ENGG 5501 Foundations of Optimization
Term 1, 2017-18

Instructor: Prof. Anthony Man-Cho So
Scribe: Zihao FU

2017.9.4

Problem

(P) $\min f(x)$
s.t. $x \in X \subseteq \mathbb{R}^n$

$x = (x_1, \cdots, x_n)$ called decision vector

is a column vector

$f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ objective function

X : feasible region

s.t. : such that

Defin: (Global minimize)

$x^* \in X$ s.t. $f(x^*) \leq f(x) \quad \forall x \in X$

x^* also called optimal solution

$f(x^*)$ is called optimal value

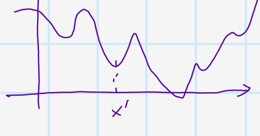
x^* is important than value

Defin: (Local minimize)

$x' \in X$, s.t. $f(x') \leq f(x) \quad \forall x \in X \cap B(x', \varepsilon)$

$B(x', \varepsilon) = \{ x \mid \|x - x'\|_2 \leq \varepsilon \}$

for some $\varepsilon > 0$



Examples

(1) $X = \mathbb{R}^n$ (unconstrained optimization)

(2) X is discrete set (discrete optimization)

e.g. $X = \{0, 1\}^n$

all fessible points are local minize

(3) Linear Programming (LP)

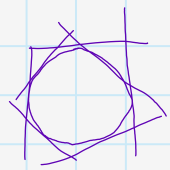
$$f(x) = c_1 x_1 + c_2 x_2 + \cdots c_n x_n = c^T x$$

(linear function will always pass (0, 0))

$$g(x) = b^T x + a \quad \text{(affine)}$$

$$X = \{x \in \mathbb{R}^n,\ a_i^T x \leq b_i,\ i = 1, \cdots, m\}$$

m should be finite, otherwise we can make X a circle



$$X = \{x : \underset{m\times n}{A} \underset{n\times 1}{x} \leq \underset{n\times 1}{b}\}$$

(a more compatable way to write)

$$A = \begin{bmatrix} \cdots & a_1^T & \cdots \\ \cdots & a_2^T & \cdots \\ & \vdots & \\ \cdots & a_m^T & \cdots \end{bmatrix}$$

Vector comparision

$$u \leq v \Leftrightarrow u_i \leq v_i \ \forall i$$

(4) Quadiatic Programming

X is as in LP (constrain is linear)

$$f(x) = \sum_{i,j=1}^{n} Q_{ij} x_i x_j$$

homogenious : no linear term

inhomogenious : have linear term

$$= \underset{1\times n}{x^T} \underset{n\times n}{Q} \underset{n\times 1}{x}$$

Rmk WLOG (without loss of generality)

Q is symmetric $(Q = Q^T)$

$$x^T Q x = x^T \left(\frac{Q+Q^T}{2}\right) x \quad ?$$

(5) Semidefinite Programming (SDP)

Def'n: Q is $n\times n$ symmetric, we say ($\to Q \in S^n$)

Q is positive semidefinite if $x^T Q x \geq 0 \ \forall x$

notation $Q \succeq 0$ $Q \in S_+^n$

$x^T Q x \geq 0 \ \forall x \iff$ all eigenvalues of $Q \geq 0$

why use symmetry matrix?

because eigenvalues of symmetry matrix is real number

Problem: SDP
(P)

$$\min b^T y$$

$$\text{s.t. } \underset{S^m}{C} - \sum_{i=1}^{m} y_i A_i \underset{S^m}{\succeq 0}$$

← Linear matrix inequality (LMI)

example

$$\begin{bmatrix} c_1 & & \\ & c_2 & \\ & & \ddots \\ & & & c_m \end{bmatrix} - y \begin{bmatrix} a_1 & & \\ & a_2 & \\ & & \ddots \\ & & & a_m \end{bmatrix} \succeq 0$$

$$\Rightarrow \begin{array}{l} c_1 - y a_1 \geq 0 \\ c_2 - y a_2 \geq 0 \\ \vdots \\ c_m - y a_m \geq 0 \end{array}$$

2017.9.6

SDP another example (S)

Consider $A_i, C \in S^n$ （对称 $C=C^T$）

$$\min \ C \bullet Z \triangleq \sum_{i,j=1}^{n} C_{ij} Z_{ij} = tr(CZ)$$

所有对应项相乘，再求和

$$\text{s.t.} \ A_i \bullet Z = b_i \quad i=1,\dots,m.$$

$$Z \succeq 0$$

优化中需要把问题转化为标准形式，下面把这个example转化为上面标准型

Convert Problem (S) into (P)

$n=2;\ m=1$

$$Z = \begin{bmatrix} z_{11} & z_{12} \\ z_{12} & z_{22} \end{bmatrix} \leftarrow Z\text{是对称、} \quad y=(z_{11}, z_{12}, z_{22})$$

$$C = \begin{bmatrix} \cancel{z_{11}}\,C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \quad b=(C_{11}, 2C_{12}, C_{22}) \Rightarrow b^T y = C \bullet Z$$

$$A \cdot Z = b \iff A_{11} z_{11} + 2A_{12} z_{12} + A_{22} z_{22}$$

注意到我们要求一个等式，而SDP一般是不等式，那么等式可以通过 $a \geq 0$ 且 $a \leq 0$ 来构造

$$0 \preceq \underbrace{\left[\begin{array}{cc|c} b & & \\ & -b & \\ \hline & & 0 \end{array}\right]}_{C} - z_{11} \underbrace{\left[\begin{array}{cc|c} A_{11} & & \\ & -A_{11} & \\ \hline & & -1 \end{array}\right]}_{A1} - z_{12} \underbrace{\left[\begin{array}{cc|cc} 2A_{12} & & & \\ & -2A_{12} & & \\ \hline & & & -1 \\ & & -1 & \end{array}\right]}_{A2} - z_{22} \underbrace{\left[\begin{array}{cc|c} A_{22} & & \\ & -A_{22} & \\ \hline & & -1 \end{array}\right]}_{A3}$$

（(P) 问题中的 C, A_1, A_2, A_3 构造出来）

以上演示了如何从(S)转化为(P)，也可以从(P)转为(S)，留做思考

分块矩阵当反对角为0时，正定条件：只需左上角正定。http://www.cis.upenn.edu/~jean/schur-comp.pdf

Example: Air Traffic Control

n planes try to land in an airport

given: i^{th} plane's arrival time interval $[a_i, b_i]$

plane come in order

Goal of the tower:

To max the minimum metering time

最大化最小间隔时间

decision variable:

Let t_i = time of the i^{th} landing

$$\max \min_{1\le i\le n-1}\{t_{i+1}-t_i\}$$

$$\text{s.t. } a_i \le t_i \le b_i \quad i=1,\dots,n$$
$$t_i \le t_{i+1} \quad i=1,\dots,n-1$$

constraints are linear

先看那个min

Let $z = \min_{1\le i\le n-1}\{t_{i+1}-t_i\}$

步骤：

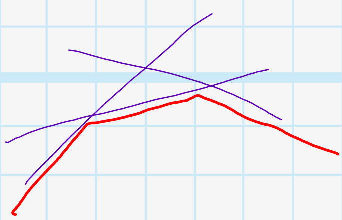
① 把等式改成不等式

② min可以去掉

$$\max z$$
$$\text{s.t. } z \le \min_{1\le i\le n-1}\{t_{i+1}-t_i\}$$
$$a_i \le t_i \le b_i$$
$$t_i \le t_{i+1}$$

≤包含=，而我们一会会对z求max

多个线性函数求min



z 比最小的那个小，意味着比所有的项都小、所以等价于：

$z \le \{t_{i+1} - t_i\}$ $i=1,\cdots,n-1$

max min 可转换，但 min min 不行

min max

Example: Data fiting Problem

Data pairs: (a_i, b_i) $i=1,\cdots,m$

$a_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$

ε_i 噪声

$a_i \longrightarrow \boxed{f} \longrightarrow \oplus \longrightarrow b_i$

Assume: f is affine: $f(y) = x^T y + t$

x在这里是系数
y是自变量

$f(a_i) = a_i^T x + t$

$b_i = f(a_i) + \varepsilon_i$

Goal: find x, t, s.t.

b_i and $f(a_i)$ are as close as possible $\forall i$

Approach 1: (MAD) $\min \sum_{i=1}^{m} |b_i - \underbrace{(a_i^T x + b)}_{f(a_i)}|$ Minimum Absolute Deviation

$x \in \mathbb{R}^n$, $t \in \mathbb{R}$

Approach 2: (LS) $\min \sum_{i=1}^{m} |b_i - (a_i^T x + t)|^2$ Least squares

如何转化为优化问题？

解: MAD: $\min \sum_{i=1}^{m} z_i$

$$\text{s.t. } z_i \geq |b_i - (a_i^T x + t)|$$ ① 把等号变成不等号

$$\Downarrow$$

$$z_i \geq b_i - (a_i^T x + t) \geq -z_i$$

现有的变量: $x \in \mathbb{R}^n$, $t \in \mathbb{R}$, $z \in \mathbb{R}^m$ 多引入了 z

以上演示了去掉绝对值的技巧

LS: 可以直接求导,然后让导数=0

留做作业

答案:

$$\begin{bmatrix} x^* \\ t^* \end{bmatrix} = (\bar{A}^T \bar{A})^{-1} \bar{A}^T b$$

如果不可逆,就用伪逆

$$\bar{A} = \begin{bmatrix} -a_1^T- & 1 \\ -a_2^T- & 1 \\ \vdots & \\ -a_m^T- & 1 \end{bmatrix} \in \mathbb{R}^{m \times (n+1)}$$

← 这列是给t用的

$a_i^T x + t = (a_i, 1)^T (x, t)$

高维统计假设之一就是数据远比特征多.

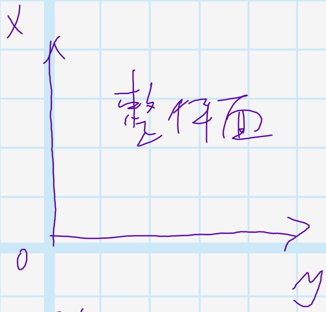
现在很多情况下特征比数据多(例如基因),这时候通常做法是加正则项.

2017.9.11

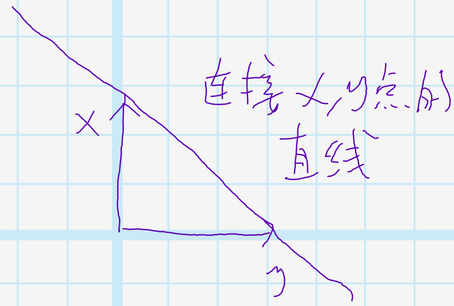
以下定义几个集合

Let $S \subseteq \mathbb{R}^n$ be a set

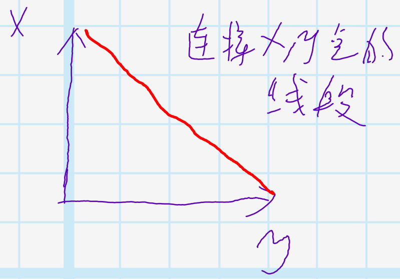
(1) S is linear
if $\alpha x + \beta y \in S$ whenever
$x, y \in S$; $\alpha, \beta \in \mathbb{R}$ β取0的时候，一定过原点



(2) S is affine
if $\alpha x + (1-\alpha) y \in S$ whenever
$x, y \in S$, $\alpha \in \mathbb{R}$



(3) S is convex
if $\alpha x + (1-\alpha) y \in S$ whenever
$x, y \in S$; $\alpha \in [0,1]$ x,y的系数都≥0



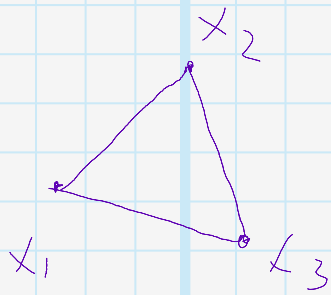
Defi. Affine combination
$x_1, \dots, x_m \in \mathbb{R}^n$
$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_m x_m$ is
$\begin{cases} \text{affine combination of } x_1, \dots, x_m \text{ if } \sum_{i=1}^{m} \alpha_i = 1, \\ \text{convex combination of } x_1, \dots, x_m \text{ if } \sum_{i=1}^{m} \alpha_i = 1,\ \alpha_i \geq 0 \end{cases}$

例:

x_2

x_1 affine combination is the whole plane

x_3



convex combination

Let $S \subseteq \mathbb{R}^n$ be a set

prop1: The following are equivalent:

(1) S is affine (定义只说了两点)

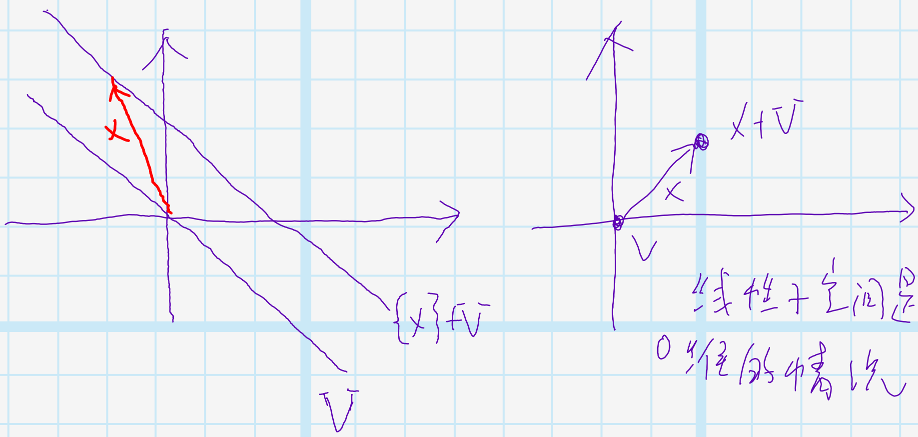
(2) Any combination of points in S belongs to S ←有可能还有多个点...

(3) S is of the form $S = \{x\} + \bar{V} = \{x + v : v \in \bar{V}\}$

for some $x \in \mathbb{R}^n$, $\bar{V}$ is a linear subspace in $\mathbb{R}^n$

仿射空间可由一个向量加上一个线性子空间构成

e.g.



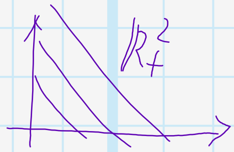
Prop2. The following are equivalent

(1) S is convex

(2) Any convex combination of points in S belongs to S

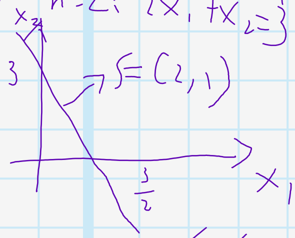
Example: (convex sets)

① Non-negative orthant $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n ; x_i \geq 0 \ \forall i\}$



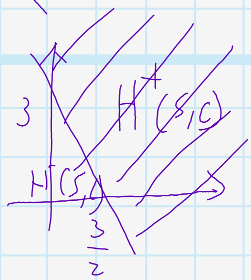
② Hyperplane $H(s,c)=\{x\in\mathbb{R}^n : s^Tx=c\}$ 线性方程的简写

$n=2$: $\{2x_1+x_2=3\}$



③ Half-space $H^+(s,c)=\{x\in\mathbb{R}^n : s^Tx\geq c\}$

$H^-(s,c)=\{x\in\mathbb{R}^n : s^Tx\leq c\}$

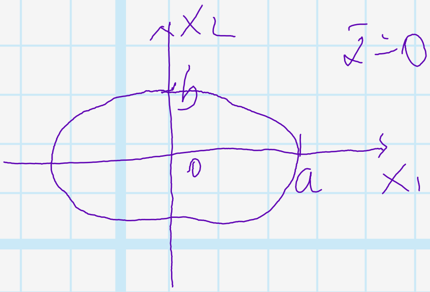


H^+是法向量指向的那侧

④ Ellipsoid,

$E(\bar{x},Q)=\{x\in\mathbb{R}^n : (x-\bar{x})^TQ(x-\bar{x})\leq 1\}$ $Q>0$

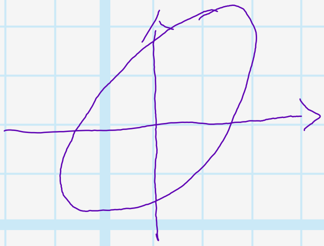
Q is positive definite: ie. $x^TQx>0$ $\forall x\neq 0$; all eigenvalues>0



$\bar{x}=0$

$\frac{x_1^2}{a^2}+\frac{x_2^2}{b^2}\leq 1$ 写成矩阵形式:

$$[x_1\ \ x_2]\underbrace{\begin{bmatrix}\frac{1}{a^2} & \\ & \frac{1}{b^2}\end{bmatrix}}_{Q}\begin{bmatrix}x_1\\x_2\end{bmatrix}\leq 0$$



arthogunal matrices $U^TU=I$

可用于旋转，∴每行跟其它行都正交

$\begin{bmatrix}\cos\theta & \sin\theta\\ -\sin\theta & \cos\theta\end{bmatrix}$ 旋转矩阵正交

$x^TU^T\underbrace{U\,Q\,U^T}_{\tilde{Q}}\underbrace{U^TU}_{=I}x$ $\tilde{Q}=UQU^T$

$y=Ux \Leftrightarrow x=U^Ty$

$=y^T\tilde{Q}y$

spectral thm

对称矩阵都能分解成

对角阵$\tilde{Q}$和正交基U

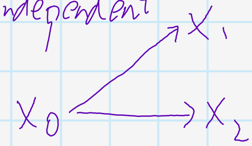
⑤ Simplex

$$\Delta=\left\{\sum_{i=0}^{n}\alpha_i x_i\in\mathbb{R}^{n} \quad \sum_{i=0}^{n}\alpha_i=1,\ \alpha_i\geq 0\right\}$$

相等

where $x_0, x_1, \cdots, x_n$ affinely independent

i.e. $x_1 - x_0, x_2 - x_0, \cdots, x_n - x_0$ linearly independent



e.g., $n=2$

$x_0 = 0, x_1 = e_1, x_2 = e_2$

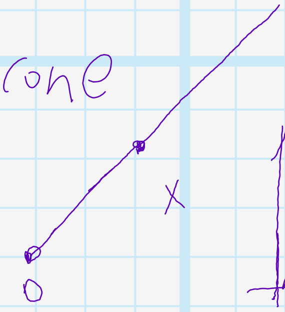
simplex有时候跟球的维度一致.

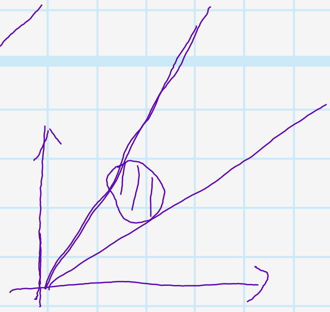
⑥ Convex Cone

Def: $K \subseteq \mathbb{R}^n$ is called a cone

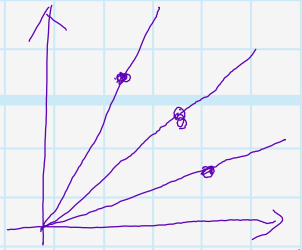
if $\{\alpha x : \alpha \geq 0\} \subseteq K$

whenever $x \in K$





Def: Convex cone $\triangleq$ cone that convex



not convex!

e.g. $S_+^n = \{X \in S^n : X \succeq 0\}$

is a convex cone

∵PSD矩阵乘一个系数仍然是PSD

这个例子说明cone定义里的x不一定是向量

也可以是矩阵.

Observe:

把这句话换成convex也成立

arbitrary intersection of affine (resp. convex) sets

is affine (resp. convex)

Given any S, there's a smallest (by inclusion)

a affine set that contains S, This is

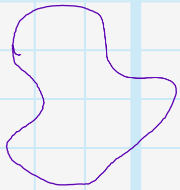
convex

called affine hull of S $(\mathrm{aff}(S))$

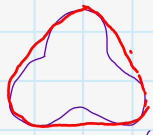
convex

怎么取smallest? 把所有满足条件的放到一起取交。

S:



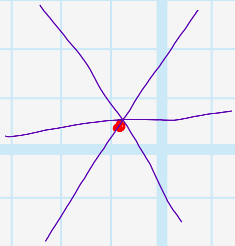
$\mathrm{aff}(S) = \mathbb{R}^2$



convex hull

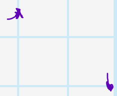
用一根橡筋崩在上面

S:

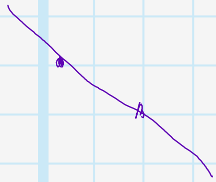


$\Rightarrow \mathrm{aff}(S) = S$

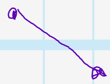
S

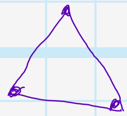


affine hull:



convex hull:





2017.9.13

Alternative characterization

① $\mathrm{aff}(S)$ = set of all affine combinations of points in S

② $\mathrm{conv}(S)$ = - - - - - - - convex - - - -

由已有的S中的点进行扩张

可以理解为 affine/convex hull 对相应的 affine/convex 组合封闭

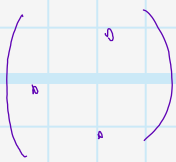
下面我们定义集合的维度

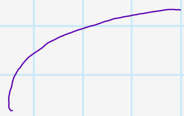
Defin. Let $S \subseteq \mathbb{R}^n$, Then

$$\dim(S) \triangleq \dim(\mathrm{affin}(S))$$

S的维度等于其 affin hull 的维度

∵仿射空间都是一个向量加上一个 linear subspace, ∴维度就是那个 linear subspace 的维度, 而线性空间维度很好定义

Exp. dim  $=2$

S:  $\dim(S)=2$

和参数表示定义的维度不同

Convexity-Preserving Operation

两个convex set经过操作仍是convex

$A \cup B$ not preserving convexity

$A \cap B$ preserve convexity

还有一种操作 affine map 也保凸

Def. Affine Maps

We say $A: \mathbb{R}^n \to \mathbb{R}^m$ is affine if

$$A(\alpha x_1+(1-\alpha)x_2)=\alpha A(x_1)+(1-\alpha)A(x_2)$$

for any $x_1, x_2 \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$

Fact: A is affine iff $\exists A_0 \in \mathbb{R}^{m\times n}$ and $y_0 \in \mathbb{R}^m$

s.t. $A(x) = A_0 x + y_0$

Linear function + constant

这个fact证明比较困难.

定义中对A的要求极强.
因此, A的形式实际上已经固定了

Prop: If A is an affine map and S is a convex set, then $A(S) \triangleq \{A(x) : x \in S\}$ is convex

e.g. ① Rigid motion A_0 orthognal

② Orthogonal projection

用一道光去照一个convex set, 影子也是convex

这个e.g.是上面那个定义的几种特殊形式

投影是仿射变换

Example: Convexity of Ellipsoids.

$$E(\bar{x}, Q) = \{x : (x-\bar{x})^T Q (x-\bar{x}) \le 1\}$$

Claim: $\exists$ affine map A, s.t.

$$A(B(0,1)) = E(\bar{x}, Q)$$

一个球经过放射变换，得到了一个椭球
都是凸的

存在一个affine map把一个标准球，映射成一个指定的椭圆。

下面看这个Affine Map怎么构造

set $y_0 = \bar{x}$, Note $Q > 0 \Rightarrow Q^{-1} > 0$

证明：Q的所有eigenvalue>0
∴ Q^{-1}的所有……>0

$$Q^{-1} = U \Lambda U^T \quad \text{谱分解}$$
$$= U \Lambda^{1/2} \Lambda^{1/2} U^T$$

$= \underbrace{U \Lambda^{1/2} U^T}\ \underbrace{U \Lambda^{1/2} U^T}$ 　　矩阵的$\frac{1}{2}$次方定义

$= Q^{-1/2}\ Q^{-1/2}$ 　　Q是对称的

$Q^{1/2}$: cholesky factor

Set $A_0 = Q^{-\frac{1}{2}}$

下面来验证 $A(B(0,1)) = E(\bar{x}, Q)$

证集合相等，只用证 $A \subseteq B$ 且 $B \subseteq A$，

"$\subseteq$" Take $x \in B(0,1)$ show $A(x) \in E(\bar{x}, Q)$

$(A(x) - \bar{x})^T Q (A(x) - \bar{x})$ 　　其中 $A(x) = Q^{-1/2} x + \bar{x}$

$= (Q^{-1/2} x)^T Q (Q^{-1/2} x)$

∵对称 ∴$(Q^{-1/2})^T = Q^{-1/2}$

$= x^T Q^{-1/2} Q^{1/2} Q^{1/2} Q^{-1/2} x$

$= x^T x$

∵ x在单位球中 ∴ $x^T x \le 1$

根据 conv(S) 的定义：

S 　　$x \in \mathrm{conv}(S)$

Then, $\exists x_1, x_2, \cdots, x_k \in S$

s.t. $x = \sum_{i=1}^{k} \alpha_i x_i$ 　convex combination

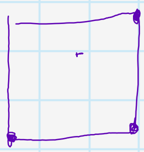
x可以是第1~k个点的凸组合

那么问题是 k 最小取多少？

Thm: (Carathéodory)

Let $S \subseteq \mathbb{R}^n$, Any $x \in \mathrm{conv}(S)$ can be expresed as a convex combination of $\le n+1$ points in S

e.g.

$S =$  $\subseteq \mathbb{R}^2$, 至少要三个点。

在线性空间中，可以固定几个基，然后所有的点都可以由基表示。但是在 convex 中不同，找不到固定的基，每个点的基可能不同

那么在什么情况下，可以选择固定的基？

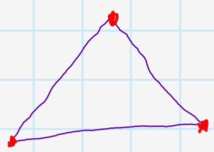
Def'n. Let $\emptyset \neq S \subseteq \mathbb{R}^n$, be convex

We say x is an extreme point of S

if (1) $x \in S$

(2) $\nexists\, x_1, x_2$, s.t. $x_1 \neq x_2$ and $x = \frac{x_1 + x_2}{2}$

e.g. $S =$



极点在顶点上

$S = B(0,1)$ the whole surface is extreme points

Claim: $\|x\|_2 = 1$ are the extreme points

Pf. Let $x_1, x_2 \in B(0,1)$, s.t. $x = \frac{x_1 + x_2}{2}$

$$1 = \|x\|_2^2 = \frac{1}{4}\|x_1 + x_2\|_2^2 = \frac{1}{2}\|x_1\|_2^2 + \frac{1}{2}\|x_2\|_2^2 - \frac{1}{4}\|x_1 - x_2\|_2^2$$

$$1 \leq \frac{1}{2} + \frac{1}{2} - \frac{1}{4}\|x_1 - x_2\|_2^2$$

$\therefore \|x_1 - x_2\|_2^2 = 0 \quad \therefore x_1 = x_2$ 矛盾

然而并不是所有 convex set 都有极点，例如 $\mathbb{R}^2$

Thm (Minkowski) Let S be convex and compact

Then $S = \text{conv}(\underbrace{\text{ext}(S)}_{\text{set of extreme points}})$

下面来定义紧集：

Defin: Let $S \subseteq \mathbb{R}^n$, We say S is compact if

(1) S is bounded ($\exists M>0$ s.t. $\|x\|_2 \le M$, $\forall x \in S$)

S can be put inside a big ball.

(2) S is closed if $x_k \in S$ and $x_k \to x$ then $x \in S$

对极限封闭

紧集就是有界闭集?

几个反例:

如果 S 仅仅 convex and closed but not bounded

the whole plane have no extreme points,

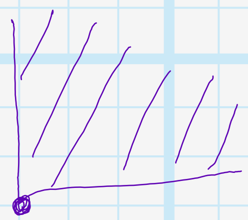
So it cannot be expanded by extrem points

S 仅仅 convex bounded but not closed

开球也没有 extreme points, 所以也不能由 extreme point 构成.

找一个反例 $\exists$ convex S, s.t.

$\mathrm{ext}(S) \neq \emptyset \quad S \neq \mathrm{conv}(\mathrm{ext}(S))$



只有一个 extreme points 不能张成.

2017.9.25

Topological Properties

Let $S \subseteq \mathbb{R}^n$

Defin: The interior of S is

$$int(S) = \{x \in S : B(x, \epsilon) \in S, \text{for some } \epsilon\}$$

就是度量空间上内部的定义.

e.g. $S = [0,1]$ $\quad int(S) = (0,1)$

注意 $S \subset \mathbb{R}$，如果 $S \subseteq \mathbb{R}^n$ 则没有内点..

若 $S = [0,1] \subseteq \mathbb{R}^2$，则 $int(S) = \phi$

我们希望不考虑这个维数的问题.

Defin. The relative interior of S is

$$relint(S) = \{x \in S : B(x, \varepsilon) \cap aff(S) \subseteq S \text{ for some } \varepsilon > 0\}$$

e.g. $S = [0,1] \subseteq \mathbb{R}^n$ $\quad aff(S) = \mathbb{R}$,

$relint(S) = (0,1)$

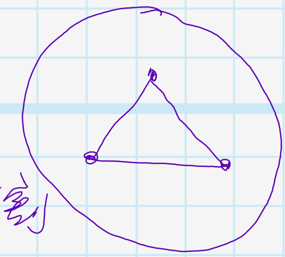
$S = \{x\}$ 单点. $\quad int(S) = \phi$

$relint(S) = S$

Prop: Let S be non-empty, convex, Then,

$\mathrm{relint}(S) \neq \phi$

证明：若有一个凸集，总能找到一些点构成 simplex，接着去证 simplex 有内点



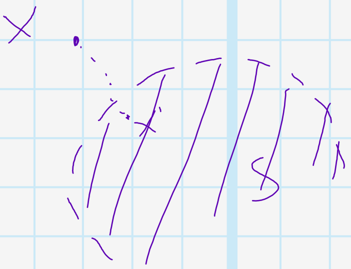
下面讲两个重要问题，第一个是投影，一个是分割

$\phi \neq S \subseteq \mathbb{R}^n,\ x \in \mathbb{R}^n \setminus S$

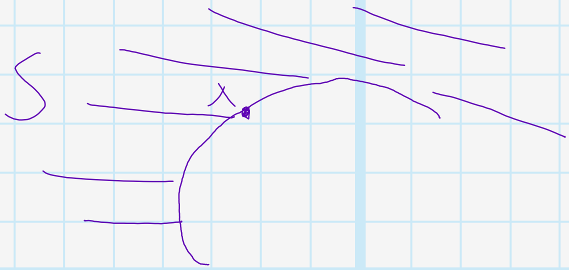
Q: Projection Problem

$\exists?\ z \in S$ s.t. z is closest to x among all points in S

对于一般任意集合的话是不成立的



如果 S 是闭的，那么存在这个点 但不一定唯一



在 S 中找不到离 x 最近的点

Thm: Let S be non-empty, closed, convex, Then $\forall x \in \mathbb{R}^n\ \exists$ unique $z^* \in S$ that closest to x

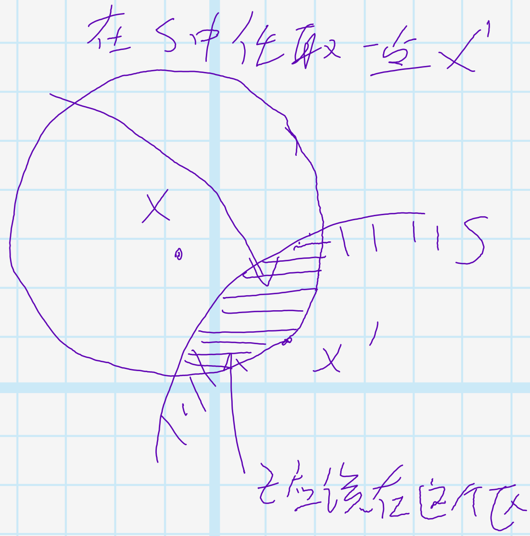
先证存在，再证唯一

Pf: (Existence)

等价 $\min\limits_{z \in S} \|x - z\|_2^2$ 有解

Let $T = \{ z \in S : \|z - x\|_2 \le \|x - x'\|_2 \}$

x'取定后是常数



$$\min_{z \in S} \|z - x\|_2^2 = \min_{z \in T} \|z - x\|_2^2$$

Observe:

① T is compact

② Weierstrass' thm:

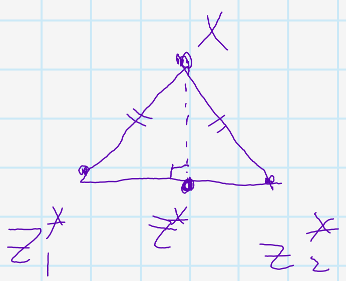
min continuous function over a compact set

$\Rightarrow$ minimizer exists

如果只是闭的话不成立, $\min\limits_{x \geq 0} \frac{1}{x}$

Pf: (Uniqueness)

假设存在 z_1^*, z_2^*



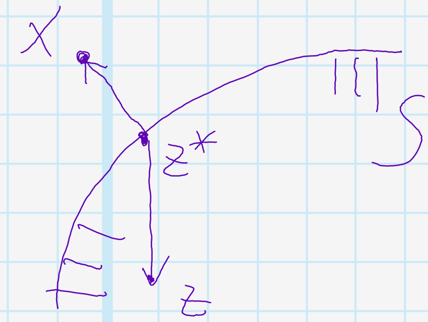
∵ S是凸的 ∴ $z^* \in S$, ∴ z^*的距离更近

Notation: $\Pi_S(x)$ = projection of x onto S

$\uparrow$
$\mathbb{R}^n$

Prop: Give $x \in \mathbb{R}^n$, $z^* = \Pi_S(x)$ iff

$(z - z^*)(x - z^*) \le 0 \quad \forall z \in S$



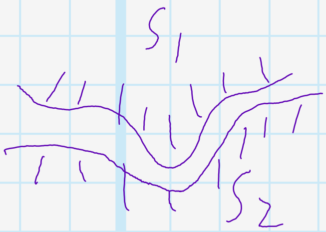
两向量夹角 大于 90°

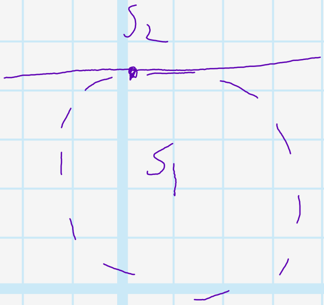
Separation

$S_1 \cap S_2 = \phi$

$\not\parallel \quad \not\parallel$

$\phi \quad \phi$



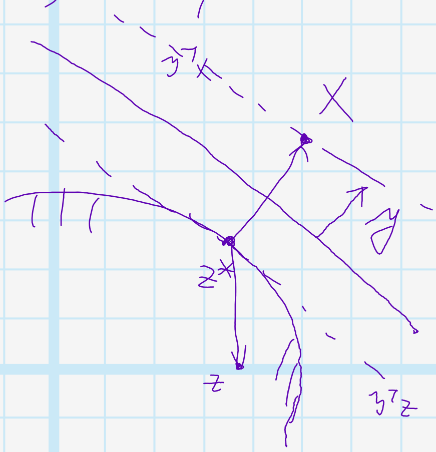


Thm (Point-Set Separation)

Let S be non-empty, closed convex, $x \notin S$

Then, $\exists y \in \mathbb{R}^n$ s.t.

$$\max_{z \in S} y^T z < y^T x$$



Pf: By assumptions on S,

We have $z^* = \Pi_S(x)$ exists and unique

Set $y = x - z^*$. Note $(z - z^*)^T y \le 0 \quad \forall z \in S$

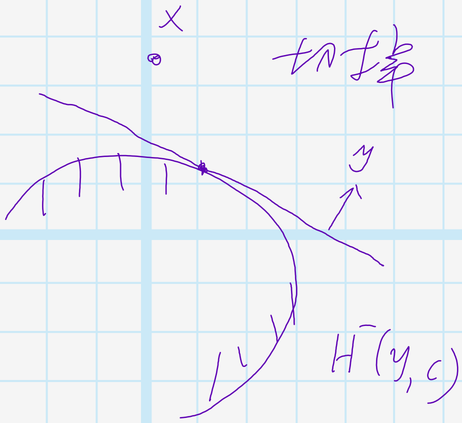
$\Rightarrow \quad y^T z \le y^T z^* = y^T x + y^T (z^* - x)$

$= y^T x - \|y\|_2^2$

$$\max_{z \in S} y^T z \le \underbrace{y^T x - \|y\|_2^2}_{\text{与z无关}} < y^T x$$

Thm, A closed convex S is the intersection of all halfspaces containing S

Pf: $\phi \subsetneq S \subsetneq \mathbb{R}^n$, let $x \notin S$,



Then, by separation thm,

$\exists y$: $\max_{z \in S} y^T z < y^T x$

i.e., $S \subseteq H^-(y, c) = \{z : y^T z \le c\}$

$c = \max_{z \in S} y^T z$

but $x \notin H^-(y, c)$

如果还存在x不在S中，则用这个x来继续切

对应到 convex function 上便是 convex function

可以用很多个 linear function 的min来构成

2017.9.27

Recall:

Thm (Point-Set Sep)

Let S be non-empty, closed, convex $x \notin S$,

Then $\exists y$ $\max_{z \in S} y^T z < y^T x$

Thm. A closed convex set is intersection of all half spaces containing it.

下面把 Point-set Sep Thm 推广

Thm: (Set-Set Sep)

Let S_1, S_2 be non-empty, closed convex,

$S_1 \cap S_2 = \phi$ Suppose S_2 is bounded

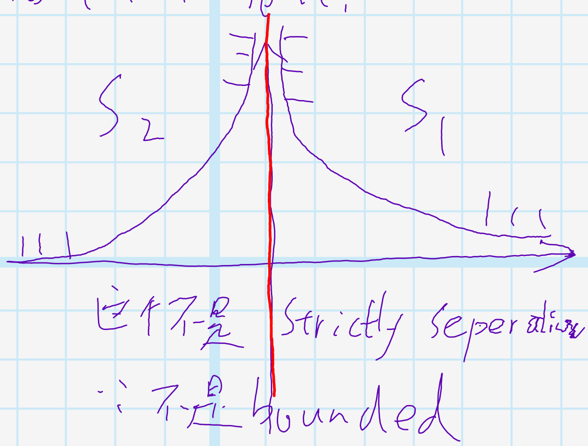
Then, $\exists y$, $\max_{z \in S_1} y^T z < \min_{z \in S_2} y^T z$

y若取负, 就把max, min, < 取反

如果取不到max, 就取上确界和下确界.

$\sup_{z \in S_1} y^T z < \inf_{z \in S_2} y^T z$

如果 S_1, S_2 都不是 bounded, 那么会出现不能严格划分的情形



Pf: Define $S_1 - S_2 = \{z - u : z \in S_1, u \in S_2\}$

如果 S_1, S_2 不相交 0 不属于 $S_1 - S_2$

Note: $0 \notin S_1 - S_2$

If $S_1 - S_2$ is non-empty closed, convex

non-empty: 显然

convex: 可证

closed: 简单证明, 一会再证

$\exists y \quad \max\limits_{v \in S_1 - S_2} y^T v < 0$ (∵ 0 不属于 $S_1 - S_2$ ∴ 可用之前的 set-point sep)

$$\max_{\substack{z \in S_1 \\ u \in S_2}} y^T (z - u) < 0$$

$$\max_{z \in S_1} y^T z + \max_{u \in S_2} (-y^T u) < 0$$

$$\max_{z \in S_1} y^T z < -\max_{u \in S_2} (-y^T u)$$

把负号提出来 max 变 min

$$\max_{z \in S_1} y^T z < \min_{u \in S_2} (y^T u)$$

下面来证 closedness

Let $x_k \in S_1 - S_2$, s.t. $x_k \to x$

证明思路: 假设 $x_k \to x$ (用一个序列收敛到 x) 然后把序列拆成在 S_1 S_2 上的序列的差, 再去证明这两个序列 z_k, u_k 分别收敛到 S_1, S_2

Q: $x \overset{?}{\in} S_1 - S_2$ 下面论证这个问题

$x_k = z_k - u_k \quad z_k \in S_1, \quad u_k \in S_2$

∵ S_2 是闭的 ∴ , $u_k \to u \in S_2$

∵ u_k in S_2 and S_2 is bounded

Bord-Weierstrass?

用那个定理：闭集上的序列存在收敛子序列

关注 u_n 收敛的子序列 $u_{n_{k_i}}$，而 x_n 收敛

∴ 推出 z_{n_k} 收敛.

$$x_{k_i} = z_{k_i} - u_{k_i}$$

$x_{k_i} \to x$, $z_{k_i} \to x+u \in S_1$, $u_{k_i} \to u \in S_2$ ∵ boundedness, closedness of S_2

∵ x_{k_i} 收敛，u_{k_i} 收敛

∴ z_{k_i} 收敛到 $x+u$

∴ x 可以被写成 $z-u$ $\quad z \in S_1, u \in S_2$

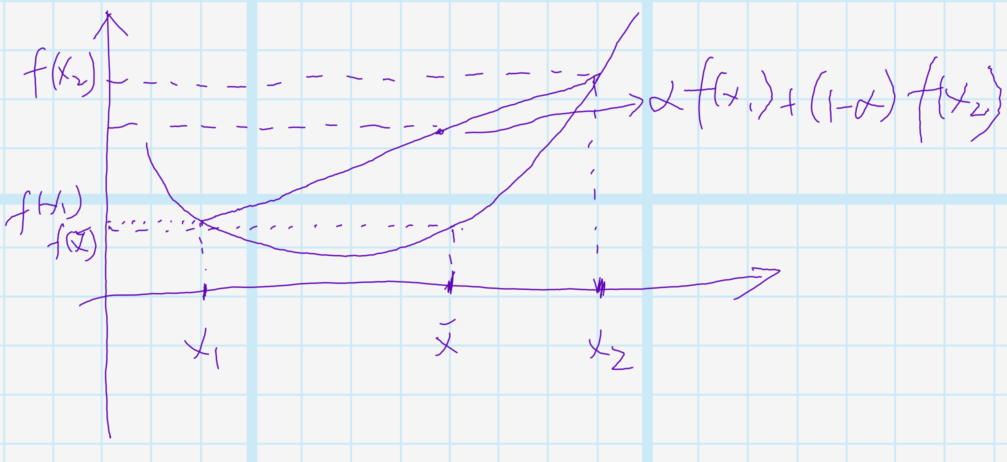
Convex Functions

不引入 $-\infty$ 就不用处理 $-\infty + \infty$

Let $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$

① f is convex if $\forall x_1, x_2 \in \mathbb{R}^n$ $0 \le \alpha \le 1$

$$f(\alpha x_1 + (1-\alpha) x_2) \le \alpha f(x_1) + (1-\alpha) f(x_2)$$



f is concave $\Leftrightarrow$ $-f$ is concave

② Epigraph of f

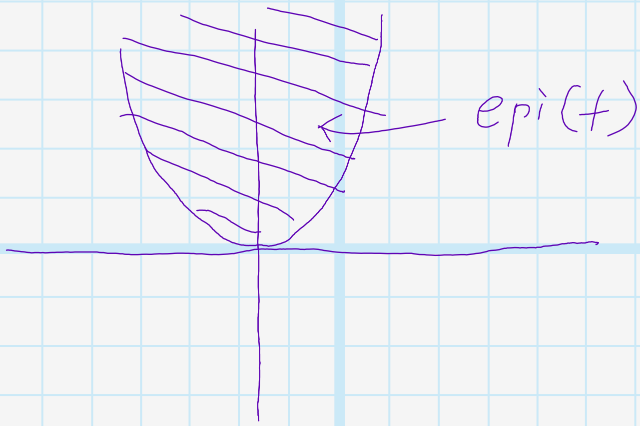
$$\mathrm{epi}(f) = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : f(x) \le t\}$$

③ Effective Domain of f

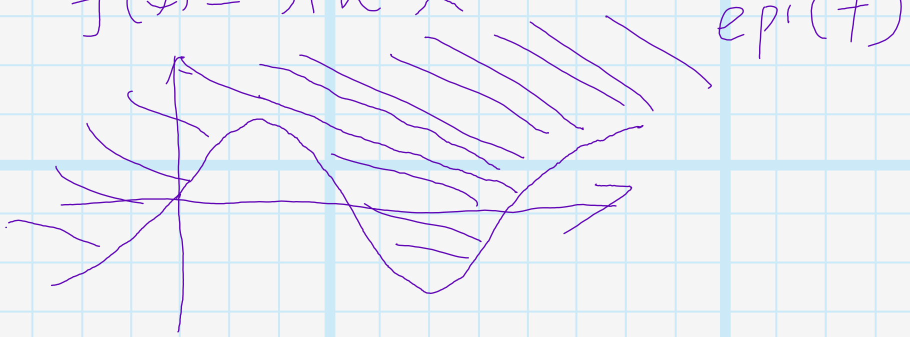
$$\mathrm{dom}(f) = \{x \in \mathbb{R}^n : f(x) < +\infty\}$$

e.g.

$f(x) = x^2$



$f(x) = \sin x$



e.g. let S be any set, Define

$$\mathbb{1}_S(x) = \begin{cases} 0 & \text{if } x \in S \\ +\infty & \text{if } x \notin S \end{cases}$$ [illegible] indicator

$$\mathrm{dom}\, \mathbb{1}_S = S$$

indicator还有个作用就是把约束的优化转为无约束的

$$\inf_{x\in S} f(x) \iff \inf_{x\in\mathbb{R}^n}\{f(x)+\mathbb{1}_S(x)\}$$

← unconstrained

Prop: Let f be as before

Then f is convex iff epi(f) is convex

这个联系了convex set 和convex function

Cor: Jensen's Inequality

f is convex iff $f\left(\sum_{i=1}^{k}\alpha_i x_i\right) \le \sum_{i=1}^{k}\alpha_i f(x_i)$

for any $\sum_i \alpha_i = 1$ $\alpha_i \ge 0$ $k\ge 1$

Pf: $(\Rightarrow)$ 用epi(f)的定义来证

observe $(x_i, f(x_i)) \in \text{epi}(f)$

$\because f$ is convex $\therefore$ epi(f) is convex

Hence take convex combination of $(x_i, f(x_i))$

$$\sum_i \alpha_i (x_i, f(x_i)) \in \text{epi}(f)$$ 仍在epi(f)中

$$\left(\sum_i \alpha_i x_i, \sum_i \alpha_i f(x_i)\right) \in \text{epi}(f)$$

$$\therefore f\left(\sum_i \alpha_i x_i\right) \le \sum_i \alpha_i f(x_i)$$

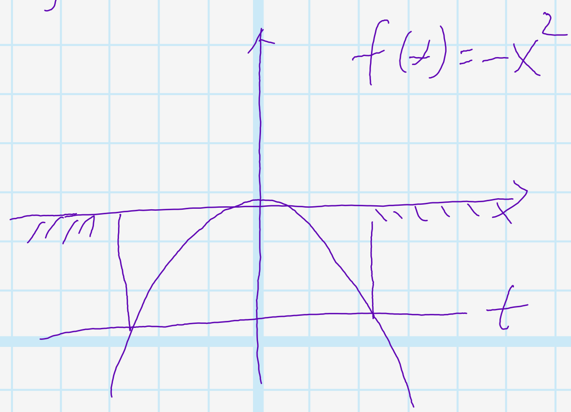
epi(f)的定义

Defin: t-level set of f

$$L_f(t)=\{x\in\mathbb{R}^n : f(x)\le t\}$$

if epi(f) is convex $\Rightarrow$ $L_f(t)$ is convex $\forall t$

反之不成立



$L_f(t)$ convex $\forall t$ 叫做 quasi-convex

所有 t 的 $L_f(t)$ convex
但是 epi(f) 不是 convex

next monday

2017.10.4

Prop: f is convex iff $epi(f)$ is convex

$$epi(f) = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : f(x) \le t\}$$

Prop: A closed convex set is an intersection of all halfspaces containing it.

Let $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex,

suppose $epi(f)$ is closed $\Rightarrow$

$epi(f)$ is intersection of halfspaces containing it

Such halfspaces take the following form:

$$H^-((y, y_0), c) = \{(x,t) \in \mathbb{R}^n \times \mathbb{R} : y^T x + y_0 t \le c\}$$

Suppose $y_0 \ne 0$ Then, $y_0 < 0$ 若$y_0 > 0$，可以找到t使不等式不成立

assume $y_0 = -1$

$$H^-((y,-1), c) = \{(x,t) : y^T x - t \le c\}$$

$$= epi(h) \qquad h(x) = y^T x - c$$

Since $epi(f) \subseteq H^-((y,-1),c) = epi(h)$

$h(x) \le f(x) \ \forall x$ 证明: Let $x \in dom(f)$

$(x, f(x)) \in epi(f) \subseteq epi(h)$

$\Rightarrow y^T x - c \le f(x)$

Since $\operatorname{epi}(f) = \bigcap_{(y,c):\, \operatorname{epi}(f) \subseteq H^-((y,-1),c)} H^-((y,-1),c)$

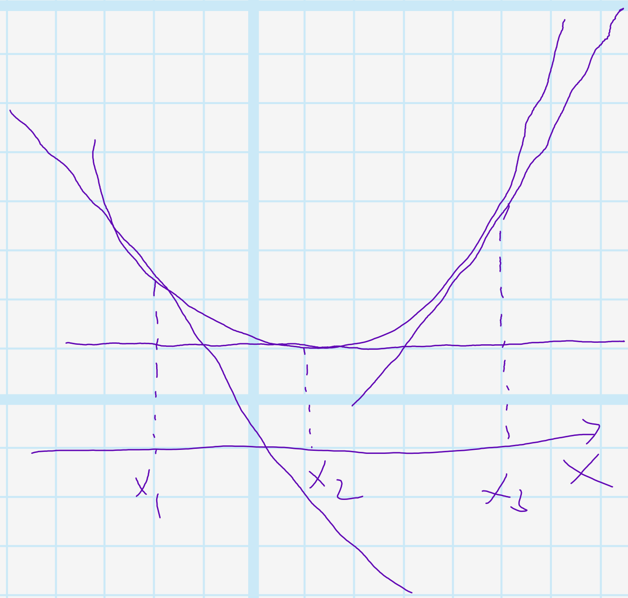
and

$$\bigcap_{(y,c)} \operatorname{epi}(h_{y,c}) = \{(x,t) : y^T x - c \le t \quad \forall (y,c)\}$$

$$= \{(x,t) : \sup_{(y,c)} \{y^T x - c\} \le t\}$$

Thm: Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be convex, $\operatorname{epi}(f)$ be closed. Then, f can be represented as the pointwise supremum of all affine functions $h \le f$, i.e.,

$$f(x) = \sup_{\substack{h \le f \\ h:\text{affine}}} h(x)$$



Define: $S_f = \{(y,c) \in \mathbb{R}^n \times \mathbb{R} : y^T x - c \le f(x) \ \forall x \in \mathbb{R}^n\}$

S_f 包含所有在 f 下的直线的参数.

Note: $y^T x - c \le f(x) \ \forall x \iff \sup_{x \in \mathbb{R}^n} \{y^T x - f(x)\} \le c$

$\Rightarrow S_f$ is the epigraph of

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

x被sup out 掉了, 是一个关于y的函数

Claim: S_f is closed + convex $\Rightarrow f^*$ is convex

Def: f^* is the conjugate of f

下面来讲保凸的操作

Convexity Preserving Ops

① (Non-neg combination)

$$\left.\begin{array}{l} f_1, \cdots, f_k \text{ convex} \\ \alpha_1, \alpha_2, \cdots \alpha_k \geq 0 \end{array}\right\} \Rightarrow \sum_{i=1}^{k} \alpha_i f_i \text{ is convex}$$

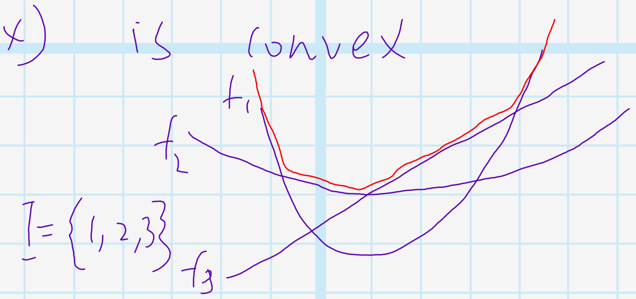
② (Pointwise Supremum)

Let I be an index set (e.g. $I = \{1, 2, \cdots n\}$, $I = \mathbb{Z}_+$, $I = \mathbb{R}$ 可以是不可数)

and $\{f_i\}_{i \in I}$ be a collection of convex functions

Then,

$$f(x) = \sup_{i \in I} f_i(x) \quad \text{is convex}$$



e.g: $f: \mathbb{R}^{m\times n} \to \mathbb{R}_+$ $\quad f(X) = \|X\| =$ largest singular value of X

$$\|X\| = \sqrt{\lambda_{max}(\underset{n\times n}{X^T X})} \qquad \lambda_{max}:\text{取最大的特征值}$$

补充一个定理

Thm (Courant - Fisher) 所有的奇异值都能写成一个优化问题

$$\|X\| = \sup_{\substack{\|u\|_2=1 \\ \|v\|_2=1}} u^T X v$$

详见 matrix analysis,
在控制系统中, singular value
通常表征一个系统的稳定性,
用这个定理来加强稳定性

下面来证明那个 e.g,

取 $I = \{(u,v) \in \mathbb{R}^m \times \mathbb{R}^n : \|u\|_2 = \|v\|_2 = 1\}$

$$f_{(u,v)}(X) = u^T X v$$

$$\|X\| = \sup_{(u,v)\in I} f_{(u,v)}(X) \qquad \text{下面只要 } f_{(u,v)}(X) \text{ 是 convex}$$

而 $f_{(u,v)}(X)$ 是 X 的线性函数 $\therefore f$ 是 convex

f 可以写成: $f_{(u,v)}(X) = tr(u^T X v) = uv^T \bullet X$

$$\underset{m\times n}{A} \bullet \underset{m\times n}{B} \triangleq tr(A^T B)$$

⑬ let $f: \mathbb{R}^n \to \mathbb{R}$, $g\ \mathbb{R} \to \mathbb{R}$ be convex

Then $g\circ f(x) = g(f(x))$

$f(x)$ 仍是 convex
$g(x) = -x$
$g(f(x))$ 不 convex
$\therefore$ 需要加条件
限制 g 为单增

Suppose g is monotonically increasing

Then $g \circ f(x) = g(f(x))$ is convex

④ (Restriction on Lines)

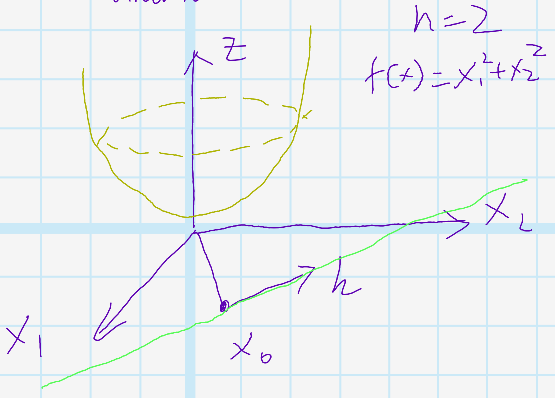
Let $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $\underset{\text{initial point}}{x_0 \in \mathbb{R}^n}$, $\underset{\text{direction}}{h \in \mathbb{R}^n}$

Define $\tilde{f}_{x_0,h}(t) = f(x_0 + th)$

Then:

f is convex iff

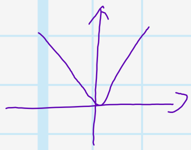
$\tilde{f}_{x_0,h}$ is convex for any $x_0, h \in \mathbb{R}^n$



只看函数在这条线上的值
类似沿着这条线切一刀

$\Downarrow$

2017.10.7

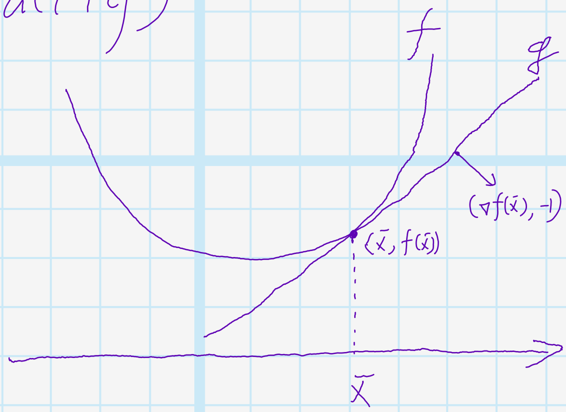
并不是所有函数都可微  下面先看可微函数

Differentiable Convex Functions

Thm: Let $f: \Omega \to \mathbb{R}$ be differentiable on open set Ω, $S \subseteq \Omega$ be convex. Then f is convex on S iff

$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T (x-\bar{x}) \quad \forall\, x, \bar{x} \in S$$

(Gredient inequality)

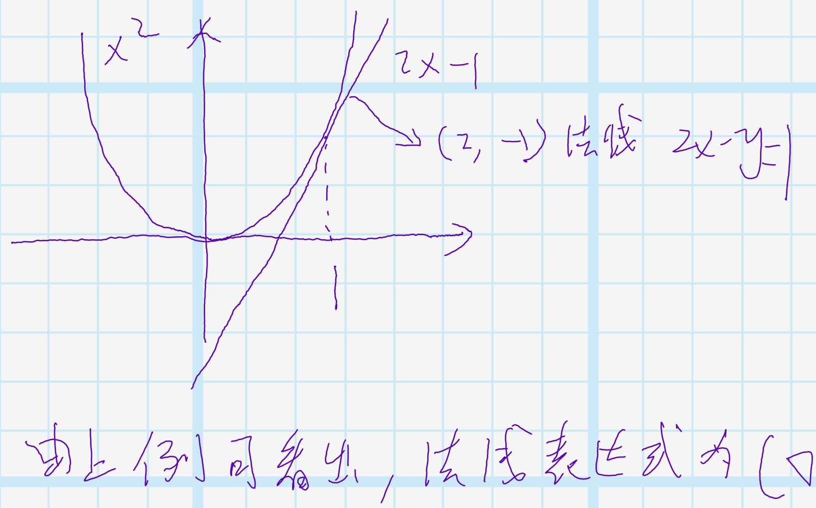


有点像泰勒展开，但只需一阶可导

e.g.: $f(x) = x^2$, $\bar{x} = 1$

Grad. Ineq. $\Rightarrow$ 取 $\bar{x} = 1$

$$x^2 \geq 1 + 2(x-1) = 2x-1$$



由上例可看出，法线表达式为 $(\nabla f(\bar{x}), -1)$

Thm: Let $f: S \to \mathbb{R}$ be twice continuously differentiable on open convex S. Then,

f is convex on S iff $\nabla^2 f(x) \succeq 0 \quad \forall x \in S$

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]$$

twice continuously differentiable

在这里需要 $\nabla^2 f(x)$ 是对称的，即求导顺序不影响

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

e.g.: $f(x, y) = x^2 - y^2$ saddle like function

Let $S = \mathbb{R} \times \{0\}$ y轴固定为 0

Claim: f is convex on S

$$\nabla^2 f(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$$

e.g. $f(x) = \frac{1}{2} x^T A x$

$$\nabla f(x) = Ax$$

$$\nabla^2 f(x) = A$$

e.g. $f: S_{++}^n \to \mathbb{R}$ $\quad f(X) = -\ln \det X$

S_{++}: positive definite，不能取 0，$\because f(x)$ 里有 log

Method 1: $\nabla f(X) = -X^{-1}$ $\quad \nabla^2 f(X) = X^{-1} \otimes X^{-1} \succ 0$

其中 $A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ \vdots & & & \vdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nn}B \end{bmatrix}$

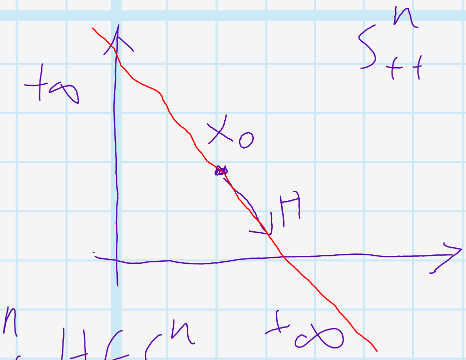
克罗内克积、

还有个结论 若 A, B 都是 PD, 则 $A \otimes B$ 是 PD

上一个方法需要涉及矩阵求导, 下一个比较通用

Method 2: Restriction on Lines

Let $X_0 \in S^n_{++}$ $\quad H \in S^n$ $\quad$ H代表直线方向



Define $\tilde{f}_{X_0,H}(t) = f(X_0 + tH)$

Goal: show $\tilde{f}_{X_0,H}$ is convex $\forall X_0 \in S^n_{++}, H \in S^n$

变形的好处: 把高维的 $f(X) = -\ln\det X$ 变成了一个一维函数

坏处: 需对所有的 f 进行说明

$$\tilde{f}_{X_0,H}(t) = -\ln\det(X_0 + tH)$$

$\because X_0$ is PD $\quad \therefore X_0 = X_0^{1/2} X_0^{1/2}$

$$= -\ln\det\left[X_0^{1/2}\left(I + tX_0^{-1/2}HX_0^{-1/2}\right)X_0^{1/2}\right]$$

$$= -\ln\left[\underbrace{\det(X_0^{1/2})^2}_{常数} \det\left[I + tX_0^{-1/2}HX_0^{-1/2}\right]\right]$$

It suffices to look at:

$$-\ln\det[I + t\underbrace{X_0^{-1/2} H X_0^{-1/2}}_{\tilde{H}}]$$

Let $\lambda_1, \dots, \lambda_n$ be eigenvalues of $\tilde{H}$ 原式 =

$$= -\sum_{i=1}^{n} \ln(1+t\lambda_i)$$

$$= \sum_{i=1}^{n} \ln \frac{1}{1+t\lambda_i}$$

consider $t \longmapsto \ln \frac{1}{1+t\lambda}$

这个函数的二阶导 >0

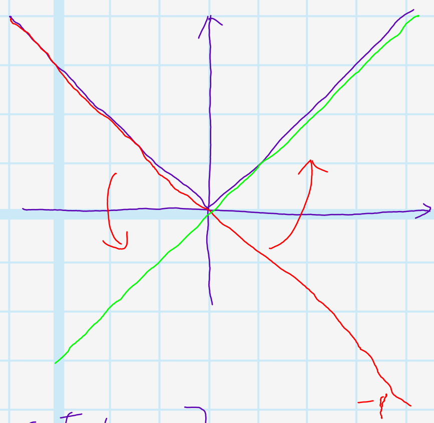
Non-differentiable Convex functions

利用 convex 的 intuition：所有 convex 的函数都可以用若干超平面来支撑

Def: Let $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, A vector $s \in \mathbb{R}^n$ is called a subgradient of f at $\bar{x}$ if

$$f(x) \geq f(\bar{x}) + s^T(x - \bar{x}) \quad \forall x \in \mathbb{R}^n$$

e.g. $f(x) = |x|$



$\bar{x} = 0$: consider

$$|x| \geq 0 + s(x-0) = sx$$

$$|x| \geq sx \quad \forall x \in \mathbb{R}$$

This holds for all $s \in [-1, 1]$

可见 subgradient 不唯一

Def: The set of subgradients of f at $\bar{x}$ is called subdifferential of f at $\bar{x}$ denoted by $\partial f(\bar{x}) = \{ s : f(x) \geq f(\bar{x}) + s^T(x-\bar{x}) \ \forall x \}$

回到 $|x|$ 那个例子

consider $\bar{x}=1$

$|x| \geq 1 + s(x-1)$

$= sx + (1-s) \quad \forall x$

s 只能取 1

Prop: Let $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ f is convex $f \not\equiv +\infty$ (不恒等于 $+\infty$) Then

(i) $\partial f(\bar{x})$ non-empty, compact, convex $\forall \bar{x} \in \text{int dom}(f)$ dom(f): where f is finite

(ii) if f is differentiable at $\bar{x}$, then $\partial f(\bar{x}) = \{ \nabla f(\bar{x}) \}$ 如果 f 在 $\bar{x}$ 可微，那么它在 $\bar{x}$ 的 subgradient 只有一个值且和 gradient 相等。

2017.10.9

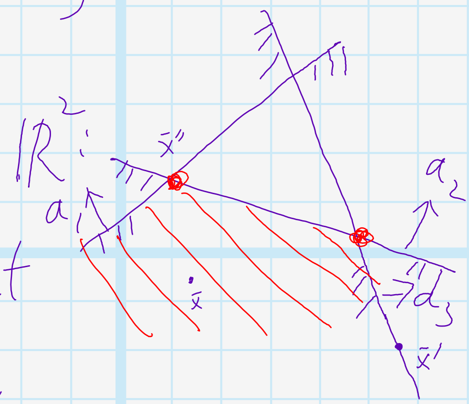
Linear Programming

Halfspace: $H^-(s,c) = \{x : s^T x \le c\}$

Def: A polyhedron is the intersection of a finite number of halfspaces.

A polytope is a bounded polyhedron

$P = \{x \in \mathbb{R}^n : a_i^T x \le b_i,\ i = 1, \cdots, m\}$



Given $\bar{x} \in P$, let

$I(\bar{x}) = \{i : a_i^T \bar{x} = b_i\}$ be the index set of active constraints 成立的约束

如图

$I(\bar{x}) = \phi$

$I(\bar{x}') = \{3\}$

Prop: The following are equivelent:

(1) $\exists$ n linearly independent vector in $\{a_i : i \in I\}$

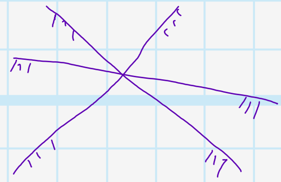
(2) $\bar{x}$ is the unique solution to

$$a_i^T x = b_i \quad \forall i \in I(\bar{x})$$

图中 x'' 有 $I(x'')=\{1,2\}$

∴只有一个交点

$\{a_i : i\in I\}$ 可以不止 n 个,
但这里要求是 independent



这个图中 $\{a_i : i\in I\}$ 有三个,
但线性无关的只有两个

Def: Let P be a polyhedron and $\bar{x}\in\mathbb{R}^n$

(1) $\tilde{x}$ is basic solution if there are n linearly indep active constrains on it.

(2) if, in addition, $\bar{x}\in P$ then it is a basic feasible solution

∵那些 constrain 可以(张成空间)的一组基,
∴ call basic
其实就是那些角点

Recall: $x\in C$ is an extreme point of the convex set C if

$$\left.\begin{array}{l} x_1, x_2 \in C \\ x=\dfrac{x_1+x_2}{2} \end{array}\right\} \Rightarrow x_1=x_2=x$$

Prop: Let P be a polyhedron $\bar{x}\in P$

The following are equivalent.

(1) $\bar{x}$ is a BFS (Basic feasible solution)

(2) $\bar{x}$ is extreme point of P

可以通过不等价来证不是 polyhedron

Example: $B(0,1)$ is not a polyhedron

Suppose not, Then

$$B(0,1) = \{x : a_i^T x \le b_i,\ i=1,\cdots,m\}$$

The number of BFS $\#BFS \le \binom{m}{n} = \frac{m!}{n!(m-n)!}$

有m个超平面,选出n个找,有可能得到一个交点,

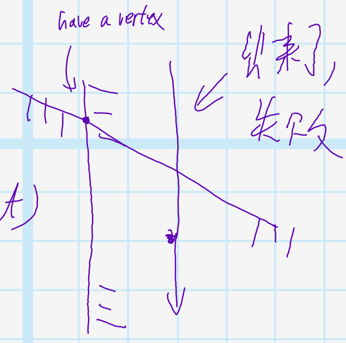
$\therefore \binom{m}{n}$ 应该是BFS点个数的上界.

但 extreame 个数是无穷个.

Def: A polyhedron P contains a line if

$\exists x \in P$ and $d \neq 0$ s.t.

$x + \alpha d \in P \quad \forall \alpha$

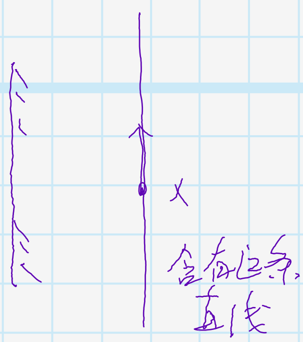


Thm: TFAE: (the following are equivalent)

(1) P has at least 1 vertex

(2) P does not contain a line

∵这个不含任何直线.



e.g.: consider

$$\phi \neq \left\{ x : \begin{array}{l} Ax = b \\ x \geq 0 \end{array} \right\} \quad \begin{array}{l} A \in \mathbb{R}^{m\times n} \\ b \in \mathbb{R}^{m} \end{array}$$

这个也是 polyhydron

∵ = 可以拆成 ≤ ≥

$Ax \geq b$

$Ax \leq b$

∵有 $x \geq 0$ ∴不能含有直线 ∴肯定有 vertex

那么 $Ax=b$ 其实没啥用

下面开始研究 LP

Consider the LP:

$$v^* = \inf_{x \in P} c^T x \qquad \text{Suppose } v^* > -\infty$$

P: polyhedron

Thm: Suppose P has at least one vertex

Then either:

(*) (1) $v^* = -\infty$

(2) $\exists$ vertex optimal solution

对于一般的结构，有可能出现存在下确界，但是域中没有点能取到下确界，∴这个 Polyhydron 的性质是比较好的.

∴如果 $v^* \neq -\infty$，那个时可以直接写成 min

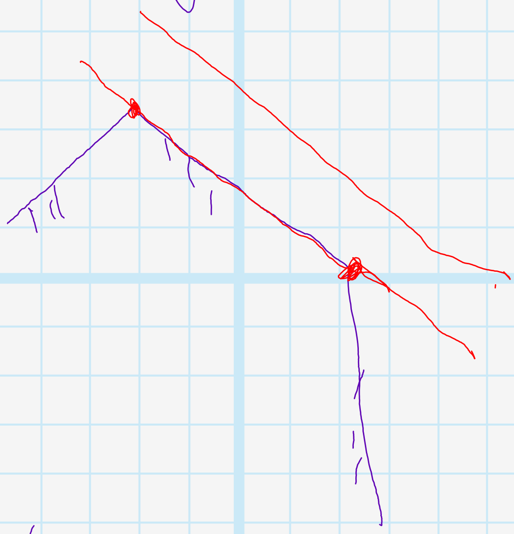
Cor: Consider (*), Suppose $P \neq \emptyset$. Then, either

(1) $v^* = -\infty$

(2) $\exists$ optimal solution

这里没有假设vertex存在，是一个更广的定理

e.g.



有两个vertex都能取到

Pf: consider

$$P = \{x \in \mathbb{R}^n : a_i^T x \le b_i,\ i = 1, \cdots, m\}$$

思路：把P embed到高维的有vertex的空间。

Introduce $x^+, x^- \in \mathbb{R}^n$, $s \in \mathbb{R}^m$

$$a_i^T x \le b_i \xleftrightarrow{\text{转换}} a_i^T (x^+ - x^-) + s_i = b_i$$

$$x_i^+, x_i^- \ge 0, \quad s_i \ge 0$$

向量x中的元素可正可负，现在转化为两个恒正向量的差

s是一个量，也大于0

(*) is equivalent to

$$\min\ c^T (x^+ - x^-)$$

$$\text{s.t.}\ a_i^T (x^+ - x^-) + s_i = b_i \quad i = 1, \cdots, m$$

$$x^+, x^-, s \ge 0 \leftarrow \text{注意看这个}$$

也可以写成更compact的形式

$$\min\ \tilde{c}^T z$$

$$\text{s.t.}\ \tilde{A} z = \tilde{b}$$

$$z \ge 0 \leftarrow \text{不等式需对所有变量都有约束}$$

$$z = (x^+, x^-, s) \in \mathbb{R}^{2n+m}$$

$$\tilde{c} = (c\ \ -c\ \ 0) \quad \tilde{A} = [A \mid -A \mid I]$$

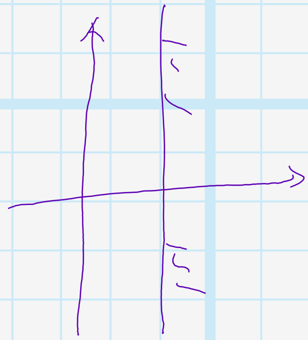
这种写法叫做LP的standard form

可见跟之前那个例子一样，变量都≥ 0，

而上面那个等式其实没有什么作用了

∴ 有一个 vertex ∴ 用 Thm 即证

e.g. $\min x$ s.t. $x \geq 1$



存在 optimal solution

但是没有 vertex.

Primal

2017.10.11

Recall: Standard Primal form of LP

$$(P) \quad \begin{aligned} V_p^* = \min \ & c^T x \\ \text{s.t.} \ & Ax = b \\ & x \geq 0 \end{aligned}$$

Fact: Consider (P). Suppose it is feasible.
Then, either
(i) $V_p^* = -\infty$
or (ii) $\exists$ optimal solution

Checking (In)feasibility

Theorem (Farkas' Lemma) (Theorems of Alternatives)
Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. Then exactly one of the following has solution:

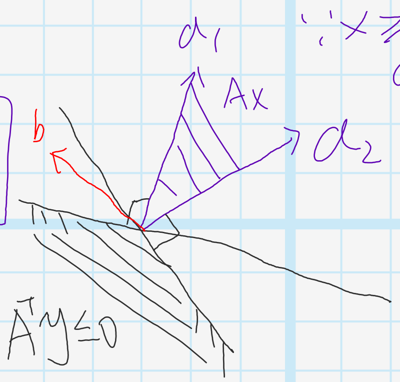
(1) $Ax = b, \quad x \geq 0$ 两个式子有且只有一个成立

(2) $A^T y \leq 0, \quad b^T y > 0$

∵ $x \geq 0$ ∴ Ax 一定在 a_1 a_2 中间.

Intuition 如图.

$A = \begin{bmatrix} | & | \\ a_1 & a_2 \\ | & | \end{bmatrix}$



当b为如图时 $Ax=b$ 在 $x \geq 0$ 上无解, 下面看(2),
$A^T y \leq 0$ 的区域如图,
∴能找到 y 使 $b^T y > 0$

Pf: Step 1: (1) and (2) cannot both have solution

Step 2: If (1) has no solution, then (2) has solution

	(1)	(2)
Step 1	✓	✓
Step 2	×	×
	× $\Longrightarrow$	✓
	✓ $\Longleftarrow$	×

为什么只用证两个？等于说：

最后一条和第三条是等价的，

∵ $\neg(\neg A \Rightarrow B)$ 等价于 $\neg B \Rightarrow A$

Step 1: Suppose both solvable

$$0 \geq \bar{y}^T A$$

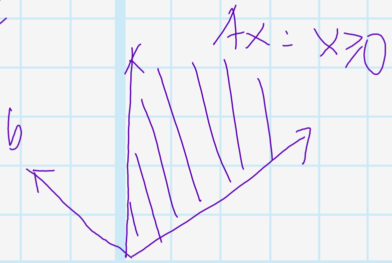
$$0 \geq (\bar{y}^T A)\bar{x} \quad (\because \bar{x} \geq 0)$$

$$\overset{A\bar{x}=b}{=} \bar{y}^T b > 0 \text{ 矛盾}$$

step 2 思路：证明当(1)不成立时(2)中的两式分别成立

Step 2: Suppose (1) is not solvable

现在有一个点 $(\vec{b})$ 和一个区域，想用 Seperation 来搞出下.



Let $S = \{Ax : x \geq 0\}$, note $b \notin S$

check S is non-empty close and convex

$\because 0 \in S \quad \therefore$ non-empty

$S = A\mathbb{R}^n_+ \quad \therefore$ convex

close 证明和之前 set 交类似.

By Separation Theorem, $\exists y$

$$\max_{z \in S} y^T z < y^T b$$

$$\Longleftrightarrow \quad y^T A x < y^T b \quad \forall x \geq 0 \qquad (*)$$

$\Rightarrow$ Take $x=0 \Rightarrow b^T y > 0$

下面来证 $A^T y \le 0$

Suppose not, Then $\exists i: (A^T y)_i > 0$

注意，这里 $A^T y \le 0$ 的反面是存在一个元素 >0，不是 $A^T y > 0$!

Take $x = \alpha e_i$, $\alpha > 0$, in $(*)$

$b^T y > y^T A x = \alpha (A^T y)_i \qquad \forall \alpha > 0$

可以把 α 取得足够大，使得不等式不成立

∴ 矛盾

注：

$$\begin{matrix} A^T y \le 1 \\ b^T y > 0 \end{matrix} \iff \begin{matrix} A^T y \le 0 \\ b^T y \ge 1 \end{matrix}$$

大和小可以任意放大，类似于等价代换

下面讲 LP Duality Theory

Suppose we find $y: A^T y \le c$

Then, for any feasible x

把 LP primal standard form 抄在这里

$$v_p^* = \min c^T x$$
$$\text{s.t. } Ax = b$$
$$x \ge 0$$

$$(c - A^T y)^T x \ge 0$$

$c^T x \ge y^T A x = b^T y$ 对所有 fessiable 的 x 都成立

这样就找到了 primal form 中 $c^T x$ 的一个下界

The best lower bound on V_p^* via this construction is given by

$$(D) \quad V_d^* = \max \ b^T y$$
$$\text{s.t. } A^T y \le c$$

去最小化一个东西 等价于去最大化它的下界

我们感兴趣的是找到这么一个最大的下界，然后来研究他和原始问题最优解的关系

This is called the standard dual of LP

Thm (LP weak Duality)

Let $\bar{x}$ be feas for (P) $\bar{y}$ feas for (D)

Then, $c^T\bar{x} \geq b^T\bar{y}$ ∴ $b^T\bar{y}$ 是 $c^T\bar{x}$ 的下界

Cor:

(1) If $v_p^* = -\infty$, then (D) is infeasible

(2) Let $(\bar{x}, \bar{y})$ be a pair of primal-dual feas solution.

Let $\Delta(\bar{x}, \bar{y}) = c^T\bar{x} - b^T\bar{y}$ be duality gap

If $\Delta(\bar{x}, \bar{y}) = 0$, then $\bar{x}$ optimal for (P)
$\bar{y}$ optimal for (D)

Thm (LP strong Duality)

Suppose (P) has an optimal solution x^*

Then (1) (D) also has an optimal solution y^*

$$(2)\quad \Delta(x^*, y^*) = 0$$

刚才那个弱对偶给出了不等式关系。这个强对偶给出了取等号的条件

Pf: By assumption,

$$Ax^* = b,\quad x^* \geq 0 \quad V_p^* = c^T x^*$$

and $\nexists x$: 不存在x让c^Tx更小

$$(\Delta)\qquad Ax = b,\ x \geq 0,\ c^T x - c^T x^* < 0$$

Idea: Apply Fa1kas

Fukars 引理

(1) $Ax = b,\ x \geq 0$

(2) $A^T y \leq 0,\ b^T y > 0$

Homogenize the system:

$$(\Delta\Delta)\ \begin{cases} Ax - bt = 0 \\ (x, t) \geq 0 \\ c^T x - (c^T x^*) t < 0 \end{cases}$$

也能写 $c^Tx - (c^Tx^*)t = -1$

加了个额外变量t.

This is equivalent to (Δ)

case 1: $t > 0$: x/t solves (Δ) 矛盾

case 2: $t = 0$: Then $Ax = 0$ $x \geq 0$ $c^T x < 0$

Consider $\bar{x} = x^* + x \geq 0$

1) $A\bar{x} = Ax^* + Ax = b$

2) $c^T \bar{x} = c^T x^* + c^T x < c^T x^*$

solve (Δ) 矛盾

$\therefore (\Delta\Delta)$

(⇐) 把最后一项写作 $c^T x - (c^T x^*) t = -1$

$$\Rightarrow \begin{bmatrix} A & -b \\ c^T & -c^T x^* \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \qquad \begin{bmatrix} x \\ t \end{bmatrix} \geq 0$$

A b is not soluable

by Fralcas!

$$\therefore \begin{bmatrix} A^T & c \\ -b^T & -c^T x^* \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \qquad -y_2 > 0$$

$(*)$

$$\begin{cases} y_2 < 0 \\ A^T y + c y_2 \leq 0 \\ -b^T y_1 - (c^T x^*) y_2 \geq 1 \end{cases}$$

$$A^T\left(-\frac{y_1}{y_2}\right) - c \leq 0$$

$$b^T\left(\frac{y_1}{y_2}\right) + c^T x^* \leq 0$$

令 $-\dfrac{y_1}{y_2} = y^*$

$$\left.\begin{cases} A^T y^* \leq c \Rightarrow (D) \text{ feasible} \\ c^T x^* \leq b^T y^* + \text{weak duality} \end{cases}\right\} \Rightarrow y^* \text{ is dual optimal}$$

Primal Form 2017.10.16

$$v_p^* = \min \ c^T x$$
$$(P) \quad \text{s.t. } Ax = b$$
$$x \geq 0$$

Dual Form

$$(D) \quad v_d^* = \max b^T y$$
$$\text{s.t. } A^T y \leq c$$

Thm: (Strong Duality)

Suppose (P) has optimal solution x^*. Then

i) (D) has an optimal solution y^*

(ii) $c^T x^* = b^T y^*$

Cor: Suppose (P) & (D) are feasible. Then, both of them have optimal solutions, and duality gap $= 0$

Observe: Solving (P) is equivalent to solve

$$\begin{cases} Ax = b, \quad x \geq 0 & \text{(Primal feasibility)} \\ A^T y \leq c, & \text{(dual feasibility)} \\ c^T x = b^T y & \text{(zero duality gap)} \end{cases}$$

把一个优化问题转化成求解一个线性系统

当duality gap达到0时，取到最优解

Thm (Complementary Slackness)

Let $\bar{x}$ be feas. for (P), $\bar{y}$ be feas for (D)

Then

$\bar{x}, \bar{y}$ are optimal iff $\underbrace{\bar{x}_i}_{\text{primal slack}} \underbrace{(c - A^T\bar{y})_i}_{\text{dual slack}} = 0 \quad \forall i$

即是说这两项若有一项非0,则另一项一定为0

Pf: $c^T\bar{x} - b^T\bar{y} = c^T\bar{x} - \bar{x}^T A^T \bar{y} = \bar{x}^T(c - A^T\bar{y})$

$$= \sum_{i=1}^{n} \bar{x}_i (c - A^T\bar{y})_i$$

$\because x \geq 0 \quad c - A^T\bar{y} \geq 0 \quad \therefore$ 每项都等于0

$\therefore$ 引入一个slack项,上面那个线性方程可写作:

$$\begin{cases} Ax = b, \ x \geq 0 & \text{(Primal feasibility)} \\ A^T y + s = c, \ s \geq 0 & \text{(dual feasibility) 变成等式了} \\ \underline{x^T s = 0} & \text{(complentarity)} \end{cases}$$

这是个非线性项,$\because x, s$ 都是未知数.

通过求解这个非线性的系统,而不是之前那个

通过以下方法求解:

Interior Point methods

把 $x^T s = 0$ 换成 $x^T s = \mu \qquad \mu > 0, \ \mu \downarrow 0$

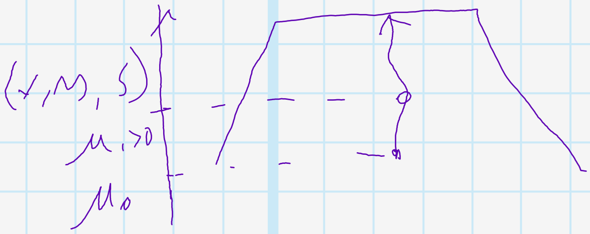
对于任意 $\mu>0$ (x,y,s) 都有唯一解

随 μ 的变化 (x,y,s) 沿着一条可微的path 移动，收敛到 optimal solution

1984: Karmarkas

83-84: Nesterov Nemirovski

94 IPM



下面讲如何根据(P)来写(D)

Example:

$$
\begin{aligned}
\min\ & x_1+2x_2+x_3 \\
\text{s.t.}\ & x_1-2x_2+x_3\geq 2 \\
& -x_1\qquad\ \ +x_3\geq 4 \\
& 2x_1\qquad\ \ +x_3\geq 6 \\
& x_1+x_2+x_3\geq 2 \\
& x\geq 0
\end{aligned}
$$

先把不等式约束转成等式约束

$$
\begin{aligned}
& x_1-2x_2+x_3-u_1=2 \\
& -x_1\qquad\ \ +x_3-u_2=4 \\
& 2x_1\qquad\ \ +x_3-u_3=6 \\
& x_1+x_2+x_3-u_4=2 \\
& x\geq 0 \qquad u\geq 0
\end{aligned}
$$

$$\min\ (1,2,1,0)^T(x_1,x_2,x_3,u)$$

$$\left[\begin{array}{ccc|c} 1 & -2 & 1 & \\ -1 & 0 & 1 & -I_4 \\ 2 & 0 & 1 & \\ 1 & 1 & 1 & \end{array}\right]\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ u \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 6 \\ 2 \end{bmatrix}$$

$$x,\ u \geq 0$$

$\therefore$ Dual 为：

$$\max\ 2y_1 + 4y_2 + 6y_3 + 2y_4$$

$$\text{s.t.}\ \left[\begin{array}{cccc} 1 & -1 & 2 & 1 \\ -2 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ \hline & -I_4 & & \end{array}\right]\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \leq \begin{bmatrix} 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}$$

Claim: $y=0$ is feasible $\therefore$ dual is feasible.

consider: $\bar{x} = \left(\frac{2}{3}, 0, \frac{14}{3}\right)$ (P)和(D)通常有一个好的最优解能直接观察.

2^{nd}: $-\frac{2}{3} + \frac{14}{3} = 4$ 这个 constrain tight

3^{rd}: $\frac{4}{3} + \frac{14}{3} = 6$

$\bar{x}_2 = 0 \quad \bar{u}_1 > 0,\ \bar{u}_4 > 0$

$$\bar{x}_1\left[1 - (y_1 - y_2 + 2y_3 + y_4)\right] = 0 \quad \text{(complementarity)}$$

$\because \bar{x}_1 = \frac{2}{3} > 0 \quad \therefore\ 1 - (y_1 - y_2 + 2y_3 + y_4) = 0$

$y_1 + y_2 + y_3 + y_4 = 1$

$\because u_1 > 0 \quad \therefore\ y_1 = 0$

$\because u_4 > 0 \quad \therefore y_4 = 0$

$y_2, y_3 \geq 0$

$\Rightarrow y_2 = \frac{1}{3} \quad y_3 = \frac{2}{3}$

$\therefore (y_1, y_2, y_3, y_4) = (0, \frac{1}{3}, \frac{2}{3}, 0)$

$\therefore$ Dual is feasible.

下面讲LP的应用场景

Vertex Cover

graph $G = (V, E)$ V: Vertex, E: Edge.

there is a cost for each vertex

$i \in V: c_i \geq 0$ cost of vertex i

Def Vertex cover, $S \subseteq V$: for each edge $e = (u, v)$, either $u \in S$ or $v \in S$

所有的边有一个点在S中

Goal: Find the miniam cost vertex cover

用relaxition方法，一个节点是否选择为0或1，把它放松限制为[0,1]

2017.10.18

接着找 Vertex cover

这也是一个组合优化问题

Let x_i be the decision variable indicating whether to pick i or not $x_i \in \{0,1\}$

$$v^* = \min \sum_i c_i x_i = c^T x$$

$$\text{s.t.} \quad x_i + x_j \geq 1 \qquad \forall (i,j) \in E$$

i, j 是一个edge的两个端点.

$$x_i \in \{0,1\} \qquad \forall i \in V$$

← 这项不是线性的

(Integer LP)

这个非线性的问题是NP-hard的.

为了解决这个非线性的问题，把 $x_i \in \{0,1\}$ 进行松弛，变成 $x_i \in [0,1]$

进行了松弛之后，原始问题的解也是新问题的解，但反之不成立。

松弛后的最优解?

LP relaxation V^*_{LP} is also opt value of LP

Note: $V^* \geq V^*_{LP}$

$$\underset{LP}{x^*} \longrightarrow \underset{ILP}{\hat{x}}$$

Algo: ① Solve LP relaxation, get optimal solution x^*

② "Round" x^* into a feasible solution $\hat{x}$ to the ILP

所以问题变成了：

Q: ①: How to round?

②: After rounding, can we say something about $V(\hat{x}) \triangleq \sum_i c_i \hat{x}_i$ vs. V^*! $\geq$

$\leq \alpha V^*$?

Thm: Consider the system

$$x_i + x_j \geq 1 \quad \forall (i,j) \in E$$

$$x_i \geq 0$$

如果 $x_i \geq 1$，那么不是最优，所以不用写：$x_i \leq 1$

(This is feasible region of LP relaxation)

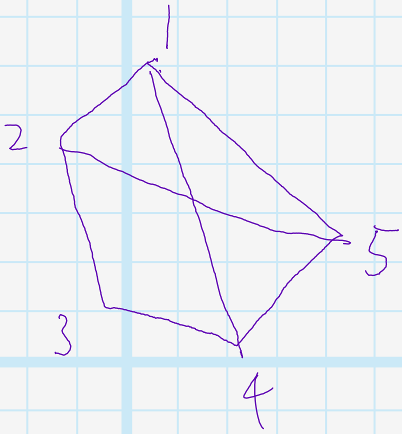
这个 region 是个 polyhedron，有 vertex

Let $\bar{x}$ be a vertex, Then,

$$\bar{x}_i \in \{0, \tfrac{1}{2}, 1\} \quad \forall i \in V$$

This called Half-integrality

eg.



$$
\begin{aligned}
x_1 + x_2 &\geq 1 \\
x_2 + x_3 &\geq 1 \\
x_3 + x_4 &\geq 1 \\
x_4 + x_5 &\geq 1 \\
x_1 + x_5 &\geq 1 \\
x_1 + x_4 &\geq 1 \\
x_2 + x_5 &\geq 1
\end{aligned}
$$

下面来找 vertex，需要五个等式确定一个点。

$$
\begin{cases}
x_1 + x_2 = 1 \\
x_2 + x_3 = 1 \\
x_1 + x_5 = 1 \\
x_1 + x_4 = 1 \\
x_2 + x_5 = 1
\end{cases}
$$

线性独立

解为 $x_1 = x_2 = x_3 = x_4 = x_5 = \frac{1}{2}$

观察到每个方程里都有两个项。

By the Thm:

① Let x^* be a vertex optimal of LP

② set $\hat{x}_i = \begin{cases} 0 & \text{if } x_i^* = 0 \\ 1 & \text{otherwise} \end{cases}$

Note: $V(\hat{x}) = \sum_i c_i \hat{x}_i \overset{*}{\leq} 2 \sum_i c_i x_i^* = 2 V_{lp}^* \leq 2V^*$

这样之前的算法中的 $\alpha = 2$

这个 α 不随 V 中点的数量而变化

下面讲那个 Thm 的证明

Pf. Let $x \in P$ $\left(P = \begin{cases} x_i + x_j \geq 1 \\ x_i \geq 0 \end{cases} \forall (i,j) \in E \right)$

$U_{-1} = \{ i : x_i \in (0, \tfrac{1}{2}) \}$

$U_1 = \{ i : x \in (\tfrac{1}{2}, 1) \}$

这样，x 分为三类

$$\underbrace{\begin{bmatrix} x_i \\ 0, \frac{1}{2}, 1 \\ \hline x_i \\ i \in U_{-1} \\ \hline x_i \\ i \in U_1 \end{bmatrix}}_{x} = \frac{1}{2} \underbrace{\begin{bmatrix} x_i \\ \hline x_i - \varepsilon \\ \hline x_i + \varepsilon \end{bmatrix}}_{y} + \frac{1}{2} \underbrace{\begin{bmatrix} x_i \\ \hline x_i + \varepsilon \\ \hline x_i - \varepsilon \end{bmatrix}}_{z}$$

下面验证 y, z is feasible

Feasibility of y_i, z_i

$$x_i + x_j \geq 1$$

Suppose $x_i \in (0, \frac{1}{2})$ $\therefore x_j \in [\frac{1}{2}, 1]$

case 1: $x_j \in (\frac{1}{2}, 1)$:

$$\underbrace{x_i - \varepsilon}_{y_i} + \underbrace{x_j + \varepsilon}_{y_j} \geq 1$$

$\therefore y_i + y_j \geq 1$ y is feasible

case 2: $x_j = 1$

$$\underbrace{x_i - \varepsilon}_{y_i} + \underbrace{1}_{y_j} \geq 1$$

If U_{-1} or $U_1 \neq \emptyset$, then $y \neq x$, $z \neq x$

下面来考虑下条件

$$\min c^T x$$
$$\text{s.t. } Ax = b$$
$$x \geq 0 \iff x_i \geq 0 \ \forall i$$

$x \geq 0$ 也可以用集合方式来定义，

$\because$ 我们没有定义过两个向量的大小如何比较

下面来讲偏序，偏序只定义在集合上，不依赖于线空间中的代数结构

"$\succeq$" define a partial order on $\mathbb{R}^n$

(a) Reflexivity: $u \succeq u \quad \forall u \in \mathbb{R}^n$

(b) Anti Symmetry: $u \succeq v$: $v \succeq u \Rightarrow u = v$

(c) Transitivity: $u \succeq v, v \succeq w \Rightarrow u \succeq w$

Compatibility w/ Linear Operations

(d) Homogeneity: $\forall u, v \in \mathbb{R}^n$ if $u \succeq v$, then $\alpha u \succeq \alpha v$
$\alpha \geq 0$

(e) Additivity: $\forall u, v, w, z \in \mathbb{R}^n$ $\left.\begin{matrix} u \succeq v \\ w \succeq z \end{matrix}\right\} u + w \succeq v + z$

有了偏序的定义，我们知道 $x_i \geq 0 \ \forall i$ 并不是 $x \geq 1$ 的唯一指定，我们可以指定新的

Recall 偏序 2017.10.23

(a). Reflexive: $v \geq v$

(b). Anti-Symmetry: $u \geq v,\ v \geq u \Rightarrow u = v$

(c). Transitivity $u \geq v,\ v \geq w \Rightarrow u \geq w$

(d). Homogeneity $\begin{matrix} u \geq v \\ \alpha \geq 0 \end{matrix} \Rightarrow \alpha u \geq \alpha v$

(e). additivity $\begin{matrix} u \geq v \\ w \geq z \end{matrix} \Rightarrow u + w \geq v + z$

E可以是vector space也可以是matrix空间

Def: Let E be a finite dim Eudidean Speace,
$\bullet$ be an inner product on E
We say an ordering $\geq$ on E is good
if it satisfies (a)-(e) above

$u \geq 0 \not\Rightarrow -u \leq 0$
这里我们没有定义这个

$\bullet$ $\geq$ 都是抽象定义.

下面的目标是把那五条转化到E的代数空间(K的空间)中去.

Obs: A good ordering is compeletely specified by
$K = \{u \in E : u \geq 0\}$
i.e. which pair of vectors u, v satifying $u \geq v$ can be deduced from K

① If $u \geq v$, Then, since $-v \geq -v$, add them $\Rightarrow$
$u - v \geq 0 \Rightarrow u - v \in K$

② If $u - v \in K$, Then $u - v \geq 0$
since $v \geq v$, we have $u \geq v$

通过以上定义，把 u, v 的代数关系转化为了和 K 的几何关系（直接转化为 $\cup$ 集合的属于关系）

obs: K is a pointed cone i.e.,

(1) K is non-empty and closed under addition: $\begin{matrix} u \in K \\ v \in K \end{matrix} \Rightarrow u+v \in K$

K 对加法封闭

(2) K is a cone: i.e.,

$\begin{matrix} u \in K \\ \alpha \geq 0 \end{matrix} \Rightarrow \alpha u \in K$

(3) K is pointed: i.e.

if $\begin{matrix} u \in K \\ -u \in K \end{matrix}$ Then $u=0$

注：这个 cone 也是 convex cone

$\because \alpha u \in K$ (由(2))

$(1-\alpha) v \in K$ (由(2))

$\therefore \alpha u+(1-\alpha) v \in K$ (由(1))

$\therefore$ 这是个 convex cone.

现在，我们已经做了：

1°: Good ordering $\longrightarrow$ pointed cone K w/ $0 \in K$

下面来做：

2°: pointed cone K w/ $0 \in K$ $\longrightarrow$ good ordering

Def: $u \geq_K v \Leftrightarrow u - v \in K$ K属于E这个空间，没有强调有序

claim: $\geq_K$ is good

逐一验证 (a)-(e) 那五条，看满不满足

以(b)为例

(b) anti symmetry

已知 $u \geq_K v$, $v \geq_K u$ 求证 $u = v$

由条件知：

$u - v \in K$, $v - u \in K$, 由(3)知： 因为是pointed

$v - u = 0$ （这里的减号就是这个空间的减号）

以(e)为例

(e) $u \geq_K v$, $w \geq_K z$ 求证 $u + w \geq v + z$

$u - v \in K$, $w - z \in K$

由(1)：$\Rightarrow$ 对加法封闭

$u + w - (v + z) \in K$

$\therefore u + w \geq_K v + z$

Example

(1) $E = \mathbb{R}^n$, $u \bullet v = u^T v$

(LP) $K = \mathbb{R}^n_+ \longleftrightarrow$ $u \geq v$ 对 $u_i \geq v_i \ \forall i$

$u \geq 0 \Leftrightarrow u \in K$

(2) Second-order cone/ice-cream cone/Lorentz cone

$E = \mathbb{R}^{n+1}$ $u \bullet v = u^T v$

$K = Q^{n+1} = \{(t, x) \in \mathbb{R} \times \mathbb{R}^n, \ t \geq \|x\|_2\}$

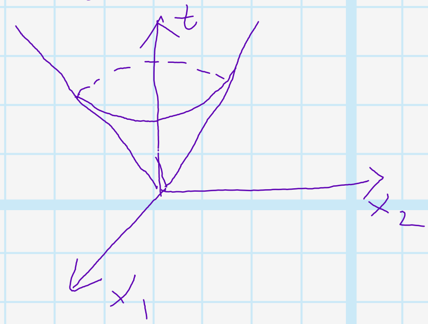
验证：则 $0\in K$ ∴ 非空

$(t_1,x_1),(t_2,x_2)\in Q^{n+1}$ 即 $t_1\ge\|x_1\|_2$，$t_2\ge\|x_2\|_2$

$(t_1+t_2,\ x_1+x_2)$ 是否在 Q^{n+1} 中？

∵ $\|x_1+x_2\|_2\le\|x_1\|_2+\|x_2\|_2\le t_1+t_2$

e.g. $n=2$ 图像是 $t\ge x_1^2+x_2^2$



(3) Semidefinite Cone

$E=S^n$ ($n\times n$ 对称矩阵)

$X\cdot Y=tr(X^TY)$ ∵对称 ∴也 $=tr(XY)$

$K=S^n_+$ (PSD matrices)

PSD 性质 $n=2$ 时 若 X 是 PSD

$$X=\begin{bmatrix}x_{11} & x_{12}\\ x_{21} & x_{22}\end{bmatrix}\quad x_{12}=x_{21}\quad x_{11}x_{22}\ge x_{12}^2\quad x_{11}\ge0\quad x_{22}\ge0$$

(4). Copositive Cone

$E=S^n$; $X\cdot Y=tr(X^TY)$

$K=C^n=\{X:\ u^TXu\ge0\ \underbrace{\forall u\ge0}\}$ → 比 PSD 那个多了这一项

Copositive Cone 是 convex cone

检验 X 是否属于 Copositive Cone 是 NP-hard 的.

Remark: $\mathbb{R}^n_+$, Q^{n+1}, S^n_+ 都有一些很好的性质

(1) closed (序列的收敛终到那个集合)

(2) non-empty interior (存在内点)

e.g. $\text{int}(\mathbb{R}^n_+) = \mathbb{R}^n_{++}$

$\text{int}(Q^{n+1}) = \{(t,x) : t > \|x\|_2\}$

$\text{int}(S^n_+) = S^n_{++}$ PD矩阵

下面讲如何构造 cone.

Let $K_i \subseteq E_i$, $i=1,\cdots,r$ be closed pointed cons w/ non-empty interior. Then So is

$$K = K_1 \times K_2 \times \cdots \times K_r = \{(u_1,\cdots,u_r) : u_i \in K_i\}$$

笛卡尔积

Conic Linear Programming

$(E, \bullet, K)$, Assume K is closed pointed cone w/ non-empty interior

下面对比 (LP) 和 (CLP)

CLP是LP的generalization版本

(LP)

$E = \mathbb{R}^n$

$u \bullet v = u^T v$

$K = \mathbb{R}^n_+$

$\inf\ c \bullet x$

s.t. $a_i \cdot x = b_i$

$x \in \mathbb{R}^n_+ \subseteq E$

$c, a_i \in E$ (∵只有在E上才定义•)

$b_i \in \mathbb{R}$

(CLP)

$\inf\ c \cdot x$

s.t. $a_i \cdot x = b_i$

$x \in K$

$c, a_i \in E$, $b_i \in \mathbb{R}$

在抽象出 CLP 之后，之前对 LP 的所有问题都可以对这个抽象的 CLP 来做。

下面来看(CLP)中的 Dual，先回忆 LP 中的对偶。

(LP) $\inf\ c^T x$

$\text{s.t. } Ax = b$

$x \geq 0$

find y: $A^T y \leq c$

强行构造下界

$c^T x \geq x^T A^T y$

$= b^T y$

$\downarrow$

$\max\ b^T y$

$\text{s.t. } A^T y \leq c$

(CLP) what's the analogous relationship for $A^T y \leq c$?

$b^T y = \sum_i b_i y_i = \sum_i (a_i \cdot x) y_i$

$= \sum_i (y_i a_i) \cdot x = (\sum_i y_i a_i) \cdot x$

现在的目标是验证这一项是否 $\leq c \cdot x$,

Define the dual cone of K

$K^* = \{ w \in E : x \cdot w \geq 0\ \ \forall x \in K \}$

and require

$\Leftrightarrow\ \underset{\in E}{c} - \underbrace{\sum_i y_i a_i}_{\in E} \in K^*$.

Then, for $x \in K$,

$(c - \sum y_i a_i) \cdot x \geq 0$

2017.10.25

Recall: E, $\bullet$, $K \subseteq E$

K is closed pointed cone with non-empty interior

$$(P)\quad V_p^* = \inf\ c\cdot x \quad \text{s.t. } a_i \cdot x = b_i,\ x \in K,\quad c, a_i \in E$$

上节课讲到 (P) 的对偶构造

$$\sum_i b_i y_i = \sum_i (a_i \cdot x) y_i = \left(\sum_i y_i a_i\right)\cdot x$$ 它是否 $\le c\cdot x$?

我们只需证明:

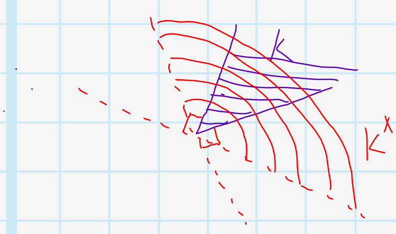
$$x\cdot\left(c - \sum_i y_i a_i\right) \ge 0 \quad (*)$$

Def Dual cone of K, $K^* = \{w \in E : x\cdot w \ge 0, \forall x \in K\}$

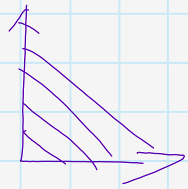
If $c - \sum_i y_i a_i \in K^*$, then $(*)$ holds

Example 下面给几个关于 K^* 的例子

① $E = \mathbb{R}^2$



② $E = \mathbb{R}^n \quad K = \mathbb{R}^n_+$



$K^* = \mathbb{R}^n_+ = K$

self-Dual

③ $(Q^{n+1})^* = Q^{n+1}$ also self dual

④ $(S^n_+)^* = S^n_+$ also self dual

Prop: Let $K \subseteq E$ be non-empty 没强调K是convex, closed等.

(a) K^* is closed convex cone 不管K是啥.

(b) If K closed convex cone, Then. $(K^*)^* = K$

反例:

$\mathbb{R}^2_{++} \xrightarrow{*} \mathbb{R}^2_+ \xrightarrow{*} \mathbb{R}^2_+$ ∴closed才行

(c). If $\mathrm{int}(K) \neq \emptyset$, Then K^* is pointed.

pointed: $\left.\begin{array}{l} u \in K \\ -u \in K \end{array}\right\} \Rightarrow u = 0$

(d). If K is closed pointed cone, Then $\mathrm{int}(K^*) \neq \emptyset$

如何用这些性质到上节课那个对偶的构造.

由(c), (d) 可知 K^* is pointed and $\mathrm{int}(K^*) \neq \emptyset$.

由(a), K^* is closed convex cone.

∴给定了K, K^* is the same type. (问题不会因引入了K^*变复杂)

(b) 证明思路:

(1) 证 $K \subset (K^*)^*$

$(K^*)^* = \{x : w \cdot x \geq 0, \forall w \in K^*\}$

$x \in K \Rightarrow w \cdot x \geq 0 \quad \forall w \in K^*$

$(K^*)^* \subset K$

If not, then, $\exists\, v \in (K^*)^*$, $v \notin K$

用分离定理 推出矛盾.

(C). Suppose K^* is not pointed.

$\Rightarrow \exists\, u \neq 0$, s.t. $u, -u \in K^*$

$\Rightarrow \quad x \cdot u = 0 \quad \forall x \in K$

$\Rightarrow K \subseteq H(u, 0) = \{x : u \cdot x = 0\}$

K 完全在一个 hyperplane 中, ∴ K 肯定没有内点.

∵由上, 我们有 $\sum_i b_i y_i \le c \cdot x$ 构成一个下界, 去找那个最大的下界

$$V_d^* = \sup \sum_i b_i y_i$$

$$\text{s.t. } c - \sum_i y_i a_i \in K^*$$

(D)

由上, K^* 和 K 一样, closed, pointed, $\text{int} \neq \phi$

Example:

① LP

(P) $\inf\ c^T x$

s.t. $a_i^T x = b_i$

$x \in \mathbb{R}_+^n$

(D) $\sup\ b^T y$

s.t. $c - \sum_i y_i a_i \in \mathbb{R}_+^n$

∵ $\mathbb{R}_+^n$ 是 self dual

回忆之前 我们写成 $A^T y \le c$, 是一样的

$$A = \begin{bmatrix} \cdots & a_1^T & \cdots \\ \cdots & a_2^T & \cdots \\ & \vdots & \\ \cdots & a_m^T & \cdots \end{bmatrix}$$

② $K = Q^{n+1}$ (SOCP) second order cone programming

$Q^{n+1} = \{(t, x) : t \geq \|x\|\}$ Q是个圆,小水滴形状

(P) $\inf\ c \cdot x$

s.t. $a_i \cdot x = b_i$

$x \in Q^{n+1}$

(D) $\sup\ b^T y$

s.t. $c - \sum_i y_i a_i \in Q^{n+1}$

∵ Q^{n+1} 也是 self-dual

下面讲 SOCP 这个名字的意思.

Let $c = (v, d)$, $v \in \mathbb{R}$, $d \in \mathbb{R}^n$

$$a_i = (u_i, \underbrace{a_{i1}, a_{i2}, \cdots, a_{in}}_{\text{记作 } \tilde{a}_i})$$

$$\sum_i y_i a_i = \sum_i y_i \begin{bmatrix} u_i \\ \hdashline \tilde{a}_i \end{bmatrix} = \begin{bmatrix} \sum_i y_i u_i \\ \hdashline \sum_i y_i \tilde{a}_i \end{bmatrix} = \begin{bmatrix} u^T y \\ \\ A^T y \end{bmatrix}$$

其中 $A^T = \begin{bmatrix} \vdots & & \vdots \\ \tilde{a}_1 & \cdots & \tilde{a}_m \\ \vdots & & \vdots \end{bmatrix}$

$\therefore c - \sum_i y_i a_i = (v - u^T y, d - A^T y)$

∴ SOCP 的 second cone 为 优化一个线性函数 $c \cdot x$, 而其约束是把变量(是y的 affine map)仿射映射到一个 second order cone Q^{n+1} 中

通常这个对偶(D) 比(P) 好解

对偶形式可写作：

$$(D)\quad \sup \sum_i b_i y_i$$

$$\text{s.t.}\quad (v_j - u_j^T y,\ d_j - A_j^T y) \in Q^{n_j+1}$$

$j=1,\dots,m$

加了 m 个约束.

所有的 LP 都可以写成 SOCP 的形式.

$$(LP)\quad \inf\ c^T x$$

$$\text{s.t.}\quad Ax=b$$

$$x \geq 0$$

下面来看怎么写成 SOCP 的形式.

先变换下 SOCP 的约束形式.

$$(v_j - u_j^T y,\ d_j - A_j^T y) \in Q^{n_j+1}$$

$$\Updownarrow$$

$$v_j - u_j^T y \geq \| d_j - A_j^T y \|_2$$

把 LP 中的那个等式约束写成不等式的形式. $Ax=b$

$$\|b - Ax\|_2 \leq 0$$

$$\Rightarrow (0,\ b-Ax) \in Q^{n+1}$$

而对于不等式约束 $x \geq 0$

$$x \geq 0 \iff (x_i,\ 0) \in Q^{n+1},\ \forall i$$

即是说 $x_i \geq \|0\|_2 = 0$

∴ 那个 LP 问题可写作如下 SOCP 的形式:

$$\inf \ c^T x$$
$$\text{s.t.} \ (0, b-Ax) \in Q^{n+1}$$
$$(x_i, 0) \in Q^2, \ i=1,\dots,n.$$

以上演示了所有LP问题都能转化成SOCP, 一般LP问题都比SOCP问题好求. ∴ 实际不用去转

③ SDP, $K = S^n_+ \subseteq S^n$

(P) $\inf \ C \bullet X$

s.t. $A_i \bullet X = b_i$

$X \in S^n_+$

(D) $\sup \ \sum_i b_i y_i$

s.t. $C - \sum_i y_i A_i \in S^n_+$

∵ S^n_+ 是 self dual

同上, SOCP 也都能写成SDP, 但SOCP更易求解

2017.10.30

(P) $V_p^* = \inf\ c \cdot x$

$\text{s.t.}\ a_i \cdot x = b_i$

$x \in K$

(D) $V_d^* = \sup\ b^T y$

$\text{s.t.}\ c - \sum_i y_i a_i \in K^*$

SOCP $\qquad Q^{n+1} = \{(t, x) : t \geq \|x\|_2\}$

$\inf\ c^T x$

$\text{s.t.}\ a_i^T x = b_i$

$x \in Q^{n_1+1} \times Q^{n_2+2} \times \cdots \times Q^{n_i+1}$ $\qquad$ x: 笛卡尔集

下面考虑最简单的形式：$x \in Q^{n+1}$

对偶问题：

$\sup\ b^T y$

$\text{s.t.}\ (v - u^T y, d - A^T y) \in Q^{n+1}$

$c = (v, d) \ldots$

SDP

$\inf\ C \cdot X$

$\text{s.t.}\ A_i \cdot X = b_i$

$X \succeq 0 \quad (X \in S_+^n)$

$$\sup\ b^T y$$
$$\text{s.t. } C - \sum_i y_i A_i \in S_+^n$$

下面演示从 SOCP 的对偶到 SDP 的对偶.

Fact: (Schur Complement)

$$\begin{array}{c|cc} & m & n \\ \hline m & A & B \\ n & B^T & C \end{array} \quad \text{i.e.}\quad \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0 \iff \begin{cases} A - BC^{-1}B^T \succeq 0 \\ A, C \succ 0 \end{cases}$$

$$\Uparrow$$
$$S^{m+n}$$

$$C - B^T A^{-1} B \succeq 0$$

Suppose y satisfies

$$(v - u^T y, d - Ay) \in Q^{n+1} \qquad (1)$$

consider

$$\begin{array}{c|cc} & n & 1 \\ \hline n & (v-u^T y)I & d - A^T y \\ 1 & (d - A^T y)^T & v - u^T y \end{array} \quad \begin{bmatrix} (v-u^T y)I & d - A^T y \\ (d - A^T y)^T & v - u^T y \end{bmatrix} \succeq 0 \qquad (2)$$

(1) $\Rightarrow$ (2)

$$v - u^T y \geq \|d - A^T y\|_2 \qquad (\text{SOCP 中的定义})$$

step1° $v - u^T y = 0 \Rightarrow d - A^T y = 0 \Rightarrow$ (2) holds.

step2° $v - u^T y > 0 \Rightarrow (v - u^T y)^2 \geq \|d - A^T y\|_2^2$

$$\Rightarrow v = u^T y \geq \frac{\|d-A^Ty\|_2^2}{v-u^Ty} = (d-A^Ty)^T\left[(v-u^Ty)I\right]^{-1}(d-A^Ty)$$

正是 (2) 的 Schur 补 ∴ (2) holds.

(2) ⇒ (1)

补充个性质：

如果 PSD 矩阵中第 i 个对角元素是 0，那么第 i 行和第 i 列都必须是 0

step 1° 若 $v-u^Ty=0$，由补充的性质 ⇒

(2) 中 $d-A^Ty=(d-A^Ty)^T=0$

step 2°

把 (1) ⇒ (2) 中的 step 2° 反过来写即可.

下面从 (2) 到 SDP

把 (2) 拆开：

$$\begin{bmatrix} vI & d \\ d^T & v \end{bmatrix} - \begin{bmatrix} (u^Ty)I & A^Ty \\ (A^Ty)^T & u^Ty \end{bmatrix} \succeq 0$$

下面看怎么把这个条件写成 $C-\sum_i y_i A_i$ 的形式.

$$= C - \sum_i y_i \begin{bmatrix} u_i & & & \vdots \\ & u_i & & \text{ith col of } A^T \\ & & \ddots & \vdots \\ & & u_i & \vdots \\ \cdots & \cdots & \cdots & u_i \end{bmatrix}$$

这个方法说明，如果矩阵中每个分块都是 y 的仿射函数，那么，这就是 (SD)P

以上说明了 (2) 也是一个 SDP

Thm: (CLP Weak Duality)

Let x be feasible for (P)

y be feasible for (D)

Then, $b^T y \le C \cdot X$

Farkus' lemma. (换一种写法)

(LP)

(I) $a_i^T x = b_i, \quad x \in \mathbb{R}^n_+$

(II) $-\sum_i y_i a_i \in \mathbb{R}^n_+, \quad b^T y > 0$ 两条有且仅有一个成立

那么它的 conic 版本是：

(I) $a_i \cdot x = b_i, \quad x \in K$

(II) $-\sum_i y_i a_i \in K^*, \quad b^T y > 0$

这个conic的版本还需要多一些的条件，先看一个例子.

"Pf":

(1) 证明不能同时有解.

Suppose both solvable

$$0 < b^T y = \sum_i y_i (a_i \cdot x) = -\Big[\underset{\in K^*}{\Big(-\sum_i y_i a_i\Big)} \cdot \underset{\in K}{x}\Big] \le 0 \text{ 矛盾}$$

K^*中的向量 · K中向量 非负

e.g. ("Failure" of Conic Farkas)

$E = S^2, \quad K = S^2_+$

$A_1=\begin{bmatrix}1&0\\0&0\end{bmatrix}$ $A_2=\begin{bmatrix}0&1\\1&0\end{bmatrix}$ $b=\begin{bmatrix}0\\2\end{bmatrix}$

(I) $A_1\cdot X=b_1$, $A_2\cdot X=b_2$ $\begin{bmatrix}x_{11}&x_{12}\\x_{21}&x_{22}\end{bmatrix}\succeq 0$

$\Downarrow$ $\Downarrow$ $\Downarrow$

$X_{11}=0$ $x_{12}=1$ $\begin{bmatrix}0&1\\1&x_{12}\end{bmatrix}\succeq 0$

$x_{00}=0$ 那必须有 $x_{10}=x_{10}=0$, ∴不是PSD

∴这个系统无解.

(II) $-y_1A_1-y_2A_2\in S_+^2$ $y_2>0$

$\begin{bmatrix}-y_1&-y_2\\-y_2&0\end{bmatrix}\succeq 0 \xRightarrow{\text{是PSD}} y_2=0$ ∴这个系统无解.

In the proof of (LP) Farkas' Lemma, we need

$S=\{Ax : x\geq 0\}$ to be closed

Q: For closed pointed cone K, is

$S=\{(a_1\cdot x, a_2\cdot x, \dots, a_m\cdot x) : x\in K\}$ closed?

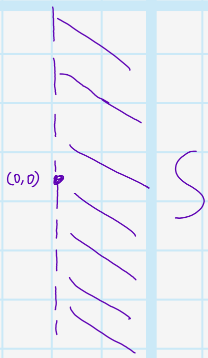
回到上面那个例子.

$S=\{(x_{11}, 2x_{12}) : X\succeq 0\}$

$(x_{11}x_{22}-x_{12}^2\geq 0)$

$S=\{(0,0)\}\cup\{(x,y): x>0, y\in\mathbb{R}\}$

S不是闭的.



∴要保证Farkas在conic问题上能工作的话需要加额外的条件来保证closeness

$\underrightarrow{\text{条件}}$ Slater condition

Conic Farkas' Lemma

$E, \bullet, K$, Suppose K is closed pointed cone with $\mathrm{int}(K)\neq\emptyset$ and the Slater condition holds:

$\exists \bar{y}$ s.t. $-\sum_i \bar{y}_i a_i \in \mathrm{int}(K^*)$, Then,

exactly one of (I) (II) has solution.

在上面那个例子中：

$$-y_1 A_1 - y_2 A_2 = \begin{bmatrix} -y_1 & -y_2 \\ -y_2 & 0 \end{bmatrix}$$

K^* 是 PD must be PD 而有0，∴这个反例对 Slater condition 不成立.

2017.11.1

Conic Farkas

Slater condition

[Suppose $\exists \bar{y}$ s.t. $-\sum_i \bar{y}_i a_i \in \text{int}(K^*)$] Then, exactly one of (I) (II) is soluvable.

(I). $a_i \cdot x = b_i \quad x \in K$

(II). $-\sum_i y_i a_i \in K^*, \quad b^T y > 0$

slater codition 的作用是保证 closeness，
这是一个充分条件，也就是说有可能不满足 slater，仍然成立 Farkas.

Thm (Strong Duality for CLP)

(P) $V_p^* = \inf \ C \cdot X$

s.t. $a_i \cdot x = b_i$

$x \in K$

(D) $V_d^* = \sup \ b^T \cdot y$

s.t. $C - \sum_i y_i a_i \in K^*$

① Suppose (p) is $\underbrace{\text{bounded below}}_{V_p^* > -\infty}$, and satisfies Slater: $\exists \bar{x}$ = $a_i \cdot \bar{x} = b_i$, $\bar{x} \in \text{int}(K)$ (aka strictly feasible)

Then, (i) $V_p^* = V_d^*$ (zero duality gap)

(ii) $\exists \bar{y}$ feasible for (D), s.t.

$b^T\bar{y} = V_d^* = V_p^*$ (dual attainment)

Primal attainment is not guaranteed

回忆 LP strong Duality

Suppose (P) (D) feasible, Then both have optimal solutions, and duality gap $= 0$.

LP中不需要 Slater, 但是 conic LP要

LP 中 (P) (D) 都存在x,y 去取到最优

注意在(ii)中, 只保证了对偶中能找到可取到最优值, 但 Primal 并不保证 (可能只能靠确界来取到最优)

上面的 Primal 和 Dual 也可以反过来, 即 dual is bounded above and satisfy slater, 这样就是保证 Primal attainment.

Slater 在哪边, attainment 取到另一边.

② Suppose (P), (D) are bounded (取不到 $\pm\infty$) and strictly feasible. Then TFAE:

(1) $\bar{x}$ opt for (P), $\bar{y}$ opt. for (D)　　两边都加了 strictly feasible (slater) ∴都能取到

(2) duality gap $=0$

(3) $\bar{x}\cdot(c-\sum_i \bar{y}_i a_i)=0$　(complementarity)

Examples:

$V_p^* = \inf\ x_{12}$

s.t. $\begin{bmatrix} 0 & x_{12} & 0 \\ x_{12} & x_{22} & 0 \\ 0 & 0 & 1+x_{12} \end{bmatrix} \succeq 0$　　$V_p^*=0$

∵ $x_{11}=0$ ∴ x_{12} 必为 0

attained by e.g. $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

∵ 第一行第一列都为 0 ∴ not strictly feasible

∴ 无法找到一个 DP

下面来写成标准的形式

$$V_p^* = \inf \underbrace{\begin{bmatrix} & \frac{1}{2} & \\ \frac{1}{2} & & \\ & & \end{bmatrix}}_{C} \cdot \underbrace{\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{12} & x_{22} & x_{23} \\ x_{13} & x_{23} & x_{33} \end{bmatrix}}_{X}$$

s.t. $\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{A_1} \cdot X = \overset{b_1}{0}$, $\underbrace{\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}}_{A_2} \cdot X = \overset{b_2}{0}$

等价 $x_{11}=0$

$\underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}}_{A_3} \cdot X = \overset{b_3}{0}$　$\underbrace{\begin{bmatrix} 0 & -\frac{1}{2} & 0 \\ -\frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{A_4} \cdot X = \overset{b_4}{1}$　$b=\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

即 $x_{33}=1+x_{12}$

$$(D) \quad V_d^* = \sup\ b^T y = \sup\ [0\,0\,0\,1]\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \sup\ y_4$$

$$\text{s.t.} \quad C - \sum_i y_i A_i \in K^*$$

$$= \begin{bmatrix} & 1/2 & \\ 1/2 & & \\ & & \end{bmatrix} - \begin{bmatrix} y_1 & -y_4/2 & y_2 \\ -y_4/2 & 0 & y_3 \\ y_2 & y_3 & y_4 \end{bmatrix} \succeq 0$$

$$-\begin{bmatrix} y_1 & -y_4/2 - \frac{1}{2} & y_2 \\ -y_4/2 - \frac{1}{2} & 0 & y_3 \\ y_2 & y_3 & y_4 \end{bmatrix} \succeq 0$$

$\therefore -y_4/2 - \frac{1}{2} = 0 \quad \therefore V_d^* = y_4 = -1$

代入 y_4 得：

$$\begin{bmatrix} -y_1 & 0 & -y_2 \\ 0 & 0 & 0 \\ -y_2 & 0 & 1 \end{bmatrix} \succeq 0$$

$\therefore V_d^*$ is attained by $y_1 = y_2 = y_3 = 0 \quad y_4 = -1$

这个例子中，(P)(D) 都不满足 Slater，但是都能取到最优值。

即不是PD.不是内点

下面给一例子满足 Slater，而 (P) 无法 attain

Example: $V_p^* = \inf\ x_1$

$$\text{s.t.} \quad (x_1 + x_2,\ 1,\ x_1 - x_2) \in Q^3$$

约束可写作：$x_1 + x_2 \geq \sqrt{1 + (x_1 - x_2)^2}$

即 $x_1 + x_2 \geq 0 \quad (x_1 + x_2)^2 \geq 1 + (x_1 - x_2)^2$

$$4x_1 x_2 \geq 1$$

bounded below: $4x_1x_2 \geq 1 \Rightarrow x_1x_2$ 同号

$x_1+x_2 \geq 0 \Rightarrow x_1, x_2$ 正号

$\therefore$ bounded below

int(Q^3) 是 $4x_1x_2>1$ $\therefore$ 存在内点 $\therefore$ strictly feasible.

$V_p^* = 0$ 当 $x_1 = \frac{1}{4x_2}$ $x_2 \to +\infty$ 时取到极限,

但是 x_1 不能取到 0.

V_p^* is not attained, because if $x_1 = 0$,

$(x_2, 1, -x_2) \in Q^3$ not possible.

这个的对偶怎么写 考试会考 ☆

在机器学习中,有时候(P)很难解,(D)很容易解.

下面来写对偶,公式原型如下:

(D) $\sup \quad b^T y$

$\text{s.t.} \quad (v - u^T y,\ d - A^T y) \in Q^{n+1}$

(P) $\inf \ c^T x \qquad c=(v,d) \quad v \in \mathbb{R}, d \in \mathbb{R}^n$

$\text{s.t.} \ h_i^T x = b_i \qquad h_i = \begin{bmatrix} u_i \\ a_{i,1} \\ \vdots \\ a_{i,n} \end{bmatrix} \begin{matrix} \}\mathbb{R} \\ \}\mathbb{R}^n \end{matrix} \qquad x \in Q^{n+1}$

例 1 中的形式,更容易写成(D), $\therefore$ 先写成 D, 再用公式转成(P)

令 $y=(y_1,y_2) \qquad v=0 \qquad u=(-1,-1) \qquad d=(1,0)$

$A^T = \begin{array}{cc} \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix} \\ a_1 \quad a_2 \end{array} \qquad b=(-1,0)$

验证: $v - u^T y = y_1 + y_2$

$d - A^T y = (1, y_1 - y_2)$

∴ (P) 是：

$$\inf \quad x_2$$

s.t. $-x_1 - x_3 = -1$

$-x_1 + x_3 = 0$

$(x_1, x_2, x_3) \in Q^3$

$h_i^T = [u_i \; a_i^T]$

$= \begin{bmatrix} u_1 & - a_1^T - \\ u_2 & - a_2^T - \end{bmatrix}$

观察发现这个可行域只有一个点，即 $x_1 = \frac{1}{2}$
$x_3 = \frac{1}{2}$ $x_2 = 0$

∴ (P) 的 $\inf x_2 = 0$

∴ 取到 optimal value.

2017.11.6

Robust Optimization

LP: $\min \tilde{c}^T x$ (*)

$\text{s.t. } \tilde{A} x \le \tilde{b}$

Data: $(\tilde{c}, \tilde{A}, \tilde{b})$ 在普通优化问题中，这三个是已知的，但实际上它们通常是被估出来的，存在 uncertainty

一些假设方法和对应的解法

① data as random vector

$\rightarrow$ stochastic optimization 随机梯度下降属于此类

② worst-case realization of data

$\rightarrow$ robust optimization

Assume WLOG no uncertainty in objective

e.g. for LP:

$\min t$

$\text{s.t. } \tilde{c}^T x \le t$

$\tilde{A} x \le \tilde{b}$

把所有的 uncertainty 都移到约束当中，目标里不含 uncertainty。这样，(*)中的 $\tilde{c}$ 可写作 $\bar{c}$

Rewrite LP (*) as

$\min c^T z$

$\text{s.t. } Az \le 0$

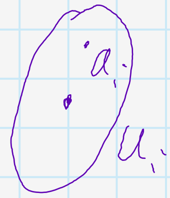
$z_{n+1} = -1$

where $c = \begin{bmatrix} \bar{c} \\ 0 \end{bmatrix}$ $z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \\ z_{n+1} \end{bmatrix} \in \mathbb{R}^{n+1}$

$A = [\tilde{A} \ \ \tilde{b}] \leftarrow$ uncertainty Model

Assumption: Each row a_i of A lies in ellipsoid $\mathcal{U}_i$

为什么用椭圆？
因为用其它形状没法解



specifically, assume

$$\mathcal{U}_i = \left\{ z \in \mathbb{R}^{n+1} : (z - u_i)^T B_i^{-2} (z - u_i) \le 1 \right\}$$

where u_i, $B_i \succ 0$, are given.

Robust Counterpart of LP: (RLP)

$$\min \ c^T z$$
$$\text{s.t. } Az \le 0 \quad \text{for all } a_i \in \mathcal{U}_i \quad A = [\tilde{A} \ \tilde{b}]$$
$$z_{n+1} = -1$$

相当于有无限个约束，
∴不是一个 LP 问题了

Note:

$$\mathcal{U}_i = \left\{ z : z = u_i + B_i v, \ \|v\|_2 \le 1 \right\}$$

since $B_i^{-1}(z - u_i) = v \quad \|v\|_2 \le 1$

$$\therefore \|B_i^{-1}(z - u_i)\|_2^2 \le 1$$

$$\Rightarrow (z - u_i)^T B_i^{-2} (z - u_i) \le 1$$

Consider ith constraint:

$$a_i^T z \le 0 \quad \forall a_i \in \mathcal{U}_i \quad \text{(robust constraint)}$$

$$\Leftrightarrow \quad (u_i + B_i v)^T z \le 0 \quad \forall \|v\|_2 \le 1$$

$$\Leftrightarrow \quad \max_{\|v\|_2 \le 1} \left\{ (u_i + B_i v)^T z \right\} \le 0$$ 现在只有一个约束了 $\|v\|_2 \le 1$

v 是引入的变量，最终还是对 z

$\Leftrightarrow \quad u_i^T z + \max\limits_{\|v\|_2 \le 1} (v^T B_i z) \le 0$ 把和v无关的提出来.

$\Leftrightarrow \quad u_i^T z + \|B_i z\|_2 \le 0$

$\max\limits_{\|v\|_2 \le 1} v^T w$ 当v和w同向的时候取到最大,最大值为 $\frac{w^T w}{\|w\|}$

$v^* = \frac{w}{\|w\|_2}$

$\Leftrightarrow \quad \|B_i z\|_2 \le -u_i^T z$ 这是一个SOC的约束.

至此一个无限约束的LP被转成了一个SOCP问题

Quad Constrained Quad Opt.

目标函数和约束都是二次

(QCQP)

$$\min \quad x^T Q x$$

$$\text{s.t.} \quad x^T A_i x \ge b_i \qquad Q, A_i \in S^n$$

e.g. 可以处理 binary constrains

$x_i \in \{0, 1\}$ 只能取二值

$$\Updownarrow$$

$$x_i(1-x_i) = 0$$

$$x_i \in \{\pm 1\}$$

$$\Updownarrow$$

$$x_i^2 = 1$$

观察到 $x^T Q x = tr(x^T Q x) = tr(Q x x^T) = Q \bullet x x^T$

$\because x^T Q x$ 是标量 $\qquad \bullet$ 的定义

∵ $x^T A_i x = A_i \bullet x x^T$

∴ 原问题转化为

$$\min \quad Q \bullet X$$

$$\text{s.t.} \quad A_i \bullet X \geq b_i$$

$$X = x x^T$$

这个不是凸的，∵最简单的 $x^2=1$，这是个圆环，不是凸的.

$X = xx^T \Rightarrow X \succeq 0 \qquad \text{rank}(X) \leq 1$

∵ $u^T(xx^T)u = (u^T x)^2$

$$\min \quad Q \bullet X$$

$$\text{s.t.} \quad A_i \bullet X \geq b_i$$

$$X \succeq 0$$

$\text{null}(\overset{xx^T}{X}) = \{u : xx^T u = 0\}$

$\text{rank}(X) \leq 1$ 非convex，两个rank=1的矩阵相加不是rank1

做一个relaxation，把 $\text{rank}(X) \leq 1$ 直接~~删~~掉了.

叫做 the semidefinite relaxation of QCQP

SDR

Let X^* be an optimal solution to (SDR)

and $\text{rank}(X^*) = 1$, Then, $X^* = x^*(x^*)^T$

and x^* is optimal for QCQP

如果能找到这么一个 x^* 的话，相当于满足那个被~~删~~掉的条件.

2017.11.8

QCQP

$V^*_{QP} = \inf \ x^T Q x$ $\xrightarrow{转化为}$ $V^*_{SDR} = \inf \ Q \cdot X$

s.t. $x^T A_i x \geq b_i$ s.t. $A_i \cdot X \geq b_i$

$Q, A_i \in S^n$ $X \succeq 0$

[$\mathrm{rank}(X) \leq 1$] drop 掉

问题：

1. 如何获得 rank=1？
2. 如果获得不了怎么办？

SDP $\inf \ C \cdot X$

s.t. $A_i \cdot X = b_i \quad i = 1, \cdots, m$

$X \succeq 0$

Thm. (Shapiro 82, Barvinok 95, Pataki 98)

Suppose (SDP) has an optimal solution, Then, $\exists$ optimal X^* with $r = \mathrm{rank}(X^*)$ and $\frac{r(r+1)}{2} \leq m$

∴ 如果 m 足够小，r 也会被限制得很小。

proof: Let $\bar{X}^*$ be opt, let $\bar{r} = \mathrm{rank}(\bar{X}^*)$.

Suppose $\frac{\bar{r}(\bar{r}+1)}{2} > m$, Let $\underset{n \times m}{\bar{X}^*} = LL^T$ be a Cholesky factorization of $\bar{X}^*$, where $L \in \mathbb{R}^{n \times \bar{r}}$

Define $\underset{S^{\bar{r}}}{\bar{C}} = L^T C L$, $\underset{S^{\bar{r}}}{\bar{A}_i} = L^T A_i L$, Consider

(ASDP) auxiliary

$$V^*_{ASDP} = \inf \ \bar{C} \bullet W$$
$$\text{s.t.} \ \bar{A}_i \bullet W = b_i$$
$$W \succeq 0$$

Note: ① $W = I_{\bar{r}}$ is feasible

$$\bar{A}_i \bullet I = (L^T A_i L) \bullet I =$$
$$\text{tr}(L^T A_i L) = A_i \bullet \bar{X}^* = b_i$$

$$\bar{C} \bullet I = C \bullet \bar{X}^* = \text{optimal value of (SDP)}$$

现在,老师把问题转化到了简单的空间 $S^{\bar{r}}$ 上,

② If $\bar{w}$ feas for (ASDP)

then $\bar{X} = L\bar{w}L^T$ is feas for (SDP)

∵ ⓐ $\bar{X} \succeq 0$

ⓑ $A_i \bullet \bar{X} = A_i \bullet L\bar{w}L^T = L^T A_i L \bullet \bar{w} = \bar{A}_i \bullet \bar{w} = b_i$

claim: ① $V^*_{SDP} = V^*_{ASDP}$

② Every feasible solution to (ASDP) is opt.

想法: 利用 $I_{\bar{r}}$ 来构造一个 rank=1 的

Construct a feas. solution of rank $< \bar{r}$ to ASDP

1°. solve $\bar{A}_i \cdot w = 0$, $w \in S^{\bar{r}} \setminus \{0\}$

call this solution $\bar{w}$

m equations

$\frac{\bar{r}(\bar{r}+1)}{2}$ variables,

since $\frac{\bar{r}(\bar{r}+1)}{2} > m$, find $\bar{w} \neq 0$

2°. Obs: $\bar{W}(\alpha) = I + \alpha \bar{w}$

satisfies $\bar{A}_i \cdot \bar{W}(\alpha) = b_i$ for any α.

3°. Pick α s.t. $I + \alpha \bar{w} \succeq 0$ and $\text{rank} < \bar{r}$

$\bar{w} \neq 0$, assume $\bar{w}$ has one non-neg eigenvalue

$\Rightarrow \lambda_{min}(\bar{w}) < 0$ $\quad \bar{w}^+ = I - \frac{1}{\lambda_{min}(\bar{w})} \bar{w}$

$\alpha = -\frac{1}{\lambda_{min}(\bar{w})}$

$\bar{w}^+$就是要找的那个减掉一个rank的矩阵.

找到这个α让 $I+\alpha\bar{w}$ 的秩能减小一个.

Take $\|u\|_2 = 1$: $u^T \bar{w}^+ u = 1 - \frac{1}{\lambda_{min}(\bar{w})} u^T \bar{w} u$

$$u^T \bar{w} u \geq \min_{\|u\|_2} u^T \bar{w} u = \lambda_{min}(\bar{w})$$

$(\lambda_{min} < 0)$

Courant-Fischer

代入 $\Rightarrow u^T \bar{w}^+ u \geq 1 - \frac{\lambda_{min}}{\lambda_{min}} = 0$

$$\bar{w}^+ = U \left\{ \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} - \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_{min} \end{bmatrix} \frac{1}{\lambda_{min}} \right\} U^T$$

$$= U \begin{bmatrix} 1 - \frac{\lambda_1}{\lambda_{min}} & & & \\ & 1 - \frac{\lambda_i}{\lambda_{min}} & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} U^T$$

以下来证 Claim 中的 ②

Every feas solution to (ASDP) is optimal

$$\sup b^T y$$

(DASDP) $$\text{s.t.} \quad \bar{C} - \sum_i y_i \bar{A}_i \succeq 0$$

① by strong duality (DASDP) has opt soln, y^*.

② $W^* = I$ is optimal for (ASDP)

③ Complementarity implies:

$$I \cdot \left(\bar{C} - \sum_i y_i^* \bar{A}_i\right) = 0$$

primal opt solution　　dual sla?

$\therefore \bar{C} - \sum_i y_i^* \bar{A}_i = 0$　　$\because \bar{C} - \sum_i y_i^* \bar{A}_i$ 所有 eigenvalue ≥ 0

$\therefore$ 把 I 换成其它的 feas. 后, complentarity 依然成立. $\therefore$ 那些 feas sol 也是 opt sol

$\therefore$ 在上节课那个 QCQP 中, 如果只有两个约束, 可以保证 rank ≤ 1
　　　　　　　　　　　　　　　　≤ 2 个都行

下面讲第一个用问题.

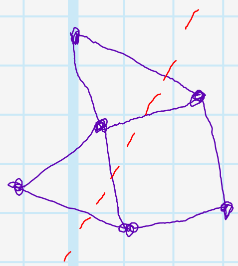
Maximum Cut (MAX-CUT)

$G = (V, E)$;　$W_{ij} \geq 0$　$(i, j) \in E$

一个cut把图分成两部分, $S \subseteq V$

value of cut:

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



两端点在不同组的边的值的和

Goal: find S to $\max w(S)$

Goemans-Williamson 94

$$x_i = \begin{cases} +1 & \text{if } i \in S \\ -1 & \text{otherwise} \end{cases}$$

$$\max \ \frac{1}{2} \sum_{(i,j) \in E} w_{ij} (1 - x_i x_j)$$

$$\text{s.t.} \quad x_i^2 = 1 \quad \forall i \in V$$

2017.11.11

MAX-CUT

$G=(V,E)$, $w_{ij} \geq 0$

Goal: Find cut $S \subseteq V$ with maximum value

$$v^* = \max \frac{1}{2} \sum_{(i,j)\in E} w_{ij}(1 - x_i x_j)$$

$$\text{s.t. } x_i^2 = 1$$

Recall: SDR semidefinite relaxiation

$$\begin{aligned} \min\ & x^T Q x \\ \text{s.t. } & x^T A_i x \geq b_i \end{aligned} \longrightarrow \begin{aligned} \min\ & Q \cdot X \\ \text{s.t. } & A_i \cdot X \geq b_i \\ & X \succeq 0 \end{aligned}$$

$X = xx^T$

$X_{ij} = x_i x_j$

$\Updownarrow$

$X \succeq 0$, $\operatorname{rank}(X) \leq 1$

relaxiation

这个SDR把非线性的x_ix_j线性化为X_{ij}，
可以借鉴这个思路。

$$V^*_{SDP} = \max \frac{1}{2} \sum_{(i,j)\in E} w_{ij}(1 - X_{ij})$$

$$\text{s.t. } X_{ii} = 1, \quad X \succeq 0$$

这里的约束也被松弛了，rank=1被删了

$\therefore v^* \leq V^*_{SDP}$

Q: Given X^* opt for (SDP), how to get a feasible solution to QCQP?

获得了松弛后的最优，
如何去求原问题的最优

GW Rounding Scheme.

① Write $X^* = U^T U$, $U \in \mathbb{R}^{n\times n}$ 乔里斯基分解.

Let u_i be the i^{th} col of U $\quad U = \begin{bmatrix} \vdots & \vdots & & \vdots \\ u_1 & u_2 & \cdots & u_n \\ \vdots & \vdots & & \vdots \end{bmatrix}$

Note: $\|u_i\|_2 = 1$ $\because X^*_{ii} = 1$ 而 $X^*_{ii} = u_i^T u_i$

② Generate random $r \in \mathbb{R}^n$, uniformly on $\{x : \|x\|_2 = 1\}$

如何产生随机的 $\{x : \|x\|_2 = 1\}$

先产生一个高维高斯 $g \sim N(0, I_n)$, set $r = \frac{g}{\|g\|_2}$

这样产生, 保证了每一维独立, 且对方向没有偏好.

③ Set $x_i' = \mathrm{sgn}(u_i^T r)$, $\mathrm{sgn}(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ -1 & \text{otherwise} \end{cases}$

Note: $x_i' \in \{\pm 1\}$ so, x' is feasible for QCQP

用 r 生成一个 hyperplane 将 u_i 分成两部分.

Thm. (GW) Let $v' = \mathbb{E}[\mathrm{val}(x')] = \mathbb{E}\left[\frac{1}{2} \sum_{(i,j)\in E} w_{ij}(1 - x_i' x_j')\right]$

Then, $0.878 v^* \leq v' \leq v^*$

Pf: $v' = \sum_{(i,j)\in E} w_{ij} \mathbb{E}\left[\frac{1 - x_i' x_j'}{2}\right] = \sum_{(i,j)\in E} w_{ij} \Pr\left[\mathrm{sgn}(u_i^T r) \neq \mathrm{sgn}(u_j^T r)\right]$

它是个0-1随机变量, 其期望等于其为1的概率

Claim: Let $\|u\|_2 = \|v\|_2 = 1$, r unif random on unit sphere.

$\Pr\left[\mathrm{sgn}(u^T r) \neq \mathrm{sgn}(v^T r)\right] = \frac{1}{\pi} \arccos(u^T v)$

Also, for $z \in [-1, 1]$, let $z = \cos\theta$. and

$\frac{1}{\pi} \arccos(z) = \frac{\theta}{\pi} = \boxed{\frac{2\theta}{\pi(1-\cos\theta)}} \cdot \frac{1}{2}(1 - \cos\theta) \geq 0.878 \cdot \frac{1}{2}(1 - z)$

$\arg\min_{\theta\in[0,\pi]} \frac{2\theta}{\pi(1-\cos\theta)} = 0.878$

By claim,

$v' = \sum_{(i,j)\in E} w_{ij} \Pr[\cdots \neq \cdots]$

$$= \sum_{(i,j)\in E} w_{ij} \frac{1}{\pi} \arccos(u_i^T u_j)$$

$$\geq 0.878 \sum_{(i,j)\in E} \frac{1}{2} w_{ij} (1 - \underbrace{u_i^T u_j}_{X_{ij}^*})$$

$$\geq 0.878\, V_{SDP}^* \geq 0.878\, V^*$$

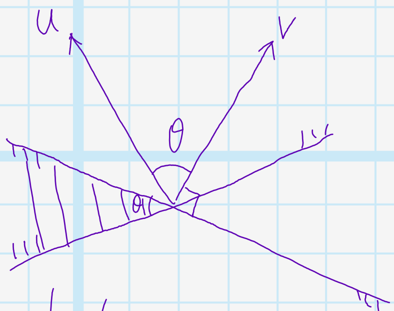
下面来证明上面那个 claim.

$$\Pr[\mathrm{sgn}(u^T r) \neq \mathrm{sgn}(v^T r)] = 2\Pr[u^T r \geq 0,\ v^T r < 0]$$

Consider the plane spanned by u, v, project r onto this plane. Let P be the projection matrix onto the plane. Then,

$$u^T r = (Pu)^T r = u^T P^T r = u^T P r$$

$\because u$ is inside the plane. $\because P$ 是 proj matrix $\therefore P = P^T$



$$2\Pr[u^T r \geq 0,\ v^T r < 0] = \frac{\theta}{\pi}$$

θ: angle between u and v

$$= \frac{\arccos(u^T v)}{\pi}$$

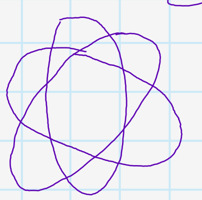
下面讲:

Extension of the Randomized Rounding Technique

Consider:

$$\max\ x^T Q x$$
$$\text{s.t. } x^T A_i x \leq b_i$$
$$Q, A_i \succeq 0$$

一般convex的问题是min一个convex目标, s.t. convex约束.
∴这个问题不是凸的.



约束是一系列椭圆的交.

SDR:

$$\max\ Q\cdot X$$
$$\text{s.t. } A_i\cdot X\le b_i$$
$$X\succeq 0$$

Let X^* be opt soln to SDR.

Generate $g\sim N(0, X^*)$ ∵ X^*是PSD的

g是随机变量, 下面的 g^TQg 都是随机变量.

$$\mathbb{E}[g^TQg]=\mathbb{E}[Q\cdot gg^T]=Q\cdot\mathbb{E}[gg^T]=Q\cdot X^*$$

$$\mathbb{E}[g^TA_ig]=A_i\cdot X^*\le b_i$$

$$\Pr\left[\left|g^TA_ig-\mathbb{E}[g^TA_ig]\right|>t\right]\le e^{-f(t)}$$

concentration inequality

尽管 $g^TA_ig\le\alpha b_i$ 不一定成立, 但它可以按一个概率来成立.

$$\Downarrow$$

$$\left(\frac{g}{\sqrt{\alpha}}\right)^T A_i\left(\frac{g}{\sqrt{\alpha}}\right) \qquad \frac{g}{\sqrt{\alpha}}\ \text{feasible}$$

2017.11.13

Recall:

$$\inf \ C \cdot X \qquad\qquad \sup \ b^T y$$
$$\text{s.t. } a_i \cdot X = b_i \qquad\qquad \text{s.t. } c - \sum y_i a_i \in K^*$$
$$X \in K$$

Optimality Conditions

(Primal feasibility) $a_i \cdot x = b_i$, $x \in K$

(dual feasibility) $C - \sum_i y_i a_i \in K^*$

(complementarity) $X \cdot (C - \sum_i y_i a_i) = 0$

Fact: If (x, y) satifies (PF) (DF) (C) then
(x, y) are optimal for (P), (D) resp.
Conversely, if regularity condition hold and
(x, y) are optimal for (P) (D), then,
(PF), (DF), (C) hold.

下面看可微多元函数的何求最小. 先看一元函数:

$\min_x f$, $f: \mathbb{R} \to \mathbb{R}$ differentiable

(1) $\frac{df}{dx} = 0$ (First-Order Necessary Condition: FONC)

(2) $\frac{d^2 f}{dx^2} > 0$ (Second-Order Sufficient condition: SOSC)

Prop: Suppose $f: \mathbb{R}^n \to \mathbb{R}$ is contiously differentiable (diffentialable and derivative is contious)

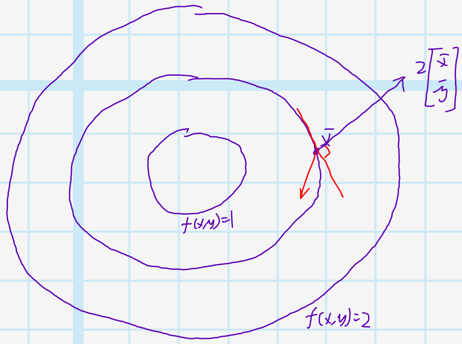
consider $\bar{x} \in \mathbb{R}^n$, If $\exists d \in \mathbb{R}^n$

s.t. $\nabla f(\bar{x})^T d < 0$, Then d is a descent direction;

i.e. $f(\bar{x} + \alpha d) < f(\bar{x})$ for $\alpha \in (0, \alpha_0)$, $\alpha_0 > 0$

e.g. $f(x, y) = x^2 + y^2$

$$\nabla f(\bar{x}) = 2\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}$$

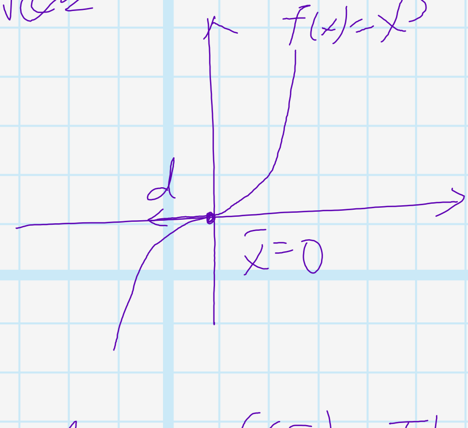


$\nabla f(\bar{x})$是增长方向,只要跟它夹角大于$90°$就会减小.

证明方法:进行泰勒展开.

Note: There could be descent direction d that violates $\nabla f(\bar{x})^T d < 0$ in general

$\therefore$ $\nabla f(\bar{x})^T d < 0$ 是一个充分条件,不是所有的下降的d都满足. 例如



Corollary: If $\bar{x}$ is local min,

Then $\nabla f(\bar{x}) = 0$ (FONC)

Pf: Suppose $\nabla f(\bar{x}) \neq 0$, Let $d = -\nabla f(\bar{x})$ Then,

$\nabla f(\bar{x})^T d = -\|\nabla f(\bar{x})\|_2^2 < 0$ 如果 $\nabla f(\bar{x})$ 不为0

一定能找到这么一个下降的方向。

Corollary: If f is in addition convex, then (不仅可微，而且凸)

$$\nabla f(\bar{x}) = 0 \iff \bar{x} \text{ is global min}$$

Pf: ($\Leftarrow$) From the previous Corollary,

($\Rightarrow$) By convexity.

$$f(x) \geq f(\bar{x}) + \underbrace{\nabla f(\bar{x})^T}_{=0}(x - \bar{x}) = f(\bar{x})$$

$$\forall x \in \mathbb{R}^n$$

$$\therefore f(x) \geq f(\bar{x}) \quad \forall x \in \mathbb{R}^n$$

这个可以推广到非凸集，只需把梯度换成 sub gradient.

Prop: Let $f: \mathbb{R}^n \to \mathbb{R}$ be arbitrary. Then

$\bar{x}$ is global min

iff $0 \in \partial f(\bar{x})$

subgradient 定义

Pf: $\partial f(\bar{x}) = \{ s : f(x) \geq f(\bar{x}) + s^T(x - \bar{x}) \ \forall x \}$

$\bar{x}$ is global min iff $f(x) \geq f(\bar{x}) \ \forall x \iff$

$f(x) \geq f(\bar{x}) + 0^T(x - \bar{x}) \iff 0 \in \partial f(\bar{x}) \ \forall x$

这条不是很有用。因为如果函数不是凸的，这样在每一点都很难去找 subgradient 或者很容易出现 subgradient 不存在的情况。

如果f是凸的，那么 subgradient 一定存在

prop: Suppose f is twice continuously diff, and let $\bar{x}$ be arbitrary. If

$$\nabla f(\bar{x})=0, \quad \nabla^2 f(\bar{x})>0 \qquad \text{(SOSC)}$$

then, $\bar{x}$ is a local min

证明：还是用泰勒展开.

Constrained Problems.

$$\inf f(x)$$

$$\text{s.t.} \quad g_i(x)\le 0 \quad \forall i=1,\dots,m_1$$

$$h_j(x)=0 \quad \forall j=1,\dots,m_2$$

$$x\in X\subseteq \mathbb{R}^n$$

X ← open, non-empty.

为什么不把 $h_j(x)=0$ 写成 $h_j(x)\ge 0$ 且 $h_j(x)\le 0$ 的形式然后用 g 来表示？这时 $h_j(x)$ 的 constrain 是一直 active 的，而 $g_i(x)\le 0$ 约束不一定 active.

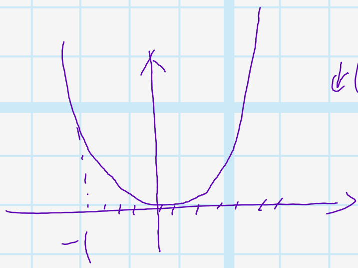
e.g.: $\min x^2$

s.t. $x\ge 1$

此时 $\left.\frac{df}{dx}\right|_{x=1}\ne 0$

$\min x^2$

s.t. $x\ge -1$



此时 $\left.\frac{df}{dx}\right|_{x=0}=0$

Thm (Fritz John Necessary Cond)

(PF)

Let $\bar{x}$ be a local min of (P). Then,

$\exists\, u \in \mathbb{R},\ v \in \mathbb{R}^{m_1},\ w \in \mathbb{R}^{m_2}$ s.t.

(DF)

$$u \nabla f(\bar{x}) + \sum_{i=1}^{m_1} v_i \nabla g_i(\bar{x}) + \sum_{j=1}^{m_2} w_j \nabla h_j(\bar{x}) = 0$$

$u,\ v_i \geq 0,\ v_i g_i(\bar{x}) = 0$ (v_i 和 g_i 的 complentarity)

$(u, v, w) \neq 0$

当无 constrain 时，等价于 $\nabla f(\bar{x}) = 0$ 就是之前的条件.

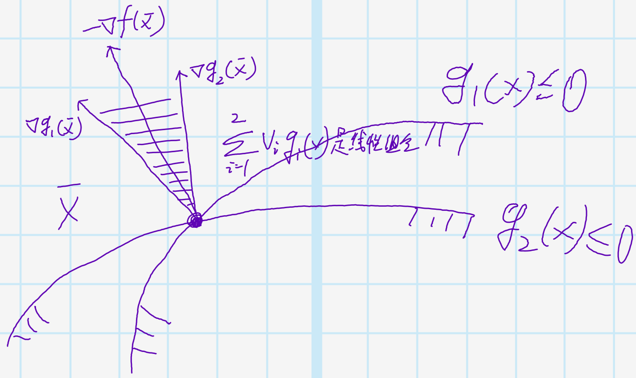
为什么没有 h_i 的 complentarity? ∵ h 恒 $=0$

$v_i = 0$ 时，说明 g_i 没什么用.

$v_i \neq 0$ 时，说明 g_i 是 tight 的，或者说是 active 的

$$v_1 \nabla g_1(\bar{x}) + v_2 \nabla g_2(\bar{x}) = -\nabla f(\bar{x})$$

(assuming $u=1$)



2017.11.15

$$
(P) \quad \begin{aligned} \inf\ & f(x) \\ \text{s.t. } & g_i(x) \le 0 \quad i=1,\cdots,m_1 \\ & h_j(x)=0 \quad j=1,\cdots,m_2 \\ & x \in X \end{aligned}
$$

Thm: (Fritz John NC)

Let $\bar{x}$ be a local minimal of (P),

Then, $\exists\, u, v, w$.

(1) $u \nabla f(\bar{x}) + \sum_i v_i \nabla g_i(\bar{x}) + \sum_j w_j \nabla h_j(\bar{x}) = 0$

(2) $u, v \geq 0$, $(u, v, w) \neq 0$

(3) $v_i g_i(\bar{x}) = 0$ complementarity

Pf: Sketch: 不把 constrains 当 hard constrain，而是当作 penalty

Penalty Approach

For $k=1,2,\cdots$

$g^+(x)=\max(0, g(x))$

$$\min F^k(x) = f(x) + \frac{k}{2}\sum_{i=1}^{m}(g_i^+(x))^2 + \frac{k}{2}\sum_{j=1}^{m_2}(h_j(x))^2 + \frac{1}{2}\|x-\bar{x}\|_2^2$$

proximal term

平方为了把函数弄成在整个定义域可微

当k↑时 惩罚越来越大

让x离$\bar{x}$不太远.

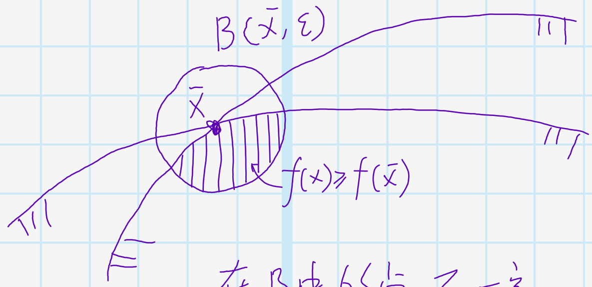
s.t. $x \in B(\bar{x}, \varepsilon) \subseteq X$, $f(\bar{x}) \le f(x)\ \forall x \in S \cap B(\bar{x}, \varepsilon)$

$S = \{x : g_i(x) \le 0\ \forall i,\ h_j(x) = 0,\ \forall j$

Fact: $F^k(\bar{x}) = f(\bar{x})$

Step1: let x^k be an opt soln to (p^k)

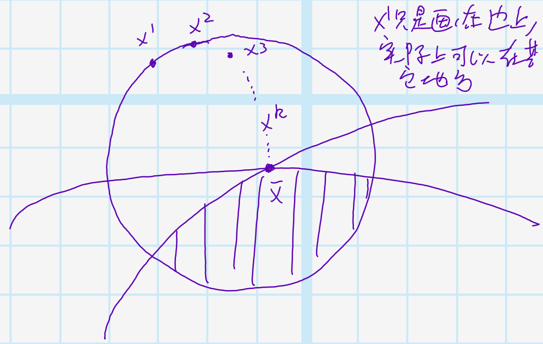
($\because$ B compact 由 Weistrass, 一定存在 x^k 取到 极值)



在B中的点不一定都满足约束. F^k 在B中取点.

$\Downarrow$

放大



Claim $x^k \to \bar{x}$

我们有一个序列 $\{x^k\}$, 我们证明其存在子序列 $\{x^{k_i}\}$ 收敛到 $\bar{x}$

进一步所有子序列收敛到 $\bar{x}$

Step2: For sufficiently large k, say $k \geq k_0$, $x^k \in \operatorname{int}(B(\bar{x}, \varepsilon))$ 是B的内点.

$\Rightarrow x^k$ is an unconstrained local min of (p^k) $\forall k \geq k_0$

那个constrain 不是在 $B(\bar{x}, \varepsilon)$ 内

$\Rightarrow \nabla F^k(x^k) = 0$

Step3:

$$0 = \nabla F^k(x^k) = \nabla f(x^k) + \sum_{i=1}^{m_1} \underbrace{k g_i^+(x^k)}_{\text{scalar}} \nabla g_i(x^k) + \sum_{j=1}^{m_2} \underbrace{k h_j(x^k)}_{\text{scalar}} \nabla h_j(x^k) + (x^k - \bar{x})$$

(Lemma: $q(x)=(\max\{x,0\})^2 \Rightarrow \frac{dq}{dx}=2\max\{x,0\}$)

Define:

$$\delta^k=\left[1+\sum_{i=1}^{m}(k g_i^+(x^k))^2+\sum_{j=1}^{m_2}(k h_j(x^k))^2\right]^{1/2} \geq 1$$

$$u^k=\frac{1}{\delta^k}>0$$

$$v_i^k=\frac{k g_i^+(x^k)}{\delta^k} \geq 0$$

$$w_j^k=\frac{k h_j(x^k)}{\delta^k}$$

$$0=\frac{\nabla F^k(x^k)}{\delta^k}=u_k\nabla f(x^k)+\sum_{i=1}^{m_1} v_i^k \nabla g_i(x^k)+\sum_{j=1}^{m_2} w_j^k \nabla h_j(x^k)+\frac{x^k-\bar{x}}{\delta^k}$$

而这样定义 u, v, w 有以下特点:

$$\|(u, v_1^k, \cdots, v_{m_1}^k, w_1^k, \cdots, w_{m_2}^k)\|_2=1 \quad \forall k \geq k_0$$

$\therefore (u,v,w)\neq 0$

$\Rightarrow \exists$ convergent subsequence

$$u_k \to u,\quad v_i^k \to v_i \quad w_j^k \to w_j$$

$$\therefore u_k \nabla f(x^k) \to u\nabla f(\bar{x})$$

$$v_i^k \nabla g_i(x^k) \to v_i \nabla g_i(\bar{x})$$

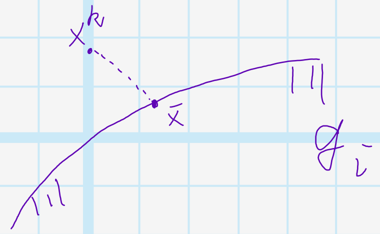
$$w_j^k \nabla h_j(x^k) \to w_j \nabla h_j(\bar{x})$$

$$\frac{x^k-\bar{x}}{\delta^k} \longrightarrow 0$$

Let $I = \{i : V_i > 0\}$ $\quad\because V_i^k$ 收敛到 V_i

$\Rightarrow$ for large enough k: $V_i V_i^k > 0 \quad \forall i \in I$

$V_i^k > 0 \Rightarrow g_i^+(x^k) > 0$



means all x^k are infeasible until the limit $\bar{x}$

$\Rightarrow g_i(\bar{x}) = 0$ $\quad\bar{x}$ 不能是 g_i 的内点.

$\therefore$ 若 $V_i > 0$ 则 $g_i(\bar{x}) = 0$ complomtarity

注意到，上面只保证 $(u, v, w) \neq 0$，但是没保证 $u \neq 0$. 当 $u = 0$ 的时候 我们的目标就没了。下面要想保证 $u \neq 0$.

Thm: (Karush-Kuhn-Tucker KKT)

Let $\bar{x}$ be a local min of (P), let

$I = \{i : g_i(\bar{x}) = 0\}$

(Regularity condition) Suppose $\{\nabla g_i(\bar{x})\}_{i \in I} \cup \{\nabla h_j(\bar{x})\}_{j=1}^{m_2}$ are linearly independent.

Then, (1)(2) in FJNC can be replaced by

(1'): $\nabla f(\bar{x}) + \sum_{i=1}^{m_1} V_i \nabla g_i(\bar{x}) + \sum_{j=1}^{m_2} w_j \nabla h_j(\bar{x}) = 0$

(2'): $V \geq 0$

(3'): stay the same.

注意到那个 linearly independent 的要求是有点强的，如果我们有很多约束，那么很容易不成立。

证明思路：

假设 $u=0$，那么(1)变成了

$$\sum u_i \nabla g_i(\vec{x}) + \sum_j w_j \nabla h_j(\vec{x}) = 0$$

线性相关，与 regular condition 矛盾。

下节课把这个较强的 linearly independent 弱化一下。

2017.11.20

$$\text{(P)}\quad \begin{aligned} \inf\ & f(x) \\ \text{s.t. } & g_i(x) \le 0 \quad i=1,\cdots,m_1 \\ & h_j(x) = 0 \quad j=1,\cdots,m_2 \\ & x \in X \subseteq \mathbb{R}^n \end{aligned}$$

Thm (KKT) Let $\bar{x}$ be a local min of (P)

(Regularity Condition) Suppose $\{\nabla g_i(\bar{x})\}_{i\in I} \cup \{\nabla h_j(\bar{x})\}_{j=1}^{m_2}$

is LI where $I = \{i : g_i(\bar{x}) = 0\}$

Then, $\exists\, v \in \mathbb{R}^{m_1}$, $w \in \mathbb{R}^{m_2}$ (Lagrangian multipliers)

s.t.

(1) $\nabla f(\bar{x}) + \sum_i v_i \nabla g_i(\bar{x}) + \sum_j w_j \nabla h_j(\bar{x}) = 0$

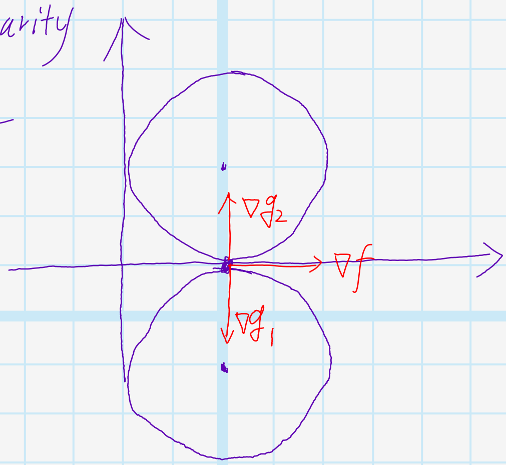
(2) $v \ge 0$

(3) $v_i g_i(\bar{x}) = 0 \quad \forall i$ (complementarity)

和F-J相比，KKT有Regularity Condition，使得在(1)中那个$u=1$不能$=0$，这样目标函数始终在里面

Regularity Condition，KKT是necessary

Example: 以下一例子说明没有Regularity condition，则KKT会不成立



$$\min x_1$$

$$\text{s.t.} \quad (x_1-1)^2+(x_2+1)^2 \le 1$$

$$(x_1-1)^2+(x_2+1)^2 \le 1$$

$x_1=1$ $\therefore \min x_1 = 1$

x_1只能取一个点.

下面来看KKT条件

$$g_1(x)=(x_1-1)^2+(x_2-1)^2-1$$

$$g_2(x)=(x_1-1)^2+(x_2+1)^2-1$$

$$\nabla f(\bar{x})+v_1\nabla g_1(\bar{x})+v_2\nabla g_1(\bar{x})$$

$$\begin{bmatrix}1\\0\end{bmatrix}+2v_1\begin{bmatrix}x_1-1\\x_2-1\end{bmatrix}+2v_2\begin{bmatrix}x_1-1\\x_2+1\end{bmatrix}=0$$

$\begin{bmatrix}x_1\\x_2\end{bmatrix}=\begin{bmatrix}1\\0\end{bmatrix}$时(即位在一交点)代入以上等式:

$$\begin{bmatrix}1\\0\end{bmatrix}+2v_1\begin{bmatrix}0\\-1\end{bmatrix}+2v_2\begin{bmatrix}0\\1\end{bmatrix} \ne 0 \quad \text{KKT条件(1)不成立}$$

原因是:

$$\nabla g_1\Big|_{x=\begin{bmatrix}1\\0\end{bmatrix}}=\begin{bmatrix}0\\-1\end{bmatrix} \qquad \nabla g_2\Big|_{x=\begin{bmatrix}1\\0\end{bmatrix}}=\begin{bmatrix}0\\1\end{bmatrix}$$

线性相关,

违背) Regularity condition

有时候, Regularity condition不成立时, KKT也有可能成立.

把刚才的问题改为

$\min \ x_2$

s.t. $(x_1-1)^2+(x_2-1)^2 \le 1$

$(x_1-1)^2+(x_2+1)^2 \le 1$

验证 KKT (1):

$$\begin{bmatrix}0\\1\end{bmatrix}+2u_1\begin{bmatrix}x_1-1\\x_2-1\end{bmatrix}+2u_2\begin{bmatrix}x_1-1\\x_2+1\end{bmatrix}=0$$

虽然 Regularity condition 不成立，但 KKT (1) 也成立.

由上面例子看出，Regularity condition 并不是很好检查，需要找出 active constrain 然后检查线性独立。以下给出另外两个 Regularity condition

(Regularity Condition) Suppose $\{g_i\}$ are convex, $\{h_j\}$ affine and $\exists x'$ feasible and $g_i(x')<0 \ \forall i$

剩余条件 Slater

(Regularity Condition) Suppose $\{g_i\}$ are concave $\{h_j\}$ affine

∴ 对于线性约束 (既 convex, 又 concave) 自动满足，∴ 不必检查 Regularity Condition

Cor: For linearly constrained problems, KKT conditions are necessary

这些 Regularity Condition 间并没有等价关系，它们都是充分条件.

Example:

$\min\ x^TAx$ 这里A是对称阵

s.t. $\|x\|_2=1$

KKT的意思是说如果一个问题有解，且满足 Regularity Condition，那么KKT成立。

通常的用法是用KKT (1) 等式去解那个 x，然后再去验证 x 是否满足 Regularity.

1) opt s/n exists, b/c min continuous fn on a compact set.

2) 那三个 Regularity Condition 中 (2) (3) 显然不能拿来用，来检查 (1):

$\|x\|_2=1 \iff h(x)=\|x\|_2^2-1=0$

$\nabla h=2x\neq 0$ on $\|x\|_2=1$

只有这一个条件，只要 $\nabla h\neq 0$ 就可以

$\Rightarrow$ LI regularity holds.

$\Rightarrow$ KKT is necessary.

3) 下面来写 KKT (1)

$\nabla f=2Ax$,

$\therefore$ KKT(1) 是:

$2Ax-2wx=0$

$\Rightarrow Ax = wx$

$\Rightarrow (x, w)$ is an eigenvector-eigenvalue pair

∴ 所有的特征向量都是局部最小值.

4) Note: 对原问题进行变形:

$x^T A x = w\|x\|_2^2 = w$

$\|x\|_2^2 = 1$

$\Rightarrow w$ is the smallest eigenvalue of A,

$x^* =$ eigenvector associated with w

Example: 把KKT用到LP上.

$\min \ c^T x$

s.t. $Ax = b$

$x \geq 0$

Assume: opt soln exists.

∵ 约束都是线性的 ∴ 可以用第三个(Regularity Condition)满足来说明KKT是 necessary 的.

$g_i(x) = -x_i \leq 0 \qquad \nabla g_i(x) = -e_i = -\begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ 只有第i个是1

$h_j(x) = b_j - a_j^T x = 0$

(1) $\underset{\nabla f(x)}{c} + \sum_{i=1}^{n} v_i \underbrace{(-e_i)}_{\nabla g_i} + \sum_{j=1}^{m} w_j \underbrace{(-a_j)}_{\nabla h_j} = 0$

(2) $v \geq 0$

(3) $v_i x_i = 0 \quad \forall i$

由(1)得:

$$c - v - A^T w = 0 \quad \because v \geq 0$$

$$\therefore \underbrace{c - A^T w \geq 0}_{\text{dual feasibility}}$$

Example:

$$\inf \ -\ln\det Z$$

s.t. $A \cdot Z \leq b$, $\longrightarrow$ 不等式约束.

$Z \in S_{++}^n \longrightarrow x \in X \subseteq \mathbb{R}^n$

$A \in S_{++}^n$, $b > 0$ 用 Regularity condition (2) 来说明

$$Z' = \frac{b}{2\,tr(A)} I \quad \text{satisfies } A \cdot Z' < b$$

$\therefore$ KKT necessary b/c (2) or (3)

KKT conditions:

$$\nabla(-\ln\det Z) = -Z^{-1}$$

$$g(Z) = A \cdot Z - b \qquad \nabla g(Z) = A$$

(1) $-Z^{-1} + vA = 0$

(2) $v \geq 0$

(3) $v(A \cdot Z - b) = 0$

由(1) $\Rightarrow$ $v > 0$ $\xRightarrow{由(3)}$ $A \cdot Z - b = 0$

(1) $\Rightarrow Z = \frac{1}{v} A^{-1}$ $\Rightarrow \frac{1}{v} \underbrace{A \cdot A^{-1}}_{=n} = b \Rightarrow v = \frac{n}{b}$

2017、11.22

Define Larangian function

$$L(x,v,w)=f(x)+\sum_i v_i g(x)+\sum_j w_j h_j(x)$$

Thm: Suppose $f, g_1, \cdots g_{m_1}$ are convex, $h_1, \cdots, h_{m_2}$ are affine, Suppose $(\bar{x}, \bar{v}, \bar{w})$ satisfies the KKT conditions, Then $\bar{x}$ is a global min。

如果满足f,gconvex，f affine，则充分性也成立

Pf: Note:

$x \mapsto L(x,\bar{v},\bar{w})$ 固定$\bar{v},\bar{w}$,变成一个只与x相关的函数，这时$L(x,\bar{v},\bar{w})$是convex,

Also, $\bar{x}$ is a global min of this function

$\because \nabla_x L(\bar{x},\bar{v},\bar{w})=\nabla f(\bar{x})+\sum_i \bar{v}_i \nabla g_i(\bar{x})+\sum_j \bar{w}_j \nabla h_j(\bar{x})=0$ and $x \mapsto L(x,\bar{v},\bar{w})$ is convex

在无约束优化问题中，导数=0 ⟺ 是最优.

Hence, $f(\bar{x})=f(\bar{x})+\underbrace{\sum_i \bar{v}_i g_i(\bar{x})}_{\text{complementarity}}+\underbrace{\sum_j \bar{w}_j h_j(\bar{x})}_{=0}=L(\bar{x},\bar{v},\bar{w})$

$$=\min_x\left\{f(x)+\sum_i \bar{v}_i g_i(x)+\sum_j \bar{w}_j h_j(x)\right\}$$

$$\le \inf_{\substack{x\\ g_i(x)\le 0\\ h_j(x)=0}}\left\{f(x)+\sum_i \bar{v}_i g_i(x)+\sum_j \bar{w}_j h_j(x)\right\}$$ 定义域变小了，值可能变大.

$$\le \inf_{\substack{x\\ g_i(x)\le 0\\ h_j(x)=0}} f(x)$$ $\therefore \bar{x}$是global min

Example. Power Allocation in Parallel AWGN Channels

n parallel channels.

有n个并行通道，有噪音.

$h_1>0$	σ_1
$h_2>0$	σ_2
$\vdots$	
$h_n>0$	σ_n

h_i: channel power gain

σ_i: noise power.

P_i: transmit power allocated on each channel

Information rate: $r_i=\log_2\left(1+\frac{h_iP_i}{\sigma_i}\right)=\underbrace{(\ln 2)^{-1}}_{\text{扔掉不影响}}\ln\left(1+\frac{h_iP_i}{\sigma_i}\right)$

注意到

$$\max \sum_{i=1}^{n} \ln\left(1+\frac{h_i P_i}{\sigma_i}\right)$$

$$\text{s.t. } \sum_{i=1}^{n} P_i \le P$$

$$P_i \geq 0 \quad i=1,\dots,n$$

$$\to g_i(P) = -P_i \le 0$$

$$\min -\sum_{i=1}^{n} \ln\left(1+\frac{h_i P_i}{\sigma_i}\right)$$

① ∵ 是 Linearly constrained problem

∴ ⇒ KKT is necessary (第三条)

② Convex opt. problem

⇒ sufficient condition applies.

③ opt. soln exists b/c Weierstrass Thm

(紧集上的连续函数取到极值)

ⓐ $$\frac{\partial f}{\partial P_i} = \frac{-h_i/\sigma_i}{1+\frac{h_i P_i}{\sigma_i}} + V_0 \cdot \underset{\frac{\partial(e^T P - P)}{\partial P_i}}{(1)} + V_i \underset{\frac{\partial g_i}{\partial P_i}}{(-1)} = 0 \quad \forall i$$

ⓑ $$V_0\left(\sum_i P_i - P\right) = 0, \quad V_i P_i = 0 \quad \forall i = 1,\dots,n$$

ⓒ $$V_0, V_i \geq 0$$

ⓐ ⇒ $$V_0 - V_i = \frac{h_i}{h_i P_i + \sigma_i} \quad ∵ h_i > 0 \ ∴ V_0 > V_i \overset{ⓒ}{\geq} 0$$

ⓑ ⇒ $$\sum_i P_i = P$$

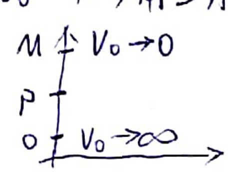
Also from ⓐ $$h_i P_i + \sigma_i = \frac{h_i}{V_0 - V_i} \Rightarrow P_i = \frac{1}{V_0 - V_i} - \frac{\sigma_i}{h_i}$$

1° $$P_i > 0 \Rightarrow V_i = 0 \Rightarrow P_i = \frac{1}{V_0} - \frac{\sigma_i}{h_i}$$ ↘

2° $$P_i = 0 \Rightarrow \frac{1}{V_0} - \frac{\sigma_i}{h_i} \le 0 \Rightarrow P_i = \left(\frac{1}{V_0} - \frac{\sigma_i}{h_i}\right)^+$$

Hence, $$P = \sum_i P_i = \sum_i \left(\frac{1}{V_0} - \frac{\sigma_i}{h_i}\right)^+$$ 现在只有一个变量 V_0 可以用二分查找

Find V_0 by bisection search



$$L(x,v,w) = f(x) + \sum_i v_i g_i(x) + \sum_j w_j h_j(x)$$

(P) $\inf f(x)$

s.t. $g_i(x) \le 0 \quad i=1,\dots,m$

$h_j(x) = 0 \quad j=1,\dots,m_2$

$x \in X \subseteq \mathbb{R}^n$

Claim: (P) is equivalent to

$$\inf_{x\in X} \sup_{\substack{v\ge 0\\ w}} L(x,v,w)$$

Pf: Take $x \in X$:

$$\sup_{\substack{v\ge 0\\ w}} \{ f(\bar{x}) + \sum_i v_i g_i(\bar{x}) + \sum_j w_j h_j(\bar{x}) \} = \begin{cases} f(\bar{x}) & \text{if } \begin{matrix} g_i(\bar{x}) \le 0 \ \forall i \\ h_j(\bar{x}) = 0 \ \forall j \end{matrix} \\ +\infty & \text{otherwise} \end{cases}$$

Take $\bar{x} \in X$, $\bar{v} \ge 0$, $\bar{w}$, Then,

$$\inf_{x\in X} L(x,\bar{v},\bar{w}) \le L(\bar{x},\bar{v},\bar{w}) \le \sup_{\substack{v\ge 0\\ w}} L(\bar{x},v,w)$$

$$\inf_{x\in X} \inf_{x\in X} L(x,\bar{v},\bar{w}) \le \inf_{x\in X} \sup_{\substack{v\ge 0\\ w}} L(\bar{x},v,w)$$

$$\sup_{\substack{v\ge 0\\ w}} \inf_{x\in X} L(x,v,w) \le \inf_{x\in X} \sup_{\substack{v\ge 0\\ w}} L(x,v,w)$$

This motivates the following Lagrange dual of (P)

$$\sup_{\substack{v\ge 0\\ w}} \theta(v,w) \qquad \text{where } \theta(v,w) = \inf_{x\in X} L(x,v,w)$$

2017.11.27

$$\inf f(x)$$

s.t. $g_i(x)\le 0 \quad i=1,\cdots,m_1$ 记作 $G(x)\le 0$, $G(x)=(g_1(x),\cdots,g_{m_1}(x))$

(P) $\quad h_j(x)=0 \quad j=1,\cdots,m_2$ 记作 $H(x)=0$, $H(x)=(h_1(x),\cdots,h_{m_2}(x))$

$x\in X\subseteq\mathbb{R}$

↳ 今天假设X是任意集合，不一定是开集

Lagrangian: $L(x,v,w)=f(x)+v^TG(x)+w^TH(x)$

(P) $\quad V_p^*=\inf\limits_{x\in X}\sup\limits_{\substack{v\ge 0\\ w}} L(x,v,w)$

其对偶问题为:

(D) $\quad V_d^*=\sup\limits_{\substack{v\ge 0\\ w}}\underbrace{\inf\limits_{x\in X} L(x,v,w)}_{\text{记作}\theta(v,w)}$

Fact: (Weak Duality)

If $\bar{x}$ feasible for (P) and $(\bar{v},\bar{w})$ faasible for (D),

Then:

$$\theta(\bar{v},\bar{w})\le f(\bar{x})$$

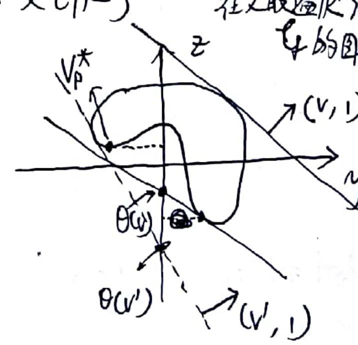
e.g. $\inf f(x)$

s.t. $g(x)\le 0$ 对偶变量v只有一个标量.

Define $\mathcal{G}=\{(y,z)\in\mathbb{R}^2,\ z=f(x),\ y=g(x)\text{ for som }x\in\mathbb{R}^n\}$

↖ 集合

在x取遍$\mathbb{R}^n$后，得到 $\mathcal{G}$ 的图像如下:



下面来看对偶:

Fix $v\ge 0$ 选一个v来看这个式子

$$\theta(v)=\inf_x\{f(x)+vg(x)\}$$

$$=\inf_{(y,z)\in\mathcal{G}}\{z+vy\}$$

Observe:

$$\theta(v,w)=\inf_{x\in X}L(x,v,w)$$

$$=\inf_{x\in X}\{f(x)+v^TG(x)+w^TH(x)\}$$

给定一个x 我们能得到一个affine function

pointwise inf is concave in (v,w)

inf affine function is concave

由此可见，对任意 f,g,h，其对偶都是 concave 的

e.g. $V_p^* = \min -x$

$\quad s.t.\ x \le 1$

$\quad x \in X = \{0,2\}$

$V_p^* = 0$ ~~由~~ $x^* = 0$

下面求 Lagrange

$$L(x,v) = \underbrace{-x}_{f(x)} + v\underbrace{(x-1)}_{g(x)}$$

对偶问题为：

$$V_d^* = \sup_{v \ge 0} \inf_{x \in \{0,2\}} \{-x + v(x-1)\}$$

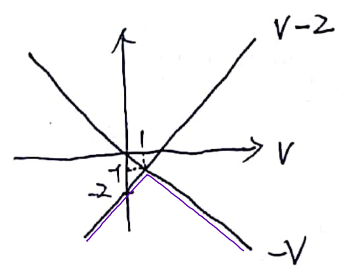
Fix $v \ge 0$

$x=0$: $\theta(v) = -v$

$x=2$: $\theta(v) = -2+v$

$\theta(v) = \min\{-v, -2+v\}$

$\therefore V_d^* = -1,\ v^* = 1$



$V_p^* \ne V_d^*$ 说明有 duality gap

那么 $x \in X = \{0,2\}$ 可以写作 $x(x-2)=0$ 原问题转化为：

$V_p^* = \min -x$

$\quad s.t.\ x \le 1$

$\quad x(x-2) = 0$

$\quad x \in X = \mathbb{R}$

$$L(x,v,w) = -x + v(x-1) + w x(x-2)$$

$$\theta(v,w) = \inf_{x \in \mathbb{R}} \{-x + v(x-1) + w(x-2)x\}$$

这个新的 dual 和刚才那个 dual 不一样的结果，自行验证。

Definition (Saddle Point)

We say $(\bar{x}, \bar{v}, \bar{w})$ is a saddle point of (P) if

(1) $\bar{x} \in X$

(2) $\bar{v} \ge 0$

(3) $\forall x \in X,\ v \ge 0,\ w,\ L(\bar{x}, v, w) \le L(\bar{x}, \bar{v}, \bar{w}) \le L(x, \bar{v}, \bar{w})$

Thm, $(\bar{x},\bar{v},\bar{w})$ is a saddle point of (P) iff 朗格朗日对偶的强对偶在鞍点处取得

$V_p^* = V_d^*$: $\bar{x}$ opt for (P); $(\bar{v},\bar{w})$ opt for (D).

Pf: ($\Rightarrow$) By(3), $L(\bar{x},v,w) \le L(\bar{x},\bar{v},\bar{w})$

By (1) $\bar{x} \in X$, This imply $\bar{x}$ feasible for (P)

By (2) $(\bar{v},\bar{w})$ feasible for (D) ($\because \bar{v} \ge 0$)

By (3) $\theta(\bar{v},\bar{w}) = \inf_{x\in X} L(x,\bar{v},\bar{w})$ θ的定义

$= L(\bar{x},\bar{v},\bar{w})$ (3)右边那个不等式.

$= \sup_{v\ge 0, w} L(\bar{x},v,w)$ (3)左边那个不等式.

$= f(\bar{x})$

($\Leftarrow$) 如果存在最优,即么存在 saddle structure.

(1) (2) 显然成立. It suffices to prove (3).

Consider

$$\theta(\bar{v},\bar{w}) \triangleq \inf_{x\in X} L(x,\bar{v},\bar{w}) \le L(\bar{x},\bar{v},\bar{w})$$

$$\le \sup_{v\ge 0, w} L(\bar{x},v,w) = f(\bar{x})$$

$\because$ 我们假设 $\theta(\bar{v},\bar{w}) = V_d^* = V_p^* = f(\bar{x})$

$\therefore$ 这一串不等式全取到等号

Hence, $\forall x \in X$; $v \ge 0$, w. $L(\bar{x},v,w) \overset{\text{取sup}}{\le} \sup_{v\ge 0, w} L(\bar{x},v,w) = L(\bar{x},\bar{v},\bar{w}) \overset{\text{上面证的}}{=} \inf_{x\in X} L(x,\bar{v},\bar{w})$

$\le L(x,\bar{v},\bar{w})$

Rmk: Suppose $(\bar{x},\bar{v},\bar{w})$ is a saddle point,

$\Rightarrow V_p^* = V_d^* \Rightarrow$

$$\inf_{x\in X} \sup_{v\ge 0, w} L(x,v,w) = \sup_{v\ge 0, w} \inf_{x\in X} L(x,v,w)$$

叫 minimax theorems $\Leftarrow$ 研究 infsup 和 supinf 何时取等号

Saddle point 是其中一种(这里假设了L的形式

minimax theorem 可以有更广义的形式

2017.11.29

$$v_p^* = \inf f(x) \qquad G: \mathbb{R}^n \to \mathbb{R}^{m_1} \quad H: \mathbb{R}^n \to \mathbb{R}^{m_2}$$

(P) s.t. $G(x) \le 0$

$H(x) = 0$

$x \in X$

(D) $v_d^* = \sup_{\substack{v \ge 0 \\ w}} \theta(v, w)$

$$\theta(v, w) = \inf_{x \in X} L(x, v, w),$$

$$L(x, v, w) = f(x) + v^T G(x) + w^T H(x)$$

Saddle point

$(\bar{x}, \bar{v}, \bar{w})$

① $\bar{x} \in X$

② $\bar{v} \ge 0$

③ $x \in X$, $v \ge 0$, w

$$L(\bar{x}, v, w) \le L(\bar{x}, \bar{v}, \bar{w}) \le L(x, \bar{v}, \bar{w})$$

Thm: $(\bar{x}, \bar{v}, \bar{w})$ saddle

$\Leftrightarrow$ $v_p^* = v_d^*$ $\bar{x}$ is opt for (P) $(\bar{v}, \bar{w})$ is opt for (D)

Thm: (Saddle Point optimality)

$(\bar{x}, \bar{v}, \bar{w})$ saddle

$\Leftrightarrow$

(a) (Primal Feasibility) $\bar{x} \in X$, $G(\bar{x}) \le 0$, $H(\bar{x}) = 0$.

(b) (Lagrangian Optimality) $\bar{v} \ge 0$, $\bar{x} = \underset{x \in X}{\operatorname{argmin}} L(x, \bar{v}, \bar{w})$

这个也叫 dual optimality

(c) (Complementarity) $\bar{v}^T G(\bar{x}) = 0$

如果有 saddle 的话, saddle 这个 $(\bar{x}, \bar{v}, \bar{w})$ 满足 (a)(b)(c).

Pf: $(\Rightarrow)$

(a)

(b) $\bar{v} \ge 0$ $\quad \bar{x} = \underset{x \in X}{\operatorname{argmin}} L(x, \bar{v}, \bar{w})$ ← 根据鞍点定义右边的等号

$L(\bar{x}, \bar{v}, \bar{w}) \le L(x, \bar{v}, \bar{w})$

(c) $f(\bar{x}) = L(\bar{x}, 0, 0) \le L(\bar{x}, \bar{v}, \bar{w}) = f(\bar{x}) + \bar{v}^T G(\bar{x}) + \bar{w}^T H(\bar{x})$

↖用鞍点不等式

$$0 \le \bar{v}^T G(\bar{x}) + \underbrace{\bar{w}^T H(\bar{x})}_{=0}$$

$\because \bar{v} \ge 0$, $G(\bar{x}) \le 0$ $\therefore \bar{v}^T G(\bar{x}) = 0$

下面讲 convex optimality 在凸优化中的含意

Suppose $f, g_1, \cdots, g_{m_1}$, convex, differentiable, $h_1, \cdots, h_{m_2}$ affine

$X = \mathbb{R}^n$

(b) $\Rightarrow$ $\bar{x} = \underset{x}{\operatorname{argmin}} L(x, \bar{v}, \bar{w}) \Leftrightarrow \nabla_x L(\bar{x}, \bar{v}, \bar{w}) = 0$

$\Leftrightarrow \nabla f(\bar{x}) + \sum_i \bar{v}_i \nabla g_i(\bar{x}) + \sum_j \bar{w}_j \nabla h_j(\bar{x}) = 0$

Cor: Under the above setting, suppose (P) satisfies Slater and suppose (P) has an opt solution. Then, (D) has optimal solution and $v_p^* = v_d^*$

Pf sketch:

Let $\bar{x}$ be opt to (P)

+ Slater + Convexity + KKT Thm

$\Rightarrow$ ① $\nabla f(\bar{x}) + \sum_i \bar{v}_i g_i(\bar{x}) + \sum_j \bar{w}_j \nabla h_j(\bar{x}) = 0$

② $\bar{v} \geq 0$

③ $\bar{v}_i g_i(\bar{x}) = 0 \quad \forall i$

① ② $\rightarrow$ (b)

③ $\rightarrow$ (c)

$\Rightarrow$ (a) (b) (c) hold $\Rightarrow (\bar{x}, \bar{v}, \bar{w})$ is a saddle

$\Rightarrow (\bar{v}, \bar{w})$ opt for (D)

and $v_p^* = v_d^*$

这里 Slater 即：

$\exists x'$ feas. $g_i(x') < 0 \ \forall i$

e.g.: (QP): $\min \frac{1}{2}x^TQx + c^Tx$ 目标是二次，约束是一次

(P) s.t. $Ax \le b$

$Q \succ 0,\ c \in \mathbb{R}^n,\ A \in \mathbb{R}^{m\times n}$.

其对偶为：

(D) $\sup\limits_{v\ge 0} \theta(v)$

$$\theta(v) = \inf_x \left\{ \frac{1}{2}x^TQx + c^Tx + v^T(Ax-b) \right\}$$

$$0 = \nabla_x \left\{ \frac{1}{2}x^TQx + c^Tx + v^T(Ax-b) \right\}$$

是里面当inf的充要条件.

$$= Qx + c + A^Tv \Rightarrow x^*(v) = -Q^{-1}(c + A^Tv)$$

Substitute into $\theta(v)$:

$$\theta(v) = \frac{1}{2}(c+A^Tv)^T Q^{-1}(c+A^Tv) - c^TQ^{-1}(c+A^Tv) - v^T(AQ^{-1}(c+A^Tv)+b)$$

$$= -\frac{1}{2}v^TAQ^{-1}A^Tv - (AQ^{-1}c+b)^Tv - \frac{1}{2}c^TQ^{-1}c$$

常数

∴对偶问题是：

(D) $\sup\limits_{v\ge 0} \left\{ \underbrace{-\frac{1}{2}v^TAQ^{-1}A^Tv}_{v\text{的二次项}} - \underbrace{(AQ^{-1}c+b)^Tv}_{v\text{的一次项}} - \frac{1}{2}c^TQ^{-1}c \right\}$

因此二次优化问题的对偶也是二次.

e.g. (Fenchel Dual)

这个问题的一个例子: LASSO

e.g.: $\inf\limits_{x}\{f(x)+g(Ax)\}$

$g(y)=\|y-b\|_2^2$

$f(x)=\|x\|_1$

$\Rightarrow \inf\limits_{x}\{\|x\|_1+\|Ax-b\|_2^2\}$

LASSO Problem.

(P) $\inf\{f_1(x)-f_2(x)\}$

s.t. $x\in X_1\cap X_2$

$$\Updownarrow$$

$\inf\{f_1(y)-f_2(z)\}$

s.t. $y=z$

$(y,z)\in X_1\times X_2$ 笛卡尔积

$L(y,z,w)=f_1(y)-f_2(z)+w^T(z-y)$

Dual:

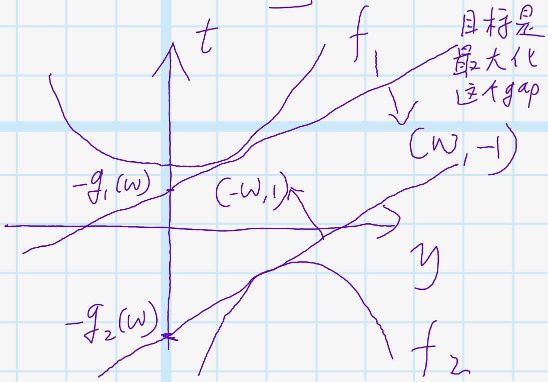
$$\sup_{w}\left\{\inf_{(y,z)\in X_1\times X_2}\{f_1(y)-f_2(z)+w^T(z-y)\}\right\}$$

观察到这个式子里, y, z 无耦合,

∴ 可以分别优化

$$=\sup_{w}\left[\inf_{y\in X_1}\{f_1(y)-w^Ty\}+\inf_{z\in X_2}\{w^Tz-f_2(z)\}\right]$$

$$=\sup_{w}\Big[\underbrace{-\sup_{y\in X_1}\{w^Ty-f_1(y)\}}_{g_1(w)\text{是}f_1(y)\text{的共轭}}+\underbrace{\inf_{z\in X_2}\{w^Tz-f_2(z)\}}_{g_2(w)}\Big]$$



$g_1(w)=\sup\limits_{y\in X_1}\{w^Ty-f_1(y)\}$

$\mathcal{G}=\{(y,t): t=f_1(y) \text{ for some } y\in X_1\}$

$$= \sup_{(y,t)\in \mathcal{E}_g} \{ w^T y - t \}$$

$$\mathcal{H} = \{(y,t) : t = f_2(y) \text{ for some } y \in X_2\}$$

$$g_2(w) = \inf_{y\in X_2} \{ w^T y - f_2(y) \}$$

$$= \inf_{(y,t)\in \mathcal{H}} \{ w^T y - t \}$$

以上给出了如何对两个函数的差求找对偶。

END

本课未包含的内容:

Algorithms 梯度相关算法

目前: 对于 Hessian 矩阵, 有时候不需要计算整个矩阵, 只需要计算部分.

Zero order 方法: 不计算梯度.

Stochastic Optimization. 找一个 x 让其期望最优. 股票.

Non-Convex Problem

2017.12.4

2016 Q1.

y is given

$$\min \|y-x\|_2^2$$
$$\text{s.t. } \|x\|_2^2=1 \quad (w)$$

First-Order Conditions 就是KKT

$$\|x\|_2^2=1$$
$$\underbrace{-2(y-x)}_{\nabla(\|y-x\|_2^2)}+w\underbrace{(2x)}_{\nabla(\|x\|_2^2-1)}=0$$

写完KKT要写是哪个Regularity在起作用.

$$\nabla(\|x\|_2^2-1)=2x\neq 0 \text{ on } \{x:\|x\|_2^2=1\}$$

$\Rightarrow$ linear indep. regularity is satisfied.

注意还要讨论 y 是否为 0.

Case 1: $y=0$: Any x is optimal

Case 2: $y\neq 0$: $w\neq -1$: $y=(1+w)x \Rightarrow \|y\|_2=|1+w|$

$$\pm\|y\|_2=1+w$$

a) $w=\|y\|_2-1 \qquad x=\dfrac{y}{\|y\|_2}$

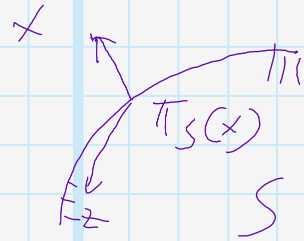
b) $w=-\|y\|_2-1 \qquad x=\dfrac{-y}{\|y\|_2}$

Compare $\left\|y-\dfrac{y}{\|y\|_2}\right\|_2$ vs. $\left\|y+\dfrac{y}{\|y\|_2}\right\|_2$

Conclude: $x=\dfrac{y}{\|y\|_2}$ is optimal

注：实际就是 Projection：可以用 Projection Lemma

$$(x-\Pi_S(x))^T(z-\Pi_S(x))\le 0 \quad \forall z\in S$$



S convex, 这里 $\|x\|_2=1$ 不是 convex

2015 Q4

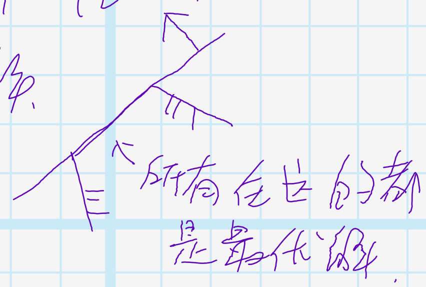
$$\begin{aligned} (P)\quad & \min\ c^Tx \\ & \text{s.t. } Ax=b \\ & \quad\ x\ge 0 \end{aligned}$$

Assume (P) has a unique optimal solution.

Q: Does this imply its dual has unique opt. soln?

通常来讲，(P)问题的最优解不唯一.

这道题限制了(P)有唯一最优解.



$$\begin{aligned} (D)\quad & \max\ b^Ty \\ & \text{s.t. } A^Ty\le c \end{aligned}$$

A^T 的每一行定义了那个 polyhedron 的边，如果 b 也是其中一条边，那么就是上图的情况。

下面构造二维情况.

$$\begin{aligned} & \min\ c_1x_1+c_2x_2 \\ & \text{s.t. } x_1+2x_2=1 \\ (P)\quad & \quad\ x_1+3x_2=1 \\ & \quad\ x\ge 0 \end{aligned}$$

b 一定要是A的一列

$(x_1,x_2)=(1,0)$ 只有一个点..

(D) $\max\ y_1+y_2$

$\text{s.t.}\quad y_1+y_2 \le C_1$

$2y_1+3y_2 \le C_2$

通过观察发现 $C_1=3$，$C_2=8$

$(1,2)$ 是一个解，$(2,1)$ 也是一个解.

2016 Q2.

$$v_m^* = \inf\ y_m$$

$$\text{s.t.}\quad \begin{bmatrix} 1 & 2 \\ 2 & y_1 \end{bmatrix} \succeq 0$$

$$\begin{bmatrix} 1 & y_{k-1} \\ y_{k-1} & y_k \end{bmatrix} \succeq 0 \quad k=2,\cdots,m.$$

Schur 补：
$$\begin{array}{c|cc} & m & n \\ \hline m & A & B \\ n & B^T & C \end{array} \succeq 0 \iff A, C \succeq 0,\ C - B^T A^{-1} B \succeq 0$$

Note: Apply Schur complement,

$y_k \ge 0 \quad y_1 \ge 2^2 \qquad y_k \ge y_{k-1}^2 \quad \Rightarrow$

$$y_m \ge 2^{2^m}$$

$$y_1 = 2^2 \quad y_2 = 2^{2^2} \quad y_k = 2^{2^k}$$

这个问题输入为 m，输出 2^m 位，∴是 NP 的问题。

表示输出就需要 2^m 位

2015 Q3

$$\max x_1 \cdots x_n$$
$$\text{s.t. } \sum_{i=1}^{n} \frac{x_i}{a_i} = 1 \quad (w)$$

需要加上 $x \geq 0$
如果不加的话,
就是那个hyperplane
上取负就是 $+\infty$

FO Conditions:

(A) $$\sum_{i=1}^{n} \frac{x_i}{a_i} = 1 \quad \text{(Primal Feasibility)}$$

ith component of KKT: KKT求导后第i个元素

(B) $$\prod_{j \neq i} x_j - \frac{w}{a_i} = 0$$

Since constraint is linear, the FO conditions are necessary

Multiply (B) by x_i and sum:

$$\prod_{j=1}^{n} x_j = \frac{w x_i}{a_i}$$

$$\sum_{i=1}^{n} \prod_{j=1}^{n} x_j = \sum_{i=1}^{n} \frac{w x_i}{a_i} = w$$

$$\parallel$$

$$n \prod_{j=1}^{n} x_j \implies \text{substitute into (B)}$$

$$\prod_{j \neq i} x_j = \frac{n}{a_i} \prod_{j=1}^{n} x_j$$

Obsere maximizing $\Rightarrow x_i \neq 0 \ \forall i$

$\because$ 上式 $\Rightarrow x_i = \frac{a_i}{n}$

2016 Q4.

$$(P) \quad \begin{aligned} &\inf f(x) \\ &\text{s.t. } h_i(x) = 0 \quad i = 1, \cdots, m \\ &x \in \mathbb{R}^n \end{aligned}$$

Lagrangian:

$$L(x, w) = f(x) + \sum_i w_i h_i(x)$$

Augmented Lagrangian: further penalize deviation

$$L_c(x, w) = f(x) + \sum_i w_i h_i(x) + \underbrace{\frac{c}{2} \sum_i h_i^2(x)}$$

Suppose $(\bar{x}, \bar{w})$ satisfies

(I) $\nabla_x L(\bar{x}, \bar{w}) = 0$

(II) $h_i(\bar{x}) = 0 \quad \forall i$

(III) $y^T \nabla^2_{xx} L(\bar{x}, \bar{w}) y > 0$

$\forall y \neq 0 : \nabla h_i^T(\bar{x}) y = 0 \ \forall i$

目前我们只学过 KKT first order necessary condition

(III) 的理解: $\nabla^2_{xx} L(\bar{x}, \bar{w})$ 只需要是 PD 在 $\nabla h_i^T(\bar{x})$ 的切线方向。只需对这些 y 正定即可

(a) compute $\nabla_x L_c(\bar{x},\bar{w})$ $\nabla^2_{xx} L_c(\bar{x},\bar{w})$

$$\nabla_x L_c(\bar{x},\bar{w}) = \nabla f(\bar{x}) + \sum_i \bar{w}_i \nabla h_i(\bar{x}) + c\sum h_i(\bar{x})\nabla h_i(\bar{x})$$

$$= \nabla f(\bar{x}) + \sum_i (\bar{w}_i + c h_i(\bar{x}))\nabla h_i(\bar{x})$$

$h_i(\bar{x})=0$ 代入了

$$\nabla_{xx} L_c(\bar{x},\bar{w}) = \nabla^2 f(\bar{x}) + \sum_i (\bar{w}_i + c h_i(\bar{x}))\nabla^2 h_i(\bar{x})$$

$$+ \sum_i c \nabla h_i(\bar{x})\nabla h_i(\bar{x})^T$$

求证：

Claim: $\exists \bar{c} \geq 0$: $\nabla^2_{xx} L_c(\bar{x},\bar{w}) > 0$ $\forall c > \bar{c}$

只要把 c 往上调到一定程度 $\nabla^2_{xx} L_c$ 就是 PD

$\Rightarrow \bar{x}$ is a local min of $L_c(\cdot,\bar{w})$ $\forall c > \bar{c}$

Hint Fact: Let $P \in S^n$, $Q \in S^n_+$, s.t. $x^T P x > 0$ $\forall x \neq 0$ satisfying $Qx=0$. Then $\exists \bar{c} \geq 0$ s.t. $x^T(P+cQ)x > 0$ $\forall x \neq 0$, $c > \bar{c}$

只要x在 Q 的 null space 中，可以把 Q 用常数拉大，使上式 >0

(III) $\Rightarrow$

$$\underbrace{y^T \nabla^2_{xx} L(\bar{x},\bar{w}) y}_{P} > 0 \quad \forall y \neq 0 \quad \underbrace{\sum_i \nabla h_i(\bar{x})\nabla h_i^T(\bar{x})}_{Q} y = 0$$

Claim: $\bar{x}$ is local min of (P)

Pf: $\exists\, \varepsilon > 0$: $L_c(\bar{x}, \bar{w}) \leq L_c(x, \bar{w})$ and

$\nabla^2_{xx} L_c(x, \bar{w}) \succ 0$ for $x \in B(\bar{x}, \varepsilon)$

Hessian 在鄰域 B 內 仍是 PD

Define: $g_d(\alpha) = L_c(\bar{x} + \alpha d, \bar{w})$, $\|d\|_2 = 1$

$$g_d(\alpha) = g_d(0) + \alpha g_d'(0) + \frac{\alpha^2}{2} g_d''(\alpha_0) \text{ for all } \alpha \in (0, \alpha_0)$$

$$= \underset{\substack{\| \\ f(\bar{x})}}{L_c(\bar{x}, \bar{w})} + \alpha \underbrace{\nabla_x L_c(\bar{x}, \bar{w})^T}_{\substack{\| \\ 0}} d$$

$\because \dfrac{\partial f(x_0 + \alpha d)}{\partial \alpha} = \nabla f^T(x_0) d$

$$+ \frac{\alpha^2}{2} d^T \nabla^2_{xx} L_c(\bar{x} + \alpha_0 d, \bar{w}) d$$

$$> f(\bar{x})$$

$$L_c(\bar{x} + \alpha d, \bar{w}) > f(\bar{x}) \text{ for } \alpha \in (0, \varepsilon)$$

Hence if $x = \bar{x} + \alpha d$ and $h_i(x) = 0$ $\forall i$, $\alpha \in (0, \varepsilon)$

then

$$f(x) = L_c(x, \bar{w}) > f(\bar{x})$$