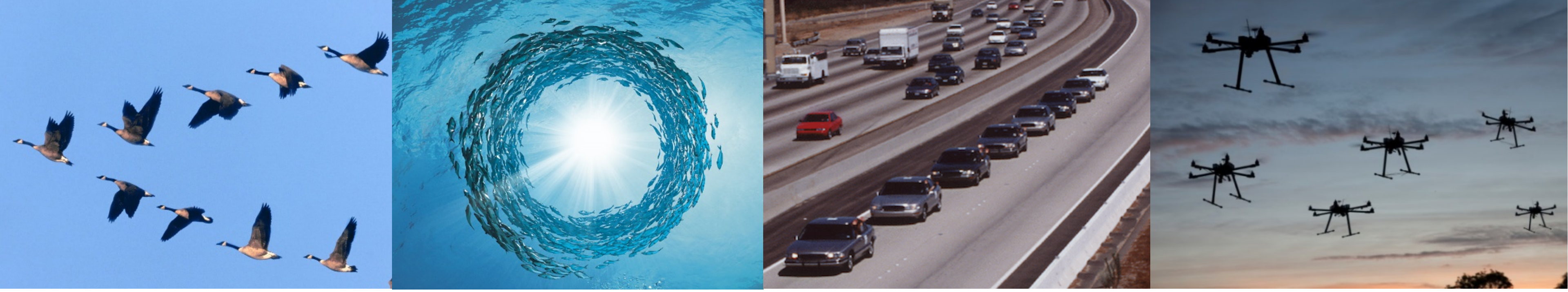
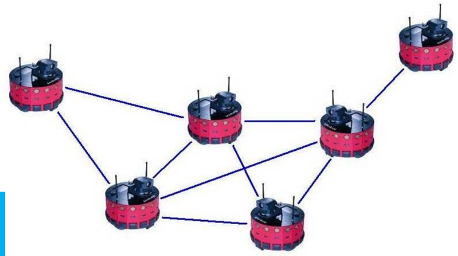
Lecture: Distributed Algorithms for Multi-Agent Optimization

Shaoshuai Mou

Assistant Professor
School of Aeronautics and Astronautics
Purdue University

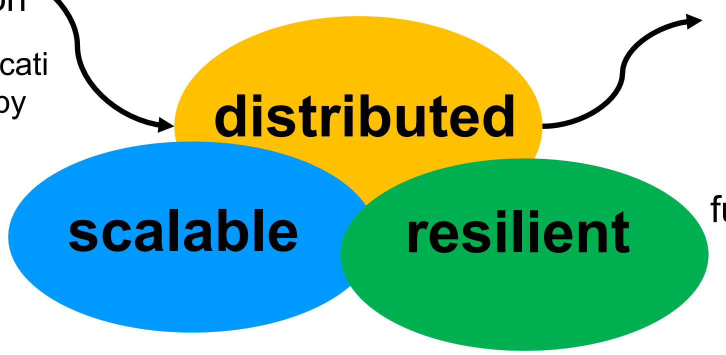


Multi-Agent Systems (MAS)



- **Features of MAS**:
  - Each individual: autonomous | low-cost | local accessibility
  - The whole team: large-scale | communication constraints | no centralized controller

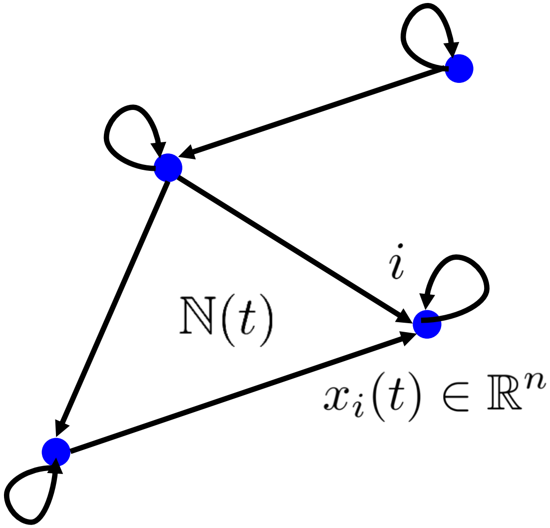
- **Algorithms for MAS:** **local** coordination (sensing/communications among nearby neighbors) → **global** objectives (exploration of the unknown, formation flight, solving large linear equations…)



scalable: with respect to mission complexity

resilient: system's ability to continue function under **sophisticated cyber-attacks**

- **Consensus:** *All agents reach an agreement on some quantity of interest.* (key enabler for MAS to work as a **cohesive whole)**



Consensus-based Distributed Optimization

(A.) *global objective* $\quad \min \sum_{i=1}^{m} f_i(x_i)$

subject to

(B.) *local constraint* $\quad x_i \in \mathcal{X}_i \subset \mathbb{R}^n$

(C.) *consensus* $\quad x_1 = x_2 = \cdots = x_m$

- **Consensus:** All agents reach the **same** value. **(C), well-studied**

$$x_i(t+1) = \sum_{j \in \mathcal{N}_i} w_{ij} x_j(t)$$

A. Jadbabaie, J. Lin, A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. IEEE Transactions on automatic control 48 (6), 988-1001, 2003

- **Constrained Consensus:** All agents reach the **same** and **specific** value. **(B)+ (C)**

  - Linear Case: $\quad \mathcal{X}_i = \{x \in \mathbb{R}^n : A_i x = b_i\}$

$$\overbrace{\boxed{x_i(t+1) = x_i(t) - P_i}}^{\text{Projection}} \left(\overbrace{\boxed{x_i(t) - \frac{1}{d_i(t)} \sum_{j \in \mathcal{N}_i(t)} x_j(t)}}^{\text{Consensus}} \right)$$

- S. Mou, J. Liu, A. S. Morse. A distributed algorithm for solving a linear algebraic equation. *IEEE Transactions on Automatic Control*, 2015, 60 (11), pp 2863-2878

- **Constrained Consensus:** All agents reach the **same** and **specific** value. **(B)+ (C)**

(B.) local constraint

$$x_i \in \mathcal{X}_i \subset \mathbb{R}^n$$

(C.) consensus

$$x_1 = x_2 = \cdots = x_m$$

- Non-Linear Case:

$$x_i(t+1) = \mathcal{P}_i\Big(\sum_{j \in \mathcal{N}_i} w_{ij} x_j(t)\Big)$$

Paracontraction
(to satisfy to the local constraint)

Consensus
*(for **combine** neighbors' information)*

A continuous nonlinear map $\mathcal{P} : \mathbb{R}^n \to \mathbb{R}^n$ is a ***paracontraction*** with respect to a given norm if

$$||\mathcal{P}(x) - y|| < ||x - y|| \quad \forall x \in \mathbb{R}^n, x \neq \mathcal{P}(x), \ \forall y$$

- **Projection** map is a paracontraction with respect to the 2-norm.

$$x \to \arg\min_{y \in \mathcal{X}} ||x - y||_2$$

- **Gradient** map is a paracontraction with respect to the 2-norm.

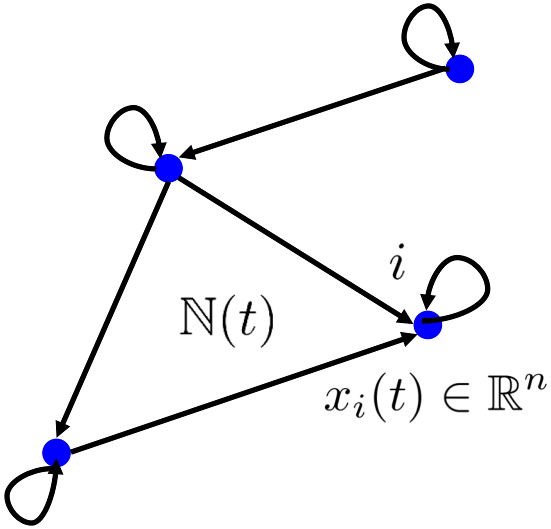
$$x \to x - \alpha \nabla f(x), \ 0 < \alpha < \frac{2}{\lambda}$$

- **Proximal** map is a paracontraction with respect to the 2-norm.

$$x \to \arg\min_{y \in \mathcal{X}} f(y) + \frac{1}{2}||x - y||_2$$

D. Fullmer, A. S. Morse. A distributed algorithm for computing a common fixed point of a finite family of paracontractions *IEEE Transactions on Automatic Control,* 2015, 60 (11), pp 2863-2878

Consensus-based Distributed Optimization



(A.) global objective $\quad \min \sum_{i=1}^{m} f_i(x_i)$

subject to

(B.) local constraint $\quad x_i \in \mathcal{X}_i \subset \mathbb{R}^n$

(C.) consensus $\quad x_1 = x_2 = \cdots = x_m$

- ***Constrained, Optimal Consensus:*** All agents reach the **same** and **specific** value, which ***minimizes*** the sum of local cost functions. **(A)+(B)+ (C)**

Not completely solved!

Key Idea for Consensus-based Distributed Optimization

$$\min \sum_{i=1}^{m} f_i(x_i)$$

subject to

Local Constraint

$$x_i \in \mathcal{X}_i$$

global consensus

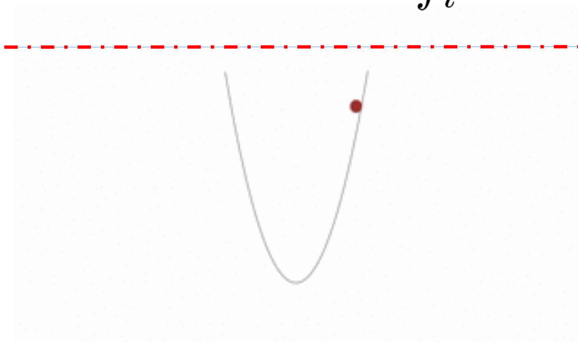
$$x_1 = x_2 = \cdots = x_m$$

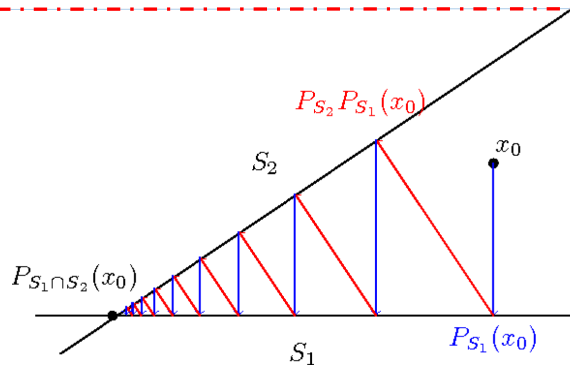
minimizes its own cost function f_i

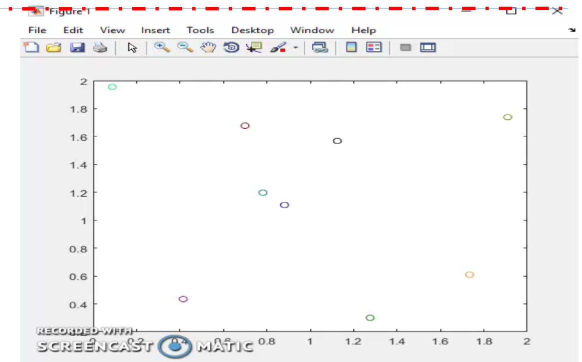
satisfies its own constraint $\mathcal{X}_i$

$$x^* \in \cap_{i=1}^{m} \mathcal{X}_i$$

reach a consensus with its nearby neighbors







Gradient Operator:

$$\mathcal{G}_f(v) = v - \alpha(t)\partial f(v)$$

Projection Operator

$$\mathcal{P}_{\mathcal{X}}(v) = \arg\min_{x \in \mathcal{X}} \|x - v\|$$

Consensus Operator:

$$\mathcal{S}(v_1, v_2, ..., v_r) = \sum_{i=1}^{r} w_{ij} v_j$$

$$y_i(t) = v_i(t) - \alpha(t)\nabla f_i(v_i(t))$$

Adapt

$$x_i(t+1) = \mathcal{P}_i\Big(y_i(t)\Big)$$

Project

$$v_i(t) = \sum_{j \in \mathcal{N}_i} w_{ij} x_j(t)$$

Combine

$$x_i(t+1) = \mathcal{P}_i\Big(\sum_{j \in \mathcal{N}_i} w_{ij} x_j(t) - \alpha(t)\nabla f_i(\sum_{j \in \mathcal{N}_i} w_{ij} x_j(t))\Big)$$

➢ An Updated for Consensus-based Distributed Optimization:

Gradient

$$x_i(t+1) = \mathcal{P}_i\left(\boxed{\sum_{j\in\mathcal{N}_i} w_{ij}x_j(t)} - \alpha(t)\nabla f_i\left(\boxed{\sum_{j\in\mathcal{N}_i} w_{ij}x_j(t)} \right)\right)$$

Projection

Average Consensus

- $\alpha(t)$ *is a diminishing step-size to be shared by all agents for asymptotic convergence*

$$\alpha(t) \to 0 \qquad \text{and} \qquad \int_0^\infty \alpha(t) \to \infty$$

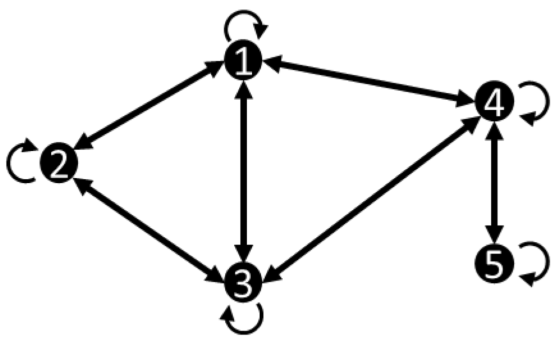
A natural choice: $\alpha(t) = \dfrac{1}{t}$

Main result:

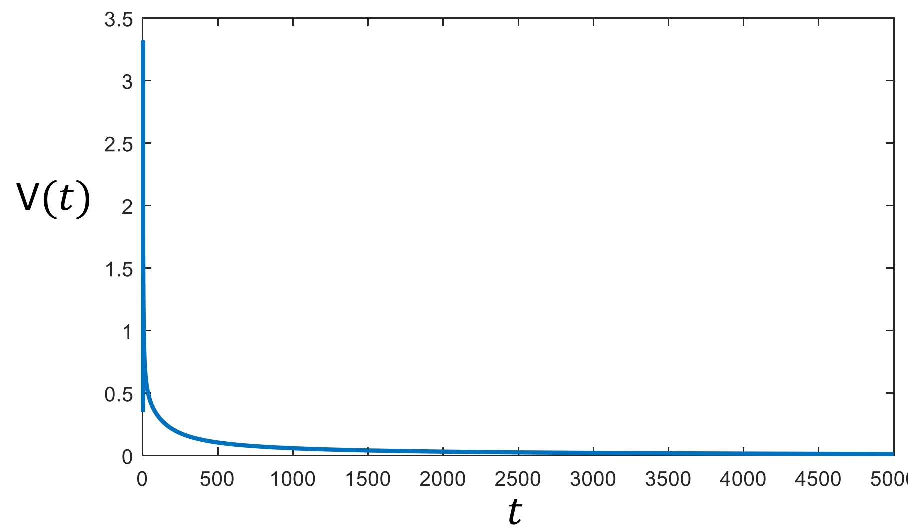
Suppose the network is *fixed, strongly connected* and *doubly stochastic*, all local constraints and objective functions are *convex*, then the update can drive the states in all agents to ***converge asymptotically*** to a same optimum point that satisfies (A+B+C) .

[1] A. Nedic, A. Ozdaglar, and P. A. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE TAC*, 2010
[2] P. Lin, W. Ren, and Y. Song, "Distributed multi-agent optimization subject to nonidentical constraints and communication delays," Automatica, 2016.

Simulation:



$x_i \in \mathbb{R}^6$



Objective functions:

$$f_i(x) = \|x - c_i\|_2^2 \quad \text{with} \quad c_i \in \mathbb{R}^6$$

Local constraints: $A_i x_i = b_i$ with $A_i \in \mathbb{R}^{1\times 20},\ b_i \in \mathbb{R}$

global information

Concern: *Diminishing step-size to be shared by all agents for asymptotic convergence*

not robust

Distributed Optimization Based on Primal-Dual

(A.) global objective $\min \sum_{i=1}^{m} f_i(x_i)$

subject to

(B.) local constraint $x_i \in \mathcal{X}_i \subset \mathbb{R}^n$

(C.) consensus $x_1 = x_2 = \cdots = x_m$

Enablers

Gradient	$x_i - \alpha \nabla f_i(x_i)$
Projection	$\mathcal{P}_i(x_i)$
Averaging	$\sum_{j \in \mathcal{N}_i} w_{ij} x_j(t)$

The Primal-Dual approach

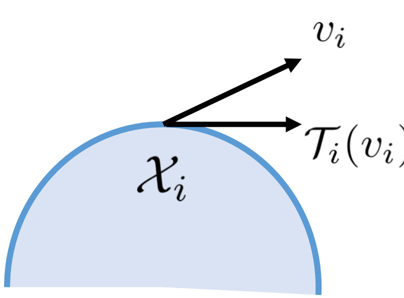
Dual vector $z_i(t)$ on **Averaging** $z_i{}^{\top} \sum_{j \in \mathcal{N}_i} w_{ij} \, (x_i - x_j)$

Saddle-point dynamics for *Primal-Dual* problem (continuous version):

$$\dot{x}_i = -\mathcal{T}_i \left(\nabla f_i(x_i) + \sum_{j \in \mathcal{N}_i} w_{ij} \, [z_i - z_j] + \sum_{j \in \mathcal{N}_i} w_{ij} \, [x_i - x_j] \right)$$

$$\dot{z}_i = \sum_{j \in \mathcal{N}_i} w_{ij} \, [x_i - x_j]$$

$\mathcal{T}_i(\cdot)$ Projection to the tangent space of the local constraint



[1] B. Gharesifard, and J. Cortes, "Distributed continuous-time convex optimization on weight-balanced digraphs," *IEEE TAC*, 2014

[2] X. Zeng, P. Yi, and Y. Hong, "Distributed Continuous-Time Algorithm for Constrained Convex Optimizations via Nonsmooth Analysis Approach" *IEEE TAC*, 2016

Distributed Optimization Based on Primal-Dual

Saddle-point dynamics for *Primal-Dual* problem (discrete version):

$$x_i(t+1) = \mathcal{P}_i\left(x_i(t) - \alpha \nabla f_i(x_i(t)) - \alpha \sum_{j \in \mathcal{N}_i} w_{ij}\,[z_i(t) - z_j(t)] - \alpha \sum_{j \in \mathcal{N}_i} w_{ij}\,[x_i(t) - x_j(t)]\right)$$

$$z_i(t+1) = z_i(t) + \beta \sum_{j \in \mathcal{N}_i} w_{ij}\,[x_i(t) - x_j(t)]$$

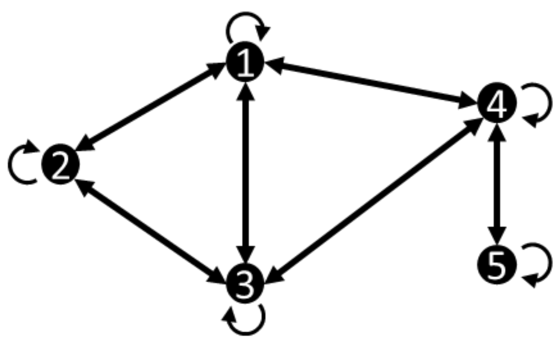
Transfer doubled state.

Main result:

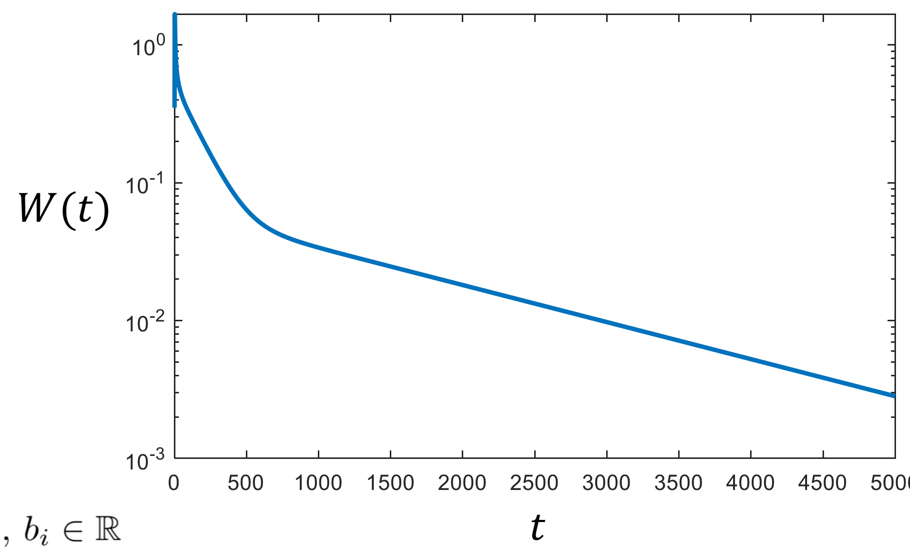
Suppose the network is *fixed, undirected and connected*, all local constraints are convex and objective functions are **strongly** *convex*, all *step-sizes* are **sufficiently** small. Then the update can drive the states in all agents to ***converge exponentially*** to a same optimum point that satisfies (A+B+C) .

- For continuous-time update, there's no need to choose step size.

Simulation:



$x_i \in \mathbb{R}^6$



Objective functions:

$$f_i(x) = \|x - c_i\|_2^2 \quad \text{with}$$

Local constraints: $A_i x_i = b_i$ with $A_i \in \mathbb{R}^{1\times 20}$, $b_i \in \mathbb{R}$

J. Lei, H. Chen, and H. Fang, Primal-Dual Algorithm for Distributed Constrained Optimization, *Systems & Control Letters, 2016*

Draw-backs of Existing Algorithms

(A.) global objective

$$\min \sum_{i=1}^{m} f_i(x_i)$$

subject to

(B.) local constraint

$$x_i \in \mathcal{X}_i \subset \mathbb{R}^n$$

(C.) consensus

$$x_1 = x_2 = \cdots = x_m$$

Enabler

Gradient $x_i - \alpha \nabla f_i(x_i)$

Projection $\mathcal{P}_i(x_i)$

Averaging $\frac{1}{d_i} \sum_{j \in \mathcal{N}_i} x_j(t)$

Diminishing step-size

$$\alpha(t) \to 0 \text{ and } \int_0^\infty \alpha(t) \to \infty$$

Common choice: $\alpha(t) = \frac{1}{t}$

Degradation on convergence $\mathcal{O}\left(\frac{1}{t}\right)$

Primal-Dual approach $z_i(t)$

Transfer doubled state

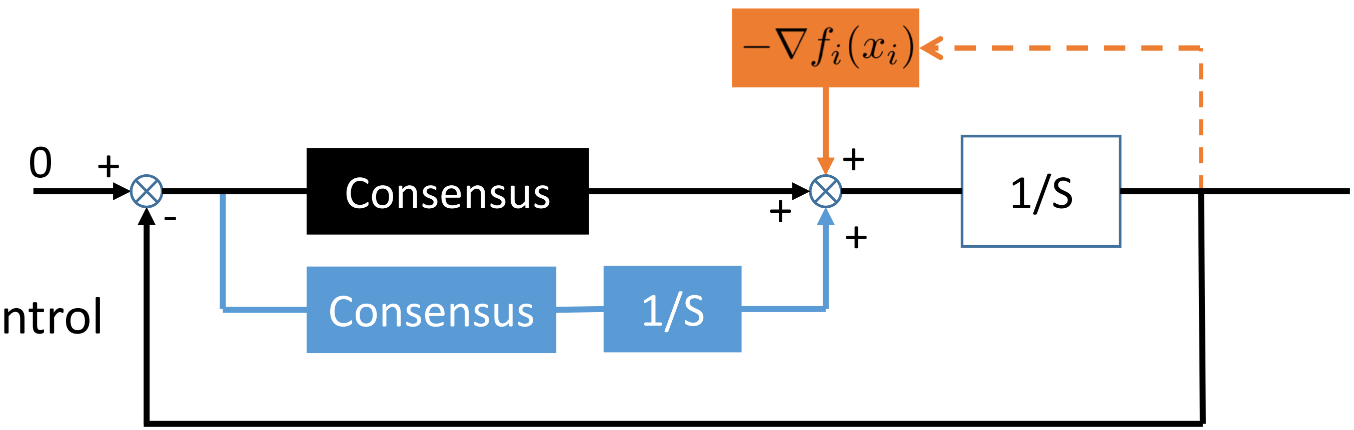
(communication burden).

Distributed Optimization based on integral-feedback

Continuous-time:

$$\dot{x}_i = \boxed{-\nabla f_i(x_i)}\underbrace{\boxed{-\sum_{j\in\mathcal{N}_i}(x_i-x_j)}\overbrace{\boxed{-\int_0^t \sum_{j\in\mathcal{N}_i}(x_i-x_j)}}^{\text{Integral-Feedback control}}}_{\text{Second order Consensus}}$$

Block Diagram:



Feedback control

Advantages:

- Elimination of Diminishing step-size
- Without doubled stated to be shared

Exponential convergence

Half of the communication

[1] X. Wang, S. Mou, and B. D. O. Anderson "Consensus-based Distributed Optimization Enhanced by Integral Feedback," *IEEE TAC, submitted*

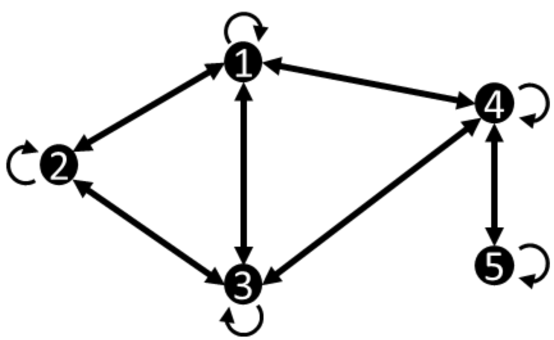
Simulation

Distributed Update based on Integral Feedback Control :

$$\dot{x}_i = P_i\Big(\boxed{-\nabla f_i(x_i)} \boxed{- \sum_{j\in\mathcal{N}_i}(x_i - x_j)} - \boxed{\int_0^t \sum_{j\in\mathcal{N}_i}(x_i - x_j)} \Big)$$

Projection | Gradient | Consensus | Integral Feedback control

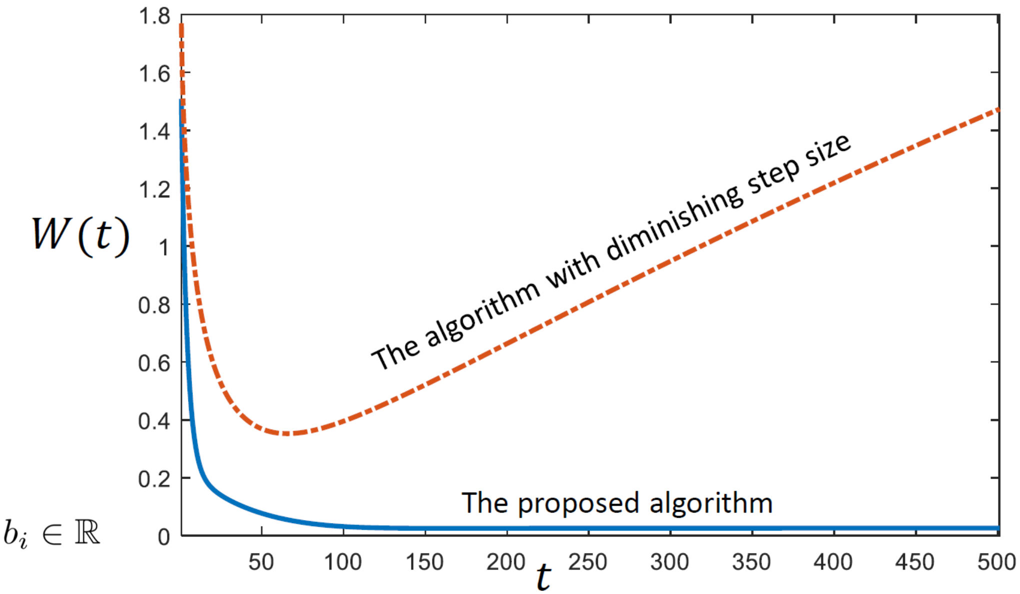
$x_i \in \mathbb{R}^6$



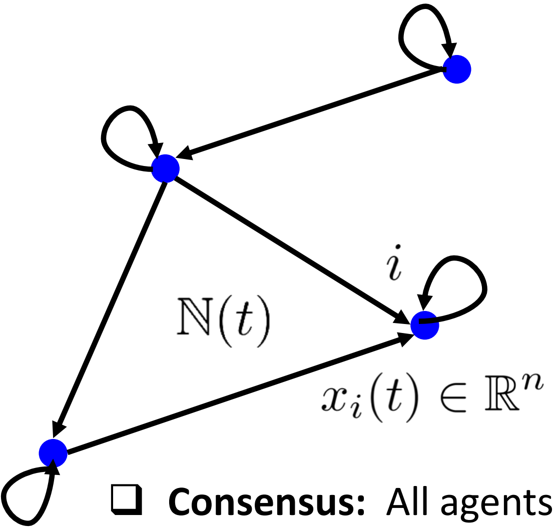
Objective functions:

$$f_i(x) = \|x - c_i\|_2^2 \quad \text{with} \quad c_i \in \mathbb{R}^6$$

Local constraints: $A_i x_i = b_i$ with $A_i \in \mathbb{R}^{1\times 20}$, $b_i \in \mathbb{R}$



Robustness Against Noise



Consensus-based Distributed Optimization

(A.) global objective $\min \sum_{i=1}^{m} f_i(x_i)$

subject to

(B.) local constraint $x_i \in \mathcal{X}_i \subset \mathbb{R}^n$

(C.) consensus $x_1 = x_2 = \cdots = x_m$

- **Consensus:** All agents reach the **same** value. **(C), well-studied** $x_i(t+1) = \sum_{j \in \mathcal{N}_i} w_{ij} x_j(t)$
- **Constrained Consensus:** All agents reach the **same** and **specific** value. **(B)+ (C)**
  - Linear Case: $x_i(t+1) = x_i(t) - P_i \left(x_i(t) - \frac{1}{d_i(t)} \sum_{j \in \mathcal{N}_i(t)} x_j(t) \right)$
  - Non-Linear Case: $x_i(t+1) = \mathcal{P}_i \left(\sum_{j \in \mathcal{N}_i} w_{ij} x_j(t) \right)$
- ***Constrained, Optimal Consensus:*** All agents reach the **same** and **specific** value, which ***minimizes*** the sum of local cost functions. **(A)+(B)+ (C)**

$$x_i(t+1) = \mathcal{P}_i \Big(\sum_{j \in \mathcal{N}_i} w_{ij} x_j(t) - \alpha(t) \nabla f_i \big(\sum_{j \in \mathcal{N}_i} w_{ij} x_j(t) \big) \Big)$$

$$x_i(t+1) = \mathcal{P}_i \left(x_i(t) - \alpha \nabla f_i(x_i(t)) - \alpha \sum_{j \in \mathcal{N}_i} w_{ij} \left[z_i(t) - z_j(t) \right] - \alpha \sum_{j \in \mathcal{N}_i} w_{ij} \left[x_i(t) - x_j(t) \right] \right)$$

$$z_i(t+1) = z_i(t) + \beta \sum_{j \in \mathcal{N}_i} w_{ij} \left[x_i(t) - x_j(t) \right]$$

$$\dot{x}_i = P_i \Big(-\nabla f_i(x_i) - \sum_{j \in \mathcal{N}_i} (x_i - x_j) - \int_0^t \sum_{j \in \mathcal{N}_i} (x_i - x_j) \Big)$$

Q&A Discussion at 4:30pm on April 1 (Wed) through Webex