Optimization Theory and Methods

Prof. Xiaoe Ruan

Email: wruanxe@xjtu.edu.cn

Tel:13279321898

2019, Fall

References：

[1] WenYu Sun, Ya-Xiang Yuan, Optimization Theory and Methods: Nonlinear Programming, Springer, 2006.

[2] Honglei, Xu, Song Wang, SoonYi Wu, Optimization Methods, Theory and Applications, Springer, 2015.

[3] 徐成贤， 陈志平，李乃成，近代优化方法，科学出版社，2002.

[4] 张可村，李换琴， 工程优化方法及其应用， 西安交通大学出版社，2007.

Chapter1 Optimization Foundations

Optimization:(dictionary)

The process to make sth. as good as it can be

or to use sth. in the possibly best way.

Optimization:(Terminology)

The process of selecting the best of many possible decisions in real-life environment, constructing computational methods to find optimal solutions, exploring the theoretical properties, and studying the computational performance of numerical algorithms implemented based on computational methods.

Outlines:

1. Models and Categories

2. Multi-variable Functional Analysis

Gradient, Hessian Matrix, Jacobi Matrix,

Taylor expansion;

Convex Sets and Convex Functions;

Separation of Convex Sets ;

Examples

(Product schedule)

Product A and B, costs and profits are as follows:

	A	B	resources
coal	1	2	30
labor	3	2	60
storehouse	0	2	24
profit	40	50	

Question: How to schedule the outputs of A and B so that the total profit is maximal?

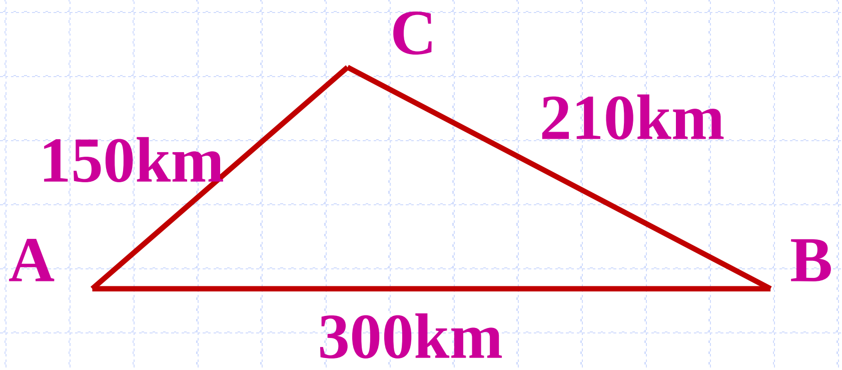
Analysis: Suppose that the outputs of A and B are x_1 and x_2, respectively, then

$$\max Z = 40x_1 + 50x_2 \quad \text{s.t.} \quad \begin{cases} x_1 + 2x_2 \le 30, \\ 3x_1 + 2x_2 \le 60, \\ 2x_2 \le 24, \\ x_1,\ x_2 \ge 0; \end{cases}$$

	A	B	resources
coal	1	2	30
labor	3	2	60
storehouse	0	2	24
profit	40	50	

Transportation schedule

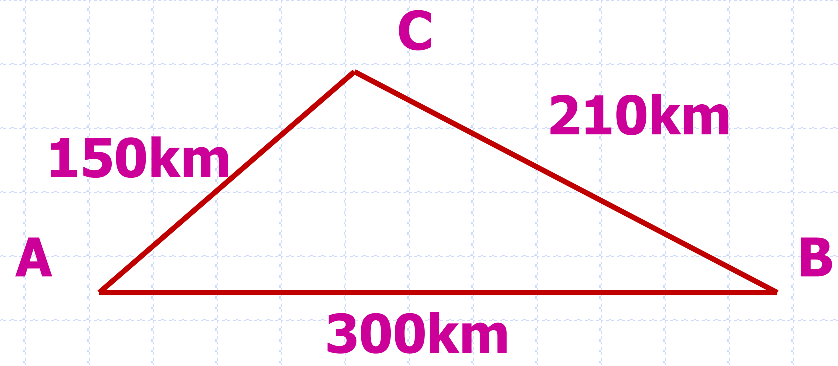
Constructing a new railway from A to B needs to ship steel tube from C to A or B by rail and then to the construction place by truck along temporary construction road. For simplicity, 1km-road steel tube is denoted as 1 unity. Assume that the transportation costs by rail is 600 Yuan/km and by road is 1000/km. The distances among A, B and C are as following.



Question： How to schedule the transportation task so that the total cost is minimal?

Analysis:

Suppose the numbers of steel tube to A and B are x_1 and x_2.



Then $x_1 + x_2 = 300$

The cost of x_1 to A is: $150*600*x_1$;

The cost of x_2 to B is: $210*600*x_2$;

The distance of x_1 steel tube from A to construction place is

$$0+1+2+\ldots+(x_1-1)=(x_1-1)\,x_1/2;$$

The distance of x_2 steel tube from B to construction place is:

$$(x_2-1)\,x_2/2;$$

Then, total cost is

$$f(x_1, x_2)= 90000x_1 +126000x_2 +500(x_1-1)\,x_1 + 500(x_2-1)\,x_2$$

1. Mathematical Description and Category

$$\begin{cases} \min\ f(x), \\ \text{s. t.} \quad c_i(x)=0, i\in E=\{1,2,\cdots m'\}, \\ \qquad\quad c_i(x)\ge 0, i\in I=\{m'+1,\cdots m\}. \end{cases}$$

where

$f(x)$ **-objective function;**

$x=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}\in R^n$ **-decision variable;**

$c_i(x)=0,(i\in E)$ **-equality constraints;**

$c_i(x)\ge 0(i\in I)$ **-inequality constraints;**

Denote $$D=\left\{x \middle| c_i(x)=0, i\in E, c_i(x)\geq 0, i\in I\right\}$$

as constraint set, constraint domain or feasible domain.

$$\begin{cases} \min \; f(x), \\ \text{s. t.} \quad c_i(x)=0, i\in E=\{1,2,\cdots m'\}, \\ \qquad\quad c_i(x)\geq 0, i\in I=\{m'+1,\cdots m\}. \end{cases} \Longleftrightarrow \min_{x\in D} \; f(x)$$

Categories:

(Un)constrained optimization

(In)equality constrained optimization

Hybrid constrained optimization

(Non)linear programming

Quadratic programming

2. Multi-variable Functional Analysis

2.1 Gradient

（1st-order derivative）：

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots, \frac{\partial f(x)}{\partial x_n} \right)^{\mathrm{T}}$$

2nd-order partial derivative :

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

2nd-order derivative matrix（Hessian matrix）：

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2}, & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2}, & \cdots, & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1}, & \frac{\partial^2 f(x)}{\partial x_2^2}, & \cdots, & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1}, & \frac{\partial^2 f(x)}{\partial x_n \partial x_2}, & \cdots, & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix} = \left[\nabla \frac{\partial f(x)}{\partial x_1}, \nabla \frac{\partial f(x)}{\partial x_2} \cdots \nabla \frac{\partial f(x)}{\partial x_n} \right]^{\mathrm{T}}$$

Exercise1: **Given matrices**

$$A \in R^{n\times n}, \qquad A^{\mathrm{T}} = A, \qquad B \in R^{1\times n}.$$

Solve gradient and Hessian matrix of the quadratic function

$$f(x) = \frac{1}{2} x^{\mathrm{T}} A x - Bx$$

Solution:

$$\nabla f(x) = Ax - B^{\mathrm{T}}$$

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

Jacobian matrix of vector-valued function

Suppose that $F(x) = \left(f_1(x), f_2(x), \cdots, f_m(x)\right)^{\mathrm{T}} \in R^m$

is differentiable.

Jacobian matrix

$$F'(x) = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1}, & \frac{\partial f_1(x)}{\partial x_2}, & \cdots, & \frac{\partial f_1(x)}{\partial x_n} \\ \frac{\partial f_2(x)}{\partial x_1}, & \frac{\partial f_2(x)}{\partial x_2}, & \cdots, & \frac{\partial f_2(x)}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1}, & \frac{\partial f_m(x)}{\partial x_2}, & \cdots, & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix}$$

2.2 Taylor Expansion

Lemma1 **Let** $\varphi(\alpha)=f(x+\alpha d),\quad \alpha\in R,\ x,d\in R^n.$

Denote $u=x+\alpha d=(x_1+\alpha d_1,\cdots,x_n+\alpha d_n)^{\mathrm{T}}=(u_1,\cdots,u_n)^{\mathrm{T}}.$

Suppose that $f(u)$

are 1st and 2nd-order continuously differentiable.

Then $$\varphi'(\alpha)=\sum_{i=1}^{n}\frac{\partial f(u)}{\partial u_i}\frac{du_i}{d\alpha}=\sum_{i=1}^{n}\frac{\partial f(u)}{\partial u_i}d_i=\nabla f(u)^{\mathrm{T}}d$$

$$\varphi''(\alpha)=\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{\partial}{\partial u_j}\left(\frac{\partial f}{\partial u_i}d_i\right)\frac{\mathrm{d}u_j}{\mathrm{d}\alpha}=\sum_{j=1}^{n}\sum_{i=1}^{n}\frac{\partial^2 f}{\partial u_i\partial u_j}d_id_j$$

$$=d^{\mathrm{T}}\nabla^2 f(x+\alpha d)d$$

Theorem1. **(1)Suppose that** $f(x), \quad x \in N(x^*) = \{x \,|\, \|x - x^*\| < \delta\}$

is 1st-order continuously differentiable

Then $\boxed{f(x) = f(x^*) + \nabla f(\xi)^{\mathrm{T}}(x - x^*)}, x \in N(x^*),$

Mean theorem

where $\xi = x^* + \theta(x - x^*), \; 0 < \theta < 1.$

Linear approximation

or $f(x) = \boxed{f(x^*) + \nabla f(x^*)^{\mathrm{T}}(x - x^*)} + o(\|x - x^*\|), x \in N(x^*).$

(2) Suppose $f(x), x \in N(x^*)$ **is 2nd-order continuously differentiable.**

Then $f(x) = f(x^*) + \nabla f(x^*)^{\mathrm{T}}(x - x^*) + \frac{1}{2}(x - x^*)^{\mathrm{T}} \nabla^2 f(\xi)(x - x^*)$

where $x \in N(x^*), \; \xi = x^* + \theta(x - x^*), \; 0 < \theta < 1.$

Quadratic approximation

or $f(x) = \boxed{f(x^*) + \nabla f(x^*)^{\mathrm{T}}(x - x^*) + \frac{1}{2}(x - x^*)^{\mathrm{T}} \nabla^2 f(x^*)(x - x^*)}$

$+ o(\|x - x^*\|^2), \qquad x \in N(x^*),$

Integral form of Taylor expansion

Th.2. **(1) Suppose that the function** $f(x),\quad x \in N(x^*)$ **is continuously differentiable. Then**

$$f(x) = f(x^*) + \int_0^1 \nabla f\left(x^* + \alpha(x - x^*)\right)^{\mathrm{T}} (x - x^*) \mathrm{d}\alpha, x \in N(x^*).$$

named as integral form mean theorem.

Proof: Let $d = x - x^*,\ \varphi(\alpha) = f\left(x^* + \alpha(x - x^*)\right) = f(x^* + \alpha d),$

Then $\varphi(0) = f(x^*),\quad \varphi(1) = f(x^* + d) = f(x).$

Denote $u = [u_1, u_2, \cdots, u_n]^{\mathrm{T}} = x^* + \alpha d = \left[x_1^* + \alpha d_1, \cdots, x_n^* + \alpha d_n\right]^{\mathrm{T}}.$

Then

$$\varphi'(\alpha) = \frac{\partial f(u)}{\partial u_1}\frac{du_1}{d\alpha} + \frac{\partial f(u)}{\partial u_2}\frac{du_2}{d\alpha} + \cdots + \frac{\partial f(u)}{\partial u_n}\frac{du_n}{d\alpha} = \nabla f(x^* + \alpha d)^{\mathrm{T}} d$$

From Newton-Leibnitz formula $\varphi(1) - \varphi(0) = \int_0^1 \varphi'(\alpha) \mathrm{d}\alpha$ **yields**

$$f(x) = f(x^*) + \int_0^1 \nabla f\left(x^* + \alpha(x - x^*)\right)^{\mathrm{T}} (x - x^*) \mathrm{d}\alpha, x \in N(x^*).$$

(2) Suppose that $f(x),\ x \in N(x^*)$ **is 2nd-order continuously differentiable. Then for any vectors** $x, d \in R^n$ **and a number** $\alpha \in R$, **we have**

$$f(x+\alpha d) = f(x) + \alpha \nabla f(x)^{\mathrm{T}} d + \alpha^2 \int_0^1 (1-t)\left[d^{\mathrm{T}} \nabla^2 f(x+t\alpha d) d\right] \mathrm{d}t.$$

Proof: **Let** $\varphi(t) = f(x+t\alpha d)$. **Then** $\varphi(0) = f(x),\ \varphi(1) = f(x+\alpha d)$,

$$\varphi'(t) = \alpha \nabla f(x+t\alpha d)^{\mathrm{T}} d, \quad \varphi'(0) = \alpha \nabla f(x)^{\mathrm{T}} d,$$

$$\varphi''(t) = \alpha^2 d^{\mathrm{T}} \nabla^2 f(x+t\alpha d) d.$$

Therefore $$\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)\mathrm{d}t = -\int_0^1 \varphi'(t)\mathrm{d}(1-t)$$

$$= -\left[\varphi'(t)(1-t)\Big|_0^1 - \int_0^1 (1-t)\varphi''(t)\mathrm{d}t\right]$$

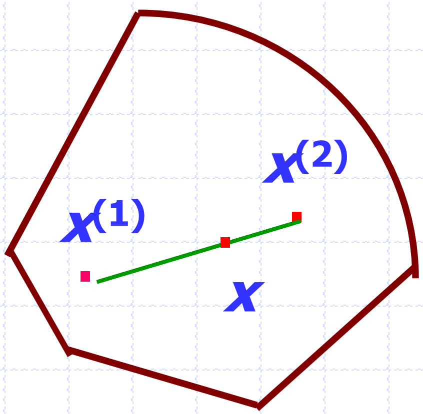
$$= \varphi'(0) + \int_0^1 (1-t)\varphi''(t)\mathrm{d}t.$$

Thus

$$f(x+\alpha d) = f(x) + \alpha \nabla f(x)^{\mathrm{T}} d + \alpha^2 \int_0^1 (1-t)\left[d^{\mathrm{T}} \nabla^2 f(x+t\alpha d) d\right] \mathrm{d}t$$

2.3 Convex Sets and Convex Functions

Convex Set:



Let set D belong to n-dimensional Euclidean space. D is convex if for any points $x^{(1)}$ and $x^{(2)} \in D$ so that the point $x = \alpha x^{(1)} + (1-\alpha) x^{(2)}$ $(0 \le \alpha \le 1)$ belongs to D.

Properties:

Suppose that the sets $D_1, D_2 \subset R^n$ are convex and the number $a \in R$.

Then (1) $D_1 \cap D_2 = \{x \mid x \in D_1, x \in D_2\}$ **is convex.**

(2) $aD_1 = \{ax \mid x \in D_1\}$ **is convex.**

(3) $D_1 + D_2 = \{x + y \mid x \in D_1, y \in D_2\}$ **is convex.**

(4) $D_1 - D_2 = \{x - y \mid x \in D_1, y \in D_2\}$ **is convex.**

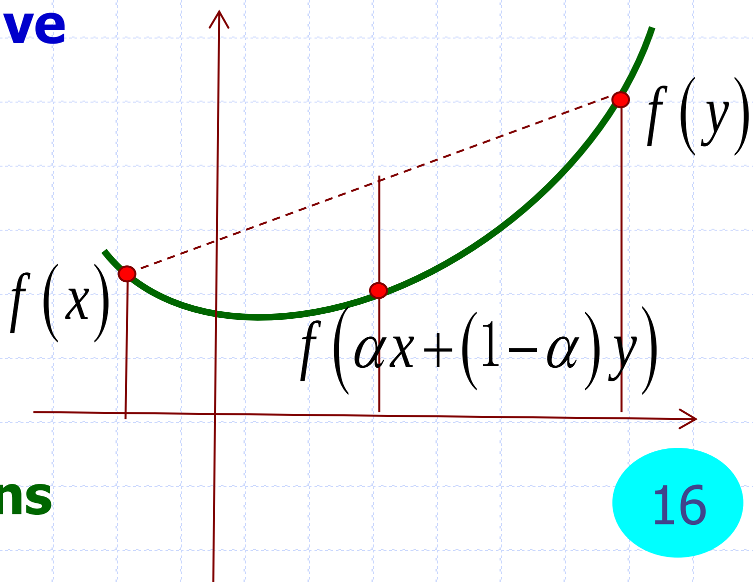
Th.3. Suppose $D \subset R^n$ is convex. Then for any points $x^{(i)} \in D$ and numbers α_i satisfying $\alpha_i \geq 0\left(i=1,\cdots,m\right), \sum_{i=1}^{m} \alpha_i = 1,$ we have $\sum_{i=1}^{m} \alpha_i x^{(i)} \in D$

Proof: By mathematical induction（omitted).

Convex function:

Function $f\left(x\right) \in R^1, x \in D$ is said to be convex on the nonempty convex set $D \subset R^n$ if for any points $x, y \in D$ and number $\alpha \in (0,1)$ we have

$$f\left(\alpha x + \left(1-\alpha\right) y\right) \leq \alpha f\left(x\right) + \left(1-\alpha\right) f\left(y\right)$$



Strictly convex function;

(strictly) concave function;

Consistently convex/concave functions

Th.4. **Suppose that the set $D \subset R^n$ is nonempty convex, the functions $f_i(x), x \in D (i = 1, \cdots, m)$ are convex and the numbers $\alpha_i \geq 0 (i = 1, \cdots, m)$.**

Then **(1) Function $\sum_{i=1}^{m} \alpha_i f_i(x), x \in D$ is convex.**

(2) Function $f(x) = \max_{1 \leq i \leq m} f_i(x), x \in D$ is convex.

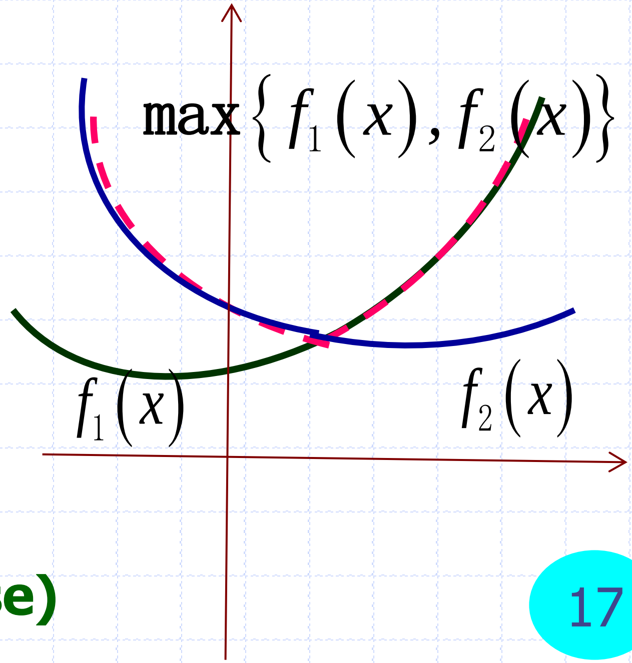
Th.5. **Suppose that the set $D \subset R^n$ is nonempty convex and the function $f(x) \in R^1, x \in D$ is convex.**

Then, for any points $x^{(i)} \in D (i = 1, \cdots, m)$

and any numbers satisfying $\alpha_i \geq 0$

and $\sum_{i=1}^{m} \alpha_i = 1,$ we have

$$f\left(\sum_{i=1}^{m} \alpha_i x^{(i)}\right) \leq \sum_{i=1}^{m} \alpha_i f\left(x^{(i)}\right)$$



Proof: By mathematical induction. (Exercise)

Judgment Theorems for Convex Functions

Th.6. **Function** $f(x)$ **is convex** $\Longleftrightarrow$ **for any points** $x, y \in R^n$ **single-variable function** $\varphi(\alpha) = f(x + \alpha y)$ **is convex wrp variable** α.

Proof: $\Longrightarrow$ **Let** $\lambda_1 \geq 0, \lambda_2 \geq 0$ **and** $\lambda_1 + \lambda_2 = 1.$ **Then**

$$\varphi(\lambda_1\alpha_1 + \lambda_2\alpha_2) = f((\lambda_1 + \lambda_2)x + (\lambda_1\alpha_1 + \lambda_2\alpha_2)y) = f(\lambda_1(x + \alpha_1 y) + \lambda_2(x + \alpha_2 y))$$
$$\leq \lambda_1 f(x + \alpha_1 y) + \lambda_2 f(x + \alpha_2 y) = \lambda_1\varphi(\alpha_1) + \lambda_2\varphi(\alpha_2).$$

$\Longleftarrow$ **Let** $\forall x, y \in R^n$, $\forall \alpha_1, \alpha_2 \in R^1$, $\lambda_1 \geq 0, \lambda_2 \geq 0$ **and** $\lambda_1 + \lambda_2 = 1$ **Then**

$$f(\lambda_1(x + \alpha_1 y) + \lambda_2(x + \alpha_2 y)) = f((\lambda_1 + \lambda_2)x + (\lambda_1\alpha_1 + \lambda_2\alpha_2)y)$$
$$= \varphi(\lambda_1\alpha_1 + \lambda_2\alpha_2) \leq \lambda_1\varphi(\alpha_1) + \lambda_2\varphi(\alpha_2) = \lambda_1 f(x + \alpha_1 y) + \lambda_2 f(x + \alpha_2 y).$$

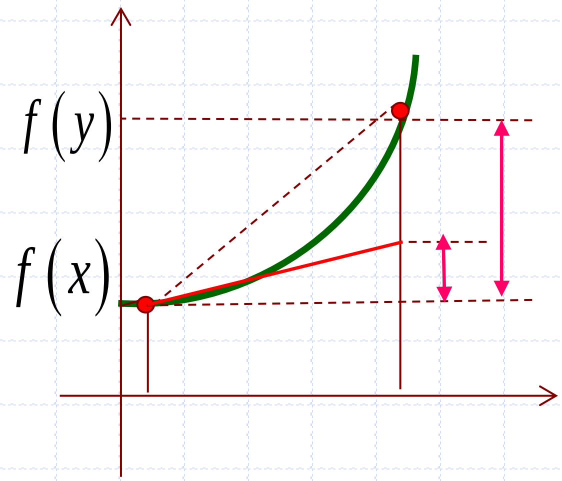
Th.7. **Suppose that set D is nonempty open convex and $f(x): D \to R$ is continuously differentiable. Then**

(1) Function $f(x)$ is convex $\Longleftrightarrow$

$$f(y)-f(x) \geq \nabla f(x)^{\mathrm{T}}(y-x), \forall x, y \in D.$$

(2) Function $f(x)$ is strictly convex $\Longleftrightarrow$

$$f(y)-f(x) > \nabla f(x)^{\mathrm{T}}(y-x), \forall x, y \in D, x \neq y.$$



(1) **Proof:** $\Longrightarrow$ **For** $\forall \alpha \in (0,1)$, **we have**

$$f(x+\alpha(y-x)) = f(\alpha y+(1-\alpha)x) \leq \alpha f(y)+(1-\alpha)f(x)$$ **that is**

$$\nabla f(x)^{\mathrm{T}} \alpha(y-x)+o(\alpha\|y-x\|) = f(x+\alpha(y-x))-f(x) \leq \alpha(f(y)-f(x))$$

Therefore $$f(y)-f(x) \geq \nabla f(x)^{\mathrm{T}}(y-x), \forall x, y \in D.$$

$\Longleftarrow$ **Let** $x^{*}=\alpha x+(1-\alpha)y.$ **Then**

$$f(x)-f(x^{*}) \geq \nabla f(x^{*})^{\mathrm{T}}(x-x^{*})$$ **and** $$f(y)-f(x^{*}) \geq \nabla f(x^{*})^{\mathrm{T}}(y-x^{*})$$

Thus $$\alpha f(x)+(1-\alpha)f(y) \geq f(x^{*}) = f(\alpha x+(1-\alpha)y).$$

Th.8. **Suppose the set D is nonempty open convex and $f(x):D\to R$ is 2nd-order continuously differentiable. Then,**

Function $f(x)$ **is convex** $\Longleftrightarrow$ $\nabla^2 f(x)=\left[\dfrac{\partial^2 f(x)}{\partial x_i \partial x_j}\right]_{n\times n}\geq 0, x\in D.$

Proof: $\Longrightarrow$ $\forall x,y\in D, \exists \delta, s.t. \alpha\in(0,\delta), x+\alpha y\in D.$ **Then**

$$\nabla f(x)^{\mathrm{T}}\alpha y\leq f(x+\alpha y)-f(x) =\nabla f(x)^{\mathrm{T}}\alpha y+\frac{1}{2}\alpha^2 y^{\mathrm{T}}\nabla^2 f(x)y+o\left(\|\alpha y\|^2\right)$$

Therefore $$y^{\mathrm{T}}\nabla^2 f(x)y=\lim_{\alpha\to 0}\frac{\alpha^2 y^{\mathrm{T}}\nabla^2 f(x)y+2o\left(\|\alpha y\|^2\right)}{\alpha^2}\geq 0.$$

$\Longleftarrow$ $$f(y)-f(x)=\nabla f(x)^{\mathrm{T}}(y-x)+\frac{1}{2}(y-x)^{\mathrm{T}}\nabla^2 f\big(x+\theta(y-x)\big)(y-x)$$

Therefore $$f(y)-f(x)\geq \nabla f(x)^{\mathrm{T}}(y-x).$$

Corallary1 **If** $\nabla^2 f(x)=\left[\dfrac{\partial^2 f(x)}{\partial x_i \partial x_j}\right]_{n\times n}>0, 0\neq x\in D,$ **then the function** $f(x)$ **is strictly convex.**

Judgment Theorem for Consistently Convex Function

Def1.(Consistently convex function) Suppose that function $f(x) \in R^1, x \in D,$ where D is a nonempty convex set.

Function $f(x)$ is said to be consistently convex if there exists a constant $\beta > 0$ and any points $x, y \in D$ so that for any $\alpha \in (0,1)$

$$\alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha) y) \geq (1-\alpha)\alpha\beta \|y - x\|^2 .$$

Corallary2. If function $f(x)$ is consistently convex.

Then the function $f(x)$ is strictly convex .

e.g. The function $f(x) = x^2, x \in R$ is consistently convex.

$(0 < \beta \leq 1)$

Th.9. Suppose the set D is nonempty open convex and $f(x): D \to R$ is 1st-order continuously differentiable. Then, function $f(x)$ is consistently convex. $\Longleftrightarrow$ There exists a constant $\beta > 0$, such that $f(y) - f(x) \geq \nabla f(x)^{\mathrm{T}}(y-x) + \beta\|y-x\|^2, \forall x, y \in D.$

Proof: $\Longrightarrow$ $\exists \beta > 0, s.t. \forall x, y \in D, \alpha \in (0,1)$, we have

$$\alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha) y) \geq (1-\alpha)\alpha\beta\|y-x\|^2.$$

That is

$$(1-\alpha) f(y) - (1-\alpha) f(x) \geq f(\alpha x + (1-\alpha) y) - f(x) + (1-\alpha)\alpha\beta\|y-x\|^2$$

Equivalently

$$f(y) - (x) \geq \frac{f(x + (1-\alpha)(y-x)) - f(x)}{1-\alpha} + \alpha\beta\|y-x\|^2$$

$$= \frac{\nabla f(x)^{\mathrm{T}}(1-\alpha)(y-x) + o(\|(1-\alpha)(y-x)\|)}{1-\alpha} + \alpha\beta\|y-x\|^2$$

$$\to \nabla f(x)^{\mathrm{T}}(y-x) + \beta\|y-x\|^2, \ (\alpha \to 1)$$

Therefore $f(y) - f(x) \geq \nabla f(x)^{\mathrm{T}}(y-x) + \beta\|y-x\|^2, \forall x, y \in D.$

$\Leftarrow$ **Let** $x^{*}=\alpha x+(1-\alpha)y.$ **Then**

$$f(x)-f(x^{*})\geq\nabla f(x^{*})^{\mathrm{T}}(x-x^{*})+\beta\left\|x-x^{*}\right\|^{2}$$

and

$$f(y)-f(x^{*})\geq\nabla f(x^{*})^{\mathrm{T}}(y-x^{*})+\beta\left\|y-x^{*}\right\|^{2}$$

Thus

$$\alpha f(x)+(1-\alpha)f(y)-f(\alpha x+(1-\alpha)y)\geq\beta\left[\alpha\left\|x-x^{*}\right\|^{2}+(1-\alpha)\left\|y-x^{*}\right\|\right]$$

$$=\beta\left[\alpha(1-\alpha)^{2}\left\|y-x\right\|^{2}+\alpha^{2}(1-\alpha)\left\|y-x\right\|^{2}\right]$$

$$=\beta\alpha(1-\alpha)\left\|y-x\right\|^{2}$$

Th.10. **Suppose the set** D **is nonempty open convex and** $f(x): D \to R$ **is 1st-order continuously differentiable. Then, function** $f(x)$ **is consistently convex.** $\Longleftrightarrow$ **There exists a constant** $\beta > 0$, **such that**

$$\left[\nabla f(y) - \nabla f(x)\right]^{\mathrm{T}}(y-x) \geq \beta \|y-x\|^2, \forall x, y \in D.$$

Proof: $\Longrightarrow$

For any $x, y \in D$, **from Th9 yields**

$$f(y) - f(x) \geq \nabla f(x)^{\mathrm{T}}(y-x) + \frac{\beta}{2}\|y-x\|^2, \forall x, y \in D.$$

and

$$f(x) - f(y) \geq \nabla f(y)^{\mathrm{T}}(x-y) + \frac{\beta}{2}\|y-x\|^2, \forall x, y \in D.$$

Thus $$\left[\nabla f(y) - \nabla f(x)\right]^{\mathrm{T}}(y-x) \geq \beta \|y-x\|^2, \forall x, y \in D.$$

Choose $\forall x, y \in D$ **& insert** $m(m>0)$ **points btw** x **and** y

denoted as $x+\frac{1}{m+1}(y-x), x+\frac{2}{m+1}(y-x), \cdots, x+\frac{m}{m+1}(y-x).$

Let $\lambda_k=\frac{k}{m+1}(k=0,1,\cdots,m+1),\ z_k=x+\lambda_k(y-x)(k=0,1,\cdots,m+1)$ **Then**

$$f(z_{k+1})-f(z_k)=\nabla f\left(z_k+\theta_k(z_{k+1}-z_k)\right)^{\mathrm{T}}(z_{k+1}-z_k) \quad (0<\theta_k<1)$$

$$=\frac{1}{m+1}\nabla f\left(x+\left(\lambda_k+\frac{\theta_k}{m+1}\right)(y-x)\right)^{\mathrm{T}}(y-x)$$

$$=\frac{1}{m+1}\nabla f\left(x+\xi_k(y-x)\right)^{\mathrm{T}}(y-x) \quad \left(\lambda_k<\xi_k=\lambda_k+\frac{\theta_k}{m+1}<\lambda_{k+1}\right)$$

Thus

$$f(y)-f(x)=\sum_{k=0}^{m}\left[f(z_{k+1})-f(z_k)\right]=\frac{1}{m+1}\sum_{k=0}^{m}\nabla f\left(x+\xi_k(y-x)\right)^{\mathrm{T}}(y-x)$$

$$=\nabla f(x)^{\mathrm{T}}(y-x)+\frac{1}{m+1}\sum_{k=0}^{m}\nabla f\left(x+\xi_k(y-x)\right)^{\mathrm{T}}(y-x)-\nabla f(x)^{\mathrm{T}}(y-x)$$

$$=\nabla f(x)^{\mathrm{T}}(y-x)+\frac{1}{m+1}\sum_{k=0}^{m}\frac{\left[\nabla f\left(x+\xi_k(y-x)\right)-\nabla f(x)\right]^{\mathrm{T}}\xi_k(y-x)}{\xi_k}$$

$$\geq\nabla f(x)^{\mathrm{T}}(y-x)+\frac{1}{m+1}\sum_{k=0}^{m}\frac{2\beta}{\xi_k}\left\|\xi_k(y-x)\right\|^2$$

$$=\nabla f(x)^{\mathrm{T}}(y-x)+\frac{2\beta\|y-x\|^2}{m+1}\sum_{k=0}^{m}\xi_k \quad >\nabla f(x)^{\mathrm{T}}(y-x)+\frac{2\beta\|y-x\|^2}{m+1}\sum_{k=0}^{m}\lambda_k$$

$$=\nabla f(x)^{\mathrm{T}}(y-x)+\frac{2\beta\|y-x\|^2}{m+1}\sum_{k=0}^{m}\frac{k}{m+1}$$

$$=\nabla f(x)^{\mathrm{T}}(y-x)+\frac{2\beta\|y-x\|^2 m(m+1)}{2(m+1)^2}$$

$$\rightarrow\nabla f(x)^{\mathrm{T}}(y-x)+\beta\|y-x\|^2 \quad (m\rightarrow\infty)$$

Therefore $$f(y)-f(x)\geq\nabla f(x)^{\mathrm{T}}(y-x)+\beta\|y-x\|^2.$$

Th.11. **Suppose the set** D **is nonempty open convex and** $f(x):D\to R$ **is 2nd-order continuously differentiable. Then, function** $f(x)$ **is consistently convex.** $\Longleftrightarrow$ **There exists a constant** $m>0$, **such that** $u^{\mathrm{T}}\nabla^2 f(x)u\ge m\|u\|^2,\forall x\in D,\forall u\in R^n.$

Proof: $\Longrightarrow$ **For** $\forall x\in D,u\in R^n,\alpha\in R,$ **we have**

$$u^{\mathrm{T}}\nabla^2 f(x)u=\left.\frac{d\nabla f(x+\alpha u)^{\mathrm{T}}}{d\alpha}\right|_{\alpha=0}u=\lim_{\alpha\to 0}\frac{\left(\nabla f(x+\alpha u)-\nabla f(x)\right)^{\mathrm{T}}u}{\alpha}$$

$$\ge\lim_{\alpha\to 0}\frac{\beta}{\alpha^2}\|\alpha u\|^2=\beta\|u\|^2\quad=m\|u\|^2\quad(m=\beta)$$

$\Longleftarrow$

$$f(y)-f(x)=\nabla f(x)^{\mathrm{T}}(y-x)+\frac{1}{2}(y-x)^{\mathrm{T}}\nabla^2 f\left(x+\theta(y-x)\right)(y-x)$$

$$\ge\nabla f(x)^{\mathrm{T}}(y-x)+\frac{1}{2}m\|y-x\|^2$$

Therefore $$f(y)-f(x)\ge\nabla f(x)^{\mathrm{T}}(y-x)+\beta\|y-x\|^2\left(\beta=\frac{m}{2}\right)$$

Level Set: Suppose that the set $D \subset R^n$ is nonempty convex and the function $f(x) \in R^1, x \in D$ is continuous and convex. Then, the set $L(x^{(0)}) = \{x \mid f(x) \le f(x^{(0)}), x \in D\}$ is called as a level set.

Th.12. Level set $L(x^{(0)})$ is convex and closed.

Proof: **Convexity:** Let $\forall x, y \in L(x^{(0)}), \alpha \in (0,1)$. Then

$$f(\alpha x + (1-\alpha) y) \le \alpha f(x) + (1-\alpha) f(y)$$
$$\le \alpha f(x^{(0)}) + (1-\alpha) f(x^{(0)}) = f(x^{(0)})$$

Therefore $\alpha x + (1-\alpha) y \in L(x^{(0)})$.

Closeness: Let x^* be an accumulation point of $L(x^{(0)})$.

Then $\exists \{x^{(k)}\} \subset L(x^{(0)}), s.t. f(x^*) = f\left(\lim_{k\to\infty} x^{(k)}\right) = \lim_{k\to\infty} f(x^{(k)}) \le f(x^{(0)})$.

Therefore $x^* \in L(x^{(0)})$.

Th.13. Suppose set $D \subset R^n$ is nonempty open convex and function $f(x): D \to R$ is 2nd-order continuously differentiable. If there exists a constant $m > 0$ such that $u^{\mathrm{T}} \nabla^2 f(x) u \geq m\|u\|^2, \forall x \in L\left(x^{(0)}\right), u \in R^n$.

Then level set $L\left(x^{(0)}\right)$ is bounded, closed and convex.

Proof: Closeness and convexity can be derived from Th12.

Boundedness: $\forall y \in L\left(x^{(0)}\right)$, then

$$
\begin{aligned}
0 &\geq f(y) - f\left(x^{(0)}\right) \\
&= \nabla f\left(x^{(0)}\right)^{\mathrm{T}}\left(y - x^{(0)}\right) + \frac{1}{2}\left(y - x^{(0)}\right)^{\mathrm{T}} \nabla^2 f\left(x^{(0)} + \theta\left(y - x^{(0)}\right)\right)\left(y - x^{(0)}\right) \\
&\geq \nabla f\left(x^{(0)}\right)^{\mathrm{T}}\left(y - x^{(0)}\right) + \frac{m}{2}\left\|y - x^{(0)}\right\|^2
\end{aligned}
$$

Therefore $\left\|y - x^{(0)}\right\| \leq \frac{2}{m}\left\|\nabla f\left(x^{(0)}\right)\right\|$.

2.4 Separation of Convex Sets

Def2: Sets $D_1, D_2 \subset R^n$ nonempty. For any $a \in R^n, \beta \in R^1$ satisfying

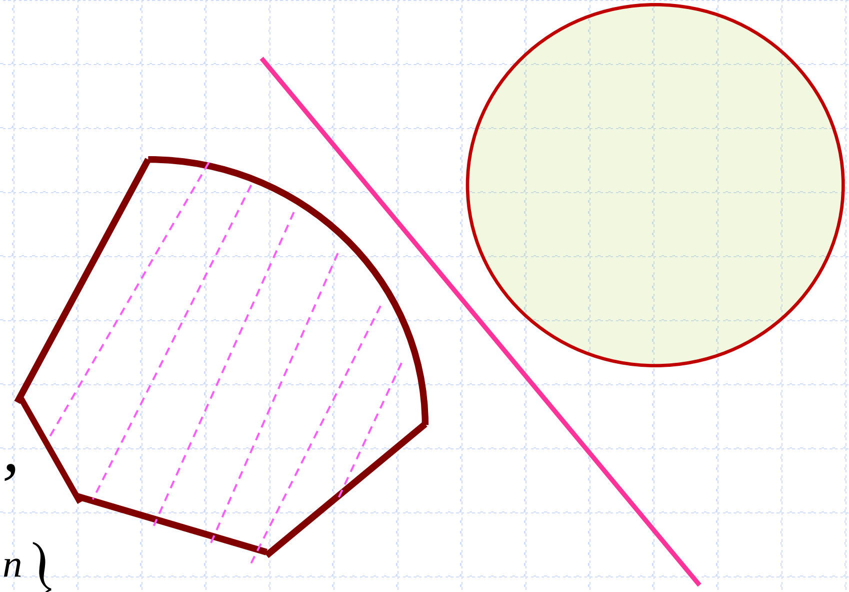
$$D_1 \subset H^+ = \{x|, a^{\mathrm{T}} x \geq \beta, x \in R^n\} \text{ and } D_2 \subset H^- = \{x|, a^{\mathrm{T}} x \leq \beta, x \in R^n\}.$$

Then superplane $H = \{x|, a^{\mathrm{T}} x = \beta, x \in R^n\}$

is said to separate set D_1 from D_2.

Superplane H is said to normally separate set D_1 from D_2

if $D_1 \cup D_2 \not\subset H$.



Superplane H is said to strictly

separate set D_1 from D_2

if $D_1 \subset \overset{0}{H}{}^+ = \{x|, a^{\mathrm{T}} x > \beta, x \in R^n\}$,

and $D_2 \subset H^- = \{x|, a^{\mathrm{T}} x \leq \beta, x \in R^n\}$.

Th.14.(Projection theorem)

Let set $D \subset R^n$ **Be nonempty closed convex.** $y \in R^n, y \notin D.$

Then **(1)** $\exists | \bar{x} \in D,$ **s.t.** $\|y-\bar{x}\| = inf\{\|y-x\|, x \in D\} > 0.$

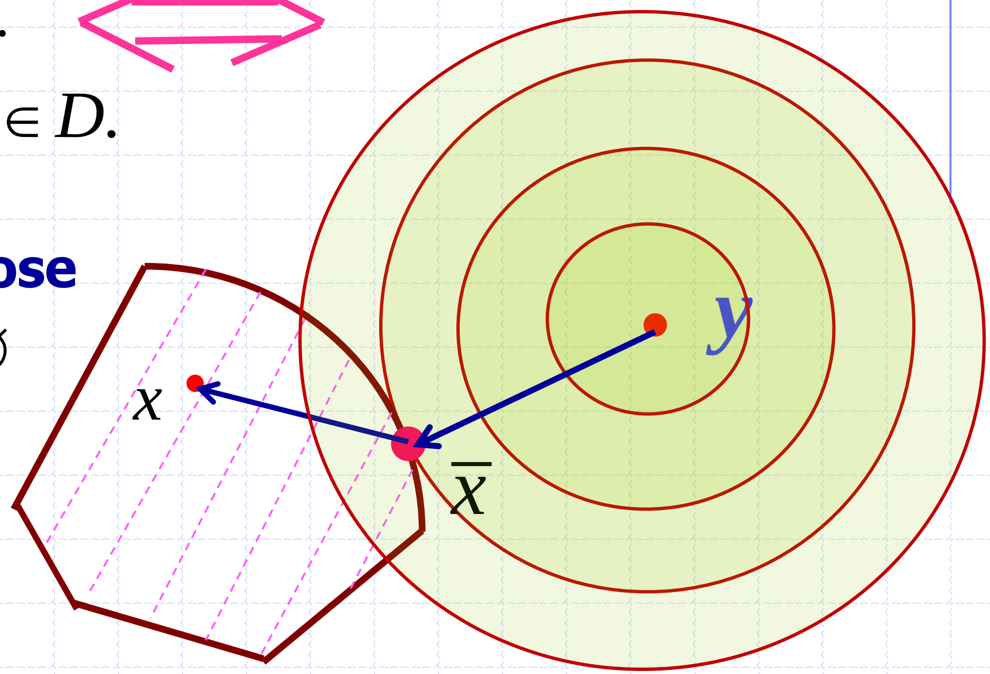
(2) $\|y-\bar{x}\| = inf\{\|y-x\|, x \in D\} > 0. \Longleftrightarrow$

$$(y-\bar{x})^T (x-\bar{x}) \le 0, \forall x \in D.$$

Proof: **(1) Existence**

Let $S = \{s \mid \|s\| \le 1, s \in R^n\}.$ **Choose** $\mu >> 0,$ **s.t.** $D \cap (y+\mu S) \neq \varnothing$

Thus $D \cap (y+\mu S)$ **is bnd closed.**



Therefore, continuous fcn $\|x-y\|$ **exists minimal point**

$\bar{x} \in D \cap (y+\mu S)$ **s.t.** $\|y-\bar{x}\| = inf\{\|y-x\|, x \in D\} > 0.$

Uniqueness: **If** $\exists \tilde{x} \in D, \tilde{x} \neq \bar{x}, s.t. \|y-\tilde{x}\| = \|y-\bar{x}\| = minimum$

Then $\dfrac{\tilde{x}+\bar{x}}{2} \in D$ **and** $\left\| y - \dfrac{\tilde{x}+\bar{x}}{2} \right\| < \dfrac{1}{2}\|y-\tilde{x}\| + \dfrac{1}{2}\|y-\bar{x}\| = minimum. \rightarrow\leftarrow$

Proof: **(2)** $\|y-\bar{x}\| = \inf\{\|y-x\|, x \in D\} > 0.$ $\Longleftrightarrow$

$$(y-\bar{x})^{\mathrm{T}}(x-\bar{x}) \le 0, \forall x \in D.$$

$\Rightarrow$ **For** $\forall x \in D,$ **we have** $\|y-\bar{x}\| \le \|y-x\|.$ **or** $\|y-\bar{x}\|^2 \le \|y-x\|^2.$

From $\bar{x} \in D$ **and** D **is convex, for** $\forall \alpha \in (0,1)$ **it yields**

$$\bar{x} + \alpha(x-\bar{x}) = \alpha x + (1-\alpha)\bar{x} \in D,$$

Then $$\|y-\bar{x}\|^2 \le \|y-\bar{x}-\alpha(x-\bar{x})\|^2 = \|(y-\bar{x})-\alpha(x-\bar{x})\|^2$$

$$= \|y-\bar{x}\|^2 - 2\alpha(y-\bar{x})^{\mathrm{T}}(x-\bar{x}) + \alpha^2\|x-\bar{x}\|^2.$$

Therefore $$(y-\bar{x})^{\mathrm{T}}(x-\bar{x}) \le \frac{\alpha}{2}\|x-\bar{x}\|^2 \to 0 \quad (\alpha \to 0).$$

$\Leftarrow$ $\forall x \in D,$ $$\|y-x\|^2 = \|(y-\bar{x})-(x-\bar{x})\|^2$$

$$= \|y-\bar{x}\|^2 - 2(y-\bar{x})^{\mathrm{T}}(x-\bar{x}) + \|x-\bar{x}\|^2 \ge \|y-\bar{x}\|^2$$

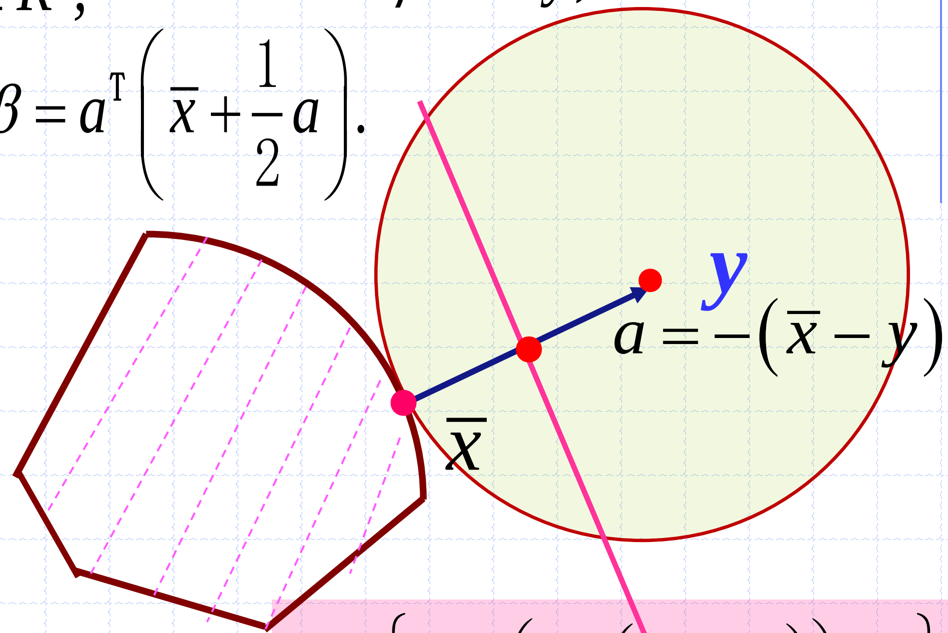
Thus $\|y-\bar{x}\| = \inf\{\|y-x\|, x \in D\} > 0.$

Th.15. (Separation theorem of point from convex set)

Let $D\subset R^n$ **be nonempty closed convex set.** $y\in R^n, y\notin D.$

Then there exist $a\in R^n, a\neq 0, \beta\in R^1,$ **s.t.** $a^{\mathrm{T}}x<\beta<a^{\mathrm{T}}y, \forall x\in D.$

Proof: **Denote** $a=-(\bar{x}-y),\ \beta=a^{\mathrm{T}}\left(\bar{x}+\frac{1}{2}a\right).$



Then for $\forall x\in D$

$$a^{\mathrm{T}}x=a^{\mathrm{T}}(x-\bar{x})+a^{\mathrm{T}}\bar{x}$$
$$=-(\bar{x}-y)^{\mathrm{T}}(x-\bar{x})+a^{\mathrm{T}}\bar{x}$$
$$\leq a^{\mathrm{T}}\bar{x}\quad<\beta$$

$$a^{\mathrm{T}}y=a^{\mathrm{T}}(a+\bar{x})=a^{\mathrm{T}}\left(\bar{x}+\frac{1}{2}a\right)+\frac{1}{2}a^{\mathrm{T}}a\quad>\beta$$

$$H=\left\{x \,\middle|\, a^{\mathrm{T}}\left(x-\left(\bar{x}+\frac{1}{2}a\right)\right)=0\right\}$$

Therefore $a^{\mathrm{T}}x<\beta<a^{\mathrm{T}}y, \forall x\in D.$

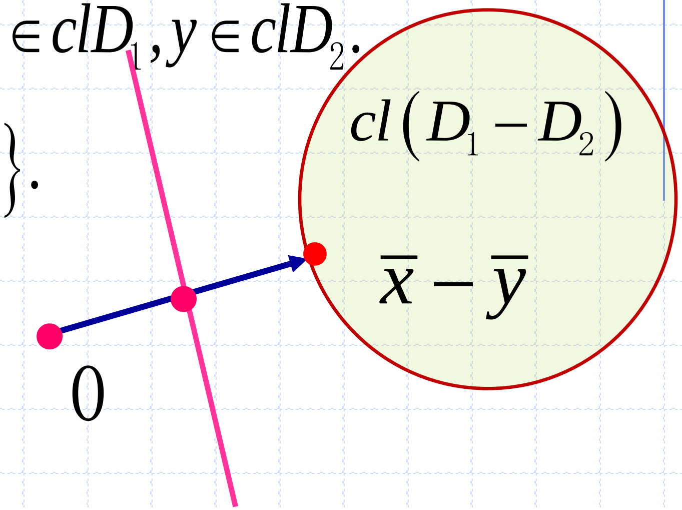
Thus, the plane $H=\left\{x \,\middle|\, a^{\mathrm{T}}x=a^{\mathrm{T}}\left(\bar{x}+\frac{1}{2}a\right)\right\}=\left\{x \,\middle|\, a^{\mathrm{T}}\left(x-\left(\bar{x}+\frac{1}{2}a\right)\right)=0\right\}$

Strictly separates y **from** D

Th.16. (Separation theorem for two convex sets)

Let $D_1, D_2 \subset R^n$ **be nonempty convex sets. If** $D_1 \cap D_2 = \varnothing$

then there exists $a \in R^n$ **s.t.** $a^{\mathrm{T}} x \le a^{\mathrm{T}} y, \forall x \in clD_1, y \in clD_2.$



Proof: **Let** $D = D_1 - D_2 = \{x - y \mid x \in D_1, y \in D_2\}.$

Then D **is convex and** $0 \notin D.$

If $0 \notin clD$ **then** $a = -(\bar{x} - \bar{y}).$

where $\|\bar{x} - \bar{y}\| = inf\,\|x - y\|, \forall x - y \in clD.$

If $0 \in clD$ **then** $0 \in \partial D.$ **Thus** $\exists 0 \ne z^{(k)} \notin clD, s.t \lim\limits_{k \to \infty} z^{(k)} = 0.$

From separation theorem of point from convex set, we have

$$\exists \left\{a^{(k)} \ne 0, \left\|a^{(k)}\right\| = 1, k = 1, 2, \cdots\right\}, s.t. a^{(k)\mathrm{T}}(x - y) < a^{(k)\mathrm{T}} z^{(k)} \to 0.$$

Denote $\left\{a^{(k_i)}\right\}$ **as the convergent subsequence. Then** $a = \lim\limits_{k_i \to \infty} a^{(k_i)}$

Question:

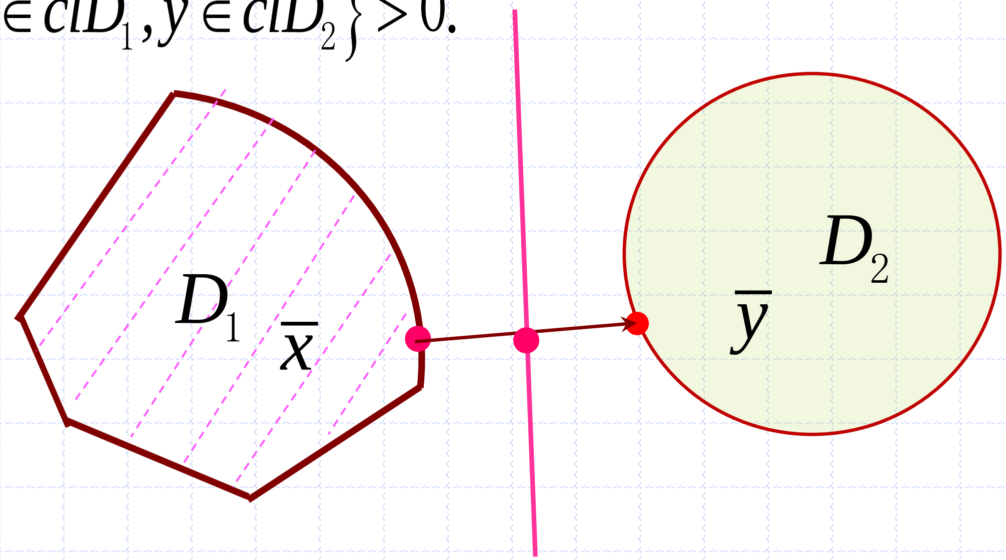
Is it possible to prove separation theorem for 2 convex sets by the method of separating a point from a convex set?

Step1. **If** $clD_1 \cap clD_2 = \varnothing$ **then there exist** $\bar{x} \in clD_1,\ \bar{y} \in clD_2,$

s.t. $\|\bar{x}-\bar{y}\| = inf\left\{\|x-y\|, x \in clD_1, y \in clD_2\right\} > 0.$

Step2. $a = -\left(\bar{x}-\bar{y}\right)$

$$\beta = a^{\mathrm{T}}\left(\bar{x}+\frac{1}{2}a\right).$$



prove the conclusion.

Step3. **Is it possible to construct sequence so that the conclusion is true for the case when** $clD_1 \cap clD_2 \neq \varnothing$ **?**

From separation theorem of a point from a convex set

Th.17. (Farkas Lemma) Let $A\in R^{m\times n}$ and $c\in R^n$. Then exactly one of the equalities $\begin{cases}Ax\le 0,\\ c^{\mathrm T}x>0.\end{cases}$ (E1) and $\begin{cases}A^{\mathrm T}y=c,\\ y\ge 0.\end{cases}$ (E2) has a solution.

If (E2) has a solution, then $\begin{cases}A^{\mathrm T}y=c,\\ y\ge 0.\end{cases} \Longrightarrow \begin{cases}c^{\mathrm T}=y^{\mathrm T}A,\\ y\ge 0.\end{cases}$

$\Longrightarrow\ 0<c^{\mathrm T}x=y^{\mathrm T}Ax\le 0.\ \Longrightarrow$ (E1) has no solution.

If (E2) has no solution, then $c\notin S=\left\{x\,\middle|\,x=A^{\mathrm T}y,y\ge 0\right\}$, where S is a polyhedral set and thus nonempty closed convex.

By Th15 there exist $0\ne p\in R^n,\alpha\in R^1$, s.t. $p^{\mathrm T}c>\alpha$ and $p^{\mathrm T}x\le\alpha$ for $\forall x\in S=\left\{x\,\middle|\,x=A^{\mathrm T}y,y\ge 0\right\}$. Since $0\in S$, $\alpha\ge p^{\mathrm T}x=p^{\mathrm T}0=0$. then $\boxed{p^{\mathrm T}c=c^{\mathrm T}p>\alpha\ge 0.}$ Note for $\forall y\ge 0$ we have

$c^{\mathrm T}p>\alpha\ge p^{\mathrm T}x=p^{\mathrm T}A^{\mathrm T}y=y^{\mathrm T}Ap.$ In specific if $y>>0$ implies $\boxed{Ap\le 0.}$

From separation theorem for 2 convex sets

Th.18. (Gordan theorem) Let $A \in R^{m \times n}$, then exactly one of the equalities $\{Ax < 0$ and $\begin{cases} A^{\mathrm{T}} y = 0, \\ y \geq 0, y \neq 0. \end{cases}$ Exists a solution.

THANK YOU FOR ATTENDING