Distributed Asynchronous Solution of Locally Coupled Optimization Problems on Agent Networks

April 28, 2017

Jianghai Hu
School of Electrical and Computer Engineering
Purdue University

Joint work with Yingying Xiao (Purdue) and Ji Liu (Stony Brook Univ)

PURDUE UNIVERSITY

Overview

- Problem formulation
- Synchronous distributed solution algorithms
- Asynchronous distributed solution algorithms
- Examples
- Extensions

Optimization Problems on Agent Networks

- A set of agents, each with a local variable and a local optimization problem
- Local problems are coupled

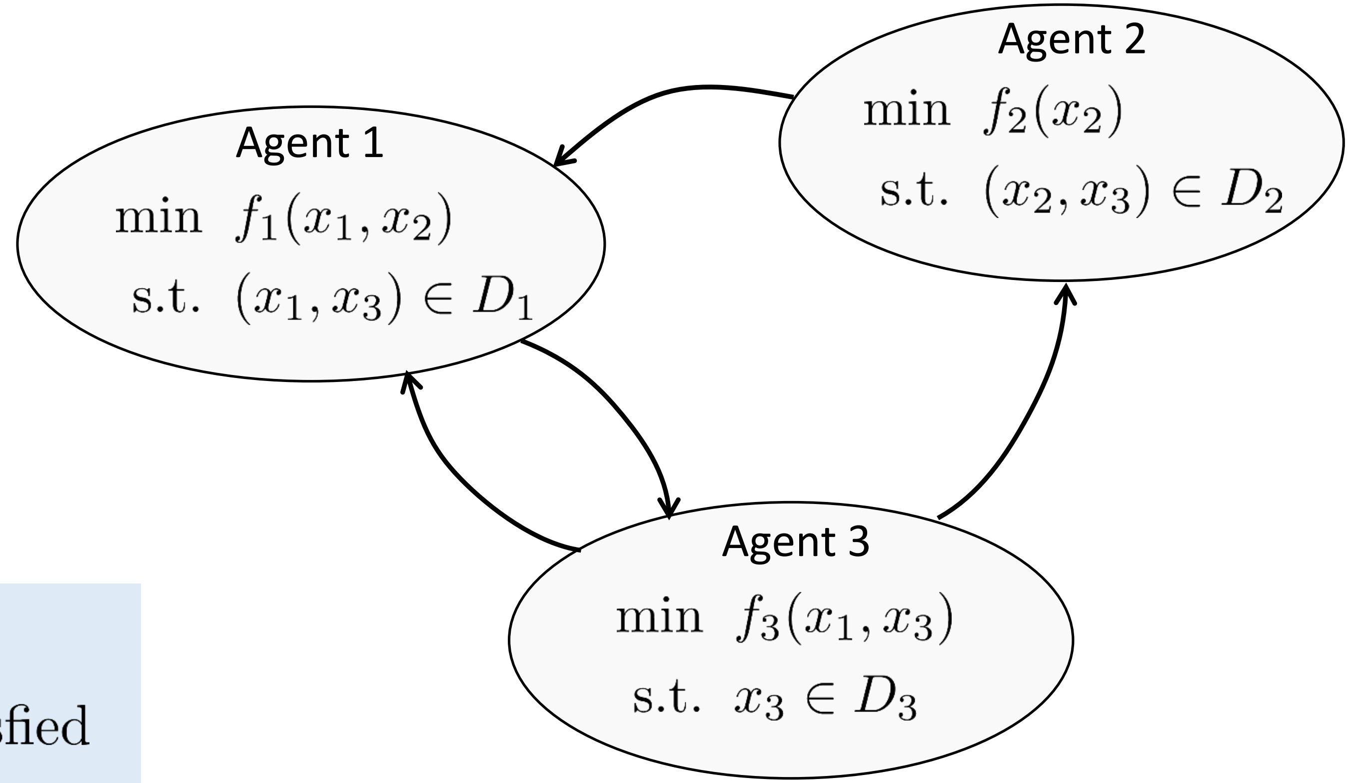
Agent 1

- Local variable x_1
- In-neighbors $\mathcal{N}_1^+ = \{2, 3\}$
- Out-neighbors $\mathcal{N}_1^- = \{3\}$

Global optimization problem

$$\text{mininize } f = f_1 + \cdots + f_m$$
$$\text{s.t. all local constraints are satisfied}$$

dependency graph



Special Cases

- Consensus optimization

$$\min\ f_1(x_1) + \cdots + f_m(x_m) \ \text{ s.t. } \ x_1 = \cdots = x_m$$

- Common feasibility problem

$$\text{find } x \in D_1 \cap D_2 \cap \cdots \cap D_m$$

- Hierarchical optimization

$$\min\ f_1(x_1) + \cdots + f_m(x_m) + f_0(x_1, \ldots, x_m)$$

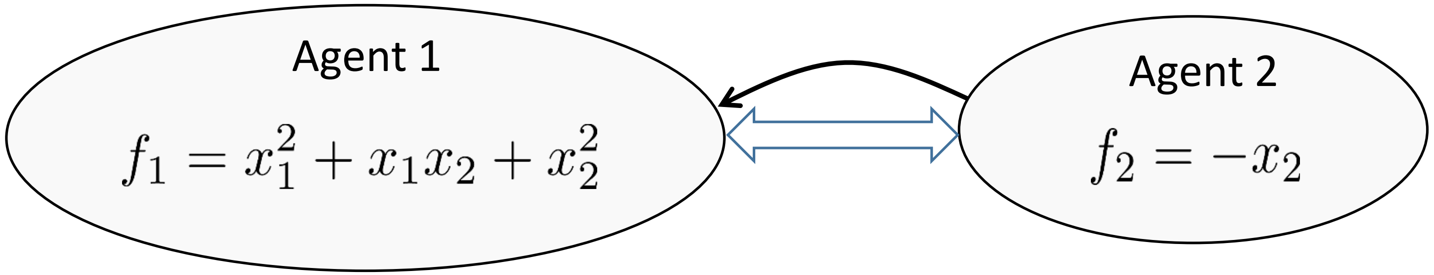
Assumptions

Convexity:

- f_i are extended-valued, closed, convex, proper (CCP) functions
- D_i are nonempty convex sets, with (convex) indicator functions $\mathbf{1}_{D_i}$

Existence of solution: global optimization problem has solutions

Communications: Neighboring agents can exchange information both ways



Relevant Approaches

- Primal approaches
  - Subgradient descent plus consensus step (e.g. [Nedic et al'09&10], ...)
  - Projected subgradient method (e.g. [Figueiredo et al'07], ...)
  - Proximal subgradient method (e.g. [Nesterov'83&07],[Shi et al'15], ...)

- Primal-dual approaches
  - Dual decomposition (e.g. [Terelius et al'11])
  - ADMM algorithms (e.g., [Gabay&Mercier'83],[Boyd et al'11],...)

- Operator splitting techniques ([Bauschke&Combettes'16])

Objectives of Our Approach

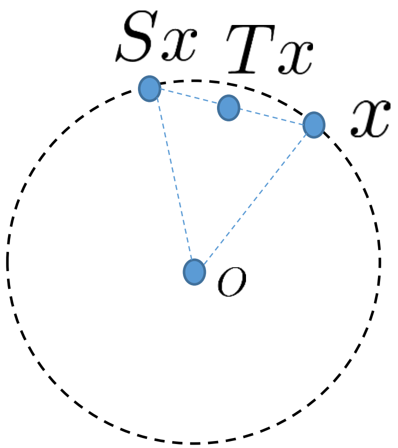
Find iteration algorithms $x^{k+1} = Tx^k$ for some operator T such that

- Fixed points of T are exactly the optimal solutions
- Starting from any x^0, $x^k \to x^* \in \mathrm{Fix}(T)$

Desired features:

- Can handle arbitrary dependency graph and state partition $x = (x_1, \ldots, x_m)$
- Can handle general convex cost function and constraints
- Distributed implementation with minimal inter-agent communications
- Can be adapted for asynchronous implementations

Averaged Operators



An operator $S : \mathbb{R}^n \to \mathbb{R}^n$ is nonexpansive if

$$\|Sx - Sy\| \leq \|x - y\|, \forall x, y$$

- May not converge to a fixed point (if exsits). e.g. a rotation

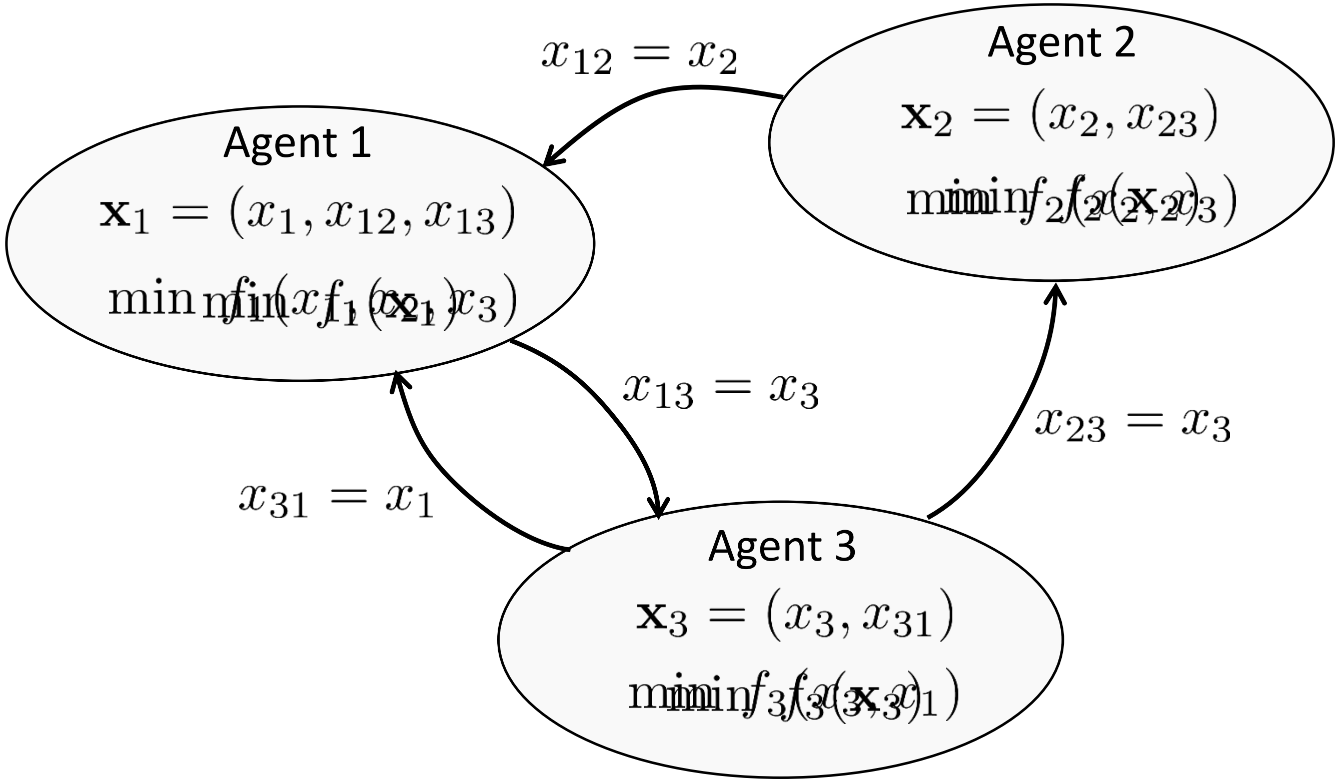
T is averaged if $T = (1-\alpha)I + \alpha S$ for a nonexpansive S and $\alpha \in (0,1)$

- Convergence to a fixed point (if exists) is guaranteed
- For any fixed point $x^* \in \mathrm{Fix}(T)$ (if exists) and any x

$$\|Tx - x^*\|^2 \leq \|x - x^*\|^2 - \frac{1-\alpha}{\alpha}\|Tx - x\|^2,$$

Problem Reformulation

- Augment agent *i*'s variable to $\mathbf{x}_i = \left(x_i, (x_{ij})_{j \in \mathcal{N}_i^+}\right)$ where x_{ij} is a local copy of x_j
- Recast local cost as $f_i(\mathbf{x}_i)$
- Impose consensus constraints



Global optimization problem

$$\min \; f(\mathbf{x}) = f_1(\mathbf{x}_1) + \cdots + f_m(\mathbf{x}_m) \quad \text{s.t. } \mathbf{x} \in \mathcal{A}$$

$$\min \; f(\mathbf{x}) + \mathbf{1}_{\mathcal{A}}(\mathbf{x})$$

Generalized consensus subspace:

$$\mathcal{A} := \bigcap_i \left\{ \mathbf{x} \,|\, x_i = x_{ji}, \; \forall j \in \mathcal{N}_i^- \right\}$$

Proximal Operators

For an extended-valued, CCP function $g(x)$, its proximal operator is $(\rho > 0)$

$$\mathrm{prox}_{\rho g}(x) = \arg\min_{z} \left(g(z) + \frac{1}{2\rho}\|z - x\|^2 \right)$$

- $2 \cdot \mathrm{prox}_{\rho g} - I$ is nonexpansive, hence $\mathrm{prox}_{\rho g}$ is (1/2)–averaged
- Fixed points of $\mathrm{prox}_{\rho g}$ are the minimizers of $g(x)$
- Proximal point algorithm ([Rockafellar'76]): $x^{k+1} = \mathrm{prox}_{\rho g}(x^k)$

Many common $g(x)$ are "proximable"

- Proximal operator of $f(\mathbf{x}) = f_1(\mathbf{x}_1) + \cdots + f_m(\mathbf{x}_m)$ is the product of $\mathrm{prox}_{\rho f_i}$
- Proximal operator of $\mathbf{1}_{\mathcal{A}}$ is the projection $\Pi_{\mathcal{A}}$ onto $\mathcal{A}$

Operator Splitting

Goal: find the minimizers of $f(x) + g(x)$ for proximable $f(x)$ and $g(x)$

Douglas-Rachford Splitting: [Douglas&Rachford'56]

1. Find a fixed point z^* of the nonexpansive map $S = (2\,\mathrm{prox}_{\rho f} - I)(2\,\mathrm{prox}_{\rho g} - I)$
2. Output $x^* = \mathrm{prox}_{\rho g}(z^*)$

- Step 1 can be accomplished by iterating the α-averaged operator:

$$T = (1-\alpha)I + \alpha S, \quad \alpha \in (0,1)$$

- Roles of f and g can be switched

Douglas-Rachford Algorithm

Goal: find the minimizers of $f(\mathbf{x}) + \mathbf{1}_{\mathcal{A}}(\mathbf{x})$

Algorithm: initialize $\mathbf{z}^0 = (\mathbf{z}_1^0, \ldots, \mathbf{z}_m^0)$

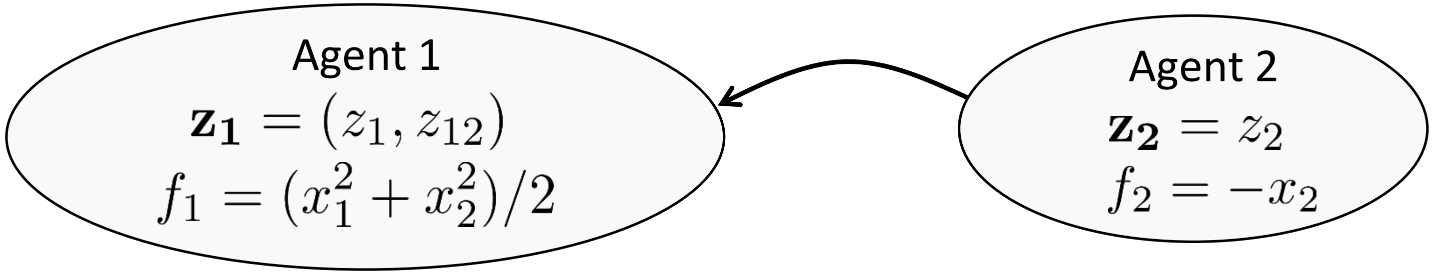
$$\mathbf{x}^{k+1} \leftarrow \Pi_{\mathcal{A}}(\mathbf{z}^k)$$

$$\mathbf{z}_i^{k+1} \leftarrow \mathbf{z}_i^k + 2\alpha \left(\mathrm{prox}_{\rho f_i}(2\mathbf{x}_i^{k+1} - \mathbf{z}_i^k) - \mathbf{x}_i^{k+1}\right), \quad \forall i$$

Output: $\mathbf{x}^k$

Theorem: $\mathbf{x}^k$ converges to an optimal solution $\mathbf{x}^*$ for any $\rho > 0$, $\alpha \in (0, 1)$

Example



- D-R algorithm with $\alpha = \frac{1}{2}$

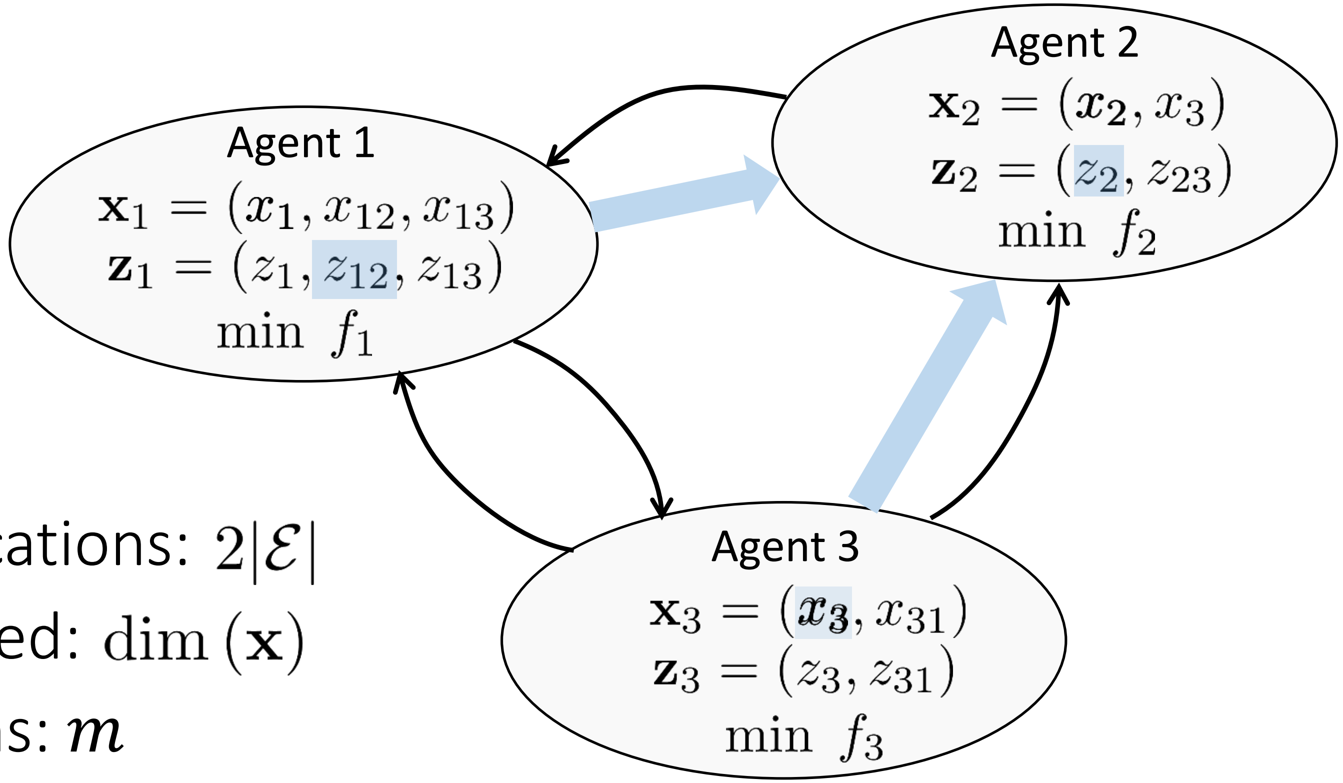
$$\begin{cases} z_1^{k+1} & = (1 - \rho/(1+\rho)) z_1^k, \\ z_{12}^{k+1} & = \frac{1}{2} z_{12}^k + \frac{1}{2}(1-\rho)/(1+\rho) z_2^k, \\ z_2^{k+1} & = (z_2^k + z_{12}^k)/2 + \rho. \end{cases}$$

- $z^k \to z^*$ with $z_1^* = 0,\ z_{12}^* = 1 - \rho,\ z_2^* = 1 + \rho$
- $x^* = \Pi_{\mathcal{A}} z^*$ with $x_1^* = 0,\ x_2^* = x_{12}^* = 1,$ is an optimal solution

Algorithm Complexity

$$\mathbf{x}^{k+1} \leftarrow \Pi_{\mathcal{A}}(\mathbf{z}^k)$$

$$\mathbf{z}_i^{k+1} \leftarrow \mathbf{z}_i^k + 2\alpha \left(\mathrm{prox}_{\rho f_i}(2\mathbf{x}_i^{k+1} - \mathbf{z}_i^k) - \mathbf{x}_i^{k+1}\right), \quad \forall i$$



In each round

- Total number of one-way communications: $2|\mathcal{E}|$
- Total number of variables transmitted: $\dim(\mathbf{x})$
- Total number of proximal evaluations: m

Dual Douglas-Rachford Algorithm

Dual Problem: Let $f^* = f_1^* + \cdots + f_m^*$ be the convex conjugate of f

$$\text{minimize } f^*(\mathbf{p}) + \mathbf{1}_{\mathcal{A}^\perp}(\mathbf{p})$$

- Moreau's decomposition relates $\text{prox}_{\rho f^*}$ to $\text{prox}_{\rho f}$

Algorithm: initialize $\mathbf{w}^0 = (\mathbf{w}_1^0, \ldots, \mathbf{w}_m^0)$

$$\mathbf{u}^{k+1} \leftarrow \Pi_{\mathcal{A}}(\mathbf{w}^k)$$

$$\mathbf{w}_i^{k+1} \leftarrow \mathbf{w}_i^k - 2\alpha \mathbf{u}_i^{k+1} - 2\alpha\rho^{-1}\text{prox}_{\rho f_i}(\rho \mathbf{w}_i^k - 2\rho \mathbf{u}_i^{k+1}), \quad \forall i$$

Output: $\Pi_{\mathcal{A}^\perp} \mathbf{w}^k$

Theorem: $\Pi_{\mathcal{A}^\perp} \mathbf{w}^k$ converges to a dual solution $\mathbf{p}^*$ for any $\rho > 0,\ \alpha \in (0,1)$

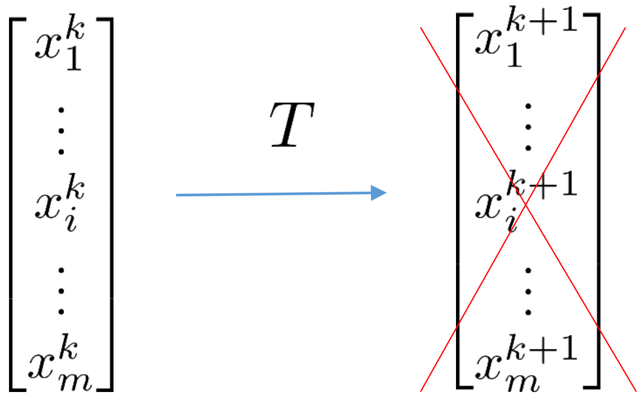
Asynchrony in Agent Networks

- Previous algorithms require multiple synchronized operations in a round
- Full synchronization may be costly or unrealistic
  - No central agent coordinating the computation
  - Heterogeneous agent computation powers and proximability
  - Blackout of agents and communication links

Asynchronous Implementation of Averaged Operators

Iteration $x^{k+1} = Tx^k$ using an averaged operator $T : \mathbb{R}^n \to \mathbb{R}^n$

- Block coordinate decomposition:



Random coordinate update

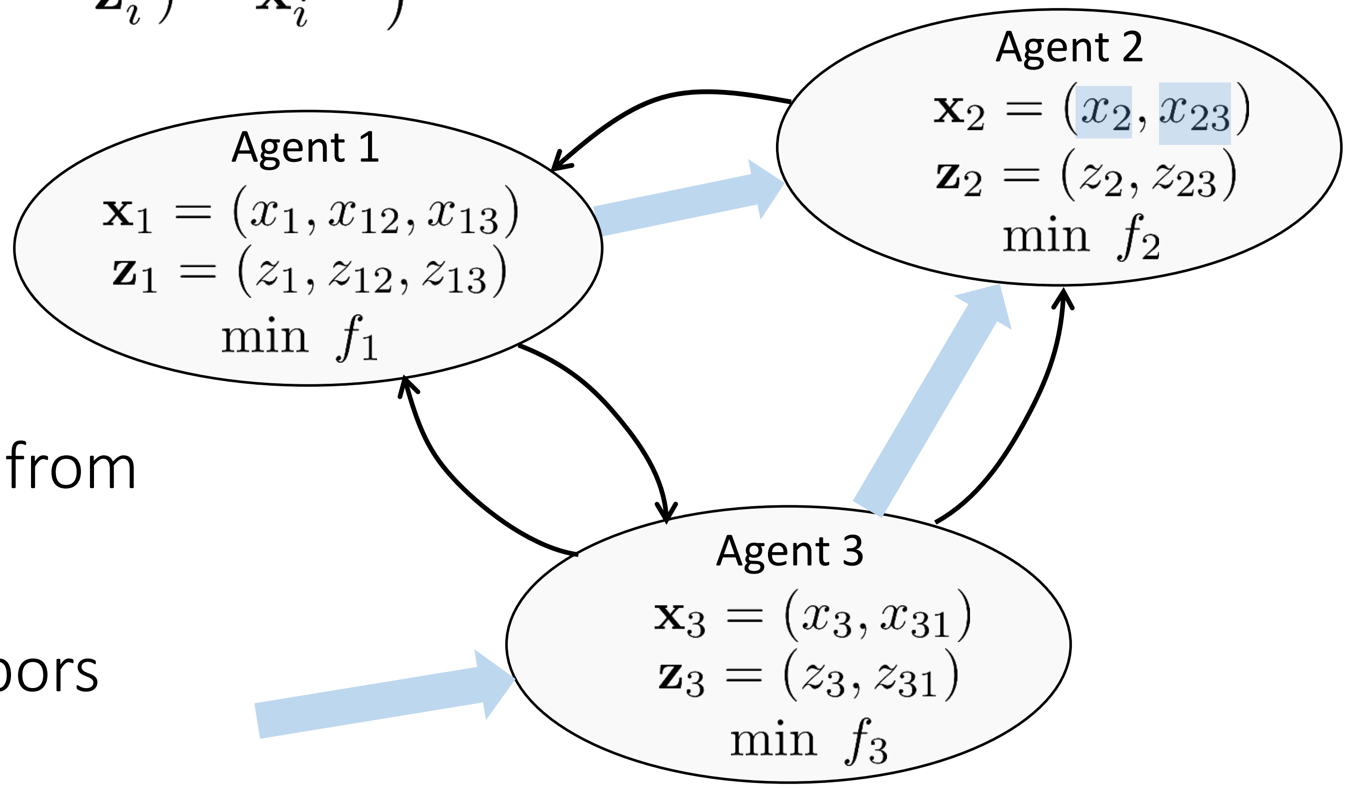
- At each round randomly activate a block i with probability p_i to update
- Under some ergodicity assumption, the iteration converges to a fixed point of T with probability one ([Wei&Ozdaglar'13] [Bianci et al'16])

Asynchronous D-R Algorithm

At each round, activate an agent i randomly with probability p_i and do

$$\mathbf{x}_i^{k+1} \leftarrow \Pi_{\mathcal{A}}(\mathbf{z}^k)$$

$$\mathbf{z}_i^{k+1} \leftarrow \mathbf{z}_i^k + 2\alpha\left(\mathrm{prox}_{\rho f_i}(2\mathbf{x}_i^{k+1} - \mathbf{z}_i^k) - \mathbf{x}_i^{k+1}\right)$$



Activated agent collects information from

- Its out-neighbors
- Its in-neighbors and their in-neighbors

Modified Asynchronous D-R Algorithm

Each agent i maintains an extra variable $\bar{z}_i$, the consensus value of z_i

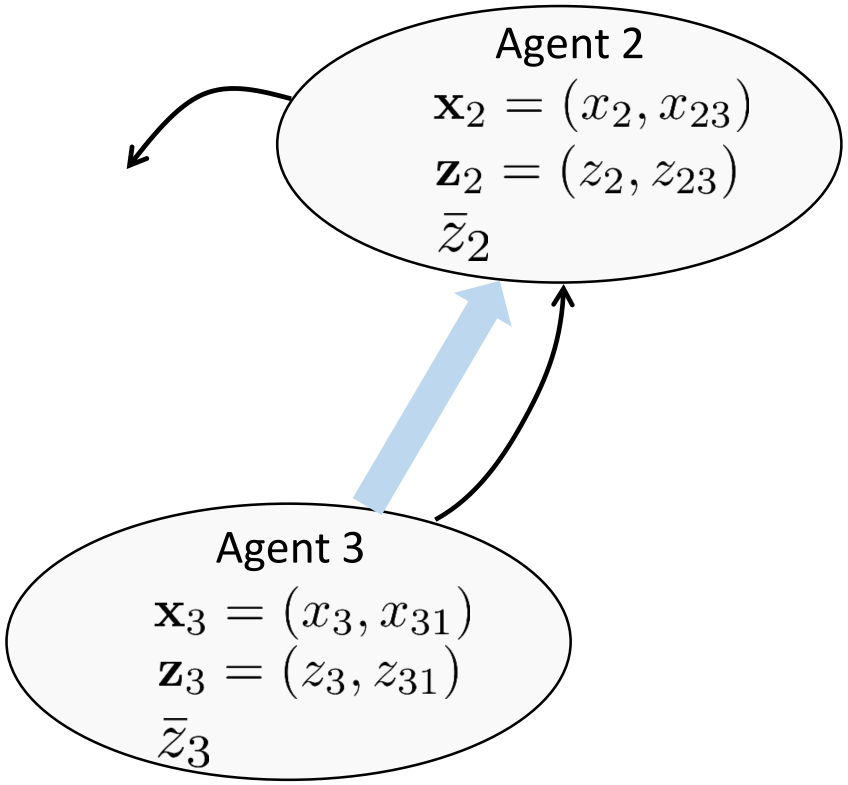
Activate randomly an agent i at each round and do

$$x_{ij}^{k+1} \leftarrow \bar{z}_j^k, \; \forall j \in \mathcal{N}_i^+$$

$$\mathbf{z}_i^{k+1} \leftarrow \mathbf{z}_i^k + 2\alpha \left(\text{prox}_{\rho f_i}(2\mathbf{x}_i^{k+1} - \mathbf{z}_i^k) - \mathbf{x}_i^{k+1} \right)$$

$$\bar{z}_i^{k+1} \leftarrow \bar{z}_i^k + (z_i^{k+1} - z_i^k)/(|\mathcal{N}_i^-| + 1)$$

$$\bar{z}_j^{k+1} \leftarrow \bar{z}_j^k + (z_{ij}^{k+1} - z_{ij}^k)/(|\mathcal{N}_j^-| + 1), \; \forall j \in \mathcal{N}_i^+$$



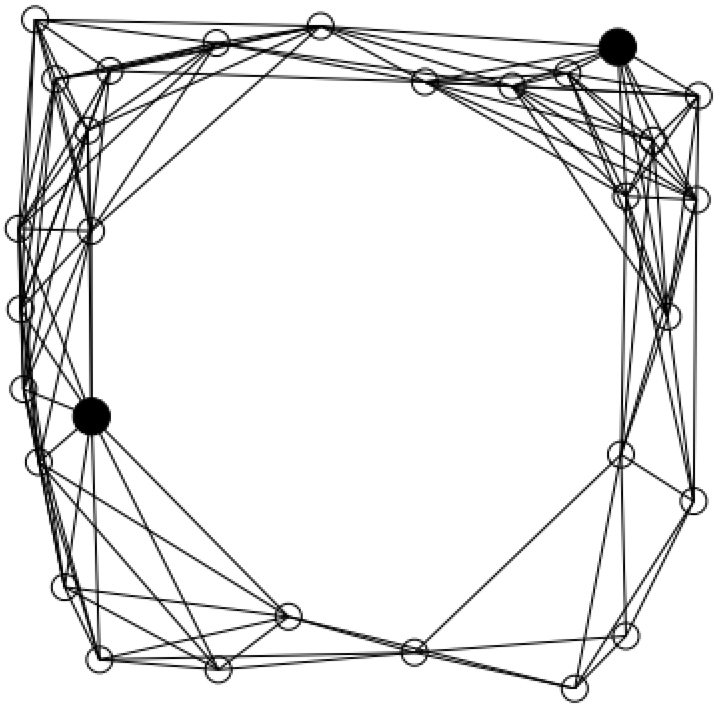
At each round

- Activated agent only communicates with in-neighbors
- Expected number of transmissions: $\sum_i 2p_i|\mathcal{N}_i^+|$

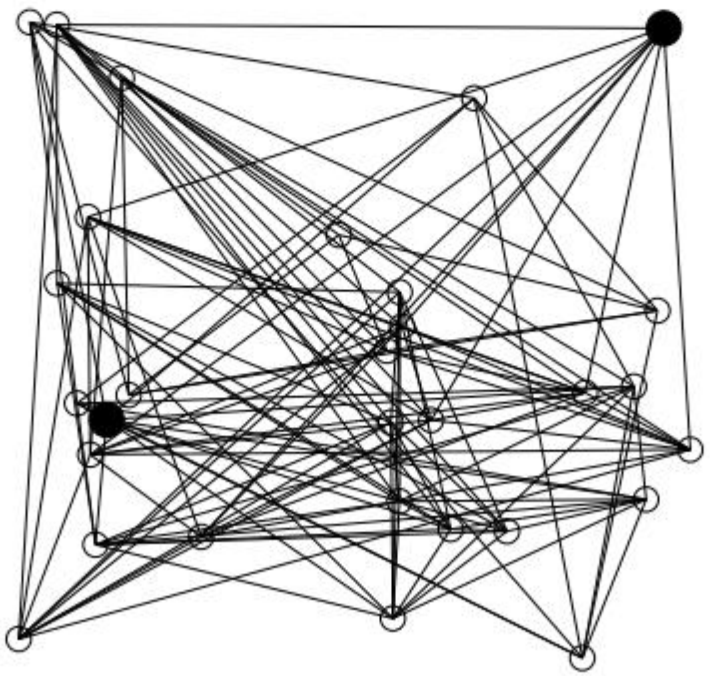
Example: Network Localization

- 28 agents with unknown positions and two anchors
- Each edge is a constraint on the relative orientation of two agents

Ground Truth

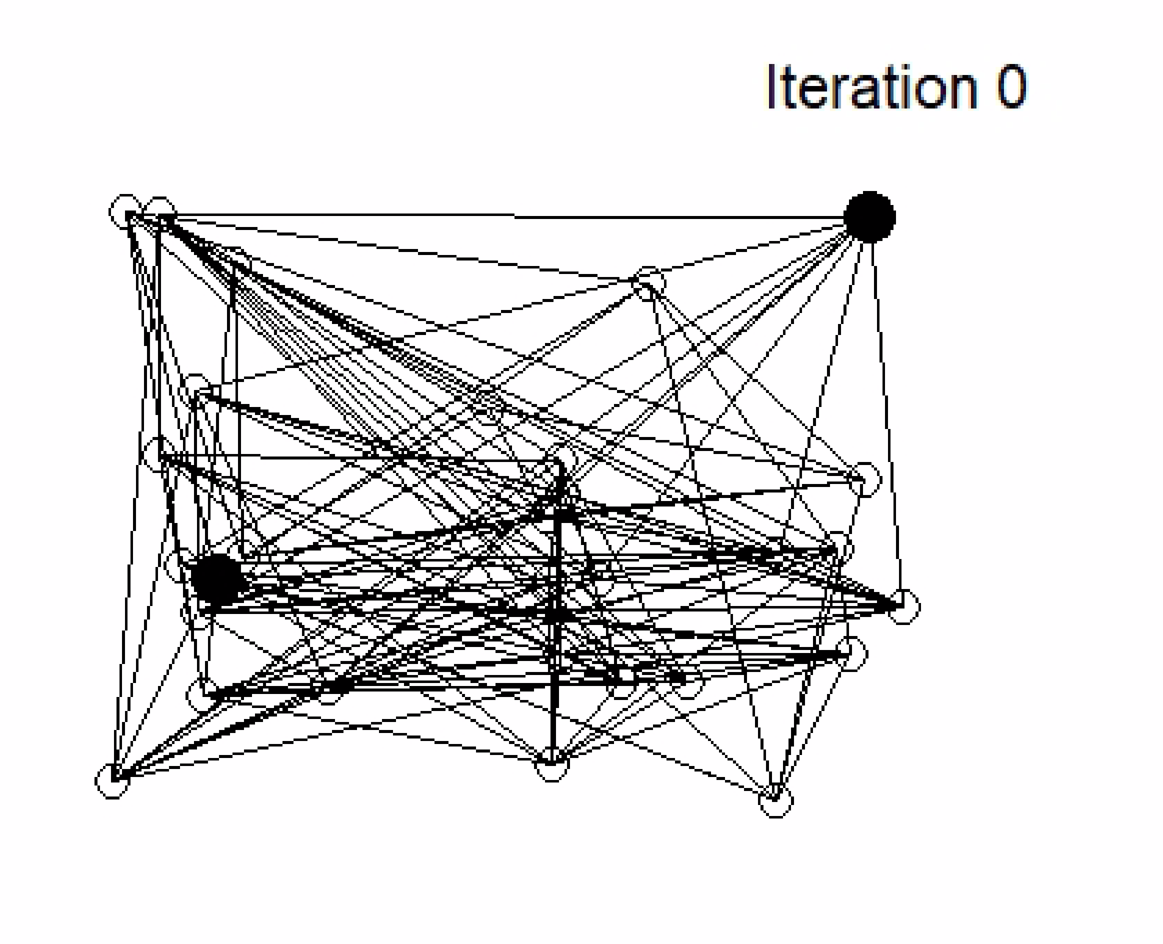


Random Initial Guess

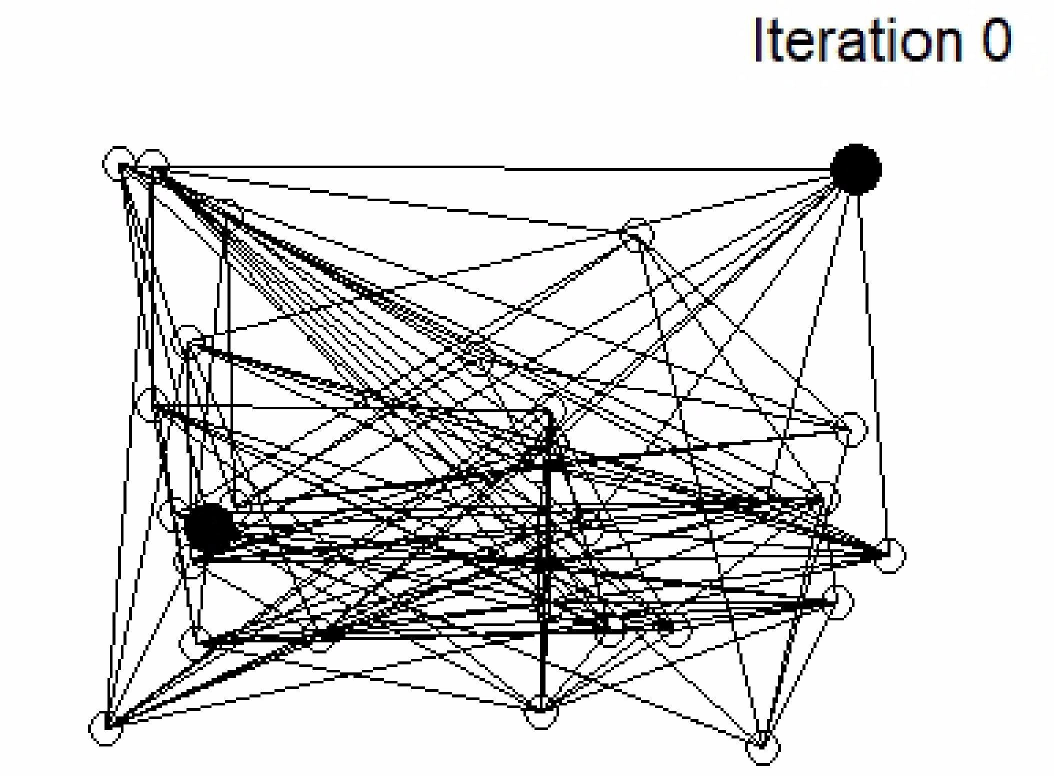


Example: Network Localization

Synchronous Algorithm



Asynchronous Algorithm



Extensions

- Local costs $f_i = g_i + h_i$ with proximable g_i, h_i
  - 3-operator splitting [Davis&Yin'15], Condat-Vu Algorithm [Condat'13] [Vu'13]
- Communication delays (e.g. ARock Algorithm [Peng et al'16])
- One-way communications on dependency graph
- Asynchronous implementation with general activation rules
- Nonconvex problems