Distributed Averaging and Optimization over Random Networks

Behrouz Touri

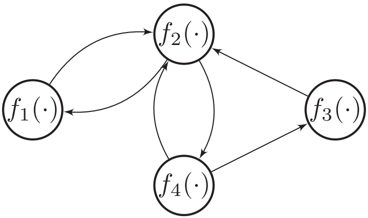
Department of Electrical and Computer Engineering
University of California, San Diego

Joint work with: **Adel Aghajan**

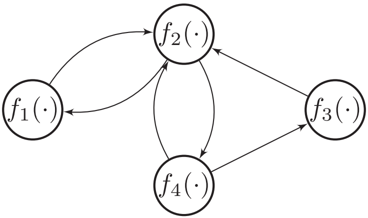
Isfahan University of Technology
December 21st 2020

Distributed Optimization

Problem Statement: Distributed Optimization



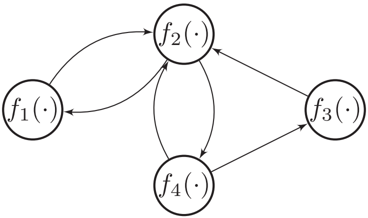
Problem Statement: Distributed Optimization



- $F(\mathbf{x}) \triangleq \frac{1}{n}\sum_{i=1}^{n} f_i(x) \triangleq \frac{1}{n}\mathbf{1}^T f(x)$ where $f(x) \triangleq \begin{bmatrix} f_1(x_1) & \cdots & f_n(x_n) \end{bmatrix}^T$

Objective: to reach and consent on some $x^* \in X := \arg\min_{x \in \mathbb{R}^m} F(x)$

Problem Statement: Distributed Optimization



- $F(\mathbf{x}) \triangleq \frac{1}{n}\sum_{i=1}^{n} f_i(x) \triangleq \frac{1}{n}\mathbf{1}^T f(x)$ where $f(x) \triangleq \begin{bmatrix} f_1(x_1) & \cdots & f_n(x_n) \end{bmatrix}^T$

  Objective: to reach and consent on some $x^* \in X := \arg\min_{x\in\mathbb{R}^m} F(x)$

- Assumptions on $F(x)$:
  - $\forall i \; f_i(x)$ is a convex function
  - $\forall i, x \;\; \|\nabla f_i(x)\| \leq L$
- For the sake of discussion: let $m = 1$

Objective

- Finding a distributed recursive algorithm (dynamics) for agents' estimate $x_i(t)$ of an optimal point:

$$\lim_{t\to\infty} x(t) = x^*\mathbf{1} \text{ and } x^* \in \arg\min_{x\in\mathbb{R}^m} F(x)$$

Objective

- Finding a distributed recursive algorithm (dynamics) for agents' estimate $x_i(t)$ of an optimal point:

$$\lim_{t\to\infty} x(t) = x^*\mathbf{1} \text{ and } x^* \in \arg\min_{x\in\mathbb{R}^m} F(x)$$

- Main and original idea: Gradient Descent + Ergodicity

Stochastic Matrices

- $W \in \mathbb{R}^{n \times n}$ is a (row)-stochastic matrix if
  1. W is non-negative,

Stochastic Matrices

- $W \in \mathbb{R}^{n\times n}$ is a (row)-stochastic matrix if
  1. W is non-negative, and
  2. each row sums up to one.

Stochastic Matrices

- $W \in \mathbb{R}^{n \times n}$ is a (row)-stochastic matrix if
  1. W is non-negative, and
  2. each row sums up to one.
- Example:

$$W = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

Stochastic Matrices

- $W \in \mathbb{R}^{n\times n}$ is a (row)-stochastic matrix if
  1. W is non-negative, and
  2. each row sums up to one.
- Example:

$$W = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix}.$$

- W is doubly stochastic if W, W^T are both stochastic.

An Algorithm

- Averaging-based distributed optimization solver:

$$x_i(t+1) = \sum_{j=1}^{n} w_{ij}(t+1)x_j(t) - \alpha(t)\nabla f_i(x_i(t)),$$

or more compactly[1]:

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t)),$$

where $\nabla f(\mathbf{x}) = [\nabla f_1(x_1), \ldots, \nabla f_n(x_n)]^T$.

[1] A. Nedić and A. Ozdaglar. " Distributed subgradient methods for multi-agent optimization", IEEE Transactions on Automatic Control, 2009.

Template for Results

Averaging-Based Distributed Optimization Solver

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t)).$$

Template for Results

Averaging-Based Distributed Optimization Solver

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t)).$$

Template for Results

If

- step-sizes $\{\alpha(t)\}$ satisfies some step-size condition,
- the local objective functions $f_i(x)$ satisfy the conditions, and
- $\{W(t)\}$ satisfies some connectivity conditions.

then $\lim_{t\to\infty} \mathbf{x}_i(t) = x^* \in X := \arg\min_{x\in\mathbb{R}^m} F(x)$.

Template for Results

Averaging-Based Distributed Optimization Solver

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t)).$$

Template for Results

Averaging-Based Distributed Optimization Solver

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t)).$$

Template for Results

If

- step-sizes $\{\alpha(t)\}$ satisfies $\alpha(t) \in [0,1]$ is decreasing, $\sum_{t=0}^{\infty} \alpha(t) = \infty$, and $\sum_{t=0}^{\infty} \alpha^2(t) < \infty$,
- the local objective functions $f_i(x)$ are convex with uniformly bounded (sub-)gradient, and
- $\{W(t)\}$ satisfies some connectivity conditions.

then $\lim_{t\to\infty} \mathbf{x}_i(t) = x^* \in X := \arg\min_{x\in\mathbb{R}^m} F(x)$.

B-Connectivity

- For a $n \times n$ matrix W, define $\mathcal{E}^\gamma(W)$ by

$$\mathcal{E}^\gamma(W) := \{(i,j) \mid w_{ij} \geq \gamma\}.$$

B-Connectivity

- For a $n \times n$ matrix W, define $\mathcal{E}^\gamma(W)$ by

$$\mathcal{E}^\gamma(W) := \{(i,j) \mid w_{ij} \geq \gamma\}.$$

- We say that $\{W(t)\}$ is (strongly) B-connected if
  - $W_{ii}(t) \geq \beta > 0$ for all $i \in [n]$ and $t \geq 0$,

B-Connectivity

- For a $n \times n$ matrix W, define $\mathcal{E}^\gamma(W)$ by

$$\mathcal{E}^\gamma(W) := \{(i,j) \mid w_{ij} \geq \gamma\}.$$

- We say that $\{W(t)\}$ is (strongly) B-connected if
  - $W_{ii}(t) \geq \beta > 0$ for all $i \in [n]$ and $t \geq 0$,
  - for every $t \geq 0$, the graph $([n], \bigcup_{k=tB}^{(t+1)B-1} \mathcal{E}^\gamma(W(k)))$ is strongly connected for some $\gamma > 0$.

Averaging-Based Distributed Optimization Results

Averaging-Based Distributed Optimization Solver

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t)).$$

Averaging-Based Distributed Optimization Results

Averaging-Based Distributed Optimization Solver

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t)).$$

Template

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ satisfies some connectivity conditions, then (for all $i \in [n]$ and initial conditions $x_i(0) \in \mathbb{R}^m$)

$$\lim_{t\to\infty} x_i(t) = x^* \in X.$$

Averaging-Based Distributed Optimization Results

Averaging-Based Distributed Optimization Solver

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t)).$$

Convergence in Deterministic Settings[2]

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is deterministic, doubly stochastic, and B-connected, then

$$\lim_{t\to\infty} x_i(t) = x^* \in X.$$

[2] A. Nedić, A. Ozdaglar, and P. Parrilo. "Constrained consensus and optimization in multi-agent networks." IEEE Transactions on Automatic Control 55.4 (2010): 922-938.

Averaging-Based Distributed Optimization Results

Averaging-Based Distributed Optimization Solver

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t)).$$

- Since $W(t)$ is row-stochastic:

$$\frac{1}{n}\mathbf{1}^T\mathbf{x}(t+1) = \frac{1}{n}\mathbf{1}^T\left[W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t))\right]$$

Averaging-Based Distributed Optimization Results

Averaging-Based Distributed Optimization Solver

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t)).$$

- Since $W(t)$ is row-stochastic:

$$\begin{aligned}\frac{1}{n}\mathbf{1}^T\mathbf{x}(t+1) &= \frac{1}{n}\mathbf{1}^T\left[W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t))\right] \\ &= \frac{1}{n}\mathbf{1}^T\mathbf{x}(t) - \alpha(t)\frac{1}{n}\mathbf{1}^T\nabla f(\mathbf{x}(t))\end{aligned}$$

Averaging-Based Distributed Optimization Results

Averaging-Based Distributed Optimization Solver

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t)).$$

- Since $W(t)$ is row-stochastic:

$$\frac{1}{n}\mathbf{1}^T\mathbf{x}(t+1) = \frac{1}{n}\mathbf{1}^T\left[W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t))\right]$$

$$= \frac{1}{n}\mathbf{1}^T\mathbf{x}(t) - \alpha(t)\frac{1}{n}\mathbf{1}^T\nabla f(\mathbf{x}(t))$$

- Well-mixing (*ergodicity*) and diminishing step-size implies: for $t \gg 0$,

$$\bar{\mathbf{x}}(t) \triangleq \frac{1}{n}\mathbf{1}^T\mathbf{x}(t) \approx x_1(t) \approx \cdots \approx x_n(t)$$

Averaging-Based Distributed Optimization Results

Averaging-Based Distributed Optimization Solver

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t)).$$

- Since $W(t)$ is row-stochastic:

$$\frac{1}{n}\mathbf{1}^T\mathbf{x}(t+1) = \frac{1}{n}\mathbf{1}^T\left[W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t))\right]$$

$$= \frac{1}{n}\mathbf{1}^T\mathbf{x}(t) - \alpha(t)\frac{1}{n}\mathbf{1}^T\nabla f(\mathbf{x}(t))$$

- Well-mixing (*ergodicity*) and diminishing step-size implies: for $t \gg 0$,

$$\bar{\mathbf{x}}(t) \triangleq \frac{1}{n}\mathbf{1}^T\mathbf{x}(t) \approx x_1(t) \approx \cdots \approx x_n(t)$$

- $\mathbf{1}^T\nabla f(\mathbf{x}(t)) \approx \nabla F(\bar{\mathbf{x}}(t))$ and $\bar{\mathbf{x}}(t+1) \approx \bar{\mathbf{x}}(t) - \bar{\alpha}(t)\nabla F(\bar{\mathbf{x}}(t))$

Averaging-Based Distributed Optimization Results

Convergence in Deterministic Settings[3]

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is deterministic and satisfies the B-connectivity condition, then

$$\lim_{t\to\infty} x_i(t) = x^* \in X.$$

Averaging-Based Distributed Optimization Results

Convergence in Deterministic Settings[3]

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is deterministic and satisfies the B-connectivity condition, then

$$\lim_{t\to\infty} x_i(t) = x^* \in X.$$

- Challenges: Link failure? $(w_{ij}(t) \to b_{ij}(t)w_{ij}(t)$ for a random binary matrix $B(t)$ and $i \neq j)$

Averaging-Based Distributed Optimization Results

Convergence in Deterministic Settings[3]

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is deterministic and satisfies the B-connectivity condition, then

$$\lim_{t\to\infty} x_i(t) = x^* \in X.$$

- Challenges: Link failure? $(w_{ij}(t) \to b_{ij}(t)w_{ij}(t)$ for a random binary matrix $B(t)$ and $i \neq j)$
  Doubly stochastic?

Previous Approaches: Random Connectivity 1

Convergence in Random Settings[4]

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is ~~deterministic~~ an independent random sequence that is doubly stochastic **almost surely**, and $\{\mathbb{E}[W(t)]\}$ is B-connected, then

$$\lim_{t \to \infty} x_i(t) = x^* \in X.$$

[4] I. Lobel and A. Ozdaglar. "Distributed subgradient methods for convex optimization over random network," IEEE Transactions on Automatic Control, 2010.

Previous Approaches: Random Connectivity 1

Convergence in Random Settings[4]

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is ~~deterministic~~ an independent random sequence that is doubly stochastic **almost surely**, and $\{\mathbb{E}[W(t)]\}$ is B-connected, then

$$\lim_{t\to\infty} x_i(t) = x^* \in X.$$

- Challenges: Link failure?

[4] I. Lobel and A. Ozdaglar. "Distributed subgradient methods for convex optimization over random network," IEEE Transactions on Automatic Control, 2010.

Previous Approaches: Random Connectivity 1

Convergence in Random Settings[4]

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is ~~deterministic~~ an independent random sequence that is doubly stochastic **almost surely**, and $\{\mathbb{E}[W(t)]\}$ is B-connected, then

$$\lim_{t \to \infty} x_i(t) = x^* \in X.$$

- Challenges: Link failure?
  Doubly stochastic?

[4] I. Lobel and A. Ozdaglar. "Distributed subgradient methods for convex optimization over random network," IEEE Transactions on Automatic Control, 2010.

Previous Approaches: Random Connectivity 2

Convergence in Random Settings[5]

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is ~~deterministic~~ independently and identically distributed random sequence, ~~doubly stochastic **almost surely**~~ row-stochastic almost surely, column-stochastic in-expectation, and ~~B-connected in-expectation~~ $\mathbb{E}[W(0)]$ is strongly connected, then

$$\lim_{t\to\infty} x_i(t) = x^* \in X := \arg\min_{x\in\mathbb{R}^m} F(x).$$

[5]K. Srivastava, A. Nedić. "Distributed asynchronous constrained stochastic optimization." IEEE Journal of Selected Topics in Signal Processing 5.4 (2011): 772-790. G. Morral, P. Bianchi, and G. Fort. "Success and failure of adaptation-diffusion algorithms with decaying step size in multiagent networks." IEEE Transactions on Signal Processing 65.11 (2017): 2798-2813.I.

Previous Approaches: Random Connectivity 2

Convergence in Random Settings[5]

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is ~~deterministic~~ independently and identically distributed random sequence, ~~doubly stochastic **almost surely**~~ row-stochastic almost surely, column-stochastic in-expectation, and ~~B-connected in-expectation~~ $\mathbb{E}[W(0)]$ is strongly connected, then

$$\lim_{t\to\infty} x_i(t) = x^* \in X := \arg\min_{x\in\mathbb{R}^m} F(x).$$

- Challenges: Link failure?
  ~~Doubly stochastic?~~

[5] K. Srivastava, A. Nedić. "Distributed asynchronous constrained stochastic optimization." IEEE Journal of Selected Topics in Signal Processing 5.4 (2011): 772-790. G. Morral, P. Bianchi, and G. Fort. "Success and failure of adaptation-diffusion algorithms with decaying step size in multiagent networks." IEEE Transactions on Signal Processing 65.11 (2017): 2798-2813.I.

Our Result

Convergence in Random Settings[6]

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is

1. a dependent random sequence,
2. row-stochastic and further $\mathbb{E}[W(t) \mid \mathcal{F}(t-1)]$ is column-stochastic almost surely, and
3. B-connected in conditional expectation, i.e., the random graph $([n], \mathcal{E}(tB))$ with

$$\mathcal{E}(tB) := \bigcup_{\tau=tB+1}^{(t+1)B} \mathcal{E}^{\gamma}(\mathbb{E}[W(\tau)|\mathcal{F}(tB)])$$

is strongly connected almost surely for all $t \geq 0$,

then $\lim_{t\to\infty} x_i(t) = x^*$ almost surely for a random vector supported on X.

[6] A. Aghajan and B. Touri. “Distributed optimization over dependent random networks.” available on Arxiv.

Our Result

Convergence in Random Settings[6]

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is

1. a dependent random sequence,
2. row-stochastic and further $\mathbb{E}[W(t) \mid \mathcal{F}(t-1)]$ is column-stochastic almost surely, and
3. B-connected in conditional expectation, i.e., the random graph $([n], \mathcal{E}(tB))$ with

$$\mathcal{E}(tB) := \bigcup_{\tau=tB+1}^{(t+1)B} \mathcal{E}^{\gamma}(\mathbb{E}[W(\tau)|\mathcal{F}(tB)])$$

   is strongly connected almost surely for all $t \geq 0$,

then $\lim_{t\to\infty} x_i(t) = x^*$ almost surely for a random vector supported on X.

[6] A. Aghajan and B. Touri. "Distributed optimization over dependent random networks." available on Arxiv.

Implication 1: Independent Setting

Convergence in Independent Settings[7]

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is an independent random sequence, row-stochastic almost surely and $\mathbb{E}[W(t)]$ is doubly-stochastic, and $\{\mathbb{E}[W(t)]\}$ is B-connected, then $\lim_{t\to\infty} x_i(t) = x^*$ almost surely for a random vector supported on X.

[7] A. Aghajan and B. Touri. "Distributed optimization over independent random networks." **somewhere on Overleaf**.

Implication 1: Independent Setting

Convergence in Independent Settings[7]

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is an independent random sequence, row-stochastic almost surely and $\mathbb{E}[W(t)]$ is doubly-stochastic, and $\{\mathbb{E}[W(t)]\}$ is B-connected, then $\lim_{t\to\infty} x_i(t) = x^*$ almost surely for a random vector supported on X.

- This implies the previously discussed results.

[7] A. Aghajan and B. Touri. "Distributed optimization over independent random networks." **somewhere on Overleaf**.

Implication 1: Independent Setting

Convergence in Independent Settings[7]

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is an independent random sequence, row-stochastic almost surely and $\mathbb{E}[W(t)]$ is doubly-stochastic, and $\{\mathbb{E}[W(t)]\}$ is B-connected, then $\lim_{t\to\infty} x_i(t) = x^*$ almost surely for a random vector supported on X.

- This implies the previously discussed results.
- Robustness of averaging-based solvers to link-failure.

[7] A. Aghajan and B. Touri. “Distributed optimization over independent random networks.” **somewhere on Overleaf**.

Implication 2: Synthesizing other distributed schemes

- Let $\mathcal{G}$ be an undirected connected graph.

Implication 2: Synthesizing other distributed schemes

- Let $\mathcal{G}$ be an undirected connected graph.
- Consider a token at the possession of agent $\ell(t)$ at time $t \geq 0$.

Implication 2: Synthesizing other distributed schemes

- Let $\mathcal{G}$ be an undirected connected graph.
- Consider a token at the possession of agent $\ell(t)$ at time $t \geq 0$.
- $\ell(t)$ asks or shares information (each w.p. 0.5) with a random neighbor $s(t)$.

Implication 2: Synthesizing other distributed schemes

- Let $\mathcal{G}$ be an undirected connected graph.
- Consider a token at the possession of agent $\ell(t)$ at time $t \geq 0$.
- $\ell(t)$ asks or shares information (each w.p. 0.5) with a random neighbor $s(t)$.
- If asking, $\ell(t)$ keeps the token ($\ell(t+1) = \ell(t)$) and if sharing, $\ell(t)$ passes the token to $s(t)$ ($\ell(t+1) = s(t)$).

Implication 2: Synthesizing other distributed schemes

- Let $\mathcal{G}$ be an undirected connected graph.
- Consider a token at the possession of agent $\ell(t)$ at time $t \geq 0$.
- $\ell(t)$ asks or shares information (each w.p. 0.5) with a random neighbor $s(t)$.
- If asking, $\ell(t)$ keeps the token ($\ell(t+1) = \ell(t)$) and if sharing, $\ell(t)$ passes the token to $s(t)$ ($\ell(t+1) = s(t)$).
- The agent with the token sets

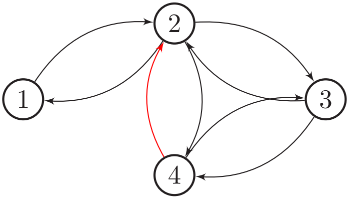
$$\mathbf{x}_{\ell(t+1)}(t+1) = \frac{1}{2}(\mathbf{x}_{s(t)}(t) + \mathbf{x}_{\ell(t)}(t)) - \alpha(t)\nabla f_{\ell(t+1)}(\mathbf{x}_{\ell(t+1)}(t)).$$

Implication 2: Synthesizing other distributed schemes

- Let $\mathcal{G}$ be an undirected connected graph.
- Consider a token at the possession of agent $\ell(t)$ at time $t \geq 0$.
- $\ell(t)$ asks or shares information (each w.p. 0.5) with a random neighbor $s(t)$.
- If asking, $\ell(t)$ keeps the token ($\ell(t+1) = \ell(t)$) and if sharing, $\ell(t)$ passes the token to $s(t)$ ($\ell(t+1) = s(t)$).
- The agent with the token sets

$$\mathbf{x}_{\ell(t+1)}(t+1) = \frac{1}{2}(\mathbf{x}_{s(t)}(t) + \mathbf{x}_{\ell(t)}(t)) - \alpha(t)\nabla f_{\ell(t+1)}(\mathbf{x}_{\ell(t+1)}(t)).$$

- The rest do simple gradient update $\mathbf{x}_i(t+1) = \mathbf{x}_i(t) - \alpha(t)\nabla f_i(\mathbf{x}_i(t))$.



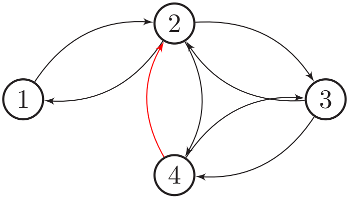
$$W(t) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Implication 2: Synthesizing other distributed schemes

- Let $\mathcal{G}$ be an undirected connected graph.
- Consider a token at the possession of agent $\ell(t)$ at time $t \geq 0$.
- $\ell(t)$ asks or shares information (each w.p. 0.5) with a random neighbor $s(t)$.
- If asking, $\ell(t)$ keeps the token ($\ell(t+1) = \ell(t)$) and if sharing, $\ell(t)$ passes the token to $s(t)$ ($\ell(t+1) = s(t)$).
- The agent with the token sets

$$\mathbf{x}_{\ell(t+1)}(t+1) = \frac{1}{2}(\mathbf{x}_{s(t)}(t) + \mathbf{x}_{\ell(t)}(t)) - \alpha(t)\nabla f_{\ell(t+1)}(\mathbf{x}_{\ell(t+1)}(t)).$$

- The rest do simple gradient update $\mathbf{x}_i(t+1) = \mathbf{x}_i(t) - \alpha(t)\nabla f_i(\mathbf{x}_i(t))$.



$$W(t) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- $W(t)$ is row-stochastic, and $\mathbb{E}[W(t) \mid \mathcal{F}(t-1)]$ is column-stochastic.
- For $B = \mathrm{d}(\mathcal{G})$, we can show that $\mathbb{E}[W(t+B) \mid \mathcal{F}(t)] \geq \gamma \mathbf{1}\mathbf{1}^T$ for some

Our Result: Sketch of the Proof

Convergence in Random Settings

If the step-size condition holds, the local objective function conditions hold, and $\{W(t)\}$ is

1. a dependent random sequence,
2. row-stochastic almost surely and $\mathbb{E}[W(t) \mid \mathcal{F}(t-1)]$ is doubly-stochastic almost surely, and
3. B-connected in conditional expectation, i.e., the random graph $([n], \mathcal{E}(tB))$ with

$$\mathcal{E}(tB) := \bigcup_{\tau=tB+1}^{(t+1)B} \mathcal{E}^{\gamma}(\mathbb{E}[W(\tau)|\mathcal{F}(tB)])$$

   is strongly connected almost surely for all $t \geq 0$.

Then $\lim_{t\to\infty} \mathbf{x}_i(t) = x^*$ almost surely for a random vector supported on X.

Sketch of the Proof: Convergence of Controlled Averaging Dynamics

- Show that $\lim_{t\to\infty} \|\bar{\mathbf{x}}(t) - x^*\| = 0$ a.s. for a random $x^* \in X$

Sketch of the Proof: Convergence of Controlled Averaging Dynamics

- Show that $\lim_{t\to\infty} \|\bar{\mathbf{x}}(t) - x^*\| = 0$ a.s. for a random $x^* \in X$
- Show that $\lim_{t\to\infty} \|\mathbf{x}_i(t) - x^*\| = 0$ a.s. for all $i \in [n]$

Sketch of the Proof: Convergence of Controlled Averaging Dynamics

- Show that $\lim_{t\to\infty} \|\bar{\mathbf{x}}(t) - x^*\| = 0$ a.s. for a random $x^* \in X$
- Show that $\lim_{t\to\infty} \|\mathbf{x}_i(t) - x^*\| = 0$ a.s. for all $i \in [n]$through the study of

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t))$$

Sketch of the Proof: Convergence of Controlled Averaging Dynamics

- Show that $\lim_{t\to\infty} \|\bar{\mathbf{x}}(t) - x^*\| = 0$ a.s. for a random $x^* \in X$
- Show that $\lim_{t\to\infty} \|\mathbf{x}_i(t) - x^*\| = 0$ a.s. for all $i \in [n]$through the study of

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t))$$

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) + u(t),$$

where $\|u(t)\| \leq \alpha(t)L$.

Sketch of the Proof: Convergence of Controlled Averaging Dynamics

- Show that $\lim_{t\to\infty} \|\bar{\mathbf{x}}(t) - x^*\| = 0$ a.s. for a random $x^* \in X$
- Show that $\lim_{t\to\infty} \|\mathbf{x}_i(t) - x^*\| = 0$ a.s. for all $i \in [n]$through the study of

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) - \alpha(t)\nabla f(\mathbf{x}(t))$$

$$\mathbf{x}(t+1) = W(t+1)\mathbf{x}(t) + u(t),$$

  where $\|u(t)\| \leq \alpha(t)L$.
- Requires showing $\lim_{t\to\infty} \mathrm{d}(\mathbf{x}(t)) = 0$ almost surely for the controlled process.

An Intermediate Result

Lemma

Suppose that $\{D(t)\}$ *is a non-negative random (scalar) process such that*

$$D(t+1) \leq a(t+1)D(t) + b(t), \quad \textit{almost surely}$$

where $\{b(t)\}$ *is a deterministic sequence and* $\{a(t)\}$ *is an adapted process (to* $\{\mathcal{F}(t)\}$*), such that* $a(t) \in [0,1]$ *and*

$$\mathbb{E}[a(t+1) \mid \mathcal{F}(t)] \leq \tilde{\lambda} < 1,$$

almost surely for all $t \geq 0$*. Then, if*

$$0 \leq b(t) \leq Kt^{-\tilde{\beta}}$$

for some $K, \tilde{\beta} > 0$*, we have* $\lim_{t\to\infty} D(t)t^{\beta} = 0$*, almost surely, for all* $\beta < \tilde{\beta}$*.*

Future Directions

- Extension to non-convex

Future Directions

- Extension to non-convex
- Extension to the social-learning

Future Directions

- Extension to non-convex
- Extension to the social-learning
- Questions?