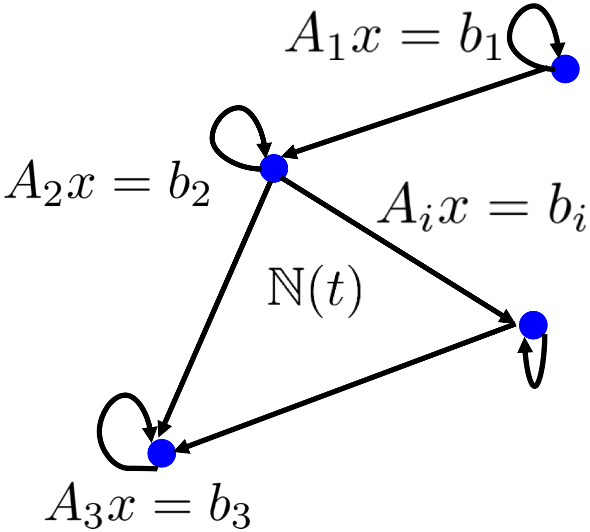
Lecture: Distributed Algorithms for Solving Large-Scale Linear Equations (II)

Shaoshuai Mou

Assistant Professor
School of Aeronautics and Astronautics
PURDUE UNIVERSITY



Review

$$\underbrace{x_i(t+1) = x_i(t) - P_i}_{\text{Satisfy its own private equation}}\left(\underbrace{x_i(t) - \frac{1}{d_i(t)}\sum_{j\in\mathcal{N}_i(t)} x_j(t)}_{\text{Consensus}}\right)$$

Satisfy its own private equation

$A_i x = b_i$

Consensus

Agreement Principle

- **Analysis: Error Dynamics** $e_i(t) = x_i(t) - x^*$

$$e_i(t+1) = P_i \frac{1}{d_i(t)} \sum_{j\in\mathcal{N}_i(t)} P_j e_j(t)$$

$$e(t+1) = P(S_t \otimes I_n)Pe(t)$$

$$P = \mathtt{diag}\{P_1, P_2, ..., P_m\}$$

$$S_t = D^{-1}_{\mathbb{N}(t)} A_{\mathbb{N}(t)}$$

Case I: Fixed Undirected Graph

$$e(t+1) = P\bar{S}e(t)$$

To prove $e(t) \to 0$, it is sufficient to show $\rho(PS) < 1$

- *Are all eigenvalues real??*
- *Are all eigenvalues in the interval (-1,1]?*
- *Prove 1 is not an eigenvalue of PS by contradiction*

$$P = \mathtt{diag}\{P_1, P_2, ..., P_m\}$$

$$\bar{S} = S \otimes I_n$$

$$S = D^{-1} A_{\mathbb{N}}$$ row stochastic

$$D = \mathtt{diag}\{d_1, d_2, ..., d_m\}$$

- **Are all eigenvalues real??**

$$P\bar{D}^{-1}\bar{A}_{\mathbb{N}} \longrightarrow PP\bar{D}^{-1/2}\bar{D}^{-1/2}\bar{A}_{\mathbb{N}}$$

$$\downarrow$$

$$\hat{S} = \bar{D}^{-1/2}\bar{A}_{\mathbb{N}}\bar{D}^{-1/2} \qquad P\hat{S}P \longleftarrow \bar{D}^{-1/2}PP\bar{D}^{-1/2}\bar{A}_{\mathbb{N}}$$

Symmetric for undirected graphs

- **Are all eigenvalues in the interval [-1,1]?**

For any nonzero eigenvalue λ of $P\hat{S}P$ with unit eigenvector v, one has

$$\begin{cases} P\hat{S}Pv = \lambda v \\ P\hat{S}Pv = \lambda Pv \\ Pv = v \end{cases}$$

$$-1 < \lambda_{\min}(S) = \lambda_{\min}(\hat{S}) \leq \lambda = v'P\hat{S}Pv = v'\hat{S}v \leq \lambda_{\max}(\hat{S}) = \lambda_{\max}(S) \leq 1$$

S is **primitive** (since its graph is $\mathbb{N}$, which is strongly connected and with self-arcs) and is **row stochastic**

$\Downarrow$ **Perron-Frobenius Theorem**

- Its largest eigenvalue is 1, which is simple, equal to its spectral radius, and strictly larger than all the other eigenvalues.
- Its eigenvector corresponding to 1 is positive and unique up to scaling by $\mathbf{1}$

❑ 1 is not an eigenvalue of $P\bar{S}$ $\quad P\hat{S}P \quad \hat{S} = \bar{D}^{-1/2}\bar{A}_{\mathbb{N}}\bar{D}^{-1/2}$

We show this by **contradiction.**

Proof: Suppose 1 is an eigenvalue of $P\hat{S}P$ with eigenvector $v \neq 0$

$$P\hat{S}Pv = v$$

multiplying P to both sides

$$P\hat{S}Pv = Pv$$

$$\left.\begin{array}{l} P\hat{S}Pv = v \\ P\hat{S}Pv = Pv \end{array}\right\} Pv = v$$

$$\left.\begin{array}{l} P\hat{S}Pv = v \\ Pv = v \end{array}\right\} v'\hat{S}v = v'v \xrightarrow{\text{symmetric}} \hat{S}v = v$$

$$\bar{D}^{-1/2}\bar{A}_{\mathbb{N}}\bar{D}^{-1/2}v = v \xrightarrow{v = \bar{D}^{1/2}u} \bar{D}^{-1}\bar{A}_{\mathbb{N}}u = u \longrightarrow (S \otimes I_n)u = u$$

stochastic PF theorem

$$u = \mathbf{1} \otimes q$$

$$P_i q = q \longleftarrow P\bar{D}^{1/2}(\mathbf{1} \otimes q) = \bar{D}^{1/2}(\mathbf{1} \otimes q) \longleftarrow u = \mathbf{1} \otimes q$$

$$q \in \cap_{i=1}^{m}\text{image}P_i = 0 \xrightarrow{\text{unique solution}} v = 0$$

Case II: Time-Varying Directed Networks

$$e(t+1) = P\bar{S}(t)e(t)$$

$P = \mathtt{diag}\{P_1, P_2, ..., P_m\}$

$\bar{S} = S(t) \otimes I_n$

$S(t) = D^{-1}(t)A_{\mathbb{N}}(t)$

Network-dependent Time-varying Discrete System

$D(t) = \mathtt{diag}\{d_1(t), d_2(t), ..., d_m(t)\}$

To prove $e(t) \to 0$, it is sufficient to show

$$\lim_{t\to\infty} P(S_t \otimes I_n)P(S_{t-1} \otimes I_n)\cdots P(S_1 \otimes I_n)P = 0$$

$$\cdots \underbrace{P(S_{2T} \otimes I_n)\cdots P(S_{T+1}}_{M_2} \otimes I_n)P \cdot \underbrace{P(S_T \otimes I_n)\cdots P(S_1 \otimes I_n)P}_{M_1}$$

$\cdots$

Find a sub-multiplicative norm to be a contraction (namely, $||M_k|| < 1$)

Key Tool 1: A norm for contraction

$$e(t+1) = P(S_t \otimes I_n) P e(t)$$

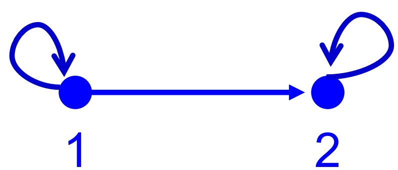
$$\lim_{t \to \infty} P(S_t \otimes I_n) P(S_{t-1} \otimes I_n) \cdots P(S_1 \otimes I_n) P = 0$$

Key Step: Find a sub-multiplicative norm to be a contraction.

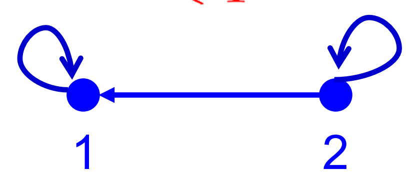
$$\cdots \left| P(S_{2T} \otimes I_n) \cdots P(S_{T+1} \otimes I_n) P \right| \cdot \left| P(S_T \otimes I_n) \cdots P(S_1 \otimes I_n) P \right| \to 0$$

< 1 $\qquad\qquad\qquad$ < 1

....



t=2



t=1

$$\begin{bmatrix} \frac{1}{2}P_1 & \frac{1}{2}P_1P_2 \\ \frac{1}{4}P_2P_1 & \frac{1}{4}P_2P_1P_2 + \frac{1}{2}P_2 \end{bmatrix} \xrightarrow[|P_i|_2 = 1]{[\,\cdot\,]_2} \begin{bmatrix} \frac{1}{2} & < \frac{1}{2} \\ < \frac{1}{4} & < (\frac{1}{4} + \frac{1}{2}) \end{bmatrix} \xrightarrow{||\cdot||_\infty} \leq 1$$

The 1, 2, ∞ norm do not work !! $\qquad$ $|P_1P_2|_2 < 1$

$\cap_{i=1}^{m} \text{image } P_i = 0$ $\qquad$ $|P_2P_1|_2 < 1 \quad |P_2P_1P_2|_2 < 1$

Mixed Matrix Norm: $|M| = || \; [\; M \;]_2 \; ||_\infty \qquad \mathbb{R}^{mn \times mn}$

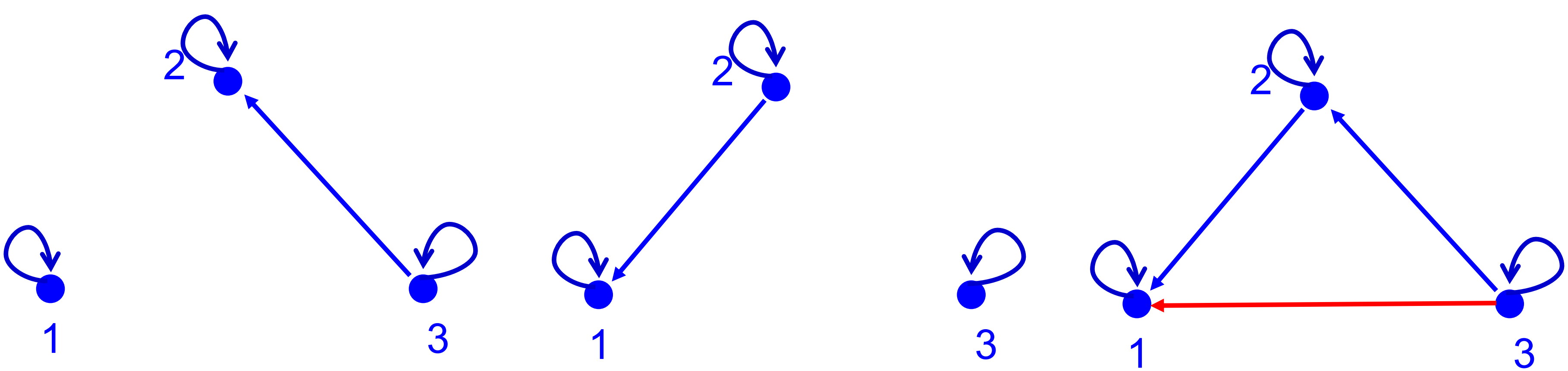
is a norm $\qquad$ sub-multiplicative $\qquad$ a contraction

Key Tool 2: Neighbor Graph Requirement

A directed graph G is strongly connected if for each pair of distinct vertices, *i* and *j*, there is a directed path in G from *i* to *j.*

Graph composition G : all directed graphs with vertex set V = {1,2,…,*m*} with self arcs

G_1 (i, k) G_2 (k, j) $G_2 \circ G_1$ (i, j)



The infinite sequence of graphs N(1), N(2), …. is repeatedly jointly strongly connected

···· N(p+q+1) N(p+q) ···· N(p+1) N(p) ···· N(2) N(1)

Composition is strongly connected.

Comparison with standard consensus problem

Given $S_1, S_2, \dots$ when does $S_t S_{t-1} \cdots S_1 \to \mathbf{1}c$

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$M_{pq} = S_p S_{p-1} \cdots S_q$$

semi-norm: $\lceil M_{pq} \rceil = \min_f ||\mathbf{1}f - M_{pq}||_\infty$ submultiplicative

contraction: $\lceil M_{pq} \rceil < 1$ for $p - q$ sufficiently large provided the graph of each S_i is rooted.

Given $S_1, S_2, \dots$ when does $P(S_t \otimes I)P(S_{t-1} \otimes I) \cdots P(S_1 \otimes I)P \to 0$

$$M_{pq} = P(S_p \otimes I)P(S_{p-1} \otimes I) \cdots P(S_q \otimes I)P$$

mixed matrix norm: $|M_{pq}| = ||\langle M_{pq} \rangle||_\infty$ submultiplicative

contraction: $|M_{pq}| < 1$ for $p - q$ sufficiently large provided the graph of each S_i is strongly connected.

Main Result:

> Suppose $Ax = b$ have solutions. If the sequence of neighbor graphs N(1), N(2), N(3), …. are repeatedly jointly strongly connected, then all $x_i(t)$ converge to a solution of the equation exponentially fast.

S. Mou, J. Liu, A. S. Morse. *IEEE Transactions on Automatic Control,* 2015, 60 (11), pp 2863-2878

Our Algorithm	Existing Results
is distributed	parallel algorithms/Conjugate Gradient
puts no requirements on A	Gauss Seidel/Jacobian Iteration
does not involve any small step-size	Gradient/SOR/[1][2][3][4]
converges exponentially fast	[1]
works for time-varying directed networks	[2],[3]
operates asynchronously	[2],[3]

Most recent results in distributed optimiztions:

[1]: A. Nedic and A. Ozdaglar. Distributed sub-gradient methods for multi-agent optimization. IEEE Trans. on Automatic Control. 2014

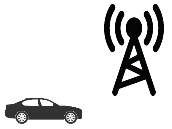
[2]: D. Jakovetic, J. Moura, J. Xavier. Fast distributed gradient method. IEEE Trans. on Automatic Control. 2014

[3]: J. Duchi, A. Agarwal, M. Wainwright. Dual averaging for distributed optimization. IEEE Trans. on Automatic Control. 2014

[4]:T. Chang, A. Nedic. Distributed constrained optimization by consensus-based primal-dual method. IEEE Trans. on Automatic Control. 2014

Applications (due to distributed natures in large networks; privacy)

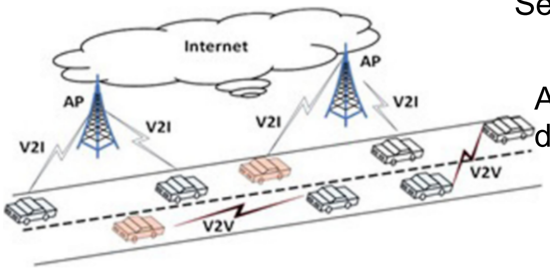
Large content distribution in vehicular networks



✓ **linear network coding**

Partition: $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$ $\quad y_i \in \mathbb{R}^n$ n small

Send out linear combination: $a_1(t)y_1 + a_2(t)y_2 + \cdots + a_m(t)y_m = b_t$



After receiving all *m* independent equations, a vehicle will be able to decode the original large y by solving large linear equations Ax=b with $A \in \mathbb{R}^{mn \times mn}$

large memory; high computational complexity.

✓ **Distributed update for solving linear equations** start iteration even when a vehicle only receives one linear equation.

- Update the linear equations by V2V at time *t*

$$\begin{aligned} (\bar{a}_1(t) \otimes I_n)'y &= b_1(t) \\ (\bar{a}_2(t) \otimes I_n)'y &= b_2(t) \\ &\vdots \\ (\bar{a}_p(t) \otimes I_n)'y &= b_p(t) \end{aligned} \qquad 1 \leq p \leq m$$

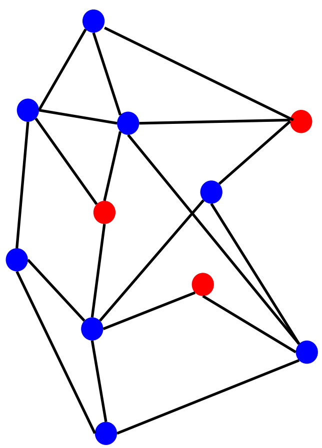
small memory

- Update the estimate at the *k*th iteration by cyclic projection to the $\bar{k}$ th row equation

$$x_i(k+1) = x_i(k) + \frac{\bar{a}_{\bar{k}}(t) \otimes I_n}{||\bar{a}_{\bar{k}}||^2} \left(b_{\bar{k}}(t) - (\bar{a}_{\bar{k}}(t) \otimes I_n)' x_i(k)\right)$$

low computational complexity

Application 2: **Distributed Network Localization.**



In sensor networks, nodes are usually deployed randomly.

Location information is valuable! detect/record events, geographic routing

GPS is costly for large networks; does not work well under obstructions

Problem : In a large sensor-network, only three agents (red) know their positions. Devise a a distributed algorithm for each (blue) to achieve its own position by communications with neighbors.

U. Khan, S. Kar, J. Moura. *IEEE Trans. on Signal Processing.* 2009

$$p_i = \sum_{j \in \mathcal{N}_i} a_{ij} p_j$$

barycentric coordinates

$$Ap = b$$

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 & 0 \\ a_{21} & a_{22} & a_{23} & 0 \\ 0 & a_{32} & a_{33} & a_{34} \\ 0 & 0 & a_{43} & a_{44} \end{bmatrix} \otimes I_2$$

- *A* is **sparse**
- **Part** of the solution is of agents' interest.

❑ Utilization of the **sparsity** for state vector reduction

$$\begin{bmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & A_{23} \\ A_{31} & 0 & A_{33} \end{bmatrix} \bar{x} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

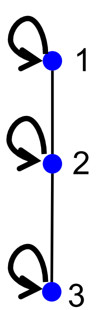
- Each agent knows one block row of *A* and is only interested in **part** of the solution corresponding to the non-zero blocks of *A*.

$$\bar{x}_i(t) \in \mathbb{R}^d$$

Satisfy its own private equation

Partial Consensus

$$\bar{x}_i(t+1) = \bar{x}_i(t) + P_i \quad (\bar{x}_i(t) - \text{consensus vector})$$



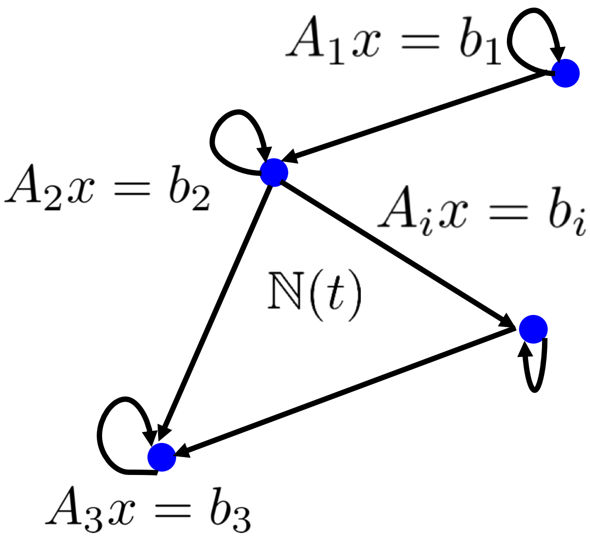
$$\begin{array}{ccc} x & y & z \end{array}$$

$$\begin{aligned} A_{11} x_1(t) + A_{12} y_1(t) \qquad\qquad &= b_1 \\ A_{21} x_2(t) + A_{22} y_2(t) + A_{23} z_2(t) &= b_2 \\ A_{31} x_3(t) \qquad\qquad + A_{33} z_3(t) &= b_3 \end{aligned}$$

$$\begin{bmatrix} x_1(t+1) \\ y_1(t+1) \end{bmatrix} = \begin{bmatrix} x_1(t) \\ y_1(t) \end{bmatrix} + P_1 \begin{bmatrix} x_1(t) - \frac{1}{2}(x_1(t) + x_2(t)) \\ y_1(t) - \frac{1}{2}(y_1(t) + y_2(t)) \end{bmatrix}$$

$$\begin{bmatrix} x_2(t+1) \\ y_2(t+1) \\ z_2(t+1) \end{bmatrix} = \begin{bmatrix} x_2(t) \\ y_2(t) \\ z_2(t) \end{bmatrix} + P_2 \begin{bmatrix} x_2(t) - \frac{1}{2}(x_2(t) + x_1(t)) \\ y_2(t) - \frac{1}{3}(y_1(t) + y_2(t) + y_3(t)) \\ z_2(t) - \frac{1}{2}(z_2(t) + z_3(t)) \end{bmatrix}$$

$$\begin{bmatrix} x_3(t+1) \\ z_3(t+1) \end{bmatrix} = \begin{bmatrix} x_3(t) \\ z_3(t) \end{bmatrix} + P_3 \begin{bmatrix} x_3(t) - \frac{1}{2}(x_3(t) + x_2(t)) \\ z_3(t) - \frac{1}{2}(z_3(t) + z_2(t)) \end{bmatrix}$$



Initialization: $A_i x_i(1) = b_i$

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

Satisfy its own private equation

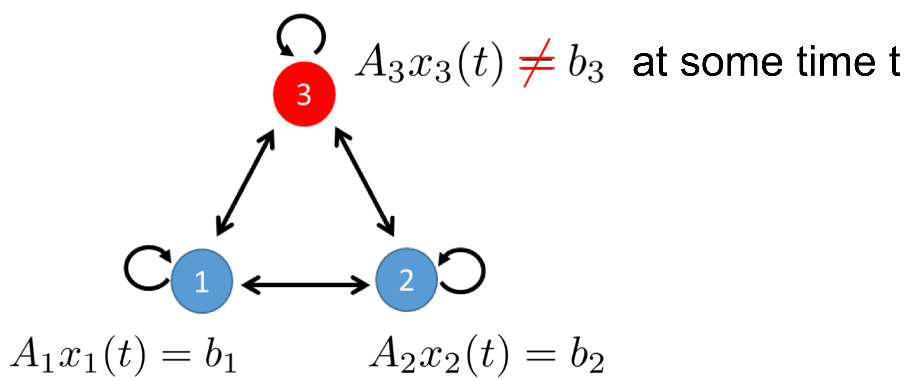
$A_i x = b_i$

Consensus

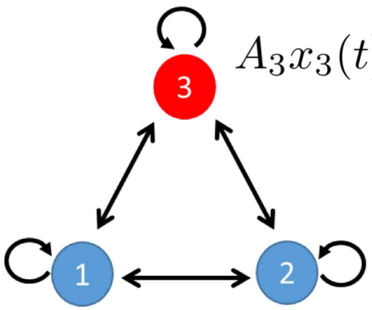
Agreement Principle

What if such principle not satisfied?

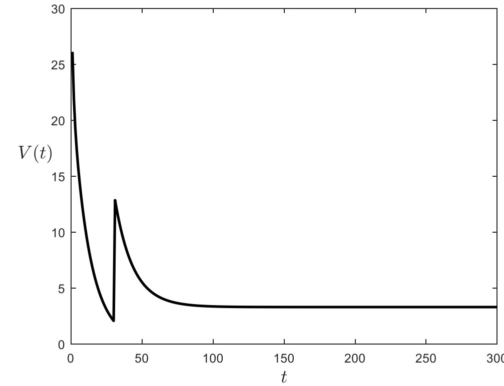
Initialization error or cyber-attacks



$A_3 x_3(t) \cancel{=} b_3$ at some time t



$A_1 x_1(t) = b_1$ $\quad$ $A_2 x_2(t) = b_2$

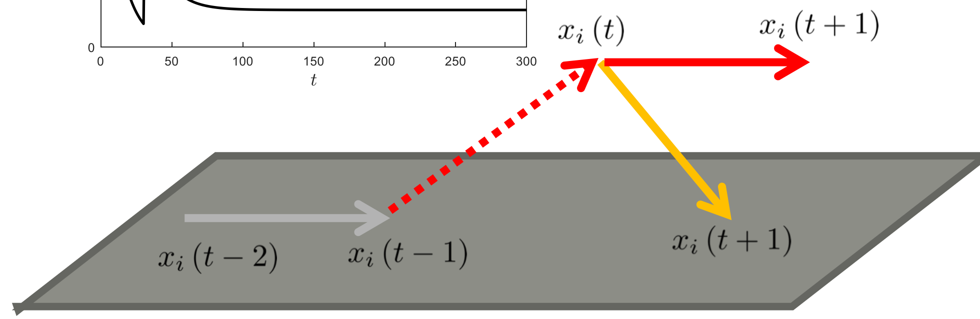


$$V(t) = \frac{1}{2}\sum_{i=1}^{m} |x_i(t) - x^*|^2$$

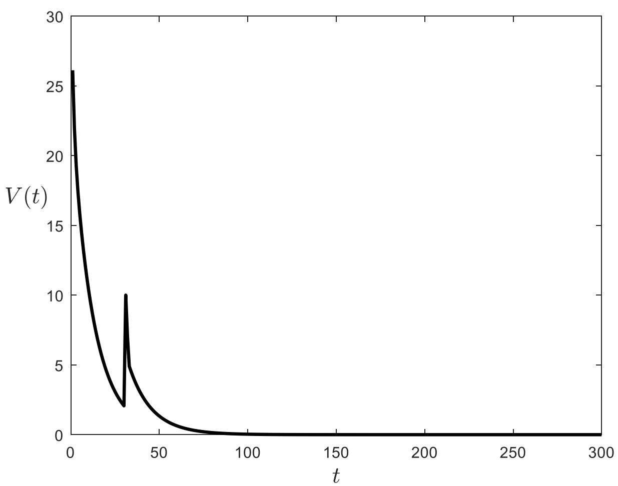
x^* is the true solution

Our Research Progress: Introduce extra control

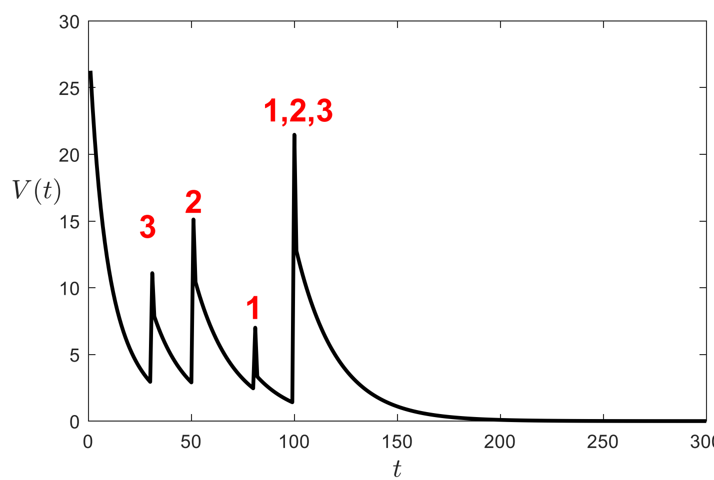
$$x_i(t+1) = x_i(t) - P_i\left(x_i(t) - \frac{1}{d_i(t)}\sum_{j\in\mathcal{N}_j(t)} x_j(t)\right)$$
$$\boxed{-{A_i}^T(A_i {A_i}^T)^{-1}(A_i x_i(t) - b_i)}$$



Force the state to satisfy its own constriants.



One-time attack at one node



Multiple times attacks at multiple nodes

* X Wang, S Mou, D Sun. Improvement of a distributed algorithm for solving linear equations. IEEE Transactions on Industrial Electronics 64 (4), 3113-3117, 2016

- All on-campus students are also encouraged to attend lectures online. Recorded lectures will be available to you through blackboard
- No need to come to Wang 2555 for AAE 590 Multi-agent Systems and Control for future lectures.
- Office hours will be replaced by emails, and/or telephone calls.