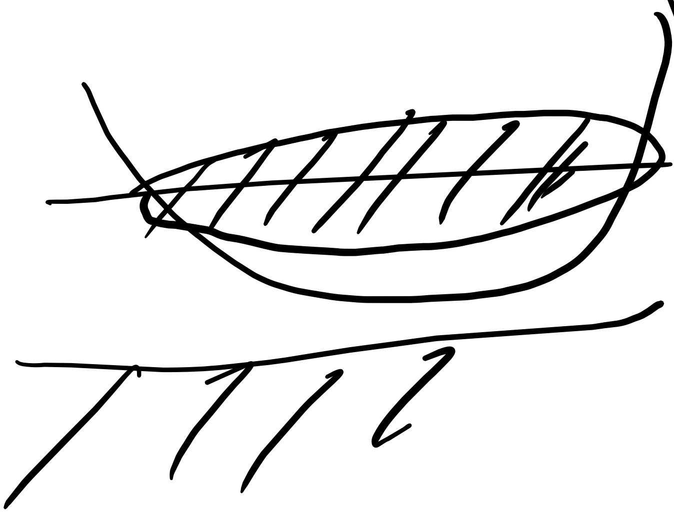
14/12/2021 Lecture - 22

Refined Slater's condition:

If the feasible set is a polyhedron and f_0 is convex (the objective function), then strong duality holds

$\boxed{g_i(\vec{x}) \le \alpha}$ → convex, if g_i is convex

$\boxed{h_i}(\vec{x}) = \alpha$ → affine



Slater's theorem

For a convex optimization problem, if (refined) Slater's condition, that is, $\exists \tilde{\vec{x}} \in \Omega = \boxed{\left(\bigcap_{i=1}^{m} \mathrm{dom}(g_i)\right)} \cap \mathrm{dom}\, f_0 \cap \left(\bigcap_{j=1}^{p} \mathrm{dom}(h_i)\right)$ ← affine function

$g_i(\tilde{\vec{x}}) < 0$ and $\boxed{h_i(\tilde{\vec{x}}) = 0}$.

(if g_i is an linear inequality constraint, we don't care whether $g_i(\tilde{\vec{x}}) < 0$ or not).

the strong duality holds and the dual optimum is attained

Proof.

$$A = \{ (\vec{u}, \vec{v}, t) \mid \exists\, \vec{x} \in \Omega = \mathrm{dom}(f_0) \cap \left(\bigcap_{i=1}^{n} \mathrm{dom}(g_i) \right)$$

$\vec{u} \in \mathbb{R}^m$, $\vec{v} \in \mathbb{R}^p$, $t \in \mathbb{R}$

$$g_i(\vec{x}) \leq u_i, \quad i = 1, \ldots, m$$

$$h_j(\vec{x}) = v_j, \quad j = 1, \ldots, p$$

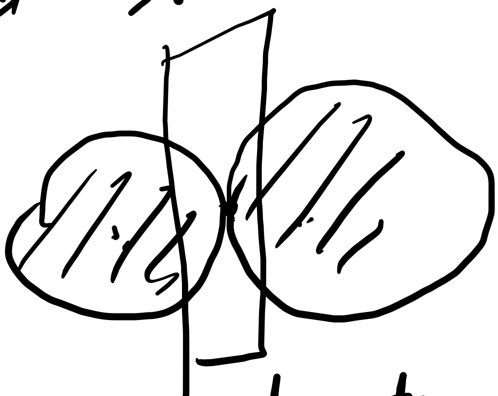
$$f_0(\vec{x}) \leq t \}.$$

$\vec{w} \geq \vec{u}$, and $(\vec{u}, \vec{v}, t) \in A$, then $(\vec{w}, \vec{v}, t) \in A$.

$$B = \{ (\vec{0}, \vec{0}, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} : s < \alpha^* \}.$$

$$A \cap B = \phi.$$

Since f_0, g_i's are convex functions and h_j's are affine functions we have that $\mathcal{A}$ is a convex set in $\mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R}$.



By the separating hyperplane theorem, $\exists\ \vec{\lambda} \in \mathbb{R}^m$, $\vec{\nu} \in \mathbb{R}^p$, $\mu \in \mathbb{R}$ |

$$\mathcal{B} \subseteq H^-_{(\underbrace{\vec{\lambda}}_{\mathbb{R}^m}, \underbrace{\vec{\nu}}_{\mathbb{R}^p}, \mu), \alpha} \qquad \mathcal{A} \subseteq H^+_{(\vec{\lambda}, \vec{\nu}, \mu)}$$

non-zero vector

$\alpha \to (\vec{0}, \vec{0}, s)$

$\mu s \leq \alpha. \ \forall s < p.$

$$\boxed{\mu t \le \alpha} \quad \forall t \text{ s.t. } t < \alpha^*.$$

$$\boxed{\begin{array}{l}(\vec{0}, \vec{0}, t) \in B. \\ (\vec{x}, \vec{y}, \mu)\end{array}}$$

$$\left[\ \underline{\vec{x}}^T \vec{u} + \underline{\vec{y}}^T \vec{v} + \mu t \ge \alpha\right.$$

for $(\vec{u}, \vec{v}, t) \in A$

Claim: $\boxed{\vec{x} \ge \vec{0},}$ $\mu \ge 0$

Proof of Claim: If $\mu < 0$, then choose $t < \alpha^*$ with $|t| > \dfrac{\alpha}{|\mu|}$. Then $\mu t = |\mu| \cdot |t| > \alpha$. So Contradiction!

If $\tilde{\lambda}_1 < 0$, we may choose $u_1 >> 0$, s.t. $\tilde{\lambda}_1 u_1 << 0$.

In particular, $\boxed{\mu\alpha^* \le \alpha}$. (affine equality constraints)

$$\left[\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \vec{\nu}^T (A\vec{x} - \vec{b}) + \mu f_0(\vec{x}) \ge \alpha \ge \underbrace{\mu\alpha^*}_{0} \right]$$

Case I: $\mu > 0$.

$$\sum_{i=1}^{m} \frac{\tilde{\lambda}_i}{\mu} f_i(x) + \left(\frac{1}{\mu}\vec{\nu}\right)^T (A\vec{x} - \vec{b}) + f_0(\vec{x}) \ge \alpha^*$$

$$L\left(\vec{x}, \frac{\vec{\lambda}}{\mu}, \frac{\vec{\nu}}{\mu}\right) \geq \alpha^*,$$

$$\Rightarrow \quad F\left(\frac{\vec{\lambda}}{\mu}, \frac{\vec{\nu}}{\mu}\right) \geq \alpha^*,$$

$\left(\frac{\vec{\lambda}}{\mu}, \frac{\vec{\nu}}{\mu}\right)$ is dual feasible.

Optimal soln to the dual problem is β^*.

$$\max \; F(\vec{\lambda}, \vec{\nu})$$

$$\text{s.t.} \quad \vec{\lambda} \geq \vec{0}, \; \vec{\nu} \in \mathbb{R}^p.$$

$\beta^* \geq \alpha^*$. But by weak duality $\beta^* \leq \alpha^*$.

In fact,

$$F\left(\tilde{\vec{\lambda}}/u,\ \tilde{\vec{\nu}}/u\right) = \alpha^*$$

dual optimal soln.

Case II: $u = 0$ (not possible due to Slater's condition)

$$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) + \tilde{\vec{\nu}}^T (A\vec{x} - \vec{b}) \geq 0$$

$\exists\ \tilde{\vec{x}}$ s.t. $f_i(\tilde{x}) < 0$ and $A\tilde{\vec{x}} = \vec{b}$.

$$\Rightarrow \tilde{\lambda}_i = 0 \quad \forall\, 1 \le i \le m.$$

$\vec{y}^T(A\vec{x}-\vec{b}) \geq \vec{0}$,

(Ω, $\mathcal{D}$ are different)

Ω → domain

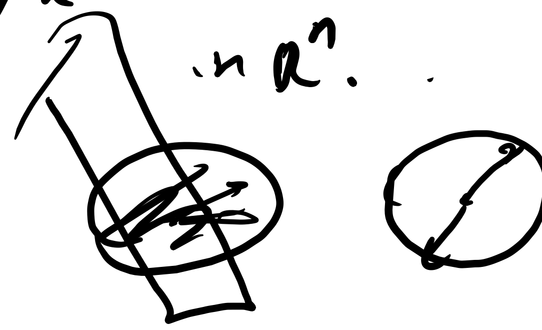
$\mathcal{D}$ → feasible set

(Wlog, assume A is full rank)

$\bar{x} \in \text{int}(\Omega)$, has dimension n.

$\vec{y}^T(A\vec{x}-\vec{b}) \geq \vec{0}$ $\forall \vec{x}$ in a ball in $\mathbb{R}^n$.

$$\boxed{\underline{\vec{y}^T A\vec{x}} \geq \vec{y}^T\vec{b}}$$



$\Rightarrow \quad \vec{y} A = \vec{0}$

$\Rightarrow \quad A^T \vec{y} = \vec{0}$

$\Rightarrow \quad \vec{y} = \vec{0}$.

Also, $\quad \vec{x} \geq \vec{0}, \quad \vec{y} \geq \vec{0}$,

Contradiction!!!

A nonconvex optimization problem with strong duality

$$\min. \quad \boxed{\langle \vec{x}, A\vec{x} \rangle} + 2\langle \vec{b}, \vec{x} \rangle, \quad \vec{x} \in \mathbb{R}^n,$$

subject to $\quad \|\vec{x}\|_2 \leq 1$

$\to$ Euclidean norm.

$A \to$ symmetric matrix in $M_n(\mathbb{R})$

$\vec{b} \in \mathbb{R}^n$,

If A is positive-semidefinite

$f(\vec{x}) = \langle \vec{x}, A\vec{x} \rangle + 2\langle \vec{b}, \vec{x} \rangle$, is convex.

If A is not positive-semidefinite,

f_b is not convex.

Detour in linear algebra:

GENERALIZED INVERSES / PSEUDO-INVERSES

Defn: A square matrix $A \in M_n(\mathbb{R})$ is said to be invertible if $\exists$ a square matrix $B \in M_n(\mathbb{R})$

s.t. $AB = BA = I_n$.

* A has full rank
* A has non-zero determinant.

The eq^n $A\vec{x} = \vec{b}$ has a unique

sol^n $\boxed{\vec{x}^* = A^{-1}\vec{b}}$.

How to solve $A\vec{x} = \vec{b}$ for general matrices

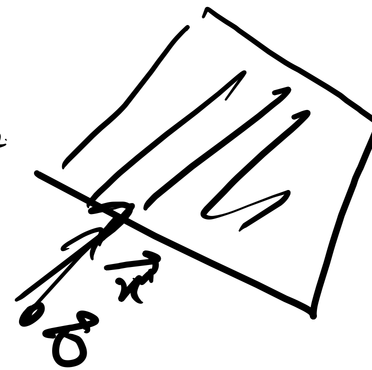
(neither ~~[illegible]~~ invertibility nor "square" ness is

assumed).

(1) If $A\vec{x} = \vec{b}$ is consistent, then we

want $\vec{x}^*$ with minimal ~~[illegible]~~ Euclidean

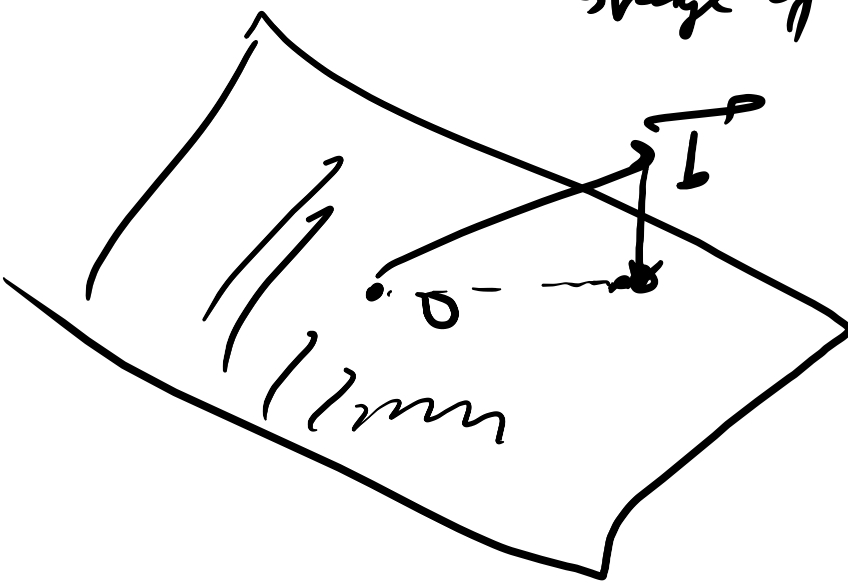
norm. ($\min \langle \vec{x}, \vec{x} \rangle$ s.t. $A\vec{x} = \vec{b}$)



(2) If $\vec{A}\vec{x}=\vec{b}$ is not consistent, we need to find $\vec{x}$ which minimizes

$$\|\vec{A}\vec{x}-\vec{b}\|_2 \qquad \vec{b}\in\mathbb{R}^m$$

range of A — linear subspace of $\mathbb{R}^m$.



$A\in M_{m,n}(\mathbb{R})$

$\boxed{\beta_0, \beta_1, \ldots, \beta_n,}$

$$\min \langle A\vec{x} - \vec{b}, A\vec{x} - \vec{b} \rangle \geq 0$$

subject to $\vec{x} \in \mathbb{R}^n$.

$$\langle \vec{x}, A^T A \vec{x} \rangle - 2 \langle A^T \vec{b}, \vec{x} \rangle + \langle \vec{b}, \vec{b} \rangle$$

positive semi-def

$$\langle \vec{x}, A \vec{x} \rangle + 2 \langle \vec{b}, \vec{x} \rangle$$

$$\boxed{\vec{x} = (A^T A)^{-1} A^T \vec{b}}$$

$$\boxed{(A^T A) \vec{x} = A^T \vec{b}}$$

slope parameters

Want to solve $A\vec{x} = \vec{b}$ without the assumption of Full column rank.

Remark: If A is of full column rank,

(Analysis)

(1) $(A^TA)^{-1}A^T$ (pseudoinverse of A)

Optimal sol^n: (pseudoinverse) $\vec{b}$,

(2) $\boxed{A(A^TA)^{-1}A^T}$ (projection matrix) — symmetric; and idempotent.

$$(A(A^TA)^{-1}A^T)(A(A^TA)^{-1}A^T)$$

$$= A(A^TA)^{-1}A^T.$$

$P^2=P$, $P^T=P$

V = range of P.

Thus, ~~Rtte~~ $Px=x$ for $x\in V$,

$x \in V^\perp \Rightarrow Px=0$

Defn. (Generalized inverse)

Let, $A \in M_{m,n}(\mathbb{R})$. Then $G \in M_{n,m}(\mathbb{R})$ is said to be a generalized inverse of A if $\boxed{AGA = A}$.

* For an invertible square matrix $\exists$ a unique generalized inverse $\boxed{G = A^{-1}}$

$\boxed{A\vec{x} = \vec{b}}$, $(AG)A\vec{x} = A(G\vec{b}) = \vec{b}$

$\Rightarrow \vec{x}^* = G\vec{b}$ is a soln to my eqn.

Theorem: $A\vec{x} = \vec{b}$, $\vec{b} \in$ column space of A.

that is the system is consistent.

Let G be a generalized inverse of A,

that is $AGA = A$. Then $G\vec{b}$ is a

particular soln to the system.

$$AGA = A \Rightarrow (AG)(AG) = AG$$

$$\Rightarrow (AG)^2 = AG$$

AG is an idempotent matrix.

Theorem: Let $A \in M_{m,n}(\mathbb{R})$ with generalized inverse $G \in M_{n,m}(\mathbb{R})$. Then $AG \in \mathbb{R}^{m\times m}$ is an idempotent matrix.

* AG and A must have the same column space. col sp. $AG \subseteq$ col sp. A. $A = (\underline{AG})A$.

* AG is projecting onto the column space of A.

GA is an idempotent matrix projecting onto the row space of A.

Pseudo-inverse (Moore-Penrose inverse)

<u>Defn</u>: Let $A \in M_{m,n}(\mathbb{R})$, $B \in M_{n,m}(\mathbb{R})$.
Then B is said to be a **pseudo-inverse** of A if it satisfies

(1) $ABA = A$ (B is a generalized inverse of A)

(2) $BAB = B$ (A is a generalized inverse of B)

(3) $(AB)^T = AB$ (AB is symmetric)

(4) $(BA)^T = BA$ (BA is symmetric)

~~Proof~~

Theorem. Every $A \in M_{m,n}(\mathbb{R})$ has a unique pseudo-inverse.

Proof. (Uniqueness). A^{+} (A dagger).

~~Let~~ Assume B, C satisfy the 4 conditions.

$$H = AB - AC \quad (\text{symmetric...})$$

$$H^2 = (AB - AC)(A(B - C))$$

$$= (ABA - ACA)(B - C)$$

$$= (A - A)(B - C) = 0$$

$\Rightarrow \quad H = 0$

$\Rightarrow \quad AB = AC$ (Similarly, $BA = CA$)

$$B \stackrel{(2)}{=} \underline{\underline{BAB}} = \underline{BAC} = \underline{CAC} \stackrel{(2)}{=} C$$

We denote the unique pseudoinverse (if it exists) by A^+.

Proof of SVD:

rank of A

$$A = \underbrace{U}_{m \times m} \underbrace{\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}}_{m \times n} \underbrace{V^T}_{n \times n}$$

orthogonal matrices

singular values of A

$:= \sqrt{\text{eigenvalues of } A^T A}$

$$(A^T A) v_i = \sigma_i^2 v_i \qquad \sigma_i \to i^{th} \text{ singular value}$$

$v_i \to$ eigenvector corresponding to σ_i^2 (with norm 1) in $\mathbb{R}^n$

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

$$u_i := \boxed{\frac{A v_i}{\sigma_i}} \quad \boxed{\text{if } \sigma_i \neq 0}$$

$$\langle u_i, u_i \rangle = \frac{\langle A^T A v_i, v_i \rangle}{\sigma_i^2} = 1.$$

$u_i \to$ unit vectors in $\mathbb{R}^m$.

$$\langle u_i, u_j \rangle = \frac{\langle A^T A v_i, v_j \rangle}{\sigma_i \sigma_j} = 0 \quad (i \neq j)$$

Add an orthonormal basis of $(\text{Range of } A)^\perp$ in $\mathbb{R}^m$.

$u_1, \dots, $ Obtain orthonormal basis of $\mathbb{R}^m$.

$T: \underline{V} \to \underline{W}$

~~P~~ Q^{-1}

Chose $\{v_i\}_{i=1}^{n}$ as basis for $\mathbb{R}^n$.

$\{u_j\}_{j=1}^{m}$ as basis for $\mathbb{R}^m$.

Then $A: \underline{\mathbb{R}^n} \to \mathbb{R}^m$ has diagonal

form in the above chosen basis

$x \in \mathbb{R}^n$, $\vec{x} = \alpha_1 \vec{v}_1 + \cdots + \alpha_n \vec{v}_n$

$A\vec{x} = \alpha_1 A\vec{v}_1 \cdots + \alpha_n A\vec{v}_n$

$= \alpha_1 \sigma_1 \underline{\vec{u}_1} \cdots + \alpha_n \sigma_n \underline{\vec{u}_n}$

$$A = U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^T$$

$$U = \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \qquad V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$$

$$B = V \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T.$$

$$AB = U \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} U^T, \quad BA = V \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} V^T.$$

$$(AB)A = \left(U \begin{bmatrix} \boxed{I_r} & 0 \\ 0 & 0 \end{bmatrix} \boxed{U^T}\right)\left(\boxed{U} \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^T\right)$$

$$= U \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} V^T = A.$$

$$BAB = B.$$

$$A = 0 \cdot E_0 + \lambda_1 E_1 + \cdots + \lambda_n E_n$$

$$A^{-1?} = 0 \cdot E_0 + \frac{1}{\lambda_1} E_1 + \cdots + \frac{1}{\lambda_n} E_n$$

$$AA^{?} = \boxed{E_1 + \cdots + E_n} + 0$$

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

Minimum-norm sol^n to a consistent linear system

Theorem – $A \in M_{m,n}(\mathbb{R})$, $\vec{b} \in \mathbb{R}^m$.

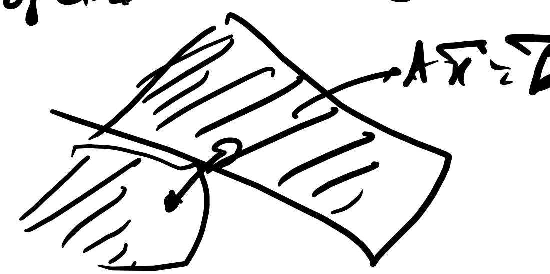
If $A\vec{x} = \vec{b}$ is consistent, then

$\vec{x}^* = A^+\vec{b}$ is an exact sol^n with

the smallest possible norm.

Proof – (A^+A) is an orthogonal projection matrix.

(AA^+) is an orthogonal projection onto the column space of A.



Since $A\vec{x}=\vec{b}$ is consistent, $\Rightarrow \vec{b} \in$ column space of A

$=$ column space of AA^T,

$\Rightarrow AA^+\vec{b} = \vec{b} \Rightarrow A^+\vec{b}$ is a soln to $A\vec{x}=\vec{b}$

$\Rightarrow A(A^+\vec{b}) = \vec{b}$

($\vec{x}$ [illegible])

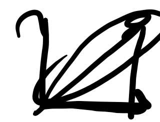
$$\vec{x} = (A^{\dagger}A)\vec{x} + (I - A^{\dagger}A)\vec{x}$$

$$= A^{\dagger}\vec{b} + (I - A^{\dagger}A)\vec{x}.$$

$$\langle A^{\dagger}\vec{b}, (I - A^{\dagger}A)\vec{x} \rangle$$

$$= \langle (I - A^{\dagger}A)A^{\dagger}\vec{b}, \vec{x} \rangle = 0$$

$$A^{\dagger}\vec{b} \perp (I - A^{\dagger}A)\vec{x}$$



$\|\vec{x}_1 + \vec{x}_2\|^2$

$= \|\vec{x}_1\|^2 + \|\vec{x}_2\|^2$

if $\langle \vec{x}_1, \vec{x}_2 \rangle = 0$

$$\|\vec{x}\|_2^2 = \|A^+\vec{b}\|_2^2 + \|(I - A^+A)\vec{x}\|_2^2$$

$$\Rightarrow \boxed{\|\vec{x}\|_2^2 \geq \|A^+\vec{b}\|_2^2}$$

Multiple linear regression

Theorem: 'A' minimizer for the problem

$$\min_{\vec{x} \in \mathbb{R}^n} \|A\vec{x} - \vec{b}\|_2 .$$

is given by $\vec{x}' = A^+\vec{b}$

Proof:

$A : \mathbb{R}^n \to \mathbb{R}^m$

$\boxed{\text{Range of } A} \subset \mathbb{R}^m$

projection onto col. sp. of A. is given by the matrix AA^+.

Thus, $A(A^+b)$ is the point on ~~[illegible]~~ column space of A. which is closest to $\vec{b}$.

$$\boxed{A^T A \vec{x} = A^T \vec{b}}$$ (normal eqn:

$$\boxed{\vec{x} = (A^T A)^{+} A^T \vec{b}}$$ is a

soln:

min $\vec{x} \in \mathbb{R}^n$ $\|A\vec{x} - \vec{b}\|^2$ ($=$) min $\vec{x} \in \mathbb{R}^n$

$\langle A\vec{x} - \vec{b}, A\vec{x} - \vec{b}\rangle$

$\langle \vec{x}, A^T A\vec{x}\rangle$

$- 2\langle A^T\vec{b}, \vec{x}\rangle$

$+ \langle \vec{b}, \vec{b}\rangle$

A nonconvex quadratic problem with strong duality

$$\min \; \langle \vec{x}, A\vec{x} \rangle + 2\langle \vec{b}, \vec{x} \rangle \qquad \vec{b} \in \mathbb{R}^n$$

$$\text{subject to} \; \langle \vec{x}, \vec{x} \rangle \leq 1. \qquad A \text{ is symmetric}$$

(Second order approximation to a function.) $\qquad M_n(\mathbb{R})$.

Lagrangian:

$$L(\vec{x}, \lambda) = \langle \vec{x}, A\vec{x} \rangle + 2\langle \vec{b}, \vec{x} \rangle + \lambda \langle \vec{x}, \vec{x} \rangle - \lambda$$

$$= \langle \vec{x}, (A + \lambda I)\vec{x} \rangle + 2\langle \vec{b}, \vec{x} \rangle - \lambda.$$

Dual function:

$$g(\lambda) = \begin{cases} -\langle \vec{b}, (A+\lambda I)^{\dagger} \vec{b} \rangle - \lambda & \text{if } A+\lambda I \succeq 0,\ b \in R(A+\lambda I), \\ -\infty, & \text{otherwise.} \end{cases}$$

Lagrange dual problem:

$$\max \ \langle -\vec{b}, (A+\lambda I)^{\dagger} \vec{b} \rangle - \lambda$$

$$\text{subject to} \quad A+\lambda I \succeq 0, \quad \vec{b} \in R(A+\lambda I).$$

~~Equivalently, max~~

Equivalent,
$$\max \quad \boxed{-\sum_{i=1}^{n} \frac{(q_i^T b)^2}{\lambda_i + \lambda} - \lambda}$$

subject to $\boxed{\lambda \geq -\lambda_{\max}(A)}$

$\lambda_i \to$ eigenvalue of A

$q_i \to$ corresponding (orthonormal) eigenvector of A.

$$\boxed{\frac{1}{\lambda_i + \lambda} = \infty \text{ if } \lambda = -\lambda_i}$$

Strong duality holds here as