Fastest Mixing Markov Chain on a Graph

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

Stanford University

SDP workshop, MSRI
Berkeley, 10/7/2002

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.*, $(i,i) \in \mathcal{E}$

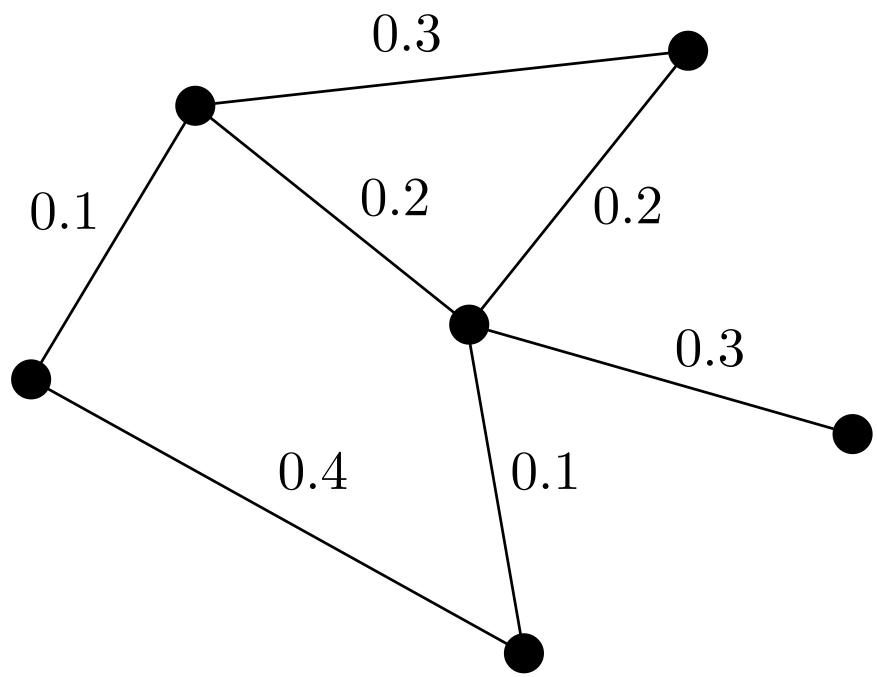
- each edge $(i,j) \in \mathcal{E}$ labeled with transition probability P_{ij}; we'll take $P_{ij} = 0$ for $(i,j) \notin \mathcal{E}$, and $P_{ij} = P_{ji}$

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

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

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

example:



self-loop transition probabilities not shown; $P_{ii} = 1 - \sum_{j \neq i} P_{ji}$

since $P = P^T$, uniform distribution $\pi_i = \mathbf{1}/n$ is stationary

Mixing rate

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

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$$

  $\lambda_1(P) = 1$; $|\lambda_i| \leq 1$ for $i \neq 1$

- asymptotic rate of convergence to equilibrium distribution determined by second largest (in magnitude) eigenvalue

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

- distribution of $X(t)$ approaches uniform as $\lambda^\star(P)^t$ (if $\lambda^\star(P) < 1$)

- the smaller $\lambda^\star(P)$ is, the faster the Markov chain mixes

Fastest mixing Markov chain problem

fastest mixing Markov chain (FMC) problem:

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

- 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)

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 $d_{\max} = \max_{i \in \mathcal{V}} d_i$

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

- Metropolis-Hastings chain

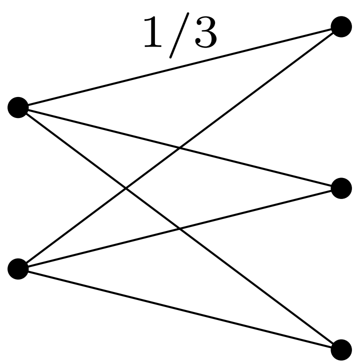
$$P_{ij} = \min\left\{\frac{1}{d_i},\ \frac{1}{d_j}\right\}, \quad i \neq j,\ (i,j) \in \mathcal{E}$$

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

A simple example

- maximum degree and Metropolis-Hastings

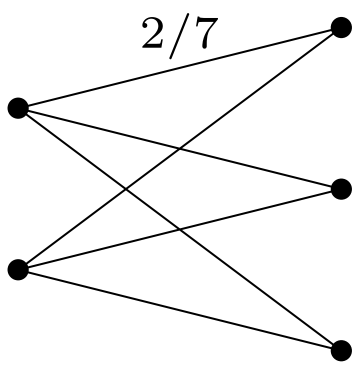
$$\lambda^\star_{\mathrm{md}} = \lambda^\star_{\mathrm{mh}} = 2/3$$



- can we do better? yes!

$$\lambda^\star_{\mathrm{opt}} = 3/7$$

is, in fact, optimal for FMC



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

Outline

- convex optimization & SDP formulation of FMC
- examples
- subgradient method
- Lagrange dual of FMC and interpretations
- optimality conditions
- extension to reversible Markov chains

Convexity of mixing rate

$\lambda^\star(P)$ is **convex function** of P

- variational characterization of $\lambda^\star(P)$:

$$\begin{aligned}\lambda^\star(P) &= \max\{\lambda_2(P), -\lambda_n(P)\} \\ &= \max\big\{\sup\{v^TPv \mid \|v\| \leq 1,\ v \in \mathbf{1}^\perp\}, \\ &\qquad \sup\{-v^TPv \mid \|v\| \leq 1,\ v \in \mathbf{1}^\perp\}\big\}\end{aligned}$$

- $\lambda^\star(P)$ is spectral norm of P on $\mathbf{1}^\perp$:

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

- 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 FMC

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

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

SDP formulation of FMC

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

a semidefinite program (SDP) in variables P, s

Extensions

can add other convex constraints on the transition probabilities

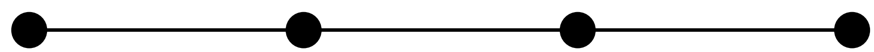
fastest local degree chain: require probability on edge to be function of degrees of vertices:

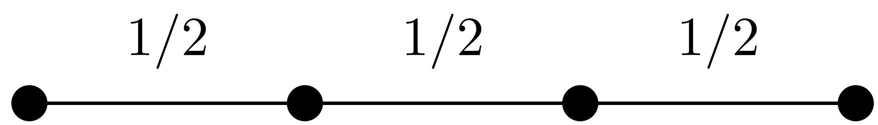
$$P_{ij} = \phi(d_i, d_j), \quad i \neq j,\ (i,j) \in \mathcal{E}$$

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

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

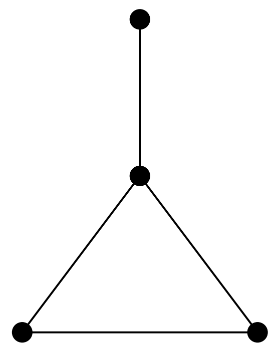
Small example (a)

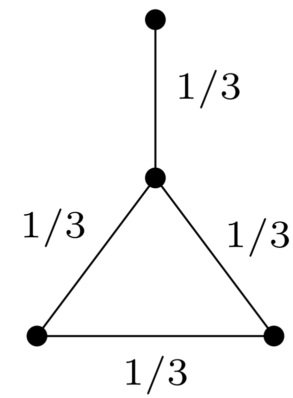




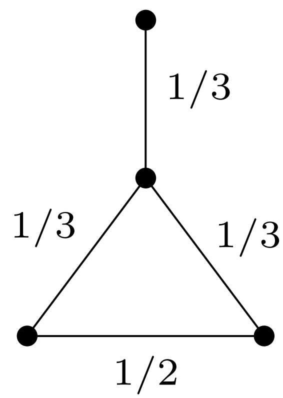
$$\lambda^\star_{\rm md} = \lambda^\star_{\rm mh} = \lambda^\star_{\rm ld} = \lambda^\star_{\rm opt} = \lambda_2 = -\lambda_n = \sqrt{2}/2$$

Small example (b)

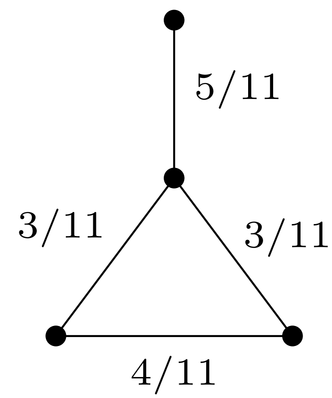




$$\lambda^\star_{\rm md} = \lambda_2 = 2/3$$

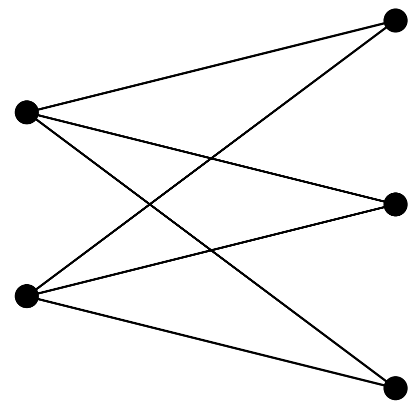


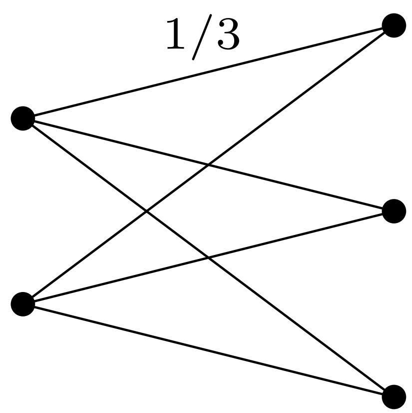
$$\lambda^\star_{\rm mh} = \lambda_2 = 2/3$$

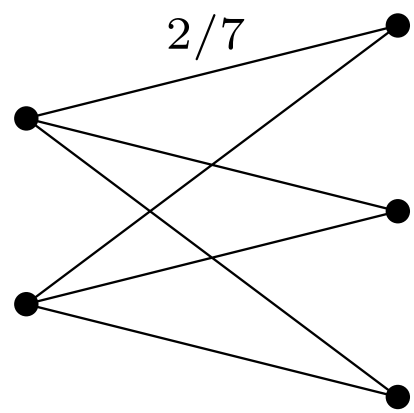


$$\lambda^\star_{\rm ld} = \lambda^\star_{\rm opt} = \lambda_2 = 7/11$$

Small example (c)

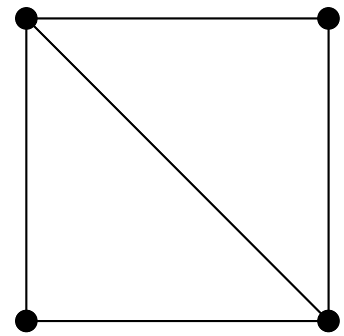


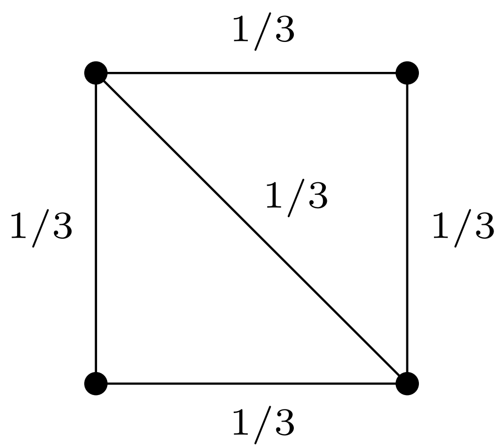


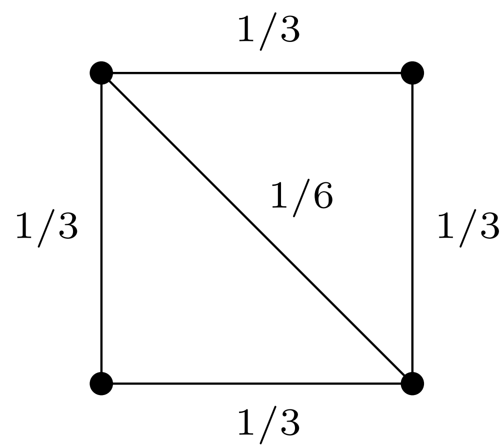


$\lambda^\star_{\rm md} = \lambda^\star_{\rm mh} = -\lambda_n = 2/3$ $\qquad$ $\lambda^\star_{\rm ld} = \lambda^\star_{\rm opt} = \lambda_2 = -\lambda_n = 3/7$

Small example (d)



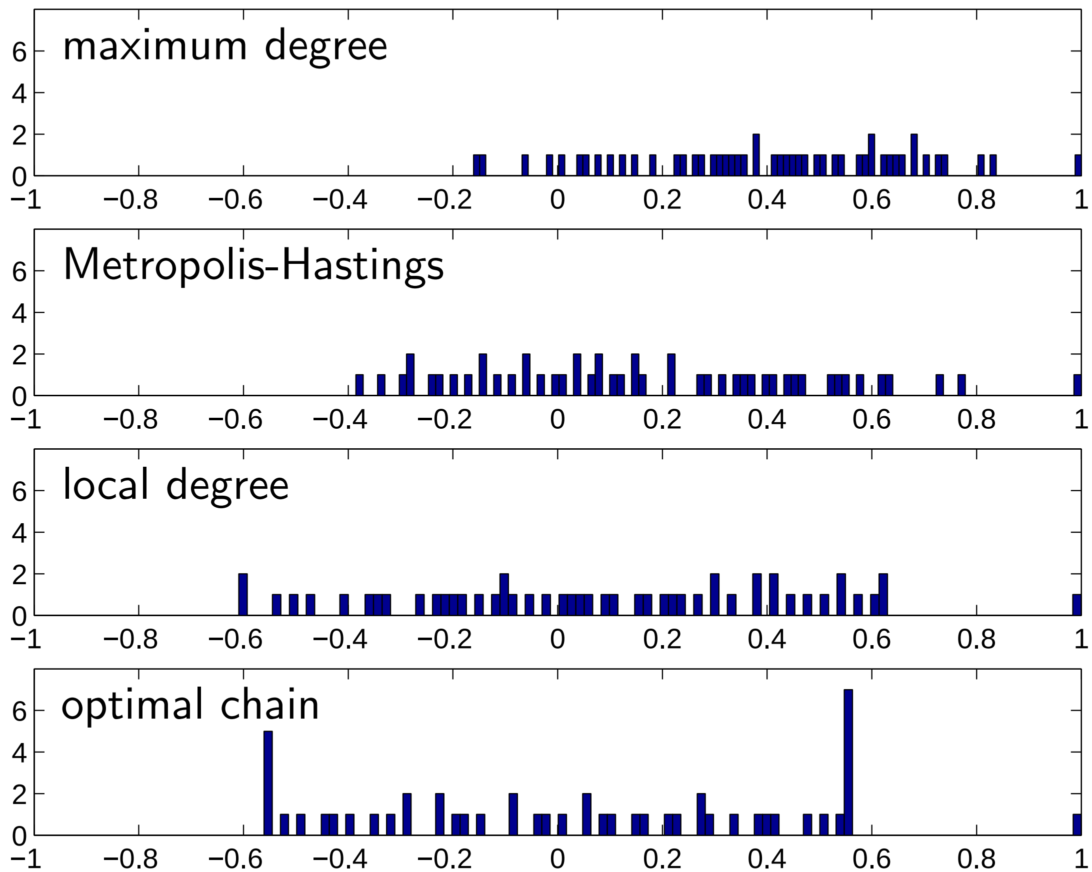




lefthand chain is Metropolis-Hastings and maximum degree; both are optimal, with $\lambda^\star = \lambda_2 = -\lambda_n = 1/3$

A larger example

random graph with 50 vertices and 226 edges (276 transition probabilities)



eigenvalue distributions

Solution methods

- for small FMC problems, up to 1000 variables: standard SDP solvers
- local degree FMC: can exploit sparsity in P, other problem structure
- large problems: subgradient method

Subdifferential of $\lambda^\star$

$G = G^T$ is a **subgradient** of $\lambda^\star$ at P if for all $\tilde{P} = \tilde{P}^T$,

$$\lambda^\star(\tilde{P}) \geq \lambda^\star(P) + \sum_{i,j} G_{ij}(\tilde{P}_{ij} - P_{ij})$$

subdifferential $\partial\lambda^\star$ at P is set of subgradients

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

Computing a subgradient

we'll use **free variables** P_{ij}, $i < j$, $(i, j) \in \mathcal{E}$ $(i.e.$, edge probabilities)

to find a subgradient w.r.t. free variable P_{ij}:

if $\lambda_2 = \lambda^\star$,

- find unit eigenvector u associated with λ_2
- $G_{ij} = -(u_i - u_j)^2$

otherwise $(i.e., -\lambda_n = \lambda^\star)$,

- find unit eigenvector u associated with λ_n
- $G_{ij} = (u_i - u_j)^2$

can use efficient method to compute λ_2, λ_n, and associated eigenvectors, for large sparse matrix

Subgradient method

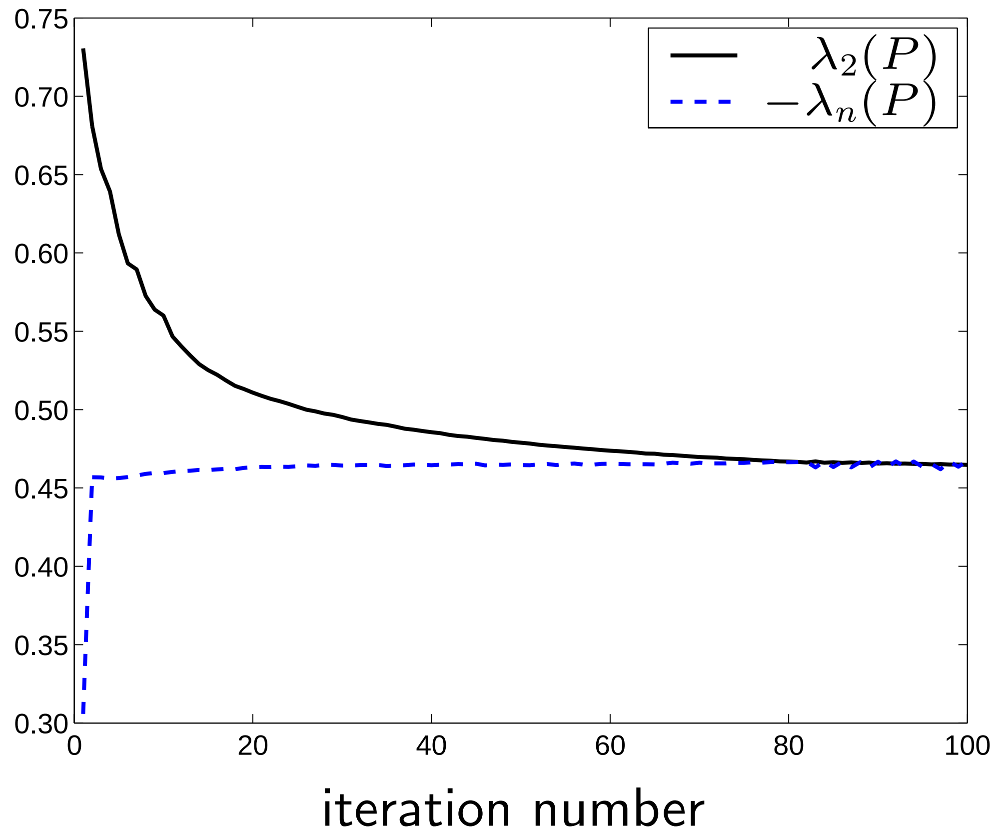
repeat:

- find a subgradient G w.r.t. free variables, at iterate $P^{(k)}$
- update: $P_{ij}^{(k+1)} = P_{ij}^{(k)} - \alpha_k G_{ij}$
- (approximately) project $P_{ij}^{(k+1)}$ back to feasible set

step lengths satisfy $\alpha_k \geq 0$, $\alpha_k \rightarrow 0$, $\sum_k \alpha_k = \infty$

A large example using subgradient method

random graph with 1000 vertices and 10000 edges; step length $\alpha_k = 1/\sqrt{k}$
starting point: Metropolis-Hastings (with $\lambda^\star = 0.73$)



Dual of FMC problem

primal FMC:

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

dual FMC (with variables Y, z):

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

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

Weak duality

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

quick proof:

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

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

Strong duality

- primal and dual FMC problems are solvable, and have same optimal value
- there are primal feasible $P^\star$, and dual feasible $Y^\star,\ z^\star$ with $\|P^\star - (1/n)\mathbf{1}\mathbf{1}^T\| = \mathbf{1}^T z^\star$

Optimality conditions

- primal feasibility

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

- dual feasibility

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

- complementary slackness

$$\left((z_i^\star + z_j^\star)/2 - Y^\star_{ij}\right) P^\star_{ij} = 0$$
$$Y^\star = V^\star - W^\star, \quad V^\star = V^{\star T} \succeq 0, \quad W^\star = W^{\star T} \succeq 0$$
$$P^\star V^\star = \lambda^\star V^\star, \quad P^\star W^\star = -\lambda^\star W^\star$$

Interpretation of dual FMC

fix variable Y in dual FMC, to obtain linear program (LP) with variable z

$$\begin{array}{ll} \text{maximize} & \mathbf{1}^T z \\ \text{subject to} & (z_i + z_j)/2 \leq Y_{ij}, \quad (i,j) \in \mathcal{E} \end{array}$$

interpretation:

- z_i: reward for visiting node i
- expected reward (uniform distribution is equilibrium):

$$\lim_{t \to \infty} \mathbf{E}\, z_{X(t)} = (1/n)\mathbf{1}^T z$$

- so problem is to choose rewards to maximize expected reward, subject to limit Y_{ij} on average reward between connected vertices dual of (maximum expected reward) LP:

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

with variable P

interpretation:

- Y_{ij}: cost of transitioning over edge (i,j)
- expected transition cost is $\lim_{t\to\infty} \mathbf{E}\, Y_{X(t+1)X(t)} = (1/n)\, \mathbf{Tr}\, PY$
- problem is to choose P to minimize expected transition cost

define MTC(Y) as optimal value; MTC is **concave** function of Y

Dual FMC in terms of minimum transition cost

can express dual FMC as

$$\begin{array}{ll} \text{maximize} & \mathsf{MTC}(Y) \\ \text{subject to} & Y\mathbf{1} = 0, \quad Y = Y^T \\ & \|Y\|_* \leq 1 \end{array}$$

- Max-min problem: choose matrix Y to maximize MTC, which is the minimum expected transition cost over all Markov chains on graph
- interpretation of $P^\star$: $P^\star$ minimizes expected transition cost for edge costs $Y^\star$

Extension: fastest mixing to nonuniform distribution

- we are given desired equilibrium distribution $\pi = (\pi_1, \dots, \pi_n)$

- we consider P with same sparsity pattern as graph, but not symmetric

- we do require **reversible** chain: $P_{ij}\pi_j = P_{ji}\pi_i$

- same as designing weights for the edges (including self-loops)

$$w_{ij} = w_{ji} = \pi_j P_{ij} = \pi_i P_{ji}$$

- random walk on weighted graph: assign transition probability as

$$P_{ij} = \frac{w_{ij}}{\sum_{(k,j)\in\mathcal{E}} w_{kj}}$$

- with $\Pi = \mathbf{diag}(\pi)$, 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

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

- asymptotic rate of convergence of distribution to π determined by

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

which is convex in P

- FMC as SDP:

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

Summary

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

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