3 Eigenvalues, Singular Values and Pseudo inverse.

3.1 Eigenvalues and Eigenvectors

For a square $n \times n$ matrix A, we have the following definition:

Definition 3.1. *If there exist (possibly complex) scalar λ and vector x such that*

$$Ax = \lambda x, \quad \text{or equivalently}, \quad (A - \lambda I)x = 0, \quad x \neq 0$$

then x is the eigenvector corresponding to the eigenvalue λ. Recall that any $n \times n$ matrix has n eigenvalues (the roots of the polynomial $\det(A - \lambda I)$).

Definition 3.2. *Matrix A is called* simple *if it has n linearly independent eigenvectors.*

Definition 3.3. *Let $A^H \triangleq \bar{A}^T$, $x^H \triangleq \bar{x}^T$ (i.e., complex conjugate transpose). Matrix A is:*
Hermitian if $A = A^H \;\Leftrightarrow x^H Ax = real$, for all $x \in C^n$
Normal if $AA^H = A^H A$
Unitary if $AA^H = A^H A = I$
Orthogonal if $AA^T = A^T A = I$, (for A real)

Definition 3.4. *Hermitian matrix D (i.e., $D = D^H$) is*
positive definite if $x^H Dx > 0$ for all $x \neq 0$
positive semi definite if $x^H Dx \geq 0$ for all $x \neq 0$
negative definite if $x^H Dx < 0$ for all $x \neq 0$
negative semi definite if $x^H Dx \leq 0$ for all $x \neq 0$
indefinite if $x^H Dx < 0$ for some nonzero x and $x^H Dx > 0$ for some other nonzero x

Definition 3.5. *If $A = QBQ^{-1}$, for some nonsingular Q, then 'A is* similar *to B' or B is obtained via a similarity transformation (Q) of A. If we had $A = QBQ^T$, then A is obtained through a 'congruent' transformation on B.*

P1. For *general* matrix A: If all e-values are distinct; i.e., $\lambda_i \neq \lambda_j$, $(i \neq j)$, then A has n linearly independent eigenvectors; i.e., it is *simple*. Furthermore, we have

$$A = Q\Lambda Q^{-1}, \quad \Lambda = Q^{-1}AQ$$

where $Q = [x_1 \ldots x_n]$ (the e-vectors) and Λ is a diagonal matrix with λ_i on the (i,i) element. (Such a matrix is sometimes called *Diagonalizable*) .

P2. For Hermitian D, its eigenvalues are real; i.e, $Imag(\lambda_i) = 0 \;\; \forall i$. Furthermore, if D is real (i.e., real symmetric) the eigenvectors are real as well.

P3. If D is Hermitian, it is also *simple* .

P4. For $D = D^H$ (i.e, Hermitian D) eigenvectors corresponding to distinct eigenvalues are orthogonal in the sense that $x_j^H x_i = 0, \; if \; \lambda_i \neq \lambda_j$.

P5. For $D = D^H$, let $x_1 \cdots x_m$ be the eigenvector corresponding to the repeated eigenvalue $\hat{\lambda}$. Show that if we replace the $x_i's$ with their Gramm-Schmidt vectors, we still have m eigenvectors for $\hat{\lambda}$.

P6. For Hermitian D, the eigenvector matrix can be written as a unitary matrix; that is

$$D = Q\Lambda Q^H, \quad QQ^H = Q^H Q = I, \; , \Lambda \; real, \quad Q \; real \; if \; D \; real \; symmetric$$

P7. If $D = D^H$ is positive (semi) definite, then $D_{ii} > (\geq) 0$, with similar result for negative (semi) definite.

P8. For a Hermitian matrix D, we have
D positive semi definite if and only if (iff or $\Longleftrightarrow$) $\lambda_i \geq 0, \quad \forall i$
D is positive definite iff $\lambda_i > 0, \quad \forall i$
D is negative semi definite iff $\lambda_i \leq 0, \quad \forall i$
D is negative definite iff $\lambda_i < 0, \quad \forall i$
D is indefinite iff $\lambda_i > 0$ for some i and $\lambda_i < 0$ for some other i

P9. For any matrix A, $x^H A^H A x \geq 0, \quad \forall x$. Sometimes we write $A^H A \geq 0$ for short.

P10. If Hermitian D is positive semi definite ($D \geq 0$), then there exist Hermitian matrices V such that

$$D = VV, \; ; \quad e.g., \quad V = Q(\Lambda)^{0.5} Q^H$$

and furthermore there exist matrices C such that

$$D = C^H C \; ; \quad e.g., \quad C = (\Lambda)^{0.5} Q^H$$

P11. If Q is unitary, all of its eigenvalues have magnitude one; i.e, $|\lambda_i(Q)| = 1$.

P12. If λ is an eigenvalue of A, it is also an eigenvalue of A^T. Also, $\bar{\lambda}$ is an eigenvalue of A^H. Therefore if A is real, eigenvalues appear in complex conjugate pairs.

P13. If A is normal, then

$$Ax = \lambda x \Longleftrightarrow A^H x = \bar{\lambda} x$$

P14. If A is normal, its eigenvectors are orthogonal, in the sense that $x_i^H x_j = 0$

P15. If $A^2 = A$ then all eigenvalues of A are either zero or one (idempotent matrix)

P16. If $A^k = 0$ for any integer k, then all eigenvalues of A are zero (nilpotent matrix)

P17. For any Hermitian matrix D

$$\lambda_{min}(D)x^H x \leq x^H D x \leq \lambda_{max}(D)x^H x \quad \forall x \in C^n$$

where λ_{min} is the smallest eigenvalue (algebraically). This inequality is often called Raleigh's inequality.

P18. For any two Hermitian matrices M and N,

$$\lambda_{min}(M+N) \geq \lambda_{min}(N)+\lambda_{min}(M) \ , \textit{and} \ \lambda_{max}(M+N) \leq \lambda_{max}(N)+\lambda_{max}(M)$$

P19. If (λ, x) are an eigenvalue/eigenvector pair of the matrix AB, with $\lambda \neq 0$, then (λ, Bx) is an eigenvalue/eigenvector pair for BA.

P20. If A and B are similar (via transformation Q), they have the same eigenvalues and their eigenvectors differ by a Q term.

3.2 Singular Value Decomposition (SVD)

For the development below, assume $A \in C^{m\times n}$, $m \geq n$, with rank r (i.e., $\rho(A) = r$). Note that $A^H A \in C^{n\times n}$ and $AA^H \in C^{m\times m}$. Also, for inner product and norm, we use $\|x\|^2 = < x, x >$, with $< x, y > = x^H y$.

We need to review the following properties

$$Range(A) = Range(AA^H), \quad and \quad Range(A^H) = Range(A^H A)$$

which implies $\rho(A) = \rho(A^H) = \rho(AA^H) = \rho(A^H A) = r$. The basic SVD can be obtained through the following

SVD1. Let $AA^H u_i = \sigma_i^2 u_i$, for $i = 1, 2, \cdots m$.

$$U \stackrel{\triangle}{=} [u_1 \;\; u_2 \;\; \cdots u_m], \quad U \in C^{m\times m}, \quad UU^H = U^H U = I_m.$$

We then have $\|A^H u_i\| = \sigma_i$ for $i = 1, 2, \cdots m$.

SVD2. Let $A^H A v_i = \hat{\sigma}_i^2 v_i$, for $i = 1, 2, \cdots n$, such that

$$V \stackrel{\triangle}{=} [v_1 \;\; v_2 \;\; \cdots v_n], \quad V \in C^{n\times n}, \quad VV^H = V^H V = I_n.$$

Then nonzero $\hat{\sigma}_i$'s are equal to nonzero σ_i's of SVD1, with $v_i = \frac{A^H u_i}{\sigma_i}$. For zero $\hat{\sigma}_i$, we have $Av_i = 0$. (To show this, use P19 of the eigenvalue handout. Show that $A^H A$ and AA^H have the same nonzero eigenvalues, with $v's$ as defined above). These v_i's are linearly independent and form a set of orthonormal vectors.

SVD3. Consider the following n equations for $i = 1, 2, \cdots n$:

$$Av_i = AA^H \frac{u_i}{\sigma_i} \;\; (or\ zero) = \sigma_i u_i \;\; (or\ zero).$$

These equations can be written as

$$AV = U\Sigma, \iff A = U\Sigma V^H \tag{3.1}$$

where U and V are the same as SVD1 and SVD2, respectively. Σ is a $m \times n$ matrix, with the top left $n \times n$ block in diagonal form with σ_i's on the diagonal and the bottom $(m - n) \times n$ rows zero. Without loss of any generality, we let $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_n \geq 0$. These σ_i's are called the *singular values* of A (or A^H). Since rank of A is assumed to be $r \leq min\{m, n\}$, there are exactly r nonzero singular values (Why? recall SVD1 and SVD2). Therefore, we can write

$$U = [U_r \;\; \bar{U}_r], \quad U_r \in C^{m\times r}, \qquad V = [V_r \;\; \bar{V}_r], \quad V_r \in C^{n\times r}, \tag{3.2}$$

and

$$\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}, \quad \Sigma_r = diag\{\sigma_1, \sigma_2, \ldots, \sigma_r\} \tag{3.3}$$

with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Or condensing (3.1),

$$A = U_r \Sigma_r V_r^H. \tag{3.4}$$

Equations (3.1) or (3.4) are often called the 'singular value decomposition of A'. If A is a real matrix, all vectors (i.e, u_i's, v_i's) will be real and the superscript 'H' is replaced by 'T' - transpose. We can now discuss some of the main properties of singular values. First we introduce the following notation

$$\overline{\sigma}(A) \stackrel{\triangle}{=} \sigma_{max}(A), \quad \underline{\sigma}(A) \stackrel{\triangle}{=} \sigma_{min}(A), \tag{3.5}$$

where σ_i i the i^{th} singular value. Recall that an $m \times n$ matrix has n singular values, of which the last $n-r$ are zero ($r = \rho(A)$).

P1-SVD. The 'principal gains' interpretation:

$$\overline{\sigma}(A)\|x\|_2 \geq \|Ax\|_2 \geq \underline{\sigma}(A)\|x\|_2, \ \forall x \tag{3.6}$$

P2-SVD. The induced 2-norm:

$$\overline{\sigma}(A) = \|A\|_2 = sup \frac{\|Ax\|_2}{\|x\|_2}, \quad x \neq 0. \tag{3.7}$$

P3-SVD. If A^{-1} exists,

$$\overline{\sigma}(A) = \frac{1}{\underline{\sigma}(A^{-1})}. \tag{3.8}$$

Extra1. Null space of $A = \text{span}\{v_{r+1} \cdots v_n\}$ and range space of $A = \text{span}\{u_1 \cdots u_r\}$.

Extra2. $U_r^H U_r = I_r$ and $U_r U_r^H$ is the orthogonal projection operator onto the Range of A. (recall $R(A) = R(AA^H)$, but $R(AA^H) = span(u_1, \cdots, u_r)$, since u_i's are orthonormal, direct calculation of the projection operator gives the result).

Extra3. $V_r^H V_r = I_r$ and $V_r V_r^H$ is the orthogonal projection operator onto the Range of A^H.

3.3 A Famous Application of SVD

Let us consider the equation

$$Ax_o = b_o \quad \Rightarrow x_o = A^{-1}b_o$$

assuming that the inverse exists and A is known accurately. Now let there be some error in our data; i.e., let $b = b_o + \delta b$, where δb is the error or noise, etc. Therefore, we are now solving

$$Ax = b_o + \delta b \ \Rightarrow x = A^{-1}b_o + A^{-1}\delta b = x_o + \delta x.$$

We are interested in investigating how small or large is this error in the answer (i.e., δx) for a given amount of error. Note that

$$\delta x = A^{-1}\delta b \ \Rightarrow \|\delta x\| \leq \|A^{-1}\| \ \|\delta b\|$$

or since $\|A^{-1}\| = \sigma_{max}A^{-1} = \frac{1}{\sigma_{min}A}$, we can write

$$\|\delta x\| \leq \frac{\|\delta b\|}{\sigma_{min}A}. \tag{3.9}$$

However, recall that $x_o = A^{-1}b_o$ and therefore

$$\|x_o\| \geq \sigma_{min}(A^{-1})\|b_o\| = \frac{\|b_o\|}{\sigma_{max}A}. \tag{3.10}$$

Combining (3.9) and (3.10)

$$\frac{\|\delta x\|}{\|x_o\|} \leq \frac{\|\delta b\|}{\sigma_{min}A} \ \frac{1}{\|x_o\|}$$

or

$$\frac{\|\delta x\|}{\|x_o\|} \leq \frac{\|\delta b\|}{\|b_o\|} \ \frac{\sigma_{max}A}{\sigma_{min}A}$$

where the last fraction is called 'the condition number of A'. This number is indicative of the magnification of error in the linear equation of interest. Similar analysis can be done regarding a great many numerical and computational issues. In most problems, a matrix with very large condition number is called ill conditioned and will result in severe numerical difficulties.

Note that by definition, the condition number is equal or larger than one. Also, note that for unitary matrices, the condition number is one (one of the main reasons these matrices are used heavily in computational linear algebra).

3.4 Important Properties of Singular Values

In the following, use $\overline{\sigma}(A)$ as the maximum singular value of A, $\underline{\sigma}(A)$ as the minimum singular value and $\sigma_i(A)$ as the generic i^{th} singular value.

In all cases, $A \in \mathcal{C}^{m\times n}$. Recall that $\sigma_i^2 = \lambda_i(A^H A) = \lambda_I(AA^H)$, and that $\sigma_i(A) \geq 0$.

P4-SVD. $\sigma_i(\alpha A) = |\alpha|\sigma_i(A), \ \forall \alpha \in \mathcal{C}$

P5-SVD. $\overline{\sigma}(AB) \leq \overline{\sigma}(A) \, . \, \overline{\sigma}(B)$

P6-SVD. $\overline{\sigma}(A+B) \leq \overline{\sigma}(A) + \overline{\sigma}(B)$

P7-SVD. $\underline{\sigma}(AB) \geq \underline{\sigma}(A) \, . \, \underline{\sigma}(B)$

P8-SVD. $\underline{\sigma}(A) \leq |\lambda_i(A)| \leq \overline{\sigma}(A) \ \ \forall \ i$

P9-SVD. $\underline{\sigma}(A) - 1 \leq \underline{\sigma}(I + A) \leq \underline{\sigma}(A) + 1$

P10-SVD. $\underline{\sigma}(A) - \overline{\sigma}(B) \leq \underline{\sigma}(A \underline{+} B) \leq \underline{\sigma}(A) + \overline{\sigma}(B)$

P11-SVD. $\overline{\sigma}(A) \leq \sqrt{trace(A^H A)} \leq \sqrt{n}\ \overline{\sigma}(A)$

P12-SVD. $TraceA^H A = \sum_1^k \sigma_i^2(A), \quad k = min(n, m)$

P13-SVD. $detA^H A = \prod_1^k \sigma_i^2(A)$

P14.-SVD In general, $\sigma_i(AB) \neq \sigma_i(BA)$

P15-SVD. $\overline{\sigma}(A)\underline{\sigma}(B) \leq \overline{\sigma}(AB) \quad A \in \mathcal{C}^{m\times n}, \ \ B \in \mathcal{C}^{n\times l} \ \ n \leq l \ only$
– $\overline{\sigma}(B)\underline{\sigma}(A) \leq \overline{\sigma}(AB) \quad A \in \mathcal{C}^{m\times n}, \ \ B \in \mathcal{C}^{n\times l} \ \ n \leq m \ only$

P16-SVD. $\underline{\sigma}(AB) \leq \overline{\sigma}(A)\underline{\sigma}(B) \quad no \ restrictions$
– $\underline{\sigma}(AB) \leq \overline{\sigma}(B)\underline{\sigma}(A) \quad no \ restrictions$

P17-SVD. $\underline{\sigma}(A)\underline{\sigma}(B) \leq \underline{\sigma}(AB) \leq \overline{\sigma}(A)\underline{\sigma}(B) \leq \overline{\sigma}(AB) \leq \overline{\sigma}(B)\overline{\sigma}(A), \ \ n \leq l$
– $\underline{\sigma}(A)\underline{\sigma}(B) \leq \underline{\sigma}(AB) \leq \overline{\sigma}(B)\underline{\sigma}(A) \leq \overline{\sigma}(AB) \leq \overline{\sigma}(B)\overline{\sigma}(A), \ \ n \leq m$

3.5 Pseudo Inverse

The basic definition of inverse of a matrix A is well known, when it is square and full rank. For non-square, but full rank, matrix $A \in R^{m\times n}$, we have the following: When $m > n$ ($n > m$) left (right) inverse of A is the matrix B in $R^{n\times m}$ (in $R^{m\times n}$) such that BA (AB) is I_n (I_m).

When the matrix is not full rank, the so called 'pseudo' inverses are used. The famous definition of Penrose is the following. The pseudo inverse of A is the unique matrix (linear operator) $A^\dagger$ that satisfies the following

1. $(A^\dagger A)^H = A^\dagger A$
2. $(AA^\dagger)^H = AA^\dagger$
3. $A^\dagger AA^\dagger = A^\dagger$
4. $AA^\dagger A{=}A$

Recalling that matrix P is a projection if $P^2 = P$ and is orthogonal projection if $P = P^2$ *and* $P = P^H$, we can see that the pseudo inverse has the following properties

- $A^\dagger A$ is the orthogonal projection onto Range of A^H
- $AA^\dagger$ is the orthogonal projection onto Range of A
- $(A^\dagger)^\dagger = A$

Now we will suggest the following candidate:

$$A = U_r \Sigma_r V_r^H \Longrightarrow A^\dagger = V_r \Sigma_r^{-1} U_r^H \tag{3.11}$$

PINV1. Show that for full rank matrices, the definition in (3.11) reduces to standard inverse (square matrices) or left or right inverse.

PINV2. Verify that $A^\dagger$ defined in (3.11) satisfies the basic properties of pseudo inverse.

To gain a better understanding of the pseudo inverse, consider the linear equation $Ax = y$. When A is square and full rank, the solution is $A^{-1}y$. In general, we say that the least squares solution of this problem is $A^\dagger y$! Let us investigate some more.

PINV3. Show that when A is a wide (or long) matrix with full row rank, the problem has infinitely many solutions, among which only one is in the range of A^H. Further, this solution has the smallest norms among all possible solutions. The solution is $x = (right\ inverse\ of\ A)y$.

PINV4. When A is a tall matrix with full column rank, then $x =$ (left inverse of A) y gives the unique solution or (if no solution exists) the solution that minimizes the 2-norm of the error $(y - Ax)$.

We can generalize this by letting A be rank deficient. Starting with y, we find y_p its projection onto range of A to minimize the norm of the error ($y_p = y$ if at least one solution exists). Now $Ax = y_p$ has one or many solutions, among which the one with minimum norm is the *unique* vector x_o such that it is in the range space of A^H. The relationship between x_o and y is $x_o = A^\dagger y$. In short, the pseudo inverse simultaneously minimizes the norm of the error as well as the norm of the solution itself.

PINV5. Show that the definition of $A^\dagger$ in (3.11) is the same as the development discussed above (i.e., show that Ax_o is equal to y_p and x_o is in the range of A^H. For this last part recall that the range of A^H is the same as range of $A^H A$ which is the same as span of the v_1 to v_r).

Another common, and equivalent, definition (see Zadeh and Desoer) for the pseudo inverse is the matrix satisfying

1. $A^\dagger Ax = x \quad \forall x \in range\ of A^H$
2. $A^\dagger z = 0 \quad \forall z \in null\ space\ of A^H$
3. $A^\dagger(y + z) = A^\dagger y + A^\dagger z \quad \forall y \in R(A) \quad, \forall z \in R(A)^\perp$

Finally, they suggest the following calculation for the inverse

$$A^\dagger = (A^H A)^\dagger A^H \tag{3.12}$$

PINV6. Show that (3.12) results in the same matrix as (3.11).