Mathematical Optimization Models and Applications	June 15,2022

Lecture 1

Lecturer: Yinyu Ye *Scriber: Bin Yu*

1 Introduction

In this section, we will give the belief introduction about the linear program and basic idea of optimization. The detail you can check out in the textbook Luenberger et al. (1984)

$$\begin{aligned} \min \quad & c^T x \\ s.t. \quad & Ax = b \\ & x \in K \end{aligned} \tag{1}$$

This is the conic linear program depending on the set K

- linear program: when K is the non-negative orthant cone
- second order cone programming: when K is the second order cone
- semidefinite cone programming: when K is the semidefinite matrix cone

$$\begin{aligned} \min \quad & 2x_1 + x_2 + x_3 \\ s.t. \quad & x_1 + x_2 + x_3 = 1 \\ & x_1, x_2, x_3 \geq 0 \end{aligned} \qquad \begin{aligned} \min \quad & 2x_1 + x_2 + x_3 \\ s.t. \quad & x_1 + x_2 + x_3 = 1 \\ & \sqrt{x_2^2 + x_3^2} \leq x_1 \end{aligned} \qquad \begin{aligned} \min \quad & 2x_1 + x_2 + x_3 \\ s.t. \quad & x_1 + x_2 + x_3 = 1 \\ & \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix} \succeq 0 \end{aligned} \tag{2}$$

They are LP, SOCP, and SDP respectively.

Definition 1. *A symmetric matrix M with real entries is positive-definite if the real number $z^T M z$ is positive for every nonzero real column vector z*

1.1 Facility Location Problem

Definition 2. *For a real number p, the p-norm or L^p-norm of x is defined by*

$$||x||_p = (\sum_i |x_i|^p)^{1/p}$$

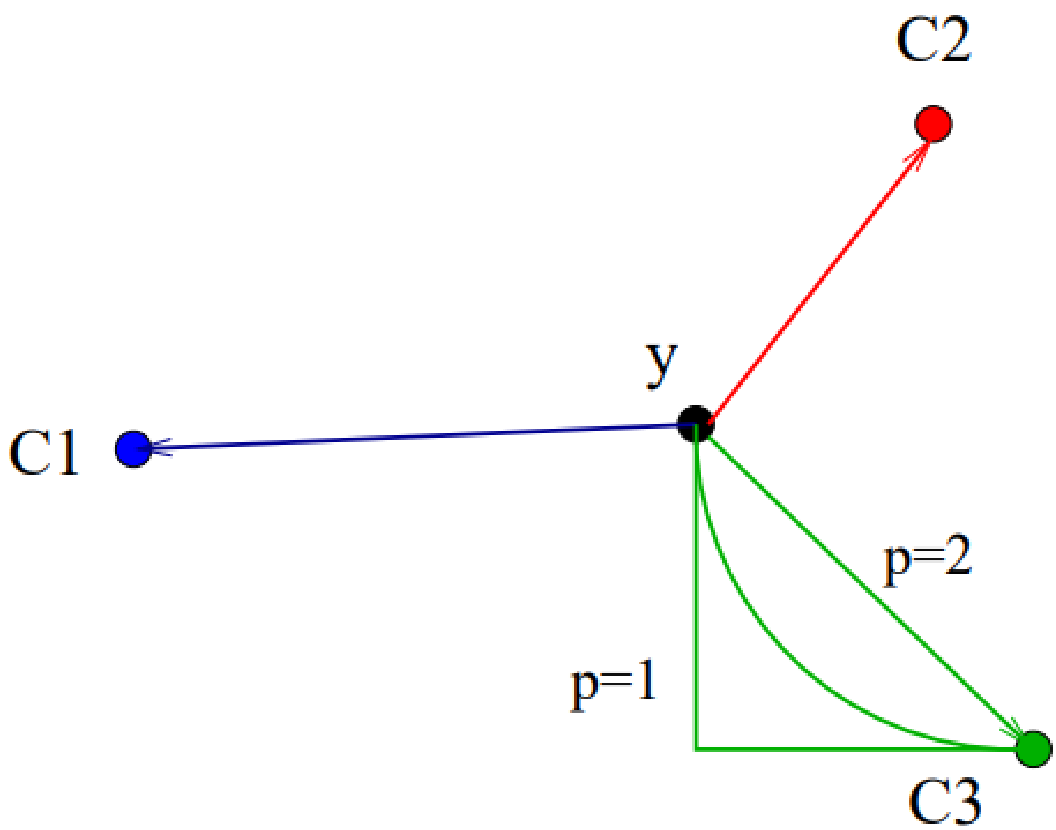

Figure 1: facility allocation

Consider this unconstrained optimization, and let c_j be the location of client $j = 1, 2, ..., m$ and y be the location decision of a facility to be built. Then we solve

$$\min \quad \sum_j ||y - c_j||_p \tag{3}$$

In the sense, we will get the different optimal solution depending the value of p. The figure will illustrate Obviously, the green lines represents the different norm $p \in [1, 2]$

1.2 Sparse Linear Regression Problems

Our target is to minimize the number of non-zero entries in x such that $Ax = b$

$$\begin{aligned} \min \quad & ||x||_0 = |\{j : x_j \neq 0\}| \\ s.t. \quad & Ax = b \end{aligned} \tag{4}$$

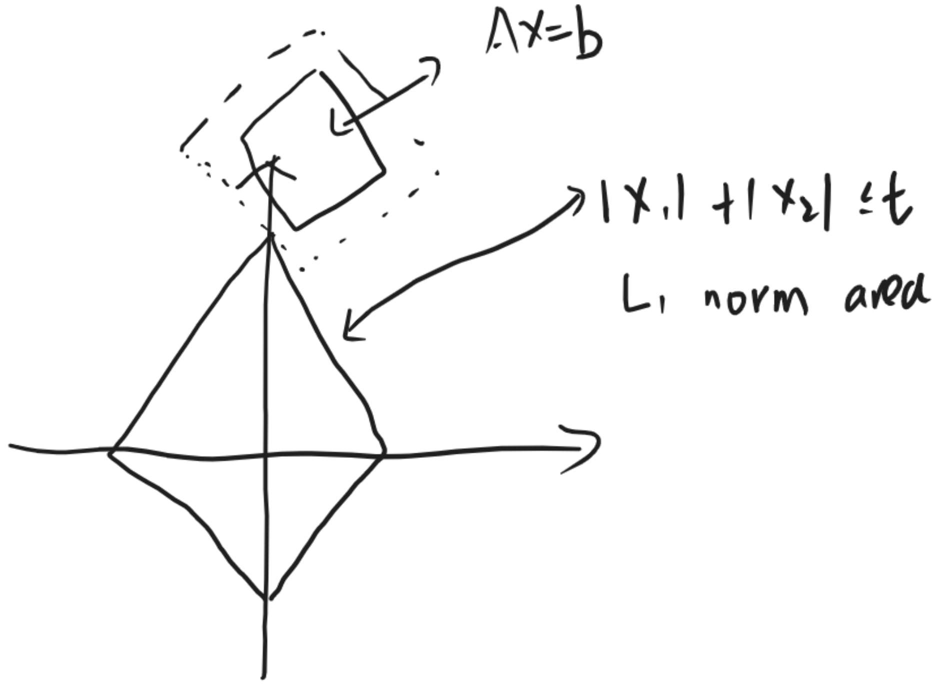

Figure 2: Approximation of Sparse Problem with 1-norm

Our target to minimize the number of zero in the vector or the rank in the matrix. Sometimes this objective can be accomplished by LASSO

$$\begin{aligned} \min \quad & ||x||_1 = \sum_{i=1}^{n} |x_j| \\ s.t. \quad & Ax = b \end{aligned} \tag{5}$$

In this figure, we can illustrate why we could approximate the sparse problem by 1-norm. In the two dim, the polyhedron will touch the vertex of the 1-norm. In the sense, the vertex will $(1,0)$ or $(0,1)$. It will reduce the number of zero in the vector. Moreover, we also use the p-norm to solve this problem

$$\min \quad ||Ax-b||^2 + \beta(\sum_{j=1}^{n} |x_j|^p) \tag{6}$$

Usually, we let $p = \frac{1}{2}$. Then we use the cross validation method to estimate β

1.3 Support Vector Machine

Denote $a_i \in \mathbb{R}^d$ and $b_j \in \mathbb{R}^d$. We like to find a hyperplant, slope vector x and intersect scalar x_0

$$\begin{aligned} s.t. \quad & \mathbf{a}_i^T\mathbf{x} + x_0 + \geq 1, \forall i \\ & \mathbf{b}_j^T\mathbf{x} + x_0 \leq -1, \forall j \end{aligned} \tag{7}$$

This is a linear program with the null objective. Frequently we add the regularization term on the slope vector

$$\begin{aligned} \min \quad & \beta + \mu\|\mathbf{x}\|^2 \\ s.t. \quad & \mathbf{a}_i^T\mathbf{x} + x_0 + \beta \geq 1, \forall i \\ & \mathbf{b}_j^T\mathbf{x} + x_0 - \beta \leq -1, \forall j \\ & \beta \geq 0 \end{aligned} \tag{8}$$

The β is the error parameter. It is the relaxation method to compute the hyperplane. Moreover, we can imagine the separating boundary is the ellipsoid. Let recall the college knowledge

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1$$

$$\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} \frac{1}{a^2} & 0 \\ 0 & \frac{1}{b^2} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \leq 1$$

Then we can have the ellipsoidal separation

$$\begin{aligned} \min \operatorname{trace}(X) + \|\mathbf{x}\|^2 & \\ \text{subject to } \mathbf{a}_i^T X \mathbf{a}_i + \mathbf{a}_i^T\mathbf{x} + x_0 \geq 1, \forall i & \\ \mathbf{b}_j^T X \mathbf{b}_j + \mathbf{b}_j^T\mathbf{x} + x_0 \leq -1, \forall j & \\ X \succeq \mathbf{0} & \end{aligned} \tag{9}$$

1.4 Transportation Problem

we consider the classic transportation problem. In the setting, there are seller and buyer, demand and supply respectively.

$$\begin{array}{lll} \min & \sum_{i=1}^m \sum_{j=1}^n c_{ij}x_{ij} & \\ \text{s.t.} & \sum_{j=1}^n x_{ij} & = s_i, \forall i = 1, \ldots, m \\ & \sum_{i=1}^m x_{ij} & = d_j, \forall j = 1, \ldots, n \\ x_{ij} & \geq 0, \forall i, j. & \end{array} \tag{10}$$

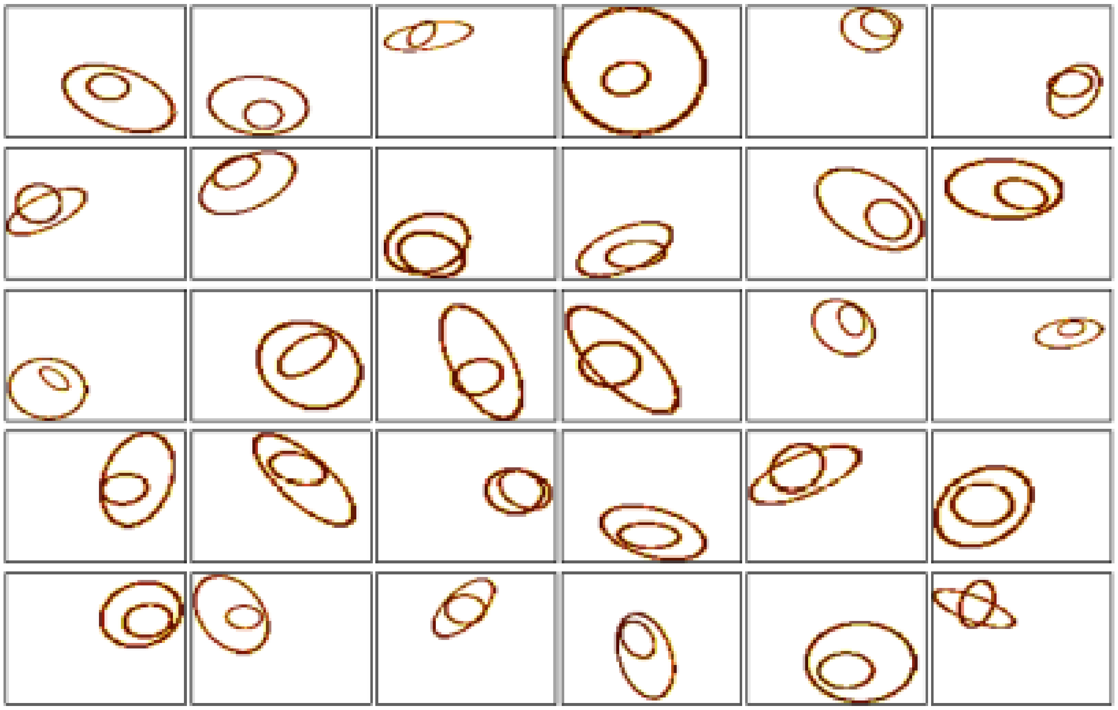

Figure 3: Sample

The minimal transportation cost is called the Wasserstein Distance (WD) between supply distribution s and demand distribution d (can be interpreted as two probability distributions after normalization). This is a linear program.

The Wasserstein Barycenter Problem is to find a distribution such that the sum of its Wasserstein Distance to each of a set of distributions would be minimized. And you can check out this example.

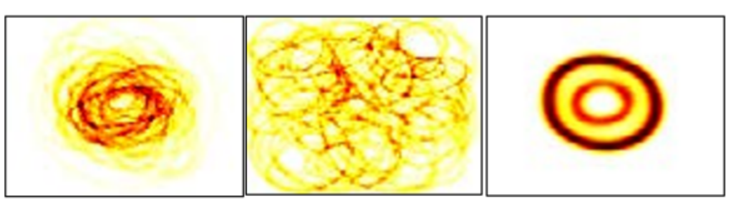

Figure 3: Mean picture constructed from the (a) Euclidean mean after re-centering images (b) Euclidean mean (c) Wasserstein Barycenter (self recenter, resize and rotate)

Mathematical Optimization Models and Applications	June 22,2022

Lecture 2

Lecturer: Yinyu Ye *Scriber: Bin Yu*

2 Optimization Application

2.1 Graph Realization and Sensor Network Localization

Given a graph $G = (V, E)$ and sets of non–negative weights, say $\{d_{ij} : (i, j) \in E\}$, the goal is to compute a realization of G in the Euclidean space $\mathbb{R}^d$ for a given low dimension d, where the distance information is preserved

$$\begin{aligned} \|\mathbf{x}_i - \mathbf{x}_j\|^2 &= d_{ij}^2, \forall (i,j) \in N_x, i < j \\ \|\mathbf{a}_k - \mathbf{x}_j\|^2 &= \hat{d}_{kj}^2, \forall (k,j) \in N_a. \end{aligned} \tag{11}$$

This is the quadratic optimization

$$\min_{\mathbf{x}_i \forall i} \sum_{(i,j)\in N_x} \left(\|\mathbf{x}_i - \mathbf{x}_j\|^2 - d_{ij}^2\right)^2 + \sum_{(k,j)\in N_a} \left(\|\mathbf{a}_k - \mathbf{x}_j\|^2 - \hat{d}_{kj}^2\right)^2 \tag{12}$$

we have two directions to relax this problem: SOCP and SDP

1. change $=$ to $\leq$
2. SDP relaxation

   Let $X = [x_1 x_2 ... x_n]$ be the d × n matrix that needs to be determined and e_j be the vector of all zero except 1 at the jth position. Then

$$\|\mathbf{x}_i - \mathbf{x}_j\|^2 = (\mathbf{e}_i - \mathbf{e}_j)^T X^T X (\mathbf{e}_i - \mathbf{e}_j) \tag{13}$$

$$\begin{aligned} \|\mathbf{a}_k - \mathbf{x}_j\|^2 = (\mathbf{a}_k; -\mathbf{e}_j)^T \left[\begin{array}{cc} I & X \end{array}\right]^T \left[\begin{array}{cc} I & X \end{array}\right] (\mathbf{a}_k; -\mathbf{e}_j) = \\ (\mathbf{a}_k; -\mathbf{e}_j)^T \begin{pmatrix} I & X \\ X^T & X^T X \end{pmatrix} (\mathbf{a}_k; -\mathbf{e}_j). \end{aligned} \tag{14}$$

Convex relaxation first and steepest-descent-search second strategy

2.2 Markov Decision Process

- An MDP problem is defined by a given number of states, indexed by i, where each state has a number of actions, A_i , to take. Each action, say $j \in A_i$, is associated with an (immeidiate) cost $c_j$of taking, and a probability distribution p_j to transfer to all possible states at the next time period

- A stationary policy for the decision maker is a function $\pi = \{\pi_1, \pi_2, ..., \pi_m\}$ that specifies an action in each state,$\pi_i \in A_i$, that the decision maker will take at any time period; which also lead to a cost-to-go value for each state

- The MDP is to find a stationary policy to minimize/maximize the expected discounted sum over the infinite horizon with a discount factor $0 \le \gamma \le 1$

$$\sum_{t=0}^{\infty} \gamma^t E\left[c^{\pi_i t}\left(i^t, i^{t+1}\right)\right]. \tag{15}$$

- If the states are partitioned into two sets, one is to minimize and the other is to maximize the discounted sum, then the process becomes a two-person turn-based zero-sum stochastic game

Mathematical Optimization Models and Applications	June 26,2022
Lecture 3	
Lecturer: Yinyu Ye	*Scriber: Bin Yu*

3 Global and Local Optimizers

A global minimizer for (P) is a vector $x^\star$ such that

$$\mathbf{x}^* \in \mathcal{X} \quad \text{and} \quad f\left(\mathbf{x}^*\right) \leq f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X} \tag{16}$$

Sometimes one has to settle for a local minimizer, that is, a vector $\bar{x}$ such that

$$\overline{\mathbf{x}} \in \mathcal{X} \quad \text{and} \quad f(\overline{\mathbf{x}}) \leq f(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X} \cap N(\bar{x}) \tag{17}$$

where $N(\bar{x})$ is a neighborhood of $\bar{x}$. Typically, $N(\bar{x}) = B_\delta(\bar{x})$, an open ball centered at $\bar{x}$ having suitably small radius $\delta > 0$.

4 Size and Complexity of Problems

- number of decision variables
- number of constraints
- bit size/number required to store the problem input data (mean the memory consumption)
- problem difficulty or complexity number
- algorithm complexity or convergence speed

the definition of L_p norm is following

$$\|\mathbf{x}\|_p = \left(\sum_{j=1}^{n} |x_j|^p\right)^{1/p} \tag{18}$$

For $x^k, x^\star \in R^n$ and $0 < \gamma < 1$ contraction sequence is

$$\left\|\mathbf{x}^{k+1} - \mathbf{x}^*\right\| \leq \gamma \left\|\mathbf{x}^k - \mathbf{x}^*\right\|, \forall k \geq 0 \tag{19}$$

inner product

$$A \bullet B = \operatorname{tr} A^T B = \sum_{i,j} a_{ij} b_{ij} \tag{20}$$

The operator norm of matrix A

$$\|A\|^2 := \max_{\mathbf{0} \neq \mathbf{x} \in \mathcal{R}^n} \frac{\|A\mathbf{x}\|^2}{\|\mathbf{x}\|^2} \tag{21}$$

The Frobenius norm of matrix A

$$\|A\|_f^2 := A \bullet A = \sum_{i,j} a_{ij}^2 \tag{22}$$

Theorem 3. *Perron-Frobenius Theorem: a real square matrix with positive entries has a unique largest real eigenvalue and that the corresponding eigenvector can be chosen to have strictly positive components.*

Stochastic Matrices: $A \leq 0$ with $e^T A = e^T$ (Column-Stochastic), or $Ae = e$ (Row-Stochastic), or Doubly-Stochastic if both. It has a unique largest real eigenvalue 1 and corresponding non-negative right or left eigenvector.

5 Affine Set

$S \subset R^n$ is affine if

$$[\mathbf{x}, \mathbf{y} \in S \text{ and } \alpha \in R] \Longrightarrow \alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in S \tag{23}$$

When $\mathbf{x}$ and $\mathbf{y}$ are two distinct points in R^n and α runs over R,

$$\{\mathbf{z} : \mathbf{z} = \alpha\mathbf{x} + (1-\alpha)\mathbf{y}\} \tag{24}$$

is the affine combination of x and y. When $0 \leq \alpha \leq 1$, it is called the convex combination of $\mathbf{x}$ and $\mathbf{y}$. More points? For multipliers $\alpha \geq 0$ and for $\beta \geq 0$

$$\{\mathbf{z} : \mathbf{z} = \alpha\mathbf{x} + \beta\mathbf{y}\}, \tag{25}$$

is called the conic combination of x and y. It is called linear combination if both α and β are "free".

5.1 Convex Set

Ω is said to be a convex set if for every $\mathbf{x}^1, \mathbf{x}^2 \in \Omega$ and every real number $\alpha \in [0, 1]$, the point $\alpha\mathbf{x}^1 + (1-\alpha)\mathbf{x}^2 \in \Omega$

Ball and Ellipsoid: for given $\mathbf{y} \in \mathcal{R}^n$ and positive definite matrix $Q:$ $E(\mathbf{y}, Q) = \left\{\mathbf{x} : (\mathbf{x}-\mathbf{y})^T Q(\mathbf{x}-\mathbf{y}) \leq 1\right\}$

The intersection of convex sets is convex, the sum-set of convex sets is convex, the scaled-set of a convext set is convex

The convex hull of a set Ω is the intersection of all convex sets containing Ω. Given column-points of A, the convex hull is $\left\{\mathbf{z} = A\mathbf{x} : \mathbf{e}^T\mathbf{x} = 1, \mathbf{x} \geq \mathbf{0}\right\}$

SVM Claim: two point sets are separable by a plane if any only if their convex hulls are separable.

An extreme point in a convex set is a point that cannot be expressed as a convex combination of other two distinct points of the set

A set is polyhedral if it has finitely many extreme points; $\{\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ and $\{\mathbf{x} : A\mathbf{x} \leq \mathbf{b}\}$ are convex polyhedral.

The dual norm

$$C^* := \{\mathbf{y} : \mathbf{x} \bullet \mathbf{y} \geq 0 \text{ for all } \mathbf{x} \in C\} \tag{26}$$

Theorem 4. *The dual is always a closed convex cone, and the dual of the dual is the closure of convex hall of C*

5.2 Cone Example

Example 1: The n-dimensional non-negative orthant, $\mathcal{R}^n_+ = \{\mathbf{x} \in \mathcal{R}^n : \mathbf{x} \geq \mathbf{0}\}$, is a convex cone. Its dual is itself.

Example 2: The set of all PSD matrices in $\mathcal{S}^n, \mathcal{S}^n_+$, is a convex cone, called the PSD matrix cone. Its dual is itself.

Example 3: The set $\left\{(t; \mathbf{x}) \in \mathcal{R}^{n+1} : t \geq \|\mathbf{x}\|_p\right\}$ for a $p \geq 1$ is a convex cone in $\mathcal{R}^{n+1}$, called the p-order cone. Its dual is the q-order cone with $\frac{1}{p} + \frac{1}{q} = 1$. The dual of the second-order cone ($p = 2$) is itself.

5.3 Convex Functions

f is a (strongly) convex function iff for $0 < \alpha < 1$,

$$f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y})(<) \leq \alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}) \tag{27}$$

The sum of convex functions is a convex function; the max of convex functions is a convex function

The Composed function $f(\phi(\mathbf{x}))$ is convex if $\phi(\mathbf{x})$ is a convex and $f(\cdot)$ is convex&non-decreasing. The (lower) level set of f is convex:

$$L(z) = \{\mathbf{x} : \quad f(\mathbf{x}) \leq z\} \tag{28}$$

Convex set $\{(z; \mathbf{x}) : f(\mathbf{x}) \leq z\}$ is called the epigraph of f. $tf(\mathbf{x}/t)$ is a convex function of $(t; \mathbf{x})$ for $t > 0$ if $f(\cdot)$ is a convex function; it's homogeneous with degree 1 Note that the difference between supreme and maximization, the maximal solution is achievable and supreme is not.

5.4 Convex Function Examples

$\|\mathbf{x}\|_p$ for $p \geq 1$

$$\|\alpha \mathbf{x} + (1-\alpha)\mathbf{y}\|_p \leq \|\alpha \mathbf{x}\|_p + \|(1-\alpha)\mathbf{y}\|_p \leq \alpha \|\mathbf{x}\|_p + (1-\alpha)\|\mathbf{y}\|_p, \tag{29}$$

from the triangle inequality.

Logistic function $\log\left(1 + e^{\mathbf{a}^T}\mathbf{x} + b\right)$ is convex.

Consider the minimal-objective function of b for fixed A and c :

$$\begin{aligned} z(\mathbf{b}) := \ & \text{minimize } f(\mathbf{x}) \\ & \text{subject to } A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned} \tag{30}$$

where $f(\mathbf{x})$ is a convex function. Show that $z(\mathbf{b})$ is a convex function in b.

Proof. There are two separated cases

$$\begin{aligned} z(\mathbf{b_1}) := \ & \text{minimize } f(\mathbf{x}_1) & \qquad z(\mathbf{b}_2) := \ & \text{minimize } f(\mathbf{x}_2) \\ & \text{subject to } A\mathbf{x}_1 = \mathbf{b}_1 & & \text{subject to } A\mathbf{x}_2 = \mathbf{b}_2 \\ & \mathbf{x}_1 \geq \mathbf{0} & & \mathbf{x}_2 \geq \mathbf{0} \end{aligned} \tag{31}$$

we need to show

$$
\begin{aligned}
b &= \alpha b_1 + (1-\alpha) b_2 \\
z(\alpha b_1 + (1-\alpha) b_2) &\le \alpha z(\mathbf{b_1}) + (1-\alpha) z(\mathbf{b_2}) \quad 0 \le \alpha \le 1 \\
\alpha x_1 + (1-\alpha) x_2 &\ge 0
\end{aligned}
$$

$$
\begin{aligned}
A(\alpha x_1 + (1-\alpha) x_2) &= \alpha A x_1 + (1-\alpha) A x_2 \\
&= \alpha b_1 + (1-\alpha) b_2
\end{aligned}
$$

Since function f is the convex function. we have done $\square$

6 Theorems on Functions

Taylor's theorem or the mean-value theorem:

Theorem 5. *Let $f \in C^1$ be in a region containing the line segment $[\mathbf{x}, \mathbf{y}]$. Then there is a $\alpha, 0 \le \alpha \le 1$, such that*

$$f(\mathbf{y}) = f(\mathbf{x}) + \nabla f(\alpha \mathbf{x} + (1-\alpha)\mathbf{y})(\mathbf{y} - \mathbf{x}) \tag{32}$$

Furthermore, if $f \in C^2$ then there is a $\alpha, 0 \le \alpha \le 1$, such that

$$f(\mathbf{y}) = f(\mathbf{x}) | \nabla f(\mathbf{x})(\mathbf{y} \quad \mathbf{x}) | (1/2)(\mathbf{y} \quad \mathbf{x})^T \nabla^2 f \left(\alpha \mathbf{x} \mid \left(1 \quad \alpha \right) \mathbf{y} \right) (\mathbf{y} \quad \mathbf{x}) \tag{33}$$

Theorem 6. *Let $f \in C^1$. Then f is convex over a convex set Ω if and only if*

$$f(\mathbf{y}) \ge f(\mathbf{x}) + \nabla f(\mathbf{x})(\mathbf{y} - \mathbf{x}) \tag{34}$$

for all $\mathbf{x}, \mathrm{y} \in \Omega$

Theorem 7. *Let $f \in C^2$. Then f is convex over a convex set Ω if and only if the Hessian matrix of f is positive semi-definite throughout Ω.*

6.1 Lipschitz Functions

The first-order β-Lipschitz function: there is a positive number β such that for any two points $\mathbf{x}$ and y :

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \le \beta \|\mathbf{x} - \mathbf{y}\| \tag{35}$$

This condition imples

$$\left| f(\mathbf{x}) - f(\mathbf{y}) - \nabla f(\mathbf{y})^T (\mathbf{x} - \mathbf{y}) \right| \le \frac{\beta}{2} \|\mathbf{x} - \mathbf{y}\|^2 \tag{36}$$

The second-order β-Lipschitz function: there is a positive number β such that for any two points $\mathbf{x}$ and $\mathbf{y}$

$$\left\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}) - \nabla^2 f(\mathbf{y})(\mathbf{x}-\mathbf{y})\right\| \leq \beta \|\mathbf{x}-\mathbf{y}\|^2 \tag{37}$$

This condition implies

$$\left| f(\mathbf{x}) - f(\mathbf{y}) - \nabla f(\mathbf{y})^T(\mathbf{x}-\mathbf{y}) - \frac{1}{2}(\mathbf{x}-\mathbf{y})^T \nabla^2 f(\mathbf{y})(\mathbf{x}-\mathbf{y}) \right| \leq \frac{\beta}{3}\|\mathbf{x}-\mathbf{y}\|^3 \tag{38}$$

7 Known Inequalities

- Cauchy-Schwarz: given $\mathbf{x}, \mathbf{y} \in \mathcal{R}^n$, $\left|\mathbf{x}^T\mathbf{y}\right| \leq \|\mathbf{x}\|_p \|\mathbf{y}\|_q$, where $\frac{1}{p} + \frac{1}{q} = 1$ and $p \geq 1$
- Triangle: given $\mathbf{x}, \mathbf{y} \in \mathcal{R}^n$, $\|\mathbf{x}+\mathbf{y}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{y}\|_p$ for $p \geq 1$
- Arithmetic-geometric mean: given $\mathbf{x} \in \mathcal{R}^n_+$,

$$\frac{\sum x_j}{n} \geq \left(\prod x_j\right)^{1/n}$$

8 Direct Solution

Given $A \in \mathcal{R}^{m \times n}$ and $\mathbf{c} \in \mathcal{R}^n$

$$\begin{array}{ll} \min & \left\|A^T\mathbf{y} - \mathbf{c}\right\|^2 \\ s.t. & \mathbf{y} \in \mathcal{R}^m \end{array} \tag{39}$$

Choleski Decomposition

$$AA^T = L\Lambda L^T, \text{ and then solve } L\Lambda L^T\mathbf{y} = A\mathbf{c}.$$

Projections Matrices: $A^T\left(AA^T\right)^{-1}A$ and $I - A^T\left(AA^T\right)^{-1}A$

Mathematical Optimization Models and Applications	June 28,2022
Lecture 4	
Lecturer: Yinyu Ye	*Scriber: Bin Yu*

9 Basic Feasible Solution and Farkas Lemma

9.1 Caratheodory's theorem

Theorem 8. *Given matrix $A \in R^{m\times n}$, let convex polyhedral cone $C = \{A\mathbf{x} : \mathbf{x} \geq 0\}$. For any $\mathbf{b} \in C$*

$$\mathbf{b} = \sum_{i=1}^{d} \mathbf{a}_{j_i} x_{j_i}, x_{j_i} \geq 0, \forall i \tag{40}$$

for some linearly independent vectors $\mathbf{a}_{j_1}, \ldots, \mathbf{a}_{j_d}$ chosen from $\mathbf{a}_1, \ldots, \mathbf{a}_n$. There is a construct proof of the theorem (page 26 of the text).

9.2 Basic Feasible Solution

Now consider the feasible set $\{\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ for given data $A \in R^{m\times n}$ and $\mathbf{b} \in R^m$. Select m linearly independent columns, denoted by the variable index set B, from A. Solve $A_B\mathbf{x}_B = \mathbf{b}$ for the m-dimension vector $\mathbf{x}_B$, and set the remaining variables, $\mathbf{x}_N$, to zero. Then, we obtain a solution $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$, that is called a basic solution to with respect to the basis A_B. If a basic solution $\mathbf{x}_B \geq \mathbf{0}$, then x is called a basic feasible solution, or BFS.

Note that the the optimal solution is the extreme point. We show the proof of this argument below

Proof.

$$\begin{aligned} \min \quad & c^T x \\ s.t. \quad & x \in X \end{aligned} \tag{41}$$

Let $\dot{x} \in$ int X. Thus we have $\{x : ||x - \dot{x}||_2 \leq r\} \in X$. Imply

$$\begin{aligned}
x &= \dot{x} - r\frac{c}{||c||_2} \\
c^T x &= c^T(\dot{x} - r\frac{c}{||c||_2}) \\
&= c^T\dot{x} - r||c|| \\
&< c^T\dot{x}
\end{aligned}$$

□

Theorem 9. *(Separating hyperplane theorem) Let C be a closed convex set in $\mathcal{R}^m$ and let b be a point exterior to C. Then there is a vector* $\mathrm{y} \in \mathcal{R}^m$ *such that*

$$\mathbf{b} \cdot \mathbf{y} > \sup_{\mathbf{x} \in C} \mathbf{x} \cdot \mathbf{y}. \tag{42}$$

Theorem 10. *(Supporting hyperplane theorem) Let C be a closed convex set and let* b *be a point on the boundary of C. Then there is a vector $y \in \mathcal{R}^m$ such that*

$$\mathbf{b} \cdot \mathbf{y} = \sup_{\mathbf{x} \in C} \mathbf{x} \cdot \mathbf{y}. \tag{43}$$

9.3 Farkas Lemma

Theorem 11. *Let $A \in \mathcal{R}^{m \times n}$ and* $\mathrm{b} \in \mathcal{R}^m$. *Then, the system* $\{\mathrm{x} : A\mathrm{x} = \mathrm{b}, \mathrm{x} \geq 0\}$ *has a feasible solution* x *if and only if that its alternative system $-A^T\mathbf{y} \geq 0$ and $\mathbf{b}^T\mathbf{y} > 0$ has no feasible solution $\mathbf{y}$.*

Geometrically, Farkas' lemma means that if a vector $b \in \mathcal{R}^m$ does not belong to the convex cone generated by $\mathbf{a}_{.1}, \ldots, \mathbf{a}_{.n}$, then there is a hyperplane separating $\mathbf{b}$ from cone $(\mathbf{a}_{.1}, \ldots, \mathbf{a}_{.n})$.

Proof. Let $\{\mathbf{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ have a feasible solution, say $\overline{\mathbf{x}}$. Then, $\left\{\mathbf{y} : A^T\mathbf{y} \leq \mathbf{0}, \mathbf{b}^T\mathbf{y} > 0\right\}$ is infeasible, since otherwise,

$$0 < \mathbf{b}^T\mathbf{y} = (A\mathbf{x})^T\mathbf{y} = \mathbf{x}^T\left(A^T\mathbf{y}\right) \leq 0$$

from $\mathbf{x} \geq \mathbf{0}$ and $A^T\mathbf{y} \leq \mathbf{0}$ Now let $\{\mathrm{x} : A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}$ have no feasible solution, or $\mathbf{b} \notin C := \{A\mathbf{x} : \mathbf{x} \geq 0\}$. We now prove that its alternative system has a solution. We first prove

Lemma 12. $C = \{A\mathbf{x} : \mathbf{x} \geq \mathbf{0}\}$ *is a closed convex set.*

That is, any convergent sequence $\mathbf{b}^k \in C, k = 1.2\ldots$ has its limit point $\overline{\mathrm{b}}$ also in C. Let $\mathbf{b}^k = A\mathbf{x}^k, \mathbf{x}^k \geq \mathbf{0}$. Then by Carathéodory's theorem, we must have $\mathbf{b}^k = A_{B^k}\mathbf{x}_{B^k}, \mathbf{x}_{B^k} \geq \mathbf{0}$ where A_{B^k} is a basis of A. Therefore, $\mathbf{x}_{B^k}$, together with zero values for the nonbasic variables, is bounded for all k, so that it has sub-sequence, say indexed by $l = 1, \ldots$, where $\mathbf{x}^l = \mathbf{x}_{B^l}$ has a limit point $\overline{\mathbf{x}}$ and $\overline{\mathbf{x}} \geq 0$. Consider this very sub-sequence $\mathbf{b}^l = A\mathbf{x}^l$ we must also have $\mathbf{b}^l \to \overline{\mathbf{b}}$. Then from

$$\|\overline{\mathbf{b}} - A\overline{\mathbf{x}}\| = \left\|\overline{\mathbf{b}} - \mathbf{b}^l + A\mathbf{x}^l - A\overline{\mathbf{x}}\right\| \leq \left\|\overline{\mathbf{b}} - \mathbf{b}^l\right\| + \left\|A\mathbf{x}^l - A\overline{\mathbf{x}}\right\| \leq \left\|\overline{\mathbf{b}} - \mathbf{b}^l\right\| + \|A\| \left\|\mathbf{x}^l - \overline{\mathbf{x}}\right\|$$

we must have $\overline{\mathrm{b}} = A\overline{\mathrm{x}}$, that is, $\overline{\mathrm{b}} \in C$; since otherwise the right-hand side of the above inequality is strictly greater than zero which is a contradiction. Now since C is a closed convex set, by the separating hyperplane theorem, there is y such that

$$\mathbf{y} \cdot \mathbf{b} > \sup_{\mathbf{c} \in C} \mathbf{y} \bullet \mathbf{c}$$

or

$$\mathbf{y} \bullet \mathbf{b} > \sup_{\mathbf{x} \geq \mathbf{0}} \mathbf{y} \bullet (A\mathbf{x}) = \sup_{\mathbf{x} \geq \mathbf{0}} A^T\mathbf{y} \bullet \mathbf{x} \tag{44}$$

From $0 \in C$ we have $\mathrm{y} \bullet > 0$. Furthermore, $A^T\mathbf{y} \leq \mathbf{0}$. Since otherwise, say $\left(A^T\mathbf{y}\right)_1 > 0$, one can have a vector $\overline{\mathbf{x}} \geq 0$ such that $\bar{x}_1 = \alpha > 0, \bar{x}_2 = \ldots = \bar{x}_n = 0$, from which

$$\sup_{\mathbf{x} \geq 0} A^T\mathbf{y} \bullet \mathbf{x} \geq A^T\mathbf{y} \bullet \overline{\mathbf{x}} = \left(A^T\mathbf{y}\right)_1 \cdot \alpha$$

and it tends to ∞ as $\alpha \to \infty$. This is a contradiction because $\sup_{\mathbf{x} \geq 0} A^T\mathbf{y} \bullet \mathbf{x}$ is bounded from above by (5). $\square$

9.3.1 Farkas Lemma Variant

Theorem 13. *Let $A \in \mathcal{R}^{m \times n}$ and $\mathrm{c} \in \mathcal{R}^n$. Then, the system $\left\{\mathbf{y} : \mathbf{c} - A^T\mathbf{y} \geq 0\right\}$ has a solution $\mathbf{y}$ if and only if that $A\mathrm{x} = 0, \mathrm{x} \geq 0$, and $\mathrm{c}^T\mathrm{x} < 0$ has no feasible solution* x.

Consider the pair:

$$\{\mathbf{x} : A\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \in K\}$$

and

$$\left\{\mathbf{y} : -A^T\mathbf{y} \in K^*, \quad \mathbf{b}^T\mathbf{y} > 0\right\}.$$

Or in operator form: given data vector or matrix $\mathbf{a}_i, i = 1, \ldots, m$, and $\mathbf{b} \in \mathcal{R}^m$, an "alternative" system pair would be

$$\mathcal{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \in K$$

and

$$-\mathcal{A}^T\mathbf{y} \in K^*, \quad \mathbf{b}^T\mathbf{y} = 1(> 0)$$

where

$$\mathcal{A}\mathbf{x} = (\mathbf{a}_1 \bullet \mathbf{x}; \ldots; \mathbf{a}_m \bullet \mathbf{x}) \in \mathcal{R}^m \text{ and } \mathcal{A}^T\mathbf{y} = \sum_i^m y_i \mathbf{a}_i$$

Let K be a closed and convexq 1 cone in the rest of the course. If there is y such that $-\mathcal{A}^T\mathbf{y} \in \operatorname{int} K^*$, then $C := \{\mathcal{A}\mathbf{x} : \mathbf{x} \in K\}$ is a closed convex cone. Consequently,

$$\mathcal{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \in K$$

and

$$-\mathcal{A}^T\mathbf{y} \in K^*, \quad \mathbf{b}^T\mathbf{y} = 1(> 0)$$

are an alternative system pair. And if there is $\mathbf{x}$ such that $\mathcal{A}^T\mathbf{x} = \mathbf{0}, \mathbf{x} \in \operatorname{int} K$, then

$$\mathcal{A}\mathbf{x} = \mathbf{0}, \quad \mathbf{x} \in K, \quad \mathbf{c} \bullet \mathbf{x} = -1(< 0)$$

and

$$\mathbf{c} - \mathcal{A}^T\mathbf{y} \in K^*$$

are an alternative system pair.

10 Conic Linear Program

$$\begin{aligned}
(\text{C L P}) \min \quad & \mathbf{c} \bullet \mathbf{x} \\
s.t. \quad & \mathbf{a}_i \bullet \mathbf{x} = b_i, i = 1, 2, \ldots, m, \mathbf{x} \in K, \\
& \left(\mathcal{A}^T\mathbf{x} = \mathbf{b}\right)
\end{aligned} \tag{45}$$

where K is a closed and pointed convex cone. Linear Programming (LP): $\mathbf{c}, \mathbf{a}_i, \mathbf{x} \in \mathcal{R}^n$ and $K = \mathcal{R}^n_+$ Second-Order Cone Programming (SOCP): $\mathbf{c}, \mathbf{a}_i, \mathbf{x} \in \mathcal{R}^n$ and $K = SOC = \{\mathbf{x} : x_1 \geq \|\mathbf{x}_{-1}\|_2\}$. Semidefinite Programming (SDP): $\mathbf{c}, \mathbf{a}_i, \mathbf{x} \in \mathcal{S}^n$ and $K = \mathcal{S}^n_+$ p-Order Cone Programming (POCP): $\mathbf{c}, \mathbf{a}_i, \mathbf{x} \in \mathcal{R}^n$ and $K = POC = \left\{\mathbf{x} : x_1 \geq \|\mathbf{x}_{-1}\|_p\right\}$. Here, $\mathbf{x}_{-1}$ is the vector $(x_2; \ldots; x_n) \in R^{n-1}$. Cone K can be also a product of different cones, that is, $\mathbf{x} = (\mathbf{x}_1; \mathbf{x}_2; \ldots)$ where $\mathbf{x}_1 \in K_1, \mathbf{x}_2 \in K_2, \ldots$ and so on with linear constraints:

$$\mathcal{A}_1\mathbf{x}_1 + \mathcal{A}_2\mathbf{x}_2 + \ldots = \mathbf{b}$$

10.1 Dual of Conic Linear Program

The dual problem to

$$\begin{array}{rll} (CLP) & \min & \mathbf{c} \bullet \mathbf{x} \\ & s.t. & \mathbf{a}_i \bullet \mathbf{x} = b_i, i = 1, 2, \ldots, m, \mathbf{x} \in K \end{array} \tag{46}$$

is

$$\begin{array}{rll} (CLD) & \min & \mathbf{b}^T \mathbf{y} \\ & s.t. & \sum_i^m y_i \mathbf{a}_i + \mathbf{s} = \mathbf{c}, \mathbf{s} \in K^*, \end{array} \tag{47}$$

where $y \in \mathcal{R}^m$, s is called the dual slack vector/matrix, and K^* is the dual cone of K. The former is called the primal problem, and the latter is called dual problem.

Mathematical Optimization Models and Applications	July 3,2022
Lecture 5	
Lecturer: Yinyu Ye	*Scriber: Bin Yu*

11 Duality of Conic Linear Program

Recall the primal and dual program

$$\begin{array}{rlrl} \min & C^T x & \max & b^T y \\ s.t. & Ax = b & s.t. & A^y + s = c \\ & x \in K & & s \in K' \end{array} \tag{48}$$

Theorem 14. *(Weak duality theorem)* $\mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y} = \mathbf{x}^T\mathrm{s} \geq 0$ *for any feasible* $\mathbf{x}$ *of* (CLP) *and* $(\mathbf{y}, \mathrm{s})$ *of* (CLD)

Corollary 15. *Let* $\mathrm{x}^* \in \mathcal{F}_p$ *and* $(\mathbf{y}^*,\ \mathrm{s}^*) \in \mathcal{F}_d$*. Then,* $\mathbf{c} \bullet \mathbf{x}^* = \mathbf{b}^T\mathbf{y}^*$ *implies that* x^* *is optimal for* (CLP) *and* $(\mathbf{y}^*,\ \mathrm{s}^*)$ *is optimal for* (CLD)

It is called the strong duality theorem, but it does not work in general. Here, operator $\mathcal{A}\mathbf{x}$ and Adjoint-Operator $\mathcal{A}^T\mathbf{y}$ minimic matrix-vector production $A\mathbf{x}$ and its transpose operation $A^T\mathbf{y}$, where

$$\mathcal{A} = (\mathbf{a}_1; \mathbf{a}_2; \ldots; \mathbf{a}_m), \quad \mathcal{A}\mathbf{x} = (\mathbf{a}_1 \bullet \mathbf{x}; \ldots; \mathbf{a}_m \bullet \mathbf{x}), \quad \text{and} \quad A^T\mathbf{y} = \sum_i y_i \mathbf{a}_i^T \tag{49}$$

Theorem 16. *The following statements hold for every pair of* (LP) *and (LD) :*

- *If* (LP) *and (LD) are both feasible, then both problems have optimal solutions and the optimal objective values of the objective functions are equal, that is, optimal solutions for both (LP) and (LD) exist and there is no duality gap*
- *If* (LP) *or* (LD) *is feasible and bounded, then the other is feasible and bounded*
- *If* (LP) *or* (LD) *is feasible and unbounded, then the other has no feasible solution*
- *If* (LP) *or* (LD) *is infeasible, then the other is either unbounded or has no feasible solution*

11.1 Farkas Lemma and Duality

The Farkas lemma concerns the system the system $\{\mathrm{x} : A\mathrm{x} = \mathbf{b}, \mathrm{x} \geq 0\}$ and its alternative $\left\{\mathbf{y} : -A^T\mathbf{y} \geq \mathbf{0}, \mathbf{b}^T\mathbf{y} > 0\right\}$ for given data $(A, \mathbf{b})$. This pair can be represented as a primal-dual LP pair

$$\begin{array}{llll} \min & \mathbf{0}^T\mathbf{x} & \max & \mathbf{b}^T\mathbf{y} \\ \text{s. t.} & A\mathbf{x} = \mathbf{b}, & \text{s.t.} & A^T\mathbf{y} \leq \mathbf{0} \\ & \mathbf{x} \geq \mathbf{0}; & & \end{array} \tag{50}$$

If the primal is infeasible, then the dual must be feasible and unbounded since it is always feasible.

11.2 Optimality Conditions for LP

$$\begin{cases} & C^T x - b^T y = 0 \\ (\mathbf{x}, \mathrm{y}, \mathrm{s}) \in (\mathcal{R}_+^n, \mathcal{R}^m, \mathcal{R}_+^n) & Ax = b \\ & A^T y + s = c \end{cases} \tag{51}$$

which is a system of linear inequalities and equations. Now it is easy to verify whether or not a pair $(\mathbf{x}, \mathrm{y}, \mathrm{s})$ is optimal.

11.3 Complementarity Condition

For feasible $\mathbf{x}$ and $(\mathbf{y}, \mathbf{s}), \mathbf{x}^T\mathbf{s} = \mathbf{x}^T\left(\mathbf{c} - A^T\mathbf{y}\right) = \mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y}$ is called the complementarity gap. If $\mathbf{x}^T\mathbf{S} = 0$, then we say x and s are complementary to each other. Since both $\mathbf{x}$ and $\mathbf{s}$ are nonnegative, $\mathbf{x}^T\mathbf{s} = 0$ implies that $\mathbf{x}. * \mathbf{s} = 0$ or $x_j s_j = 0$ for all $j = 1, \ldots, n$.

$$\begin{aligned} \mathrm{x} \cdot * &= 0 \\ A\mathrm{x} &= \mathbf{b} \\ -A^T\mathbf{y} - \mathrm{s} &= -\mathbf{c}. \end{aligned} \tag{52}$$

This system has total $2n + m$ unknowns and $2n + m$ equations including n nonlinear equations. Interpretation of $s_j = 0$: the j th inequality constraint of the dual is "binding" or "active".

11.4 Duality of Conic Program

The strong duality theorem may not hold for general convex cones:

$$\mathbf{c} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \mathbf{a}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \mathbf{a}_2 = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

and

$$\mathbf{b} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

The problem is

$$\begin{aligned} \min \quad & x_1 + x_2 \\ s.t. \quad & \begin{bmatrix} x_2 \\ x_1 + x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \end{bmatrix} \\ & \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix} \succeq 0 \end{aligned} \qquad \begin{aligned} \max \quad & y_2 \\ s.t. \quad & \begin{bmatrix} 0 & -y_2 & 0 \\ -y_2 & y_1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + s = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ & s \succeq 0 \end{aligned}$$

Theorem 17. *The following statements hold for every pair of (CLP) and (CLD):*

- *If (CLP) and (CLD) both are feasible, and furthermore one of them have an interior, then there is no duality gap between (CLP) and (CLD). However, one of the optimal solution may not be attainable.*
- *If (CLP) and (CLD) both are feasible and have interior, then, then both have attainable optimal solutions with no duality gap.*
- *If (CLP) or (CLD) is feasible and unbounded, then the other has no feasible solution.*
- *If (CLP) or (CLD) is infeasible, and furthermore the other is feasible and has an interior, then the other is unbounded.*

11.4.1 Construct the Dual Cone

Consider set

$$\{(\tau, x) : \tau > 0, \tau c_i(x/\tau) \leq 0\}$$

The dual cone is the set of all points $(\kappa; s)$ such that

$$\kappa\tau + s^T x \geq 0, \quad \forall(\tau; x) \quad s.t. \tau > 0, \tau c_i(x/\tau) \leq 0, i = 1, ..., m$$

Without loss of generality, we can set $\tau = 1$ and the condition becomes

$$\kappa + s^T x \geq 0, \quad \forall x \ s.t. c_i(x) \leq 0, i = 1, ..., m$$

Then, consider the optimization problem

$$\begin{aligned} \psi(s) := inf \quad & s^T x \\ s.t. \quad & c_i(x) \leq 0, i = 1, 2, ..., m \end{aligned}$$

Then, the dual cone coan be represented as

$$K^\star = \{(\kappa, s) : \kappa + \psi(x) \geq 0\}$$

12 Combinationrial Auction Pricing

Given the m different states that are mutually exclusive and exactly one of them will be true at the maturity. A contract on a state is a paper agreement so that on maturity it is worth a notional \$1 if it is on the winning state and worth \$0 if is not on the winning state. There are n orders betting on one or a combination of states, with a price limit and a quantity limit

Order:	#1	#2	#3	#4	#5
Argentina	1	0	1	1	0
Brazil	1	0	0	1	1
Italy	1	0	1	1	0
Germany	0	1	0	1	1
France	0	0	1	0	0
Bidding Prize: π	0.75	0.35	0.4	0.95	0.75
Quantity limit:q	10	5	10	10	5
Order fill:x	x_1	x_2	x_3	x_4	x_5

Lethal x_j be the number of contacts awarded to the jth order. Then jth better will pay the amount

$$\pi_j \times x_j$$

and the total collected amount is $\sum_{i=1}^{n} \pi_j \times x_j = \pi^T x$. If the ith state is the winning state, then the auction organizer need to pay.

$$\sum_{j=1}^{n} a_{ij} x_j$$

We can formulate the primal and dual problem

$$
\begin{array}{llll}
\max & \pi^T \mathbf{x} - z & \min & \mathbf{q}^T \mathbf{y} \\
\text{s.t.} & A\mathbf{x} - \mathbf{e} \cdot z \leq \mathbf{0} & \text{s.t.} & A^T \mathbf{p} + \mathbf{y} \geq \pi, \\
 & \mathbf{x} \leq \mathbf{q} & & \mathbf{e}^T \mathbf{p} = 1, \\
 & \mathbf{x} \geq 0 & & (\mathbf{p}, \mathbf{y}) \geq 0.
\end{array}
\tag{53}
$$

12.1 Online Linear Programming

The main idea of linear program is we don't know the coefficient matrix. We need to make the decision and reveal the information sequentially.

$$
\begin{aligned}
\max \quad & \sum_{t=1}^{n} \pi_t x_t \\
s.t. \quad & \sum_{t=1}^{n} a_{it} x_t \leq b_i, \quad \forall i = 1, \ldots, m \\
& 0 \leq x_t \leq 1, \quad \forall t = 1, \ldots, n
\end{aligned}
\tag{54}
$$

Each bid/activity t requests a bundle of m resources, and the payment is π_t. Online Decision Making: we only know $(n, \mathbf{b})$ at the start, but - the (bounded) order-data of each variable x_t is revealed sequentially. - an irrevocable decision must be made as soon as an order arrives without observing or knowing the future data.

The algorithm/mechanism quality is evaluated on the expected performance over all the permutations comparing to the offline optimal solution, i.e., an algorithm $\mathcal{A}$ is c-competitive if and only if

$$
E_\sigma \left[\sum_{t=1}^{n} \pi_t x_t(\sigma, \mathcal{A}) \right] \geq c \cdot OPT(A, \pi), \forall (A, \pi). \tag{55}
$$

Then we will introduce the algorithm how to solve the online linear program.

1. Set $x_t = 0$ for all $1 \leq t \leq \epsilon n$

2. Solve the ϵ portion of the problem

$$
\begin{array}{lll}
\text{maximize}_{\mathbf{x}} & \sum_{t=1}^{\epsilon n} \pi_t x_t & \\
\text{subject to} & \sum_{t=1}^{\epsilon n} a_{it} x_t \leq \epsilon b_i \quad i = 1, \ldots, m & \\
 & 0 \leq x_t \leq 1 & t = 1, \ldots, \epsilon n
\end{array}
$$

   and get the optimal dual solution $\hat{\mathbf{p}}$ of the sample LP;

3. Determine the future allocation x_t as:

$$x_t = \begin{cases} 0 & \text{if } \pi_t \leq \hat{\mathbf{p}}^T \mathbf{a}_t \\ 1 & \text{if } \pi_t > \hat{\mathbf{p}}^T \mathbf{a}_t \end{cases}$$

as long as $a_{it}x_t \leq b_i - \sum_{j=1}^{t-1} a_{ij}x_j$ for all i; otherwise, set $x_t = 0$. Online Learning: Periodically resolve the sample LP with all arrived orders and update the ”ideal” prices...

Mathematical Optimization Models and Applications	July 8,2022
Lecture 6&7	
Lecturer: Yinyu Ye	*Scriber: Bin Yu*

13 Optimality Conditions

Recall the primal and dual program

$$\left\{ (\mathbf{x},\mathbf{y},\mathbf{s}) \in \left(\mathcal{R}^n_+, \mathcal{R}^m, \mathcal{R}^n_+\right) : \begin{array}{rl} \mathbf{c}^T\mathbf{x} - \mathbf{b}^T\mathbf{y} & = \mathbf{0} \\ A\mathbf{x} & = \mathbf{b} \\ -A^T\mathbf{y} - \mathbf{s} & = -\mathbf{c} \end{array} \right\} \tag{56}$$

Let $x^\star$ and $s^\star$ be optimal solutions with zero duality gap

$$|support(x^\star)| + |support(s^\star)| \leq n \tag{57}$$

Note that support size means the number of non-zero components. Then we will have

Theorem 18. *If (LP) and (LD) are both feasible, then there exists a pair of strictly complementary solutions $\mathbf{x}^* \in \mathcal{F}_p$ and $(\mathbf{y}^*, \mathbf{s}^*) \in \mathcal{F}_d$ such that*

$$\mathbf{x}^* \cdot \mathbf{s}^* = \mathbf{0} \text{ and } |\operatorname{supp}(\mathbf{x}^*)| + |\operatorname{supp}(\mathbf{s}^*)| = n. \tag{58}$$

Moreover, the supports

$$P^* = \{j : x^*_j > 0\} \quad \text{and} \quad Z^* = \{j : s^*_j > 0\}$$

are invariant for all strictly complementary solution pairs.

13.1 Uniqueness Theorem for Linear Program

If we have the optimal solution $x^\star$ and how to clarify the uniqueness of $x^\star$

Theorem 19. *An LP optimal solution $\mathbf{x}^*$ is unique if and only if the size of supp $(\mathbf{x}^*)$ is maximal among all optimal solutions and the columns of $A_{Supp}(x^*)$ are linear independent.*

Proof. It is easy to see both conditions are necessary, since otherwise, one can find an optimal solution with a different support size. To see sufficiency, suppose there there is another optimal solution $\mathbf{y}^*$ such that $\mathbf{x}^* - \mathbf{y}^* \neq \mathbf{0}$. We must have $\operatorname{supp}(\mathbf{y}^*) \subset \operatorname{supp}(\mathbf{x}^*)$, since, otherwise, $(0.5\mathbf{x}^* + 0.5\mathbf{y}^*)$ remains optimal and its support size is greater than that of x^* which is a contradiction. Then we see

$$\mathbf{0} = A\mathbf{x}^* - A\mathbf{y}^* = A(\mathbf{x}^* - \mathbf{y}^*) = A_{\operatorname{supp}(\mathbf{x}^*)}(\mathbf{x}^* - \mathbf{y}^*)\operatorname{supp}(\mathbf{x}^*)$$

which implies that columns of $A_{\operatorname{Supp}(\mathbf{x}^*)}$ are linearly dependent. Think the $x^\star = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ and $y^\star = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ □

13.2 Uniqueness Theorem for Semidefinite Program

Theorem 20. *An SDP optimal and complementary solution X^* is unique if and only if the rank of X^* is maximal among all optimal solutions and $V^* A_i (V^*)^T, i = 1, \ldots, m$, are linearly independent, where $X^* = (V^*)^T V^*, V^* \in \mathcal{R}^{r \times n}$, and r is the rank of X^*.*

14 Relaxation Example

14.1 Sensor Localization Problem

Given $\mathbf{a}_k \in \mathbf{R}^d, d_{ij} \in N_x$, and $\hat{d}_{kj} \in N_a$, find $\mathbf{x}_i \in \mathbf{R}^d$ such that

$$\begin{aligned} \|\mathbf{x}_i - \mathbf{x}_j\|^2 &= d_{ij}^2, \forall (i,j) \in N_x, i < j, \\ \|\mathbf{a}_k - \mathbf{x}_j\|^2 &= \hat{d}_{kj}^2, \forall (k,j) \in N_a, \end{aligned}$$

We can transform to the matrix form and SDP relaxation

$$\begin{aligned} &(\mathbf{e}_i - \mathbf{e}_j)^T Y (\mathbf{e}_i - \mathbf{e}_j) = d_{ij}^2, \forall i,j \in N_x, i < j \\ &(\mathbf{a}_k; -\mathbf{e}_j)^T \begin{pmatrix} I & X \\ X^T & Y \end{pmatrix} (\mathbf{a}_k; -\mathbf{e}_j) = \hat{d}_{kj}^2, \forall k,j \in N_a \\ &\begin{pmatrix} I & X \\ X^T & Y \end{pmatrix} \succeq \mathbf{0} \end{aligned} \tag{59}$$

Theorem 21. *Let $\bar{Z}$ be a feasible solution for SDP and $\bar{U}$ be an optimal slack matrix of the dual. Then,*

1. *complementarity condition holds: $\bar{Z} \bullet \bar{U} = 0$ or $\bar{Z}\bar{U} = 0$*
2. $\mathrm{Rank}(\bar{Z}) + \mathrm{Rank}(\bar{U}) \leq 2 + n$
3. $\mathrm{Rank}(\bar{Z}) \geq 2$ *and* $\mathrm{Rank}(\bar{U}) \leq n$

15 Rank-Reduction for SDP

In the most SDP case, it is difficult to find the rank-minimal SDP solution

$$\begin{array}{lll} (SDP) & \min & C \bullet X \\ & \text{subject to} & A_i \bullet X = b_i, i = 1, 2, \ldots, m, X \succeq 0 \end{array} \tag{60}$$

where $C, A_i \in S^n$

Theorem 22. *(Carathéodory's theorem)*

- *If there is a minimizer for (LP), then there is a minimizer of (LP) whose support size r satisfying $r \leq m$*
- *If there is a minimizer for (SDP), then there is a minimizer of (SDP) whose rank r satisfying $\frac{r(r+1)}{2} \leq m$. Moreover, such a solution can be find in polynomial time.*

Then if we simply the SDP feasiblility problem

$$A_i \bullet X = b_i \quad i = 1, .., m, \quad X \succeq 0$$

we try to find an approximate $\hat{X} \succeq 0$ of rank at most d

$$\beta(m, n, d) \cdot b_i \leq A_i \bullet \hat{X} \leq \alpha(m, n, d) \cdot b_i \quad \forall i = 1, \ldots, m \tag{61}$$

Here, $\alpha \geq 1$ and $\beta \in (0, 1]$ are called the distortion factors. Clearly, the closer are both to 1 , the better.

Theorem 23. *Let $r = \max\{\mathrm{rank}\,(A_i)\}$ and $\bar{X}$ be a feasible solution. Then, for any $d \geq 1$, the randomly generated*

$$\hat{X} = \sum_i^d \xi_i \xi_i^T, \quad \xi_i \in N\left(\mathbf{0}, \frac{1}{d}\bar{X}\right)$$

$$\alpha(m, n, d) = \begin{cases} 1 + \frac{12\ln(4mr)}{d} & \textit{for } 1 \leq d \leq 12\ln(4mr) \\ 1 + \sqrt{\frac{12\ln(4mr)}{d}} & \textit{for } d > 12\ln(4mr) \end{cases}$$

and

$$\beta(m,n,d) = \begin{cases} \frac{1}{e(2m)^{2/d}} & \textit{for } 1 \le d \le 4\ln(2m) \\ \max\left\{\frac{1}{e(2m)^{2/d}}, 1-\sqrt{\frac{4\ln(2m)}{d}}\right\} & \textit{for } d > 4\ln(2m) \end{cases}$$

Here is the some remarks from theorem 9

- There is always a low-rank, or sparse, approximate SDP solution with respect to a bounded relative residual distortion. As the allowable rank increases, the distortion bounds become smaller and smaller.
- The lower distortion factor is independent of n and the rank of A_i s.
- The factors can be improved if we only consider one-sided inequalities.
- This result contains as special cases several well-known results in the literature.
- Can the distortion upper bound be improved such that it's independent of rank of A_i ?
- Is there deterministic rank-reduction procedure? Choose the largest d eigenvalue component of X ?
- General symmetric A_i ?
- In practical applications, we see much smaller distortion, why?

16 Max-Cut Problem

This is the Max-Cut problem on an undirected graph $G = (V, E)$ with non-negative weights w_{ij} for each edge in E (and $w_{ij} = 0$ if $(i,j) \notin E$), which is the problem of partitioning the nodes of V into two sets S and $V \backslash S$ so that

$$w(S) := \sum_{i \in S, j \in V \backslash S} w_{ij}$$

is maximized. A problem of this type arises from many network planning, circuit design, and scheduling applications.

$$\begin{aligned} w^* := \quad & \text{Maximize} \quad w(\mathbf{x}) := \frac{1}{4}\sum_{i,j} w_{ij}\left(1 - x_i x_j\right) \\ & s.t. \quad (x_j)^2 = 1, j = 1, \ldots, n \end{aligned} \tag{62}$$

Then we do the semidefinite relaxation reformulation. Let $Z_{ij} = x_i x_j$

$$\begin{aligned} z^{SDP} := \ \text{Maximize} \ & \frac{1}{4} \sum_{i,j} w_{ij} \left(1 - Z_{ij}\right) \\ \textit{Subject to} \quad & Z_{ii} = 1, \quad i = 1, \ldots, n \\ & Z \succeq \mathbf{0}, \operatorname{rank}(Z) = 1 \end{aligned} \tag{63}$$

If we remove the rank-one constraint, it will be SDP relaxation problem

Theorem 24. *(Goemans and Williamson)*

$$\mathrm{E}[w(\hat{\mathbf{x}})] \geq .878 z^{SDP} \geq .878 w^* \tag{64}$$

Mathematical Optimization Models and Applications	July 8,2022
Lecture 8	
Lecturer: Yinyu Ye	*Scriber: Bin Yu*

17 Optimality Conditions for Nonlinear Optimization

17.1 KKT Optimality Condition

A differentiable function f of one variable defined on an interval $F = [ae]$. If an interior-point $\bar{x}$ is a local/global minimizer, then $f'(\bar{x}) = 0$; if the left-end-point $\bar{x} = a$ is a local minimizer, then $f'(a) \geq 0$; if the right-end-point $\bar{x} = e$ is a local minimizer, then $f'(e) \leq 0$. first-order necessary condition (FONC) summarizes the three cases by a unified set of optimality/complementarity slackness conditions

If $f'(\bar{x}) = 0$, then it is also necessary that $f(x)$ is locally convex at $\bar{x}$ for it being a local minimizer. How to tell the function is locally convex at the solution? It is necessary $f''(\bar{x}) \geq 0$, which is called the second-order necessary condition (SONC), which we would explored further.

These conditions are still not, in general, sufficient. It does not distinguish between local minimizers, local maximizers, or saddle points.

If the second-order sufficient condition (SOSC): $f''(\bar{x}) > 0$, is satisfied or the function is strictly locally convex, then $\bar{x}$ is a local minimizer. Thus, if the function is convex everywhere, the first-order necessary condition is already sufficient.

Then we want to explore the second order optimality under unconstrained optimization.

Theorem 25. *(First-Order Necessary Condition) Let $f(\mathbf{x})$ be a C^1 function where $\mathbf{x} \in R^n$. Then, if $\overline{\mathbf{x}}$ is a minimizer, it is necessarily $\nabla f(\overline{\mathbf{x}}) = 0$.*

Theorem 26. *(Second-Order Necessary Condition) Let $f(\mathbf{x})$ be a C^2 function where $\mathbf{x} \in R^n$. Then, if $\overline{\mathbf{x}}$ is a minimizer, it is necessarily*

$$\nabla f(\overline{\mathbf{x}}) = \mathbf{0} \textit{ and } \nabla^2 f(\overline{\mathbf{x}}) \succeq \mathbf{0}.$$

Note that the Hessian matrix is the semidefinite

Furthermore, if $\nabla^2 f(\overline{\mathbf{x}}) \succ 0$, then the condition becomes sufficient. The proofs would be based on 2nd-order Taylor's expansion at $\overline{\mathrm{x}}$

Proof. we have the second order Taylor's expansion

$$f(x) = f(\bar{x}) + (x-\bar{x})^T \bigtriangledown f(\bar{x}) + (x-\bar{x})^T \bigtriangledown^2 f(\bar{x})(x-\bar{x}) + o(||x-\bar{x}||^2)$$

If $\bigtriangledown f(\bar{x}) = 0$ and $\bigtriangledown^2 f(\bar{x}) = 0$, the $\bar{x}$ is not the local optimal. □

Such that if these conditions are not satisfied, then one would be find a descent-direction d and a small constant $\bar{\alpha} > 0$ such that $f(\overline{\mathbf{x}} + \alpha\mathbf{d}) < f(\overline{\mathbf{x}}), \forall 0 < \alpha \le \bar{\alpha}$

For example, if $\nabla f(\overline{\mathbf{x}}) = \mathbf{0}$ and $\nabla^2 f(\overline{\mathbf{x}}) \not\succeq 0$, the eigenvector of a negative eigenvalue of the Hessian would be a descent direction from $\overline{\mathrm{x}}$. Again, they may still not be sufficient, e.g., $f(x) = x^3$.

17.2 Descent Direction

Let f be a differentiable function on R^n. If point $\overline{\mathrm{x}} \in R^n$ and there exists a vector d such that

$$\nabla f(\overline{\mathbf{x}})\mathbf{d} < 0$$

then there exists a scalar $\bar{\tau} > 0$ such that

$$f(\overline{\mathbf{x}} + \tau\mathbf{d}) < f(\overline{\mathbf{x}}) \text{ for all } \tau \in (0, \bar{\tau}).$$

Note that $\exists\tau, \frac{f(\overline{\mathbf{x}}+\tau\mathbf{d})-f(\overline{\mathbf{x}})}{\tau\mathbf{d}}\mathbf{d} < 0$

The vector $\mathbf{d}$ (above) is called a descent direction at $\overline{\mathbf{x}}$. If $\nabla f(\overline{\mathbf{x}}) \neq 0$, then $\nabla f(\overline{\mathbf{x}})$ is the direction of steepest ascent and $-\nabla f(\overline{\mathbf{x}})$ is the direction of steepest descent at $\overline{\mathbf{x}}$. Denote by $\mathcal{D}^d_{\overline{\mathrm{x}}}$ the set of descent directions at $\overline{\mathrm{x}}$, that is,

$$\mathcal{D}^d_{\overline{\mathrm{x}}} = \{\mathbf{d} \in R^n : \nabla f(\overline{\mathbf{x}})\mathbf{d} < 0\}$$

At feasible point $\overline{\mathbf{X}}$, a feasible direction is

$$\mathcal{D}^f_{\overline{\mathbf{x}}} := \{\mathbf{d} \in R^n : \mathbf{d} \neq \mathbf{0}, \overline{\mathbf{x}} + \lambda\mathbf{d} \in \mathcal{F} \text{ for all small } \lambda > 0\}.$$

17.3 Optimality Condition of Problem

Roughly the optimality condition of unconstrained problem is

Theorem 27. *Let* $\overline{\mathrm{x}}$ *be a (local) minimizer of (UP). If the functions* f *is continuously differentiable at* $\overline{\mathrm{x}}$*, then*

$$\nabla f(\overline{\mathbf{x}}) = \mathbf{0}.$$

This condition is also sufficient for global optimality if f *is a convex function.*

Consider the linear equality-constrained problem, where f is differentiable on R^n,

$$\begin{aligned} \text{(LEP)} \qquad & \min f(\mathbf{x}) \\ & \text{s.t. } A\mathbf{x} = \mathbf{b}. \end{aligned}$$

Theorem 28. *(the Lagrange Theorem) Let* $\overline{\mathrm{x}}$ *be a (local) minimizer of (LEP). If the functions* f *is continuously differentiable at* $\overline{\mathrm{x}}$*, then*

$$\nabla f(\overline{\mathbf{x}}) = \overline{\mathbf{y}}^T A$$

for some $\overline{\mathbf{y}} = (\bar{y}_1; \ldots; \bar{y}_m) \in R^m$*, which are called Lagrange or dual multipliers. This condition is also sufficient for global optimality if* f *is a convex function.*

The Lagrange formula is

$$L(x, y) = f(x) + y^T b - y^T Ax$$

Then take the derivative

$$\bigtriangledown L(\bar{x}, y) = \bigtriangledown f(\bar{x}) - \bar{y}^T A = 0$$

The geometric interpretation: the objective gradient vector is perpendicular to or the objective level set tangents the constraint hyperplanes. Let $v(\mathbf{b})$ be the minimal value function of $\mathbf{b}$ of (LEP). Simliarly

$$\bigtriangledown L(b) = \bar{y} = 0$$

Proof. Consider feasible direction space

$$\mathcal{F} = \{\mathbf{x} : A\mathbf{x} = \mathbf{b}\} \Rightarrow \mathcal{D}^f_{\overline{\mathrm{x}}} = \{\mathrm{d} : A\ \mathrm{d} = 0\}$$

If $\overline{\mathbf{x}}$ is a local optimizer, then the intersection of the descent and feasible direction sets at $\bar{x}$ must be empty or

$$A\mathbf{d} = \mathbf{0}, \nabla f(\overline{\mathbf{x}})\mathbf{d} \neq 0$$

has no feasible solution for d. By the Alternative System Theorem it must be true that its alternative system has a solution, that is, there is $\bar{y} \in R^n$ such that

$$\nabla f(\bar{\mathbf{x}}) = \bar{\mathbf{y}}^T A = \sum_{i=1}^{m} \bar{y}_i A_i$$

$\square$

17.4 Barrier Optimization

Consider the problem

$$\begin{array}{ll} \min & -\sum_{j=1}^{n} \log x_j \\ \text{s.t.} & A\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{array} \tag{65}$$

The non-negativity constraint can be removed if the feasible region has an "interior", that is, there is a feasible solution such that x > 0. Thus, if a minimizer $\bar{\text{x}}$ exists, then $\bar{\text{x}} > 0$ and

$$-\mathbf{e}^T \bar{X}^{-1} = \bar{\mathbf{y}}^T A = \sum_{i=1}^{m} \bar{y}_i A_i.$$

18 KKT Condition

Let us now consider the inequality-constrained problem

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & A\mathbf{x} \geq \mathbf{b}. \end{array}$$

Theorem 29. *(the KKT Theorem) Let $\bar{\text{x}}$ be a (local) minimizer of (LIP). If the functions f is continuously differentiable at $\bar{\text{x}}$, then*

$$\nabla f(\bar{\mathbf{x}}) = \bar{\mathbf{y}}^T A, \bar{\mathbf{y}} \geq \mathbf{0}$$

for some $\bar{\mathbf{y}} = (\bar{y}_1; \ldots; \bar{y}_m) \in R^m$, which are called Lagrange or dual multipliers, and $\bar{y}_i = 0, \quad if\ i \notin \mathcal{A}(\bar{\mathbf{x}})$. These conditions are also sufficient for global optimality if f is a convex function.

Then we can have the KKT constraint. We now consider optimality conditions for problems having both inequality and equality constraints. These can be denoted

$$\begin{array}{lll} \min & f(\mathbf{x}) & \\ \text{s.t.} & \mathbf{a}_i\mathbf{x} \quad (\leq, =, \geq) & b_i, i = 1, \ldots, m, \end{array}$$

For any feasible point $\bar{x}$ of (P) we have the sets

$$\mathcal{A}(\overline{\mathbf{x}}) = \{i : \mathbf{a}_i\mathbf{x} = b_i\}$$
$$\mathcal{D}^d_{\overline{\mathbf{x}}} = \{\mathbf{d} : \nabla f(\overline{\mathbf{x}})\mathbf{d} < 0\}$$

Theorem 30. *Let* $\overline{\mathrm{x}}$ *be a local minimizer for* (P)*. Then there exist multipliers* $\overline{\mathrm{y}}$ *such that*

$$\begin{array}{rccl}
\mathbf{a}_i\overline{\mathbf{x}} & (\leq,=,\geq) & b_i, i = 1,\dots,m, & \\
\textit{(Original Problem Constraints } (OPC)) & & & \\
\nabla f(\overline{\mathbf{x}}) & = & \overline{\mathbf{y}}^T A & \\
\textit{(Lagrangian Multiplier Conditions (LMC))} & & & \\
\bar{y}_i \quad (\leq,' \textit{ free }',\geq) & 0, i = 1,\dots,m, & & \\
\textit{Multiplier Sign Constraints } (MSC) & & & \\
\bar{y}_i & = & 0 & \textit{if } i \notin \mathcal{A}(\overline{\mathbf{x}}) \\
\textit{(Complementarity Slackness Conditions (CSC)).} & & &
\end{array} \tag{66}$$

These conditions are also sufficient for global optimality if f *is a convex function.*

And we will emphasize again to sufficient and necessary the convexity. Like $f(x) = x^3$

- Hessian matrix is PSD in the feasible region
- Epigraph is a convex set

18.1 LCOP Examples: Linear Optimization

$$(LP) \quad \min \quad \mathbf{c}^T\mathbf{x}$$
$$s.t. A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq 0$$
$$\text{s}$$

For any feasible x of (LP), it's optimal if for some y,

$$x_j s_j = 0, \forall j = 1,\dots,n$$
$$A\mathbf{x} = \mathbf{b}$$
$$\nabla\left(\mathbf{c}^T\mathbf{x}\right) = \mathbf{c}^T = \mathbf{y}^T A + \mathbf{s}^T$$
$$\mathbf{x},\mathbf{s} \geq \mathbf{0}.$$

Here, **y** (shadow prices in LP) are Lagrange or dual multipliers of equality constraints, and **s** (reduced gradient/costs in LP) are Lagrange or dual multipliers for x $\geq \mathbf{0}$.

18.2 LCOP Examples : Quadratic Optimization

$$\begin{array}{lll} (QP) & \min & \mathbf{x}^T Q\mathbf{x} - 2\mathbf{c}^T\mathbf{x} \\ & \text{s.t.} & A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0} \end{array}$$

Optimality Conditions:

$$\begin{aligned} x_j s_j &= 0, \forall j = 1, \ldots, n \\ A\mathbf{x} &= \mathbf{b} \\ 2Q\mathbf{x} - 2\mathbf{c} - A^T\mathbf{y} - \mathbf{s} &= \mathbf{0} \\ \mathbf{x}, \mathbf{s} &\geq \mathbf{0} \end{aligned}$$

18.3 LCOP Examples: Linear Barrier Optimization

$$\min f(\mathbf{x}) = \mathbf{c}^T\mathbf{x} - \mu \sum_{j=1}^{n} \log(x_j), \text{ s.t. } A\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}$$

for some fixed $\mu > 0$. Assume that interior of the feasible region is not empty:

$$\begin{aligned} A\mathbf{x} &= \mathbf{b} \\ c_j - \frac{\mu}{x_j} - \left(\mathbf{y}^T A\right)_j &= 0, \forall j = 1, \ldots, n \\ \mathbf{x} &> \mathbf{0}. \end{aligned}$$

Let $s_j = \frac{\mu}{x_j}$ for all j (note that this s is not the s in the KKT condition of $f(\mathbf{x})$). Then

$$\begin{aligned} x_j s_j &= \mu, \forall j = 1, \ldots, n, \\ A\mathbf{x} &= \mathbf{b}, \\ A^T\mathbf{y} + \mathbf{s} &= \mathbf{c}, \\ (\mathbf{x}, \mathbf{s}) &> \mathbf{0}. \end{aligned}$$

19 Inverse Optimization

We know that the KKT theorem could be applied into the equilibrium, and the dual variable reflects the econ perproties. We can explore which are we can use the KKT theorem, such as the inverse optimization. This area is introduced by Ahuja and Orlin (2001). In this paper, we study inverse optimization problems defined as follows. Let $\mathbf{S}$ denote the set of feasible solutions of an optimization problem P, let c be a specified cost vector, and x^0 be a given feasible solution. The solution x^0 may or may not be an optimal solution of $\mathbf{P}$ with respect to the cost vector c. The inverse optimization problem is to perturb the cost vector c to d so that x^0 is an optimal solution of $\mathbf{P}$ with respect to d and $\|d-c\|_p$ is minimum, where $\|d-c\|_p$ is some selected L_p norm. In this paper, we consider the inverse linear programming problem under L_1 norm (where $\|d-c\|_p = \sum_{i\in J} w_j\,|d_j - c_j|$, with J denoting the index set of variables x_j and w_j denoting the weight of the variable j) and under L_∞ norm (where $\|d-c\|_p = \max_{j\in J}\{w_j\,|d_j - c_j|\}$). We prove the following results:

1. If the problem $\mathbf{P}$ is a linear programming problem, then its inverse problem under the L_1 as well as L_∞ norm is also a linear programming problem.

2. If the problem $\mathbf{P}$ is a shortest path, assignment or minimum cut problem, then its inverse problem under the L_1 norm and unit weights can be solved by solving a problem of the same kind. For the nonunit weight case, the inverse problem reduces to solving a minimum cost flow problem.

3. If the problem $\mathbf{P}$ is a minimum cost flow problem, then its inverse problem under the L_1 norm and unit weights reduces to solving a unit-capacity minimum cost flow problem. For the nonunit weight case, the inverse problem reduces to solving a minimum cost flow problem.

4. If the problem $\mathbf{P}$ is a minimum cost flow problem, then its inverse problem under the L_∞ norm and unit weights reduces to solving a minimum mean cycle problem. For the nonunit weight case, the inverse problem reduces to solving a minimum cost-to-time ratio cycle problem.

5. If the problem $\mathbf{P}$ is polynomially solvable for linear cost functions, then inverse versions of $\mathbf{P}$ under the L_1 and L_∞ norms are also polynomially solvable.

References

Ahuja, Ravindra K, James B Orlin. 2001. Inverse optimization. *Operations Research* **49**(5) 771–783.

Luenberger, David G, Yinyu Ye, et al. 1984. *Linear and nonlinear programming*, vol. 2. Springer.