Lecture: Convex Optimization Problems

http://bicmr.pku.edu.cn/~wenzw/opt-2019-fall.html

Acknowledgement: this slides is based on Prof. Lieven Vandenberghe's lecture notes

Introduction

- optimization problem in standard form
- convex optimization problems
- quasiconvex optimization
- linear optimization
- quadratic optimization
- generalized inequality constraints
- semidefinite programming
- composite program

Optimization problem in standard form

$$\begin{array}{lll} \min & f_0(x) & \\ \text{s.t.} & f_i(x) \leq 0, & i = 1, ..., m \\ & h_i(x) = 0, & i = 1, ..., p \end{array}$$

- $x \in \mathbb{R}^n$ is the optimization variable
- $f_0 : \mathbb{R}^n \rightarrow \mathbb{R}$ is the objective or cost function
- $f_i : \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, ..., m$, are the inequality constraint functions
- $h_i : \mathbb{R}^n \rightarrow \mathbb{R}$ are the equality constraint functions

optimal value:

$$p^* = \inf\{f_0(x) | f_i(x) \leq 0, i = 1, ..., m, h_i(x) = 0, i = 1, ..., p\}$$

- $p^* = \infty$ if problem is infeasible (no x satisfies the constraints)
- $p^* = -\infty$ if problem is unbounded below

Optimal and locally optimal points

x is **feasible** if $x \in \mathbf{dom}\, f_0$ and it satisfies the constraints

a feasible x is **optimal** if $f_0(x) = p^*$; X_{opt} is the set of optimal points

x is **locally optimal** if there is an $R > 0$ such that x is optimal for

$$\begin{array}{ll} \min(\text{over} z) & f_0(z) \\ \text{s.t.} & f_i(z) \leq 0, \quad i = 1, ..., m \\ & h_i(z) = 0, \quad i = 1, ..., p \\ & \|z - x\|_2 \leq R \end{array}$$

examples (with $n = 1$, $m = p = 0$)

- $f_0(x) = 1/x$, $\mathbf{dom}\, f_0 = \mathbb{R}_{++} : p^* = 0$, no optimal point
- $f_0(x) = -\log x$, $\mathbf{dom}\, f_0 = \mathbb{R}_{++} : p^* = -\infty$
- $f_0(x) = x \log x$, $\mathbf{dom}\, f_0 = \mathbb{R}_{++} : p^* = -1/e, x = 1/e$ is optimal
- $f_0(x) = x^3 - 3x$, $p^* = -\infty$, local optimum at $x = 1$

Implicit constraints

the standard form optimization problem has an **implicit constraint**

$$x \in \mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i,$$

- we call $\mathcal{D}$ the **domain** of the problem
- the constraints $f_i(x) \leq 0$, $h_i(x) = 0$ are the explicit constraints
- a problem is **unconstrained** if it has no explicit constraints ($m = p = 0$)

example:

$$\text{min} \quad f_0(x) = -\sum_{i=1}^{k} \log(b_i - a_i^T x)$$

is an unconstrained problem with implicit constraints $a_i^T x < b_i$

Feasibility problem

$$\begin{array}{ll} \text{find} & x \\ \text{s.t.} & f_i(x) \le 0, \quad i = 1, ..., m \\ & h_i(x) = 0, \quad i = 1, ..., p \end{array}$$

can be considered a special case of the general problem with $f_0(x) = 0$:

$$\begin{array}{ll} \min & 0 \\ \text{s.t.} & f_i(x) \le 0, \quad i = 1, ..., m \\ & h_i(x) = 0, \quad i = 1, ..., p \end{array}$$

- $p^* = 0$ if constraints are feasible; any feasible x is optimal
- $p^* = \infty$ if constraints are infeasible

Convex optimization problem

standard form convex optimization problem

$$\begin{array}{lll} \min & f_0(x) & \\ \text{s.t.} & f_i(x) \leq 0, & i = 1, \ldots, m \\ & a_i^T x = b_i, & i = 1, \ldots, p \end{array}$$

- $f_0, f_1, \ldots, f_m$ are convex; equality constraints are affine
- problem is *quasiconvex* if f_0 is quasiconvex (and $f_1, \ldots, f_m$ convex)

often written as

$$\begin{array}{lll} \min & f_0(x) & \\ \text{s.t.} & f_i(x) \leq 0, & i = 1, \ldots, m \\ & Ax = b & \end{array}$$

important property: feasible set of a convex optimization problem is convex

example

$$\begin{aligned} \min \quad & f_0(x) = x_1^2 + x_2^2 \\ \text{s.t.} \quad & f_1(x) = x_1/(1+x_2^2) \leq 0 \\ & h_1(x) = (x_1+x_2)^2 = 0 \end{aligned}$$

- f_0 is convex; feasible set $\{(x_1, x_2) | x_1 = -x_2 \leq 0\}$ is convex
- not a convex problem (according to our definition): f_1 is not convex, h_1 is not affine
- equivalent (but not identical) to the convex problem

$$\begin{aligned} \min \quad & x_1^2 + x_2^2 \\ \text{s.t.} \quad & x_1 \leq 0 \\ & x_1 + x_2 = 0 \end{aligned}$$

Local and global optima

any locally optimal point of a convex problem is (globally) optimal

proof : suppose x is locally optimal and y is optimal with $f_0(y) < f_0(x)$
x locally optimal means there is an $R > 0$ such that

$$z \text{ feasible}, \quad \|z - x\|_2 \leq R \quad \Longrightarrow \quad f_0(z) \geq f_0(x)$$

consider $z = \theta y + (1 - \theta)x$ with $\theta = R/(2\|y - x\|_2)$

- $\|y - x\|_2 > R$, so $0 < \theta < 1/2$
- z is a convex combination of two feasible points, hence also feasible
- $\|z - x\|_2 = R/2$ and

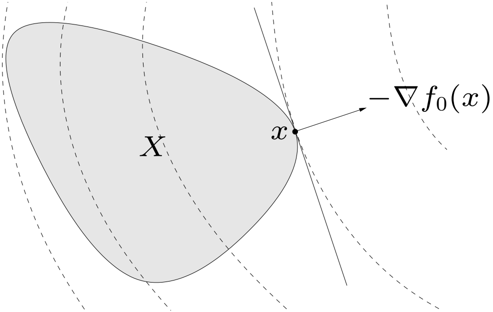
$$f_0(z) \leq \theta f_0(x) + (1 - \theta)f_0(y) < f_0(x)$$

which contradicts our assumption that x is locally optimal

Optimality criterion for differentiable f_0

x is optimal if and only if it is feasible and

$$\nabla f_0(x)^T(y - x) \geq 0 \quad \text{for all feasible } y$$



if nonzero, $\nabla f_0(x)$ defines a supporting hyperplane to feasible set X at x

- **unconstrained problem**: x is optimal if and only if

$$x \in \operatorname{dom} f_0, \quad \nabla f_0(x) = 0$$

- **equality constrained problem**

$$\min \quad f_0(x) \qquad \text{s.t.} \quad Ax = b$$

x is optimal if and only if there exists a υ such that

$$x \in \operatorname{dom} f_0, \quad Ax = b, \quad \nabla f_0(x) + A^T \upsilon = 0$$

- **minimization over nonnegative orthant**

$$\min \quad f_0(x) \qquad \text{s.t.} \quad x \succeq 0$$

x is optimal if and only if

$$x \in \operatorname{dom} f_0, \quad x \succeq 0, \quad \begin{cases} \nabla f_0(x)_i \geq 0 & x_i = 0 \\ \nabla f_0(x)_i = 0 & x_i > 0 \end{cases}$$

Equivalent convex problems

two problems are (informally) **equivalent** if the solution of one is readily obtained from the solution of the other, and vice-versa

some common transformations that preserve convexity:

- **eliminating equality constraints**

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, \quad i = 1, ..., m \\ & Ax = b \end{array}$$

is equivalent to

$$\begin{array}{ll} \min(\text{over} z) & f_0(Fz + x_0) \\ \text{s.t.} & f_i(Fz + x_0) \leq 0, \quad i = 1, ..., m \end{array}$$

where F and x_0 are such that

$$Ax = b \quad \Longleftrightarrow \quad x = Fz + x_0 \text{ for some } z$$

Equivalent convex problems

- **introducing equality constraints**

$$\begin{array}{ll} \min & f_0(A_0x + b_0) \\ \text{s.t.} & f_i(A_ix + b_i) \leq 0, \quad i = 1, ..., m \end{array}$$

is equivalent to

$$\begin{array}{ll} \min(\text{over } x, y_i) & f_0(y_0) \\ \text{s.t.} & f_i(y_i) \leq 0, \quad i = 1, ..., m \\ & y_i = A_ix + b_i, \quad i = 0, 1, ..., m \end{array}$$

- **introducing slack variables for linear inequalities**

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & a^Tx \leq b_i, \quad i = 1, ..., m \end{array}$$

is equivalent to

$$\begin{array}{ll} \min(\text{over } x, s) & f_0(x) \\ \text{s.t.} & a^Tx + s_i = b_i, \quad i = 1, ..., m \\ & s_i \geq 0, \quad i = 1, ...m \end{array}$$

Equivalent convex problems

- **epigraph form**: standard form convex problem is equivalent to

$$\begin{array}{ll} \min(\text{over } x, t) & t \\ \text{s.t.} & f_0(x) - t \leq 0 \\ & f_i(x) \leq 0, \quad i = 1, ..., m \\ & Ax = b \end{array}$$

- **minimizing over some variables**

$$\begin{array}{ll} \min & f_0(x_1, x_2) \\ \text{s.t.} & f_i(x_1) \leq 0, \quad i = 1, ..., m \end{array}$$

is equivalent to

$$\begin{array}{ll} \min & \tilde{f}_0(x_1) \\ \text{s.t.} & f_i(x_1) \leq 0, \quad i = 1, ..., m \end{array}$$

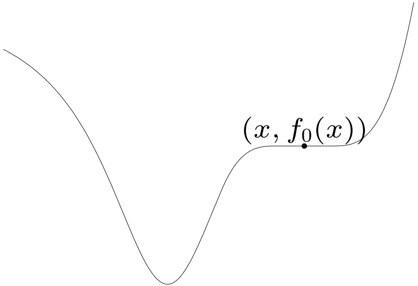
where $\tilde{f}_0(x_1) = \inf_{x_2} f_0(x_1, x_2)$

Quasiconvex optimization

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array}$$

with $f_0 : \mathbb{R}^n \to \mathbb{R}$ quasiconvex, $f_1, \ldots, f_m$ convex

can have locally optimal points that are not (globally) optimal



convex representation of sublevel sets of f_0

if f_0 is quasiconvex, there exists a family of functions ϕ_t such that:

- $\phi_t(x)$ is convex in x for fixed t
- t-sublevel set of f_0 is 0-sublevel set of ϕ_t ,$i.e.$,

$$f_0(x) \leq t \quad \Longleftrightarrow \quad \phi_t(x) \leq 0$$

example

$$f_0(x) = \frac{p(x)}{q(x)}$$

with p convex, q concave, and $p(x) \geq 0$, $q(x) > 0$ on $\operatorname{dom} f_0$
can take $\phi_t(x) = p(x) - tq(x)$:

- for $t \geq 0$, ϕ_t convex in x
- $p(x)/q(x) \leq t$ if and only if $\phi_t(x) \leq 0$

quasiconvex optimization via convex feasibility problems

$$\phi_t(x) \le 0, \qquad f_i(x) \le 0, \quad i = 1, \dots, m, \qquad Ax = b \tag{1}$$

- for fixed t, a convex feasibility problem in x
- if feasible, we can conclude that $t \ge p^*$; if infeasible, $t \le p^*$

Bisection method for quasiconvex optimization

given $l \le p^*$, $u \ge p^*$, tolerance $\epsilon > 0$.
repeat

1. $t := (l + u)/2$.
2. Solve the convex feasibility problem (1).
3. **if** (1) is feasible, $u := t$; **else** $l := t$.

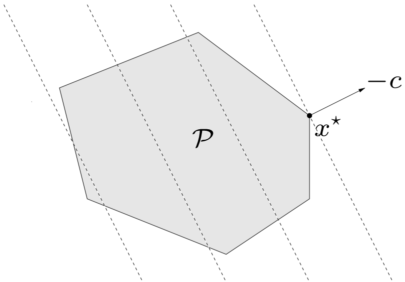
until $u - l \le \epsilon$.

requires exactly $\lceil \log_2((u - l)/\epsilon) \rceil$ iterations (where u, l are initial values)

Linear program (LP)

$$\begin{array}{ll} \min & c^T x + d \\ \text{s.t.} & Gx \leq h \\ & Ax = b \end{array}$$

- convex problem with affine objective and constraint functions
- feasible set is a polyhedron



Examples

diet problem: choose quantities $x_1, \ldots, x_n$ of n foods

- one unit of food j costs c_j , contains amount a_{ij} of nutrient i
- healthy diet requires nutrient i in quantity at least b_i

to find cheapest healthy diet,

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax \geq b, \quad x \geq 0 \end{array}$$

piecewise-linear minimization

$$\min \quad \max_{i=1,\ldots,m}(a_i^T x + b_i)$$

equivalent to an LP

$$\begin{array}{ll} \min & t \\ \text{s.t.} & a_i^T x + b_i \leq t, \quad i = 1, \ldots, m \end{array}$$

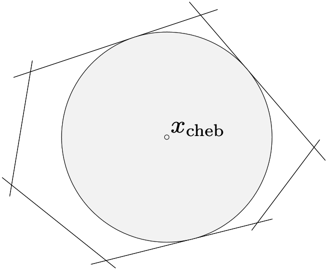
Chebyshev center of a polyhedron

Chebyshev center of

$$\mathcal{P} = \{x | a_i^T x \leq b_i, i = 1, ..., m\}$$

is center of largest inscribed ball

$$\mathcal{B} = \{x_c + u | \ \|u\|_2 \leq r\}$$



- $a_i^T x \leq b_i$ for all $x \in \mathcal{B}$ if and only if

$$\sup\{a_i^T(x_c + u) | \ \|u\|_2 \leq r\} = a_i^T x_c + r\|a_i\|_2 \leq b_i$$

- hence, x_c, r can be determined by solving the LP

$$\begin{array}{ll} \max & r \\ \text{s.t.} & a_i^T x_c + r\|a_i\|_2 \leq b_i, \quad i = 1, ..., m \end{array}$$

Problems with absolute values

$$\begin{aligned} \min \quad & \sum_i c_i |x_i|, \qquad \text{assume } c \geq 0 \\ \text{s.t.} \quad & Ax \geq b \end{aligned}$$

- Reformulation 1:

$$\begin{aligned} \min \quad & \sum_i c_i z_i \\ \text{s.t.} \quad & Ax \geq b \\ & |x_i| \leq z_i \end{aligned} \iff \begin{aligned} \min \quad & \sum_i c_i z_i \\ \text{s.t.} \quad & Ax \geq b \\ & -z_i \leq x_i \leq z_i \end{aligned}$$

- Reformulation 2: $x_i = x_i^+ - x_i^-$, $x_i^+, x_i^- \geq 0$. Then $|x_i| = x_i^+ + x_i^-$

$$\begin{aligned} \min \quad & \sum_i c_i (x_i^+ + x_i^-) \\ \text{s.t.} \quad & Ax^+ - Ax^- \geq b, x^+, x^- \geq 0 \end{aligned}$$

Problems with absolute values

- data fitting:

$$\min_x \quad \|Ax - b\|_\infty$$

$$\min_x \quad \|Ax - b\|_1$$

- Compressive sensing

$$\begin{aligned}
\min \quad & \|x\|_1, \text{ s.t. } \quad Ax = b \quad (LP) \\
\min \quad & \mu\|x\|_1 + \tfrac{1}{2}\|Ax + b\|^2 \quad (QP, SOCP) \\
\min \quad & \|Ax - b\|, \text{ s.t. } \quad \|x\|_1 \leq 1
\end{aligned}$$

Linear-fractional program

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & Gx \leq h \\ & Ax = b \end{aligned}$$

linear-fractional program

$$f_0(x) = \frac{c^T x + d}{e^T x + f}, \quad \operatorname{dom} f_0(x) = \{x | e^T x + f > 0\}$$

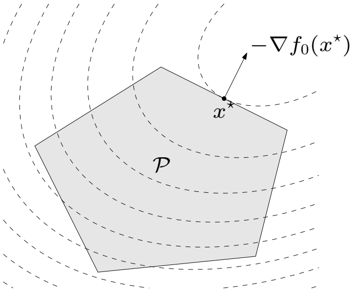
- a quasiconvex optimization problem; can be solved by bisection
- also equivalent to the LP (variables y, z)

$$\begin{aligned} \min \quad & c^T y + dz \\ \text{s.t.} \quad & Gy \leq hz \\ & Ay = bz \\ & e^T y + fz = 1 \\ & z \geq 0 \end{aligned}$$

Quadratic program (QP)

$$
\begin{array}{ll}
\min & (1/2)x^T P x + q^T x + r \\
\text{s.t.} & Gx \leq h \\
& Ax = b
\end{array}
$$

- $P \in \mathbb{S}_+^n$, so objective is convex quadratic
- minimize a convex quadratic function over a polyhedron



Examples

least-squares

$$\min \quad \|Ax - b\|_2^2$$

- analytical solution $x^* = A^\dagger b$ ($A^\dagger$ is pseudo-inverse)
- can add linear constraints, $e.g.$, $l \leq x \leq u$

linear program with random cost

$$\begin{array}{ll} \min & \bar{c}^T x + \gamma x^T \Sigma x = \mathbf{E} c^T x + \gamma \mathrm{var}(c^T x) \\ \text{s.t.} & Gx \leq h, \quad Ax = b \end{array}$$

- c is random vector with mean $\bar{c}$ and covariance Σ
- hence, $c^T x$ is random variable with mean $\bar{c}^T x$ and variance $x^T \Sigma x$
- $\gamma > 0$ is risk aversion parameter; controls the trade-off between expected cost and variance (risk)

Quadratically constrained quadratic program (QCQP)

$$
\begin{aligned}
\min \quad & (1/2)x^T P_0 x + q_0^T x + r_0 \\
\text{s.t.} \quad & (1/2)x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \dots, m \\
& Ax = b
\end{aligned}
$$

- $P_i \in \mathbb{S}^n_+$; objective and constraints are convex quadratic
- if $P_1, \dots, P_m \in \mathbb{S}^n_{++}$, feasible region is intersection of m ellipsoids and an affine set

Generalized inequality constraints

convex problem with generalized inequality constraints

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

- $f_0 : \mathbb{R}^n \to \mathbb{R}$ convex; $f_i : \mathbb{R}^n \to \mathbb{R}^{k_i}$ K_i-convex w.r.t. proper cone K_i
- same properties as standard convex problem (convex feasible set, local optimum is global, etc.)

conic form problem: special case with affine objective and constraints

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Fx + g \preceq_K 0 \\ & Ax = b \end{aligned}$$

extends linear programming ($K = \mathbb{R}^m_+$) to nonpolyhedral cones

Conic quadratic programming

- second-order (quadratic) cone (SOC):

$$Q = \left\{ x \in \mathbb{R}^{n+1} \mid \|\bar{x}\|_2 \le x_1, \quad x = \begin{bmatrix} x_1 \\ \bar{x} \end{bmatrix} \right\}$$

- rotated quadratic cone

$$Q = \left\{ x \in \mathbb{R}^{n+1} \mid \|\bar{x}\|_2^2 \le 2x_1x_2, x_1, x_2 \ge 0, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \bar{x} \end{bmatrix} \right\}$$

- rotated quadratic cone $\|\bar{x}\|^2 \le x_1x_2$, where $x_1, x_2 \ge 0$, is equivalent to

$$\left\| \begin{pmatrix} x_1 - x_2 \\ 2\bar{x} \end{pmatrix} \right\| \le x_1 + x_2$$

Second-order cone programming

$$\begin{array}{ll} \min & f^T x \\ \text{s.t.} & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m \\ & Fx = g \end{array}$$

$(A_i \in \mathbb{R}^{n_i \times n}, F \in \mathbb{R}^{p \times n})$

- the inequalities:

$$(A_i x + b_i, c_i^T x + d_i) \in \text{ second-order cone in } \mathbb{R}^{n_i+1}$$

- for $n_i = 0$, reduces to an LP; if $c_i = 0$, reduces to a QCQP
- more general than QCQP and LP

Quadratic Programming (QP)

$$\begin{array}{ll} \min & q(x) = x^\top Q x + a^\top x + \beta \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array} \qquad \text{assume} \quad Q \succ 0, Q = Q^\top$$

- $q(x) = \|\bar{u}\|^2 + \beta - \frac{1}{4} a^\top Q^{-1} a$, where $\bar{u} = Q^{1/2} x + \frac{1}{2} Q^{-1/2} a$.
- equivalent SOCP

$$\begin{array}{ll} \min & u_0 \\ \text{s.t.} & \bar{u} = Q^{1/2} x + \dfrac{1}{2} Q^{-1/2} a \\ & Ax = b \\ & x \geq 0, \quad (u_0, \bar{u}) \succeq_{\mathcal{Q}} 0 \end{array}$$

Quadratic constraints

$$q(x) = x^\top B^\top Bx + a^\top x + \beta \leq 0$$

is equivalent to

$$(u_0, \bar{u}) \succeq_{\mathcal{Q}} 0,$$

where

$$\bar{u} = \begin{pmatrix} Bx \\ \frac{a^\top x+\beta+1}{2} \end{pmatrix} \quad \text{and } u_0 = \frac{1 - a^\top x - \beta}{2}$$

Norm minimization problems

Let $\bar{v}_i = A_i x + b_i \in \mathbb{R}^{n_i}$.

- $\min_x \quad \sum_i \|\bar{v}_i\|$ is equivalent to

$$\begin{aligned} \min \quad & \sum_i v_{i0} \\ \text{s.t.} \quad & \bar{v}_i = A_i x + b_i \\ & (v_{i0}, \bar{v}_i) \succeq_{\mathcal{Q}} 0 \end{aligned}$$

- $\min_x \quad \max_{1 \le i \le r} \|\bar{v}_i\|$ is equivalent to

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & \bar{v}_i = A_i x + b_i \\ & (t, \bar{v}_i) \succeq_{\mathcal{Q}} 0 \end{aligned}$$

Norm minimization problems

Let $\bar{v}_i = A_i x + b_i \in \mathbb{R}^{n_i}$.

- $\|\bar{v}_{[1]}\|, \ldots, \|\bar{v}_{[r]}\|$ are the norms $\|\bar{v}_1\|, \ldots, \|\bar{v}_r\|$ sorted in nonincreasing order
- $\min_x \quad \sum_{i=1}^k \|\bar{v}_{[i]}\|$ is equivalent to

$$
\begin{aligned}
\min \quad & \sum_{i=1}^m u_i + kt \\
\text{s.t.} \quad & \bar{v}_i = A_i x + b_i, \quad i = 1, \ldots, m \\
& \|\bar{v}_i\| \leq u_i + t, \quad i = 1, \ldots, m \\
& u_i \geq 0, \quad i = 1, \ldots, m
\end{aligned}
$$

Rotated Quadratic Cone

- Minimize the harmonic mean of positive affine functions

$$\min \quad \sum_i 1/(a_i^\top x + \beta_i), \text{ s.t.} \quad a_i^\top x + \beta_i > 0$$

is equivalent to

$$\begin{aligned} \min \quad & \sum_i u_i \\ \text{s.t.} \quad & \bar{v}_i = a_i^\top x + \beta_i \\ & 1 \leq u_i v_i \\ & u_i \geq 0 \end{aligned}$$

- Logarithmic Tchebychev approximation

$$\min_{x} \quad \max_{1 \le i \le r} \quad |\ln(a_i^\top x) - \ln b_i|$$

Since $|\ln(a_i^\top x) - \ln b_i| = \ln \max(a_i^\top x/b_i, b_i/a_i^\top x)$, the problem is equivalent to

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & 1 \le (a_i^\top x/b_i)t \\ & a_i^\top x/b_i \le t \\ & t \ge 0 \end{aligned}$$

- Inequalities involving geometric means

$$\left(\prod_{i=1}^{n}(a_i^\top x + b_i)\right)^{1/n} \ge t$$

- n=4

$$\max \prod_{i=1}^{4}(a_i^\top x - b_i) \iff \begin{array}{ll} \max & w_3 \\ \text{s.t.} & a_i^\top x - b_i \geq 0 \\ & (a_1^\top x - b_1)(a_2^\top x - b_2) \geq w_1^2 \\ & (a_3^\top x - b_3)(a_4^\top x - b_4) \geq w_2^2 \\ & w_1 w_2 \geq w_3^2 \\ & w_i \geq 0 \end{array}$$

- This can be extended to products of rational powers of affine functions

Robust linear programming

the parameters in LP are often uncertain

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & a_i^\top x \le b_i \end{aligned}$$

There can be uncertainty in c, a_i, b.

two common approaches to handling uncertainty (in a_i, for simplicity)

- deterministic model: constraints must hold for all $a_i \in \mathcal{E}_i$

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & a_i^\top x \le b_i, \ \text{for all } a_i \in \mathcal{E}_i \end{aligned}$$

- stochastic model: a_i is random variable; constraints must hold with probability η

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & \mathrm{prob}(a_i^\top x \le b_i) \ge \eta \end{aligned}$$

deterministic approach via SOCP

- Choose an ellipsoid as $\mathcal{E}_i$:

$$\mathcal{E}_i = \{\bar{a}_i + P_i u \mid \|u\|_2 \leq 1\}, \quad \bar{a}_i \in \mathbb{R}^n, P_i \in \mathbb{R}^{n \times n}$$

- Robust LP

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & a_i^\top x \leq b_i, \text{ for all } a_i \in \mathcal{E}_i \end{aligned}$$

  is equivalent to the SOCP

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & \bar{a}_i^\top x + \|P_i^\top x\|_2 \leq b_i \end{aligned}$$

  since

$$\sup_{\|u\|_2 \leq 1} (\bar{a}_i + P_i u)^\top x = \bar{a}_i^\top x + \|P_i^\top x\|_2$$

stochastic approach via SOCP

- a_i is Gaussian with mean $\bar{a}_i$, covariance Σ_i ($a_i \sim \mathcal{N}(\bar{a}_i, \Sigma_i)$
- $a_i^\top x$ is Gaussian r.v. with mean $\bar{a}_i^\top x$, variance $x^\top \Sigma_i x$; hence

$$\mathrm{prob}(a_i^\top x \le b_i) = \Phi\left(\frac{b_i - \bar{a}_i^\top x}{\|\Sigma^{1/2} x\|_2}\right)$$

  where $\Phi(x) = (1/\sqrt{2\pi}) \int_{-\infty}^{x} e^{-t^2/2} dt$ is CDF of $\mathcal{N}(0,1)$
- robust LP

$$\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & \mathrm{prob}(a_i^\top x \le b_i) \ge \eta \end{array}$$

  is equivalent to the SOCP

$$\begin{array}{ll} \min & c^\top x \\ \text{s.t.} & \bar{a}_i^\top x + \Phi^{-1}(\eta) \|\Sigma^{1/2} x\|_2 \le b_i \end{array}$$

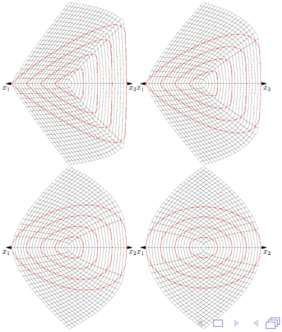
Power cone

https://docs.mosek.com/modeling-cookbook/powo.html

- For $m < n$ and $\alpha_1 + \cdots + \alpha_m = 1$, the power cone is

$$P_n^{\alpha_1,\cdots,\alpha_m} = \left\{ x \in \mathbb{R}^n \mid \Pi_{i=1}^m x_i^{\alpha_i} \geq \sqrt{\sum_{i=m+1}^n x_i^2}, \quad x_1, \cdots, x_m \geq 0 \right\}$$

  The left-hand side is the weighted geometric mean of the x_i
- The boundary of $P_3^{\alpha,1-\alpha}$ for $\alpha = 0.1, 0.2, 0.35, 0.5$



Examples of power cone: x^p

powers: x^p

- for $p > 1$, $t \geq |x|^p$ is equivalent to $t^{1/p} \geq |x|$:

$$t \geq |x|^p \iff (t, 1, x) \in P_3^{1/p, 1-1/p}$$

- for $0 < p < 1$, x^p is concave for $x \geq 0$:

$$t \leq |x|^p, x \geq 0 \iff (x, 1, t) \in P_3^{p, 1-p}$$

- for $p < 0$, x^p is convex for $x > 0$

$$t \geq |x|^p \iff t^{1/(1-p)} x^{-p/(1-p)} \iff (t, x, 1) \in P_3^{1/(1-p), -p/(1-p)}$$

Examples of power cone: p-norm

p-norm: $\|x\|^p = \sum_i \|x_i|^p$

- p-norm cones for $p \geq 1$:

$$\{(t, x) \in \mathbb{R}^{n+1} \mid t \geq \|x\|_p\}$$

  Since $t \geq \|x\|_p$ can be written as $t \geq \sum_i |x_i|^p / t^{p-1}$, we obtain:

$$r_i t^{p-1} \geq |x_i|^p \quad (i.e., (r_i, t, x_i) \in P_3^{1/p, 1-1/p}))$$
$$\sum_i r_i = t$$

- For $0 \neq p < 1$, we can consider:

$$\{(t, x) \in \mathbb{R}^{n+1} \mid 0 \leq t \leq (\sum x_i^p)^{1/p}, x_i \geq 0\}$$

Splitting of general power cone

- Consider the general power cone: $\sum_i \alpha_i = 1$,

$$P_{m+1}^{\alpha_1,\cdots,\alpha_m} = \{x \mid x_1^{\alpha_1} \cdots x_m^{\alpha_m} \geq |z|, x_1, \cdots, x_m \geq 0\}.$$

  Let $s = \alpha_1 + \cdots \alpha_{m-1}$. we can split $P_{m+1}^{\alpha_1,\cdots,\alpha_m}$ as

$$x_1^{\alpha_1/s} \cdots x_m^{\alpha_m/s} \geq |t|, x_1, \cdots, x_{m-1} \geq 0$$
$$t^s x_m^{\alpha_m} \geq |z|, x_m \geq 0.$$

  Hence, $P_{m+1}^{\alpha_1,\cdots,\alpha_m}$ can be expressed using two power cones $P_m^{\alpha_1/s,\cdot,\alpha_{m-1}/s}$ and P_3^{s,α_m}.

- Proceeding by induction shows that it can be expressed the basic three-dimensional cones $P_3^{\alpha,1-\alpha}$.

Non-homogenous constraints

- Consider $\sum_i \alpha_i < \beta$ and $\alpha_i > 0$ and

$$x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_m^{\alpha_m} \geq |z|^{\beta}, x_i \geq 0$$

- equivalent to

$$(x_1, x_2, \cdots, x_m, 1, z) \in P_{m+1}^{\alpha_1/\beta, \cdots, \alpha_m/\beta, s},$$

where $s = 1 - \sum_i \alpha_i / \beta$

Exponential cone

https://docs.mosek.com/modeling-cookbook/expo.html

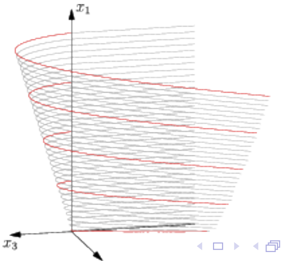
- The exponential cone is a convex subset of $\mathbb{R}^3$ defined as

$$K_{\text{exp}} = \left\{(x_1, x_2, x_3) \mid x_1 \geq x_2 e^{x_3/x_2}, x_2 > 0\right\} \cup \{(x_1, 0, x_3) \mid x_1 \geq 0, x_3 \leq 0\}.$$

  It is the closure of the set of points which satisfy

$$x_1 \geq x_2, x_2 e^{x_3/x_2}, x_1, x_2 > 0$$

- The boundary of the exponential cone. The red isolines are graphs of $x_2 \to x_2 \log(x_1/x_2)$ for fixed x_1



Examples of exponential cone

- Exponential: $t \geq e^x \iff (t, 1, x) \in K_{\exp}$
- Logarithm: $t \leq \log x \iff (x, 1, t) \in K_{\exp}$
- Entropy

$$t \leq -x \log x \iff t \leq x \log(1/x) \iff (1, x, t) \in K_{\exp}$$

- Relative entropy $D(x, y) = x \log(x/y)$

$$t \geq D(x, y) \iff -t \leq x \log(y/x) \iff (y, x, -t) \in K_{\exp}$$

- Softplus function

$$t \geq \log(1 + e^x) \iff e^{x-t} + e^{-t} \leq 1,$$

can be written as

$$u + v \leq 1, \ (u, 1, x - t) \in K_{\exp}, \ (v, 1, -t) \in K_{\exp}$$

Examples of exponential cone

- Log-sum-exp

$$t \geq \log(e^{x_1} + \cdots + e^{x_n}) \Longleftrightarrow e^{x_1 - t} + \cdots + e^{x_n - t} \leq 1$$

can be written as

$$\sum_i u_i \leq 1, \quad (u_i, 1, x_i - t) \in K_{\exp}, i = 1, \cdots, n$$

- Log-sum-inv

$$t \geq \log\left(1/x_1 + \cdots + 1/x_n\right), \quad x_i > 0$$

can be written as

$$t \geq \log(e^{y_1} + \cdots + e^{y_n}) \quad x_i \geq e^{-y_i},\ i = 1, \cdots, n.$$

Examples of exponential cone

- Arbitrary exponential: let a_i be arbitrary positive constants

$$t \geq \alpha_1^{x_1} \alpha_2^{x_2} \cdots \alpha_n^{x_n}$$

which can be written as

$$t \geq \exp \sum_i x_i \log \alpha_i \Longleftrightarrow (t, 1, \sum_i x_i \log \alpha_i) \in K_{\exp}$$

- Lambert W-function is the unique function satisfying

$$W(x)e^{W(x)} = x.$$

We have:

$$\{(x,t) \mid 0 \leq x, 0 \leq t \leq W(x)\} = \{(x,t) \mid x \geq te^t = te^{t^2/t}\}$$

can be written as

$$(x,t,u) \in K_{\exp} \quad (i.e., x \geq t \exp(u/t))$$
$$(1/2, u, t) \in Q \quad (i.e., u \geq t^2)$$

Geometric programming

- monomial function: $f(x) = cx_1^{\alpha_1}x_2^{\alpha_2}\cdots x_n^{\alpha_n}$ with $c > 0$ and α_i are arbitrary real numbers. Let $x_j = e^{y_j}$, then

$$ce^{\alpha_1 y_1}\cdots e^{\alpha_n y_n} = \exp(a^\top y + \log c)$$

- A posynomial (positive polynomial) is a sum of monomials.
- A geometric program (GP): $f_0, \cdots, f_m$ are posynomials:

$$\begin{array}{lll} \min & f_0(x) & \\ \text{s.t.} & f_i(x) \le 1, & i = 1,\dots,m, \\ & x_j > 0, & j = 1,\dots,n. \end{array} \iff \begin{array}{l} \min \;\; t \\ \text{s.t. } \log(\sum_k \exp(a_{0,k}^T y + \log c_{0,k})) \le t, \\ \log(\sum_k \exp(a_{i,k}^T y + \log c_{i,k})) \le 0 \\ i = 1,\dots,m, \end{array}$$

SDP Standard Form

- $\mathcal{S}^n = \{X \in \mathbb{R}^{n\times n} \mid X^\top = X\}$, $\mathcal{S}^n_+ = \{X \in \mathcal{S}^n \mid X \succeq 0\}$,
  $\mathcal{S}^n_{++} = \{X \in \mathcal{S}^n \mid X \succ 0\}$
- Define linear operator $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$:

$$\mathcal{A}(X) = \left(\langle A_1, X\rangle, \ldots, \langle A_m, X\rangle\right)^\top, \quad X \in \mathcal{S}^n.$$

  Since $\mathcal{A}(X)^\top y = \sum_{i=1}^m y_i \langle A_i, X\rangle = \langle \sum_{i=1}^m y_i A_i, X\rangle$, the adjoint of $\mathcal{A}$:

$$\mathcal{A}^*(y) = \sum_{i=1}^m y_i A_i$$

- The SDP standard form:

$$\textbf{(P)} \quad \begin{array}{ll} \min & \langle C, X\rangle \\ \text{s.t.} & \mathcal{A}(X) = b \\ & X \succeq 0 \end{array} \qquad \textbf{(D)} \quad \begin{array}{ll} \max & b^\top y \\ \text{s.t.} & \mathcal{A}^*(y) + S = C \\ & S \succeq 0 \end{array}$$

Facts on matrix calcuation

- If $A, B \in \mathbb{R}^{m\times n}$, then $\mathrm{Tr}(AB^\top) = \mathrm{Tr}(B^\top A)$
- If $U, V \in \mathcal{S}^n$ and Q is orthogonal, then $\langle U, V\rangle = \langle Q^\top UQ, Q^\top UQ\rangle$
- If $X \in \mathcal{S}^n$, then $U = Q^\top \Lambda Q$, where $Q^\top Q = I$ and Λ is diagonal.
- Matrix norms: $\|X\|_F = \|\lambda(X)\|_2$, $\|X\|_2 = \|\lambda(X)\|_\infty$, $\lambda(X) = \mathrm{diag}(\Lambda)$
- $X \succeq 0 \iff v^\top Xv \geq$ for all $v \in \mathbb{R}^n \iff \lambda(X) \geq 0 \iff X = B^\top B$
- The dual cone of $\mathcal{S}^n_+$ is $\mathcal{S}^n_+$
- If $X \succeq 0$, then $X_{ii} \geq 0$. If $X_{ii} = 0$, then $X_{ik} = X_{ki} = 0$ for all k.
- If $X \succeq 0$, then $PXP^\top \succeq 0$ for any P of approriate dimensions
- If $X = \begin{pmatrix} X_{11} & X_{12} \\ X_{12}^\top & X_{22} \end{pmatrix} \succeq 0$, then $X_{11} \succeq 0$.
- $X \succeq 0$ iff every principal submatrix is positive semidefinite (psd).

Facts on matrix calcuation

- Let $U = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}$ with A and C symmetric and $A \succ 0$. Then

$$U \succeq 0 \text{ (or } \succ 0) \quad \Longleftrightarrow \quad C - B^\top A^{-1} B \succeq 0 \text{ (or } \succ 0).$$

  The matrix $C - B^\top A^{-1} B$ is the **Schur complement** of A in U:

$$\begin{pmatrix} A & B \\ B^\top & C \end{pmatrix} = \begin{pmatrix} I & 0 \\ B^\top A^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & C - B^\top A^{-1} B \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix}$$

- If $A \in \mathcal{S}^n$, then $x^\top A x = \langle A, xx^\top \rangle$
- If $A \succ 0$, then $\langle A, B \rangle > 0$ for every nonzero $B \succeq 0$ and $\{B \succeq 0 \mid \langle A, B \rangle \leq \beta\}$ is bounded for $\beta > 0$
- If $A, B \succeq 0$, then $\langle A, B \rangle = 0$ iff $AB = 0$
- $A, B \in \mathcal{S}^n$, then A and B are commute iff AB is symmetric, iff A and B can be simultaneously diagonalized

Semidefinite program (SDP)

$$\begin{array}{ll} \min & b^T y \\ \text{s.t.} & y_1 A_1 + y_2 A_2 + \cdots + y_m A_m \preceq C \\ & By = d \end{array}$$

with $A_i, C \in \mathbb{S}^k$

- inequality constraint is called linear matrix inequality (LMI)
- includes problems with multiple LMI constraints: for example,

$$x_1 \hat{F}_1 + \cdots + x_n \hat{F}_n + \hat{G} \preceq 0, \quad x_1 \tilde{F}_1 + \cdots + x_n \tilde{F}_n + \tilde{G} \preceq 0$$

  is equivalent to single LMI

$$x_1 \begin{bmatrix} \tilde{F}_1 & 0 \\ 0 & \tilde{F}_1 \end{bmatrix} + x_2 \begin{bmatrix} \tilde{F}_2 & 0 \\ 0 & \tilde{F}_2 \end{bmatrix} + \cdots + x_n \begin{bmatrix} \tilde{F}_n & 0 \\ 0 & \tilde{F}_n \end{bmatrix} + \begin{bmatrix} \tilde{G} & 0 \\ 0 & \tilde{G} \end{bmatrix} \preceq 0$$

LP and SOCP as SDP

LP and equivalent SDP

LP: $\min \quad c^T x$ s.t. $Ax \leq b$

SDP: $\min \quad c^T x$ s.t. $\mathrm{diag}(Ax - b) \preceq 0$

(note different interpretation of generalized inequality $\preceq$)

SOCP and equivalent SDP

SOCP:
$$\begin{aligned} \min \quad & f^T x \\ \text{s.t.} \quad & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \dots, m \end{aligned}$$

SDP:
$$\begin{aligned} \min \quad & f^T x \\ \text{s.t.} \quad & \begin{bmatrix} (c_i^T x + d_i) I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d_i \end{bmatrix} \succeq 0, \quad i = 1, \dots, m \end{aligned}$$

Eigenvalue optimization

- minimizing the largest eigenvalue $\lambda_{\max}(A_0 + \sum_i x_i A_i)$:

$$\min \quad \lambda_{\max}(A_0 + \sum_i x_i A_i)$$

can be expressed as an SDP

$$\begin{aligned} \min \quad & z \\ \text{s.t.} \quad & zI - \sum_i x_i A_i \succeq A_0 \end{aligned}$$

and its dual is

$$\begin{aligned} \max \quad & \langle A_0, Y \rangle \\ \text{s.t.} \quad & \langle A_i, Y \rangle = k \\ & \langle I, Y \rangle = 1 \\ & Y \succeq 0 \end{aligned}$$

- follows from

$$\lambda_{\max}(A) \leq t \iff A \preceq tI$$

Eigenvalue optimization

- Let $A_i \in \mathbb{R}^{m\times n}$. Minimizing the 2-norm of $A(x) = A_0 + \sum_i x_i A_i$:

$$\min_{x} \quad \|A(x)\|_2$$

can be expressed as an SDP

$$\begin{aligned} \min_{x,t} \quad & t \\ \text{s.t.} \quad & \begin{pmatrix} tI & A(x) \\ A(x)^\top & tI \end{pmatrix} \succeq 0 \end{aligned}$$

- Constraint follows from

$$\begin{aligned} \|A\|_2 \le t \quad & \Longleftrightarrow \quad A^\top A \preceq t^2 I, \quad t \ge 0 \\ & \Longleftrightarrow \quad \begin{pmatrix} tI & A(x) \\ A(x)^\top & tI \end{pmatrix} \succeq 0 \end{aligned}$$

Eigenvalue optimization

- Let $\Lambda_k(A)$ indicate sum of the k largest eigenvalues of A. Then minimizing $\Lambda_k(A_0 + \sum_i x_i A_i)$:

$$\min \quad \Lambda_k(A_0 + \sum_i x_i A_i)$$

can be expressed as an SDP

$$\begin{aligned} \min \quad & kz + \mathrm{Tr}(X) \\ \text{s.t.} \quad & zI + X - \sum_i x_i A_i \succeq A_0 \\ & X \succeq 0 \end{aligned}$$

since

$$\Lambda_k(A) \le t \iff t - kz - \mathrm{Tr}(X) \ge 0, zI + X \succeq A, X \succeq 0$$

The following problems can be expressed as SDP

- maximizing sum of the k smallest eigenvalues of $A_0 + \sum_i x_i A_i$
- minimizing sum of the k absolute-value-wise largest eigenvalues
- minimizing sum of the k largest singular values of $A_0 + \sum_i x_i A_i$

Quadratically Constrained Quadratic Programming

Consider QCQP

$$\begin{aligned} \min \quad & x^\top A_0 x + 2b_0^\top x + c_0 \qquad \text{assume } A_i \in \mathcal{S}^n \\ \text{s.t.} \quad & x^\top A_i x + 2b_i^\top x + c_i \le 0, \quad i = 1, \dots, m \end{aligned}$$

- If $A_0 \succ 0$ and $A_i = B_i^\top B_i$, $i = 1, \dots, m$, then it is a SOCP
- If $A_i \in \mathcal{S}^n$ but may be indefinite

$$x^\top A_i x + 2b_i^\top x + c_i = \left\langle A_i, xx^\top \right\rangle + 2b_i^\top x + c_i$$

- The original problem is equivalent to

$$\begin{aligned} \min \quad & \mathrm{Tr} A_0 X + 2b_0^\top x + c_0 \\ \text{s.t.} \quad & \mathrm{Tr} A_i X + 2b_i^\top x + c_i \le 0, \quad i = 1, \dots, m \\ & X = xx^\top \end{aligned}$$

QCQP

- If $A_i \in \mathcal{S}^n$ but may be indefinite

$$x^\top A_i x + 2b_i^\top x + c_i = \left\langle \begin{pmatrix} A_i & b_i \\ b_i^\top & c_i \end{pmatrix}, \begin{pmatrix} X & x \\ x^\top & 1 \end{pmatrix} \right\rangle := \langle \bar{A}_i, \bar{X} \rangle$$

$\bar{X} \succeq 0$ is equivalent to $X \succeq xx^\top$

- The SDP relaxation is

$$\begin{aligned} \min \quad & \mathrm{Tr} A_0 X + 2b_0^\top x + c_0 \\ \text{s.t.} \quad & \mathrm{Tr} A_i X + 2b_i^\top x + c_i \le 0, \quad i = 1, \ldots, m \\ & X \succeq xx^\top \end{aligned}$$

- Maxcut: $\max \; x^\top W x,$ s.t. $x_i^2 = 1$
- Phase retrieval: $|a_i^\top x| = b_i$, the value of $a_i^\top x$ is complex

Max cut

- For graph (V, E) and weights $w_{ij} = w_{ji} \geq 0$, the maxcut problem is

$$(Q) \quad \max_{x} \ \frac{1}{2} \sum_{i<j} w_{ij}(1 - x_i x_j), \ \text{s.t.} \ x_i \in \{-1, 1\}$$

- Relaxation:

$$(P) \quad \max_{v_i \in \mathbb{R}^n} \ \frac{1}{2} \sum_{i<j} w_{ij}(1 - v_i^\top v_j), \ \text{s.t.} \ \|v_i\|_2 = 1$$

- Equivalent SDP of (P):

$$(SDP) \quad \max_{X \in S^n} \ \frac{1}{2} \sum_{i<j} w_{ij}(1 - X_{ij}), \ \text{s.t.} \ X_{ii} = 1, X \succeq 0$$

Max cut: rounding procedure

Goemans and Williamson's randomized approach

- Solve (SDP) to obtain an optimal solution X. Compute the decomposition $X = V^\top V$, where

$$V = [v_1, v_2, \ldots, v_n]$$

- Generate a vector r uniformly distributed on the unit sphere, i.e., $\|r\|_2 = 1$
- Set

$$x_i = \begin{cases} 1 & v_i^\top r \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

Max cut: theoretical results

- Let W be the objective function value of x and $E(W)$ be the expected value. Then

$$E(W) = \frac{1}{\pi} \sum_{i<j} w_{ij} \arccos(v_i^\top v_j)$$

- Goemans and Williamson showed:

$$E(W) \geq \alpha \frac{1}{2} \sum_{i<j} w_{ij} (1 - v_i^\top v_j)$$

where

$$\alpha = \min_{0 \leq \theta \pi} \frac{2}{\pi} \frac{\theta}{1 - \cos\theta} > 0.878$$

- Let $Z^*_{(SDP)}$ and $Z^*_{(Q)}$ be the optimal values of (SDP) and (Q)

$$E(W) \geq \alpha Z^*_{(SDP)} \geq \alpha Z^*_{(Q)}$$

复合优化问题一般可以表示为如下形式：

$$\min_{x \in \mathbb{R}^n} \quad \psi(x) = f(x) + h(x),$$

其中$f(x)$ 是光滑函数，$h(x)$ 可能是非光滑的（比如ℓ_1范数正则项，约束集合的示性函数，或他们的线性组合）。令$h(x) = \mu\|x\|_1$:

- ℓ_1 范数正则化回归分析问题：$f(x) = \|Ax - b\|_2^2$或$\|Ax - b\|_1$.
- ℓ_1 范数正则化逻辑回归问题：$f(x) = \sum_{i=1}^m \log(1 + \exp(-b_i \cdot a_i^T x))$.
- ℓ_1 范数正则化支持向量机：$f(x) = C\sum_{i=1}^m \max\{1 - b_i a_i^T x, 0\}$.
- ℓ_1范数正则化精度矩阵估计：$f(x) = -(\log\det(X) - \mathrm{tr}XS)$.
- 矩阵分离问题：$f(X) = \|X\|_*$.

令Ω 是矩阵M 中所有已知元素的下标的集合

- 低秩矩阵恢复

$$\begin{aligned} \min_{X\in\mathbb{R}^{m\times n}} \quad & \operatorname{rank}(X) \\ \text{s.t.} \quad & X_{ij} = M_{ij},\ (i,j)\in\Omega. \end{aligned}$$

- 核范数松弛问题：

$$\begin{aligned} \min_{X\in\mathbb{R}^{m\times n}} \quad & \|X\|_*, \\ \text{s.t.} \quad & X_{ij} = M_{ij},\ (i,j)\in\Omega. \end{aligned}$$

- 二次罚函数形式：

$$\min_{X\in\mathbb{R}^{m\times n}} \quad \mu\|X\|_* + \frac{1}{2}\sum_{(i,j)\in\Omega}(X_{ij}-M_{ij})^2.$$

随机优化问题

- 随机优化问题可以表示成以下形式：

$$\min_{x\in\mathcal{X}} \quad \mathbb{E}_{\xi}[f(x,\xi)] + h(x),$$

其中$\mathcal{X} \subseteq \mathbb{R}^n$ 表示决策变量x 的可行域，ξ 是一个随机变量（分布一般是未知的）。对于每个固定的ξ，$f(x,\xi)$ 表示样本ξ 上的损失或者奖励。正则项$h(x)$ 用来保证解的某种性质。由于变量ξ 分布的未知性，其期望$\mathbb{E}_{\xi}[f(x,\xi)]$一般是不可计算的。为了得到目标函数值的一个比较好的估计，实际问题中往往利用ξ 的经验分布来代替其真实分布。

- 假设有N 个样本$\xi_1,\xi_2,\ldots,\xi_N$，令$f_i(x) = f(x,\xi_i)$，我们得到下面的优化问题

$$\min_{x\in\mathcal{X}} \quad f(x) := \frac{1}{N}\sum_{i=1}^{N} f_i(x) + h(x),$$

其也被称作经验风险极小化问题或者采样平均极小化问题。