbb{E}[\nabla_{x_i}\psi_i(x_i, x_{-i}; \xi(\omega))]$;
- Bounded second moments: There exists $M_i > 0$ such that for all $x \in X$,

$$\mathbb{E}[\|\nabla_{x_i}\psi_i(x_i, x_{-i}; \xi(\omega))\|^2] \leq M_i^2.$$

Stochastic variational inequality problems

- Consider a game $\mathscr{G} \in \mathcal{G}$ where player problems are expectation-valued. Then x^* is an NE of $\mathscr{G}$ if and only if x^* is a solution of $\text{VI}(X, F)$ where

$$F(x) \triangleq \begin{pmatrix} \nabla_{x_1} \mathbb{E}[f_1(x, \xi)] \\ \vdots \\ \nabla_{x_N} \mathbb{E}[f_N(x, \xi)] \end{pmatrix}.$$

- Consequently, algorithms for stochastic VIs or SVIs are closely related to schemes for computing NE in stochastic regimes.

Introduction   **BR for SNEPs**   **Distributed BR and GR schemes**
○○○○○○○○○○○○○○○●○○   ○○○○○○○○○○○○○○○○○○○○○   ○○○○○○○○○○○○○○○○○○

Background

Extension of GR to Stochastic regimes

Algorithm 6 Synchronous Stochastic GR scheme

Set $k = 0$, $x_{i,0} \in X_i$ be a given sequence.

(1) For $i = 1, \ldots, N$, let $x_{i,k+1} \in X_i$ be $\widehat{x}_{i,k+1}$ defined as follows

$$\boxed{\widehat{x}_{i,k+1} := \Pi_{X_i} \left[x_{i,k} - \gamma_{i,k} \nabla_{x_i} f_i(x_k, \omega_{i,k}) \right].}$$

(2) $k := k + 1$; If $k < K$, return to (1); else STOP.

Introduction   **BR for SNEPs**   **Distributed BR and GR schemes**
○○○○○○○○○○○○○●○○   ○○○○○○○○○○○○○○○○○○○○   ○○○○○○○○○○○○○○○○○○○

Background

Non-exhaustive review for SGR schemes (and SVIs)

- a.s. convergence for strongly monotone and Lipschitz maps **[Jiang and Xu, 2008]**
- Rate statements for monotone and Lipschitz maps via extragradient schemes **[Juditsky, Nemirovski, and Tauvel, 2011], [Dang and Lan, 2015]**
- a.s. convergence for monotone and Lipschitz maps under single projection regularized schemes **[Koshal, Nedić and Shanbhag (2013)]**
- Non-Lipschitzian regimes via random smoothing **[Yousefian, Nedić and Shanbhag (2016)]**
- Related work on non-monotone regimes (cf. **[Thompson, Jofré, Iusem (2017)], [Kannan and Shanbhag (2019)]**)

Background

Motivation

Part I: Best-response in stochastic regimes

◆ Closed-form expression of proximal best-response map is unavailable in finite time since the objective is expectation-valued

◆ Part I develops several *inexact* proximal best-response schemes

- best-response solutions are approximated via stochastic approximation (SA),
- and inexactness $\downarrow$ zero by an increasing number of projected gradient steps.
- Extensions to asynchronous and randomized regimes

Part II: Distributed BR and GR in stochastic regimes

◆ Part II develops distributed schemes for a subclass of games, i.e. "aggregative" where player objectives are coupled via the aggregate strategy

- Add a consensus layer for players to "learn" aggregate
- Examine BR and GR with variance reduction+multiple communication rounds
- Goal: Under what conditions, can linear convergence rates be achieved?

Introduction

Slightly stricter assumptions

◆ Convexity of subproblems

- X_i is a closed, compact, convex set.
- $f_i(x_i, x_{-i})$ is convex and C^2 in x_i over an open set containing X_i for any given $x_{-i} \in \prod_{j \neq i} X_j$.

Introduction
○○○○○○○○○○○○○○○○

BR for SNEPs
○●○○○○○○○○○○○○○○○○○○○○

Distributed BR and GR schemes
○○○○○○○○○○○○○○○○○○○○

Background on Proximal Best-response

◆ Proximal best-response map

$$\widehat{x}(y) \triangleq \underset{x \in X}{\mathrm{argmin}} \left[\sum_{i=1}^{N} \mathbb{E}[\psi_i(x_i, y_{-i}; \omega)] + \frac{\mu}{2} \|x - y\|^2 \right], \quad \mu > 0$$

The objective function is separable in x_i, player i's subproblem is

$$\widehat{x}_i(y) \triangleq \underset{x_i \in X_i}{\mathrm{argmin}} \left[\mathbb{E}[\psi_i(x_i, y_{-i}; \omega)] + \frac{\mu}{2} \|x_i - y_i\|^2 \right].$$

Synchronous algorithm

Algorithm Design

Algorithm 7 Synchronous inexact proximal best-response scheme

---

Set $k = 0$, $x_{i,0} \in X_i$; Let $\{\alpha_{i,k}\}_{k \geq 1}$ be a given sequence.

(1) For $i = 1, \ldots, N$, let $x_{i,k+1} \in X_i$ be defined as follows:

$$x_{i,k+1} = \widehat{x}_i(x_k) + \varepsilon_{i,k+1}$$

with $\{\varepsilon_{i,k+1}\}$ satisfying $\mathbb{E}\left[\|\varepsilon_{i,k+1}\|^2 \big| \mathcal{F}_k\right] \leq \alpha_{i,k}^2$ a.s., where $\mathcal{F}_k = \sigma\{x_0, \cdots, x_k\}$.

(2) $k := k + 1$; If $k < K$, return to (1); else STOP.

$$\boxed{\widehat{x}_i(x_k) \triangleq \underset{x_i \in X_i}{\operatorname{argmin}} \left[\mathbb{E}[\psi_i(x_i, x_{-i,k}; \omega)] + \frac{\mu}{2} \|x_i - x_{i,k}\|^2 \right].}$$

◆ Stochastic approximation (SA) to obtain an inexact best-response.

$$z_{i,t+1} := \Pi_{X_i} \left[z_{i,t} - \gamma_t \left(\nabla_{x_i} \psi_i(z_{i,t}, x_{-i,k}; \xi_{i,k}^t) + \mu(z_{i,t} - x_{i,k}) \right) \right], \quad \text{(SA}_{i,k})$$

where $z_{i,1} = x_{i,k}$, $\gamma_{i,t} = 1/\mu(t+1)$. Set $x_{i,k+1} = z_{i,j_{i,k}}$.

Lemma (**Error Bounds of SA [Nemirovski et al., 2009]**)

Define $\xi_{i,k} = (\xi_{i,k}^1, \cdots, \xi_{i,k}^{j_{i,k}})$, and $\mathcal{F}_k = \sigma\{x_0, \xi_{i,l}, i \in \mathcal{N}, 0 \le l \le k-1\}$. Assume that for any $i \in \mathcal{N}$, the random variables $\{\xi_{i,k}^t\}_{1 \le t \le j_{i,k}}$ are iid and the random vector $\xi_{i,k}$ is independent of $\mathcal{F}_k$. Then for any $t \ge 1$ we have

$$\mathbb{E}\left[\|z_{i,t} - \widehat{x}_i(x_k)\|^2 \big| \mathcal{F}_k\right] \le \frac{Q_i}{(t+1)}, \quad a.s.$$

where $Q_i \triangleq \frac{2M_i^2}{\mu^2} + 2D_{X_i}^2$, and $D_{X_i} = \sup\{d(x_i, x_i') : x_i, x_i' \in X_i\}$.

$$\mathbb{E}\left[\|\varepsilon_{i,k+1}\|^2 \big| \mathcal{F}_k\right] = \mathbb{E}\left[\|x_{i,k+1} - \widehat{x}_i(y_k)\|^2 \big| \mathcal{F}_k\right] \le \frac{Q_i}{j_{i,k}} =: \alpha_{i,k}^2$$

Almost Sure Convergence

Let the sequence $\{x_k\}_{k \geq 0}$ be generated by the synchronous algorithm. Assume that $\|\Gamma\| < 1$, and $\alpha_{i,k} \geq 0$ with $\sum_{k=1}^{\infty} \alpha_{i,k} < \infty$ for any $i \in \mathcal{N}$. Then for any $i \in \mathcal{N}$,

$$\lim_{k \to \infty} x_{i,k} = x_i^* \quad a.s.$$

Convergence in Mean and of the Variance

Let the sequence $\{x_k\}_{k=1}^{\infty}$ be generated by the synchronous algorithm. Assume that $\|\Gamma\| < 1$, and that $0 \leq \alpha_{i,k} \to 0$ as $k \to \infty$ for any $i \in \mathcal{N}$. Then for any $i \in \mathcal{N}$,

(a) (**convergence in mean**) $\lim_{k \to \infty} \mathbb{E}[\|x_{i,k} - x_i^*\|] = 0$.

(b) (**convergence of the variance of** x_k) $\lim_{k \to \infty} \mathbb{V}\text{ar}(x_k) = 0$.

Geometric Convergence

Consider the synchronous scheme where $\mathbb{E}[\|x_{i,0} - x_i^*\|] \leq C \;\forall i \in \mathcal{N}$. Assume that $a = \|\Gamma\| < 1$, and that $\alpha_{i,k} = \eta^k \;\forall i \in \mathcal{V}$ with $\eta \in (0,1)$. Define

$$u_k = \mathbb{E}\left[\left\|\begin{pmatrix} \|x_{1,k} - x_1^*\| \\ \vdots \\ \|x_{N,k} - x_N^*\| \end{pmatrix}\right\|\right].$$

Then, the following holds for $k \geq 0$

(a) If $\eta = a$, $q > a$ and $D \triangleq 1/\ln((q/a)^e)$, then $u_k \leq (u_0 + \sqrt{N}k)a^k \leq \sqrt{N}(C + D)q^k$.

(b) If $\eta \in (a,1)$, then $u_k \leq \left(\sqrt{N}C + \frac{\sqrt{N}\eta}{\eta - a}\right) q^k$ with $q = \eta$.

(c) If $0 < \eta < a$, then $u_k \leq \left(\sqrt{N}C + \frac{\sqrt{N}a}{a - \eta}\right) q^k$ with $q = a$.

Overall iteration complexity

Consider the synchronous scheme and let inexact solutions be computed via SA, where $\mathbb{E}[\|x_{i,0} - x_i^*\|^2] \leq C^2$. Assume that $a = \|\Gamma\| < 1$ and $\alpha_{i,k} = \eta^k \ \forall i \in \mathcal{V}$ with $\eta \in (0,1)$. Then the number of projected gradient steps[a] for i to achieve an ϵ–NE is no greater than $\mathcal{O}\left(\frac{\sqrt{N}}{\epsilon}\right)^2 + \left(\ln\left(\frac{\sqrt{N}}{\epsilon}\right)\right)$.

---

[a]Suppose $\ell_i(\eta) = \sum_{k=1}^{K(\epsilon)} j_{i,k}$ with $j_{i,k} = \lceil \frac{Q_i}{\eta^{2(k+1)}} \rceil$. If $\eta \leq a$, then

$\ell_i(\eta) \leq \frac{Q_i}{\eta^4 \ln(1/\eta^2)} \left(\frac{\sqrt{N}(C+D)}{\epsilon} \right)^{\frac{\ln(1/\eta^2)}{\ln(1/q)}} + \frac{\ln\left(\sqrt{N}(C+D)/\epsilon\right)}{\ln(1/q)}$, where $Q_i \triangleq \frac{2M_i^2}{\mu^2} + 2D_{X_i}^2$, $q > a = \|\Gamma\|$, and $D = 1/\ln((q/c)^e)$.

◆ The bound grows slowly in N, a desirable feature of equilibrium computation with a large collection of players.

Comparison with Stochastic Gradient Response: Competitive portfolio Investment C. OCinneide, B. Scherer, and X. Xu (2006)
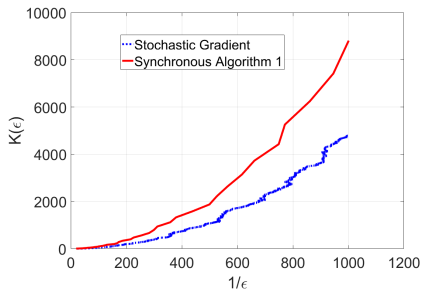

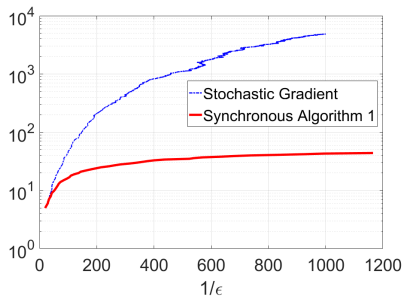Figure: Empirical Iteration Complexity

Figure: Empirical Communication Complexity

- The iteration complexity is of the same order as stochastic gradient response (SGR); but the constant of SG is superior to that of the synchronous BR scheme.
- Significant decrease in communication overhead compared to SGR; communication overhead often crucial in rendering a scheme impractical.

Introduction
○○○○○○○○○○○○○○○○

BR for SNEPs
○○○○○○○○●○○○○○○○○○○

Distributed BR and GR schemes
○○○○○○○○○○○○○○○○○○○○

Asynchronicity and delay

Asynchronous Scheme: Algorithm Design Bertsekas and Tsitsiklis (1989)

◆ Motivation:
In a large-scale network, players might not be able to make simultaneous updates nor have access to their rivals' latest information.

◆ Description:

- $T_i \subset T = \{0, 1, 2, \cdots\}$: the set of times player i updates x_i
- $y_k^i \triangleq (x_{1, k - d_1^i(k)}, \cdots, x_{n, k - d_N^i(k)})$ is available to player i if $k \in T_i$, where $d_j^i(k)$ denotes the communication delay

◆ Assumptions

- Almost Cyclic Rule: There exists an integer $B_1 > 0$ such that each player updates its decision at least once during any time interval of length B_1
- Partial Asynchronism: There exists an integer $B_2 \geq 0$ such that

$$0 \leq d_j^i(k) \leq B_2 \quad \forall i, j = 1, \cdots, N, \ k \geq 0$$

Algorithm 8 Asynchronous inexact proximal BR scheme

Let $k := 0$, $x_{i,0} \in X_i$ for $i = 1, \ldots, N$.

(1) For $i = 1, \ldots, N$, if $k \in T_i$, then set $y_k^i \triangleq (x_{1, k - d_1^i(k)}, \cdots, x_{n, k - d_N^i(k)})$.

(2) For $i = 1, \ldots, N$, if $k \in T_i$, then updates $x_{i,k+1} \in X_i$ as follows:

$$x_{i,k+1} = \widehat{x}_i(y_k) + \varepsilon_{i,k+1}$$

with $\varepsilon_{i,k+1}$ satisfying $\mathbb{E}\left[\|\varepsilon_{i,k+1}\|^2 \big| \mathcal{F}_k\right] \leq \alpha_{i,k}^2$ *a.s.*, where $\mathcal{F}_k = \sigma\{x_0, \cdots, x_k\}$.
Otherwise, if $k \notin T_i$, then $x_{i,k+1} := x_{i,k}$.

(3) $k := k + 1$; If $k < K$, return to (1); else STOP.

Define $n_0 = \lceil \frac{B_2}{B_1} \rceil$, let $\beta_{i,k}$ denote the number of elements in T_i that are not larger than k.

Lemma (**Linear Rate of Convergence**)

Let the asynchronous inexact proximal best-response scheme be applied to the N-player stochastic Nash game, where $\alpha_{i,k+1} = \eta^{\beta_{i,k}}$ for some $\eta \in (0,1)$, and $\mathbb{E}[\|x_{i,0} - x_i^*\|] \le C \; \forall i \in \mathcal{N}$. Assume $a = \|\Gamma\|_\infty < 1$. If $q > c \triangleq \rho^{\frac{1}{B_1}}$ and $D > 1/\ln((q/c)^e)$,

$$\max_{i \in \mathcal{N}} \mathbb{E}[\|\widehat{x}_{i,k} - x_i^*\|] \le \rho^{-\frac{B_1-1}{B_1}}(C+D)q^k \;\; \forall k \ge 0,$$

Iteration Complexity (Impact of delay and asynchronicity)

Consider the asynchronous algorithm and let the inexact proximal solutions be computed via SA, where $\alpha_{i,k+1} = \eta^{\beta_{i,k}}$ for $\eta \in (0,1)$. Suppose $a = \|\Gamma\|_\infty < 1$. Then the number of projected gradient steps[a] for i to compute an $\epsilon-$NE is no greater than $\mathcal{O}\left((1/\epsilon)^{2B_1(1+\lceil \frac{B_2}{B_1} \rceil)+\delta} \right)$.

---

[a] where $\max_{i \in \mathcal{N}} \mathbb{E}[\|x_{i,0} - x_i^*\|^2] \le C^2$, $Q_i \triangleq \frac{2M_i^2}{\mu^2} + 2\{D_{X_i}^2\}$, $\rho^{n_0+1} = \max\{a, \eta\}$ with $n_0 = \lceil \frac{B_2}{B_1} \rceil$ and $\eta \in (0,1), q > c \triangleq \rho^{\frac{1}{B_1}}$, and $D > 1/\ln((q/c)^e)$.

update	delay	complexity bound
B_1	B_2	$\mathcal{O}\left((1/\epsilon)^{2B_1(1+\lceil \frac{B_2}{B_1} \rceil)+\delta} \right)$
1	B_2	$\mathcal{O}\left((1/\epsilon)^{2(1+B_2)+\delta} \right)$
1	0	$\mathcal{O}\left((1/\epsilon)^{2+\delta} \right)$

Set $B_1 = 1$, the communication delays $k - \tau_j^i(k)$ are independently generated from a uniform distribution on the set $\{0, 1, \cdots, B_2\}$.
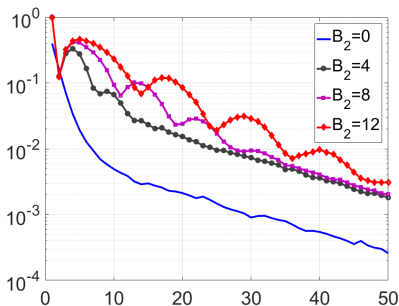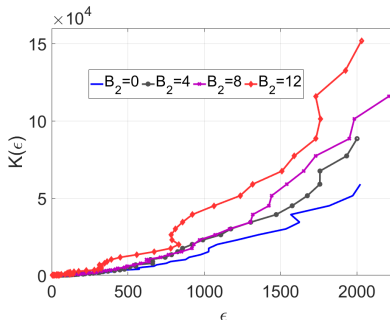

Figure: Linear Convergence

Figure: Empirical Iteration Complexity

Introduction   **BR for SNEPs**   **Distributed BR and GR schemes**
○○○○○○○○○○○○○○○○○   ○○○○○○○○○○○○○○○●●○○○○○○   ○○○○○○○○○○○○○○○○○○○○

Randomized BR: Algorithm Design

Randomized Best-Response Scheme I

Literature review

- The randomized block-coordinate descent method [Y. Nestrerov (2012)] partitions the coordinates into several blocks and randomly choses a single block to update while the other blocks keep invariant at each iteration.

- Generalized to the fixed point problem by [P. L. Combettes and J. C Pesquet (2015)], in which a subset of block variables is randomly updated

Randomized Best-response: For any $i \in \mathcal{N}$, let $\chi_{i,k} = 1$ (or 0) if player i updates at iteration k (or not).

◆ Assumption: For any $i \in \mathcal{N}$, $\mathbb{P}(\chi_{i,k} = 1) = p_i > 0$ and $\chi_{i,k}$ is independent of $\mathcal{F}_k$.

Randomized Best-Response Scheme II

Algorithm 9 Randomized inexact proximal best-response scheme

Let $k := 0$, $x_{i,0} \in X_i$ for $i = 1, \ldots, N$.

(1) If $\chi_{i,k} = 1$, then $x_{i,k+1} \in X_i$ is defined as follows:

$$x_{i,k+1} = \widehat{x}_i(x_k) + \varepsilon_{i,k+1}$$

with $\varepsilon_{i,k+1}$ satisfying $\mathbb{E}\left[\|\varepsilon_{i,k+1}\|^2 \big| \mathcal{F}_k\right] \leq \alpha_{i,k}^2$ *a.s.*, where $\mathcal{F}_k = \sigma\{x_0, \cdots, x_k\}$. Otherwise, $x_{i,k+1} = x_{i,k}$ when $\chi_{i,k} = 0$.

(2) $k := k + 1$; If $k < K$, return to (1); else STOP.

Almost Sure Convergence

Let the sequence $\{x_k\}_{k \geq 0}$ be generated by the randomized algorithm. Assume that $a = \|\Gamma\| < 1$ and for any $i \in \mathcal{N}$, $0 \leq \alpha_{i,k} < 1$ and $\sum_{k=0}^{\infty} \alpha_{i,k} < \infty$ *a.s.* Then for any $i \in \mathcal{N}$, $\lim_{k \to \infty} x_{i,k} = x_i^*$ *a.s.*

Geometric Convergence

Let the sequence $\{x_k\}_{k \geq 0}$ be generated by the randomized algorithm.[a] Then the following holds for $k \geq 0$,

$$\mathbb{E}\left[\|x_k - x^*\|\right] \leq \sqrt{N}(\widetilde{C} + \tilde{D})\tilde{q}^k.$$

[a] $\mathbb{E}[\|x_{i,0} - x_i^*\|] \leq C \; \forall i \in \mathcal{N}$ and $\alpha_{i,k} = \eta^{\beta_{i,k}+1}$ for some $\eta \in (0, 1)$. Define $\beta_{i,0} = 0$ and $\beta_{i,k} = \sum_{p=0}^{k-1} \chi_{i,p}$ for all $k \geq 1$, $\tilde{c} \triangleq \max\{\tilde{a}, \tilde{\eta}\}$ with $\tilde{a} = \sqrt{1 - p_{\min}(1 - a^2)}$, $\tilde{\eta} = \sqrt{1 - p_{\min}(1 - \eta^2)}$ and $p_{\min} = \min_{i \in \mathcal{N}} p_i$, $\tilde{q} > \tilde{c}$, $D \triangleq 1/\ln((\tilde{q}/\tilde{c})^e)$, $\widetilde{C} = C\left(\sum_{l=1}^{N} N^{-1}p_{\min}^{-1}\right)^{1/2}$, and $\tilde{D} = D\eta\tilde{\eta}^{-1}$

Overall Iteration Complexity

Let the randomized algorithm be applied with inexact solutions computed via SA, where $\alpha_{i,k} = \eta^{\beta_{i,k}+1}$ for some $\eta \in (0,1)$. Suppose $a = \|\Gamma\| < 1$. Then expected number of projected gradient steps[a] for i to compute an $\epsilon-$NE is no greater than

$$\mathcal{O}\left(\frac{\sqrt{N}p_{\max}}{\tilde{\epsilon}}\right)^{\frac{\ln(1/\tilde{\eta}_0^2)}{\ln(1/\tilde{q})}} + \left\lceil \frac{\ln(1/\tilde{\epsilon})}{\ln(1/\tilde{q})} \right\rceil.$$

---

[a]The expected number of gradient steps is bounded by $\bar{\ell}_i(\eta)$ where $\widetilde{\ell}_i(\eta) \triangleq \frac{p_i Q_i}{\eta^2 \tilde{\eta}_0^2 \ln(1/\tilde{\eta}_0^2)} \left(\frac{1}{\tilde{\epsilon}}\right)^{\frac{\ln(1/\tilde{\eta}_0^2)}{\ln(1/\tilde{q})}} + \left\lceil \frac{\ln(1/\tilde{\epsilon})}{\ln(1/\tilde{q})} \right\rceil$, where $\tilde{\eta}_0^2 = \left(p_{\max}(\eta^{-2}-1)+1\right)^{-1}$, $\tilde{\epsilon} \triangleq \frac{\epsilon}{(Np_{\max})^{1/2}(\widetilde{C}+\widetilde{D})}$, $\tilde{q} > \bar{c} \triangleq \max\{\tilde{a}, \tilde{\eta}\}$, $D \triangleq 1/\ln((\tilde{q}/\bar{c})^e)$, and $\widetilde{D} = D\eta\tilde{\eta}^{-1}$.

Introduction
ooooooooooooooooo

BR for SNEPs
oooooooooooooooooo●○○

Distributed BR and GR schemes
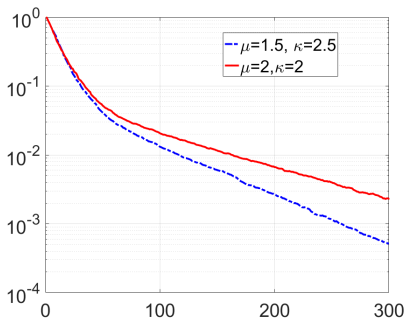oooooooooooooooooooooo

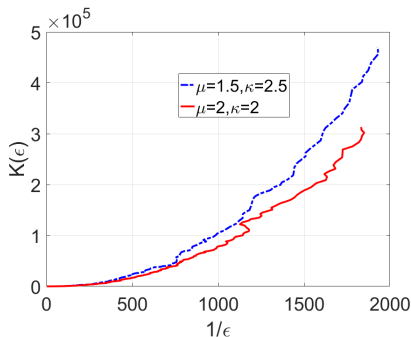**Convergence Analysis**

Figure: Linear Convergence

Figure: Empirical Iteration Complexity

- $j_{i,k} = \left\lceil \frac{1}{\eta^{2(\beta_{i,k}+1)}} \right\rceil$ steps of SA are taken to get an inexact solution.
- The randomized algorithm still displays linear convergence but its empirical iteration complexity is larger than the synchronous algorithm, *a less surprising observation*.

parameters		synchronous		randomized		asynchronous	
		empirical	theoretical	empirical	theoretical	empirical	theoretical
$\mu = 1$	$\eta = a^{0.5}$	2e-03	1.89	2.64e-03	1.98e+01	1.43e-03	1.36e+01
	$\eta = a^{0.75}$	4.76e-04	7.18e-01	7.42e-04	1.73e+01	3.24e-04	1.36e+01
	$\eta = a$	1.08e-04	3.18e-01	2.27e-04	1.53e+01	7.94e-05	1.37e+01
$\mu = 2$	$\eta = a^{0.5}$	6.1e-03	6.98	7.88e-03	2.49e+01	4.33e-03	3.69e+01
	$\eta = a^{0.75}$	2.26e-03	3.89	3.3e-03	2.27e+01	1.72e-03	3.69e+01
	$\eta = a$	9.39e-04	2.33	1.39e-03	2.09e+01	6.9e-04	3.69e+01
$\mu = 5$	$\eta = a^{0.5}$	1.3e-02	2.6e+01	2.06e-02	3.16e+01	1.11e-02	9.62e+01
	$\eta = a^{0.75}$	7.5e-03	2.01e+01	1.24e-02	3.01e+01	6.55e-03	9.62e+01
	$\eta = a$	4.8e-03	1.58e+01	8.4e-03	2.89e+01	4.2e-03	9.62e+01

Table: Comparison of theoretical and empirical error

Introduction
ooooooooooooooooo

BR for SNEPs
ooooooooooooooooooooo●

Distributed BR and GR schemes
ooooooooooooooooooooo

Summary of findings

| Update scheme	Asymptotic convergence	Rate of convergence	Iteration complexity
Synchronous Algorithm (using $\|.\|_2$ norm)	a.s. convergence convergence in mean	geometric	ϵ-NE$_2$: $\mathcal{O}\left((\sqrt{N}/\epsilon)^{2+\delta}\right)$ $\eta \in (a, 1)$: $\mathcal{O}(N/\epsilon^2)$
Randomized Algorithm (using $\|.\|_2$ norm)	a.s. convergence convergence in mean	geometric	ϵ-NE$_2$: $\mathcal{O}\left((\sqrt{N}/\epsilon)^{2\ln(\bar{\eta}_0^{-1})/\ln(\bar{\eta}^{-1})+\delta}\right)$
Asynchronous Algorithm (using $\|.\|_\infty$ norm)	convergence in mean	geometric	ϵ-NE$_\infty$: $\mathcal{O}\left((1/\epsilon)^{2B_1\left(1+\lceil\frac{B_2}{B_1}\rceil\right)+\delta}\right)$ $\mathcal{O}\left((1/\epsilon)^{2\left(1+\lceil\frac{B_2}{N}\rceil\right)+\delta}\right)$

Table: Summary of Contributions

◆ Key findings: the iteration complexity is $\mathcal{O}(1/\epsilon^{2(1+c)+\delta})$

- $c = 0$ for the synchronous scheme
- $c > 0$ represents the positive cost of randomization in the randomized scheme
- $c > 0$ represents the positive cost of asynchronicity and delay

Background

- Game-theoretic models and tools are extensively used in networks since
  - they enable a flexible control paradigm where agents autonomously control their resource usage to optimize their own selfish objectives;
  - provide potentially tractable decentralized algorithms for network control based on "designed games" [Marden and Shamma, 2007], [Marden 2009].
- Noncooperative games have wide application in capturing networked systems, such as power systems, markets, communication and transportation networks

Distributed schemes for aggregative games
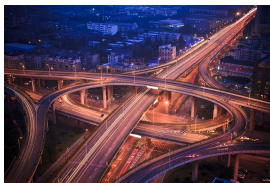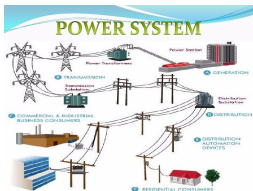
Problem Statement

Aggregative** Stochastic Nash Games

$$\min_{x_i \in X_i} f_i(x_i, \bar{x}) \triangleq \mathbb{E}\left[\psi_i(x_i, \bar{x}; \xi)\right]$$

- $\mathcal{N} = \{1, \cdots, n\}$ is a group of n players, indexed by i;
- X_i denotes the strategy set of player i while $x \triangleq (x_1, \cdots, x_N)$ denotes a *strategy profile*;
- player i has an objective $f_i(x_i, \bar{x})$, where $\bar{x} \triangleq \sum_{i=1}^n x_i$ is the aggregate;
- $\xi : \Omega \to \mathbb{R}^m$ defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

** Aggregative games first discussed in [Jensen, 2010]

- Convexity of subproblems: X_i is a closed, compact, convex set; For any $y \in \mathbb{R}^d$, $f_i(x_i, y)$ is C^1 and convex in $x_i \in X_i$.
- Existence of a stochastic oracle returning a sampled gradient $\nabla_{x_i}\psi_i(x_i, y; \xi)$, $\nabla_{x_i} f_i(x_i, y) = \mathbb{E}[\nabla_{x_i} \psi_i(x_i, y; \xi)]$ and $\mathbb{E}[\|\nabla_{x_i} f_i(x_i, y) - \nabla_{x_i} \psi_i(x_i, y; \xi)\|^2] \leq \nu_i^2$.

Previous Works

- Non-exhaustive summary of consensus and distributed optimization.
  [Tsitsiklis, 1984], [Olfati-Saber and Murray, 2004] [Ren, Beard and Atkins, 2005] [Nedić and Ozdaglar, 2009], [Nedić, Ozdaglar and Parrilo, 2010], [Nedić and Olshevsky, 2015]
- Distributed schemes for Nash games
  - Gradient response+consensus for aggregative games [Koshal, Nedić, and UVS, 2016]
  - Aggregative games with coupling constraints [Paccagnan et al., 2017] [Belgioioso et al., 2017], a semi-dencentralized algorithm, requiring a *central node* for the update of the common multiplier.
  - Generalized Nash equilibrium problems:
    - Distributed primal-dual algorithms [Zhu and Frazzoli, 2017; Yi and Pavel, 2017].
    - Distributed stochastic gradient scheme with constant stepsize [Yu et al., 2017], mean-squared convergence to a neighborhood of the GNE.

Introduction
ooooooooooooooooo

BR for SNEPs
oooooooooooooooooooooo

Distributed BR and GR schemes
ooooooooooooooooooooo

Distributed schemes for aggregative games

Our Work

- The players cannot observe rival strategies, while interacting through a communication graph (connected) $\mathcal{G} = (\mathcal{N}, \mathcal{E}, A)$:
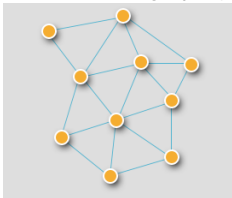
  - $\mathcal{E}$ is a collection of undirected edges;
  - Neighbor set $\mathcal{N}_i = \{j : (i, j) \in \mathcal{E}\}$;
  - The adjacency matrix $A = [a_{ij}]_{i,j=1}^n$, where $a_{ij} > 0$ if $j \in \mathcal{N}_i$ and $a_{ij} = 0$ otherwise such that A is doubly stochastic.

- We aim to design a fully distributed algorithm to compute an NE only through local communications and computation.

- Can we achieve the best known deterministic rates?

Distributed schemes for aggregative games

Distributed VS-PGR

$x_{i,k}$: its equilibrium strategy, $v_{i,k}$: the average of the aggregate.

Distributed Variable Sample-size Projected Gradient-response Scheme

Initialize: Set $k = 0$, and $v_{i,0} = x_{i,0} \in X_i$ for any $i \in \mathcal{N}$.
Iterate until convergence

Consensus (average among neighbors). $\hat{v}_{i,k} := v_{i,k}$ and repeat τ_k times

$$\hat{v}_{i,k} := \sum_{j \in \mathcal{N}_i} a_{ij} \hat{v}_{j,k} \quad \forall i \in \mathcal{N} \qquad \text{or compact form } \widehat{V}_k = A^{\tau_k} V_k.$$

Strategy Update (move along the negative gradient of the payoff).

$$x_{i,k+1} := \Pi_{X_i} \left[x_{i,k} - \frac{\alpha}{S_k} \sum_{p=1}^{S_k} \nabla_{x_i} \psi_i \left(x_{i,k}, n\hat{v}_{i,k}; \xi_k^p \right) \right]$$

reduce the noise variance by increasing S_k

$$v_{i,k+1} := v_{i,k} + x_{i,k+1} - x_{i,k}.$$

Analysis Sketch

- Consensus error: based on $\left| \left[A^k \right]_{ij} - \frac{1}{n} \right| \leq \theta \beta^k$ for a constant $\theta > 0$ and $\beta \in (0, 1)$, by defining $y_k \triangleq \sum_{i=1}^n v_{i,k}/n$ and $D_X \triangleq \sum_{j=1}^n \max_{x_j \in X_j} \|x_j\|$,

$$\|y_k - \hat{v}_{i,k}\| \leq \theta D_X \beta^{\sum_{p=0}^k \tau_p} + 2\theta D_X \sum_{s=1}^k \beta^{\sum_{p=s}^k \tau_p} \quad \forall k \geq 0.$$

- Suppose $\phi(x) \triangleq \left(\nabla_{x_i} f_i(x_i, \sum_{i=1}^n x_i) \right)_{i=1}^n$ is η_ϕ-strongly monotone and L_ϕ-Lipschitz continuous. Recursion on the conditional mean-squared error:

$$\mathbb{E}[\|x_{k+1} - x^*\|^2 | \mathcal{F}_k] \leq \underbrace{\left(1 - 2\alpha\eta_\phi + 2\alpha^2 L_\phi^2 \right)}_{\text{contraction property}} \|x_k - x^*\|^2 + \alpha^2 \underbrace{\sum_{i=1}^n \nu_i^2 / S_k}_{\text{noise}}$$

$$\underbrace{+ 4\alpha n D_X \sum_{i=1}^n L_i \|\hat{v}_{i,k} - y_k\| + 2\alpha^2 n^2 \sum_{i=1}^n L_i^2 \|\hat{v}_{i,k} - y_k\|^2}_{\text{consensus error}}.$$

Convergence Results–Geometric

Theorem 1: Linear rate of convergence

Set $\tau_k = k + 1$, $S_k = \left\lceil \rho^{-(k+1)} \right\rceil$ for some $\rho \in (0, 1)$. Suppose $\alpha \in \left(0, \eta_\phi / L_\phi^2\right)$, define $\varrho_\phi \triangleq 1 - 2\alpha\eta_\phi + 2\alpha^2 L_\phi^2$ and $\gamma \triangleq \max\{\rho, \beta\}$. Then
$$\mathbb{E}[\|x_k - x^*\|^2] = \mathcal{O}(\max\{\varrho_\phi, \gamma\}^k).$$

Theorem 2: Complexity Bounds

Set $\tau_k = k + 1$, $\alpha = \frac{\eta_\phi}{2L_\phi^2}$ and $S_k = \left\lceil \rho^{-(k+1)} \right\rceil$ with $\rho \triangleq \max\left\{1 - \frac{\eta_\phi^2}{2L_\phi^2}, \beta\right\}$. For obtaining ϵ−NE such that $\mathbb{E}[\|x_K - x^*\|^2] \leq \epsilon$, the iteration complexity $K = \mathcal{O}(\ln(1/\epsilon))$ (optimal, deterministic), communication complexity $\sum_{k=0}^{K} \tau_k = \mathcal{O}(\ln^2(1/\epsilon))$, and the oracle complexity is $\sum_{k=0}^{K} S_k = \mathcal{O}(1/\epsilon)$ (optimal, SGD).

less projections and communications than SGD $\mathcal{O}(1/\epsilon)$
best known comm. comp. in dis. opt. is $K \ln(K)$[Jakovetic, Xavier, and Moura, 14]

Convergence Results–Polynomial

- We need not increase the samples too fast when the oracle is costly.
- Explore the performance with slower rates of growth of sample-size?

Proposition 1: Polynomial rate of convergence

Set $\tau_k = \lceil (k+1)^u \rceil$ and $S_k = \lceil (k+1)^v \rceil$ for some $u \in (0,1)$ and $v > 0$. Let $\alpha \in (0, \eta_\phi/L_\phi^2)$ and define $\varrho_\phi \triangleq 1 - 2\alpha\eta_\phi + 2\alpha^2 L_\phi^2$. Then we obtain a polynomial rate of convergence $\mathbb{E}[\|x_{k+1} - x^*\|^2] = \mathcal{O}\left((k+1)^{-v}\right)$,

Proposition 2: Complexity Bounds

Set $\tau_k = \lceil (k+1)^u \rceil$ and $S_k = \lceil (k+1)^v \rceil$ for some $u \in (0,1)$ and $v > 0$. Then the iteration, communication, and oracle complexity to obtain an $\epsilon-$NE are bounded by $\mathcal{O}((1/\epsilon)^{1/v})$, $\mathcal{O}((1/\epsilon)^{(u+1)/v})$, and $\mathcal{O}\left((1/\epsilon)^{1+1/v}\right)$, respectively.

Distributed schemes for aggregative games

An Example: Nash Cournot Competition

- Firms compete on the amount of output they will produce and sell in the markets, where the aggregate is the sum of production of all firms.
- Consider a *stochastic* environment in which n firms competing over L markets, where firm i's production quantity is $x_i = (x_i^1, \cdots, x_i^L) \in \mathbb{R}^L$.
- There exists a random linear production cost of firm i: function $c_i(x_i; \xi_i) = (c_i + \xi_i) \sum_{l=1}^{L} x_i^l$ for $c_i > 0$ and random disturbance ξ_i with mean zero.
- The price of products sold in market $l \in \mathcal{L}$ is determined by a random linear inverse demand (or price) function $p_l(\bar{x}_l; \zeta_l) = a_l + \zeta_l - b_l \bar{x}_l$, where the aggregate $\bar{x}_l = \sum_{i=1}^{n} x_i^l$, $a_l > 0$, $b_l > 0$, and ζ_l is zero-mean.
- Firm i has a payoff: $F_i(x) = \mathbb{E}\left[c_i(x_i; \xi_i) - \sum_{l=1}^{L} p_l(\bar{x}_l; \zeta_l) x_i^l\right]$.

Introduction
○○○○○○○○○○○○○○○○

BR for SNEPs
○○○○○○○○○○○○○○○○○○○○○○○

Distributed BR and GR schemes
○○○○○○○○○○●○○○○○○○○

Distributed schemes for aggregative games

Numerical Validation: Distributed vs Centralized

- Implement the distributed algorithm over Erdős–Rényi graph and the centralized algorithm, where $\alpha = 0.01, \tau_k = \lceil \log(k) \rceil, S_k = \left\lceil 0.98^{-(k+1)} \right\rceil$.
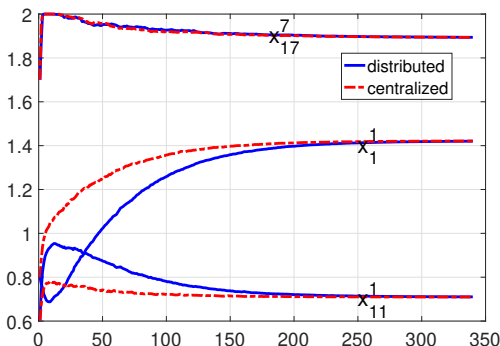
The empirical error $\frac{\mathbb{E}[\|x_k - x^*\|]}{\|x^*\|}$ by averaging across 50 sample paths.

Distributed schemes for aggregative games

Numerical Validation: Network Connectivity

Run the algorithm with $\tau_k = k + 1$, $\alpha = 0.01$, and $S_k = \left\lceil \beta^{-(k+1)} \right\rceil$ over the cycle, **star**, and Erdős–Rényi graphs, where β are 0.967, **0.95**, 0.986 respectively.
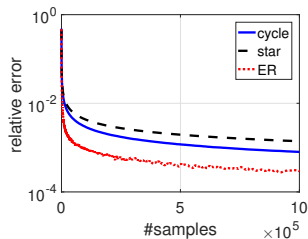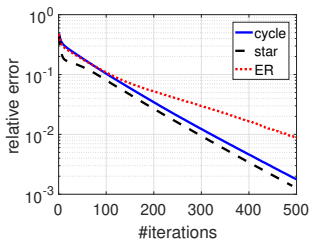
- The **star** graph has the **fastest convergence rate**, which is consistent with Theorem 1 that smaller β may lead to faster rate of convergence.
- The ER graph has the best oracle complexity, which reinforces the theoretical findings that larger β may lead to better oracle complexity.

Introduction
○○○○○○○○○○○○○○

BR for SNEPs
○○○○○○○○○○○○○○○○○○○○

Distributed BR and GR schemes
○○○○○○○○○○○○○●○○○○○○

Distributed schemes for aggregative games

Numerical Validation: Geometric vs Polynomial

- Set $n = 20, L = 13$ and run the algorithm over the complete graph with geometric and polynomial increasing sample-sizes.
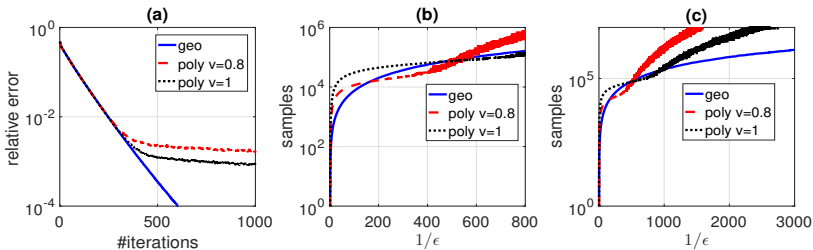
- with low accuracy ϵ the poly with smaller degree v appears to have better oracle complexity, while for a high accuracy ϵ, the geo and poly (with larger v) may have better oracle complexity.

Over the complete graph, the iteration and oracle complexity to obtain an $\epsilon-$NE are $\mathcal{O}(v(1/\epsilon)^{1/v})$ and $\mathcal{O}\left(e^v v^v (1/\epsilon)^{1+1/v}\right)$, respectively.

Introduction
BR for SNEPs
Distributed BR and GR schemes

◦◦◦◦◦◦◦◦◦◦◦◦◦◦◦◦◦
◦◦◦◦◦◦◦◦◦◦◦◦◦◦◦◦◦◦◦◦◦
◦◦◦◦◦◦◦◦◦◦◦◦◦●◦◦◦◦◦

Distributed schemes for aggregative games

Distributed VS-PBR

Distributed Variable Sample-size Proximal Best-response Scheme

Initialize: Set $k = 0$, and $v_{i,0} = x_{i,0} \in X_i$ for any $i \in \mathcal{N}$.

Iterate until convergence

Consensus. $\hat{v}_{i,k} := v_{i,k} \; \forall i \in \mathcal{N}$ and repeat τ_k times

$$\hat{v}_{i,k} := \sum_{j \in \mathcal{N}_i} a_{ij} \hat{v}_{j,k} \quad \forall i \in \mathcal{N}.$$

Strategy Update (sample average objective), for any $i \in \mathcal{N}$

$$x_{i,k+1} = \underset{x_i \in X_i}{\operatorname{argmin}} \left[\frac{1}{S_k} \sum_{p=1}^{S_k} \psi_i(x_i, n\hat{v}_{i,k}; \xi_k^p) + \frac{\mu}{2} \|x_i - x_{i,k}\|^2 \right],$$

$$v_{i,k+1} := v_{i,k} + x_{i,k+1} - x_{i,k}.$$

Main Results

- Assumption: proximal BR map is contractive with parameter $a \in (0, 1)$.

$$T_i(y) \triangleq \mathrm{argmin}_{x_i \in X_i} \left[f_i(x_i, \bar{y}) + \frac{\mu}{2} \|x_i - y_i\|^2 \right] \quad \mu > 0.$$
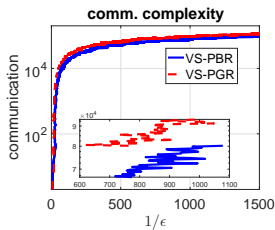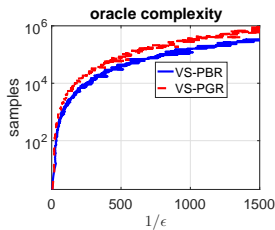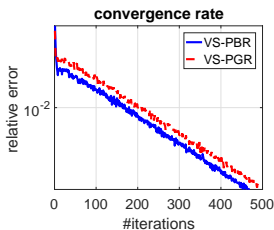
- Geometric Convergence. Set $\tau_k = k + 1$ and $S_k = \lceil \eta^{-2k} \rceil$ with $\eta \in (0, 1)$. Then $\mathbb{E}[\|x_k - x^*\|^2] = \mathcal{O}\left(\max\{a, \gamma\}^{2k}\right)$, where $\gamma \triangleq \max\{\eta, \beta\}$. The iteration, oracle, and communication complexity to compute an ϵ-NE are $\mathcal{O}(\ln(1/\epsilon))$, $\mathcal{O}\left(1/\epsilon\right)$, and $\mathcal{O}\left(\ln^2(1/\epsilon)\right)$, respectively.

- Often computing a sampled gradient is costly and geometric growth is impractical.

> Polynomial growth in sample-size represents a "dial".

- Polynomial Rate of Convergence. Set $\tau_k = \lceil (k + 1)^u \rceil$ and $S_k = \lceil (k + 1)^v \rceil$ for $u \in (0, 1)$ and $v > 0$. Then $\mathbb{E}[\|x_{k+1} - x^*\|^2] = \mathcal{O}\left((k + 1)^{-v}\right)$, the iteration, communication, and oracle complexity to obtain an $\epsilon-$NE are $\mathcal{O}((1/\epsilon)^{1/v})$, $\mathcal{O}((1/\epsilon)^{(u+1)/v})$, and $\mathcal{O}\left((1/\epsilon)^{1+1/v}\right)$, respectively.

Introduction
○○○○○○○○○○○○○○○

BR for SNEPs
○○○○○○○○○○○○○○○○○○○○○○

Distributed BR and GR schemes
○○○○○○○○○○○○○○○○○●○○○

Distributed schemes for aggregative games

Numerical Validation: Distributed VS-PGR and VS-PBR

Run both algorithms over a Erdős–Rényi graph with $\alpha = 0.04$, $\tau_k = k + 1$ and $S_k = \left\lceil 0.98^{-(k+1)} \right\rceil$, and $\mu = 30$.

Introduction
○○○○○○○○○○○○○○○

BR for SNEPs
○○○○○○○○○○○○○○○○○○○○

Distributed BR and GR schemes
○○○○○○○○○○○○○○○○○●○○

Distributed schemes for aggregative games

Summary of Contributions

$$\min_{x_i \in \mathbb{R}^{d_i}} F_i(x_i, x_{-i}) \triangleq \mathbb{E}\left[\psi_i(x; \xi)\right] + r_i(x_i).$$

Algorithm	S_k	Rate $\mathbb{E}[\|x_k - x^*\|^2]$	Iter. Comp.	Oracle Comp.	Ass.
VS-PGR	$\lceil \rho^{-(k+1)} \rceil$	Linear: $\mathcal{O}(\rho^k)$	$\mathcal{O}(\ln(1/\epsilon))$	$\mathcal{O}(1/\epsilon)$	SM
	$\lceil (k+1)^v \rceil$	$\mathcal{O}(q^k) + \mathcal{O}(k^{-v})$	$\mathcal{O}((1/\epsilon)^{1/v})$	$\mathcal{O}(1/\epsilon)^{(1+1/v)}$	SM
VS-PBR	$\lceil \rho^{-(k+1)} \rceil$	$\mathcal{O}(\rho^k)$	$\mathcal{O}(\ln(1/\epsilon))$	$\mathcal{O}(1/\epsilon)$	CPM
	$\lceil (k+1)^v \rceil$	$\mathcal{O}(a^k) + \mathcal{O}(k^{-v})$	$\mathcal{O}(1/\epsilon^{1/v})$	$\mathcal{O}(1/\epsilon^{1+1/v})$	CPM

SM: Strongly monotone, CPM: Contract. prox. BR Map

Introduction
ooooooooooooooooo

BR for SNEPs
ooooooooooooooooooooo

Distributed BR and GR schemes
ooooooooooooooooooo●o

Distributed schemes for aggregative games

Algorithm	S_k	Comm. τ_k	Rate $\mathbb{E}[\|x_k - x^*\|^2]$	Iter. Comp.	Oracle Comp.	Comm. Comp
d-VS-PGR	$\lceil \rho^{-(k+1)} \rceil$	$k+1$	Linear: $\mathcal{O}(\rho^k)$	$\mathcal{O}(\ln(1/\epsilon))$	$\mathcal{O}(1/\epsilon)$	$\mathcal{O}(\ln^2(1/\epsilon))$
	$\lceil (k+1)^v \rceil$	$\lceil (k+1)^u \rceil$	$\mathcal{O}((k+1)^{-v})$	$\mathcal{O}((1/\epsilon)^{1/v})$	$\mathcal{O}((1/\epsilon)^{1+1/v})$	$\mathcal{O}((1/\epsilon)^{\frac{1+u}{v}})$
d-VS-PBR	$\lceil \rho^{-(k+1)} \rceil$	$k+1$	Linear: $\mathcal{O}(\rho^k)$	$\mathcal{O}(\ln(1/\epsilon))$	$\mathcal{O}(1/\epsilon)$	$\mathcal{O}(\ln^2(1/\epsilon))$
	$\lceil (k+1)^v \rceil$	$\lceil (k+1)^u \rceil$	$\mathcal{O}((k+1)^{-v})$	$\mathcal{O}((1/\epsilon)^{1/v})$	$\mathcal{O}((1/\epsilon)^{1+1/v})$	$\mathcal{O}((1/\epsilon)^{\frac{1+u}{v}})$

(d-VS-PGR) and (d-VS-PBR) schemes for Aggregative games ($v > 0$, $u \in (0,1)$)

Introduction **BR for SNEPs** **Distributed BR and GR schemes**
○○○○○○○○○○○○○○○ ○○○○○○○○○○○○○○○○○○○○○ ○○○○○○○○○○○○○○○○○●

Distributed schemes for aggregative games

Summary and related work I

◆ Part I: Synch., asynch., and randomized BR schemes for stochastic Nash games

- UVS, Jong-Shi Pang, and Suvrajeet Sen, *Inexact best-response schemes for stochastic Nash games: Linear convergence and iteration complexity*, CDC 2016
- Jinlong Lei, UVS, Jong-Shi Pang, and Suvrajeet Sen, *On Synchronous, Asynchronous, and Randomized Best-Response Schemes for Stochastic Nash Games*, Mathematics of Operations Research (to appear, 2019).

◆ Part II: Distributed schemes for Stochastic Nash games over graphs

- Jinlong Lei and UVS, *Linearly Convergent Variable Sample-Size Schemes for Stochastic Nash Games: Best-Response Schemes and Distributed Gradient-Response Schemes*, CDC 2018: 3547-3552
- Jinlong Lei and UVS, *Distributed Variable Sample-Size Gradient-response and Best-response Schemes for Stochastic Nash Games over Graphs*, `arXiv:1811.11246`.

Introduction
○○○○○○○○○○○○○○○○○○

BR for SNEPs
○○○○○○○○○○○○○○○○○○○○○○○

Distributed BR and GR schemes
○○○○○○○○○○○○○○○○○○○●

Distributed schemes for aggregative games

📄 Nemirovski, A., Juditsky, A., Lan, G., and Shapiro, A. (2009).
Robust stochastic approximation approach to stochastic programming.
SIAM Journal on Optimization, 19(4):1574–1609.