EE 227B Lecture Notes

Tom Hu

October 30, 2018

This notes is scribed from the lecture notes of EE 227B by Professor El Ghaoui and Sojoudi. All typos are on me.

If you find any typos, or if you are interested in any extra topics, you are welcomed to shoot an email to zyhu95@berkeley.edu

Contents

1 Lecture 01 (Aug 23)

Overview

- Optimization problem
- Conic optimization and semi-definite programming
- Conic duality
- Optimality conditions
- Midterm
- Robust optimization
- Applications

Optimization

$$\min_x \quad f_0(x) \qquad \text{s.t. } f_i(x) \leq 0, i = 1, 2, \ldots, m$$

$x \in \mathbb{R}^n$ is the optimization variable. $f_0(x) : \mathbb{R}^n \to \mathbb{R}$ is the objective function. $f_i(x) : \mathbb{R}^n \to \mathbb{R}$ are the constraint functions. We define feasible set as $T := \{x \mid f_i(x) \leq 0, i = 1, 2, \ldots, m\}$. Hence our problem of interest can also be written as:

$$\min_x \quad f_0(x) \qquad \text{s.t. } x \in T$$

Different type of points

- Feasible x: $x \in T$
- Global minimum: x^*. Giving lowest cost among all feasible points.
- Local minimum

How can we measure the complexity of a problem?

- Solution methods:

  Analytical solution: Formula

  Numerical solution: Algorithm

  Example: Least squares problem. Define x: decision, b: measurement and Ax: model. Thus the problem is:

  $$\min_x \|Ax - b\|_2^2 \Rightarrow x^* = (A^T A)^{-1} A^T b \approx \mathcal{O}(n^3)$$

  Now if the problem gets more complicated:

  $$\min_{x : c_i^T x \leq d_i, i=1,\ldots,m} \|Ax - b\|_2^2$$

This problem does not have a closed form solution. Hence we need an algorithm: give an initial guess $x^{(0)}$, then we form a sequence following some updates rules $x^{(1)}$, $x^{(2)}$, ... that hopefully converge to the optimal solution x^*. The algorithm iteratively designs $x^{(i)}$ converging to x^*. Now the complexity of the algorithm depends on number of basic operations. It is a function of n (number of variables), m (number of constraint), F (hidden cost for evaluating gradient / hessian etc. etc.).

- $\mathcal{O}(n)$: easy, $\mathcal{O}(2^n)$: hard.
  Example:
  $$\min_{x:x_i^2=1} \quad c^T x$$
  We suppose $n = 500$, then this combinatorial problem will become intractable, i.e. 2^{500} is too big!
- In this class, we consider complexity in **polynomial** degree of m and n as easy problems.

Class of convex optimization problems

LP (Linear programming) $\subset$ QP (Quadratic programming) $\subset$ QCQP (Quadratically constraint quadratic programming) $\subset$ SOCP (Second-order cone programming) $\subset$ SDP (Semi-definite programming) $\subset$ Conic

Reformulation and approximation

Some examples include:

- Circuits: device sizing
- Control theory: optimal control and system identification
- Communication networks: TCP (transmission-control protocol)
- Power system: scheduling of generators
- Signal processing: compress sensing

Compressed sensing

Some theorem states that sampling rate for exact recovery should $\geq 2\times$ highest frequency of the signal. But what if we cannot sample that much?
We suppose x is a sparse vector and after some linear measurement, we have y, which is rather dense. i.e. We have a matrix A such that $y = Ax$, thus A is a "fat" matrix since the size of y is way smaller than the size of x. We want to **design** A such that given y, we can recover x.
We give the **sparse recovery** problem:

$$\min_{x\in\mathbb{R}^n:y=Ax} \|x\|_0$$

We use 0-norm primarily for the cardinality. But then this problem is **non-convex** and hence hard. Thus, to solve this, we need to use a convex approximation of the problem:

$$\min_{x:y=Ax} \|x\|_1$$

1.1 Operations and convexity

Suppose we have $x, y \in \mathbb{R}^n$ two vectors, we define the following operations:

- Affine combination:
  $$(x, y) \to \{\alpha x + \beta y \mid \alpha + \beta = 1\}$$
  Geometry: any points on a straight line cross two points.
- Convex combination:
  $$(x, y) \to \{\alpha x + \beta y | \alpha + \beta = 1, \alpha, \beta \geq 0\}$$
  Geometry: any points on the **line segment** starting from x ending at y.
- Conic combination:
  $$(x, y) \to \{\alpha x + \beta y \mid \alpha, \beta \geq 0\}$$
  Geometry:
- General case: $(x_1, \ldots, x_k) \to \sum_{i=1}^{k} \alpha_i x_i$. Then:
  Affine: $\sum \alpha_i = 1$
  Convex: $\sum \alpha_i = 1, \alpha_i \geq 0, \forall i$
  Conic: $\alpha_i \geq 0, \forall i$

Now we define the sets:

- Affine set S: $(x_1, \ldots, x_k) \in S$, then affine combination of the points $x_1, \ldots, x_k$ should $\in S$.
- Convex set S: $(x_1, \ldots, x_k) \in S$, then convex combination of the points $x_1, \ldots, x_k$ should $\in S$.
- Convex hull: $C = \{\sum \alpha_i x_i \mid \sum_{i=1}^{k} \alpha_i = 1, \alpha_i \geq 0, x_i \in C\}$. It is the smallest convex set containing C.
- Cone S: $(x_1, \ldots, x_k) \in S$, then conic combination of the points $x_1, \ldots, x_k$ should $\in S$.

Example: **hyperplane** is the set of all vector x such that $a^T x = b$. This is an affine set. **half-space** is the set of all vectors x such that $a^T x \leq b$. This is a convex set.

2 Lecture 02 (Aug 28)

This lecture is scribed by Calvin Chi, only minimum edit performed.

2.1 Example of convex sets

- **Half-spaces**. Let $a^T x = b$ define a hyperplane, then $a^T x \geq b$ or $a^T x \leq b$ is the corresponding half-space.
- **Polyhedron**. A polyhedron can be described as the set

$$P = \{x \mid Ax \preceq b, Cx = d\}$$

where $A \in \mathbb{R}^{m \times n}$ and $C \in \mathbb{R}^{p \times n}$.

- **Norm balls**. A norm ball can be described as the set

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\}$$

To prove this, use the triangle inequality and positive homogeneity of norms.

- **Ellipsoids**. An ellipsoid can be described as the set

$$\mathcal{E} = \{x \mid (x - x_c)^T P^{-1}(x - x_c) \leq 1\}$$

where $P \in \mathbb{S}^n_{++}$[1].

2.2 Set Operations Preserving Convexity

Let $f : \mathbb{R}^n \to \mathbb{R}^m$, then dom f is the set of values x where $f(x)$ is defined. The range of f is the set of all values $f(x)$ where $x \in \text{dom} f$. The following are operations on convex sets that preserve convexity.

- **Intersection**. Intersection of convex sets are also convex. However, the union of convex sets is generally not convex.
- **Affine transformation**. Let $f(x) = Ax + b$ be a function where $f : \mathbb{R}^n \to \mathbb{R}^m$. Let S denote a convex set, then the image of S ($f(s) = \{f(x) | x \in S\}$) under f is also a convex set. The inverse image of S under f is also convex ($f^{-1}(s) = \{x | f(x) \in S\}$).
- **Projection**: the projection of members of a convex set to a lower dimensional space results in another convex set.
- **Linear fractional transformation**. Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be of the form

$$f(x) = \frac{Ax + b}{c^T x + d}$$

For $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, and $d \in \mathbb{R}$. dom $f = \{x | c^T x + d > 0\}$. Then if S is a convex set, then the image $f(S)$ is also a convex set. The inverse image $f^{-1}(S)$ is also convex.

[1] Here $\mathbb{S}_{++}$ denotes the set of positive definite matrices and $\mathbb{S}_{+}$ denotes the set of positive semidefinite matrices

2.3 Convex Functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if $\forall x, y$ and $\forall \alpha \in [0, 1]$,

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y) \tag{2.1}$$

where f has a convex domain[2]. Strict convexity is achieved if the inequality of Equation 1 is strict. The geometric interpretation of a convex function is shown in Figure 1

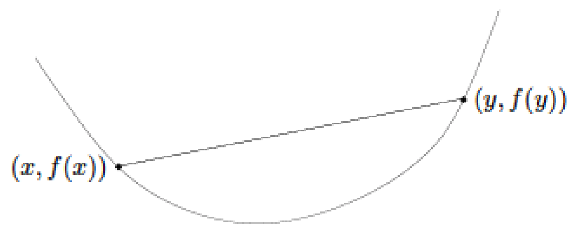


Figure 1: The line segment $\alpha f(x) + (1 - \alpha)f(y)$ is above $f(\alpha x + (1 - \alpha)y)$.

If the function f is furthermore continuous, then the midpoint theorem states that checking Equation 1 is true for $\alpha = \frac{1}{2}$ is sufficient to establishing convexity of f.

First Order Condition for Convexity. Suppose a function $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable and continuous, then f is convex if and only if dom f is a convex set and for all $x, y \in \text{dom} f$,

$$f(y) \geq f(x) + \nabla_x f(x)^T (y - x) \tag{2.2}$$

The geometric interpretation is shown in Figure 2.

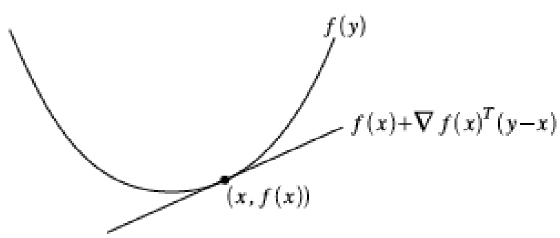


Figure 2: Illustration of the condition that $f(y) \geq f(x) + \nabla_x f(x)^T (y - x)$.

Second Order Condition for Convexity. Suppose a function $f : \mathbb{R}^n \to \mathbb{R}$ is twice differentiable and continuous.Then f is convex if and only if dom f is convex and its Hessian is positive semidefinite (PSD).

$$\nabla_x^2 f(x) \succeq 0 \tag{2.3}$$

[2]The requirement that the domain of f be a convex set is just to ensure that $f(\alpha x + (1 - \alpha)y)$ is defined.

However, a positive definite Hessian is a sufficient but not necessary condition for strict convexity. For instance $f(x) = x^4$ is strictly convex but $\nabla_x^2 f(x) = 12x^2 = 0$ at $x = 0$.

Below are some examples of convex functions that can be verified using the above conditions for convexity.

- **Exponential**. A function of the form $f(x) = e^{ax}$ is convex for $a \in \mathbb{R}$ and strictly convex if $a \neq 0$. This is easily checked with the second order condition for convexity, where $\nabla f^2(x) = a^2 e^{ax} \geq 0$.
- **Powers**.
  1. x^a is convex on $\mathbb{R}_{++}$ if $a \geq 1$ or $a \leq 0$[3].
  2. $-x^a$ is convex on $\mathbb{R}_+$ if $0 \leq a \leq 1$.
- **Logs**. The negative log determinant $-\log \det X$ is convex on domain of PSD matrices. This is a generalization of the statement that $-\log x$ is convex on $\mathbb{R}_{++}$. The negative entropy $x \log x$ is also convex on $\mathbb{R}_{++}$.

Below are additional conditions for establishing convexity.

- Pointwise maximum of a set of convex functions is convex. If $f(x, y)$ is a convex function of x, then for every y in the index set $\mathcal{D}$, $g(x) = \sup_{y \in \mathcal{D}} f(x, y)$ is a convex function in x. Figure 3 illustrates this concept.
- If $f(x)$ is convex, then $g(x) = f(Ax + b)$ is also convex for arbitrary $A \in \mathbb{R}^{m \times n}$ and $b \times \mathbb{R}^m$.
- Assume $f(x, y)$ is jointly convex in x and y and $\mathcal{D}$ is a convex set. Then $g(x) = \inf_{y \in \mathcal{D}} f(x, y)$ is convex if $g(x)$ is always greater than $-\infty$.

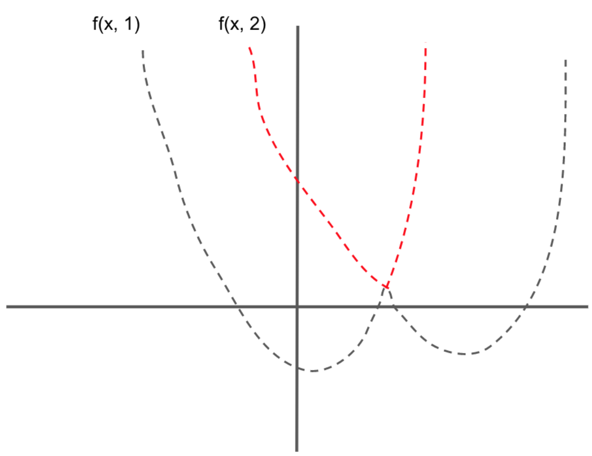


Figure 3: Illustration of the condition that the pointwise maximum of a set of convex functions is convex. Here the index set $\mathcal{D} = \{1, 2\}$ and the red portion of the graph is $g(x) = \sup_{y \in \mathcal{D}} f(x, y)$.

[3] x^a needs to be defined on $\mathbb{R}_{++}$ because otherwise $x^{-1/2}$ is undefined for $x = 0$.

3 Lecture 03 (Aug 30)

Theorem 3.1 (Line restriction theorem). *f is convex if and only if its restriction to any line is convex, i.e. $g(t) = f(x+ty)$ is convex for every $x \in dom(f)$, $y \in \mathbb{R}^n$. $\{t \mid x + ty \in dom(f)\}$.*

Theorem 3.2 (Conic combination). *Let $\alpha_1, \alpha_2, \ldots, \alpha_k \geq 0$, if $f_1, \ldots f_k$ is convex. Then $\sum_{i=1}^{k} \alpha_i f_i(x)$ is also convex.*

Theorem 3.3 (Composite theorem). *Let $f : D_1 \to \mathbb{R}$, $g : D_2 \to \mathbb{R}$, $range(f) \subseteq D_2$. If f and g are convex and g is non-decreasing, then $g(f_i(x))$ is also convex.*

Examples:

- **Quadratic function:** $f(x) = x^T Ax + b^T x + c$, $\nabla f(x) = 2Ax + b$, $\nabla^2 f(x) = 2A$. f is convex iff $A \geq 0$. f is strictly convex **if** $A > 0$.
- **Indicator function** Given D,

$$I_D(x) \begin{cases} 0, & x \in D \\ \infty, & \text{otherwise} \end{cases}$$

  I_D is convex iff D is convex. We can transfer problem using indicator function, i.e. we have $\min_{f_i(x)\leq 0} f_0(x)$, we can transfer it to $\min f_0(x) + I_D(x)$ where D is the feasible region.

- **Distance to a set**

$$g(x) = \inf_{y \in M} \|x - y\|$$

  $\|x - y\|$ is jointly convex in (x, y), if $z = \begin{bmatrix} x \\ y \end{bmatrix}$ and M is convex, by previous theorem $g(x)$ is convex.

- **Norm** $f(x) = \|x\|$, arbitrary norms. Note that the 0-norm is actually not a norm.
- **Dual norm** $\|\cdot\| \in \mathbb{R}^n$, then:

$$\|u\|_* := \sup\{u^T x | \|x\| \leq 1\}$$

3.1 Fentchel conjugate

Definition 3.1 (Fentchel conjugate). *The Fenchel conjugate of f is defined as:*

$$f^*(z) = \sup_{x \in \ dom\ f} \quad x^T z - f(x)$$

point-wise supremum of a set of affine functions in $z \to$ convex.

Theorem 3.4. *$f^*(z)$ is convex even when $f(x)$ is not*

Theorem 3.5 (Fenchel inequality)**.** $f(x) + f^*(z) \geq x^T z, \quad \forall x, z$

Example:

- $f(x) = ax + b$. Then

$$f^*(z) = \sup_x zx - (ax+b) = \sup_x (z-a)x - b = \begin{cases} -b & z = a \\ \infty & \text{otherwise} \end{cases} \rightarrow f^*(z) = -b, \text{dom}(f^*) = \{a\}$$

- $f(x) = \frac{1}{2}x^T Ax + b^T x + c, A > 0$, then: $f^*(z) = \frac{1}{2}(z-b)^T A^{-1}(z-b) - c$.

- $f(x) = \sum_{i=1}^n x_i \log x_i$, then $f^*(z) = \sum_{i=1}^n e^{z_i - 1}$. Both ones use the trick that $f(x)$ itself is non-negative, hence would be upper-bounded by $x^T z$. Calculation is eliminated.

- $f(X) = -\log\det(X), \text{dom}(f) = S^n_{++}$, then $f^*(Z) = -\log\det(-Z^{-1}) - n$. The optimization problem is:

$$f^*(Z) = \sup_{X \in \mathbb{S}^n_{++}} \text{Tr}(XZ) + \log\det(X)$$

Note that we can simply construct $X = I + tvv^T$, then if $Z \not< 0$:

$$\text{tr}(XZ) + \log\det(X) = \text{tr}(Z) + t\lambda + \log\det(I + vv^T) \rightarrow \infty$$

If $Z < 0$, we take the gradient with respect to X and yield $Z + X^{-1} = 0$. Thus $X^* = -Z^{-1}$, plug it in yield desired solution.

- If $f(x) = \|x\|$, then:

$$f^*(z) = I_{\|\cdot\|_* \leq 1}(z) = \begin{cases} 0, & \text{if } \|z\|_* \leq 1 \\ +\infty, & \text{if } \|z\|_* > 1 \end{cases}$$

Proof. Recall $\|z\|_* = \sup_{\|x\| \leq 1} x^T z$.
Case 1: if $\|z\|_* \leq 1$, then:

$$z^T \left(\frac{x}{\|x\|} \right) \leq \|z\|_* \leq 1 \Rightarrow z^T x - \|x\| \leq 0 \Rightarrow f^*(z) = \sup z^T x - \|x\| \leq 0$$

Case 2: If $\|z\|_* > 1$, then:

$$\exists x \text{ s.t. } \|x\| \leq 1, x^T z > 1 \Rightarrow f^*(z) \geq z^T(tx) - \|tx\| = t \underbrace{(z^T x - \|x\|)}_{>0} \rightarrow \infty$$

□

Theorem 3.6. *If f is convex and the domain of f is closed, then $f^{**} = f, \forall x$. In general, $f^{**}(x) \leq f(x), \forall x$.*

3.2 Subgradients

Definition 3.2. *A vector $g \in \mathbb{R}^n$ is called a subgradient at a point y, if $\forall x \in dom(f)$, the following inequality holds:*

$$f(x) \geq f(y) + \langle g, x - y \rangle$$

Recall the first order condition for convex functions:

$$f(x) \geq f(y) + \langle \nabla f, x - y \rangle$$

Differentiable convex functions:

Non-differentiable convex functions:

Subgradient is a global linear underestimator of function at $(y, f(y))$. For convex functions, any local underestimator is a global underestimator.

Definition 3.3 (Subdifferentials)**.** *Subdifferentials is the set of all subgradients at y and is denoted by $\partial f(y)$*

Theorem 3.7. *$\partial f(x)$ is a closed, convex set and possibly empty.*

Proof. $f(x) - f(y) - \langle f, x - y \rangle \geq 0$. Note that y is fixed, if fix x, set of all g's is a half-space. Intersection of infinitely many half-spaces is still convex. □

Theorem 3.8. *If x is in the interior of $dom(f)$, f convex, then subdifferential is non-empty and bounded.*

Example: $f(x) = |x|$, $\partial f(x) = \begin{cases} 1, & x > 0 \\ -1, & x < 0 \\ [-1, 1], & x = 0 \end{cases}$

Theorem 3.9. *If f is differentiable, then $\partial f(x) = \{\nabla f(x)\}$.*

Theorem 3.10. *If $\partial f(x) = \{g\}$, i.e. a single element, then f is differentiable and $g = \nabla f$.*

Theorem 3.11. *f convex, if $f(x) = \max\{f_1(x), \ldots, f_k(x)\}$, then $\partial f(x)$ is the convex hull of union of $\partial f_i(x)$, where i stands for active functions.*

Example: $f(x) = \max_{1 \leq i \leq k}(a_i^T x + b_i)$, then $\partial(a_i^T x + b_i) = \{a_i\}$, $\partial f(x)$ is the convex hull of $\{a_i \mid i \in I_{(u)}\}$, and

$$I_{(u)} = \{i \mid a_i^T u + b_i = f(x)\}$$

4 Lecture 04 (Sep 4)

In general

$$\min_{\substack{x\in\mathbb{R}^n \\ f_i(x)\le 0, i=1,2,\dots,m \\ h_j(x)=0, j=1,2\dots,p}} f_0(x)$$

4.1 Re-formulations

- $\max g(x) \iff \min -g(x)$
- $g(x) = 0 \iff \begin{cases} g(x) \le 0 \\ -g(x) \le 0 \end{cases}$
- $g(x) \le 0 \iff g(x) + z^2 = 0$, z is the slack variable
- Change of variables: $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is an one-to-one mapping. $x = \phi(z)$. New problem would be: $\min_{\substack{z\in\mathbb{R}^n \\ \tilde{f}_i(z)\le 0 \\ \tilde{h}_j(z)=0}} \tilde{f}_0(z)$. $\tilde{f}_i = f_i \cdot \phi$ and $\tilde{h}_j = h_j \cdot \phi$.
- Transformation:

$$\min_{x\in\mathbb{R}^n, \psi_i(f_i(x))\le 0, i=1,\dots,m, \psi_{j+m}(h_j(x))=0, j=1,\dots,p} \psi_0(f_0(x))$$

  - $\psi_0 : \mathbb{R} \to \mathbb{R}$: strictly increasing
  - $\psi_i : \mathbb{R} \to \mathbb{R}, i = 1, 2, \dots, m$ s.t. $\psi_i(y) \le 0 \iff y \le 0$
  - $\psi_i : \mathbb{R} \to \mathbb{R}$, $i = m+1, \dots, m+p$ s.t. $\psi_i(y) = 0 \iff y = 0$.
- Epigraph. Any optimization problem can be converted to another optimization problem with linear objective. New problem:

$$\min_{x\in\mathbb{R}^n, t\in\mathbb{R}, f_0(x)-t\le 0, f_i(x)\le 0, h_j(x)=0} t$$

  At optimality, $t^* = f(x^*)$

4.2 Solutions of optimization problems

- Feasible solution: x is a feasible solution if it satisfies all the constraints:

$$dom = dom(f_0) \cap dom(f_i) \cap dom(h_j)$$

- Optimal value: Infimum of $f_0(x)$ over all feasible x's $\begin{cases} \infty : \text{infeasible} \\ \text{finite} \\ -\infty : \text{unbounded from below} \end{cases}$
- Global min: x^* is feasible and gives the lowest value possible for $f_0(x)$
- Local min: x^* if $\exists R > 0$ such that $f(x^*) \le f(x), \forall x \in \text{feasible set} \cap \{y \| \|y - x^*\| \le R\}$

4.2.1 Non-convex optimization

- 1st order necessary condition
- 2nd order necessary condition
- 2nd order sufficient condition

4.2.2 Convex optimization

- 1st order necessary condition: global min

4.3 Convex optimization

$f_0, f_1, \ldots, f_m$ convex functions, $h_1, \ldots, h_p$: affine. We have:

$$\min_{\substack{x \in \mathbb{R}^n \\ f_i(x) \leq 0 \\ Ax=b}} f_0(x)$$

Exercise: show that the feasible set of a convex optimization is a convex set.
Overall: convex optimization is the minimization of a convex function over a convex set.
Example:

$$\min_{\substack{x_1^3 \leq 0 \\ x_1+x_2 \geq 1}} x_1^2 + x_2^2$$

is not convex, but we can change $x_1^3 \leq 0$ to $x_1 \leq 0$ to yield an equivalent convex optimization problem.

Theorem 4.1. *Every local min of a convex optimization is a global mean.*

Proof. Assume x^* is a local min but not a global min, and $\tilde{x}$ is the global min. Assume the feasible set of the problem is M. Pick $z = (1-t)x^* + t\tilde{x}$, t small. $z \in M$ but $f_0(z) < f_0(x^*)$, hence x^* can't be a local mean, thus a contradiction. □

Theorem 4.2. *Consider a point x^* that is in the interior of the feasible set. Then x^* is a global min if and only if $\nabla f_0(x^*) = 0$.*

Theorem 4.3. *Consider a point x^* that is in the feasible set. Then x^* is a global min if and only if $\nabla f_0(x^*)(x - x^*) \geq 0$ for all x in the feasible set, X.*

Proof. $\Rightarrow$: Use first order condition, trivial.
$\Leftarrow$: By first order convexity:

$$\begin{aligned} f_0(x) &\geq f_0(x^*) + \nabla f_0^T(x - x^*) \\ &\geq f_0(x^*) \end{aligned}$$

Thus x^* is a global min. □

Definition 4.1. *The level sets of $f_0(x)$ with constant c:*

$$\{x \mid f_0(x) = c\}$$

5 Lecture 05 (Sep 6)

Example 5.1.

$$\min_{x \in \mathbb{R}^n, x \geq 0} f_0(x)$$

convex optimization, x^* is a global min iff $\nabla f_0(x^*)^T(x - x^*) \geq 0, \forall x \geq 0$, this is equivalent to $\sum_{i=1}^n \frac{\partial f_0(x^*)}{\partial x_i}(x_i - x_i^*) \geq 0, \forall x \geq 0$ Pick $x = \begin{bmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_i^* + 1 \\ \vdots \\ x_n^* \end{bmatrix}$, and $\frac{\partial f_0(x^*)}{\partial x_i} \geq 0, \forall i$ If $x_i^* > 0$ for some i, then: $x = \begin{bmatrix} x_1^* \\ x_2^* \\ \vdots \\ x_i^*/2 \\ \vdots \\ x_n^* \end{bmatrix}$ then $\frac{\partial f_0(x^*)}{\partial x_i}(-\frac{x_i^*}{2}) \geq 0 \Rightarrow \frac{\partial f_0(x^*)}{\partial x_i} = 0$.

Theorem 5.1. *Set of global mins of a function $f_0(x) = \{x^* \in \mathbb{R}^n \mid 0 \in \partial f_0(x^*)\}$.*

Proof. x^* is a global min if and only if $f_0(x) \geq f_0(x^*), \forall x \in X$, equivalent to $f_0(x) \geq f_0(x^*) + \langle 0, x - x^* \rangle, \forall x \in X \iff 0 \in df(x^*)$ □

5.1 Different classes of convex optimization

5.1.1 Linear program (LP)

Minimization of a linear function subject to linear equality and inequality constraints.

$$\min_{Gx \leq h, Ax = b} c^T x + d$$

Canonical form:

$$\min_{\tilde{x} \geq 0, \tilde{A}\tilde{x} = \tilde{b}} \tilde{c}^T \tilde{x}$$

Facts:

- $Gx \leq h \Rightarrow Gx + s = h, s \geq 0$

- x can be written as $x = x^+ - x^-, x^+ \geq 0, x^- \geq 0$

Example:

- Piecewise linear

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

$$\min_{x,t} t, \quad a_i^T x + b_i \leq t, i = 1, \ldots, m$$

- Linear fractional

$$\min_{\substack{Gx \leq h \\ Ax=b}} \frac{c^T x + d}{e^T x + f}$$

Note that $\text{dom}(f_0) = \{e^T x + f > 0\}$, define $y = \frac{x}{e^T x + f}, z = \frac{1}{e^T x + f}$. Replace x with y and z subject to: $z \geq 0, e^T y + fz = 1$.

- Absolute value, also use the epigraph idea:

$$\min_{\substack{Gx \leq h \\ Ax=b}} \|x\|_1$$

$$\min_{\substack{Gx \leq h \\ Ax=b \\ -t_i \leq x_i \leq t_i}} \quad t_1 + \ldots + t_n$$

5.1.2 Quadratic program (QP)

Minimization of a quadratic function subject to a linear equality and inequality constraints.

$$\min_{Gx \leq h, Ax=b, P \geq 0} x^T P x + q^T x + r$$

QP includes LP by setting $P = 0$.

5.1.3 Quadratically constrained quadratic program (QCQP)

$$\min_{\substack{\frac{1}{2} x^T P_i x + q_i^T x + r_i \leq 0 \\ Fx=g \\ P_0, P_i \geq 0}} \frac{1}{2} x^T P_0 x + q_0^T x + r_0$$

QCQP includes QP by setting $P_i = 0$

- Constrained least squares:

$$\min_{\|x\|_2 \leq 1} \|Ax - b\|_2^2$$

5.1.4 Second order cone program (SOCP)

Minimization of linear function subject to linear equality constraints and 2-norm inequality constraints:

$$\min_{\substack{\|A_i x + b_i\|_2 \le c_i^T x + d_i \\ Fx = g}} f^T x$$

SOCP includes QCQP. Show it as an exercise! (Hint: Epigraph idea)

Definition 5.1. *Second order cone is:*

$$\{(x,t) | x \in \mathbb{R}^n, t \in \mathbb{R}, \|x\|_2 \le t\}$$

- Robust LP.
  Recall the general LP:
  $$\min_{a_i^T x \le b_i} c^T x$$
  a_i is not known exactly, but we know $a_i \in \epsilon_i = \{\bar{a}_i + P_i u \mid \|u\|_2 \le 1\}$. The robust LP is:
  $$\min_{a_i^T x \le b_i, \forall a_i \in \epsilon_i} c^T x$$
  Note that this is infinitely dimensional convex optimization, very hard to solve.
  We show that Robust LP $\iff$ SOCP:
  $$\min_{\|P_i^T x\|_2 \le b_i - \bar{a}_i x} c^T x$$

  Proof. Note that:
  $$\begin{aligned} &\sup_{a_i \in \{\bar{a}_i + P_i u \mid \|u\|_2 \le 1\}} a_i^T x \\ &= \bar{a}_i^T x + \sup_{\|u\|_2 \le 1} u^T P_i^T x \\ &= \bar{a}_i^T x + \|P_i^T x\|_2 \end{aligned}$$
  □

- LP with random constraints.
  Consider an alternative of LP:
  $$\min_{\mathbb{P}(a_i^T x \le b_i) \ge \eta} c^T x$$
  where the parameters a_i are independent Gaussian random vectors, with mean $\bar{a}_i$ and covariance Σ_i. And $\eta \ge 0.5$.
  Let $u = a_i^T x$, standardize the constraints we yield $\mathbb{P}\left(\frac{u - \bar{u}}{\sigma} \le \frac{b_i - \bar{u}}{\sigma}\right) \ge \eta$. Define $\Phi(z)$ the CDF of unit variance Gaussian. Then we can express the probability constraint as: $\frac{b_i - \bar{u}}{\sigma} \ge \Phi^{-1}(\eta)$. Thus the problem can be expressed as an SOCP:
  $$\min_{\bar{a}_i^T x + \Phi^{-1}(\eta) \|\Sigma_i^{1/2} x\|_2 \le b_i} c^T x$$

5.1.5 Semi-definite program (SDP)

Replace $Gx \leq h$ as linear matrix inequality (LMI) in LP. $F : S^n \to S^n$ s.t. $F(x)$ is linear in every entry $F(x) \leq 0$ is an LMI.
Example: LMI, set $n = 2$, we have:

$$F(x) = \begin{bmatrix} 1 + x_1 & x_1 - 2x_2 \\ x_1 - 2x_2 & x_2 + 5x_1 \end{bmatrix}, F(x) \leq 0 \to LMI$$

A canonical form is:

$$F(x) = x_1F_1 + x_2F_2 + \ldots + x_nF_n + F_0$$

where F_i's are all symmetric.

$$\min_{\substack{x_1F_1+x_2F_2+\ldots+F_0 \leq 0 \\ Ax=b}} c^Tx$$

Canonical LP	Canonical SDP
$x \in \mathbb{R}^n$	$x \in S^n$
$\min c^Tx$	$\min\langle C, X\rangle$
$Ax = b$	$\langle A_i, X\rangle = b_i$
$x \geq 0$	$X \geq 0$ PSD

Note that $\langle A_i, X\rangle = tr(A_iX)$, the canonical SDP is:

$$\min_{\substack{tr(A_iX)=b_i \\ X \geq 0}} tr(CX)$$

Trick: change of variables:

$$X = -F(x) = -x_1F_1 - x_2F_2 \ldots F_0$$

$c^Tx, A^Tx = b$ can be rewritten in terms of X.

6 Lecture 06 (Sep 11)

6.1 Examples of SDP

6.1.1 Eigenvalues

$$\begin{bmatrix} 1 + x_1 & x_2 - 2x_1 & x_1 \\ x_2 - 2x_1 & x_1 - x_2 + 1 & 2x_1 + 1 \\ x_1 & 2x_1 + 1 & x_2 + 3 \end{bmatrix}$$

1. How to minimize the maximum eigenvalue of this matrix?

2. How to minimize sum of two largest eigenvalues?

3. How to minimize sum of all eigenvalues?

4. How to maximize the minimum eigenvalue of the matrix?

Recall the general form:

$$A(x) = A_0 + x_1 A_1 + \dots x_n A_n, A_i \in \mathbb{S}_+^m, \lambda_1(A_i) \leq \lambda_2(A_i) \leq \dots \leq \lambda_m(A_i)$$

1.

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \lambda_{\max}(A(x)) \iff \\ \min_{\substack{x \in \mathbb{R}^n \\ t \in \mathbb{R} \\ A(x) \leq t I_m}} \quad & t \end{aligned}$$

2.

$$\begin{aligned} \min \quad & \textstyle\sum_{i=k}^m \lambda_i(A(x)) \iff \\ \min_{\substack{x \in \mathbb{R}^n, S \in \mathbb{R} \\ Z \in \mathbb{S}^m, Z \geq 0 \\ Z - A(x) + s I_m \geq 0}} \quad & tr(Z) + s(m-k+1) \end{aligned}$$

EVD:

$$\begin{aligned} A &= U \Sigma U^T \\ &= U \text{ diag}(\lambda_i) U^T \\ &= U \underbrace{\begin{bmatrix} \lambda_m - \lambda_k & 0 & \dots & 0 \\ 0 & \lambda_2 - \lambda_k & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots \lambda_k - \lambda_k & \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}}_{Z} U^T + U \underbrace{\begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & \lambda_{k-1} - \lambda_k & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \lambda_1 - \lambda_k \end{bmatrix}}_{NSD} U^T + \lambda_k I \end{aligned}$$

That is $A(x) - Z - \lambda_k I \leq 0$ and $tr(Z) = \sum_{i=k}^m (\lambda_i - \lambda_k) = \sum_{i=k}^m \lambda_i - \underbrace{\lambda_k}_{s}(m-k+1)$:

$$\sum_{i=k}^m \lambda_i = tr(Z) + s(m-k+1)$$

3.

$$\begin{aligned} \max \quad & \textstyle\sum_{i=1}^k \lambda_i(A(x)) \iff \\ \min_{B(x) = -A(x)} \quad & \textstyle\sum_{i=m-k+1}^m \lambda_i(B(x)) \end{aligned}$$

6.1.2 Max Cut

How to partition the vertices of a graph into two sets, such that the number of edges between the two sets is maximal.

- Each edge (i, j) has a weight $w_{ij} \geq 0$.
- Optimization formulation: x_i is a variable associated with note i, and $\begin{cases} +1 & \text{if } u_i \in S_1 \\ -1 & \text{if } u_i \in S_2 \end{cases}$.
- Cut: Define ϵ the edge set, then:

$$\sum_{(i,j)\in\epsilon, i\in S_1, j\in S_2} = \sum_{i,j=1}^{n} \frac{-(x_i+1)(x_j-1)}{4} w_{ij} = \sum_{i,j=1}^{n} \frac{1}{4} \underbrace{w_{ij}(1-x_ix_j)}_{quadratic}$$

  Max-cut can be formulated as:

$$\max_{x\in\mathbb{R}^n, x_i^2-1=0} x^TQx$$

  Matirx reformulation:

$$x^TQx = tr(x^TQx) = tr(Qxx^T) = tr(QX)$$

  Hence the problem:

$$\max_{X\in\mathbb{S}^n, X\geq 0, X_{ii}=1, \text{rank}(X)=1} QX$$

  However, we need to drop the rank constraint in order to make it an SDP, i.e. it's a convex relaxation of a non-convex set. Some proofs show:

$$\frac{1}{0.87}(x^*)^TQx^* \geq tr(QX^*) \geq (x^*)^TQ(x^*)$$

  (Geomans and Williamson in 1995). A sketch of proof is:

  - X^* is not always rank 1, i.e. we cannot always decompose it into $(u^*)^Tu^*$
  - Consider $\mathcal{N}_n(0, X^*)$
  - Define a new probability distribution:

$$\hat{x}_i = \begin{cases} 1 & x_i \geq 0, i = 1, \ldots, n \\ -1 & \text{otherwise} \end{cases}$$

    We then show (see paper for details):

    * $\mathbb{E}[\hat{x}^TQ\hat{x}] \leq (x^*)^TQx^*$
    * $\mathbb{E}[\hat{x}^TQ\hat{x}] \geq 0.87tr(QX^*)$.

    The general idea on simulation is that:

    * Solve SDP relaxation
    * Using X^*, generate a probability distribution
    * Sample random variable $\hat{u}$
    * Get a cut that is 13% away from optimal.

6.1.3 Polynomial optimization

$$\min_{x\in\mathbb{R}^n, p_i(x)\leq 0} p_0(x)$$

where $p(x)$ stands for a polynomial. For example, suppose we want:

$$\min x_1^4 + x_2^2 x_1 + x_1^6$$

These can always be converted into a non-convex QCQP, i.e. the above can be written as:

$$\min_{x_1^2-x_3=0, x_3^2-x_4=0, x_2^2-x_5=0} x_3^2 + x_5 x_1 + x_3 x_4$$

$$\min_{x^T A_i x + b_i^T x + c_i \leq 0} x^T A_0 x + b_0^T x + c_0$$

Note that $A_0, A_1, \ldots$ might not be PSD. It's hence a non-convex QCQP.

6.1.4 General SDP relaxation

By defining $X = \begin{bmatrix} x \\ 1 \end{bmatrix} \begin{bmatrix} x^T & 1 \end{bmatrix}$, drop the rank constraint, we get SDP relaxation.

Theorem 6.1 (Pataki's Theorem)**.** *Consider a canonical SDP:*

$$\min_{\langle M_i, X\rangle = a_i, X\geq 0} tr\,(M_0 X)$$

There is a solution whose rank is upper-bounded by:

$$\left\lfloor \frac{\sqrt{8k+1}-1}{2} \right\rfloor$$

i.e. when $k = 1$, $rank = 1$, $k = O(n)$, $rank = O(\sqrt{n})$.

6.2 Conic optimization

Definition 6.1 (Proper cone)**.** *We say a cone $K \subseteq \mathbb{R}^n$ proper if:*

- *K is convex*
- *K is closed*
- *K has non-empty interior*
- *K is pointed, i.e. if $x \in K, -x \in K$, then $x = 0$. "It cannot contain a line."*

Linear canonical form of Conic optimization:

Canonical LP	Canonical SDP
$\min c^T x$	$\min c^T x$
$Ax = b$	$Ax = b$
$x \geq 0$	$x \geq_K 0$

Note that $x \geq_K 0$ means $x \in K$, a convex cone. Hence $x \geq_K y$ means $x - y \in K$. If $K = \mathbb{R}^n_+, x \geq_K 0 \iff x \geq 0$, so conic becomes LP.

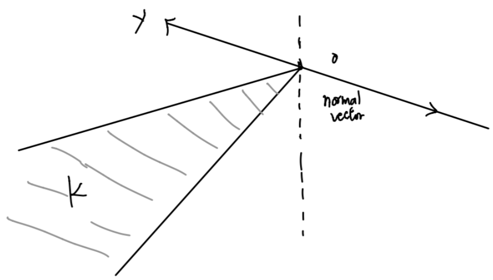
6.2.1 General form of linear conic optimization

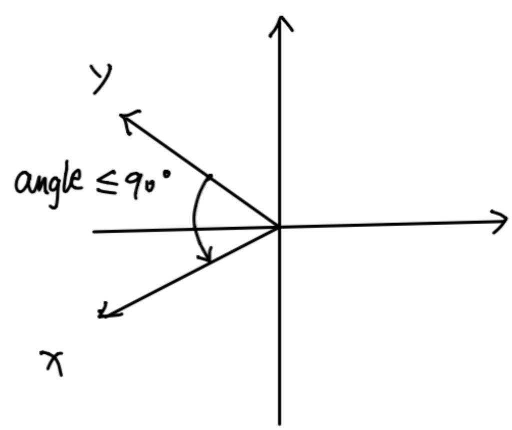
$$\min_{A_i x - b_i \leq_K 0, Fx = g} c^T x$$

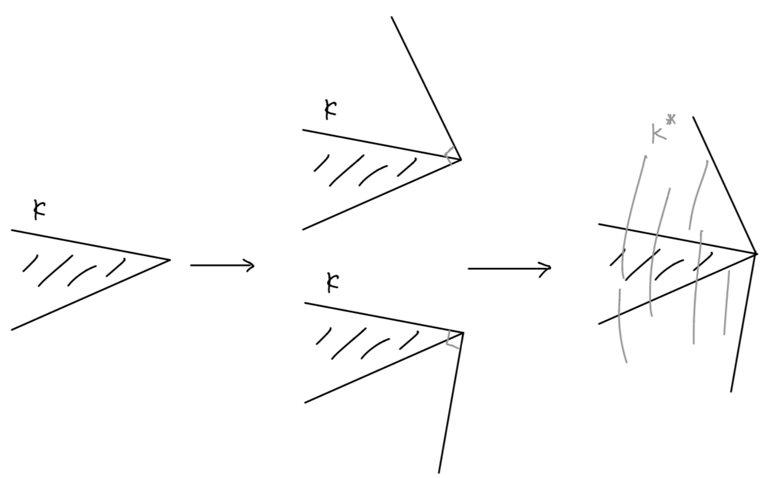
Definition 6.2. *Given a cone K, the dual cone K^* is:*

$$K^* = \{y \mid x^T y \geq 0, \forall x \in K\}$$

Note that K^* is always a cone even if K is not. $y \in K^*$ iff $-y$ is a normal vector of the hyperplane that supports K at the origin.







6.3 Tractable conic optimization

- The non-negative orthant, i.e. $\mathbb{R}_+^n$
- The second-order cone: $\mathbb{Q}^n := \{(x,t) \in \mathbb{R}_+^n \times \mathbb{R}_+ : t \geq \|x\|_2\}$.
  - Rotated second-order cone: $\mathbb{Q}_r^n := \{(x,y,z) \in \mathbb{R}^{n+2} : 2yz \geq \|x\|_2^2, y \geq 0, z \geq 0\}$, which is equivalent to:
  $$(y+z) \geq \left\| \begin{matrix} (y-z) \\ \sqrt{2}x \end{matrix} \right\|_2$$
  Note that this is useful to show that QP is a special case of SOCP, especially the constraint $t \geq x^T Q x$ can be interpreted as $(Q^{1/2}x, t, 1) \in \mathbb{Q}_r^n$.
- The semi-definite cone: $\mathbb{S}_+^n := \{X = X^T \succeq 0\}$.

Note that the below demonstrates that semi-definite cone actually contains second-order cone (also, soc contains nno):

$$\|x\|_2 \leq t \iff \begin{bmatrix} t & x_1 & \dots & x_t \\ x_1 & t & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_n & 0 & \dots & t \end{bmatrix} \succeq 0$$

Theorem 6.2 (Schur's complement)**.** *Given a matrix* $X = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$, *and define Schur complement of* A *as:*

$$X/A = C - B^T A^{-1} B$$

and Schur complement of C *as:*

$$X/C = A - B^T C^{-1} B$$

Then the following hold true:

- *X is positive definite if and only if A (resp C) and X/A (resp X/C) are both positive definite.*
- *if A (resp. C) is positive definite, then X is positive semi-definite if and only if X/A (resp. X/C) is positive semi-definite.*

We can hence show the iff equation from Schur's complement.

7 Lecture 07 (Sep 13)

7.1 Review Problems

1. Give an SDP relaxation of non-convex quadratic optimization of the form:

$$\min_{x^T Q_i x + 2q_i x + p_i \leq 0} x^T Q_0 x + 2q_0 + p_0$$

> We denote the variable as $X = xx^T, x$, then the problem becomes:
>
> $$\min_{\substack{\text{Tr}(XQ_0)+2q_0 x+p_0 \leq 0 \\ \text{Rank}(X)=1}} \text{Tr}(XQ_i) + 2q_i x + p_i$$
>
> Drop the rank 1 we get a SDP relaxation.

2. Recall that $\mathbb{S}_+^n$ is the PSD cone. Given an arbitrary matrix $A \in \mathbb{S}^n$, we want to find the closest point in $\mathbb{S}_+^n$ to A with respect to Frobenious norm (distance between A and X is $\|A - X\|_F$).

   a. Formulate the problem as a smooth conic optimization.

   > The problem can be explicitly formed as:
   >
   > $$\min_{X \geq 0} \|A - X\|_F^2$$
   >
   > Recall that Frob norm can be written as $\|X\|_F = \sqrt{tr(X^T X)}$, thus the problem above can be formulated as
   >
   > $$\min_{X \geq 0} \quad tr(A^T A + X^T X - 2A^T X)$$

   b. Using the optimality condition, find the point.

   > Recall the optimality condition is when $\nabla f(x^*)(x^* - x) \leq 0, \forall x$ in feasible set, i.e.
   >
   > $$\langle (X^* - A), (X^* - X) \rangle \leq 0, \forall X \in \mathbb{S}^n$$
   >
   > Decompose $A = A^+ - A^-$, where $A^+ = U\text{diag}(\lambda_1, \ldots, \lambda_k, 0, \ldots, 0)$, similarly we construct A^-, note that $\langle A^+, A^- \rangle = 0$ hence the above becomes
   >
   > $$\langle X^*, X^* - X \rangle - \langle A^+ - A^-, X \rangle \leq 0$$

3. Consider an integer k between 1 and n.

a. Find the convex hull of the set

$$M = \{XX^T | X \in \mathbb{R}^{n\times k}\}$$

It's the set itself. Let $K_1, \dots, K_k \in M$, then

$$\mathcal{H} := \left\{ \sum_{i=1}^{k} \alpha_i K_i \geq 0 | K_i \in XX^T \right\}$$

Thus $\mathcal{H} \subseteq \mathbb{S}_+^n$. We proceed to prove the other side, suppose we have $A \in \mathbb{S}_+^n$, $A = \sum_{i=1}^n \lambda_i u_i u_i^T$. Define a unit, say

$$\epsilon_i = \sqrt{n}\sqrt{\lambda_i} \begin{bmatrix} u_i & 0 & \dots & 0 \end{bmatrix}$$

Then we can see

$$\sum_{i=1}^{n} \frac{1}{n} \epsilon_i \epsilon_i^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T = A$$

b. Find the convex hull of the set

$$\{XX^T | X \in \mathbb{R}^{n\times k}, \text{rank}(X) = k\}$$

- If $A \geq 0$, but $\text{rank}(A) < k$, then $A \notin \mathcal{H}$. Suppose $A \in \mathcal{H}$, then $A = \sum_{i=1}^m \alpha_i X_i X_i^T$ s.t. $\sum \alpha_i = 1$, then since rank $< k$, $\exists N$ such that $\text{rank}(N) > n-k$, i.e. $N^T A N = 0 \Rightarrow \sum \alpha_i (N^T X_i)(X_i^T N) = 0$. At least one $\alpha_j \neq 0$, then $(N^T X_j)(X_j^T N) = 0$, $X_j^T N = 0$, then this demonstrates $\text{rank}(X_j^T) < n - n(n-k) = k$, a contradiction.
- $\mathcal{H} \in \mathbb{S}_+^n$.
- Convex hull is $\{\mathbb{S}_+^n | \text{rank} \geq k\}$. Assume that $A \in T$ and rank$= k$, then $A = XX^T$ for some X, A is in the convex hull. We show the other way. The other way around: assume $A \in T$ and rank$k = k+1$, then

$$\begin{aligned} A &= \sum_{i=1}^{k} \lambda_i u_i u_i^T \\ &= \frac{2\lambda_1 u_1 u_1^T + \lambda_2 u_2 u_2^T \dots \lambda_k u_k u_k^T}{2} + \frac{\lambda_2 u_2 u_2^T \dots \lambda_k u_k u_k^T + \lambda_{k+1} u_{k+1} u_{k+1}^T}{2} \end{aligned}$$

  We can complete by induction.

4. Find the sub-differential of the function $f(x) = \|x\|_\infty$.

According to Denskin's theorem, $\partial f(x)$ is the convex hull of union of $\text{sgn}(x_i)$ for $i \in D$, where D is the set of index such that the maximum is attained. Or in other words, we can express it as:

5. A matrix $X \in \mathbb{S}^n$ is called copositive if $z^T X z \geq 0, \forall z \geq 0$. Find the dual cone of the cone of copositive matrices.

By definition, dual cone is defined as :

$$\langle Y, X\rangle \geq 0, \forall X \text{ copositive}$$

Claim: $vv^T \in C^*$ if $v \geq 0, v \in \mathbb{R}^n$, note that

$$\langle vv^T, X\rangle = v^T X v \geq 0$$

Thus $C^* \supseteq \alpha_i v_i v_i^T$.
C: copositive, C^*: dual cone. Let:

$$T = \left\{\sum \alpha_i v_i v_i^T \mid \alpha_i \geq 0, v_i \in \mathbb{R}^n_+\right\}$$

Suppose there exists B such that $B \in T^*$ but $B \notin C$. Note that $T \subset C^*$, then $T^* \supset (C^*)^*$, then there exists B such that $B \in T^*$ but $B \notin C$, since $B \notin C$, $\exists z \geq 0$ such that $z^T B z < 0$, i.e. $\langle zz^T, B\rangle < 0$. But $zz^T \in T, B \in T^*$, by definition, the inner product of dual and dual cone must be ≥ 0, thus a contradiction.

6. Find the dual cone of a positive semi-definite cone.

We claim it's self dual and show by contrapositive. We want to establish the fact:

$$Y \mid \text{Tr}(XY) \geq 0, \forall X \geq 0 \iff Y \geq 0$$

Suppose $Y \notin \mathbb{S}^n_+$, then there exists $q \in \mathbb{R}^n$ with:

$$q^T Y q = \text{Tr}(qq^T Y) < 0$$

Hence the psd matrix $X = qq^T$ satisfies $\text{Tr}(XY) < 0$, thus $Y \notin (\mathbb{S}^n_+)^*$.
Now suppose $X, Y \in \mathbb{S}^n_+$, we can express X as $X = \sum_{i=1}^n \lambda_i q_i q_i^T$ where (the eigenvalues) $\lambda_i > 0$, then we have:

$$\text{Tr}(YX) = \text{Tr}\left(Y \sum_{i=1}^n \lambda_i q_i q_i^T\right) = \sum_{i=1}^n \lambda_i q_i^T Y q_i \geq 0$$

This shows $Y \in (\mathbb{S}^n_+)^*$.

7. Find the dual cone of a norm cone, i.e. $K = \{(x,t) \in \mathbb{R}^{n+1} \mid \|u\|_* \le v\}$.

We claim the following:

$$x^T u + tv \ge 0, \forall \|x\| \le t \iff \|u\|_* \le v$$

$\Leftarrow$: Suppose $\|u\|_* \le v, \|x\| \le t, t > 0$, then since we know $\|-x/t\| \le 1$, we have:

$$u^T(-x/t) \le \|u\|_* \le v$$

Thus $u^T x + vt \ge 0$ for x, t such that $\|x\| \le t$.
$\Rightarrow$: We show that if RHS does not hold then LHS does not hold. i.e. Suppose $\|u\|_* > v$, then by definition of the dual norm, $\exists x$ such that $\|x\| \le 1$ and $x^T u < v$, take $t = 1$, we have:

$$u^T(-x) + v < 0$$

which contradicts LHS.

8 Lecture 08 (Sep 18)

8.1 Separating Hyperplanes

Theorem 8.1 (Separating hyperplane theorem)**.** *Suppose C and D are two convex sets that do not intersect, i.e. $C \cap D = \emptyset$, then $\exists a \ne 0$ and b such that:*

$$a^T x \le b, \forall x \in C$$

$$a^T x \ge b, \forall x \in D$$

Note that the inequality is not always strict. A simple counter-example would be $y \le 0$ and $y \ge e^x$, note that $y = 0$ is a separating hyperplane.

Proof. Sketch $\text{dist}(C, D) = \inf\{\|u - v\|_2) \mid u \in C, v \in D\}$. Find the best pair (c, d) to minimize the length.

- Find (c, d)
- Find the mid-point of the segment.
- Draw an orthogonal hyperplane going through the midpoint.

Define: $a = d - c, b = \frac{\|d\|_2^2 - \|c\|_2^2}{2}$. It's easy to show that this is a separating hyperplane. □

Theorem 8.2. *If D is a single point, (we denote as b) and C is a closed set then there is a strict separation.*

Proof. WLOG, we assume $b = 0$ (If not, change variable $x' = x - b$).
There exists $x^0 \in C$ with $\|x^0\|_2 = \delta = \min\{\|x\|_2 \quad | \quad x \in C\}$.
Let a point $m = \frac{x^0}{2}$ and a vector $a = \frac{x^0}{\delta}$ ($\|a\|_2 = 1$), $a^T(x - m) = a^T x + \beta$.
Claim: $a^T x + \beta = 0$ strictly separate C and $b = 0$.

(1) We first consider the case starting from $b = 0$, then we have:

$$\begin{aligned} a^T b + \beta = \beta &= -a^T m \\ &= -\frac{x^0}{\delta} \cdot \frac{x^0}{2} \\ &= -\frac{\|x^0\|_2^2}{2\delta} \\ &= -\frac{\delta^2}{2\delta} = -\frac{\delta}{2} < 0 \end{aligned}$$

(2) Suppose $x \in C$, then we have: $\|x^0\|_2 \leq \|(1-\theta)x^0 + \theta x\|_2$, $\delta^2 = \|x^0\|_2^2 \leq \|x^0 + \theta(x - x^0)\|_2^2 = \|x^0\|_2^2 + 2\theta x^0(x - x^0) + \theta^2 \|x - x^0\|_2^2$, so we have: $0 \leq 2x^0(x - x^0) + \theta\|x - x^0\|_2^2$. Now let $\theta \to 0$, we have: $x^0(x - x^0) \geq 0$. $0 \leq x^0(x - x^0) = (\delta a)(x - 2m)$, since $\delta > 0$, we have $0 < a^T(x - 2m)$. Then:

$$0 < \frac{\delta}{2} = a^T m < a^T(x - m) = a^T x + \beta$$

Hence $a^T x + \beta \geq \frac{\delta}{2} > 0, \forall x \in C$.

□

Theorem 8.3 (Converse separating hyperplane theorem)**.** *Any two convex sets C and D, with at least one being open are disjoint iff there is a separating hyperplane.*

Note that existance of hyperplane does not necessarily imply the sets are disjoint. A sufficient condition is that one set should be open. For example: find necessary and sufficient conditions on $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^n$ such that the inequality $Ax < b$ has no solution.

Define $C = \{b - Ax \mid x \in \mathbb{R}^n\}$, $D = \mathbb{R}^m_{++}$. The problem is infeasible iff $C \cap D$ is empty.

C and D are convex and D is open. By previous theorem,

$$C \cap D = \emptyset \iff \exists \lambda \in \mathbb{R}^m, \mu \in \mathbb{R}, \begin{cases} \lambda^T y \leq M, \text{on } C \\ \lambda^T y \geq M, \text{on } D \end{cases}$$

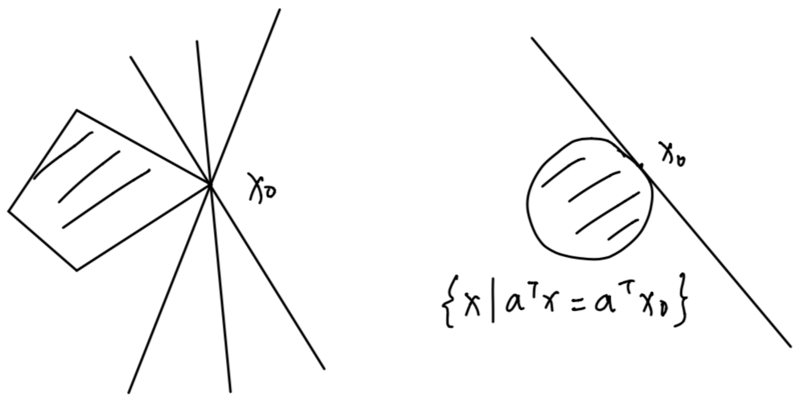
$\iff \exists \lambda \in \mathbb{R}^m$ such that $\lambda \neq 0, \lambda \geq 0, A^T \lambda = 0, \lambda^T b \leq 0$. (left as an exercise).

Summary: The set of inequalities $Ax < b$ is infeasible if the set of inequalities $\lambda \geq 0, A^T \lambda = 0, \lambda^T b \leq 0$ has a non-zero solution.

* This is known as Farkas Lemma. (Special case of theorem of alternatives)

8.2 Supporting Hyperplanes

Definition 8.1. *If C is a closed convex set, and D is a single point x_0 on the boundary of C, then there is a separating hyperplane which is called a supporting hyperplane.*



As the picture shows, at a smooth point, supporting hyperplane is unique and tangential to the set.

8.3 Generalized inequalities

Theorem 8.4. *Assume K is a proper cone, then $\nexists x \in \mathbb{R}^n$ such that $Ax <_K b \iff \exists \lambda \neq 0, \lambda_{K^*} \geq 0, A^T\lambda = 0, \lambda^T b \leq 0$*

Proof. Use separation between two sets. □

Theorem 8.5. *Dual cones satisfy several properties, such as:*

- *K^* is convex and closed*
- *$K_1 \subseteq K_2$ implies $K_2^* \subseteq K_1^*$*
- *If K has non-empty interior, then K^* is pointed.*
- *If the closure of K is pointed, then K^* has non-empty interior.*
- *K^{**} is the closure of the convex hull of K. (Hence if K is convex and closed, $K^{**} = K$.*

Theorem 8.6.

- *If $K = \mathbb{R}^n_+$, then $K^* = K$ by Farkas lemma.*
- *If K is a second order cone, then $K^* = K$*
- *If K is a PSD cone, then $K^* = K$*

Theorem 8.7. *If K is a proper cone, then $K^{**} = K$.*

9 Lecture 09 (Sep 20)

9.1 Duality

Consider the optimization problem:

$$\min_{\substack{f_i(x)\leq 0, i=1,\dots,m \\ h_j(x)=0, j=1,\dots,p}} f_0(x)$$

We associate a scalar variable to each constraint,

$$\lambda_i : f_i(x) \leq 0, \quad \nu_j : h_j(x) = 0$$

Then the Largrangian is defined as $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$:

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{p} \nu_j h_j(x)$$

Note that $\text{dom}(\mathcal{L}) = D \times \mathbb{R}^m \times \mathbb{R}^p$, where D is the domain of the original optimization problem, and λ_i, ν_j's are called Lagrangian multipliers. Let p^* denote the optimal objective value for the original problem.

Theorem 9.1. *As long as $\lambda \geq 0$, we have $g(\lambda, \nu) \leq p^*$, where*

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

Proof. Assume x^* is a global solution of the original problem.

$$\begin{aligned} p^* = f_0(x^*) &\geq f_0(x^*) + \sum_{i=1}^{m} \underbrace{\lambda_i}_{\geq 0} \underbrace{f_i(x^*)}_{\leq 0} + \underbrace{\sum_{j=1}^{p} \nu_j h_j(x^*)}_{=0} \\ &\geq \min_x \mathcal{L}(x, \lambda, \nu) = g(\lambda, \nu) \end{aligned}$$

□

Theorem 9.2. *$-g(\lambda, \nu)$ is always convex for arbitrary functions, f_i's and h_j's.*

Proof. $-g \to \max_x -L$, hence convex. □

- What is the dual function of LP?
  Lagrangian: $\mathcal{L}(x, \lambda, \nu) = c^T x - \sum_{i=1}^n \lambda_i x_i + \nu^T(Ax - b) = -b^T\nu + (c + A^T\nu - \lambda)^T x$.
  Dual function: $g(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu) = -b^T\nu + \inf_x (c + A^T\nu - \lambda)^T x$ Thus $(c + A^T\nu - \lambda) = 0$, (since $-\infty$ otherways), thus:

$$g(\lambda, \nu) = \begin{cases} -b^T\nu & \text{if } c + A^T\nu - \lambda = 0 \\ -\infty & \text{otherwise} \end{cases}$$

- Dual function for max-cut. Primal:

$$\max_{x_i^2 = 1} x^T W x$$

  Note that we can form $Q = -W$, then the problem becomes a minimization problem.

$$\begin{aligned} \mathcal{L}(x, \nu) &= x^T Q x + \sum_{i=1}^n \nu_i (x_i^2 - 1) \\ &= u^T \underbrace{(Q + \text{diag}(\nu_1, \ldots, \nu_n))}_{*} u - \sum_{i=1}^n \nu_i \end{aligned}$$

$$\min_x * = \begin{cases} 0, & \text{if } Q + \text{diag}(\nu_1, \ldots, nu_n) \geq 0 \\ -\infty, & \text{otherwise} \end{cases}$$

$$g(\nu) = \begin{cases} -\sum_{i=1}^n \nu_i & \text{if } Q + \text{diag}(\nu_1, \ldots, nu_n) \geq 0 \\ -\infty, & \text{otherwise} \end{cases}$$

9.2 Lagrangian Dual

Logic: since $g(\lambda, \nu)$ is a lower bound on p^* for every $\lambda \geq 0$, maximize it!

Theorem 9.3. *Weak duality: $p^* \geq d^*$, p^* is the primal solution and d^* is the dual solution.*

Theorem 9.4. *Dual problem is always convex even when primal is not convex.*

- Dual of LP is still LP

- QP:

$$\min_{Ax\leq b, Cx=d} x^T px + q^T x + r$$

$$\begin{aligned}\mathcal{L}(x,\lambda,\nu) &= x^T px + q^T x + \nu + \lambda^T(Ax-b) + \nu^T(Ax-d)\\ &= x^T px + (q^T + \lambda^T A + \nu^T c)x + (r - \lambda^T b - \nu^T d)\end{aligned}$$

  Assume $p \geq 0$. To minimize $\mathcal{L}$, take the derivative:

$$2px + (q^T + \lambda^T A + \nu^T c) = 0$$

  Hence x is linear in λ, ν, plug x in L, then we are maximizing $x^T px$, and x is linear in constraint, i.e. a QP again.

- Exercise: Dual of QCQP is SOCP.

- Exercise: Dual of SOCP is SOCP.

9.3 Connection to Conjugate Functions

For an objective with linear constraints, dual function is:

$$\begin{aligned}g(\lambda,\nu) &= \min_x(f_0(x) + \lambda^T(Ax-b) + \nu^T(cx-d))\\ &= \min_x(f_0(x) + (\lambda^T A + \nu^T c)x) + (-\lambda^T b - \nu^T d)\end{aligned}$$

Note that:

$$\begin{cases} f_0^*(y) & = \sup(y^T x - f_0(x))\\ -f_0^*(y) & = \min(f_0(x) - y^T x)\end{cases}$$

Exactly the conjugate, hence :

$$g(\lambda,\nu) = -f_0^*(-\lambda^T A - \nu^T c)$$

Thus the dual optimization is:

$$\max_{\lambda\geq 0} \; -f_0^*(z) + (-\lambda^T b - \nu^T d)$$

Note $y \in \text{dom}(f_0^*)$.

-

$$\min_{Ax=b} \|x\|$$

$$f_0^*(y) = \begin{cases} 0 & \|y\|_* \leq 1\\ \infty & \text{otherwise}\end{cases}$$

  Dual:

$$\max_{\|-A^T\nu\|_*\leq 1} -f_0^*(-A^T\nu) - b^T\nu$$

9.4 Duality Gap

$p^* - d^*$ is the duality gap. When duality gap = 0, strong duality holds. **How to analyze the duality gap?**
Suppose we have a generic optimization problem (no assumption on convexity).

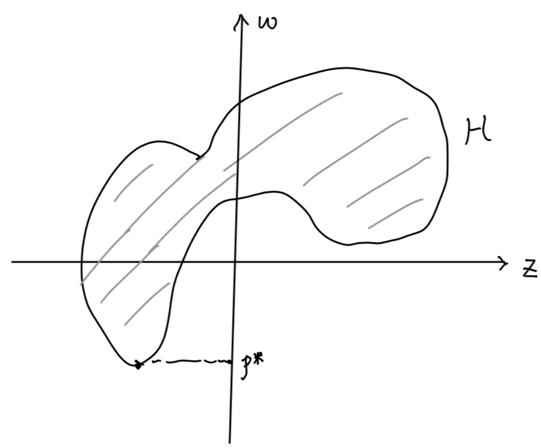
Definition 9.1. $\lambda^* \geq 0$ *is called a geometric multiplier if*

$$p^* = \min_x \mathcal{L}(x, \lambda^*)$$

Consider the space:

$$H = \left\{ (z \in \mathbb{R}^m, w \in \mathbb{R}) \mid z = \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix}, w = f_0(x), \text{for some } x \right\} \in \mathbb{R}^{m+1}$$

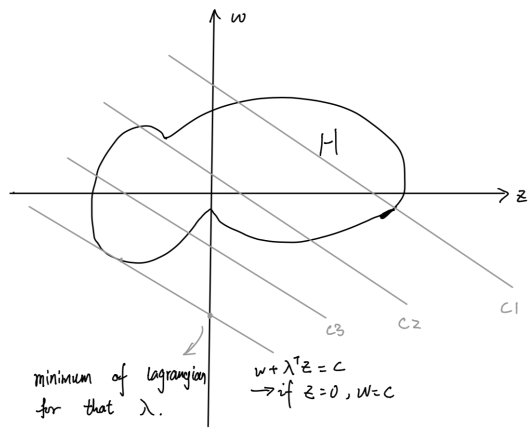
Assuming $n = 1$ for simplicity.



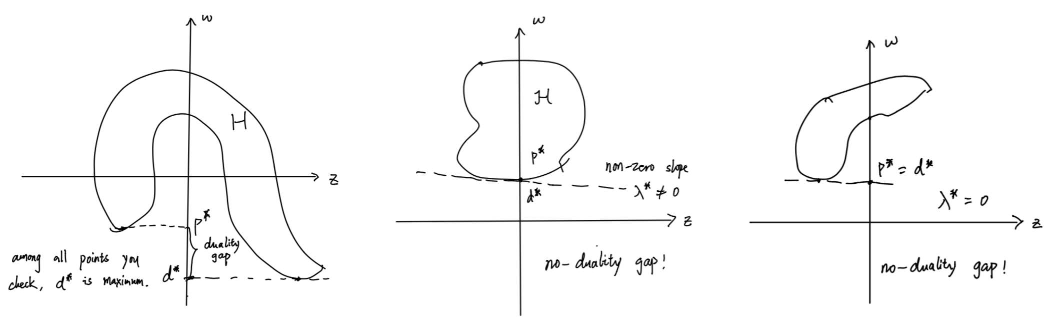
$$\mathcal{L}(x, \lambda) = f_0(x) + \lambda^T f(x) = w + \lambda^T z$$

This is a line in $\mathbb{R}^2$ with the coefficients $\begin{bmatrix} \lambda^T & 1 \end{bmatrix}$, i.e. $w + \lambda^T z = \begin{bmatrix} \lambda^T & 1 \end{bmatrix} \begin{bmatrix} z \\ w \end{bmatrix}$.

$$\min_x \mathcal{L}(x, \lambda) = \min w + \lambda^T z \text{ s.t.} (z, w) \in H$$



Based on the graph above, we can see among all λ's, redo the same thing and find the maximum point that gives you d^*.



Theorem 9.5.

- *If there is no duality gap, the set of geometric multipliers is equal to the set of optimal dual solutions.*
- *If there is a duality gap, even though the set of dual multipliers may not be empty, the set of geometric multipliers might be empty.*

10 Lecture 10 (Sep 25)

10.1 Strong duality

Consider

$$\min_{f_i(x)\leq 0} f_0(x)$$

Assume λ^* is a geometric multiplier (see previous lecture for definition), and the solution exists. Then $p^* = \min_x \mathcal{L}(x, \lambda^*) = f_0(x^*)$. Note that

$$\mathcal{L}(x, \lambda^*) = f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x)$$

Thus $f_0(x^*) \leq \mathcal{L}(x^*, \lambda^*) = f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(u^*)$, which implies $\lambda_i^* f_i(x^*) = 0, i = 1, \ldots, m$, which is **complementary slackness**.

Theorem 10.1. *Existence of a geometric multiplier means no duality gap.*

Proof. By weak duality, we know $p^* \geq d^*$, and $p^* = \min \mathcal{L}(x, \lambda^*) \leq \max_\lambda \min_x \mathcal{L}(x, \lambda) = d^*$. Thus $p^* = d^*$. □

10.1.1 Necessary and sufficient condition for zero duality gap

$\exists(x^*, \lambda^*)$ such that:

1. Primal feasibility: $f_i(x^*) \leq 0, i = 1, \ldots, m$
2. Dual feasibility: $\lambda_i^* \geq 0, i = 1, \ldots, m$
3. Complementary slackness: $\lambda_i^* f_i(x^*) = 0, i = 1, \ldots, m$
4. Lagrangian minimization: $\mathcal{L}(x^*, \lambda^*) = f_0(x^*) = \min_x \mathcal{L}(x, \lambda^*)$.

Unfortunately, 4 is very hard to check. Note that a necessary condition for 4 in the differentiable case is:

$$\nabla \mathcal{L}(x, \lambda^*)\Big|_{x=x^*} = 0$$

Hence 4 can be placed by the **stationarity condition**:

$$\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) = 0 \tag{10.1}$$

It turns out that 1, 2, 3, 10.1 also sufficient for strong duality for certain problems (i.e. convex optimization, under constraint qualification).

10.2 Optimality conditions for equality constraints

We now consider a class of problems that only has equality constraints, i.e.

$$\min_{h_i(x)=0, i=1,\ldots,p} f_0(x)$$

assume h_i's are differentiable. To find a local solution, we need to do a local analysis. Consider a point x^*, we want to check if this is a local minimum. We need to study the feasible set around x^*, which is related to the notion of tangent plane.
Example:

- $f(x) = x_1^2 + x_2^2 - 1$
- $h(x) = x_1^2 + x_2^2 + x_3^2 - 1$

- $h_1(x) = x_1^2 + x_2^2 - 1, x \in \mathbb{R}^3$
- $h_2(x) = x_3 - 1$

How can you find tangent plane?

Definition 10.1 (Regular point)**.** *A point x^* is called regular if vectors $\nabla h_1(x^*), \nabla h_2(x^*), \ldots, \nabla h_p(x^*)$ are linearly independent at that single point.*

If x^* is regular, then the tangent plane of feasible set at x^* is:

$$\left\{\Delta x \in \mathbb{R}^n \mid \nabla h_i(x^*)^T \Delta x = 0, i = 1, \ldots, p\right\}$$

Intuition: $h_i(x^* + \Delta x) = h_i(x^*) + \nabla h_i(x^*)^T \Delta x + \mathcal{O}(x^{*2})$. If a point is regular, then the higher order term is really not important in this equation.

Example:

- Tangent plane for $h(x) = x_1^2 + x_2^2 + x_3^2 - 1$. $\nabla h(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \end{bmatrix}$, this is linearly independent unless $x = 0$, but $x = 0$ is not a feasible point, because $h(0) \neq 0$, which implies that all feasible points are regular. Tangent plane at x^* is:

$$\{\Delta x \in \mathbb{R}^3 \mid x_1^* \Delta x_1 + x_2^* \Delta x_2 + x_3^* \Delta x_3 = 0\}$$

- $h_1(x) = x_1^2 + x_2^2 - 1, h_2(x) = x_3 - 1$, then $\nabla h_1(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \\ 0 \end{bmatrix}$, $\nabla h_2(x) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, and linearly independent, unless $x = 0$, where x is not feasible, thus all points are regular, thus tangent plane at x^* is:

$$\{\Delta x \in \mathbb{R}^3 \mid x_1^* \Delta x_1 + x_2^* \Delta x_2 = 0, \Delta x_3 = 0\}$$

10.2.1 First order optimality conditions

$\min_{x \in D} f_0(x)$ by a local analysis around x^* as: $\min f_0(x^* + \Delta x)$, such that $\Delta x \in$ tangent plane at x^* and Δx small. Note that:

$$\begin{aligned} f_0(x^* + \Delta x) &= f_0(x^*) + \nabla f_0(x^*)^T \Delta x + \text{ h.o.t} \\ &\geq f_0(x^*) \end{aligned}$$

Theorem 10.2. *If x^* is regular and a local min, then $\nabla f_0(x^*)^T \Delta x \geq 0$ for every Δx such that $\nabla h_i(x^*)^T \Delta x = 0$.*

Theorem 10.3. *Under the conditions of previous theorem, $\exists \nu_1^*, \dots, \nu_p^*$ such that*

$$\nabla f_0(x^*) + \nu_1^* \nabla h(x^*) + \dots \nu_p^* \nabla h_p(x^*) = 0$$

This is called ***first order necessary optimality condition.***

10.2.2 Second order optimality conditions

$$\begin{aligned}
f_0(x^* + \Delta x) &= f_0(x^*) + \nabla f_0(x^*)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f_0(x^*) \Delta x + \dots \\
h_1(x^* + \Delta x) &= h_1(x^*) + \nabla h_1(x^*)^T \nabla x + \frac{1}{2}\Delta x^T \nabla^2 h_1(x^*) \Delta x + \dots \\
&\vdots \\
h_p(x^* + \Delta x) &= h_p(x^*) + \nabla h_p(x^*)^T \nabla x + \frac{1}{2}\Delta x^T \nabla^2 h_p(x^*) \Delta x + \dots
\end{aligned}$$

$$f_0(x^* + \Delta x) = f_0(x^*) + \frac{1}{2}\Delta x^T \nabla^2 f_0(x^*) \Delta x + \dots + \frac{1}{2}\sum_{i=1}^{p} \nu_i^* \Delta x^T \nabla^2 h_i(x^*) \Delta x + \text{ h.o.t} \geq f_0(x^*)$$

by local optimality. This concludes

$$\Delta x^T \left(\sum_{i=1}^{p} \nu_i^* \nabla^2 h_i(x^*) + \nabla^2 f_0(x^*) \right) \Delta x \geq 0$$

for all Δx in the tangent plane, which brings us the following theorem:

Theorem 10.4 (2nd order necessary condition)**.** *Under the conditions of previous theorem:*

$$M = \Delta x^T \left(\sum_{i=1}^{p} \nu_i^* \nabla^2 h_i(x^*) + \nabla^2 f_0(x^*) \right) \Delta x \geq 0$$

for every Δx such that $\nabla h_i(x^*)^T \Delta x = 0, i = 1, \dots, p$. If no constraint:

$$\Delta x^T \nabla^2 f_0(x^*) \Delta x \geq 0 \Rightarrow \nabla^2 f_0(x^*) \geq 0, \ \textit{PSD}$$

$M = 0$, to compare $f_0(x^* + \Delta x), f_0(x^*)$ go to third order condition.

Theorem 10.5 (2nd order sufficient condition)**.** *If x^* is regular and feasible, for which $\exists \nu^*$ such that f.o.c. is satisfied and $M > 0$ for every Δx in tangent plane such that $\Delta x \neq 0$, then x^* is a local min. In unconstrained case, x^* is a local min if:*

$$\nabla f_0(x^*) = 0, \nabla^2 f_0(x^*) > 0$$

How to check second order condition? Tangent plane:

$$\{\Delta x \in \mathbb{R}^n \mid \nabla h_i(x^*)^T \Delta x = 0, i = 1, \dots, p\}$$

Dimension of this set is in $n-p$. Pick $n-p$ linearly independent vectors in tangent plane, called $E_1, E_2, \dots, E_{n-p}$. Define $E = \begin{bmatrix} E_1 & E_2 & \dots & E_{n-p} \end{bmatrix}$, we define the tangent plane:

$$\{Ey \mid y \in \mathbb{R}^{n-p}\}$$

2nd order necessary condition:

$$E^T \left(\nabla^2 f_0(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla^2 h_i(x^*) \right) E \geq 0$$

2nd order sufficient condition:

$$E^T \left(\nabla^2 f_0(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla^2 h_i(x^*) \right) E > 0$$

11 Lecture 11 (Sep 27)

11.1 Optimality conditions with inequality constraints

$$\min_{f_1(x) \leq 0, h_1(x)=0} f_0(x) \iff \min_{f_1(x)+z^2=0, h_1(x)=0} f_0(x)$$

Define: $\tilde{x} = \begin{bmatrix} x \\ z \end{bmatrix} \in \mathbb{R}^{n+1}$, $\tilde{h}_2(\tilde{x}) = f_1(x) + z^2$, $\tilde{f}_0(\tilde{x} = f_0(x), h_1(\tilde{x}) = h_1(x)$.

First order necessary condition

$\nabla \tilde{f}_0(\tilde{x}^*) + \nu_1^* \nabla \tilde{h}_1(\tilde{x}^*) + \nu_2^* \tilde{h}_2(\tilde{x}^*) = 0$, which is:

$$\begin{bmatrix} \nabla f_0(x^*) \\ 0 \end{bmatrix} + \nu_1^* \begin{bmatrix} \nabla h_1(x^*) \\ 0 \end{bmatrix} + \nu_2^* \begin{bmatrix} \nabla f_1(x^*) \\ 2z^* \end{bmatrix} = 0$$

Note that since $\nu_2^* = \lambda_1^*$, thus we have $\lambda_1^* z^* = 0 \Rightarrow \lambda_1^* z^{*2} = 0 \Rightarrow \lambda_1^* f(x^*) = 0$, which is the complementary slackness.

Second order conditions

$$\Delta \tilde{x}^T \left(\underbrace{\nabla^2 \tilde{f}_0(\tilde{x}^*)}_{\begin{bmatrix} \nabla^2 f_0(x^*) & 0 \\ 0 & 0 \end{bmatrix}} + \nu_1^* \underbrace{\nabla^2 \tilde{h}_1(\tilde{x}^*)}_{\begin{bmatrix} \nabla^2 h_1(x^*) & 0 \\ 0 & 0 \end{bmatrix}} + \nu_2^* \underbrace{\nabla^2 \tilde{h}_2(\tilde{x}^*)}_{\begin{bmatrix} \nabla^2 f_1(x^*) & 0 \\ 0 & 2 \end{bmatrix}} \right) \Delta \tilde{x} \geq 0$$

which is exactly:

$$\Delta \tilde{x}^T \left(\begin{bmatrix} \nabla^2 f_0(x^*) + \nu_1^* \nabla^2 h_1(x^*) + \lambda_1^* \nabla f_1(x) & 0 \\ 0 & 2 \end{bmatrix} \right) \Delta \tilde{x} \geq 0$$

By tangent plane:

(1) If x^* is regular, then $\tilde{x}^*$ is regular for the new problem.

(2) $2\lambda_1^* \geq 0$

Theorem 11.1 (1st order necessary condition). *Consider* $\min_x f_0(x)$, *such that* $f_i(x) \leq 0, i = 1, \ldots, m$, $h_j(x) = 0, j = 1, \ldots, p$. *If* x^* *is regular and a local min, then* $\exists \lambda_1^*, \ldots, \lambda_m^*, \nu_1^*, \ldots, \nu_p^*$ *such that:*

(1) $\lambda_i^* \geq 0, i = 1, \ldots, m$

(2) $\lambda_i^* f_i(x^*) = 0$ *(complementary slackness)*

(3) $\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(x^*) = 0$

Theorem 11.2 (2nd order necessary condition). *Consider* $\min_x f_0(x)$, *such that* $f_i(x) \leq 0, i = 1, \ldots, m$, $h_j(x) = 0, j = 1, \ldots, p$. *If* x^* *is regular and a local min, then* $\exists \lambda_1^*, \ldots, \lambda_m^*, \nu_1^*, \ldots, \nu_p^*$ *such that:*

(1) $\lambda_i^* \geq 0, i = 1, \ldots, m$

(2) $\lambda_i^* f_i(x^*) = 0$ *(complementary slackness)*

(3) $\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(x^*) = 0$

(4) $\Delta x^T \left(\nabla^2 f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 f_i(x^*) + \sum_{i=1}^p \nu_i^* \nabla^2 h_i(x^*)\right) \Delta x \geq 0$ *for every* Δx *in tangent plane at* x^*, $\Delta x \geq 0$

Theorem 11.3 (2nd order sufficient condition). *If* x^* *is feasible and a regular point for which* $\exists \lambda^*, \nu^*$ *such that (1), (2), (3) are satisfied, and (4) is satisfied in a strict way wherever* $\Delta x \neq 0, \Delta x \in T$, *then* x^* *is a local min.* T *is a set bigger than tangent plane, i.e.:*

$$T = \left\{\Delta x \mid \nabla h_i(x^*)^T \Delta x = 0, i = 1, \ldots, p; \nabla f_i(x^*)^T \Delta x = 0 \text{ if } f_i(x^*) = 0 \& \lambda_i^* > 0\right\}$$

As it has fewer constraints than the tangent plane

Definition 11.1. $f_i(x) \leq 0$ *is called binding (active) at* x^* *is* $f_i(x^*) = 0$

Definition 11.2. x^* *is called regular if gradients of equality and all active inequality constraints are linearly independent.*

Definition 11.3. *Tangent plane at* x^* *is a set of all* $\Delta x \in \mathbb{R}^n$ *that are orthogonal to the gradients of equality and active inequality.*

Note: 2nd order sufficient condition guarantees a strict local optimality.

11.2 Sensitivity analysis

$$\min_{f_i(x)\leq 0, h_j(x)=0} \quad f_0(x) \tag{11.1}$$

$$\min_{f_i(x)\leq c_i, h_j(x)=d_j} \quad f_0(x) \tag{11.2}$$

Suppose $p^*(c,d)$ is the solution to 11.2, then $p^*(0,0)$ would be the solution to 11.1. Note: for convex optimization, $p^*(c,d)$ is still convex.
Assume second order sufficient condition is satisfied for x^* and no constraint is degenerating, then $\exists$ a ball around $(0,0)$ such that for every $(c,d) \in$ ball, we have:

(1) $p^*(c,d)$ exists.

(2) There is a solution $x^*(c,d)$ such that $x^*(0,0) = x^*$ and continuous.

(3) $\nabla_c p^*(c,d)|_{(0,0)} = -\lambda^*$, $\nabla_d p^*(c,d)|_{(0,0)} = -\nu^*$.

For small perturbations (c,d), we have:

$$p^*(c,d) \approx p^* - \sum_{i=1}^{m} \lambda_i^* c_i - \sum_{j=1}^{p} \nu_j^* d_j \tag{11.3}$$

11.3 Optimality conditions for convex optimization

2nd order necessary condition is automatically satisfied for convex optimization. Recall:

$$\Delta x^T \left(\nabla^2 f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla^2 f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla^2 h_i(x^*) \right) \Delta x \geq 0$$

Since f_0, f_i convex, h_j linear, we have $\geq 0 + \geq 0 + 0$, i.e. ≥ 0.

Although second order sufficient condition may not be satisfied, we don't care about second order sufficient condition for convex optimization.

Theorem 11.4.

$$f(x+\Delta x) = f(x) + \nabla f(x)^T \Delta x + \frac{1}{2}\Delta x^T \nabla^2 f(x) \Delta x + \ h.o.t.$$

$\exists y$ such that $\frac{1}{2}\Delta x^T \nabla^2 f(y) \Delta x \geq 0$

Summary:

1. If x^* is regular and a local min, then first order optimality condition is satisfied.

2. If x^* is regular and feasible and satisfies first order conditions, then it's a global min for convex optimization.

which implies FOC is both necessary and sufficient under regularity assumptions.

Theorem 11.5 (Optimality condition for convex optimization)**.**

1) Primal feasibility

2) $\lambda^* \geq 0$

3) Complementary slackness

4) Stationarity of Lagrangian, i.e. $\nabla_x \mathcal{L}(x, \lambda^*, \nu^*) = 0$

Which is called KKT condition.

$$\mathcal{L}(x, \lambda^*, \nu^*) = f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x)$$

Thus $\begin{cases} \nabla_x \mathcal{L}(x^*, \lambda^*, \nu^*) = 0 \\ \mathcal{L}(x^*, \lambda^*, \nu^*) = f_0(x^*) \end{cases}$, which implies x^* is a solution to $\min_x \mathcal{L}(x, \lambda^*, \nu^*)$. Which implies (λ^*, ν^*) is a geometric multiplier. (2) in KKT is called dual feasibility.

For convex optimization, the regularity condition can be replaced with Slater's condition:

Definition 11.4 (Slater)**.** *Slater's condition is satisfied if* $\exists x \in \mathbb{R}^n$ *that is feasible and* $f_i(\bar{x}) < 0$ *(satisfies inequality constraint in a strict way).*
Note that $\bar{x}$ *is an arbitrary point, not need to be the optimal solution.*

Definition 11.5 (Weaker Slater)**.** *Need a feasible point that satisfies all non-linear inequalities in a strict way.*

Theorem 11.6. *Weak duality always holds. Strong duality holds for convex optimization under weaker Slater's condition. If objective value is finite, then there is a dual solution.*

Theorem 11.7. *If Slater is satisfied:* $p^* = d^*$

(1) If p^* *finite and Slater is satisfied for* (p)*, then* $\exists(\lambda^*, \nu^*)$ *finite*

(2) If d^* *finite and Slater is satisfied for* (d)*, then* $\exists x^*$ *finite.*

Example 11.1 (Scenarios for primal-dual for LP)**.**

1. $p^* = d^* = +\infty$*,* (p) *is infeasible*

2. $p^* = d^* = -\infty$*,* (d) *is infeasible*

3. $p^* = d^* =$ finite, which is both $(p), (d)$ have solutions.

The only way we don't have strong duality is when both $(p), (d)$ are infeasible, i.e. $p^* = +\infty, d^* = -\infty$, gap is infinity.

Theorem 11.8. *LP has a zero duality gap unless (p) and (d) are both infeasible. Same for QP.*

12 Lecture 12 (Oct 2)

Consider an example:

$$\min_{\sum_{i=1}^n x_i^2 = n, i \neq j} x_j x_j$$

We know:

$$2\sum_{i\neq j} x_j x_j + \sum_{i=1}^n x_i^2 = \left(\sum_{i=1}^n x_i\right)^2$$

Assume we do not know this, use the optimality condition to solve the problem:

$$\mathcal{L}(x,\nu) = \sum_{i\neq j} x_i x_j + \nu\left(\sum_{i=1}^n x_i^2 - n\right)$$

1st order necessary condition:

- $\nabla_x \mathcal{L}(x,\nu) = 0$, i.e. $\sum_{i\neq j} x_j + 2\nu x_i = 0, i = 1,\ldots,n$
- $\sum_{i=1}^n x_i^2 = n$, which implies $\sum_j x_j + (2\nu - 1)x_i = 0, i = 1,\ldots,n$
  2 possibilities: $\begin{cases} 2\nu - 1 = 0 \Rightarrow \nu = 1/2 \Rightarrow \sum_{j=1}^n x_i = 0 \\ 2\nu - 1 \neq 0 \Rightarrow x_1 = x_2 = \ldots x_n = -\frac{\sum_{i=1}^n x_i}{2\nu - 1} \Rightarrow n = \sum_{i=1}^n x_i^2 \end{cases}$ Now there are two more possibilities: $\begin{cases} x_1 = x_2 = \ldots = x_n = 1 \Rightarrow \nu = -\frac{n-1}{2} \\ x_1 = x_2 = \ldots = x_n = -1 \Rightarrow \nu = -\frac{n-1}{2} \end{cases}$

Second order condition:

- Start with $x_i = 1, i = 1,\ldots,n$
- Regularity: $\nabla_x h(x) = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{bmatrix}$
- The vector is linearly independent if $x \neq 0$. But $x = 0$ is not feasible, thus all feasible points are regular.

- Tangent plane (at x). To find E, we need $n-m=n-1$ linearly independent vectors, one choice:

$$E_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, E_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, E_{n-1} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ -1 \end{bmatrix}$$

-

$$\begin{aligned} & E^T \left(\nabla^2 f_0(x^*) + \nu^* \nabla^2 h_1(x^*)\right) E \\ &= E^T \left(\begin{bmatrix} 0 & 1 & \ldots 1 & \\ 1 & 0 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & 0 \end{bmatrix} - \frac{n-1}{2} \text{diag}(2) \right) E \\ &= -E^T E - (n-1) E^T E = -n E^T E \\ &= -n \left(\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix} + I \right) < 0 \end{aligned}$$

Thus 2nd order condition is satisfied for maximization. and point $[1, \ldots, 1]$ is a strict local max.

- If $n \geq 3$, intersection of a sphere and a hyperplane have infinitely many points. Thus points are not isolated. i.e. second order sufficient condition cannot be satisfied for $n \geq 3$. But one can show 2nd order necessary condition is satisfied:

$$2 \sum_{i \neq j} x_i x_j = \left(\sum x_i\right)^2 - \left(\sum x_i^2\right) \geq 0 - n$$

We can show all these points are global minimum. Also, previously found local max are global max:

  - both give same optimal objective value $n(n-1)/2$
  - All points are regular and we have analyzed every possible stationary point
  - global solution exists

Theorem 12.1. *If feasible set is closed and compact, there exists global min and global max.*

12.1 Example 2

$$\min_{Ax=b} \frac{1}{2}x^T px + q^T x + r$$

$A \in \mathbb{R}^{m\times n}, p \in \mathbb{S}^n_+$. If the problem is feasible, Slater holds for QP and optimality is equivalent to KKT:

- Primal feasibility: $Ax^* = b$
- Stationarity: $0 = \nabla_x \mathcal{L}(x^*, \nu^*) = \nabla_x(f_0(x^*) + \nu^{*T} h_1(x^*))$, by a simple calculation:

$$\begin{aligned} &\Rightarrow 0 = px^* + q + A^T\nu^* \\ &\Rightarrow \underbrace{\begin{bmatrix} p & A^T \\ A & 0 \end{bmatrix}}_{M} \begin{bmatrix} x^* \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix} \end{aligned}$$

Note that M has $n+m$ variables and $n+m$ conditions. Possibilities include:

- If M is invertible, then there exists a unique solution
- Otherwise, it might have zero or infinitely many solutions.

12.2 Example 3: QCQP

$$\min_{\frac{1}{2}x^T p_i x + q_i^T x + r_i \leq 0} \frac{1}{2}x^T p_0 x + q_0^T x + r_0$$

Theorem 12.2 (s-procedure). *If $m = 1$ and Slater holds, then duality gap is 0.*

Consider possibly non-convex equations

$$\begin{cases} f_i(x) \leq 0, i = 1, \dots, m \\ h_j(x) = 0, j = 1, \dots, p \end{cases}$$

We are interested in checking feasibility/infeasibility of this problem. Consider minimizing 0 over this constraint. Then consider the dual function:

$$g(\lambda, \nu) = \inf_x \left(\sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^p \nu_j h_j(x) \right)$$

If $g(\lambda, \nu)$ is strictly positive for some (λ, ν), then $g(\alpha\lambda, \alpha\nu) = \alpha g(\lambda, \nu)$, as $\alpha \to \infty$ then $g(\alpha, \lambda, \nu) \to \infty$. Thus:

$$d^* = \begin{cases} +\infty & \exists(\lambda, \nu) \text{ s.t. } \lambda \geq 0, g(\lambda, \nu) > 0 \\ 0 & \text{otherwise} \end{cases}$$

By weak duality $p^* \geq d^*$. Thus

- If $d^* = +\infty \Rightarrow p^* = +\infty$, means the primal is infeasible.
- If $p^* = 0$, then $d^* = 0$.
- Cannot talk about the case when $p^* = \infty, d^* = 0$ due to duality gap.

Theorem 12.3 (Weak alternatives). *p and condition 1 of d^* cannot be feasible at the same time.*

Convex case: equality should be linear and since need to satisfy Slater, focus on $f_i(x) < 0$.

Theorem 12.4. $(p) : \begin{cases} f_i(x) < 0, i = 1, \dots, m \\ Ax = b \end{cases}$, $\Theta = \begin{cases} \lambda > 0 \\ \lambda \neq 0 \\ g(\lambda, \nu) \geq 0 \end{cases}$. *If f_i's are convex, then strong alternatives hold meaning (p) is feasible iff Θ is infeasible.*

13 Lecture 13 (Oct 4)

13.1 Conic duality

Linear conic duality:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & a_0^T x && (13.1) \\ \text{s.t.} \quad & Ax = b, A_i x - b_i \leq_{K_i} 0, i = 1, \dots, m && (13.2) \end{aligned}$$

Duality:

1. Define a Lagrange multiplier ν for $Ax - b = 0$
2. Define a Lagrange multiplier λ_i for $A_i x - b \leq_{k_i} 0$. (Note that if $A_i x - b_i \in \mathbb{R}^{n_i}$, then $\lambda_i \in \mathbb{R}^{n_i}$.
3. Define $g(\lambda, \nu)$ as:

$$g(\lambda, \nu) = \inf_{x \in D} \mathcal{L}(x, \lambda, \nu) = \inf_{x \in D} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^T f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

Dual optimization of Conic form:

$$\begin{aligned} \max \quad & g(\lambda, \nu) && (13.3) \\ \text{s.t.} \quad & \lambda_i \geq_{K_i^*} 0, i = 1, \dots, m && (13.4) \end{aligned}$$

13.1.1 SDP

$$\min_{\text{Tr}(M_1 X)=a_i, i=1,\ldots,m, X\geq 0} \text{Tr}(M_0 X)$$

M_i's symmetric. Then:

1. $\text{Tr}(M_i X) = a_i$ associate with ν_i
2. $-X \leq 0$ associate with $w \geq 0$ (recall that PSD is self-dual)

$$\begin{aligned}\mathcal{L}(X, \nu, w) &= \text{Tr}(M_0 X) + \sum_{i=1}^{m} \nu_i (\text{Tr}(M_i X) - a_i) + \text{Tr}((-X)w) \\ &= \text{Tr}\left(\left(M_0 + \sum_{i=1}^{m} \nu_i M_i - w\right) X\right) - \sum_{i=1}^{m} \nu_i a_i\end{aligned}$$

Recall that $\text{vec}(X)^T \text{vec}(w) = \text{Tr}(XW)$.
$g(\nu, w) = \min \mathcal{L}(x, \nu, w)$, then

$$X = \begin{cases} -\infty \\ -\sum_{i=1}^{m} \nu_i a_i & \text{if } M_0 + \sum_{i=1}^{m} \nu_i M_i - w = 0 \end{cases}$$

Dual:

$$\max_{\substack{M_0 t + \sum_{i=1}^{m} \nu_i M_i - w = 0 \\ w \geq 0}} -\sum_{i=1}^{m} \nu_i a_i$$

Which indicates the dual of SDP is also SDP.

13.1.2 SDP Canonical

$$\min_{F_0 + x_1 F_1 + \ldots + x_n F_n} a^T x$$

Lagrange multiplier: Z needs to be PSD, then

$$\mathcal{L}(X, Z) = a^T x + \text{Tr}((F_0 + x_1 F_1 + \ldots + x_n F_n) Z)$$

$g(Z) = \min_x \mathcal{L}(X, Z) = \min \sum_{i=1}^{n} (a_i + \text{Tr}(F_i Z)) X_i + \text{Tr}(F_0 Z)$, then:

$$X = \begin{cases} -\infty \\ -\text{Tr}(F_0 Z) \text{ if } a_i + \text{Tr}(F_i Z) = 0, \forall i \end{cases}$$

Dual:

$$\max_{a_i + \text{Tr}(F_i Z) = 0, Z \geq 0} \text{Tr}(F_0 Z)$$

13.1.3 SOCP

$$\min_{\|A_i x+b_i\|_2 \leq c_i^T x+d_i} c^T x$$

Use linear conic duality: 2nd order cone is defined as:

$$K = \{(u,v) \mid u \in \mathbb{R}^n, v \in \mathbb{R}, \|u\|_2 \leq v\}$$

$$\min_{-\begin{bmatrix} A_i x + b_i \\ c_i^T x + d_i \end{bmatrix} \leq_K 0} \quad c^T x$$

Define dual parameters $\begin{bmatrix} u_i \\ v_i \end{bmatrix}$, where $u_i \in \mathbb{R}^{n_i}, v_i \in \mathbb{R}$. Recall the second order cone is self-dual, then $\|u_i\|_2 \leq v_i$.
Thus dual is:

$$\begin{aligned}
\mathcal{L}(x,u,v) &= c^T x - \sum_{i=1}^m \begin{bmatrix} u_i^T & v_i \end{bmatrix} \begin{bmatrix} A_i x + b_i \\ c_i^T x + d_i \end{bmatrix} \\
&= (c^T - \sum_{i=1}^m u_i^T A_i - \sum_{i=1}^m v_i c_i^T) x - \sum_{i=1}^m u_i^T b_i - \sum_{i=1}^m v_i d_i
\end{aligned}$$

Dual: $g(u,v) = \min_X \mathcal{L}(x,u,v)$.

$$g(u,v) = \begin{cases} -\infty \\ -\sum_{i=1}^m u_i^T b - \sum_{i=1}^m v_i d_i \text{ if } c^T - \sum_{i=1}^m u_i^T A_i - \sum_{i=1}^m v_i c_i^T = 0 \end{cases}$$

Dual optimization is:

$$\max_{\substack{c^T - \sum_{i=1}^m u_i^T A_i - \sum_{i=1}^m v_i c_i^T = 0 \\ \|u_i\|_2 \leq v_i}} - \sum_{i=1}^m u_i^T b - \sum_{i=1}^m v_i d_i$$

This is also SOCP.

13.2 KKT Condition for Conic Programming

$$\min_{\substack{f_i(x) \leq_{k_i} 0 \\ h_i(x) = 0}} f_0(x)$$

KKT conditions for conic programming:

1. Primal feasibility: $f_i(x^*) \leq_{K_i} 0, h_i(x^*) = 0$
2. Dual feasibility: $\lambda_i^* \geq_{K_i^*} 0$
3. Complementary slackness:
$$\lambda_i^{*T} f_i(x^*) = 0$$
Note: this **does not** mean either f_i or λ_i is zero.
4. $\nabla_x \mathcal{L}(x^*, \lambda^*, \nu^*) = 0$

13.2.1 KKT for SDP

$$\min_{\substack{\mathrm{Tr}(M_i X) = a_i, i=1,\dots,m \\ X \geq 0}} \mathrm{Tr}(M_0 X)$$

1. $\mathrm{Tr}(M_i X) = a_i, \forall i$ and $X^* \geq 0$
2. $w^* \geq 0$
3. $\mathrm{Tr}((-X^*)w^*) = 0$. Note: since $X^* \geq 0, w^* \geq 0$, then $X^* w^* = 0$.
4. $\mathcal{L}(x, \nu, w) = \mathrm{Tr}(M_0 X) + \sum_{i=1}^m \nu_i (\mathrm{Tr}(M_i X) - a_i) + \mathrm{Tr}((-X)w)$, and
$$\nabla_x \mathcal{L}(x^*, \nu^*, w^*) = 0$$

13.3 Strong duality

Theorem 13.1. *If $\exists x$ in the relative interior of domain of optimization such that the conic inequalities are satisfied in a strict sense:*

$$f_i(x) <_{K_i} 0$$

Then we say, Slater's condition is satisfied. If Slater's is satisfied, then strong duality holds for linear conic program. Also, the dual optimization has a solution and KKT is equivalent to optimality conditions.

Example 13.1. *Slater for SDP:*

$$\begin{cases} Tr(M_i, \bar{X}) = a_i, \forall i \\ \bar{X} > 0 \end{cases}$$

13.3.1 Fentchel's duality

$$\min_{x \in X_1 \cap X_2} f_1(x) - f_2(x)$$

In this scenario, f_1, f_2 are arbitrary function, can replace it by:

$$\min_{y \in X_1, z \in X_2, y=z} f_1(y) - f_2(z)$$

Dualize the constraint $z - y = 0$:

$$g(\nu) = \min_{y \in X_1, z \in X_2} f_1(y) - f_2(z) + \nu^T(z - y)$$

Define:

$$\begin{cases} g_1(\nu) = \sup_{x \in X_1}\{x^T\nu - f_1(x)\} \\ g_2(\nu) = \inf_{x \in X_2}\{x^T\nu - f_2(x)\} \end{cases}$$

Dual:

$$\max_{\nu \in \Lambda_1 \cap \Lambda_2} g_2(\nu) - g_1(\nu)$$

$$\begin{cases} \Lambda_1 = \{\nu \mid g_1(\nu) < +\infty\} \\ \Lambda_2 = \{\nu \mid g_2(\nu) > -\infty\} \end{cases}$$

Theorem 13.2. *If the following assumption is satisfied:*

- f_1 *convex*
- f_2 *concave*
- $\exists$ *a point in relative interior of both* X_1 *and* X_2.

Under the assumptions strong duality holds, so:

$$\inf_{x \in X_1 \cap X_2} f_1(x) - f_2(x) = \max_{x \in \Lambda_1 \cap \Lambda_2} g_2(\nu) - g_1(\nu)$$

Dual has a solution.

14 Lecture 14 (Oct 11)

Nominal Problem:

$$\min_{a_i^T x \le b_i, i=1,\ldots,m} c^T x$$

Robust LP:

$$\min_{\substack{a_i^T x \le b_i, i=1,\ldots,m \\ a_i \in \mathcal{U}_i := \{\hat{a}_i + u : \|u\|_2 \le \rho_i\}}} c^T x$$

Robust counterpart:

$$\min_{\substack{a_i^T x \le b_i, \forall a_i \in \mathcal{U}_i \\ i=1,\dots,m}} c^T x$$

$$\min_{a_i^T x \le b_i} \max_{a_i \in \mathcal{U}_i} c^T x$$

Example (SVM): (x_i, y_i): $x_i \in \mathbb{R}^n$ and $y_i = +1$ or -1. Then predicted $\hat{y}(x) = \text{sgn}(w^T x)$, which implies:

$$\min_{\substack{w \\ \|x_i - \hat{x}_i\|_2 \le \rho_i}} \sum_{i=1}^n (1 - y_i w^T x_i)_+$$

$$\min_w \left[\max_{\|x_i - \hat{x}_i\|_\infty \le \rho_i} \sum_{i=1}^n (1 - y_i w^T x_i)_+ \right]$$

$$\min_w \sum_{i=1}^n \left[\max_{\|x_i - \hat{x}_i\|_\infty \le \rho_i} (1 - y_i w^T x_i)_+ \right]$$

$$\min_w \sum_{i=1}^n \left[(1 - y_i w^T \hat{x}_i + \rho_i \|w\|_1)_+ \right]$$

which is

$$\le \min_w \sum_{i=1}^n \left[(1 - y_i w^T \hat{x}_i \right] + \sum_{i=1}^n \rho_i \|\omega\|_1$$

which is exactly ℓ_1 regularization!
It's not clear if robust optimization is a practical method! But if we can get:

$$x \to \max_{a_i \in \mathcal{U}_i} a_i^T x = \phi_i(x)$$

(note that the robust counterpart is convex). We can replace $\mathcal{U}_i$ as its convex hull anyway.
RC is intractable in general:

$$\max_{\|u\|_\infty \le 1} \|Au\|_2$$

where $A = A_0 + \sum x_i A_i$ is the decision variable.
Simple tractable case:

$$\max_{\substack{a \in \mathcal{U} \\ \mathcal{U} := \{a : \|a - \hat{a}\|_\infty \le \rho\}}} a^T x \le b$$

Note that we can write $a = \hat{a} + \rho u$, $\|u\|_\infty \le 1$, then:

$$\phi(x) = \hat{a}^T x + \rho \max_{\|u\|_\infty \le 1} u^T x = \hat{a}^T x + \rho \|x\|_1$$

15 Lecture 15 (Oct 16)

Robust counterpart:

$$\min_x \max_{u\in\mathcal{U}} f_0(x,u)$$

$$\max_{u\in\mathcal{U}} f_i(x,u) \le 0, i=1,\dots,m$$

Assume $f_i(\cdot,u)$ is convex $\forall u\in\mathcal{U}$.
Linear programming:
Nominal problem:

$$\min_{a_i^T x\le b_i} c^T x$$

Uncertainty set:

$$u_i = a_i - \hat{a}_i, \|u_i\|_\infty \le \epsilon$$

Robust counterpart:

$$\min_x c^T x$$

s.t.

$$(\hat{a}_i + u_i)^T x \le b_i, \forall u_i : \|u_i\|_\infty \le \epsilon$$

which is:

$$\max_{\|u_i\|_\infty\le\epsilon} (\hat{a}_i + u_i)^T x \le b_i$$

Note that:

$$\max \sum_{i=1}^n u_i x_i = \sum_{i=1}^n \max_{|u_i|\le\epsilon} u_i x_i = \epsilon|x_i| = \epsilon\|x\|_1$$

as we can select $u_i^* = \epsilon\text{sgn}(x_i)$. Thus the problem becomes:

$$\min_{\hat{a}_i^T x + \epsilon\|x\|_1 \le b_i} c^T x$$

SVM

$$\min_{w,b} \sum (y_i - \text{sgn}(w^T x_i + b))^2$$

But it's highly non-convex, so instead we have the hinge loss:

$$\min_{w,b} \sum_i (1 - y_i(w^T x_i + b))_+$$

Consider the robust SVM:

$$\min_{w,b} \frac{1}{m} \sum_{i=1}^m (1 - y_i(w^T \hat{x}_i + b + \epsilon\|w\|_1))_+$$

$$\le \min_{w,b} \frac{1}{m} \sum (1 - y_i(w^T \hat{x}_i + b))_+ + \epsilon\|w\|_1$$

$u_i = x_i - \hat{x}_i$, $\|u_i\|_\infty \le \epsilon$.

15.1 Intersect two sets of uncertainty

$a^T x = b$, and $\mathcal{A} = \{a : \|a\|_{(1)} \le \alpha, \|a\|_{(2)} \le \epsilon\}$, want to find $\max_{a\in\mathcal{A}} a^T x$.

$$\min_x \max_{\substack{y_i \in \{-1,1\} \\ \|y-\hat{y}\|_1 \le 2k}} \sum_{i=1}^m (1 - y_i(w_i^T x + b))_+$$

16 Lecture 16 (Oct 18)

nominal: LP: $\min_x c^T x : a_i^T x \le b_i, i = 1, \dots, m$
uncertainty: $a_i \in \mathcal{U}_i, i = 1, \dots, m$
robust counterpart: $\min_x c^T x : a_i^T x \le b_i, \forall a_i \in \mathcal{U}_i$, which is equivalent to: $\min_x c^T x : \max_{a_i \in \mathcal{U}_i} a_i^T x \le b_i$
simple case: $\mathcal{U} = \{a \in \mathbb{R}^n : \|a - \hat{a}\| \le \rho\}$, we know $\psi_u(x) = \max a^T x : \|a - \hat{a}\| \le \rho = \hat{a}^T x + \rho \|x\|_*$.

17 Lecture 17 (Oct 23)

How to compute a function $\psi(x) = \max_{a\in\mathcal{A}} a^T x \le b$ as $\min_{(u,x)\in\mathcal{C}} c^T x + d^T u \le b \iff \exists u : c^T x + d^T u \le b$.

17.1 Chance Programming

$a^T x \le b, a \sim \mathcal{N}(\hat{a}, \Sigma)$.

$$\mathbb{P}\{a : a^T x \le b\} \ge 1 - \epsilon$$

which is called the "chance constraints" on x. ϵ: reliability level.

$$\hat{a}^T x + \kappa(\epsilon) \|\Sigma^{1/2} x\|_2 \le b$$

Deterministic interpretation: assume $(a - \hat{a})^T \Sigma^{-1} (a - \hat{a}) \le \kappa$, $a = \hat{a} + \kappa \Sigma^{-1/2} u$, $\|u\|_2 \le 1$. If we want $a^T x \le b, \forall a$ in this set.

17.1.1 Large deviation theory

Distributional robustness:

$$\inf_{p\in\mathcal{P}} \mathbb{P}_{\mathcal{P}}\{a : a^T x \le b\} \ge 1 - \epsilon$$

Generalized Chebyshev's bound:

$$\mathbb{P}\{a^T x \le b\} \ge 1 - \epsilon, \forall p \text{ s.t. } \mathbb{E}[a] = \hat{a}, \mathrm{Var} = \Sigma$$

17.1.2 Approaches

$a^T x \leq b$, a random:

- $$\mathbb{P}_{\mathcal{P}}\{a : a^T x \leq b\} \geq 1 - \epsilon$$
  distribution $\mathcal{P}$ known.

- $$\inf_{p \in \mathcal{P}} \mathbb{P}_{\mathcal{P}}\{a : a^T x \leq b\} \geq 1 - \epsilon$$
  distribution $\mathcal{P}$ unknown.

18 Lecture 18 (Oct 25)

Assume distribution of a is known. i.e. $a \sim \mathcal{N}(\hat{a}, \Gamma)$: mean $\hat{a}$, Γ: covariance matrix. Then

$$\mathbb{P}\{a : a^T x \leq b\} \geq 1 - \epsilon, \epsilon \ll 1$$

$$Var(a^T x) = \mathbb{E}(x^T (a - \hat{a})(a - \hat{a})^T x) = x^T \Gamma x$$

Assuming distribution of a is partially known,

$$\inf_{p \in \mathcal{P}} \mathbb{P}_{\mathcal{P}}(a : a^T x \leq b) \tag{18.1}$$

e.g. $a \sim (\hat{a}, \Gamma)$ where $\mathbb{E}a = \hat{a}$ and $\mathbb{E}(a - \hat{a})(a - \hat{a})^T = \Gamma$. Then 18.1 is equivalent to:

$$\hat{a}^T x + \kappa(\epsilon)\sqrt{x^T \Gamma x} \leq b$$

where $\kappa(\epsilon) = \sqrt{\frac{1-\epsilon}{\epsilon}}$.
Approach based on moment generating function:

$$\sup_{p \in \mathcal{P}} \mathbb{P}_{\mathcal{P}}(a : a^T x \leq b)$$

Define a "generator" γ such that

$$\begin{cases} \gamma(s) \to 0, s \to -\infty \\ \gamma(0) \geq 1 \end{cases}$$

$\mathbb{E}\gamma\left(\omega^T \begin{bmatrix} x \\ 1 \end{bmatrix}\right) \geq \mathbb{P}\left(\omega^T \begin{bmatrix} 0 \\ 1 \end{bmatrix} > 0\right)$, $\mathbb{P}(\bar{\omega} > 0) = \mathbb{E}(1_+(\bar{\omega}))$, we have:

$$\sup_{p \in \mathcal{P}} \mathbb{E}(\gamma(a^T x - b)) = \alpha \Psi \left(\frac{(x, 1)}{\alpha}\right)$$

$x \to f(x)$ convex, and:

$$(x, \alpha) \to \begin{cases} \alpha f(x/\alpha) : x \in \text{ dom}(f), \alpha > 0 \\ +\infty \text{ otherwise} \end{cases}$$

Special case: $\gamma(s) = \exp(s)$, consider the constraint

$$\sup_{p \in \mathcal{P}} \mathbb{P}_p(a^T x - b > 0) \leq \epsilon$$

Assumption: a is zero-mean, support of a is in $[-1, 1]^n$. a_i's are independent of each other. Set $\alpha = 1$, calculating:

$$\mathbb{E}[\exp(a^T x + b)] = \prod \exp a_i x_i e^{-b} \leq \cosh(t) \leq e^{t^2/2}$$

Which is induced by considering the function $f(s) = e^{ts} - s \sinh(t)$, since it's zero mean, we have:

$$\begin{aligned} \mathbb{E}(e^{ts}) &= \int e^{ts} dQ(s) = \mathbb{E}(f(s)) \\ &= \max_{s \in [-1,1]} f(s) = \cosh(t) \leq e^{t^2/2} \end{aligned}$$

$$\mathbb{E}[\exp(a^T x + b)] \leq \sum_i \exp\left(\frac{1}{2} \Sigma x_i^2 - b\right) \leq \epsilon$$

Assume $a = \hat{a} + \rho u$, u: random, u_i's independent, $\mathbb{E}u = 0$, $u \in [-1, 1]$. Use $\gamma(s) = \max(1 + s, 0)$ (this gives you the best bound). In order for you to ensure

$$\mathbb{P}\{a : a^T x > b\} \leq \epsilon$$

for all distribution on $a \in p$, sufficient condition:

$$\min_p \beta + \frac{1}{\epsilon} \mathbb{E}(a^T x - b - \beta)_+ \leq 0$$

19 Lecture 19 (Oct 30)

19.1 Moment bound on probabilities

1. $X \geq 0$: random variable:

$$\mathbb{P}(X \geq a) \leq \mathbb{E}(X)/a$$