EE 520: Topics – Compressed Sensing
Linear Algebra Review

Notes scribed by Kevin Palmowski, Spring 2013, for Namrata Vaswani's course
Notes on matrix spark courtesy of Brian Lois
More notes added by Namrata Vaswani

Notes based primarily on Horn and Johnson, *Matrix Analysis*, 1e and 2e, as well as Dr. Namrata Vaswani's in-class lectures.

Chapter 0 – Miscellaneous Preliminaries

Unless otherwise noted, all vectors are elements of $\mathbb{R}^n$, although results extend to complex vector spaces.

Let $S = \{v_i\}_1^k \subseteq \mathbb{C}^n$. We define the **span** of S by

$$\text{span}\,\{S\} = \left\{ x \in \mathbb{C}^n \mid \exists \{\alpha_i\}_1^k \subseteq \mathbb{C} \text{ with } x = \sum_1^k \alpha_i v_i \right\}.$$

The set S is **linearly independent** if $\sum_1^k \alpha_i v_i = \mathbf{0}$ if and only if $\alpha_i = 0$ for all i.

S is a **spanning set** for the vector space V if $\text{span}\,\{S\} = V$. A linearly independent spanning set for a vector space V is called a **basis**. The **dimension** of a vector space V, $\dim(V)$, is the size of the smallest spanning set for V.

The **rank** of a matrix $A \in \mathbb{R}^{m\times n}$, $\text{rank}(A)$, is the size of the largest linearly independent set of columns of A. Rank satisfies $\text{rank}(A) \leq \min\{m, n\}$. For matrices $A \in \mathbb{R}^{m\times k}$ and $B \in \mathbb{R}^{k\times n}$, we have

$$\text{rank}(A) + \text{rank}(B) - k \quad \leq \quad \text{rank}(AB) \quad \leq \quad \min\{\text{rank}(A),\ \text{rank}(B)\}.$$

For two matrices of the same size, we have $\text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B)$.

The **trace** of a matrix is the sum of its main diagonal entries, that is, $\text{trace}(A) = \sum a_{ii}$. For two matrices A and B, $\text{trace}(AB) = \text{trace}(BA)$.

The **support** of a vector, $\text{supp}(x)$, is the set of indices i such that $x_i \neq 0$. The size of the support of x (that is, the number of nonzero entries of x), $|\text{supp}(x)|$, is often denoted $\|x\|_0$, although $\|\cdot\|_0$ is not a vector norm (see the section on Chapter 5).

The **range** of a matrix $A \in \mathbb{R}^{m\times n}$, $\text{range}(A)$, is the set

$$\text{range}(A) = \{b \in \mathbb{R}^m \mid \exists x \in \mathbb{R}^n \text{ with } Ax = b\}.$$

Equivalently, the range of A is the set of all linear combinations of the columns of A. The **nullspace** of a matrix, $\mathrm{null}(A)$ (also called the kernel of A), is the set of all vectors x such that $Ax = \mathbf{0}$.

Suppose we have the matrix-vector equation $Ax = b$ with $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$. The equation is **consistent** if there exists a solution x to this equation; equivalently, we have $\mathrm{rank}([A, b]) = \mathrm{rank}(A)$, or $b \in \mathrm{range}(A)$. The equation has a **unique solution** if $\mathrm{rank}([A, b]) = \mathrm{rank}(A) = n$. The equation has **infinitely many solutions** if $\mathrm{rank}([A, b]) = \mathrm{rank}(A) < n$. The equation has **no solution** if $\mathrm{rank}([A, b]) > \mathrm{rank}(A)$.

We say a matrix $A \in \mathbb{R}^{m\times n}$ is **nonsingular** if $Ax = \mathbf{0}$ if and only if $x = \mathbf{0}$. When $m \geq n$ and A has full rank ($\mathrm{rank}(A) = n$), A is nonsingular. If $m < n$, then A must be singular, since $\mathrm{rank}(A) \leq \min\{m, n\} = m < n$. If $m = n$ and A is nonsingular, then there exists a matrix A^{-1} with $AA^{-1} = I_n = A^{-1}A$ and we call A invertible. The matrix-vector equation $Ax = b$ has the unique solution $x = A^{-1}b$ in this case.

The **Euclidean inner product** is a function defined (using the engineering convention – see Chapter 5) by $\langle x, y\rangle = x^*y = \sum_1^n \bar{x}_i y_i$, where the vectors in use are the same size and could be complex-valued. The **Euclidean norm** is a function defined by $\|x\| = \|x\|_2 = (\sum_1^n |x_i|^2)^{\frac{1}{2}}$ and satisfies $\langle x, x\rangle = \|x\|_2^2$. Note that $\langle Ax, y\rangle = \langle x, A^*y\rangle$ for all $A \in \mathbb{C}^{m\times n}$, $x \in \mathbb{C}^n$, $y \in \mathbb{C}^m$. For more information on inner products and norms, refer to the section on Chapter 5.

Two vectors x and y are **orthogonal** if $\langle x, y\rangle = 0$. Two vectors are **orthonormal** if they are orthogonal and $\|x\|_2 = \|y\|_2 = 1$. This concept can be extended to a set of vectors $\{v_i\}$.

Given a set $S = \{v_i\}_1^k \subseteq \mathbb{R}^n$, we define the **orthogonal complement** of S ("S perp") by

$$S^\perp = \{x \in \mathbb{R}^n \mid \langle x, v_i\rangle = 0 \text{ for all } i\ \}\,.$$

Note that $\left(S^\perp\right)^\perp = \mathrm{span}\,\{S\}$.

For a matrix $A \in \mathbb{R}^{m\times n}$, we have the relation $\mathrm{range}(A)^\perp = \mathrm{null}(A^*)$. *Proof:* ($\subseteq$) Let $y \in \mathrm{range}(A)^\perp$. Then for all $b = Ax \in \mathrm{range}(A)$, $0 = \langle b, y\rangle = \langle Ax, y\rangle = \langle x, A^*y\rangle$. In particular, if $x \equiv A^*y$, then we have $\|A^*y\|_2^2 = 0$, so that $A^*y \equiv \mathbf{0}$. Thus $y \in \mathrm{null}(A^*)$. ($\supseteq$) Let $y \in \mathrm{null}(A^*)$. Then for all $x \in \mathbb{R}^n$, we have $0 = \langle x, A^*y\rangle = \langle Ax, y\rangle$. As this holds for all choices of x, we conclude that $y \in \mathrm{range}(A)^\perp$. Set equality follows. $\square$

Projection theorem: Let $A \in \mathbb{R}^{m\times n}$. Then for all $y \in \mathbb{R}^m$, there exist unique vectors y_A and $y_\perp$ in $\mathbb{R}^m$ such that $y = y_A + y_\perp$, where $y_A \in \mathrm{range}(A)$ and $y_\perp \in \mathrm{range}(A)^\perp \equiv \mathrm{null}(A^*)$.

A **normal matrix** is a matrix N such that $NN^* = N^*N$. A **Hermitian matrix** is one such that $A^* = A$ and a **skew Hermitian matrix** is one such that $A^* = -A$. A real-valued Hermitian matrix is called a **symmetric matrix**. A **unitary matrix** is a square matrix with $UU^* = I = U^*U$. A real-valued unitary matrix is called an **orthogonal matrix**. An

idempotent matrix satisfies $A^2 = A$.

A **projection matrix** P is one which satisfies $P^2 = P$ (P is idempotent). If $P = P^*$, then P is called an **orthogonal projection**. Projection matrices project vectors onto specific subspaces. For any projection P which projects onto a subspace S, the projector onto the subspace $S^\perp$ is given by $(I - P)$. Given a matrix U with orthonormal columns, the (orthogonal) projector onto the column space of U is given by $P = UU^*$.

The **(Classical) Gram-Schmidt algorithm** is a theoretical tool which takes a set of vectors $\{v_i\}_1^k$ and creates a set of orthonormal vectors $\{q_i\}_1^k$ which span the same space as the original set. In practice, the Gram Schmidt algorithm is numerically unstable due to round-off and cancellation error and should not be implemented (numerically stable algorithms which accomplish the same goal, such as the Modified Gram-Schmidt algorithm, are freely available); however, as a theoretical device, it can be used to justify the existence of the QR factorization (see the section on Chapter 2). The original set of vectors need not be linearly independent to implement the algorithm (a minor modification can handle this situation), but for the span $\{q_i\} = \text{span}\,\{v_i\}$ property to hold, linear independence is required.

Classical Gram Schmidt Algorithm: Let $\{v_i\}_1^k$ be a linearly independent set of vectors. Initialize $z_1 = \frac{v_1}{\|v_1\|_2}$. For $\ell = 2..k$, compute $y_\ell = \left(v_\ell - \sum_{i=1}^{\ell-1} \langle z_i, v_\ell\rangle\, z_i\right)$ and then let $z_\ell = \frac{y_\ell}{\|y_\ell\|_2}$.

A **permutation matrix** is a matrix obtained by permuting rows and/or columns of an identity matrix. Permutation matrices satisfy $P^2 = I$, so that a permutation matrix is its own inverse. Permutation matrices are symmetric and orthogonal. Left-multiplication by a permutation matrix interchanges rows of the matrix being multiplied; right-multiplication interchanges columns.

A **circulant matrix** has the general form

$$\begin{bmatrix} a_1 & a_2 & \dots & a_{n-1} & a_n \\ a_n & a_1 & a_2 & \ddots & a_{n-1} \\ a_{n-1} & a_n & a_1 & \dots & a_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ a_2 & a_3 & \dots & a_n & a_1 \end{bmatrix}.$$

Each row of a circulant matrix is a cyclic permutation of the first row.

A **Toeplitz matrix** has the general form

$$\begin{bmatrix} a_0 & a_1 & a_2 & \dots & a_{n-1} & a_n \\ a_{-1} & a_0 & a_1 & \ddots & a_{n-2} & a_{n-1} \\ a_{-2} & a_{-1} & a_0 & \ddots & \ddots & a_{n-2} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ a_{-n+1} & a_{-n+2} & \ddots & \ddots & a_0 & a_1 \\ a_{-n} & a_{-n+1} & a_{-n+2} & \dots & a_{-1} & a_0 \end{bmatrix}$$

where $\{a_i\}_{-n}^{n}$ is any collection of scalars. With this notation, the general (ij)-entry of a Toeplitz matrix is given by $[A]_{ij} = a_{j-i}$. Notice that the entries of each diagonal of a Toeplitz matrix are constant.

An **upper (lower) triangular matrix** is a matrix whose entries below (above) the main diagonal are all zero. The remaining entries can be anything. A diagonal matrix is one whose only nonzero entries (if any) lie on the main diagonal. The eigenvalues of a triangular or diagonal matrix (see notes on Chapter 1) are the diagonal entries.

Let T be an upper-triangular matrix. T is invertible if and only if its diagonal entries are nonzero (since these are its eigenvalues). The matrix T^{-1} is also upper-triangular. Given any two upper-triangular matrices T_1 and T_2, their sum $T_1 + T_2$ and their product $T_1 T_2$ are also upper-triangular. A Hermitian upper-triangular matrix is necessarily diagonal (and real-valued). More generally, any normal upper-triangular matrix is diagonal. These results also hold for lower-triangular matrices.

Chapter 1 – Eigenvalues and Similarity

Suppose $A \in \mathbb{C}^{n\times n}$ and there exist $\lambda \in \mathbb{C}$ and $x \in \mathbb{C}^n$ (with $x \neq \mathbf{0}$) such that $Ax = \lambda x$. Then we call λ an **eigenvalue** of A with corresponding **eigenvector** x. The **spectrum** of A, $\sigma(A)$, is the set of all eigenvalues of A. The **spectral radius** of A, $\rho(A)$, is defined as $\rho(A) = \max_{\lambda\in\sigma(A)}\{|\lambda|\}$.

If $Ax = \lambda x$, then $A^k x = \lambda^k x$ for $k \in \mathbb{N} \cup \{0\}$. *Proof:* We proceed by induction. For the base case $k = 0$, the result is obvious (note $A^0 = I$). Suppose the result is true for $k = m \geq 0$. Consider $A^{m+1}x = A^m(Ax) = A^m(\lambda x) = \lambda(A^m x) = \lambda \cdot \lambda^m x = \lambda^{m+1}x$. By the Principle of Mathematical Induction, the desired result follows. $\square$

A consequence of the previous result: if $p(z) = \sum_0^k \alpha_i z^i$ be a polynomial. Then $p(A)x = p(\lambda)x$ for an eigenvector x of A with eigenvalue λ. *Proof:* $p(A)x = \left(\sum_0^k \alpha_i A^i\right)x = \sum_0^k \alpha_i\left(A^i x\right) = \sum_0^k \alpha_i\left(\lambda^i x\right) = \left(\sum_0^k \alpha_i \lambda^i\right)x = p(\lambda)x$. $\square$

If A is a Hermitian matrix, then $\sigma(A) \subset \mathbb{R}$. *Proof:* Let $Ax = \lambda x$. Then $\langle Ax, x\rangle = \langle \lambda x, x\rangle = \overline{\lambda}\langle x, x\rangle$. However, $\langle Ax, x\rangle = \langle x, A^* x\rangle = \langle x, Ax\rangle = \langle x, \lambda x\rangle = \lambda\langle x, x\rangle$. Since x is an eigenvector, $x \neq \mathbf{0}$, so $\langle x, x\rangle = \|x\|_2^2 \neq 0$. Thus $\lambda = \overline{\lambda}$, so $\lambda \in \mathbb{R}$. $\square$

A square matrix A is invertible if and only if 0 is not an eigenvalue of A.

Matrix trace satisfies $\text{trace}(A) = \sum_i \lambda_i(A)$ (see the next section).

Two matrices A and B are **similar** if there exists a nonsingular matrix S such that $A = S^{-1}BS$. This relation is often denoted $A \sim B$.

Similar matrices have the same eigenvalues, that is, $A \sim B$ implies $\sigma(A) = \sigma(B)$. *Proof:* Let $\lambda \in \sigma(A)$. Then there exists $x \neq \mathbf{0}$ such that $Ax = \lambda x$. Applying similarity, we have $Ax = S^{-1}BSx = \lambda x$, which implies that $B(Sx) = \lambda(Sx)$. Since S is invertible and $x \neq \mathbf{0}$, $Sx \neq \mathbf{0}$, so $\lambda \in \sigma(B)$. A similar argument (no pun intended) proves the reverse containment and thus equality. $\square$

Let $A \in \mathbb{C}^{n\times n}$. Then A is **diagonalizable** if A is similar to a diagonal matrix Λ whose (diagonal) entries are the eigenvalues of A, that is, there exists S such that $A = S^{-1}\Lambda S$. A matrix A is **unitarily diagonalizable** if S is a unitary matrix: $A = U^*\Lambda U$ for U unitary.

A matrix $A \in \mathbb{C}^{n\times n}$ is diagonalizable if and only if A has n linearly independent eigenvectors. *Proof:* Suppose $A = S^{-1}\Lambda S$. Then $AS^{-1} = S^{-1}\Lambda$. The matrix AS^{-1} has columns $(AS^{-1})_j = A(S^{-1})_j$ and the matrix $S^{-1}\Lambda$ has columns $(S^{-1}\Lambda)_j = \lambda_j(S^{-1})_j$. Therefore, the columns of S^{-1} are the eigenvectors of A. Since S^{-1} is invertible, it has n linearly independent columns, which proves the result (as all steps used are if and only ifs). $\square$

Chapter 2 – Triangularization and Factorizations

Two matrices A and B in $\mathbb{C}^{n\times n}$ are **unitarily equivalent** if there exist unitary matrices U and V such that $A = UBV$. The matrices are **unitarily similar** if $A = U^*BU$ for some unitary U.

Schur's Unitary Triangularization Theorem: Every matrix $A \in \mathbb{C}^{n\times n}$ is unitarily similar to an upper-triangular matrix T whose diagonal entries are the eigenvalues of A; that is, there exist U unitary and T upper-triangular with $t_{ii} = \lambda_i(A)$ such that $A = U^*TU$. If A is real and has only real eigenvalues, then U can be chosen real (orthogonal).

A consequence of Schur's theorem is that if A is normal, then T is also normal. *Proof:* Let $A = U^*TU$. Then $A^*A = U^*T^*UU^*TU = U^*T^*TU$ and $AA^* = U^*TUU^*T^*U = U^*TT^*U$. Therefore, $U^*T^*TU = A^*A = AA^* = U^*TT^*U$, so $T^*T = TT^*$, as desired. $\square$

Another consequence of Schur's theorem is that $\mathrm{trace}(A) = \sum_{\sigma(A)} \lambda_i$. *Proof:* There exist U unitary and T upper-triangular such that $A = U^*TU$ with $t_{ii} = \lambda_i$. So $\mathrm{trace}(A) = \mathrm{trace}(U^*TU) = \mathrm{trace}(TUU^*) = \mathrm{trace}(T) = \sum t_{ii} = \sum_{\sigma(A)} \lambda_i$. $\square$

The following are equivalent: (1) A is normal; (2) A is unitarily diagonalizable; (3) A has n orthonormal eigenvectors; (4) $\sum |a_{ij}|^2 = \sum |\lambda_i|^2$.

Proof: ($1 \Rightarrow 2$) If A is normal, then $A = U^*TU$ implies T is also normal. A normal triangular matrix is diagonal, so A is unitarily diagonalizable.

($2 \Leftarrow 1$) Let $A = U^*\Lambda U$. Since diagonal matrices commute, $\Lambda^*\Lambda = \Lambda\Lambda^*$, so $A^*A = U^*\Lambda^*\Lambda U = U^*\Lambda\Lambda^*U = AA^*$, and thus A is normal.

($2 \Leftrightarrow 3$) $A = U^*\Lambda U$ if and only if $AU^* = U^*\Lambda$. As we saw in the section on eigenvalues, this is true if and only if the columns of U^* are the eigenvectors of A. These eigenvectors are orthonormal since U^* is unitary.

($2 \Rightarrow 4$) Let $A = U^*\Lambda U$. Consider

$$\sum |a_{ij}|^2 = \mathrm{trace}(A^*A) = \mathrm{trace}(U^*\Lambda^*\Lambda U) = \mathrm{trace}(\Lambda^*\Lambda UU^*) = \mathrm{trace}(\Lambda^*\Lambda) = \sum |\lambda_i|^2.$$

($4 \Rightarrow 2$) Suppose that $\sum |a_{ij}|^2 = \mathrm{trace}(A^*A) = \sum |\lambda_i|^2$. By Schur's thereom, $A = U^*TU$ for some upper-triangular T. We have $\mathrm{trace}(A^*A) = \mathrm{trace}(T^*T) = \sum_{i,j} |t_{ij}|^2 = \sum |t_{ii}|^2 + \sum_{i\neq j} |t_{ij}|^2$. Since $t_{ii} \equiv \lambda_i$, we see that $\sum_{i\neq j} |t_{ij}|^2 = 0$, which implies that $t_{ij} \equiv 0$ for all $i \neq j$. Thus T is diagonal and A is therefore unitarily diagonalizable. $\square$

A matrix A is Hermitian if and only if $A = U\Lambda U^*$ with Λ diagonal and real. Further, a normal matrix whose eigenvalues are real is necessarily Hermitian. *Proof:* $A = A^* \Leftrightarrow U^*TU = U^*T^*U \Leftrightarrow T = T^* \Leftrightarrow T = \Lambda$ is diagonal and real-valued, which proves the first result. Since normal matrices are unitarily diagonalizable, the second result follows. $\square$

QR factorization: Let $A \in \mathbb{C}^{m\times n}$ with $m \geq n$. There exist matrices $Q \in \mathbb{C}^{m\times m}$ unitary and $R \in \mathbb{C}^{m\times n}$ upper-triangular such that $A = QR$. If A is nonsingular, then the diagonal entries of R can be chosen positive and the resulting QR factorization is unique. R is invertible in this case. If $m > n$, then we can form the reduced QR factorization $A = \hat{Q}\hat{R}$, where $\hat{Q} \in \mathbb{C}^{m\times n}$ has orthonormal columns and $\hat{R} \in \mathbb{C}^{n\times n}$ is upper-triangular. Lastly, if A is nonsingular, then the columns of Q span the same space as the columns of A.

Cholesky factorization: Suppose $B = A^*A$ for some matrix A. Then B has a Cholesky factorization $B = LL^*$, where L is lower-triangular. *Proof:* Since A has a full QR factorization, $B = A^*A = R^*Q^*QR = R^*R = LL^*$, where $L = R^*$. $\square$

Chapter 4 – Variational Characteristics of Hermitian Matrices

In this section, all matrices are $(n \times n)$ Hermitian matrices unless otherwise noted. Since the eigenvalues of Hermitian matrices are real-valued, we can order the eigenvalues, $\lambda_{\min} = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n = \lambda_{\max}$.

For $x \neq \mathbf{0}$, the value $\frac{x^*Ax}{x^*x} = \frac{\langle x, Ax\rangle}{\langle x, x\rangle}$ is called a **Rayleigh quotient**.

Rayleigh-Ritz Theorem: we have the following relations:

$$\lambda_{\max} = \max_{x \neq 0} \frac{x^*Ax}{x^*x} = \max_{\|x\|_2=1} x^*Ax$$

$$\lambda_{\min} = \min_{x \neq 0} \frac{x^*Ax}{x^*x} = \min_{\|x\|_2=1} x^*Ax$$

Courant-Fisher Theorem: Let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Let $\{w_i\}$ be arbitrary sets of linearly independent vectors in $\mathbb{C}^n$. Then the following characterizations of λ_k hold:

$$\lambda_k(A) = \min_{\{w_1,\ldots,w_{n-k}\}} \; \max_{x\neq 0;\; x\perp\{w_1,\ldots,w_{n-k}\}} \frac{x^*Ax}{x^*x}$$

$$\lambda_k(A) = \max_{\{w_1,\ldots,w_{k-1}\}} \; \min_{x\neq 0;\; x\perp\{w_1,\ldots,w_{k-1}\}} \frac{x^*Ax}{x^*x}$$

To simplify notation, we can equivalently express this in terms of an arbitrary subspace S:

$$\lambda_k(A) = \min_{\dim(S)=n-k} \; \max_{x\neq 0;\; x\in S^\perp} \frac{x^*Ax}{x^*x}$$

$$\lambda_k(A) = \max_{\dim(S)=k-1} \; \min_{x\neq 0;\; x\in S^\perp} \frac{x^*Ax}{x^*x}$$

One final equivalent version of the theorem (Horn and Johnson 2e) is given by:

$$\lambda_k(A) = \min_{\dim(S)=k} \; \max_{x\neq 0;\; x\in S} \frac{x^*Ax}{x^*x}$$

$$\lambda_k(A) = \max_{\dim(S)=n-k+1} \; \min_{x\neq 0;\; x\in S} \frac{x^*Ax}{x^*x}$$

Weyl's Inequality (simpler special case): let A, B be Hermitian matrices. Then

$$\lambda_k(A) + \lambda_{\min}(B) \quad \leq \quad \lambda_k(A+B) \quad \leq \quad \lambda_k(A) + \lambda_{\max}(B).$$

Using the fact that for a Hermitian matrix, $\|B\|_2 = \max(|\lambda_{\min}(B)|, |\lambda_{\max}(B)|)$, we have that $-\|B\|_2 \leq \lambda_{\min}(B) \leq \lambda_{\max}(B) \leq \|B\|_2$. Using this, Weyl implies that

$$\lambda_k(A) - \|B\|_2 \quad \leq \quad \lambda_k(A+B) \quad \leq \quad \lambda_k(A) + \|B\|_2.$$

In general, we have

$$\begin{aligned} \lambda_{j+k-n}(A+B) \quad &\leq \quad \lambda_j(A) + \lambda_k(B) \\ \lambda_{j+k-1}(A+B) \quad &\geq \quad \lambda_j(A) + \lambda_k(B). \end{aligned}$$

Ostrowski's Theorem: Let A be Hermitian and S be nonsingular. Then there exists $\theta_k \in [\lambda_{\min}(SS^*), \lambda_{\max}(SS^*)]$ such that $\lambda_k(SAS^*) = \theta_k \lambda_k(A)$.

Corollary: $\lambda_{\min}(SS^*)\lambda_k(A) \quad \leq \quad \lambda_k(SAS^*) \quad \leq \quad \lambda_{\max}(SS^*)\lambda_k(A)$.

Interlacing of eigenvalues: Let A be Hermitian and z be a vector. Then

$$\lambda_k(A + zz^*) \quad \leq \quad \lambda_{k+1}(A) \quad \leq \quad \lambda_{k+2}(A + zz^*)$$

Bordered matrices: Let $B \in \mathbb{C}^{n \times n}$, $a \in \mathbb{R}$, $y \in \mathbb{C}^n$ and define

$$A = \begin{bmatrix} B & y \\ y^* & a \end{bmatrix}.$$

Then

$$\lambda_1(A) \quad \leq \quad \lambda_1(B) \quad \leq \quad \lambda_2(A) \quad \leq \quad \lambda_2(B) \quad \leq \quad \ldots \quad \leq \quad \lambda_n(B) \quad \leq \quad \lambda_{n+1}(A).$$

If no eigenvector of B is orthogonal to y, then every inequality is a strict inequality.

Theorem: if $\mu_1 \leq \lambda_1 \leq \mu_2 \leq \ldots \leq \mu_n \leq \lambda_n \leq \mu_{n+1}$ then there exist $y \in \mathbb{R}^n$ and $a \in \mathbb{R}$ such that

$$M = \begin{bmatrix} \Lambda & y \\ y^* & a \end{bmatrix}$$

has the eigenvalues $\{\mu_i\}_1^{n+1}$, where $\Lambda = \text{diag}(\{\lambda_i\}_1^n)$.

Chapter 5 – Norms and Inner Products

A function $\|\cdot\| : \mathbb{C}^n \to \mathbb{R}$ is a **vector norm** if it satisfies the following 3 properties:

1. $\|x\| \geq 0$ for all $x \in \mathbb{C}^n$ and $\|x\| = 0$ if and only if $x \equiv \mathbf{0}$;
2. $\|\alpha x\| = |\alpha|\|x\|$ for all $\alpha \in \mathbb{C}$, $x \in \mathbb{C}^n$;
3. $\|x + y\| \leq \|x\| + \|y\|$ for all x, $y \in \mathbb{C}^n$ (Triangle Inequality).

Another useful form of the Triangle Inequality is $\|x - y\| \geq \big|\|x\| - \|y\|\big|$. *Proof:* Let $z = x - y$. Then

$$\|x\| = \|z + y\| \leq \|z\| + \|y\| = \|x - y\| + \|y\|,$$

so that $\|x\| - \|y\| \leq \|x - y\|$. Swapping the roles of x and y, we see that $\|y\| - \|x\| \leq \|y - x\| = \|x - y\|$; equivalently, $\|x\| - \|y\| \geq -\|x - y\|$. Thus,

$$-\|x - y\| \leq \|x\| - \|y\| \leq \|x - y\|,$$

which proves the result by definition of absolute value. $\square$

Common vector norms:

- $\|x\|_1 = \sum_i |x_i|$
- $\|x\|_2 = \sqrt{x^* x} = \sqrt{\sum_i |x_i|^2}$
- $\|x\|_\infty = \max_i\{|x_i|\}$
- $\|x\|_p = (\sum |x_i|^p)^{\frac{1}{p}}$ (the ℓ_p-norm, $p \in \mathbb{N}$; these norms are convex)

  $\diamondsuit$ The three norms above are the ℓ_1, ℓ_2 and ℓ_∞ norms.
- $\sqrt{\langle x, x\rangle}$ for any inner product $\langle \cdot, \cdot \rangle$
- $\|x\|_A = \|Ax\|$ for A nonsingular, $\|\cdot\|$ any vector norm

To denote the magnitude of the support of a vector x, $|\mathrm{supp}(x)|$, we often write $\|x\|_0$. The notation is widely used but is misleading because this function is NOT actually a norm; property (2) is not satisfied.

Equivalence of norms: Let $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ be any two norms on $\mathbb{C}^n$. Then there exist constants m and M such that $m\|x\|_\alpha \le \|x\|_\beta \le M\|x\|_\alpha$ for all $x \in \mathbb{C}^n$. The best attainable bounds for $\|x\|_\alpha \le C_{\alpha,\beta}\|x\|_\beta$ are given in the table below for α, $\beta \in \{1, 2, \infty\}$:

		β		
	$C_{\alpha,\beta}$	1	2	∞
	1	1	$\sqrt{n}$	n
α	2	1	1	$\sqrt{n}$
	∞	1	1	1

A **pseudonorm** on $\mathbb{C}^n$ is a function $f(\cdot)$ that satisfies all of the norm conditions except that $f(x)$ may equal 0 for a nonzero x (i.e. (1) is not totally satisfied).

A **pre-norm** is a continuous function $f(\cdot)$ which satisfies $f(x) \ge 0$ for all x, $f(x) = 0$ if and only if $x \equiv \mathbf{0}$ and $f(\alpha x) = |\alpha| f(x)$ (that is, f satisfies (1) and (2) above, but not necessarily (3)). Note that all norms are also pre-norms, but pre-norms are not norms.

Given any pre-norm $f(x)$, we can define the **dual norm** of f as

$$f^D(y) = \max_{f(x)=1} |y^* x|.$$

Note that f could be a vector norm, as all norms are pre-norms. In particular, $\|\cdot\|_1^D = \|\cdot\|_\infty$, and vice-versa. Further, $\|\cdot\|_2^D = \|\cdot\|_2$.

The dual norm of a pre-norm is always a norm, regardless of whether $f(x)$ is a (full) norm or not. If $f(x)$ is a norm, then $(f^D)^D = f$.

An **inner product** is a function $\langle \cdot, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ such that the following properties hold:

1. $\langle x, x \rangle \in \mathbb{R}$ with $\langle x, x \rangle \ge 0$ for all $x \in \mathbb{C}^n$ and $\langle x, x \rangle = 0$ if and only if $x \equiv \mathbf{0}$;
2. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ for all x, y, $z \in \mathbb{C}^n$;
3. $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$ for all $\alpha \in \mathbb{C}$, x, $y \in \mathbb{C}^n$;
4. $\langle x, y \rangle = \overline{\langle y, x \rangle}$ for all x, $y \in \mathbb{C}^n$.

Note that condition (4) and (3) together imply that $\langle x, \alpha y \rangle = \overline{\alpha} \langle x, y \rangle$.

It should be noted that the engineering convention of writing $x^* y = \langle x, y \rangle$ (as opposed to the mathematically accepted notation $\langle x, y \rangle = y^* x$) results in property (3) being re-defined as $\langle x, \alpha y \rangle = \alpha \langle x, y \rangle$.

Cauchy-Schwartz Inequality: $|\langle x, y\rangle|^2 \le \langle x, x\rangle \langle y, y\rangle$. *Proof:* Let $v = ax - by$, where $a = \langle y, y\rangle$ and $b = \langle x, y\rangle$. WLOG assume $y \neq \mathbf{0}$. Consider

$$\begin{aligned}
0 &\le \langle v, v\rangle \\
&= \langle ax, ax\rangle + \langle ax, -by\rangle + \langle -by, ax\rangle + \langle -by, -by\rangle \\
&= |a|^2 \langle x, x\rangle - a\bar{b} \langle x, y\rangle - \bar{a}b\overline{\langle x, y\rangle} + |b|^2 \langle y, y\rangle \\
&= \langle y, y\rangle^2 \langle x, x\rangle - \langle y, y\rangle \overline{\langle x, y\rangle} \langle x, y\rangle - \overline{\langle y, y\rangle} \langle x, y\rangle \overline{\langle x, y\rangle} + |\langle x, y\rangle|^2 \langle y, y\rangle \\
&= \langle y, y\rangle^2 \langle x, x\rangle - \langle y, y\rangle |\langle x, y\rangle|^2
\end{aligned}$$

Adding $\langle y, y\rangle |\langle x, y\rangle|^2$ to both sides and dividing by $\langle y, y\rangle$ proves the result. $\square$

The most common inner product is the ℓ_2-inner product, defined (in engineering terms) by $\langle x, y\rangle = x^*y = \sum \overline{x}_i y_i$. This inner product induces the ℓ_2-norm: $\|x\|_2 = \sqrt{x^*x} = \sqrt{\langle x, x\rangle}$.

If A is a Hermitian positive definite matrix, then $\langle x, y\rangle_A = \langle x, Ay\rangle = x^*Ay$ is also an inner product.

Any vector norm induced by an inner product must satisfy the **Parallelogram Law:**

$$\|u+v\|^2 + \|u-v\|^2 = 2\left(\|u\|^2 + \|v\|^2\right)$$

A **matrix norm** is a function $\|\cdot\| : \mathbb{C}^{n\times n} \to \mathbb{R}$ which satisfies:

1. $\|A\| \ge 0$ for all $A \in \mathbb{C}^{n\times n}$ and $\|A\| = 0$ if and only if $A \equiv 0$;
2. $\|\alpha A\| = |\alpha|\|A\|$ for all $\alpha \in \mathbb{C}$, $A \in \mathbb{C}^{n\times n}$;
3. $\|A + B\| \le \|A\| + \|B\|$ for all A, $B \in \mathbb{C}^{n\times n}$ (Triangle Inequality);
4. $\|AB\| \le \|A\|\|B\|$ for all A, $B \in \mathbb{C}^{n\times n}$.

Common matrix norms:

- $\|A\|_1 = \max_j \sum_i |a_{ij}|$ = maximum absolute column sum
- $\|A\|_2 = \sigma_1(A) = \sqrt{\lambda_{\max}(A^*A)}$
- $\|A\|_\infty = \max_i \sum_j |a_{ij}|$ = maximum absolute row sum
- Matrix norms induced by vector norms: $\|A\|_\beta = \max_{\|x\|_\beta=1} \|Ax\|_\beta = \max_{x\neq\mathbf{0}} \frac{\|Ax\|_\beta}{\|x\|_\beta}$

  ♢ The three norms above are alternate characterizations of the matrix norms induced by the vector norms $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$, respectively

- $\|A\|_* = \sum_i \sigma_i(A) = \sum_i \sqrt{\lambda_i(A^*A)}$
- $\|A\|_F = \sqrt{\sum |a_{ij}|^2} = \sqrt{\text{trace}(A^*A)} = \sqrt{\sum_i \lambda_i(A^*A)}$, sometimes denoted $\|A\|_{2,\text{vec}}$

For any invertible matrix A, we have $\|A^{-1}\| \geq \|A\|^{-1}$. *Proof:*

$$I = AA^{-1} \quad \Rightarrow \quad 1 \leq \|I\| = \|AA^{-1}\| \leq \|A\|\|A^{-1}\| \quad \Rightarrow \quad \|A\|^{-1} = \frac{1}{\|A\|} \leq \|A^{-1}\| \qquad \square$$

For any matrix A, we have $\|A\|_2 \leq \|A\|_F$. *Proof:*

$$\|A\|_2^2 = \max_i \sigma_i^2(A) \leq \sum_i \sigma_i^2(A) = \sum_i \lambda_i(A^*A) = \text{trace}(A^*A) = \|A\|_F^2. \qquad \square$$

A matrix B is an **isometry** for the norm $\|\cdot\|$ if $\|Bx\| = \|x\|$ for all x.

If U is a unitary matrix, then $\|Ux\|_2 = \|x\|_2$ for all vectors x; that is, any unitary matrix is an isometry for the 2-norm. *Proof:* $\|Ux\|_2^2 = \langle Ux, Ux\rangle = \langle x, U^*Ux\rangle = \langle x, x\rangle = \|x\|_2^2$. $\square$

If $A = U^*BU$, then $\|A\|_F = \|B\|_F$. *Proof:* $\|A\|_F^2 = \text{trace}(A^*A) = \text{trace}(U^*B^*BU) = \text{trace}(B^*BUU^*) = \text{trace}(B^*B) = \|B\|_F^2$. $\square$

Recall that $\rho(A) = \max_i |\lambda_i(A)|$. For any matrix A, matrix norm $\|\cdot\|$, and eigenvalue $\lambda = \lambda(A)$, we have $\lambda \leq \rho(A) \leq \|A\|$. *Proof:* Let $\lambda \in \sigma(A)$ with corresponding eigenvector x and let $X = [x, x, \ldots, x]$ (n copies of x). Consider $AX = [Ax, Ax, \ldots, Ax] = [\lambda x, \lambda x, \ldots, \lambda x] = \lambda X$. Therefore, $|\lambda|\|X\| = \|\lambda X\| = \|AX\| \leq \|A\|\|X\|$. Since x is an eigenvector, $x \neq \mathbf{0}$, so $\|X\| \neq 0$. Dividing by $\|X\|$, we obtain $|\lambda| \leq \|A\|$. Since λ is arbitrary, we conclude that $|\lambda| \leq \rho(A) \leq \|A\|$, as desired. $\square$

Let $A \in \mathbb{C}^{n\times n}$ and let $\varepsilon > 0$ be given. Then there exists a matrix norm $\|\cdot\|$ such that $\rho(A) \leq \|A\| \leq \rho(A) + \varepsilon$. As a consequence, if $\rho(A) < 1$, then there exists some matrix norm such that $\|A\| < 1$.

Let $A \in \mathbb{C}^{n\times n}$. If $\|A\| < 1$ for some matrix norm, then $\lim_{k\to\infty} A^k = 0$. Further, $\lim_{k\to\infty} A^k = 0$ if and only if $\rho(A) < 1$.

Chapter 7 – SVD, Pseudoinverse

Let $A \in \mathbb{C}^{m\times n}$. Then a **singular value decomposition (SVD)** of A is given by $A = U\Sigma V^*$, where $U \in \mathbb{C}^{m\times m}$ is unitary, $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p) \in \mathbb{R}^{m\times n}$ is diagonal with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$ [$p = \min(m,n)$], and $V \in \mathbb{C}^{n\times n}$ is unitary.

The values $\sigma_i = \sigma_i(A)$ are called the **singular values** of A and are uniquely determined as the positive square roots of the eigenvalues of A^*A.

If $\operatorname{rank}(A) = r$, then $\sigma_1 \geq \ldots \geq \sigma_r > 0$ and $\sigma_{r+1} = \sigma_{r+2} = \ldots = \sigma_p = 0$. In this case, a **reduced SVD** of A is given by $A = \hat{U}\hat{\Sigma}\hat{V}^*$, where $\hat{U} \in \mathbb{C}^{m\times r}$ has orthonormal columns, $\hat{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r\times r}$ is diagonal, and $\hat{V} \in \mathbb{C}^{n\times r}$ has orthonormal columns. In particular, given an SVD $A = U\Sigma V^*$, $\hat{U}$ and $\hat{V}$ in the reduced SVD are the first r columns of U and V.

One useful identity is that if $A = U\Sigma V^*$ with $\operatorname{rank}(A) = r$, then $A = \sum_1^r \sigma_i u_i v_i^*$, where u_i and v_i are columns of U and V (or, $\hat{U}$ and $\hat{V}$), respectively.

The first r columns of U in the SVD span the same space as the columns of A. *Proof:* Let $x \in \mathbb{C}^n$. Then $Ax = \sum_1^n a_i x_i$ is in the column space of A. However, $Ax = \left(\sum_1^r \sigma_i u_i v_i^*\right) x = \sum_1^r \sigma_i u_i v_i^* x = \sum_1^r \beta_i u_i$, where $\beta_i = \sigma_i v_i^* x$. This last expression lies in the span of the first r columns of U. Thus, every element in either column space can be equivalently expressed as an element in the other. $\square$

A matrix $A \in \mathbb{C}^{n\times n}$ has 0 as an eigenvalue if and only if 0 is also a singular value of A. *Proof:* ($\Rightarrow$) Suppose that 0 is an eigenvalue of A, that is, $Ax = \mathbf{0}$ for some nonzero x. Then $A^*Ax = \mathbf{0}$, so 0 is also an eigenvalue of A^*A. Therefore, $0 = \sqrt{0}$ is a singular value of A. ($\Leftarrow$) Suppose 0 is a singular value of A. Then there exists some nonzero x such that $A^*Ax = \mathbf{0}$. This implies that $x^*A^*Ax = 0 = (Ax)^*(Ax) = \|Ax\|_2^2$. By the properties of norms, we must have $Ax = \mathbf{0}$, which completes the proof. $\square$

Let $A \in \mathbb{C}^{m\times n}$ have SVD $A = U\Sigma V^*$. The **Moore-Penrose pseudoinverse** of A is the matrix $A^\dagger = V\Sigma^\dagger U^*$ ("A dagger"), where $\Sigma^\dagger$ is obtained by replacing the nonzero singular values of A (in Σ) with their inverses and then transposing the resulting matrix.

In terms of a reduced SVD, $A^\dagger = \hat{V}\hat{\Sigma}^{-1}\hat{U}^*$.

The pseudoinverse is uniquely determined by the following three properties:

1. $AA^\dagger$ and $A^\dagger A$ are Hermitian;

2. $AA^\dagger A = A$;

3. $A^\dagger AA^\dagger = A^\dagger$.

Additionally, $(A^\dagger)^\dagger = A$. If A is square and nonsingular, then $A^\dagger = A^{-1}$. If A has full column rank, then $A^\dagger = (A^*A)^{-1}A^*$.

One use of the pseudoinverse is to compute least-squares solutions of $Ax = b$. A least-squares solution x satisfies $\|x\|_2$ is minimal among all z such that $\|Az - b\|_2$ is also minimal. In this setup, the unique minimizer is computed as $x = A^\dagger b$.

More topics: Spark of a matrix

Let $A \in \mathbb{R}^{m\times n}$. Then we define the **spark** of A by

$$\text{spark}(A) = \begin{cases} \min_{x\neq 0} \|x\|_0 \ \text{ s.t. } Ax = 0 & \text{rank}(A) < n \\ \infty & \text{rank}(A) = n \end{cases}$$

Equivalently, spark(A) is the size of the smallest set of linearly dependent columns of A.

Equivalently, every set of spark$(A) - 1$ columns of A is linearly independent.

(Recall: rank(A) is the size of the largest linearly independent set of columns of A)

Two facts about spark: (a) spark$(A) = 1$ if and only if A has a column of zeros; (b) if spark$(A) \neq \infty$ then spark$(A) \leq \text{rank}(A) + 1$. *Proof:* (a) This follows immediately from the fact that the zero vector is always linearly dependent. (b) Let rank$(A) = r < n$. Then every set of $r + 1$ columns of A is linearly dependent, so spark$(A) \leq r + 1$. $\square$

Given a square matrix of size n and rank $r < n$, any value between 1 and $r+1$ is possible for spark(A). For example, all 3 matrices below have rank 3, but spark$(A_i) = i$.

$$A_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad A_2 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad A_3 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad A_4 = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

More Topics

A matrix A is positive semi-definite (p.s.d.) iff A is Hermitian and $x^*Ax \geq 0$ for all x. A is positive definite (p.d.) iff $x^*Ax > 0$ for all x. Similarly definite negative semi-definite and negative definite. Some results

1. A is p.s.d. iff A is Hermitian and $\lambda_i(A) \geq 0$. Proof: use variational characterization of $\lambda_{\min}$.
2. If A is p.s.d., then $\|A\|_2 = \lambda_{\max}(A)$.

More simple results on the spectral norm and the max/min eigenvalue. Recall that $\|B\|_2 := \lambda_{\max}(B^*B)$.

1. For any matrix B, $\|B\|_2 = \lambda_{\max}(BB^*)$.
   Proof: Use the fact that the set of nonzero eigenvalues of BB^* is equal to the set of nonzero eigenvalues of B^*B
2. For any matrix B, $\|B\|_2 = \max_{x \neq 0} \frac{\|Bx\|_2}{\|x\|_2}$.

   Proof: Use definition of $\|B\|_2$ and Rayleigh-Ritz (variational characterization) for $\lambda_{\max}(B^*B)$
3. For a Hermitian matrix A, $\|A\|_2 = \max(|\lambda_{\max}(A)|, |\lambda_{\min}(A)|)$. If A is p.s.d. then $\|A\|_2 = \lambda_{\max}(A)$.

   Proof: Use $\lambda_{\max}(A^*A) = \lambda_{\max}(A^2)$ and $\lambda_{\max}(A^2) = \max(\lambda_{\max}(A)^2, \lambda_{\min}(A)^2)$.
4. Let A be a block diagonal Hermitian matrix with blocks A_1 and A_2 both of which are Hermitian. Then $\lambda_{\max}(A) = \max(\lambda_{\max}(A_1), \lambda_{\max}(A_2))$.

   Proof: Let EVD of $A_1 = U_1 \Lambda_1 U_1^*$ and of $A_2 = U_2 \Lambda_2 U_2^*$, then EVD of A is $\begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} \Lambda_1 & \\ 0 & \Lambda_2 \end{bmatrix} \begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}^*$.
   This is EVD of A because it is easy to see that $\begin{bmatrix} U_1 & 0 \\ 0 & U_2 \end{bmatrix}$ is unitary.
5. Let B be a block diagonal matrix with blocks B_1 and B_2. Then $\|B\|_2 = \max(\|B_1\|_2, \|B_2\|_2)$.

   Proof: Use the fact that B^*B is block diagonal Hermitian with blocks $B_1^*B_1$ and $B_2^*B_2$
6. Let C be a 3x3 block matrix with the (1,2)-th block being C_1 and (2,3)-th block being C_2 and all other blocks being zero. Then $\|C\|_2 = \max(\|C_1\|_2, \|C_2\|_2)$. Similar result extends to general block matrices which have nonzero entries on just one diagonal.

   Proof: Use the fact that C^*C is block diagonal.

7. The above two results can be used to bound the spectral norm of block tri-diagonal matrices or their generalizations. [reference: Brian Lois and Namrata Vaswani, Online Matrix Completion and Online Robust PCA]

8. For a matrix B, we can define the dilation of B as $dil(B) := \begin{bmatrix} 0 & B \\ B^* & 0 \end{bmatrix}$. Clearly $dil(B)$ is Hermitian. Also, $\|B\|_2 = \|dil(B)\|_2 = \lambda_{\max}(dil(B))$. [reference: Joel Tropp, User friendly tail bounds for random matrices]

   Proof: First equality is easy to see using $dil(B)^* dil(B) = \begin{bmatrix} BB^* & 0 \\ 0 & B^*B \end{bmatrix}$. Thus $\|dil(B)\|_2^2 = \|dil(B)dil(B)^*\|_2 = \lambda_{\max}(BB^*) = \|B\|_2^2$.

   Let $B = USV'$. Second equality - follows by showing that $\begin{bmatrix} u_i \\ v_i \end{bmatrix}$ is an eigenvector of $dil(B)$ with eigenvalue σ_i and $\begin{bmatrix} -u_i \\ v_i \end{bmatrix}$ is an eigenvector with eigenvalue $-\sigma_i$. Thus the set of eigenvalues of $dil(B)$ is $\{\pm\sigma_i\}$ and so its maximum eigenvalue is equal to its minimum eigenvalue which is equal to maximum singular value of B.

9. For A Hermitian, $\|A\|_2 = \max_x \frac{|x^* A x|}{x^* x}$

   Proof - follows from Rayleigh-Ritz (variational characterization)

10. Dilation and the fact that the eigenvalues of $dil(B)$ are $\pm\sigma_i(B)$ is a very powerful concept to extend various results for eigenvalues to results for singular values.

    (a) This is used to get **Weyl's inequality for singular values of non-Hermitian matrices.** (reference: Terry Tao's blog or `http://comisef.wikidot.com/concept:eigenvalue-and-singular-value-inequalities`). Given $Y = X + H$,

    $$\sigma_{i+j-1}(Y) \leq \sigma_i(X) + \sigma_j(H)$$

    for all $1 \leq i, j \leq r$ and $i + j \leq r + 1$ where r is the rank of Y. Using $j = 1, i = r$, we get the special case

    $$\sigma_{\min}(Y) = \sigma_r(Y) \leq \sigma_r(X) + \sigma_1(H) = \sigma_{\min}(X) + \sigma_{\max}(H)$$

    Using $j = 1, i = 1$, we get the special case

    $$\sigma_1(Y) \leq \sigma_1(X) + \sigma_1(H)$$

    (b) It is also used to get bounds on min and max singular values of sums of non-Hermitian matrices (reference: Tropp's paper on User friendly tail bounds).

Topics to add

- ϵ-nets from Vershynin's tutorial
  - restricted isometry
  - order notation
  - Stirling approximation
  - bound min and max singular value of a tall matrix with random entries.

Proofs to add

add in proof of Rayleigh-Ritz, basic idea for Courant Fisher.