Multi-agent constrained optimization of a strongly convex function

Necdet Serhat Aybat
Industrial & Manufacturing Engineering
Penn State University

Joint work with Erfan Yazdandoost Hamedani

Research supported by NSF CMMI-1635106 and ARO W911NF-17-1-0298

PennState

Motivation

Many real-life networks are

- large-scale
- composed of agents with local information
- agents willing to collaborate without sharing their private data

This motivated huge interest in designing decentralized methods for optimization of multi-agent systems

Motivation

Many real-life networks are

- large-scale
- composed of agents with local information
- agents willing to collaborate without sharing their private data

This motivated huge interest in designing decentralized methods for optimization of multi-agent systems

Examples:

- Routing and congestion control in wired and wireless networks
- parameter estimation in sensor networks
- multi-agent cooperative control and coordination
- processing distributed big-data in (online) machine learning

Decentralized Consensus Optimization

Compute an optimal solution for

$$(P): \min_{x} \bar{\varphi}(x) \triangleq \sum_{i \in \mathcal{N}} \varphi_i(x) \text{ s.t. } x \in \bigcap_{i \in \mathcal{N}} \mathcal{X}_i$$

Decentralized Consensus Optimization

Compute an optimal solution for

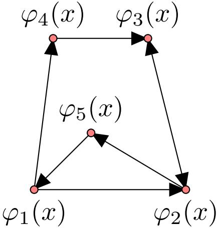
$$(P): \min_{x} \ \bar{\varphi}(x) \triangleq \sum_{i \in \mathcal{N}} \varphi_i(x) \ \text{s.t.} \ x \in \bigcap_{i \in \mathcal{N}} \mathcal{X}_i$$

- $\mathcal{N} = \{1, \ldots, N\}$ processing nodes

Decentralized Consensus Optimization

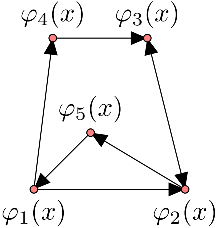
Compute an optimal solution for

$$(P): \min_{x} \ \bar{\varphi}(x) \triangleq \sum_{i \in \mathcal{N}} \varphi_i(x) \ \text{s.t.} \ x \in \bigcap_{i \in \mathcal{N}} \mathcal{X}_i$$



- $\mathcal{N} = \{1, \ldots, N\}$ processing nodes on a time-varying $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$

Decentralized Consensus Optimization

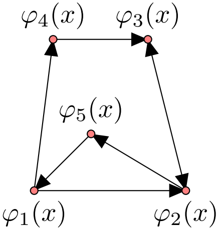


Compute an optimal solution for

$$(P): \min_x \ \bar{\varphi}(x) \triangleq \sum_{i\in\mathcal{N}} \varphi_i(x) \ \text{s.t.} \ x \in \bigcap_{i\in\mathcal{N}} \mathcal{X}_i$$

- $\mathcal{N} = \{1, \ldots, N\}$ processing nodes on a time-varying $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$
  - Node i can transmit data to j at time t only if $(i, j) \in \mathcal{E}^t$

Decentralized Consensus Optimization

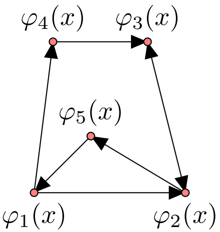


Compute an optimal solution for

$$(P): \min_{x} \ \bar{\varphi}(x) \triangleq \sum_{i\in\mathcal{N}} \varphi_i(x) \ \text{s.t.} \ x \in \bigcap_{i\in\mathcal{N}} \mathcal{X}_i$$

- $\mathcal{N} = \{1,\ldots,N\}$ processing nodes on a time-varying $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$
  - Node i can transmit data to j at time t only if $(i,j) \in \mathcal{E}^t$
- $\bar{\varphi}(x)$: strongly convex ($\bar{\mu} > 0$)
- $\varphi_i(x) \triangleq \rho_i(x) + f_i(x)$ locally known ($\underline{\mu} \triangleq \min_{i\in\mathcal{N}} \mu_i \geq 0$)
- $\mathcal{X}_i \triangleq \{x : G_i(x) \in -\mathcal{K}_i\}$ locally known, $\mathcal{K}_i$ closed convex cone.

Decentralized Consensus Optimization

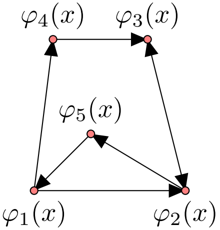


Compute an optimal solution for

$$(P): \min_x \ \bar{\varphi}(x) \triangleq \sum_{i\in\mathcal{N}} \varphi_i(x) \ \text{s.t.} \ x \in \bigcap_{i\in\mathcal{N}} \mathcal{X}_i$$

- $\mathcal{N} = \{1,\ldots,N\}$ processing nodes on a time-varying $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$
  - Node i can transmit data to j at time t only if $(i,j) \in \mathcal{E}^t$
- $\bar{\varphi}(x)$: strongly convex ($\bar{\mu} > 0$)
- $\varphi_i(x) \triangleq \rho_i(x) + f_i(x)$ locally known ($\underline{\mu} \triangleq \min_{i\in\mathcal{N}} \mu_i \geq 0$)
  - f_i: convex + Lipschitz continuous gradient (constant L_i)
  - ρ_i: convex (possibly non-smooth) + efficient prox-map
  - $\mathbf{prox}_{\rho_i}(x) \triangleq \text{argmin}_{\xi\in\mathbb{R}^n} \left\{\rho_i(\xi) + \frac{1}{2}\|\xi - x\|_2^2\right\}$
- $\mathcal{X}_i \triangleq \{x : G_i(x) \in -\mathcal{K}_i\}$ locally known, $\mathcal{K}_i$ closed convex cone.

Decentralized Consensus Optimization



Compute an optimal solution for

$$\color{red}{(P)}: \min_x \bar{\varphi}(x) \triangleq \sum_{i\in\mathcal{N}} \varphi_i(x) \text{ s.t. } x \in \bigcap_{i\in\mathcal{N}} \mathcal{X}_i$$

- $\mathcal{N} = \{1,\ldots,N\}$ processing nodes on a time-varying $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$
  - Node i can transmit data to j at time t only if $(i,j) \in \mathcal{E}^t$
- $\bar{\varphi}(x)$: strongly convex ($\bar{\mu} > 0$)
- $\varphi_i(x) \triangleq \rho_i(x) + f_i(x)$ locally known ($\underline{\mu} \triangleq \min_{i\in\mathcal{N}} \mu_i \geq 0$)
  - f_i: convex + Lipschitz continuous gradient (constant L_i)
  - ρ_i: convex (possibly non-smooth) + efficient prox-map
  - $\mathbf{prox}_{\rho_i}(x) \triangleq \operatorname{argmin}_{\xi\in\mathbb{R}^n} \left\{\rho_i(\xi) + \frac{1}{2}\|\xi - x\|_2^2\right\}$
- $\mathcal{X}_i \triangleq \{x : G_i(x) \in -\mathcal{K}_i\}$ locally known, $\mathcal{K}_i$ closed convex cone.
  - G_i: $\mathcal{K}_i$-convex + Lip. cont. (C_G) + Lip. cont. Jacobian ∇G_i (L_G)

Examples

- **Constrained Lasso**

$$\min_{x\in\mathbb{R}^n}\{\lambda\|x\|_1+\|Cx-d\|_2^2:\ Ax\le b\},\quad \mathcal{K}=-\mathbb{R}^p_+$$

  - distributed data: $C_i\in\mathbb{R}^{m_i\times n}$ and $d_i\in\mathbb{R}^{m_i}$ for $i\in\mathcal{N}$

    $C=[C_i]_{i\in\mathcal{N}}\in\mathbb{R}^{m\times n},\quad d=[d_i]_{i\in\mathcal{N}}\in\mathbb{R}^m,\quad m=\sum_{i\in\mathcal{N}}m_i$
  - $\varphi_i(x)=\frac{\lambda}{|\mathcal{N}|}\|x\|_1+\|C_ix-d_i\|_2^2$ merely convex $(m_i<n)$
  - $\bar\varphi(x)=\sum_{i\in\mathcal{N}}\varphi_i(x)$ strongly convex when **rank**$(C)=n$ $(m\ge n)$

Examples

- **Constrained Lasso**

$$\min_{x\in\mathbb{R}^n}\{\lambda\|x\|_1+\|Cx-d\|_2^2:\ Ax\le b\},\quad \mathcal{K}=-\mathbb{R}^p_+$$

  - distributed data: $C_i\in\mathbb{R}^{m_i\times n}$ and $d_i\in\mathbb{R}^{m_i}$ for $i\in\mathcal{N}$

    $C=[C_i]_{i\in\mathcal{N}}\in\mathbb{R}^{m\times n},\quad d=[d_i]_{i\in\mathcal{N}}\in\mathbb{R}^m,\quad m=\sum_{i\in\mathcal{N}}m_i$
  - $\varphi_i(x)=\frac{\lambda}{|\mathcal{N}|}\|x\|_1+\|C_ix-d_i\|_2^2$ merely convex $(m_i<n)$
  - $\bar{\varphi}(x)=\sum_{i\in\mathcal{N}}\varphi_i(x)$ strongly convex when **rank**$(C)=n$ $(m\ge n)$

- **Closest point in the intersection**

$$\min_{x\in\cap_{i\in\mathcal{N}}\mathcal{X}_i}\sum_{i\in\mathcal{N}}\|x-\bar{x}\|_2^2\quad \text{s.t.}\quad G_i(x)\in-\mathcal{K}_i,\ i\in\mathcal{N}.$$

Related Work - Constrained

- Chang, Nedich, Scaglione'14: primal-dual method
  - $\min_{x \in \cap_{i \in \mathcal{N}} \mathcal{X}_i} \mathcal{F}\left(\sum_{i \in \mathcal{N}} f_i(x)\right)$ s.t. $\sum_{i \in \mathcal{N}} g_i(x) \leq 0$
  - $\mathcal{F}$ and f_i smooth, $\mathcal{X}_i$ compact, and time-varying directed $\mathcal{G}$
  - no rate result, can handle non-smooth constraints

Related Work - Constrained

- Chang, Nedich, Scaglione'14: primal-dual method
  - $\min_{x \in \cap_{i \in \mathcal{N}} \mathcal{X}_i} \mathcal{F}\left(\sum_{i \in \mathcal{N}} f_i(x)\right)$ s.t. $\sum_{i \in \mathcal{N}} g_i(x) \leq 0$
  - $\mathcal{F}$ and f_i smooth, $\mathcal{X}_i$ compact, and time-varying directed $\mathcal{G}$
  - no rate result, can handle non-smooth constraints

- Núñez, Cortés'15: $\min \sum_{i \in \mathcal{N}} \varphi_i(\xi_i, x)$ s.t. $\sum_{i \in \mathcal{N}} g_i(\xi_i, x) \leq 0$
  - φ_i and g_i convex; time-varying directed $\mathcal{G}$
  - $\mathcal{O}(1/\sqrt{k})$ rate for $\mathcal{L}(\bar{\boldsymbol{\xi}}^k, \bar{x}^k, \bar{\mathbf{y}}^k) - \mathcal{L}(\boldsymbol{\xi}^*, x^*, \mathbf{y}^*)$
  - no error bounds on infeasibility, and suboptimality

Related Work - Constrained

- Aybat, Yazdandoost Hamedani'16: primal-dual method
  - $\min \sum_{i \in \mathcal{N}} \varphi_i(x)$ s.t. $A_i x - b_i \in \mathcal{K}_i,\ i \in \mathcal{N}$
  - time-varying undirected and directed $\mathcal{G}$
  - $\mathcal{O}(1/k)$ ergodic rate for infeasibility, and suboptimality
  - Convergence of the primal-dual iterate sequence (without rate)

Related Work - Constrained

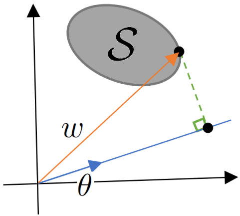
- Aybat, Yazdandoost Hamedani'16: primal-dual method
  - $\min \sum_{i\in\mathcal{N}} \varphi_i(x)$ s.t. $A_i x - b_i \in \mathcal{K}_i,\ i \in \mathcal{N}$
  - time-varying undirected and directed $\mathcal{G}$
  - $\mathcal{O}(1/k)$ ergodic rate for infeasibility, and suboptimality
  - Convergence of the primal-dual iterate sequence (without rate)

- Chang'16: (primal-dual method)
  - $\min_{x_i\in\mathcal{X}_i} \sum_{i\in\mathcal{N}} \rho_i(x_i) + f_i(C_i x_i)$ s.t. $\sum_{i\in\mathcal{N}} A_i x_i = b$
  - f_i smooth and strongly convex; time-varying undirected $\mathcal{G}$
  - $\mathcal{O}(1/k)$ ergodic rate

Notation

- $\|.\|$: Euclidean norm
- $\sigma_{\mathcal{S}}(.)$: Support function of set $\mathcal{S}$,

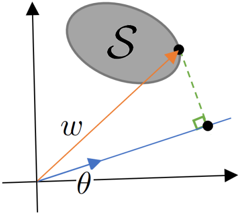
$$\sigma_{\mathcal{S}}(\theta) \triangleq \sup_{w \in \mathcal{S}} \langle \theta,\ w \rangle$$



Notation

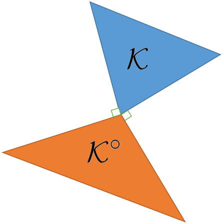
- $\|.\|$: Euclidean norm
- $\sigma_{\mathcal{S}}(.)$: Support function of set $\mathcal{S}$,

$$\sigma_{\mathcal{S}}(\theta) \triangleq \sup_{w \in \mathcal{S}} \langle \theta, \ w \rangle$$



- $\mathcal{P}_{\mathcal{S}}(w) \triangleq \operatorname{argmin}\{\|v - w\| : \ v \in \mathcal{S}\}$: Projection onto $\mathcal{S}$
- $d_{\mathcal{S}}(w) \triangleq \|\mathcal{P}_{\mathcal{S}}(w) - w\|$: Distance function
- $\mathcal{K}^*$: Dual cone of $\mathcal{K}$,
- $\mathcal{K}^\circ$: Polar cone of $\mathcal{K}$,

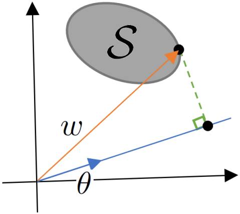
$$\mathcal{K}^\circ \triangleq \{\theta \in \mathbb{R}^m : \ \langle \theta, w \rangle \leq 0 \ \ \forall w \in \mathcal{K}\},$$



Notation

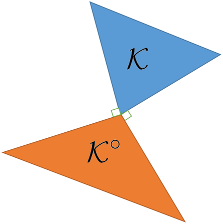
- $\|.\|$: Euclidean norm
- $\sigma_{\mathcal{S}}(.)$: Support function of set $\mathcal{S}$,

$$\sigma_{\mathcal{S}}(\theta) \triangleq \sup_{w \in \mathcal{S}} \langle \theta,\ w \rangle$$



- $\mathcal{P}_{\mathcal{S}}(w) \triangleq \operatorname{argmin}\{\|v - w\| :\ v \in \mathcal{S}\}$: Projection onto $\mathcal{S}$
- $d_{\mathcal{S}}(w) \triangleq \|\mathcal{P}_{\mathcal{S}}(w) - w\|$: Distance function
- $\mathcal{K}^*$: Dual cone of $\mathcal{K}$,
- $\mathcal{K}^\circ$: Polar cone of $\mathcal{K}$,

$$\mathcal{K}^\circ \triangleq \{\theta \in \mathbb{R}^m :\ \langle \theta, w \rangle \leq 0 \ \ \forall w \in \mathcal{K}\},$$



- $\otimes$: Kronecker product
- Π: Cartesian product

Preliminaries: Primal-dual Algorithm (PDA)

PDA for convex-concave saddle-point problem by Chambolle and Pock'16

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} \mathcal{L}(\mathbf{x},\mathbf{y}) \triangleq \Phi(\mathbf{x}) + \langle T\mathbf{x}, \mathbf{y}\rangle - h(\mathbf{y}),$$

- $\Phi(\mathbf{x}) \triangleq \rho(\mathbf{x}) + f(\mathbf{x})$ strongly convex ($\mu > 0$), h convex, T linear map

Preliminaries: Primal-dual Algorithm (PDA)

PDA for convex-concave saddle-point problem by Chambolle and Pock'16

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} \mathcal{L}(\mathbf{x},\mathbf{y}) \triangleq \Phi(\mathbf{x}) + \langle T\mathbf{x}, \mathbf{y}\rangle - h(\mathbf{y}),$$

- $\Phi(\mathbf{x}) \triangleq \rho(\mathbf{x}) + f(\mathbf{x})$ strongly convex ($\mu > 0$), h convex, T linear map

PDA:

$$\mathbf{y}^{k+1} \leftarrow \operatorname*{argmin}_{\mathbf{y}} h(\mathbf{y}) - \left\langle T\left(\mathbf{x}^k + \eta^k(\mathbf{x}^k - \mathbf{x}^{k-1})\right), \mathbf{y}\right\rangle + D_k(\mathbf{y}, \mathbf{y}^k),$$

$$\mathbf{x}^{k+1} \leftarrow \operatorname*{argmin}_{\mathbf{x}} \rho(\mathbf{x}) + f(\mathbf{x}^k) + \left\langle \nabla f(\mathbf{x}^k),\ \mathbf{x} - \mathbf{x}^k\right\rangle + \left\langle T\mathbf{x}, \mathbf{y}^{k+1}\right\rangle + \frac{1}{2\tau^k}\|\mathbf{x} - \mathbf{x}^k\|,$$

- D_k is Bregman distance function $D_k(\mathbf{y}, \bar{\mathbf{y}}) \geq \frac{1}{2\kappa^k}\|\mathbf{y} - \bar{\mathbf{y}}\|^2$

Preliminaries: Primal-dual Algorithm (PDA)

PDA for convex-concave saddle-point problem by Chambolle and Pock'16

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} \mathcal{L}(\mathbf{x},\mathbf{y}) \triangleq \Phi(\mathbf{x}) + \langle T\mathbf{x},\mathbf{y}\rangle - h(\mathbf{y}),$$

- $\Phi(\mathbf{x}) \triangleq \rho(\mathbf{x}) + f(\mathbf{x})$ strongly convex ($\mu > 0$), h convex, T linear map

PDA:

$$\mathbf{y}^{k+1} \leftarrow \operatorname*{argmin}_{\mathbf{y}} h(\mathbf{y}) - \left\langle T\left(\mathbf{x}^k + \eta^k(\mathbf{x}^k - \mathbf{x}^{k-1})\right), \mathbf{y}\right\rangle + D_k(\mathbf{y},\mathbf{y}^k),$$

$$\mathbf{x}^{k+1} \leftarrow \operatorname*{argmin}_{\mathbf{x}} \rho(\mathbf{x}) + f(\mathbf{x}^k) + \left\langle \nabla f(\mathbf{x}^k),\ \mathbf{x} - \mathbf{x}^k\right\rangle + \left\langle T\mathbf{x}, \mathbf{y}^{k+1}\right\rangle + \frac{1}{2\tau^k}\|\mathbf{x} - \mathbf{x}^k\|,$$

- D_k is Bregman distance function $D_k(\mathbf{y},\bar{\mathbf{y}}) \geq \frac{1}{2\kappa^k}\|\mathbf{y} - \bar{\mathbf{y}}\|^2$

Theorem: If $\tau^k, \kappa^k, \eta^k > 0$ such that $\frac{1}{\tau^k} - \mu \geq L + \|T\|^2\kappa^k$, $\kappa^k = \kappa^{k+1}\eta^{k+1}$ and $\eta^{k+1} \geq \tau^k\left(\frac{1}{\tau^{k+1}} - \mu\right)$ for all $k \geq 0$ then

$$\mathcal{L}(\bar{\mathbf{x}}^K, \mathbf{y}) - \mathcal{L}(\mathbf{x}, \bar{\mathbf{y}}^K) \leq \frac{1}{N_K}\left(\frac{1}{2\tau^0}\|\mathbf{x} - \mathbf{x}^0\|^2 + D_0(\mathbf{y},\mathbf{y}^0)\right) \quad \forall \mathbf{x},\mathbf{y}$$

where $N_K = \sum_{k=1}^K \frac{\kappa^k}{\kappa^0} = \mathcal{O}(1/K^2)$ and $(\bar{\mathbf{x}}^K, \bar{\mathbf{y}}^K) \triangleq N_K^{-1}\sum_{k=1}^K \frac{\kappa^k}{\kappa^0}(\mathbf{x}^k,\mathbf{y}^k)$.

Extension to nonlinear constraints

Consider a more general convex-concave saddle-point problem

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} \mathcal{L}(\mathbf{x},\mathbf{y}) \triangleq \Phi(\mathbf{x}) + \boxed{\langle G(\mathbf{x}), \mathbf{y}\rangle} - h(\mathbf{y}),$$

- G is $\mathcal{K}$-convex, Lipschitz continuous (C_G)
- Jacobian ∇G is Lipschitz continuous (L_G)

Extension to nonlinear constraints

Consider a more general convex-concave saddle-point problem

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} \mathcal{L}(\mathbf{x},\mathbf{y}) \triangleq \Phi(\mathbf{x}) + \langle G(\mathbf{x}), \mathbf{y}\rangle - h(\mathbf{y}),$$

- G is $\mathcal{K}$-convex, Lipschitz continuous (C_G)
- Jacobian ∇G is Lipschitz continuous (L_G)

General **PDA**

$$\mathbf{y}^{k+1} \leftarrow \operatorname*{argmin}_{\mathbf{y}} h(\mathbf{y}) - \Big\langle G(\mathbf{x}^k) + \eta^k \Big(G(\mathbf{x}^k) - G(\mathbf{x}^{k-1})\Big), \mathbf{y}\Big\rangle + D_k(\mathbf{y}, \mathbf{y}^k)$$

$$\mathbf{x}^{k+1} \leftarrow \operatorname*{argmin}_{\mathbf{x}} \rho(\mathbf{x}) + f(\mathbf{x}^k) + \Big\langle \nabla f(\mathbf{x}^k),\ \mathbf{x} - \mathbf{x}^k\Big\rangle + \Big\langle \nabla G(\mathbf{x}^k)\mathbf{x}, \mathbf{y}^{k+1}\Big\rangle + \frac{1}{2\tau^k}\|\mathbf{x} - \mathbf{x}^k\|^2$$

Extension to nonlinear constraints

Consider a more general convex-concave saddle-point problem

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} \mathcal{L}(\mathbf{x},\mathbf{y}) \triangleq \Phi(\mathbf{x}) + \langle G(\mathbf{x}),\mathbf{y}\rangle - h(\mathbf{y}),$$

- G is $\mathcal{K}$-convex, Lipschitz continuous (C_G)
- Jacobian ∇G is Lipschitz continuous (L_G)

General **PDA**

$$\mathbf{y}^{k+1} \leftarrow \operatorname*{argmin}_{\mathbf{y}} h(\mathbf{y}) - \left\langle G(\mathbf{x}^k) + \eta^k \left(G(\mathbf{x}^k) - G(\mathbf{x}^{k-1})\right), \mathbf{y}\right\rangle + D_k(\mathbf{y},\mathbf{y}^k)$$

$$\mathbf{x}^{k+1} \leftarrow \operatorname*{argmin}_{\mathbf{x}} \rho(\mathbf{x}) + f(\mathbf{x}^k) + \left\langle \nabla f(\mathbf{x}^k),\ \mathbf{x}-\mathbf{x}^k\right\rangle + \left\langle \nabla G(\mathbf{x}^k)\mathbf{x}, \mathbf{y}^{k+1}\right\rangle + \frac{1}{2\tau^k}\|\mathbf{x}-\mathbf{x}^k\|^2$$

Theorem: If $\tau^k,\kappa^k,\eta^k>0$ such that $\frac{1}{\tau^k}-\mu \geq L + 2C_G^2\kappa^k + L_G\|\mathbf{y}^{k+1}\|$, $\kappa^k = \kappa^{k+1}\eta^{k+1}$ and $\eta^{k+1} \geq \tau^k(\frac{1}{\tau^{k+1}}-\mu)$ for all $k\geq 0$ then

$$\mathcal{L}(\bar{\mathbf{x}}^K,\mathbf{y}) - \mathcal{L}(\mathbf{x},\bar{\mathbf{y}}^K) \leq \frac{1}{N_K}\left(\frac{1}{2\tau^0}\|\mathbf{x}-\mathbf{x}^0\|^2 + D_0(\mathbf{y},\mathbf{y}^0)\right) \quad \forall \mathbf{x},\mathbf{y}$$

where $(\bar{\mathbf{x}}^K,\bar{\mathbf{y}}^K) \triangleq N_K^{-1}\sum_{k=1}^K \frac{\kappa^k}{\kappa^0}(\mathbf{x}^k,\mathbf{y}^k)$, and $N_K = \sum_{k=1}^K \frac{\kappa^k}{\kappa^0} = \mathcal{O}(1/K^2)$.

Extension to nonlinear constraints

We also obtain the following bounds:

$$\frac{1}{2}\|\mathbf{x}^* - \mathbf{x}^K\|^2 \leq \kappa^0 \frac{\tau^K}{\kappa^K}\left(\frac{1}{2\tau^0}\|\mathbf{x}^* - \mathbf{x}^0\|^2 + D_0(\mathbf{y}^*, \mathbf{y}^0)\right)$$

$$\|\mathbf{y}^K\| \leq \|\mathbf{y}^*\| + \sqrt{4\kappa^0\left(\frac{1}{2\tau^0}\|\mathbf{x}^* - \mathbf{x}^0\|^2 + D_0(\mathbf{y}^*, \mathbf{y}^0)\right)}$$

and $\tau^K/\kappa^K = \mathcal{O}(1/K^2)$.

Extension to nonlinear constraints

We also obtain the following bounds:

$$\frac{1}{2}\|\mathbf{x}^* - \mathbf{x}^K\|^2 \leq \kappa^0 \frac{\tau^K}{\kappa^K} \left(\frac{1}{2\tau^0}\|\mathbf{x}^* - \mathbf{x}^0\|^2 + D_0(\mathbf{y}^*, \mathbf{y}^0) \right)$$

$$\|\mathbf{y}^K\| \leq \|\mathbf{y}^*\| + \sqrt{4\kappa^0 \left(\frac{1}{2\tau^0}\|\mathbf{x}^* - \mathbf{x}^0\|^2 + D_0(\mathbf{y}^*, \mathbf{y}^0) \right)}$$

and $\tau^K/\kappa^K = \mathcal{O}(1/K^2)$.

Specific stepsize sequence:

Initialization: $\eta^0 = 0,\ \kappa^0 = 1$

For $k \geq 0:$

$$\tau^k = \frac{1}{2C_G^2\kappa^k + L_G\|\mathbf{y}^{k+1}\| + L + \mu}$$

$$\eta^{k+1} = \sqrt{1 - \mu\tau^k}, \quad \kappa^{k+1} = \kappa^k/\eta^{k+1}$$

Comparison with related works

- Our proximal gradient primal-dual Alg.
  - $\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} \Phi(\mathbf{x}) + \langle G(\mathbf{x}), \mathbf{y}\rangle - h(\mathbf{y})$
  - Φ composite strongly convex, h convex, G is $\mathcal{K}$-convex, Lipschitz, such that ∇G is Lipschitz continuous
  - $\mathcal{L}(\bar{\mathbf{x}}^K, \mathbf{y}) - \mathcal{L}(\mathbf{x}, \bar{\mathbf{y}}^K) \leq \mathcal{O}(1/K^2)$
  - $\|\mathbf{x}^* - \mathbf{x}^K\|^2 \leq \mathcal{O}(1/K^2)$ and $\|y^K\| \leq \mathcal{O}(1)$

Comparison with related works

- Our proximal gradient primal-dual Alg.
  - $\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} \Phi(\mathbf{x}) + \langle G(\mathbf{x}), \mathbf{y}\rangle - h(\mathbf{y})$
  - Φ composite strongly convex, h convex, G is $\mathcal{K}$-convex, Lipschitz, such that ∇G is Lipschitz continuous
  - $\mathcal{L}(\bar{\mathbf{x}}^K, \mathbf{y}) - \mathcal{L}(\mathbf{x}, \bar{\mathbf{y}}^K) \le \mathcal{O}(1/K^2)$
  - $\|\mathbf{x}^* - \mathbf{x}^K\|^2 \le \mathcal{O}(1/K^2)$ and $\|y^K\| \le \mathcal{O}(1)$
- Proximal gradient primal-dual alg. by Chambolle and Pock'16
  - $\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} \Phi(\mathbf{x}) + \langle T\mathbf{x}, \mathbf{y}\rangle - h(\mathbf{y})$
  - Φ composite strongly convex, h convex, and T linear map
  - $\mathcal{L}(\bar{\mathbf{x}}^K, \mathbf{y}) - \mathcal{L}(\mathbf{x}, \bar{\mathbf{y}}^K) \le \mathcal{O}(1/K^2)$
  - $\|\mathbf{x}^* - \mathbf{x}^K\|^2 \le \mathcal{O}(1/K^2)$

Comparison with related works

- Our proximal gradient primal-dual Alg.
  - $\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} \Phi(\mathbf{x}) + \langle G(\mathbf{x}), \mathbf{y}\rangle - h(\mathbf{y})$
  - Φ composite strongly convex, h convex, G is $\mathcal{K}$-convex, Lipschitz, such that ∇G is Lipschitz continuous
  - $\mathcal{L}(\bar{\mathbf{x}}^K, \mathbf{y}) - \mathcal{L}(\mathbf{x}, \bar{\mathbf{y}}^K) \le \mathcal{O}(1/K^2)$
  - $\|\mathbf{x}^* - \mathbf{x}^K\|^2 \le \mathcal{O}(1/K^2)$ and $\|y^K\| \le \mathcal{O}(1)$
- Proximal gradient primal-dual alg. by Chambolle and Pock'16
  - $\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} \Phi(\mathbf{x}) + \langle T\mathbf{x}, \mathbf{y}\rangle - h(\mathbf{y})$
  - Φ composite strongly convex, h convex, and T linear map
  - $\mathcal{L}(\bar{\mathbf{x}}^K, \mathbf{y}) - \mathcal{L}(\mathbf{x}, \bar{\mathbf{y}}^K) \le \mathcal{O}(1/K^2)$
  - $\|\mathbf{x}^* - \mathbf{x}^K\|^2 \le \mathcal{O}(1/K^2)$
- Proximal gradient primal-dual alg. by Yu and Neely'17
  - $\min_{\mathbf{x}} f(\mathbf{x}) \;\; \mathbf{s.t.} \;\; G(\mathbf{x}) \le 0$
  - f composite convex, G composite convex and Lipschitz continuous
  - $f(\bar{\mathbf{x}}^K) - f(\mathbf{x}^*) \le \mathcal{O}(1/K)$ and $G(\bar{\mathbf{x}}^K) \le \mathcal{O}(1/K)$

Comparison with related works

- mirror-prox for $\min_{\mathbf{x}} \max_{\mathbf{y}} \phi(\mathbf{x}, \mathbf{y})$ (Nemirovski'04)
  - $\phi(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ and concave in $\mathbf{y}$,
  - $\nabla\phi$ is Lipschitz continuous in $(\mathbf{x}, \mathbf{y})$
  - $\phi(\bar{\mathbf{x}}^K, \mathbf{y}) - \phi(\mathbf{x}, \bar{\mathbf{y}}^K) \leq \mathcal{O}(1/K)$

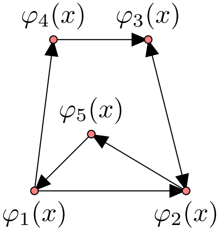
Comparison with related works

- mirror-prox for $\min_{\mathbf{x}} \max_{\mathbf{y}} \phi(\mathbf{x}, \mathbf{y})$ (Nemirovski'04)
  - $\phi(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ and concave in $\mathbf{y}$,
  - $\nabla \phi$ is Lipschitz continuous in $(\mathbf{x}, \mathbf{y})$
  - $\phi(\bar{\mathbf{x}}^K, \mathbf{y}) - \phi(\mathbf{x}, \bar{\mathbf{y}}^K) \leq \mathcal{O}(1/K)$
- primal-dual algorithm with linesearch (Malitsky and Pock'16)
  - $\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbf{y} \in \mathcal{Y}} \Phi(\mathbf{x}) + \langle T\mathbf{x}, \mathbf{y} \rangle - h(\mathbf{y})$
  - Φ and h^* are convex and at least one is strongly convex
  - Proximal primal and dual steps with linesearch determining stepsizes and accelerating parameter
  - $\mathcal{L}(\bar{\mathbf{x}}^K, \mathbf{y}) - \mathcal{L}(\mathbf{x}, \bar{\mathbf{y}}^K) \leq \mathcal{O}(1/K^2)$

Comparison with related works

- mirror-prox for $\min_{\mathbf{x}} \max_{\mathbf{y}} \phi(\mathbf{x}, \mathbf{y})$ (Nemirovski'04)
  - $\phi(\mathbf{x}, \mathbf{y})$ is convex in $\mathbf{x}$ and concave in $\mathbf{y}$,
  - $\nabla \phi$ is Lipschitz continuous in $(\mathbf{x}, \mathbf{y})$
  - $\phi(\bar{\mathbf{x}}^K, \mathbf{y}) - \phi(\mathbf{x}, \bar{\mathbf{y}}^K) \leq \mathcal{O}(1/K)$
- primal-dual algorithm with linesearch (Malitsky and Pock'16)
  - $\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbf{y} \in \mathcal{Y}} \Phi(\mathbf{x}) + \langle T\mathbf{x}, \mathbf{y} \rangle - h(\mathbf{y})$
  - Φ and h^* are convex and at least one is strongly convex
  - Proximal primal and dual steps with linesearch determining stepsizes and accelerating parameter
  - $\mathcal{L}(\bar{\mathbf{x}}^K, \mathbf{y}) - \mathcal{L}(\mathbf{x}, \bar{\mathbf{y}}^K) \leq \mathcal{O}(1/K^2)$
- Accelerated primal-dual algorithm (Chen, Lan and Ouyang'13)
  - $\min_{\mathbf{x} \in \mathcal{X}} \max_{\mathbf{y} \in \mathcal{Y}} \Phi(\mathbf{x}) + \langle T\mathbf{x}, \mathbf{y} \rangle - h(\mathbf{y})$
  - Φ convex function with Lipshcitz continuous gradient, h convex and T linear map
  - Proximal-gradient steps with several accelerating steps
  - $\mathcal{L}(\bar{\mathbf{x}}^K, \mathbf{y}) - \mathcal{L}(\mathbf{x}, \bar{\mathbf{y}}^K) \leq \mathcal{O}(1/K^2 + 1/K)$

Decentralized Consensus Optimization



Compute an optimal solution for

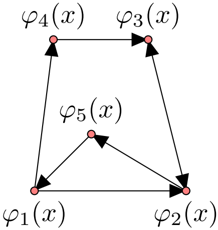
$$(P): \min_x \ \bar{\varphi}(x) \triangleq \sum_{i\in\mathcal{N}} \varphi_i(x) \ \text{s.t.} \ x \in \bigcap_{i\in\mathcal{N}} \mathcal{X}_i$$

- $\mathcal{N} = \{1, \ldots, N\}$ processing nodes on a time-varying $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$
- $\bar{\varphi}(x)$: strongly convex ($\bar{\mu} > 0$)
- $\varphi_i(x) \triangleq \rho_i(x) + f_i(x)$ locally known ($\underline{\mu} \triangleq \min_{i\in\mathcal{N}} \mu_i \geq 0$)
- $\mathcal{X}_i \triangleq \{x : G_i(x) \in -\mathcal{K}_i\}$ locally known, $\mathcal{K}_i$ closed convex cone.

Decentralized Consensus Optimization

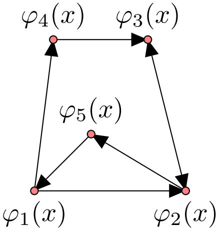
Compute an optimal solution for

$$(P): \ \min_x \ \bar{\varphi}(x) \triangleq \sum_{i \in \mathcal{N}} \varphi_i(x) \ \text{s.t.} \ x \in \bigcap_{i \in \mathcal{N}} \mathcal{X}_i$$



- $\mathcal{N} = \{1, \ldots, N\}$ processing nodes on a time-varying $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$
- $\bar{\varphi}(x)$: strongly convex ($\bar{\mu} > 0$)
- $\varphi_i(x) \triangleq \rho_i(x) + f_i(x)$ locally known ($\underline{\mu} \triangleq \min_{i \in \mathcal{N}} \mu_i \geq 0$)
  - f_i: convex + Lipschitz continuous gradient (constant L_i)
  - ρ_i: convex (possibly non-smooth) + efficient prox-map
  - $\mathbf{prox}_{\rho_i}(x) \triangleq \operatorname{argmin}_{\xi \in \mathbb{R}^n} \left\{\rho_i(\xi) + \frac{1}{2}\|\xi - x\|_2^2\right\}$
- $\mathcal{X}_i \triangleq \{x : G_i(x) \in -\mathcal{K}_i\}$ locally known, $\mathcal{K}_i$ closed convex cone.

Decentralized Consensus Optimization



Compute an optimal solution for

$$(P): \min_x \ \bar{\varphi}(x) \triangleq \sum_{i \in \mathcal{N}} \varphi_i(x) \ \text{s.t.} \ x \in \bigcap_{i \in \mathcal{N}} \mathcal{X}_i$$

- $\mathcal{N} = \{1, \ldots, N\}$ processing nodes on a time-varying $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$
- $\bar{\varphi}(x)$: strongly convex ($\bar{\mu} > 0$)
- $\varphi_i(x) \triangleq \rho_i(x) + f_i(x)$ locally known ($\underline{\mu} \triangleq \min_{i \in \mathcal{N}} \mu_i \geq 0$)
  - f_i: convex + Lipschitz continuous gradient (constant L_i)
  - ρ_i: convex (possibly non-smooth) + efficient prox-map
  - $\mathbf{prox}_{\rho_i}(x) \triangleq \text{argmin}_{\xi \in \mathbb{R}^n} \left\{ \rho_i(\xi) + \frac{1}{2} \|\xi - x\|_2^2 \right\}$
- $\mathcal{X}_i \triangleq \{x : G_i(x) \in -\mathcal{K}_i\}$ locally known, $\mathcal{K}_i$ closed convex cone.
  - G_i: $\mathcal{K}_i$-convex + Lip. cont. (C_G) + Lip. cont. Jacobian ∇G_i (L_G)

Methodology: Time-varying Topology

Suppose $\bar{f}$ strongly convex, and f_i's merely convex, i.e.,

- $\bar{f}(x) = \sum_{i \in \mathcal{N}} f_i(x)$ strongly convex ($\bar{\mu} > 0$)
- $\underline{\mu} = \min_{i \in \mathcal{N}} \mu_i = 0$

Methodology: Time-varying Topology

Suppose $\bar{f}$ strongly convex, and f_i's merely convex, i.e.,

- $\bar{f}(x) = \sum_{i \in \mathcal{N}} f_i(x)$ strongly convex ($\bar{\mu} > 0$)
- $\underline{\mu} = \min_{i \in \mathcal{N}} \mu_i = 0$

Define the consensus cone $\mathcal{C}$:

- $\mathcal{C} \triangleq \{\mathbf{x} = [x_i]_{i \in \mathcal{N}} \in \mathbb{R}^{n|\mathcal{N}|} : \ \exists \bar{x} \in \mathbb{R}^n \ \mathbf{s.t.}\ x_i = \bar{x} \quad \forall\, i \in \mathcal{N}\}$
- $f(\mathbf{x}) \triangleq \sum_{i \in \mathcal{N}} f_i(x_i)$ is strongly convex on $\mathcal{C}$

Methodology: Time-varying Topology

Suppose $\bar{f}$ strongly convex, and f_i's merely convex, i.e.,

- $\bar{f}(x) = \sum_{i\in\mathcal{N}} f_i(x)$ strongly convex ($\bar{\mu} > 0$)
- $\underline{\mu} = \min_{i\in\mathcal{N}} \mu_i = 0$

Define the consensus cone $\mathcal{C}$:

- $\mathcal{C} \triangleq \{\mathbf{x} = [x_i]_{i\in\mathcal{N}} \in \mathbb{R}^{n|\mathcal{N}|} : \ \exists \bar{x} \in \mathbb{R}^n \ \mathbf{s.t.}\ x_i = \bar{x} \quad \forall\, i \in \mathcal{N}\}$
- $f(\mathbf{x}) \triangleq \sum_{i\in\mathcal{N}} f_i(x_i)$ is strongly convex on $\mathcal{C}$

Lemma: Let $f_\alpha(\mathbf{x}) \triangleq f(\mathbf{x}) + \frac{\alpha}{2} d_{\mathcal{C}}^2(\mathbf{x})$. Then f_α is strongly convex with

$$\mu_\alpha \triangleq \frac{\bar{\mu}/|\mathcal{N}| \ + \alpha}{2} - \sqrt{\left(\frac{\bar{\mu}/|\mathcal{N}| \ - \alpha}{2}\right)^2 + 4\bar{L}^2} > 0$$

for any $\alpha > \frac{4}{\bar{\mu}}|\mathcal{N}|\bar{L}^2$, where $\bar{L} = \sqrt{\frac{\sum_{i\in\mathcal{N}} L_i^2}{|\mathcal{N}|}}$.

Methodology: Time-varying Topology

Suppose $\bar{f}$ strongly convex, and f_i's merely convex, i.e.,

- $\bar{f}(x) = \sum_{i\in\mathcal{N}} f_i(x)$ strongly convex ($\bar{\mu} > 0$)
- $\underline{\mu} = \min_{i\in\mathcal{N}} \mu_i = 0$

Define the consensus cone $\mathcal{C}$:

- $\mathcal{C} \triangleq \{\mathbf{x} = [x_i]_{i\in\mathcal{N}} \in \mathbb{R}^{n|\mathcal{N}|} : \ \exists \bar{x} \in \mathbb{R}^n \ \mathbf{s.t.}\ x_i = \bar{x} \quad \forall\ i \in \mathcal{N}\}$
- $f(\mathbf{x}) \triangleq \sum_{i\in\mathcal{N}} f_i(x_i)$ is strongly convex on $\mathcal{C}$

Lemma: Let $f_\alpha(\mathbf{x}) \triangleq f(\mathbf{x}) + \frac{\alpha}{2} d^2_{\mathcal{C}}(\mathbf{x})$. Then f_α is strongly convex with

$$\mu_\alpha \triangleq \frac{\bar{\mu}/|\mathcal{N}| \ + \alpha}{2} - \sqrt{\left(\frac{\bar{\mu}/|\mathcal{N}| \ - \alpha}{2}\right)^2 + 4\bar{L}^2} > 0$$

for any $\alpha > \frac{4}{\bar{\mu}}|\mathcal{N}|\bar{L}^2$, where $\bar{L} = \sqrt{\frac{\sum_{i\in\mathcal{N}} L_i^2}{|\mathcal{N}|}}$.

Note: The result in (Shi et al.'15) uses mixing matrices (static $\mathcal{G}$)

Methodology: Time-varying Topology

Let $\rho(\mathbf{x}) \triangleq \sum_{i\in\mathcal{N}} \rho_i(x_i)$, $\quad f(\mathbf{x}) \triangleq \sum_{i\in\mathcal{N}} f_i(x_i)$, $\quad f_\alpha(\mathbf{x}) \triangleq f(\mathbf{x}) + \frac{\alpha}{2} d_{\mathcal{C}}^2(\mathbf{x})$

Methodology: Time-varying Topology

Let $\rho(\mathbf{x}) \triangleq \sum_{i\in\mathcal{N}} \rho_i(x_i)$, $\quad f(\mathbf{x}) \triangleq \sum_{i\in\mathcal{N}} f_i(x_i)$, $\quad f_\alpha(\mathbf{x}) \triangleq f(\mathbf{x}) + \frac{\alpha}{2} d_{\mathcal{C}}^2(\mathbf{x})$

- $x_i \in \mathbb{R}^n$: local copy of x at $i \in \mathcal{N}$, and $\mathbf{x} \triangleq [x_i]_{i\in\mathcal{N}} \in \mathbb{R}^{n|\mathcal{N}|}$

Methodology: Time-varying Topology

Let $\rho(\mathbf{x}) \triangleq \sum_{i\in\mathcal{N}} \rho_i(x_i)$, $\quad f(\mathbf{x}) \triangleq \sum_{i\in\mathcal{N}} f_i(x_i)$, $\quad f_\alpha(\mathbf{x}) \triangleq f(\mathbf{x}) + \frac{\alpha}{2} d_{\mathcal{C}}^2(\mathbf{x})$

- $x_i \in \mathbb{R}^n$: local copy of x at $i \in \mathcal{N}$, and $\mathbf{x} \triangleq [x_i]_{i\in\mathcal{N}} \in \mathbb{R}^{n|\mathcal{N}|}$
- $\Delta_i \triangleq \max_{x_i, x_i' \in \mathbf{dom}\,\varphi_i} \|x - x'\|$, $\quad \Delta \triangleq \max_{i\in\mathcal{N}} \Delta_i < \infty$

Methodology: Time-varying Topology

Let $\rho(\mathbf{x}) \triangleq \sum_{i\in\mathcal{N}} \rho_i(x_i)$, $\quad f(\mathbf{x}) \triangleq \sum_{i\in\mathcal{N}} f_i(x_i)$, $\quad f_\alpha(\mathbf{x}) \triangleq f(\mathbf{x}) + \frac{\alpha}{2} d_{\mathcal{C}}^2(\mathbf{x})$

- $x_i \in \mathbb{R}^n$: local copy of x at $i \in \mathcal{N}$, and $\mathbf{x} \triangleq [x_i]_{i\in\mathcal{N}} \in \mathbb{R}^{n|\mathcal{N}|}$
- $\Delta_i \triangleq \max_{x_i, x_i' \in \mathbf{dom}\,\varphi_i} \|x - x'\|$, $\quad \Delta \triangleq \max_{i\in\mathcal{N}} \Delta_i < \infty$
- $\mathcal{B}_0 \triangleq \{x \in \mathbb{R}^m : \ \|x\| \le 2\Delta\}$, $\quad \mathcal{B} \triangleq \Pi_{i\in\mathcal{N}} \mathcal{B}_0$, $\quad \widetilde{\mathcal{C}} \triangleq \mathcal{C} \cap \mathcal{B}$

Methodology: Time-varying Topology

Let $\rho(\mathbf{x}) \triangleq \sum_{i\in\mathcal{N}} \rho_i(x_i)$, $f(\mathbf{x}) \triangleq \sum_{i\in\mathcal{N}} f_i(x_i)$, $f_\alpha(\mathbf{x}) \triangleq f(\mathbf{x}) + \frac{\alpha}{2} d_{\mathcal{C}}^2(\mathbf{x})$

- $x_i \in \mathbb{R}^n$: local copy of x at $i \in \mathcal{N}$, and $\mathbf{x} \triangleq [x_i]_{i\in\mathcal{N}} \in \mathbb{R}^{n|\mathcal{N}|}$
- $\Delta_i \triangleq \max_{x_i, x_i' \in \mathbf{dom}\,\varphi_i} \|x - x'\|$, $\Delta \triangleq \max_{i\in\mathcal{N}} \Delta_i < \infty$
- $\mathcal{B}_0 \triangleq \{x \in \mathbb{R}^m : \|x\| \le 2\Delta\}$, $\mathcal{B} \triangleq \Pi_{i\in\mathcal{N}} \mathcal{B}_0$, $\widetilde{\mathcal{C}} \triangleq \mathcal{C} \cap \mathcal{B}$
- For any $\alpha \ge 0$, an equivalent formulation:

$$\begin{array}{ll} \min\limits_{x\in\mathbb{R}^n} & \sum_{i\in\mathcal{N}} \rho_i(x) + f_i(x) \\ \text{s.t.} & G_i(x) \in -\mathcal{K}_i,\ i \in \mathcal{N} \end{array} \quad \equiv \quad \begin{array}{ll} \min\limits_{\mathbf{x}\in\widetilde{\mathcal{C}}} & \rho(\mathbf{x}) + f_\alpha(\mathbf{x}) \\ \text{s.t.} & G_i(x_i) \in -\mathcal{K}_i,\ i \in \mathcal{N} \end{array}$$

Methodology: Time-varying Topology

Let $\rho(\mathbf{x}) \triangleq \sum_{i\in\mathcal{N}} \rho_i(x_i)$, $f(\mathbf{x}) \triangleq \sum_{i\in\mathcal{N}} f_i(x_i)$, $f_\alpha(\mathbf{x}) \triangleq f(\mathbf{x}) + \frac{\alpha}{2} d_{\mathcal{C}}^2(\mathbf{x})$

- $x_i \in \mathbb{R}^n$: local copy of x at $i \in \mathcal{N}$, and $\mathbf{x} \triangleq [x_i]_{i\in\mathcal{N}} \in \mathbb{R}^{n|\mathcal{N}|}$
- $\Delta_i \triangleq \max_{x_i, x_i' \in \mathbf{dom}\, \varphi_i} \|x - x'\|$, $\Delta \triangleq \max_{i\in\mathcal{N}} \Delta_i < \infty$
- $\mathcal{B}_0 \triangleq \{x \in \mathbb{R}^m : \ \|x\| \leq 2\Delta\}$, $\mathcal{B} \triangleq \Pi_{i\in\mathcal{N}} \mathcal{B}_0$, $\widetilde{\mathcal{C}} \triangleq \mathcal{C} \cap \mathcal{B}$
- For any $\alpha \geq 0$, an equivalent formulation:

$$\begin{array}{cl} \min\limits_{x\in\mathbb{R}^n} & \sum_{i\in\mathcal{N}} \rho_i(x) + f_i(x) \\ \text{s.t.} & G_i(x) \in -\mathcal{K}_i, \ i \in \mathcal{N} \end{array} \quad \equiv \quad \begin{array}{cl} \min\limits_{\mathbf{x}\in\widetilde{\mathcal{C}}} & \rho(\mathbf{x}) + f_\alpha(\mathbf{x}) \\ \text{s.t.} & G_i(x_i) \in -\mathcal{K}_i, \ i \in \mathcal{N} \end{array}$$

- Saddle Point Formulation:

$$\min_{\mathbf{x}} \max_{\mathbf{y}} \ \mathcal{L}(\mathbf{x}, \mathbf{y}) \triangleq \rho(\mathbf{x}) + f_\alpha(\mathbf{x}) + \langle \boldsymbol{\lambda}, \ \mathbf{x} \rangle - \sigma_{\widetilde{\mathcal{C}}}(\boldsymbol{\lambda}) + \sum_{i\in\mathcal{N}} \langle \theta_i, G_i(x_i) \rangle - \sigma_{-\mathcal{K}_i}(\theta_i)$$

$\mathbf{y} = [\boldsymbol{\theta}^\top \ \boldsymbol{\lambda}^\top]^\top$, $\boldsymbol{\theta} = [\theta_i]_{i\in\mathcal{N}} \in \mathbb{R}^m$ and $\boldsymbol{\lambda} = [\lambda_i]_{i\in\mathcal{N}} \in \mathbb{R}^{n|\mathcal{N}|}$

Methodology: Time-varying Topology

- Implementing **PDA** on the saddle-point formulation:

$$\theta_i^{k+1} \leftarrow \underset{\theta_i}{\operatorname{argmin}}\, \sigma_{-\mathcal{K}_i}(\theta_i) - \langle G_i(x_i^k) + \eta^k(G_i(x_i^k) - G_i(x_i^{k-1})),\ \theta_i\rangle + \frac{1}{2\kappa^k}\|\theta_i - \theta_i^k\|_2^2$$

$$\boldsymbol{\lambda}^{k+1} \leftarrow \underset{\boldsymbol{\lambda}}{\operatorname{argmin}}\, \sigma_{\widetilde{\mathcal{C}}}(\boldsymbol{\lambda}) - \langle \mathbf{x}^k + \eta^k(\mathbf{x}^k - \mathbf{x}^{k-1}),\ \boldsymbol{\lambda}\rangle + \frac{1}{2\gamma^k}\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^k\|_2^2,$$

$$\mathbf{x}^{k+1} \leftarrow \underset{\mathbf{x}}{\operatorname{argmin}}\, \rho(\mathbf{x}) + \langle \nabla f_\alpha(\mathbf{x}^k),\ \mathbf{x}\rangle + \langle \nabla G(\mathbf{x}^k)\mathbf{x}, \boldsymbol{\theta}^{k+1}\rangle + \langle \mathbf{x},\ \boldsymbol{\lambda}^{k+1}\rangle + \frac{1}{2\tau^k}\|\mathbf{x} - \mathbf{x}^k\|^2$$

Methodology: Time-varying Topology

- Implementing **PDA** on the saddle-point formulation:

$$\theta_i^{k+1} \leftarrow \underset{\theta_i}{\operatorname{argmin}}\, \sigma_{-\mathcal{K}_i}(\theta_i) - \langle G_i(x_i^k) + \eta^k (G_i(x_i^k) - G_i(x_i^{k-1})),\ \theta_i\rangle + \frac{1}{2\kappa^k}\|\theta_i - \theta_i^k\|_2^2$$

$$\boldsymbol{\lambda}^{k+1} \leftarrow \underset{\boldsymbol{\lambda}}{\operatorname{argmin}}\, \sigma_{\widetilde{\mathcal{C}}}(\boldsymbol{\lambda}) - \langle \mathbf{x}^k + \eta^k(\mathbf{x}^k - \mathbf{x}^{k-1}),\ \boldsymbol{\lambda}\rangle + \frac{1}{2\gamma^k}\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^k\|_2^2,$$

$$\mathbf{x}^{k+1} \leftarrow \underset{\mathbf{x}}{\operatorname{argmin}}\, \rho(\mathbf{x}) + \langle \nabla f_\alpha(\mathbf{x}^k),\ \mathbf{x}\rangle + \langle \nabla G(\mathbf{x}^k)\mathbf{x}, \boldsymbol{\theta}^{k+1}\rangle + \langle \mathbf{x},\ \boldsymbol{\lambda}^{k+1}\rangle + \frac{1}{2\tau^k}\|\mathbf{x} - \mathbf{x}^k\|^2$$

Note: $\boldsymbol{\lambda}$, $\mathbf{x}$ updates require $\mathcal{P}_{\mathcal{C}}(\boldsymbol{\omega}) = \mathbf{1}_{|\mathcal{N}|} \otimes \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} \omega_i$ and $\mathcal{P}_{\widetilde{\mathcal{C}}}(\boldsymbol{\omega}) = \mathcal{P}_{\mathcal{B}}(\mathcal{P}_{\mathcal{C}}(\boldsymbol{\omega}))$

Methodology: Time-varying Topology

- Implementing **PDA** on the saddle-point formulation:

$$\theta_i^{k+1} \leftarrow \operatorname*{argmin}_{\theta_i} \sigma_{-\mathcal{K}_i}(\theta_i) - \langle G_i(x_i^k) + \eta^k (G_i(x_i^k) - G_i(x_i^{k-1})),\ \theta_i \rangle + \frac{1}{2\kappa^k} \|\theta_i - \theta_i^k\|_2^2$$

$$\boldsymbol{\lambda}^{k+1} \leftarrow \operatorname*{argmin}_{\boldsymbol{\lambda}} \sigma_{\widetilde{\mathcal{C}}}(\boldsymbol{\lambda}) - \langle \mathbf{x}^k + \eta^k (\mathbf{x}^k - \mathbf{x}^{k-1}),\ \boldsymbol{\lambda} \rangle + \frac{1}{2\gamma^k} \|\boldsymbol{\lambda} - \boldsymbol{\lambda}^k\|_2^2,$$

$$\mathbf{x}^{k+1} \leftarrow \operatorname*{argmin}_{\mathbf{x}} \rho(\mathbf{x}) + \langle \nabla f_\alpha(\mathbf{x}^k),\ \mathbf{x} \rangle + \langle \nabla G(\mathbf{x}^k)\mathbf{x}, \boldsymbol{\theta}^{k+1} \rangle + \langle \mathbf{x},\ \boldsymbol{\lambda}^{k+1} \rangle + \frac{1}{2\tau^k} \|\mathbf{x} - \mathbf{x}^k\|^2$$

Note: $\boldsymbol{\lambda}$, $\mathbf{x}$ updates require $\mathcal{P}_{\mathcal{C}}(\boldsymbol{\omega}) = \mathbf{1}_{|\mathcal{N}|} \otimes \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} \omega_i$ and $\mathcal{P}_{\widetilde{\mathcal{C}}}(\boldsymbol{\omega}) = \mathcal{P}_{\mathcal{B}}(\mathcal{P}_{\mathcal{C}}(\boldsymbol{\omega}))$

- $\boldsymbol{\lambda}$ update: $\boldsymbol{\lambda}^{k+1} = \gamma^k \left(\boldsymbol{\omega}^k - \mathcal{P}_{\widetilde{\mathcal{C}}}(\boldsymbol{\omega}^k) \right)$, $\boldsymbol{\omega}^k \triangleq \frac{1}{\gamma^k} \boldsymbol{\lambda}^k + \mathbf{x}^k + \eta^k (\mathbf{x}^k - \mathbf{x}^{k-1})$
- $\mathbf{x}$ update: $\nabla f_\alpha(\mathbf{x}^k) = f(\mathbf{x}^k) + \alpha(\mathbf{x}^k - \mathcal{P}_{\mathcal{C}}(\mathbf{x}^k))$

Methodology: Time-varying Topology

- Implementing **PDA** on the saddle-point formulation:

$$\theta_i^{k+1} \leftarrow \operatorname*{argmin}_{\theta_i} \sigma_{-\mathcal{K}_i}(\theta_i) - \langle G_i(x_i^k) + \eta^k(G_i(x_i^k) - G_i(x_i^{k-1})),\ \theta_i\rangle + \frac{1}{2\kappa^k}\|\theta_i - \theta_i^k\|_2^2$$

$$\boldsymbol{\lambda}^{k+1} \leftarrow \operatorname*{argmin}_{\boldsymbol{\lambda}} \sigma_{\widetilde{\mathcal{C}}}(\boldsymbol{\lambda}) - \langle \mathbf{x}^k + \eta^k(\mathbf{x}^k - \mathbf{x}^{k-1}),\ \boldsymbol{\lambda}\rangle + \frac{1}{2\gamma^k}\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^k\|_2^2,$$

$$\mathbf{x}^{k+1} \leftarrow \operatorname*{argmin}_{\mathbf{x}} \rho(\mathbf{x}) + \langle \nabla f_\alpha(\mathbf{x}^k),\ \mathbf{x}\rangle + \langle \nabla G(\mathbf{x}^k)\mathbf{x}, \boldsymbol{\theta}^{k+1}\rangle + \langle \mathbf{x},\ \boldsymbol{\lambda}^{k+1}\rangle + \frac{1}{2\tau^k}\|\mathbf{x} - \mathbf{x}^k\|^2$$

Note: $\boldsymbol{\lambda}$, $\mathbf{x}$ updates require $\mathcal{P}_{\mathcal{C}}(\boldsymbol{\omega}) = \mathbf{1}_{|\mathcal{N}|} \otimes \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} \omega_i$ and $\mathcal{P}_{\widetilde{\mathcal{C}}}(\boldsymbol{\omega}) = \mathcal{P}_{\mathcal{B}}(\mathcal{P}_{\mathcal{C}}(\boldsymbol{\omega}))$

- $\boldsymbol{\lambda}$ update: $\boldsymbol{\lambda}^{k+1} = \gamma^k\left(\boldsymbol{\omega}^k - \mathcal{P}_{\widetilde{\mathcal{C}}}(\boldsymbol{\omega}^k)\right)$, $\boldsymbol{\omega}^k \triangleq \frac{1}{\gamma^k}\boldsymbol{\lambda}^k + \mathbf{x}^k + \eta^k(\mathbf{x}^k - \mathbf{x}^{k-1})$
- $\mathbf{x}$ update: $\nabla f_\alpha(\mathbf{x}^k) = f(\mathbf{x}^k) + \alpha(\mathbf{x}^k - \mathcal{P}_{\mathcal{C}}(\mathbf{x}^k))$
- Approximate averaging operator $\mathcal{R}^k(\boldsymbol{\omega}) = [\mathcal{R}_i^k(\boldsymbol{\omega})]_{i\in\mathcal{N}}$ s.t.

$\|\mathcal{P}_{\mathcal{C}}(\boldsymbol{\omega}) - \mathcal{R}^k(\boldsymbol{\omega})\| = \mathcal{O}(\beta^{q_k}\|\boldsymbol{\omega}\|)$ $\forall\boldsymbol{\omega}$ for some $\beta \in (0,1)$, increasing $\{q_k\}_{k\geq 0}$

Methodology: Time-varying Topology

- Implementing **PDA** on the saddle-point formulation:

$$\theta_i^{k+1} \leftarrow \underset{\theta_i}{\operatorname{argmin}}\, \sigma_{-\mathcal{K}_i}(\theta_i) - \langle G_i(x_i^k) + \eta^k (G_i(x_i^k) - G_i(x_i^{k-1})),\ \theta_i\rangle + \frac{1}{2\kappa^k}\|\theta_i - \theta_i^k\|_2^2$$

$$\boldsymbol{\lambda}^{k+1} \leftarrow \underset{\boldsymbol{\lambda}}{\operatorname{argmin}}\, \sigma_{\widetilde{\mathcal{C}}}(\boldsymbol{\lambda}) - \langle \mathbf{x}^k + \eta^k(\mathbf{x}^k - \mathbf{x}^{k-1}),\ \boldsymbol{\lambda}\rangle + \frac{1}{2\gamma^k}\|\boldsymbol{\lambda} - \boldsymbol{\lambda}^k\|_2^2,$$

$$\mathbf{x}^{k+1} \leftarrow \underset{\mathbf{x}}{\operatorname{argmin}}\, \rho(\mathbf{x}) + \langle \nabla f_\alpha(\mathbf{x}^k),\ \mathbf{x}\rangle + \langle \nabla G(\mathbf{x}^k)\mathbf{x}, \boldsymbol{\theta}^{k+1}\rangle + \langle \mathbf{x},\ \boldsymbol{\lambda}^{k+1}\rangle + \frac{1}{2\tau^k}\|\mathbf{x} - \mathbf{x}^k\|^2$$

Note: $\boldsymbol{\lambda}$, $\mathbf{x}$ updates require $\mathcal{P}_{\mathcal{C}}(\boldsymbol{\omega}) = \mathbf{1}_{|\mathcal{N}|} \otimes \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} \omega_i$ and $\mathcal{P}_{\widetilde{\mathcal{C}}}(\boldsymbol{\omega}) = \mathcal{P}_{\mathcal{B}}(\mathcal{P}_{\mathcal{C}}(\boldsymbol{\omega}))$

- $\boldsymbol{\lambda}$ update: $\boldsymbol{\lambda}^{k+1} = \gamma^k\left(\boldsymbol{\omega}^k - \mathcal{P}_{\widetilde{\mathcal{C}}}(\boldsymbol{\omega}^k)\right)$, $\boldsymbol{\omega}^k \triangleq \frac{1}{\gamma^k}\boldsymbol{\lambda}^k + \mathbf{x}^k + \eta^k(\mathbf{x}^k - \mathbf{x}^{k-1})$
- $\mathbf{x}$ update: $\nabla f_\alpha(\mathbf{x}^k) = f(\mathbf{x}^k) + \alpha(\mathbf{x}^k - \mathcal{P}_{\mathcal{C}}(\mathbf{x}^k))$
- Approximate averaging operator $\mathcal{R}^k(\boldsymbol{\omega}) = [\mathcal{R}_i^k(\boldsymbol{\omega})]_{i\in\mathcal{N}}$ s.t.

$\|\mathcal{P}_{\mathcal{C}}(\boldsymbol{\omega}) - \mathcal{R}^k(\boldsymbol{\omega})\| = \mathcal{O}(\beta^{q_k}\|\boldsymbol{\omega}\|)$ $\forall\boldsymbol{\omega}$ for some $\beta \in (0,1)$, increasing $\{q_k\}_{k\geq 0}$

$$\boxed{\boldsymbol{\lambda}^{k+1} \leftarrow \gamma^k\left(\boldsymbol{\omega}^k - \mathcal{P}_{\mathcal{B}}\big(\mathcal{R}^k(\boldsymbol{\omega}^k)\big)\right)}, \quad \boxed{\mathbf{x}^{k+1} \leftarrow \mathbf{prox}_{\tau^k\rho}\left(\mathbf{x}^k - \tau^k\mathbf{s}^k\right)},$$

$$\mathbf{s}^k \leftarrow \nabla f(\mathbf{x}^k) + \nabla G(\mathbf{x}^k)^\top\boldsymbol{\theta}^{k+1} + \boldsymbol{\lambda}^{k+1} + \alpha\,\boxed{\left(\mathbf{x}^k - \mathcal{R}^k(\mathbf{x}^k)\right)}$$

Methodology: Time-varying Topology

If $\underline{\mu} > 0$, then $\alpha \leftarrow 0$ and $\mu \leftarrow \underline{\mu}$; else, $\alpha > \frac{4}{\bar{\mu}} \sum_{i \in \mathcal{N}} L_i^2$ and $\mu \leftarrow \mu_\alpha$

Algorithm DPDA-TV ($\mathbf{x}^0, \boldsymbol{\theta}^0, \alpha, \delta, \gamma, \mu$)

Initialization: $\mathbf{x}^{-1} \leftarrow \mathbf{x}^0$, $\mathbf{s}^0 \leftarrow \mathbf{0}$,
$\delta, \gamma > 0, \quad \gamma^0 \leftarrow \gamma, \quad \mu \in (0,\ \max\{\underline{\mu},\ \mu_\alpha\}], \quad \eta^0 \leftarrow 0, \quad \kappa^0 \leftarrow \gamma \frac{\delta}{2C_G^2} \quad i \in \mathcal{N}$

Step k: $(k \geq 0,\ i \in \mathcal{N})$

1. $\theta_i^{k+1} \leftarrow \mathcal{P}_{\mathcal{K}_i^*}\Big(\theta_i^k + \kappa^k \big(G_i(x_i^k) + \eta^k (G_i(x_i^k) - G_i(x_i^{k-1}))\big)\Big)$,
2. $\omega_i^k \leftarrow \frac{1}{\gamma^k} \lambda_i^k + x_i^k + \eta^k (x_i^k - x_i^{k-1})$,
3. $\lambda_i^{k+1} \leftarrow \gamma^k \omega_i^k - \gamma^k \mathcal{P}_{\mathcal{B}_0}\Big(\mathcal{R}_i^k(\boldsymbol{\omega}^k) \Big)$,
5. $s_i^k \leftarrow \nabla f_i(x_i^k) + \nabla G_i(x_i^k)^\top \theta_i^{k+1} + \lambda_i^{k+1} + \alpha\big(x_i^k - \mathcal{R}_i^k(\mathbf{x}^k)\big)$,
4. $x_i^{k+1} \leftarrow \mathbf{prox}_{\tau^k \rho_i}\Big(x_i^k - \tau^k s_i^k\Big)$,
5. τ^{k+1}, η^{k+1}, γ^{k+1}, κ^{k+1} update by step-size condition rule

Methodology: Time-varying Topology

If $\underline{\mu} > 0$, then $\alpha \leftarrow 0$ and $\mu \leftarrow \underline{\mu}$; else, $\alpha > \frac{4}{\bar{\mu}} \sum_{i \in \mathcal{N}} L_i^2$ and $\mu \leftarrow \mu_\alpha$

- **Step-size condition:** given $\delta > 0$, choose $\tau^k, \eta^k, \kappa^k, \gamma^k > 0$ such that

$$\eta^{k+1} \geq \tau^k \left(\frac{1}{\tau^{k+1}} - \mu \right), \quad \frac{1}{\tau^k} - (L_i + \alpha + \mu + L_G \|\theta_i^{k+1}\|) \geq \gamma^k \eta^{k+1} (2 + \delta)$$

$$\frac{2\kappa^k C_G^2}{\gamma^k} \leq \delta, \quad \kappa^{k+1} \geq \frac{\kappa^k}{\eta^{k+1}}, \quad \gamma^{k+1} = \frac{\gamma^k}{\eta^{k+1}}$$

- **A possible way of choosing:**

Initialization: $\eta^0 \leftarrow 0, \ \gamma^0 \leftarrow \gamma, \ \kappa^0 \leftarrow \gamma \dfrac{\delta}{2C_G^2}$

For $k \geq 0$:

$$\tilde{\tau}^k \leftarrow \frac{1}{\gamma^k(2+\delta) + L_{\max} + \alpha + L_G \max_{i \in \mathcal{N}} \|\theta_i^{k+1}\|}, \quad \tau^k \leftarrow \left(\frac{1}{\tilde{\tau}^k} + \mu\right)^{-1}$$

$$\gamma^{k+1} \leftarrow \gamma^k \sqrt{1 + \mu \tilde{\tau}^k}, \quad \eta^{k+1} \leftarrow \gamma^k / \gamma^{k+1}, \quad \kappa^{k+1} \leftarrow \gamma^{k+1} \frac{\delta}{2C_G^2}$$

We have $\eta^k \in (0,1)$, $\tau^k = \Theta(1/k)$, $\gamma^k = \Theta(k)$, $\kappa^k = \Theta(k)$

Main result

Theorem: Let $\boldsymbol{\lambda}^0 = \mathbf{0}$, $\boldsymbol{\theta}^0 = \mathbf{0}$. Suppose step-size condition holds. $q_k \geq \lceil (5+c) \log_{1/\beta}(k+1) \rceil$ communication rounds at iteration-k.

Then $\mathbf{x}^k \to \mathbf{x}^*$ such that $\mathbf{x}^* = \mathbf{1}_{|\mathcal{N}|} \otimes x^*$, i.e., $x_i^* = x^*$ for $i \in \mathcal{N}$.

Within $\mathcal{O}(K \log_{1/\beta}(K))$ communication rounds,

Main result

Theorem: Let $\boldsymbol{\lambda}^0 = \mathbf{0}$, $\boldsymbol{\theta}^0 = \mathbf{0}$. Suppose step-size condition holds.
$q_k \geq \lceil (5+c) \log_{1/\beta}(k+1) \rceil$ communication rounds at iteration-k.

Then $\mathbf{x}^k \to \mathbf{x}^*$ such that $\mathbf{x}^* = \mathbf{1}_{|\mathcal{N}|} \otimes x^*$, i.e., $x_i^* = x^*$ for $i \in \mathcal{N}$.

Within $\mathcal{O}(K \log_{1/\beta}(K))$ communication rounds,

Dual iterate bound: $\|\theta_i^K\| \leq \|\theta_i^*\| + \sqrt{\frac{\delta\gamma^0}{C_G^2}\Lambda(K)}, \quad \forall i \in \mathcal{N}$

Main result

Theorem: Let $\boldsymbol{\lambda}^0 = \mathbf{0}$, $\boldsymbol{\theta}^0 = \mathbf{0}$. Suppose step-size condition holds.
$q_k \geq \lceil (5+c)\log_{1/\beta}(k+1) \rceil$ communication rounds at iteration-k.

Then $\mathbf{x}^k \to \mathbf{x}^*$ such that $\mathbf{x}^* = \mathbf{1}_{|\mathcal{N}|} \otimes x^*$, i.e., $x_i^* = x^*$ for $i \in \mathcal{N}$.

Within $\mathcal{O}(K \log_{1/\beta}(K))$ communication rounds,

Dual iterate bound: $\|\theta_i^K\| \leq \|\theta_i^*\| + \sqrt{\frac{\delta\gamma^0}{C_G^2}\Lambda(K)}, \quad \forall i \in \mathcal{N}$

Suboptimality: $|\varphi(\bar{\mathbf{x}}^K) - \varphi(\mathbf{x}^*)| \leq \frac{\Lambda(K)}{N_K}$, where $\varphi(\mathbf{x}) = \rho(\mathbf{x}) + f(\mathbf{x})$

Main result

Theorem: Let $\boldsymbol{\lambda}^0 = \mathbf{0}$, $\boldsymbol{\theta}^0 = \mathbf{0}$. Suppose step-size condition holds.
$q_k \geq \lceil (5+c) \log_{1/\beta}(k+1) \rceil$ communication rounds at iteration-k.

Then $\mathbf{x}^k \to \mathbf{x}^*$ such that $\mathbf{x}^* = \mathbf{1}_{|\mathcal{N}|} \otimes x^*$, i.e., $x_i^* = x^*$ for $i \in \mathcal{N}$.

Within $\mathcal{O}(K \log_{1/\beta}(K))$ communication rounds,

Dual iterate bound: $\|\theta_i^K\| \leq \|\theta_i^*\| + \sqrt{\frac{\delta\gamma^0}{C_G^2}\Lambda(K)}, \quad \forall i \in \mathcal{N}$

Suboptimality: $|\varphi(\bar{\mathbf{x}}^K) - \varphi(\mathbf{x}^*)| \leq \frac{\Lambda(K)}{N_K}$, where $\varphi(\mathbf{x}) = \rho(\mathbf{x}) + f(\mathbf{x})$

Infeasiblity: $d_{\mathcal{C}}(\bar{\mathbf{x}}^K) + \sum_{i\in\mathcal{N}} \|\theta_i^*\| d_{-\mathcal{K}_i}(G_i(\bar{x}_i^K)) \leq \frac{\Lambda(K)}{N_K}$

Main result

Theorem: Let $\boldsymbol{\lambda}^0 = \mathbf{0}$, $\boldsymbol{\theta}^0 = \mathbf{0}$. Suppose step-size condition holds. $q_k \geq \lceil (5+c) \log_{1/\beta}(k+1) \rceil$ communication rounds at iteration-k.

Then $\mathbf{x}^k \to \mathbf{x}^*$ such that $\mathbf{x}^* = \mathbf{1}_{|\mathcal{N}|} \otimes x^*$, i.e., $x_i^* = x^*$ for $i \in \mathcal{N}$.

Within $\mathcal{O}(K \log_{1/\beta}(K))$ communication rounds,

Dual iterate bound: $\|\theta_i^K\| \leq \|\theta_i^*\| + \sqrt{\frac{\delta\gamma^0}{C_G^2}\Lambda(K)}, \quad \forall i \in \mathcal{N}$

Suboptimality: $|\varphi(\bar{\mathbf{x}}^K) - \varphi(\mathbf{x}^*)| \leq \frac{\Lambda(K)}{N_K}$, where $\varphi(\mathbf{x}) = \rho(\mathbf{x}) + f(\mathbf{x})$

Infeasiblity: $d_{\mathcal{C}}(\bar{\mathbf{x}}^K) + \sum_{i \in \mathcal{N}} \|\theta_i^*\| d_{-\mathcal{K}_i}(G_i(\bar{x}_i^K)) \leq \frac{\Lambda(K)}{N_K}$

Solution error: $\|\mathbf{x}^K - \mathbf{x}^*\|^2 \leq 2\gamma^0 \, \frac{\tilde{\tau}^K}{\gamma^K} \, \Lambda(K)$

Main result

Theorem: Let $\boldsymbol{\lambda}^0 = \mathbf{0}$, $\boldsymbol{\theta}^0 = \mathbf{0}$. Suppose step-size condition holds. $q_k \geq \lceil (5+c) \log_{1/\beta}(k+1) \rceil$ communication rounds at iteration-k.

Then $\mathbf{x}^k \to \mathbf{x}^*$ such that $\mathbf{x}^* = \mathbf{1}_{|\mathcal{N}|} \otimes x^*$, i.e., $x_i^* = x^*$ for $i \in \mathcal{N}$.

Within $\mathcal{O}(K \log_{1/\beta}(K))$ communication rounds,

Dual iterate bound: $\|\theta_i^K\| \leq \|\theta_i^*\| + \sqrt{\frac{\delta\gamma^0}{C_G^2}\Lambda(K)}, \quad \forall i \in \mathcal{N}$

Suboptimality: $|\varphi(\bar{\mathbf{x}}^K) - \varphi(\mathbf{x}^*)| \leq \frac{\Lambda(K)}{N_K}$, where $\varphi(\mathbf{x}) = \rho(\mathbf{x}) + f(\mathbf{x})$

Infeasiblity: $d_{\mathcal{C}}(\bar{\mathbf{x}}^K) + \sum_{i \in \mathcal{N}} \|\theta_i^*\| d_{-\mathcal{K}_i}(G_i(\bar{x}_i^K)) \leq \frac{\Lambda(K)}{N_K}$

Solution error: $\|\mathbf{x}^K - \mathbf{x}^*\|^2 \leq 2\gamma^0 \, \frac{\tilde{\tau}^K}{\gamma^K} \, \Lambda(K)$

- $\Lambda(K) = \mathcal{O}(\sum_{k=1}^K \beta^{q_{k-1}} k^4)$ and $\sup_{K \geq 1} \Lambda(K) < \infty$
- $N_K = \mathcal{O}(K^2)$, $\tilde{\tau}^K / \gamma^K = \mathcal{O}(1/K^2)$ and $\kappa^K / \gamma^K = \mathcal{O}(1)$

Main result

Theorem: Let $\boldsymbol{\lambda}^0 = \mathbf{0}$, $\boldsymbol{\theta}^0 = \mathbf{0}$. Suppose step-size condition holds. $q_k \geq \lceil (5+c)\log_{1/\beta}(k+1) \rceil$ communication rounds at iteration-k.

Then $\mathbf{x}^k \to \mathbf{x}^*$ such that $\mathbf{x}^* = \mathbf{1}_{|\mathcal{N}|} \otimes x^*$, i.e., $x_i^* = x^*$ for $i \in \mathcal{N}$.

Within $\mathcal{O}(K \log_{1/\beta}(K))$ communication rounds,

Dual iterate bound: $\|\theta_i^K\| \leq \|\theta_i^*\| + \sqrt{\frac{\delta\gamma^0}{C_G^2}\Lambda(K)}, \quad \forall i \in \mathcal{N}$

Suboptimality: $|\varphi(\bar{\mathbf{x}}^K) - \varphi(\mathbf{x}^*)| \leq \frac{\Lambda(K)}{N_K}$, where $\varphi(\mathbf{x}) = \rho(\mathbf{x}) + f(\mathbf{x})$

Infeasiblity: $d_{\mathcal{C}}(\bar{\mathbf{x}}^K) + \sum_{i\in\mathcal{N}} \|\theta_i^*\| d_{-\mathcal{K}_i}(G_i(\bar{x}_i^K)) \leq \frac{\Lambda(K)}{N_K}$

Solution error: $\|\mathbf{x}^K - \mathbf{x}^*\|^2 \leq 2\gamma^0 \, \frac{\tilde{\tau}^K}{\gamma^K} \, \Lambda(K)$

- $\Lambda(K) = \mathcal{O}(\sum_{k=1}^K \beta^{q_{k-1}} k^4)$ and $\sup_{K\geq 1} \Lambda(K) < \infty$
- $N_K = \mathcal{O}(K^2)$, $\tilde{\tau}^K/\gamma^K = \mathcal{O}(1/K^2)$ and $\kappa^K/\gamma^K = \mathcal{O}(1)$

Note: For static undirected $\mathcal{G}$, $q_k = 1$, $\Lambda(K) \leq \frac{|\mathcal{N}|\Delta}{2\tau^0} + \frac{1}{2\kappa^0}\sum_{i\in\mathcal{N}} \|\theta_i^0 - \theta_i^*\|^2$

Approximate Averaging Operator $\mathcal{R}^k(\cdot)$

We adopt the following information exchange models

Undirected $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Ozdaglar'09 and Chen & Ozdaglar'12

Directed $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Olshevsky'15

Approximate Averaging Operator $\mathcal{R}^k(\cdot)$

We adopt the following information exchange models

Undirected $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Ozdaglar'09 and Chen & Ozdaglar'12

Directed $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Olshevsky'15

- In one communication round, every node talks to its neighbors

Approximate Averaging Operator $\mathcal{R}^k(\cdot)$

We adopt the following information exchange models

Undirected $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Ozdaglar'09 and Chen & Ozdaglar'12

Directed $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Olshevsky'15

- In one communication round, every node talks to its neighbors
- communication among neighbors occurs instantaneously

Approximate Averaging Operator $\mathcal{R}^k(\cdot)$

We adopt the following information exchange models

Undirected $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Ozdaglar'09 and Chen & Ozdaglar'12

Directed $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Olshevsky'15

- In one communication round, every node talks to its neighbors
- communication among neighbors occurs instantaneously
- Synchronous nodes: one communication round in every unit time

Approximate Averaging Operator $\mathcal{R}^k(\cdot)$

We adopt the following information exchange models

Undirected $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Ozdaglar'09 and Chen & Ozdaglar'12

Directed $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Olshevsky'15

- In one communication round, every node talks to its neighbors
- communication among neighbors occurs instantaneously
- Synchronous nodes: one communication round in every unit time
- q_k: $\#$ of communication rounds within the k-th DPDA iteration

Approximate Averaging Operator $\mathcal{R}^k(\cdot)$

We adopt the following information exchange models

Undirected $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Ozdaglar'09 and Chen & Ozdaglar'12

Directed $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Olshevsky'15

- In one communication round, every node talks to its neighbors
- communication among neighbors occurs instantaneously
- Synchronous nodes: one communication round in every unit time
- q_k: # of communication rounds within the k-th DPDA iteration
- t_k: # of communication rounds before the k-th DPDA iteration

Approximate Averaging Operator $\mathcal{R}^k(\cdot)$

We adopt the following information exchange models
Undirected $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Ozdaglar'09 and Chen & Ozdaglar'12
Directed $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Olshevsky'15

- In one communication round, every node talks to its neighbors
- communication among neighbors occurs instantaneously
- Synchronous nodes: one communication round in every unit time
- q_k: # of communication rounds within the k-th DPDA iteration
- t_k: # of communication rounds before the k-th DPDA iteration
- $V^t \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{N}|}$ be a mixing matrix for $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$:

Approximate Averaging Operator $\mathcal{R}^k(\cdot)$

We adopt the following information exchange models

Undirected $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Ozdaglar'09 and Chen & Ozdaglar'12

Directed $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Olshevsky'15

- In one communication round, every node talks to its neighbors
- communication among neighbors occurs instantaneously
- Synchronous nodes: one communication round in every unit time
- q_k: # of communication rounds within the k-th DPDA iteration
- t_k: # of communication rounds before the k-th DPDA iteration
- $V^t \in \mathbb{R}^{|\mathcal{N}|\times|\mathcal{N}|}$ be a mixing matrix for $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$:
  - $\forall\ i \in \mathcal{N}, \qquad V^t_{ij} = 0 \quad \Leftrightarrow \quad j \notin \mathcal{N}^t_i$
  - $\sum_{j\in\mathcal{N}^t_i} V^t_{ij} x^t_j$ can be computed at $i \in \mathcal{N}$ with local communication

Approximate Averaging Operator $\mathcal{R}^k(\cdot)$

We adopt the following information exchange models

Undirected $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Ozdaglar'09 and Chen & Ozdaglar'12

Directed $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$: Nedich & Olshevsky'15

- In one communication round, every node talks to its neighbors
- communication among neighbors occurs instantaneously
- Synchronous nodes: one communication round in every unit time
- q_k: # of communication rounds within the k-th DPDA iteration
- t_k: # of communication rounds before the k-th DPDA iteration
- $V^t \in \mathbb{R}^{|\mathcal{N}| \times |\mathcal{N}|}$ be a mixing matrix for $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$:
  - $\forall\ i \in \mathcal{N}, \qquad V^t_{ij} = 0 \quad \Leftrightarrow \quad j \notin \mathcal{N}^t_i$
  - $\sum_{j \in \mathcal{N}^t_i} V^t_{ij} x^t_j$ can be computed at $i \in \mathcal{N}$ with local communication
- $W^{t,s} \triangleq V^t V^{t-1} ... V^{s+1}$ for $t \geq s+1$

$\mathcal{R}^k(\cdot)$ for directed communication networks

$\mathcal{R}^k(\mathbf{w}) = [\mathcal{R}_i^k(\mathbf{w})]_{i\in\mathcal{N}}$ s.t. $\|\mathcal{P}_{\mathcal{C}}(\mathbf{w}) - \mathcal{R}^k(\mathbf{w})\| = \mathcal{O}(\beta^{q_k}\|\mathbf{w}\|) \quad \forall \mathbf{w},\ k \geq 0$

Definition:

- $\mathcal{N}_i^{t,\mathrm{out}} \triangleq \{j \in \mathcal{N} :\ (i,j) \in \mathcal{E}^t\} \cup \{i\}$ and $d_i^t \triangleq |\mathcal{N}_i^{t,\mathrm{out}}|$
- $\mathcal{N}_i^{t,\mathrm{in}} \triangleq \{j \in \mathcal{N} :\ (j,i) \in \mathcal{E}^t\} \cup \{i\}$
- $V^t \in \mathbb{R}^{|\mathcal{N}|\times|\mathcal{N}|}$ s.t. $V_{i,j}^t = 1/d_i^t$ for $j \in \mathcal{N}_i^{t,\mathrm{in}}$ and 0 otherwise.

$\mathcal{R}^k(\cdot)$ for directed communication networks

$\mathcal{R}^k(\mathbf{w}) = [\mathcal{R}_i^k(\mathbf{w})]_{i\in\mathcal{N}}$ s.t. $\|\mathcal{P}_{\mathcal{C}}(\mathbf{w}) - \mathcal{R}^k(\mathbf{w})\| = \mathcal{O}(\beta^{q_k}\|\mathbf{w}\|) \quad \forall \mathbf{w},\ k \geq 0$

Definition:

- $\mathcal{N}_i^{t,\mathrm{out}} \triangleq \{j \in \mathcal{N} :\ (i,j) \in \mathcal{E}^t\} \cup \{i\}$ and $d_i^t \triangleq |\mathcal{N}_i^{t,\mathrm{out}}|$
- $\mathcal{N}_i^{t,\mathrm{in}} \triangleq \{j \in \mathcal{N} :\ (j,i) \in \mathcal{E}^t\} \cup \{i\}$
- $V^t \in \mathbb{R}^{|\mathcal{N}|\times|\mathcal{N}|}$ s.t. $V_{i,j}^t = 1/d_i^t$ for $j \in \mathcal{N}_i^{t,\mathrm{in}}$ and 0 otherwise.

Assumption: $\exists\ M > 1$ s.t. $(\mathcal{N}, \mathcal{E}_{k,M})$ is strongly connected for $k \geq 1$, $\mathcal{E}_{k,M} \triangleq \bigcup_{t=kM}^{(k+1)M-1} \mathcal{E}^t$.

$\mathcal{R}^k(\cdot)$ for directed communication networks

$\mathcal{R}^k(\mathbf{w}) = [\mathcal{R}_i^k(\mathbf{w})]_{i\in\mathcal{N}}$ s.t. $\|\mathcal{P}_{\mathcal{C}}(\mathbf{w}) - \mathcal{R}^k(\mathbf{w})\| = \mathcal{O}(\beta^{q_k}\|\mathbf{w}\|) \quad \forall \mathbf{w},\ k \geq 0$

Definition:

- $\mathcal{N}_i^{t,\mathrm{out}} \triangleq \{j \in \mathcal{N} :\ (i,j) \in \mathcal{E}^t\} \cup \{i\}$ and $d_i^t \triangleq |\mathcal{N}_i^{t,\mathrm{out}}|$
- $\mathcal{N}_i^{t,\mathrm{in}} \triangleq \{j \in \mathcal{N} :\ (j,i) \in \mathcal{E}^t\} \cup \{i\}$
- $V^t \in \mathbb{R}^{|\mathcal{N}|\times|\mathcal{N}|}$ s.t. $V_{i,j}^t = 1/d_i^t$ for $j \in \mathcal{N}_i^{t,\mathrm{in}}$ and 0 otherwise.

Assumption: $\exists\ M > 1$ s.t. $(\mathcal{N}, \mathcal{E}_{k,M})$ is strongly connected for $k \geq 1$, $\mathcal{E}_{k,M} \triangleq \bigcup_{t=kM}^{(k+1)M-1} \mathcal{E}^t$.

Lemma: $\exists \beta \in (0,1)$ and $\Gamma > 0$ s.t. for any $t > s \geq 0$ and $\mathbf{w} = [w_i]_{i\in\mathcal{N}}$

$$\left\| \left(\mathbf{diag}\left(W^{t,s}\mathbf{1}\right)^{-1} W^{t,s} \otimes \mathbf{I}_m\right)\mathbf{w} - \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} w_i \right\| \leq \Gamma \beta^{t-s}\|\mathbf{w}\|$$

$\mathcal{R}^k(\cdot)$ for directed communication networks

$\mathcal{R}^k(\mathbf{w}) = [\mathcal{R}_i^k(\mathbf{w})]_{i\in\mathcal{N}}$ s.t. $\|\mathcal{P}_{\mathcal{C}}(\mathbf{w}) - \mathcal{R}^k(\mathbf{w})\| = \mathcal{O}(\beta^{q_k}\|\mathbf{w}\|) \quad \forall \mathbf{w},\ k \geq 0$

Definition:

- $\mathcal{N}_i^{t,\text{out}} \triangleq \{j \in \mathcal{N} :\ (i,j) \in \mathcal{E}^t\} \cup \{i\}$ and $d_i^t \triangleq |\mathcal{N}_i^{t,\text{out}}|$
- $\mathcal{N}_i^{t,\text{in}} \triangleq \{j \in \mathcal{N} :\ (j,i) \in \mathcal{E}^t\} \cup \{i\}$
- $V^t \in \mathbb{R}^{|\mathcal{N}|\times|\mathcal{N}|}$ s.t. $V_{i,j}^t = 1/d_i^t$ for $j \in \mathcal{N}_i^{t,\text{in}}$ and 0 otherwise.

Assumption: $\exists\ M > 1$ s.t. $(\mathcal{N}, \mathcal{E}_{k,M})$ is strongly connected for $k \geq 1$, $\mathcal{E}_{k,M} \triangleq \bigcup_{t=kM}^{(k+1)M-1} \mathcal{E}^t$.

Lemma: $\exists \beta \in (0,1)$ and $\Gamma > 0$ s.t. for any $t > s \geq 0$ and $\mathbf{w} = [w_i]_{i\in\mathcal{N}}$

$$\left\| \left(\mathbf{diag}\left(W^{t,s}\mathbf{1}\right)^{-1} W^{t,s} \otimes \mathbf{I}_m\right)\mathbf{w} - \frac{1}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} w_i \right\| \leq \Gamma\beta^{t-s}\|\mathbf{w}\|$$

Hence, $\mathcal{R}^k(\mathbf{w}) \triangleq \left(\mathbf{diag}\left(W^{t_k+q_k,t_k}\mathbf{1}_{|\mathcal{N}|}\right)^{-1} W^{t_k+q_k,t_k} \otimes \mathbf{I}_m\right)\mathbf{w}$

Numerical Experiments

Distributed Isotonic LASSO:

- $\mathbf{x} \in \mathbb{R}^{n|\mathcal{N}|}$, $C = [C_i]_{i\in\mathcal{N}} \in \mathbb{R}^{m|\mathcal{N}|\times n}$,
- $d = [d_i]_{i\in\mathcal{N}} \in \mathbb{R}^{m|\mathcal{N}|}$, and $A \in \mathbb{R}^{n-1\times n}$

$$\min_{\substack{\mathbf{x}=[x_i]_{i\in\mathcal{N}}\in\mathcal{C},\\ Ax_i\leq\mathbf{0}\ i\in\mathcal{N}}} \ \frac{1}{2}\sum_{i\in\mathcal{N}} \|C_i x_i - d_i\|^2 + \frac{\lambda}{|\mathcal{N}|}\sum_{i\in\mathcal{N}} \|x_i\|_1,$$

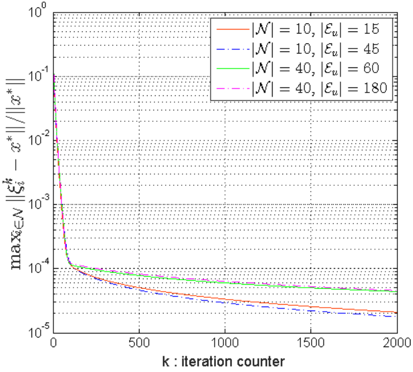
- $n = 20$, $m = n + 2$
- Random C_i with standard Gaussian entries, $d_i = C_i(x^\circ + \epsilon)$
- $\epsilon \in \mathbb{R}^n$ random with Gaussian of zero mean and std. deviation 10^{-3}
- Random $x^\circ \in \mathbb{R}^{n-1}$:

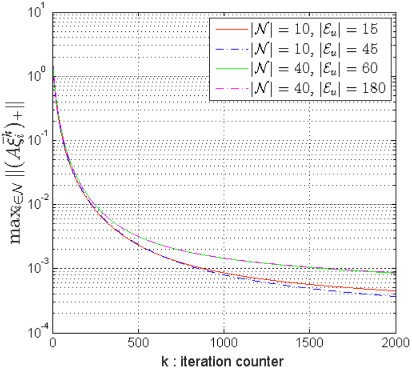
$$x^\circ = \underbrace{\left[\ \underbrace{U[-10,0]^5}_{\text{first 5 components}}\ ,\ \underbrace{0,\ 0,\ \ldots,\ 0,}_{n-11},\ \underbrace{U[0,10]^5}_{\text{last 5 components}}\ \right]^\top}_{\text{ascending order}}$$

Numerical Experiments

- $\mathcal{G}_0 = (\mathcal{N}, \mathcal{E}_0)$ small-world network
- $\mathcal{G}^t = (\mathcal{N}, \mathcal{E}^t)$ generated by sampling 80% of $\mathcal{E}_0$ s.t. $M = 5$

Effect of Network Topology (time-varying undirected):

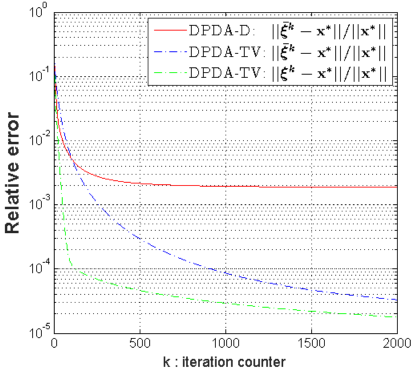


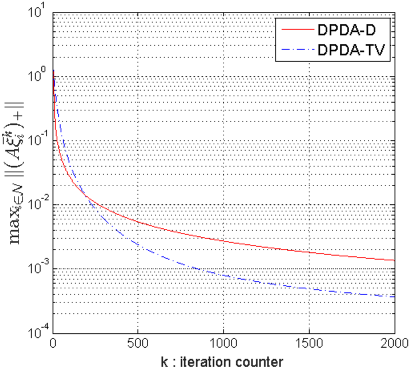


Numerical Experiments

Compared against DPDA-D (Aybat et al.'16) – $\mathcal{O}(1/k)$ ergodic rate

- Time-varying undirected Network: $\mathcal{G}_u = (\mathcal{N}, \mathcal{E}_u)$, $|\mathcal{N}| = 10$, $|\mathcal{E}_u| = 45$
- $\mathcal{G}_u^t = (\mathcal{N}, \mathcal{E}_u^t)$ generated by sampling 80% of $\mathcal{E}_u$ s.t. $M = 5$





Numerical Experiments

Compared against DPDA-D (Aybat et al.'16) – $\mathcal{O}(1/k)$ ergodic rate

- Time-varying directed Network: $\mathcal{G}_d = (\mathcal{N}, \mathcal{E}_d)$, $|\mathcal{N}| = 12$, $|\mathcal{E}_d| = 24$
- $\mathcal{G}_d^t = (\mathcal{N}, \mathcal{E}_d^t)$ generated by sampling 80% of $\mathcal{E}_d$ s.t. $M = 5$ (Nedich et al.'17)