Introduction

Tsung-Ming Huang

Department of Mathematics
National Taiwan Normal University

September 8, 2011

Outline

Vectors and matrices

$A \in \mathbb{F}$ with

$$A = [a_{ij}] = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}, \quad \mathbb{F} = \mathbb{R} \text{ or } \mathbb{C}.$$

- **Product of matrices**: $C = AB$, where $c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$, $i = 1, \ldots, m$, $j = 1, \ldots, p$.
- **Transpose**: $C = A^T$, where $c_{ij} = a_{ji} \in \mathbb{R}$.
- **Conjugate transpose**: $C = A^*$ or $C = A^H$, where $c_{ij} = \bar{a}_{ji} \in \mathbb{C}$.
- **Differentiation**: Let $C = (c_{ij}(t))$. Then $\dot{C} = \frac{d}{dt} C = [\dot{c}_{ij}(t)]$.

- **Outer product** of $x \in \mathbb{F}^m$ and $y \in \mathbb{F}^n$:

$$xy^* = \begin{bmatrix} x_1\bar{y}_1 & \cdots & x_1\bar{y}_n \\ \vdots & \ddots & \vdots \\ x_m\bar{y}_1 & \cdots & x_m\bar{y}_n \end{bmatrix} \in \mathbb{F}^{m\times n}.$$

- **Inner product** of $x \in \mathbb{F}^n$ and $y \in \mathbb{F}^n$:

$$\langle y, x\rangle := x^T y = \sum_{i=1}^{n} x_i y_i = y^T x \in \mathbb{R},$$

$$\langle y, x\rangle := x^* y = \sum_{i=1}^{n} \bar{x}_i y_i = \overline{y^* x} \in \mathbb{C}.$$

- **Sherman-Morrison Formula:**
  Let $A \in \mathbb{R}^{n \times n}$ be nonsingular, u, $v \in \mathbb{R}^n$. If $v^T A^{-1} u \neq -1$, then

$$(A + uv^T)^{-1} = A^{-1} - A^{-1}uv^T A^{-1}/(1 + v^T A^{-1}u). \tag{1}$$

- **Sherman-Morrison-Woodbury Formula**:
  Let $A \in \mathbb{R}^{n \times n}$, be nonsingular U, $V \in \mathbb{R}^{n \times k}$. If $(I + V^T A^{-1} U)$ is invertible, then

$$(A + UV^T)^{-1} = A^{-1} - A^{-1}U(I + V^T A^{-1}U)^{-1}V^T A^{-1}.$$

Proof of (1)**:**

$$\begin{aligned}
& (A + uv^T)[A^{-1} - A^{-1}uv^T A^{-1}/(1 + v^T A^{-1}u)] \\
= \quad & I + \frac{1}{1 + v^T A^{-1}u}[uv^T A^{-1}(1 + v^T A^{-1}u) - uv^T A^{-1} - uv^T A^{-1}uv^T A^{-1}] \\
= \quad & I + \frac{1}{1 + v^T A^{-1}u}[u(v^T A^{-1}u)v^T A^{-1} - u(v^T A^{-1}u)v^T A^{-1}] \\
= \quad & I.
\end{aligned}$$

Example 1

$$\tilde{A} = \begin{bmatrix} 3 & -1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 & 2 \\ 0 & 0 & 4 & 1 & 1 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & -1 & 0 & 0 & 3 \end{bmatrix} = A + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \end{bmatrix},$$

where

$$A = \begin{bmatrix} 3 & -1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 2 & 2 \\ 0 & 0 & 4 & 1 & 1 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 3 \end{bmatrix}.$$

Rank and orthogonality

Let $A \in \mathbb{R}^{m\times n}$. Then

- $\mathcal{R}(A) = \{y \in \mathbb{R}^m \mid y = Ax \text{ for some } x \in \mathbb{R}^n\} \subseteq \mathbb{R}^m$ is the **range space of** A.
- $\mathcal{N}(A) = \{x \in \mathbb{R}^n \mid Ax = 0\} \subseteq \mathbb{R}^n$ is the **null space of** A.
- rank$(A) =$ dim$[\mathcal{R}(A)] =$ the number of maximal linearly independent columns of A.
- rank$(A) =$ rank(A^T).
- dim$(\mathcal{N}(A))+$ rank$(A) = n$.
- If $m = n$, then A is nonsingular $\Leftrightarrow \mathcal{N}(A) = \{0\} \Leftrightarrow$ rank$(A) = n$.

- Let $\{x_1, \cdots, x_p\}$ in $\mathbb{R}^n$. Then $\{x_1, \cdots, x_p\}$ is said to be **orthogonal** if
$$x_i^T x_j = 0, \quad \text{for } i \neq j$$
and **orthonormal** if
$$x_i^T x_j = \delta_{ij},$$
where $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ij} = 1$ if $i = j$.
- $S^\perp = \{y \in \mathbb{R}^m \mid y^T x = 0, \text{ for } x \in S\}$ = orthogonal complement of S.
- $\mathbb{R}^m = \mathcal{R}(A) \oplus \mathcal{N}(A^T)$.
- $\mathbb{R}^n = \mathcal{R}(A^T) \oplus \mathcal{N}(A)$.
- $\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$.
- $\mathcal{R}(A^T)^\perp = \mathcal{N}(A)$.

Special matrices

$A \in \mathbb{R}^{n\times n}$
Symmetric: $A^T = A$ skew-symmetric: $A^T = -A$ positive definite: $x^TAx > 0, x \neq 0$ non-negative definite: $x^TAx \geq 0$ indefinite: $(x^TAx)(y^TAy) < 0$, for some x, y orthogonal: $A^TA = I_n$ normal: $A^TA = AA^T$ positive: $a_{ij} > 0$ non-negative: $a_{ij} \geq 0$.

$A \in \mathbb{C}^{n\times n}$
Hermitian: $A^* = A$ ($A^H = A$)
skew-Hermitian: $A^* = -A$
positive definite: $x^*Ax > 0, x \neq 0$
non-negative definite: $x^*Ax \geq 0$
indefinite: $(x^*Ax)(y^*Ay) < 0$, for some x, y
unitary: $A^*A = I_n$
normal: $A^*A = AA^*$

Let $A \in \mathbb{F}^{n\times n}$. Then the matrix A is

- **diagonal** if $a_{ij} = 0$, for $i \neq j$. Denote $D = diag(d_1, \cdots, d_n) \in \mathbf{D}_n$;
- **tridiagonal** if $a_{ij} = 0, |i-j| > 1$;
- **upper bi-diagonal** if $a_{ij} = 0, i > j$ or $j > i+1$;
- **(strictly) upper triangular** if $a_{ij} = 0, i > j \quad (i \geq j)$;
- **upper Hessenberg** if $a_{ij} = 0, i > j+1$.
  (Note: the lower case is the same as above.)

Sparse matrix: n^{1+r}, where $r < 1$ (usually between $0.2 \sim 0.5$). If $n = 1000$, $r = 0.9$, then $n^{1+r} = 501187$.

Eigenvalues and Eigenvectors

Definition 2

Let $A \in \mathbb{C}^{n\times n}$. Then $\lambda \in \mathbb{C}$ is called an **eigenvalue** of A, if there exists $x \neq 0$, $x \in \mathbb{C}^n$ with $Ax = \lambda x$ and x is called an **eigenvector** corresponding to λ.

Notations:

$\sigma(A) :=$ spectrum of $A =$ the set of eigenvalues of A.
$\rho(A) :=$ radius of $A = \max\{|\lambda| : \lambda \in \sigma(A)\}$.

- $\lambda \in \sigma(A) \Leftrightarrow det(A - \lambda I) = 0$.
- $p(\lambda) = det(\lambda I - A) =$ characteristic polynomial of A.
- $p(\lambda) = \prod_{i=1}^{s}(\lambda - \lambda_i)^{m(\lambda_i)}, \lambda_i \neq \lambda_j$ (for $i \neq j$) and $\sum_{i=1}^{s} m(\lambda_i) = n$.
- $m(\lambda_i) =$ algebraic multiplicity of λ_i.
- $n(\lambda_i) = n -$ rank$(A - \lambda_i I) =$ geometric multiplicity of λ_i.
- $1 \leq n(\lambda_i) \leq m(\lambda_i)$.

If there is some i such that $n(\lambda_i) < m(\lambda_i)$, then A is called degenerated.

The following statements are equivalent:

(1) There are n linearly independent eigenvectors;

(2) A is diagonalizable, i.e., there is a nonsingular matrix T such that $T^{-1}AT \in \mathbf{D}_n$;

(3) For each $\lambda \in \sigma(A)$, it holds $m(\lambda) = n(\lambda)$.

If A is degenerated, then eigenvectors plus principal vectors derive Jordan form.

Theorem 3 (Schur decomposition)

(1) *Let* $A \in \mathbb{C}^{n\times n}$*. There is a unitary matrix* U *such that* U^*AU *is upper triangular.*

(2) *Let* $A \in \mathbb{R}^{n\times n}$*. There is an orthogonal matrix* Q *such that* Q^TAQ *is quasi-upper triangular, i.e., an upper triangular matrix possibly with nonzero subdiagonal elements in non-consecutive positions.*

(3) A *is normal if and only if there is a unitary* U *such that* $U^*AU = D$ *is diagonal.*

(4) A *is Hermitian if and only if* A *is normal and* $\sigma(A) \subseteq \mathbb{R}$*.*

(5) A *is symmetric if and only if there is an orthogonal* U *such that* $U^TAU = D$ *is diagonal and* $\sigma(A) \subseteq \mathbb{R}$*.*

Norms and eigenvalues

Let X be a vectorspace over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

Definition 4 (Vector norms)

Let N be a real-valued function defined on X $(N : X \longrightarrow \mathbb{R}_+)$. Then N is a (vector) norm, if

N1: $N(\alpha x) = |\alpha| N(x)$, $\alpha \in \mathbb{F}$, for $x \in X$;

N2: $N(x + y) \le N(x) + N(y)$, for x, $y \in X$;

N3: $N(x) = 0$ if and only if $x = 0$.

The usual notation is $\|x\| = N(x)$.

Example 5

Let $X = \mathbb{C}^n$, $p \geq 1$. Then $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ is an l_p-norm. Especially,

$$\|x\|_1 = \sum_{i=1}^n |x_i| \qquad (\ l_1\text{-norm}),$$

$$\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2} \qquad (\ \text{Euclidean-norm}),$$

$$\|x\|_\infty = \max_{1\leq i\leq n} |x_i| \qquad (\ \text{maximum-norm}).$$

Lemma 6

$N(x)$ is a continuous function in the components $x_1, \cdots, x_n$ of x.

Proof:

$$|N(x) - N(y)| \leq N(x-y) \leq \sum_{j=1}^{n} |x_j - y_j| \, N(e_j) \leq \|x - y\|_\infty \sum_{j=1}^{n} N(e_j).$$

$\square$

Theorem 7 (Equivalence of norms)

Let N and M be two norms on $\mathbb{C}^n$. Then there exist constants $c_1, c_2 > 0$ such that

$$c_1 M(x) \leq N(x) \leq c_2 M(x), \text{ for all } x \in \mathbb{C}^n.$$

▸ Proof of Theorem 7

Remark: Theorem 7 does not hold in infinite dimensional space.

Norms and eigenvalues

Definition 8 (Matrix-norms)

Let $A \in \mathbb{C}^{m\times n}$. A real value function $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R}_+$ satisfying

N1: $\|\alpha A\| = |\alpha|\|A\|$;

N2: $\|A + B\| \leq \|A\| + \|B\|$;

N3: $\|A\| = 0$ if and only if $A = 0$;

N4: $\|AB\| \leq \|A\|\|B\|$;

N5: $\|Ax\|_v \leq \|A\|\|x\|_v$.

If $\|\cdot\|$ satisfies N1 to N4, then it is called a matrix norm. In addition, matrix and vector norms are compatible for some $\|\cdot\|_v$ in N5.

Example 9 (Frobenius norm)

Let $\|A\|_F = \left\{\sum_{i,j=1}^{n} |a_{i,j}|^2\right\}^{1/2}$.

$$\begin{aligned}
\|AB\|_F &= \left(\sum_{i,j}\left|\sum_k a_{ik}b_{kj}\right|^2\right)^{1/2} \\
&\le \left(\sum_{i,j}\left\{\sum_k |a_{ik}|^2\right\}\left\{\sum_k |b_{kj}|^2\right\}\right)^{1/2} \quad \text{(Cauchy-Schwartz Ineq.)} \\
&= \left(\sum_i \sum_k |a_{ik}|^2\right)^{1/2}\left(\sum_j \sum_k |b_{kj}|^2\right)^{1/2} = \|A\|_F \|B\|_F .
\end{aligned}$$

This implies that N4 holds.

$$\|Ax\|_2 = \left(\sum_i\left|\sum_j a_{ij}x_j\right|^2\right)^{1/2} \le \left\{\sum_i\left(\sum_j |a_{ij}|^2\right)\left(\sum_j |x_j|^2\right)\right\}^{1/2} = \|A\|_F\|x\|_2. \tag{2}$$

This implies N5 holds. Also, N1, N2 and N3 hold obviously. ($\|I\|_F = \sqrt{n}$) $\square$

Example 10 (Operator norm)

Given a vector norm $\|\cdot\|$. An associated matrix norm is defined by

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{\|x\|=1} \{\|Ax\|\}.$$

N5 holds immediately. On the other hand,

$$\begin{aligned} \|(AB)x\| &= \|A(Bx)\| \leq \|A\| \, \|Bx\| \\ &\leq \|A\| \, \|B\| \, \|x\| \end{aligned}$$

for all $x \neq 0$. This implies that

$$\|AB\| \leq \|A\| \, \|B\|.$$

Thus, $N4$ holds. ($\|I\| = 1$). □

Three useful matrix norms:

$$\|A\|_1 = \sup_{x \neq 0} \frac{\|Ax\|_1}{\|x\|_1} = \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}| \tag{3}$$

$$\|A\|_\infty = \sup_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty} = \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}| \tag{4}$$

$$\|A\|_2 = \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sqrt{\rho(A^* A)} \tag{5}$$

Proof of (3)-(5)

Example 11 (Dual norm)

Let $\frac{1}{p} + \frac{1}{q} = 1$. Then $\|\cdot\|_p^* = \|\cdot\|_q$, $(p = \infty, q = 1)$. (It concluds from the application of the Hölder inequality, i.e. $|y^* x| \le \|x\|_p \|y\|_q$.)

Theorem 12

Let $A \in \mathbb{C}^{n\times n}$. Then for any operator norm $\|\cdot\|$, it holds

$$\rho(A) \leq \|A\|.$$

Moreover, for any $\epsilon > 0$, there exists an operator norm $\|\cdot\|_\epsilon$ such that

$$\|\cdot\|_\epsilon \leq \rho(A) + \epsilon.$$

▸ Proof of Theorem 12

Lemma 13

Let U and V are unitary. Then

$$\|UAV\|_F = \|A\|_F, \quad \|UAV\|_2 = \|A\|_2$$

From

$$\begin{aligned} \|UA\|_F &= \sqrt{\|Ua_1\|_2^2 + \cdots + \|Ua_n\|_2^2}, \\ \rho(A^*A) &= \rho(AA^*). \end{aligned}$$

Theorem 14 (Singular Value Decomposition (SVD))

Let $A \in \mathbb{C}^{m\times n}$. Then there exist unitary matrices $U = [u_1, \cdots, u_m] \in \mathbb{C}^{m\times m}$ and $V = [v_1, \cdots, v_n] \in \mathbb{C}^{n\times n}$ such that

$$U^*AV = diag(\sigma_1, \cdots, \sigma_p) = \Sigma, \tag{6}$$

where $p = \min\{m, n\}$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$. (Here, σ_i denotes the i-th largest singular value of A).

▸ Proof of Theorem 14

Remark: From (6), we have $\|A\|_2 = \sqrt{\rho(A^*A)} = \sigma_1$, which is the maximal singular value of A, and

$$\|ABC\|_F = \|U\Sigma V^*BC\|_F = \|\Sigma V^*BC\|_F \leq \sigma_1\|BC\|_F = \|A\|_2\|BC\|_F.$$

This implies

$$\|ABC\|_F \leq \|A\|_2\|B\|_F\|C\|_2. \tag{7}$$

In addition, by (2) and (7), we get

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\|A\|_2.$$

Theorem 15

Let $A \in \mathbb{C}^{n\times n}$. The statements are equivalent:

(1) $\lim_{m\to\infty} A^m = 0$;

(2) $\lim_{m\to\infty} A^m x = 0$ *for all* x;

(3) $\rho(A) < 1$.

Proof:

(1) $\Rightarrow$ (2): Trivial.

(2) $\Rightarrow$ (3): Let $\lambda \in \sigma(A)$, i.e., $Ax = \lambda x$, $x \neq 0$. This implies $A^m x = \lambda^m x \to 0$, as $\lambda^m \to 0$. Thus $|\lambda| < 1$, i.e., $\rho(A) < 1$.

(3) $\Rightarrow$ (1): There is a norm $\|\cdot\|$ with $\|A\| < 1$ (by Theorem 12). Therefore, $\|A^m\| \leq \|A\|^m \to 0$, i.e., $A^m \to 0$. $\square$

Theorem 16

$$\rho(A) = \lim_{k\to\infty} \left\|A^k\right\|^{1/k}.$$

Proof: Since

$$\rho(A)^k = \rho(A^k) \le \left\|A^k\right\| \;\Rightarrow\; \rho(A) \le \left\|A^k\right\|^{1/k},$$

for $k = 1, 2, \ldots$. If $\epsilon > 0$, then $\tilde{A} = [\rho(A) + \epsilon]^{-1} A$ has spectral radius < 1 and $\left\|\tilde{A}^k\right\| \to 0$ as $k \to \infty$. There is an $N = N(\epsilon, A)$ such that $\left\|\tilde{A}^k\right\| < 1$ for all $k \ge N$. Thus,

$$\left\|A^k\right\| \le [\rho(A) + \epsilon]^k, \text{ for all } k \ge N$$

or

$$\left\|A^k\right\|^{1/k} \le \rho(A) + \epsilon, \quad \text{for all } k \ge N.$$

Since $\rho(A) \le \left\|A^k\right\|^{1/k}$, and k, ϵ are arbitrary, $\lim_{k\to\infty} \left\|A^k\right\|^{1/k}$ exists and equals $\rho(A)$.

Theorem 17

Let $A \in \mathbb{C}^{n\times n}$, *and* $\rho(A) < 1$. *Then* $(I-A)^{-1}$ *exists and*

$$(I-A)^{-1} = I + A + A^2 + \cdots .$$

Proof: Since $\rho(A) < 1$, the eigenvalues of $(I-A)$ are nonzero. Therefore, by Theorem 15, $(I-A)^{-1}$ exists and

$$(I-A)(I + A + A^2 + \cdots + A^m) = I - A^m \to I.$$

$\square$

Corollary 18

If $\|A\| < 1$, *then* $(I-A)^{-1}$ *exists and*

$$\left\|(I-A)^{-1}\right\| \le \frac{1}{1-\|A\|}.$$

Proof: Since $\rho(A) \le \|A\| < 1$ (by Theorem 12),

$$\left\|(I-A)^{-1}\right\| = \left\|\sum_{i=0}^{\infty} A^i\right\| \le \sum_{i=0}^{\infty} \|A\|^i = (1-\|A\|)^{-1}.$$

Theorem 19 (without proof)

For $A \in \mathbb{F}^{n\times n}$ the following statements are equivalent:

(1) *There is a multiplicative norm p with $p(A^k) \leq 1, k = 1, 2, \ldots$.*

(2) *For each multiplicative norm p the power $p(A^k)$ are uniformly bounded, i.e., there exists a $M(p) < \infty$ such that $p(A^k) \leq M(p)$, $k = 0, 1, 2, \ldots$.*

(3) *$\rho(A) \leq 1$ and all eigenvalue λ with $|\lambda| = 1$ are not degenerated. (i.e., $m(\lambda) = n(\lambda)$.)*

(See Householder: The theory of matrix, pp.45-47.)

In the following, we prove some important inequalities of vector norms and matrix norms.

$$1 \leq \frac{\|x\|_p}{\|x\|_q} \leq n^{(q-p)/pq}, \quad (p \leq q). \tag{8}$$

▸ Proof of (8)

$$1 \leq \frac{\|x\|_p}{\|x\|_\infty} \leq n^{\frac{1}{p}}. \tag{9}$$

▸ Proof of (9)

$$\max_{1 \leq j \leq n} \|a_j\|_p \leq \|A\|_p \leq n^{(p-1)/p} \max_{1 \leq j \leq n} \|a_j\|_p, \tag{10}$$

where $A = [a_1, \cdots, a_n] \in \mathbb{R}^{m \times n}$.

▸ Proof of (10)

$$\max_{i,j} |a_{ij}| \leq \|A\|_p \leq n^{(p-1)/p} m^{1/p} \max_{i,j} |a_{ij}|\,, \tag{11}$$

where $A \in \mathbb{R}^{m \times n}$.

Proof of (11)**:** By (9) and (10) immediately. □

$$m^{(1-p)/p} \|A\|_1 \leq \|A\|_p \leq n^{(p-1)/p} \|A\|_1\,. \tag{12}$$

Proof of (12)**:** By (10) and (8) immediately. □

Hölder inequality:

$$\left|x^T y\right| \leq \|x\|_p \|y\|_q \text{, where } \frac{1}{p} + \frac{1}{q} = 1. \tag{13}$$

Proof of (13)**:** Let $\alpha_i = \frac{x_i}{\|x\|_p}$, $\beta_i = \frac{y_i}{\|y\|_q}$. Then

$$(\alpha_i^p)^{1/p}(\beta_i^q)^{1/q} \leq \frac{1}{p}\alpha_i^p + \frac{1}{q}\beta_i^q. \quad \text{(Jensen Inequality)}$$

Since $\|\alpha\|_p = 1$, $\|\beta\|_q = 1$, it follows that

$$\sum_{i=1}^{n} \alpha_i \beta_i \leq \frac{1}{p} + \frac{1}{q} = 1.$$

Then we have $\left|x^T y\right| \leq \|x\|_p \|y\|_q$. $\square$

$$\max\{|x^T y| : \|x\|_p = 1\} = \|y\|_q . \tag{14}$$

Proof of (14)**:** Take $x_i = y_i^{q-1} / \|y\|_q^{q/p}$. Then we have

$$\|x\|_p^p = \frac{\sum |y_i|^q}{\|y\|_q^{q/p}} = \frac{\|y\|_q^q}{\|y\|_q^{q/p}} = 1. \quad (\because (q-1)p = 1)$$

It follows

$$\left| \sum_{i=1}^{n} x_i^T y_i \right| = \frac{\sum |y_i|^q}{\|y\|_q^{q/p}} = \frac{\|y\|_q^q}{\|y\|_q^{q/p}} = \|y\|_q .$$

$\square$

Remark: $\exists \hat{z}$ with $\|\hat{z}\|_p = 1$ s.t. $\|y\|_q = \hat{z}^T y$. Let $z = \hat{z} / \|y\|_q$. Then we have $\exists z$ s.t. $z^T y = 1$ with $\|z\|_p = \frac{1}{\|y\|_q}$.

$$\|A\|_p = \left\|A^T\right\|_q \tag{15}$$

▸ Proof of (15)

$$n^{-\frac{1}{p}} \|A\|_\infty \leq \|A\|_p \leq m^{\frac{1}{p}} \|A\|_\infty . \tag{16}$$

▸ Proof of (16)

$$\|A\|_2 \leq \sqrt{\|A\|_p \|A\|_q}, \quad (\frac{1}{p} + \frac{1}{q} = 1). \tag{17}$$

▸ Proof of (17)

$$n^{(p-q)/pq} \|A\|_q \leq \|A\|_p \leq m^{(q-p)/pq} \|A\|_q , \tag{18}$$

where $A \in \mathbb{R}^{m \times n}$ and $q \geq p \geq 1$.

▸ Proof of (18)

Backward error and Forward error

Let $x = F(a)$. We define backward and forward errors in Figure 1. In Figure 1, $\hat{x} + \Delta x = F(a + \Delta a)$ is called a mixed forward-backward error, where $|\Delta x| \leq \varepsilon |x|$, $|\Delta a| \leq \eta |a|$.

Definition 20

(i) An algorithm is **backward stable**, if for all a, it produces a computed $\hat{x}$ with a small backward error, i.e., $\hat{x} = F(a + \Delta a)$ with Δa small.

(ii) An algorithm is **numerical stable**, if it is stable in the mixed forward-backward error sense, i.e., $\hat{x} + \Delta x = F(a + \Delta a)$ with both Δa and Δx small.

(iii) If a method which produces answers with forward errors of similar magnitude to those produced by a backward stable method, is called a **forward stable**.

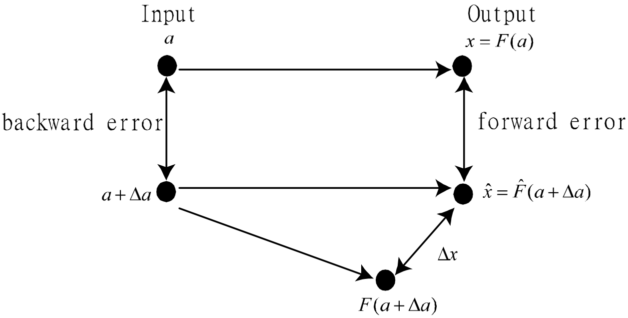

Figure: Relationship between backward and forward errors.

Remark:

(i) Backward stable $\Rightarrow$ forward stable, no vice versa!

(ii) Forward error $\leq$ condition number $\times$ backward error

Consider

$$\hat{x} - x = F(a + \Delta a) - F(a) = F'(a)\Delta a + \frac{F''(a + \theta\Delta a)}{2}(\Delta a)^2, \ \theta \in (0, 1).$$

Then we have

$$\frac{\hat{x} - x}{x} = \left(\frac{aF'(a)}{F(a)}\right)\frac{\Delta a}{a} + O\left((\Delta a)^2\right).$$

The quantity $C(a) = \left|\frac{aF'(a)}{F(a)}\right|$ is called the condition number of F. If x or F is a vector, then the condition number is defined in a similar way using norms and it measures the maximum relative change, which is attained for some, but not all Δa.

$$\text{Backward error:} \quad \begin{cases} \text{Àpriori error estimate !} \\ \text{Àposteriori error estimate !} \end{cases}$$

Lemma 21

$$\begin{cases} Ax = b \\ (A + \Delta A)\hat{x} = b + \Delta b \end{cases}$$

with $\|\Delta A\| \leq \delta \|A\|$ *and* $\|\Delta b\| \leq \delta \|b\|$. *If* $\delta\kappa(A) = r < 1$ *then* $A + \Delta A$ *is nonsingular and* $\frac{\|\hat{x}\|}{\|x\|} \leq \frac{1+r}{1-r}$.

Proof: Since $\|A^{-1}\Delta A\| < \delta \|A^{-1}\| \|A\| = r < 1$, it follows that $A + \Delta A$ is nonsingular. From $(I + A^{-1}\Delta A)\hat{x} = x + A^{-1}\Delta b$, we have

$$\begin{aligned} \|\hat{x}\| &\leq \left\|(I + A^{-1}\Delta A)^{-1}\right\| \left(\|x\| + \delta \left\|A^{-1}\right\| \|b\|\right) \\ &\leq \frac{1}{1-r}\left(\|x\| + \delta \left\|A^{-1}\right\| \|b\|\right) \\ &= \frac{1}{1-r}\left(\|x\| + r\frac{\|b\|}{\|A\|}\right) \end{aligned}$$

$\Rightarrow \|\hat{x}\| \leq \frac{1}{1-r}\left(\|x\| + r\|x\|\right)$. $\quad (\because \|b\| = \|Ax\| \leq \|A\| \|x\|)$ $\square$

Normwise Forward Error Bound

Theorem 22

If the condition of Lemma 21 hold, then

$$\frac{\|x-\hat{x}\|}{\|x\|} \le \frac{2\delta}{1-r}\kappa(A).$$

Proof: Since $\hat{x}-x = A^{-1}\Delta b - A^{-1}\Delta A\hat{x}$, we have

$$\|\hat{x}-x\| \le \delta \left\|A^{-1}\right\| \|b\| + \delta \left\|A^{-1}\right\| \|A\| \|\hat{x}\|.$$

So, by Lemma 21, we have

$$\begin{aligned}
\frac{\|\hat{x}-x\|}{\|x\|} &\le \delta\kappa(A)\frac{\|b\|}{\|A\|\,\|x\|} + \delta\kappa(A)\frac{\|\hat{x}\|}{\|x\|} \\
&\le \delta\kappa(A)\left(1+\frac{1+r}{1-r}\right) = \frac{2\delta}{1-r}\kappa(A).
\end{aligned}$$

Componentwise Forward Error Bounds

Theorem 23

Let $Ax = b$ and $(A + \Delta A)\hat{x} = b + \Delta b$. Let $|\Delta A| \leq \delta |A|$ and $|\Delta b| \leq \delta |b|$. If $\delta \kappa_\infty(A) = r < 1$ then $(A + \Delta A)$ is nonsingular and

$$\frac{\|\hat{x} - x\|_\infty}{\|x\|_\infty} \leq \frac{2\delta}{1 - r} \left\| |A^{-1}| \, |A| \right\|_\infty .$$

Proof: Since $\|\Delta A\|_\infty \leq \delta \|A\|_\infty$ and $\|\Delta b\|_\infty \leq \delta \|b\|_\infty$, the conditions of Lemma 21 are satisfied in ∞-norm. Then $A + \Delta A$ is nonsingular and $\frac{\|\hat{x}\|_\infty}{\|x\|_\infty} \leq \frac{1+r}{1-r}$.

Since $\hat{x} - x = A^{-1}\Delta b - A^{-1}\Delta A \hat{x}$, we have

$$\begin{aligned} |\hat{x} - x| &\leq \left|A^{-1}\right| |\Delta b| + \left|A^{-1}\right| |\Delta A| \, |\hat{x}| \\ &\leq \delta \left|A^{-1}\right| |b| + \delta \left|A^{-1}\right| |A| \, |\hat{x}| \leq \delta \left|A^{-1}\right| |A| \left(|x| + |\hat{x}|\right). \end{aligned}$$

Taking ∞-norm, we get

$$\begin{aligned}
\|\hat{x} - x\|_\infty &\leq \delta \left\| |A^{-1}| \, |A| \right\|_\infty \left(\|x\|_\infty + \frac{1+r}{1-r} \|x\|_\infty \right) \\
&= \frac{2\delta}{1-r} \, \| \underbrace{|A^{-1}| \, |A|}_{\text{Skeel condition number}} \|_\infty \, .
\end{aligned}$$

$\square$

Condition Number by First Order Approximation

$$(A+\epsilon F)x(\epsilon) = b+\epsilon f, \qquad x(0) = x$$

$$\dot{x}(0) = A^{-1}(f - Fx)$$

$$x(\epsilon) = x + \epsilon\dot{x}(0) + o(\epsilon^2)$$

$$\frac{\|x(\epsilon)-x\|}{\|x\|} \le \epsilon \left\|A^{-1}\right\| \left\{\frac{\|f\|}{\|x\|} + \|F\|\right\} + o(\epsilon^2)$$

Condition number $\kappa(A) := \|A\| \left\|A^{-1}\right\|$

$$\|b\| \le \|A\| \, \|x\| \, ,$$

$$\frac{\|x(\epsilon)-x\|}{\|x\|} \le \kappa(A)(\rho_A + \rho_b) + o(\epsilon^2).$$

$$\rho_A = \epsilon\frac{\|F\|}{\|A\|}, \quad \rho_b = \epsilon\frac{\|f\|}{\|b\|}, \quad \kappa_2(A) = \frac{\sigma_1(A)}{\sigma_n(A)}.$$

Normwise Backward Error Bound

Theorem 24

Let $\hat{x}$ be the computed solution of $Ax = b$. Then the normwise backward error bound

$$\eta(\hat{x}) := \min\left\{\epsilon | (A + \Delta A)\hat{x} = b + \Delta b, \quad \|\Delta A\| \leq \epsilon \|A\|, \quad \|\Delta b\| \leq \epsilon \|b\|\right\}$$

is given by

$$\eta(\hat{x}) = \frac{\|r\|}{\|A\| \, \|\hat{x}\| + \|b\|}, \tag{19}$$

where $r = b - A\hat{x}$ is the residual.

Proof: The right hand side of (19) is a upper bound of $\eta(\hat{x})$. This upper bound is attained for the perturbation (by construction)

$$\Delta A_{\min} = \frac{\|A\| \, \|\hat{x}\| \, r z^T}{\|A\| \, \|\hat{x}\| + \|b\|}, \quad \Delta b_{\min} = -\frac{\|b\|}{\|A\| \, \|\hat{x}\| + \|b\|} r,$$

where z is the dual vector of $\hat{x}$, i.e. $z^T \hat{x} = 1$ and $\|z\|_* = \frac{1}{\|\hat{x}\|}$.
Check:

$$\|\Delta A_{\min}\| = \eta(\hat{x}) \, \|A\| \, ,$$

or

$$\|\Delta A_{\min}\| = \frac{\|A\| \, \|\hat{x}\| \, \left\|r z^T\right\|}{\|A\| \, \|\hat{x}\| + \|b\|} = \left(\frac{\|r\|}{\|A\| \, \|\hat{x}\| + \|b\|} \right) \|A\| \, ,$$

i.e. claim

$$\left\| r z^T \right\| = \frac{\|r\|}{\|\hat{x}\|}.$$

Since

$$\left\| r z^T \right\| = \max_{\|u\|=1} \left\| (r z^T) u \right\| = \|r\| \max_{\|u\|=1} \left| z^T u \right| = \|r\| \, \|z\|_* = \|r\| \, \frac{1}{\|\hat{x}\|},$$

we have done. Similarly, $\|\Delta b_{\min}\| = \eta(\hat{x}) \, \|b\|$.

Componentwise Backward Error Bound

Theorem 25

The componentwise backward error bound

$$\omega(\hat{x}) := \min \{\epsilon | (A + \Delta A)\hat{x} = b + \Delta b, \quad |\Delta A| \le \epsilon |A|, \quad |\Delta b| \le \epsilon |b|\}$$

is given by

$$\omega(\hat{x}) = \max_i \frac{|r|_i}{(A|\hat{x}| + b)_i}, \tag{20}$$

where $r = b - A\hat{x}$. (note: $\xi/0 = 0$ if $\xi = 0$; $\xi/0 = \infty$ if $\xi \neq 0$.)

Proof: The right hand side of (20) is a upper bound for $\omega(\hat{x})$. This bound is attained for the perturbation

$$\Delta A = D_1 A D_2, \quad \Delta b = -D_1 b,$$

where

$$D_1 = \mathrm{diag}(r_i/(A|\hat{x}| + b)_i) \text{ and } D_2 = \mathrm{diag}(\mathrm{sign}(\hat{x}_i)).$$

Determinents and Nearness to Singularity

$$
\begin{aligned}
B_n &= \begin{bmatrix} 1 & -1 & \cdots & -1 \\ & 1 & \ddots & \vdots \\ & & 1 & -1 \\ 0 & & & 1 \end{bmatrix}, \quad B_n^{-1} = \begin{bmatrix} 1 & 1 & \cdots & 2^{n-2} \\ & \ddots & \ddots & \vdots \\ & & \ddots & 1 \\ 0 & & & 1 \end{bmatrix}, \\
\det(B_n) &= 1, \ \kappa_\infty(B_n) = n2^{n-1}, \ \sigma_n(B_n) \approx 10^{-8} (n = 30). \\
D_n &= \begin{bmatrix} 10^{-1} & & 0 \\ & \ddots & \\ 0 & & 10^{-1} \end{bmatrix}, \\
\det(D_n) &= 10^{-n}, \quad \kappa_p(D_n) = 1, \quad \sigma_n(D_n) = 10^{-1}.
\end{aligned}
$$

Appendix

Proof of Theorem 7: Without loss of generality (W.L.O.G.) we can assume that $M(x) = \|x\|_\infty$ and N is arbitrary. We claim

$$c_1\|x\|_\infty \leq N(x) \leq c_2\|x\|_\infty$$

or

$$c_1 \leq N(z) \leq c_2, \text{ for } z \in S = \{z \in \mathbf{C}^n | \|z\|_\infty = 1\}.$$

From Lemma 6, N is continuous on S (closed and bounded). By maximum and minimum principle, there are $c_1, c_2 \geq 0$ and $z_1, z_2 \in S$ such that

$$c_1 = N(z_1) \leq N(z) \leq N(z_2) = c_2.$$

If $c_1 = 0$, then $N(z_1) = 0$. Thus, $z_1 = 0$. This contradicts that $z_1 \in S$. $\square$

Return

Proof of (3):

$$\|Ax\|_1 = \sum_i \left| \sum_j a_{ij} x_j \right| \le \sum_i \sum_j |a_{ij}| \, |x_j| = \sum_j |x_j| \sum_i |a_{ij}| \, .$$

Let

$$\mathcal{C} := \sum_i |a_{ik}| = \max_j \sum_i |a_{ij}| \, .$$

Then $\|Ax\|_1 \le \mathcal{C} \, \|x\|_1$, thus $\|A\|_1 \le \mathcal{C}$. On the other hand, $\|e_k\|_1 = 1$ and $\|Ae_k\|_1 = \sum_{i=1}^n |a_{ik}| = \mathcal{C}$. $\square$

Proof of (4):

$$\begin{aligned} \|Ax\|_\infty &= \max_i \left| \sum_j a_{ij} x_j \right| \le \max_i \sum_j |a_{ij} x_j| \\ &\le \max_i \sum_j |a_{ij}| \, \|x\|_\infty \equiv \sum_j |a_{kj}| \, \|x\|_\infty \equiv \hat{\mathcal{C}} \, \|x\|_\infty . \end{aligned}$$

This implies, $\|A\|_\infty \le \hat{\mathcal{C}}$. If $A = 0$, then there is nothing to prove. Assume $A \neq 0$. Thus, the k-th row of A is nonzero. Define $z = [z_i] \in \mathbb{C}^n$ by

$$\begin{cases} z_i = \frac{\bar{a}_{ki}}{|a_{ki}|} & \text{if} \quad a_{ki} \neq 0, \\ z_i = 1 & \text{if} \quad a_{ki} = 0. \end{cases}$$

Then $\|z\|_\infty = 1$ and $a_{kj} z_j = |a_{kj}|$, for $j = 1, \ldots, n$. It follows

$$\|A\|_\infty \ge \|Az\|_\infty = \max_i \left| \sum_j a_{ij} z_j \right| \ge \left| \sum_j a_{kj} z_j \right| = \sum_{j=1}^n |a_{kj}| \equiv \hat{\mathcal{C}}.$$

Then, $\|A\|_\infty \ge \max\limits_{1 \le i \le n} \sum_{j=1}^n |a_{ij}| \equiv \hat{\mathcal{C}}$. □

Proof of (5)**:** Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$ be the eigenvalues of A^*A. There are muturally orthonormal vectors v_j, $j = 1, \ldots, n$ such that $(A^*A)v_j = \lambda_j v_j$. Let $x = \sum_j \alpha_j v_j$. Since $\|Ax\|_2^2 = (Ax, Ax) = (x, A^*Ax)$,

$$\|Ax\|_2^2 = \left(\sum_j \alpha_j v_j, \sum_j \alpha_j \lambda_j v_j \right) = \sum_j \lambda_j \left|\alpha_j\right|^2 \leq \lambda_1 \|x\|_2^2 .$$

Therefore, $\|A\|_2^2 \leq \lambda_1$. Equality follows by choosing $x = v_1$ and $\|Av_1\|_2^2 = (v_1, \lambda_1 v_1) = \lambda_1$. So, we have $\|A\|_2 = \sqrt{\rho(A^*A)}$. $\square$

Return

Proof of Theorem 12: Let $|\lambda| = \rho(A) \equiv \rho$ and x be the associated eigenvector with $\|x\| = 1$. Then,

$$\rho(A) = |\lambda| = \|\lambda x\| = \|Ax\| \leq \|A\| \, \|x\| = \|A\| \, .$$

<u>Claim</u>: $\|\cdot\|_\epsilon \leq \rho(A) + \epsilon$. There is a unitary U such that $A = U^* R U$, where R is upper triangular.
Let $D_t = diag(t, t^2, \cdots, t^n)$. For $t > 0$ large enough, the sum of all absolute values of the off-diagonal elements of $D_t R D_t^{-1}$ is less than ϵ. So, it holds $\left\|D_t R D_t^{-1}\right\|_1 \leq \rho(A) + \epsilon$ for large $t(\epsilon) > 0$. Define $\|\cdot\|_\epsilon$ for any B by

$$\begin{aligned} \|B\|_\epsilon &= \left\|D_t U B U^* D_t^{-1}\right\|_1 \\ &= \left\|(U D_t^{-1})^{-1} B (U D_t^{-1})\right\|_1 . \end{aligned}$$

This implies,

$$\|A\|_\epsilon = \left\|D_t R D_t^{-1}\right\| \leq \rho(A) + \epsilon.$$

Return

Proof of Theorem 14: There are $x \in \mathbb{C}^n$, $y \in \mathbb{C}^m$ with $\|x\|_2 = \|y\|_2 = 1$ such that $Ax = \sigma y$, where $\sigma = \|A\|_2$ $(\|A\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2)$. Let $V = [x, V_1] \in \mathbb{C}^{n\times n}$, and $U = [y, U_1] \in \mathbb{C}^{m\times m}$ be unitary. Then

$$A_1 \equiv U^* A V = \left[\begin{array}{cc} \sigma & w^* \\ 0 & B \end{array} \right].$$

Since $\left\| A_1 \begin{pmatrix} \sigma \\ w \end{pmatrix} \right\|_2^2 \geq (\sigma^2 + w^* w)^2$, it follows

$$\|A_1\|_2^2 \geq \sigma^2 + w^* w \quad \text{from} \quad \frac{\left\| A_1 \begin{pmatrix} \sigma \\ w \end{pmatrix} \right\|_2^2}{\left\| \begin{pmatrix} \sigma \\ w \end{pmatrix} \right\|_2^2} \geq \sigma^2 + w^* w.$$

But $\sigma^2 = \|A\|_2^2 = \|A_1\|_2^2$, it implies $w = 0$. Hence, the theorem holds by induction. $\square$

Return

Proof of (8)**:** Claim $\|x\|_q \leq \|x\|_p$, $(p \leq q)$: It holds

$$\|x\|_q = \left\| \|x\|_p \frac{x}{\|x\|_p} \right\|_q = \|x\|_p \left\| \frac{x}{\|x\|_p} \right\|_q \leq \mathcal{C}_{p,q} \|x\|_p ,$$

where

$$\mathcal{C}_{p,q} = \max_{\|e\|_p = 1} \|e\|_q , \quad e = (e_1, \cdots, e_n)^T.$$

We now show that $\mathcal{C}_{p,q} \leq 1$. From $p \leq q$, we have

$$\|e\|_q^q = \sum_{i=1}^{n} |e_i|^q \leq \sum_{i=1}^{n} |e_i|^p = 1 \quad (\text{by } |e_i| \leq 1).$$

Hence, $\mathcal{C}_{p,q} \leq 1$, thus $\|x\|_q \leq \|x\|_p$.

To prove the second inequality: Let $\alpha = q/p > 1$. Then the Jensen ineqality holds for the convex function $\varphi(x) \equiv x^{\alpha}$:

$$\int_{\Omega} |f|^q \, dx = \int_{\Omega} (|f|^p)^{q/p} \, dx \geq \left(\int_{\Omega} |f|^p \, dx \right)^{q/p}$$

with $|\Omega| = 1$. Consider the discrete measure $\sum_{i=1}^{n} \frac{1}{n} = 1$ and $f(i) = |x_i|$. It follows that

$$\sum_{i=1}^{n} |x_i|^q \frac{1}{n} \geq \left(\sum_{i=1}^{n} |x_i|^p \frac{1}{n} \right)^{q/p} .$$

Hence, we have

$$n^{-\frac{1}{q}} \|x\|_q \geq n^{-\frac{1}{p}} \|x\|_p .$$

Thus,

$$n^{(q-p)/pq} \|x\|_q \geq \|x\|_p .$$

Return

Proof of (9)**:** Let $q \to \infty$ and $\lim\limits_{q\to\infty} \|x\|_q = \|x\|_\infty$:

$$\|x\|_\infty = |x_k| = (|x_k|^q)^{\frac{1}{q}} \leq \left(\sum_{i=1}^{n} |x_i|^q \right)^{\frac{1}{q}} = \|x\|_q \,.$$

On the other hand,

$$\|x\|_q = \left(\sum_{i=1}^{n} |x_i|^q \right)^{\frac{1}{q}} \leq (n \|x\|_\infty^q)^{\frac{1}{q}} \leq n^{\frac{1}{q}} \|x\|_\infty \,.$$

It follows that $\lim\limits_{q\to\infty} \|x\|_q = \|x\|_\infty$. □

Return

To prove the second inequality: Let $\alpha = q/p > 1$. Then the Jensen ineqality holds for the convex function $\varphi(x) \equiv x^\alpha$:

$$\int_\Omega |f|^q \, dx = \int_\Omega (|f|^p)^{q/p} \, dx \geq \left(\int_\Omega |f|^p \, dx \right)^{q/p}$$

with $|\Omega| = 1$.
Consider the discrete measure $\sum_{i=1}^n \frac{1}{n} = 1$ and $f(i) = |x_i|$. It follows that

$$\sum_{i=1}^n |x_i|^q \frac{1}{n} \geq \left(\sum_{i=1}^n |x_i|^p \frac{1}{n} \right)^{q/p}.$$

Hence, we have

$$n^{-\frac{1}{q}} \|x\|_q \geq n^{-\frac{1}{p}} \|x\|_p .$$

Thus,

$$n^{(q-p)/pq} \|x\|_q \geq \|x\|_p .$$

Return

Proof of (10)**:** The first inequality holds obviously. Now, for the second inequality, we have

$$
\begin{aligned}
\|Ay\|_p &\leq \sum_{j=1}^{n} |y_j| \, \|a_j\|_p \\
&\leq \sum_{j=1}^{n} |y_j| \max_j \|a_j\|_p \\
&= \|y\|_1 \max_j \|a_j\|_p \\
&\leq n^{(p-1)/p} \max_j \|a_j\|_p . \quad \text{(by (8))}
\end{aligned}
$$

$\square$

Return

Proof of (15)**:** It holds

$$\begin{aligned}
\max_{\|x\|_p=1} \|Ax\|_p &= \max_{\|x\|_p=1} \max_{\|y\|_q=1} \left|(Ax)^T y\right| \\
&= \max_{\|y\|_q=1} \max_{\|x\|_p=1} \left|x^T (A^T y)\right| \\
&= \max_{\|y\|_q=1} \left\|A^T y\right\|_q \\
&= \left\|A^T\right\|_q .
\end{aligned}$$

$\square$

Proof of (16)**:** By (12) and (15), we get

$$\begin{aligned}
m^{\frac{1}{p}} \|A\|_\infty &= m^{\frac{1}{p}} \left\|A^T\right\|_1 = m^{1-\frac{1}{q}} \left\|A^T\right\|_1 \\
&= m^{(q-1)/q} \left\|A^T\right\|_1 \geq \left\|A^T\right\|_q = \|A\|_p .
\end{aligned}$$

$\square$

Return

Proof of (17)**:** It holds

$$\|A\|_p \|A\|_q = \|A^T\|_q \|A\|_q \geq \|A^T A\|_q \geq \|A^T A\|_2 .$$

The last inequality holds by the following statement: Let S be a symmetric matrix. Then $\|S\|_2 \leq \|S\|$, for any matrix operator norm $\|\cdot\|$. Since $|\lambda| \leq \|S\|$,

$$\|S\|_2 = \sqrt{\rho(S^* S)} = \sqrt{\rho(S^2)} = \max_{\lambda \in \sigma(S)} |\lambda| = |\lambda_{\max}| .$$

This implies, $\|S\|_2 \leq \|S\|$.

□

Proof of (18)**:** By (8), we get

$$\|A\|_p = \max_{\|x\|_p = 1} \|Ax\|_p \leq \max_{\|x\|_q \leq 1} m^{(q-p)/pq} \|Ax\|_q = m^{(q-p)/pq} \|A\|_q .$$

□

Return