Topics in Matrix Theory

Singular Value Decomposition

Tang Chiu Chak

The University of Hong Kong

3rd July, 2014

Contents

1 Introduction

In linear algebra, we have come across some useful theories related to matrices. For instance, eigenvalues of squared matrices, which in some sense translate complicated matrix multiplication to simple scalar multiplication and provide us a way to factorise some matrices into a nice form.
However, most of the discussions in the courses are restricted in squared matrix with real entries. Will the theorems become different if we allow complex matirces? Can we extend the concept of eigenvalue to rectangular matrices? In response to these question, here we would like to first have a quick review on what we learned in linear algebra (but in the contest of complex matrices) and then study upon the idea of **singular value**–the "eigenvalue" for non-squared matrices.

2 Review on linear algebra

2.1 Definition

An **m by n matrix** is a rectantgular array with m rows and n columns (and hence mn entries in total) and sometimes we will entriewise write an m by n matrix whose (i,j)-entry is a_{ij} as $[a_{ij}]_{m\times n}$.

Example 2.1.0.1 : $\begin{bmatrix} i & e & \pi \\ 4 & 5 & 6 \end{bmatrix}$ is a 2 by 3 matrix.

Column vectors are m by 1 matrices and **row vectors** are 1 by n matrices.

2.1.1 Operation for Matrices

Let $A = [a_{ij}]_{m\times n}$ and $B = [b_{ij}]_{l\times k}$. Matrix addition and matrix multiplication are only defined between matrices of appropriate size–

Addition between matrices A and B, denoted $A + B$, is their entrywise addition, and hence is only defined for matrices of the same size ($m = l$ and $n = k$)

$$[a_{ij}]_{m\times n} + [b_{ij}]_{m\times n} = [a_{ij} + b_{ij}]_{m\times n}$$

Multiplication between matrices, denoted as AB, is defined when $n = l$. The product will be a m by k matrix C, where $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$

$$[a_{ij}]_{m\times n}[b_{ij}]_{n\times k} = \left[\sum_{k=1}^{n} a_{ik}b_{kj}\right]_{m\times k}$$

Scalar multiplication for matrix is simpler:

$$\text{For any scalar } \lambda\ ,\ \lambda[a_{ij}]_{m\times n} = [\lambda a_{ij}]_{m\times n}$$

The **transpose** of $A = [a_{ij}]_{m\times n}$, denoted as A^T, is an n by m matrix whose (i, j)-entry is a_{ji}

For complex matrix $A = [a_{ij}]$, we can also talk about its **conjugate** $\overline{A} = [\overline{a_{ij}}]$ and its **conjugate transpose** $A^H = \overline{A^T}$

For real vectors $\mathbf{x} = [x_1, \ldots, x_n]^T, \mathbf{y} = [y_1, \ldots, y_n]^T$, the **inner product** between them is defined to be $< \mathbf{x}, \mathbf{y} >= \mathbf{y}^T\mathbf{x} = \sum x_i y_i$. However, for complex vector, the definition will be slightly different: for complex vectors $\mathbf{u} = [u_1, \ldots, u_n]^T, \mathbf{v} = [v_1, \ldots, v_n]^T$,

$$< \mathbf{u}, \mathbf{v} >= \mathbf{v}^H\mathbf{u} = \sum u_i \overline{v_i}$$

When a matrix has as many rows as columns, i.e. $m = n$, the matrix is called **square matrix**. For a square matrix A, we can talk about its **determinant**, denoted as $det(A)$, which is defined inductively:

1. For 1 by 1 matrix $A = [a_{ij}]$, $det(A) = a_{ij}$
2. For larger matrix, $det(A) = \sum_{i=1}^{n}(-1)^{i+1}a_{1i}\ det(A_{1i})$
   where A_{1i} is a matrix obtained from A by deleting the first row and the i-th column

The **trace** of A, denoted as $tr(A)$, is the sum of all entries in the diagonal of A, i.e. $tr(A) = \sum a_{ii}$

2.1.2 Notation and terminology

Linear combination of a set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is the weighted sum of the vectors inside, i.e. $\sum \lambda_i \mathbf{x}_i$, where the λs are scalar.

The vector space spanned by the set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, denoted as $Span\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, is a collection of all linear combinations of the vectors in the set.

The set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is said to be **linear independent** iff $\lambda_1 = \lambda_2 = \cdots = 0$ is the only solution for the equation $\sum \lambda_i \mathbf{x}_i = \bar{0}$, where $\bar{0}$ denote the **zero vector**, i.e. a vector with all entries zero.

A **basis** B for a vector space V is a linear independent set of vectors which span V. The **dimension** of V is the number of elements in B. $col(A)$ is a vector space spanned by the columns of A and the **rank** of A, denoted as $rank(A)$, is the dimension of $col(A)$

$||\mathbf{x}|| = \sqrt{<\mathbf{x},\mathbf{x}>}$ is the **norm** of $\mathbf{x}$. When $||\mathbf{x}|| = 1$, $\mathbf{x}$ is an **unit vector**. When $<\mathbf{x},\mathbf{y}>= 0$, we say that $\mathbf{x}$ and $\mathbf{y}$ are orthogonal to each other. A set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is called **an orthogonal set** if $<\mathbf{x}_i, \mathbf{x}_j>= 0$ for $i \neq j$. It is further called an **orthonormal set** if it also satisfies that $||\mathbf{x}_i|| =<\mathbf{x}_i, \mathbf{x}_i>= 1$ for all i.

The **zero matrices** O is a matrices with all entries zero. The **identity matrix** I is a sqaure matrix whose etries on the main diagonal are 1 and otherwise 0. Usually, their sizes are not specified as they will be clear in the context.

When a square matrix A has a non-zero deternimant, A will be **invertible**, i.e. there exists a matrix of the same size,call the **inverse** of A and denoted as A^{-1}, such that $A^{-1}A = AA^{-1} = I$

The **eigenvalue** λ of a square matrix A and the corresponding (to λ) eigenvector $\mathbf{x}$ is a pair of scalar and non-zero vector, such that

$$A\mathbf{x} = \lambda\mathbf{x}$$

λ is the eigenvalue of A iff it satisfies the equation $det(A - \lambda I) = 0$, which is also called the **characteristic polynomial** of A. Moreover, the corresponding λ-eigenvector is the non-tivial solution to the system $(A - \lambda I)\mathbf{x} = 0$.

A is said to be **diagonalisable** iff A is **similar** to a diagonal matrix, i.e. there exists an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$. In linear algebra, we know that a matrix is diagonalisable if and only if it has n linearly independent eigenvectors.

A square matrix $A = [a_{ij}]$ is said to be

- **upper** (lower)- **triangular** if $a_{ij} = 0$ for $i > j$ $(i < j)$
- **diagonal** if $a_{ij} = 0$ for $i \neq j$
- **symmetric** if $A^T = A$
- **skew-symmetric** if $A^T = -A$
- **Hermitian** if $A^H = A$
- **normal** if $A^H A = AA^H$
- **orthogonal** if $A^T A = AA^T = I$
- **unitary** if $A^H A = AA^H = I$

2.2 Some basic results

Recall that the standard inner product between complex vectors are:

$$< \mathbf{x}, \mathbf{y} >= \mathbf{y}^H \mathbf{x} = \sum x_i \overline{y_i}$$

Then, we have

1. $< \mathbf{x}, \mathbf{y} >= \overline{< \mathbf{y}, \mathbf{x} >}$
2. $< \mathbf{x}_1 + \mathbf{x}_2, \mathbf{y} >=< \mathbf{x}_1, \mathbf{y} > + < \mathbf{x}_2, \mathbf{y} >$ and
$< \mathbf{x}, \mathbf{y}_1 + \mathbf{y}_2 >=< \mathbf{x}, \mathbf{y}_1 > + < \mathbf{x}, \mathbf{y}_2 >$
3. $< \lambda\mathbf{x}, \mathbf{y} >= \lambda < \mathbf{x}, \mathbf{y} >=< \mathbf{x}, \overline{\lambda}\mathbf{y} >$
4. $< \mathbf{x}, \mathbf{x} >\geq 0$ and the equality holds iff $\mathbf{x} = 0$
($||\mathbf{x}|| = \sqrt{< \mathbf{x}, \mathbf{x} >}$ is the **norm** of $\mathbf{x}$)

On the other hand, for conjugate transpose, we have,

1. $(A^H)^H = A$
2. $(A + B)^H = A^H + B^H$
3. $(\lambda A)^H = \overline{\lambda} A^H$
4. $(AB)^H = B^H A^H$

Recall that a matrix A is called hermitian when $A^H = A$. One may be capable to see that:

$$A^H = A \text{ if and only if } < A\mathbf{x}, \mathbf{y} >=< \mathbf{x}, A\mathbf{y} > \forall \mathbf{x}, \mathbf{y} \ \in \mathbb{C}^n$$

The verification of the above properties is simple and is left as an exercise.

The following result ensures the existence of an orthogonal basis, which is important for our later discussion.

Gram-Schmidit Orthogonalization Algorithm

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be any basis of a vector space V. We successively construct $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n\}$ as follows:

$$\mathbf{y}_1 = \mathbf{x}_1$$

$$\mathbf{y}_k = \mathbf{x}_k - \sum_{i=2}^{k-1} \frac{< \mathbf{x}_k, \mathbf{y}_i >}{||\mathbf{y}_i||} \mathbf{y}_i \qquad \text{(for k=2, \ldots,n)}$$

Then

1. $Span\{\mathbf{y}_1, \ldots, \mathbf{y}_k\} = Span\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$
2. $< \mathbf{y}_i, \mathbf{y}_j >= 0$ for $i \neq j$

In linear algebra, we have learned about the **Principal Axis Theorem**,which claim that the following statements are equivalent for matrices with real entries:

1. A is orthogonally diagonalizable
2. A is symmetric, i.e. $A^T = A$
3. A has an orthonomal set of n eigenvectors

In next section, we aim to generalise some of the results above to complex matrices and extend the theorems to non-square matrices.

3 Singular value Decomposition

3.1 Schur's Theorem and Spectral Theorem

The following lemma is a simple fact that will be useful in our later proves.

Lemma 1
The following are equivant for an n by n complex matrix A

1. A is unitary, i.e. A is invertible with $A^{-1} = A^H$
2. The rows (columns) of A are an orthonormal set in $\mathbb{C}^n$

Now, before we start trying to diagonalise a complex matrix, we would like to introduce the Schur's Theorem, which states that every sqaure matrix is unitarily similar to an upper triangular.

Schur's Theorem

Let A be a n by n complex matrix. There exist a unitary matrix U, i.e. $U^{-1} = U^H$, such that

$$U^H AU = T$$

where T is upper triangular whose entries on main diagonal are eigenvalues $\lambda_1, \ldots, \lambda_n$ of A (counting with multiplicity)

Proof
We will prove this by induction on n.

Firstly, When n=1, A is already an upper triangular with eigenvalue on main diagonal and we just need to let $U = I$, done.
Now, suppose that for $n = k$, where $k = 1, 2, \dots$, there exist such a U, so that $U^H AU = T$.
Next, for $n = k + 1$, we let λ_1 be an eigenvalue of A and $\mathbf{x}_1$ is the corresponding **unit** eigenvector. Extend $\{\mathbf{x}_1\}$ to an orthonormal basis $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ of $\mathbb{C}^n$. Then by Lemma 1, we know that $U_1 = [\mathbf{x}_1 \dots \mathbf{x}_n]$ is unitary and in block form we can see that

$$U_1^H AU_1 = \begin{bmatrix} \lambda_1 & * \\ \bar{0} & A_1 \end{bmatrix}$$

Then by I.H., there there exist such a W_1, so that $W_1^H A_1 W_1 = T_1$.
Let $U_2 = \begin{bmatrix} 1 & \bar{0} \\ \bar{0} & W_1 \end{bmatrix}$. Then U_2 is a unitary matrix and so do $U = U_1 U_2$.
Finally, we have

$$\begin{aligned} U^H AU =& U_2^H (U_1^H AU_1) U_2 \\ =& \begin{bmatrix} 1 & \bar{0} \\ \bar{0} & W_1^H \end{bmatrix} \begin{bmatrix} \lambda_1 & * \\ \bar{0} & A_1 \end{bmatrix} \begin{bmatrix} 1 & \bar{0} \\ \bar{0} & W_1 \end{bmatrix} \\ =& \begin{bmatrix} \lambda_1 & ** \\ \bar{0} & T_1 \end{bmatrix} \end{aligned}$$

which is a upper triangular. The result follows. ††

Notice that A and T will have the same rank, traces, determinants and eigenvalues as they are **similar**, i.e. $U^{-1}AU = T$ for some invertible U.

When A is hermitian, we will obtain a even nicer result:

Spectral Theorem

If A is hermitian, then A is unitarily diagonalizable, i.e. there exist a unitary matrix U such that $U^H AU = D$ is diagonal.

Proof

As $A^H = A$, we have

$$T^H = (U^H AU)^H = (U)^H (A)^H (U^H)^H = U^H AU = T$$

That is, T is both upper and lower triangular and hence diagonal. ††

3.2 Unitarily diagonalizable

Recall that in linear algebra, we learned that a matrix is said to be orthogonally diagonalisable if there exist an orthogonal matrix U (i.e. $U^T U = UU^T = I$) such that $U^T AU$ is diagonal. We also learned that real symmetric matrices must be orthogonally diagonalisable. However, this does not holds for complex matrices. For instance, $\begin{bmatrix} 1 & i \\ i & -1 \end{bmatrix}$ is symmetric but not orthogonally diagonalisable.

In fact, for complex matrices, we are more concern about unitarily diagonalizable than orthogonally diagonalisable. By unitarily diagonalizable, we mean that there exist an unitary matrix U (i.e. $U^H U = UU^H = I$) such that $U^T AU$ is diagonal.

However, the sufficient condition for a complex matrix to be unitarily diagonalizable provided in the last subsection turns out to be not a necessary condition. For instance, let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ is not hermitian but unitarily diagonalizable.

It turns out that the necessary and sufficient condition for a matrix A to be unitarily diagonalizable is A is normal , i.e. $AA^H = A^H A$.

Theorem 1

An n by n complex matrix A is unitarily diagonalizable if and only if A is normal.

Proof

(”only if” part)

Firstly, if A is unitarily diagonalizable, then $A = UDU^H$ for some unitary U and diagonal D. It is easy to verify that $D^H D = DD^H$. Hence, we have A is normal as

$$\begin{aligned} AA^H &= UDU^H (UDU^H)^H \\ &= UDU^H UD^H U^H \\ &= UDD^H U^H \\ &= UD^H DU^H \\ &= (UDU^H)^H UDU^H = A^H A \end{aligned}$$

(”if” part)

On the other hand, by Schur’s thoerem, write $U^H AU = T$, we have if A is normal, then T is normal too. The reason is that

$$TT^H = U^H AUU^H A^H U = U^H (AA^H)U = U^H (A^H A)U = U^H A^H UU^H AU = T^H T$$

It suffices to show that this implies T is diagonal. However, notice that $T = [t_{ij}]$ is upper-triangular, i.e. $t_{ij} = 0$ for $i > j$, if we consider the (i, i)-enties of TT^H and $T^H T$ respectively, we have

$$\sum_{k=1}^{n} |t_{ik}| = \sum_{k=1}^{n} |t_{ki}|$$

That is, the sum of the modulus of the entries in the i-th row of T equals to the sum of the modulus of the entries in the i-th column of T.

For instance, when $i = 1$, we have

$$|t_{11}| + \sum_{k=2}^{n} |t_{1k}| = |t_{11}| + 0$$

which gives $t_{12} = \cdots = t_{1n} = 0$ as $|t_ij| \geq 0$. Continue with simular argument on $i = 2, 3, \ldots$, we have that T is diagonal. ††

3.3 Singular value decomposition

To extend our discussion to non-square matrix, we need to introduce the concept of **singular value**:

Singular Value – For any m by n matrix A, the **singular value** σ of A are defined to be the non-negative of the n eigenvalues of $A^H A$

Note that $(A^H A)^H = A^H (A^H)^H = A^H A$ and hence $A^H A$ is hermitian. Thus, let λ be an eigenvalue of $A^H A$ and $\mathbf{x}$ is a corresponding eigenvector, consider that

$$\lambda ||\mathbf{x}|| = \lambda(\mathbf{x}^H \mathbf{x}) = \mathbf{x}^H \lambda \mathbf{x} = \mathbf{x}^H (A^H A)\mathbf{x} = (A\mathbf{x})^H A\mathbf{x} = ||A\mathbf{x}||$$

We have $\lambda \geq 0$. Hence, singular values are always real (and non-negative).

Now, we would like to introduction a useful way to decompose complex matrices of any sizes– Singular value decomposition.

Singular Value Decomposition

Let A be an m by n matrix with non-zero singular values $\sigma_1, \sigma_2, \ldots, \sigma_r$. Then there exist an m by m unitary matrix U and an n by n unitary matrix V, such that

$$A = U\Sigma V$$

where $\Sigma = \begin{bmatrix} D & O \\ O & O \end{bmatrix}$, $D = diag(\sigma_1, \sigma_2, \ldots, \sigma_r)$ and O denote zero matirces of appropriate size.

Proof

First, notice that we may assume $A \neq O$ since for otherwise, we simply let $U = I$, $V = I$, $\Sigma = O$, then done. Now, for $A \neq O$, we would like to prove by induction on the size of A (on n).

When $n = 1$, then A is a column vector and its singular value σ will simply be the norm of A. Let U be a unitary matrix with the unit vector $\frac{1}{\sigma}A$ the first column (Which is possible by Lemma 1 and orthogonalization algorithm), $\Sigma = \begin{bmatrix} \sigma \\ 0 \end{bmatrix}$, $V = [1]$. Then we have $A = U\Sigma V$

Now, suppose that this also holds up to $n = k$, i.e. assume that $A = U\Sigma V$ for m by k matrix A. Then, for $n = k+1$, let σ_1 be a singular value of A, $\mathbf{v}_1$ be an unit eigenvector of $A^H A$ correspoding to σ_1^2 and let $\mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1$.
Computation gives

$$||\mathbf{u}_1|| = \frac{1}{\sigma_1^2}\mathbf{v}^H A^H A\mathbf{v} = 1 \tag{1}$$

$$A^H\mathbf{u}_1 = \sigma_1\mathbf{v}_1 \tag{2}$$

$$(A\mathbf{v}_1)^H = \sigma_1\mathbf{u}_1^H \tag{3}$$

Let V_1 be a unitary matrix with $\mathbf{v}_1^H$ as the first row and U_1 be a unitary matrix with $\mathbf{u}_1$ as the first column. Then in block form we have

$$V_1 A^H U_1 = \begin{bmatrix} \sigma_1 & X_1 \\ Y_1 & A_1^H \end{bmatrix}$$

Notice that by (2) we have $Y_1 = O$ and by (3) we have $X_1 = O$. Hence we have

$$A = U_1 \begin{bmatrix} \sigma_1 & O \\ O & A_1 \end{bmatrix} V_1$$

Here we have produce an smaller matrix A_1 with k columns.
So by induction, we have $A_1 = U_2\Sigma_1 V_2$

Let $U = U_1 \begin{bmatrix} 1 & O \\ O & U_2 \end{bmatrix}$, $\Sigma = \begin{bmatrix} \sigma_1 & O \\ O & \Sigma_1 \end{bmatrix}$, $V = \begin{bmatrix} 1 & O \\ O & V_2 \end{bmatrix} V_1$
Then we have

$$U\Sigma V = U_1 \begin{bmatrix} 1 & O \\ O & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & O \\ O & \Sigma_1 \end{bmatrix} \begin{bmatrix} 1 & O \\ O & V_2 \end{bmatrix} V_1 = A$$

as desire.

3.4 Brief disscussion on SVD

3.4.1 Singular vector

Let $A = U\Sigma V$. Observe that:

$$A^H A = V^H \Sigma^H U^H U \Sigma V = V^H \Sigma^2 V$$
$$AA^H = U\Sigma^H V V^H \Sigma U^H = U\Sigma^2 U^H$$

We have

1. columns in U, called the left-singular vectors of A, are eigenvectors of AA^H
2. columns in V^H, called the right-singular vectors of A, are eigenvectors of $A^H A$

3. $A^H A$ and AA^H have the same non-zero eigenvalues

4. Since U, V are invertible, rank of A = rank of Σ

3.4.2 Reduced SVD

Let $(\sigma_1, \ldots \sigma_r)$ be a sequence of singular values of $A = U\Sigma V$. Let E_ij denote a matrix with the (i,j)-entry equals 1 and otherwise 0, e_i^k denote a column vector in $\mathbb{C}^k$ with the $(i,1)$-entry equals 1 and otherwise 0, u_i be the i-th columns of U, v_i be the i-th row of V, then we have

$$\begin{aligned} A &= U\ (D \bigoplus O)V & (D = diag(\sigma_1, \ldots \sigma_r)) \\ &= U\ (\sum_{i=1}^{r} \sigma_i E_{ii} + O)V \\ &= \sum_{i=1}^{r} \left[\sigma_i U\ e_i^m (e_i^n)^H V\right] \\ &= \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i \end{aligned}$$

This may give us a sense that it is not a must to compute all the singular vectors so as to decompose A. In reality, we normally compute the reduced SVDs. For example, compact SVD, in which we have $A = U_r \Sigma_r V_r^H$, where $\Sigma_r = diag(\sigma_1, \ldots \sigma_r)$ namely that we will only compute those singular vectors correpsond to non-zero singular values. This way, we can still decompose A into nice form with less complicated computation.

3.4.3 Pseudoinverse

A **pseudoinverse** A^+ of an m by n matrix A is a generalization of the inverse matrix, which is defined to be an n by m matrix satisfying all of the followings:

1. $AA^+A = A$

2. $A^+AA^+ = A^+$

3. $(AA^+)^H = AA^+$

4. $(A^+A)^H = A^+A$

In fact, such a matrix always exists and we can obtain it easily via SVD:

If $A = U\ (D \bigoplus O)V$ where $D = diag(\sigma_1, \ldots \sigma_r)$, then we will have $A^+ = V^H\ (D^+ \bigoplus O)U^H$ where $D^+ = diag(\frac{1}{\sigma_1}, \ldots \frac{1}{\sigma_r})$
One can verify that the matrix obtained will satisfy the above criteria.

4 Reference

1. Zhang, Fuzhen. *Matrix theory: basic results and techniques.* Springer, 2011.
2. Nicholson, W. Keith. *Linear algebra with applications.* McGraw-Hill Ryerson, 2006.