

Linear Algebra Self-Study

Kansapat Udomchokpaiboon (Hu Jinlin)

August 4, 2025

Contents

1	Vector and Matrices	4
1.1	My bad guys	4
2	Solving Linear Equation $Ax = b$	5
2.1	Elimination and Back Substitution	5
2.2	Elimination Matrices and Inverse Matrices	5
2.3	Matrix Computation and $A = LU$	5
2.4	Permutation and Tranposes	5
2.5	Derivatives and Finite Difference Matrices	5
2.5.1	Properties of K	6
2.5.2	Free-fixed Matrice T_n	6
2.5.3	The Free-Free Matrices B_n are Singular	7
3	The Four Fundamental Subspaces	8
3.1	Vector Spaces and Subspaces	8
3.1.1	Examples of Vector Spaces	9
3.1.2	Subspaces of Vector Spaces	9
3.1.3	The Column Space of A	9
3.1.4	The Row Space of A	9
3.2	Computing the Nullspace by Elimination: $A = CR$	9
3.2.1	Elimination from A to $\text{rref}(A)$: Reduced Row Echelon Form	10
3.2.2	Elimination Column by Column: The Steps from A to R_0	11
3.2.3	The Matrix Factorization $A = CR$ and the Nullspace	11
3.2.4	Block Elimination in Three Steps: Final Thoughts	13
3.3	The Complete Solution to $Ax = b$	13
3.3.1	One Particular Solution $Ax_p = b$	14
3.3.2	Full Row Rank and the Complete Solution	14
3.4	Independence, Basis, Dimension	15
3.4.1	Linear Independence	16
3.4.2	Vectors that Span a Subspace	16
3.4.3	A Basis for a Vector Space	17
3.4.4	Dimension of a Vector Space	17
3.4.5	Bases for Matrix Spaces and Function Spaces	17
3.5	Dimensions of the Four Subspaces	18
3.5.1	The Four Subspaces for R_0	19
3.5.2	The Four Subspaces for A	19
3.5.3	Rank Two Matrices = Rank One plus Rank One	20
4	Orthogonality	22
4.1	Orthogonality of Vectors and Subspaces	22
4.1.1	Combing Bases from Subspaces	23
4.2	Projections onto Lines and Subspaces	24
4.2.1	Projection onto Lines and Subspaces	24
4.2.2	Projection Onto a Line	25
4.2.3	Projection Onto a Subspace	26
4.3	Least Squares Approximations	28
4.3.1	Minimizing the error	28

4.3.2	The Big Picture for Least Square	29
4.3.3	Fitting a Straight Line	29
4.3.4	Dependent Columns in A: What is \hat{x} ?	30
4.3.5	Fitting by a Parabola	30
4.4	Orthogonal Matrices and Gram-Schmidt	31
4.4.1	Projection QQ^T Using Orthonormal Bases: Q Replaces A	32
4.4.2	The Gram-Schmidt Process	33
4.4.3	The Factorization $A = QR$	34
4.5	The Pseudoinverse of a Matrix	35
4.5.1	The Pseudoinverse A^+ ($n \times m$) of a Matrix A ($m \times n$)	36

Chapter 1

Vector and Matrices

1.1 My bad guys

So, the issue, right? I picked up LaTeX after I already finished chapter 1 , and nearly finished chapter 2. So to future me and any poor sods reading this, good luck lol.

Chapter 2

Solving Linear Equation $Ax = b$

2.1 Elimination and Back Substitution

I fucked up

2.2 Elimination Matrices and Inverse Matrices

This too.

2.3 Matrix Computation and $A = LU$

This one too.

2.4 Permutation and Tranposes

Also this shit.

2.5 Derivatives and Finite Difference Matrices

Second difference matrices includes K, T, B . They all have the $-1, 2, -1$ pattern.

$$K_4 = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}$$

Now we can approximate $-\frac{d^2u}{dx^2} = f(x)$. So, we want to compute $-\frac{d^2u}{dx^2}$ with a computer, but the computer can't understand derivative. So what we do is we turn $\frac{d^2u}{dx^2}$ into the matrix $\frac{K^2}{h}$, function $u(x)$ into vector u , and function $f(x)$ into F . We also need the boundary conditions, which are given where $u(0) = 0$ and $u(1) = 0$. We can't pick out the infinite space between 0 and 1, so we pick N equally spaced points at a regular interval. The space between each points (and the first and the last point) becomes meshwidth (h). If we have N internal points u_0, u_1, u_2, \dots plus two boundary points u_0 and u_{N+1} , we divide the total length into $N+1$ segments. Therefore the spacing is $h = \frac{1}{N+1}$. If we have 4 N , then the spacing is $h = \frac{1}{5}$. So instead of finding the continuous function $u(x)$, we will find the value at each internal points, and they becomes the unknown vector $U = [u_1, u_2, u_3, u_4]^T$.

$$-\frac{d^2u}{dx^2} = f(x) \text{ becomes } \frac{KU}{h^2} = F, \frac{1}{h^2} \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} f(h) \\ f(2h) \\ f(3h) \\ f(4h) \end{bmatrix}$$

The key point is that when divide the meshpoint into 4, therefore $N = 4$. Row 1 times U is $2u_1 - u_2$, and we already got the boundary where $u_0 = 0$ and $u_5 = 0$, making a typical row $\frac{(-u_1 + 2u_2 - u_3)}{h^2} = f(h)$. The division by h^2 makes $\frac{K}{h^2}$ a second difference matrix, replacing $-\frac{d^2u}{dx^2}$.

2.5.1 Properties of K

K has 4 properties. For the sake of example, we will use K for $N = 4$.

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}$$

1. K is symmetrical, as in $K_{ij} = K_{ji}$
2. K is banded. All the non-zeros $(-1, 2, -1)$ lie in a band around the main diagonal. The band has three diagonals, so K is a tridiagonal matrix.
3. K has a constant diagonals. A diagonal of -1, then 2, then -1 again. The matrix is called "shift-invariant", because the differential equation always have a constant coefficient of -1. The approximation to $-\frac{d^2u}{dx^2}$ is always $-1, 2, -1$ at every X .
4. X is invertible. It has an inverse matrix K^{-1} then $K^{-1}K = I$ and $KK^{-1} = I$.

$$K_4^{-1} = \frac{1}{5} \begin{bmatrix} 4 & 3 & 2 & 1 \\ 3 & 6 & 4 & 2 \\ 2 & 4 & 6 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

K^{-1} is also symmetric but it is no diagonal. It is dense matrix, meaning no zeros.

5. Symmetric K_n matrices are positive definite.

Invertible, positive definite symmetric matrix, and semidefinite matrix are defined by their pivots.

1. Invertible matrices has nonzero pivots.
2. Positive definite symmetric matrices has positive nonzero pivots.
3. Positive semidefinite symmetric matrices has nonnegative pivots.

2.5.2 Free-fixed Matrice T_n

T_n and B_n are variations on K_n , where the variation comes from changing the boundary conditions. Think of it as an elastic band that are fixed at both ends, one end, or totally free at both ends. For example, T_n is very similar to K_n except that input $(1, 1)$ is switched from 2 to 1., representing a free boundary condition where $\frac{du}{dx} = 0$

$$\text{Free-fixed boundary conditions, still positive definite } T_4 = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}$$

T is no longer Toeplitz because its main constant, though it does have a simpler factorization than K; every pivot of T equals 1.

$$T = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ 0 & -1 & 1 & \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 0 & 0 \\ & 1 & -1 & 0 \\ & & 1 & -1 \\ & & & 1 \end{bmatrix} = LU$$

Note that $U = L^T$. Notice that U is a forward difference while L is a backward difference. Together, they add up to be a second difference, meaning that $x_{i+1} - 2x_i + x_{i-1}$ correspond to $[-1, 2, -1]$, meaning that T is a second difference.

2.5.3 The Free-Free Matrices B_n are Singular

In this context, "singular" means "not invertible". One test is simply seeing if determinant equals zero or not.

Theorem 2.5.1. *If B multiplies a nonzero vector x to produce $Bx = 0$, then B can't be invertible.*

For example, free-free matrix has 1 (and not 2) in its (1, 1) and (3, 3) input.

$$B_3 = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \text{ has } B_3 x = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

B_n is singular while T_n and K_n are invertible because T_n and K_n have a fixed end, which allows them to adjust, whereas B_n is free on both end so it can't adjust.

Chapter 3

The Four Fundamental Subspaces

How can we define "vector space"? Well, if we are talking about R^3 , the key operation are $v + w$ and cv . Notice that v and w could be matrices, so we could have matrix spaces and function spaces. Then inside R^n we could only allow x that satisfies $Ax = 0$, which will produce "nullspace of A ". All combination of solution to $Ax = 0$ are also solutions, meaning that the nullspace is a subspace. To point it simply, nullspace is just x in Ax that crushes it down to 0. Why is it called a space? Because it has structures and rules.

1. It must contain zero vector, $A \times 0 = 0$ always.
2. It must be closed under addition, meaning that if you take two vectors from the nullspace then add them together, they must still be in the nullspace.
3. It must be closed under scalar addition, meaning that if you take any vector from the nullspace then multiplies them by a constant, the result must still be in the nullspace.

Then, lastly, there are basis. A set of vectors that perfectly describes the space. Very important, some considers it the fundamental theorem. So, what it means is that basis are just sets of movement vectors for each dimension.

1. For a 3D space, you will need forward, right, and up.
2. For a 2D space, you will need right and up.
3. For a 1D space (line of sight, nullspace), you only need the direction of the line.

And so, $n - r$ special solutions to $Ax = 0$ are a basis for $N(A)$. It's called rank-nullity theorem.

Dimension of Input Space = Dimension of Column Space + Dimension of Nullspace, $n = r + (n - r)$

Note that $n - r$ is the dimension of the nullspace. One last note: THIS CHAPTER IS VERY IMPORTANT.

3.1 Vector Spaces and Subspaces

Here are a few fundamental points:

1. All linear combination of $cv + dw$ must stay in the vector space (What is a vector space? Very simply, all the possible spaces that can be achieved by your given vector with vector addition or scalar multiplication), where c and d are scalar, and v and w are vectors.
2. The row space (meaning all possible linear combination of the row vectors) is "spanned" (made up of) rows of A . While the column of A spans $C(A)$.
3. Matrices can be filled by more than just numbers. As long as it obeys the rules of a vector space, it can be treated as a vector. For example, we have two equations, $f(x) = x^2$ and $g(x) = 2x$. Can they be added together? Yes! $h(x) = x^2 + 2x$. Now if we fill it with the likes of \sin, \cos, x, x^2 , we can "span" and build a lot of other functions. For example, all quadratic polynomials are "spanned" by the functions $f_1(x) = 1$, $f_2(x) = x$, and $f_3(x) = x^2$

R^n contains all column vector v to the length of n . For this case, the components from v_1 to v_n are all real numbers. However, if they allow for complex numbers (i), the R^n becomes C^n . To reiterate, all linear combination of $cv + dw$ must be in the vector space R^n . For example, all positive the set with all positive (meaning no vector consist of ANY nonpositive numbers) vectors (v_1, \dots, v_n) are NOT a vector space. Why? Simply take one simple vector, say $(1, 2)$ then multiply it by scalar of, say, $c = -1$. $(-1, -2)$ is NOT in our set, therefore it is not a vector space. Or for another example, a set of solution for $Ax = (1, \dots, 1)$ is not a vector space because a line in R^n is not a vector space unless it goes the central point $(0, \dots, 0)$.

3.1.1 Examples of Vector Spaces

Here are some examples of a neat vector space, the Z (zero vector) where $0 = (0, 0, \dots, 0)$. Combinations of $c0 + d0$ are all still 0, so still in the subspace. How about vector space of matrices? We can do that. $R^{3 \times 3}$ is a space that contains all 3×3 matrices. It does satisfy all eight rules, so why not? It's also a vector space. How about a vector space of functions? Sure can. The line of functions $y = ce^x$ (any c) is a line in a function space. This line contains all solutions to the differential equations of $\frac{dy}{dx} = y$. Yet another function space contains all quadratics $y = a + bx + cx^2$, where they are the solutions to $\frac{d^3y}{dx^3}$. And to reiterate, space in this context means all possible linear combination of the vectors or matrices or functions, and they all stay inside it.

3.1.2 Subspaces of Vector Spaces

What ar subspaces? To put it simply, they are a flat plane inside the dimensional space, however, they are still the same dimension. Let's say, we got a R^3 space. We can make a plane any way we want as long as it passes $(0, 0, 0)$, what we get may look like a 2D plane, but it's still 3D. Therefore, the plane is a subspace of the full vector space R^3 .

Here is a list of possible subspaces of R^3 :

1. Any line through $(0, 0, 0)$
2. Any plane through $(0, 0, 0)$
3. The whole space R^3
4. The zero vector $(0, 0, 0)$

3.1.3 The Column Space of A

What we are trying to solve here is $Ax = b$. We want to know b , right? Well, b are a column space of A . Ax is just a combination of A , and to get every possible b , we need all possible x , which is just all linear combination of A , which is the column space of A , as written earlier. To build on that, vector space is made up of column vectors.

A crucial point to understand is that, **to solve $Ax = b$ is just to express b as a combination of the columns**. b got to be in the column space of A , otherwise, it doesn't exist!

Caution: columns of A do not form a subspace. Neither do invertible matrices, or singular matrices. Only all linear combinations.

3.1.4 The Row Space of A

The rows of A are the column of A^T , why do we do this? Because we like working with columns, so we use the column of A^T

The row space of A is just the column space of A^T

3.2 Computing the Nullspace by Elimination: $A = CR$

1. The **nullspace $N(A)$** in R^n contains all solutions x to $Ax = 0$, including $x = 0$.
2. CONTINUE THIS LATER

The goal of this section is to find all solutions for $Ax = 0$. If A is an invertible matrix, then the only solution is $x = 0$. In general, A has r independent columns, the other $n - r$ are a linear combination. Here is a matrix R with rank $r = 2$, with $n = 4$ columns. This means we have $n - r = 4 - 2 = 2$ independent solutions to $Rx = 0$. So the nullspace $N(R)$ will have 2 dimensions.

Example 1: $R = [IF]P = \begin{bmatrix} 1 & 0 & 3 & 5 \\ 0 & 1 & 4 & 6 \end{bmatrix}$

Which means $Rx = 0$ is $x_1 + 3x_3 + 5x_4 = 0$ and $x_2 + 4x_3 + 6x_4 = 0$. We find the special solutions by just letting x_3 and x_4 equal 1 and 0 or 0 and 1. Set $x_3 = 1, x_4 = 0$ the equations will give $x_1 = -3, x_2 = -4$. Set $x_3 = 0, x_4 = 1$ the equations will give $x_1 = -5, x_2 = -6$. This gives us two special solutions: $s_1 = (-3, -4, 1, 0)$ and $s_2 = (-5, -6, 0, 1)$. They are both in the nullspace of R , as we can also see, $cs_1 + ds_2$ is still in the the nullspace. And so, s_1 and s_2 are the basis of nullspace.

Example 2: $R_0 = \begin{bmatrix} 1 & 7 & 0 & 8 \\ 0 & 0 & 1 & 9 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Which means $x_1 + 7x_2 + 0x_3 + 8x_4 = 0$, $x_3 + 9x_4 = 0$, and $0 = 0$ (wow on the last one). The matrix identity is inside column 1 and 3, and row 3 is all zero, which makes it a reduced row echelon form, even elimination can't make it simpler. We still have free variables for the special solution, namely, x_2 and x_4 . Set $x_2 = 1, x_4 = 0$, the equations give $x_1 = -7, x_3 = 0$. Set $x_2 = 0, x_4 = 1$, the equations give $x_1 = -8, x_3 = -9$. The special solutions are noe $s_1 = (-7, 1, 0, 0)$ and $s_2 = (-8, 0, -9, 1)$.

First, we start with any m by n matrix A , then apply elimination. That changes A into its reduced row echelon form, $R_0 = \text{rref}(A)$. Removing all zero rows of R_0 leaves R .

$r, m, n = 2, 2, 4$ Simplest Case $R = [IF]$ as in $\begin{bmatrix} 1 & 0 & 3 & 5 \\ 0 & 1 & 4 & 6 \end{bmatrix}$

$r, m, n = 2, 3, 4$ General Case $R_0 = \begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix} P$ as in $\begin{bmatrix} 1 & 7 & 0 & 8 \\ 0 & 0 & 1 & 9 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Hold up, what is rref, I, F here? For something to be rref, you need to satisfy 4 conditions.

1. The zero row is at the bottom
2. The first non-zero in row 1 is 1, the first non-zero entry in row 2 is 1.
3. Each pivot is to the right of the pivots in the rows above it.
4. Every pivot is only non-zero number in its entire column.

I in this context is still identity, but only if you take the pivot columns and align them chronologically as they were. F is free matrix, the other non-pivot columns, still in the same chronologically order.

In this case (R_0),

$$I \text{ is } \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, F \text{ is } \begin{bmatrix} 7 & 8 \\ 0 & 9 \\ 0 & 0 \end{bmatrix}$$

3.2.1 Elimination from A to $\text{rref}(A)$: Reduced Row Echelon Form

Refresher, how does elimination works?

1. Subtract a multiple of one row from another row
2. Multiple a row by nonzero number
3. Exchange any rows

For demonstration,

$$A = \begin{bmatrix} 1 & 2 & 11 & 17 \\ 3 & 7 & 37 & 57 \end{bmatrix} \text{ then } \begin{bmatrix} 1 & 2 & 11 & 17 \\ 0 & 1 & 4 & 6 \end{bmatrix} \text{ then } \begin{bmatrix} 1 & 0 & 3 & 5 \\ 0 & 1 & 4 & 6 \end{bmatrix}$$

So, what did elimination actually do? It inverted the leading 2 by 2 matrix, which we will call W .

$$W = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix} \text{ into } \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

We multiplied $W^{-1}A = W^{-1}[WH]$ to produce $R = [IW^{-1}H] = [IF]$. We always knew that free columns (H) is some combination of independent columns (W), but we now know that $H = WF$.

$$H = \begin{bmatrix} 11 & 17 \\ 37 & 57 \end{bmatrix} = WF = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix} \times \begin{bmatrix} 3 & 5 \\ 4 & 6 \end{bmatrix}$$

However you compute R from A , you will always get the same R . R is completely controlled by A .

For **example 2**, let us rref another A .

$$A = \begin{bmatrix} 1 & 7 & 3 & 35 \\ 2 & 14 & 6 & 70 \\ 2 & 14 & 9 & 97 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 7 & 3 & 35 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 27 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 7 & 0 & 8 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 27 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 7 & 0 & 8 \\ 0 & 0 & 1 & 9 \\ 0 & 0 & 0 & 0 \end{bmatrix} = R_0$$

3.2.2 Elimination Column by Column: The Steps from A to R_0

We will now reduce what we learn to an easily applied algorithm. The big question is **does this new column $k+1$ join with I_k or F_k ?**

If l is all zero, the new column is **dependent** on the first k columns. Then u joins with F_k to become F_{k+1} .

If l is not all zero, then it is independent of the first k columns. Use the **largest** number, preferably, in l as the pivot. **Important** thing to remember here is that the column are talking about means the columns **UNDERNEATH** all pivots, not **ALL** columns. Then do elimination. Whatever is left becomes part of I as I_{k+1} .

From example 2, we can see that the combination of independent and dependent comes out to

$$C \times F = \begin{bmatrix} 1 & 3 \\ 2 & 6 \\ 2 & 9 \end{bmatrix} \begin{bmatrix} 7 & 8 \\ 0 & 9 \end{bmatrix} = \begin{bmatrix} 7 & 35 \\ 14 & 70 \\ 14 & 97 \end{bmatrix} = \text{depend columns of 2 and 4 of } A$$

Right back to where we came from, showing that C are almost like the ingredients and F are almost like the method.

3.2.3 The Matrix Factorization $A = CR$ and the Nullspace

In chapters prior, we know that $A = CR$ but we have no systemic way to find them, now we do. We apply elimination to reduce A to R_0 . Then I in R_0 locates the matrix C of independent columns in A . Removing zero row in R_0 produces R for $A = CR$

We have two special solution s_1 and s_2 for every column of F in R .

$$Rs_1 = 0, \begin{bmatrix} 1 & 7 & 0 & 8 \\ 0 & 0 & 1 & 9 \end{bmatrix} \begin{bmatrix} -7 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$Rs_1 = 0, \begin{bmatrix} 1 & 7 & 0 & 8 \\ 0 & 0 & 1 & 9 \end{bmatrix} \begin{bmatrix} -8 \\ 0 \\ -9 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

s_1 and s_2 are the easiest to see using the matrices $-F$ and I and P^T .

As a reminder, the two special solutions to $[IF]Px = 0$ are the columns of $P^T \begin{bmatrix} -F \\ I \end{bmatrix}$

It more correct than the other option because PP^T is the identity matrix of permutation matrix P:

$$\mathbf{R}\mathbf{x} = \mathbf{0}, [IF]P \times P^T \begin{bmatrix} -F \\ I \end{bmatrix} \text{ reduces to } [IF] \begin{bmatrix} -F \\ I \end{bmatrix} = [0]$$

Review Say, the m by n matrix A has rank r . We can find $n - r$ special solution to $Ax = 0$ by computing the rref R_0 of A . Remove the $m - r$ zero rows of R_0 to produce $R = [IF]P$ and $A = CR$. Then the special solutions to $Ax = 0$ are the $n - r$ columns of $P^T[-F, I]^T$

Example 3: Elimination of A gives R_0 and R . R reveals the nullspace of A

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 5 \\ 3 & 6 & 9 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 3 \\ 0 & 0 & 6 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = R_0 \text{ with rank 2}$$

$$R = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The independent columns are 1 and 3.

To solve $Ax = 0$ and $Rx = 0$, set $x_2 = 1$, which will get you $x_1 = -2, x_3 = 0$. Leave us special solution:

$$\mathbf{s} = (-2, 1, 0)$$

All solutions $x = (-2c, c, 0)$. And here it is, $A = CR$

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 5 \\ 3 & 6 & 9 \end{bmatrix} = CR = \begin{bmatrix} 1 & 1 \\ 2 & 5 \\ 3 & 9 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \text{columns basis in } C \times \text{row basis in } R.$$

For a lot matrices, the only solution to $Ax = 0$ is $x = 0$. Simply, all columns of A are independent. The nullspace $N(A)$ contains only the zero vector, no special solution. This case zero nullspace is **important** because it means that all columns of A is independent. But this can't happen if $n > m$ (column > row) because you can have n independent column in R^m .

Important Say A has more columns than rows ($n > m$), there will be at least one free variable. Meaning that $Ax = 0$ has at least one non-zero solution. Or to put it more specifically, there must be more than $n - m$ free columns. $Ax = 0$ must have nonzero solutions in $N(A)$.

Example 4: Find the nullspace of A, B, M and the two special solutions to $Mx = 0$

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 8 \end{bmatrix}, B = \begin{bmatrix} A \\ 2A \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 8 \\ 2 & 4 \\ 6 & 16 \end{bmatrix}, M = [A \quad 2A] = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 3 & 8 & 6 & 16 \end{bmatrix}$$

Solution The equation $Ax = 0$ has only the zero solution $x = 0$. The nullspace is only Z .

$$Ax = \begin{bmatrix} 1 & 2 \\ 3 & 8 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = R = I$$

No free variables, meaning A is invertible; therefore, no special solution.

The M matrix is different. It has extra columns instead of rows. That means that, with 4 columns and 2 rows, there will be 2 free columns leftover.

$$M = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 3 & 8 & 6 & 16 \end{bmatrix}, R = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix} = [I \quad F]$$

Again, to get special solutions out of it, we will let $x_3 = 1, x_4 = 0$ and $x_3 = 0, x_4 = 1$. What we will get is the special solution for the nullspace of M .

$$Mx = 0R = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix} s_1 = \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix} \text{ and } s_2 = \begin{bmatrix} 0 \\ -2 \\ 0 \\ 1 \end{bmatrix}$$

3.2.4 Block Elimination in Three Steps: Final Thoughts

We will conclude nicely with three steps to block elimination.

Step 1 Exchange the columns and rows of P_C and P_R so that that r independent columns and rows come first in $P_R A P_C$

$$P_R A P_C = \begin{bmatrix} W & H \\ J & K \end{bmatrix}, C = \begin{bmatrix} W \\ J \end{bmatrix} \text{ and } B = \begin{bmatrix} W & H \end{bmatrix}$$

Step 2 Multiple the top rows by W^{-1} to produce $W^{-1}B = [I, W^{-1}H] = [I, F]$.

Step 3 Subtract $J[I, W^{-1}H]$ from $[J, K]$ to produce $[0, 0]$.

The results of the steps should be an rref form of R_0

$$P_R A P_C = \begin{bmatrix} W & H \\ J & K \end{bmatrix} \rightarrow \begin{bmatrix} I & W^{-1}H \\ J & K \end{bmatrix} \rightarrow \begin{bmatrix} I & W^{-1}H \\ 0 & 0 \end{bmatrix} = R_0$$

There are two things that need to be remembered.

1. W is invertible

2. The block satisfies $JW^{-1}H = K$

1. We must think back to $A = CR$. We can see that $B = WR$, and since B and R have the rank of r and W is also r by r , that means that W must have a rank of r and be invertible.

2. We know that the first row $[I, W^{-1}H]$ is linearly independent. Since A has the rank r , it means that the lower row $[J, K]$ must be a combination of the upper rows. This means for the combination to be valid, $JJ = J$ and $JW^{-1}H = K$.

$$\text{The conclusion is that } P_R A P_C = \begin{bmatrix} W \\ J \end{bmatrix} W^{-1} \begin{bmatrix} W & H \end{bmatrix} = C W^{-1} B.$$

3.3 The Complete Solution to $Ax = b$

Our goal in this section will be about:

1. The complete solution to $Ax = b$: $x = x_p + x_n$, where p starts for any particular x and n nullspace.
2. Elimination from $Ax = b$ to $R_0x = d$: Solvable when zero rows of R_0 have zero in d .
3. When $R_0x = d$ is solvable, one x_p has all free variable equal to zero.
4. A has full column rank $r = n$ when its nullspace $N(A) = \text{zero vector}$: no free variables.
5. A has full row rank $r = m$ when its column space $C(A)$ is R^m : $Ax = b$ is always solvable.

The biggest thing that changed is that the b of $Ax = b$ is now not zero. Therefore, the row operation will also act on the right side, the b side. $Ax = b$ is reduced to a simpler $R_0x = d$ with the same, if any, solutions. One way to organize that is by augmenting (adding another column to the right side of the matrix). We can augment A with the right side of $(b_1, b_2, b_3) = (1, 6, 7)$ to produce augmented matrix $[Ab]$

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 7 \end{bmatrix} \text{ has the augmented matrix } \begin{bmatrix} 1 & 3 & 0 & 2 & 1 \\ 0 & 0 & 1 & 4 & 6 \\ 1 & 3 & 1 & 6 & 7 \end{bmatrix} = [Ab]$$

Now if we turn it into its rref form

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix} \text{ has the augmented matrix } \begin{bmatrix} 1 & 3 & 0 & 2 & 1 \\ 0 & 0 & 1 & 4 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = [R_0d]$$

The last row is very important. Third equation became $0 = 0$, each means it can be solved. In the original matrix, the first row plus the second row equals the third row. Meaning to solve $Ax = b$ we need $b_1 + b_2 = b_3$, which led to $0 = 0$ in the third equation.

3.3.1 One Particular Solution $Ax_p = b$

To get an easy x_p , let the free variables be zeros: $x_2 = x_4 = 0$, and the two nonzero equations be the two pivot variables $x_1 = 1, x_3 = 6$. So our x_p is $x_p = (1, 0, 6, 0)$. To put it simply, **free variables = zero, pivots = variable from d**.

For a solution to exist, zero rows in R_0 must also be zero in d .

$$R_0 x_p = \begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix}$$

Notice how the complete solution includes all x_n :

$$x = x_p + x_n = \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -2 \\ 0 \\ -4 \\ 1 \end{bmatrix}$$

Example 1 Find the condition on b_1, b_2, b_3 for $Ax = b$ for

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ -2 & -3 \end{bmatrix} \text{ and } \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Solution Elimination using augmented matrix $[A, b]$

$$\begin{bmatrix} 1 & 1 & b_1 \\ 1 & 2 & b_2 \\ -2 & -3 & b_3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & b_1 \\ 0 & 1 & b_2 - b_1 \\ 0 & -1 & b_3 + 2b_1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 2b_1 - b_2 \\ 0 & 1 & b_2 - b_1 \\ 0 & 0 & b_3 + b_1 + b_2 \end{bmatrix} = [R_0 d]$$

And since there is no special solution ($n - r = 2 - 2 = 0$), the nullspace solution is $x_n = 0$. So the complete solution is

$$x = x_p + x_n = \begin{bmatrix} 2b_1 - b_2 \\ b_2 - b_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Every matrix A with full column rank ($r = n$) has these properties :

1. All columns are independent, no free variables.
2. The nullspace $N(A)$ include only the zero vector $x = 0$
3. If $Ax = b$ has a solution, then it has only one solution.

With full column rank, $Ax = b$ will have **one or no solution only**.

3.3.2 Full Row Rank and the Complete Solution

Another extreme case is the full row rank. Now $Ax = b$ has one or infinitely many solutions. Full row rank requires that the matrix be a short and wide one ($m > n$, row $q > \text{col } q$), and every row is independent. For **example 2**, $Ax = b$ has 3 n but only two m .

$$\begin{bmatrix} 1 & 1 & 1 & 3 \\ 1 & 2 & -1 & 4 \end{bmatrix} \text{ rank } r = m = 2$$

Imagine them as plane in xyz space. There are two planes, colliding into a line. The particular solution is a spot on that line, and the nullspace vector will move us along that line. $x = x_p + \text{all } x_n$ gives us the whole line solution.

Fast forward a bit, getting ex. 2 into the $[Ab]$ form gives us:

$$\begin{bmatrix} 1 & 0 & 3 & 2 \\ 0 & 1 & -2 & 1 \end{bmatrix} = [R, d]$$

This particular solution $(2, 1, 0)$ has free variable $x_3 = 0$. The special solution has $s_3 = 1$, and the $-x_1$ and $-x_2$ comes from the free column of R .

Check that x_p and s satisfies $Ax_p = b$ and $As = 0$

$$2 + 1 = 3, 2 + 2 = 4 - 3 + 2 + 1 = 0, -3 + 4 - 1 = 0$$

Remove that the nullspace solution x_n is just any multiple of s .

$$\text{Computer solution: } x = x_p + x_n = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix}$$

Every matrix A with a full row rank ($r = m$) has all these properties:

1. All rows have pivot and R_0 has no zero rows: $R_0 = R$
2. $Ax = b$ has a solution for every right side b .
3. The column space of A is the whole space R^m .
4. If $m < n$, the equation has many solutions (called underdetermined in formal language).
5. The rows are linearly independent.

There are **four** possibilities for linear equation depend on rank r

1. $r = m$ and $r = n$ Square and invertible, $Ax = b$ has 1 solution.
2. $r = m$ and $r < n$ Short and wide, $Ax = b$ has infinite solutions.
3. $r < m$ and $r = n$ Tall and thin, $Ax = b$ has 0 or 1 solutions.
4. $r < m$ and $r < n$ Not full rank, $Ax = b$ has 0 or infinite solutions.

The reduced R_0 will fall in the same category as matrix A . For $R_0x = d$ and $Ax = b$ to be solvable, d must end in $m - r$ zeros.

Four types for R_0

$$\begin{bmatrix} I \end{bmatrix}, r = m = n$$

$$\begin{bmatrix} I & F \end{bmatrix} r = m < n$$

$$\begin{bmatrix} I \\ 0 \end{bmatrix} r = n < m$$

$$\begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix} r < m, r, n$$

Case 1 and 2 have full row rank $r = m$. Case 1 and 3 have full column rank $r = n$.

3.4 Independence, Basis, Dimension

How big is the true size of a subspace? A matrix might be m by n but the column space is not necessarily n . The column space $C(A)$ is measured by independent columns. This will be clarified later.

Our goal here is to understand a **basis: independent vectors that "spans" a space**. Every vector in the space is a unique combination of the basis vectors. Some vague explanation of the terms:

1. Independent vectors, no extra vectors
2. Spanning a space, enough vectors to produce the rest.
3. Basis for a space, not too many and not too few
4. Dimension of a space, the number of vectors in every basis.

3.4.1 Linear Independence

Here's an odd definition for independence:

The columns of A are linearly independent when the only solution of $Ax = 0$ is $x = 0$. No other combination of Ax gives

To illustrate why: it is impossible for three vectors that are not in a plane to be in a plane **unless** x is 0, i.e.,

$$0v_1 + 0v_2 + 0v_3 + \dots$$

Or to put it in other words:

Linear independence only happens when $x_1v_1 + x_2v_2 + \dots = 0$ and all $x_i = 0$.

One point to drive is that vectors are either dependent or independent, no in between. For some examples:

1. $(1, 0)$ and $(1, 0.000001)$ are independent.
2. $(1, 1)$ and $(-1, -1)$ are dependent.
3. $(1, 1)$ and $(0, 0)$ are dependent because of the zero vector.
4. In \mathbb{R}^2 : Any three vectors $(a, b), (c, d), (e, f)$ are dependent.

Example 1: The columns of this A are dependent, $Ax = 0$ has a nonzero solution:

$$Ax = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 5 \\ 1 & 0 & 3 \end{bmatrix} \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \text{ is } -3 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 3 \\ 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Question, how do we fix this? Just do elimination.

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 2 & 1 & 5 \\ 1 & 0 & 3 \end{bmatrix} \rightarrow R_0 = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}. F = \begin{bmatrix} 3 \\ 3 \\ -1 \end{bmatrix}, x = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$$

Remember that **columns of A are independent exactly when $r = n$** , there are n pivots and zero free variable. $x = 0$ is the only nullspace.

Also remember that any set of n vectors must be linearly dependent if $n > m$.

3.4.2 Vectors that Span a Subspace

To reiterate, column space consists of all combination of $x_1v_1 + \dots + x_nv_n$. The word "span" describes $C(A)$.

The columns of a matrix span its column space. They might be dependent.

Example 2: We will try to describe the column space and row space of A .

$$m = 3, n = 2 : A = \begin{bmatrix} 1 & 4 \\ 2 & 7 \\ 3 & 5 \end{bmatrix} \text{ and } A^T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 7 & 5 \end{bmatrix}$$

We could say that the column space of A is a plane in R^3 spanned by two columns of A . The row space of A is spanned by three rows of A (which are columns in A^T) in R^2 . Oh, and the rows span R^n and columns span R^m . Yes, they are swapped.

3.4.3 A Basis for a Vector Space

To reiterate, the basis means just right. Two independent vectors can't span R^3 , four vectors can't be all independent even if they span R^3 . Three independent vectors for R^3 is just right.

Basis vectors are independent and they span the space

Note that there is **only one way to write v as a combination of the basis vector**

Example 3: The column of $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ produce the standard basis for R^2 .

$$i = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 1 \end{bmatrix} \text{ are independent, and they span } R^2$$

Example 4: The column of every invertible n by n matrix give a basis for R^n :

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 1 & 2 \end{bmatrix}$$

We can see that A is **invertible, independent**, with a rank = 3. B is singular matrix, which also have dependent columns, and it is not full column rank (column space $\neq R^3$). The only solution to $Ax = 0$ is $x = A^{-1}0 = 0$. However, for $Ax = b$ can always be solved by $x = A^{-1}b$. Everything comes together for invertible matrices. **The vector v_1, \dots, v_n are a basis for R^n exactly when they are the column of an n by n invertible matrix. R^n also has infinitely many bases** Or to put it more compactly, **every set of independent vectors can be extended to a basis; the spanning set of vectors can be reduced to a basis.**

Example 5: The matrix is not invertible so its column are not basis for anything.

$$A = \begin{bmatrix} 2 & 4 \\ 3 & 6 \end{bmatrix} \rightarrow R_0 = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$$

Example 6: Find the bases for the column and row spaces of this rank two matrix.

$$R_0 = \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Like us see, so for column space of R_0 , we got two pivots in a R^3 , where $m = 3$; we got a subspace in R^3 . And for row space, we got two non-zero rows and a rank of R^4 where $n = 4$.

Question! Given five vectors in R^7 , how do we find their a basis that they span on? Two ways,

1. Make them into a row and eliminate to find their non-zero.
2. Make them into columns, and eliminate to find pivots. The pivot columns are the basis.

Another question, can another basis for the same vector space have more or less vectors? Answer, no! Number of vectors are tied to the space.

3.4.4 Dimension of a Vector Space

This chapter is proof heavy, but the gist of it is simply that there can be many vector bases to choose from but no matter what, to the same vector space, there will always be the same amount of vectors.

That is what dimension means, the number of basis vectors.

3.4.5 Bases for Matrix Spaces and Function Spaces

We should also know that **independence, basis, dimension** are not just limited to column vectors. We can also ask whether specific matrices A_1, A_2, A_3 are independent or not. Say, a 3×4 matrix space, what dimension is that? 12 dimensions since you need twelve matrices in every basis.

In differential equation, $\frac{d^2y}{dx^2} = y$ has a space of solutions. One of the basis is $y = e^x$ and $y = e^{-x}$. Oh, and the basis function gives dimension = 2 for this solution space because the linear equation starts with the second derivative.

Matrix space The vector space \mathbf{M} contains all 2 by 2 matrices, it's dimension is 4.

$$A_1, A_2, A_3, A_4 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = A$$

These matrices are linearly independent, they can produce any matrix in \mathbf{M} .

We can see from $c_1A_1 + \dots + c_4A_4$ that the only way to get 0 is if all coefficients are 0. We can use this basis for many things. A_1, A_2, A_4 are a basis for the upper triangle matrices, A_1, A_4 for diagonal matrices. What about for symmetric matrices? $A_1, A_4, A_2 + A_3$.

If list out a few more dimensions:

1. The dimension of whole n by n matrix space is n^2
2. The dimension of the subspace of upper triangular matrices is $\frac{1}{2}n^2 + \frac{1}{2}n$
3. The dimension of diagonal matrices is n .
4. The dimension of the subspace of symmetric matrices is $\frac{1}{2}n^2 + \frac{1}{2}n$

Oh, and remember that Z is a vector that contains only zero vector. Its basis is a **empty set**, the one with no vector. NEVER allow zero vector into the basis because then nothing is independent.

For summary,

1. Columns of A are independent if $x = 0$ is the only solution to $Ax = 0$.
2. v_1, \dots, v_r **span** a space if their combinations fill that space.
3. Basis consists of linearly independent vectors that span the space. Ever vector in the space is a unique combination of the basis vector.
4. The number of basis vectors per basis is entirely dependent on the space. The number of vectors in a basis is the dimension of the space.
5. Pivot columns are one basis for the column space. The dimension of $C(A)$ is r

3.5 Dimensions of the Four Subspaces

1. The column space $C(A)$ and the row space $C(A^T)$ both have dimension r , the rank of A .
2. The nullspace $N(A)$ has dimension $n - r$. The left nullspace $N(A^T)$ has dimension $m - r$.
3. Elimination from A to R_0 changes $C(A)$ and $N(A^T)$ but their dimension doesn't change.

We must separate the difference between rank and dimension. The rank of a matrix counts independent columns, the dimension counts the number of vectors in a basis. We can count both pivots and basis vector. Rank reveals the dimension of the four fundamental subspaces.

1. Row space, $C(A^T)$, a subspace of R^n , its dimension is r
2. Column space, $C(A)$, a subspace of R^m , its dimension is r
3. Nullspace, $N(A)$, a subspace of R^n , its dimension is $n - r$
4. Left nullspace, $N(A^T)$, a subspae of R^m , its dimension is $m - r$

Keep in mind that row space is simply just the column space of A^T . Remember again that the row space and column space have the same dimension r .

3.5.1 The Four Subspaces for R_0

We will now try to identify the four subspaces in R_0 . Though the main point is that **the four dimensions are the same for A and R_0**

$$\begin{bmatrix} 1 & 3 & 5 & 0 & 7 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We can that see That

1. The pivot row is row 1 and 2.
2. The pivot column is column 1 and 3.
3. $r = 2$

Now, finding the nullspace, $n - r = 5 - 2 = 3$. We will need to find 3 special solutions.

1. Special solution 1 = $(-3, 1, 0, 0, 0)$
2. Special solution 2 = $(-5, 0, 1, 0, 0)$
3. Special solution 3 = $(-7, 0, 0, -2, 1)$

Now we will find the nullspace of $(R_0^T$ aka the left nullspace. Its dimension should be $m - r = 3 - 2 = 1$

$$(R_0^T y = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 0 & 0 \\ 5 & 0 & 0 \\ 0 & 1 & 0 \\ 7 & 2 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \text{ is solved by } y = \begin{bmatrix} 0 \\ 0 \\ y_3 \end{bmatrix}$$

Wonder why it's called **left nullspace**? It's because we can transpose $(R_0^T y = 0$ to $y^T R_0 = 0^T$, the y^T is now a row vector to the left of R .

3.5.2 The Four Subspaces for A

A few things to remember,

1. A has the same row space as R_0 and R . Same dimension r and same basis. Elimination changes the row, but the row space remains untouched.
2. The column space of A has dimension r . The column rank **equals** the row ranks. The same combination (as in the same column number) are zero or otherwise (meaning both get zero or none get zero) for both A and R_0 .
3. A has the same nullspace as R_0 , same dimension $(n - r)$ and same basis. Elimination doesn't change the solutions to $Ax = 0$, including the special solutions.
4. The left nullspace of A (the nullspace of A^T) has dimension $m - r$. The counting rule for A was $r + (n - r) = n$ and the counting rule for A^T is $r + (m - r) = m$

Very IMPORTANT, This is a fundamental theorem of linear algebra, part 1: The column space and row space both have dimension r . The nullspaces have dimension $n - r$ and $m - r$

Imagine we have a 11 by 17 matrix with 187 nonzero entries, there are two key facts:

1. dimension of $C(A) = \text{dimension of } C(A^T) = \text{rank of } A$
2. dimension of $C(A) + \text{dimension of } N(A) = 17$

Example 1:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix}$$

It has $m = 2, n = 3, r = 1$. The row space is the line through $(1, 2, 3)$, the nullspace is the plane $x_1 + 2x_2 + 3x_3 = 0$. The dimension of the row space is 1 since it is a line, and the dimension of the

nullspace is 2 since it is a plane. We can do the equation $r + (n - r)$, where $n - r$ is the dimension of the nullspace and we get $1 + 2 = 3$ where $3 = n$, the number of column.

We can do the same for left nullspace and column space. The column space is $[1, 2]^T$ since we can see that everything is linearly dependent on it. And nullspace is perpendicular to column space, so try dotting it and see if it gets a 0 or not. In this case, the left nullspace is a line through $[2, 1]^T$

The y's in the left nullspace combine the rows of A to give the zero row,

Example 2: Time for a more practical question (finally!) Say, we have five equations with four unknowns, one for every nodes.

$$\begin{bmatrix} -1 & 1 & & & \\ -1 & & 1 & & \\ & -1 & 1 & & \\ & -1 & & 1 & \\ & & -1 & 1 & \end{bmatrix}$$

To find the nullspace $N(A)$, set $b = 0$. The first $x_1 = x_2$, the second $x_1 = x_3$, the third $x_2 = x_3$, the fourth $x_2 = x_4$, the fifth $x_3 = x_4$. So all of them are equal, $x = (c, c, c, c)$. Just one vector is enough to fill the nullspace of A , there it is a line in R^4 . We can say $(1, 1, 1, 1)$ is a basis for $N(A)$. $N(A)$ has a dim of 1, then r of A must be 3 since $4 - 3 = 1$. We now know the dimension of all four subspaces.

The column space $C(A)$ must be $r = 3$. We can again easily find out the independent column by rrefing it.

$$R_0 = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The left nullspace $N(A^T)$, now we can solve $A^T y = 0$. Note that the rows give zero, because of that we can say row 3 = row 2 - row 1, so one of the solution is $y = (1, -1, 1, 0, 0)$. In case of the lower loop, row 3 = row 4 - row 5, $y = (0, 0, -1, 1, 1)$. The dimension of $m - r = 5 - 3 = 2$, giving us the basis for the left nullspace.

1. 1, 2, 3 forms a loop in the graph, (1, 2, 3) are dependent.
2. 1, 2, 4 forms a tree in the graph, (1, 2, 4) are independent

So, where the hell did loop and tree come from? Well, they didn't need to happen, we could've used elimination. But it is more elegant. Note that $Ax = b$ gives voltage at x_1, x_2, x_3, x_4 at the four nodes and $A^T y = 0$ gives currents y_1, y_2, y_3, y_4, y_5 on the five edges. These two equations are Kirchhoff's Voltage Law and Kirchhoff's Current Law. Now, I should've said this earlier, but we need a graph for this. I don't know how to make one **yet**. Check page **135** in Strang's Linear Algebra. Graphs are the most important model in discrete applied mathematics.

For summary, **Incidence matrix A** comes from a connected graph with n nodes and m edges. The row space and column space have dimensions of $r = n - 1$, the nullspaces of A and A^T have dimensions 1 and $m - n + 1$.

1. $N(A)$ the constant vectors (c, c, \dots, c) make up the nullspace of A : $\dim = 1$.
2. $C(A^T)$ the edge of any tree give r independent rows of A : $r = n - 1$.
3. $C(A)$ Voltage Law: The components of Ax add to zero around all loops: $\dim = n - 1$.
4. $N(A^T)$ (Current Law): $A^T y = \text{flow in} - \text{flow out} = 0$, which is solved by loop currents. There are $m - r = m - n + 1$ independent small loops in the graph.

3.5.3 Rank Two Matrices = Rank One plus Rank One

This one is a doozy. What it means is then any rank r matrix can be decomposed into r amount of rank 1 matrices.

For example, a rank 2 matrix:

$$A = \begin{bmatrix} 1 & 0 & 3 \\ 1 & 1 & 7 \\ 4 & 2 & 20 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 4 \end{bmatrix}$$

For reminder, C is literally just pivot columns and R is just nonzero-only rref.

If we put it in letters for columns and rows, we can see **rank 2 = rank 1 + rank 1**.

$$A = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} v_1^T \\ v_2^T \\ \text{zero row} \end{bmatrix} = u_1 v_1^T + u_2 v_2^T$$

To put in Strang's word, **Columns of C times rows of R , every rank r is a sum of r rank one matrices.**

Chapter 4

Orthogonality

Two vectors are orthogonal (translated as right-angled from Greek) when their dot product is zero: $v \cdot w = v^T w = 0$. The vectors in the two subspaces, the vectors in a basis, the column vectors in Q . Orthogonal vectors also have a curious behavior where it is also like pythagoras's theorem.

$$v^T w = 0 \text{ and } \|v\|^2 + \|w\|^2 = \|v + w\|^2$$

Important part, **the fundamental subspaces are orthogonal**

1. $N(A)$ contains all vectors orthogonal to the row space $C(A^T)$.
2. $N(A^T)$ contains all vectors orthogonal to the column space $C(A)$.

$Ax = 0$ makes x orthogonal to each row, $A^T y = 0$ make y orthogonal to each column.

A key idea in this chapter is **projection**: If b is outside the column space of A , then we must find the closest point p that is still inside. The line from b to p shows the error e , and that line is perpendicular to the column space. The **least squares equation** $A^T Ax = A^T b$ produces the closest $p = Ax$ and smallest possible e , it also gives the best e when $Ax = b$ is unsolvable. The best x makes $\|Ax - b\|$ as small as possible, the least squares. $A^T Ax = A^T b$ is easy when $A^T A = I$. Then A has orthonormal columns perpendicular unit vector. Remember that Q has $Q^T Q = I$ and $QR = A$, where the R is upper triangle. Orthogonal matrices are perfect for computations, $A = QR$ is even better than $A + LU$

4.1 Orthogonality of Vectors and Subspaces

1. Orthogonal vectors have $v^T w = 0$, then $\|v\|^2 + \|w\|^2 = \|v + w\|^2$ as in $a^2 + b^2 = c^2$
2. Subspace V and W are orthogonal when $v^T w = 0$ for every v in V and every w in W .
3. Row space of A is orthogonal to nullspace, column space of A is orthogonal to left nullspace.
4. The dimensions add to $r + (n - r) = n$ and $r + (m - r) = m$: orthogonal complements.
5. If n vectors in R^n are independent, they span R^n . If n vectors span R^n , they are independent.

How can we prove that the entirety of the nullspace of A is orthogonal to the row space of A ? Look at $Ax = 0$.

$$Ax = \begin{bmatrix} \text{row 1 of } A \\ \vdots \\ \text{row } m \text{ of } A \end{bmatrix} [x] = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

Notice how they are all zero? That means all of the row space is orthogonal to the nullspace.

But given that we like to work with columns, another way you can prove it is: $x^T(A^T y) = (Ax)^T y = 0^T y = 0$

Importantly, column space $C(A)$ and left nullspace $N(A^T)$ is also a perpendicular pair. The proof is more or less the same except we use A^T instead of A .

Example 1: The two rows of A are perpendicular to x in the nullspace of A :

$$Ax = \begin{bmatrix} 1 & -2 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

or

$$A^T y = \begin{bmatrix} 1 & 1 \\ -2 & 0 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

This is a fairly extreme case since $C(A)$ is all of R^2 and the nullspace of A^T is a zero vector. The total dimension is $2 + 1 = 3$. The subspaces accounted for all vectors in $R^3 = R^n$. The column space and left nullspace have a dimension of $2 + 0 = 2$, accounting for all vectors in $R^2 = R^m$.

Note that **if V and W are orthogonal subspaces in R^n then $\dim V + \dim W \leq n$** . Of course, imagine our house. We got a wall and a floor, they can't be orthogonal subspaces since they are both R^2 and our world is R^3 , $2 + 2 \geq 3$. So this is wrong. Some vector will lie in both the wall and the floor, which is the line where the wall meets the floor in both subspaces.

Two orthogonal subspaces that account for the whole space have a special name, **orthogonal complements**. Orthogonal complement of V^\perp of V contains all vectors orthogonal to V . So the two pairs of subspace in linear algebra are actually orthogonal complements.

1. Row space and null space, $r + (n - r) = n$
2. Column space and left nullspace $r + (m - r) = m$

Any vector x in R^n is the sum $x = x_{\text{row}} + x_{\text{null}}$ of its row space and component and its null space component. Same goes for y in R^m is the sum $y = y_{\text{col}} + y_{\text{null}}$, between its column space component and its component in $N(A^T)$.

Fundamental theorem of Linear Algebra, Part 2:

1. $N(A)$ is the orthogonal complement of the row space $C(A^T)$ in R^n
2. $N(A^T)$ is the orthogonal complement of the column space $C(A)$ in R^m

Check the figure on page 146 of Strang's LA. It will us that the complete solution to $Ax = b$ is $x = \text{one } x_r + \text{any } x_n$. Then the minimum norm solution to $Ax = b$ is $x = x_r$ from the row space plus $x_n = 0$ from the nullspace, from $\|x\|^2 = \|x_r\|^2 + \|x_n\|^2$

Every vector Ax is in the column space. Multiplying by A cannot do anything else. More importantly, every b in the column space comes from exactly one vector x_r in the row space.

Example 2: Every matrix of rank r has an r by r invertible submatrix. A has rank = 2:

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 4 & 5 & 6 \\ 1 & 2 & 4 & 5 & 6 \end{bmatrix} \text{ contains } \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \text{ in the pivot rows and pivot columns.}$$

So if the submatrix C has r independent columns, then C and C^T has r independent columns. This locates an r by r invertible submatrix of A .

4.1.1 Combing Bases from Subspaces

Basis have two properties:

1. They are linearly independent
2. They span the space

However, the two properties implies eachother. **If there are n columns of independent vector in A , then they span R^n , $Ax = b$ is solvable. If n vectors span R^n , they must be independent; $Ax = b$ has one solution. If $AB = I$ for square matrix, then $BA = I$ too.** We can also start from the opposite side, say $Ax = b$ can be solved for every b , forcing **existence of solution**. This means elimination produced no zero rows. There are n and not free variables. The nullspace contains only $x = 0$, ending in **uniqueness of solutions**.

Example 3:

$$\text{For } A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \text{ split } x = \begin{bmatrix} 4 \\ 3 \end{bmatrix} \text{ into } x_r + x_n = \begin{bmatrix} 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

The vector $(2, 4)$ is in the row space, while the orthogonal vector is from the nullspace $(2, -1)$.

Example 4: Suppose S a six dimensional subspace of nine-dimensional space R^9

1. What are the possible dimensions of subspaces orthogonal to S ? **0, 1, 2, 3** since S already took **6 of the total 9**.
2. What are the possible dimension of the orthogonal complement S^\perp of S ? **3, because they are asking for THE orthogonal complement, the one that contains everything else. So the biggest one, 3, is the answer.**
3. What is the smallest size of matrix A that has row space S ? **6 by 9, because A is a matrix of $m \times n$ and given $R^n = R^9$, $m \times 9$. For m , the dimension of the row space is the rank of m , so $m = 6$. Therefore, 6 times 9**
4. What is the smallest possible size of a matrix B that has nullspace S^\perp ? **6 by 9. We are told that R^9 have an n of 9. And that $S^\perp = 3$ from the second question, so we know that $r + \text{null} = n$, we got $r + 3 = 9$, $r = 6$. 6 \times 9**

Notice how question 3 and 4 gives the same matrix? It is telling us that $C(A^T) = (N(A))^\perp$

4.2 Projections onto Lines and Subspaces

4.2.1 Projection onto Lines and Subspaces

1. The projection of b onto the line through a is the closest point to b : $p = a(\frac{a^T b}{a^T a})$.
2. The error $e = b - p$ is perpendicular to a : the right triangle bpe has $\|p\|^2 + \|e\|^2 = \|b\|^2$.
3. The projection of b onto to asubspace S is the closest vector p in S ; $b - p$ is orthogonal to S .
4. $A^T A$ is invertible and symmetric when A has independent columns: $N(A^T A) = N(A)$.
5. The projection of b onto $C(A)$ is the vector $p = A(A^T A)^{-1} A^T b$.
6. The projection matrix onto $C(A)$ is $P = A(A^T A)^{-1} A^T$. It has $p = Pb$ and $P^2 = P = P^T$.

We can visual projection on a subspace S by having each vector b go the closest point p in S . The error vector $e = b - p$. If A has independent columns, the projection of b onto $C(A)$ is $p = A(A^T A^{-1}) A^T b$. The projection matrix is $P = A(A^T A)^{-1} A^T$. Its special property is $P^2 = P$, a second projection changes nothing, because it projects onto itself.

For some really nice subspaces, we can directly see their projection matrices.

1. What are the projections of $b = (2, 3, 4)$ onto the z axis and the xy plane.
2. What matrix P_1 and P_2 produce those projections onto a line and a plane?

When b is projected onto a line, its projection p is the part of the b along that line. If b is projected onto a plane, p is the part in that plane. The projection p is Pb .

We will call projection onto the z axis p_1 . Where projection p_2 drop straight down to the xy plane. To with $b = (2, 3, 4)$. $p_1 = (0, 0, 4)$, $p_2 = (2, 3, 0)$. We can see that

$$\text{Onto the } z \text{ axis: } P_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ Onto the } xy \text{ plane: } P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$p_1 = P_1 b = \begin{bmatrix} 0 \\ 0 \\ z \end{bmatrix}, p_2 = P_2 b = \begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$$

In this case, the projections p_1 and p_2 are perpendicular. The xy plane and the z axis are orthogonal subspaces. More than that, they are orthogonal complements; their dimensions add to $1 + 2 = 3$. Every vector b in the whole space is the sum of its part in two subspaces.

1. The vectors give $p_1 + p_2 = b$.
2. The matrices give $P_1 + P_2$ and $P_1 P_2 = 0$.

For this example, our goals have been reached, and it is the same for any line and any plane and any n -dimensional subspaces in m dimension. The objective is to find the part p in each subspaces, and the projection matrix P that produces that part $p = Pb$. Every subspaces of R^m has its own m by m projection matrix P .

So, if the best description of a subspace is a basis, by putting the basis vectors into the columns A , we are now projecting onto the column space of A . We can say that the z axis is the column space of the $3 \times 1 A_1$. The xy plane is the column space of A_2 . That xy plane too is also the column space of A_3 . So $p_2 = p_3$ and $P_2 = P_3$.

$$A_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, A_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, A_3 = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 0 \end{bmatrix}$$

Our problem is to project any b onto the column space of any m by n matrix. Starting with a line, dimension $n = 1$, the matrix will have only one column; we will call it a .

4.2.2 Projection Onto a Line

The thing we must remember is how the line from b to p called e is always perpendicular to a . The projection p will be some multiple of a , calling it $p = \hat{x}a$. So first we need to find \hat{x} , then vector p , then matrix P . Oh, and now $e = b - \hat{x}a$ too.

$$\hat{x} = \frac{a \cdot b}{a \cdot a} = \frac{a^T b}{a^T a}$$

The tranpose version is better because we can use it with matrices too.

$$\text{For line, } e = b - pp = \hat{x}a = \frac{a^T b}{a^T a} a$$

$$\text{For plane project, } p = A\hat{x} = A(A^T A)^{-1} A^T b = Pb$$

Example one: Find the $p = \hat{x}a$

$$b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ onto } a = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

\hat{x} is a ratio between $\frac{a^T b}{a^T a} = \frac{5}{9}$. So $p = \frac{5}{9}a = (\frac{5}{9}, \frac{10}{9}, \frac{10}{9})$ and $e = b - p = (\frac{4}{9}, -\frac{1}{9}, \frac{1}{9})$. And to check that e is perpendicular to $a = (1, 2, 2)$, which it is: $e^T a = \frac{4}{9} - \frac{2}{9} - \frac{2}{9} = 0$.

So, about the projection matrix. In the formula for p , what matrix is the multiplying b ?

$$p = a\hat{x} = a \frac{a^T b}{a^T a} = Pb \text{ when } P = \frac{aa^T}{a^T a}$$

We can see that it is a column timed a row, where a is a column and a^T is a row. The projection matrix P is $m \times m$ but its rank is **one**. We are just projecting onto a one-dimensional subspace, the line through a . That line is the column space of P . This proves that p will always be on a .

Example 2: Find the projection matrix $P = \frac{aa^T}{a^T a}$ onto the line $a = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$.

$$\text{Projection matrix } P = \frac{aa^T}{a^T a} = \frac{1}{9} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 2 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix}$$

This matrix projects **any** vector b onto a . Check $p = Pb$ for $b = (1, 1, 1)$ in example 1:

$$p = Pb = \frac{1}{9} \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 5 \\ 10 \\ 10 \end{bmatrix}$$

$$e = b - p = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{9} \begin{bmatrix} 5 \\ 10 \\ 10 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} 4 \\ -1 \\ -1 \end{bmatrix}$$

Even if vector a is doubled, the P matrix stays the same. Again, for reminder, projecting a second line doesn't change anything since it is already at the closest spot.

Another thing could be a projection, $I - P$, which projects onto the perpendicular subspace, ie a plane perpendicular to a , the orthogonal complement.

4.2.3 Projection Onto a Subspace

Start with n vectors a_1, \dots, a_n in R^m . Assume the a 's are linearly independent. Now, **find the combination** $p = \hat{x}_1 a_1 + \dots + \hat{x}_n a_n$ **closest to a given vector b** . We are project each b in R^m onto the n -dimensional subspace spanned by the a 's. Where with just $n = 1$, it's just a line, which is a column space of A ; where it is just one column. You should know that the hat on \hat{x} means the best. So when we say we are looking for particular solution $p = A\hat{x}$, we are saying we are looking for the best approximation to b . And for what we know, $\hat{x} = \frac{a^T b}{a^T a}$ when $n = 1$. For the next part, we will try to find \hat{x} when $n > 1$.

We can compute projections onto n -dimensions subspaces in three steps as before:

1. Find the vector \hat{x} in (S) .
2. Find the projection $p = A\hat{x}$ in (C) .
3. Find the projection matrix P in (T) .

Remember that the error vector $b - A\hat{x}$ is perpendicular to the subspace, where it makes right angle with all the vector in the subspace.

$$\begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} [b - A\hat{x}] = [0] \text{ or } A^T A\hat{x} = A^T b$$

We can see that $A^T(b - A\hat{x})$, which can be rewritten was $A^T A\hat{x} = A^T b$. This is the equation for \hat{x} , where we can see $A^T A$ as a coefficient matrix. With this, we can now find \hat{x} , p , and P in that order. **REMEMBER THE NEXT THREE EQUATIONS!**

1. Find $\hat{x}(n \times 1)$ $A^T(b - A\hat{x} = 0)$ **or** $A^T A\hat{x} = A^T b$.
2. The symmetric matrix $A^T A$ is $n \times n$. It is invertible if the a 's are independent. The solution is $\hat{x} = (A^T A)^{-1} A^T b$ we can find the projection of b onto the subspace is p : **Find** $p(m \times 1)p = A\hat{x} = A(A^T A)^{-1} A^T b$.
3. The projection matrix P is multiplying b in the last example, the one with four A 's. **Find** $P(m \times m)$, $P = A(A^T A)^{-1} A^T$

For $n = 1$

1. $\hat{x} = \frac{a^T b}{a^T a}$
2. $p = a \frac{a^T b}{a^T a}$
3. $P = \frac{a a^T}{a^T a}$

It is very similar to the prior 3 equations. Where the number $a^T a$ becomes $A^T A$. When it is a number, we divide by it, when it becomes a matrix, we invert it. Since a_1, \dots, a_n is independent, that means $A^T A$ is invertible. From $A^T(b - A\hat{x})$, we can tell that e is orthogonal to each a .

1. Our subspace $C(A)$ is the column space of A
2. The error $e = b - A\hat{x}$ is in the perpendicular subspace $N(A^T)$

That means that $A^T(b - A\hat{x}) = 0$ and that left nullspace $N(A^T)$ is important in projection since it contains the error vector $e = b - A\hat{x}$. The vector b is split into the projection p and the error e , where the production produces a right triangle with sides p, e, b , remember we are talking about vector, the line begin from origin.

Example 3: If $A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}$ and $b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$, find \hat{x}, p, P

1. Find $A^T A$ and $A^T b$
2. Solve the normal equation $A^T A \hat{x} = A^T b$ to find \hat{x}
3. The combination $p = A\hat{x}$ is the projection of b onto the column space of A .

$$A^T A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \text{ and } A^T b = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix} \text{ gives } \hat{x} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}$$

$$p = 5 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - 3 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}, \text{ The error is } e = b - p = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

We can now try to find P matrix. The projection is $P = A(A^T A)^{-1} A^T$. The determinant of $A^T A$ is $15 - 9 = 6$, then also inverse the 2×2 . Multiple $A \times (A^T A)^{-1} \times A = P$

$$(A^T A)^{-1} = \frac{1}{6} \begin{bmatrix} 5 & -3 \\ -3 & 3 \end{bmatrix} \text{ and } P = \frac{1}{6} \begin{bmatrix} 5 & 2 & -1 \\ 2 & 2 & 2 \\ -1 & 2 & 5 \end{bmatrix}$$

And $P^2 = P$ must be true because we discuss before, it projects onto itself.

Warning! $P = A(A^T A)^{-1} A^T$ can be deceptive. We can't split $(A^T A)^{-1}$ into $A^{-1} \times (A^T)^{-1}$. If we do that, we will find that $P = A A^{-1} (A^T)^{-1} A^T$ collapses into $P = I$, which means this is wrong. Why? **Matrix A is rectangular, which means it has no inverse matrix (as in, A^{-1}).** We can split $(A^T A)^{-1}$ into $A^{-1} \times (A^T)^{-1}$ when $m > n$. Here is something we need to remember:

$A^T A$ is invertible if and only if A has full-rank column.

When A has independent columns, $A^T A$ is square, symmetric, and invertible.

For emphasis, $A^T A$ is $n \times m$ times $m \times n$, $A^T A$ is square $n \times n$. We just again prove $A^T A$ is invertible, provided A is full-rank column.

$$\begin{bmatrix} 1 & 1 & 0 \\ 2 & 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 0 \\ 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 4 & 9 \end{bmatrix}$$

The first example has dependent A and singular output, where the second example has full-rank columns and the output is invertible.

Very brief summary, to find $p = x_1 a_1 + \dots + \hat{x}_n a_n$, solve $A^T A \hat{x} = A^T b$. This gives us \hat{x} . The projection is $p = A\hat{x}$ and the error is $e = b - p = b - A\hat{x}$. The projection matrix gives $P = A(A^T A)^{-1} A^T$ gives $p = Pb$. P is invertible only if $P = I$. This matrix satisfies $P^2 = P$ and the distance from b to subspace $C(A)$ is $\|e\|$.

Here is a review of a few key ideas,

1. Projection of b onto a is $p = a\hat{x} = a\left(\frac{a^T b}{a^T a}\right)$.
2. Rank one projection is $P = \frac{aa^T}{a^T a}$ multiples with b to produce p .
3. Projecting b onto a subspace leaves $e = b - p$ perpendicular to the subspace.
4. When A has full rank n , the equation $A^T A \hat{x} = A^T b$ leads to \hat{x} and $p = A\hat{x}$.
5. Projection matrix $P = A(A^T A)^{-1} A^T$ has $P^T = P$ and $P^2 = P$ and $Pb = p$

4.3 Least Squares Approximations

A lot of $Ax = b$ has no solution due to there being *too* many equations, which happens when there is more rows than columns ($m > n$). The n columns spans only a **small** part of the m -dimensional space

When $Ax = b$ has no solution, multiply by A^T and solve $A^T A \hat{x} = A^T b$

Example 1: One of the biggest use for least square is fitting a straight line to m points. Say we have 3 points, we need to find the line closest to them $(0, 6), (1, 0), (2, 0)$ No straight line go through these points, but we can turn them into three equations ($n = 2, m = 3$).

1. $t = 0$, the first point is on the line $b = C + Dt$ if $C \cdot D = 6$
2. $t = 1$, the second point is on the line $b = C + Dt$ if $C \cdot D = 0$
3. $t = 2$, the third point is on the line $b = C + Dt$ if $C \cdot D = 0$

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}, x = \begin{bmatrix} C \\ D \end{bmatrix}, b = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}, Ax = b \text{ is not solvable}$$

This is the same question as example 3 in the last section, we computed that $\hat{x} = (5, -3)$. So the line $5 - 3t$ will be the best for the three points.

4.3.1 Minimizing the error

How do we make $e = b - Ax$ as small as possible is a very important question. \hat{x} can be found by geometry, where e is perpendicular to A . Calculus also gives the same \hat{x} , the derivative of the error $\|Ax - b\|^2$ is zero \hat{x} .

By geometry, every Ax lies in the plane of the column $(1, 1, 1)$ and $(0, 1, 2)$. In that plane, we will need to look for the closest point to b , which is p . While all three points can't be connected with one line, they are all still within the column space of A . Therefore, in a fitting straight line, $\hat{x} = (C, D)$ is the best choice.

By algebra, b splits into two parts, the part in column space is p , the perpendicular part is e . We can not solve $Ax = b$, but we can solve $A\hat{x} = p$ (by removing e and solving $A^T A \hat{x} = A^T b$)

$$Ax = b = p + e \text{ is impossible}$$

$$A\hat{x} = p \text{ is solvable}$$

$$\hat{x} = (A^T A)^{-1} A^T b$$

The solution to $A\hat{x} = p$ leaves the least possible error, e :

$$\text{Squared error for any } x, \|Ax - b\|^2 = \|Ax - p\|^2 + \|e\|^2$$

We reduce $Ax - p$ to zero by choosing $x = \hat{x}$, so we get the smallest $e = (e_1, e_2, e_3)$, which we can't reduce further.

The least square solutions \hat{x} makes $E = \|Ax - b\|^2$ as small as possible.

Look along at figure 4.6 on page 165. We can see that the best line is $b = 5 - 3t$, and the closest point is $p = 5a_1 - 3a_2$.

Notice how the errors are $(1, -2, 1)$, adding to zero. The reason is that the error $e = (e_1, e_2, e_3)$ is perpendicular to the first column (and the second column too) $(1, 1, 1)$ in A . The dot product gives $e_1 + e_2 + e_3 = 0$.

$$e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} \rightarrow e \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = e_1 + e_2 + e_3 = 0$$

Most functions are minimized by calculus, where the graph bottoms out and the derivative in every direction is **zero**. Where E is minimized is a sum of squares.

$$E = \|Ax - b\|^2 = (C + D \cdot 0 - 6)^2 + (C + D \cdot 1)^2 + (C + D \cdot 2)^2$$

Since there are two unknowns, where we want the two derivative at zero, we will need to do partial derivative.

$$\frac{\partial E}{\partial C} = 2(C + D \cdot 0 - 6)^2 + 2(C + D \cdot 1)^2 + 2(C + D \cdot 2)^2 = 0$$

$$\frac{\partial E}{\partial D} = 2(C + D \cdot 0 - 6)^2(0) + 2(C + D \cdot 1)^2(1) + 2(C + D \cdot 2)^2(2) = 0$$

$\frac{\partial E}{\partial D}$ contains factor 0, 1, 2 from the chain rule. It is not an accidental that the factor (1, 1, 1) and (0, 1, 2) is in the derivative of $\|Ax - b\|^2$ are in the columns of A . Now cancel 2 from every term and we get:

1. C derivative is zero: $3C + 3D = 6$

2. D derivative is zero: $3C + 5D = 0$

$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \text{ is } A^T A$$

Those two equations are $A^T A \hat{x} = A^T b$. The best C and D are components of \hat{x} . The equations from calculus are the same as the normal equations from linear algebra. These equation are to minimize $\|Ax - b\|^2 = x^T A^T A x - 2x^T A^T b + b^T b$

The partial derivatives of $\|Ax - b\|^2$ are zero when $A^T A \hat{x} = A^T b$

And the solution are $C = 5$ and $D = -3$ therefore $b = 5 - 3t$ is the best line. This is the vector of e .

4.3.2 The Big Picture for Least Square

This correspond to page 166 (176), figure 4.7. We can remember how on an older "big picture" figure, that we needed to split x into $x_p + x_n$. That was for when there were **many** solutions. This is for cases where there is **no** answer to $Ax = b$, however. Instead of splitting x we are splitting $b = p + e$. Instead of $Ax = b$, we solve $A\hat{x} = p$. From the figure 4.7, we can see that the nullspace $N(A)$ is very small. Remember that $A^T A$ is an invertible matrix. The equation $A^T A \hat{x} = A^T b$ determines \hat{x} . The error $e = b - p$ has $A^T e = 0$.

4.3.3 Fitting a Straight Line

Fitting a line requires $m > 2$ points; at time t_1, \dots, t_m , those points are at height b_1, \dots, b_m . The best line $C + Dt$ misses the point by vertical distance e_1, \dots, e_m . No line is perfect, so we want to minimize $E = e_1^2 + \dots + e_m^2$. We used $m = 3$ for the last example, but really, m can be a very large number; though we will still have the component of \hat{x} be C and D . When $Ax = b$ is solved exactly, the line goes through the m points, which rarely happens. Two unknown (C, D) determines a line, so A has $n = 2$ columns.

$$Ax = b \text{ is } \begin{bmatrix} C + Dt_1 = b_1 \\ C + Dt_2 = b_2 \\ \vdots \\ C + Dt_m = b_m \end{bmatrix} \text{ with } A = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}$$

The column space is so so thin that it is almost certain that b is outside of it. When b **does** lie inside of it, that means the points are on the line. In that case, $b = p$, therefore $Ax = b$ and $e = (0, \dots, 0)$.

The closest line $C + Dt$ has heights p_1, \dots, p_m with errors e_1, \dots, e_m .

Solve $A^T A = A^T b$ for $\hat{x} = (C, D)$. The errors are $e_i = b_i - C - Dt_i$.

$$\text{Dot-product matrix } A^T A = \begin{bmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ \dots & \dots \\ 1 & t_m \end{bmatrix} = \begin{bmatrix} m & \Sigma t_i \\ \Sigma t_i & \Sigma t_i^2 \end{bmatrix}$$

On the right side of the normal equation is the 2×1 vector $A^T b$:

$$A^T A \begin{bmatrix} C \\ D \end{bmatrix} = A^T A \hat{x} = A^T b = \begin{bmatrix} 1 & \dots & 1 \\ t_1 & \dots & t_m \end{bmatrix} \begin{bmatrix} b_1 \\ \dots \\ b_m \end{bmatrix} = \begin{bmatrix} \Sigma b_i \\ \Sigma t_i b_i \end{bmatrix}$$

The best $\hat{x} = (C, D)$ is $(A^T A)^{-1} A^T b$:

$$A^T A \hat{x} = A^T b \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}$$

The vertical errors at the m points on the line are the components of the $e = b - p$. The error vector is perpendicular to the column of A , in term of geometry. The error is in the nullspace of A^T , in term of linear algebra. The best $\hat{x} = (C, D)$ minimizes the total error E , the sum of squares, in term of calculus

$$E(x) = \|Ax - b\|^2 = (C + Dt_1 - b_1)^2 + \dots + (C + Dt_m - b_m)^2$$

Calculus sets the derivative $\frac{\partial E}{\partial C}$ and $\frac{\partial E}{\partial D}$ to zero, and recovers $A^T A \hat{x} = A^T b$. A lot of other problem have more than two unknowns. In general, we are fitting m data points by n parameters x_1, \dots, x_n . The matrix A . The matrix A has n columns and $n < m$. The derivative of $\|Ax - b\|^2$ gives the n equations $A^T A \hat{x} = A^T b$, the derivative of a square is linear.

Example 2: A has an **orthogonal** columns when the measurement times t_i add to zero. Suppose $b = 1, 2, 4$ when $t = -2, 0, 2$, where the time adds to zero. The columns of A have zero dot product: $(1, 1, 1)$ is orthogonal to $(0, 2, 0, 2)$:

$$\begin{bmatrix} C + D(-2) = 1 \\ C + D(0) = 2 \\ C + D(2) = 4 \end{bmatrix} \text{ or } Ax = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

Since $A^T A$ is diagonal, we can easily solve separate that $C = \frac{7}{3}$ and $D = \frac{6}{8}$. The zeros in $A^T A$ are dot products of perpendicular columns in A . With the lower right input, we can see that $t_1^2 + t_2^2 + t_3^2 = 8$.

Orthogonal columns are very helpful so much so that it can be worth shifting the times by subtracting the average time \hat{t} . If the average times were 1, 3, 5 then their average is $\hat{t} = 3$. The shifted times $T = t - \hat{t} = t - 3$ adds up to zero.

$$\begin{bmatrix} T_1 = 1 - 3 = -2 \\ T_2 = 3 - 3 = 0 \\ T_3 = 5 - 3 = 2 \end{bmatrix}, A_{new} = \begin{bmatrix} 1 & T_1 \\ 1 & T_2 \\ 1 & T_3 \end{bmatrix}, A_{new}^T A = \begin{bmatrix} 3 & 0 \\ 0 & 8 \end{bmatrix}$$

4.3.4 Dependent Columns in A: What is \hat{x} ?

So far, we have assumed that the columns in A are independent. What if they aren't? For this example:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} = b, Ax = b$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} = p, Ax = b$$

We can see a figure on page 169 (179), at the same T , $b_1 = 3$ and $b_2 = 1$. A straight line $C + Dt$ can't go through both points. If we can first try to see the column space of A but it still provides us with 3 solutions. We could do $\hat{x}_1 = C = 2, \hat{x}_2 = D = 0$. We could also do $\hat{x}_1 = C = 1, \hat{x}_2 = D = 1$. But we can also make a line as $b = ct + d$, where $\hat{x}_1 = C = 0, \hat{x}_2 = D = 2$ is also correct. We will have a different answer when we get to the pseudoinverse but here, our shortest solution will be $x^+ = (1, 1)$. This is because x^+ has the length of $\sqrt{2}$, whereas the solution $\hat{x} = (0, 2)$ or $(2, 0)$ have a length of 2.

4.3.5 Fitting by a Parabola

When we throw a ball, the most accurate approximation would be a parabola $b = C + Dt + Et^2$ so that it can come up and down. But even with non-linear function like t^2 , the unknowns C, D, E still appear linearly.

Problem Try to fit heights b_1, \dots, b_m at times t_1, \dots, t_m by parabola $C + Dt + Et^2$

Solution

$$\begin{bmatrix} C + Dt_1 + Dt_1^2 = b_1 \\ \vdots \\ C + Dt_m + Dt_m^2 = b_m \end{bmatrix} \text{ is } Ax = b \text{ with } \begin{bmatrix} 1 & t_1 & t_1^2 \\ \vdots & \vdots & \vdots \\ 1 & t_m & t_m^2 \end{bmatrix}$$

So what we are dealing with here is a 3 dimensional column space of A . Where the projection of b is $p = A\hat{x}$, with three columns of coefficient C, D, E . The error at the data points are $e_m = b_m - C - Dt_m - Et_m^2$. We can minimize it by calculus by taking partial derivative of $e_1^2 + \dots + e_m^2$ with respect to C, D, E .

Example 3: For the the parabola $b = C + Dt + Et^2$ to go through the three heights $b = 6, 0, 0$ when $t = 0, 1, 2$ the equation for C, D, E are

$$\begin{bmatrix} C + D \cdot 0 + E \cdot 0^2 = 6 \\ C + D \cdot 1 + E \cdot 1^2 = 0 \\ C + D \cdot 2 + E \cdot 2^2 = 0 \end{bmatrix}$$

We can solve this directly for $x = (C, D, E) = (6, -9, 3)$. The parabola through the three points is b . Here are a few important points to remember:

1. The least square solution \hat{x} minimizes the $\|Ax - b\|^2 = x^T A^T A x - 2x^T A^T b + b^T b$. That is E, the sum of squares of the errors e_1, \dots, e_m in the m equations $m > n$.
2. The best \hat{x} comes from the normal equations $A^T A \hat{x} = A^T b$. E is a minimum.
3. To fit m points by a line $b = C + Dt$, the normal equations give C and D .
4. The heights of the best line are $p = p_1, \dots, p_m$. The vertical distance to the data points are the errors $e = e_1, \dots, e_m$. A key equation is $A^T e = 0$.
5. If we try to fit m by a combination of $n < m$, the m equations $Ax = b$ are generally unsolvable. The equation $A^T A \hat{x} = A^T b$ give the least square solution, the combination with the smallest mean square error (MSE).

4.4 Orthogonal Matrices and Gram-Schmidt

1. The columns q_1, \dots, q_n are orthogonal if $q_i^T q_j = \begin{bmatrix} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j \end{bmatrix}$. Then $Q^T Q = I$.
2. If Q is also square, then $Q Q^T = I$ and $Q^T = Q^{-1}$. Now Q is an orthogonal matrix.
3. The least squares solution to $Qx = b$ is $\hat{x} = Q^T b$. Projection of $b : p = Q Q^T b = P b$.
4. The **Gram-Schmidt** process takes independent a_i to orthonormal q_i . Start with $q_1 = \frac{a_1 - \text{its projection } p_1}{\|a_1 - p_1\|}$; projection $p_i = (a_i^T q_1 + \dots + a_i^T q_{i-1}) q_{i-1}$.
5. Each a_i will be a combination of q_1 to q_i . Then $A = QR$ orthogonal Q and triangular R .

We have two goal in this section, **why** and **how**. We will first see why orthogonal column in A are good. The second goal is to construct orthogonal vectors q_i . The orthonormal basis vectors q_i will be the columns of a new matrix Q . The vector q_1, \dots, q_n are orthogonal when their dot products $q_i \cdot q_j$ are zero, more exactly $q_i^T q_j = 0$ whenever $i \neq j$. With one more step, we can divide each vector by its length for the vectors to become *orthogonal unit vector*. All their length are 1 (called **normal**), then the basis is called **orthonormal**.

Definition the n vectors q_1, \dots, q_n are orthogonal if

$$q_i^T q_j = \begin{cases} 0, & \text{if } i \neq j \text{ (orthogonal vectors);} \\ 1, & \text{if } i = j \text{ (unit vectors, } \|q_i\| = 1) \end{cases}$$

A matrix Q with orthonormal columns has $Q^T Q = I$, typically $m > n$.

Matrix Q is easy to work with because $Q^T Q = I$, which means that the columns q_1, \dots, q_n are orthonormal. Q is not required to be square.

$$Q^T Q = \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \\ q_n^T \end{bmatrix} \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = I$$

When row i of Q^T multiplies column j of Q , the dot product is $q_i^T q_j$. Off the diagonal ($i \neq j$), the dot is zero by orthogonality. On the diagonal ($i = j$) the unit vector gives $q_i^T q_i = \|q_i\|^2 = 1$.

When Q is square, $Q^T Q = I$ means that $Q^T = Q^{-1}$: transpose = inverse.

If the columns are only orthogonal and not unit vectors, they still give a non-1 diagonal matrix. The inverse is the transpose of Q^T . In the case it's square, we call it an orthogonal matrix.

Example 1 (Rotation): Q rotates every vector in the plane by the angle θ :

$$Q = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \text{ and } Q^T = Q^{-1} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$$

The columns of Q are orthogonal. They are unit vectors because $\sin^2 \theta + \cos^2 \theta = 1$, giving us an orthonormal basis for the plane R^2 . The standard basis vectors i and j are rotated through θ . Then Q^{-1} rotates back through $-\theta$. It does agree with Q^T because $\cos -\theta = \cos \theta$ and $\sin -\theta = -\sin \theta$. We have $Q Q^T = I = Q^T Q$.

Example 2 (Permutation): The matrices change the order to (y, z, x) and (y, x) :

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \\ x \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$$

The columns of these Q are orthogonal and also unit vectors. *The inverse of a permutation matrix is its transpose:* $Q^{-1} = Q^T$.

$$\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ z \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

Every permutation matrix is an orthogonal matrix.

Example 3 (Reflection): If u is any unit vector, set $Q = I - 2uu^T$. Notice that uu^T is a matrix while $u^T u$ is the number $\|u\|^2 = 1$. Then Q^T and Q^{-1} both equal Q :

$$Q^T = I - 2uu^T = Q \text{ and } Q^T Q = I - 4uu^T + 4uu^T uu^T = I$$

Reflection matrix $I - 2uu^T$ are symmetric and orthogonal. If we square them, we get the identity matrix: $Q^2 = Q^T Q = I$, where reflecting the mirror twice brings us back to the original $(-1)^2 \rightarrow 1$. Check figure 4.9, page 178. We achieve rotation by $Q = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$ and reflection by $Q = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

Say, we choose the direction $u = (\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, compute $2uu^T$ and subtract it from I to get reflection Q :

$$Q = I - 2 \begin{bmatrix} .5 & -.5 \\ -.5 & .5 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} y \\ x \end{bmatrix}$$

So when we change x, y to y, x , vectors like 3, 3 or on the diagonal don't move because it is on the mirror line. It tells us that rotation, reflection, and permutations don't change the length of any vectors.

Proof $\|Qx\|^2 = \|x\|^2$ because $(Qx)^T(Qx) = x^T Q^T Q x = x^T I x = x^T x$.

If Q has orthonormal columns, it leaves length unchanged. Same length for Qx too. $\|Qx\| = \|x\|$ for every x .

Q also preserves dot products, $(Qx)^T(Qy) = x^T Q^T Q y = x^T y$, or just use $Q^T Q = I$.

4.4.1 Projection $Q Q^T$ Using Orthonormal Bases: Q Replaces A

For projection onto subspaces all far, all formula involves $A^T A$, where the entries are dot products $a_i^T a_j$ of the basis vector a_1, \dots, a_n . Say, *what if these basis vectors are orthonormal?* The a 's become q 's. Then $A^T A$ simplifies to $Q^T Q = I$. Look at the improvement for \hat{x}, p, P .

$$\hat{x} = Q^T b, p = Q \hat{x}, P = Q Q^T$$

The least squares solution of $Qx = b$ is $\hat{x} = Q^T b$. The projection matrix is $Q Q^T$.

$$\text{Projection onto } q\text{'s}, p = \begin{bmatrix} q_1 & \dots & q_n \end{bmatrix} \begin{bmatrix} q_1^T b \\ \vdots \\ q_n^T b \end{bmatrix} = q_1(q_1^T b) + \dots + q_n(q_n^T b)$$

Important, when Q is square and $m = n$, the subspace is whole space. Meaning that $Q^T = Q^{-1}$ and $\hat{x} = Q^T b$ is the same as $x = Q^{-1} b$, the solution are exact. The projection of b onto the whole space is b itself. So $p = b$ and $P = QQ^T = I$.

Now that doesn't sound like much, but when $p = b$, then a formula assembles b out of its 1-dimensional projections. If q_1, \dots, q_n is an orthonormal basis for the whole space, then Q is square. Every $b = QQ^T b$ is the sum of its component along the q 's:

$$b = q_1(q_1^T b) + \dots + q_n(q_n^T b)$$

Transform $QQ^T = I$ is the foundation of Fourier series and all the great transform. It breaks $f(x)$ into perpendicular pieces then adding the pieces like the equation above, inverse transform puts b and $f(x)$ back together.

Example 4: The columns of this orthogonal Q are orthonormal vectors q_1, q_2, q_3 :

$$m = n = 3, Q = \frac{1}{3} \begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & 2 \\ 2 & 2 & -1 \end{bmatrix} \text{ has } Q^T Q = QQ^T = I$$

Separate projection b onto q_1, q_2, q_3 are p_1, p_2, p_3 :

$$q_1(q_1^T b) = \frac{2}{3}q_1, q_2(q_2^T b) = \frac{2}{3}q_2, q_3(q_3^T b) = -\frac{1}{3}q_3$$

The first two sum of projection $p_1 + p_2$ projects b onto the **plane** of q_1 and q_2 . The sum of all three is the projection of b onto the whole space, where $p_1 + p_2 + p_3 = b$:

$$\frac{2}{3}q_1 + \frac{2}{3}q_2 + \frac{1}{3}q_3 = \frac{1}{9} \begin{bmatrix} -2 & +4 & -2 \\ 4 & -2 & -2 \\ 4 & +4 & 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = b$$

4.4.2 The Gram-Schmidt Process

To make orthonormal vectors, we will need to start with independent vectors. Say a, b, c . Will we turn these into orthogonal vectors A, B, C . Then we will turn them orthonormal by $q_1 = \frac{A}{\|A\|}, q_2 = \frac{B}{\|B\|}, q_3 = \frac{C}{\|C\|}$, and so on for other cases.

We start by choosing $A = a$, as the first direction. Next, B must be perpendicular to A . We can do that by making b minus its projection on A , which is just its e , which is of course perpendicular.

$$B = b - \frac{A^T b}{A^T A} A$$

We should confirm that they are orthogonal by $A^T B = 0$. If B is 0, that means A and B is dependent. The direction for A and B is now set and we can continue to C .

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

The core idea of Gram-Schmidt process is that we *subtract from every new vector its projections in the direction already set*. If we had D , we would minus d with its projection on A, B, C . Then to turn them orthonormal, we divide them by their norm length.

Example, suppose we have three independent non-orthogonal vectors a, b, c :

$$a = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \text{ and } b = \begin{bmatrix} 2 \\ 0 \\ -2 \end{bmatrix} \text{ and } \begin{bmatrix} 3 \\ -3 \\ 3 \end{bmatrix}$$

The math says, $A = a, A^T A = 2, A^T b = 2$. Subtract b from its projection on A .

$$B = b - \frac{A^T b}{A^T A} A = b - \frac{2}{2} A = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$$

Check whether $A^T B = 0$ or not, it should. Now, we do the same to c on A, B for C .

$$C = c - \frac{A^T C}{A^T A} A - \frac{B^T C}{B^T B} B = c - \frac{6}{2} A + \frac{6}{6} B = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Check that C is perpendicular to A and B . Then we convert them to unit, A, B, C 's length should be $\sqrt{2}, \sqrt{6}, \sqrt{3}$ respectively.

$$q_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \text{ and } q_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \text{ and } q_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

A, B, C usually contains fraction, q_1, q_2, q_3 usually contains square roots. That is just the price we have to pay for this tool.

4.4.3 The Factorization $A = QR$

Since we moved from A to Q , and found that they are related, then what is the third matrix for their relation? It's R in $A = QR$, not the same R from chapter 1. If look back, we will see how the later q 's are not involved in the prior q 's

1. The vector a, A, q_1 are all along a single line.
2. The vector a, b, A, B, q_1, q_2 are all in the same place.
3. The vector $a, b, c, A, B, C, q_1, q_2, q_3$ are in one 3D subspace.

We can see that the R is a triangular matrix.

$$\begin{bmatrix} a & b & c \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} q_1^T a & q_1^T b & q_1^T c \\ q_2^T a & q_2^T b & q_2^T c \\ q_3^T a & q_3^T b & q_3^T c \end{bmatrix} \text{ or } A = QR$$

Remember that $R = Q^T A$ is upper triangular because later q 's are orthogonal to earlier a 's.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & -3 \\ 0 & -2 & 3 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{18} \\ 0 & \sqrt{6} & -\sqrt{6} \\ 0 & 0 & \sqrt{3} \end{bmatrix} = QR.$$

We look closely at the diagonal of R , we can see the length of $A, B, C = \sqrt{2}, \sqrt{6}, \sqrt{3}$ there. We must remember that this is useful for least square.

$$A^T A = (QR)^T QR = R^T Q^T QR = R^T R, A^T A \hat{x} = A^T b \text{ simplifies to } R^T R \hat{x} = R^T Q^T b, \text{ and finally } R \hat{x} = Q^T b$$

$$\text{Least squares: } R^T R \hat{x} = R^T Q^T b \text{ or } R \hat{x} = Q^T b \text{ or } \hat{x} = R^{-1} Q^T b$$

While $Ax = b$ is impossible, $R \hat{x} = Q^T b$ is very fast. Page 183 will have a code and logic worth looking through.

Key ideas:

1. If the orthonormal vectors q_1, \dots, q_n are the columns of Q , then $q_i^T q_j = 0$ and $q_i^T q_i = 1$, which translate to $Q^T Q = I$.
2. If Q is square then $Q^T = Q^{-1}$, transpose = inverse.
3. Length of Qx equals the length of x , $\|Qx\| = \|x\|$.
4. The projection onto the column space of Q spanned by the q 's is $P = QQ^T$.
5. If Q is square then $P = QQ^T = I$ and every $b = q_1(q_1^T b) + \dots + q_n(q_n^T b)$.
6. Gram-Schmidt produces orthonormal vectors q_1, q_2, q_3 from independent a, b, c . In matrix form, it is the factorization $A = QR$.

4.5 The Pseudoinverse of a Matrix

$$r = m = n = 2I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} I = I^{-1}, (I)(I) = I$$

$$r = m < n = 3I_L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} I_L = \text{left inverse of } I_R I_R = \text{right inverse of } I_L$$

$$r = n < m = 3I_R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} (I_L)(I_R) = I(I_R)(I_L) \neq I$$

Only the first I is truly invertible. I_L and I_R have one-sided inverses but not true inverses. Every matrix A have a **pseudoinverse** A^+ . Which includes left inverse and right inverse.

A has a left inverse:

1. $A^+A = I$
2. A has independent columns
3. A can be tall and thin, $r = n$
4. $Ax = b$ can have no solution or one solution.
5. $N(A) = \text{zero vector}$
6. $A^T A$ is $n \times n$ and invertible.
7. The left inverse is $A^+(A^T A)^{-1} A^T$

A has a right inverse:

1. $AA^+ = I$
2. A has independent rows
3. A can be short and wide, $r = m$
4. $Ax = b$ can have one or many solutions
5. Left nullspace $N(A)^T = \text{zero vector}$
6. AA^T is $m \times m$ and invertible
7. $A^+ = A^T(AA^T)^{-1}$

$A^+A = I$ describes the matrices in the least square chapter. The rank of $r = n < m$ where $Ax = b$ might be unsolvable. Then $\hat{x} = A^+b$ is the vector that solves $A^T A \hat{x} = A^T b$ and makes the error $e = b - A\hat{x}$ as small as possible, that is least square.

$AA^+ = I$ describes the opposite problem, $r = m < n$, which means $Ax = b$ has infinitely many solutions. In this case $x^+ = A^+b$ is the minimum length solution to $Ax = b$. The solution is in the row space of A and all other solutions $x = x^+ + x_n$ have a nullspace component x_n , which increases the length.

x^+ is the minimum norm least squares solution to $Ax = b$

$x^+ = A^+b$ also minimizes $\|x\|^2$. x^+ has nullspace component = 0. If A has dependent columns, we need A^+ .

4.5.1 The Pseudoinverse A^+ ($n \times m$) of a Matrix A ($m \times n$)

To get this, we must remember a few things.

1. Every y in m dimensions has two perpendicular parts, $y = b + z$.
2. b is in the column space of A where z is in the nullspace of A^T .
3. We can invert $Ax = b$ to find $A^+b = x$.

$$\text{Pseudoinverse of } A \text{ has } A^+b = x, A^+z = 0, A^+y = A^+b + A^+z = x$$

If I were to put it into simple words, pseudoinverse is a practical, flawed, and yet the best undo button we have. It is guaranteed to exist, unlike a perfect inverse. If you look on page 191 of Strang's, you will see that A^+ looks at $y = b + z$, and ignores z then send it off to an appropriate x . However, in a case where $y = b + 0$, it is the same since no matter what the z is, it is ignored.

If we look at it, it is basically projection onto the row or col space. Therefore, each projection onto the row and col space have an equation.

$$P_{row} = A^+A = (A^+A)^2 = (A^+A)^T, P_{col} = AA^+ = (AA^+)^2 = (AA^+)^T$$

Example 1:

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ y_2 \end{bmatrix} = b + z, A^+ = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix}$$

$$b = \begin{bmatrix} y_1 \\ 0 \end{bmatrix} \text{ is in the column space of } A, Ax = b \text{ is } \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{y_1}{2} \\ 0 \end{bmatrix} = \begin{bmatrix} y_1 \\ 0 \end{bmatrix}$$

$$z = \begin{bmatrix} 0 \\ y_2 \end{bmatrix} \text{ in the nullspace of } A^T, A^T z = 0, \text{ you get the jazz.}$$

$$\text{Then it ends at } A^+y = A^+b + A^+z = x + 0 = \begin{bmatrix} \frac{y_1}{2} \\ 0 \end{bmatrix}. \text{ Then } A^+ = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix} \text{ where } A^+A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Another useful usecase is diagonal matrix D , square or rectangular:

$$\text{Pseudoinverse of } D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{bmatrix} \rightarrow D^+ = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Remember though, $\frac{1}{0}$ must be zero. This also extends from a diagonal D to any matrix $A = BDC$. If B and C are invertible then $A^+ = B^{-1}D^+C^{-1}$, and for completion, if only B is invertible, then $A^+ = B^+D^{-1}C^{-1}$

Is there a general solution Yes, SVD. But that is still far off. For now, we are only concerned with solution for specific cases. Tall thin matrices, and short wide matrices. Most matrices do not have a A^{-1} where $A^{-1}A = I$ and $AA^{-1} = I$, only square invertible matrices can do that.

For tall thin matrices where the columns are independent, $A^+ = (A^T A)^{-1} A^T$ is the left inverse of A . Which makes $A^+A = I$, but $AA^+ \neq I$.

Then for short wide matrices where the rows are independent, $A^+ = A^T (AA^T)^{-1}$, as the right inverse of A . Here, $AA^+ = I$, but $A^+A \neq I$.