Convex Optimization

Kontonis Vasilis

24 February, 2017

Corelab, NTUA

Contents

Generalized Inequalities & SDP

Generalized Inequalities

Cone & Convex Cone

- K is a cone if for every $x \in K$ and $\theta \geqslant 0$, $\theta x \in K$.
- K is a **convex** cone if for every $x_1, x_2 \in K$ and $\theta_1, \theta_2 \geqslant 0$, $\theta_1 x_1 + \theta_2 x_2 \in K$.

Proper Cone

A cone $K \subseteq \mathbb{R}^n$ is a proper cone if:

- K is convex.
- K is closed.
- $\text{int}K \neq \varnothing$
- K is pointed $\Leftrightarrow x \in K, \ -x \in K \implies x = 0$

Proper Cones can be used to define partial orderings on $\mathbb{R}^n$

$$x \leqslant_K y \Leftrightarrow y - x \in K$$

Examples: $\mathbb{R}_+, \ \mathbb{R}^n_+, S^n_+$

Generalized Monotonicity & Convexity

$f : U \to \mathbb{R}$ is K-nondecreasing if $x \preccurlyeq_K y \implies f(x) \leqslant f(y)$
Examples:

- $\mathrm{tr}(WX),\ W \in S^n$ is matrix nondecreasing if $W \geqslant 0$, matrix decreasing if $W \leqslant 0$.
- $\mathrm{tr}(X^{-1})$ is matrix decreasing on S^n_{++}.
- $\det(X)$ is matrix increasing on S^n_{++}.

$f : U \to F$ is K-convex if $f(\theta x + (1-\theta)y) \preccurlyeq_K \theta f(x) + (1-\theta) f(y)$.
If $f : U \to S^m$ then we can deduce that f is **matrix**-convex using the equivalent condition that the real valued function $z^T f(x) z$ is convex.

- $f(X) = XX^T$ is matrix convex.
- $f(X) = X^2$ is matrix convex.
- $f(X) = e^X$ is not matrix convex.

Generalized Constrained Problem

- $f_0 : \mathbb{R}^n \to \mathbb{R}$.
- $K_i \subseteq \mathbb{R}^{k_i}$ are proper cones.
- $f_i : \mathbb{R}^n \to \mathbb{R}^{k_i}$ are K_i-convex.

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

- Feasible, Sublevel, Optimal Sets are convex.
- Locally optimal point is globally optimal.
- If f_0 is differentiable, the usual optimality condition holds.
- Often solved as easily as ordinary convex optimization problems.

Cone Programs

Cone programs are generalized linear programs.

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Fx + g \preceq_K 0 \\ & Ax = b \end{aligned}$$

Constraint function is affine thus K-convex.
Standard form conic problem:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & x \preceq_K 0 \\ & Ax = b \end{aligned}$$

SOCP

SOCP is a Cone Program.

$$
\begin{aligned}
\text{minimize} \quad & c^T x \\
\text{subject to} \quad & -(A_i x + b_i, c_i^T x + d_i) \preccurlyeq_{K_i} 0, \; i = 1, \ldots, m \\
& Fx = g
\end{aligned}
$$

- $K_i = \{(y, t) \in \mathbb{R}^{n_i+1} \mid \|y\|_2 \leqslant t\}$ is a second-order cone in $\mathbb{R}^{n_i+1}$.

Semidefinite Programming

K is the cone of semidefinite $k \times k$ matrices, $K = S_+^k$.

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

- The Constraint is a Linear Matrix Inequality (LMI).
- Is SDP a generalization of LP?

Multiple LMI Constraints

A SDP can have more than one LMI constraints

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & F^i(x) = x_1 F_1^i + \ldots + x_n F_n^i + G^i \preccurlyeq 0, \ i = 1, \ldots, m \\ & Ax = b. \end{aligned}$$

We can use the fact that a block diagonal matrix is positive semi-definite iff all its blocks are positive semi-definite to form a large block diagonal LMI constraint

$$\text{diag}(F^1(x), \ldots, F^m(x)) \preccurlyeq 0$$

LMIs

The (strict) LMI

$$F(x) \coloneqq F_0 + \sum x_i F_i > 0$$

is equivalent to a set of n polynomial inequalities since $u^T F(x) u > 0$ for all $u \in \mathbb{R}^n$.

- The solution set of an LMI is convex. Consider the affine map $F_0 + \sum x_i F_i$.
- A set of convex **non-linear** inequalities can be represented as an LMI. Let $Q(x) = Q(x)^T,\ R(x) = R(x)^T$ and $S(x)$ depend affinely on x then

$$\begin{pmatrix} Q(x) & S(x) \\ S(x)^T & R(x) \end{pmatrix} > 0 \Leftrightarrow \begin{array}{r} R(x) > 0 \\ Q(x) - S(x)R(x)^{-1}S(x)^T > 0 \end{array}$$

Matrix norm Minimization

Let $A(x) = A_0 + x_1 A_1 + \ldots + x_n A_n$.

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

$\|\cdot\|_2$ is the spectral norm.
Equivalent SDP

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & \begin{pmatrix} tI & A(x) \\ A^T(x) & tI \end{pmatrix} \succcurlyeq 0 \end{aligned}$$

- Is SDP a generalization of SOCP?
- Should we solve SOCPs with SDP solvers?

Fastest mixing Markov Chain

In probability theory, the mixing time of a Markov chain is the time until the Markov chain is "close" to its steady state distribution.

- $G(V, E)$ is an undirected graph.
- $X(t)$ is the state of the MC.
- Each edge has a probability $P_{ij} = \Pr[X(t+1) = i \mid X(t) = j)$, $P_{ij} = 0$, if $(i, j) \notin E$.
- $P_{ij} \geqslant 0, \ 1^T P = 1^T, \ P = P^T$.
- $(1/n)\mathbf{1}$ is an equilibrium distribution of the MC.
- Eigenvalues of P: $1 = \lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$
- Convergence is deteremined by the mixing rate $r = \max\{\lambda_2, -\lambda_n\}$

Fastest mixing Markov Chain

We want to reach as fast as possible the uniform distribution, thus we minimize the mixing time r.

$$
\begin{array}{ll}
\text{minimize} & r \\
\text{subject to} & P_{ij} \geqslant 0 \\
& \mathbf{1}^T P = \mathbf{1}^T
\end{array}
$$

The equivalent **SDP**

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

GW MaxCut

GW MaxCut

Approximation & SDP
SDP can be solved in polynomial time, up to accuracy ϵ.

MaxCut Problem

- Undirected graph $G = (V, E)$.
- $z_i \in \{-1, 1\}$ corresponds to i-th vertex.
- A cut $(S,\ V \setminus S)$, where $S = \{i \in V\ :\ z_i = 1\}$.

$$
\begin{aligned}
&\text{maximize} \quad && \sum_{(i,j)\in E} \frac{1 - z_i z_j}{2} \\
&\text{subject to} \quad && z_i \in \{-1, 1\},\ i = 1, \ldots, n
\end{aligned}
$$

SDP Relaxation

We replace the real variables z_i with vectors $u_i \in S^{n-1}$.

$$\begin{aligned} \text{maximize} \quad & \sum_{(i,j)\in E} \frac{1 - \mathbf{u}_i^T \mathbf{u}_j}{2} \\ \text{subject to} \quad & \mathbf{u}_i \in S^{n-1},\ i = 1, \dots, n \end{aligned}$$

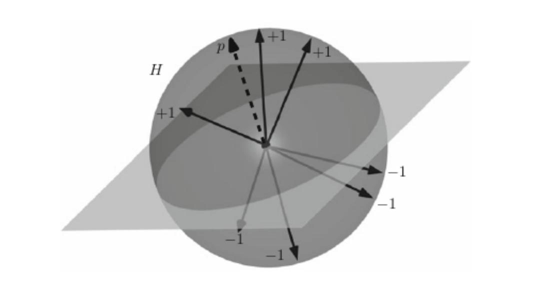
Equivalent Problem:

$$\begin{aligned} \text{maximize} \quad & \sum_{(i,j)\in E} \frac{1 - x_{ij}}{2} \\ \text{subject to} \quad & x_{ii} = 1,\ i = 1, 2, \dots, n \\ & X \succcurlyeq 0 \end{aligned}$$

Rounding the Vector Solution

Chose randomly $\mathbf{p} \in S^{n-1}$ and consider the mapping

$$\mathbf{u} \mapsto \begin{cases} 1, & \text{if } \mathbf{p}^\top \mathbf{u} \geqslant 0, \\ -1, & \text{otherwise.} \end{cases}$$



The probability that this rounding maps $\mathbf{u}$ and $\mathbf{u}'$ to different values is

$$\frac{\arccos \mathbf{u}^\top \mathbf{u}'}{\pi}$$

Getting the Bound

The Expected Number of edges in the resulting cut equals

$$\sum_{(i,j)\in E} \frac{\arccos({u_i^*}^T u_j^*)}{\pi}$$

We know that

$$\sum_{(i,j)\in E} \frac{1 - {u_i^*}^T u_j^*}{2} \geqslant \mathrm{Opt}(G) - \epsilon$$

It holds that

$$\frac{\arccos(z)}{\pi} \geqslant 0.87856 \frac{1-z}{2}$$

Questions?

S. Boyd, P. Diaconis, and L. Xiao.
Fastest Mixing Markov Chain on a Graph.
SIAM Review, 46, Dec. 2004.

S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan.
Linear Matrix Inequalities in System and Control Theory.
Society for Industrial and Applied Mathematics, Philadelphia, June 1997.

S. Boyd and L. Vandenberghe.
Convex Optimization.
Cambridge University Press, Cambridge, UK ; New York, Mar. 2004.

M. X. Goemans and D. P. Williamson.
Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming.
J. ACM, 42(6):1115–1145, Nov. 1995.