Recitation Notes: 801 recitation II optimization ~~types~~ models/examples. Duality, deriving the dual.

Prepared by Yu-Xiang Wang.

Part I: Different types of optimization programs/ mathematical programming

$$LP \subset \overset{(\text{convex})}{QP} \subset \overset{(\text{convex})}{QCQP} \subset SOCP \subset SDP$$

- restrictive, but still versatile
- widely used/ easy to formulate
- Simplex: worst case not polynomial but "usually" fast
- Cplex free for academic users.
- $10^5 - 10^7$ variables/ constraints are quite routine.

- general
- hard to think/formulate
- Computationally intensive
- Usually need customized solvers though generic tools CVX, Sedumi, SDPT3, exists, and work well
- small scale problem usually

- Technology ⟷ Art (turning technology)

LP: $\boxed{\begin{aligned} &\min c^T x + d \\ &\text{s.t. } Gx \le h \\ &\quad Ax = b \end{aligned}}$

Examples: ① minimize cost while meeting demand by choosing the right decision x

e.g. $x_1 \cdots x_n$ are quantities of n foods

$c_1 \cdots c_n$ are the cost of them

A_{ij} is the ith type of nutrition jth item cotains.

b is the requirement of some diet.

$$\min c^T x \quad \text{s.t. } Ax \ge b, \quad x \ge 0$$

② Max-flow min-cut. s: source t: terminal $G(V,E)$

$\max F$

s.t. $f_{ij} \le C_{ij}$ for all $(i,j) \in E$

$\sum_j f_{ij} - \sum_k f_{jk} = 0$ for all $j \in V/\{s,t\}$

$\sum_j f_{ij} - \sum_k f_{jk} = F$ for s

$\sum_j f_{ij} - \sum_k f_{jk} = -F$ for t

$f_{ij} \ge 0$

min-cut is its dual!

⊛ LP is sometimes more powerful that it appears

③ Linear Fractional Programming

Ratio = $\frac{\text{Cost}}{\text{Profit}}$

$$\min \frac{c^T x + d}{e^T x + f} \quad \text{s.t. } Gx \le h,\ Ax = b \text{ and } e^T x + f > 0$$

A non-convex optimization problem!

$\frac{x+1}{x+2} = 1 - \frac{1}{x+2}$

Concave

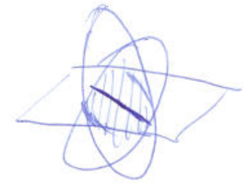
can be reformulated as an LP!

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

we get an LP:

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

then $x^* = \frac{y^*}{z^*}$



QP (quadratic programming)

$$\min \tfrac{1}{2}x^TPx + q^Tx + r$$
$$\text{s.t. } Gx \leq h,\ Ax = b$$

This is convex when $P \succeq 0$

EX ① Least squares
$$\min_x \|Ax-b\|^2$$

② soft-thresholding operator
$$\min \tfrac{1}{2}\|x-b\|^2 + \lambda\|x\|_1$$

③ Lasso, ridge regression

④ SVM, soft-margin SVM
$$\min_{w,\xi,b} \tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i$$
$$\text{s.t. } y_i(w^Tx_i - b) \geq 1-\xi_i,\ \xi_i \geq 0$$

QCQP (quadratically constrained QP)

$$\min \tfrac{1}{2}x^TP_0x + q_0^Tx + r_0$$
$$\text{s.t. } \tfrac{1}{2}x^TP_ix + q_i^Tx + r_i \leq 0 \text{ for } \forall i$$
$$Ax = b$$

Ellipsoids!

We need $P_i \in S_+^n$

EX Constrained version of ~~[illegible]~~

① Elastic-net.
$$\min \tfrac{1}{2}\|Ax-b\|^2$$
$$\text{s.t. } \|x\|_1 + \tfrac{1}{2}\|x\|^2 < \delta$$

② Sparse Representation under fixed tolerance δ.
$$\min |x|_1$$
$$\text{s.t. } \|Ax-b\|^2 \leq \delta$$

SOCP

③ Group Sparsity

polytope

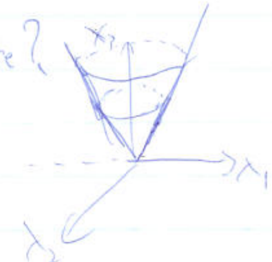
$$\min_x \sum_{i=1}^{k} \|x_i\|_2 \text{ s.t. } x \in P$$
$$\Downarrow$$
$$\min_{x,y} \sum_{i=1}^{k} y_i \quad \text{(A SOCP!)}$$
$$\text{s.t. } x \in P$$
$$y_i \geq \|x_i\|_2 \text{ for } \forall i$$

SOCP (second order cone programming)

$$\min f^Tx$$
$$\text{s.t. } \|A_ix + b_i\|_2 \leq c_i^Tx + d_i,\ i=1,\dots,m$$
$$Fx = g$$

when $c_i = 0$, QCQP.

How does it look like? $\left\|\begin{pmatrix}x_1\\x_2\end{pmatrix}\right\|_2 \leq x_3$



EX The most important application is probably ① Robust Optimization. / ② Stochastic Optimization

LP with ∞ # of constraints

R.O. $\min c^Tx$
$$\text{s.t. } a_i^Tx \leq b_i \text{ for all } a_i \in \mathcal{E}_i,\ i=1,\dots,m$$

uncertainty set!

Books: Ben-Tal, Nemirovski, Papers: Bertsimas & Sim

S.O. $\min c^Tx$
$$\text{s.t. } P_r(a_i^Tx \leq b_i) \geq 1-\delta,\ i=1,\dots,m$$

chance constraint!

Books: Birge & Louveaux Intro to Stochastic programming

Remark: R.O. and S.O. are often equivalent ... Both use SOCP!!!

Choose $\mathcal{E}_i = \{\bar{a}_i + P_iu \mid \|u\|_2 < 1\}$

This is an ellipsoidal uncertainty set.

When $P_i = \lambda I$, this is a ball.

Robust to perturbation, feasible/meet the demand with high margin.

SOCP: $\min c^Tx$
$$\text{s.t. } \bar{a}_i^Tx + \|P_i^Tx\|_2 \leq b_i$$

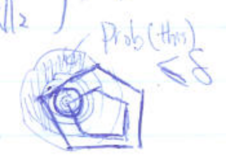
$\because \sup_{\|u\|_2 \leq 1} (\bar{a}_i + P_iu)^Tx = \bar{a}_i^Tx + \|P_i^Tx\|_2$

$a_i \sim N(\bar{a}_i, \Sigma_i)$ assume Gaussian model

$$P_r(a_i^Tx \leq b_i) = \Phi\left(\frac{b_i - \bar{a}_i^Tx}{\|\Sigma_i^{\frac{1}{2}}x\|_2}\right)$$

Φ is CDF

SOCP: $\min c^Tx$
$$\text{s.t. } \bar{a}_i^Tx + \Phi^{-1}(1-\delta)\|\Sigma_i^{\frac{1}{2}}x\|_2 \leq b_i$$



SDP (Semidefinite Programming)

$$\boxed{\begin{array}{l} \min_x \ c^T x \qquad\qquad \text{LMI} \\ \text{s.t. } x_1F_1 + x_2F_2 + \cdots + x_nF_n + G \preceq 0 \\ \quad Ax = b, \ F_i, G \in S^k \end{array}}$$

$Ax=b$ can be formulated as an LMI too!
How?

$$\left\{ \underbrace{\boxed{\diagdown}}_{A_1} x_1 + \underbrace{\boxed{\diagdown}}_{A_2} x_2 + \cdots + \underbrace{\boxed{\diagdown}}_{A_n} x_n \ \substack{\preceq \\ \succeq} \ \overset{b}{\boxed{\diagdown}} \right.$$

Examples ⓪ MaxCUT (discussed with details in class)

① Eigenvalue minimization

$$\min \lambda_{max}(A(x)) \iff \min t \quad \text{s.t. } A(x) \preceq tI \qquad \because \lambda_{max}(A(x)) \le t \iff A \preceq tI$$

$A(x) = A_0 + A_1x_1 + A_2x_2 + \cdots + A_nx_n; \ A_i \in S^k$

② Nuclear norm minimization

$$\|X\|_* = \min_{W_1, W_2} \frac{1}{2}\left(tr(W_1) + tr(W_2)\right) \quad \text{s.t. } \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0$$

Matrix Completion:
$$\min_X \|X\|_* \quad \text{s.t. } P_\Omega(X) = \underbrace{P_\Omega}_{\text{mask}}(\overset{\text{observation}}{Y}) \iff \min_{X, W_1, W_2} \frac{1}{2}(tr W_1 + tr W_2) \quad \text{s.t. } P_\Omega(X) = P_\Omega(Y), \ \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0$$

Robust PCA
$$\min_{L,S} \|L\|_* + \|S\|_1 \quad \text{s.t. } L + S = \overset{\text{observation}}{Y} \iff \text{SDP similarly.}$$

* Note: SDP is powerful in formulating almost all interesting convex optimization into its standard form, and for analyzing a problem, or providing a lower bound.
But it is often too slow to use for real problems in practice. There are way faster algorithms to solve MC and RPCA.

** SDP does not always have strong duality. Need to check Slater's Condition.

Nuclear norm identity for cases when X is square. (Correcting what I've written wrongly on the black board.)

Again: change-of-variable. Idea ① $\dfrac{trW_1 + trW_2}{2} > tr\left(\sqrt{W_1W_2}\right)$

② $tr\sqrt{XX^T} = \|X\|_*$

③ $\begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0 \Rightarrow W_1W_2 \succeq XX^T \Rightarrow \sqrt{W_1W_2} \succeq \sqrt{XX^T} \Rightarrow tr\sqrt{W_1W_2} \geq tr\sqrt{XX^T}$