Fastest Mixing Markov Chain on a Graph

Stephen Boyd (Elec. Engr.)
Persi Diaconis (Stat. & Math)
Lin Xiao (Aero/Astro.)

Stanford University

MS&E 314, Semidefinite Programming
Stanford, 3/11/2003

Markov chain on a graph

- connected undirected graph $\mathcal{G} = (\mathcal{V},\ \mathcal{E})$

$$\mathcal{V} = \{1, \ldots, n\}, \qquad \mathcal{E} = \{(i,j) \mid i \text{ and } j \text{ connected}\}$$

  we'll assume each vertex has self-loop, *i.e.*, $(i,i) \in \mathcal{E}$

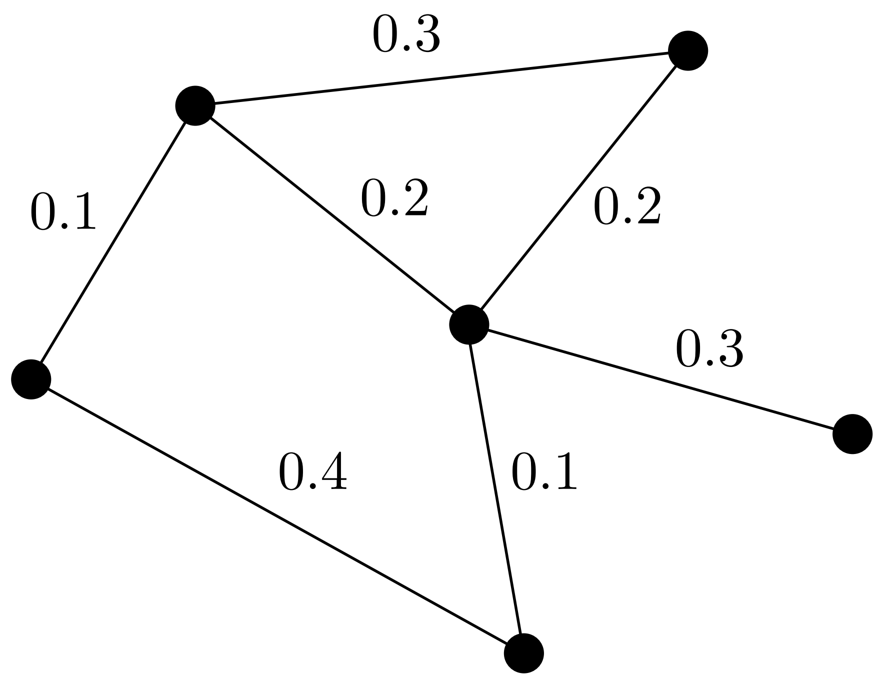
- define Markov chain on vertices $X(t) \in \{1, \ldots, n\}$, with transition probabilities

$$P_{ij} = \mathbf{Prob}(X(t+1) = j \mid X(t) = i)$$

  - each edge $(i,j) \in \mathcal{E}$ labeled with transition probability P_{ij}
  - we'll take $P_{ij} = 0$ for $(i,j) \notin \mathcal{E}$, and $P_{ij} = P_{ji}$

- P must satisfy $P_{ij} \geq 0$, $P\mathbf{1} = \mathbf{1}$, $P = P^T$, $P_{ij} = 0$ for $(i,j) \notin \mathcal{E}$

example:



self-loop transition probabilities not shown, given by

$$P_{ii} = 1 - \sum_{j \neq i} P_{ij}$$

Stationary distribution

- let $\pi_i(t) = \mathbf{Prob}(X(t) = i)$, then $\pi(t) = (\pi_1(t), \ldots, \pi_n(t))$ is the probability distribution at time t

$$\pi(t+1)^T = \pi(t)^T P \qquad\qquad \pi(t)^T = \pi(0)^T P^t$$

- stationary distribution π_{st} satisfies

$$\pi_{\mathrm{st}}^T P = \pi_{\mathrm{st}}^T$$

  the *global balance equation*

- since $P = P^T$ and $P\mathbf{1} = \mathbf{1}$, uniform distribution $\pi_{\mathrm{st}} = \mathbf{1}/n$ is stationary

$$\lim_{t\to\infty} \|\pi(t) - \mathbf{1}/n\| = 0$$

Mixing rate

- since $P = P^T$, all eigenvalues are real; can order as

$$1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq -1$$

- asymptotic rate of convergence to stationary distribution determined by second largest (in magnitude) eigenvalue, the *mixing rate*

$$\mu(P) = \max_{i=2,\ldots,n} |\lambda_i| = \max\{\lambda_2(P), -\lambda_n(P)\}$$

- distribution of $X(t)$ approaches uniform as μ^t (if $\mu < 1$), $e.g.$,

$$\sup_{\pi(0)} \|\pi(t) - \mathbf{1}/n\|_1 \leq \sqrt{n}\mu^t$$

  the smaller μ is, the faster the Markov chain mixes

Mixing time

- how long to wait for the Markov chain to be close to stationary?
- define *mixing time*

$$\tau = \frac{1}{\log(1/\mu)}$$

gives (asymptotic) time for norm of error to decrease by factor $1/e$

Fastest mixing Markov chain problem

fastest mixing Markov chain (FMMC) problem:

$$\begin{array}{ll} \text{minimize} & \mu(P) \\ \text{subject to} & P \geq 0, \quad P\mathbf{1} = \mathbf{1}, \quad P = P^T \\ & P_{ij} = 0, \quad (i,j) \notin \mathcal{E} \end{array}$$

- optimization variable is P; problem data is graph
- can add other constraints

another interpretation: find fastest mixing symmetric Markov chain with fixed sparsity pattern ($i.e.$, allowed transitions)

Background

- *Markov chain Monte Carlo simulation*, with applications in statistics, physics, chemistry, biology, computer science . . .
  - random sampling of a huge state space with a specified distribution
  - construct a Markov chain that converges asymptotically to the desired distribution
  - simulate the Markov chain until close to stationary, then use states of the chain as random samples
- efficiency of simulation determined by mixing rate
- **previous work**: bound the mixing rate with various techneques, and derive heuristics to obtain faster mixing chains
- **this talk**: find the fastest mixing Markov chain (and the mixing rate)
- limited by the size of pratical problems

Two common suboptimal schemes

let d_i be degree of vertex i, $i.e.$, number of edges connected to vertex i (not counting self-loops)

- maximum degree chain: with $d_{\max} = \max_{i \in \mathcal{V}} d_i$

$$P_{ij}^{\mathrm{md}} = \frac{1}{d_{\max}}, \quad i \neq j,\ (i,j) \in \mathcal{E}$$

- Metropolis-Hastings chain

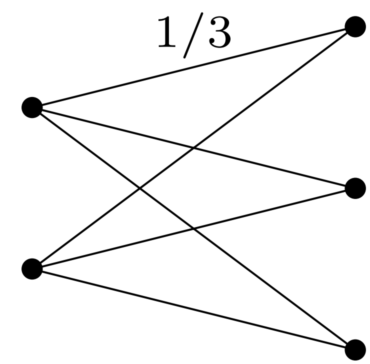
$$P_{ij}^{\mathrm{mh}} = \frac{1}{\max\{d_i, d_j\}}, \quad i \neq j,\ (i,j) \in \mathcal{E}$$

diagonal entries determined by $P_{ii} = 1 - \sum_{j \neq i} P_{ij}$

A simple example

- maximum degree and Metropolis-Hastings

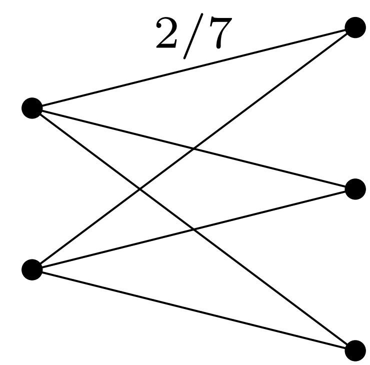
$$\mu^{\mathrm{md}} = \mu^{\mathrm{mh}} = 2/3$$



- can we do better? yes!

$$\mu^{\star} = 3/7$$

is, in fact, optimal for FMMC



- can we always find the best? how difficult is it?
  how suboptimal is maximum degree or Metropolis-Hastings?

Outline

- **convex optimization & SDP formulation of FMMC**
- examples
- Lagrange dual of FMMC and optimality conditions
- exploit structure in interior-point methods
- subgradient method
- extension to reversible Markov chains

Convexity of mixing rate

$\mu(P)$ is **convex function** of P

- variational characterization of $\mu(P)$:

$$\begin{aligned}
\mu(P) &= \max\{\lambda_2(P), -\lambda_n(P)\} \\
\lambda_2(P) &= \sup\{v^T P v \mid \|v\|_2 \le 1,\ \mathbf{1}^T v = 0\} \\
\lambda_n(P) &= \sup\{-v^T P v \mid \|v\|_2 \le 1,\ \mathbf{1}^T v = 0\}
\end{aligned}$$

- $\mu(P)$ is spectral norm of P on $\mathbf{1}^\perp = \{v \mid \mathbf{1}^T v = 0\}$:

$$\mu(P) = \left\|\left(I - (1/n)\mathbf{1}\mathbf{1}^T\right) P \left(I - (1/n)\mathbf{1}\mathbf{1}^T\right)\right\|_2 = \left\|P - (1/n)\mathbf{1}\mathbf{1}^T\right\|_2$$

- for $X = X^T$, $\lambda_1(X) + \lambda_2(X)$ and $-\lambda_n(X)$ are convex; here $\lambda_1 = 1$, so $\max\{\lambda_2(X), -\lambda_n(X)\}$ is convex

Convex optimization formulation of FMMC

$$
\begin{array}{ll}
\text{minimize} & \mu(P) = \left\| P - (1/n)\mathbf{1}\mathbf{1}^T \right\|_2 \\
\text{subject to} & P \geq 0, \quad P\mathbf{1} = \mathbf{1}, \quad P = P^T \\
& P_{ij} = 0, \quad (i,j) \notin \mathcal{E}
\end{array}
$$

- **convex optimization** problem
- nondifferentiable objective function, linear constraints
- hence, can solve efficiently; have duality theory, . . .

SDP formulation of FMMC

introducing a scalar variable s to bound the norm of $P - (1/n)\mathbf{1}\mathbf{1}^T$

$$
\begin{array}{ll}
\text{minimize} & s \\
\text{subject to} & -sI \preceq P - (1/n)\mathbf{1}\mathbf{1}^T \preceq sI \\
& P \geq 0, \quad P\mathbf{1} = \mathbf{1}, \quad P = P^T \\
& P_{ij} = 0, \quad (i,j) \notin \mathcal{E}
\end{array}
$$

a semidefinite program (SDP) in variables P, s

- $X \preceq Y$ means $Y - X$ is positive semidefinite
- $P \geq 0$ denotes elementwise inequality

Extensions

can add other convex constraints on the transition probabilities

fastest local degree chain: require probability on edge to be function of degrees of vertices:

$$P_{ij}^{\mathrm{ld}} = \phi(d_i, d_j), \quad i \neq j, \ (i,j) \in \mathcal{E}$$

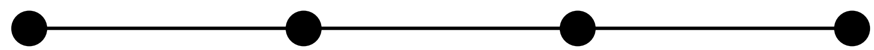
- diagonal entries determined by $P_{ii} = 1 - \sum_{j \neq i} P_{ij}$
- includes Metropolis-Hastings as special case $P_{ij} = 1/\max\{d_i, d_j\}$
- for convex/SDP formulation, add linear equality constraints

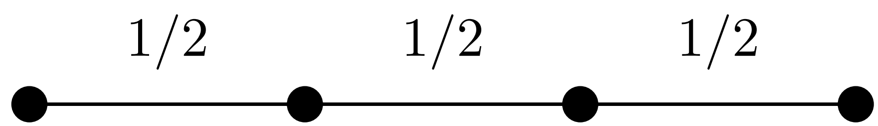
$$P_{ij} = P_{kl} \text{ whenever } d_i = d_k < d_j = d_l$$

Outline

- convex optimization & SDP formulation of FMMC
- **examples**
- Lagrange dual of FMMC and optimality conditions
- exploit structure in interior-point methods
- subgradient method
- extension to reversible Markov chains

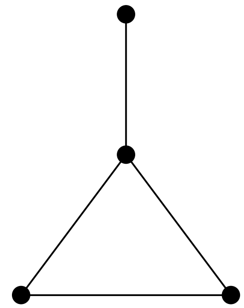
Small example (a)

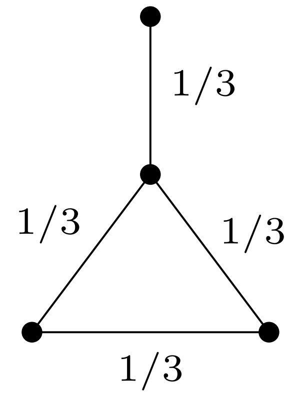




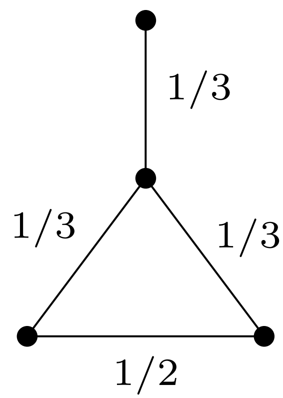
$$\mu^{\mathrm{md}} = \mu^{\mathrm{mh}} = \mu^{\mathrm{ld}} = \mu^{\star} = \lambda_2 = -\lambda_n = \sqrt{2}/2$$

Small example (b)

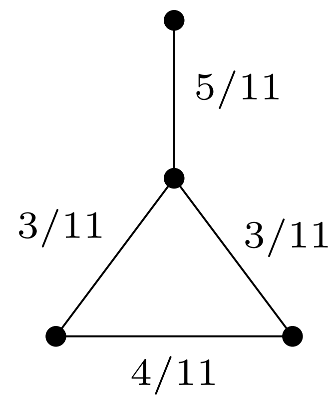




$$\mu^{\mathrm{md}} = \lambda_2 = 2/3$$

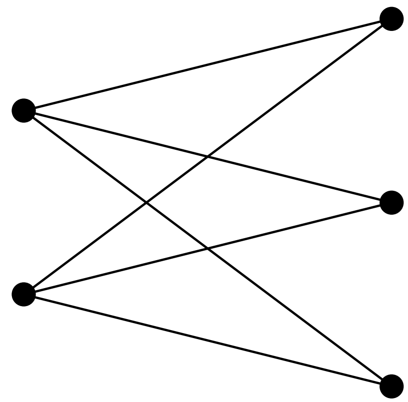


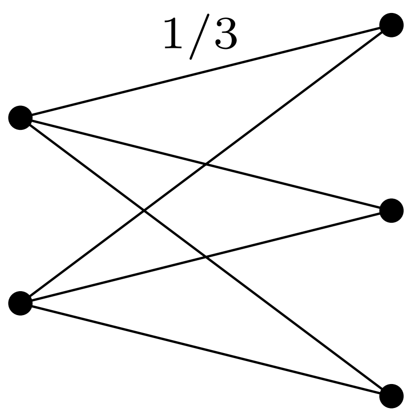
$$\mu^{\mathrm{mh}} = \lambda_2 = 2/3$$

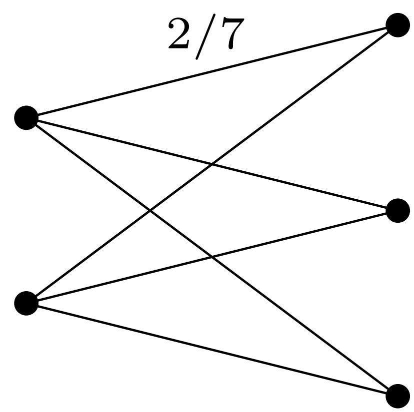


$$\mu^{\mathrm{ld}} = \mu^{\star} = \lambda_2 = 7/11$$

Small example (c)

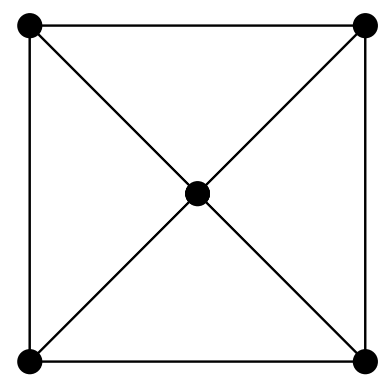


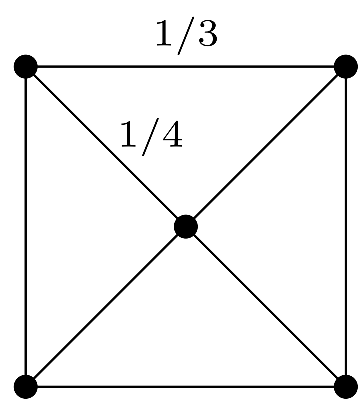


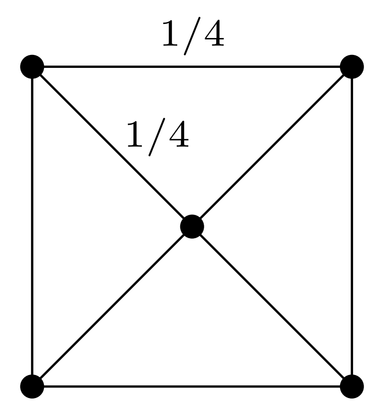


$$\mu^{\rm md} = \mu^{\rm mh} = -\lambda_n = 2/3 \qquad \mu^{\rm ld} = \mu^{\star} = \lambda_2 = -\lambda_n = 3/7$$

Small example (d)





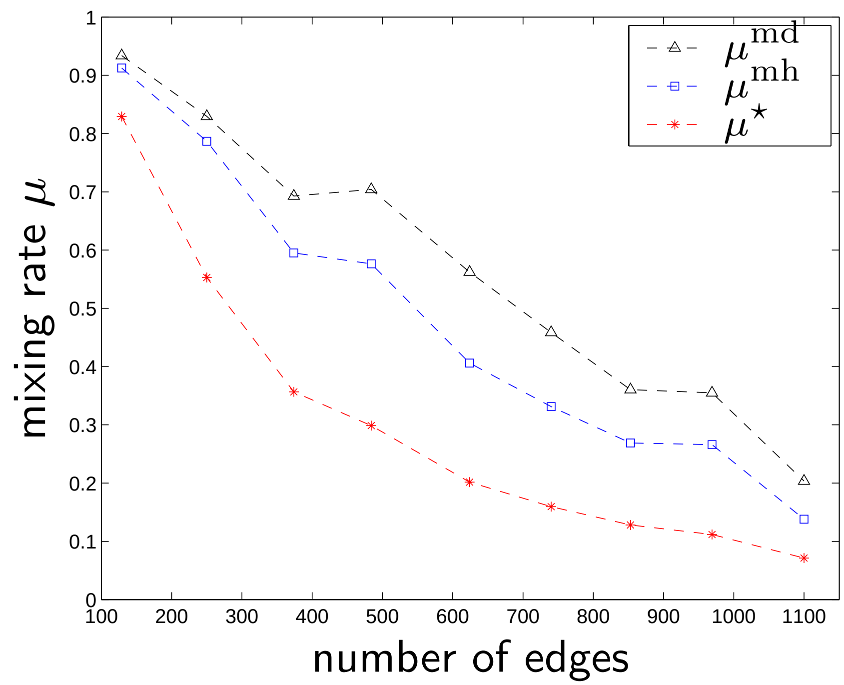


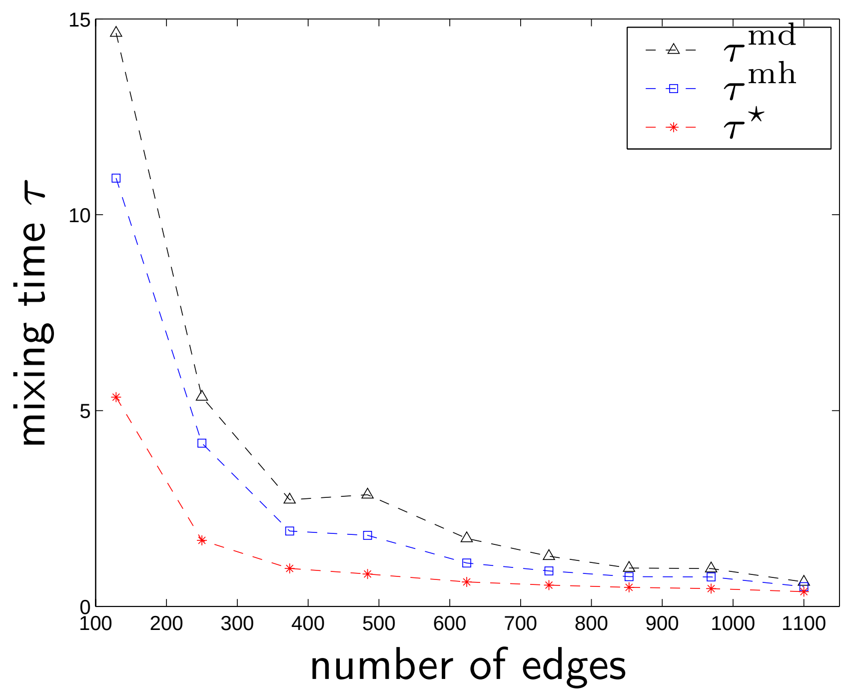
$$\mu^{\mathrm{mh}} = 7/12$$

$$\mu^{\mathrm{md}} = \mu^{\star} = 1/4$$

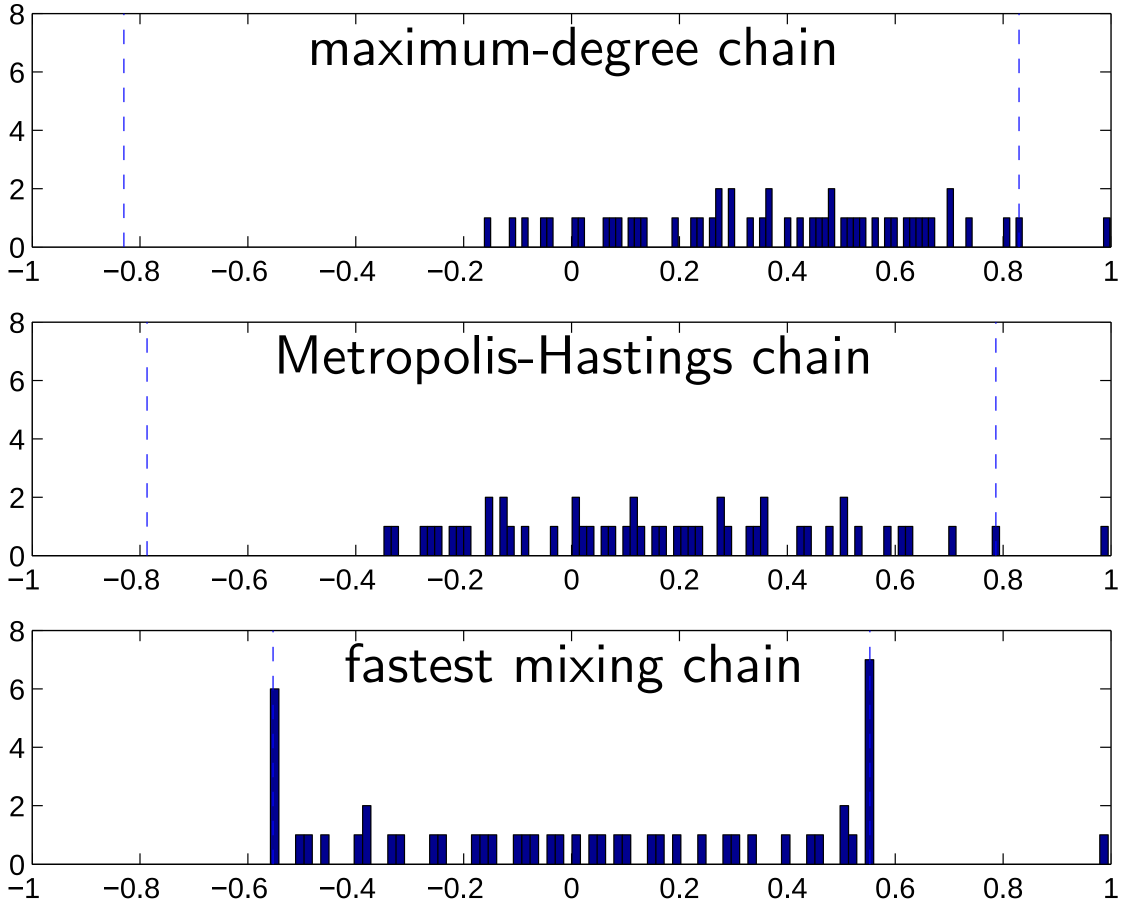
A random family with 50 vertices

- randomly generate symmetric matrix $A \in \mathbf{R}^{50\times 50}$, where A_{ij}, for $i \leq j$, are i.i.d. uniformly distributed on interval $[0,1]$
- for each threshold $c \in [0,1]$, place an edge between i and j if $A_{ij} \leq c$
- let $c = 0.1, 0.2, \ldots, 0.9$, obtain a montone family of graphs





Eigenvalue distribution



for the graph with $c = 0.2$. The dashed lines indicate $\pm\mu$ for each chain.

Outline

- convex optimization & SDP formulation of FMMC
- examples
- **Lagrange dual of FMMC and optimality conditions**
- exploit structure in interior-point methods
- subgradient method
- extension to reversible Markov chains

Dual of FMMC problem

primal FMMC:

$$\begin{array}{ll} \text{minimize} & \mu(P) = \left\|P - (1/n)\mathbf{1}\mathbf{1}^T\right\|_2 \\ \text{subject to} & P \geq 0, \quad P\mathbf{1} = \mathbf{1}, \quad P = P^T \\ & P_{ij} = 0, \quad (i,j) \notin \mathcal{E} \end{array}$$

dual FMMC (with variables Y, z):

$$\begin{array}{ll} \text{maximize} & \mathbf{1}^T z \\ \text{subject to} & Y\mathbf{1} = 0, \qquad Y = Y^T \\ & \|Y\|_* = \sum_{i=1}^n |\lambda_i(Y)| \leq 1 \\ & (z_i + z_j)/2 \leq Y_{ij}, \quad (i,j) \in \mathcal{E} \end{array}$$

($\|\cdot\|_*$ is indeed the dual of the spectral norm, *nuclear norm*)

Weak duality

if P primal feasible, and $Y,\ z$ dual feasible, then $\mathbf{1}^T z \leq \mu(P)$

quick proof:

$$\begin{array}{rcl}
\mathbf{Tr}\, Y \left(P - (1/n)\mathbf{1}\mathbf{1}^T\right) & \leq & \|Y\|_* \|P - (1/n)\mathbf{1}\mathbf{1}^T\|_2 \\
& \leq & \|P - (1/n)\mathbf{1}\mathbf{1}^T\|_2 \\
& = & \mu(P)
\end{array}$$

$$\begin{array}{rcl}
\mathbf{Tr}\, Y \left(P - (1/n)\mathbf{1}\mathbf{1}^T\right) & = & \mathbf{Tr}\, Y P = \sum_{i,j} Y_{ij} P_{ij} \\
& \geq & \sum_{i,j} (1/2)(z_i + z_j) P_{ij} \\
& = & (1/2)(z^T P \mathbf{1} + \mathbf{1}^T P z) \\
& = & \mathbf{1}^T z
\end{array}$$

Strong duality

- primal and dual FMMC problems are solvable, and have same optimal value

- there are primal feasible $P^\star$, and dual feasible $Y^\star,\ z^\star$ with

$$\|P^\star - (1/n)\mathbf{1}\mathbf{1}^T\|_2 = \mathbf{1}^T z^\star$$

Optimality conditions

- primal feasibility

$$P \geq 0, \quad P = P^T, \quad P\mathbf{1} = \mathbf{1}, \quad P_{ij} = 0 \text{ for } (i,j) \notin \mathcal{E}$$

- dual feasibility

$$Y = Y^T, \quad Y\mathbf{1} = 0, \quad \|Y\|_* \leq 1, \quad (z_i + z_j)/2 \leq Y_{ij} \text{ for } (i,j) \in \mathcal{E}$$

- complementary slackness

$$\begin{aligned}
&((z_i + z_j)/2 - Y_{ij})\, P_{ij} = 0 \\
&Y = Y_+ - Y_-, \quad Y_+ = Y_+^T \succeq 0, \quad Y_- = Y_-^T \succeq 0 \\
&\mathbf{Tr}\, Y_+ + \mathbf{Tr}\, Y_- = 1 \\
&PY_+ = \mu(P) Y_+, \quad PY_- = -\mu(P) Y_-
\end{aligned}$$

Outline

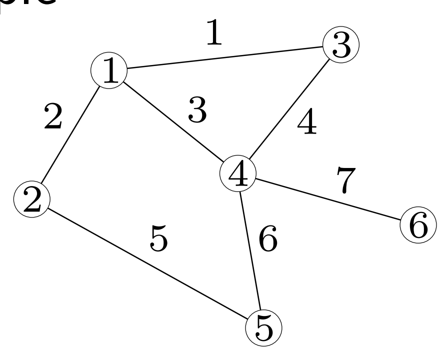
- convex optimization & SDP formulation of FMMC
- examples
- Lagrange dual of FMMC and optimality conditions
- **exploit structure in interior-point methods**
- subgradient method
- extension to reversible Markov chains

The incidence matrix

- label edges (not including self-loops) by $l = 1, \ldots, m$; and define vertex-edge *incidence matrix* $A \in \mathbf{R}^{n \times m}$

$$A_{il} = \begin{cases} 1 & \text{if edge } l \text{ starts from vertex } i \\ -1 & \text{if edge } l \text{ ends at vertex } i \\ 0 & \text{otherwise.} \end{cases}$$

- example



	1	2	3	4	5	6	7
1	1	1	-1	0	0	0	0
2	0	-1	0	0	-1	0	0
3	-1	0	0	1	0	0	0
4	0	0	1	-1	0	1	1
5	0	0	0	0	1	-1	0
6	0	0	0	0	0	0	-1

- the columns of A

$$a_l = e_i - e_j, \qquad l \sim (i, j)$$

An alternative representation

- introduce $p \in \mathbf{R}^m$, with p_l being transition probability on edge l, then

$$P(p) = I - A\,\mathbf{diag}(p)A^T = I - \sum_{l=1}^{m} p_l a_l a_l^T$$

$$p \geq 0, \qquad |A|p \leq \mathbf{1}$$

$|A|$: elementwise absolute value; edge directions make no difference

- FMMC problem in terms of new variable $p \in \mathbf{R}^m$

$$\begin{array}{ll} \text{minimize} & \left\| I - A\,\mathbf{diag}(p)A^T - (1/n)\mathbf{1}\mathbf{1}^T \right\|_2 \\ \text{subject to} & p \geq 0, \quad |A|p \leq \mathbf{1} \end{array}$$

The centering problem

- SDP formulation of FMMC

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & -sI \preceq I - A\,\mathbf{diag}(p)A^T - (1/n)\mathbf{1}\mathbf{1}^T \preceq sI \\ & p \geq 0, \quad |A|p \leq \mathbf{1} \end{array}$$

- the centering problem with logarithmic barrier functions

$$\begin{array}{ll} \text{minimize} & ts - \log\det\left(sI - (I - A\,\mathbf{diag}(p)A^T) + (1/n)\mathbf{1}\mathbf{1}^T\right) \\ & -\log\det\left(sI + (I - A\,\mathbf{diag}(p)A^T) - (1/n)\mathbf{1}\mathbf{1}^T\right) \\ & -\displaystyle\sum_{l=1}^{m} \log p_l - \sum_{i=1}^{n} \log\left(1 - (|A|p)_i\right) \end{array}$$

as t becomes large enough, get good approximation of FMMC

Exploit sparse structure

- use Newton's method to solve the centering problem
- frequent computing of $(sI - P + \mathbf{1}\mathbf{1}^T/n)^{-1}$ and $(sI + P - \mathbf{1}\mathbf{1}^T/n)^{-1}$
- structure: **sparse** + **rank-one**
- use Sherman-Woodbury-Morrison formula, *e.g.*,

$$(sI - P + \mathbf{1}\mathbf{1}^T/n)^{-1} = (sI - P)^{-1} - \frac{(sI - P)^{-1}\mathbf{1}\mathbf{1}^T(sI - P)^{-1}}{n + \mathbf{1}^T(sI - P)^{-1}\mathbf{1}}$$

can efficiently compute $(sI - P)^{-1}$ by sparse factorization

Assembly gradient and Hessian

- Newton step v computed by solving $Hv = -g$
- for convenience, define two matrices (note $P = I - \sum_{l=1}^m p_l a_l a_l^T$)

$$
\begin{aligned}
U &= \left(sI + P - (1/n)\mathbf{1}\mathbf{1}^T\right)^{-1} \\
V &= \left(sI - P + (1/n)\mathbf{1}\mathbf{1}^T\right)^{-1}
\end{aligned}
$$

- gradient $g = (g_0, g_1, \ldots, g_m)$, subscript 0 indicates variable s

$$
\begin{aligned}
g_0 &= t - \mathbf{Tr}\, U - \mathbf{Tr}\, V \\
g_l &= \mathbf{Tr}(U a_l a_l^T) - \mathbf{Tr}(V a_l a_l^T) = a_l^T U a_l - a_l^T V a_l \\
&= (U_{ii} + U_{jj} - 2U_{ij}) - (V_{ii} + V_{jj} - 2V_{ij})
\end{aligned}
$$

- Hessian $H \in \mathbf{R}^{(m+1)\times(m+1)}$

$$
\begin{aligned}
H_{00} &= \mathbf{Tr}\, U^2 + \mathbf{Tr}\, V^2 \\
H_{0l} &= -\,\mathbf{Tr}(U a_l a_l^T U) + \mathbf{Tr}(V a_l a_l^T V) \\
&= -a_l^T U^2 a_l + a_l^T V^2 a_l \\
&= -((U^2)_{ii} + (U^2)_{jj} - 2(U^2)_{ij}) + ((V^2)_{ii} + (V^2)_{jj} - 2(V^2)_{ij}) \\
H_{ll'} &= \mathbf{Tr}(U a_l a_l^T U a_{l'} a_{l'}^T) + \mathbf{Tr}(V a_l a_l^T V a_{l'} a_{l'}^T) \\
&= (a_l U a_{l'})^2 + (a_l V a_{l'})^2 \\
&= (U_{ii'} + U_{jj'} - U_{ij'} - U_{i'j})^2 + (V_{ii'} + V_{jj'} - V_{ij'} - V_{i'j})^2
\end{aligned}
$$

edges denoted by $l \sim (i,j)$, $l' \sim (i',j')$

Outline

- convex optimization & SDP formulation of FMMC
- examples
- Lagrange dual of FMMC and optimality conditions
- exploit structure in interior-point methods
- **subgradient method**
- extension to reversible Markov chains

Solution methods

- for small FMMC problems, up to 1000 variables: standard SDP solvers
- can exploit structure to gain efficiency, $e.g.$, sparsity in P
- large problems up to 100000 edges: subgradient method

Subgradient of μ

- mixing rate $\mu(P) = \|P - (1/n)\mathbf{1}\mathbf{1}^T\|_2$ nondifferentiable, convex
- $G = G^T$ is a **subgradient** of μ at P if for all $\tilde{P} = \tilde{P}^T$ and $\tilde{P}\mathbf{1} = \mathbf{1}$

$$\mu(\tilde{P}) \geq \mu(P) + \mathbf{Tr}\, G(\tilde{P} - P)$$

- for example, if $\mu(P) = \lambda_2(P)$ and u is associated unit eigenvector, then

$$\mu(P) = \lambda_2(P) = u^T P u$$

$$\mu(\tilde{P}) \geq \lambda_2(\tilde{P}) \geq u^T \tilde{P} u$$

subtracting the both sides of equality from inequality, we have

$$\mu(\tilde{P}) \geq \mu(P) + u^T(\tilde{P} - P)u = \mu(P) + \mathbf{Tr}(uu^T)(\tilde{P} - P)$$

so $G = uu^T$ is a subgradient of μ at P

- if $\mu(P) = -\lambda_n(P)$ and v is associated unit eigenvector, then a subgradient is given by $G = -vv^T$

Subdifferential of μ

subdifferential $\partial\mu$ at P is the set of subgradients

$$\begin{aligned}
\partial\lambda^\star(P) &= \mathbf{Co}(\{vv^T \mid Pv = \mu v,\ \|v\|_2 = 1\} \\
&\qquad \cup\{-vv^T \mid Pv = -\mu v,\ \|v\|_2 = 1\}) \\
&= \{Y \mid Y = Y_+ - Y_-,\ Y_+ = Y_+^T \succeq 0,\ Y_- = Y_-^T \succeq 0, \\
&\qquad \mathbf{Tr}\, Y_+ + \mathbf{Tr}\, Y_- = 1,\ PY_+ = \mu Y_+,\ PY_- = -\mu Y_-\}
\end{aligned}$$

related to optimality conditions (KKT)

Subgradient of μ as function of p

- mixing rate $\mu(p) = \|I - \sum_{l=1}^{m} p_l a_l a_l^T - (1/n)\mathbf{1}\mathbf{1}^T\|_2$
- if G is a subgradient of $\mu(P)$, then a subgradient of $\mu(p)$ is

$$g(p) = (-a_1^T G a_1, \ldots, -a_m^T G a_m)$$

  if $\mu(P) = \lambda_2(P)$ and u is associated unit eigenvector, then $G = uu^T$

$$g_l(p) = -(u_i - u_j)^2, \qquad l \sim (i,j), \quad l = 1, \ldots, m$$

  if $\mu(P) = -\lambda_n(P)$ and v is associated unit eigenvector, then $G = -vv^T$

$$g_l(p) = (v_i - v_j)^2, \qquad l \sim (i,j), \quad l = 1, \ldots, m$$

- for large sparse matrix, can compute λ_2, λ_n and associated eigenvectors very efficiently by Lanczos methods

Subgradient method

given a feasible $p^{(1)}$ at $k = 1$ ($e.g.$, maximum-degree or M-H chain)

repeat

1. Compute a subgradient $g^{(k)}$ of μ at $p^{(k)}$, and set

$$p^{(k+1)} = p^{(k)} - \alpha_k g^{(k)} / \|g^{(k)}\|$$

2. *approximately project* $p^{(k+1)}$ onto $\{p \mid p \geq 0, \ |A|p \leq \mathbf{1}\}$
3. $k := k + 1$

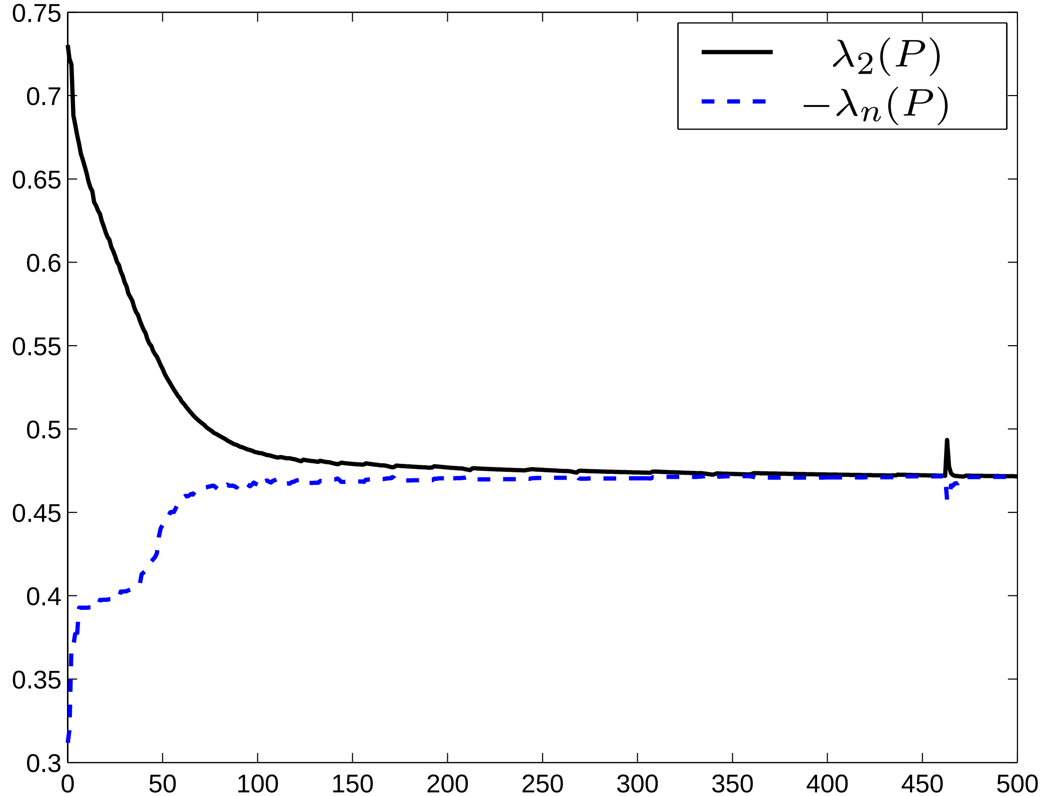
- step lengths satisfy the *diminishing rule*:

$$\alpha_k \geq 0, \qquad \lim_{k \to \infty} \alpha_k = 0, \qquad \sum_{k=1}^{\infty} \alpha_k = \infty$$

A large example using subgradient method

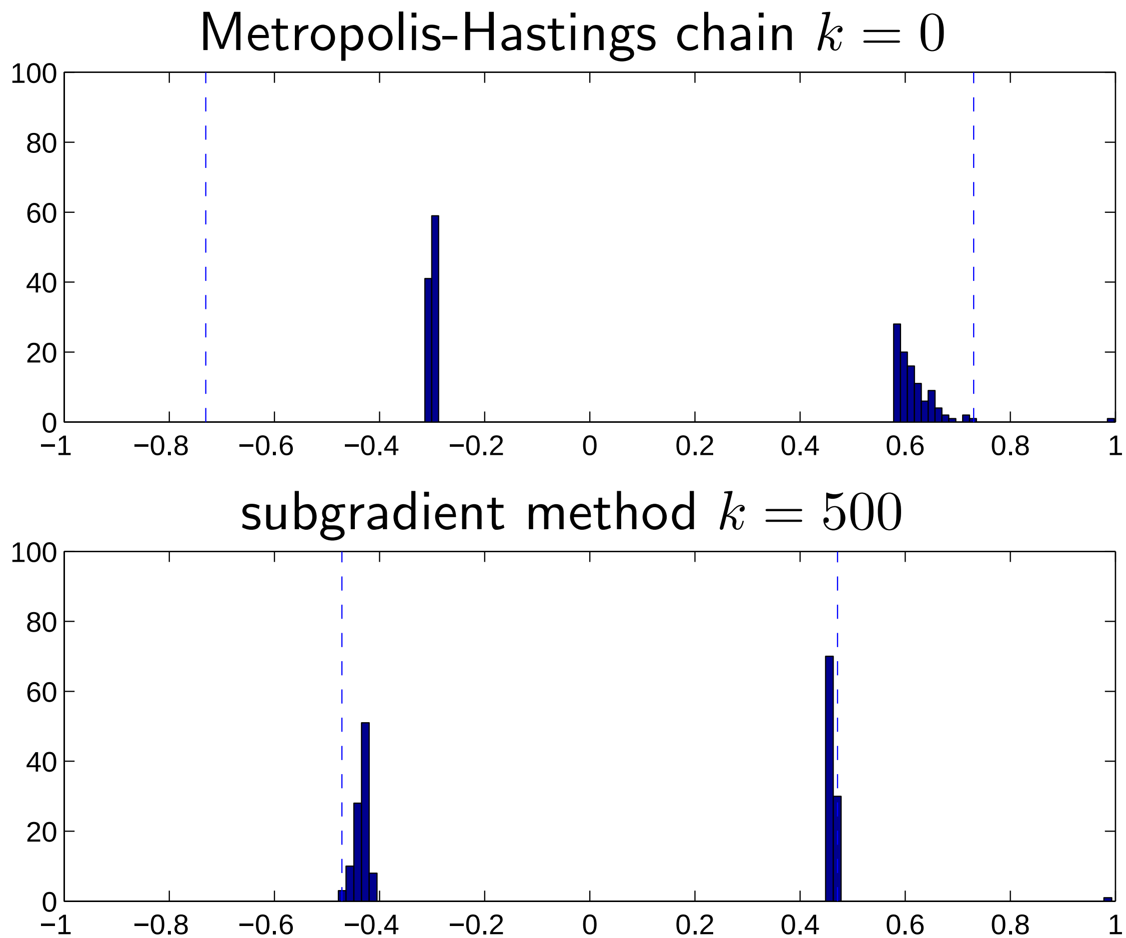
random graph with 10000 vertices and 100000 edges; $\alpha_k = 1/\sqrt{k}$

starting point: Metropolis-Hastings chain (with $\mu = 0.73$)



iteration number

Eigenvalue distribution



Distribution of the 101 largest and 100 smallest eigenvalues

Outline

- convex optimization & SDP formulation of FMMC
- examples
- Lagrange dual of FMMC and optimality conditions
- exploit structure in interior-point methods
- subgradient method
- **extension to reversible Markov chains**

Extension: fastest mixing to nonuniform distribution

- we are given desired equilibrium distribution $\pi = (\pi_1, \ldots, \pi_n)$
- we consider P with same sparsity pattern as graph, but not symmetric
- we do require **reversible** chain:

$$\pi_i P_{ij} = \pi_j P_{ji}, \qquad i, j = 1, \ldots, n$$

  the *detailed balance equation*

- let $\Pi = \mathbf{diag}(\pi)$, then detailed balance is equivalent to

$$\Pi P = P^T \Pi$$

- the matrix $\Pi^{-1/2} P \Pi^{1/2}$ is symmetric, with same eigenvalues as P

- eigenvector of $\Pi^{-1/2} P \Pi^{1/2}$ associated with maximum eigenvalue (which is one) is

$$q = (\sqrt{\pi_1}, \ldots, \sqrt{\pi_n})$$

- asymptotic rate of convergence of distribution to π determined by

$$\mu(P) = \left\| \Pi^{-1/2} P \Pi^{1/2} - qq^T \right\|_2$$

which is convex in P

- SDP formulation of fastest mixing reversible Markov chain

$$\begin{array}{ll}
\text{minimize} & s \\
\text{subject to} & -sI \preceq \Pi^{-1/2} P \Pi^{1/2} - qq^T \preceq sI \\
& P \geq 0, \quad P\mathbf{1} = \mathbf{1}, \quad \Pi P = P^T \Pi \\
& P_{ij} = 0, \quad (i,j) \notin \mathcal{E}
\end{array}$$

Summary

FMMC problem (and many variations) are convex problems, in fact SDPs

- can solve modest problems exactly and easily
- can solve larger problems via subgradient method

Current research

- exploit symmetry: reduce number of variables and matrix dimensions
- fast distributed averaging, fast resource allocation over networks