2: Cholesky and matrix norms

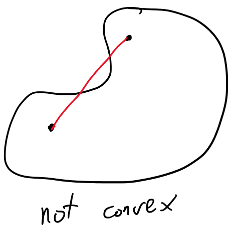
Sunday, September 6, 2020 11:11 PM

SPD matrices and Cholesky decomposition

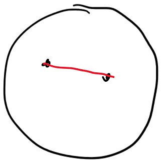
let S_{++}^n be the set of symmetric positive definite matrices

A few important properties

S_{++}^n is convex; i.e. if $A, B \in S_{++}^n$, then $\forall\ 0 \leq \lambda \leq 1$, $(1-\lambda)A + \lambda B \in S_{++}^n$

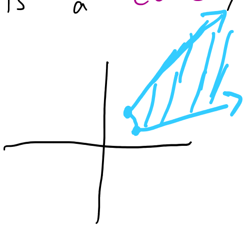


not convex



convex

S_{++}^n is a cone; i.e. if $\lambda > 0$, and $A \in S_{++}^n$, then $\lambda A \in S_{++}^n$



All rays from a starting point are in a cone

Prop. 1.7.5/1.7.9

If $A \in S_{++}^n$, then $A(1:k, 1:k) \in S_{++}^n$ for $k = 1, \ldots, n$.

proof. let $w \in \mathbb{R}^k$. Let $x = \begin{bmatrix} w \\ 0 \end{bmatrix} \in \mathbb{R}^n$.

Then $w^T A(1:k, 1:k) w = x^T A x > 0$.

▨

Thm 1.7.4/1.7.10 Let $A \in S_{++}^n$. Then $\exists$ real lower-triangular B s.t. $A = BB^T$.
Furthermore, B can be chosen so its diagonal elements are strictly positive.

Idea: Use $A = LDL^T$ decomposition, possible because all $A(1:k, 1:k)$ are invertible and A ~~B~~ symmetric.

If D is positive, then let $D^{\frac{1}{2}} = \begin{bmatrix} \sqrt{d_{11}} & & 0 \\ & \sqrt{d_{22}} & \\ 0 & & \ddots & \sqrt{d_{nn}} \end{bmatrix}$.

Then $A = LD^{\frac{1}{2}}D^{\frac{1}{2}}L^T$, so $B = LD^{\frac{1}{2}}$ works.

$L \quad U \qquad \sqrt{d_{nn}}\,]$

Then $A = LD^{\frac{1}{2}}D^{\frac{1}{2}}L^T$, so $B = LD^{\frac{1}{2}}$ works.

This almost works, and tells us that we could just reuse LU and would get the right answer but does not prove positivity.

Before we actually prove the theorem, we need to introduce Schur complements.

Consider finding the inverse of

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

using Gaussian elimination

$$\left(\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right)$$

$$\left(\begin{array}{cc|cc} 1 & b/a & 1/a & 0 \\ c & d & 0 & 1 \end{array}\right) \quad \text{Suppose } a \neq 0$$

$$\left(\begin{array}{cc|cc} 1 & b/a & 1/a & 0 \\ 0 & d-\frac{bc}{a} & -\frac{c}{a} & 1 \end{array}\right)$$

$$\left(\begin{array}{cc|cc} 1 & b/a & 1/a & 0 \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right)$$

$$\left(\begin{array}{cc|cc} 1 & 0 & \frac{1}{a}+\frac{bc}{a(ad-bc)} & \frac{-b}{ad-bc} \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right)$$

$$\Rightarrow \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

What about a 2×2 block matrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

$A \in \mathbb{R}^{p\times p}$
$B \in \mathbb{R}^{p\times q}$
$C \in \mathbb{R}^{q\times p}$
$D \in \mathbb{R}^{q\times q}$

$$\left(\begin{array}{cc|cc} A & B & I & 0 \\ C & D & 0 & I \end{array}\right)$$

Suppose A^{-1} exists.

$$\left(\begin{array}{cc|cc} I & A^{-1}B & A^{-1} & 0 \\ C & D & 0 & I \end{array}\right)$$

$$\left(\begin{array}{cc|cc} I & A^{-1}B & A^{-1} & 0 \\ 0 & D-CA^{-1}B & -CA^{-1} & I \end{array}\right)$$

$$\left(\begin{array}{cc|cc} I & A^{-1}B & A^{-1} & 0 \\ 0 & I & -(D-CA^{-1}B)CA^{-1} & (D-CA^{-1}B)^{-1} \end{array}\right)$$

$$\left(\begin{array}{cc|cc} I & 0 & A^{-1}+A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\ 0 & I & -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1} \end{array}\right)$$

$$\Rightarrow \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}B(D-CA^{-1}B) \\ 0 & (D-CA^{-1}B)^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}$$

$$= \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix}\begin{bmatrix} A^{-1} & 0 \\ 0 & (D-CA^{-1}B)^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix}$$

Prop. 2.7.1

If $\exists A^{-1}$ $\Rightarrow \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix}\begin{bmatrix} A & 0 \\ 0 & D-CA^{-1}B \end{bmatrix}\begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$

Note that we can directly check this expression, and only require invertibility of A.

Def 2.7.1 Given any block matrix $M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$ where A and D are square, if A is invertible, then $D - CA^{-1}B$ is the **Schur complement** of A in M.
if D " " " $A - BD^{-1}C$ " " " " of D in M.

if A is invertible, then $D-CA^{-1}B$ is the Schur complement
if D " " " $A-BD^{-1}C$ " " " " of D in M.

Prop. 2.7.2 If A, D and both Schur complements $A-BD^{-1}C$ and $D-CA^{-1}B$ are all invertible, then

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A-BD^{-1}C)^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1} \\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1} \end{pmatrix}$$

Matrix inversion lemma Setting $D=I$ and changing B to $-B$,

$$(A+BC)^{-1} = A^{-1} - A^{-1}B(I-CA^{-1}B)^{-1}CA^{-1}.$$

When we get to the optimization half of this class, we will see that we can often use Schur complements to convert nonlinear problems into semidefinite programs.

Back to Cholesky factorization and **Thm 1.7.4/1.7.10**

(if $A \in S^n_{++}$, then $A=BB^T$, where B is a unique lower triangular matrix with positive diagonal)

proof. Induction on the dim n of A.

Base case: $n=1$. Then $a_{11}>0$. Let $\alpha=\sqrt{a_{11}}$, $B=(\alpha)$.

Induction: $n \geq 2$. $a_{11}>0$, so

$$A = \begin{pmatrix} a_{11} & W^T \\ W & C \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha & 0 \\ W/\alpha & I \end{pmatrix}}_{B_1} \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & \underbrace{C - WW^T/a_{11}}_{\text{Schur complement}} \end{pmatrix}}_{A_1} \underbrace{\begin{pmatrix} \alpha & W^T/\alpha \\ 0 & I \end{pmatrix}}_{B_1^T} = B_1 A_1 B_1^T, \quad \alpha = \sqrt{a_{11}}.$$

Note A_1 is also symmetric positive definite because $A_1 = B_1^{-1}A(B_1^{-1})^T$.

$\Rightarrow C - \frac{WW^T}{a_{11}}$ is symm pos. def. (by restriction to vectors with $x_1=0$, $x \neq 0$.)

Thus, we can apply the induction hypo to the $(n-1)\times(n-1)$ $C - \frac{WW^T}{a_{11}}$.

$\Rightarrow C - WW^T/a_{11} = LL^T$, L unique with pos. diagonal entries

So, we get $A = \begin{pmatrix} \alpha & 0 \\ W/\alpha & I \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & LL^T \end{pmatrix} \begin{pmatrix} \alpha & W^T/\alpha \\ 0 & I \end{pmatrix}$

So, we get $$A = \begin{pmatrix} \alpha & 0 \\ w/\alpha & I \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & LL^T \end{pmatrix} \begin{pmatrix} \alpha & w/\alpha \\ 0 & I \end{pmatrix}$$
$$= \begin{pmatrix} \alpha & 0 \\ w/\alpha & I \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & L \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & L^T \end{pmatrix} \begin{pmatrix} \alpha & w^T/\alpha \\ 0 & I \end{pmatrix}$$
$$= \begin{pmatrix} \alpha & 0 \\ w/\alpha & L \end{pmatrix} \begin{pmatrix} \alpha & w^T/\alpha \\ 0 & L^T \end{pmatrix}$$

Let $B = \begin{pmatrix} \alpha & 0 \\ w/\alpha & L \end{pmatrix}$, a unique lower-triangular matrix with pos. diagonal and $A = BB^T$. ☐

Note: Cholesky is a special case of LU and uniqueness can be proven throug LU decompositions.

However, Cholesky requires half the number of operators / space and is also numerically stable.

Prop. 1.7.6 The following are equivalent for a symmetric n×n matrix

1.7.11

(1) $A \in S^n_{++}$

(2) Sylvester's criterion. All principal minors are positive. $\left(\det(A(1:k, 1:k)) > 0\right)$

(3) A has an LU factorization and all pivots are positive.

(4) A has an LDL^T factorization and all pivots in D are positive.

(5) All eigenvalues of A are strictly positive

Matrix norms

Def. 1.8.1 Let E be a vector space over a field K (e.g. $\mathbb{R}$ or $\mathbb{C}$).

A **norm** on E is a function $\| \ \| : E \to \mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$ s.t.

$\forall\ x, y, z \in E$ and $\lambda \in K$,

(N1) $\|x\| \geq 0$, $\|x\| = 0$ iff $x = 0$

(N2) $\|\lambda x\| = |\lambda| \, \|x\|$

(N3) $\|x+y\| \leq \|x\| + \|y\|$

$(E, \| \ \|)$ is a **normed vector space**.

Prop 1.8.1 If $E = \mathbb{C}^n$ or $E = \mathbb{R}^n$, for $p \geq 1$, the ℓ^p-norm

$$\|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}$$ is a norm.

Aside: $\lim_{p\to\infty} \|x\|_p \equiv \|x\|_\infty = \max_i |x_i|$ is a commonly used norm

The "zero-norm" $\|x\|_0$ denoting the number of nonzero components is often used, but it is not a norm.

Hölder's inequalities For any $p, q \in \mathbb{R}$ s.t. $p, q \geq 1$ and $\frac{1}{p} + \frac{1}{q} = 1$, (with $q = +\infty$ if $p = 1$ and $p = +\infty$ if $q = 1$),

$$\sum_{i=1}^n |u_i v_i| \leq \left(\sum_{i=1}^n |u_i|^p \right)^{1/p} \left(\sum_{i=1}^n |v_i|^q \right)^{1/q}$$

and $$|\langle u, v\rangle| \leq \|u\|_p \, \|v\|_q, \quad u, v \in \mathbb{C}^n$$

where $\langle\, ,\, \rangle$ is the Hermitian inner product

$$\langle u, v\rangle = \sum_{i=1}^n u_i \overline{v_i} .$$

Def 1.8.2 Given any real or complex vector space E, two norms are equivalent iff $\exists\, C_1, C_2 > 0$ s.t.

$$\|u\|_a \leq C_1 \|u\|_b \quad \text{and} \quad \|u\|_b \leq C_2 \|u\|_a, \quad \forall\, u \in E.$$

Aside: Two norms are equivalent iff they induce the same topology on E.

Thm 1.8.1/1.8.5 If E is any real or complex finite-dimensional vector space, then any two norms are equivalent.

proof. More of an analysis/topology result, so we won't cover the proof in this class.

Def. 1.8.3 A matrix norm $\|\ \|$ on the space of square $n \times n$ matrices in $K^{n\times n}$ with $K = \mathbb{R}$ or $K = \mathbb{C}$ is a norm on the vector space $K^{n\times n}$ with the additional property of that $\|AB\| \leq \|A\|\|B\| \quad \forall\, A, B \in K^{n\times n}$.

the vector space $K^{n\times n}$ with the additional property of submultiplicativity that $\|AB\| \le \|A\|\|B\|$, $\forall A, B \in K^{n\times n}$.
A norm satisfying the above prop. is often called a submultiplicative matrix norm.

Note: $\|I\| = \|I^2\| \le \|I\|^2 \Rightarrow \|I\| \ge 1$ for every matrix norm.

$\forall A \in \mathbb{C}^{n\times n}$, let $\lambda_1, \ldots, \lambda_n$ be the spectrum/eigenvalues of A

Def. The spectral radius $\rho(A) = \max_{1\le i\le n} |\lambda_i|$

Prop. 1.8.3/8.6 For any matrix norm $\|\ \|$ on $\mathbb{C}^{n\times n}$ and $A \in \mathbb{C}^{n\times n}$,

$$\rho(A) \le \|A\|$$

proof. Let λ be some eigenvalue of A s.t. $|\lambda| = \rho(A)$.
If $u \neq 0$ is an eigenvector associated with λ, let

$$U = [u \cdots u] \in \mathbb{C}^{n\times n},$$

(so $\|U\| \neq 0$)

Then $AU = \lambda U$.

$$\Rightarrow |\lambda|\|U\| = \|\lambda U\| = \|AU\| \le \|A\|\ \|U\|$$

$$\Rightarrow \rho(A) = |\lambda| \le \|A\|.$$

$\square$

Note: the proposition also holds for real matrices, but the proof is a little more complicated and we'll need the machinery of induced norms.

Def. 1.8.6 The Frobenius norm $\|\ \|_F$ is defined $\forall A \in \mathbb{C}^{n\times n}$,

$$\|A\|_F = \left(\sum_{i,j=1}^{n} |a_{ij}|^2\right)^{\frac{1}{2}} = \sqrt{tr(AA^*)} = \sqrt{tr(AA^*)}$$

Note that this is just the norm $\|\ \|_2$ on $\mathbb{C}^{n^2}$.

Prop. 1.8.4/8.6 The Frobenius norm satisfies the following

(1) It is a matrix norm, i.e. $\|AB\|_F \le \|A\|_F\|B\|_F$.

(2) It is unitarily invariant, i.e. $\forall$ unitary matrices U, V,

$$\|A\|_F = \|UA\|_F = \|AV\|_F = \|UAV\|_F.$$

Recall, U is unitary iff $UU^* = U^*U = I$

(3) $\sqrt{\rho(A^*A)} \le \|A\|_F \le \sqrt{n}\sqrt{\rho(A^*A)}$, $\forall A \in \mathbb{C}^{n\times n}$.

(3) $\sqrt{\rho(A^*A)} \le \|A\|_F \le \sqrt{n}\sqrt{\rho(A^*A)}$, $\forall A \in \mathbb{C}^{n\times n}$.

(Hölder's ineq. for $p=2$)

proof. (1) Apply Cauchy-Schwarz inequality

$$\|AB\|_F^2 = \sum_{i,j=1}^{n}\Big|\sum_{k=1}^{n} a_{ik}b_{kj}\Big|^2$$

$$\le \sum_{i,j=1}^{n}\Big(\sum_{k=1}^{n}|a_{ik}b_{kj}|\Big)^2$$

$$\le \sum_{i,j=1}^{n}\Big[\Big(\sum_{h=1}^{n}|a_{ih}|^2\Big)^{\frac{1}{2}}\Big(\sum_{k=1}^{n}|b_{kj}|^2\Big)^{\frac{1}{2}}\Big]^2$$

(Cauchy-Schwarz)

$$= \sum_{i,j=1}^{n}\Big(\sum_{h=1}^{n}|a_{ih}|^2\Big)\Big(\sum_{k=1}^{n}|b_{kj}|^2\Big)$$

$$= \Big(\sum_{i,h=1}^{n}|a_{ih}|^2\Big)\Big(\sum_{k,j=1}^{n}|b_{kj}|^2\Big) = \|A\|_F^2\,\|B\|_F^2 .$$

(2) $\|A\|_F^2 = tr(A^*A) = tr(VV^*A^*A) = tr(V^*A^*AV) = \|AV\|_F^2$.

$\uparrow$ similarity

$\|A\|_F^2 = tr(A^*A) = tr(A^*U^*UA) = \|UA\|_F^2$.

$\Rightarrow \|A\|_F = \|UAV\|_F$.

(3) Note that A^*A is Hermitian positive semi-definite, so it has non-negative real eigenvalues

(spectral theorems, Chapter 1.16)

$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n \ge 0$, $\rho(A^*A) = \lambda_1$.

$\Rightarrow \rho(A^*A) = \lambda_1 \le \sum_i \lambda_i = tr(A^*A) \le n\lambda_1 = n\rho(A^*A)$. $\boxed{\diagup}$

Subordinate norms

Prop. 1.8.5/8.8 For every norm $\|\ \|$ on $\mathbb{C}^n$ (or $\mathbb{R}^n$), for every $n\times n$ matrix A, $\exists\, C_A \ge 0$ s.t.

$$\|Au\| \le C_A\|u\|,$$

$\forall u \in \mathbb{C}^n$ (or $u \in \mathbb{R}^n$ if A is real)

proof. For every basis $(e_1,\dots,e_n)$ of $\mathbb{C}^n$, $\forall u = u_1e_1 + \cdots + u_ne_n$,

proof. For every basis $(e_1, \dots, e_n)$ of $\mathbb{C}^n$, $\forall\, u = u_1 e_1 + \dots + u_n e_n$,

$$\begin{aligned} \|Au\| &= \|u_1 A e_1 + \dots + u_n A e_n\| \\ &\le |u_1|\,\|Ae_1\| + \dots + |u_n|\,\|Ae_n\| \\ &\le C_1 (|u_1| + \dots + |u_n|) = C_1 \|u\|_1, \qquad C_1 = \max_i \|Ae_i\|. \end{aligned}$$

But $\| \ \|$ and $\| \ \|_1$ are equivalent, so $\exists\, C_2 > 0$ s.t.

$$\|u\|_1 \le C_2 \|u\| \implies \|Au\| \le C_A \|u\|, \qquad C_A = C_1 C_2.$$

Def. 8.7 If $\| \ \|$ is any norm on $\mathbb{C}^n$, we define the function $\| \ \|_{op}$ on matrices $\mathbb{C}^{n\times n}$ by

$$\|A\|_{op} = \sup_{\substack{x \in \mathbb{C}^n \\ x \ne 0}} \frac{\|Ax\|}{\|x\|} = \sup_{\substack{x \in \mathbb{C}^n \\ \|x\| = 1}} \|Ax\|$$

($\le C_A$
So finite)

$\| \ \|_{op}$ is called the subordinate matrix norm or induced operator norm induced by the norm $\| \ \|$.

Note: By definition, $\|Ax\| \le \|A\|_{op} \|x\|$ for all $x \in \mathbb{C}^n$.

So $\|ABx\| \le \|A\|_{op} \|Bx\| \le \|A\|_{op} \|B\|_{op} \|x\| \quad \forall x \in \mathbb{C}^n$

$\implies \|AB\|_{op} \le \|A\|_{op} \|B\|_{op}$ (by choosing x to maximize $\|ABx\|$)

Thus, $\| \ \|_{op}$ has submultiplicativity and is a matrix norm.

Def. 8.8 If $K = \mathbb{R}$ or $K = \mathbb{C}$, for any norm $\| \ \|$ on $K^{m\times n}$, and for any two norms $\| \ \|_a$ on K^n and $\| \ \|_b$ on K^m, we say that $\| \ \|$ is subordinate to the norms $\| \ \|_a$, $\| \ \|_b$ if

$$\|Ax\|_b \le \|A\| \, \|x\|_a \quad \forall\, A \in K^{m\times n} \text{ and all } x \in K^n.$$

Prop 8.7/8.10 For every square matrix $A = (a_{ij}) \in \mathbb{C}^{n\times n}$, we have

$$\|A\|_1 = \sup_{\substack{x \in \mathbb{C}^n \\ \|x\|_1 = 1}} \|Ax\|_1 = \max_j \sum_{i=1}^n |a_{ij}|$$

(max of l^1-norm of columns)

(Note, we are dropping op from the

(are dropping op from the notation $\| \|_{op}$)

$$x \in \mathbb{C}^n \qquad i=1$$
$$\|x\|_1 = 1$$

$$\|A\|_\infty = \sup_{\substack{x \in \mathbb{C}^n \\ \|x\|_\infty = 1}} \|Ax\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$$ (max of ℓ^1-norm of rows)

$$\|A\|_2 = \sup_{\substack{x \in \mathbb{C}^n \\ \|x\|_2 = 1}} \|Ax\|_2 = \sqrt{\rho(A^*A)} = \sqrt{\rho(AA^*)}.$$

(Note: $\|A\|_2 = \|A^*\|_2$ and $\|A\|_2 = \|UAV\|_2$, for U, V unitary matrices.
If A is normal, then $\|A\|_2 = \rho(A)$)
$(AA^* = A^*A)$

proof. $\forall u$, $\|Au\|_1 = \sum_i \left| \sum_j a_{ij} u_j \right| \le \sum_j |u_j| \sum_i |a_{ij}| \le \left(\max_j \sum_i |a_{ij}| \right) \|u\|_1$.

Alternately, $u = [0, \ldots, 1, \ldots, 0]$ $\uparrow$ pos $j_0 = \arg\max_j \sum_i |a_{ij}|$, proving the claim for $\| \|_1$.

Similarly, $\|Au\|_\infty = \max_i \left| \sum_j a_{ij} u_j \right| \le \left| \max_i \sum_j |a_{ij}| \right| \|u\|_\infty$.

Alternately, let $i_0 = \operatorname{argmax}_i \sum_j |a_{ij}|$, and let $u_j = \begin{cases} \dfrac{\overline{a_{i_0,j}}}{|a_{i_0,j}|} & \text{if } a_{i_0,j} \neq 0 \\ 1 & \text{if } a_{i_0,j} = 0. \end{cases}$

It can be shown that such a vector u works.

For $\| \|_2$, it's quite a bit harder.

$$\|A\|_2^2 = \sup_{\substack{x \in \mathbb{C}^n \\ x^*x = 1}} \|Ax\|_2^2 = \sup_{\substack{x \in \mathbb{C}^n \\ x^*x = 1}} x^*A^*Ax.$$

Since A^*A is symmetric, it has real eigenvalues and can be diagonalized w.r.t. a unitary matrix.

It turns out that this implies $x \mapsto x^*A^*Ax$ has a maximum on the sphere $x^*x = 1$, and that max value is $\rho(A^*A)$.

Proof of this fact is dependent on some quadratic ... later in the class
which we will get to

Proof of this fact is dependent on some quadratic optimization, which we will get to later in the class.

$$\Rightarrow \|A\|_2 = \sqrt{\rho(A^*A)}.$$

Claim: $\rho(A^*A) = \rho(AA^*)$.

proof. Assume $\rho(A^*A) > 0$. Then $\exists$ eigenvector $u \neq 0$ s.t.

$$A^*Au = \rho(A^*A)u.$$

$$\Rightarrow AA^*(Au) = \rho(A^*A)(Au)$$

$\Rightarrow \rho(A^*A)$ is an eigenvalue of AA^*.

$$\Rightarrow \rho(A^*A) \leq \rho(AA^*).$$

Similarly, $\rho(AA^*) \leq \rho(A^*A) \Rightarrow \rho(AA^*) = \rho(A^*A)$ ▨

▨