Tutorial optimization

10th June 21

Ex. 2 Inequality constraint

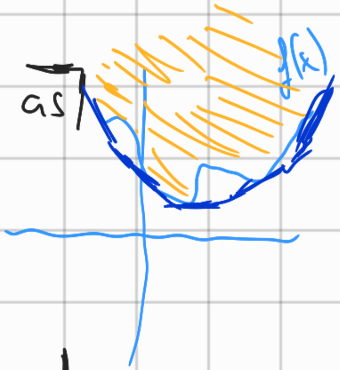
$$\begin{aligned} &\underset{x}{\text{minimize}} \quad c^T x \\ &\text{subject to} \quad f_1(x) \leq 0 \end{aligned} \qquad (2)$$

i) Express the dual problem of the primal problem given in (2) with $c \neq 0$ in terms of the conjugate.

[Conjugate of a function $f(x)$ is $f^*(x)$ defined as

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

f^* is always convex (even if f is not)]



First we write the Lagrangian of (2):

$$L(x, \lambda, v) = \underbrace{c^T x}_{f_0(x)} + \lambda f_1(x)$$

$$L(x,\lambda,v) = f_0(x) + \lambda^T f(x) + v^T h(x)$$

The dual function is defined as $\left[g(\lambda, v) = \min_x L(x, \lambda, v) \right]$

And the dual problem of (2) is

$$\begin{aligned} &\underset{\lambda, v}{\text{maximize}} \quad g(\lambda, v) \\ &\text{subject to} \quad \lambda \geq 0 \end{aligned} \qquad (3)$$

$$\min f(x) = -\max -f(x)$$

Now, note that in our case

$$g(\lambda) = \min_x \{ c^T x + \lambda f_1(x) \} = \lambda \min_x \left\{ \frac{c^T}{\lambda} x + f_1(x) \right\} =$$

$$= - \lambda \max_x \left\{ - \frac{c^T}{\lambda} x - f_1(x) \right\} = -\lambda f_1^* \left(\frac{-c}{\lambda} \right)$$

Therefore, we can write the dual problem (3) in terms of the conjugate of $f_1(x)$:

$$\begin{aligned} &\underset{\lambda}{\text{maximize}} \quad -\lambda f_1^* \left(\frac{-c}{\lambda} \right) \\ &\text{subject to} \quad \lambda \geq 0 \end{aligned} \qquad (4)$$

$\max f(x)$

$\min -f(x)$

ii) Explain why (4) is convex. f does not need to be convex.

We can write (4) in standard form, i.e.,

$$-\underset{\lambda}{\text{minimize}} \quad \lambda f_1^*(-c/\lambda) \qquad (5)$$
$$\text{subject to} \quad -\lambda \leq 0$$

Here, $f_0(\lambda) = \lambda f_1^*(-c/\lambda)$ $\qquad f_1(\lambda) = -\lambda$

The perspective of a function $f: \mathbb{R}^n \to \mathbb{R}$ is the function $g: \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ $\quad g(x,t) = t f(x/t)$

- If f is convex, g is convex.

- f_0 is the perspective of $f_1^*(-y)$ $\Rightarrow$ convex iff $f_1^*(-y)$ convex
- $f_1^*(-y)$ convex $\Leftrightarrow$ $f_1^*(y)$ convex
- $f_1^*(y)$ convex since is the point-wise supremum of a family of convex functions w.r.t. y.

Thus f_0 is convex. $f_1(\lambda) = -\lambda$ is convex. Then (5) is convex.

Ex. 3 KKT conditions

$$\underset{X \in S^n}{\text{minimize}} \quad tr(X) - \log\det X \qquad (6)$$
$$\text{subject to} \quad Xs = y$$

$\text{dom } f_0(X) = S^n_{++}$

i) Derive the KKT conditions for (6) with $y, s \in \mathbb{R}^n$ s.t. $y^T s = 1$

1) Primal constraints: $h_1(X) = Xs - y = 0$; $X \succ 0$
2) Dual constraints: ϕ (there is no λ)
3) Complementary slackness: ϕ $\quad \lambda_i f_i(x) = 0$
4) Gradient Lagrangian w.r.t. x vanishes: $[\nabla_X L(x, \lambda, v) = 0]$

$L(x, \lambda, v) = tr(X) - \log\det X + v^T(Xs - y)$ with $v \in \mathbb{R}^n$

$\nabla_X L(x, \lambda, v) = I - X^{-T} + \frac{1}{2}(sv^T + vs^T) = 0$

$\Rightarrow X^{-T} = I + \frac{1}{2}(sv^T + vs^T)$

$\Rightarrow X^{-1} = I + \frac{1}{2}(vs^T + sv^T)$ $\quad (X^{-1})^{-1} = (X^{-1})^{-1} = X$

ii) Verify that the optimal solution is given by

$$X^* = I + yy^T - \frac{1}{s^Ts} ss^T$$

If (6) is a convex problem, then KKT is a sufficient condition for optimality, that is, if X^* holds KKT, then it is optimal.

Problem (6) is convex since $Xs - y = h_1(x)$ is affine and both $tr(x)$ and $-\log \det X$ are convex. $\rightarrow tr(X) - \log\det X$ $\quad f_0(x)$

$f((-\log) \circ \det X)$

$\boxed{\times}$ Condition $X^*s - y = 0$

$$(I + yy^T - \frac{1}{s^Ts} ss^T)s - y = s + y(\cancel{y^Ts})^{1} - \frac{1}{\cancel{s^Ts}} s(\cancel{s^Ts}) - y = s - s + y - y = 0$$

$\boxed{\times}$ Condition $X^{-1} = I + \frac{1}{2}(sv^T + vs^T) \Rightarrow (I + \frac{1}{2}(sv^T + vs^T))X^* = I$

what's v?

$XS = y$

$s = X^{-1}y$

$v^Ty \in \mathbb{R}$

$\langle v, y \rangle$

$\langle y, v \rangle$

y^Tv

$X^{-1} = I + \frac{1}{2}(vs^T + sv^T) \overset{y}{\Rightarrow} X^{-1}y = s = y + \frac{1}{2}v\cancel{s^Ty}^{1} + \frac{1}{2}sv^Ty \quad *$

$* \quad y^T \Rightarrow y^Ts = 1 = y^Ty + \frac{1}{2}y^Tv + \frac{1}{2}\cancel{y^Ts}v^Ty$

$\Rightarrow v^Ty = 1 - y^Ty$

$* \Rightarrow s = y + \frac{1}{2}v + \frac{1}{2}s(1 - y^Ty) \Rightarrow \frac{1}{2}v = \frac{1}{2}s - y + \frac{1}{2}sy^Ty$

$\Rightarrow v = -2y + (1 + y^Ty)s$

Replace v in X^{-1}

$$X^{-1} = I + \frac{1}{2}(sv^T + vs^T) = I + \frac{1}{2}\left(s(-2y + (1+y^Ty)s)^T + (-2y + (1+y^Ty)s)s^T\right) =$$

$$= I - sy^T + \frac{1}{2}ss^T + \frac{1}{2}sy^Tys^T - ys^T + \frac{1}{2}ss^T + \frac{1}{2}y^Tyss^T$$

sy^Tys^T

$$= I - ys^T - sy^T + ss^T + y^Tyss^T = I - ys^T - sy^T + (1 + y^Ty)ss^T = X^{-1}$$

We now need to check that $X^{-1}X^* = I$

$$[I - ys^T - sy^T + (1 + y^Ty)ss^T](I + yy^T - \frac{1}{s^Ts}ss^T) = 1) + 2) + 3)$$

1) $I - ys^T - sy^T + (1 + y^Ty)ss^T = X^{-1}$

2) ...

$2)\ yy^T - y\cancel{s^Ty}y^T - sy^Tyy^T + (1+y^Ty)s\cancel{s^Ty}y^T = \cancel{yy^T} - \cancel{yy^T} - \cancel{yy^Tsy^T} + \cancel{yy^Tsy^T} + sy^T$

$y^Ts = s^Ty = 1$

$ys^T \neq sy^T = (s_i y_j) \in \mathbb{R}^{n\times n} = sy^T$

$3)\ \frac{-1}{s^Ts} ss^T + \frac{1}{\cancel{s^Ts}} y \cancel{s^Ts} s^T + \frac{1}{s^Ts} s \cancel{y^Ts} s^T - \frac{(1+y^Ty)}{\cancel{s^Ts}} s \cancel{s^Ts} s^T =$

$= \cancel{\frac{-1}{s^Ts} ss^T} + \cancel{\frac{1}{s^Ts} ss^T} + ys^T - (1+y^Ty)ss^T = ys^T - (1+y^Ty)ss^T$

$= 1)+2)+3) = I \underbrace{- \cancel{ys^T} - \cancel{sy^T} + \cancel{(1+y^Ty)ss^T}}_{1)} \underbrace{+ \cancel{sy^T}}_{2)} \underbrace{+ \cancel{ys^T} - \cancel{(1+y^Ty)ss^T}}_{3)} = I$

$\boxed{x}$ Condition $X^* > 0$

We show that $a^TX^*a > 0$ for any $a \in \mathbb{R}^n$ $a \neq 0$

$a^Ty = y^Ta$

$(a^Ty)(y^Ta) = (a^Ty)^2$

$a^T(I + yy^T - \frac{1}{s^Ts}ss^T)a = a^Ta + (a^Ty)(y^Ta) - \frac{1}{s^Ts}(a^Ts)(s^Ta) =$

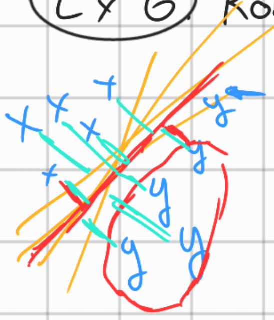
$= \underbrace{a^Ta}_{>0} + \underbrace{(a^Ty)^2}_{\geq 0} - \underbrace{\frac{(a^Ts)^2}{s^Ts}}_{>0}$ ⟹ lowest value when $s = ka$ with $k \neq 0$, $k \in \mathbb{R}$

$a^Ts = a^Tka$

Then $a^Ta + (a^Ty)^2 - \frac{k^2(a^Ta)^2}{k^2(a^Ta)} = \cancel{a^Ta} + (a^Ty)^2 - \cancel{a^Ta} = (a^Ty)^2 = \frac{1}{k^2} > 0$

$ka^Ty = s^Ty = 1 \to a^Ty = \frac{1}{k}$

Ex 6 Robust linear classification



$$\begin{aligned} &\underset{t,w,b}{\text{maximize}} && t \\ &\text{subject to} && w^Tx_i - b \geq t \quad i = 1,2,\dots,N \\ & && w^Ty_i - b \leq -t \quad i = 1,2,\dots,M \\ & && \|w\|_2 \leq 1. \end{aligned} \qquad (7)$$

$w^Tx_i \geq b + t$

$w^Ty_i \leq b - t$

i1) Show that $t^* > 0$ iff $\{x_i\}$ and $\{y_i\}$ are linearly separated

if $t^* > 0$ then

$w^{*T}x_i \geq b^* + t^* > b^* > b^* - t^* \geq w^{*T}y_j$) $-b^*$

$[w^{*T}x_i - b^* > 0 > w^{*T}y_j - b^*]$ $\forall i$ $\forall j$

Thus $f(x) = w^{*T}x - b^*$ linearly separates $\{x_i\}$ and $\{y_j\}$.

if $\exists w, b$ such that $[w^Tx_i - b > 0 > w^Ty_j - b]$

max t

$t \leq t^*$

then

$0 < t = \min_{i,j}(wx_i - b, b - wy_j) \cdot \frac{1}{2} < t^*$ and t holds (7) i.2) Show that the optimal w^* always $\|w^*\| = 1$.

Assume w^*, b^*, t^* optima of (7) with $\|w^*\| < 1$.

Dividing by $\|w^*\| = c < 1$ $\quad \frac{w^*}{c}, \frac{b^*}{c}, \frac{t^*}{c}$

$$\left.\begin{array}{l} w^* x_i - b^* \geq t^* \\ w^* y_j - b^* \leq t^* \end{array}\right\} \Rightarrow \begin{array}{l} \left(\frac{w^*}{c}\right) x_i - \left(\frac{b^*}{c}\right) \geq \left(\frac{t^*}{c}\right) \\ \left(\frac{w^*}{c}\right) y_j - \left(\frac{b^*}{c}\right) \leq \left(\frac{t^*}{c}\right) \end{array} \text{ also holds (7)}$$

and $t^*/c > t^* \Rightarrow w^*, b^*, t^*$ were not optimal ↯

ii)

$$\begin{array}{ll} \max & t \\ \text{s.t} & w^T x_i - b \geq t \\ & w^T y_j - b \leq -t \\ & \|w\| \leq 1 \end{array} \quad \boxed{\begin{array}{l} \tilde{w} = \frac{w}{t} \\ \tilde{b} = \frac{b}{t} \end{array}} \Longleftrightarrow \quad \begin{array}{ll} \min & \|\tilde{w}\| \\ \text{s.t} & \tilde{w}^T x_i - \tilde{b} \geq 1 \\ & \tilde{w}^T y_j - \tilde{b} \leq -1 \end{array}$$

w, t, b feasible iff $\tilde{w}, \tilde{b}$ feasible $\quad \tilde{w}, \tilde{b}, 1$

$\max t \equiv \min \frac{w}{t} = \min \tilde{w}$ $\quad C = \|\tilde{w}\| > 1$

$\square$