

You can also find the official video lecture summary on <https://ocw.mit.edu/courses/mathematics/18-06sc-linear-algebra-fall-2011/ax-b-and-the-four-subspaces/the-geometry-of-linear-equations/> See the left navigation bar on this page, you will get all summaries.

1

Row picture, column picture

Example: For

$$2x - y = 0$$

$$-x + 2y = 3$$

Row picture is that draw $2x - y = 0$, $-x + 2y = 3$ on coordinate and find a point satisfying the two equations. That point is the solution.

Column picture is to solve $x(2, -1)^T + y(-1, 2)^T = (0, 3)^T$. This means combine two vectors $((2, -1), (-1, 2))$ to get the final $(0, 3)$ point. In addition, $x(2, -1)^T + y(-1, 2)^T$ is called **linear combination** of columns. In column picture, for $Ax=b$, in view of x , every row of A makes a transform on x , resulting every new dimension value which assemble b ; in view of A , combination of all its columns weighted by x , resulting b .

So when calculate Ax , we can use row picture or column picture.

2

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} c & d \\ a & b \end{bmatrix}$$

You can get the left 0, 1, 1, 0 matrix by imagining column picture. The result re-permutes the origin in row. If you want to re-permute it in column, write the 0, 1, 1, 0 matrix at the right-hand side of $[a, b; c, d]$.

3

Part 1.

There are four ways to calculate $AB = C$.

1. The commonest way is $C_{ij} = \sum A_{ik} B_{kj}$.
 2. But you can also focus on one column of B or C at one time. In this view, $C_{i-col} = A * B_{i-col}$. It's exactly column picture in section 1. So all columns in A are combined, weighted by one column of B at one time, resulting many columns, and these columns assemble C.
 3. The third way is to focus on one row of A or C at one time. In this view, $A_{i-row} = A_{i-row} * B$. You can think it with imagining transpose of $C_{i-col} = A * B_{i-col}$ in the second way.
 4. $C = \sum A_{i-col} * B_{i-row}$, the last way. I think this is the most magical way:)
- In addition, you can also use trick of block.

Part 2.

Inverse: $AA^{-1} = A^{-1}A = I$.

When is a matrix A invertible?

An invertible matrix is called **non-singular/invertible**. From the first term, you can calculate A's determinant. If it's 0, A is singular.

But there is also a magical way, if you can find a vector $x \neq 0$, satisfies $Ax = 0$, A is singular. Why? If $Ax = 0$, $A^{-1}Ax = x = 0$, but $x \neq 0$. Why it's magical? Imagine when you can get $x \neq 0$, $Ax = 0$? By column picture, it means that we can get at least one (minus) column of A by combining all of the rest columns. That's called **linear dependence**!

When A is invertible, how to get its inverse?

Concatenate A and I, write as $[A \ I]$, and transform A to I with elementary transformation. The result is $[I \ A^{-1}]$. This is called **Gaussian-Jordan elimination**. Why does it work? If $A \rightarrow I$, the \rightarrow must be multiplied by A^{-1} . So

$I \rightarrow ?$, the ? must be $I * A^{-1} = A^{-1}$. Of course, you can use other ways to get A^{-1} , but I think this is the easiest way.

4

$A = LU = LDU$, here L is lower, D is middle, U is upper.

Specially, how many operations do we need to get $A_{n*n} = LU$?

We can use elementary transformation to transform A to U, the elementary matrix is L^{-1} . To get U, first just eliminate the first column(except for the first row), so we need about n^2 operations(take multiply and add as one operation and let the first row also operate on itself). Then only focus on the sub-matrix $A_{(n-1)*(n-1)}$, here we need about $(n-1)^2$ operations. The final number is $\sum_1^n i^2 = (n * (n+1) * (2n+1))/6$, simply taking it as $n^3/3$.

How many operations do we need to get $A_{n*n} = LDU$?

We can get $A = LU'$ first, and then get $U' = DU$. The first step take about $n^3/3$ operations, and the second step take $n(n+1)/2$ operations. You can get this number similar to $A = LU'$ step.

5

Part 1. Permutation matrix

Permutation matrix is identity matrix I with reordering rows, including I itself.

- There are $n!$ permutation matrixes for I_{n*n} .
- For Permutation matrix P, $P^{-1} = P^T$.
- $P^{-1} = P^T$ is also permutation matrix.

Part 2. Transpose

For $A = RR^T$, A is symmetric, $A = A^T$. Because $A^T = (RR^T)^T = RR^T = A$.

Part 3. Vector space, Column space, Vector subspace

A vector space is defined as result of addition of any two vectors in this space or multiply by a scalar still lies in this space.

Column space is defined as the set of all possible linear combinations of a matrix's column vectors.

A subspace(vector subspace) is a subspace of vector space, and it also is a vector space(vector space inside vector space). For example, subspace of R^3 is set of R^3 , any plane R^2 through origin, any line R through origin. You can see that any subspace must through origin, because we can multiply a vector in the space by 0, resulting origin.

You can see that a vector space or a subspace must be R^n . It can't be part of R^n .

6

Part 1.

Intersection of subspaces S and T, $S \cap T$, is also a subspace, because a subspace must be R^n , so the intersection is $\min(S, T)$.

Part 2. Nullspace

For $Ax = b$, we can get $\{x | Ax = b\}$, but is it a vector space?

When b is not 0, it's not a vector space. Because x can't be 0, and a vector space must contains 0. In detail, $\{x | Ax = b\}$ is a point, line, plate or space that doesn't go through origin. For example, you get $x(1, 0, 0)^T + y(0, 1, 0)^T + z(0, 0, 0)^T = (1, 1, 0)$ in column picture. The solution is $(1, 1, n)$, n can be any value, but the solution always can't be $(0, 0, 0)^T$.

But when b is 0, it's a vector space. Why? Suppose w, v are two solutions of $Ax = b$, $A(aw + bv) = a(Aw) + b(Av) = 0$, so $aw + bv$ is also in $\{x\}$. (a, b are scalars, w, v are vectors). Specially, for A, $\{x | Ax = 0\}$ is called A's **nullspace**, $N(A)$.

7

To get nullspace of A, (i) Use **Reduced Row-Echelon Form(RREF)** and reorder columns to get $[I, F; 0, 0]$. (ii) take inverse-reorder of (i) on rows of $[-F, I]^T$. Nullspace is column space of $[-F, I]^T$.

Explanation and example

For $Ax = 0$, take three steps. $\{x\}$ is nullspace.

$$\begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{bmatrix} \xrightarrow[\text{change}]{\{x\} \text{ doesn't change}} \begin{bmatrix} 1 & 2 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[\text{changes}]{\{x\}} \begin{bmatrix} 1 & 0 & 2 & -2 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$$

The first arrow doesn't change $\{x\}$, because we only take elementary transformation on rows, $EAx = E0 = 0 = Ax$ (see section 2 if you forget it should be AEx or EAx). This process is called **Reduced Row-Echelon Form, RREF** at matrix view.

The second arrow changes $\{x\}$, because we reordered the columns. In view of elementary transformation on columns, $AEx \neq (0 = Ax)$. In another way, think about column picture, if we reorder the columns, we have to inverse-reorder the dimensions(rows) of x to get correct solution.

If we rewrite the third matrix in blocks, as the final matrix above, we can immediately give some solutions of $A^{third}x = 0$. They are $[-F, I]^T$. Note here $[-F, I]^T$ is size of 4×2 because F and I are both size of 2×2 . So reordering the rows (here 2,3) of $[-F, I]^T$ gives us two **special solutions** of $Ax = 0$. And magically, the nullspace is just linear combination of the columns of $[-F, I]^T$, i.e., column space.

Why the column space of $[-F, I]^T$ is nullspace of A?

See the second matrix, we know here are four variables, but when we give any values to any two variables, the left two are determined. So we only need to choose two variables, and assign all possible values to the two variables, we

get nullspace. That's exactly column space of $[-F, I]^T$ with choosing the two variables as dimensions of column of I. The chose variables are called **free variables**. Free variables corresponds to the columns that can be linear combination of the rest columns.

8

Part 1. Get $\{x|Ax = b\}$

$\{x|Ax = b\}$ can be get through an arbitrary solution(formally, **particular solution**) of $Ax = b$ plus every vector in nullspace of A.

Explanation:

For given x^p and an arbitrary x^q , which are both in $\{x|Ax = b\}$, $x^q - x^p$ is a vector v lying in $N(A)$, because $A(x^q - x^p) = Ax^q - Ax^p = b - b = 0$. So an arbitrary x^q in $\{x|Ax = b\}$ can be rewritten as $x^p + v, v \in N(A)$, and in this process, you would not get more solutions($x^p + v$ is not a vector in $\{x|Ax = b\}$), because $A(x^p + v) = Ax^p + Ax^v = b + 0 = b$.

You can get that $N(A)$ and $\{x|Ax = b\}$ are two parallel planes from the explanation above.

Part 2. How many x exist for $A_{m \times n}x = b$? $r = \text{rank}(A)$.

- If $r = n < m$, there may be 0 or 1 solution.
- If $r = m < n$, there may be 1 or infinite solutions.
- If $r < m, r < n$, there may be 0 or infinite solutions.

Explanation:

In summary:

First, using rref, convert A in $Ax = b$ to a combination of $I, F, 0$, and then think the equation by column picture, you will get the answer quickly.

In detail:

When $r = n < m$, take rref on both side of $Ax = b$, you get $\hat{A}x = \hat{b}$, rewrite \hat{A}

in blocks $\hat{A} = [I, 0]^T$. For $[I, 0]^T x = \hat{b}$, think about column picture, there may be 1 or 0 solution, which depends on the positions of 0s in \hat{b} . When $r = m < n$, $\hat{A} = [I, F]$, the number of solutions depends on whether $b = 0$. Here F means free, as all columns in F can be a linear combination of A, and so $r(A) < n$. You may be wondering why $\hat{A} = [I, F]$, not $[I, 0]$. That's because rref only allow us to do operations on rows, not columns. When $r < m, r < n$, $\hat{A} = [I, F; 0, 0]$, it's a combination of the two situations above. I believe you can derivate the answer yourself.

9

Part 1. Convenient ways to classify linear dependence

We have known the definition of linear dependence in section 3, part 2. But how can we easily know whether a bunch of vectors are linear dependence or independence? The easy ways are through $\#(\text{free variables})$ or $N(A)$. If there is not free variable(in x of $Ax = 0$) or $N(A)$ only contains zero vector, these vectors are linear independence.

Explanation:

Definitions of linear dependence in 3.2 and free variable in 7 immediately give that: if there is not free variable, these vectors are linear independence.

The definitions of linear dependence in 3.2 also equals to that: vectors x_1, x_2, \dots, x_n are independent if no combination gives zero vector(except the zero vector), i.e., $c_1x_1 + c_2x_2 + \dots + c_nx_n \neq 0$. Think about the two definitions and you will quickly know they are equal. Then the latter definition immediately equals to $N(A)$ only contains zero vector.

Part 2. Span, basis,

The space consists all of combinations of some vectors are commonly used. So give it a shorter name. Vectors v_1, v_2, \dots, v_n **span** a space.

So for one space, we can have many set of vectors to span it. It's natural that we are interested to sets that contain least vectors. Give any set contains least vectors a name, **basis**.

Part 3. understand rank, $\text{rank}(A) + \# \text{dimensions of } N(A) = (\text{rows of } A_{m \times n} \text{ or } \# \text{variables})$

I know you know rank. And rank is associated to many and many theorems in linear algebra. But what's the best way to understand rank? I think the best way is understand it as: $\# \text{pivots}(\text{dimensions})$ of a matrix A 's column space. Why I think this is the best way? Remember that in this notes, I tend to, and used to use column picture to depict everything I can. So this understanding is coherent and compatible with column picture. You can make connections with different concepts using this understanding quickly and easily.

$$(\# \text{rows of } A_{m \times n} \text{ or } \# \text{variables}) = \text{rank}(A) + \# \text{dimensions of } N(A).$$

It's nice but, why? Think about $Ax = 0$, if $\text{rank}(A) = m$, it's easy to show this rule is correct. if $\text{rank}(A) < m$, A can be converted to $\hat{A} = [I, 0]^T$ with rref. Here $\# \text{dimensions of } I$ is $\text{rank}(A)$ and think about column picture, $\# \text{dimensions of } 0$ is $\# \text{dimensions of } N(A)$.

10

A matrix $A_{m \times n}$ has four subspaces, column space $C(A)$, nullspace $N(A)$, rowspace $C(A^T)$, nullspace of A^T $N(A^T)$ (also called **left nullspace**).

Dimensions of $C(A), C(A^T)$ are both $r(A)$. Dimension of $N(A), N(A^T)$ are $n - r(A), m - r(A)$.

All above make sense if you understand all previous sections:)

How to get $C(A), N(A), C(A^T), N(A^T)$?

First of all, to get a space, you just need to find one of its basis.

For $C(A^T)$, take rref, the non-zero rows are one basis.

For $C(A)$, take reduction on columns, the non-zero columns are one basis.
 For $N(A)$ and $N(A^T)$, see section 7. This course also taught a other way to get $N(A)$ and $N(A^T)$. But I think it's more complicated compared to section 7 as we need to write out an elementary matrix.

11

Part 1. Matrix space, solution space

If we define a space as result of addition of any two elements in this space or multiply by a scalar still lies in this space, we can get matrix space and solution space. Similarly, we can also get basis and dimension of these spaces.

For example, for all 3*3 symmetric matrix, it's a (matrix) space. One kind of basis is as below, and its dimension is 6 as there are 3 free variables.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \dots \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

For example, all solutions of $\frac{d^2y}{dx^2} + y = 0$ is a (solution) space. One kind of basis is $\cos x$ and $\sin x$, and its dimension is 2.

For example, the subset of M containing all rank 4 matrices is not a space, even if we include the zero matrix, because the sum of two rank 4 matrices may not have rank 4.

Part 2. Rank 1 matrices

For a matrix A with $\text{rank}(A) = 1$, it can be decomposed as $A = A_{col1}A_{row1}$.

For example,

$$\begin{bmatrix} 1 & 4 & 5 \\ 2 & 8 & 10 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 4 & 5 \end{bmatrix}$$

Part 3. Graph

In this class, a **graph** G is a collection of nodes joined by edges: $G = \{nodes, edges\}$.

12

Application of graph.

13

Review and exam.

14

$\|x\| = x^T x$, this is useful when you want to take some operation on x , for example, derivation.

Part 1. Orthogonal space

Subspace S is **orthogonal**(\perp) to subspace T is defined as every vector in S is orthogonal to every vector in T . This is the same as two plane are orthogonal, which we learned in high school. Note that here S and T can have different dimension.

For a matrix A , its row space is orthogonal to its nullspace.

Explanation and example:

For $Ax = 0$ like this,

$$\begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Immediately, $[1, 2, 2, 2][x_1, x_2, x_3, x_4]^T = [2, 4, 6, 8][x_1, x_2, x_3, x_4]^T = [3, 6, 8, 10][x_1, x_2, x_3, x_4]^T = 0$. And this can easily generalize to the two whole space.

Furthermore, row space and nullspace are not only orthogonal, but also **complementary** in R^n . Complementary here means that nullspace contains all vectors that \perp row space, and vice versa.

Part 2. $A^T Ax = A^T b$

Sometimes there is no solution for $Ax = b$, but we still want to get a solution.

So we calculate $A^T Ax = A^T b$. More detail is in next section.

$N(A^T A) = N(A)$, because $Ax = 0 \leftrightarrow A^T Ax = A^T 0 \leftrightarrow A^T Ax = 0$. So $A^T A$ is invertible if $N(A)$ only contains zero vector, i.e., columns of A are independent.

15

Part 1. Why is it $A^T Ax = A^T b$ in 14.2?

In summary

First, in $Ax = b$, as it has no solution, replace b with p (the vector in $C(A)$ closest to b), get $Ax = p$. Second, column vectors in $A \perp (b - Ax)$, so $A^T(b - Ax) = 0$, $A^T Ax = A^T b$.

In detail

Okay, suppose size of A, x, b are $3 \times 2, 2 \times 1, 3 \times 1$, columns of A are a_1, a_2 . $Ax = b$ has no solution means b is not in column space of A, $C(A)$. So we replace b with a vector p (projection) in $C(A)$ closest to b . Now we have $Ax = p$ and it has a

solution. What's more, we know that $a_1, a_2 \perp (b - p)$, because only in this way p is a vector in $C(A)$ closest to b . So $a_1^T(b - p) = a_2^T(b - p) = a_1^T(b - Ax) = a_2^T(b - Ax) = 0$, write it in matrix $A^T(b - Ax) = 0$, i.e., $A^T Ax = A^T b$.

Note

- a. We never explicitly write out p above. We have $Ax = p$, but we don't know x and p . We join $Ax = p$ and the orthogonal case to calculate x and p .
- b. If you think p is $A^T b$, you are wrong. Because $p = Ax \neq (A^T b = A^T Ax)$. We need to calculate x first to get p .
- c. If you think $A^T Ax = A^T b$ then $Ax = b$, you are wrong again. Because to get the latter equation, you need to left multiply $(A^T)^{-1}$ at both side of the former equation. But $(A^T)^{-1}$ doesn't exist.

Part 2. Projection matrix

In part 1, we project b to p with a **projection matrix** P . $P = A(A^T A)^{-1} A^T$ in part 1 and $P = P^T$, $P = P^2 = P^n$ for this P . You can verify it easily.

Part 3. Application of $A^T Ax = A^T b$, least mean square error

If there is noise in our data, when apply all points to a fitting function and write all these fitting equations with $Ax = b$, it has no solution. But $A^T Ax = A^T b$ has a solution, what's more, it's the best solution with least mean square error. Magical! More detail and explanation is in next section.

~~(Redundant) Here x is the variables of our hyperplane, A is coefficient of the fitting function, b is function value. For example, For $Ax = b$ and fitting function $y = ax_1 + b_1$, x is x_1 , A is a and b_1 , b is y .~~