Convex Optimization and Modeling

Convex Optimization

Fourth lecture, 05.05.2010

Jun.-Prof. Matthias Hein

Reminder from last time

Convex functions:

- first-order condition: $f(y) \geq f(x) + \langle \nabla f|_x, y - x \rangle$,
- second-order condition: Hessian Hf positive semi-definite,
- convex functions are continuous on the relative interior,
- a function f is convex $\Longleftrightarrow$ the epigraph of f is a convex set.

Extensions:

- quasiconvex functions have convex sublevel sets,
- log-concave/convex f: $\log f$ is concave/convex.

Optimization:

- general definition and terminology
- convex optimization
- quasiconvex optimization
- linear optimization (linear programming (LP))
- quadratic optimization (quadratic programming (QP))
- geometric programming
- generalized inequality constraints
- semi-definite and cone programming

Definition 1. *A **general optimization problem** has the form*

$$\min_{x \in D} f(x),$$

$$\textit{subject to: } g_i(x) \leq 0, \; i = 1, \ldots, r$$

$$h_j(x) = 0, \; j = 1, \ldots, s.$$

- x is the optimization variable, f the objective (cost) function,
- $x \in D$ is **feasible** if the inequality and equality constraints hold at x.
- the **optimal value** p^* of the optimization problem

$$p^* = \inf\{ f(x) \,|\, x \text{ feasible } \}.$$

$p^* = -\infty$: problem is unbounded from below,
$p^* = \infty$: problem is infeasible.

Terminology:

- A point x is called **locally optimal** if there exists $R > 0$ such that

$$f(x) = \inf\{f(z) \mid \|z - x\| \leq R, \ z \text{ feasible } \}.$$

- x is feasible,
  $g_i(x) = 0$: inequality constraint is **active** at x.
  $g_i(x) < 0$: is **inactive**.
  A constraint is **redundant** if deleting it does not change the feasible set.
- If $f \equiv 0$ then the optimal value is either zero (feasible set is nonempty) or ∞ (feasible set is empty). This problem is the **feasibility problem**.

$$\begin{aligned} &\text{find } x \\ \text{subject to: } & g_i(x) \leq 0, \ i = 1, \dots, r \\ & h_j(x) = 0, \ j = 1, \dots, s \end{aligned}$$

Equivalent problems: Two problems are called **equivalent** if one can obtain from the solution of one problem the solution of the other problem and vice versa.

Transformations which lead to equivalent problems:

- **Slack variables:** $g_i(x) \le 0 \iff \exists\, s_i \ge 0$ such that $g_i(x) + s_i = 0$.

$$\min_{x \in \mathbb{R}^n,\ s \in \mathbb{R}^r} f(x),$$

$$\text{subject to: } g_i(x) + s_i = 0,\ i = 1, \ldots, r$$

$$s_i \ge 0,\ i = 1, \ldots, r$$

$$h_j(x) = 0,\ j = 1, \ldots, s,$$

which has variables $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^r$.

Transformations which lead to equivalent problems II:

- **Epigraph problem form of the standard optimization problem:**

$$\begin{aligned} \min_{x \in \mathbb{R}^n,\, t \in \mathbb{R}} \quad & t, \\ \text{subject to:} \quad & f(x) - t \leq 0, \\ & g_i(x) \leq 0, \; i = 1, \ldots, r \\ & h_j(x) = 0, \; j = 1, \ldots, s, \end{aligned}$$

which has variables $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$.

Definition 2. *A* ***convex optimization problem*** *has the standard form*

$$\min_{x \in D} f(x),$$

$$\textit{subject to: } g_i(x) \leq 0, \; i = 1, \ldots, r$$

$$\langle a_j, x \rangle = b_j, \; j = 1, \ldots, s,$$

where f, $g_1, \ldots, g_r$ *are convex functions.*

Difference to the general problem:

- the objective function must be convex,
- the inequality constraint functions must be convex,
- the equality constraint functions must be linear.

$\Longrightarrow$ **The feasible set of a convex optimization problem is convex.**

Local and global minima

Theorem 1. *Any locally optimal point of a convex optimization problem is globally optimal.*

Proof. Suppose x is locally optimal, that means x is feasible and $\exists\, R > 0$,

$$f(x) = \inf\{f(z) \,|\, \|z - x\| \leq R,\ z \text{ feasible }\}.$$

Assume x is not globally optimal $\Longrightarrow \exists$ feasible y such that $f(y) < f(x)$.

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y) < f(x),$$

for any $0 < \lambda < 1 \implies x$ is not locally optimal ↯. □

Locally optimal points of **quasiconvex problems** are not generally globally optimal.

First-Order Condition for Optimality:

Theorem 2. *Suppose f is convex and continuously differentiable, Then x is optimal if and only if x is feasible and*

$$\langle \nabla f|_x, y - x \rangle \geq 0, \quad \forall\, y \in X.$$

Proof: Suppose $x \in X$ and $\langle \nabla f|_x, y - x \rangle \geq 0, \quad \forall\, y \in X \Longrightarrow f(y) \geq f(x)$ for all $y \in X$ (first order condition).
Suppose that x is optimal but there is $y \in X$ such that

$$\langle \nabla f|_x, y - x \rangle < 0.$$

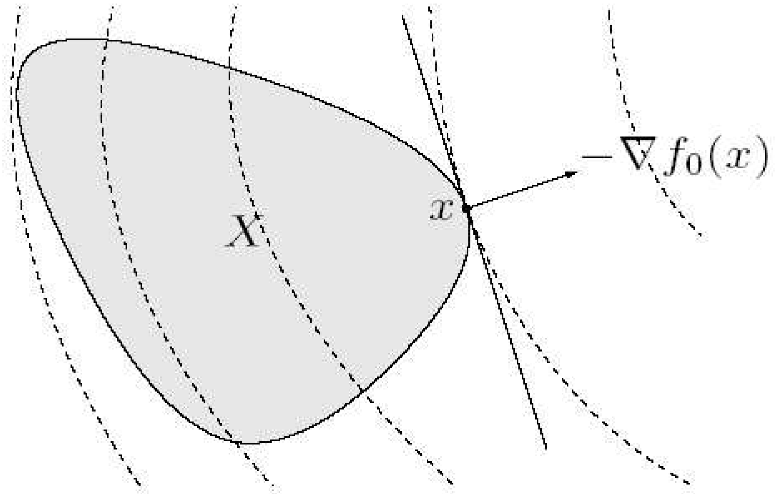
Let $z = ty + (1-t)x$ with $t \in [0,1]$. Then $z(t)$ is feasible for all $t \in [0,1]$ and,

$$\frac{\partial f}{\partial t}\Big|_{t=0} = \langle \nabla f|_x, y - x \rangle < 0,$$

so that for $t \ll 1$ we have $f(y) < f(x)$ ↯.

First-Order Condition for Optimality

Geometric Interpretation:



- x^* on the boundary of the feasible set: $\nabla f|_{x^*}$ defines a supporting hyperplane at x^*.
- x^* in the interior of the feasible set, $\nabla f|_x = 0$,
- Problem only with equality constraint $Ax = b$, then x is optimal if

$$\langle \nabla f|_x, v\rangle = 0, \quad \forall\, v \in \ker(A), \iff \exists\, \nu \in \mathbb{R}^s \text{ such that } \nabla f|_x + A^T\nu = 0.$$

Equivalent convex problems

$$\min_{x \in D} f(x),$$

$$\text{subject to: } g_i(x) \leq 0, \; i = 1, \ldots, r$$

$$\langle a_j, x \rangle = b_j, \; j = 1, \ldots, s,$$

- **Elimination of equality constraints:**
  Let $F \in \mathbb{R}^{n \times k}$ and $x_0 \in \mathbb{R}^n$ such that

$$Ax = b \quad \Longleftrightarrow x = Fz + x_0, \quad z \in \mathbb{R}^k,$$

$$\min f(Fz + x_0),$$

$$\text{subject to: } g_i(Fz + x_0) \leq 0, \; i = 1, \ldots, r$$

This problem has only $n - \dim(\mathrm{ran}(A))$ or $\dim(\ker A)$ variables.

Transformations which preserve convexity

- **Introduction of slack variables,**
- **Introduction of new linear equality constraints,**
- **Epigraph problem formulation,**
- **Minimization over some variables.**

Definition 3. *A* ***quasiconvex optimization problem*** *has the standard form*

$$\min_{x \in D} f(x),$$

$$\textit{subject to: } g_i(x) \leq 0, \; i = 1, \ldots, r$$

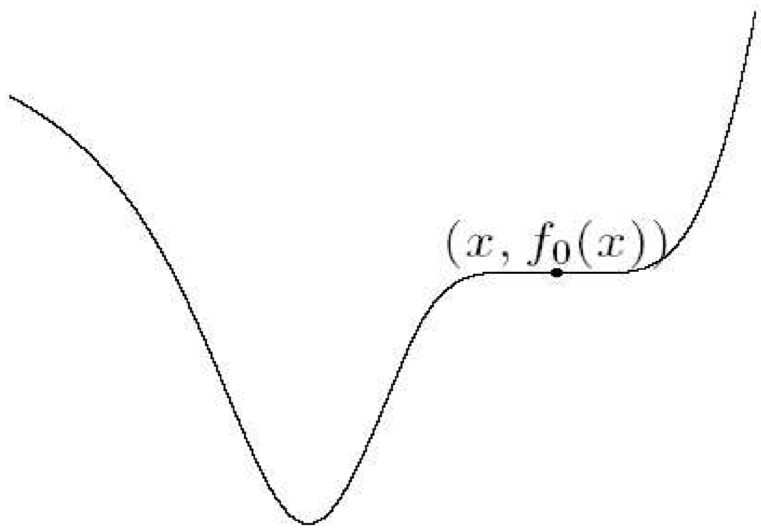
$$\langle a_j, x \rangle = b_j, \; j = 1, \ldots, s,$$

where f is quasiconvex and $g_1, \ldots, g_r$ are convex functions.

Quasiconvex inequality functions can be reduced to convex inequality functions with the same 0-sublevel set.

Theorem 3. *Let X denote the feasible set of a quasiconvex optimization problem with a differentiable objective function f. Then $x \in X$ is optimal if*

$$\langle \nabla f|_x, y - x \rangle > 0, \quad \forall\, y \in X,\, y \neq x.$$



A quasi-convex function with $\nabla f|_{x_0} = 0$ but x_0 is not optimal.

How to solve a quasiconvex optimization problem ?

Representation of the sublevel sets of a quasiconvex functions via sublevel sets of convex functions.

For $t \in \mathbb{R}$ let $\phi_t : \mathbb{R}^n \to \mathbb{R}$ be a family of convex functions such that

$$f(x) \leq t \quad \Longleftrightarrow \quad \phi_t(x) \leq 0,$$

and for each x in the domain

$$\phi_s(x) \geq \phi_t(x) \quad \text{for all} \quad s \leq t.$$

Solve the convex feasibility problem:

$$\begin{aligned} & \text{find } x \\ \text{subject to: } & \phi_t(x) \le 0 \\ & g_i(x) \le 0, \; i = 1, \ldots, r \\ & Ax = b, \end{aligned}$$

Two cases:

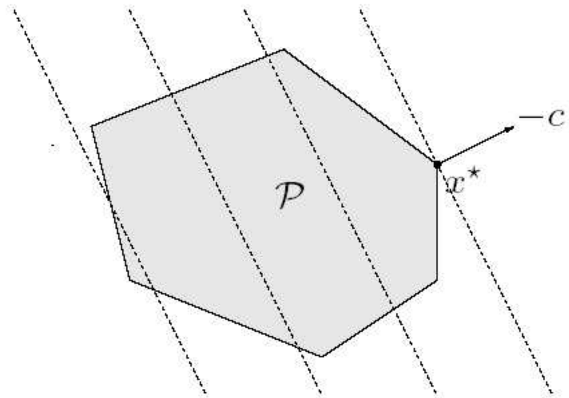
- a feasible point exists $\quad \Longrightarrow \quad$ optimal value $p^* \le t$
- problem is infeasible $\quad \Longrightarrow \quad$ optimal value $p^* \ge t$

Solution procedure:

- assume $p^* \in [a, b]$ and use bisection $t = \frac{b+a}{2}$,
- after k-th iteration interval has length $\frac{b-a}{2^k}$,
- $k = \log_2 \frac{b-a}{\epsilon}$ iterations in order to find an ϵ-approximation of p^*.

Definition 4. *A general* ***linear optimization problem (linear program (LP))*** *has the form*

$$\begin{aligned} \min \ & \langle c, x \rangle \\ \textit{subject to: } Gx & \preceq h, \\ Ax & = b, \end{aligned}$$



where $c \in \mathbb{R}^n$, $G \in \mathbb{R}^{r\times n}$ *with* $h \in \mathbb{R}^r$ *and* $A \in \mathbb{R}^{s\times n}$ *with* $b \in \mathbb{R}^s$.

A linear program is a convex optimization problem with

- affine cost function and linear inequality constraints
- The feasible set is a polyhedron.

The standard form of an LP

$$\min_{x \in \mathbb{R}^n} \; \langle c, x \rangle$$
$$\text{subject to: } Ax = b,$$
$$x \succeq 0.$$

Conversion of a general linear program into the standard form:

- introduce slack variables,
- decompose $x = x^+ - x^-$ with $x^+ \succeq 0$ and $x^- \succeq 0$.

$$\min_{x \in \mathbb{R}^n,\, s \in \mathbb{R}^r} \; \langle c, x \rangle$$
$$\text{subject to: } Gx + s = h,$$
$$Ax = b,$$
$$s \succeq 0.$$

$$\min_{x^+ \in \mathbb{R}^n,\, x^- \in \mathbb{R}^n,\, s \in \mathbb{R}^r} \; \big\langle c, x^+ \big\rangle - \big\langle c, x^- \big\rangle$$
$$\text{subject to: } Gx^+ - Gx^- + s = h,$$
$$Ax^+ - Ax^- = b,$$
$$s \succeq 0,\; x^+ \succeq 0,\; x^- \succeq 0.$$

The diet problem:

- A healthy diet has m different nutrients in quantities at least equal to $b_1, \ldots, b_m$,
- n different kind of food and $x_1, \ldots, x_n$ is the amount of them and has costs $c_1, \ldots, c_n$,
- The food j contains an amount of a_{ij} of nutrient i.
- **Goal:** find the cheapest diet that satisfies the nutritional requirements

$$\begin{aligned} &\min \ \langle c, x\rangle \\ \text{subject to: } &Ax \succeq b, \\ &x \succeq 0. \end{aligned}$$

Chebychev center of a polyhedron:

- find the largest Euclidean ball and its center which fits into a polyhedron

$$P = \{x \in \mathbb{R}^n \mid \langle a_i, x \rangle \leq b_i, \; i = 1, \ldots, r\}.$$

- constraint that the ball $B = \{x_c + u \mid \|u\| \leq R\}$ lies in one half-space

$$\forall u \in \mathbb{R}^n, \|u\| \leq R \quad \Longrightarrow \quad \langle a_i, x_c + u \rangle \leq b_i.$$

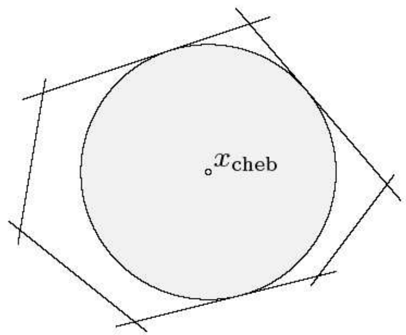
With $\sup\{\langle a_i, u \rangle \mid \|u\| \leq R\} = R \, \|a_i\|_2$ the constraint can be rewritten as

$$\langle a_i, x_c \rangle + R \, \|a_i\|_2 \leq b_i.$$

Thus the problem can be reformulated as

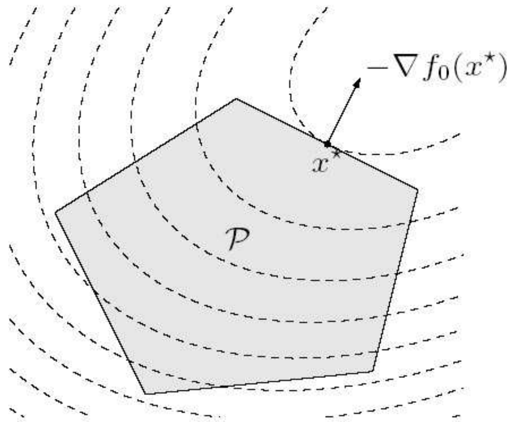
$$\max_{x \in \mathbb{R}^n, \, R \in \mathbb{R}} \quad R$$

subject to: $\langle a_i, x_c \rangle + R \, \|a_i\|_2 \leq b_i, \quad i = 1, \ldots, r.$



Definition 5. *A general* ***quadratic program (QP)*** *has the form*

$$\min \frac{1}{2}\langle x, Px\rangle + \langle q, x\rangle + c$$
$$\textit{subject to: } Gx \preceq h,$$
$$Ax = b,$$



where $P \in S_+^n$, $G \in \mathbb{R}^{r\times n}$ *and* $A \in \mathbb{R}^{s\times n}$.

With quadratic inequality constraints:

$$\frac{1}{2}\langle x, P_i x\rangle + \langle q_i, x\rangle + c_i \leq 0, \qquad \textit{with } P_i \in S_+^n,\ i = 1, \dots, r$$

we have a ***quadratically constrained quadratic program (QCQP)***.

$$\textbf{LP} \quad \subset \quad \textbf{QP} \quad \subset \quad \textbf{QCQP}.$$

Examples for a QP

- **Least Squares:** Minimizing $\|Ax - b\|_2^2 = \langle x, A^T Ax \rangle - 2\langle b, Ax \rangle + \langle b, b \rangle$ is an unconstrained QP. Analytical solution: $x = A^{\dagger} b$.
- **Linear Program with random cost:**

$$\begin{aligned} \min \ & \langle c, x \rangle \\ \text{subject to: } & Gx \preceq h, \\ & Ax = b, \end{aligned}$$

  - c is **random** with: $\bar{c} = \mathbb{E}[c]$, and covariance $\Sigma = \mathbb{E}[(c - \bar{c})(c - \bar{c})^T]$.
  - the cost $\langle c, x \rangle$ is **random** with mean $\mathbb{E}[\langle c, x \rangle] = \langle \bar{c}, x \rangle$ and variance

$$\mathrm{Var}[\langle c, x \rangle] = \langle x, \Sigma x \rangle .$$

Risk-sensitive cost: $\mathbb{E}[\langle c, x \rangle] + \gamma \, \mathrm{Var}[\langle c, x \rangle] = \langle \bar{c}, x \rangle + \gamma \langle x, \Sigma x \rangle$,
We get the following QP:

$$\begin{aligned} \min \ & \langle \bar{c}, x \rangle + \gamma \langle x, \Sigma x \rangle \\ \text{subject to: } & Gx \preceq h, \\ & Ax = b. \end{aligned}$$

Definition 6. *A* ***second-order cone problem (SOCP)*** *has the form*

$$
\begin{aligned}
&\min \ \langle f, x \rangle \\
&\textit{subject to:} \ \ \|A_i x + b_i\| \leq \langle c_i, x \rangle + d_i, \ \ i = 1, \ldots, r \\
&F x = g,
\end{aligned}
$$

where $A_i \in \mathbb{R}^{n_i \times n}, b \in \mathbb{R}^{n_i}$ *and* $F \in \mathbb{R}^{p \times n}$.

$$\|Ax + b\|_2 \leq \langle c, x \rangle + d,$$

with $A \in \mathbb{R}^{k \times n}$ is a **second-order cone constraint**. The function

$$\mathbb{R}^n \to \mathbb{R}^{k+1}, \quad x \mapsto (Ax + b, \langle c, x \rangle + d)$$

is required lie in the second order cone in $\mathbb{R}^{k+1}$.

$c_i = 0, \ 1 \leq i \leq r$: reduces to a QCQP, $\quad A_i = 0, \ 1 \leq i \leq r$: reduces to a LP.

$$\textbf{QCQP} \quad \subset \quad \textbf{SOCP.}$$

Example for SOCP

Robust linear programming: robust wrt to uncertainty in parameters,

- Consider the linear program

$$\min \ \langle c, x \rangle$$
$$\text{subject to: } \ \langle a_i, x \rangle \leq b_i,$$

- $a_i \in E_i = \{\bar{a}_i + P_i u \,|\, \|u\|_2 \leq 1\}$ where $P_i \in S^n_+$,

$$\min \ \langle c, x \rangle$$
$$\text{subject to: } \ \langle a_i, x \rangle \leq b_i, \quad \forall\, a_i \in E_i$$

- $\sup\{\langle a_i, x \rangle \,|\, a_i \in E_i\} = \langle \bar{a}_i, x \rangle + \|P_i x\|_2$. Thus,

$$\langle \bar{a}_i, x \rangle + \|P_i x\|_2 \leq b_i \qquad \text{(second-order constraint)}\ .$$

$$\min \ \langle c, x \rangle$$
$$\text{subject to: } \ \langle \bar{a}_i, x \rangle + \|P_i x\|_2 \leq b_i.$$

Example for SOCP II

Linear Programming with random constraints: $a_i \sim N(\bar{a}_i, \Sigma_i)$.

The linear program with random constraints

$$\begin{aligned} \min \ & \langle c, x \rangle \\ \text{subject to: } & \mathrm{P}(\langle a_i, x \rangle \leq b_i) \geq \eta, \quad i = 1, \ldots, r, \end{aligned}$$

can be expressed as SOCP

$$\begin{aligned} \min \ & \langle c, x \rangle \\ \text{subject to: } & \langle \bar{a}_i, x \rangle + \Phi^{-1}(\eta) \left\| \Sigma_i^{\frac{1}{2}} x \right\|_2 \leq b_i, \quad i = 1, \ldots, r, \end{aligned}$$

where $\phi(z) = \mathrm{P}(X \leq z)$ with $X \sim N(0, 1)$.

Definition 7. *A* ***convex optimization problem with generalized inequality constraints*** *has the standard form*

$$\min_{x \in D} f(x),$$
$$\textit{subject to: } g_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, r$$
$$Ax = b,$$

where f is convex, $K_i \subseteq \mathbb{R}^{k_i}$ are proper cones, $g_i : \mathbb{R}^n \to \mathbb{R}^{k_i}$ are K_i-convex.

Properties:

- The feasible set and the optimal set are convex,
- Any locally optimal point is also globally optimal,
- The optimality condition for differentiable f holds without change.

Definition 8. *A* ***conic-form problem or conic program*** *has the form*

$$\min \ \langle c, x \rangle$$
$$\textit{subject to: } Fx + g \preceq_K 0,$$
$$Ax = b,$$

where $F \in \mathbb{R}^{r \times n}$ *with* $g \in \mathbb{R}^r$ *and* K *is a proper cone in* $\mathbb{R}^r$*.*

K = positive-orthant $\Rightarrow$ the conic program reduces to a linear program.

Definition 9. *A* ***semi-definite program (SDP)*** *has the form*

$$\begin{aligned} &\min \ \langle c, x \rangle \\ \textit{subject to: } &\sum_{i=1}^{n} x_i F_i + G \preceq_{S_+^k} 0, \\ &Ax = b, \end{aligned}$$

where $G, F_1, \ldots, F_r \in S^k$ *and* $A \in \mathbb{R}^{s \times n}$.

The **standard form of an SDP** (similar to the LP standard form):

$$\begin{aligned} &\min_{X \in S^n} \ \mathrm{tr}(CX) \\ \text{subject to: } &\mathrm{tr}(A_i X) = b_i, \quad i = 1, \ldots, s \\ &X \succeq 0, \end{aligned}$$

where $C, A_1, \ldots, A_s \in S^n$ and $A \in \mathbb{R}^{s \times n}$.

Fastest mixing Markov chain on an undirected graph:

- Let $G = (V, E)$ where $|V| = n$ and $E \subset \{1, \ldots, n\} \times \{1, \ldots, n\}$,
- Markov chain on the graph with states $X(t)$ with transition probabilities

$$P_{ij} = \mathrm{P}\Big(X(t+1) = i \,|\, X(t) = j\Big),$$

  from vertex j to vertex i (note that (i, j) has to be in E).
  The matrix P should satisfy $P_{ij} = 0$ for all $(i, j) \notin E$ and

$$P_{ij} \geq 0, \; i, j = 1, \ldots, n, \quad \mathbf{1}^T P = \mathbf{1}^T, \quad P = P^T.$$

- Since P is symmetric and $\mathbf{1}^T P = \mathbf{1}^T$ we have $P\mathbf{1} = \mathbf{1}$.
  Uniform distribution $p_i = \frac{1}{n}$ is an equilibrium of the Markov chain.
  Convergence rate is determined by $r = \max\{\lambda_2, -\lambda_n\}$, where

$$1 = \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n,$$

  are the eigenvalues of P.

Fastest mixing Markov chain on an undirected graph: We have

$$r = \|QPQ\|_{2,2} = \left\|(\mathbb{1} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)P(\mathbb{1} - \frac{1}{n}\mathbf{1}\mathbf{1}^T)\right\|_{2,2} = \left\|P - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right\|_{2,2},$$

where $Q = \mathbb{1} - \frac{1}{n}\mathbf{1}\mathbf{1}^T$ is the projection matrix on the subspace orthogonal to $\mathbf{1}$. Thus the mixing rate r is a convex function of P.

$$\begin{aligned} \min_{P \in S^n} \quad & \left\|P - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right\|_{2,2} \\ \text{subject to:} \quad & P\mathbf{1} = \mathbf{1}, \\ & P_{ij} \geq 0, \quad i,j = 1,\ldots,n, \\ & P_{ij} = 0, \quad (i,j) \notin E \end{aligned}$$

$$\begin{aligned} \min_{t \in \mathbb{R}, P \in S^n} \quad & t \\ \text{subject to:} \quad & -t\mathbb{1} \preceq P - \frac{1}{n}\mathbf{1}\mathbf{1}^T \preceq t\mathbb{1} \\ & P\mathbf{1} = \mathbf{1}, \\ & P_{ij} \geq 0, \quad i,j = 1,\ldots,n, \\ & P_{ij} = 0, \quad (i,j) \notin E \end{aligned}$$

The right problem is an SDP.