Convex Optimization

$$\text{Minimize} \quad f_0(\underline{x})$$

$$\text{ST} \quad f_i(\underline{x}) \le 0 \qquad i = 1, 2, \dots, m$$

$$h_j(\underline{x}) = 0 \qquad j = 1, 2, \dots p$$

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

KKT conditions :

1. $f_i(\underline{x}) \le 0 \qquad \forall i$
2. $h_j(x) = 0 \qquad \forall j$
3. $\lambda_i f_i(x) = 0 \qquad \forall i$
4. $\underline{\lambda} \succeq 0$
5. $\nabla_{\underline{x}} L(\underline{x}, \lambda, \underline{\nu}) = 0$

Entropy

$(p_1 \, - - \, p_k)$ $\qquad p \geq 0$

$$\sum_{i=1}^{k} p_i = 1$$

$(0.9, \; 0.05, \; 0.01 \, - - \quad)$

$$H(p) = \sum_{i=1}^{k} p_i \log \frac{1}{p_i}$$

$$\text{Maximize} \sum_{i=1}^{k} p_i \log \frac{1}{p_i} \equiv \text{Minimize} \sum_{i=1}^{k} p_i \log p_i$$

ST $\quad p \geq 0$

$$\sum_i p_i = 1$$

$$\sum_{j=1}^{k} \alpha_{lj} \, p_j \leq \beta_l \qquad l = 1, 2, \sim m$$

$$p_i = \Pr[X = i] \qquad i = 0, 1, 2, \dots k-1$$

$$\mathbb{E}X = \sum_{i=0}^{k-1} p_i \, i$$

$$\mathbb{E}\psi(X) = \sum_{i=0}^{k-1} \psi(i) p_i$$

$$\mathbb{E}X^2 = \sum_{i=0}^{k-1} i^2 p_i$$

$$H(p) = \sum_{i=0}^{k-1} p_i \log \frac{1}{p_i}$$ is concave

$$f(x) = -x \log x \qquad f''(x) = -\frac{1}{x} \leq 0$$

$$f'(x) = -(1 + \log x)$$

$$
\begin{aligned}
H(p) &= \sum_{i=0}^{k-1} p_i \log \frac{1}{p_i} = \mathbb{E} \log \frac{1}{p(X)} \\
&\le \log \mathbb{E}\left(\frac{1}{p_X}\right) \\
&= \log\left(\sum_{i=0}^{k-1} p_i \frac{1}{p_i}\right) \\
&= \log k
\end{aligned}
$$

$$Y = X + Z$$

Problem :

$$\boxed{\begin{array}{lll} \text{Minimize} & f_0(\underline{x}) & \\ \text{ST} & f_i(\underline{x}) \le 0 & 1 \le i \le m \\ & h_j(\underline{x}) = 0 & 1 \le j \le p \end{array}} \left.\vphantom{\begin{array}{l} a \\ b \end{array}}\right\} C$$

$$L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$$

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

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

$$f^* = \min f_0(\underline{x}) \quad \text{ST } \underline{x} \in C$$

For $\underline{\lambda} \succeq \underline{0}$ & any $\underline{\nu}$,

$$g(\underline{\lambda}, \underline{\nu}) \le f^*$$

$$\Rightarrow \sup_{\substack{\lambda \succeq 0 \\ \nu \in \mathbb{R}^p}} g(\underline{\lambda}, \underline{\nu}) = d^* \le f^*$$

Eg: Minimize $x^2 + x$ $\equiv$ Min $x^2 + x = f_0(x)$

ST $x \geq 1$ $\qquad$ $1 - x \leq 0$

$$L(\underline{x}, \lambda) = x^2 + x + \lambda(1 - x)$$

$$g(\lambda) = \min_{x \in \mathbb{R}} x^2 + x + \lambda(1 - x)$$

$$x = \frac{\lambda - 1}{2}$$

$$= \frac{(\lambda - 1)^2}{4} + \frac{(\lambda - 1)}{2} + \lambda - \frac{\lambda(\lambda - 1)}{2}$$

$$= \frac{-\lambda^2 + 6\lambda - 1}{4} \longrightarrow \text{concave}$$

$$\max_{\lambda \geq 0} g(\lambda) = 2$$

Minimize $f_0(\underline{x})$

St $f_i(\underline{x}) \leq 0$

$h_j(\underline{x}) = 0$

$$\tilde{f}(\underline{x}) = f_0(\underline{x}) + \sum_{i=1}^{m} I_-(f_i(\underline{x})) + \sum_{j=1}^{p} I_0(h_j(\underline{x}))$$

$$I_-(y) = \begin{cases} 0 & \text{if } y \leq 0 \\ \infty & y > 0 \end{cases}$$

$$I_0(y) = \begin{cases} 0 & \text{if } y = 0 \\ \infty & y \neq 0 \end{cases}$$

$$\tilde{f}(\underline{x}) = \begin{cases} f_0(\underline{x}) & \text{for } \underline{x} \in C \\ \infty & \text{for } \underline{x} \notin C \end{cases}$$

Original pros $\equiv \inf_{\underset{\sim}{x} \in \mathcal{D}} \tilde{f}(\underset{\sim}{x})$

$$\bar{f}(\underset{\sim}{x}) = L(\underset{\sim}{x}, \lambda, \nu) = f_0(\underset{\sim}{x}) + \sum_i \lambda_i f_i(x) + \sum_j \nu_j h_j(x)$$

$$\mathcal{A} = \{ \underbrace{(f_1(\underline{x}), f_2(\underline{x}) \cdots, f_m(\underline{x}), h_1(\underline{x}), \cdots h_p(\underline{x}), f_0(\underline{x}))}_{\in \mathbb{R}^{m+p+1}} : \underline{x} \in \mathcal{D} \}$$

$$= \{ (\underline{f}, \underline{h}, f_0) \}$$

$$f^* = \inf \{ t : (\underline{u}, \underline{v}, t) \in \mathcal{A}, \quad \underline{u} \leq 0 \rightarrow \text{inq. constraints}$$
$$\underline{v} = \underline{0} \rightarrow \text{eq. const} \}$$

$$\text{Min } f_0(x)$$
$$\text{ST } f_1(x) \leq 0$$

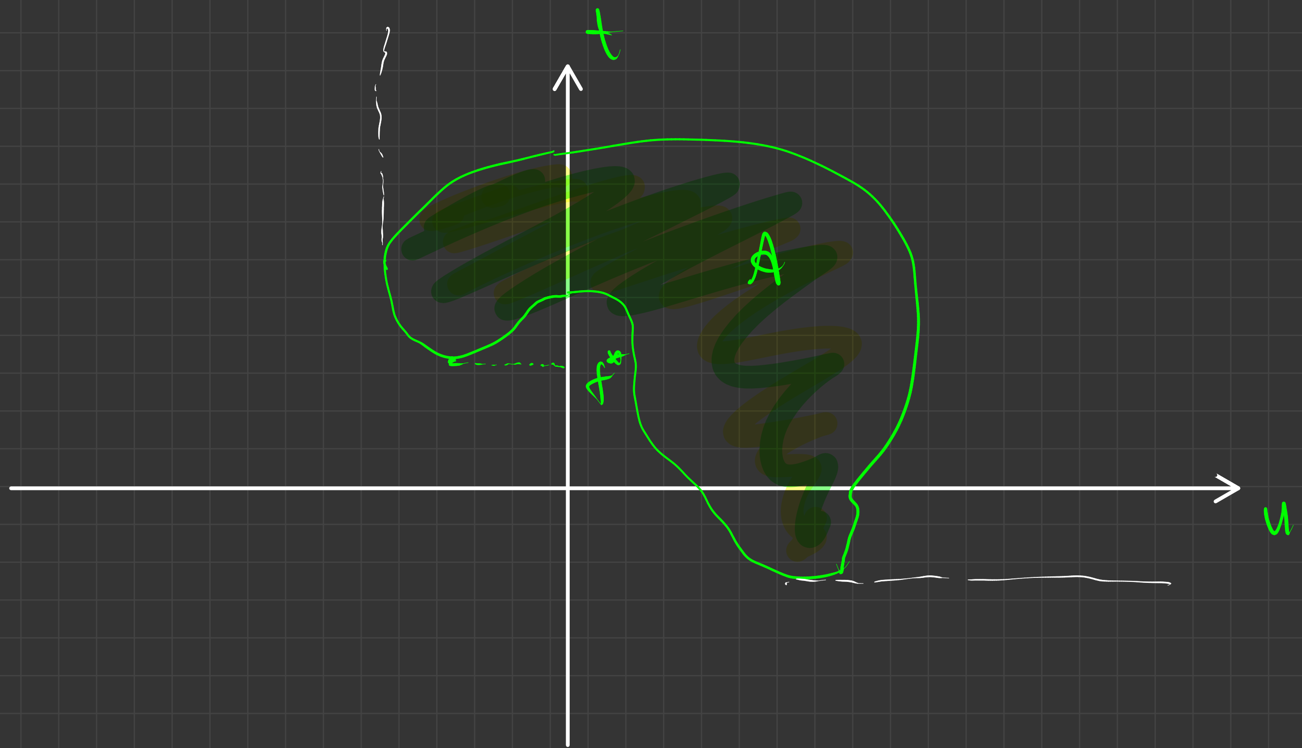
$$\mathcal{A} = \{ (f_1(x), f_0(x)) : x \in \mathbb{R} \}$$
$$(u, t)$$

In the example,

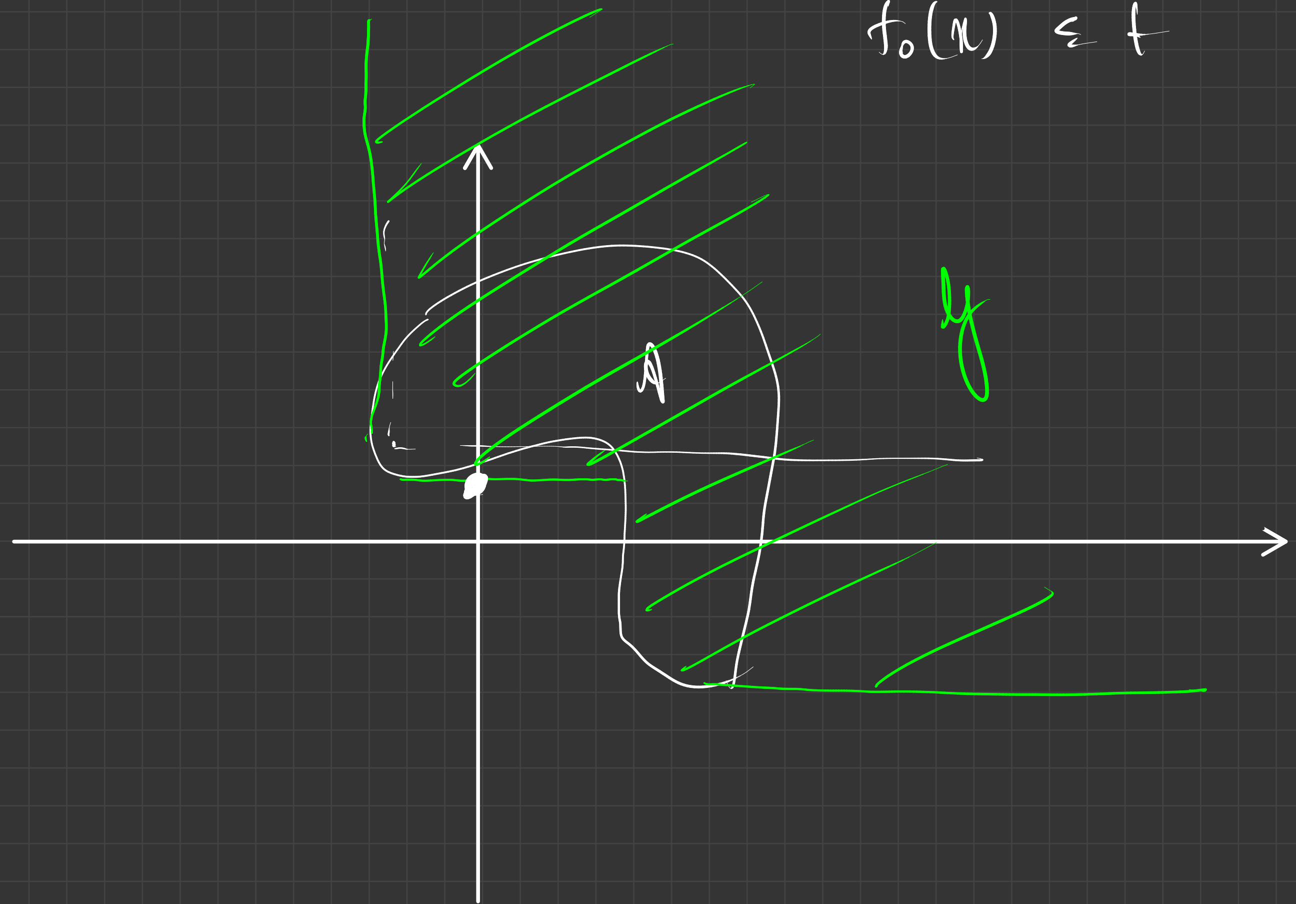
$$\min \; x^2+x$$
$$\text{st } 1-x \le 0$$

$$\mathcal{A} = \{ (\underbrace{1-x}_{u}, \; x^2+x) : x \in \mathbb{R} \}$$

$$= \{ (u, \; (1-u)^2+(1-u)) : u \in \mathbb{R} \}$$



$$\mathcal{G} = \left\{ (u, v, t) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} : \quad \exists\ x \in \mathcal{D} \quad \begin{array}{l} f_i(x) \le u_i \\ h_j(x) = v_j \\ f_0(x) \le t \end{array} \right\}$$



$$\mathcal{A} = \left\{ (\underline{u}, \underline{v}, t) \ : \ \exists\, \underline{x} \ \text{st} \quad \begin{aligned} f_i(x) &= u_i \\ h_j(x) &= v_j \\ f_0(x) &= t \end{aligned} \right\}$$

$$\mathcal{G} = \left\{ (\underline{u}, \underline{v}, t) \ : \ \exists\, \underline{x} \ \text{st} \quad \begin{aligned} f_i(x) &\le u_i \\ h_j(x) &= v_j \\ f_0(x) &\le t \end{aligned} \right\}$$

$$L(\underline{x}, \underline{\lambda}, \underline{\nu}) = \begin{bmatrix} \underline{\lambda} \\ \underline{\nu} \\ 1 \end{bmatrix}^T \begin{bmatrix} \underline{u} \\ \underline{v} \\ t \end{bmatrix}$$

$$g(\lambda, \nu) = \inf \left\{ (\underline{\lambda}, \underline{\nu}, 1) \begin{bmatrix} \underline{u} \\ \underline{v} \\ t \end{bmatrix} \ : \ \begin{bmatrix} u \\ v \\ t \end{bmatrix} \in \mathcal{G} \right\}$$

For a given $\underline{\lambda} \succeq 0$, $\underline{\nu} \in \mathbb{R}$,

$$[\underline{\lambda}, \underline{\nu}, 1] \begin{bmatrix} \underline{u} \\ \underline{v} \\ t \end{bmatrix} \geq \underbrace{g(\lambda, \nu)} \quad \forall \begin{bmatrix} u \\ v \\ t \end{bmatrix} \in \mathcal{G}$$

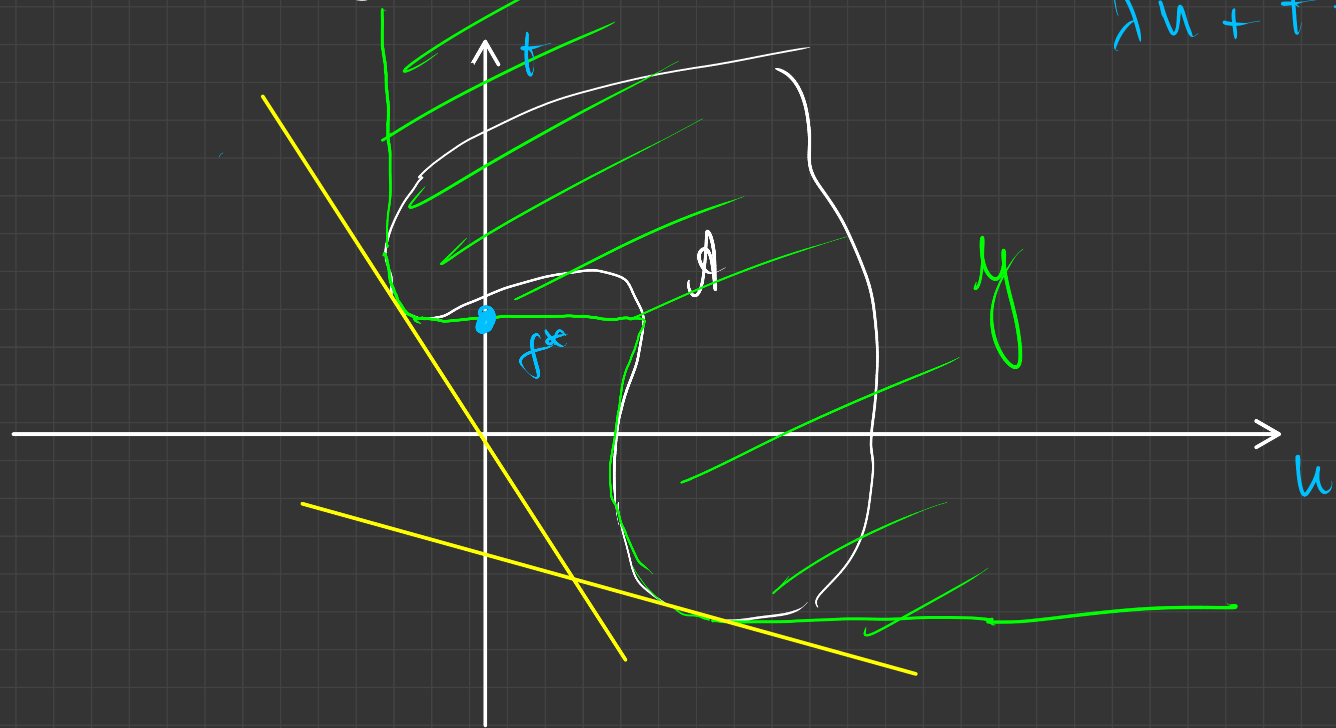
Since this is the minimum value of the inner product

$$\therefore \left\{ \underline{\Psi} : [\underline{\lambda}\ \underline{\nu}\ 1]\, \underline{\Psi} = g(\underline{\lambda}, \nu) \right\}$$ is a supporting hyperplane for $\mathcal{G}$

If we have $\min f_0(x)$

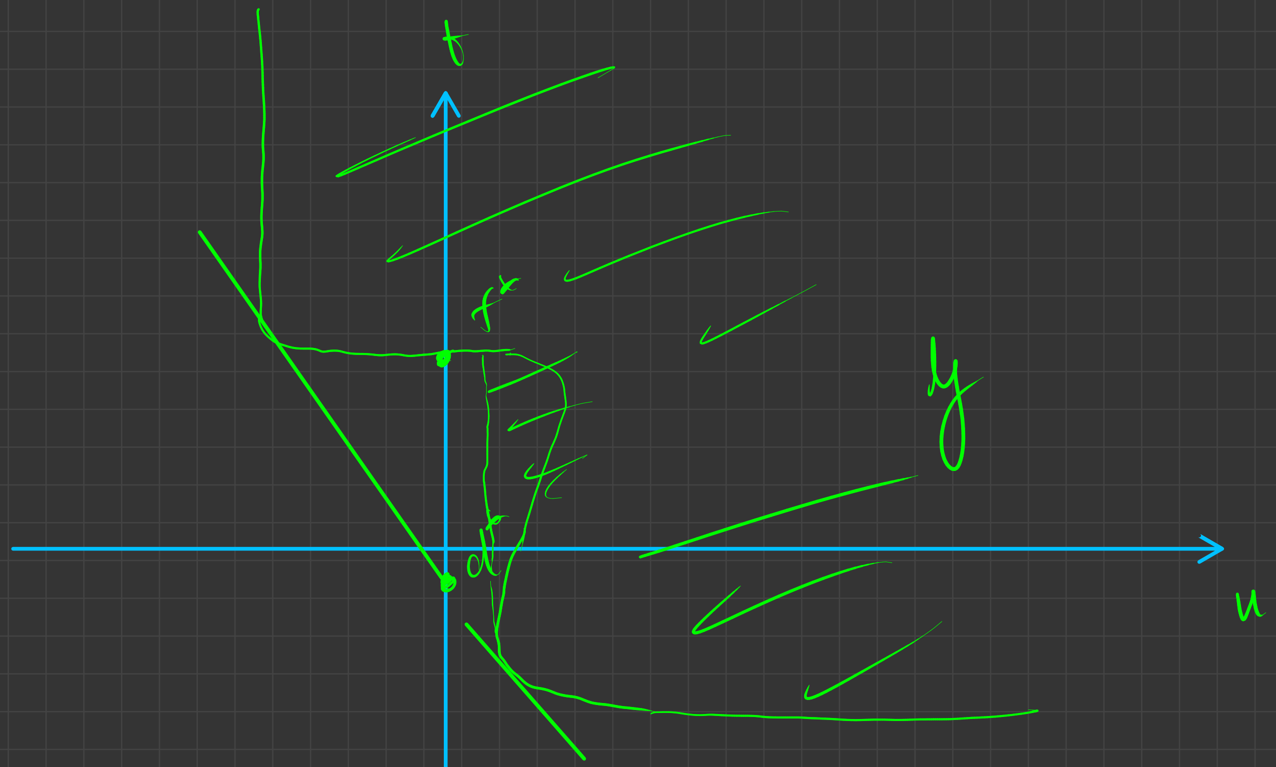
s.t. $f_1(x) \le 0$

$$\left\{ [\lambda, 1] \begin{bmatrix} u \\ t \end{bmatrix} = g(\lambda) \right\} \to \{ (u,t) : \lambda u + t = g(\lambda)$$



$g(\lambda)$ = Intercept of supporting hyperplane with t-axis

$$d^* = \sup_{\lambda \ge 0} g(\lambda)$$



$d^* < f^*$

$$f^* = \min \{ t : (\underline{0}, \underline{0}, t) \in \mathcal{G} \}$$

A sufficient condition for strong duality (Slater's condition)

Constraint qualification

If prob is convex,

① $\exists$ some $x \in \text{relint}(\mathcal{D})$

s.t. $f_i(x) < 0 \quad i = 1, 2, \dots m$

then the prob satisfies strong duality.

Eg:

Minimize e^{-x}

s.t. $\dfrac{x^2}{y} \leq 0$ $\qquad \mathcal{D} = \{(x,y) : y > 0\}$

$f^* = 1$

$L(x,y,\lambda) = e^{-x} + \lambda \dfrac{x^2}{y}$

$$g(\lambda) = \inf_{(x,y) \in \mathcal{D}} \left\{ e^{-x} + \lambda \frac{x^2}{y} \right\} = 0$$

$$y = x^4$$

$$d^* = 0$$

$$\begin{aligned} \text{Min} \quad & f_0(\underline{x}) \\ \text{st} \quad & f_i(\underline{x}) \leq 0 \qquad 1 \leq i \leq m \\ & h_j(\underline{x}) = 0 \qquad 1 \leq j \leq p \end{aligned}$$

$$L(\underline{x}, \underline{\lambda}, \underline{\nu}) = f_0(\underline{x}) + \sum_i \lambda_i f_i(\underline{x}) + \sum_j \nu_j h_j(\underline{x})$$

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

If strong duality holds,

$$f^* = f_0(x^*) = g(\lambda^*, \nu^*) \qquad \text{for } x^* \in C \quad \underline{\lambda}^* \geq \underline{0}$$

$$= \inf_{x \in \mathcal{D}} \left\{ f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{j=1}^{p} \nu^* h_j(x) \right\}$$

$$\leq f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \underbrace{f_i(x^*)}_{\leq 0} + \sum_{j=1}^{p} \nu^* \underbrace{h_j(x^*)}_{\to 0}$$

$$\leq f_0(x^*)$$

If strong duality holds, then

KKT conditions

① $\lambda_i^* f_i(x^*) = 0 \qquad 1 \le i \le m$ → Necessary

(Complementary slackness conditions)

② $f_0(x^*) = \inf_{x \in \mathcal{D}} f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_j \nu_j^* h_j(x)$

$$\Rightarrow \nabla f_0(x^*) + \sum_i \lambda_i^* \nabla f_i(x^*) + \sum_j \nu_j^* \nabla h_j(x^*) = \underline{0}$$

(Necessary)

③ $\lambda_i^* \geq 0 \quad \forall i$,

④ $f_i(x^*) \leq 0 \quad \forall i$ ⑤ $h_j(x^*) = 0 \quad \forall j$

$$g(\lambda^*, \nu^*) = f^*$$

Consider an algorithm that produces iterates $(\underline{x}^{(k)}, \underline{\lambda}^{(k)}, \underline{\nu}^{(k)})$

$$g(\lambda^{(k)}, \nu^{(k)}) \le f^*$$

$$f_0(x^{(k)}) + g(\lambda^{(k)}, \nu^{(k)}) \le f_0(x^{(k)}) + f^*$$

$$\underbrace{f_0(x^{(k)}) - f^*}_{\text{Optimality gap}} \le \underbrace{f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)})}_{\text{Duality gap}}$$

Stopping rule : $f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)}) \le \epsilon$

Consider the generalization:

$$\text{Minimize} \quad f_0(\underline{x})$$
$$\text{ST} \quad h_j(\underline{x}) = 0 \qquad 1 \leq j \leq p$$
$$f_i(\underline{x}) \preccurlyeq_{K_i} 0 \qquad 1 \leq i \leq m$$

generalized inequality wrt K_i

Lagrangian:

$$L(\underline{x}, \underline{\lambda}_1, \underline{\lambda}_2 \cdots \underline{\lambda}_m, \underline{\nu}) = f_0(\underline{x}) + \sum_i \langle \underline{\lambda}_i, f_i(\underline{x}) \rangle + \sum_j \nu_j h_j(\underline{x})$$

$$g(\underline{\lambda}_1 \cdots \underline{\lambda}_m, \underline{\nu}) = \inf_{\underline{x} \in \mathcal{D}} L(\quad)$$

Dual optimization problem

maximize $g(\lambda_1, \dots \lambda_m, \nu)$

st $\lambda_i \succeq_{K_i^*} 0$

generalized inequality
wrt dual cone K^*

Examples SDP (Semidefinite program)

$$\text{Minimize} \quad \underline{c}^T \underline{x}$$

$$\text{ST} \quad G + \sum_{i=1}^{n} x_i F_i \quad \text{is PSD}$$

Symmetric

$$\equiv \quad \text{Min} \quad \underline{c}^T \underline{x}$$

$$\text{ST} \quad G + \sum_{i=1}^{n} x_i F_i \succeq_K 0$$

$$K = S_+^{n} \quad (\text{set of } PSD^{\,n\times n} \text{ matrices})$$

Lagrange multipliers (n×n, symmetric)

$$L(\underline{x}, Z) = f_0(\underline{x}) - \operatorname{tr}\Big(Z\big(G + \sum_i x_i F_i\big)\Big)$$

$$= f_0(\underline{x}) - \operatorname{tr}\Big(ZG + \sum_i x_i Z F_i\Big)$$

$$= \sum_{i=1}^{n} c_i x_i - \operatorname{tr}(ZG) - \sum_{i=1}^{n} x_i \operatorname{Tr}(ZF_i)$$

$$= -\operatorname{Tr}(ZG) + \sum_{i=1}^{n} x_i\big(c_i - \operatorname{tr}(ZF_i)\big)$$

$$g(Z) = \inf_{x \in \mathbb{R}^n} L(x, Z)$$

$$= \begin{cases} -\operatorname{tr}(ZG) & \text{if } \operatorname{tr}(ZF_i) = c_i \ \forall i \\ -\infty & \text{else} \end{cases}$$

$$\text{Maximize} \quad -\operatorname{tr}(ZG)$$

$$\text{s.t.} \quad Z \succeq_{K^*} 0 \equiv Z \text{ is PSD}$$

$$\operatorname{tr}(ZF_i) = c_i \qquad i = 1, 2 \ldots n$$

Conic program:

$$\text{Minimize} \quad \underline{c}^T \underline{x}$$
$$\text{ST} \quad A\underline{x} = \underline{b}$$
$$\underline{x} \preceq_K 0$$

K is any proper cone

$$L(\underline{x}, \underline{\lambda}, \underline{\nu}) = \underline{c}^T \underline{x} + \underline{\lambda}^T \underline{x} + \underline{\nu}^T (A\underline{x} - \underline{b})$$
$$= -\underline{\nu}^T \underline{b} + (\underline{c} + \underline{\lambda} + A^T \underline{\nu})^T \underline{x}$$

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

Dual program:

$$\left.\begin{array}{l} \text{Maximize} \quad -\nu^T \underline{b} \\ \text{ST} \quad \underline{\lambda} \succeq_{K^*} 0 \\ \underline{c} + \underline{\lambda} + A^T \underline{\nu} = 0 \end{array}\right\} = \begin{array}{l} \max \quad -\nu^T b \\ \text{ST} \\ -A^T \nu - c \succeq_{K^*} 0 \end{array}$$

Classification (Linear)

R^n

$\mathcal{X} = \{\underline{x}_1, \dots, \underline{x}_N\}$ $\qquad$ $\mathcal{Y} = \{\underline{y}_1, \dots, \underline{y}_M\}$

$\underline{x}_i \in R^n$ $\qquad$ $\underline{y}_i \in R^n$

Linear classification / discrimination

Given $\mathcal{X}$ & $\mathcal{Y}$, can we construct a hyper plane

$(\underline{a}, b)$ ST $\quad \underline{a}^T \underline{x}_i > b \quad \forall x_i \in \mathcal{X}$

$\underline{a}^T \underline{y}_i < b \quad \forall y_i \in \mathcal{Y}$

If yes, linearly separable.

$A = \text{Conv}(\mathcal{X})$ $\quad\quad B = \text{Conv}(\mathcal{Y})$

$\exists\ \underline{x} \in A$

$$\underline{x} = \sum_i \alpha_i y_i$$

$$\underline{a}^T y > b \quad \forall\ y \in \mathcal{Y}$$

<u>Prove</u>: If a hyperplane separates $\mathcal{X}, \mathcal{Y}$ then it also separates A & B.

Linear discrimination as an optimization problem

Variables : $(\underline{a}, b) \in \mathbb{R}^n \times \mathbb{R}$

(1) $\underline{a}^T \underline{x}_i < b \quad \forall\ x_i \in \mathcal{X}$

$a^T y_i > b \quad \forall\ y_i \in \mathcal{Y}$

$$\left.\begin{array}{ll} \text{Min} & 1 \\ \text{ST} & a^T \underline{x}_i < b \ ;\ x_i \in \mathcal{X} \\ & a^T y_i > b \ ;\ y_i \in \mathcal{Y} \end{array}\right\} \text{Satisfiability}$$

$\{\underline{x} : \underline{a}^T \underline{x} = b\}$

$$\min_{(\underline{a}, b)} \quad 1$$

ST $\quad \underline{a}^T \underline{x}_i \geq b+1 \qquad x_i \in \mathcal{X}$

$\quad \underline{a}^T \underline{y}_i \leq b-1 \qquad y_i \in \mathcal{Y}$

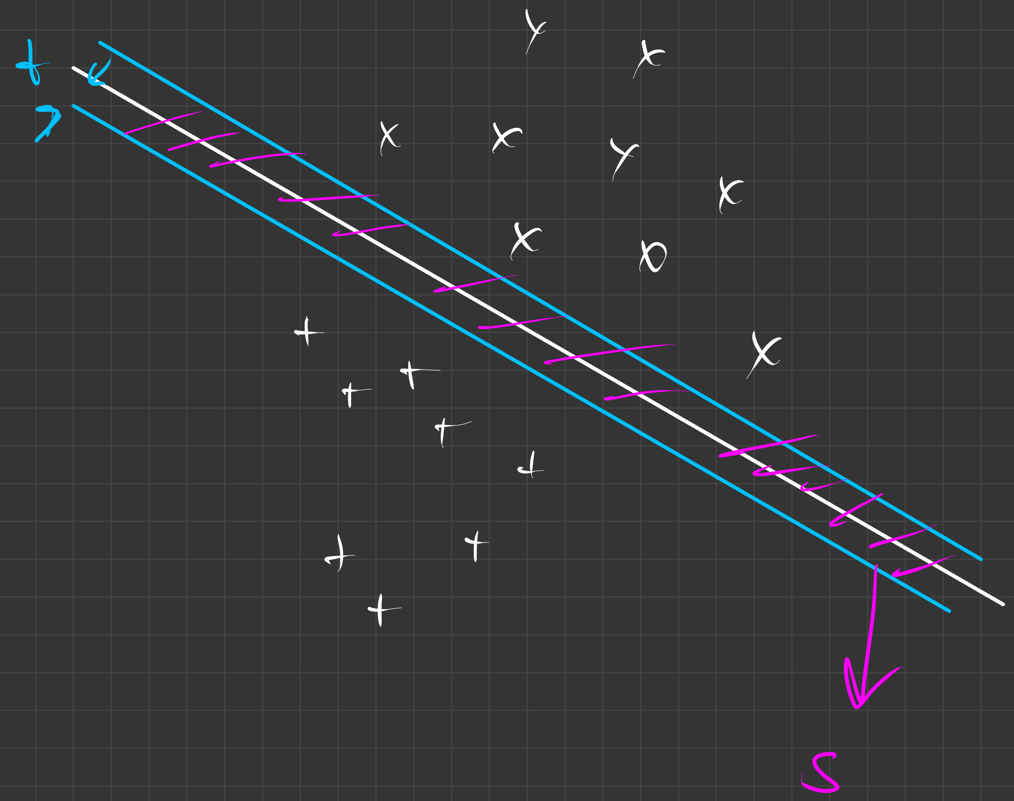
Another approach:

$$a^T x_i \leq b - t \quad \forall\ x_i \in X$$

$$a^T y_i \geq b + t \quad \forall\ y_i \in Y$$

$$t \geq 0$$

$$\|a\|_2 \leq 1$$



Min $\ -t$

ST

$$a^T x_i \leq b - t$$

$$a^T y_i \geq b + t$$

$$t \geq 0$$

$$\|a\|_2 \leq 1$$

Consider $\{ \underline{x} : -t \leq \underline{a}^T \underline{x} - b \leq t \} = S$

$\| \underline{a} \| = 1$

<u>Exercise</u>: Find min dist b/w S^c & the hyperplane

Problem:

$$\text{Min} \quad -t$$
ST

$$\underline{a}^T \underline{x}_i \leq b - t \qquad \underline{a}^T y_i \geq b + t$$

$$t \geq 0 \qquad \|a\|_2 \leq 1$$

$$L(\underline{\lambda}, \underline{\tilde{\lambda}}, \tau, \alpha, \underline{a}, b, t) = -t + \sum_{i=1}^{N} \lambda_i (\underline{a}^T \underline{x}_i - b + t)$$

$$+ \sum_{j=1}^{m} \tilde{\lambda}_j \left(-\underline{a}^T y_j + b + t \right)$$

$$+ \tau(-t) + \alpha(\|\underline{a}\| - 1)$$

$$= t\left(-1 + \sum_i \lambda_i + \sum \tilde{\lambda}_j - \tau\right) + b\left(\sum_j \tilde{\lambda}_j - \sum_i \lambda_i\right)$$

$$+ a^T \left(\sum_i \lambda_i \underline{x}_i - \sum_j \tilde{\lambda}_j y_j \right) + \alpha \|a\| - \alpha$$

$$\inf_{b,t} L(\quad) = \begin{cases} a^T\left(\sum_i \lambda_i x_i - \sum_j \tilde{\lambda}_j y_j\right) + \alpha\|a\| - \alpha & \text{if } \sum_j \tilde{\lambda}_j = \sum_i \lambda_i \ \& \\ & \sum_i \lambda_i + \sum_j \lambda_j = 1 + 2 \\ -\infty, & \text{otherwise} \end{cases}$$

$$f(\underline{a}) = \underline{a}^T\underline{u} + \alpha\|a\|_2 \quad , \quad \alpha \geq 0$$

What is $\inf_{\underline{a}} f(\underline{a})$?

$$f(\underline{a}) = a^T\underline{u} + \alpha\sqrt{\underline{a}^T\underline{a}}$$

$$\underline{a} = \beta\,\underline{u} \quad , \quad \beta < 0$$

$$\tilde{f}(\beta) = \beta u^T u + \alpha \sqrt{\beta^2 \underline{u}^T \underline{u}}$$

$$= \beta \sqrt{u^T u}\left(\sqrt{\underline{u}^T \underline{u}} - \alpha\right)$$

$$\inf_{\beta < 0} \tilde{f}(\beta) = \begin{cases} -\infty & \text{if } \|\underline{u}\| > \alpha \\ 0 & \text{else} \end{cases}$$

Suppose $\|u\| \leq \alpha$

$$f(\underline{a}) = a^T \underline{u} + \alpha \|a\|_2$$

$$\geq -\|\underline{u}\| \|a\| + \alpha \|a\|$$

$$= \|a\| \left(\alpha - \|\underline{u}\|\right)$$

$$\geq 0$$

Cauchy-Schwartz

$$|\underline{a}^T \underline{u}| \leq \|\underline{u}\| \|\underline{a}\|$$

$$-|\underline{a}^T \underline{u}| \geq -\|u\| \|a\|$$

$$\underline{a}^T \underline{u} \geq -\|\underline{u}\| \|\underline{a}\|$$

$$\inf_{\underset{\sim}{a}} \left(\underset{\sim}{a}^T \underset{\sim}{u} + \alpha \|\underset{\sim}{a}\| \right) = \begin{cases} -\infty & \text{if } \|u\| > \alpha \\ 0 & \text{if } \|u\| \leq \alpha \end{cases}$$

$$g(\underset{\sim}{\lambda}, \tilde{\lambda}, \alpha, \tau) = \begin{cases} -\alpha & \text{if } \left\| \sum_i \lambda_i x_i - \sum_j \tilde{\lambda}_j y_i \right\| \leq \alpha \\ & \sum_i \lambda_i = \sum_j \lambda_j, \ \sum_i \lambda_i + \sum_j \lambda_j = 1 + \tau \\ -\infty & \text{else} \end{cases}$$

Dual:

$$\max \quad -\alpha$$

s.t.

$$\Big\| \sum_i \lambda_i x_i - \sum_j \tilde{\lambda}_j y_j \Big\| \le \alpha$$

$\longrightarrow$ min dist b/w Conv(X) & Conv(Y)

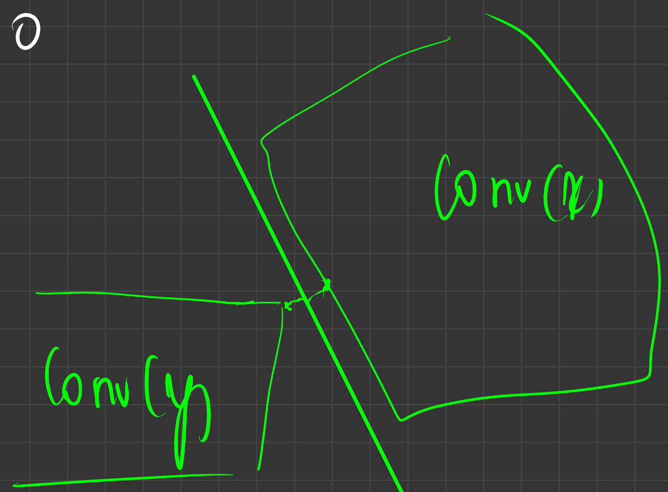
$$\sum_i \lambda_i = \sum_j \tilde{\lambda}_j, \quad \sum_i \lambda_i + \sum_j \tilde{\lambda}_j = 1 + \tau$$

$$u_i = \lambda_i \left(\frac{1+\tau}{2}\right)$$

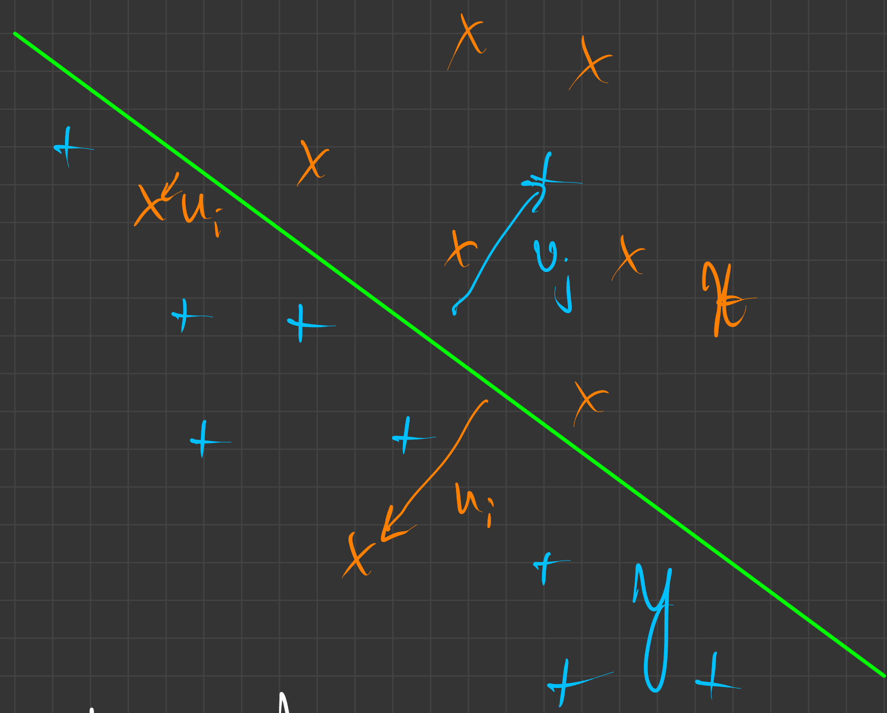
$$v_j = \tilde{\lambda}_j \left(\frac{1+\tau}{2}\right)$$

$$\lambda \ge 0, \quad \tilde{\lambda} \ge 0, \quad \tau \ge 0,$$

$$\alpha \ge 0$$



What if the two sets are NOT linearly separable?



Natural problem :

Minimize "misclassifications"

$\downarrow$

combinatorial optimization problem

$\forall\, \underline{x}_i \in \mathcal{X} \qquad \underline{a}^T \underline{x}_i \geq b + 1$

$\forall\, y_j \in \mathcal{Y}, \qquad \underline{a}^T y_j \leq b - 1$

} Can be satisfied only if l.s.

$\forall\, \underline{x}_i \in \mathcal{X} \qquad \underline{a}^T \underline{x}_i \geq b + 1 - u_i$

$\forall\, y_j \in \mathcal{Y}, \qquad \underline{a}^T y_j \leq b - 1 + v_j$

Support Vector Classifier

$$\begin{cases} \text{Minimize} \quad \displaystyle\sum_{i=1}^{M} u_i + \sum_{j=1}^{N} v_j \\ \text{ST} \\ \underline{a}^T \underline{x}_i \geq b + 1 - u_i \quad \forall x_i \in \mathcal{X} \\ \underline{a}^T \underline{y}_j \leq b - 1 + v_j \quad \forall y_j \in \mathcal{Y} \\ \underline{u} \geq 0 \\ \underline{v} \geq 0 \end{cases}$$

Logistic Modeling

$$f(z) = \frac{e^z}{1+e^z} = \frac{1}{1+e^{-z}}$$

Given any point $\underline{u} \in \mathbb{R}^n$,

$$\Pr[\underline{u} \in \text{Class } 1] = \frac{e^{\underline{a}^T\underline{u} - b}}{1 + e^{\underline{a}^T\underline{u} - b}}$$

Given $\{\underline{u}_1, \ldots \underline{u}_k\}$, randomly label 0 & 1 acc the above probability

You only observe $\{\underline{u}_1 - \underline{u}_k\}$ & labels

$\underline{u}_1, \underline{u}_2 \ldots u_M \to$ label 1
$u_{M+1} \ldots u_{M+N} \to$ label 0
$\Big\}$ Configuration C

$$\Pr\left[C \mid \underset{\underline{a},\, b}{\underline{u}_1 \ldots u_{M+N}} \right] = \prod_{i=1}^{M} \Pr[u_i \text{ is labeled } 1] \prod_{j=M+1}^{M+N} \Pr[u_j \text{ is labeled } 0]$$

$$= \prod_{i=1}^{M} \frac{1}{1+e^{-(\underline{a}^T \underline{u}_i - b)}} \prod_{j=M+1}^{M+N} \frac{e^{-(\underline{a}^T u_j - b)}}{1+e^{-\underline{a}^T u_j - b}}$$

Choose $(\underline{a}^*, b^*) = \text{argmax} \log \Pr\left[C \mid \underline{u}_1 \ldots u_{M+N}, \underline{a}, b \right]$

Nonlinear classifier : $f: \mathbb{R}^n \to \mathbb{R}$

st $f(\underline{x}_i) > 0 \quad \forall \; \underline{x}_i \in \mathcal{X}$

$f(\underline{y}_j) < 0 \quad \forall \; \underline{y}_j \in \mathcal{Y}$

Numerically solving unconstrained minimization problems

① Gradient descent

$$\underline{x}_t = \underline{x}_{t-1} + \delta_t \, \underline{u}_t \rightarrow \text{General descent}$$

δ_t: step size

$\underline{u}_t$: descent direction

Choose $\underline{u}_t = -\nabla f(\underline{x}_{t-1})$

How do we choose δ_t?

Fixed beforehand
- Constant
- $\delta_t \rightarrow 0$ as $t \rightarrow \infty$
  $\delta_t = 1/(t+1)$

Line search :

① Exact line search : Choose σ_t s.t.

$$f(x_{t+1} - \sigma_t \nabla f(x_{t+1}))$$

is min

② Backtracking line search

Fix $\underline{x}_t$, direction $\underline{u}$, α, β.

$\alpha \in (0,1)$, $\beta \in (0, \frac{1}{2})$

Fix $\sigma = 1$

While $f(\underline{x}_t + \sigma \underline{u}) > f(\underline{x}_t) + \beta \sigma \nabla f(x_t)^T \underline{u}$

$\sigma = \alpha \sigma$

For any convex fn,

$$f(y) \geq f(x) + \nabla f(x)^T (y-x)$$

$$f(x_t + \delta u) \geq f(x_t) + \nabla f(x_t)^T (\delta u)$$

Want to choose δ st

$$f(x_t + \delta u) \leq f(x_t) + \beta \nabla f(x_t)^T (\delta u)$$

Least squares problem

$$\min_{\underset{\sim}{x} \in \mathbb{R}^n} \|A\underset{\sim}{x} - \underset{\sim}{b}\|_2^2$$

$$f_0(\underset{\sim}{x}) = \|A\underset{\sim}{x} - \underset{\sim}{b}\|_2^2 = (A\underset{\sim}{x} - \underset{\sim}{b})^T (A\underset{\sim}{x} - \underset{\sim}{b})$$

$$\nabla f_0(\underset{\sim}{x}) = 2A^TA\underset{\sim}{x} - 2A^T\underset{\sim}{b}$$

$$\nabla^2 f_0(\underset{\sim}{x}) = 2A^TA$$

$$\underset{\sim}{x}^* = (A^TA)^{-1} A^T \underset{\sim}{b}$$

Method of steepest descent

Descent direction :

$$\underline{u}_{NSD} = \underset{\|\underline{u}\| = 1}{\operatorname{argmin}} \left(\underline{u}^T \nabla f(\underline{x}) \right)$$

NSD : Normalized Steepest Descent

Any norm

Different choices of norm :

① Euclidean norm :

$$\underline{u}_{NSD} = \frac{-\nabla f(\underline{x})}{\|\nabla f(x)\|_2}$$

② L_1 norm :

$$\left.\begin{aligned} i^{*} &= \underset{i}{\operatorname{argmax}} \left(|\nabla f(x)| \right)_i \\ \underline{u}_{NSD} &= -e_i \ \operatorname{sgn}\left((\nabla f(x))_i \right) \end{aligned}\right\} \text{Coordinate descent}$$

③ $\|\underline{x}\|_P = \|P^{1/2}\underline{x}\|_2$ for some P.D matrix P.

$$\|x\|_P^2 = x^T P x$$

$$\underline{u}_{NSD} = \frac{-P^{-1/2}\underline{x}}{\|P^{-1/2}\underline{x}\|_2} \longrightarrow \text{Coordinate transformation + GD}$$

Second-order method: Newton's method

Basic principle: 2nd order Taylor series approx at $\underline{x}_t$

$$\tilde{f}_{\underline{x}_t}(\underline{v}) = f(\underline{x}_t) + \underline{v}^T \nabla f(\underline{x}_t) + \frac{1}{2}\underline{v}^T \nabla^2 f(\underline{x}_t)\,\underline{v}$$

$$\shortparallel$$

$$\approx f(\underline{x}_t + \underline{v})$$

$\downarrow$

convex if f is convex

$$\min_{\underline{v}} \tilde{f}_{\underline{x}_t}(\underline{v})$$

$$\underline{v}^* = -(\nabla^2 f(\underline{x}_t))^{-1} (\nabla f(\underline{x}_t))$$

$\downarrow$

$v^*(x_t)$

Newton update :

$$x_t = x_{t-1} + \delta_t \, v^*(x_{t-1})$$

$\delta_t = 1 \rightarrow$ Pure Newton

$\delta_t \neq 1 \rightarrow$ Damped/guarded Newton method.

Newton's method for equality-constrained minimization

$$\text{Minimize } f(\underline{x})$$
$$\text{ST } A\underline{x} = \underline{b}$$

Assume that we have x_t s.t

$$A\underline{x}_t = \underline{b}$$

Want to choose $\underline{x}_{t+1}$ st ① $A\underline{x}_{t+1} = \underline{b}$

② $f(\underline{x}_{t+1}) \leq f(\underline{x}_t)$

$$\tilde{f}_{\underline{x}_t}(\underline{v}) = f(\underline{x}_t) + \underline{v}^T \nabla f(\underline{x}_t) + \frac{1}{2} \underline{v}^T \nabla^2 f(\underline{x}_t) \underline{v}$$

$$\text{Minimize } \tilde{f}_{\underline{x}_t}(\underline{v})$$
$$\underline{v} : A(\underline{x}_t + \underline{v}) = \underline{b}$$

OR $A\underline{v} = 0$

Solve KKT conditions :

$$L(\underline{v}, \underline{w}) = \tilde{f}_{\underline{x}_t}(\underline{v}) + \underline{w}^T (A\underline{v})$$

KKT conditions : (i) $A\underline{v} = 0$

(ii) $\nabla_{\underline{v}} L(\underline{v}, \underline{w}) = 0$

$\downarrow$

$$\nabla f(\underline{x}_t) + \nabla^2 f(x_t)\, \underline{v} + A^T \underline{w} = 0$$

$$\nabla^2 f(x_t)\underline{v} + A^T \underline{w} = -\nabla f(x_t)$$

$$\begin{bmatrix} \nabla^2 f(x_t) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \underline{v} \\ \underline{w} \end{bmatrix} = \begin{bmatrix} -\nabla f(\underline{x}_t) \\ 0 \end{bmatrix}$$

$$\underline{v}^*(x_t) = \left(\begin{bmatrix} \nabla^2 f(x_t) & A^T \\ A & 0 \end{bmatrix}^{-1} \begin{bmatrix} -\nabla f(x_t) \\ 0 \end{bmatrix} \right)_{1:n}$$

<u>Update</u> :

$$x_{t+1} = x_t + \delta_t \, v^*(x_t)$$

Inequality constraints

$$\min \; f_0(\underline{x})$$
$$f_1(\underline{x}) \leq 0$$

Suppose we have $\underline{x}_t$ s.t. $f_1(\underline{x}) < 0$

find $\underline{x}_{t+1}$ s.t. $f_1(x_{t+1}) < 0$

& $f_0(x_{t+1}) \leq f_0(\underline{x}_t)$

$$\bar{f}_t(\underline{x}) = f_0(\underline{x}) + \epsilon_t g(f_1(\underline{x}))$$

g: Barrier fn

$$g(u) = \begin{cases} \text{finite} & \text{for } u < 0 \\ \infty & \text{for } u = 0 \end{cases}$$

Problem :

$$\text{Minimize} \quad f_0(\underline{x})$$
$$\text{ST} \quad f_i(\underline{x}) \leq 0 \qquad i = 1, 2 \cdots m$$
$$A\underline{x} = \underline{b}$$

$$\equiv \text{Minimize} \quad f_0(\underline{x}) + \sum_{i=1}^{m} I_-(f_i(\underline{x}))$$
$$\text{ST} \quad A\underline{x} = b$$

$$I_-(z) = \begin{cases} 0 & \text{if } z \leq 0 \\ \infty & \text{else} \end{cases}$$

Replace $I_-(z)$ with a barrier function

$$L_t(z) = -\frac{1}{t}\log(-z)$$

$$\phi(x) = -\sum_{i=1}^{m} \log(-f_i(x))$$ → log barrier fn for the problem

Solve for $\min_{x : Ax = b} \; t f_0(x) + \phi(x)$

$$\nabla \phi(x) = -\sum_{i=1}^{m} \frac{1}{f_i(x)} \nabla f_i(x)$$

$$\nabla^2 \phi(x) = -\sum_{i=1}^{m} \frac{1}{f_i(x)} \nabla^2 f_i(x) + \sum_{i=1}^{m} \frac{1}{f_i^2(x)} (\nabla f_i(x))(\nabla f_i(x))^T$$

Barrier Method:

Initialize $\underline{x}$ that satisfies $A\underline{x} = \underline{b}$

$f_i(\underline{x}) < 0 \quad \forall i$

$\mu > 1, \quad \epsilon > 0, \quad t > 0$

While not stopping condn reached: $\quad \frac{m}{t} \leq \epsilon$

Outer loop

Run Newton's method starting at $\underline{x}$ for $\min_{A\underline{x} = \underline{b}} \; t f_0(x) + \phi(\underline{x})$ — Inner loop

Call the o/p $x^*(t)$

Set $t \leftarrow \mu t$

$\underline{x} \leftarrow x^*(t)$

In each outer iteration,

$$\text{solve} \quad \min_{x \,:\, Ax = b} \; t f_0(x) + \phi(x)$$

$$\downarrow$$

$$x^*(t)$$

How much does $f(x^*(t))$ differ from f^*

$$\downarrow$$

$$\begin{aligned} &\min \; f_0(x) \\ &\text{s.t. } Ax = b \\ &\quad f_i(x) \le 0 \;\; \forall i \end{aligned}$$

<u>Claim</u> : $f(x^*(t)) - f^* \le \dfrac{m}{t}$

Eg 1

Minimize $\|A\underline{x} - \underline{b}\|^2$

ST $\|\underline{x}\|^2 \le 1$

$$= \text{Minimize } \overbrace{\underline{x}^T A^T A \underline{x} - 2 b^T A x + b^T b}^{f_0(x)}$$

ST $x^T x - 1 \le 0$

$$\nabla f_0(\underline{x}) = 2 A^T A \underline{x} - 2 A^T \underline{b}$$

$$\nabla^2 f_0(\underline{x}) = 2 A^T A$$

$$\phi(\underline{x}) = -\log(-(x^T x - 1)) = -\log(1 - \underline{x}^T \underline{x})$$

$$\nabla \phi(\underline{x}) = + \frac{2\underline{x}}{(1 - \underline{x}^T \underline{x})}$$

$$\nabla^2 \phi(x) = \frac{2I}{1 - x^T x} - \frac{2}{(1 - x^T x)^2} (-2 x x^T)$$

$$= \frac{2I}{1 - x^T x} + \frac{4 x x^T}{(1 - x^T x)^2}$$