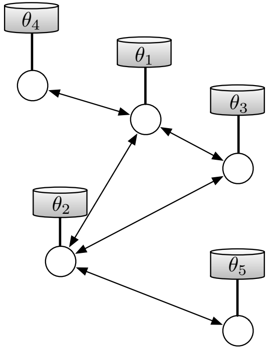
Network Science
Models and Distributed Algorithms

IST-CMU Phd course
João Xavier
TA: João Martins

October 3, 2016

Consensus in static undirected networks



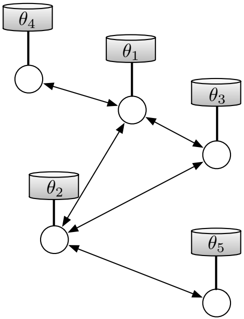
- n agents; agent i holds $\theta_i \in \mathbf{R}$
- communication network is static and undirected
- communication happens in discrete time $t = 0, 1, 2, 3, \ldots$
- goal: compute the average

$$\overline{\theta} = \frac{\theta_1 + \cdots + \theta_n}{n}$$

- we model the network as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$:
  - $\mathcal{V} = \{1, 2, \ldots, n\}$ is set of agents
  - $\mathcal{E}$ is set of communication channels

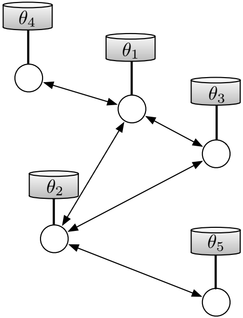
- agents i and j can communicate if and only if $\{i, j\} \in \mathcal{E}$

- example:



$$\mathcal{V} = \{1, 2, 3, 4, 5\} \quad \mathcal{E} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 4\}, \{2, 5\}\}$$

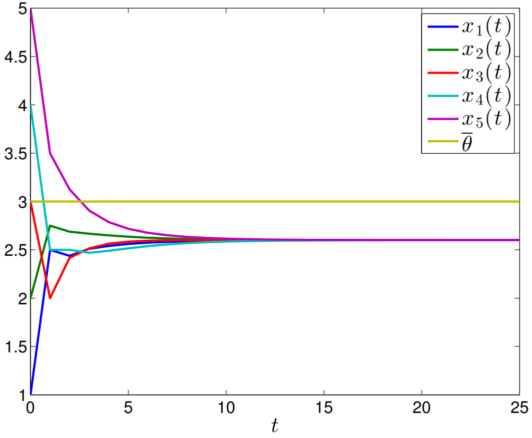
- we also use the notation: $i \sim j \quad \Leftrightarrow \quad \{i, j\} \in \mathcal{E}$



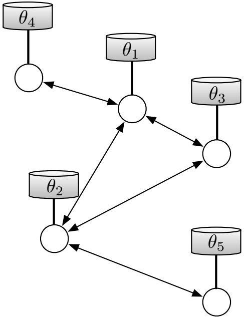
- naive scheme: agents repeatedly compute local averages
- $x_i(0) := \theta_i$ and

$$\begin{cases} x_1(t+1) &= \frac{x_1(t)+x_2(t)+x_3(t)+x_4(t)}{4} \\ x_2(t+1) &= \frac{x_1(t)+x_2(t)+x_3(t)+x_5(t)}{4} \\ x_3(t+1) &= \frac{x_1(t)+x_2(t)+x_3(t)}{3} \\ x_4(t+1) &= \frac{x_1(t)+x_4(t)}{2} \\ x_5(t+1) &= \frac{x_2(t)+x_5(t)}{2} \end{cases}$$

- naive scheme doesn't work:



- how can we fix this?



- naive scheme in matrix form: $x(0) = \underbrace{(\theta_1, \theta_2, \theta_3, \theta_4, \theta_5)}_{\theta}$

$$\underbrace{\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \\ x_3(t+1) \\ x_4(t+1) \\ x_5(t+1) \end{bmatrix}}_{x(t+1)} = \underbrace{\begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{bmatrix}}_{W} \underbrace{\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \\ x_5(t) \end{bmatrix}}_{x(t)}$$

- important: in distributed algorithms, $W_{ij} = 0$ whenever $i \not\sim j$

- consider a more general iterative scheme: $x(0) = \theta$ and

$$x(t+1) = Wx(t), \quad t = 0, 1, 2, \ldots$$

where

$$W \in \mathcal{W}_{\mathcal{G}} := \left\{ W \in \mathbf{R}^{n \times n} \, : \, W_{ij} = 0 \text{ if } i \nsim j, W = W^T \right\}$$

- can this scheme work for some weight matrix W?

- "works" means agents' states converge to $\overline{\theta} = \frac{\theta_1 + \cdots + \theta_n}{n}$:

$$x(t) \underset{t \to \infty}{\rightarrow} \overline{\theta} \mathbf{1}$$

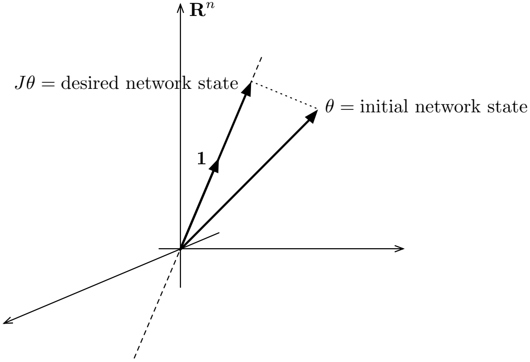
where $\mathbf{1} := (1, 1, \ldots, 1) \in \mathbf{R}^n$

- equivalent to

$$x(t) \underset{t \to \infty}{\longrightarrow} J\theta$$

where

$$J := \frac{1}{n}\mathbf{1}\mathbf{1}^T = \begin{bmatrix} \frac{1}{n} & \cdots & \frac{1}{n} \\ \vdots & \cdots & \vdots \\ \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix} \in \mathbf{R}^{n \times n}$$



- J is the orthogonal projector onto span$(\mathbf{1})$

- unrolling the recursion $x(t+1) = Wx(t)$ gives

$$x(t) = W^t \theta$$

  for all $t \geq 0$

- we conclude that

$$x(t) \underset{t \to \infty}{\to} J\theta$$

  for all $\theta \in \mathbf{R}^n$ if and only if

$$W^t \underset{t \to \infty}{\to} J$$

- analysis boils down to analyzing the powers of W

- an useful tool for analyzing the powers of a matrix is the EVD

- **EVD:** for any symmetric $W \in \mathbf{R}^{n \times n}$, there exist $Q, \Lambda \in \mathbf{R}^{n \times n}$ such that

$$
\begin{aligned}
W &= \underbrace{\begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} q_1^T \\ \vdots \\ q_n^T \end{bmatrix}}_{Q^T} \\
&= \lambda_1 q_1 q_1^T + \cdots + \lambda_n q_n q_n^T
\end{aligned}
$$

  with Q:orthogonal ($Q^T Q = I$) and Λ:diagonal ($\lambda_1 \geq \cdots \geq \lambda_n$)

- (q_i, λ_i) is an eigenpair:

$$W q_i = \lambda_i q_i, \quad i = 1, \ldots, n$$

- EVD implies $W^t = Q\Lambda^t Q$, where

$$\Lambda^t = \begin{bmatrix} \lambda_1^t & & & \\ & \lambda_2^t & & \\ & & \ddots & \\ & & & \lambda_n^t \end{bmatrix}$$

- we reduced the analysis from powers of matrices to powers of scalars

- conclusion: $W^t \underset{t\to\infty}{\to} J$ if and only if W satisfies

$$(q_1, \lambda_1) = \left(\frac{1}{\sqrt{n}}\mathbf{1}, 1\right) \text{ and } |\lambda_i| < 1, \text{ for } i = 2, \ldots, n$$

- in terms of the EVD of W:

$$\begin{aligned} W &= \underbrace{\begin{bmatrix} \frac{1}{\sqrt{n}}\mathbf{1} & \widetilde{Q} \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} 1 & \\ & \widetilde{\Lambda} \end{bmatrix}}_{\Lambda} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{n}}\mathbf{1}^T \\ \widetilde{Q}^T \end{bmatrix}}_{Q^T} \\ &= J + \underbrace{\widetilde{Q}\widetilde{\Lambda}\widetilde{Q}^T}_{\widetilde{W}} \end{aligned}$$

where

$$\widetilde{Q} := \begin{bmatrix} q_2 & \cdots & q_n \end{bmatrix} \quad \text{and} \quad \widetilde{\Lambda} := \begin{bmatrix} \lambda_2 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

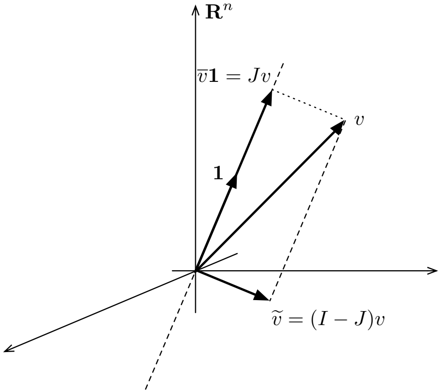
- note that $\left\| \widetilde{W} \right\| = \max \{ |\lambda_i| \, : \, i = 2, \ldots, n \} < 1$

- for such W, we can interpret $x(t+1) = Wx(t)$ geometrically

- for $v = (v_1, \ldots, v_n) \in \mathbf{R}^n$, consider the orthogonal decomposition

$$v = \overline{v}\mathbf{1} + \widetilde{v}$$

where $\overline{v} := \frac{v_1 + \cdots + v_n}{n} = \frac{1}{n}\mathbf{1}^T v \in \mathbf{R}$ and $\widetilde{v} := (I - J)v \in \mathbf{R}^n$



- $\overline{v}\mathbf{1}$ is the in-consensus component; $\widetilde{v}$ is the off-consensus component

- since $W = W^T$ and $(\mathbf{1}, 1)$ is an eigenpair of W:
  - $W\mathbf{1} = \mathbf{1}, \mathbf{1}^T W = \mathbf{1}^T$
  - $JW = WJ = J$
  - $(I - J)W = W(I - J) = W - J = \widetilde{W}$

- on one hand:

$$\begin{aligned}\overline{x}(t+1) &= \frac{1}{n}\mathbf{1}^T x(t+1) \\ &= \frac{1}{n}\mathbf{1}^T W x(t) \\ &= \frac{1}{n}\mathbf{1}^T x(t) \\ &= \overline{x}(t)\end{aligned}$$

- interpretation: algorithm preserves the in-consensus component

$$\overline{x}(t)\mathbf{1} = \overline{x}\mathbf{1}, \quad \text{for all } t \geq 0$$

- on the other hand:

$$\begin{aligned}\widetilde{x}(t+1) &= (I-J)x(t+1) \\ &= (I-J)Wx(t) \\ &= (I-J)W(I-J)x(t) \\ &= \widetilde{W}\widetilde{x}(t)\end{aligned}$$

- it follows that

$$\|\widetilde{x}(t)\| \leq \left\|\widetilde{W}\right\|^t \|\widetilde{x}\|$$

- interpretation: algorithm shrinks the off-consensus component to 0 geometrically fast

- **key question**: can we find in $\mathcal{W}_{\mathcal{G}}$ a matrix $W = Q\Lambda Q^T$ such that

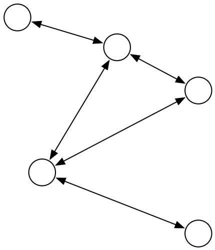
$$(q_1, \lambda_1) = \left(\frac{1}{\sqrt{n}}\mathbf{1}, 1\right) \text{ and } |\lambda_i| < 1, \text{ for } i = 2, \ldots, n?$$

- answer: YES if and only if $\mathcal{G}$ is connected

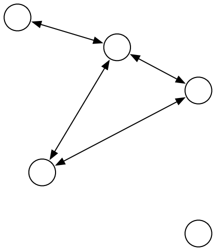
- proof:
  - ($\Rightarrow$): trivial
  - ($\Leftarrow$): we will show two examples of W that work:
    - Laplacian weights
    - Metropolis weights

Undirected graphs

- a graph is connected if there exists a path between any two nodes

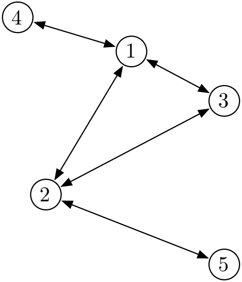


connected graph



disconnected graph

- the degree d_i of an agent i is its number of neighbors
- $d = (d_1, \dots, d_n) \in \mathbf{R}^n$ is the degree vector
- $D = \mathrm{diag}\,(d) \in \mathbf{R}^{n \times n}$ is the degree matrix
- example:

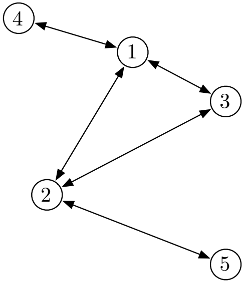


$$d = \begin{bmatrix} 3 \\ 3 \\ 2 \\ 1 \\ 1 \end{bmatrix} \qquad D = \begin{bmatrix} 3 & & & & \\ & 3 & & & \\ & & 2 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix}$$

- the (i, j) entry of the adjacency matrix $A = (a_{ij}) \in \mathbf{R}^{n \times n}$ is

$$a_{ij} = \begin{cases} 1, & \text{if } i \sim j \\ 0, & \text{otherwise} \end{cases}$$

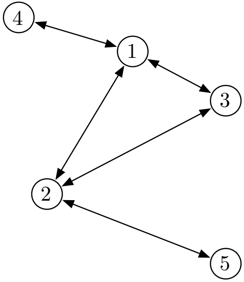
- example:



$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

- Elementary properties:
  - A is symmetric: $A = A^T$
  - $d = A\mathbf{1}$

- the laplacian is $L = D - A \in \mathbf{R}^{n \times n}$
- example:



$$L = \begin{bmatrix} 3 & -1 & -1 & -1 & 0 \\ -1 & 3 & -1 & 0 & -1 \\ -1 & -1 & 2 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{bmatrix}$$

- Properties of the laplacian:
  - as a linear operator:

$$(Lx)_i = d_i x_i - \sum_{j \sim i} x_j = d_i \left(x_i - \frac{1}{d_i} \sum_{j \sim i} x_j \right)$$

    interpretation: quantifies local disagreement between each agent and its neighbors' average
  - as a quadratic form:

$$x^T L x = \sum_{i \sim j} (x_i - x_j)^2$$

    interpretation: quantifies global disagreement between agents
  - $L\mathbf{1} = 0$
  - $L \succeq 0$
  - letting $\sigma(L) = \{\mu_1, \mu_2, \ldots, \mu_n\}$ with $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$:
    - $\mu_1 = 0$
    - $\mu_2 > 0$ if and only if $\mathcal{G}$ is connected

- μ_2 is the Fiedler eigenvalue of $\mathcal{G}$

- **key question**: can we find in $\mathcal{W}_\mathcal{G}$ a matrix $W = Q\Lambda Q^T$ such that

$$(q_1, \lambda_1) = \left(\frac{1}{\sqrt{n}}\mathbf{1}, 1\right) \text{ and } |\lambda_i| < 1, \text{ for } i = 2, \ldots, n?$$

- answer: YES if and only if $\mathcal{G}$ is connected

- proof: ($\Leftarrow$) the matrix $W = I - \alpha L$ works, if $0 < \alpha < \frac{2}{\mu_n}$

- interpretation 1: the iterative algorithm

$$x(t+1) = Wx(t) = x(t) - \alpha Lx(t)$$

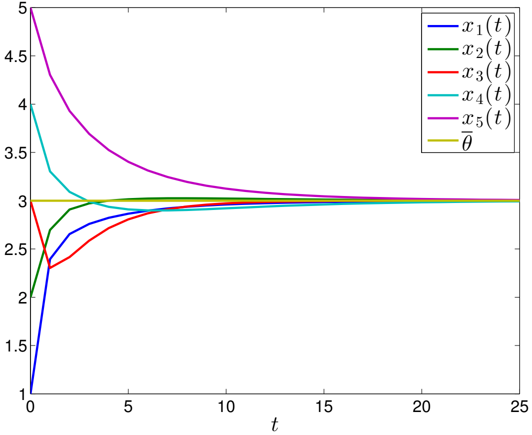
  is gradient method applied to $f(x) = \frac{1}{2}x^T Lx$ with stepsize α

- interpretation 2: in each iteration

$$x_i(t+1) = x_i(t) + \alpha d_i \left(\frac{1}{d_i}\sum_{j\sim i} x_j(t) - x_i(t)\right)$$

  agent i moves a bit toward its local average

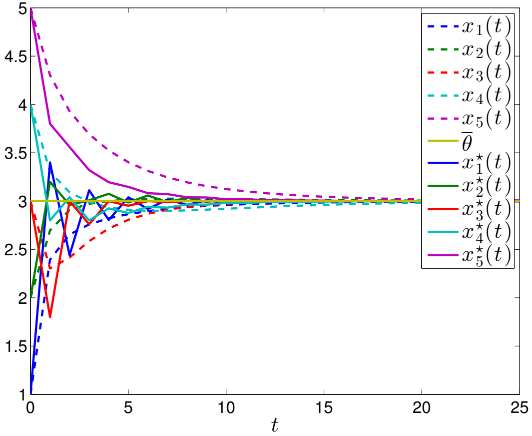
- back to example in page 5 with $W = I - \alpha L$ and $\alpha = \frac{1}{\mu_n}$:



- we fixed the problem

- which $0 < \alpha < \frac{2}{\mu_n}$ gives the fastest network?
- answer:

$$\alpha^\star = \frac{2}{\mu_2 + \mu_n}$$



Jordan canonical form

- for any $A \in \mathbf{C}^{n\times n}$, there exist $S, J \in \mathbf{C}^{n\times n}$ such that

$$A = SJS^{-1}$$

  with S:non-singular and

$$J = \begin{bmatrix} J_{\lambda_1} & & & \\ & J_{\lambda_2} & & \\ & & \ddots & \\ & & & J_{\lambda_p} \end{bmatrix}, \qquad J_{\lambda_i} = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ & & & & \lambda_i \end{bmatrix}$$

- each λ_i is an eigenvalue of A (the λ_i's may be repeated)

- some columns of S are eigenvectors of A

- example:

$$A = \underbrace{\begin{bmatrix} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} \lambda_1 & 1 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 & 0 & 0 \\ 0 & 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_3 & 1 \\ 0 & 0 & 0 & 0 & 0 & \lambda_3 \end{bmatrix}}_{J} \underbrace{\begin{bmatrix} \widetilde{s}_1^T \\ \widetilde{s}_2^T \\ \widetilde{s}_3^T \\ \widetilde{s}_4^T \\ \widetilde{s}_5^T \\ \widetilde{s}_6^T \end{bmatrix}}_{S^{-1}}$$

  with $J_{\lambda_1} \in \mathbf{C}^{3\times 3}, J_{\lambda_2} \in \mathbf{C}^{1\times 1}, J_{\lambda_3} \in \mathbf{C}^{2\times 2}$

- spectrum of A is $\sigma(A) = \{\lambda_1, \lambda_1, \lambda_1, \lambda_2, \lambda_3, \lambda_3\}$

- s_1, s_4, s_5 are right-eigenvectors:

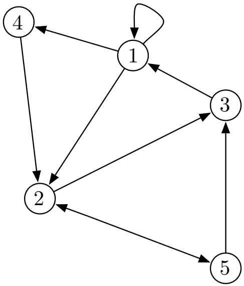
$$As_1 = \lambda_1 s_1, \quad As_4 = \lambda_2 s_4, \quad As_5 = \lambda_3 s_5$$

- $\widetilde{s}_3^T, \widetilde{s}_4^T, \widetilde{s}_6^T$ are left-eigenvectors:

$$\widetilde{s}_3^T A = \lambda_1 \widetilde{s}_3^T, \quad \widetilde{s}_4^T A = \lambda_2 \widetilde{s}_4^T, \quad \widetilde{s}_6^T A = \lambda_3 \widetilde{s}_6^T$$

Directed graphs

- example:



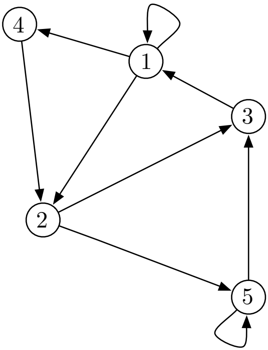
$\mathcal{V} = \{1, 2, 3, 4, 5\} \quad \mathcal{E} = \{(1,1), (1,2), (1,4), (2,3), (2,5), (3,1), (4,2), (5,2), (5,3)\}$

- we also use the notation: $i \to j \quad \Leftrightarrow \quad (i,j) \in \mathcal{E}$

- the graph induced by a matrix $A \in \mathbf{C}^{n\times n}$ is $\mathcal{G}(A) = (\mathcal{V}, \mathcal{E})$
  - $\mathcal{V} = \{1, 2, \ldots, n\}$
  - $\mathcal{E} = \{(i, j) : A_{ij} \neq 0\}$

- example:

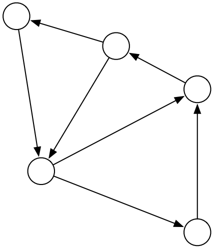
$$A = \begin{bmatrix} -0.1 & 0.7 & 0 & 1.3 & 0 \\ 0 & 0 & -0.4 & 0 & 0.2 \\ 1.5 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0.9 & 0 & -0.1 \end{bmatrix}$$



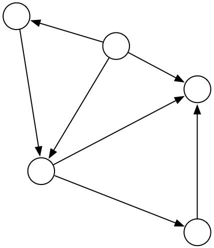
$\mathcal{G}(A)$

- Perron-Frobenius theorem reveals spectral properties of A from $\mathcal{G}(A)$

- a directed graph is connected[1] if there exists a path between any two nodes



connected graph



disconnected graph

[1] Some authors use the term *strongly connected*.

Perron-Frobenius theorem[2]

- if $A \geq 0$ and $\mathcal{G}(A)$ is connected, then:
  - $\rho(A)$ is an eigenvalue of A
  - there exists $v > 0$ such that

$$Av = \rho(A)v$$

  - there exists $w > 0, w^T v = 1$ such that

$$w^T A = \rho(A)w^T$$

  - if there are K eigenvalues $\{\lambda_1, \ldots, \lambda_K\}$ on the spectral circle, then $\lambda_k = \rho(A)e^{i\frac{2\pi}{K}(k-1)}$ and J_{λ_k} is 1×1 for all $k = 1, \ldots, K$.

[2] C. Meyer, *Matrix Analysis and Applied Linear Algebra*, ch. 8, p. 673 and p.676; R. Horn, C. R. Johnson, *Matrix Analysis*, ch. 8, Theorem 8.4.4, p. 508.

- interpretation: in terms of the Jordan canonical form of A,

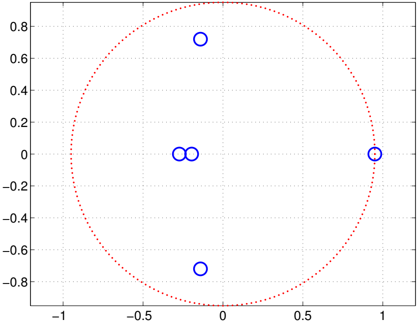
$$A = \underbrace{\begin{bmatrix} v & s_2 & \cdots & s_n \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} \rho(A) & & & \\ & J_{\lambda_2} & & \\ & & \ddots & \\ & & & J_{\lambda_p} \end{bmatrix}}_{J} \underbrace{\begin{bmatrix} w^T \\ \widetilde{s}_2^T \\ \vdots \\ \widetilde{s}_n^T \end{bmatrix}}_{S^{-1}}$$

with $\lambda_i \neq \rho(A)$ for $i = 2, \ldots, p$

- example:

$$A = \begin{bmatrix} 0.1 & 0.7 & 0 & 1.3 & 0 \\ 0 & 0 & 0.4 & 0 & 0.2 \\ 1.5 & 0 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0.9 & 0 & 0.1 \end{bmatrix}$$

- spectrum of A in the complex plane:



- Jordan canonical form: $A = SJS^{-1}$

$$
\begin{aligned}
S &= \begin{bmatrix} 0.3694 & * & * & * & * \\ 0.3754 & * & * & * & * \\ 0.5831 & * & * & * & * \\ 0.0395 & * & * & * & * \\ 0.6173 & * & * & * & * \end{bmatrix} \\
J &= \begin{bmatrix} 0.9593 & & & & \\ & * & & & \\ & & * & & \\ & & & * & \\ & & & & * \end{bmatrix} \\
S^{-1} &= \begin{bmatrix} 0.8248 & 0.7263 & 0.4675 & 1.1283 & 0.1708 \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}
\end{aligned}
$$

- definition: A is a primitive matrix if $A \geq 0$, $\mathcal{G}(A)$ is connected and $\rho(A)$ is the unique eigenvalue in the spectral circle of A

- if $A \geq 0$, $\mathcal{G}(A)$ is connected and $A_{ii} > 0$ for all i then A is primitive

Metropolis weights

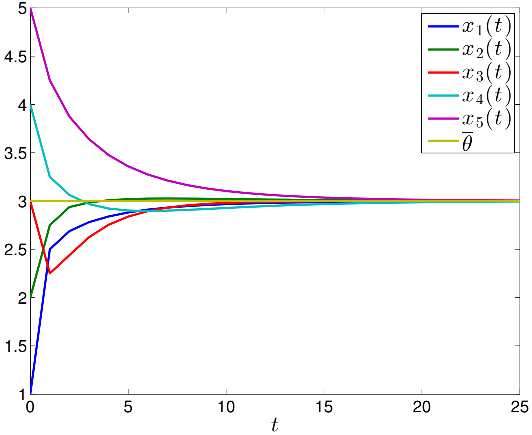
- for an undirected graph $\mathcal{G}$:

$$w_{ij} = \begin{cases} \frac{1}{1+\max\{d_i,d_j\}} & \text{, if } j \sim i \\ 1 - \sum_{j\sim i} w_{ij} & \text{, if } i = j \end{cases}$$

- properties of the matrix $W = (w_{ij})$:
  - $W = W^T$
  - $W\mathbf{1} = \mathbf{1}$
  - $W \geq 0$
  - $\rho(W) = 1$
  - $\sigma(W) = \{1, \lambda_2, \ldots, \lambda_n\}$ with $|\lambda_i| < 1$ for $i = 2, \ldots, n$

- conclusion: $W \in \mathcal{W}_{\mathcal{G}}$

- back to example in page 5 with a Metropolis matrix W:



To know more

- Optimizing the weight matrix for fast consensus
  - L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Sys. and Control Lett.*, 53, 2004
  - S. You, "A fast linear consensus protocol on an asymmetric directed graph," *American Control Conf.*, 2014.
- Laplacians
  - F. Chung, *Spectral graph theory*, ch. 1
  - M. Fiedler, "Algebraic connectivity of graphs", *Czech. Math. Journal*, 23 (98) 1973.
- Matrix analysis (Jordan forms, EVD, SVD)
  - R. Horn, C. R. Johnson, *Matrix Analysis*, ch. 1–5.
- Perron-Frobenius theory
  - C. Meyer, *Matrix Analysis and Applied Linear Algebra*, ch. 8.
  - R. Horn, C. R. Johnson, *Matrix Analysis*, ch. 8.