- Standard form

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g_i(x) \le 0 \ (i = 1, 2 \cdots m)$$
$$h_j(x) = 0 \ (i = 1, 2 \cdots p)$$

$$\Omega = \left(\bigcap_{i=1}^{m} \{x : g_i(x) \le 0\} \right) \cap \left(\bigcap_{j=1}^{p} \{x ; h_j(x) = 0\} \right)$$

proof: ① $x \in \Omega \Rightarrow g_i(x) \le 0 \ (\forall i) \ \ h_j(x) = 0 \ (\forall j) \Rightarrow$ x belongs to every set in the intersections.

② x belongs to the intersection $\Rightarrow$ it satisfies all inequalities and equalities simultaneously. $\Rightarrow x \in \Omega$

- Optimization Modeling of Support Vector Machine

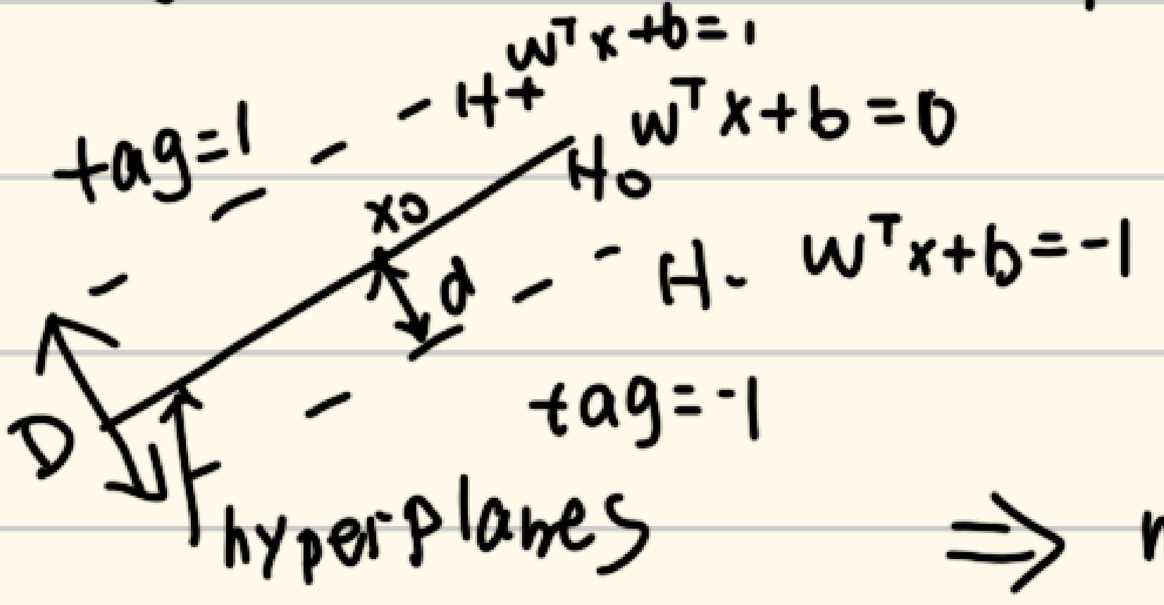
correctly classified

$y_i \in \{-1, 1\}$

linear function $u(x) = w^T x + b$ Find w, b. s.t. $y_i(w^T x_i + b) \ge 1 \quad \forall i = 1, 2 \cdots m$

$\begin{cases} w & w \perp \text{boundary} \quad \text{Changing } w \text{ rotates/tilts the boundary} \\ b & \text{shifts the boundary without changing its orientation} \end{cases}$

[$y_i(w^T x_i + b) \ge 0$ 存在自由度，有无穷解 (w_i, b_i) works, (cw_i, cb_i) also works]



$u = \frac{w}{\|w\|}$, $x_+ = x_0 + du$

$\Rightarrow w^T(x_0 + du) + b = d\frac{w^T \cdot w}{\|w\|} = 1 \Rightarrow d = \frac{1}{\|w\|} \Rightarrow D = \frac{2}{\|w\|}$

$\Rightarrow \underset{w, b}{\text{maximize}} \ \frac{2}{\|w\|} \quad \text{s.t. } y_i(w^T x_i + b) \ge 1, \forall i$

[所有点分类正确，且距超平面边界有一定距离]

$\Rightarrow \underset{w, b}{\text{minimize}} \ \|w\| \Rightarrow \underset{w, b}{\text{minimize}} \ \frac{1}{2}\|w\|^2$ ($\|w\|$ is not differentiable at 0)

$\hookrightarrow \nabla f(w) = w$

$\Rightarrow$ Soft-margin SVM $\quad \min_{w, b} \frac{\lambda}{2}\|w\|^2 + \sum_{i=1}^{n} \max\{0, 1 - y_i(w^T x_i + b)\}$

·Linear Programming

1. trasformation

* standard form (in course) $\min_{x\in\mathbb{R}^n} c^Tx$ s.t. $Ax=b$ $x\ge 0$

① max ⟺ min $\max\ c^Tx \Leftrightarrow \min\ \min(-c)^Tx$

② inequality ⟺ equality slack variable $a^Tx\le b \Leftrightarrow a^Tx+s=b, s\ge 0$ / $a^Tx\ge b \Leftrightarrow a^Tx-s=b$. $s\ge 0$

③ sign restriction $x_i\le 0 \Rightarrow x_i=-y_i \Rightarrow y_i\ge 0$

④ free variable $x_i=x_i^+-x_i^-$, $x_i^+\ge 0, x_i^-\ge 0$

⑤ auxiliary variable

$\underset{x}{\text{minimize}} \sum_{i=1}^m f_i(x)$ s.t. $x\in\Omega \Longleftrightarrow \underset{x,t}{\text{minimize}} \sum_{i=1}^m t_i$ s.t. $x\in\Omega$, $f_i(x)\le t_i$ $\forall i$

Proof: 1° let $t_i^*=f_i(x^*)$ $\Rightarrow$ $\begin{cases}\sum_{i=1}^m t_i^*=\sum_{i=1}^m f_i(x^*)\le\sum_{i=1}^m f_i(x)\le\sum_{i=1}^m t_i & \text{optimality}\\ f_i(x^*)=t_i^*\le t_i^* & \text{feasibility}\end{cases}$

2° let $x\in\Omega$, then $(x, f_1(x)\cdots f_m(x))$ is feasible

if (x^*, t^*) if optimal . $x^*\in\Omega$. $f_i(x^*)\le t_i^*$

$\Rightarrow \sum_{i=1}^m f_i(x^*)\le\sum_{i=1}^t t_i^*$ since $(x, f_1(x)\cdots f_m(x))$ is feasible

$\Rightarrow \sum_{i=1}^m t_i^*\le\sum_{i=1}^m f_i(x) \Rightarrow \sum_{i=1}^m f_i(x^*)\le\sum_{i=1}^m f_i(x)$

⑥ Linear fractional Programming

minimize $\dfrac{c^Tx+d}{e^Tx+a}$ s.t $Ax\le b$ assume $e^Tx+a>0$

let $z=\dfrac{1}{e^Tx+a}$, $y=\dfrac{x}{e^Tx+a}$ $\underset{y,z}{\text{minimize}}$ c^Ty+dz s.t. $\begin{array}{l} Ay-bz\le 0 \Leftrightarrow Ax\le b\\ e^Ty+az=1 \Leftrightarrow \text{relationship}(y,z)\\ z\ge 0 \Leftrightarrow e^Tx+a>0\end{array}$

2. ployhedron $\{x\in\mathbb{R}^n: Ax\ge b\}$

$Ax=b \Leftrightarrow Ax\ge b$ and $Ax\le b$ | $x\ge 0 \Leftrightarrow Ix\ge 0$

☆ standard form feasible set is the intersection of finitely many half-spaces, hence a ployhedron.

3. convex set $S\subseteq\mathbb{R}^n$ is convex if for any $x,y\in S$ and $\lambda\in[0,1]$ $\lambda x+(1-\lambda)y\in S$

if $\lambda_1+\cdots+\lambda_k=1$ ($\lambda_i\ge 0$) $\sum_{i=1}^k \lambda_i x_i$ is a convex combination of $x_1,\cdots,x_k$

$x\in P$ is an extreme point if we cannot find $y,z\in P$ ($y\ne x$, $z\ne x$) . $\lambda\in[0,1]$, s.t. $x=\lambda y+(1-\lambda)z$

4. Basic Solution for $Ax=b$ ($x\ge 0$)   m是 $Ax=b$ 约束个数 $A\in\mathbb{R}^{m\times n}$

① choose any m linearly independently columns of A $[A_{B(1)}\ A_{B(2)}\cdots A_{B(m)}]$

② set $x_i=0$ for all $i\ne B(1),\cdots,B(m)$

③ solve $Ax=b$ for the remaining variables $x_{B(1)}\cdots x_{B(m)}$

$x_B=A_B^{-1}b$ [A linearly independent]

5. Basic Feasible Solution (BFS) $x \geq 0$

* simplex method swap one basic and one nonbasic variable to find the optimal BFS

for standard LP feasible set $P = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$

x is an extreme point of P (a corner) $\Leftrightarrow$ x is a basic feasible solution

proof:

① extreme $\Rightarrow$ BFS

let x be an extreme point of P, then $Ax = b$ $(x \geq 0)$

let $B = \{i : x_i > 0\}$. Assume x is not BFS, $\{A_i : i \in B\}$ linearly dependent

$\Rightarrow \exists \alpha_{i \, i \in B} \neq 0$ s.t. $\sum_{i \in B} \alpha_i A_i = 0$ $\Rightarrow$ $A\alpha = 0$ $(i \notin B, \alpha_i = 0)$

$\Rightarrow \forall \epsilon$ $A(x \pm \epsilon\alpha) = Ax \pm \epsilon A\alpha = b \pm 0 = b$

Since $i \in B$. $x_i > 0$, $\exists \epsilon > 0$. $x \pm \epsilon\alpha > 0$ then $x = \frac{1}{2}(x + \epsilon\alpha) + \frac{1}{2}(x - \epsilon\alpha)$ is a convex combination of two different feasible points $\Rightarrow$ contradiction

② BFS $\Rightarrow$ extreme Assume x is not an extreme point of P

let $B = \{B_1, \cdots B_m\} \Rightarrow [A_{B_1} \cdots A_{B_m}]$ is linearly independent, $i \notin B$, $x_i = 0$

an extreme point x $\Rightarrow$ $x = \lambda y + (1-\lambda) z$ $(\lambda \in [0,1], y_j \neq x_j, z_j \neq x_j)$

$x_j = 0 = \lambda y + (1-\lambda) z$, $y_j \geq 0, z_j \geq 0 \Rightarrow y_j = z_j = 0$ for $j \notin B$

and $j \in B$ $Ax = B$ is invertible $\Rightarrow y_j = z_j = x_j$

$\Rightarrow y = z = x$ $\Rightarrow$ contradiction

6. Fundamental Theorem of Linear Programming

$P = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$ $(\mathrm{row}(A) = m)$

① $P \neq \emptyset$, P has a BFS

② opt. solution exists $\Rightarrow$ an optimal BFS

proof ①

Similar to 6.① $A(x - \epsilon\alpha) = Ax - \epsilon A\alpha = b$

consider $\alpha_i > 0$. define $\epsilon^* = \min_{i \in B, \alpha_i > 0} \frac{x_i}{\alpha_i} \Rightarrow x' = x - \epsilon^*\alpha \geq 0$ and at least one component becomes 0

proof ② let x be an optimal feasible solution, but not basic.

Similar to 6.① $A\alpha = 0$, for small ϵ $x \pm \epsilon\alpha$ still holds $Ax = b$

for a coefficient c. $c^T\alpha = 0$ ($c^T\alpha < 0$, $x + \epsilon\alpha$ has smaller objective value; $c^T\alpha > 0$, $x - \epsilon\alpha$ has smaller objective value $\Rightarrow$ contradict with x is optimal)

$\Rightarrow$ walk along $\pm x$ direction to find the optimal BFS [e.g.] whether it is a polyhedron

(1) $S=\{x\in\mathbb{R}^n \mid x\geq 0,\ x^Ty\leq 1,\ \text{for all } y \text{ with } \|y\|_2=1\}$

$S'=\{x\in\mathbb{R}^n \mid x\geq 0,\ \|x\|_2\leq 1\}$

① $S\subset S'$ for $x\neq 0$, choose $y=\frac{x}{\|x\|_2}$ satisfying $\|y\|_2=1$

$x^Ty=\frac{x^Tx}{\|x\|_2}=\|x\|_2\leq 1$

② $S'\subset S$ $\quad x^Ty\leq \|x\|_2\|y\|_2\leq 1$

$\Rightarrow S'=S$ the answer is no

(2) $S=\{x\in\mathbb{R}^n \mid x\geq 0,\ x^Ty\leq 1.\ \text{for all } y \text{ with } \|y\|_1=1\}$

① $\|x\|\leq 1 \Rightarrow |x^Ty| = \sum \|x_i\|\,\|y_i\|\leq \sum\|y_i\|=1$

② Let k be an index which $|x_k|=\max_i |x_i|$

$y_k=1$ if $x_k>0$; $y_k=-1$ if $x_k<0$, and $y_i=0$ for $i\neq k$

$\Rightarrow x^Ty=y_kx_k=|x_k|=\max_i |x_i|\leq 1$

$\Rightarrow 0\leq x_i\leq 1 \Rightarrow$ the answer is yes

7. Simplex Method [单纯形法]

① two basic solutions are neighboring if their bases differ by exactly one index

② Method: 1° Start from a BFS

2° move to a neighbor BFS (pick one non-basic variable x_j, increase it [enter the basis], adjust basic variable to hold $Ax=b$, Stop when some basic variable hits 0 [leave the basis])

original LP: $\min\ c^Tx$ $\quad x(\theta)=x+\theta d \Rightarrow c^Tx(\theta)-c^Tx=\theta c^Td=\theta(c_B^Td_B+c_N^Td_N)$

$=\theta(c_B^Td_B+c_j\cdot 1)$

$=\theta[c_B^T(-A_B^{-1}A_j)+c_j]$

the reduced cost

$\Rightarrow \bar{c}_j=c^Td=\boxed{c_j-c_B^TA_B^{-1}A_j}$ [the slope of objective]

$\begin{cases} \bar{c}_j<0 & \text{the objective decrease} \Rightarrow \text{promising direction} \\ \bar{c}_j>0 & \text{not chosen} \\ \bar{c}_j=0 & x_j \text{ doesn't change} \end{cases}$

reduced costs matter only for nonbasic variables

$\bar{c}_{B(i)}=c_{B(i)}-c_B^TA_B^{-1}A_{B(i)}=c_{B(i)}-c_B^TA_B^{-1}(A_Be_i)=c_{B(i)}-c_{B(i)}=0$

let $y^T=c_B^TA_B^{-1}$, $\bar{c}_j=c_j-y^TA_j$ ☆ $(A_B^Ty=c_B)$

Pick any j $(\bar{c}_j<0)$

consider a line $x(\theta)=x+\theta d$ ($\overset{\text{非负}}{\theta\geq 0}$) choose $j\in N$ to enter the basis $d_j=1$, $d_{j'}|_{j'\neq j}=0$

$Ax(\theta)=b \Leftrightarrow Ad=0$ [移动方向不会破坏约束] 对$\forall\theta$

$d=[d_B;d_N]$ $\quad Ad=0\Leftrightarrow A_Bd_B+A_j=0$ $\quad \boxed{d_B=-A_B^{-1}A_j}$

$x_j\leftarrow x_j+\theta \Rightarrow \begin{cases} x_B(\theta)=x_B+\theta d_B \\ x_j(\theta)=0+\theta\cdot 1=\theta \end{cases}$

$\boxed{\theta^* = \min_{i\in B: d_i<0} \frac{x_i}{-d_i}}$ The index attaining the minimum is the leaving variable — no $d_i<0 \Rightarrow$ unbounded $\Rightarrow$ stop

[$d_i<0$, we need to choose θ^* to keep $x_i+\theta d_i \ge 0$] Set $y = x+\theta^* d$. $B \leftarrow (B\setminus\{i^*\})\cup\{j\}$

3° Stop when all reduced costs are nonnegative

Proof Assume x is a BFS with basis B and reduced cost $\bar{c}\ge 0$. $Az=b\ (z\ge 0)$

$u=z-x \Rightarrow Au=0=A_Bu_B+A_Nu_N \qquad u_B=-A_B^{-1}A_Nu_N$

$c^T(z-x)=c_B^T(-A_B^{-1}A_Nu_N)+c_N^Tu_N=\sum_{j\in N}(c_j-c_B^TA_B^{-1}A_j)u_j=\sum_{j\in N}\bar{c}_ju_j$

x is a BFS $\Rightarrow x_N=0 \Rightarrow u_j=z_j-0\ge 0$

$\bar{c}_j\ge 0, u_j\ge 0 \Rightarrow c^T(z-x)\ge 0 \Rightarrow c^Tz\ge c^Tx \Rightarrow x$ is optimal

8. degeneracy A BFS is degeneracy if at least one basic variable equals to 0

i.e. $\exists i\in B$ s.t. $x_i=0$ [extra constraints are tight / ratio Test Tie]

• Pivoting rules

① Bland's Rule (smallest index rule) { Entering Choose the smallest index j with $\bar{c}_j<0$
Leaving If the ratio test has multiple minimizers, choose the smallest index among them.

② Most Negative Rule Choose the smallest $\bar{c}_j$

③ Most Descent Rule Choose j with the smallest $\theta^*\bar{c}_j$

9. Feasibility Theorem

*Phase I auxiliary LP: $\min_{x,y} 1^Ty$ s.t. $Ax+y=b$ $x\ge 0$, $y\ge 0$

The original problem is feasible iff the optimal value of phase I is 0

Proof

① $\Rightarrow$ feasible $\rightarrow \exists \hat{x}\ge 0$ s.t. $A\hat{x}=b$ set $\hat{y}=0$ $1^T\hat{y}=0$.

$A\hat{x}+\hat{y}=b\ (\hat{x}\ge 0, \hat{y}\ge 0)$ is feasible

② $\Leftarrow$ Phase I opt. Value $=0$ (x^*, y^*) $y^*\ge 0$, $1^Ty^*=0 \Rightarrow y^*=0$

$Ax^*=b$, $x^*\ge 0 \Rightarrow x^*$ is feasible for original LP

10. (1) Two-phase Method

Phase I Find feasibility ① ensure $b\ge 0$ ② $\min 1^Ty$ s.t. $Ax+y=b$, $x,y\ge 0$

③ minimum >0, LP is infeasible / $=0$ proceed to phase II

Phase II Optimize the original objective ① Start from phase I where $y^*=0$

② some y is still basic (degenerate case) $\Rightarrow$ pivot it out (or drop a redundant row) to form a full basis in x ③ run Simplex on $\min c^Tx$

(2) Big M Method $\min_{x,y} c^Tx+M1^Ty$ s.t. $Ax+y=b$ $x,y\ge 0$

$y^*=0 \Rightarrow$ feasible

11. Simplex Tableau

$c^T - c_B^T A_B^{-1} A$	$-c_B^T A_B^{-1} b$ x_B $[x_N=0]$	目标行
$A_B^{-1} A$	$A_B^{-1} b$	约束行

• $Ax=b \Rightarrow A_B^{-1}Ax = A_B^{-1}b$ $A=[A_B\ A_N]$ $A_B^{-1}A = [A_B^{-1}A_B\ \ A_B^{-1}A_N] = [I\ \ A_B^{-1}A_N]$

$\Rightarrow A_B^{-1}A \cdot \begin{bmatrix} x_B \\ x_N \end{bmatrix} = x_B + (A_B^{-1}A_N)x_N = A_B^{-1}b$ $x_N = 0 \Rightarrow$ $x_B = A_B^{-1}b$ (bfs)

• $\begin{cases} z = c_B^T x_B + c_N^T x_N \\ A_B x_B + A_N x_N = b \end{cases}$ $\Rightarrow z = c_B^T(A_B^{-1}b - A_B^{-1}A_N x_N) + c_N^T x_N$

$= (c_B^T A_B^{-1} b) + (c_N^T - c_B^T A_B^{-1} A_N) x_N$

$= c_B^T x_B + \bar{c}_N^T x_N$

省略

$\Rightarrow -z + \bar{c}_N^T x_N = -c_B^T x_B$

* Canonical Tableau Form $[B; N]$

0_m^T	$c_N^T - c_B^T A_B^{-1} A_N$	$-c_B^T x_B$
$I_m (A_B^{-1}A_B)$	$A_B^{-1} A_N$	$A_B^{-1} b = x_B$

1. $x_B + (A_B^{-1}A_N)x_N = A_B^{-1}b$ $z = c_B^T x_B + \sum_{j \in N} \bar{c}_j \cdot x_j$

entering variable x_j $d_B = -(A_B^{-1}A_j)$

12. Dual

$\begin{cases} \text{primal unique solution} \Leftrightarrow \text{dual nondegenerate} \\ \text{primal nondegenerate} \Leftrightarrow \text{dual unique solution} \end{cases}$

minimize $c^T x$
subject to $Ax = b$
$x \geq 0$
$\Longleftrightarrow$
maximize $b^T y$
subject to $A^T y \leq c$

$[x \in \mathbb{R}^n]$ $[y \in \mathbb{R}^m]$

$\min_{x \geq 0} \max_{y} [c^T x + y^T(b - Ax)] \Longleftrightarrow \max_{y} \min_{x \geq 0} [b^T y + x^T(c - A^T y)]$

If $b \neq Ax$, then $\exists y^T$ s.t. $y^T(b-Ax) \to \infty$ if $c - A^T y < 0$, $\exists x^T$ s.t. $x^T(c - A^T y) \to -\infty$, otherwise choose $x^T = 0$

对于某一可行解

* $g(y) = \min_{x \geq 0} c^T x + y^T(b - Ax) \leq c^T x^* + y^T(b - Ax^*) = c^T x^*$ [dual $\max g(y)$ no constraints]

$= y^T b + \min_{x \geq 0} (c^T - y^T A) x \begin{cases} y^T b + 0 & \text{if } c^T - y^T A \geq 0 \\ y^T b + (-\infty) & \text{otherwise} \end{cases}$

找到最紧的下界

$\Rightarrow \max\ y^T b$
s.t. $A^T y \leq c$

$Ax \leq b,\ x \geq 0$

$\max_{y \geq 0} -b^T y + \min_{x \geq 0} x^T(c + A^T y)$ $\Rightarrow$ $\max_y b^T y$
s.t. $A^T y \leq c$
$y \leq 0$

Primal	minimize	maximize	Dual	
constraints	$\geq b_i$ $\leq b_i$ $= b_i$	≥ 0 ≤ 0 free	variables	①
variables	≥ 0 ≤ 0 free	$\leq c_j$ $\geq c_j$ $= c_j$	constraints	②

若 $\overset{P}{\max} \Leftrightarrow \overset{D}{\min}$，则需要变号

① $Ax \geq b \Rightarrow y^T(b-Ax) \leq 0 \Rightarrow y^T \geq 0$

② $\min\limits_{x\geq 0} (c-A^Ty)\cdot x \Rightarrow c-A^Ty \geq 0$ (avoid going to $-\infty$) $\Rightarrow A^Ty \leq c$

- Weak Duality Theorem

Let x be any feasible solution to the Primal, y be any feasible solution to the Dual, then $b^Ty \leq c^Tx$

proof:

$b^Ty = (Ax)^Ty = x^T(A^Ty) \leq x^Tc = c^Tx$ $\begin{cases} Ax=b, x\geq 0 \text{ (Primal feasibility)} \\ A^Ty \leq c \text{ (Dual feasibility)} \end{cases}$

$\begin{cases} \text{If Primial is unbounded } (-\infty) \rightarrow \text{Dual infeasible} \\ \text{If Dual is unbounded } (+\infty) \rightarrow \text{Primal infeasible} \end{cases}$

- Strong Duality Theorem

If a linear program (Primal) has an optimal solution, then its dual also has an optimal solution, and $c^Tx^* = b^Ty^*$

Proof: Assume the Primal has an optimal solution x^* that is a BFS with optimal basis B. The reduced cost $c_j - c_B^T A_B^{-1} A_j \geq 0$

define $y^T = c_B^T A_B^{-1} \Rightarrow c_j \geq A_j^T y \Rightarrow A^Ty \leq c$ [y is dual feasible]

$b^Ty = b^T(A_B^{-T} c_B) = c_B^T A_B^{-1} b = c_B^T x_B = c^Tx^*$

P \ D	Finite Optimum	Unbounded	Infeasible
Finite Optimum	✓	×	×
Unbound	×	$\boxed{\times}$	✓
Infeasible	×	✓	✓

for $\min c^Tx \Leftrightarrow \max b^Ty$ $\quad b^Ty \leq c^Tx$ ① $c^Tx \to -\infty$，b^Ty 不可能有解
② c^Tx 有 min，b^Ty 有上界

- Complementarity Conditions

Let x, y be feasible points of the primal and dual problems, respectively.
Then x and y are optimal solutions iff. $\begin{cases} x_j (c_j - A_j^T y) = 0 & \forall j \\ y_i (a_i^T x - b_i) = 0 & \forall i \end{cases}$

① $c^T x = b^T y \Rightarrow c^T x - b^T y = c^T x - y^T A x = \sum_{i=1}^{n} (c_i - A_i^T y) x_i$

x, y feasible $\Rightarrow c_i - A_i^T y \geq 0, x_i \geq 0 \quad \Rightarrow (c_i - A_i^T y) x_i = 0 \quad (i = 1, \cdots, n)$

② $c^T x - b^T y = y^T A x - b^T y = y^T (Ax - b) = 0$

\+ $\min \ c^T x$ s.t. $Ax = b \ (x \geq 0)$ $\rightarrow$ $\max \ b^T y$ s.t. $A^T y + s = c \ (s \geq 0)$ $\rightarrow$ $x_i \cdot s_i = 0$

[e.g.] minimize $\sum_{i=1}^{m} t_i$
subject to $y_i (x_i^T w + b) + t_i \geq 1 \quad \forall i = 1, \cdots, m$
$t_i \geq 0 \rightarrow u \geq 0$

$(\operatorname{diag}(y) x^T \ \ y \ \ I) \begin{pmatrix} w \\ b \\ t \end{pmatrix} \geq \mathbb{1}$

$\begin{pmatrix} x \operatorname{diag}(y) \\ y^T \\ I \end{pmatrix} u \geq 1$

$= 0$ (free)

$\rightarrow = 0$ (free)

$\leq 1 \ (t \geq 0)$

$\Rightarrow$ maximize $\sum_{i=1}^{m} u_i$
subject to $x \operatorname{diag}(y) u = 0$
$y^T u = 0$
$0 \leq u_i \leq 1 \quad \forall i = 1, \cdots, m$

- Theorem of Alternatives

If we can find a vector x s.t. $Ax = 0, x \geq 0, c^T x < 0$, then $A^T y \leq c$ cannot have a solution.

Assume $\exists y$ s.t. $A^T y \leq c$ $\quad x^T A^T y \leq x^T c \quad (Ax)^T y \leq c^T x$

$Ax = 0 \quad$ LHS $= 0 \leq$ RHS < 0 $\Rightarrow$ contradiction

13. Sensitive analysis

(1) local sensitivity (linear approximations)

[Unique solution x^* / y^* otherwise not differentiable]

① $V(b) = \max \{ c^T x \mid Ax = b, x \geq 0 \}$ $\Rightarrow \nabla_b V(b) = y^*$

dual problem $\quad \underset{y}{\text{maximize}} \ b^T y$
s.t. $A^T y \leq c$

$V(b) = b^T y^* \Rightarrow \Delta V = \sum_i (\Delta b_i) y_i^*$

② $V(c) = \max \{ c^T x \mid Ax = b, x \geq 0 \}$ $\quad V(c) = c^T x^* \Rightarrow \nabla_c V(c) = x^*$

(2) gobal sensitivity

① $\tilde{b} = b + \Delta b \quad \tilde{x}_B = A_B^{-1} \tilde{b} = x_B^* + A_B^{-1} \Delta b$ [feasibility]

- Assume $\tilde{x}_B \geq 0$, B was optimal before change

proof:
$A \tilde{x} = A_B \tilde{x}_B + A_N \tilde{x}_N = A_B (A_B^{-1} \tilde{b}) = \tilde{b}$, $\tilde{x} \geq 0$ because $\tilde{x}_B \geq 0, \tilde{x}_N = 0$

Reduced cost : $c^T - c_B^T A_B^{-1} A \geq 0 \quad \Rightarrow B$ is still optimal

- $V(\tilde{b}) = \tilde{b}^T y^* = V^* + (y^*)^T \Delta b$

A_B^{-1} 原基变量对应内容 $A_B^{-1} \cdot I = A_B^{-1}$

$A_B^{-1}A_N$ 将非基变量对应内容 $x_B = A_B^{-1}b - \boxed{A_B^{-1}A_N x_N}$

$A_B x_B + A_N x_N = b$

- let $\Delta b = \lambda e_i$ $\quad x_B = x_B^* + \lambda A_B^{-1} e_i$

let $d = A_B^{-1} e_i$ $\Rightarrow x_B^* + \lambda d \geq 0$ i.e. $(x_B^*)_k + \lambda d_k \geq 0$

if $d_k > 0$, $\lambda \geq -\frac{(x_B)^*_k}{d_k}$; if $d_k < 0$ $\lambda \leq -\frac{(x_B^*)_k}{d_k}$; if $d_k = 0$ no constraints

$\Rightarrow \lambda \in [\max\{\text{lower bounds}\}, \min\{\text{upper bounds}\}]$

② $\tilde{c} = c + \Delta c$ $\quad \Delta c = \lambda e_j$ $\quad C_B = 0$ only consider $C_N = C_N^T - C_B^T A_B^{-1} A_N$

1° $j \in B$ $\quad C_N^T - (C_B^T + \lambda e_j^T) A_B^{-1} A_N = r_N^T - \lambda e_j^T A_B^{-1} A_N \geq 0$

2° $j \in N$ $\quad (C_N^T + \lambda e_j^T) - C_B^T A_B^{-1} A_N = r_N^T + \lambda e_j^T \geq 0$ $\Rightarrow$ maintain the optimality

③ change A $\quad$ recompute $\bar{C}_j$ $\quad$ if $\bar{C}_j < 0 \Rightarrow$ simplex method

for maximum, e_j 中 1 变 -1

14. The Interior Point Method

$Ax = b$. $x \geq 0$

$A^T y + s = c$. $s \geq 0$

$x_i s_i = 0$. $\forall i$

$x_i = 0 / s_i = 0$ $\xrightarrow{\text{The central Path}}$

$Ax = b$. $x \geq 0$ $\quad u > 0$

$A^T y + s = c$, $s \geq 0$ $\quad$ the complementarity gap

$x_i \cdot s_i = u$

Start from a certain u.

decrease u until find a solution of LP

The Interior Point Method will always find the optimal solution with the maximum possible number of non-zeros. [hight-rank]

- Integer Optimization

1. thoughts of relaxation $\quad V^{Rounding} \le V^{IP} \le V^{LP}$ (for maximization)

2. Total Unimodularity (TU)

A is total unimodularity if the determinant of every square submatrix of A is either 1, 0, -1.

- If A is TU, b is integer $\Rightarrow$ all BFs are integers $\Rightarrow$ Simplex Method
- Matrix A is TU if all these are true:

The "Row Partition" Test

Matrix A is TU if **all** these are true:

1. **Values:** Every entry is $0, +1,$ or -1.
2. **Counts:** Every column has at most **two** non-zero entries.
3. **Partition:** You can split the rows into two teams (Sets B and C) such that for every column:
   - 3.1 **Same Sign Case:** If the two non-zeros have the same sign (e.g., both $+1$), their rows must be in different teams (one in B, one in C).
   - 3.2 **Diff Sign Case:** If the two non-zeros have opposite signs $(+1, -1)$, their rows must be in the same team (both B or both C).

3. Divide and Conquer

① Pruning rule $\quad$ If the LP relaxation of a branch is less than or equal to current incumbent
全局下界
we prune this branch entirely.

② If the solution is an integer, then it is optimal to the IP

① Initialize the Incumbent $= -\infty$
② Subproblem $\quad 1^\circ$ infeasible [prune] $\quad 2^\circ$ LP value $\le$ Incumbent [prune by bound]
3° solution is integer (> Incumbent, update) stop $\quad 4^\circ$ branch
③ add constraints $\quad x_i \le \lfloor x_i^* \rfloor, \quad x_i \ge \lceil x_i^* \rceil$

- Nonlinear Optimization

1. First order Necessary Conditions (FONC)

If x^* is a local minimizer of the unconstrained problem $\min_{x \in \mathbb{R}^n} f(x)$. then $\nabla f(x^*) = 0$

2. Second Order Necessary Conditions (SONC)

$$f(x+td) = f(x) + t\nabla f(x)^T d + \frac{1}{2} t^2 d^T \nabla^2 f(x) d + O(t^2)$$

If x^* is a local minimizer of f, then it holds that $\begin{cases} \nabla f(x^*) = 0 \\ \forall d \in \mathbb{R}^n \quad d^T \nabla^2 f(x^*) d \ge 0 \end{cases}$ $\quad (\nabla^2 f(x^*) \succeq 0)$

[PSD matrices : $A^T A$ $\quad x^T A x = \frac{1}{2} x^T (A + A^T) x$ if A is not symmetric]

3. Second Order Sufficient Conditions (SOSC) $\begin{cases} \nabla f(x^*) = 0 \\ \forall d \in \mathbb{R}^n \setminus \{0\} \quad d^T \nabla^2 f(x^*) d > 0 \end{cases}$ (PD)

Proof: Lemma: $\lambda_{min}(A)\|x\|^2 \le x^T A x \le \lambda_{max}(A)\|x\|^2$ λ eigenvalues (A is symmetric)

$f(x^*+d) = f(x^*) + \frac{1}{2} d^T \nabla^2 f(x^*) d + o(\|d\|^2)$ $\|d\| \to 0$

when $\nabla^2 f(x^*)$ is PD $d^T \nabla^2 f(x^*) d \ge \mu \|d\|^2$ ($\mu > 0$, the smallest eigenvalue of $\nabla^2 f(x^*)$)

$f(x^*+d) \ge f(x^*) + \frac{\mu}{2}\|d\|^2 + o(\|d\|^2) = f(x^*) + \|d\|^2 \left(\frac{\mu}{2} + \frac{o(\|d\|^2)}{\|d\|^2}\right)$

Since $\|d\| \to 0$ $\frac{o(\|d\|^2)}{\|d\|^2} \to 0 \Rightarrow f(x^*+d) > f(x^*)$

4. $\nabla f(x) = 0$ critical point/stationary point

$\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is indefinite $\Rightarrow$ saddle point (indefinite 有正有负)

5. FONC for constrained problem

① feasible direction $x \in \Omega$ d is a feasible direction at x if $\exists \bar{t} > 0$ s.t. $x + td \in \Omega$ $\forall\, 0 \le t \le \bar{t}$

- $\Omega = \{x : Ax \ge b\}$

for $a_i^T x > b$, already have some slack ; for $a_i^T x = b$ $a_i^T(x+td) = a_i^T x + a_i^T t d \ge 0$

$\Rightarrow a_i^T d \ge 0$

② tagent cone $T_\Omega(x) := \{d \in \mathbb{R}^n : \exists t_k \downarrow 0, \exists d^k \to d, \text{ s.t. } x + t_k d^k \in \Omega\ \forall k\}$

(t_k 趋近于0, 步长越来越小) 切锥

- If $\Omega = \{x \in \mathbb{R}^n : Ax = b\}$, then for any $x \in \Omega$, $T_\Omega(x) = \{d \in \mathbb{R}^n : Ad = 0\}$ $b + tAd = b \Rightarrow Ad = 0$
  - $\Omega = \{Ax \ge b\}$ $T_\Omega(x) = \{d \in \mathbb{R}^n : a_i^T d \ge 0 \text{ for active } i\}$ active $\mathcal{A}(x) = \{i : a_i^T x = b_i\}$

③ Normal Cone Let $\Omega \in \mathbb{R}^n$ and let $x \in \Omega$ $N_\Omega(x) = \{v \in \mathbb{R}^n \ v^T d \le 0 \text{ for all } d \in T_\Omega(x)\}$

if $\Omega = \{Ax \ge b\}$ $N_\Omega(x) = \{-\sum_{i \in \mathcal{A}(x)} \lambda_i a_i : \lambda_i \ge 0\}$ $\Rightarrow \begin{cases} N_\Omega(x) = \{0\} & \text{if } a^T x > b \\ N_\Omega(x) = -\lambda a \ (\lambda \ge 0) & \text{if } a^T x = b \end{cases}$

active constraints

钝角, 不合法移动

$\underline{-\lambda(a^T d) \le 0}$

* x^* is a local minimum of $\min_{x \in \Omega} f(x)$, $f(x)$ is continuously differentiable, then for any feasible direction d at x^*, $\nabla f(x^*)^T d \ge 0$

(for unconstrained problem $\nabla f(x^*)^T d \ge 0$ and $\nabla f(x^*)^T(-d) \ge 0 \Rightarrow f(x^*) = 0$]

proof Since d is a feasible direction at x^* $\exists \bar{t} > 0$ s.t. $x^* + td \in \Omega$ $\forall\, 0 \le t \le \bar{t}$

$\Rightarrow f(x^*+td) \ge f(x^*)$

$f(x^*+td) = f(x^*) + t \nabla f(x^*)^T d + o(t)$ $(t \to 0)$ $f(x^*+td) - f(x^*) = t \nabla f(x^*)^T d + o(t) \ge 0$

- $\nabla f(x^*)^T d \ge 0$ for all $d \in T_\Omega(x^*)$ $-\nabla f(x^*)^T d \le 0 \Rightarrow -\nabla f(x^*) \in N_\Omega(x^*)$

6. FONC for Linearly Constrained Problems

$\min_x f(x)$ s.t. $Ax \geq b$ $\Rightarrow$ If x^* is a local minimum, $\exists y \in \mathbb{R}^m$ $\begin{cases} \nabla f(x^*) - A^T y = 0 \\ y_i(a_i^T x^* - b_i) = 0 \end{cases}$

proof: $\nabla f(x^*)^T d \geq 0 \quad \forall d \in S_\Omega(x^*)$

$\Leftrightarrow \left[\min_{d \in \mathbb{R}^n} \nabla f(x^*)^T d \text{ s.t. } d \in S_\Omega(x^*)\right] \geq 0 \Leftrightarrow \left[\min_{d \in \mathbb{R}^n} \nabla f(x^*)^T d \text{ s.t. } a_i^T d \geq 0 \ \forall i \in A(x^*)\right] \geq 0$

define $C = \begin{bmatrix} c_1^T \\ \vdots \\ c_m^T \end{bmatrix}$ $c_i^T = a_i^T$ if $i \in A(x^*)$; $c_i^T = 0$ if $i \in \mathcal{I}(x^*)$

$\Leftrightarrow \left[\min_{d \in \mathbb{R}^n} \nabla f(x^*)^T d \text{ s.t. } Cd \geq 0\right] \geq 0$

the dual of the problem is $\max_y 0$ s.t. $y \geq 0$, $C^T y = \nabla f(x^*)$

$\nabla f(x^*) = \sum_{i \in A(x^*)} c_i y_i + \sum_{i \in \mathcal{I}(x^*)} c_i y_i \quad \Rightarrow \quad y \geq 0.\ \nabla f(x^*) = A^T y \quad y_i(a_i^T x^* - b_i) = 0$

- $\min_x f(x)$ s.t. $Ax = b \Rightarrow \nabla f(x^*) = A^T y$

define $\begin{cases} A(x^*) : \{i : g_i(x^*) = 0\} \\ \mathcal{I}(x^*) : \{i : g_i(x^*) < 0\} \end{cases}$

7. Linearized Tangent cone

$\mathcal{L}_\Omega(x^*) = \{d \in \mathbb{R}^n \ \nabla g_i(x^*)^T d \leq 0, \ \forall i \in A(x^*)\}$

① for the feasible set $\Omega = \{x : g_i(x) \leq 0 \ \forall i\}$. always have $T_\Omega(x^*) \subseteq \mathcal{L}_\Omega(x^*)$

$\forall i \in A(x^*)$. $g_i(x^*) = 0 \geq g_i(x^* + t_k d^k)$ [let $d \in T_\Omega(x^*)$]

$\frac{g_i(x^* + t_k d^k) - g_i(x^*)}{t_k} = \frac{\nabla g_i(x^*)^T (t_k d^k) + o(t_k)}{t_k} \leq 0 \quad (k \to \infty,\ d^k \to d,\ \frac{o(t_k)}{t_k} \to 0)$

$\Rightarrow \nabla g_i(x^*)^T d^k \leq 0 \Rightarrow d \in \mathcal{L}_\Omega(x^*)$

② If x^* is a local minimizer, there does not exist a vector $d \in \mathbb{R}^n$ s.t.

$\nabla f(x^*) d < 0$ and $\nabla g_i(x^*)^T d < 0 \quad \forall i \in A(x^*)$.

Assume d exists. By taloyer expansion, $f(x^* + td) < f(x^*)$

for each $i \in A(x^*)$. $g_i(x^* + td) = g_i(x^*) + t \nabla g_i(x^*) d + o(t) < 0$ for small $t > 0$

for each $i \in \mathcal{I}(x^*)$ $g_i(x^*) < 0$

$\Rightarrow$ contradict with x^* is a local minimizer

8. Linearized Normal Cone

$= \{v \in \mathbb{R}^n : v^T d \leq 0, \forall d \in \mathcal{L}_\Omega(x^*)\} = \left\{\sum_{i \in A(x^*)} \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{p} \mu_j \nabla h_j(x^*) : \lambda_i \geq 0, \mu_j \in \mathbb{R}\right\}$

$N_\Omega^{lin}(x^*) = \left\{\sum_{i=1}^m \lambda_i \nabla g_i(x^*) : \lambda_i \geq 0, \lambda_i g_i(x^*) = 0\right\}$

$\nabla g_i(x^*)^T d \leq 0\ [\forall i \in A(x^*)]$

for any $v = \sum_{i=1}^m \lambda_i \nabla g_i(x^*)$ with $\lambda_i \geq 0$, $\lambda_i g_i(x^*) = 0$, for any $d \in \mathcal{L}_\Omega(x^*)$ $v^T d = \sum_{i \in A(x^*)} \lambda_i \nabla g_i(x^*)^T d \leq 0$

9. Fritz-John Conditions

If x^* is a local minimizer. $\exists \lambda_0 \geq 0$. $\lambda_1, \cdots \lambda_m \geq 0$ (not all zero). s.t.

$\lambda_0 \nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) = 0$ and $\lambda_i g_i(x^*) = 0 \quad i = 1, \cdots, m$

$\Rightarrow -\lambda_0 \nabla f(x^*) = -\sum_{i=1}^m \lambda_i \nabla g(x^*) \Rightarrow -\lambda_0 \nabla f(x^*) \in N_\Omega^{lin}(x^*)$

10. KKT [When $T_\Omega(x^*) = \mathcal{L}(x^*) \Rightarrow$ CQ点. 可用 KKT]

$\{\nabla g_i(x^*) : i \in A(x^*)$ are linearly independent$\}$

$\min_{x\in\mathbb{R}^n} f(x)$ s.t. $g_i(x) \le 0$ x^* is a local minimizer and LICQ holds at x^*.

there exist $\lambda_1, \cdots \lambda_m \ge 0$ s.t. $\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) = 0$, $\lambda_i \nabla g_i(x^*) = 0$

claim in 9. $\lambda_0 \neq 0$ Assume the opposite $\lambda_0 = 0$, $\sum_{i=1}^m \lambda_i \nabla g_i(x^*) = 0 \Rightarrow \sum_{i\in A(x^*)}^m \lambda_i \nabla g(x^*) = 0$

since LICQ hold ats x^*. $\lambda_1 = \cdots \lambda_m = 0$

$\Rightarrow \lambda_0 > 0$. $\bar{\lambda}_i = \frac{\lambda_i}{\lambda_0} \Rightarrow \nabla f(x^*) + \sum_{i=1}^m \bar{\lambda}_i \nabla g_i(x^*) = 0$, $\bar{\lambda}_i \nabla g_i(x^*) = 0$

$\Rightarrow \min_{x\in\mathbb{R}^n} f(x)$ s.t. $g_i(x) \le 0$, $h_j(x) = 0$

$N_\Omega^{lin}(x^*) = \left\{ \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{j=1}^{?} M_j \nabla h_j(x^*) : \lambda_i \ge 0, \lambda_i g_i(x^*) = 0, M_j \in \mathbb{R} \right\}$

- Lagrangian: $L(x, \lambda, M) = f(x) + \sum_{i=1}^m \lambda_i g_i(x) + \sum_{j=1}^p M_j h_j(x)$
- KKT in Lagrangian form: x^* local minimum, there exist $\lambda_i \ge 0$, $M_j \in \mathbb{R}$ s.t.

$\nabla_x L(x^*, \lambda, M) = 0$. $g_i(x^*) \le 0$. $h_j(x^*) = 0$. $\lambda_i g_i(x^*) = 0$

$\hookrightarrow -\nabla f(x^*) \in N_\Omega^{lin}(x^*)$

* if a point satisfy the constraint qualification (CQ), then KKT condition is a necessary condition for it to be a minimizer. [for convex problem, KKT is sufficient for a global minimizer]

- Convexity

1. linear restriction viewpoint

f is convex $\Leftrightarrow$ every line segment contained in Ω, the one variable function $\phi(t) = f(x+td)$ is convex in t.

2. $f: \Omega \to \mathbb{R}$ be differentiable on a convex set Ω. f is convex iff. $f(y) \ge f(x) + \nabla f(x)^T(y-x)$ $\forall x, y \in \Omega$

- $\nabla f(x^*) = 0 \Rightarrow f(y) \ge f(x^*)$ $\forall y \in \Omega$ x^* is a global minimizer

proof:

① f is convex $\Rightarrow f(y) \ge f(x) + \nabla f(x)^T(y-x)$

$g(t) = f(x + t(y-x))$ $t \in [0,1]$

f is convex $\Rightarrow$ g(t) is convex $g((1-t)\cdot 0 + t\cdot 1) \le (1-t)g(0) + tg(1)$

$\Rightarrow \frac{g(t) - g(0)}{t - 0} \le g(1) - g(0)$ $g'(0) \le g(1) - g(0)$

$\nabla f(x)^T (y-x) \le f(y) - f(x)$

② $f(y) \geq f(x) + \nabla f(x)^T (y-x) \Rightarrow f$ is convex

let $z = \lambda x + (1-\lambda) y$ $\quad \lambda \in [0,1]$ $\quad (x, y, z \in \Omega)$

$$\Rightarrow \begin{cases} f(x) \geq f(z) + \nabla f(z)^T (x-z) & ① \\ f(y) \geq f(z) + \nabla f(z)^T (y-z) & ② \end{cases}$$

$① \times \lambda + ② \times (1-\lambda) \Rightarrow \lambda f(x) + (1-\lambda) f(y) \geq f(z) + \nabla f(z)^T [(\lambda x + (1-\lambda) y) - z] = f(z)$

i.e. $\lambda f(x) + (1-\lambda) f(y) \geq f(\lambda x + (1-\lambda) y)$

3. f is convex on $\Omega \Leftrightarrow \nabla^2 f(x) \succeq 0$, $\forall x \in \Omega$

$\phi(t) = f(x+td)$, $\phi''(t) = d^T \nabla^2 f(x+td) d$ $\quad$ ① $\nabla^2 f(\cdot) \succeq 0 \Rightarrow \phi''(t) \geq 0 \Rightarrow f$ is convex

② f is convex $\Rightarrow \phi''(0) \geq 0 \Rightarrow d^T \nabla^2 f(x) d \geq 0 \Rightarrow \nabla^2 f(x) \succeq 0$

- $\nabla^2 f(x) \succ 0 \Rightarrow$ strictly convex $\Rightarrow$ minimizer is unique

· Convex optimization

1. $\begin{cases} \text{minimization form : minimize a convex function over a convex feasible region} \\ \text{maximization form : maximize a concave function over a convex feasible region} \end{cases}$

$* \begin{cases} xyz \text{ not concave} \longrightarrow \log(xyz) \text{ concave} \\ x^3 \leq 1 \text{ not concave} \longrightarrow x \leq 1 \text{ concave} \end{cases}$

2. KKT $\min\limits_{x \in \Omega} f(x)$. $\Omega = \{x : g_i(x) \leq 0, h_j(x) = 0\}$ where f and g_i are convex and h_j are affine, then KKT points is a global minimizer.

Let $(\bar{x}, \bar{\lambda}, \bar{\mu})$ be a KKT

$L(x) = \underbrace{f(x)}_{[\text{convex}]} + \sum_i \bar{\lambda}_i \underbrace{g_i(x)}_{[\text{convex}]} + \sum_j \bar{\mu}_j \underbrace{h_j(x)}_{[\text{affine}]}$ $\quad$ is convex

$L(\bar{x}, \bar{\lambda}, \bar{\mu}) = \min\limits_{x' \in \mathbb{R}^n} L(x', \bar{\lambda}, \bar{\mu}) \leq L(\underset{\text{feasible}}{\underset{\uparrow}{x}}, \bar{\lambda}, \bar{\mu}) \leq f(x)$

$\hookrightarrow \bar{\lambda}_i g_i(x) \leq 0$

· Unconstrainted problems

1. basic algorithm $\min\limits_{x \in \mathbb{R}^n} f(x)$ $\qquad x^{k+1} = x^k + \alpha_k d^k$ $\qquad \alpha_k = \operatorname{argmin}\limits_{\alpha \geq 0} f(x^k + \alpha d^k)$

2. bisection Method $[g(x) = 0]$ $\quad$ find an ϵ-approximation of x^*, we need at most $\left\lceil \log_2 \frac{x_r - x_l}{\epsilon} \right\rceil$ iterations.

① $x_m = \frac{x_l + x_r}{2}$ $\quad (x_l < 0 < x_r)$

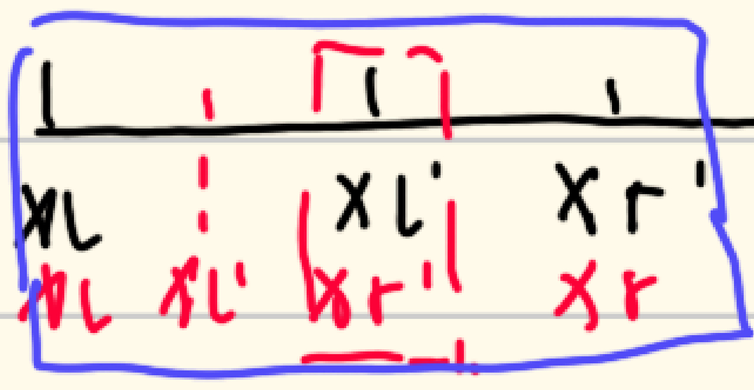
② if $|g(x_m)| < \epsilon$, stops. otherwise $g(x_m) > 0$ $x_r \Rightarrow x_m$ / $g(x_m) < 0$ $x_l \Rightarrow x_m$

③ $|x_r - x_l| < \epsilon$, Stop and output $\frac{x_l + x_r}{2}$, Otherwise go to step 1.

3. golden section Method $\quad$ [可解决单峰]

$[x_l, x_r]$ contains a unique minimizer, $0 < \phi < \frac{1}{2}$ $\qquad x_l' = x_l + \phi(x_r - x_l)$, $x_r' = x_l + (1-\phi)(x_r - x_l)$

s.t. $x_l < x_l' < x_r' < x_r$ $\qquad \begin{cases} f(x_l') < f(x_r') & x_r \leftarrow x_r' \\ f(x_l') \geq f(x_r') & x_l \leftarrow x_l' \end{cases}$ $\quad |x_r - x_l| < \epsilon$, Stop and output $\frac{x_l + x_r}{2}$, Otherwise go to step 1.



xr 复用一个点的值

$$x r'_{new} = x_L + (1-\phi) L' = x_L + (1-\phi)(1-\phi) L = x_L' = x_L + \phi L$$

$\Rightarrow (1-\phi)^2 = \phi \qquad \phi = \frac{3-\sqrt{5}}{2}$

4. Gradient Descent Method $\to d^k = -\nabla f(x^k)$

① $x^0 \in \mathbb{R}^n$ ② Pick a descent d^k at x^k, find a step size $\alpha_k > 0$ s.t. $f(x^k + \alpha_k d^k) < f(x^k)$

$x^{k+1} = x^k + \alpha_k d^k$ ③ stop (e.g. $\|\nabla f(x^{k+1})\| \le \epsilon$, $|f(x^{k+1}) - f(x^k)| \le \epsilon$, $|x^{k+1} - x^k| \le \epsilon \cdots$) and output x^{k+1}

· backtracking / Armijo Line Search [$f'(x^k) \neq 0 \quad f(x^{k+1}) < f(x^k)$]

$\sigma, \gamma \in (0,1)$ be given, choose α_k as the largest element in $\{1, \sigma, \sigma^2, \cdots\}$ s.t.

$f(x^k + \alpha_k d^k) \le f(x^k) + \gamma \alpha_k \cdot \nabla f(x^k)^T d^k \qquad [\alpha = 1 \Rightarrow \alpha_k = \alpha,\ \alpha = \sigma \alpha]$

$f(x^k + \alpha_k d^k) \approx f(x^k) + \alpha_k \underbrace{\nabla f(x^k)^T d^k}_{<0} < f(x^k) + \gamma \alpha \nabla f(x^k)^T d^k$

let $\phi_k(\alpha) = f(x^k + \alpha_k d^k) - f(x^k) \qquad \phi_k'(\alpha) = \nabla f(x^k + \alpha_k d^k) d^k$

$\phi_k'(0) = \nabla f(x^k)^T d^k < 0$

$\Rightarrow f(x^k + \alpha_k d^k) \le f(x^k) + \gamma \alpha_k \phi_k'(0)$

Gradient with line search:

$\phi(\alpha) = f(x^k - \alpha \nabla f(x^k)) - f(x^k) \le \gamma \alpha_k \phi_k'(0) \qquad (\text{let } d^k = -\nabla f(x^k))$

then set $x^{k+1} = x^k - \alpha_k \nabla f(x^k)$, if $\|\nabla f(x^{k+1})\| \le \epsilon$ stop

· convergence

(1) Accumulation Point

x is an accumulation point of $\{x^k\}_k$ if for every $\epsilon > 0$, there are infinitely many numbers k with $x^k \in B_\epsilon(x)$. 以x为中心，ε为半径的开球

[A bounded sequence always has at least one accumulation point.]

(2) Global Convergence

Let $f: \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable use exact line search / Armijo line seach to get $\{x^k\}_k$, then $\{f(x^k)_k\}$ is nonincreasing and every accumulation point of $\{x^k\}_k$ is a stationary point of f.

① If $\forall k\ \nabla f(x^k) \neq 0 \Rightarrow$ the accumulation points can only be local/global minima or saddle points

② f is a ploynomial function, $\{x^k\}_k$ is bounded $\Rightarrow$ $\{x^k\}_k$ converges to a stationary point x^*

③ x^* be an accumulation point of $\{x^k\}_k$ and $\nabla f(x^*) = 0$, $\nabla^2 f(x^*)$ is PD $\Rightarrow$ $\{x^k\}_k$ converges to the strict local minimizer x^*.

B1 Local convergence $\begin{cases} \text{Lipschitz continuous} \\ \text{Strict convex} \end{cases}$ $[C_L^{1,1}]$

- ∇f is Lipschitz continuous over $\mathbb{R}^n$: $\|\nabla f(x) - \nabla f(y)\| \le L\|x-y\|$ ($L>0$ is a constant)

$\|A\|_{op} = \sqrt{\lambda_{max}(A^T A)} = \max_{\|d\|=1} \|Ad\|$ operator/spectral form

- If f twice continuously differetiable, then ∇f is Lipschitz continuous with L

$\Longleftrightarrow$ $\|\nabla^2 f(x)\|_{op} \le L$ for any $x \in \mathbb{R}^n$

- Descent Lemma suppose ∇f is Lipschitz continuous with parameter L, then

$f(y) \le f(x) + \nabla f(x)^T (y-x) + \frac{L}{2}\|y-x\|^2 \quad \forall x, y \in \mathbb{R}^n$

* fix $x = x'$ $f(y) \le g(y) := f(x') + \nabla f(x')^T (y-x') + \frac{L}{2}\|y-x'\|^2 \quad \forall y \in \mathbb{R}^n$

for gradient descent: $x^{k+1} = x^k - \bar{\alpha}\nabla f(x^k) \Rightarrow x^{k+1} - x^k = -\bar{\alpha}\nabla f(x^k)$

$$f(x^{k+1}) < f(x^k) \quad \Leftarrow \quad f(x^{k+1}) \le f(x^k) + \nabla f(x^k)^T(-\bar{\alpha}\nabla f(x^k)) + \frac{L}{2}\|-\bar{\alpha}\nabla f(x^k)\|^2$$

$$= f(x^k) - \underbrace{\bar{\alpha}\left(1 - \frac{L\bar{\alpha}}{2}\right)\|\nabla f(x^k)\|^2}_{<0} \qquad 0 < \bar{\alpha} < \frac{2}{L}$$

- Linear Convergence

$\{x^k\}_k$ converges linearly with rate $\eta \in (0,1)$ to $x^* \in \mathbb{R}^n$ if there is $l \ge 0$ s.t.

$\|x^{k+1} - x^*\| \le \eta \cdot \|x^k - x^*\|, \ \forall k \ge l$

* Strong convexity $d^T \nabla^2 f(x) d \ge \mu\|d\|^2$ for all d, x ($\mu > 0$)

$\mu\|d\|^2 \le d^T \nabla^2 f(x) d \le L\|d\|^2$ [有二次上下界]

$\Rightarrow \{x^k\}_k$ converges linearly to x^* and $f(x^{k+1}) - f(x^*) \le \eta\,(f(x^k) - f(x^*))\ \forall k$,

$\eta = 1 - 2m\mu \in (0,1)$

$$m = \begin{cases} \bar{\alpha}\left(1 - \frac{L\bar{\alpha}}{2}\right) & \text{constant step size } \bar{\alpha} \in (0, \frac{2}{L}) \\ \frac{1}{2L} & \text{exact line search} \\ \gamma \min\left\{1, \frac{2\sigma(1-\gamma)}{L}\right\} & \text{Armijo line search} \end{cases} \qquad f(x^{k+1}) \le f(x^k) - m\|\nabla f(x^k)\|^2$$

- Exact Line Search Property $(d^{k+1})^T d^k = 0$

$\phi(\alpha) = f(x^k + \alpha d^k) \quad \phi'(\alpha_k) = \nabla f(x^k + \alpha_k d^k)^T d^k = \nabla f(x^{k+1})^T d^k = 0$

$d^{k+1} = -\nabla f(x^{k+1}) \Rightarrow (d^{k+1})^T d^k = 0$

• Newton's Method [local] Sensitive to the initial point

$\underbrace{\nabla^2 f(x^k)}_{A} \underbrace{d^k}_{x} = \underbrace{-\nabla f(x^k)}_{B}$ $O(n^3)$

1. consider the equation $g(x)=0 \Leftrightarrow \nabla f(x)=0$

$g(x)=g(x^k)+g'(x^k)(x-x_k)=0 \Rightarrow x^{k+1}=x^k-\frac{g(x^k)}{g'(x^k)}$ [Stop when $|g(x^k)|$ is sufficiently small].

$x^{k+1}=x^k-(\nabla^2 f(x^k))^{-1}\nabla f(x^k)$

Newton direction d^k: $\nabla^2 f(x^k)d^k=-\nabla f(x^k) \Rightarrow d^k=-(\nabla^2 f(x^k))^{-1}\nabla f(x^k)$

$\nabla f(x)^T d=-\nabla f(x^k)^T(\nabla^2 f(x^k))^{-1}\nabla f(x^k)<0$ only $\nabla^2 f(x^k)\succ 0$

2. convergence of Newton's Method

g is twice continuously differentiable and x^* is a root of g ($g'(x^*)\neq 0$)

Provided that $|x^0-x^*|$ is sufficiently small, then $x^{k+1}=x^k-\frac{g(x^k)}{g'(x^k)}$ will satisfy $|x^{k+1}-x^k|\le C|x^k-x^*|^2$ with $C=\sup_{x\in I}|g''(x)|\cdot\sup_{x\in I}\left(\frac{1}{|g'(x)|}\right)$ $I:\{x:|x-x^*|\le|x^0-x^*|\}$

3. (Damped) Newton's Method

$\nabla^2 f(x^k)d^k=-\nabla f(x^k) \Rightarrow$ Choose α_k via backtracking line search $x^{k+1}=x^k+\alpha_k d^k$

4. Local Quadratic Convergence (x^* is a local minimum)

$\|x^{k+1}-x^*\|\le\frac{L}{2\mu}\|x^k-x^*\|^2$ Assume for some $\epsilon>0$, $\nabla^2 f(x)\succeq\mu I$,

$\|\nabla^2 f(x)-\nabla^2 f(y)\|\le L\|x-y\|$ for all $x,y\in B_\epsilon(x^*)$

if $\|x^0-x^*\|\le\frac{\mu\min\{1,\epsilon\}}{L}$, then $\|x^k-x^*\|\le\frac{2\mu}{L}\left(\frac{1}{2}\right)^{2^k}$, $k=0,1,2\cdots$

• Project Gradient Method

1. Eucliden/Orthogonal Projection $\min_x \frac{1}{2}\|x-y\|^2$ s.t. $x\in\Omega$ $y^*=P_\Omega(y)$

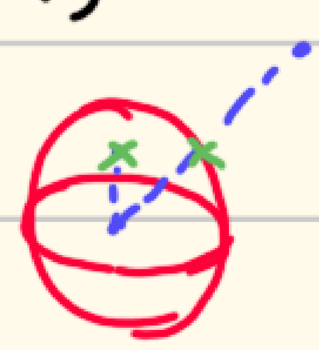
[e.g.1] $\Omega=\{x:Ax=b\}$

$P_\Omega(y)\Leftrightarrow\min_x\frac{1}{2}\|x-y\|^2$ s.t. $Ax=b$

KKT: $\mathcal{L}=\frac{1}{2}\|x-y\|^2+\mu(Ax-b)=0 \Rightarrow \frac{\partial\mathcal{L}}{\partial x}=x-y+A^T\mu=0 \Rightarrow Ax-Ay+AA^T\mu=0$

$\Rightarrow AA^T\mu=Ay-b$ $\mu=(AA^T)^{-1}(Ay-b) \Rightarrow P_\Omega(y)=x=y-A^T(AA^T)^{-1}(Ay-b)$

[e.g.2] $\Omega=\{x\in\mathbb{R}^n:\|x-m\|\le r\}$ $P_\Omega(y)=\begin{cases}y & \text{if } \|y-m\|\le r\\ m+\frac{r}{\|y-m\|}(y-m) & \text{if } \|y-m\|\ge r\end{cases}$



2. Projection Theorem

Let Ω be a nonempty, closed, and convex set. Then a point y^* is the projection of y onto Ω, i.e., $y^*=P_\Omega(y)$ iff. $(y^*-y)^T(x-y^*)\ge 0$ $\forall x\in\Omega$ ($P_\Omega:\mathbb{R}^n\to\mathbb{R}^n$ is Lipschitz continuous with $L=1$)

prove ① $\Rightarrow P_\Omega(y)=\arg\min_{x\in\Omega}\frac{1}{2}\|x-y\|^2$ $\nabla\left(\frac{1}{2}\|x-y\|^2\right)=x-y$

$\nabla f(y^*)^T(x-y^*)=(y^*-y)^T(x-y^*)\ge 0$ By FNOC

② $\Leftarrow$ choose $z = P_\Omega(x)$ $\begin{cases} (P_\Omega(y)-y)^T(z-P_\Omega(y)) \geq 0 & ① \\ (P_\Omega(y)-x)^T(z-P_\Omega(x)) \geq 0 & ② \end{cases}$

choose $z = P_\Omega(y)$

①+②. $(P_\Omega(x)-P_\Omega(y))^T(P_\Omega(y)-P_\Omega(x)+x-y) = -\|P_\Omega(x)-P_\Omega(y)\|^2 + (P_\Omega(x)-P_\Omega(y))^T(x-y) \geq 0$

$\Rightarrow \|P_\Omega(x)-P_\Omega(y)\|^2 \leq (P_\Omega(x)-P_\Omega(y))^T(x-y) \leq \|P_\Omega(x)-P_\Omega(y)\| \cdot \|x-y\|$

3. FONC for $\min_x f(x)$ s.t. $x \in \Omega$

$\forall \lambda > 0$ x^* is a stationary point iff. $x^* - P_\Omega(x^* - \lambda \nabla f(x^*)) = 0$

Proof: $x^* = P_\Omega(x^* - \lambda \nabla f(x^*)) \Leftrightarrow [x^* - (x^* - \lambda \nabla f(x^*))][x - x^*] \geq 0$

$\rightarrow \lambda \nabla f(x^*)^T(x - x^*) \geq 0$ $\lambda > 0 \Rightarrow \nabla f(x^*)^T(x - x^*) \geq 0$

4. Project Gradient Method

$x^{k+1} = P_\Omega(x^k - \lambda \nabla f(x^k))$

- $x^{k+1} = x^k + \alpha_k [P_\Omega(x^k - \lambda \nabla f(x^k)) - x^k] = (1-\alpha_k)x^k + \alpha_k P_\Omega(x^k - \lambda_k \nabla f(x^k))$

is feasible if $x^k \in \Omega$

- $d := P_\Omega(x - \lambda \nabla f(x)) - x$ $\nabla f(x)^T d \leq -\frac{1}{\lambda}\|d\|^2 < 0$

Proof $\nabla f(x)^T d = \nabla f(x)^T [P_\Omega(x - \lambda \nabla f(x)) - P_\Omega(x)] = -\frac{1}{\lambda}[(x - \lambda \nabla f(x)) - x]^T [P_\Omega(x - \lambda \nabla f(x)) - P_\Omega(x)]$

$\leq -\frac{1}{\lambda}\|P_\Omega(x - \lambda \nabla f(x)) - P_\Omega(x)\|^2 \leq -\frac{1}{\lambda}\|P_\Omega(x - \lambda \nabla f(x)) - x\|^2 = -\frac{1}{\lambda}\|d\|^2$

5. Project Gradient Method with Line Search

$d^k = P_\Omega(x^k - \lambda_k \nabla f(x^k)) - x^k$

$f(x^k + \alpha_k d^k) - f(x^k) \leq r \alpha_k \nabla f(x^k)^T d^k$ $\alpha_k = \{1, 6, 6^2 \cdots\}$

$x^{k+1} = x^k + \alpha d^k$