Lecture #10

Minimize 2-norm of a matrix:

$$\min_{\underline{x} \in \mathbb{R}^n} \| A(\underline{x}) \|_2 = \lambda_{max}\left(A^T_{(\underline{x})} A_{(\underline{x})}\right)$$

Matrix 2 - norm (induced-2 norm)

(Matrix norm induced by vector norm)

$$\| \underbrace{A\underline{u}}_{\underline{v}} \|_2$$

$$\sup_{\underline{u} \neq \underline{0} \in \mathbb{R}^n} \frac{\|\underline{v}\|_2}{\|\underline{u}\|_2}$$

$$= \sup_{\|\underline{u}\|_2 = 1} \|\underline{v}\|_2$$

$$= \sup_{\|\underline{u}\|_2 = 1} \|A\underline{u}\|_2 = \sup_{\|\underline{u}\|=1} \|A\underline{u}\|_2$$

$= \lambda_{max}(A^T A)$ (from earlier lectures)

$$\|A\|_F^2 \quad \text{Frobenius norm}$$

$$= tr(A^T A)$$

$$= \sum_{i,j} a_{ij}^2$$

$$\|A\|_2 \le t \iff \|A\|_2^2 \le t^2$$

$$\iff A^TA \preceq t^2 I \ \& \ t \ge 0$$

(Schur complement lemma) (see Lecture 8, p. 10)

$$\|\underline{u}\|_2 \le t \iff u^Tu \le t^2$$

$$\iff \begin{bmatrix} tI & \underline{u}_{n\times 1} \\ \underline{u}^T_{1\times n} & t_{1\times 1} \end{bmatrix} \succeq 0$$

Prev. problem (minimize 2-norm of symmetric matrix) $\iff$

$$\boxed{\begin{aligned} &\min_{\underline{x},\, t} \ t \\ &\text{s.t.} \begin{bmatrix} tI & A \\ A^T & t \end{bmatrix} \succeq 0 \end{aligned}}$$

SDP

Feasibility Problem : find x

$$\text{s.t. } f_i(\underline{x}) \leq 0, \quad i=1,\dots,m$$
$$h_j(\underline{x}) = 0, \quad j=1,\dots,p.$$

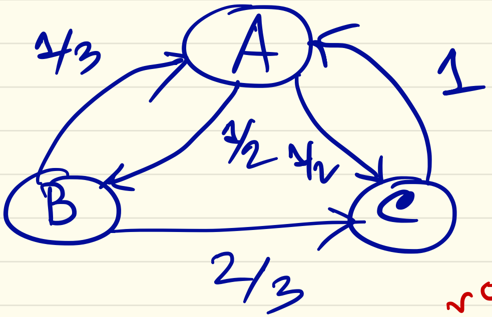
Slack Variable (s_i)

Inequality constraints $\longrightarrow$ Equality constraints + non-neg. constraints

$$f_i(x) \leq 0 \iff \begin{array}{l} s_i \geq 0 \\ f_i(\underline{x}) + s_i = 0, \\ \quad i=1,\dots,m \end{array}$$

We call s_i as "<u>slack variables</u>"

Search Engine Design (Application example)



$$\underbrace{\begin{array}{cc} & \begin{matrix} A & B & C \end{matrix} \\ \begin{matrix} A \\ B \\ C \end{matrix} & \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/3 & 0 & 2/3 \\ 1 & 0 & 0 \end{bmatrix} \end{array}}_{\text{State transition matrix}}$$

row vector

row vector

square matrix

$$\underbrace{\pi_{k+1}}_{1\times 3} = \underbrace{\pi_k}_{1\times 3} \underbrace{P}_{3\times 3}$$

(column form) $\Rightarrow$ $\pi_{k+1}^T = P^T \pi_k$

π_k

Occupation probability vector

$\mathbb{1}^T P = 1$ (row-sum = 1)

$\Rightarrow$ $\boxed{\pi_k = \pi_0 P^k}$

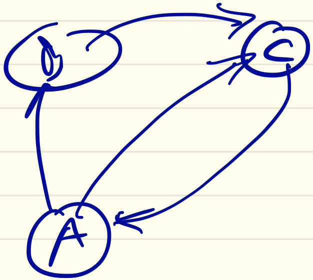
$\boxed{\pi_\infty = \pi_\infty P}$ Page rank

eig. vector for eig. value = 1

If $P = P^T$ then $\pi_\infty = (1/n)\mathbb{1}$ (uniform probability distribution)

↳ all eig. values are real

$$1 = \lambda_1(P) \geq \lambda_2(P) \geq \cdots \geq \lambda_n(P) \geq -1.$$



$$\underbrace{\mu(P)}_{\text{SLEM}} = \max_{i=2,\ldots,n} |\lambda_i(P)|$$

SLEM

(Second largest eig. value modulus) $= \max\{\lambda_2(P), -\lambda_n(P)\}$

→ can show that rate-of-convergence is governed by $\mu(P)$

$\therefore$ $\mu(P)$ small $\Longleftrightarrow$ fast convergence

$\Leftrightarrow$ Optimization problem (Design fastest mixing Markov chain)

$$\text{minimize } \mu(P)$$
$$\text{s.t. } P_{ij} \geq 0,\ P\mathbb{1} = \mathbb{1},\ P = P^T$$
$$P_{ij} = 0 \ \forall (i,j) \notin \mathcal{E} \ (\text{edge set})$$

$$\lambda_2(P) = \sup_{\|u\|_2 = 1} \{ \underline{u}^T P \underline{u} \mid \mathbb{1}^T \underline{u} = 0 \}$$

$$-\lambda_n(P) = \sup_{\|u\|_2 = 1} \{ -u^T P u \} \leftarrow \text{convex}$$

$$\therefore \text{SLEM} = \mu(P) = \max\{\lambda_2(P), -\lambda_n(P)\}$$
$$= \text{convex in } P$$

$\Updownarrow$

(SDP)
$$\boxed{\begin{aligned} &\min\ s \\ &\text{s.t. } \operatorname{diag}\left(P - \left(\tfrac{1}{n}\right)\mathbb{1}\mathbb{1}^T + sI,\ sI - P + \left(\tfrac{1}{n}\right)\mathbb{1}\mathbb{1}^T,\ \operatorname{vech}(P)\right) \succeq 0 \end{aligned}}$$

Matrix Calculus (How to take derivative of matrix valued functions)

$\underbrace{f}_{\text{convex } f^n} : \mathbb{R}^{n \times n} \mapsto \mathbb{R}$ (e.g. $tr(\cdot)$, $det(\cdot)$)

Question: How to consistently solve/find $\frac{\partial f}{\partial X}$

Fundamental: Directional derivative (for vectors)

$$D_{\underline{z}} f(\underline{x}) = \boxed{\lim_{h \to 0} \frac{f(\underline{x} + h\underline{z}) - f(\underline{x})}{h}}$$

Derivative/gradient of $f(\cdot)$ at $\underline{x}$ in the direction $\underline{z}$

$$= \boxed{\langle \nabla f(\underline{x}), \underline{z} \rangle}$$

$$= \cancel{\|\nabla f\| \, \|\underline{z}\| \cos\theta}$$

$$= (\nabla f)^T \underline{z} \quad \text{(vector inner product)}$$

$$\langle \nabla f, \underline{z} \rangle \quad \left\langle \frac{\partial f}{\partial X}, Z \right\rangle = tr\left(\left(\frac{\partial f}{\partial X}\right)^T Z \right) \quad \text{(matrix inner product)}$$

Example: $f(X) = \operatorname{tr}(AX)$

$$\underbrace{D_Z f(X)}_{\langle \frac{\partial f}{\partial x}, Z\rangle} = \lim_{h\to 0} \frac{\operatorname{tr}(A(X+hZ)) - \operatorname{tr}(AX)}{h}$$

$$\langle \frac{\partial f}{\partial x}, Z\rangle = \boxed{\operatorname{tr}\left(\left(\frac{\partial f}{\partial x}\right)^T Z\right)}$$

$$= \lim_{h\to 0} \frac{\operatorname{tr}(AhZ)}{h}$$

$$= \lim_{h\to 0} \frac{\cancel{h}}{\cancel{h}} \operatorname{tr}(AZ)$$

$$= \boxed{\operatorname{tr}(AZ)}$$

$$\Downarrow$$

$$\left(\frac{\partial f}{\partial x}\right)^T = A \iff \boxed{\frac{\partial f}{\partial x} = A^T}$$