

# Vector space, subspaces

- A vector space is a set that is closed under finite vector addition and scalar multiplication.

## Four subspaces ¶

For  $m \times n$  matrix  $A$ ,  $\text{rank}(A) = r$ , we have:

- row space  $C(A^T) \in \mathbb{R}^n$ ,  $\dim C(A^T) = r$
- null space  $N(A) \in \mathbb{R}^n$ ,  $\dim N(A) = n - r$
- column  $C(A) \in \mathbb{R}^m$ ,  $\dim C(A) = r$
- left null  $N(A^T) \in \mathbb{R}^m$ ,  $\dim N(A^T) = m - r$ . The reason why it is called left null space:  $A^T y = 0 \rightarrow (A^T y)^T = 0^T \rightarrow y^T A = 0^T$
- For  $A_{mn}$ , column space is the subspace of  $\mathbb{R}^m$  and null space is the subspace of  $\mathbb{R}^n$ .
- A null space example. For a particular matrix  $A_{mn} \equiv A_{24}$ , we obtain a system of two equations but four variables (more variables than equations). For example  $x + 2y - z + s = 2$  and  $2x + 4y - 2z + 2s = 4$ . We have two free variables  $z$  and  $s$  and can be set with two independent values  $(z,s) = (0,1)$  or  $(1,0)$ . Whenever, we have free variables, we have non-zero dimension null space. This happens when  $m > n$ .
- We have now two ways of creating a subspace. (1) First we use the linear combination of columns of a matrix and obtain the so called column space  $C(A)$ . Second we can use the solutions to  $Ax = 0$  for form a null space, which is a subspace. The key point is that the two ways both guarantee that the zero vector is included. For row space and left null space, it is similar.
- Note we cannot use the solution of  $Ax = b$ ,  $b \neq 0$  to form a subspace. (1) In this case  $x = 0$  is not necessarily the solution to the equation. However, we require any vector subspace must have zero vector. (2) The solution to  $Ax = b$  is  $x = x_n + x_p = x_p$ , which is given by a null subspace vector shifted by a particular vector, and thus no longer is a subspace (not going through zero vector anymore).
- (i) row operations (such as eliminations) do not change the row space and null space (see the link below); (ii) column operations do not change the column space (and left-null space?). But how to understand row operations does not affect null space.  
<https://math.stackexchange.com/questions/108041/linear-algebra-preserving-the-null-space>  
(<https://math.stackexchange.com/questions/108041/linear-algebra-preserving-the-null-space>).
- The sum dimensions  $C(A)$  and  $N(A)$  is  $n$  for the  $m \times n$  matrix.
- A vectors such as  $v \in \mathbb{R}^4$ , called a 4-dimensional vector. However, we cannot call a matrix contains such vectors as a four-dimensional space. As studied earlier, the dimension of four subspaces for this matrix can be very different.
- When plotting two vectors  $a$  and  $b$  on a paper, NEVER think that they must be 2D vectors ( $a \text{ or } b \in \mathbb{R}^2$ ). In fact, they might be true that  $a \text{ or } b \in \mathbb{R}^4$ , etc.

- A very important relation:  $\dim C(A) = \dim C(A^T)$ . This means that the rank or dimension of column space is equal to that of row space. Be careful, never judge whether a matrix is a basis or invertible only from relations of columns.
- Full rank matrix is not necessarily a square matrix. We need separate column full rank, row full rank, and both row and column full rank.

## Solution structure of $Ax = b$

- $Ax = b$  has solution if and only if  $b \in C(A)$

$r = m = n$	$r = n < m$	$r = m < n$	$r < m, r < n$
$R = I$	$R = \begin{bmatrix} I \\ 0 \end{bmatrix}$	$R = \begin{bmatrix} I & F \end{bmatrix}$	$R = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$
1 solution	0 or 1 solution	$\infty$ solution	0 or $\infty$ solution

- There are two major groups: (1) Both column and row full rank. (2) Other case which include 2-a: full-column rank, 2-b: full-row rank, 2-c: both row and column are not full rank. All these cases can be understood with geometric interpretation.
- 1-a.  $r = m = n$ :  $Ax = 0$  only have a solution  $x = 0$ . Or we can understand as the number of free variable is zero and thus null space is only  $x = 0$ . So the complete solution is  $x = x_n + x_p = x_p$ . In other words, we have solution and only one particular solution. I focused on only this type of solution for many years. The unique solution is  $x = A^{-1}b$ . The reason why we must have a unique solution is that  $b$  and column space are with the same dimension, and thus  $b$  can sure be expanded in the column space.
- 2-a. Only column full rank (memorize as tall matrix). Imagine a  $4 \times 2$  matrix. Full column rank indicates two independent column vectors in  $\mathbb{R}^4$  and thus  $Ax = 0$  has only  $x = 0$  solution. If  $Ax = b$  has solution, then it requires that the  $b$  must lie in the  $C(A)$ . If this is the case, then we have ONE solution. However, it is possible that two column vectors of  $A$  cannot linearly combine to any four-dimensional vector  $b$ . In this later case, there is no solution.
- 2-b. Only row full rank (memorize as fat matrix). Imagine a  $2 \times 4$  matrix. We thus know  $C(A) \in \mathbb{R}^2$ . As shown in lecture 10,  $\dim C(A) = \dim C(A^T)$ , so if there are two independent rows, then there must be two independent columns. Therefore the four columns (two of them are independent) in  $A$  can linearly combine into any  $b$  vector in  $\mathbb{R}^2$ . In other words, we can always find at least one particular solution  $x_p$  to  $Ax = b$ . Now we turn to the solution to  $Ax = 0$ , i.e., the  $N(A)$ . The null space dimension is  $4-2 = 2$  and thus we have two free variables, or have two independent vector in null space. We can obtain the  $x_n$ , which is linear combination of two  $\mathbb{R}^4$  vectors. Then the final solution will be  $x = x_n + x_p$ . The key is this fat matrix case, in 2-dimensional space, there are at least two independent column vectors, and thus we can expand any vector  $b$ , and thus we have at least one particular solution. By the way, the  $N(A)$  is a combination of two independent vectors, and thus is a plane in  $\mathbb{R}^4$  space. Finally, from above, the non-zero  $N(A)$  and the existing  $x_p$  give infinite number of solutions.
- 2-c If column is not full rank, it is like the case of 2-b and we may have a particular solution if  $b \in C(A)$ . Otherwise no nonzero  $x_p$ . If row is not full rank, it is like the case of 2-a, we have nonzero  $N(A)$  and thus may have infinity number of solution. So in this case,

there might be no solution or  $\infty$  solutions.

## Orthogonality

### Orthogonality of four subspaces

For  $m \times n$  matrix  $A$  with rank  $r$ , the row space ( $\dim C(A^T) = r$ ) and null space ( $\dim N(A) = n - r$ ) are orthogonal to each other and both belong to  $\mathbb{R}^n$ . The column space ( $\dim C(A) = r$ ) and left null space ( $\dim N(A^T) = m - r$ ) are also orthogonal to each other and belong to  $\mathbb{R}^m$ .

A best imagination to understand the orthogonality of the four subspaces. (A) in the matrix form of  $Ax = 0$ , note the rows of  $A$  (horizontal direction) are perpendicular to the  $x$  vector (vertical direction). Thus the  $N(A)$  where  $x$  resides is orthogonal to the row space of  $A$ . (B) In the matrix form of  $x^T A = 0$ , the  $x$  is perpendicular to the column vector. So left null space is equal to column space of  $A$ .

When subspace  $S$  and subspace  $T$  are orthogonal, then any vector in  $S$  is orthogonal to any vectors in  $T$ . Two walls cannot be two orthogonal spaces.

Example: For  $A = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \end{bmatrix}$ , we have  $m = 2, n = 3, \text{rank}(A) = 1, \dim N(A) = 2$  From  $Ax = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ , we

obtain a basis of null space:  $x_1 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$   $x_2 = \begin{bmatrix} -5 \\ 0 \\ 1 \end{bmatrix}$ . The basis in row space is  $r = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$ . The null space is orthogonal to the row

space. In this example, row space is the normal vector to the null space.

Row space and null space are orthogonal complement in  $n$  dimensional space, i.e., null space contains all the vectors that are perpendicular to row space. Similarly, column space and left null space are orthogonal complement in  $m$  dimensional space.

## Projection to subspaces

### General picture

- First have a clear picture on how column space  $C(A) \perp N(A^T)$ . Review the imagination picture earlier.
- We want to solve  $Ax = b$ , where  $A$  is a tall matrix. We do not have a solution because  $C(A)$  cannot expand vector  $b$ . For tall matrix,  $b$  might be a four-dimensional vector but  $C(A)$  has a dimension of three. Note even for fat matrix,  $C(A)$  might not expand vector  $b$ . **So**

**it is not just for tall matrix that  $Ax = b$  might have no solution.**

- Because there are no solution in  $C(A)$ , we want an approximate solution  $\hat{x}$ . From the projection picture,  $b$ 's projection to  $C(A)$  is  $p = A\hat{x}$ .
- The error or difference between  $b$  and  $p$  is  $e = b - p = b - A\hat{x}$ .
- Also from the projection picture,  $e$  is normal to all the columns of  $A$ . Thus we have  $a_1^T(b - p) = a_1^T e = 0$ ,  $a_2^T(b - p) = a_2^T e = 0 \dots$  These can be written as  $A^T e = A^T(b - A\hat{x}) = 0$ , or  $e$  is in the  $N(A^T)$ , the left null space of  $A$ . From this we can obtain  $\hat{x} = (A^T A)^{-1} A^T b$ , which is the approximate solution of  $Ax = b$ .
- The above picture of  $e$  perpendicular to  $C(A)$  is extremely important. All the ideas of introducing normal matrix, normal equation, etc. are all from this geometric picture.

## 2D Projection

In projection, vector  $e$  is the error between two vectors  $b, p$ ,  $e = b - p$ ,  $e \perp p$ . Projection  $p$  is on  $a$  and  $p = ax$ . Thus we have  $a^T e = a^T(b - p) = a^T(b - ax) = 0$ . We have the very important equation,  $a^T(b - xa) = 0$ ,  $xa^T a = a^T b$ ,  $x = \frac{a^T b}{a^T a}$ ,  $p = a \frac{a^T b}{a^T a}$ .

$P = \frac{aa^T}{a^T a}$ . If  $a$  is  $n$  dimensional column vector, then  $P$  is a  $n \times n$  matrix. The column space of projection matrix  $C(P)$  is a line going through  $a$  with  $\text{rank}(P) = 1$ .  $a$  is the basis of this matrix.

Projection is a symmetric matrix  $P = P^T$  and thus can always be diagonalizable. Also we have  $P^2 = P$ .

## 3D projection

The projection matrix in 3D can be derived with the similar way in 2D case (see details in course notes). However, here we will use a simple way to understand. The 2D case  $P = \frac{aa^T}{a^T a}$  becomes  $P = aa^T$  when  $a$  is a unit vector. Then  $Px = aa^T x = a_i a_i$ , where  $a_i$  is the projection of vector  $x$  on the vector  $a$ . This is exactly same as the projection operator in physics which usually described by Dirac notation. Extending to 3D case  $P = AA^T$  gives  $Px = AA^T x$ . Each row of  $A^T$  dot multiply with vector  $x$  give the value, which is the projection of  $x$  on this row vector of  $A^T$  or column vector of  $A$ . This gives all the projections of  $x$  on all the column vectors of  $A$ . The projections is denoted by a vector  $y$  and thus  $Px = Ay$ . From the column picture of matrix multiplication,  $Ay$  is just expand with the columns of  $A$ . In other words,  $Px$  projects  $x$  into the column space of  $A$ . If the column vectors of  $A$  is not unit vectors, then  $P = A(A^T A)^{-1} A^T$ .

**Important conclusion:**

- If  $A$  contains **unit column vectors**, then  $AA^T x$  projects  $x$  onto the column space of  $A$ .
- If  $A$  contains **unit row vectors** (required?), then  $A^T Ax$  projects  $x$  onto the row space of  $A$ . So  $Ax$  may have interpreted as: (1) column expansion with coefficients in  $x$ . (2).  $Ax = y$ , then  $y$  is the projections of  $x$  on the row space (the row vectors need to be unit vector).

## A summary to memorize

- $Ax = b \Rightarrow A\hat{x} = p \Rightarrow A\hat{x} = Pb = A(A^T A)^{-1} A^T b \Rightarrow \hat{x} = (A^T A)^{-1} A^T b$  From the pseudo-inverse part, we know that the projection matrix here projecting  $\mathbb{R}^m$  to column space  $C(A)$ . The projection matrix can be written as  $A$  multiplied by left inverse matrix  $A_{left}^{-1} = (A^T A)^{-1} A^T$ . Also from  $Ax = b$  we know that the linear transformation  $A$  transform a vector  $x$  in row space  $C(A^T)$  to column space  $C(A)$ .
- $A$  left multiplying left inverse gives projection matrix which projects  $\mathbb{R}^m$  to  $C(A)$ . If we right multiply  $A$  to right inverse, then it will give a projection matrix projecting  $\mathbb{R}^n$  to row space  $C(A^T)$ .  $A_{right}^{-1} = A^T (AA^T)^{-1}$ . Thus the projecting matrix is  $p = A^T (AA^T)^{-1} A$ .

## Minimum Least Square

The minimum least square method is exactly mapped to the problem of solving  $Ax = b$ . So I can understand everything here with the conclusions arrived earlier. In the following results, we need find the clear geometric significance of terms such as  $p, e, \hat{x}$ . Also what is  $C(A), N(A^T)$ , etc.

We will find a line  $b = C + Dt$  that has minimum deviation from three points  $(1, 1), (2, 2), (3, 2)$ . From this we have the following

equations  $\begin{cases} C + D = 1 \\ C + 2D = 2 \\ C + 3D = 2 \end{cases}$ . The matrix form is  $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ . That is  $Ax = b$ . Obviously, there is no solution to the system of

equations. However, there is solution to  $A^T A \hat{x} = A^T b$ . Multiplying  $A^T$  in both sides gives  $A^T A \hat{x} = A^T b$ , which is the fundamental equation for minimum least square.

Now we are finding the solutions  $\hat{x} = \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix}$  with  $p = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix}$ .

$$A^T A \hat{x} = A^T b$$

$$A^T A = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \quad A^T b = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$$

Converting to equations gives  $\begin{cases} 3\hat{C} + 16\hat{D} = 5 \\ 6\hat{C} + 14\hat{D} = 11 \end{cases}$ , which are also called normal equations. The solutions are  $\hat{C} = \frac{2}{3}, \hat{D} = \frac{1}{2}$ , corresponding to the 'best line'  $y = \frac{2}{3} + \frac{1}{2}t$ . Plugging into the original equations gives  $p_1 = \frac{7}{6}, p_2 = \frac{5}{3}, p_3 = \frac{13}{6}$ . That is,  $e_1 = -\frac{1}{6}, e_2 = \frac{1}{3}, e_3 = -\frac{1}{6}$ .

Thus we have  $p = \begin{bmatrix} \frac{7}{6} \\ \frac{5}{3} \\ \frac{13}{6} \end{bmatrix}, e = \begin{bmatrix} -\frac{1}{6} \\ \frac{1}{3} \\ -\frac{1}{6} \end{bmatrix}$ . Obviously,  $b = p + e$  and  $p \cdot e = 0$ , i.e.,  $p \perp e$ . The error vector  $e$  is not only perpendicular to  $p$

but also to the whole column space. For example,  $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ .

## Why projection is introduced?

- Projection is strongly related to orthogonality.
- Projection is also strongly related to optimization, i.e. minimization or maximization. From the geometrical picture, we know a that a projection is automatically the optimal solution is the space projected into. The key reason is as follows:  $e$  is the error between two vectors, and  $e$  is minimized when vertically (orthogonally) projected to another space.
- To let the normal equations solvable, we need the  $A$  is full column rank, as shown below.

If the columns of matrix  $A$  are linearly independent, then  $A^T A$  is invertible.

First assume  $A^T A x = 0$ , and then multiplying  $x^T$  in both sides gives  $x^T A^T A x = 0$ , i.e.,  $(Ax)^T (Ax) = 0$ . Thus  $Ax = 0$ . Because columns of  $A$  are independent, thus the null space of  $A$  only has zero vector.

## Orthonormal matrix and Gram-Schmidt

orthonormal:  $q_i^T q_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

$$Q = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix}$$

$$Q^T Q = I \quad Q^T = Q^{-1}$$

Advantage of orthonormal matrix:

For ordinary matrix, the projection matrix to its column space is complicated. For orthonormal matrix, if we want to project vector  $b$  to the column space of matrix  $Q$ , then we have a much simpler projection matrix,  $P = Q(Q^T Q)^{-1} Q^T = QQ^T$ . Therefore, when columns of a matrix is orthonormal, then  $QQ^T$  is the projection matrix. In special case of square matrix, then if its columns are orthonormal, then its column space is the whole vector space, and the projection matrix to this whole space is identity matrix, i.e.,  $QQ^T = I$ .

Steps of Gram-Schmidt:

For two linearly independent vectors  $a, b$ , first transfer them into two perpendicular vectors  $A, B$ , and then normalize them with

$$q_1 = \frac{A}{\|A\|}, q_2 = \frac{B}{\|B\|}:$$

- Let  $a = A$
- Projecting  $b$  onto the direction of normal vector of  $A$  gives vector  $B$ . This is just the error vector introduced earlier  $e = b - p$ . That is  $B = b - \frac{A^T b}{A^T A} A$ . Now verify whether  $A \perp B$  is true.  $A^T B = A^T b - A^T \frac{A^T b}{A^T A} A = A^T b - \frac{A^T A}{A^T A} A^T b = 0$  ( $\frac{A^T b}{A^T A} A$  is  $A\hat{x} = p$ ).

For three independent vectors  $a, b, c$ , we need first find their corresponding orthogonal vectors  $A, B, C$  and then normalize them with

$$q_1 = \frac{A}{\|A\|}, q_2 = \frac{B}{\|B\|}, q_3 = \frac{C}{\|C\|}:$$

- We have shown how to obtain the first two orthogonal vectors. Now we need find the third vector which is orthogonal to both  $A, B$ .
- Following the similar approach, we first calculate the projection of  $c$  on  $A, B$ , and then subtract this projection from  $c$ .

$$C = c - \frac{A^T c}{A^T A} A - \frac{B^T c}{B^T B} B.$$

## Properties of orthogonal matrix (unitary matrix)

- It preserves the norm of a vector. An example is rotational matrix. This is easy to prove:

$$\|Ux\|^2 = (Ux)^T (Ux) = x^T U^T U x = x^T x = \|x\|^2$$