MN1

1.

(Lec 9)

(Recht, Fazel + Parrilo) p. 477–482

NOT NEC SYM.

Matrix Norms

Given matrix norm $\|\cdot\|$ on $\mathbb{R}^{m\times n}$, the dual norm $\|\cdot\|_d$ is defined as

$$\|X\|_d = \sup\{\langle X,Y\rangle : Y\in\mathbb{R}^{m\times n},\ \|Y\|\le 1\}$$
$$(\langle X,Y\rangle = \operatorname{tr} X^T Y)$$

For vector p-norms; the dual of lp norm is lq norm, with $\frac{1}{p}+\frac{1}{q}=1$ (Hölder's inequality) and dual of l_1 norm is l_∞ norm.

Consider $\|X\| = \|X\|_F = \langle X,X\rangle^{1/2} = (\operatorname{tr} X^T X)^{1/2}$

Then $\|X\|_d = \|X\|_F$ (just as dual of l_2 is l_2).

How about the dual of $\underbrace{\|X\|_2}_{=\text{MAX } \sigma_i(X)}$? (operator norm, spectral norm).

Thm The dual of $\|\cdot\|_2$ is the NUCLEAR NORM

$$\|X\|_* = \sum_{i=1} \sigma_i(X)$$ (Schatten 1-norm) ("trace" norm)

To prove this we'll characterize $\|X\|_2$ and $\|X\|_*$ by SDPs.

FIRST: Characterization of $\|Z\|_2$:

$$\|Z\|_2 \le t \Leftrightarrow t^2 I_m - ZZ^T \succeq 0 \Leftrightarrow t^2 I_n - Z^T Z \succeq 0$$

$$\Leftrightarrow \begin{bmatrix} tI_m & Z \\ Z^T & tI_n \end{bmatrix} \succeq 0 \Leftrightarrow \begin{bmatrix} tI_n & Z^T \\ Z & tI_m \end{bmatrix} \succeq 0$$

MN2
2.

Pf Use question 1 in HW
or use Schur complement (see BV p.650):

Let $M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$.

If $A \succ 0$, then Schur complement $S = C - B^T A^{-1} B$
$\succeq 0$ iff $M \succeq 0$.

(block Gauss elimination: subtract $B^T A^{-1}$ * 1st row from 2nd row:
$\begin{bmatrix} A & B \\ 0 & -B^T A^{-1} B + C \end{bmatrix}$

Hence
$\|Z\|_2 = \min \left\{ t : \begin{bmatrix} tI & Z \\ Z^T & tI \end{bmatrix} \succeq 0 \right\}$
an SDP.

Characterization of $\|X\|_*$.

Let $X = U \Sigma V^T$

$m \left[\underset{r}{} \right] \left[\right] \left[\underset{r}{} \right] n$

U $m \times r$ $U^T U = I$
V $n \times r$ $V^T V = I$
Σ $r \times r$ diagonal.
$r = \mathrm{rank}(X)$.

Then by def'n, $\|X\|_* = \mathrm{tr}\, \Sigma$.

Note $\|UV^T\|_2 = \max_{\|q\|_2 = 1} \|UV^T q\| = 1$

(the SVD of UV^T is UIV^T).

So

$$\|X\|_{2,d} = \sup \{ \langle X, Y \rangle : \|Y\|_2 \le 1 \}$$
$$\ge \mathrm{tr}\, X^T U V^T = \mathrm{tr}\, V \Sigma \cancel{U^T U} V^T$$
$$= \mathrm{tr}\, \Sigma V^T V$$
$$= \mathrm{tr}\, \Sigma = \|X\|_*.$$

MN3. To find $\|X\|_{2,d}$ we need

$$\max_{Y} \langle X, Y\rangle \quad \text{s.t. } \|Y\|_2 \le 1$$

MINUS SIGN IS FOR CONVENIENCE BELOW

$$\equiv \quad \max_{Y} \operatorname{tr} X^T Y$$

(D) $$\text{s.t. } \underbrace{\begin{bmatrix} I_m & -Y \\ -Y^T & I_n \end{bmatrix}}_{\in S^{m+n}} \succeq 0$$

Var. $Y \in \mathbb{R}^{m\times n}$
Fixed: $X \in \mathbb{R}^{m\times n}$

This is an SDP in Dual form: Let's write $B \equiv X$

(D') $$\max \sum_{\substack{i=1,\dots,m\\ j=1,\dots,n}} b_{ij}\, y_{ij}$$

$$\text{s.t. } \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} - \sum_{\substack{i=1,\dots,m\\ j=1,\dots,n}} y_{ij} \underbrace{\begin{array}{c|c} & m+j \\ \hline i & \left[\begin{array}{c|c} & -1 \\ \hline -1 & \end{array}\right] \end{array}}_{2E_{ij} \in S^{m+n}} \succeq 0$$

(D'') i.e. $$\sum_{i,j} y_{ij} E_{ij} \preceq \frac{1}{2}\begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

The Primal SDP is

(P') $$\min_{W \in S^{m+n}} \frac{1}{2}\left\langle \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}, \overbrace{\begin{bmatrix} W_1 & W_3 \\ W_3^T & W_2 \end{bmatrix}}^{W} \right\rangle$$

$$\text{s.t. } \langle E_{ij}, W\rangle = b_{ij} \quad i=1,\dots,m,\ j=1,\dots,n$$

$$W \succeq 0$$

$$\langle E_{ij}, W\rangle = (W_3)_{ij}, \quad b_{ij} \equiv x_{ij}$$

MN4.

This has the feasible point

$$W = \begin{bmatrix} U\Sigma U^T & U\Sigma V^T \\ V\Sigma U^T & V\Sigma V^T \end{bmatrix} = \begin{bmatrix} U \\ V \end{bmatrix} \Sigma \begin{bmatrix} U^T & V^T \end{bmatrix}$$

$$\succeq 0$$

since $W_3 = U\Sigma V^T \equiv X \equiv B$.

Also the corresponding primal objective value is

$$\tfrac{1}{2}(\operatorname{tr} W_1 + \operatorname{tr} W_2) = \operatorname{tr} \Sigma = \|X\|_*.$$

Now any feasible point for (P) is an upper bound for the optimal solution of (D); so

$$\|X\|_{2,d} \leq \|X\|_*.$$

Combining this with $\|X\|_{2,d} \geq \|X\|_*$ (p. MN2)

we have $\|X\|_{2,d} = \|X\|_*$.

Hence by SDP duality (as (P), (D) both have strictly feasible points),

$$\|X\|_* \equiv \text{solution of the SDP}$$

$$\text{(P)} \quad \min_{\substack{W_1 \in S^m \\ W_2 \in S^n}} \tfrac{1}{2}(\operatorname{tr} W_1 + \operatorname{tr} W_2) \quad \text{s.t.} \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0.$$

MN5. <u>Matrix Completion</u>

The Netflix Problem.

Given $X = \Big[\qquad\qquad \Big]$

only some entries known

Believe X to be low rank.

Would like to solve

$$\min_{X \in \mathbb{R}^{m \times n}} \operatorname{rank}(X)$$
$$\text{s.t. } X_{ij} = m_{ij} \quad (i,j) \in \Omega$$

(given values)

NP-hard!

But, just as l_1 minimization for vectors "encourages" sparsity, nuclear norm minimization for matrices "encourages" low rank — so solve

$$\min_{X} \|X\|_*$$
$$\text{s.t. } X_{ij} = m_{ij} \quad (i,j) \in \Omega$$

i.e., the <u>SDP</u>

$$\min_{\substack{W_1 \in S^m \\ W_2 \in S^n \\ X \in \mathbb{R}^{m \times n}}} \tfrac{1}{2}(\operatorname{tr} W_1 + \operatorname{tr} W_2)$$
$$\text{s.t. } \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0$$
$$X_{ij} = m_{ij} \in \Omega.$$

MN6

More generally

$$\min_{X \in \mathbb{R}^{m \times n}} \operatorname{rank}(X)$$
$$\text{s.t. } \mathcal{A}(X) = b \qquad \mathcal{A}: \mathbb{R}^{m \times n} \to \mathbb{R}^p$$

← LINEAR MAP.

Can represent as

$$\langle A_k, X \rangle = b_k \quad i = 1, \dots, p$$

Earlier case

$$A_k = \begin{bmatrix} & \vdots & \\ & 1 & \\ & \vdots & \end{bmatrix} i$$

(like constraints of primal standard form SDP, except there A_i, X are symmetric; here they are not).

N.N. Relaxation:

Primal: $$\min_{\substack{X \in \mathbb{R}^{m \times n} \\ W_1 \in S^m \\ W_2 \in S^n}} \tfrac{1}{2}(\operatorname{tr} W_1 + \operatorname{tr} W_2) \quad \text{s.t. } \begin{bmatrix} W_1 & X \\ X^T & W_2 \end{bmatrix} \succeq 0$$

$$\& \ \mathcal{A}(X) = b$$

Dual: see HW.

$$\max b^T z$$
$$\text{s.t. } \begin{bmatrix} I_m & A^*(z) \\ (A^*(z))^T & I_n \end{bmatrix} \succeq 0$$

where

$$A^*(z) = \sum_{k=1}^{p} z_k A_k.$$

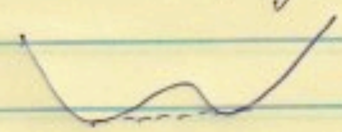
Lecture continues at bottom of p. 478 of Recht, Fazel & Parrilo

MN7

The Convex Envelope

Let $f : C \to \mathbb{R}$
C — convex set in $\mathbb{R}^n$.

not convex
The convex envelope of f is the pointwise max of all convex functions that lie below f.



Clearly, for $\|X\|_2 \le 1$, we have

$$\operatorname{rank}(X) \ge \|X\|_* = \sigma_1 + \dots + \sigma_r$$

(sum of r numbers each ≤ 1)

So $\|\cdot\|_*$ is a convex lower bound for the rank on the unit ball $\|\cdot\|_2 \le 1$

Thm $\|\cdot\|_*$ is the convex envelope of the rank on the unit ball.

Furthermore, for $\|X\|_2 \le M$ we have

$$\operatorname{rank}(X) \ge \frac{\|X\|_*}{M} = \frac{\sigma_1 + \dots + \sigma_r}{M}$$

& $\frac{\|\cdot\|_*}{M}$ is the convex envelope of the rank on $\{X : \|X\|_2 \le M\}$.

Now let $X_0 = \arg\min\{\operatorname{rank}(X) : A(X) = b\}$ (1)

$X_* = \arg\min\{\|X\|_* : A(X) = b\}$, (2)

and suppose $\|X_0\|_2 = M$.

$$\frac{\|X_*\|_*}{M} \underset{\text{by (2)}}{\le} \frac{\|X_0\|_*}{M} \underset{\text{by convex envelope}}{\le} \operatorname{rank}(X_0) \underset{\text{by (1)}}{\le} \operatorname{rank}(X_*)$$

MW8

Additivity of Rank & Nuclear Norm.

f is subadditive if $f(x+y) \le f(x) + f(y)$.

For vectors, cardinality (# nonzeros) (l_0 norm) and l_1 norm are subadditive.

$$\|x+y\|_1 \le \|x\|_1 + \|y\|_1$$

f is additive if $f(x+y) = f(x) + f(y)$.

l_0 is additive when x, y have disjoint support

l_1 " " " " " " " " .

For matrices, rank is subadditive (& so is $\|\cdot\|$)

& rank is additive for A, B

$$\Longleftrightarrow \begin{cases} R(A) \cap R(B) = \{0\} \\ R(A^T) \cap R(B^T) = \{0\} \end{cases} \quad (1972)$$

$\|\cdot\|_*$ is additive for A, B.

$$\Longleftarrow \begin{cases} R(A) \perp R(B) \\ R(A^T) \perp R(B^T) \end{cases} \quad (A^TB = 0 \ \& \ AB^T = 0)$$

Pf Consider the SVD's

$$\underset{m\times n}{A} = \underset{m\times r}{U_A} \underset{r\times r}{\Sigma_A} \underset{r\times n}{V_A^T} \qquad \Sigma_A > 0$$

i.e. rank$(A) = r$

$U_A^T U_A = I_r$

$V_A^T V_A = I_r$

$$B = U_B \Sigma_B V_B^T$$

$$V_A^T V_B = \Sigma_A^{-1} U_A^T \underbrace{A B^T}_{0} U_B \Sigma_B^{-1} = 0$$

Likewise $U_A^T U_B = 0$ because $A^T B = 0$

So $A+B = \begin{bmatrix} U_A & U_B \end{bmatrix} \begin{bmatrix} \Sigma_A & \\ & \Sigma_B \end{bmatrix} \begin{bmatrix} V_A & V_B \end{bmatrix}^T$

is a *valid* SVD of $A+B$, so $\|A+B\|_* = \|A\|_* + \|B\|_*$.

*(valid as $[U_A \ U_B]$ has orthonormal columns; also $[V_A \ V_B]$)

(Still following Recht, Fazel + Parrilo)

MN9. <u>Restricted Isometry Property + Low Rank Recovery</u>

Let $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$.
Let X_0 have rank r, + let $b = \mathcal{A}(X_0)$.

Define $X_* = \arg\min_X \{\|X\|_* : \mathcal{A}(X) = b\}$

Want to provide conditions under which it can be guaranteed that $X_* = X_0$.

<u>RIP</u> (Restricted Isometry Property)
(adapted from Candes + Tao from Compressed sensing: approximating ℓ_0 by ℓ_1)

Def: Let $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$, $m \le n$. For every integer r with $1 \le r \le m$, define the RIP constant $\delta_r(A)$ as the smallest no. s.t.

$$(1 - \delta_r(\mathcal{A}))\|X\|_F \le \|\mathcal{A}(X)\|_2 \le (1 + \delta_r(\mathcal{A}))\|X\|_F$$

holds for all matrices X of rank at most r.

<u>Extreme Examples</u>

(1) $\mathcal{A}(X) = [x_{11} \ldots x_{1n}, \ldots, x_{mn}]^T \equiv \mathrm{vec}(X)$
so $\|\mathcal{A}(X)\|_2 = \|X\|_F$ so $\delta_r = 0$

(2) $\mathcal{A}(X) = x_{11}$. Tells us very little about X.
Even for $r=1$, $m=2$ there is no $\delta_r(\mathcal{A}) < 1$.

For other examples, see annotated copy of RF&P.

MN10. Note: $\delta_r(A) \leq \delta_{r'}(A)$ if $r \leq r'$

Thm 1

Suppose $\delta_{2r} < 1$ for some $r \geq 1$. Then X_0 is the only matrix of rank $\leq r$ with $A(X) = b$.

Pf If not, let X have rank r, $X \neq X_0$, $A(X) = b$. Let $Z = X_0 - X$. We have $\text{rank}(Z) \leq 2r$, by rank subadditivity, with $A(Z) = 0$. Then

$$0 = \|A(Z)\| \geq (1 - \delta_{2r}(A)) \|Z\|_F > 0$$

contradiction.

Thm 2 Suppose $r \geq 1$ is such that $\delta_{5r}(A) < \frac{1}{10}$. Then $X_* = X_0$

$\leftarrow \min \|\cdot\|_*$ over $A(X) = b$: see top of p. MN9.

Pf. takes 2 pages.

Relies on Lemma

Lemma Let $A, B \in \mathbb{R}^{m \times n}$. Then $\exists B_1, B_2$ s.t.

1. $B = B_1 + B_2$
2. $\text{rank}(B_1) \leq 2\,\text{rank}(A)$
3. AB_2^T and $A^T B_2$ equal 0.

   (so $\|A + B\|_* = \|A\|_* + \|B\|_*$
4. $\text{tr}\, B_1^T B_2 = 0$

   (i.e. $\langle B_1, B_2 \rangle = 0$).

Pf: see next page.

MN11

<u>Pf of lemma</u>

Let $A = \underset{m\times m}{U} \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \underset{n\times n}{V^T}$

Let $\hat{B} = U^T B V = \begin{bmatrix} \hat{B}_{11} & \hat{B}_{12} \\ \hat{B}_{21} & \hat{B}_{22} \end{bmatrix}$

Let $B_1 = U \begin{bmatrix} \hat{B}_{11} & \hat{B}_{12} \\ \hat{B}_{21} & 0 \end{bmatrix} V^T$,

$B_2 = U \begin{bmatrix} 0 & 0 \\ 0 & \hat{B}_{22} \end{bmatrix} V^T$.

Done.

Note: since $\delta_r(A) \le \delta_{5r}(A) < 1$,
Thm 1 <u>also</u> applies so X_0 is the <u>only</u>
matrix with rank $\le r$ satisfying $\mathcal{A}(X) = b$,
+ hence <u>is</u> the minimal rank solution