Semidefinite Programming

Chee Wei Tan

Convex Optimization and its Applications to Computer Science

Outline

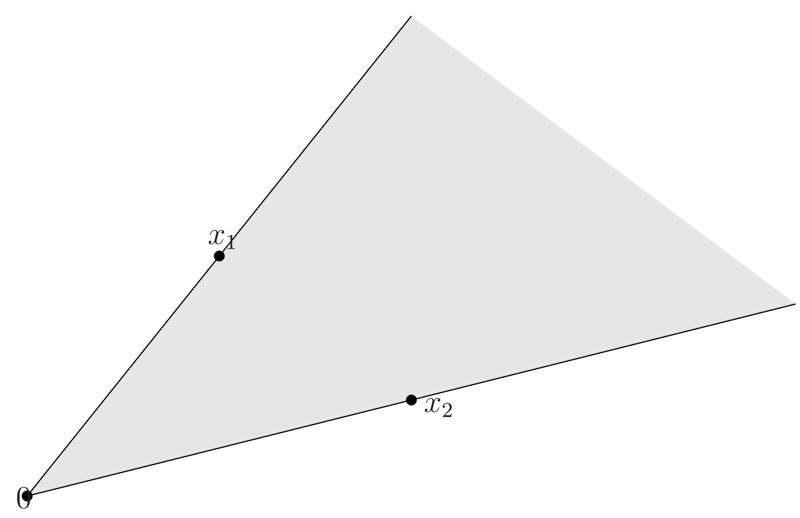
- Cones and dual cones
- Generalized inequality and convexity
- Conic programming
- Semidefinite programming (SDP)
- SDP Applications: Max-Cut Approximation Algorithm, Graph Laplacian optimization

Cones and Convex Cones

C is a cone if for every $x \in C$ and $\theta \geq 0$, we have $\theta x \in C$

C is a convex cone if it is convex and a cone: for any $x_1, x_2 \in C$ and $\theta_1, \theta_2 \geq 0$

$$\theta_1 x_1 + \theta_2 x_2 \in C$$



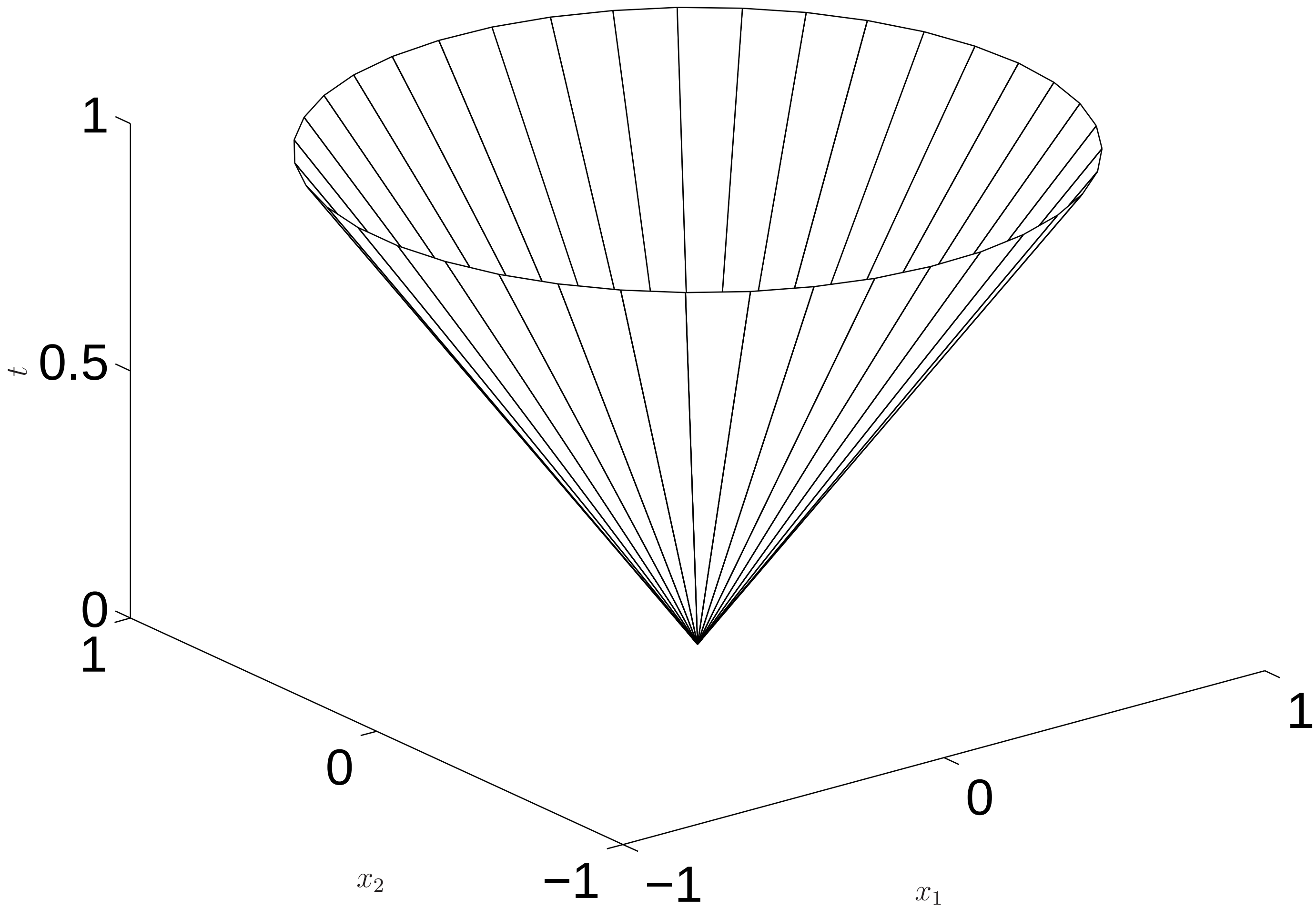
Norm Cones

Given a norm, norm cone is a convex cone:

$$C = \{(x,t) \in \mathbf{R}^{n+1} | \|x\| \leq t\}$$

Example: second order cone:

$$\begin{aligned} C &= \{(x,t) \in \mathbf{R}^{n+1} | \|x\|_2 \leq t\} \\ &= \left\{ \begin{bmatrix} x \\ t \end{bmatrix} \middle| \begin{bmatrix} x \\ t \end{bmatrix}^T \begin{bmatrix} I & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} \leq 0, t \geq 0 \right\} \end{aligned}$$


1
0.5
0
t
1
0
−1
x_2
−1
0
1
x_1

Positive Semidefinite Cone

Matrix $A \in \mathbf{R}^{n \times n}$ is positive semidefinite $A \succeq 0$ if for all $x \in \mathbf{R}^n$,

$$x^T A x \geq 0$$

Matrix $A \in \mathbf{R}^{n \times n}$ is positive definite $A \succ 0$ if for all $x \in \mathbf{R}^n$,

$$x^T A x > 0$$

Set of symmetric positive semidefinite matrices:

$$\mathbf{S}_+^n = \{X \in \mathbf{R}^{n \times n} | X = X^T, X \succeq 0\}$$

$\mathbf{S}_+^n$ is a convex cone: if $\theta_1, \theta_2 \geq 0$ and $A, B \in \mathbf{S}_+^n$, then $\theta_1 A + \theta_2 B \in \mathbf{S}_+^n$, since for all $x \in \mathbf{R}^n$:

$$x^T(\theta_1 A + \theta_2 B)x = \theta_1 x^T A x + \theta_2 x^T B x \geq 0$$

Proper Cones and Generalized Inequalities

A cone K is a proper cone if

- K is convex
- K is closed
- K has nonempty interior
- K has no lines ($x \in K, -x \in K \Rightarrow x = 0$)

Proper cone K induces a generalized inequality (partial ordering on $\mathbf{R}^n$):

$$x \preceq_K y \quad \Leftrightarrow \quad y - x \in K$$

$$x \prec_K y \quad \Leftrightarrow \quad y - x \in \textbf{int } K$$

Examples

- Nonnegative orthant and componentwise inequality:

  $K = \mathbf{R}^n_+$ is a proper cone

  $x \preceq_K y$ means $x_i \le y_i, \ \ i = 1, \ldots, n$

  $x \prec_K y$ means $x_i < y_i, \ \ i = 1, \ldots, n$

- Positive semidefinite cone and matrix inequality:

  $K = \mathbf{S}^n_+$ is a proper cone in the set of symmetric matrices $\mathbf{S}^n$

  $X \preceq_K Y$ means $Y - X$ is positive semidefinite

  $X \prec_K Y$ means $Y - X$ is positive definite.

Properties of Generalized Inequalities

- If $x \preceq_K y$ and $u \preceq_K v$, then $x + u \preceq_K y + v$
- If $x \preceq_K y$ and $y \preceq_K z$, then $x \preceq_K z$
- If $x \preceq_K y$ and $\alpha \geq 0$, then $\alpha x \preceq_K \alpha y$
- If $x \preceq_K y$ and $y \preceq_K x$, then $x = y$
- If $x_i \preceq_K y_i$ for $i = 1, \ldots,$ and $x_i \to x$ and $y_i \to y$ as $i \to \infty$, then $x \preceq_K y$

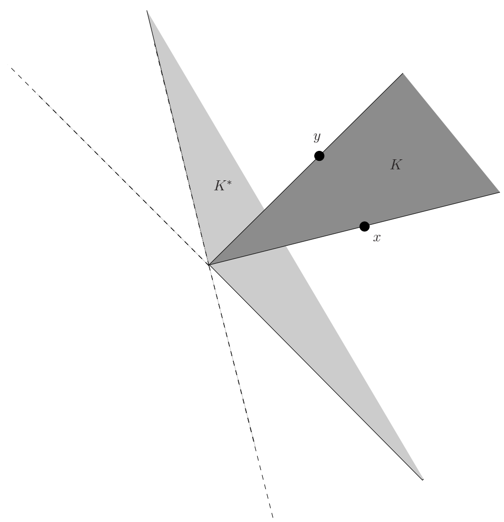
Dual Cones

Given a cone K. Dual cone of K:

$$K^* = \{y | x^T y \geq 0 \; \forall x \in K\}$$

K^* is always a convex cone

K is proper cone $\Rightarrow K^*$ is a proper cone and $K^{**} = K$



Nonnegative orthant cone is self-dual

Dual PSD Cone

Consider inner product $X^T Y = \text{Tr}(XY) = \sum_{i,j} X_{ij} Y_{ij}$ on $\mathbf{S}^n$. PSD cone $\mathbf{S}_+^n$ is self-dual:

$$\text{Tr}(XY) \geq 0 \ \forall X \succeq 0 \Leftrightarrow Y \succeq 0$$

Proof: Forward direction: suppose $Y \not\succeq 0$ and $\text{Tr}(XY) \geq 0, \forall X \succeq 0$. Then $\exists q \in \mathbf{R}^n$ such that $q^T Y q = \text{Tr}(qq^T Y) < 0$. Therefore, $X = qq^T \succeq 0$ satisfies $\text{Tr}(XY) < 0$, thus contradicting the assumption.

Reverse direction: suppose $X, Y \succeq 0$. Express X by eigenvalue decomposition: $X = \sum_{i=1}^n \lambda_i q_i q_i^T$ where $\lambda_i \geq 0$, $i = 1, \ldots, n$. Then

$$\text{Tr}(XY) = \text{Tr}\left(Y \sum_{i=1}^n \lambda_i q_i q_i^T \right) = \sum_{i=1}^n \lambda_i q_i^T Y q_i \geq 0.$$

Dual Generalized Inequality

Given proper cone K and generalized inequality $\succeq_K$. Dual cone K^* is also proper and induces dual generalized inequality $\succeq_{K^*}$

Relationship between $\succeq_K$ and $\succeq_{K^*}$:

1. $x \preceq_K y$ if and only if $\lambda^T x \leq \lambda^T y$ for all $\lambda \succeq_{K^*} 0$
2. $x \prec_K y$ if and only if $\lambda^T x < \lambda^T y$ for all $\lambda \succeq_{K^*} 0, \lambda \neq 0$

Since $K^{**} = K$, these properties hold if $\succeq_K$ and $\succeq_{K^*}$ are interchanged

Generalized Inequality Induced Monotonicity

$f : \mathbf{R}^n \to \mathbf{R}$ is K-nondecreasing (increasing) if

$$x \preceq_K y \Rightarrow f(x) \leq (<) f(y)$$

First order condition for differentiable f with convex domain: f is K-nondecreasing if and only if:

$$\nabla f(x) \succeq_{K^*} 0$$

f is K-increasing if:

$$\nabla f(x) \succ_{K^*} 0$$

Example: For PSD cone, Tr(X^{-1}) is matrix decreasing on $\mathbf{S}^n_{++}$ and $\det X$ is matrix increasing on $\mathbf{S}^n_+$

Generalized Inequality Induced Convexity

$f : \mathbf{R}^n \to \mathbf{R}^m$ is K-convex if for all x, y and $\theta \in [0, 1]$,

$$f(\theta x + (1-\theta)y) \preceq_K \theta f(x) + (1-\theta)f(y)$$

$f : \mathbf{R}^n \to \mathbf{R}^m$ is strictly K-convex if for all $x \neq y$ and $\theta \in (0, 1)$,

$$f(\theta x + (1-\theta)y) \prec_K \theta f(x) + (1-\theta)f(y)$$

First order condition: For differentiable f with convex domain, f is K-convex if and only if for all $x, y \in \mathbf{dom}\, f$,

$$f(y) \succeq_K f(x) + Df(x)(y - x)$$

Matrix Convexity

f is a symmetric-matrix-valued function. $f : \mathbf{R}^{n\times m} \to \mathbf{S}^m$. f is convex with respect to matrix inequality if for any $x, y \in \mathbf{dom}\, f$ and $\theta \in [0, 1]$,

$$f(\theta x + (1-\theta)y) \preceq \theta f(x) + (1-\theta) f(y)$$

Equivalently, f is matrix convex if and only if scalar-valued function $z^T f(x) z$ is convex for all z

- $f(X) = XX^T$ is matrix convex
- $f(X) = X^p$ is matrix convex for $p \in [1, 2]$ or $p \in [-1, 0]$, and matrix concave for $p \in [0, 1]$
- $f(X) = e^X$ is not matrix convex (unless X is a scalar)

Generalized Inequality Constraints

Convex optimization with generalized inequality constraints on vector-valued functions:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, \quad i = 1, 2, \ldots, m \end{array}$$

$f_0 : \mathbf{R}^n \to \mathbf{R}$, $f_i : \mathbf{R}^n \to \mathbf{R}^{k_i}$ are K_i-convex for some proper cones K_i

- Feasible set is convex
- Local optimality $\Rightarrow$ global optimality

Conic Programming

Linear programming with linear generalized inequality constraint:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Fx + G \preceq_K 0 \\ & Ax = b \end{array}$$

- When K is nonnegative orthant, conic program reduces to LP
- When K is PSD cone, write inequality constraints as Linear Matrix Inequalities (LMI):

$$x_1 F_1 + \ldots + x_n F_n + G \preceq 0$$

  where $F_i, G \in \mathbf{S}^k$. When they are diagonal, LMI reduces to linear inequalities

SDP

SDP: Minimize linear objective over linear equalities and LMI on variables $x \in \mathbf{R}^n$

$$
\begin{array}{ll}
\text{minimize} & c^T x \\
\text{subject to} & x_1 F_1 + \ldots + x_n F_n + G \preceq 0 \\
& Ax = b
\end{array}
$$

SDP in standard form: Minimize a matrix inner product over equality constraints on inner products on variables $X \in \mathbf{S}^n$

$$
\begin{array}{ll}
\text{minimize} & \mathrm{Tr}(CX) \\
\text{subject to} & \mathrm{Tr}(A_i X) = b_i, \quad i = 1, 2, \ldots, p \\
& X \succeq 0
\end{array}
$$

LP and SOCP as SDP

LP as SDP:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \mathbf{diag}(Gx - h) \preceq 0 \\ & Ax = b \end{array}$$

SOCP:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \|A_i x + b_i\|_2 \le c_i^T x + d_i, \quad i = 1, \ldots, N \\ & Fx = g \end{array}$$

SOCP as SDP:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \begin{bmatrix} (c_i x + d_i) I & A_i x + b_i \\ (A_i x + b_i)^T & (c_i x + d_i) I \end{bmatrix} \succeq 0, \quad i = 1, \ldots, N \\ & Fx = g \end{array}$$

Matrix Norm Minimization

$A(x) = A_0 + x_1 A_1 + \ldots x_n A_n$ where $A_i \in \mathbf{R}^{p \times q}$. Consider unconstrained spectral norm (max. singular value) minimization over x:

$$\text{minimize } \|A(x)\|_2$$

which is equivalent to convex optimization with LMI on (x, t)

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & A(x)^T A(x) \preceq t^2 I \end{array}$$

which is equivalent to SDP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \left[\begin{array}{cc} tI & A(x) \\ A(x)^T & tI \end{array} \right] \succeq 0 \end{array}$$

Related Problems and SDP Applications

SDP problems:

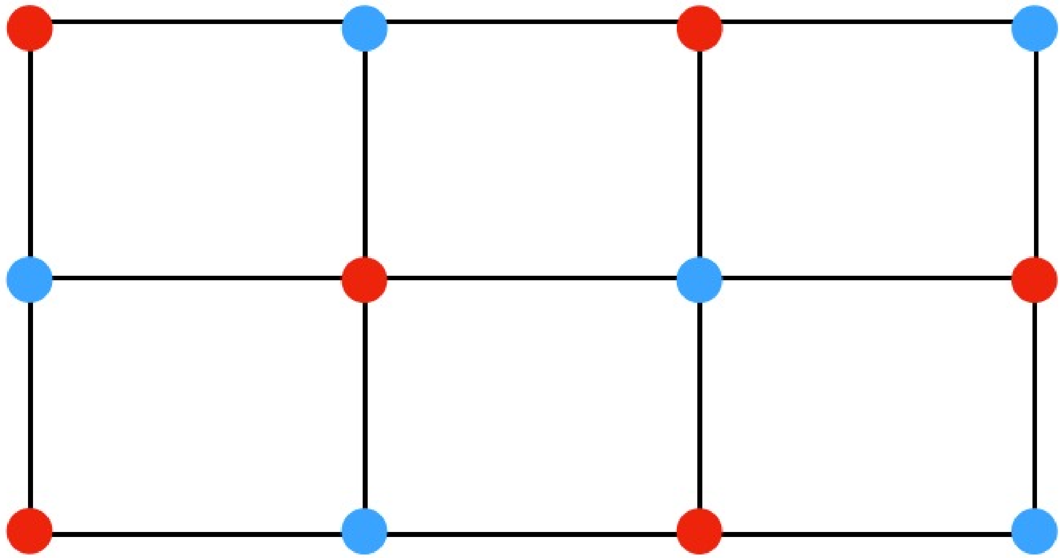
- Minimize largest eigenvalue
- Minimize sum of r largest eigenvalues
- Maximize log determinant

SDP approximations:

- Combinatorial optimization such as Max-Cut in graph theory
- Rank minimization
- Sum-of-squares optimization

Max-Cut Problem: Approximation Algorithms

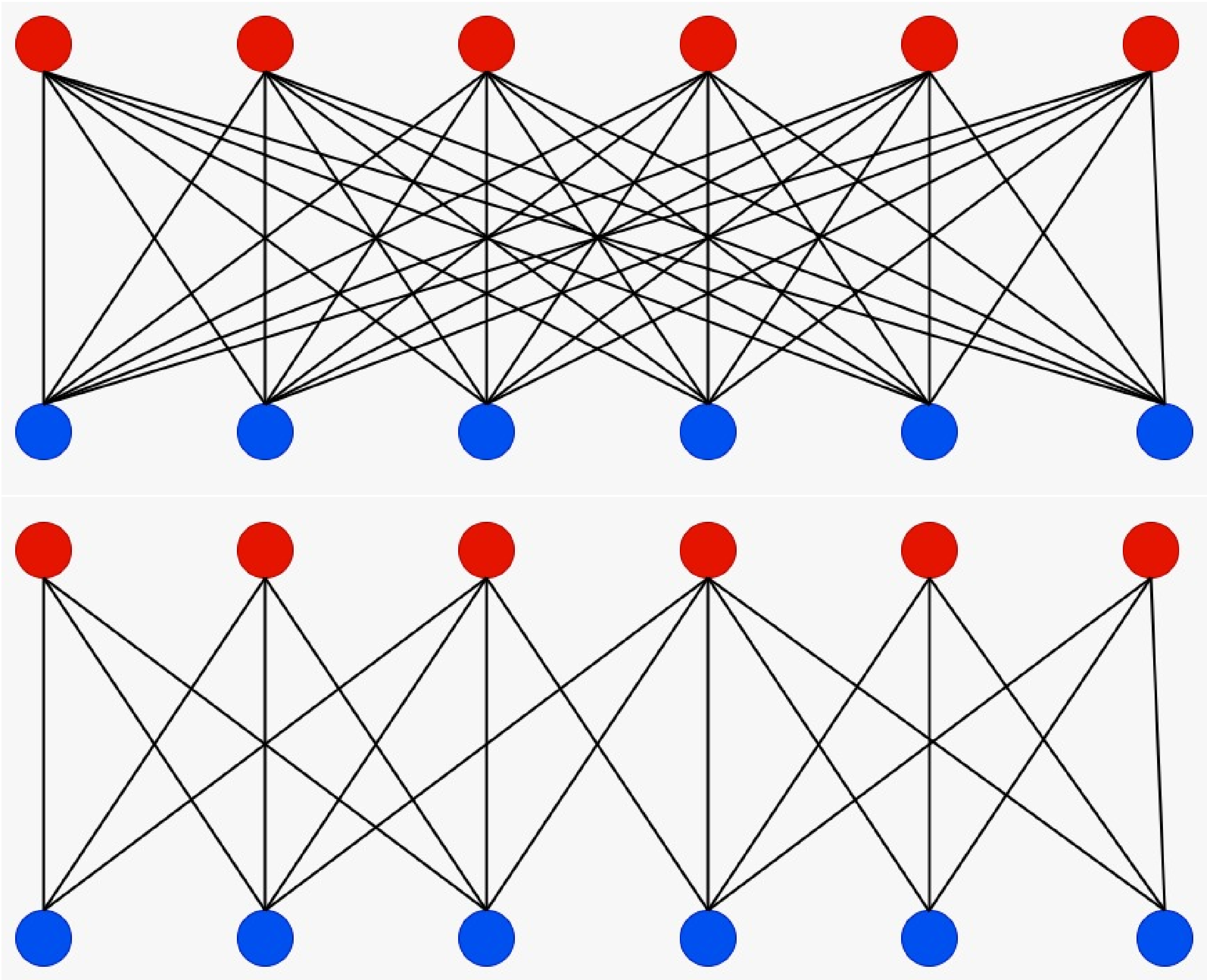
Given an undirected graph with vertex set $V = \{1, ..., n\}$ and edge set $E \subset \{\{i, j\} \mid i, j \in V, i \neq j\}$ with no self-loops



Max-Cut problem: For a subset $S \subset V$, the value of a cut is the number of edges connecting a vertex in S to a vertex not in S, find $S \subset V$ with maximum cut.

In the above grid graph, vertices shaded red are in S. What is the cut value?

Max-Cut Problem: Approximation Algorithms



What is the cut value of a fully bipartite graph and the Laman graph?

Max-Cut Problem Formulation

Consider the adjacency matrix of the graph

$$W_{ij} = \begin{cases} 1, & \text{if } \{i,j\} \in E \\ 0, & \text{otherwise} \end{cases}$$

and denote a cut S by a vector $x \in \mathbb{R}^n$

$$x_i = \begin{cases} +1, & \text{if } i \in S \\ -1, & \text{otherwise} \end{cases}$$

then $1 - x_i x_j = 2$ if $\{i,j\}$ is a cut, so the value of a cut induced by x is

$$\frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij}(1 - x_i x_j)$$

thus maximizing the cut is equivalent to minimizing $\displaystyle\sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} x_i x_j$

Max-Cut Problem Formulation

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 - 1 = 0,\ i = 1, \ldots, n. \end{array}$$

- This is a Nonconvex QCQP that is NP-hard
- Denote the optimal value by p^*, and thus the maximum cut is

$$c_{max} = \frac{1}{4} \sum_{i=1}^{n} \sum_{j=1}^{n} W_{ij} - \frac{1}{4} p^*$$

- Seminal work on SDP convex relaxations to combinatorial optimization: M. X. Goemans and D. P. Williamson, Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, Journal of the ACM, 42(6), pp. 1115–1145, 1995. http://www-math.mit.edu/~goemans/PAPERS/maxcut-jacm.pdf

Lagrange Dual Relaxation

Assume the existence of Lagrange dual solution. The Lagrangian is

$$L(x, \lambda) = x^T W x - \sum_{i=1}^{n} \lambda_i (x_i^2 - 1) = x^T (W - \Lambda) x + \mathsf{Tr}(\Lambda)$$

where $\Lambda = \mathbf{diag}(\lambda_1, ..., \lambda_n)$ and is bounded below if $W - \Lambda \succeq 0$ giving

$$\boxed{\begin{array}{ll} \text{maximize} & \mathsf{Tr}(\Lambda) \\ \text{subject to} & W \succeq \Lambda \end{array}}$$

which lower bounds p^*, since for any feasible x (including the optimal solution),

$$x^T W x \geq x^T \Lambda x = \sum_{i=1}^{n} \Lambda_{ii} x_i^2 = \mathsf{Tr}(\Lambda)$$

as the first inequality is due to $W \succeq \Lambda$ and the last equality from $x_i \in \{+1, -1\}$

SDP Relaxation

Consider the Max-Cut problem:

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1,\ i = 1, \ldots, n. \end{array}$$

Let $X := xx^T$. Then $x^T W x = \mathrm{Tr}(Wxx^T) = \mathrm{Tr}(WX)$ Therefore, $X \succeq 0$, has rank one and $X_{ii} = x_i^2 = 1$. Conversely, any rank-one matrix X with $X \succeq 0, X_{ii} = 1$ implies $X = xx^T$ for some ± 1 vector x. Thus, a reformulation:

$$\begin{array}{ll} \text{minimize} & \mathrm{Tr}(WX) \\ \text{subject to} & X_{ii} = 1,\ i = 1, \ldots, n, \\ & X \succeq 0,\ \textbf{rank}(X) = 1. \end{array}$$

Dropping the nonconvex rank constraint leads to a convex relaxation. If the solution X of this relaxed problem has rank one, the original problem is solved. Otherwise, we project this solution back to the original problem domain *while still retaining some form of performance guarantees*: Approximation Algorithm.

Duality of SDP Relaxation

We now have two SDP problems that give lower bounds to p^*

Lagrange Dual Relaxation	SDP Relaxation
$$\begin{array}{ll} \text{maximize} & \mathrm{Tr}(\Lambda) \\ \text{subject to} & W \succeq \Lambda \\ & \Lambda \text{ diagonal} \end{array}$$	$$\begin{array}{ll} \text{minimize} & \mathrm{Tr}(WX) \\ \text{subject to} & X \succeq 0 \\ & X_{ii} = 1,\ i = 1, \ldots, n \end{array}$$

- How are they related? Which one is closer to p^*?
- Lagrange Dual relaxations give certified bounds
- SDP relaxations provide hints of possible feasible primal solution
- Both problems can be solved by primal-dual SDP solvers (e.g., SeDuMi, SDPT3, Mosek, available in CVX software)

Projection via Randomized Rounding

Suppose the optimal solution X is rank r $(r > 1)$ after solving SDP relaxation:

$$\begin{array}{rl} \text{minimize} & \mathrm{Tr}(WX) \\ \text{subject to} & X \succeq 0 \\ & X_{ii} = 1,\ i = 1, \ldots, n. \end{array}$$

Consider a randomized algorithm to project X to rank-one space

- Factorize $X = V^T V$, where $V = [v_1 \ldots v_n] \in \mathbb{R}^{r \times n}$ (X is the Gram matrix)
- Since $X_{ij} = v_i^T v_j$ and $X_{ii} = 1$, the Gram matrix factorization gives n vectors on the unit sphere in $\mathbb{R}^r$, i.e., $v_i^T v_i = 1$ for all i
- Instead of assigning either +1 or -1 to each vertex of the graph, a point on the unit sphere in $\mathbb{R}^r$ is assigned to each vertex

Projection via Randomized Rounding

Pick a random unit vector $q \in \mathbb{R}^r$, and choose a cut

$$S = \{i \mid v_i^T q \geq 0\}$$

then the probability that $\{i, j\}$ is a cut edge (i belongs to S and j not) is

$$\text{Prob}(v_i, v_j \,\text{separated}) = \frac{\text{angle between } v_i \text{ and } v_j}{\pi} = \frac{1}{\pi}\text{arccos}\, v_i^T v_j$$
$$= \frac{1}{\pi}\text{arccos}\, X_{ij}$$

But, observe the following:

$$\frac{\text{Prob}(v_i, v_j \,\text{separated})}{(1 - v_i^T v_j)/2} \geq \frac{\theta/\pi}{(1 - \cos\theta)/2} = \frac{2}{\pi}\frac{\theta}{1 - \cos\theta}$$
$$\geq \frac{2}{\pi} \min_{\theta \in [0,\pi]} \frac{\theta}{1 - \cos\theta}$$
$$> 0.878$$

Projection via Randomized Rounding

Summing up over all vertices, we have an expected cut value of

$$\frac{1}{2}\sum_{i,j\in E} W_{ij}\text{Prob}(v_i, v_j \text{ separated}) = \frac{1}{2}\sum_{i,j\in E} W_{ij}\frac{\arccos v_i^T v_j}{\pi} = \frac{1}{4}\sum_{i,j\in E} W_{ij}\frac{2\arccos X_{ij}}{\pi}$$

$$= \frac{1}{4}\sum_{i,j\in E} W_{ij}(1 - X_{ij})\frac{2\arccos X_{ij}}{\pi(1 - X_{ij})}$$

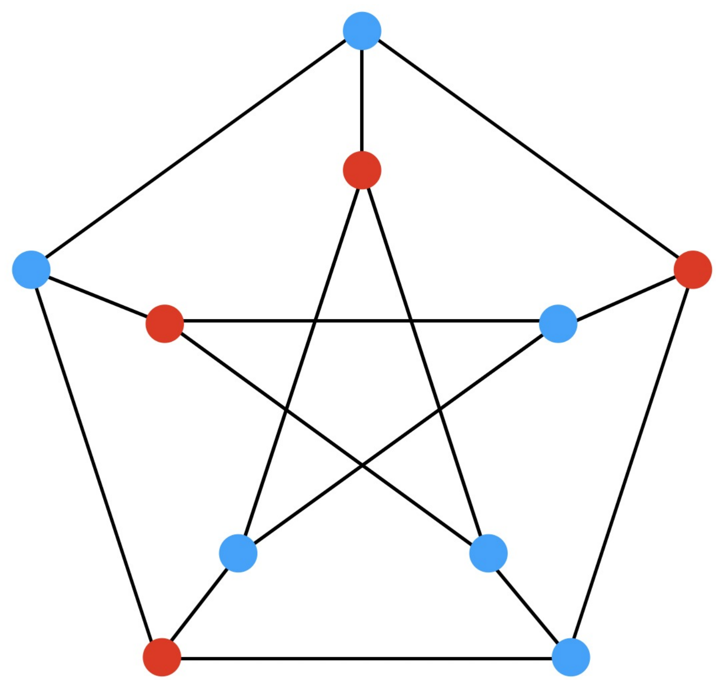
$$> \frac{1}{4}\sum_{i,j\in E} W_{ij}(1 - X_{ij})\; 0.878$$

$$> 0.878\; p^*$$

Thus, we have established that SDP Relaxation with Randomized Rounding Projection is a 0.878-approximation algorithm for Max-Cut

Max-Cut Problem: Approximation Algorithms

An easy heuristic for the Max-Cut problem is a 1/2-approximation algorithm that simply labels each vertex of the graph as $+1$ or -1 uniformly at random. Compare the performance of these two approximation algorithms for the Petersen Graph.



In the above Petersen Graph, vertices shaded red are in S. What is the cut value? How good does the Goemans-Williamson SDP relaxation and randomized rounding projection perform? Can you find a cut achieving 75% of optimality?

MAP Inference in Markov Random Networks

Consider a pairwise Markov Random Field (e.g., Ising model) which is defined for a graph $G = (V, E)$ with n vertices. Associate a binary variable $x_i \in \{-1, +1\}$ with each vertex $i \in V$. Let $\theta_i : \{\pm 1\} \to \mathbb{R}$ and $\theta_{ij} : \{\pm 1\}^2 \to \mathbb{R}$ be defined for each vertex and edge of the graph, respectively, as the vertex and pairwise potential. Thus, $a\ posterior$ distribution of x follows the Gibbs distribution:

$$p(x|\theta) = \frac{e^{U(x|\theta)}}{Z(\theta)},$$

with $U(x;\theta) = \sum_{i \in V} \theta_i(x_i) + \sum_{(i,j) \in E} \theta_{ij}(x_i, x_j)$ and $Z(\theta)$ the normalization.

The Maximum $a\ Posterior$ (MAP) estimate is given by

$$\hat{x} = \operatorname*{arg\,max}_{x \in \{-1,+1\}^n} p(x|\theta) = \operatorname*{arg\,max}_{x \in \{-1,+1\}^n} U(x;\theta)$$

which, for an indefinite matrix W, can be rewritten as the Max-Cut problem:

$$\hat{x} = \operatorname*{arg\,min}_{x \in \{-1,+1\}^n} x^T W x.$$

Graph Laplacian

- graph $G = (V, E)$ with $n = |V|$ nodes, $m = |E|$ edges
- edge weights $w_1, ..., w_m \in R$
- $l \sim (i, j)$ means edge l connects nodes i, j
- incidence matrix: $A_{il} = \begin{cases} 1, & \text{edge } l \text{ enters node } i \\ -1, & \text{edge } l \text{ enters node } i \\ 0, & \text{otherwise} \end{cases}$
- (weighted) Laplacian: $L = A\mathrm{diag}(w)A^T$
- $L_{ij} = \begin{cases} -w_l, & l \sim (i, j) \\ \sum_{l \sim (i,k)} w_l, & i = j \\ 0, & \text{otherwise} \end{cases}$

Laplacian Eigenvalues

- L is symmetric; $L\mathbf{1} = 0$
- we'll be interested in case when $L \succeq 0$ (i.e., L is PSD)(always the case when weights nonnegative)
- Laplacian eigenvalues (eigenvalues of L):

$$0 = \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$$

- spectral graph theory connects properties of graph, and λ_i (with $w = 1$) e.g.: G connected if and only if $\lambda_2 > 0$ (with $w = 1$)

Convex Spectral Functions

- suppose ϕ is a symmetric convex function in $n-1$ variables
- then $\psi(w) = \phi(\lambda_2, ..., \lambda_n)$ is a convex function of weight vector w
- examples:

$$-\phi(u) = 1^T u (\text{i.e., the sum}) :$$

$$\psi(w) = \sum_{i=2}^{n} \lambda_i = \sum_{i=1}^{n} \lambda_i = \mathbf{Tr}\ L = 2\mathbf{1}^T w \ (\text{twice the total weight})$$

$$-\phi(u) = \max_i u_i :$$

$$\psi(w) = \max\{\lambda_2, ..., \lambda_n\} = \lambda_n \quad (\text{spectral radius})$$

Random Walk on Graph

- Markov chain on nodes of G, with transition probabilities on edges

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

- we'll focus on symmetric transition probability matrices P (everything extends to reversible case, with fixed equilibrium distr.)

- identifying P_{ij} with w_l for $l \sim (i,j)$, we have $P = I - L$

- same as linear averaging matrix W, but here $W_{ij} \geq 0$ (i.e., $w \geq 0, \mathbf{diag}(L) \leq 1$)

Mixing Rate

- probability distribution $\pi_i(t) = \textbf{Prob}(X(t) = i)$ satisfies $\pi(t+1)^T = \pi(t)^T P$
- since $P = P^T$ and $P\mathbf{1} = \mathbf{1}$, uniform distribution $\pi = (1/n)\mathbf{1}$ is stationary, i.e., $((1/n)\mathbf{1})^T P = ((1/n)\mathbf{1})^T$
- $\pi(t) \to (1/n)\mathbf{1}$ for any $\pi(0)$ if and only if

$$\mu = \|P - (1/n)\mathbf{1}\mathbf{1}^T\| = \|I - L - (1/n)\mathbf{1}\mathbf{1}^T\| < 1$$

  μ is called second largest eigenvalue modulus (SLEM) of MC

- SLEM determines convergence (mixing) rate, e.g.,

$$\sup_{\pi(0)} \|\pi(t) - (1/n)\mathbf{1}\|_{tv} \leq (\sqrt{n}/2)\mu^t$$

- associated mixing time is $\tau = 1/\log(1/\mu)$

Fastest Mixing Markov Chain Problem

$$\begin{array}{ll} \text{minimize} & \mu = \|I - L - (1/n)\mathbf{1}\mathbf{1}^T\| = \max\{1-\lambda_2, \lambda_n - 1\} \\ \text{subject to} & w \geq 0, \textbf{diag}(L) \leq 1 \end{array}$$

- optimization variable is w; problem data is graph G
- same as fast linear averaging problem, with additional nonnegativity constraint $W_{ij} \geq 0$ on weights
- convex optimization problem (indeed, SDP), hence efficiently solved
- `https://web.stanford.edu/~boyd/papers/pdf/icm06_talk.pdf`

Markov Process on a Graph

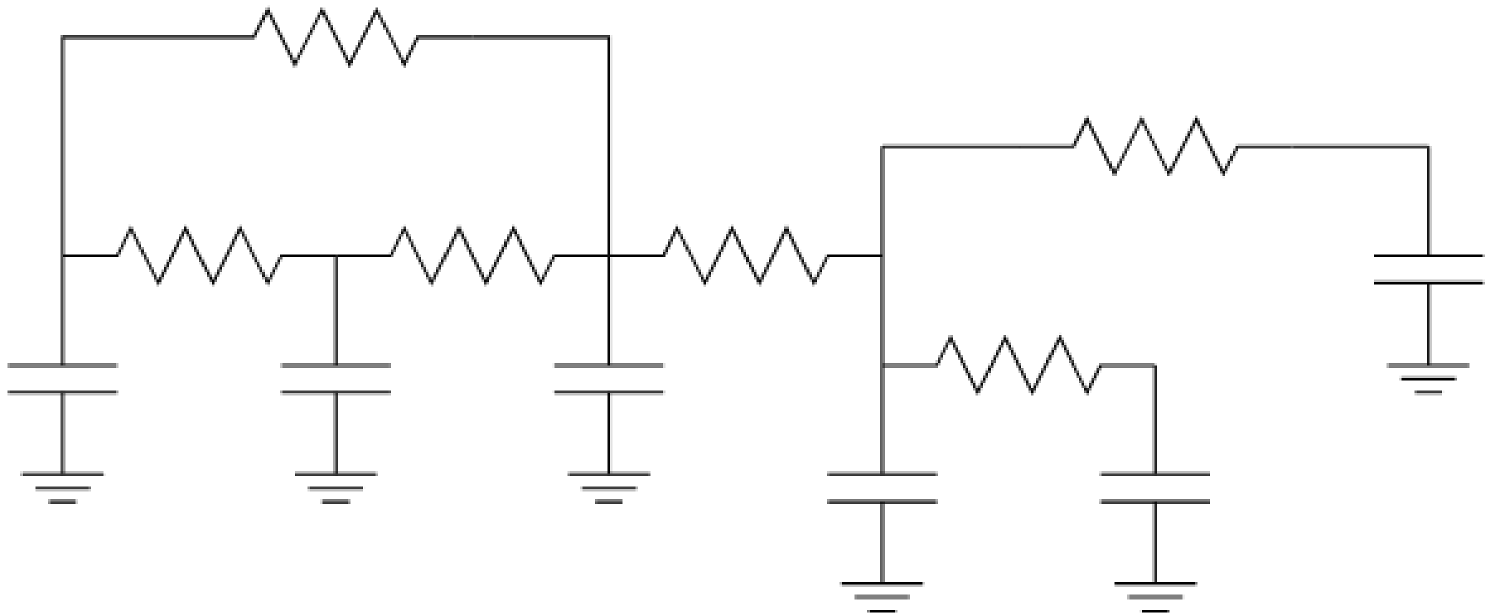
- (continuous-time) Markov process on nodes of G, with transition rate $w_l \geq 0$ between nodes i and j, for $l \sim (i,j)$
- probability distribution $\pi(t) \in R^n$ satisfies heat equation $\dot{\pi}(t) = -L\pi(t)$
- $\pi(t) = e^{-tL}\pi(0)$
- $\pi(t)$ converges to uniform distribution $(1/n)\mathbf{1}$, for any $\pi(0)$, if and only if $\lambda_2 > 0$
- (asymptotic) convergence as $e^{-\lambda_2 t}$; λ_2 gives mixing rate of process
- λ_2 is concave, homogeneous function of w (come from symmetric concave function $\phi(u) = \min_i u_i$)

Fastest Mixing Process on a Graph

$$\begin{array}{ll} \text{maximize} & \lambda_2 \\ \text{subject to} & \sum_l d_l^2 w_l \leq 1, \quad w \geq 0 \end{array}$$

- variable is $w \in R^m$; data is graph, normalization constants $d_l > 0$
- a convex optimization problem, hence easily solved
- allocate rate across edges so as maximize mixing rate
- constraint is always tight at solution, i.e., $\sum_l d_l^2 w_l = 1$
- when $d_l^2 = 1/m$, optimal value is called absolute algebraic connectivity

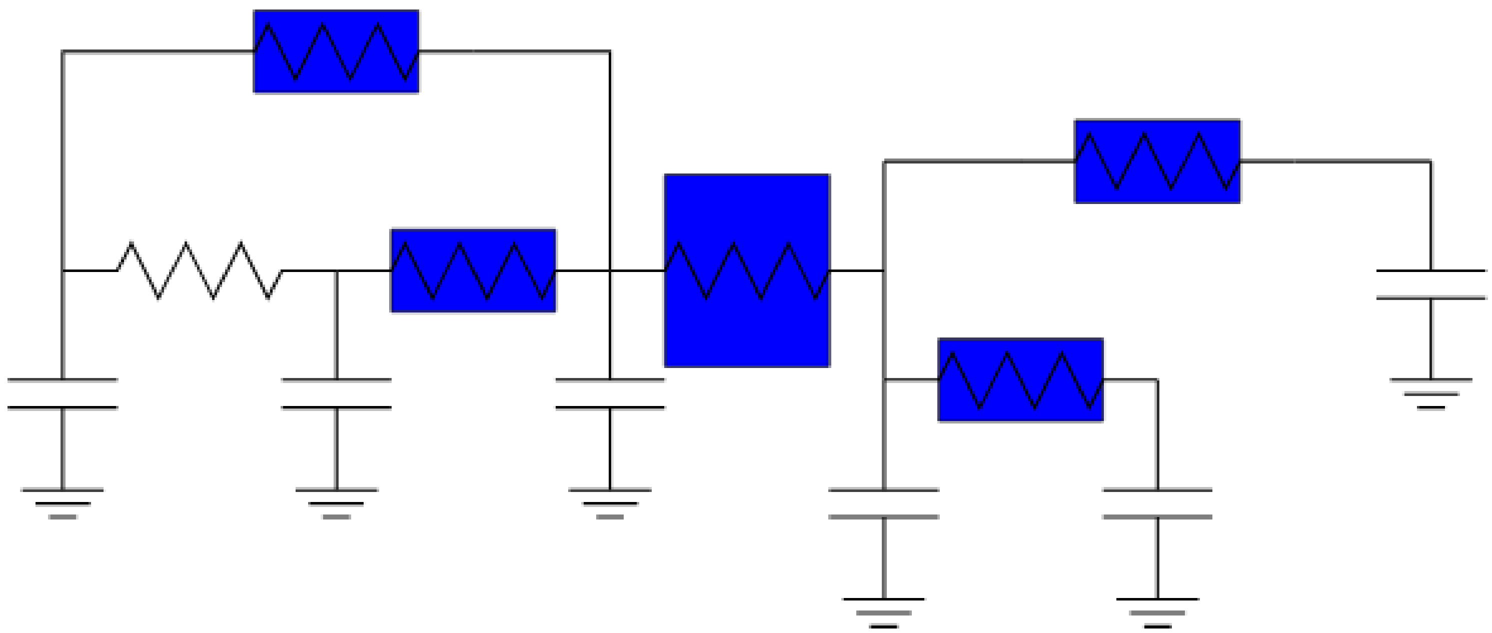
Grounded Unit Capacitor RC Circuit View



- charge vector $q(t)$ satisfies $\dot{q}(t) = -Lq(t)$, with edge weights given by conductances, $w_l = g_l$
- charge equilibrates (i.e., converges to uniform) at rate determined by λ_2
- with conductor resistivity ρ, length d_l, and cross-sectional area a_l, we have $g_l = a_l/(\rho d_l)$

Graph Laplacian

- total conductor volume is $\sum_l d_l a_l = \rho \sum_l d_l^2 w_l$
- problem is to choose conductor cross-sectional areas, subject to a total volume constraint, so as to make the circuit equilibrate charge as fast as possible



optimal λ_2 is .105; uniform allocation of conductance gives $\lambda_2 = .073$

SDP Formulation and Dual

alternate formulation:

$$\begin{array}{ll} \text{minimize} & \sum d_l^2 w_l \\ \text{subject to} & \lambda_2 \geq 1, \quad w \geq 0 \end{array}$$

SDP formulation:

$$\begin{array}{ll} \text{minimize} & \sum d_l^2 w_l \\ \text{subject to} & L \succeq I - (1/n)\mathbf{1}\mathbf{1}^T, \quad w \geq 0 \end{array}$$

dual problem:

$$\begin{array}{ll} \text{maximize} & \mathbf{Tr}(X) \\ \text{subject to} & X_{ii} + X_{jj} - X_{ij} - X_{ji} \leq d_l^2, \; l \sim (i,j) \\ & \mathbf{1}^T X \mathbf{1} = 0, X \succeq 0 \end{array}$$

with variable $X \in R^{n \times n}$

Maximum Variance Unfolding Problem

- use variables $x_1, ..., x_n \in R^n$, with $X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix} \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$

- dual problem becomes **maximum variance unfolding** (MVU) problem

$$\begin{array}{ll} \text{maximize} & \sum_i \|x_i\|^2 \\ \text{subject to} & \|x_i - x_j\| \leq d_l, \quad l \sim (i,j) \\ & \sum_i x_i = 0 \end{array}$$

- position n points in R^n to maximize variance, while respecting local distance constraints

Summary

- Proper cones induce generalized inequalities in $\mathbf{R}^n$ and $\mathbf{S}^n$, which induces generalized convex inequality constraints
- Optimization with generalized inequalities: conic programming
- SDP is conic programming over PSD cone with LMI, includes LP, QP, QCQP, SOCP as special cases
- SDP Applications: Max-Cut, Graph Laplacian optimization

Reading assignment: Sections 2.1, 2.6, 3.6, 4.6, 5.9 of textbook.

- L. Vandenberghe and S. Boyd, "Semidefinite programming," SIAM Review, March 1996.