

## Basics

- In most cases,  $Ax$  and  $x$  point to different directions. When  $Ax$  is parallel to  $x$  (stretched), we call  $x$  the eigenvector of  $A$ . This is described by  $Ax = \lambda x$
- From  $Ax = 0$ , we know that the eigenvector corresponding to  $\lambda = 0$  lies in the null space of  $A$ . In other words, if a matrix is singular, then it has at least one zero eigenvalue.
- We examine the eigenvector and eigenvalues of projection matrix  $P = A(A^T A)^{-1} A^T$ .  $Pb$  and  $b$  are parallel only when  $b$  is already in the projection plane.  $b$  is unchanged before and after projection ( $Pb = 1 \cdot b$ ). That is, all the vector in the projection plane are eigenvectors of the projection matrix, with 1 as their eigenvalues. The projection plane are the column space of  $A$ . This indicates that the projection operator defined this way is to project to the column space of matrix  $A$ . Let's examine the normal vector to the projection plane,  $e$ . Because  $e \perp C(A)$ , we have  $Pe = 0e$ . That is the eigenvalue of eigenvector  $e$  is 0. In summary, the eigenvalues of projection matrix are  $\lambda = 1, 0$ . Due to the zero eigenvalue, then projection matrix is usually singular, unless it is an identity matrix.
- Properties of eigenvalues:
  - $\sum_{i=1}^n \lambda_i = \sum_{i=1}^n a_{ii}$ .
  - $\prod_{i=1}^n \lambda_i = \det A$ .
  - The eigenvalues of  $AB$  and  $BA$  are same. See proof in [Are the eigenvalues of AB equal to the eigenvalues of BA? \(Citation needed!\)](http://math.stackexchange.com/questions/124888/are-the-eigenvalues-of-ab-equal-to-the-eigenvalues-of-ba-citation-needed) (<http://math.stackexchange.com/questions/124888/are-the-eigenvalues-of-ab-equal-to-the-eigenvalues-of-ba-citation-needed>).

## Find eigenvalue and eigenvectors from solving $Ax = \lambda x$

- Solving eigenvalue problem

$$Ax = \lambda x \Rightarrow (A - \lambda I)x = 0 \Rightarrow (A - \lambda I) \text{ is singular if } x \text{ has non-zero solution} \Rightarrow |A - \lambda I| = 0$$

From  $|A - \lambda I| = 0$  for eigenvalues  $\lambda$ , and then substituting  $\lambda$  to  $(A - \lambda I)x = 0$  to solve for the eigenvector  $x$ . **Eigenvectors are just the null space of shifted matrix of  $A$ .**

- Example 1

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix} \Rightarrow \det(A - \lambda I) = \begin{vmatrix} 3 - \lambda & 1 \\ 1 & 3 - \lambda \end{vmatrix} \Rightarrow (3 - \lambda)^2 - 1 = \lambda^2 - 6\lambda + 8 = 0, \lambda_1 = 4, \lambda_2 = 2$$

Note the first-order coefficient  $-6$  and constant  $8$  are respectively related to the trace and determinant of the matrix  $A$ . It is easy to verify that  $A - 4I = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$  is singular. Otherwise there must be error in the calculation of eigenvalues. The null space of  $A - 4I = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$  is  $x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ . Another solution  $x_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$  can be obtained from the null space of  $A - 2I = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ .  $x_1, x_2$  are thus the eigenvectors of matrix  $A$ .

- Example 2

For matrix  $A' = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ , we have  $\lambda_1 = 1, x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \lambda_2 = -1, x_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ . The relation of  $A'$  and  $A$  is  $A = A' + 3I$ . So if adding  $3I$  to a matrix, then the eigenvalues will be added a constant  $3$  and eigenvectors keep unchanged (eigenvectors are unique to within a negative sign).

Note the above conclusion is only valid when we add an scaled identity matrix. That is, for  $Ax = \lambda x, Bx = \alpha x$ , the relation  $(A + B)x = (\lambda + \alpha)x$  is only valid when at least  $A$  or  $B$  is identity matrix multiplied by a constant. In general case, we should write as  $Ax = \lambda x, By = \alpha y$ , which cannot be added at all.

- Example 3

Consider rotation matrix  $Q = \begin{bmatrix} \cos 90^\circ & -\sin 90^\circ \\ \sin 90^\circ & \cos 90^\circ \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ , we have  $\begin{cases} \lambda_1 + \lambda_2 = 0 \\ \lambda_1 \cdot \lambda_2 = 1 \end{cases}$ . Also considering the requirement of eigenvectors in this case: Rotating a vector by  $90^\circ$  and then still parallel with itself. Obviously, there is no such vector. Also there are no real eigenvalues satisfy the above system of equations.

Solving  $\det(Q - \lambda I) = \begin{vmatrix} \lambda & -1 \\ 1 & \lambda \end{vmatrix} = \lambda^2 + 1 = 0$  gives  $\lambda_1 = i, \lambda_2 = -i$ . The more a matrix is like a symmetric matrix, the more likely we have real eigenvalues. The more a matrix is away from a symmetric matrix, the more likely we have complex or pure imaginary eigenvalues. The  $Q$  is not a symmetric matrix but an anti-symmetric matrix. General matrix are in the between of symmetric and antisymmetric.

- Example 4

Consider an even worse case,  $A = \begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$ . This is triangular matrix and its eigenvalues are just the diagonal elements. This can be verified as  $\det(A - \lambda I) = \begin{vmatrix} 3 - \lambda & 1 \\ 0 & 3 - \lambda \end{vmatrix} = (3 - \lambda)^2 = 0$ ,  $\lambda_1 = 3, \lambda_2 = 3$ . Due the same eigenvalues, we cannot obtain two independent eigenvectors. In this case,  $A$  is a degenerate matrix.

## Diagonalization of matrix: application of eigenvectors

Using  $n$  independent eigenvectors of matrix  $A$  to form a new matrix  $S$ , then we have

$$AS = A \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 & \lambda_2 x_2 & \cdots & \lambda_n x_n \end{bmatrix} = \begin{bmatrix} x_1 \lambda_1 & x_2 \lambda_2 & \cdots & x_n \lambda_n \end{bmatrix}.$$

Using column picture, we have  $\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} = S\Lambda$ . when using column picture, note the  $x_1, x_2, \cdots, x_n$  are column

vectors but not components of a vector. From the above derivation, be familiar with all the following forms of equations,

$$AS = S\Lambda S^{-1}AS = \Lambda A = S\Lambda S^{-1}$$

Note we require  $S$  is an invertible matrix in the above derivation. If a matrix has  $n$  different eigenvalues, then it must have  $n$  different eigenvectors and thus can be diagonalized. If a matrix has same eigenvalues, it might (e.g. identity matrix) or might not have different eigenvectors. The key to diagonalize a matrix is whether it is invertible.

Application of the above matrix diagonalization. For  $A^2$ , eigenvectors are unchanged while eigenvalues should be squared. Similarly we have  $A^k = S\Lambda^k S^{-1}$ .

If  $k \rightarrow \infty$ , then what is the condition for  $A^k \rightarrow 0$  to be stable? From  $S\Lambda^k S^{-1}$ , we obtain  $|\lambda_i| < 1$ .

## Symmetric matrix

### Two fundamental properties of symmetric matrix (or Hermitian matrix)

- Eigenvalues are real.

$$Ax = \lambda x \Rightarrow \bar{x}^T Ax = \bar{x}^T \lambda x$$

$$Ax = \lambda x \Rightarrow \bar{A}\bar{x} = \bar{\lambda}\bar{x} \Rightarrow \bar{x}^T \bar{A}^T = \bar{x}^T \bar{\lambda} \Rightarrow \bar{x}^T \bar{A}^T x = \bar{x}^T \bar{\lambda} x$$

If we have  $\bar{A}^T = A$ , Then the LHS of the two equations above are same. Thus we have

$$\bar{x}^T \lambda x = \bar{x}^T \bar{\lambda} x$$

$$\bar{\lambda} = \lambda$$

Note we have used the condition  $\bar{x}^T x \neq 0$ .

- Eigenvectors are orthogonal.

$\bar{x}^T \lambda_1 y = (\lambda_1 \bar{x})^T y = (\bar{A}\bar{x})^T y = \bar{x}^T \bar{A}^T y = \bar{x}^T A y = \bar{x}^T \lambda_2 y$  Because  $\lambda_1 \neq \lambda_2$ , we have inner product  $\bar{x}^T y = 0$ .

Instead of  $\bar{A}$ , the notation  $A^*$  is sometimes also used, which can lead to confusion since this symbol is also used to denote the conjugate transpose.

### Important notes

- A symmetric matrix is always diagonalizable.  $A = Q\Sigma Q^T$ . For general matrix, it requires that  $A$  is an invertible matrix to be diagonalizable. This is because  $A = S\Sigma S^{-1}$ . If  $A$  is not invertible, then it does not have full eigenvectors and thus  $S$  is not invertible.
- Note the  $Q$  obtained by finding eigenvectors of  $A$  is only a special basis. It can make  $A$  diagonal and thus easy to handle. Like other basis (square or not, orthogonal or not)  $Q$  can be used to do basis transformation for either vector or operator. Any independent columns can form a basis. Such a basis can be linear transformed to another basis by another linear operator.

## Matrix diagonalization with symmetric (Hermitian) matrix

In the case of symmetric matrix, we have

$$A = S\Lambda S^{-1} = Q\Lambda Q^T$$

This is called spectral theorem or principle axis theorem. It requires that a matrix is symmetric (or Hermitian conjugate), and hence there are orthogonal eigenvectors. These orthogonal eigenvectors can form an orthonormal matrix  $Q$  satisfying  $QQ^T = I$  or  $Q^{-1} = Q^T$

We further write  $A = QAQ^T = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix} = \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T + \cdots + \lambda_n q_n q_n^T \cdot \frac{qq^T}{q^T q} = qq^T$  is a projection matrix.

Thus each symmetric matrix can be expanded as a series of orthogonal projection matrices.

## Positive definite matrix and minimum value

### What is positive definite matrix?

Positive definite matrix should be a symmetric matrix in the first place. Are there matrices with positive eigenvalues but are not symmetric? Check the entry "positive definite matrix" in Wikipedia. It says: "Some authors use more general definitions of "positive definite", including some non-symmetric real matrices, or non-Hermitian complex ones." From above we know that positive definite is commonly defined as a subset of symmetric matrix. And there should be some all-positive-eigenvalue non-symmetric matrix.

### How to judge whether a matrix is positive definite

$A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$  is positive definite if:

1.  $\lambda_1 > 0, \lambda_2 > 0$ ;
2. The determinant of all leading principal submatrices are all  $> 0$ :  $a > 0, ac - b^2 > 0$ ;
3. After elimination, pivots  $> 0$ :  $a > 0, \frac{ac - b^2}{a} > 0$ ;
4.  $x^T Ax > 0$ ;

Consider the following matrix  $\begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}$ .

- All pivots are positive. First pivot is 5. The second pivot need not to do elimination. We know that the product of two pivots are  $\det A$  which is easy to calculate in  $2 \times 2$  matrix, then the second pivot is  $\frac{5 \times 3 - 2 \times 2}{5} = \frac{11}{5}$ . So two pivots are all positive, and thus the matrix is positive definite (note it is already a symmetric matrix).
- All eigenvalues are positive. Using  $|\lambda I - A| = 0$  to obtain the two eigenvalues and find they are also positive. Thus from this we know that the symmetric matrix is positive definite. Note when calculating eigenvalues, we can use the two properties for eigenvalues: trace of matrix equals to the sum of eigenvalues, product of eigenvalues equal to  $\det A$ .
- All sub-determinants are positive. First sub-determinant is 5, the second is 11. All positive, and therefore positive definite.

## Other ways to determine whether a matrix is positive definite

- $A = S\Lambda S^{-1} \Rightarrow A^{-1} = S\Lambda^{-1}S^{-1}$ . The inverse matrix of a positive definite matrix is also positive definite.
- If  $A, B$  are both positive definite, then we have  $x^T A x > 0, x^T B x > 0 \Rightarrow x^T (A + B) x > 0$ . Thus  $A + B$  is positive definite.
- For  $m \times n$  matrix  $A$ ,  $A^T A$  is symmetric at least semi positive definite.  $x^T A^T A x \Rightarrow (Ax)^T (Ax) = |Ax|^2 \geq 0$ . If we further assume  $A$  has full column rank, or when  $A$  has independent columns, then  $A$  is positive definite.

## Relation of positive definite and minimum

In the matrix  $A = \begin{bmatrix} 2 & 6 \\ 6 & ? \end{bmatrix}$ , what value of  $?$  can make it positive definite?

- First try  $? = 18$ , which gives  $A = \begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix}$ ,  $\det A = 0$ . Thus  $A$  is positive semi-definite.  $A$  is singular and one eigenvalue must be zero.

From trace, we know the other eigenvalue must be 20. Calculating  $x^T A x$  gives  $\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_1^2 + 12x_1x_2 + 18x_2^2$ . Thus we

obtain a function of  $x_1, x_2, f(x_1, x_2) = 2x_1^2 + 12x_1x_2 + 18x_2^2$ . In this example, it is a pure quadratic function. When  $? = 18$ , the first three conditions for positive definite are just failed.

- An example that three conditions must fail. Let  $? = 7$  and then  $A = \begin{bmatrix} 2 & 6 \\ 6 & 7 \end{bmatrix}$ . The function  $f(x_1, x_2) = 2x_1^2 + 12x_1x_2 + 7x_2^2$ . If take  $x_1 = 1, x_2 = -1$ , then  $f(1, -1) = 2 - 12 + 7 < 0$ . In the  $(x, y, z)$  coordinate system,  $z(0, 0) = 0$ . When  $y = 0$  or  $x = 0$  or  $x = y$ , the function is a parabola opening up. So the function has positive values along some direction and negative along others. We actually have a

saddle curve where  $(0, 0, 0)$  is called saddle point. This point can be either maximum or minimum, depending along which direction.

- An example that three conditions must be satisfied. Let  $\gamma = 20$ , then  $A = \begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix}$ , with determinant  $\det A = 4$  and trace  $\text{trace}(A) = 22$ .

Thus all eigenvalues are positive. The function becomes  $f(x_1, x_2) = 2x_1^2 + 12x_1x_2 + 20x_2^2$ . The function takes positive values except at  $(0, 0)$ . Consider the surface described by  $z = 2x^2 + 12xy + 20y^2$ , which is a paraboloid. At  $(0, 0)$ , the first partial derivatives are zero while second partial derivatives are positive.

- The 2nd-order partial derivative matrix must be positive definite in order for the multi-variable function  $f(x_1, x_2, \dots, x_n)$  to have a minimum. In the examples above, if  $\gamma = 7$ , then  $f(x, y) = 2(x + 3y)^2 - 11y^2$ . If  $\gamma = 18$ , then  $f(x, y) = 2(x + 3y)^2$ . Let  $z = 1$ , we can respectively obtain a hyperbola and an ellipse.

- The second derivative matrix mentioned above has the form  $\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$ . Obviously, all the diagonal elements must be positive. Also

these diagonal elements must offset the effect from off-diagonal terms if they are negative. Because the order of taking derivatives in calculating second derivative does not affect the results, we conclude that the matrix must be symmetric.

## Similar matrix

- $A, B$  are similar to each other if there is a relation  $B = M^{-1}AM$  for a matrix  $M$ . We can also write it in other form  $A = MBM^{-1}$ . In the matrix diagonalization, we have  $S^{-1}AS = \Lambda$ , thus  $A$  is similar to  $\Lambda$ .
- Jordan's form was once the climax of the subject of similar matrix or even linear algebra. This is no longer the case any more. Now it becomes the SVD.
- Jordan's form is try to diagonalize matrix that cannot be diagonalizable. In other words, it tries to make in the best form (most like diagonalized matrix, although it is not). Nowadays, this can be easily and efficiently done by SVD.
- Similar matrices have same eigenvalues. This can be proved as follows:  $Ax = \lambda x$ ,  $B = M^{-1}AM$ . Modifying the first equation to be  $AMM^{-1}x = \lambda x$  and then left multiplying  $M^{-1}$  in both sides give  $M^{-1}AMM^{-1}x = \lambda M^{-1}x$ . Regrouping the terms gives

$$(M^{-1}AM)M^{-1}x = \lambda M^{-1}x, \text{ i.e., } BM^{-1}x = \lambda M^{-1}x. \text{ } BM^{-1} = \lambda M^{-1}x. \text{ Thus } \lambda \text{ is still the eigenvalue of } B, \text{ with eigenvector } M^{-1}x.$$

- Example 1:  $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ , the corresponding diagonalized matrix is  $\Lambda = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$ . Take  $M = \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}$ , then

$$B = M^{-1}AM = \begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & -15 \\ 1 & 6 \end{bmatrix}. \text{ Now calculate the eigenvalues of these matrices (use the properties of}$$

determinant and trace) and have  $\lambda_A = 3, 1, \lambda_B = 3, 1$ . Therefore, all the matrices with eigenvalues 3, 1 are similar matrices.

For example,  $\begin{bmatrix} 3 & 7 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 7 \\ 0 & 3 \end{bmatrix}$ . The most special one is  $\Lambda$ .

- Example 2: Let  $\lambda_1 = \lambda_2 = 4$ . Write out two matrices with these eigenvalues  $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}, \begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ . There are in fact two families of these matrices. The first family has only one matrix which is  $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ . It is similar to itself as  $M^{-1} \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} M = 4M^{-1}IM = 4I = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ . So whatever the form of  $M$ , the diagonal matrix is similar to only itself. The other family includes matrices like  $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ . In fact, this is the 'best' matrix in the family called Jordan's form. Other matrices in the family include  $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}, \begin{bmatrix} 5 & 1 \\ -1 & 3 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 17 & 4 \end{bmatrix}$ . We can always construct a matrix satisfying  $\text{trace}(A) = 8, \det A = 16$ , and this matrix must belong to the family.

## Singular Value Decomposition (SVD)

### Equivalent statements for $n \times n$ square matrix $A$

- $A$  is invertible.
- $A$  has linearly independent column vectors.
- $A$  has no zero eigenvalue. Otherwise,  $Ax = 0$  and  $x$  in null space. In this case, we have a singular matrix. However, the singular here seems not related to singular value.
- $A$ 's determinant is nonzero.
- $A$  has  $n$  non-zero singular values. From here, we know that we cannot use whether it is zero to determine eigenvalue or singular value. Eigenvalue can be zero, while singular values can be all non-zero. While eigenvalue is defined as the  $\lambda$  in  $Ax = \lambda x$ , singular values are defined as the elements of  $\Sigma$  in the definition of singular value decomposition.
- $Ax = 0$  has only the trivial solution  $x = 0$ .
- Line transformation  $x$  to  $Ax$  is one-to-one.
- $A^T$  is invertible.
- The null space of  $A$  is  $\{0\}$ .



- Other equivalents are in <http://mathworld.wolfram.com/InvertibleMatrixTheorem.html>  
(<http://mathworld.wolfram.com/InvertibleMatrixTheorem.html>)

## Comparison to Eigen-decomposition / spectral decomposition

In eigen-decomposition of matrix  $A$ , we need many constraints:

- $A$  is a square matrix.
- $A$  has independent columns, or  $A$  has non-zero eigenvalues, or  $A$  is invertible.
- The above constraints also apply to the spectral decomposition for symmetric matrix.

## Proof of the SVD

SVD arises from finding an orthogonal basis for the row space that gets transformed into an orthogonal basis for the column space:  
 $Av_i = \sigma_i u_i$  (**see later how  $\sigma_i$  is arrived**). In the above equation, we assume  $A$  is  $m \times n$  matrix, while  $v_i$  is  $n \times 1$  row space vector and  $u_i$  is  $m \times 1$  column space vector. It's not hard to find an orthogonal basis for the row space – the GramSchmidt process gives us one right away. But in general, there is no reason to expect  $A$  to transform that basis to another orthogonal basis. However, we **can achieve this if  $v$  and  $u$  are eigenvectors of  $A^T A$  and  $AA^T$  respectively**. The following is a proof from p.368 of MIT linear algebra book. We now explain why  $Av_i$  falls in the direction of  $u_i$ . In other words, why  $A$  transform an orthogonal basis in row space of  $A$  to the column space of  $A$ . The last  $v$ 's and  $u$ 's (in the nullspaces) are easier.

Using the assumption that  $v_i$  is eigenvectors of  $A^T A$ , we have  $A^T Av_i = \sigma_i^2 v_i$ . **Note because  $A^T A$  is positive semi-definite, we can assume the eigenvalues to be  $\sigma_i^2 \geq 0$ .** However, we here consider only the blocks where  $\sigma_i > 0$  and the  $\sigma_i = 0$  part is trivial. First, we show:

$v_i^T A^T Av_i = \sigma_i^2 v_i^T v_i$  gives  $\|Av_i\|^2 = \sigma_i^2$  so that  $\|Av_i\| = \sigma_i$  (The last equation for norm will be used to normalize vector).

To prove that  $Av_i = \sigma_i u_i$ , the key step is to multiply the above equations by  $A$ :

$AA^T Av_i = \sigma_i^2 Av_i$  gives  $u_i = Av_i / \sigma_i$  as a unit eigenvector of  $AA^T$ .  $u_i = Av_i / \sigma_i$  is just what we want  $Av_i = \sigma_i u_i$ . Note here we consider only the case  $\sigma_i > 0$  and thus divided by  $\sigma_i$  is always well defined. In the derivation, we use the trick of placing parentheses. In the second equation, we divide  $Av_i$  by its length  $\sigma_i$  to get the unit vector  $u_i = Av_i / \sigma_i$ . These  $u$ 's are orthogonal because  $(Av_i)^T (Av_j) = v_i^T (A^T Av_j) = v_i^T (\sigma_j^2 v_j) = 0$ .

Therefore, we transform one orthogonal basis in row space to another orthogonal basis in column space of matrix  $A$ . Once we obtain this result  $Av_i = \sigma_i u_i$ , then constructing SVD below is trivial.

## Construction of SVD

From above, we can find in row space of  $A$  a set of special orthonormal basis  $v_1, v_2, \dots, v_r$ , which can be transformed by  $A$  into another special orthonormal basis  $u_1, u_2, \dots, u_r$  in the column space of  $A$ . That is, we have:

$$A \begin{bmatrix} v_1 & v_2 & \dots & v_r \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \dots & \sigma_r u_r \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \dots & u_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{bmatrix}, \text{ i.e., } Av_1 = \sigma_1 u_1, Av_2 = \sigma_2 u_2, \dots, Av_r = \sigma_r u_r.$$

We further consider the orthonormal bases in null and left null spaces and have

$$A \begin{bmatrix} v_1 & v_2 & \dots & v_r & v_{r+1} & \dots & v_m \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \dots & u_r & u_{r+1} & \dots & u_n \end{bmatrix} \begin{bmatrix} \sigma_1 & & & & & & \\ & \ddots & & & & & \\ & & \sigma_r & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & \\ & & & & & & 0 \end{bmatrix}, \text{ where } U \text{ is } m \times m \text{ orthonormal matrix, } \Sigma \text{ } m \times n \text{ is diagonal matrix,}$$

and  $V^T$  is  $n \times n$  orthonormal matrix.

Finally we have  $AV = U\Sigma$ , which is an equation similar to matrix diagonalizing. Matrix  $A$  is transformed into a diagonal matrix  $\Sigma$ . We also note  $U, V$  are two different orthonormal bases. We therefore have  $A = U\Sigma V^{-1}$ . Because  $V$  is orthonormal, we have  $A = U\Sigma V^T$ .

- $v_1, \dots, v_r$  orthonormal basis of row space
- $u_1, \dots, u_r$  orthonormal basis of column space
- $v_{r+1}, \dots, v_n$  orthonormal basis in null space
- $u_{r+1}, \dots, u_m$  orthonormal basis in left null space

### Comments

$A^T A$  and  $AA^T$  is not only symmetric, but also at least positive semi-definite. So the eigenvectors of either  $A^T A$  or  $AA^T$  are orthogonal. Moreover, the eigenvalues of the two matrices are same, as generally the eigenvalues of  $AB$  and  $BA$  are same.

## SVD calculation example 1

We now calculate the SVD for matrix  $A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$ . From the proof part of SVD, we can calculate the eigenvalues and eigenvectors of  $A^T A$  to obtain  $v_i$  and  $\sigma_i^2$ . This can also be understood as follows (although not necessary).  $A^T A = V \Sigma^T U^T U \Sigma V^T$ .  $U$  is orthonormal matrix and thus  $U^T U = I$ , while  $\Sigma^T \Sigma$  is diagonal matrix with element  $\sigma^2$ .

Now we have  $A^T A = V \begin{bmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \ddots \\ & & & \sigma_n^2 \end{bmatrix} V^T$ , where  $V$  is a matrix from the eigenvectors of  $A^T A$  and  $\Sigma^2$  consists eigenvalues.

$A^T A = \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} = \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}$ . For simple matrix, it is straightforward to obtain its eigenvectors.

$A^T A \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 32 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ ,  $A^T A \begin{bmatrix} 1 \\ -1 \end{bmatrix} = 18 \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ . Transforming to unit vector gives  $\sigma_1^2 = 32$ ,  $v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ ,  $\sigma_2^2 = 18$ ,  $v_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$ . \*\*Be careful the

eigenvalue of  $A^T A$  is  $\sigma^2$  but not  $\sigma$ .

To calculate  $u_i$ , we can calculate the eigenvectors of  $A A^T$ . However, this is not necessary. We can directly use its definition  $u_i = A v_i / \sigma_i$ , i.e.  $A v_i = \sigma_i u_i$ . For example:

$$A v_2 = \begin{bmatrix} 0 \\ -\sqrt{18} \end{bmatrix} = u_2 \sigma_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \sqrt{18}. \text{ 则 } u_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

## SVD Calculation example 2

See more details in SVD\_examples.pdf.  $A = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}$  is a rank(1) matrix and thus has null space. The row space of  $A$  is the multiples of  $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$ , while the column space of  $A$  is the multiples of  $\begin{bmatrix} 4 \\ 8 \end{bmatrix}$ .

- Normalizing the above vectors give  $v_1 = \begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix}$ ,  $u_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ .
- $A^T A = \begin{bmatrix} 4 & 8 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} = \begin{bmatrix} 80 & 60 \\ 60 & 45 \end{bmatrix}$ . Because  $A$  is rank(1),  $A^T A$  is not full rank either. Thus it must have a zero eigenvalue. According to the eigenvalue property involving trace, we know the other eigenvalue is 125.
- The eigenvectors for null spaces are  $v_2 = \begin{bmatrix} 0.6 \\ -0.8 \end{bmatrix}$ ,  $u_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ .

Finally we have  $\begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & -1 \\ - & - \end{bmatrix} \begin{bmatrix} \sqrt{125} & 0 \\ 0 & 0 \\ - & - \end{bmatrix} \begin{bmatrix} 0.8 & 0.6 \\ 0.6 & -0.8 \\ - & - \end{bmatrix}$ , where the underscored elements are related to null spaces.

**Comments:** In the above SVD, note the singular values matrix have zero diagonals. In eigen decomposition, eigenvalues cannot be zero.

## Inverse, left /right, pseudo inverse

### Two-sided inverse

$A^{-1}A = I = AA^{-1}$ .  $m \times n$  matrix  $A$  satisfies the relation  $m = n = \text{rank}(A)$ . Or, it is a full-rank square matrix.

### Left inverse (for full column rank matrix)

For full column rank matrix  $A$  with  $m > n = \text{rank}(A)$ ,  $A^T A$  is a full-rank  $n \times n$  matrix. Therefore  $A^T A$  is invertible:  $(A^T A)^{-1} A^T A = I$ . The portion in the bracket  $(A^T A)^{-1} A^T$  is called the left inverse of  $A$ .

$$A_{\text{left}}^{-1} = (A^T A)^{-1} A^T.$$

### Right inverse (for full row rank matrix)

For full row rank  $m \times n$  matrix  $A$  with  $n > m = \text{rank}(A)$ ,  $AA^T$  is a full-rank  $m \times m$  matrix. So  $AA^T$  is invertible:  $AA^T(AA^T)^{-1} = I$ . The portion in the bracket  $A^T(AA^T)^{-1}$  is called right inverse of  $A$ :  $A_{\text{right}}^{-1} = A^T(AA^T)^{-1}$ .

Right multiplying  $A$  to the right inverse gives a row-space projection matrix  $P = A^T(AA^T)^{-1}A$ , which is different from the column space projection matrix.

### Pseudo inverse (both column and row space are not full rank)

- Considering two vectors  $x, y \in C(A^T)$ , then two vectors  $Ax \neq Ay$  in  $C(A)$ . This indicates that from  $C(A^T)$  to  $C(A)$ , transformation  $A$  is a one-to-one mapping. If constrained on these two spaces,  $A$  is an 'invertible' matrix, or it has an pseudo-inverse. In the proof of SVD, we also show this one-to-one mapping from  $Av_i = \sigma_i u_i$ , where  $v_i$  and  $u_i$  are respectively from  $C(A^T)$  and  $C(A)$ .
- $A = U\Sigma V^T x = y \Rightarrow A^\dagger = (U\Sigma V^T)^\dagger = V\Sigma^\dagger U^T$ , where  $\Sigma^\dagger = \begin{bmatrix} \Sigma_1^{-1} & 0 \\ 0 & 0 \end{bmatrix}$  inverses only non-zero singular values.

### Projections and pseudo-inverse

- Comparing to the projection matrix to the column space of  $A$  and left inverse of  $A$ , we have

$$P = AA_{\text{left}}^{-1} = A(A^T A)^{-1} A^T$$

- For the pseudo-inverse  $A^\dagger$ , we also have

$$P = AA^\dagger = AV\Sigma^\dagger U^T$$

- For real inverse matrix  $A^{-1}$ , we have

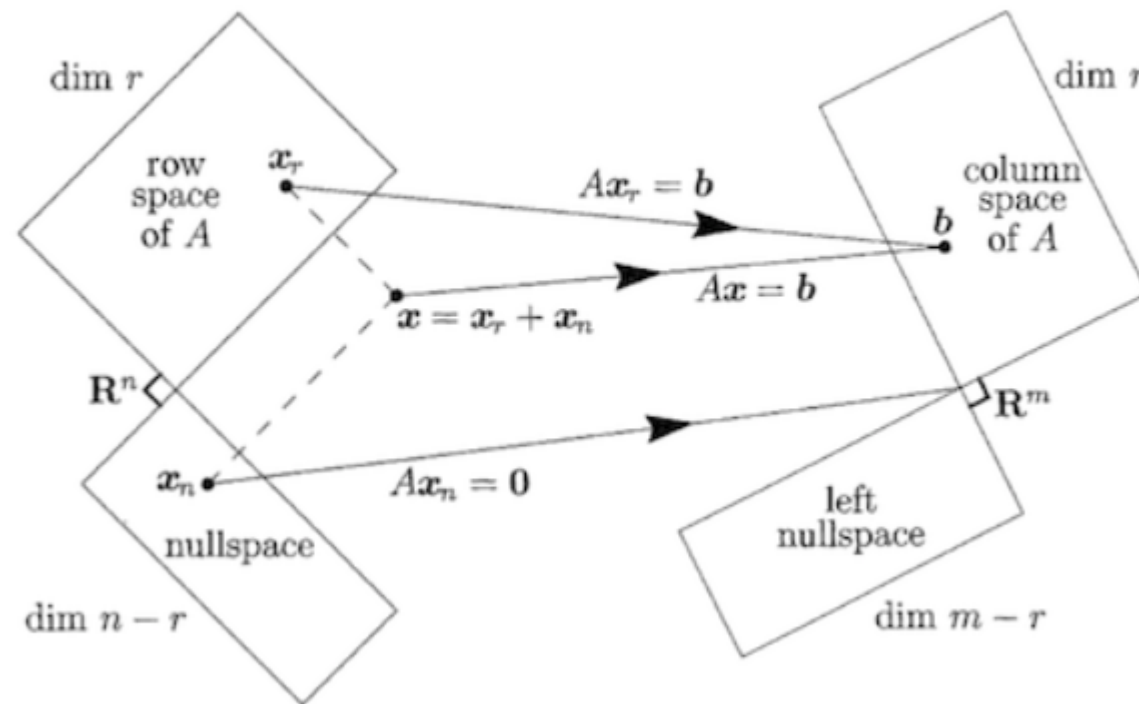
$$P = AA^{-1} = I$$

This also suggest that an invertible matrix cannot be a normal projection matrix unless it is an identity. see.  
<https://math.stackexchange.com/questions/2086814/the-matrix-of-a-projection-can-never-be-invertible>  
<https://math.stackexchange.com/questions/2086814/the-matrix-of-a-projection-can-never-be-invertible>.

### Geometric picture

- Matrix  $A$  has null space and it transform some vectors in  $\mathbb{R}^n$  into zero vector. Any  $\mathbb{R}^n$  vector can be regarded with  $N(A)$  and  $C(A^T)$  portions.  $Ax$  eliminates the  $N(A)$  portion of the  $x$ . However, if we consider only the vectors in row space  $C(A^T)$ , then they are all transformed into the column space  $C(A)$ .

**3F** From the row space to the column space,  $A$  is actually invertible. Every vector  $b$  in the column space comes from exactly one vector  $x_r$  in the row space.



**Figure 3.4** The true action  $Ax = A(x_{\text{row}} + x_{\text{null}})$  of any  $m$  by  $n$  matrix.

### Summary

- In some literatures, left (or may be right) inverse is called pseudo-inverse.
- Three equivalent concepts: generalized inverse; pseudo-inverse; Moore–Penrose inverse. Also in the book of MIT linear algebra, it says that we can have different right and left inverse. The ones we introduced earlier are only one of versions.
- It is the non-trivial null space that causes trouble in inverting a matrix. If  $x$  has non-zero solution in  $Ax = 0$ , then we cannot obtain a  $A^{-1}$  to have  $x = A^{-1} \times 0$ . For pseudo-inverse, it is because we constrain ourselves within the  $r$ -dim row and column spaces where all singular values are not zero.
- There is a formula for the inverse of a partitioned or block matrix.

## Compact forms of SVD

Suppose  $A \in \mathbb{R}^{m,n}$ ,  $A = U\Sigma V^T$  is the SVD of  $A \in \mathbb{R}^{m,n}$  and let  $r$  is the rank matrix  $A$ . Partition  $U$  and  $V$  as follows

$$U = [U_1, U_2], U_1 \in \mathbb{R}^{m,r}, U_2 \in \mathbb{R}^{m,m-r} V = [V_1, V_2], V_1 \in \mathbb{R}^{n,r}, V_2 \in \mathbb{R}^{n,n-r}$$

Thus we have the compact form as below

$$A = [U_1, U_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix} = U_1 \Sigma_1 V_1^T$$

Writing as an outer product form gives

$$A = U_r \Sigma_r V_r^T = \sum_{i=1}^r \sigma_i u_i v_i^T$$

When using outer product form, be careful the meaning of  $v_i^T$ . See earlier notes. Here the meaning of  $v_i^T$  should be different from the  $b_i^T$  introduced before.

## Geometric interpretation of SVD

Because any matrix  $A$  can have a SVD of  $A = U\Sigma V^T$ , we can interpret the effect of  $A$  as rotating the axes first, then scaling (stretching), and then rotating again. Each action is a linear transformation.

## A fundamental problem and the way out

### Rank deficiency of a matrix

- The knowledge of the effective rank of a matrix is of vital importance in solving many linear algebra problems. The rank structure of a matrix fully determine the solution structure of  $Ax = b$ . For example, full rank, full-column rank, full-row rank and rank-deficient in both column and row will give very different solution structure of the equation above ( see details in early chapters).



- Rank deficiency comes from the strongly correlated data (variables, features) in a matrix. In data analysis and machine learning, the strongly correlated data are not only redundant and thus increase calculation burden, but also cause numerical instability and give big variance for predictive models. This is can be seen, e.g., in the pseudo-inverse solution to linear regression problems where tiny singular values as denominator can significantly amplify the errors in the experimental data.

## Eigenvalues and eigenvectors of a matrix

- There are many ways to handle the problem arising from rank deficiency. However, most of them are the variants of eigenvalue/eigenvector analysis. When the experimental data are put in its eigenbasis, it is easy for us to eliminate the tiny singular values and the unnecessary columns in singular vectors.
- Principal component analysis(PCA), usually achieved by singular value decomposition(SVD), is one way to choose major principal components and thus reduce the effect of tiny or zero singular values.

## Regularization

- For problems that cannot be solved with explicit matrix calculations, then regularization is the way to handle the matrix deficiency problem.

In [ ]: