Convex Optimization

Kontonis Vasilis

March 10, 2017

Corelab, NTUA

Contents

Convex Problems

Optimization Problem

General Optimization Problem:

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leqslant 0,\ i = 1, \ldots, m \\ & h_i(x) = 0,\ i = 1, \ldots, p \end{aligned}$$

- Domain of the problem:

$$\mathcal{D} = \bigcap_{i=0}^{m} \text{dom} f_i \cap \bigcap_{i=1}^{p} \text{dom} h_i$$

- When a point x is feasible?
- $p^* = \inf\{f_0(x) \mid x \text{ is feasible}\}$
- A feasible point x with $f_0(x) \leqslant p^* + \epsilon$ is called ϵ-**suboptimal**.
- A feasible point x is locally optimal if it is optimal in a norm ball of radius $R > 0$.

Convex Optimization Problem (CP)

We minimize a **convex objective** over a **convex set**.

Generic Convex Problem:

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i \leqslant 0, \ i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

- $f_0, \ f_i$ must be convex.
- The **equality constraints** must be affine.
- The feasible set of this problem is convex. Why?
- If f_0 is quasiconvex the problem is called quasiconvex.
- For convex and quasiconvex problems the optimal set, and the ϵ-suboptimal sets are convex.
- Locally optimal points are globally optimal.

Optimality Criteria

A point x is optimal iff $x \in X$ and

$$\nabla f_0(x)^T(y - x) \geqslant 0, \text{ for all } y \in X.$$

- Unconstraint problems $\nabla f_0(x) = 0$.
  Example: $f_0(x) = (1/2)x^T P x + q^T x + r,\ \nabla f_0(x) = x^T P + q^T$
- Equality Constraints

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & Ax = b \end{aligned}$$

$\nabla f_0(x) \perp \mathcal{N}(A) \implies \nabla f_0(x) \in \mathcal{R}(A^T)$. Therefore we obtain the Langrange Multiplier optimality condition:

$$\nabla f_0(x) + A^T \nu = 0$$

Quasiconvex Optimization

A quasiconvex problem can have locally optimal solutions that are **not** globally optimal.

Optimality Condition: A point x is optimal if

$$\nabla f_0(x)^T(y-x) > 0, \text{ for all } y \in X \setminus \{x\}.$$

Quasiconvexity of f_0 implies that there exist a family of convex functions $\phi_t : \mathbb{R}^n \to \mathbb{R},\ t \in \mathbb{R}$ such that $f_0(x) \leqslant t \Leftrightarrow \varphi_t(x) \leqslant 0$, moreover it holds $s \geqslant t \implies \varphi_s(x) \leqslant \varphi_t(x)$.

Consider the feasibility problem

$$\begin{aligned} \text{find} \quad & x \\ \text{subject to} \quad & \varphi_t(x) \leqslant 0 \\ & f_i(x) \leqslant 0,\ i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

Can we use this feasibility problem to derive an approximation algorithm for the Quasiconvex Optimization problem?

Quasiconvex Optimization

A quasiconvex problem can have locally optimal solutions that are **not** globally optimal.

Optimality Condition: A point x is optimal if

$$\nabla f_0(x)^T(y-x) > 0, \text{ for all } y \in X \setminus \{x\}.$$

Quasiconvexity of f_0 implies that there exist a family of convex functions $\phi_t : \mathbb{R}^n \to \mathbb{R},\ t \in \mathbb{R}$ such that $f_0(x) \leqslant t \Leftrightarrow \varphi_t(x) \leqslant 0$, moreover it holds $s \geqslant t \implies \varphi_s(x) \leqslant \varphi_t(x)$.

Consider the feasibility problem

$$\begin{aligned} \text{find} \quad & x \\ \text{subject to} \quad & \varphi_t(x) \leqslant 0 \\ & f_i(x) \leqslant 0,\ i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

Bisection

Linear Programming

Chebyshev Inequalities

Let X be a RV with values $\{u_1, \ldots, u_n\}$, $p_i = \Pr(X = u_i)$.
Assume that the distribution of X, namely the p_i, is unknown.
Assume that we have upper and lower bounds on expected values of some function of X and probabilities of some subsets of $\mathbb{R}$.

$$\mathbb{E}[f(X)] = \sum_{i=1}^{n} p_i f(u_i), \quad \Pr[x \in S] = \sum_{u_i \in S} p_i$$

LP:

$$\begin{array}{ll} \text{minimize} & a_0^T p \\ \text{subject to} & p \geqslant 0,\ 1^T p = 1, \\ & \alpha_i \leqslant a_i^T p \leqslant \beta_i,\ i = 1, \ldots, m \end{array}$$

Linear Fractional Programming

Suppose we want to minimize a ratio of linear functions
$f_0(x) = \frac{c^T x + d}{e^T x + f}$, $\mathrm{dom} f_0 = \{x \mid e^T x + f > 0\}$.

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & Gx \leqslant h \\ & Ax = b \end{aligned}$$

Equivalent **LP**:

$$\begin{aligned} \text{minimize (y,z)} \quad & c^T y + dz \\ \text{subject to} \quad & Gy - hz \leqslant 0 \\ & Ay - bz = 0 \\ & e^T y + fz = 1 \\ & z \geqslant 0 \end{aligned}$$

To show equivalence consider the pair

$$y = \frac{x}{e^T x + f}, \quad z = \frac{1}{e^T x + f}$$

Quadratic Programming

QP & QCQP

Basic **QP** Problem

$$\begin{aligned} \text{minimize} \quad & (1/2)x^T P x + q^T x + r \\ \text{subject to} \quad & Gx \leqslant h \\ & Ax = b \end{aligned}$$

If we allow **quadratic** inequality constraints we have a **QCQP** problem

$$\begin{aligned} \text{minimize} \quad & (1/2)x^T P x + q^T x + r \\ \text{subject to} \quad & (1/2)x^T P_i x + q_i^T x + r_i \leqslant 0, \; i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

- The feasible region is the intersection of ellipsoids.
- Generalizes QP and LP.

Bounded Least Squares

The well-known least squares problem

$$\text{minimize} \quad \|Ax - b\|_2^2 = x^T(A^TAx) - 2b^TAx + b^Tb$$

In the unconstraint case we can obtain the **normal** equations

$$A^TAx = A^Tb.$$

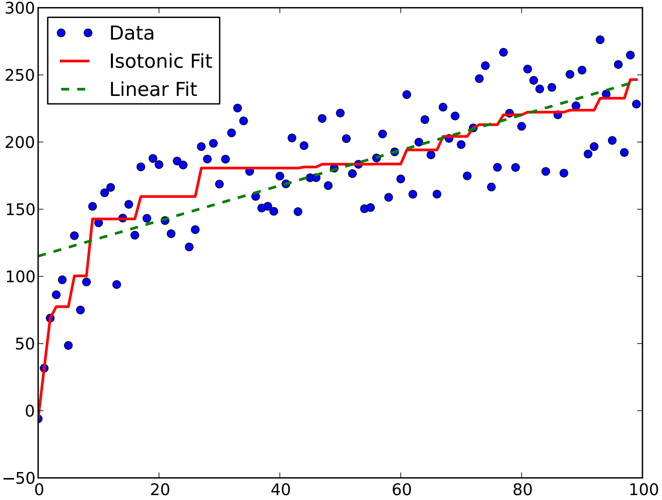
The QP for constraint Least Squares:

$$\begin{aligned} \text{minimize} \quad & \|Ax - b\|_2^2 \\ \text{subject to} \quad & l_i \leqslant x_i \leqslant u_i,\ i = 1, \ldots, n \end{aligned}$$

Examples:

- Estimation of non-negative parameters.
- Isotonic (or Monotonic) Regression, $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$.

Isotonic Regression



Polyhedra Distance

Let $P_1 = \{x \mid A_1 x \leqslant b_1\}$ and $P_2 = \{x \mid A_2 x \leqslant b_2\}$ be two polyhedra in $\mathbb{R}^n$.

$$\text{dist}(P_1, P_2) = \inf\{\|x_1 - x_2\|_2 \mid x_1 \in P_1,\ x_2 \in P_2\}$$

The QP:

$$\begin{aligned} \text{minimize} \quad & \|x_1 - x_2\|_2^2 \\ \text{subject to} \quad & A_1 x_1 \leqslant b_1,\ A_2 x_2 \leqslant b_2 \end{aligned}$$

Bounding Variance

We want to bound the variance of a function f of the RV of Chebyshev Inequalities problem.

$$\mathsf{Var}[f(X)] = \mathbb{E}[f^2(X)] - (\mathbb{E}[f(X)])^2 = \sum f_i^2 p_i - \left(\sum f_i p_i\right)^2$$

QP:

$$\begin{array}{ll} \text{maximize} & \mathsf{Var}[f(X)] \\ \text{subject to} & \alpha_i \leqslant a_i^T p \leqslant \beta_i, \; i = 1, \ldots, m \\ & p \geqslant 0, \; \mathbf{1}^T p = 1 \end{array}$$

Linear Program with Random Cost

Let $c \in \mathbb{R}^n$ be a Random Vector, with mean $\bar{c}$ and covariance $\mathbb{E}(c-\bar{c})(c-\bar{c})^T = \Sigma$.

Basic LP:

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

- Trade-off between small expected cost and small cost variance.
- Define the risk-sensitive cost $\mathbb{E}[c^T x] + \gamma \text{Var}(c^T x)$, where γ is the **risk-aversion** parameter. Is the covariance matrix PSD?

QP:

$$\begin{aligned} \text{minimize} \quad & \bar{c}^T x + \gamma x^T \Sigma x \\ \text{subject to} \quad & Gx \leqslant h \\ & Ax = b \end{aligned}$$

Markowitz portfolio Optimization

- n assets held over a period of time.
- x_i(dollars) amount of asset i held throughout the period.
- p_i relative change in the price of asset i over the period, $r = p^T x$ return of the portfolio.
- We do not allow "shorting" assets, $x \geqslant 0$.
- Total budget is assumed to be 1, $\mathbf{1}^T x = 1$.

We assume p to be a Random Vector with mean $\bar{p}$ and covariance Σ.
QP:

$$\begin{aligned} \text{minimize} \quad & x^T \Sigma x \\ \text{subject to} \quad & \bar{p}^T x \geqslant r_{min} \\ & \mathbf{1}^T x = 1, \ x \geqslant 0 \end{aligned}$$

Markowitz portfolio Optimization

Extensions:

- To allow short positions $x_i < 0$ we introduce x_{long}, x_{short} s.t.

  $x_{long} \geqslant 0,\ x_{short} \geqslant 0,\ x = x_{long} - x_{short},\ \mathbf{1}^T x_{short} \geqslant \eta \mathbf{1}^T x_{long}$

- Include linear transaction costs to go from an initial portfolio x_{init} to a desired portfolio x, which then is held over the period.

$$x = x_{init} + u_{buy} - u_{sell},$$

$$u_{buy} \geqslant 0, u_{sell} \geqslant 0.$$

  Initial buying and selling involves zero net cach:

$$(1 - f_{sell})\mathbf{1}^T u_{sell} = (1 + f_{buy})\mathbf{1}^T u_{buy}$$

$$f_{buy}, f_{sell} > 0.$$

Second-Order Cone Programming

Dual Spaces

Linear Maps

Let $X,\ Y$ be two normed spaces.

- A map $T : X \to Y$ s.t $T(\lambda x_1 + \mu x_2) = \lambda T(x_1) + \mu T(x_2)$ is a **linear map**.
- T is bounded if there is a constant c s.t. $\|Tx\|_Y \leqslant c\|x\|_X$.
  $\|T\| = \min\{c \geqslant 0 : \forall x \in X,\ \|Tx\| \leqslant c\|x\|\}$.
- Operator Norm $\|T\| = \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|} = \sup_{\|x\|=1} \|Tx\|$.
- $\|F\|_2 = \sup\{\|Fx\|_2 \mid \|u\|_2 \leqslant 1\} = \sqrt{\lambda_{\max}(F^T F)}$

Linear Functional

A Linear functional is a Linear Map $F : X \to \mathbb{R}$.

Dual Space

Let X be a normed space. The space X^* of the bounded linear functionals $F : X \to \mathbb{R}$ is the dual space of X.

Dual Norms

Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Its **dual** norm is defined

$$\|z\|_* = \sup\{z^\mathsf{T} x \mid \|x\| \leqslant 1\} = \sup\{|z^\mathsf{T} x| \mid \|x\| \leqslant 1\}$$

- $\|x\|_{**} = \|x\|$. Does not hold in infinite-dimensional vector spaces.
- The ℓ_2 norm is self-dual.
- The dual of ℓ_∞-norm is the ℓ_1-norm.

Definition

Norm Cone: $C = \{(x, t) \mid \|x\| \leqslant t\} \subseteq \mathbb{R}^{n+1}$.
SOCP Definition:

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

- SOCP is a generalization of LP and QCQP.
  $x^T P_0 x + 2q_0^T x + r_0 = \|P_0^{1/2} x + P_0^{-1/2} q_0\|^2 + r_0 - q_0^T P_0^{-1} q_0$
  The optimal values of the QCQP and the SOCP are equal up to a square root and a constant.
- The second-order cone constraint requires that the affine function $(Ax + b, c^T x + d)$ lies in the second-order cone in $\mathbb{R}^{k+1}$

Robust Linear Programming

Often we only know approximations of the coefficients the usual LP:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \leqslant b_i,\ i = 1, \ldots, m \end{array}$$

Assume that c, b_i are known exactly but a_i are known to lie in ellipsoids $\mathcal{E}_i = \{\bar{a}_i + P_i u \mid \|u\|_2 \leqslant 1\}$.
Robust SOCP:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \bar{a}_i{}^T x + \|P_i^T x\|_2 \leqslant b_i,\ i = 1, \ldots, m \end{array}$$

Linear Programming with Random Constraints

Statistical framework for the robust LP.
Each constraint a_i is a Gaussian Random Vector with mean $\bar{a}_i$ and covariance Σ_i and the constraints must hold with confidence at least $\eta \geqslant 1/2$ (Why?)

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \Pr[a_i^T x \leqslant b_i] \geqslant \eta \end{array}$$

Equivalent **SOCP**:

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \bar{a}_i^T x + \Phi^{-1}(\eta)\|\Sigma^{1/2} x\|_2 \leqslant b_i, \; i = 1, \dots, m \end{array}$$

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 $x \preceq_K y \implies f(x) \leqslant f(y)$
Examples:

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

$f : U \to F$ is K-convex if $f(\theta x + (1-\theta)y) \preceq_K \theta f(x) + (1-\theta)f(y)$.
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.

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

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

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

SOCP

SOCP is a Cone Program.

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

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

Semidefinite Programming

K is the cone of semidefinite $k \times k$ matrices, $K = \mathcal{S}_+^k$.

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

- 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

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

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

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

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

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 $u_i \in S^{n-1}$.

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

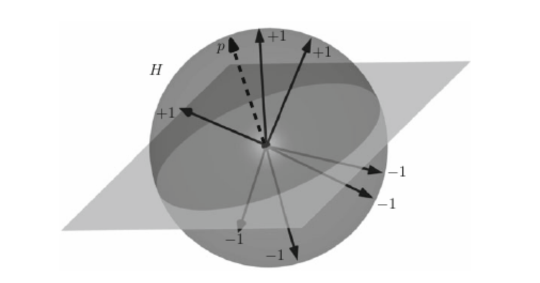
Equivalent Problem:

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

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

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

Vector Optimization

Dual Inequalities

Dual Cone

Let X be a vector space and X^* be its dual

- If $K \subseteq X$ is a cone then its dual cone is the set

$$K^* = \{y \in X^* \mid y^T x \geqslant 0, \text{ for all } x \in K\}$$

- $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$
- $(S^n_+)^* = S^n_+$
- K^* is always convex.
- K proper $\implies$ K^* proper.

Minimal Elements

Dual Inequalities

$x \preceq_K y \Leftrightarrow \lambda^T x \leq \lambda^T y$ for all $\lambda \succeq_{K^*} 0$.

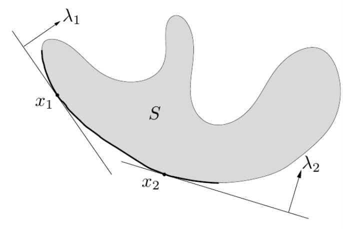
Minimum Element

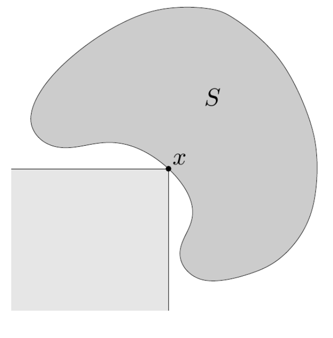
x is minimum in $S \Leftrightarrow$ for all $\lambda \succ_{K^*} 0$, x is the unique minimizer of $\lambda^T z$ over $z \in S \Leftrightarrow$ The hyperplane $\{z \mid \lambda^T (z - x) = 0\}$ is a strict supporting hyperplane to S at x for all $\lambda \in K^*$.

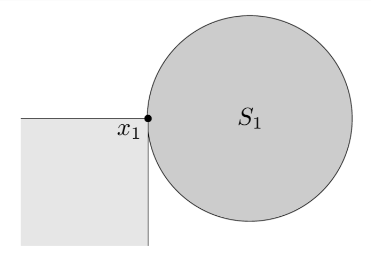
Minimal Elements

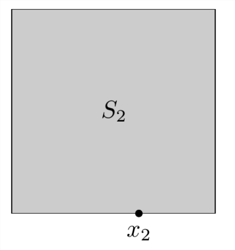
- If $\lambda^T \succ_{K^*} 0$ and x minimizes $\lambda^T z$ over $z \in S$, then x is minimal.
- If S is convex, for any minimal element x there exists nonzero $\lambda \succeq_{K^*} 0$ s.t. x minimizes $\lambda^T z$ over $z \in S$.

Counterexamples









Convex Vector Optimization Problem

Let $f_0 : \mathbb{R}^n \to \mathbb{R}^q$, $K \subseteq \mathbb{R}^q$ a proper cone.

$$\begin{array}{ll} \text{minimize (with respect to K)} & f_0(x) \\ \text{subject to} & f_i(x) \leqslant 0 \\ & h_i(x) = 0 \end{array}$$

- f_0 is K-convex.
- f_i are convex.
- h_i are affine.

A point x^* is optimal iff it is feasible and

$$f_0(D) \subseteq f_0(x^*) + K$$

Scalarization

Pareto Optimal Points

- A point x is Pareto optimal iff it is feasible and $(f_0(x) - K) \cap f_0(D) = \{f_0(x)\}$
- The set of Pareto optimal values, $\mathcal{P}$ satisfies $\mathcal{P} \subseteq f_0(D) \cap \partial f_0(D)$

Scalarization

Let $\lambda \succcurlyeq_{K^*} 0$ be the weight vector.

$$\begin{array}{rl} \text{minimize} & \lambda^T f_0(x) \\ \text{subject to} & f_i(x) \leqslant 0 \\ & h_i(x) = 0 \end{array}$$

If the problem is convex then **every** pareto optimal point is attainable via scalarization.

Minimal Matrix Upper Bound

$$\begin{aligned} \text{minimize (w.r.t } S^n_+) \quad & X \\ \text{subject to} \quad & X \succcurlyeq A_i, \ i = 1, \dots, m \end{aligned}$$

Let $W \in S^n_{++}$ and form the equivalent **SDP**

$$\begin{aligned} \text{minimize (w.r.t } S^n_+) \quad & \mathrm{tr}(WX) \\ \text{subject to} \quad & X \succcurlyeq A_i, \ i = 1, \dots, m \end{aligned}$$

Ellipsoids and Positive Definiteness

$\mathcal{E}_A = \{u \mid u^T A^{-1} u \leqslant 1\}$
$A \preccurlyeq B \Leftrightarrow \mathcal{E}_A \subseteq \mathcal{E}_B$

Duality

Langrangian

Langrangian

$L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, with $\text{dom} L = D \times \mathbb{R}^m \times \mathbb{R}^p$.

$$L(x, \lambda, \mu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \mu_i h_i(x)$$

Dual function

$g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$g(\lambda, \mu) = \inf_{x \in D} L(x, \lambda, \mu)$$

Dual function for $\lambda \geqslant 0$ underestimates the optimal value $g(\lambda, \mu) \leqslant p^*$.

Multicriterion Interpretation

Primal Problem without equality constraints:

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leqslant 0, \ i = 1, \ldots, m \end{aligned}$$

Scalarization of the multicreterion problem:

$$\text{minimize} \quad F(x) = (f_0(x),\ f_1(x), \ldots, f_m(x))$$

Take $\widetilde{\lambda} = (1, \lambda)$ and then minimize

$$\widetilde{\lambda}^T F(x) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x)$$

which is the Langrangian of the Primal Problem.

Nonconvex QCQP

Let $A \in S^n, A \not\succeq 0, b \in \mathbb{R}^n$.

$$\begin{array}{ll} \text{maximize} & x^T A x + 2b^T x \\ \text{subject to} & x^T x \leqslant 1 \end{array}$$

Langrangian:

$$L(x, \lambda) = x^T A x + 2b^T x + \lambda(x^T x - 1) = x^T (A + \lambda I) x + 2b^T x - \lambda$$

Dual Function:

$$g(\lambda) = \begin{cases} -b^T (A + \lambda I)^\dagger b - \lambda, & A + \lambda I \geqslant 0, \quad b \in \mathcal{R}(A + \lambda I) \\ -\infty & \text{otherwise} \end{cases}$$

Nonconvex QCQP

Dual Problem

$$\begin{aligned} \text{maximize} \quad & -b^T(A+\lambda I)^\dagger b - \lambda \\ \text{subject to} \quad & A + \lambda I \succcurlyeq 0, b \in \mathcal{R}(A+\lambda I) \end{aligned}$$

We can find an equivalent **concave** problem

$$\begin{aligned} \text{maximize} \quad & -\sum_{i=1}^{n} \frac{(q_i^T b)^2}{\lambda_i + \lambda} - \lambda \\ \text{subject to} \quad & \lambda \geqslant -\lambda_{\min}(A) \end{aligned}$$

For these problems strong duality obtains.

Rayleigh Quotient

Let $A \in S^n$

$$\text{maximize} \quad \frac{x^T A x}{x^T x}$$

Equivalent problem:

$$\begin{aligned} \text{maximize} \quad & x^T A x \\ \text{subject to} \quad & x^T x \leqslant 1 \end{aligned}$$

Lagrangian: $L(x, \mu) = x^T A x + \lambda(x^T x - 1)$

Derivative

Let E, F be **Banach Spaces**, that is complete normed spaces.

Derivative is a Linear Map

Let U be open in E, and let $x \in U$. Let $f : U \to F$ be a map. f is **differentiable** at x if there exists a continuous linear map $\lambda : E \to F$ and a map ψ defined for all sufficiently small h in E, with values in F, such that

$$\lim_{h \to 0} \psi(h) = 0, \text{ and } f(x+h) = f(x) + \lambda(h) + |h|\psi(h).$$

log(det(X))

$$f(X) : S^n_{++} \to \mathbb{R}, \ \ f(X) = \log\det(X)$$

$$\begin{aligned}
\log\det(X+H) &= \log\det(X+H) \\
&= \log\det\left(X^{1/2}(I + X^{-1/2}HX^{-1/2})X^{1/2}\right) \\
&= \log\det X + \log\det(I + X^{-1/2}HX^{-1/2}) \\
&= \log\det X + \sum_{i=1}^{n} \log(1+\lambda_i) \\
&\simeq \log\det X + \sum_{i=1}^{n} \lambda_i \\
&= \log\det X + \mathrm{tr}(X^{-1/2}HX^{-1/2}) \\
&= \log\det X + \mathrm{tr}(X^{-1}H)
\end{aligned}$$

$$\nabla f(X) = X^{-1}$$

Conjugate of logdet

Conjugate function:

$$f^*(y) = \sup_{x \in D} (y^T x - f(x))$$

$f(X) = \log\det X^{-1}, X \in \mathcal{S}^n_{++}$
The conjugate of f is

$$f^*(Y) = \sup_{X > 0} \left(\operatorname{tr}(YX) + \log\det X\right)$$

- $\operatorname{tr}(YX) + \log\det X$ is unbounded if $Y \nless 0$.
- If $Y < 0$ then setting the gradient with respect to X to zero yields $X_0 = -Y^{-1}$

$f^*(Y) = \log\det(-Y)^{-1} - n = -\log\det(-Y) - n$
dom $f^* = -\mathcal{S}^n_{++}$

Dual of Affine Constraints

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & Ax \leqslant b \\ & Cx = d \end{aligned}$$

$$\begin{aligned} g(\lambda, \mu) &= \inf_x (f_0(x) + \lambda^T (Ax - b) + \mu^T (Cx - d)) \\ &= -b^T\lambda - d^T\mu + \inf_x (f_0(x) + (A^T\lambda - C^T\mu)) \\ &= -b^T\lambda - d^T\mu - f_0^*(-A^T\lambda - C^T\mu) \end{aligned}$$

with $\text{dom}\, g = \{(\lambda, \mu) \mid -A^T\lambda - C^T\mu \in \text{dom}\, f_0^*\}$

Minimum Volume Covering Ellipsoid

Primal

$$\begin{array}{ll} \text{minimize} & f_0(X) = \log\det(X^{-1}) \\ \text{subject to} & a_i^T X a_i \leqslant 1,\ i = 1, \ldots, m \end{array}$$

$a_i^T X a_i \Leftrightarrow \mathrm{tr}(a_i a_i^T X) \leqslant 1$

Dual Function

$$g(\lambda, \nu) = \begin{cases} \log\det\left(\sum_{i=1}^m \lambda_i a_i a_i^T\right) - \mathbf{1}^T\lambda + n, & \sum_{i=1}^m \lambda_i a_i a_i^T > 0 \\ -\infty, & \text{otherwise} \end{cases}$$

Dual

$$\begin{array}{ll} \text{minimize} & \log\det\left(\displaystyle\sum_{i=1}^m \lambda_i a_i a_i^T\right) - \mathbf{1}^T\lambda + n \\ \text{subject to} & \lambda \geqslant 0 \end{array}$$

The weaker Slater condition is satisfied ($\exists X \in S_{++}^n, a_i^T X a_i \leqslant 1, i \in [m]$) and therefore Strong Duality obtains.

The Perturbed Problem

The perturbed version of the convex problem:

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leqslant u_i, \; i = 1, \ldots, m \\ & h_i(x) = v_i, \; i = 1, \ldots, p \end{aligned}$$

The optimal value:

$$p^*(u, v) = \inf\{f_0(x) \mid \exists x \in D, f_i(x) \leqslant u_i, h_i(x) = v_i\}$$

- The optimal value of the unperturbed problem is $p^*(0, 0) = p^*$
- When the perturbations result in infeasibility we have $p^*(u, v) = \infty$.
- $p^*(u, v)$ is convex when the original problem is convex.

A Global Inequality

Assume that the original problem is **convex** and Slater's condition is satisfied.
Let (λ^*, μ^*) be optimal for the dual of the original problem. Then

$$p^*(u, v) \geqslant p^*(0, 0) - {\lambda^*}^T u - {\mu^*}^T v$$

Proof.

$$\begin{aligned} p^*(0,0) &= g(\lambda^*, \mu^*) \\ &\leqslant f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \mu_i^* h_i(x) \\ &\leqslant f_0(x) + {\lambda^*}^T u + {\mu^*}^T v \end{aligned}$$

□

Interpretation of the Global Inequality

$$p^*(u,v) \geqslant p^*(0,0) - {\lambda^*}^T u - {\mu^*}^T v$$

- λ_i^* is large, $u_i < 0$ then $p^*(u,v)$ will increase greatly.
- μ_i^* is large and positive, $v_i < 0$ OR μ_i^* is large and negative, $v_i > 0$ then $p^*(u,v)$ will increase greatly.
- If λ_i^* is small, $u_i > 0$ then $p^*(u,v)$ will not decrease too much.
- If μ_i^* is small and positive ,$v_i > 0$ OR μ_i^* is small and negative and $v_i < 0$
  then $p^*(u,v)$ will not decrease too much.

These results are **not symmetric** with respect to tightening or loosening a constraint.

Local Sensitivity Analysis

Assume strong duality and differentiability of $p^*(u, v)$ at $(0, 0)$.

$$\lambda_i^* = -\frac{\partial p^*}{\partial u_i}\bigg|_{(0,0)}, \quad \mu_i^* = -\frac{\partial p^*}{\partial v_i}\bigg|_{(0,0)}$$

Differentiability of p^* allows a symmetric sensitivity result.

Proof.

$$\frac{\partial p^*}{\partial u_i}\bigg|_{(0,0)} = \lim_{t \to 0} \frac{p^*(te_i, 0) - p^*(0, 0)}{t}$$

From the global inequality we have

$$\frac{p(u, v) - p^*(0, 0)}{t} \geqslant -\lambda_i \text{ if } t > 0 \text{ and } \frac{p(u, v) - p^*(0, 0)}{t} \leqslant -\lambda_i \text{ if } t < 0$$

□

Duality in SDP

Primal SDP:

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

Then

$$\begin{aligned} L(x, Z) &= c^T x + \text{tr}((x_1 F_1 + \ldots + x_n F_n + G) Z) \\ &= x_1 (c_1 + \text{tr}(F_1 Z)) + \ldots + x_n (c_n + \text{tr}(F_n Z)) + \text{tr}(GZ) \end{aligned}$$

Dual function:

$$g(Z) = \inf_x L(x, Z) = \begin{cases} \text{tr}(GZ), & \text{tr}(F_i Z) + c_i = 0, \ i = 1, \ldots, n \\ -\infty, & \text{otherwise} \end{cases}$$

Duality in SDP

Dual Problem:

$$\begin{aligned} \text{minimize} \quad & \mathrm{tr}(GZ) \\ \text{subject to} \quad & \mathrm{tr}(F_i Z) + c_i = 0, \ i = 1, \ldots, n \\ & Z \geqslant 0 \end{aligned}$$

Strong Duality obtains if the SDP is strictly feasible, namely there exists an x with

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

Questions?

References I