Homework 11

1. Dual of a QP and differentiability

(a) consider $p^* = \max_x c^T x - \frac{1}{2} x^T Q x : Ax \le b$

$= \min_x \left(-c^T x + \frac{1}{2} x^T Q x\right)$

$Q \in S^n_{++}$, $Q \succ 0$ } symmetric positive definite

$\mathcal{L}(\vec{x}, \vec{\lambda}) = -c^T \vec{x} + \frac{1}{2} \vec{x}^T Q \vec{x} + \vec{\lambda}^T (A\vec{x} - \vec{b})$

Lagrangian

dual function

$g(\vec{\lambda}) = \min_x \mathcal{L}(\vec{x}, \vec{\lambda})$

Then the dual problem

$$\boxed{d^* = \max_{\vec{\lambda} \ge 0} g(\vec{\lambda})}$$

Primal problem in standard form

$p^* = \min_x -c^T x + \frac{1}{2} x^T Q x$

s.t. $Ax - b \le 0$ $\qquad A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ $\quad c \in \mathbb{R}^n$

$Q \in S^n_{++}$

$f_0 = -c^T x + \frac{1}{2} x^T Q x$

$f_1 = Ax - b$

1) Hession $\nabla^2 f_0 = Q \succ 0$ is positive definite $\therefore$ convex

2) constraint[s] f_1 is affine and $\exists x_0 \in \text{Relint}(\mathcal{D})$ s.t. $Ax_0 - b \le 0$ $\rightsquigarrow$ sol'n in half space

? therefore feasible, satisfying weak slaters condition

By 1) and 2) Strong Duality holds

(b) $p^* = \min_x -c^T x + \frac{1}{2} x^T Q x$

$$= \max_{\vec{\lambda} \geq 0} \underbrace{\min_x -\vec{c}^T \vec{x} + \frac{1}{2} \vec{x}^T Q \vec{x} + \vec{\lambda}^T (A\vec{x} - \vec{b})}_{} = d^*$$

$\lambda \in \mathbb{R}^m$

$$\Delta_x \mathcal{L}(\vec{x}, \vec{\lambda}) = -\vec{c} + Qx + A^T \vec{\lambda} = 0$$

$A^T_{n \times m}$

$$\Rightarrow \vec{x} = Q^{-1}(\vec{c} - A^T \vec{\lambda})$$

minimum

$Q^{-1} \in S^n_{++}$

plugging in $\vec{x}$

$$p^* = \max_{\vec{\lambda} \geq 0} \underbrace{-c^T Q^{-1}(\vec{c} - A^T \vec{\lambda}) + \frac{1}{2}(\vec{c}^T - \vec{\lambda}^T A) Q^{-1} Q Q^{-1} (\vec{c}^T - A^T \vec{\lambda}) + \vec{\lambda}^T (A Q^{-1}(\vec{c} - A^T \vec{\lambda}) - \vec{b})}_{}$$

let $h(c) = -\vec{c}^T Q^{-1} \vec{c} + c^T A^T \vec{\lambda} + \frac{1}{2}(\vec{c}^T Q^{-1} - \lambda^T A Q^{-1})(\vec{c}^T - A^T \vec{\lambda}) + \vec{\lambda} A Q^{-1}(\vec{c} - A^T \vec{\lambda}) - \vec{\lambda}^T \vec{b}$

$$\nabla_c^2 h(c) = -2Q^{-1} + Q^{-1}$$

$$= -Q^{-1} \prec 0 \qquad \text{since } Q \succ 0 \Leftrightarrow Q^{-1} \succ 0$$

so p^* is concave function of c

(c) $p^*(c) = \max_x f_x(c)$

where $f_x(c) = c^T x - \frac{1}{2} x^T Q x$

where $f(c) = \max_x f_x(c)$

since $p^*(c)$ is concave, in order to find $f(c)$ we take the gradient

$$\nabla_x f_x(c) = 0$$

$$c - Qx = 0$$

$$x = Q^{-1} c \quad \Rightarrow \quad x^T = c^T Q^{-1} \qquad Q \in S_+^n \Rightarrow Q^{-1} \in S_+^n$$

$$\Rightarrow f(c) = c^T Q^{-1} c - \frac{1}{2} c^T Q^{-1} c$$

$$= \frac{1}{2} c^T Q^{-1} c$$

Then we can take any subgradient of this function

subgradient g at x s.t.

$$\forall z \quad f(z) \geq f(x) + g^T (z - x)$$

Taylor expand around c

$$f(z) \simeq \frac{1}{2} c^T Q^{-1} c + (Q^{-1} c)^T (z - c)$$

then subgradient $\boxed{g = c^T Q^{-1}}$

(d). Assume 2 subgradients g_1, g_2 at every point x

$$f(z) \geq f(x) + g_1^T(z-x) = \frac{1}{2}x^TQ^{-1}x + g_1^T(z-x)$$
$$f(z) \geq f(x) + g_2^T(z-x) = \frac{1}{2}x^TQ^{-1}x + g_2^T(z-x)$$

since this is true for all z

2. KKT conditions

$$\min_{x \in \mathbb{R}^n} \sum^n \left(\frac{1}{2} d_i x_i^2 + r_i x_i\right)$$

s.t. $a^T x = 1$, $x_i \in [-1, 1]$, $i = 1, \ldots, n$

where $|a_i| \geq 1$ and $d_i > 0$ for $i = 1, \ldots, n$

(a) $D = \mathrm{diag}(d_1, \ldots, d_n)$, $r = (r_1, \ldots, r_n)$

$$\left.\begin{aligned} \frac{1}{2} x^T D x &= \sum^n \frac{1}{2} d_i x_i^2 \\ r^T x &= \sum^n r_i x_i \end{aligned}\right\} \text{by definition}$$

so $\min_x \sum \frac{1}{2} d_i x_i^2 + r_i x$ ...

can be written as

$x_i \in [-1, 1]$ is encompassed in $x_i^2 \leq 1$

$$\begin{aligned} &\min_x \frac{1}{2} x^T D x + r^T x \\ &\text{s.t. } a^T x \leq 1 \\ &\qquad x_i^2 \leq 1 \end{aligned}$$

(b) • The objective and constraints are differentiable

let $f_0 = \frac{1}{2}x^T D x + r^T x$

$h_1 = a^T x - 1$ equality constraint function

$f_1 = x_i^2 - 1$

$\Rightarrow$ $\nabla f_0 = Dx + r$

$\nabla h_1 = a$

$\frac{\partial f_1}{\partial x_i} = 2x_i$

$\therefore$ objective and constraints are differentiable

• Strong duality holds

$\nabla^2 f_0 = D \succ 0$ since $d_i > 0$ $\forall i$

$\therefore$ objective function is convex

equality constraint

$a^T x - 1 = 0$ is satisfied if we let

let $x^T = (0, \cdots, \frac{1}{a_k}, \cdots, 0)$ for some $|a_k| > 1$

strict

then inequality constraint

$x_i^2 - 1 < 0$ for $i = 1, \dots, n$ except for $i = k$

strict

and for $x_k^2 - 1 = \frac{1}{a_k^2} - 1 < 0$ since $|a_k| > 1$

Therefore the problem is strictly feasible, ie satisfies slaters condition

and since the the problem is convex

f_0, f_1 are convex and h_1 is affine

$\boxed{\text{Strong duality holds}}$

Scanned with CamScanner

- The optimization problem is convex

$f_0 = \frac{1}{2}x^TDx + r^Tx$ is convex

$\nabla^2 f_0 = D \succ 0$ since hessian is positive definite

$f_1 = x_i^2 - 1$ is convex since

$f_1'' = 2 > 0$ second derivative is positive

and

$h_p = a^T x - 1$ is affine

$\therefore$ this problem is convex

Scanned with CamScanner

(c) show $\mathcal{L}(x,\lambda,\mu) = \frac{1}{2}x^T(D+\Lambda)x + (r+\mu a)^T x - (\mu + \frac{1}{2}\sum \lambda_i)$

where $\Lambda = \mathrm{diag}(\lambda)$

$\lambda = (\lambda_1, \dots, \lambda_n)$: inequality μ: equality

$$\mathcal{L}(x,\lambda,\mu) = \frac{1}{2}x^T D x + r^T x + \underbrace{x^T \Lambda x - \lambda 1\!\!1}_{\text{inequality}} + \underbrace{\mu a^T x - \mu}_{\text{equality}}$$

If we write inequality constraint as $\frac{1}{2}x_i^2 \le \frac{1}{2}$

$$\Rightarrow \mathcal{L}(x,\lambda,\mu) = \frac{1}{2}x^T D x + r^T x + \frac{1}{2}x^T \Lambda x - \frac{1}{2}\lambda 1\!\!1 + \mu a^T x - \mu$$

$$= \frac{1}{2}x^T(D+\Lambda)x + (r+\mu a)^T x - \left(\mu + \frac{1}{2}\lambda 1\!\!1\right)$$

where $1\!\!1^T = (1, \dots, 1)$ ☒

KKT conditions

Primal feasibility

$a^T x \le 1$

$x_i^2 \le 1$

Dual feasibility

$\lambda_i \ge 0$

complementary slackness

$$\frac{1}{2} x^T \Lambda x - \frac{1}{2} \lambda \mathbb{1} = 0$$

Stationarary

$$\nabla_x \mathcal{L}(x, \lambda, \mu) = 0$$

$$= (D + \Lambda) x + r + \mu a = 0$$

$$x = -(D + \Lambda)^{-1} (r + \mu a)$$

(d) from (c)

$$\nabla_x \mathcal{L}(x, \lambda, \mu) = 0$$

$$\Rightarrow x^* = -\underbrace{(D + \Lambda^*)^{-1}}(r + \mu^* a)$$ in terms of optimal $\lambda : \lambda^*$ and $\mu : \mu^*$

componentwise

$$x^* = -\frac{(r_i + \mu_i^* a_i)}{d_i + \lambda_i^*}$$ since $(D + \Lambda)$ is diagonal

☒

(e) plugging in previous x^* to $\mathcal{L}(x, \lambda, \mu)$

matrix form

$$\mathcal{L}(x^*, \lambda, \mu) := g(\lambda, \mu)$$

$$= \frac{1}{2} x^{*T}(D + \Lambda) x^* + (r + \mu a)^T x^* - \left(\mu + \frac{1}{2} \lambda \mathbb{1}\right)$$

$$= +\frac{1}{2}(r + \mu a)^T (D + \Lambda)^{-1} (r + \mu a)$$

$$- (r + \mu a)^T (D + \Lambda)^{-1} (r + \mu a) \quad - \left(\mu + \frac{1}{2} \lambda \mathbb{1}\right)$$

$$= -\frac{1}{2}(r + \mu a)^T (D + \Lambda)^{-1} (r + \mu a) - \left(\mu + \frac{1}{2} \lambda \mathbb{1}\right)$$

componentwise (rearranging)

$$= \boxed{-\mu - \frac{1}{2} \sum^{n} \left[\frac{(r_i + \mu a_i)^2}{d_i + \lambda_i} + \lambda_i \right]}$$

☒

(f) Find $\max_{\lambda \geq 0} g(\lambda, \mu)$ for fixed μ

note $g(\lambda,\mu)$ is concave so we need only take the gradient and set it equal to zero

$$\nabla_\lambda g(\lambda,\mu) = 0$$

$$= -\mu - \frac{1}{2} \sum^{n} \left[\frac{(r_i + \mu a_i)^2}{d_i + \lambda_i} + \lambda_i \right] = 0$$

$$= \frac{1}{2} \sum \left[\frac{(r_i + \mu a_i)^2}{(d_i + \lambda_i)^2} + 1 \right] = 0$$

$$\frac{(r_i + \mu a_i)^2}{(\lambda_i + \lambda_i)^2} = -1$$

$$|r_i + \mu a_i| - d_i = \lambda_i$$

where $|r_i + \mu a_i| \geq d_i$ - in order to satisfy $\lambda_i \geq 0$

(g) $x^*(\mu) = g(\lambda^*, \mu)$ where $\lambda_i^* = |r_i + \mu a_i| - d_i$

$$= \boxed{-\mu - \frac{1}{2}\sum^{n} \frac{(r_i+\mu a_i)^2}{|r_i+\mu a_i|} + |r_i+\mu a_i| - d_i}$$

(h)

3. A matrix problem with strong duality

$p^* = \min_{\Delta} c^T(A+\Delta)^{-1}b \ : \ \|\Delta\| \le 1$ $\qquad$ $\|\cdot\|$: largest singular value norm

$\sigma_{min}(A) > 1$ strict $\qquad$ smallest singular value of A

(a) objective function

$f_0(\Delta) = c^T(A+\Delta)^{-1}b$ is well-defined for $\|\Delta\| \le 1$

if $(A+\Delta)^{-1}$ exists

since $\sigma_{min}(A) > 1$ A has no singular values equal to zero

Thus implies A is full rank, since A is square, A is invertible

proof

$$\Sigma = \begin{pmatrix} \sigma_{max} & & 0 \\ & \ddots & \\ 0 & & \sigma_{min} > 1 \end{pmatrix}_{n \times n}$$

by assumption $\sigma_{min}(A) \ge 1$

since $\Sigma y = 0 \Rightarrow y = 0$

and U and V^T are orthogonal matrices

$\Rightarrow A$ is invertible

$\Rightarrow \det(A) \ne 0$

since $\det(A+\Delta) \ge \det(A) + \det(\Delta)$

and $\det(A) \ne 0$

$\det(A+\Delta) \geq \det(A) + \det(\Delta) > 0$

is long as $\det(A) \neq -\det(\Delta)$

than $\det(A+\Delta) \neq 0$

$$\det(A) = \det(U\Sigma V^T) = \det(U)\det(\Sigma)\det(V^T)$$
$$= \det(U)\det(\Sigma)\det(V)$$
$$= \pm\det(\Sigma)$$

since $\det(I) = \det(U^TU) = \det(U)^2 = 1$

$\det(U) = \pm 1$

U, V are orthogonal

$\det(\Delta) = \det(\Sigma_\Delta)$

since $\sigma_{max}(\Delta) \leq 1$

and $\sigma_{min}(A) > 1$

$\det(A) = \det(\Sigma_A) \neq \det(\Sigma_\Delta) = \det(\Delta)$

$\therefore \det(A+\Delta) \neq 0$

and is therefor invertible

(b) Not convex

(c) $p^* = \min_{\Delta, t} t$

s.t $\quad b^T (A+\Delta)^{-1} b \leq t$

$\|\Delta\| \leq 1$

$t - b^T (A+\Delta)^{-1} b \geq 0$

let $M = \begin{pmatrix} A+\Delta & b^T \\ b & t \end{pmatrix} \geq 0$

$\Rightarrow \quad A+\Delta > 0$

4. (a) the problem

$$\min_w \|Xw - y\|_2^2 + \lambda \|w\|_2^2$$

Id change labels from $\{0, 1\}^n$ to $\{-1, 1\}$

then the problem will find a w that will bring the data points close to either -1, or 1 then any $w_i > 0$ classifies the corresponding data point to 1 and vice versa, essentially classifying the data.