18.06: Linear Algebra

Personal notes from MIT OpenCourseWare 18.06

Himanshu Dongre

Based on lectures by Gilbert Strang

These are personal study notes written while following MIT OpenCourseWare 18.06: Linear Algebra lectures by Gilbert Strang. They are not official MIT course materials and are not endorsed by MIT. Quotations are attributed to the lecturer; all organization, annotations, and any errors are my own.

1. The Geometry of Linear Equations

"Hi. This is the first lecture in MIT's course 18.06 linear algebra, and I'm Gilbert Strang."

Outline.

- n linear equations, n unknowns
- Row picture
- $\star$ Column picture
- Matrix form

$\star$ The fundamental problem of linear algebra is to solve a system of linear equations.

Definition. A *matrix* is a rectangular array of numbers.

Example 1. Consider the linear equations:-
(1) $2x - y = 0$
(2) $-x + 2y = 3$
`Matrix form:` $\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$

- The *coefficient matrix* denoted by A.
- The *vector* of unknowns denoted by x.
- The RHS vector denoted by b.

$\star$ Linear equations are always $\boxed{Ax = b}$.

`Row picture:` The X-Y plot of the two equations (the two lines, with their intersection point denoting the solution).

`Algebra for column picture:` $x \begin{bmatrix} 2 \\ -1 \end{bmatrix} + y \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$. (This is called a *linear combination* and is the most fundamental operation of the whole course).
`Column picture:` Column 1 would be $2\hat{\imath} - \hat{\jmath}$, column 2 would be $-\hat{\imath} + 2\hat{\jmath}$. If I take 1 of column 1 and 2 of column 2, I get $3\hat{\jmath}$. This picture is the column picture.

Question. If I took *all* the x's and *all* the y's, what would be *all* the results of the linear combinations of columns 1 and 2?
Answer. I could get any RHS at all. The combinations of the two columns would fill the whole plane.

Example 2. Consider the linear equations:-
(1) $2x - y = 0$
(2) $-x + 2y - z = -1$
(3) $-3y + 4z = 4$

`Matrix form` $(Ax = b)$`:` $A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -3 & 4 \end{bmatrix}$, $b = \begin{bmatrix} 0 \\ -1 \\ 4 \end{bmatrix}$.

`Row picture:` 3 planes on a X-Y-Z plot, one for each equation. They meet at a point, which is the solution.

`Algebra for column picture:` $x \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} + y \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix} + z \begin{bmatrix} 0 \\ -1 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 4 \end{bmatrix}$.

`Column picture:` Column 1 would be $2\hat{\imath} - \hat{\jmath}$, column 2 would be $-\hat{\imath} + 2\hat{\jmath} - 3\hat{k}$, column 3 would be $-\hat{\jmath} + 4\hat{k}$.

In this case, the RHS is same as column 3. Hence $x = 0$, $y = 0$, $z = 1$, or $(0, 0, 1)$ is the solution.

Question. Can I solve $Ax = b$ for *every* RHS b?
Rephrase. Do the linear combinations of the 3 columns fill 3-D space?
Answer. For the A in example 2, the answer is YES. The matrix is non-singular and invertible.

Question. What could go wrong?
Rephrase. When could the linear combinations of 3 columns not be able to produce some b?
Answer. If the 3 columns lie in the same plane, then their combinations will lie in the same plane, and hence not fill up 3-D space. This would be a singular case, the matrix would be not invertible.

Example 3. Imagine that we have vectors with 9 components. If we had 9 equations in 9 unknowns, then we would have 9 columns, and each would be a vector in 9-D space. We would be finding linear combinations of 9 vectors in 9-D space, for the one that would hit the RHS b. If the matrix were non-singular and invertible (*beautiful*), the linear combinations would fill up 9-D space. If the 9$^{\text{th}}$ column were same as the 8$^{\text{th}}$ column, then the linear combinations would fill up an 8-D plane in 9-D space.

Definition. *Multiplication of a matrix by a vector.*
`Method 1, Columns:` $\begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 2 \begin{bmatrix} 5 \\ 3 \end{bmatrix} = \begin{bmatrix} 12 \\ 7 \end{bmatrix}$.
`Method 2, Dot Product:` $\begin{bmatrix} 2 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 12,\ \begin{bmatrix} 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = 7$.

⋆ Ax is a (linear) combination of the columns of A.

2. Elimination with Matrices

"Gauss thought of it before we did, but only because he was born earlier."

Outline.

- Elimination - success, failure.
- Back-substitution.
- Elimination matrices.
- Matrix multiplication.

Example 1. Consider the system:-
(1) $x + 2y + z = 2$
(2) $3x + 8y + z = 12$
(3) $4y + z = 2$

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 3 & 8 & 1 \\ 0 & 4 & 1 \end{bmatrix} \xrightarrow[\text{(2,1) step}]{(2)-3(1)} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix} \xrightarrow[\text{(3,1) step}]{\text{Already 0}} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 4 & 1 \end{bmatrix} \xrightarrow[\text{(3,2) step}]{(3)-2(2)} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 5 \end{bmatrix} = U(\text{upper-triangular})$$

The three *pivots* are **1**, **2**, and **5**.
⋆ Pivots can't be 0. 0 is not a pivot. We could have a 0 in a pivot position. "I'll never utter the words *0 pivot*."

Question. How could elimination have failed?
Answer. If I don't have 3 pivots after any necessary row exchanges, then elimination fails. A 0 in a pivot position may only mean that a row exchange is needed.

⋆ When elimination cannot produce all pivots even after row exchanges, the matrix is singular and not invertible.

Elimination on augmented matrix:-

$$Ax = b \Rightarrow \begin{bmatrix} A & b \end{bmatrix} = \begin{bmatrix} 1 & 2 & 1 & | & 2 \\ 3 & 8 & 1 & | & 12 \\ 0 & 4 & 1 & | & 2 \end{bmatrix} \xrightarrow[\text{(2,1) step}]{(2)-3(1)} \begin{bmatrix} 1 & 2 & 1 & | & 2 \\ 0 & 2 & -2 & | & 6 \\ 0 & 4 & 1 & | & 2 \end{bmatrix} \xrightarrow[\text{(3,1) step}]{\text{Already 0}} \begin{bmatrix} 1 & 2 & 1 & | & 2 \\ 0 & 2 & -2 & | & 6 \\ 0 & 4 & 1 & | & 2 \end{bmatrix} \xrightarrow[\text{(3,2) step}]{(3)-2(2)}$$
$$\begin{bmatrix} 1 & 2 & 1 & | & 2 \\ 0 & 2 & -2 & | & 6 \\ 0 & 0 & 5 & | & -10 \end{bmatrix} = \begin{bmatrix} U & c \end{bmatrix} \Rightarrow Ux = c$$

⇒ (1') $x + 2y + z = 2$
(2') $2y - 2z = 6$

(3') $5z = -10 \Rightarrow z = -2$

Back-substitute in (2') to obtain $y = 1$

Back-substitute in (1') to obtain $x = 2$

$(\mathbf{2}, \mathbf{1}, -\mathbf{2})$ is the solution.

Elimination matrices:-

1. Subtract 3(1) from (2): $E_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
2. Do nothing: $E_{31} = I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
3. Subtract 2(2) from (3): $E_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}$

$\Rightarrow E_{32}E_{31}E_{21}A = U \Rightarrow EA = U$, where $E = E_{32}E_{31}E_{21}$.

$\star$ To do column operations, a matrix multiplies on the right. To do row operations, a matrix multiplies on the left.

Inverses. What matrix *undoes* the operation performed by E_{21}? The matrix that adds 3(1) to (2).

$\Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

$\Rightarrow E_{21}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

3. Multiplication and Inverse Matrices

"All key equations in mathematics have a zero on the right hand side."

Outline.

- Matrix multiplication.
- Inverse of A.
- Gauss-Jordan / Find A^{-1}.

1. Let $C = AB$. Then the entry c_{ij} comes from row i of A and column j of B.
   $c_{ij} =$ (row i of A) $\bullet$ (column j of B) $= \sum_{k=1}^{n} a_{ik}b_{kj}$. If A is of order $m \times n$, then the number of rows in B must be n, i.e. B must be of the order $n \times p$. Then, C is of the order $m \times p$.

2. Let $A_{m\times n}$ and $B_{n\times p}$ be matrices. Let $C_{m\times p} = AB$. Then, A times the i^{th} column of B gives the i^{th} column of C. Hence, every column of C is a combination of the columns of A, since it's a product of A and a vector.

3. Let $A_{m\times n}$ and $B_{n\times p}$ be matrices. Let $C_{m\times p} = AB$. Then, the i^{th} row of A times B gives the i^{th} row of C. Hence, every row of C is a combination of the rows of B, since it's a product of a row vector and B.

$\star$ (column of A)$\times$(row of B) gives a matrix of order $m \times p$, in which every column is a multiple of the multiplicand and every row is a multiple of the multiplier.

4. Let $A_{m\times n}$ and $B_{n\times p}$ be matrices. Let $C_{m\times p} = AB$. Then, $C = \sum_{k=1}^{n}$(column k of A)$\times$(row k of B).

5. `Block multiplication:` Suppose A and B are square matrices of the same size, we can split both into blocks. Let $C = AB$. Then, the block $C_{ij} =$ (block row i of A) $\bullet$ (block column j of B).

Inverses (square matrices). Let A be a matrix. Then, the *left inverse* is a matrix A_{L} such that $A_{\mathrm{L}}A = I$, and the *right inverse* is a matrix A_{R} such that $AA_{\mathrm{R}} = I$. If A is a square matrix, then $A_{\mathrm{L}} = A_{\mathrm{R}} = A^{-1}$ such that $A^{-1}A = AA^{-1} = I$. The inverse of a square matrix may not exist. Square matrices that don't have an inverse are *singular* or *non-invertible*.

Example 1. The matrix $A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$ is singular (has no inverse), because A times any vector will result in a vector whose columns are the multiples of $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$. Hence, there is no matrix which can multiply A to produce the identity.

$\star$ A matrix A is singular if there exists a vector $x \neq 0$ such that $Ax = 0$.
For the A in example 1, such an x is $\begin{bmatrix} 3 \\ -1 \end{bmatrix}$.

$\star$ A times column j of A^{-1} gives column j of the identity.

Let $A = \begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix}$. Let $A^{-1} = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$. We know that $AA^{-1} = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Finding A^{-1} is as good as solving the system:-
(1) $\begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$
(2) $\begin{bmatrix} 1 & 3 \\ 2 & 7 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

Gauss-Jordan (solve two equations at once):-

$$\begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} 1 & 3 & | & 1 & 0 \\ 2 & 7 & | & 0 & 1 \end{bmatrix} \xrightarrow{(2)-2(1)} \begin{bmatrix} 1 & 3 & | & 1 & 0 \\ 0 & 1 & | & -2 & 1 \end{bmatrix} \xrightarrow{(1)-3(2)} \begin{bmatrix} 1 & 0 & | & 7 & -3 \\ 0 & 1 & | & -2 & 1 \end{bmatrix} = \begin{bmatrix} I & A^{-1} \end{bmatrix}.$$

When Gauss-Jordan succeeds, there is an elimination matrix E such that $E\begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} I & E \end{bmatrix}$
$\Rightarrow EA = I \Rightarrow E$ must be A^{-1}
$\Rightarrow E\begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} I & A^{-1} \end{bmatrix}$.

4. Factorization into $A = LU$

"It's not that I don't like 18.01, but 18.06 is better."

Outline.

- Inverse of AB, A^T
- Product of elimination matrices
- $A = LU$ (no row exchanges)

Suppose A and B are invertible matrices, then the inverse of AB is $B^{-1}A^{-1}$.
Suppose A is a square invertible matrix $\Rightarrow AA^{-1} = I \Rightarrow (AA^{-1})^\mathsf{T} = I^\mathsf{T} = I \Rightarrow (A^{-1})^\mathsf{T}A^\mathsf{T} = I \Rightarrow (A^\mathsf{T})^{-1} = (A^{-1})^\mathsf{T}$.

Example 1. Let $A = \begin{bmatrix} 2 & 1 \\ 8 & 7 \end{bmatrix}$. Then $E_{21} = \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix}$ is such that $E_{21}A = U = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$
$\Rightarrow A = E_{21}^{-1}U$. $E_{21}^{-1} = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix}$. Call it L (lower-triangular).
$\Rightarrow A = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix} = LU$.
For a good matrix, L has 1's on the diagonal, and U has the pivots on the diagonal.
Sometimes, we may want to separate out the pivots as $A = \begin{bmatrix} 1 & 0 \\ 4 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix} = LDU$, and now both L and U have 1's on the diagonal.

Example 2. Let A be a (good) 3×3 matrix. By elimination, $E_{32}E_{31}E_{21}A = U$ (no row exchanges) $\Rightarrow A = (E_{32}E_{31}E_{21})^{-1}U = E_{21}^{-1}E_{31}^{-1}E_{32}^{-1}U = LU$. L is this product of inverses.

$\star$ $A = LU$. If no row exchanges, the multipliers go directly into L.

Question. How many operations in elimination on an $n \times n$ matrix A?
Answer. $\sum_{i=1}^{n} i^2 \approx \frac{1}{3}n^3$

We need to do row exchanges if a 0 shows up in a pivot position.

Permutations. All 3×3 permutation matrices:-

1. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$: Permutation matrix that doesn't do anything. The identity. $(1, 2, 3)$
2. $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$: Permutation matrix that exchanges rows 1 and 2. $(2, 1, 3)$
3. $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$: Permutation matrix that exchanges rows 1 and 3. $(3, 2, 1)$
4. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$: Permutation matrix that exchanges rows 2 and 3. $(1, 3, 2)$
5. $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$: Permutation matrix that cycles row $2 \to 3 \to 1$. $(2, 3, 1)$
6. $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$: Permutation matrix that cycles rows $3 \to 1 \to 2$. $(3, 1, 2)$

They form a *group*(!) under the operation of matrix multiplication.
$\star$ Their inverses are their transposes. $P^{-1} = P^{\mathsf{T}} \Rightarrow PP^{\mathsf{T}} = P^{\mathsf{T}}P = I$.

5. Transposes, Permutations, Spaces $\mathbb{R}^n$

"Okay, thanks!"

Outline.

- $PA = LU$
- Vector spaces and subspaces

In bad case of elimination when 0's are found in pivot positions, row exchanges are required, which are carried out using permutation matrices. In this case, $A = LU$ becomes $PA = LU$. In the good case, $P = I$. Hence, for any invertible A, $PA = LU$.

Permutations. A permutation matrix is the identity with reordered rows (includes the case of the rows being in the same order). Hence there are $n!$ $n \times n$ permutation matrices.
A permutation matrix P has the property $P^{-1} = P^{\mathsf{T}} \Rightarrow PP^{\mathsf{T}} = P^{\mathsf{T}}P = I$.

Transposes. $(A^{\mathsf{T}})_{ij} = A_{ji}$.
Let $A = \begin{bmatrix} 1 & 3 \\ 2 & 3 \\ 4 & 1 \end{bmatrix}$. Then, $A^{\mathsf{T}} = \begin{bmatrix} 1 & 2 & 4 \\ 3 & 3 & 1 \end{bmatrix}$.
If the order of a matrix is $m \times n$, then the order of its transpose is $n \times m$.

Symmetric matrices. $A^{\mathsf{T}} = A$.
$A = \begin{bmatrix} 1 & 3 & 7 \\ 3 & 2 & 9 \\ 7 & 9 & 4 \end{bmatrix}$ is a symmetric matrix.

$R^\mathsf{T}R$ is symmetric for any rectangular matrix R, since $(R^\mathsf{T}R)^\mathsf{T} = R^\mathsf{T}(R^\mathsf{T})^\mathsf{T} = R^\mathsf{T}R$.

Vector spaces. Spaces of vectors.

Example 1. $\mathbb{R}^2$: all 2-D real vectors such as $\begin{bmatrix}3\\2\end{bmatrix}, \begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\pi\\ \mathrm{e}\end{bmatrix} \dots$ = the X-Y plane.

Example 2. $\mathbb{R}^3$: all 3-D vectors (vectors with 3 real components).

Example 3. $\mathbb{R}^n$: all vectors with n real components.

⋆ A vector space must be *closed* under addition, and multiplication by scalars, and hence under linear combinations.
⋆ A vector space must contain the zero vector (denoted by *0*).

Subspaces. Subspaces of $\mathbb{R}^2$:-

1. All of $\mathbb{R}^2$
2. Any line through $\begin{bmatrix}0\\0\end{bmatrix}$
3. Zero vector only (denoted by Z)

Example. Let $A = \begin{bmatrix}1 & 3\\2 & 3\\4 & 1\end{bmatrix}$.
The columns of A are in $\mathbb{R}^3$, and all their linear combinations form a subspace called the *column space* (denoted by $\mathrm{C}(A)$). This subspace forms a plane in $\mathbb{R}^3$.

6. Column Space and Nullspace

> "… anybody's going to say, 'No, you dope, you can't usually solve four equations with only three unknowns.' "

Outline.

- Vector spaces and subspaces
- Column space of A: solving $Ax = b$
- Nullspace of A

Vector space. Vector space requirements: $v + w$ and $\mathsf{c}v$ are *in the space*, for any scalar $\mathsf{c} \Rightarrow$ all combinations $\mathsf{c}v + \mathsf{d}w$ are *in the space*.

Example 1. $\mathbb{R}^3$

Subspace. A vector space within a vector space. All subspaces have got to contain *0*.

Example 1. P: A plane through $\begin{bmatrix}0\\0\\0\end{bmatrix}$ is a subspace of $\mathbb{R}^3$.

Example 2. L: A line through $\begin{bmatrix}0\\0\\0\end{bmatrix}$ is a subspace of $\mathbb{R}^3$.

Though P and L are subspaces of $\mathbb{R}^3$, $P \cup L$ is *not* a subspace of $\mathbb{R}^3$ unless one contains the other. However, $P \cap L$ *is* a subspace of $\mathbb{R}^3$.

This can be generalized to any two subspaces. Suppose S and T are subspaces, then $S \cap T$ is a subspace but $S \cup T$ is not necessarily one.

Column space of a matrix. The column space of a matrix A, denoted by $\mathrm{C}(A)$, is all linear combinations of the columns of A.

Example 1. Let $A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix}$. Then, $\mathrm{C}(A)$ is a subspace of $\mathbb{R}^4$, because A is a 4×3 matrix.

Question. Does $Ax = b$ have a solution for every RHS b?
Answer. *No*, since $Ax = b$ is 4 equations and only 3 unknowns.

The system looks like: $\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$.

Question. Which vectors b allow this system to be solved?
Answer. Can solve $Ax = b$ exactly when b is in the column space of A.

Nullspace of a matrix. The nullspace of a matrix A, denoted by $\mathrm{N}(A)$, contains all the solutions x to the equation $Ax = \mathit{0}$.

Example 1. Let $A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix}$. Then, $\mathrm{N}(A)$ contains all vectors $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ such that $Ax = \mathit{0}$.

The system looks like: $\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$.

$\mathrm{N}(A)$ trivially contains $\mathit{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. It also contains $\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 6 \\ 6 \\ -6 \end{bmatrix}, \begin{bmatrix} -4 \\ -4 \\ 4 \end{bmatrix}$, etc. Hence, $\mathrm{N}(A)$ contains all vectors of the form $\begin{bmatrix} \mathsf{c} \\ \mathsf{c} \\ -\mathsf{c} \end{bmatrix} = \mathsf{c} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$, where c is any scalar (including 0).
Hence, $\mathrm{N}(A)$ is a line through $\mathit{0}$ in $\mathbb{R}^3$, and therefore a subspace.

$\star$ The nullspace is *always* a vector space.
Check: If $Av = \mathit{0}$ and $Aw = \mathit{0}$, then $A(v + w) = Av + Aw = \mathit{0} + \mathit{0} = \mathit{0}$.
If $Av = \mathit{0}$, then $A(\mathsf{c}v) = \mathsf{c}(Av) = c\mathit{0} = \mathit{0}$.

Example. Consider the system $\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \\ 4 & 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$. Then, $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ is a solution.

Question 1. Are there more solutions?

Answer 1. Yes (eg. $\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$).

Question 2. Do all the solutions form a vector space?

Answer 2. No, because $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ is not a solution.

$\star$ Two important ways to construct subspaces:-

1. Take linear combinations of a few given vectors, then, by definition, they form a subspace (like the column space of a matrix).
2. Find all vectors that satisfy some given requirement and show that they form a subspace (like the nullspace of a matrix).

7. Solving $Ax = 0$: Pivot Variables, Special Solutions

"Elimination better discover that."

Outline.

- Computing the nullspace ($Ax = 0$)
- Pivot variables, free variables
- Special solutions — rref(A) $= R$

Example 1. Let $A = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{bmatrix}$.

⋆ The nullspace remains unchanged by elimination, since it consists of the solutions to $Ax = 0$, and elimination keeps the solutions to $Ax = b$ unchanged. However, elimination changes the column space.

By elimination:-

$$A = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 8 & 10 \end{bmatrix} \xrightarrow[\text{Do Column 1}]{\text{(2) - 2(1); (3) - 3(1)}} \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 2 & 4 \end{bmatrix} \xrightarrow[\text{Do Column 2}]{\text{Nothing to do}} \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 2 & 4 \end{bmatrix} \xrightarrow[\text{Do Column 3}]{(3)-(2)} \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} = U$$

- The resultant matrix, U, is in the so-called *echelon* (staircase) form.

- The *rank* of a matrix is equal to the number of pivots, which, in this case, is 2.

- The *pivot columns* are the columns with the pivots. In this case, there are 2 pivot columns viz. columns 1 and 3.

- The *free columns* are those which are not pivot columns. In this case, there are 2 free columns viz. columns 2 and 4. They are so-called because any value can be (freely) assigned to their multipliers in the solution.

Suppose $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ is a solution to $Ux = 0$. Then x_2 and x_4 can be assigned values freely because they multiply free columns, and the system:-
(1) $x_1 + 2x_2 + 2x_3 + 2x_4 = 0$
(2) $2x_3 + 4x_4 = 0$
can then be solved for x_1 and x_3.
Suppose we systematically assign the values $x_2 = 1$ and $x_4 = 0$. Then, $x_3 = 0$ from (2) $\Rightarrow x_1 = -2$ from (1).
Hence, $x = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$ is a solution to $Ax = 0$ and in turn a vector in the nullspace of A. Then, any multiple of x, $c \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, is also in the nullspace.
Suppose we now assign the values $x_2 = 0$ and $x_4 = 1$. Then, $x_3 = -2$ from (2) $\Rightarrow x_1 = 2$ from (1). Hence,

$x = \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}$ is in the nullspace. Then, any multiple of x, $\mathsf{d}\begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}$ is also in the nullspace.

$\therefore$ Solutions to $Ax = \mathit{0}$ (we really solved $Ux = \mathit{0}$):-

- *Special* solution(s): $\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}$
- *General* solution: Any combination of the special solution(s) $= \mathsf{c}\begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \mathsf{d}\begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}$

$\star\star\star$ The nullspace contains exactly all the combinations of the special solutions. There is one special solution for every free variable. If the matrix is of size $m \times n$ (here, $m = 3$, $n = 4$), and has rank r (here, $r = 2$), then there are $n - r$ free variables and hence $n - r$ special solutions.

Reduced row echelon form. Work harder on $U = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ to get 0's above *and* below the pivots.

$\star$ When we get a row of 0's, it tells us that one row was a combination of the other rows, and elimination found it and knocked it out.

$U = \begin{bmatrix} 1 & 2 & 2 & 2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[\text{Get 0's above the pivots}]{(1)-(2)} \begin{bmatrix} 1 & 2 & 0 & -2 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow[\text{Make pivots equal to 1}]{(2)\div 2} \begin{bmatrix} 1 & 2 & 0 & -2 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} = R =$ rref(A) (reduced row echelon form).

- Notice $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ in the pivot rows and pivot columns of R. The *free part* of the matrix in the free rows and free columns, $F = \begin{bmatrix} 2 & -2 \\ 0 & 2 \end{bmatrix} \Rightarrow R = \begin{bmatrix} I & F \\ O & O \end{bmatrix}$. Let N be the nullspace matrix, whose columns are special solutions to $Rx = \mathit{0}$; that is, RN gives $\mathit{0}$ columns. Then, N is simply $\begin{bmatrix} -F \\ I \end{bmatrix}$. Another way to look at $Rx = \mathit{0}$ is $\begin{bmatrix} I & F \end{bmatrix}\begin{bmatrix} x_{\text{pivot}} \\ x_{\text{free}} \end{bmatrix} = \mathit{0} \Rightarrow x_{\text{pivot}} + Fx_{\text{free}} = \mathit{0} \Rightarrow x_{\text{pivot}} = -Fx_{\text{free}}$. Hence, if the free variables x_{free} are systematically chosen to be the identity, then the pivot variables $x_{\text{pivot}} = -F$.
- The solutions to $Ax = \mathit{0}$, $Ux = \mathit{0}$, and $Rx = \mathit{0}$ are all the same.

Example 2. Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 2 & 6 & 8 \\ 2 & 8 & 10 \end{bmatrix}$. By elimination:-

$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 2 & 6 & 8 \\ 2 & 8 & 10 \end{bmatrix} \xrightarrow{(2)\text{ - }2(1);\ (3)\text{ - }2(1);\ (4)\text{ - }2(1)} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 4 & 4 \end{bmatrix} \xrightarrow{\text{Swap (2) and (3)}} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 4 & 4 \end{bmatrix} \xrightarrow{(4)-2(2)} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} =$

$U \xrightarrow{(1)-(2)} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{(2)\div 2} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = R.$

- Rank $r = 2 \Rightarrow 2$ pivot columns and 1 free columns $\Rightarrow 1$ special solution of $Ax = \mathit{0}$.

$\star$ The number of pivot columns (rank) of A and A^{T} are the same.

- $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $F = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow x = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$ is a special solution of $Ax = 0 \Rightarrow$ the whole nullspace is the line $\mathsf{c}\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$.

8. Solving $Ax = b$: Row Reduced Form R

"No prize for that one."

Outline.

- Complete solution of $Ax = b$
- Rank r
- $x = x_p + x_n$

Example 1. Consider the system:-
(1) $x_1 + 2x_2 + 2x_3 + 2x_4 = b_1$
(2) $2x_1 + 4x_2 + 6x_3 + 8x_4 = b_2$
(3) $3x_1 + 6x_2 + 8x_3 + 10x_4 = b_3$

⋆ If some combination on the LHS gives all zeros, then the same combination on the RHS must give zero.

$$\begin{bmatrix} A & b \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 & 2 & b_1 \\ 2 & 4 & 6 & 8 & b_2 \\ 3 & 6 & 8 & 10 & b_3 \end{bmatrix} \xrightarrow{\text{(2) - 2(1), (3) - 3(1)}} \begin{bmatrix} 1 & 2 & 2 & 2 & b_1 \\ 0 & 0 & 2 & 4 & b_2 - 2b_1 \\ 0 & 0 & 2 & 4 & b_3 - 3b_1 \end{bmatrix} \xrightarrow{(3)-(2)} \begin{bmatrix} 1 & 2 & 2 & 2 & b_1 \\ 0 & 0 & 2 & 4 & b_2 - 2b_1 \\ 0 & 0 & 0 & 0 & b_3 - b_2 - b_1 \end{bmatrix}$$

- Columns 1 and 3 are the pivot columns.
- Last equation (represented by the zero row) says $0 = b_3 - b_2 - b_1$: condition for solvability.
- Suppose $b = \begin{bmatrix} 1 \\ 5 \\ 6 \end{bmatrix}$. This is an OK b. Then, after elimination, $\begin{bmatrix} U & b \end{bmatrix} = \begin{bmatrix} 1 & 2 & 2 & 2 & 1 \\ 0 & 0 & 2 & 4 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$.

Solvability. $Ax = b$ is *solvable* exactly when b is in $\mathrm{C}(A) \Leftrightarrow$ if a combination of the rows of A gives 0, then the same combination of entries of b must give 0.

To find complete solutions to $Ax = b$:-

1. $x_{\text{particular}}$: set all free variables to zero, solve $Ax = b$ for pivot variables.
   Set $x_2 = 0$, $x_4 = 0$
   $\Rightarrow x_1 + 2x_3 = 1$ and $2x_3 = 3 \Rightarrow x_3 = \frac{3}{2}$, $x_1 = -2$.
   $\Rightarrow x_p = \begin{bmatrix} -2 \\ 0 \\ \frac{3}{2} \\ 0 \end{bmatrix}$.
2. $x_{\text{nullspace}}$: add any vector, x_n, from the nullspace.

⋆ The complete solution is $x = x_p + x_n$, since $A(x_p + x_n) = Ax_p + Ax_n = b + 0 = b$.

$$x_{\text{complete}} = x_{\text{particular}} + x_{\text{nullspace}} = \begin{bmatrix} -2 \\ 0 \\ \frac{3}{2} \\ 0 \end{bmatrix} + \mathsf{c}_1 \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \mathsf{c}_2 \begin{bmatrix} 2 \\ 0 \\ -2 \\ 1 \end{bmatrix}.$$

To plot all solutions x in $\mathbb{R}^4$:-

1. The nullspace is a 2-D subspace in $\mathbb{R}^4$

2. Hence, the plot of all solutions x is this subspace *shifted* to contain x_p instead of $\mathit{0}$ (which doesn't remain a subspace).

Consider an $m \times n$ matrix A of rank r (know $r \leq m$, $r \leq n$).

Case 1: Full column rank. $r = n < m$: Every column has a pivot $\Rightarrow$ No free variables
$\Rightarrow$ $\mathrm{N}(A) = \{\mathit{0}\}$ $\Rightarrow$ Solution to $Ax = b$: $x = x_p$
$\Rightarrow$ *Unique* solution if it exists $\Rightarrow$ **Either 0 or 1 solution.**

`Example:` For the matrix $A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \\ 6 & 1 \\ 5 & 1 \end{bmatrix}$, $m = 4$, $r = n = 2$.

- The reduced row echelon form $R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.
- There are 4 equations and 2 unknowns, so there is not always a solution to $Ax = b$.
- If $b = \begin{bmatrix} 4 \\ 3 \\ 7 \\ 6 \end{bmatrix}$, then $x = x_p = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. In this case of a "good" RHS, a unique solution exists.

Case 2: Full row rank. $r = m < n$: Every row has a pivot $\Rightarrow$ No zero rows
$\Rightarrow$ No requirements for b $\Rightarrow$ Can solve $Ax = b$ for *every* b.
Left with $r = m$ pivot variables and $n - r = n - m$ free variables.

`Example:` For the matrix $A = \begin{bmatrix} 1 & 2 & 6 & 5 \\ 3 & 1 & 1 & 1 \end{bmatrix}$, $r = m = 2$, $n = 4$.

- The reduced row echelon form $R = \begin{bmatrix} 1 & 0 & f_1 & f_2 \\ 0 & 1 & f_3 & f_4 \end{bmatrix}$.
- There are 2 equations and 4 unknowns, so $\mathrm{N}(A)$ contains at least one non-zero vector, and there exist **infinitely many solutions** to $Ax = b$.

Case 3: Full rank. $r = m = n$: Square matrix with no free variables and no zero rows
$\Rightarrow$ $Ax = b$ has a **unique solution** for *every* b.

`Example:` For the matrix $A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$, $r = m = n = 2$.

- The reduced row echelon form $R = I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.
- $\mathrm{N}(A) = \{\mathit{0}\}$.

Case 4. $r < m$, $r < n$: Free variables *and* zero rows exist.
$\Rightarrow$ **0 or infinitely many solutions.**

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

9. Independence, Basis, and Dimension

"That would be a goof."

Outline.

- Linear independence
- Spanning a space
- Basis and dimension

⋆ Suppose A is an $m \times n$ matrix with $m < n$ (more unknowns than equations). Then there are *non-zero* solutions to $Ax = 0$ (Reason: There will be free variables!!).

Independence. Vectors $v_1, v_2, ..., v_n$ are linearly *independent* if no combination gives the zero vector (except the zero combination).

- If one of these vectors is the zero vector, then independence is dead.
- Repeat when $v_1, v_2, ..., v_n$ are columns of a matrix A:-
  - They are independent if N(A) is $\{0\}$. Rank $= n$. No free variables.
  - They are dependent if $Ac = 0$ for some non-zero vector $c \in \mathrm{N}(A)$. Rank $< n$. Yes free variables.

Spanning a space. Vectors $v_1, v_2, ..., v_l$ *span* a vector space means: The space consists of *all* combinations of those vectors.

- The columns of a matrix span its column space (by definition of column space).

Basis. A *basis* for a vector space is a sequence of vectors $v_1, v_2, ..., v_d$ that has two properties:-

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

`Example 1.` Space is $\mathbb{R}^3$. One basis is: $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}$. Another basis is: $\begin{bmatrix}1\\1\\2\end{bmatrix}, \begin{bmatrix}2\\2\\5\end{bmatrix}, \begin{bmatrix}0\\4\\\pi\end{bmatrix}$.

⋆ For $\mathbb{R}^n$, n vectors give a basis if the $n \times n$ matrix with those columns is *invertible*.

⋆ Given a space, the basis for the space is not unique. However, every basis has the same number of vectors.

Dimension. The number of vectors in every basis of a space is the *dimension* of that space.

`Example 1.` Let $A = \begin{bmatrix}1 & 2 & 3 & 1\\1 & 1 & 2 & 1\\1 & 2 & 3 & 1\end{bmatrix}$. Then:-

- The four columns of A are dependent, they span the column space of A, and they are *not* a basis for $\mathrm{C}(A)$.
- The first two columns of A form a basis of $\mathrm{C}(A) \Rightarrow$ rank of A is 2.
- Another basis for $\mathrm{C}(A)$ is $\begin{bmatrix}2\\2\\2\end{bmatrix}, \begin{bmatrix}7\\5\\7\end{bmatrix}$.
- The special solutions to $Ax = 0$ are $\begin{bmatrix}-1\\-1\\1\\0\end{bmatrix}$ and $\begin{bmatrix}-1\\0\\0\\1\end{bmatrix}$. They form a basis for $\mathrm{N}(A) \Rightarrow \dim(\mathrm{n}(A)) = 2$.

⋆ $\dim(\mathrm{N}(A)) = \#$ free variables $= n - r$. The special solutions to $Ax = 0$ form a basis for $\mathrm{N}(A)$.

⋆ $\dim(\mathrm{C}(A)) = \#$ pivot columns $= r$. The pivot columns form a basis for $\mathrm{C}(A)$.

⋆ Given a space and its dimension m, then any m independent vectors in the space would form a basis for it.

10. Four Fundamental Subspaces

"... left nullspace, left out practically ..."

Outline.

- Four fundamental subspaces (of a matrix A)

4 subspaces. The four fundamental subspaces associated with a matrix A are:-

1. Column space $\mathrm{C}(A)$
2. Nullspace $\mathrm{N}(A)$
3. Row space = All combinations of the rows of A = All combinations of the columns of $A^\top = \mathrm{C}(A^\top)$
4. Left nullspace = Nullspace of $A^\top = \mathrm{N}(A^\top)$

Let A be an $m \times n$ matrix with rank r. Then:-

- $\mathrm{C}(A)$ and $\mathrm{N}(A^\top)$ are in $\mathbb{R}^m$.
- $\mathrm{N}(A)$ and $\mathrm{C}(A^\top)$ are in $\mathbb{R}^n$.
- A basis for $\mathrm{C}(A)$ is the pivot columns and $\dim(\mathrm{C}(A)) = r$.
- $\dim(\mathrm{C}(A^\top)) = r$.
- A basis for $\mathrm{N}(A)$ is the special solutions to $Ax = \mathit{0}$ and $\dim(\mathrm{N}(A)) = n - r$.
- $\dim(\mathrm{N}(A^\top)) = m - r$
- $\dim(\mathrm{C}(A)) + \dim(\mathrm{N}(A)) = \#$ columns of A
- $\dim(\mathrm{C}(A^\top)) + \dim(\mathrm{N}(A^\top)) = \#$ rows of A.

Example 1. Let $A = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{bmatrix}$. By elimination and row reduction:-

$$A = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 1 \end{bmatrix} \xrightarrow{\text{(2) - (1), (3) - (1)}} \begin{bmatrix} 1 & 2 & 3 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = U \xrightarrow{(1)-(-2)(2)} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \xrightarrow{-1\times(2)} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = R.$$

$\star$ Elimination and row reduction do row operations and hence preserve the row space but change the column space; $\mathrm{C}(R) \neq \mathrm{C}(A)$.

- A basis for $\mathrm{C}(A^\top)$ and $\mathrm{C}(R^\top)$ is the first r rows (all non-zero rows) of R.
- $\mathrm{N}(A^\top)$ contains all vectors y such that $A^\top y = \mathit{0} \Rightarrow (A^\top y)^\top = y^\top A = \mathit{0}^\top$

Suppose $\mathrm{rref}(\begin{bmatrix} A_{m\times n} & I_{m\times m} \end{bmatrix}) \rightarrow \begin{bmatrix} R_{m\times n} & E_{m\times m} \end{bmatrix}$ for some matrix E. Then we know that $EA = R$. (When A is an invertible square matrix, $R = I$ and $E = A^{-1}$). By elimination and row reduction (again):-

$$\begin{bmatrix} A & I \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 1 & 0 & 1 & 0 \\ 1 & 2 & 3 & 1 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\text{(2) - (1), (3) - (1)}} \begin{bmatrix} 1 & 2 & 3 & 1 & 1 & 0 & 0 \\ 0 & -1 & -1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix} \xrightarrow{(1)-(-2)(2)} \begin{bmatrix} 1 & 0 & 1 & 1 & -1 & 2 & 0 \\ 0 & -1 & -1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix} \xrightarrow{-1\times(2)} \begin{bmatrix} 1 & 0 & 1 & 1 & -1 & 2 & 0 \\ 0 & 1 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} R & E \end{bmatrix}$$

$\Rightarrow E = \begin{bmatrix} -1 & 2 & 0 \\ 1 & -1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$. The last row of E is the combination of the rows of A that produce the zero row (evident since the last row of R is the zero row and $EA = R$). Hence, the transpose of the last row of E is a special solution to $A^\top y = \mathit{0}$.

- A basis for $\mathrm{N}(A^\top)$ is the transposes of the last $m - r$ rows of E.

Subspace	Notation	Dimension	Basis	Superspace
Column space	$\mathrm{C}(A)$	r	Pivot columns of A	$\mathbb{R}^m$
Nullspace	$\mathrm{N}(A)$	$n - r$	Special solutions to $Ax = \mathit{0}$	$\mathbb{R}^n$
Row space	$\mathrm{C}(A^\top)$	r	First r rows of R	$\mathbb{R}^n$

Subspace	Notation	Dimension	Basis	Superspace
Left nullspace	$\mathrm{N}(A^\mathsf{T})$	$m-r$	Last $m-r$ rows of E	$\mathbb{R}^m$

Matrix spaces. All 3×3 matrices form a "vector" space with the matrices as "vectors". They thus form a *matrix* space.

Example 1. Let M denote the matrix space of all 3×3 matrices. Then, the space S of all 3×3 upper-triangular matrices, and the space T of all 3×3 symmetric matrices are subspaces of M. Then, $S \cap T$ is also a subspace of M, which contains all 3×3 diagonal matrices. Let $D = S \cap T$. Then, a basis for D is $\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix}$, $\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix}$, $\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}$. Hence, the dimension of D is 3.

11. Matrix Spaces; Rank 1; Small World Graphs

Outline.

- Bases of new vector spaces
- Rank one matrices

Matrix spaces. Let M be the matrix space of all 3×3 matrices.

- A basis for M is $\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix}$, $\begin{bmatrix}0&1&0\\0&0&0\\0&0&0\end{bmatrix}$, $\begin{bmatrix}0&0&1\\0&0&0\\0&0&0\end{bmatrix}$,..., $\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}$.
- $\dim(M) = 9$. M is practically the same as 9-D space.
- The space of all 3×3 symmetric matrices S is a subspace of M; $\dim(S) = 6$.
- The space of all 3×3 upper-triangular matrices T is a subspace of M; $\dim(T) = 6$.
- The space of all 3×3 diagonal matrices $D = S \cap T$ is a subspace of M; $\dim(D) = 3$.
- $S \cup T$ is *not* a subspace.
- $S + T = \{\text{any element of } S + \text{any element of } T\} = M \Rightarrow \dim(S+T) = 9$.

⋆ For any two subspaces, S and T, $\dim(S) + \dim(T) = \dim(S \cap T) + \dim(S+T)$.

Example 1. Consider the ODE $\frac{\mathrm{d}^2y}{\mathrm{d}x^2} + y = 0$. Then $y = \cos x$ and $y = \sin x$ are solutions. Any of their combinations are also solutions (eg:- e^{ix}, allowing complex coefficients) since we're looking at the nullspace of the ODE, which is the solution space. Hence, the complete solution is $y = \mathsf{c}_1\cos x + \mathsf{c}_2\sin x$. The set of solutions to the ODE form a "vector" space whose basis is $\cos x$, $\sin x$. The dimension of the solution space is 2, which it has to be because we are given a 2^{nd} order ODE. "There is 18.03, and 5 minutes of 18.06 is enough to take care of 18.03."

Rank 1 matrices. Matrices with rank = 1.

Example 1. Consider a 2×3 matrix A with $r = 1$. Suppose its first row is $\begin{bmatrix}1&4&5\end{bmatrix}$. Then, the 2^{nd} row must be a multiple of $\begin{bmatrix}1&4&5\end{bmatrix}$, say $\begin{bmatrix}2&8&10\end{bmatrix}$. A basis for the row space of A is $\begin{bmatrix}1&4&5\end{bmatrix}$. The dimension of the column space $= r = 1$. A basis for the column space of A is the first column $\begin{bmatrix}1\\2\end{bmatrix}$, and the other columns are multiples of that.

⋆ Rank 1 matrices can be written as 1 column times 1 row, where the column is the basis for the column space and the row is the basis for the row space. If A is a rank 1 matrix, then $A = uv^\mathsf{T}$, where u and v are some column vectors.

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

⋆ Any matrix with rank r can be written as a combination of r rank 1 matrices.

Example 2. Given a 5×17 matrix with rank $= 4$. It can be written as a combination of 4 rank 1 matrices.

⋆ $\text{rank}(A+B) \le \text{rank}(A) + \text{rank}(B)$.

Question. Is the space of all matrices with rank ≤ 4 a subspace?
Answer. No.

Example 3. In $\mathbb{R}^4$, let $S = \{v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} \mid v_1 + v_2 + v_3 + v_4 = 0\}$. Then, S is a subspace. It is the nullspace of the 1×4 matrix $A = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$. The rank of A is $1 \Rightarrow \dim(\text{N}(A)) = n - r = 4 - 1 = \dim(S) = 3$. A basis for S is the special solutions to $Av = 0$, which are $\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$. The column space of A is a subspace of $\mathbb{R}$. It is all multiples of 1, which is $\mathbb{R}$. The dimension of the left nullspace of A is $m - r = 1 - 1 = 0$. Hence, $\text{N}(A^{\mathsf{T}}) = \{\mathit{0}\}$.

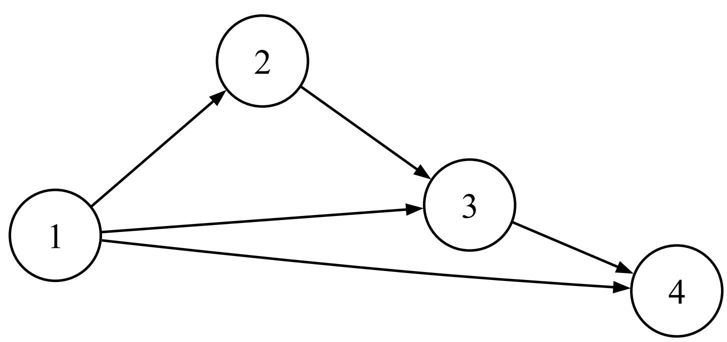
12. Graphs, Networks, Incidence Matrices

"The equation $A^{\mathsf{T}} y = \mathit{0}$ is probably the most fundamental equation of applied mathematics."

Outline.

- Graphs and networks
- Incidence matrices
- Kirchoff's laws

Example 1. Let $G = \{V = \{1,2,3,4\}, E = \{(1,2),(2,3),(1,3),(1,4),(3,4)\}\}$ be a directed graph. Then G has $n = 4$ nodes and $m = 5$ edges. This graph can be interpreted in many ways but consider it to be an electrical network with current flowing through the edges in the direction of the arrow.



`Incidence matrix:` $A = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$.

- The matrix is of size $m \times n = 5 \times 4$: 1 row for each edge, and 1 column for each node.

- The first row says that edge 1 goes out from node 1 and into node 2. Similarly the second row says that edge 2 is from node 2 to node 3, and so on.
- Edges that form a loop in the graph correspond to linearly dependent rows (eg:- rows 1, 2, 3).
- The number of non-zeros is $2m$.

$$Ax = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_2 - x_1 \\ x_3 - x_2 \\ x_3 - x_1 \\ x_4 - x_1 \\ x_4 - x_3 \end{bmatrix}.$$

Hence, if x_1, x_2, x_3, x_4 are *potentials at the nodes*, then Ax computes the *potential differences* across the edges. When all those differences are 0, we are looking at the nullspace of A.

- If the potentials are *constant*; that is, if $x_1 = x_2 = x_3 = x_4$, then $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$ is in $\mathrm{N}(A)$.
- The nullspace is actually 1-D (it's a line in 4-D space), and a basis for it is $\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$. The whole nullspace is $\mathsf{c}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ $\Rightarrow$ The columns are dependent.
- Grounding a node and setting its potential to 0 is a way to get that column out of the matrix and to be left with independent columns. This will also tell us the other potentials, just like setting the free variables tells us the pivot variables.
- $\operatorname{rank}(A) = 3$ and any 3 columns form a basis and we need to ground the 4^{th} one.

$$A^{\mathsf{T}}_{4\times 5} = \begin{bmatrix} -1 & 0 & -1 & -1 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}. \text{ Consider the equation } A^{\mathsf{T}}y = \begin{bmatrix} -1 & 0 & -1 & -1 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

- $\dim(\mathrm{N}(A^{\mathsf{T}})) = m - r = 5 - 3 = 2$.
- y_1, y_2, y_3, y_4, and y_5 are the *currents* on the edges.
- The equation $A^{\mathsf{T}}y = \mathit{0}$ represents Kirchoff's current law (KCL).
- Elimination on A^{T} will give a last row of 0's, since rank $= 3$.
- Any current in a loop satisfies KCL, which is what we have to use to get a basis for $\mathrm{N}(A^{\mathsf{T}})$. Suppose the first basis vector sends current across the loop $1 \to 2 \to 3$, and the second one sends current across the loop $1 \to 3 \to 4$. One such basis is $\begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}$.

$\star$ All rows of the incidence matrix of a tree are independent since trees are acyclic graphs.

$\star$ For a graph with n nodes and c connected components, its incidence matrix has rank $n - c$. In particular, a connected graph has rank $n - 1$: once the potential at one node is set, the potentials at other nodes can be found $\Rightarrow$ There is only one free variable.

$\star$ **Euler's formula (connected graph).** In a connected graph, #loops $= \dim(\mathrm{N}(A^{\mathsf{T}}))$, #edges $= m$, and #nodes $= n = r + 1$. But,
$\dim(\mathrm{N}(A^{\mathsf{T}})) = m - r \Rightarrow$ #nodes $-$ #edges $+$ #loops $= 1$.
In a graph, nodes are 0-D, edges are 1-D, and loops are 2-D, and their counts are related by Euler's formula.

The great picture of applied math (in equilibrium):-

$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix}$: potentials at nodes $\xrightarrow{e=Ax}$ $x_2 - x_1$, etc.: potential differences $\xrightarrow[\text{Ohm's law}]{y=Ce}$ $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \end{bmatrix}$: currents on edges $\xrightarrow{A^\mathsf{T}}$ $A^\mathsf{T} y = f$: Kirchoff's current law (KCL)

- C is the matrix of resistances (Ohm's law says it must be diagonal).
- f is external current; i.e. current from external sources.

Putting the three equations together, $A^\mathsf{T} y = A^\mathsf{T} Ce = \boxed{A^\mathsf{T} CAx = f}$ ← basic equation of applied math.

- The matrix $A^\mathsf{T} CA$, like $A^\mathsf{T} A$, is symmetric.

13. Quiz 1 Review

"I don't know why the author put these in here."

Outline.

- Review for part 1: $Ax = b$

Question. Suppose u, v, w are non-zero vectors in $\mathbb{R}^7$. What are the possible dimensions of the subspace spanned by u, v, w?
Answer. 1, 2, or 3.

Example 1. Let U be a 5×3 matrix with $r = 3$.

- $\mathrm{N}(U) = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right\}$.
- Consider the 10×3 matrix $B = \begin{bmatrix} U \\ 2U \end{bmatrix}$. It's echelon form would be $\begin{bmatrix} U \\ 0 \end{bmatrix}$. It's rank is 3.
- Consider the 10×6 matrix $C = \begin{bmatrix} U & U \\ U & O \end{bmatrix}$. It's echelon form would be $\begin{bmatrix} U & U \\ O & -U \end{bmatrix}$. It's reduced row echelon form would be $\begin{bmatrix} U & O \\ O & U \end{bmatrix}$. If U itself is in reduced row echelon form, it would have some zero rows which should propagate to the bottom in $\mathrm{rref}(C)$. $\mathrm{rank}(C) = 6$.
- $\dim(\mathrm{N}(C^\mathsf{T})) = m - r = 10 - 6 = 4$.

Example 2. Given the complete solution $x = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} + \mathsf{c} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + \mathsf{d} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ to the equation $Ax = \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix}$.

- Size of the matrix is 3×3.
- $\dim(\mathrm{N}(A)) = 2 \Rightarrow \mathrm{rank}(A) = 1$.
- $A = \begin{bmatrix} 1 & -1 & 0 \\ 2 & -2 & 0 \\ 1 & -1 & 0 \end{bmatrix}$.
- $Ax = b$ can be solved if b has the form $b = \mathsf{c} \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$.

⋆ If the nullspace of a square matrix is $\{0\}$, then the nullspace of its transpose is also $\{0\}$.

⋆ If a matrix can be multiplied by itself, it must be square.

Question. T/F: If $B^2 = O$, then $B = O$.
Answer. False. $B = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \neq O$, but $B^2 = O$.

Example 3. $B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

- Size of B is 3×4.
- $\mathrm{N}(B)$ is a subspace of $\mathbb{R}^4$. A basis for $\mathrm{N}(B)$ is $\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$.

$\star$ $\mathrm{N}(CD) = \mathrm{N}(D)$ if C is invertible.

- The complete solution to $Bx = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ is $x = x_p + x_n = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \mathsf{c} \begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix} + \mathsf{d} \begin{bmatrix} -2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$.

$\star$ *Any* solution could be the particular solution.

Question. T/F: If A and B have the same four subspaces, then $A = \mathsf{c}B$.
Answer. False. Counterexample: A and B could be any invertible 6×6 matrices.

Question. Why can't the vector $v = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ be in the nullspace *and* be a row of a matrix A?

Answer. If $\begin{bmatrix} 1 & 2 & 3 \\ * & * & * \\ * & * & * \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, then the first entry on the RHS would not be 0, it would be 14.

$\star$ The intersection of the nullspace and row space is only the zero vector.

14. Orthogonal Vectors and Subspaces

"This is a 90° chapter."

Outline.

- Orthogonal vectors and subspaces
- Nullspace $\perp$ row space
- $\mathrm{N}(A^\mathsf{T} A) = \mathrm{N}(A)$

Orthogonal vectors. Two (column) vectors x and y are *orthogonal* if $x^\mathsf{T} y = 0$. This comes from the Pythagorean test for orthogonality:-
Vectors x and y are orthogonal if $\|x\|^2 + \|y\|^2 = \|x + y\|^2$

$$\begin{aligned}
&\Rightarrow x^\mathsf{T} x + y^\mathsf{T} y = (x+y)^\mathsf{T}(x+y) \\
&\Rightarrow x^\mathsf{T} x + y^\mathsf{T} y = x^\mathsf{T} x + x^\mathsf{T} y + y^\mathsf{T} x + y^\mathsf{T} y \\
&\Rightarrow x^\mathsf{T} y + y^\mathsf{T} x = 0 \\
&\Rightarrow 2x^\mathsf{T} y = 0 \\
&\Rightarrow x^\mathsf{T} y = 0
\end{aligned}$$

$\star$ The zero vector is orthogonal to everybody.

Orthogonal subspaces. Subspace S is orthogonal to subspace T means: *every* vector in S is orthogonal to *every* vector in T.

$\star$ If two subspaces intersect at any non-zero vector, then they are *not* orthogonal, since that non-zero vector is surely not orthogonal to itself.

$\star$ Row space is orthogonal to the nullspace. Their dimensions add up to n when both are in $\mathbb{R}^n$. Hence, they are *orthogonal complements* in $\mathbb{R}^n$.

Key matrix. $A^\top A$.

- Symmetric.
- $\mathrm{N}(A^\top A) = \mathrm{N}(A)$.
- Rank of $A^\top A$ = rank of A.
- Invertible exactly if A has independent columns.

15. Projections Onto Subspaces

> "Gosh, 8 A's in a row is, like, obscene."

Outline.

- Projections!
- Least squares
- Projection Matrix

Projection. Suppose we have two vectors a and b and p is the *projection* of b onto a. Let the error $e = b - p$. Since p is along a, it can be written as a scalar x times $a \Rightarrow p = \mathsf{x}a$. Also e and a are orthogonal

$\Rightarrow a^\top e = a^\top(b - \mathsf{x}a) = 0 \Rightarrow \mathsf{x}a^\top a = a^\top b$

$\therefore \mathsf{x} = \frac{a^\top b}{a^\top a} \Rightarrow p = a\frac{a^\top b}{a^\top a}$.

"There is a matrix here." There is a matrix P that multiplies b to output the projection $p = Pb \Rightarrow P = \frac{aa^\top}{a^\top a}$.

- The column space of P is the line through a since multiplication by P always "lands" on that line in the form of a projection.
- $\mathrm{rank}(P) = 1$.
- P is symmetric ($\because P^\top = P$).
- If I project twice, I get the same answer as the first projection $\Rightarrow P^2 = P$.

Why project? Because $Ax = b$ may have no solution. Solve $A\hat{x} = p$ instead, where p is the projection of b onto the column space of A.

- When the columns of A are independent, $P = A(A^\top A)^{-1}A^\top$. We can't write $(A^\top A)^{-1} = A^{-1}(A^\top)^{-1}$ because A is not necessarily a square matrix. If A were square with independent columns, then it would be invertible and $P = I$.
- $\hat{x} = (A^\top A)^{-1}A^\top b,\ p = A\hat{x} = A(A^\top A)^{-1}A^\top b = Pb$.
- $P^\top = P = P^2$.
- The error $e = b - p$ is orthogonal to the column space of A and hence lies in the nullspace of $A^\top$.

$\star$ When $Ax = b$ has no solutions, the magic equation to solve is $\boxed{A^\top A\hat{x} = A^\top b}$.

> "I will venture to call that the most important equation in statistics, and in estimation, and wherever you've got error and noise, this is the estimate that you use first. Whenever you're fitting things by a few parameters, that's the equation to use."

16. Projection Matrices and Least Squares

"I promise not to put a single letter more on that picture."

Outline.

- Projections
- Least squares and best straight line

Projection matrix. $P = A(A^\intercal A)^{-1}A^\intercal$.

- If $b \in \mathrm{C}(P)$ then $Pb = b$.
- If $b \perp \mathrm{C}(P)$ then $Pb = \mathit{0}$.

$\star$ If P is a projection, then $I - P$ is a projection onto the perpendicular space.

Example: best straight line. Suppose we have three points $(1,1)$, $(2,2)$, and $(3,2)$, we, want to find the best straight line $y = C + Dt$ that *fits* the three points. Our three equations are:-

1. $C + D = 1$
2. $C + 2D = 2$
3. $C + 3D = 2$

We can write it in the form $Ax = b$ as $\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. This has no solutions. Instead, we are looking to minimize some kind of error which we choose to be the squared error $\|e\|^2 = \|Ax - b\|^2$. Find $\hat{x} = \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix}$ and p such that $A\hat{x} = p$ and $\|e\|^2$ is minimized:

$A^\intercal A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 \\ 6 & 14 \end{bmatrix} \leftarrow$ symmetric and invertible

$A^\intercal b = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$

The equations are:

1. $3\hat{C} + 6\hat{D} = 5$
2. $6\hat{C} + 14\hat{D} = 11$

These are called the **normal equations**. These can be obtained from calculus too:

$\|e\|^2 = (C + D - 1)^2 + (C + 2D - 2)^2 + (C + 3D - 2)^2$

Taking partial derivatives of $\|e\|^2$ with respect to C and D and equating them to 0 will give the normal equations.

Solving the equations gives: $\hat{C} = \frac{2}{3},\ \hat{D} = \frac{1}{2} \Rightarrow \hat{x} = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{2} \end{bmatrix}$.

- The best line is $y = \frac{2}{3} + \frac{1}{2}t$.
- The three points on the line are:
  1. $p_1 = \frac{2}{3} + \frac{1}{2}(1) = \frac{7}{6}$.
  2. $p_2 = \frac{2}{3} + \frac{1}{2}(2) = \frac{5}{3}$.
  3. $p_3 = \frac{2}{3} + \frac{1}{2}(3) = \frac{13}{6}$.

  The projected vector is $p = A\hat{x} = \frac{1}{6}\begin{bmatrix} 7 \\ 10 \\ 13 \end{bmatrix}$.

- The errors are:
  1. $e_1 = b_1 - p_1 = 1 - \frac{7}{6} = -\frac{1}{6}$.
  2. $e_2 = b_2 - p_2 = 2 - \frac{5}{3} = \frac{1}{3}$.
  3. $e_3 = b_3 - p_3 = 2 - \frac{13}{6} = -\frac{1}{6}$.

  The error vector is $e = \frac{1}{6}\begin{bmatrix}-1\\2\\-1\end{bmatrix}$.
- Verify $b = p + e$: $\frac{1}{6}\begin{bmatrix}7\\10\\13\end{bmatrix} + \frac{1}{6}\begin{bmatrix}-1\\2\\-1\end{bmatrix} = \begin{bmatrix}1\\2\\2\end{bmatrix}$.
- $p \perp e$. Moreover, p is in the column space of $A = \begin{bmatrix}1&1\\1&2\\1&3\end{bmatrix}$ and is perpendicular to the nullspace of $A^\mathsf{T} = \begin{bmatrix}1&1&1\\1&2&3\end{bmatrix}$, while e is in that nullspace and is perpendicular to the column space of A.

⋆ If A has independent columns then $A^\mathsf{T}A$ is invertible.

Proof sketch. Suppose $A^\mathsf{T}Ax = \mathit{0}$, we want to show that $x = \mathit{0}$.

- **IDEA.** ("I shouldn't have written TRICK, that makes it sound like just a dumb idea. 'Brilliant idea!' I should've put. Okay.") Multiply both sides by x^T: $x^\mathsf{T}A^\mathsf{T}Ax = \mathit{0} \Rightarrow (Ax)^\mathsf{T}Ax = \mathit{0} \Rightarrow Ax = \mathit{0}$. But A has independent columns $\Rightarrow x = \mathit{0}$.

⋆ Perpendicular unit (orthonormal) vectors are definitely independent. "Those are the best columns you could ask for."

Example. $\begin{bmatrix}\cos\theta\\\sin\theta\end{bmatrix}$ and $\begin{bmatrix}-\sin\theta\\\cos\theta\end{bmatrix}$.

17. Orthogonal Matrices and Gram-Schmidt

> "There's only one idea here. If Gram had the idea, I don't know what Schmidt did. ... Maybe that's what Schmidt did, he—brilliant Schmidt—thought divide by the length."

Outline.

- Orthogonal basis $q_1, \ldots, q_n$
- Orthogonal matrix Q
- Gram-Schmidt $A \to Q$

Orthonormality. Vectors $q_1, \ldots, q_n$ are *orthonormal* if ${q_i}^\mathsf{T} q_j = \delta_{ij}$.

Let's put them as columns of a matrix $Q = \begin{bmatrix}q_1 & \ldots & q_n\end{bmatrix}$. Then $Q^\mathsf{T}Q = \begin{bmatrix}{q_1}^\mathsf{T}\\\vdots\\{q_n}^\mathsf{T}\end{bmatrix}\begin{bmatrix}q_1 & \ldots & q_n\end{bmatrix} = I_n$.

⋆ $A^\mathsf{T}A$ asks for all the row times column dot products.

If Q is square then $Q^\mathsf{T}Q = I \Rightarrow Q^{-1} = Q^\mathsf{T}$.

A **Hadamard matrix** is a square matrix of 1's and -1's with mutually orthogonal rows; after normalization it is an orthogonal matrix. E.g.: $Q_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}1&1\\1&-1\end{bmatrix}$

$$Q_4 = \frac{1}{\sqrt{2}}\begin{bmatrix} Q_2 & Q_2 \\ Q_2 & -Q_2 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

It is an open problem as to which sizes n allow Hadamard matrices Q_n.

Suppose we want to project onto the column space of an orthonormal matrix Q.

- The projection matrix is $P = Q(Q^\intercal Q)^{-1}Q^\intercal = QIQ^\intercal = QQ^\intercal$.
- The normal equation is $Q^\intercal Q\hat{x} = Q^\intercal b \Rightarrow \hat{x} = Q^\intercal b \Rightarrow \hat{x}_i = {q_i}^\intercal b$.

"Here comes Gram-Schmidt…"

Gram-Schmidt. "With elimination, our goal was 'make the matrix triangular,' here our goal is, 'make the columns orthonormal.' " Suppose we have two independent vectors a and b, our first step is to get orthogonal vectors v_a and v_b, and then to get orthonormal vectors $q_a = \frac{v_a}{\|v_a\|}$ and $q_b = \frac{v_b}{\|v_b\|}$.

1. Simply set $v_a = a$. "No complaint about that."
2. We're looking for a vector orthogonal to $v_a = a$ and our starting point is b. This is nothing but the error e when we project b onto $a \Rightarrow v_b = b - p = b - Pb = b - \frac{{v_a}^\intercal b}{{v_a}^\intercal v_a} v_a$.

Suppose now we add a third initial vector c, and we're looking for v_c that's perpendicular to v_a *and* v_b that ultimately gives $q_c = \frac{v_c}{\|v_c\|}$.

3. **IDEA.** Subtract the components of c along the v_a and v_b directions from c, and the resulting vector will be orthogonal to both.

   $v_c = c - \frac{{v_a}^\intercal c}{{v_a}^\intercal v_a} v_a - \frac{{v_b}^\intercal c}{{v_b}^\intercal v_b} v_b$.

Example. $a = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$.

- $v_a = a = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.
- $v_b = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} - \frac{3}{3}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$.
- $Q = \begin{bmatrix} q_a & q_b \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & 0 \\ 1/\sqrt{3} & -1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{2} \end{bmatrix}$.
- The column space of $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 2 \end{bmatrix}$ and C(Q) are "one and the same" since we are taking combinations of the original columns.

Analogous to how elimination can be summarized as $A = LU$ using *elimination matrices*, Gram-Schmidt can be summarized as $A = QR$.

$\begin{bmatrix} a & b \end{bmatrix} = \begin{bmatrix} q_a & q_b \end{bmatrix}\begin{bmatrix} a^\intercal q_a & b^\intercal q_a \\ a^\intercal q_b & b^\intercal q_b \end{bmatrix}$, where $a^\intercal q_b = 0$.

$\star$ R is upper-triangular.

19. Properties of Determinants

"How do I get it onto tape that I'm not saying that."

Outline

- Determinants $\det A$
- Properties 1, 2, 3, 4 - 10
- $\pm$ signs

Determinant. A *determinant* is a number associated with every square matrix. Properties:

1. $\star \det I = 1$.
2. $\star$ Exchange rows $\to$ reverse sign of the determinant.

   `Corollary.` For a permutation matrix P, $\det P = \begin{cases} +1 & \text{if even row exchanges} \\ -1 & \text{if odd row exchanges} \end{cases}$.

3. $\star$ Key property about linear combinations of rows:

   a. $\begin{vmatrix} \vdots \\ \mathtt{t}r_i \\ \vdots \end{vmatrix} = \mathtt{t} \begin{vmatrix} \vdots \\ r_i \\ \vdots \end{vmatrix}$.

   `Corollary.` $\det \mathtt{k}A_{n\times n} = \mathtt{k}^n \det A$.

   b. $\begin{vmatrix} \vdots \\ u+v \\ \vdots \end{vmatrix} = \begin{vmatrix} \vdots \\ u \\ \vdots \end{vmatrix} + \begin{vmatrix} \vdots \\ v \\ \vdots \end{vmatrix}$.

4. Two equal rows $\to$ determinant $= 0$.
5. Subtract $\mathtt{t}r_i$ from r_j $\to$ determinant doesn't change.

$$\begin{vmatrix} \vdots \\ r_i \\ \vdots \\ r_j \\ \vdots \end{vmatrix} = \begin{vmatrix} \vdots \\ r_i \\ \vdots \\ r_j - \mathtt{t}r_i \\ \vdots \end{vmatrix}$$

   `Corollary.` Row-replacement steps in elimination don't change the determinant. Row exchanges reverse its sign, and row scalings scale it; thus ordinary row reduction need not preserve the determinant.

6. Row of zeros $\to$ determinant $= 0$.
7. The determinant of a triangular matrix is the product of its diagonal entries. For pivots produced from A by elimination, their product equals $\det A$ if there were no row exchanges, and differs by the sign of the row permutation otherwise.

   `Corollary.` The determinant of a diagonal matrix is the product of the diagonal entries.

8. $\det A = 0 \Leftrightarrow A$ is singular.

   `Proof sketch.` Elimination on a singular A will give a row of 0's in U and $\det A = \det U$ by (5). By (6), $\det U = 0$. $\therefore \det A = 0$.

9. $\det AB = (\det A)(\det B)$.

   `Corollary.` $\det A^{-1} = \frac{1}{\det A}$.

10. $\det A^{\mathsf{T}} = \det A$.

    `Corollary.` Everything true about row operations holds true for column operations as well w.r.t. determinants.

    `Proof sketch.` $|A| = |LU| = |L||U| = |U^{\mathsf{T}}||L^{\mathsf{T}}| = |U^{\mathsf{T}}L^{\mathsf{T}}| = |A^{\mathsf{T}}|$.

$\star\star\star$ The determinant is well-defined by properties (1), (2), and (3).

20. Determinant Formulae and Cofactors

"I was a little scared actually as that apparition went by."

Outline.

- Formula for $\det A$ ($n!$ terms)
- Cofactor formula
- Tridiagonal matrices

$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$ can be split up into 27 determinants by property (3), out of which many would be 0 (because of a row/column of zeros). The survivors would be the ones having an entry in each row and each column. Let's write the survivors:

$$\begin{vmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ 0 & 0 & a_{23} \\ a_{31} & 0 & 0 \end{vmatrix} + \begin{vmatrix} 0 & 0 & a_{13} \\ 0 & a_{22} & 0 \\ a_{31} & 0 & 0 \end{vmatrix} + \begin{vmatrix} a_{11} & 0 & 0 \\ 0 & 0 & a_{23} \\ 0 & a_{32} & 0 \end{vmatrix} + \begin{vmatrix} 0 & a_{12} & 0 \\ a_{21} & 0 & 0 \\ 0 & 0 & a_{33} \end{vmatrix}.$$

The nonzero elements can be factored out from each survivor to turn it into a permutation matrix with determinant ± 1.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}.$$

The Big Formula:

$$\boxed{\det A = \sum_{n!\text{ terms}} \pm a_{1\alpha}a_{2\beta}a_{3\gamma}\ldots a_{n\omega}, \text{ where } (\alpha, \beta, \gamma, \ldots, \omega) \text{ is some permutation of } (1, 2, 3, \ldots, n).}$$

Cofactors. Consider the 3×3 determinant we have computed:
$\det A = a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{12}(-a_{21}a_{33} + a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31})$,
the terms in the parens are the *cofactors*. The cofactor of any a_{ij} is the coefficient of a_{ij} in the big formula.

$\star$ Cofactor of $a_{ij} = C_{ij} = (-1)^{i+j} \det(n-1$ matrix with row i, col j erased).

The cofactor without the built-in sign is called the **minor** of a_{ij} "but I don't care about those."

Cofactor formula (along row i):

$$\boxed{\det A = \sum_{j=1}^{n} a_{ij}C_{ij}.}$$

Tridiagonal example. $A_4 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$ is a tridiagonal matrix of size 4. We know $|A_1| = 1$, $|A_2| = 0$, and $|A_3| = -1$. Then, $|A_4| = 1 \cdot |A_3| + 1 \cdot C_{12}$ (by CF along row 1). But $C_{12} = -\begin{vmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{vmatrix} = -1 \cdot |A_2|$ (by CF along column 1) $\Rightarrow |A_4| = |A_3| - |A_2| = -1$.

$\star$ In general, $|A_n| = |A_{n-1}| - |A_{n-2}|$.

The sequence goes like $1, 0, -1, -1, 0, 1, 1, \ldots$. These determinants have period 6. "It's pretty fantastic."

21. Cramer's Rule, Inverse Matrix, and Volume

"Let me call that matrix A_S: A screwed up."

Outline.

- Formula for A^{-1}
- Cramer's rule for $x = A^{-1}b$
- $|A| =$ volume of box

Cofactor matrix. The *cofactor matrix* C of a matrix A is a matrix such that $c_{ij} =$ cofactor of a_{ij} in A.

The Big Inverse Formula. $\boxed{A^{-1} = \frac{1}{\det A} C^{\mathsf{T}}.}$

- `Proof sketch.` Check $AC^{\mathsf{T}} = (\det A)I$. This is true by CF and the fact that $\sum_{j=1}^{n} a_{kj}C_{ij} = 0$ for $k \neq i$ (since it is the determinant of the matrix with the same k^{th} and i^{th} rows).

- The determinant $\det A_{n \times n}$ involves products of n entries and elements of the cofactor matrix (the cofactors) involve products of $n-1$ entries.

Cramer's rule. $Ax = b \Rightarrow x = A^{-1}b = \frac{1}{\det A} C^{\mathsf{T}} b$. This says: $x_1 = \frac{\det B_1}{\det A}$, $x_2 = \frac{\det B_2}{\det A}, \ldots, x_n = \frac{\det B_n}{\det A}$, where B_i is the matrix obtained by replacing column i in A by b.

- This requires computing $n+1$ determinants, which is disastrous because it takes "approximately forever."

Volume application. Consider a box in 3D space whose defining edges are given by the rows of $A_{3\times 3}$: (a_{11}, a_{12}, a_{13}), (a_{21}, a_{22}, a_{23}), (a_{31}, a_{32}, a_{33}). The box is completed by drawing parallelograms for each face. "The proper word, of course, is parallelopiped, but for obvious reasons, I wrote 'box.' "

$\boxed{|\det A| = \text{volume of box.}}$

- `Proof sketch.` Check properties (1), (2), and (3) w.r.t. the volume. For instance, if the defining matrix $A = I$, then the box is nothing but the unit cube, and $V(\text{unit cube}) = 1$. If a side of the box is doubled, the volume also doubles.

- The sign of the determinant tells me whether the box is left-handed or right-handed.

22. Eigenvalues and Eigenvectors

"Complex numbers are gonna have to enter 18.06 at this moment. Boo. ... If I just choose good matrices that have real eigenvalues we can postpone that evil day."

Outline.

- Eigenvalues - Eigenvectors
- $\det[A - \lambda I] = 0$
- Trace $= \lambda_1 + \lambda_2 + \ldots + \lambda_n$

Eigenvalues and Eigenvectors. Think of Ax as a function or as an operator (!). We send in a vector x and out comes Ax. The vectors which come out in the same direction that they went in are the *eigenvectors* $\rightarrow$

Ax parallel to x. This can be written as: $\boxed{Ax = \lambda x}$. The scalar multiples λ corresponding to the eigenvectors are the *eigenvalues*.

$\star$ Only square matrices have eigenvalues and eigenvectors since otherwise Ax won't have the same number of components as x.

- If $\lambda = 0$, $Ax = \mathit{0}$. In this case, the eigenvectors are the "guys in the nullspace."

$\star$ A is singular if and only if $\lambda = 0$ is an eigenvalue of A.

- `Question.` What are the x's and λ's for a projection matrix P?
  `Answer.` For a vector x to be an eigenvector of P, Px must be in the same direction as x; i.e. the projection of x onto a space must be in the same direction as x. Hence, all the vectors in the projection space are eigenvectors. For all these x's, $Px = x$ because they are projected onto themselves $\Rightarrow \lambda = 1$. But projection matrices are singular because any vector perpendicular to the projection space is projected to $\mathit{0}$. Hence, another eigenvalue is 0, and the corresponding x's are all vectors perpendicular to the projection space.

Eigenvector	Equation	Eigenvalue
Any x in the projection space	$Px = x$	$\lambda = 1$
Any $x \perp$ the projection space	$Px = \mathit{0}$	$\lambda = 0$

- `Question.` What are the x's and λ's for the permutation matrix $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$?
  `Answer.` The question asks which are the vectors $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ such that $\begin{bmatrix} x_2 \\ x_1 \end{bmatrix}$ is in the same direction as x. One possibility is $x_2 = x_1$ such as $x = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, in which case, $Ax = x \Rightarrow \lambda = 1$. Another possibility is $x_2 = -x_1$ such as $x = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, in which case, $Ax = -x \Rightarrow \lambda = -1$.

$\star$ Sum of the eigenvalues of a matrix is equal to its trace.

How to solve $Ax = \lambda x$?

1. Rewrite: $(A - \lambda I)x = \mathit{0}$. Also, $x \neq \mathit{0}$ since it is an eigenvector. This implies that the matrix $A - \lambda I$ is singular.
2. Since $A - \lambda I$ is singular, $\det(A - \lambda I) = 0$ This is called the **characteristic** or the **eigenvalue equation**. From this, find λ's.
3. Once λ is found, solve for the corresponding x's by finding the nullspace of $A - \lambda I$.

$\star$ In the 2×2 case, the characteristic equation becomes $\lambda^2 - \mathrm{tr}[A]\lambda + \det A = 0$.

$\star$ Eigenvectors of a symmetric matrix corresponding to different eigenvalues are orthogonal. Symmetric matrices have real eigenvalues.

$\star$ The eigenvalues of $A + nI$ are n bigger than those of A, and the eigenvectors are the same.

Example. $Q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ is the 90° rotation matrix. By definition, Qx is never in the same direction as x. Yet, it must have two eigenvalues with corresponding eigenvectors. Solving the characteristic equation gives $\lambda_1 = \iota$ and $\lambda_2 = -\iota$.

$\star$ Skew-symmetric matrices ($A^\mathsf{T} = -A$) have purely imaginary eigenvalues.

$\star$ The possibility of repeated eigenvalues opens the further possibility of a shortage of eigenvectors, since the eigenvectors for the different eigenvalues should ideally be independent. Such a matrix is **degenerate**.

$\star$ The eigenvalues of a triangular matrix are its diagonal entries. These diagonal entries may be called its pivots, but pivots obtained by ordinary elimination on the original matrix are not generally its eigenvalues, because elimination does not preserve eigenvalues.

23. Diagonalization and Powers of A

"It's the idea that's central."

Outline.

- Diagonalizing a matrix $X^{-1}AX = \Lambda$
- Powers of A
- Equation $u_{k+1} = Au_k$

Diagonalization. Suppose we have n linearly independent eigenvectors of $A_{n\times n}$. Put them in the columns of a matrix X, called the **eigenvector matrix**.

$$AX = A\begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} = \begin{bmatrix} \lambda_1 x_1 & \lambda_2 x_2 & \dots & \lambda_n x_n \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}.$$

Call the diagonal matrix the **eigenvalue matrix** and denote it by Λ. Then we have got $AX = X\Lambda$. Since $x_1, x_2, \dots, x_n$ are independent, X is invertible $\Rightarrow$ $\boxed{X^{-1}AX = \Lambda.}$ $\leftarrow$ diagonalized. We can also write this as $A = X\Lambda X^{-1}$, which is another factorization for a matrix.

$\star$ The diagonalization factorization is $A = X\Lambda X^{-1}$.

The existence of n independent eigenvectors is an assumption here as it is the case for most matrices.

Powers of A. If $Ax = \lambda x$, then $A^2x = \lambda Ax = \lambda^2 x$. The eigenvalues of A^2 are squared but the eigenvectors are the same. Another way:

$A^2 = X\Lambda X^{-1}X\Lambda X^{-1} = X\Lambda^2 X^{-1}$.

This also tells us, but in matrix form, that the eigenvalues of A^2 are squared but the eigenvectors are the same since the eigenvector matrix is the same.

In general, $\boxed{A^k = X\Lambda^k X^{-1}}$. The eigenvalues of A^k are the k^{th} powers but the eigenvectors remain the same.

$\star$ **Theorem.** $A^k \to O$ as $k \to \infty$ if $|\lambda| < 1$ for all eigenvalues λ.

Note that the assumption of n independent eigenvectors is still in play because we are using diagonalization, which is only possible under this assumption.

$\star$ $A_{n\times n}$ is sure to have n independent eigenvectors (and be diagonalizable) if all eigenvalues are distinct.

$\star$ $A_{n\times n}$ with repeated eigenvalues may or may not have n independent eigenvectors.

Difference equations. Start with a given vector u_0 and use the given relation $u_{k+1} = Au_k$ to get the equation $u_k = A^k u_0$. To really solve this, write u_0 in terms of the eigenvectors (assumed independent and normalized) as $u_0 = \sum_{i=1}^n c_i x_i$. The coefficients can be the components of a vector c, and we can write $u_0 = Xc \Rightarrow u_k = A^k u_0 = X\Lambda^k X^{-1}Xc \Rightarrow \boxed{u_k = X\Lambda^k c.}$

Fibonacci example. Start with $F_0 = 0$ and $F_1 = 1$ and $F_{k+2} = F_{k+1} + F_k$ for all $k \geq 0$. The questions are what is F_{100} and how fast are they growing?

`TRICK`. Let $u_k = \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix}$. Then, $u_{k+1} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_{k+1} \\ F_k \end{bmatrix} = Au_k$.

The eigenvalues of A are $\lambda_\pm = \frac{1\pm\sqrt{5}}{2}$. Solving the equations $(A - \lambda_\pm I)x_\pm = \begin{bmatrix} 1-\lambda_\pm & 1 \\ 1 & -\lambda_\pm \end{bmatrix} x_\pm = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

gives the eigenvectors $x_\pm = \begin{bmatrix}\lambda_\pm \\ 1\end{bmatrix} \Rightarrow X = \begin{bmatrix}\lambda_+ & \lambda_- \\ 1 & 1\end{bmatrix} \Rightarrow X^{-1} = \frac{1}{\sqrt{5}}\begin{bmatrix}1 & -\lambda_- \\ -1 & \lambda_+\end{bmatrix}$(by BIF). We know $u_0 = \begin{bmatrix}F_1 \\ F_0\end{bmatrix} = \begin{bmatrix}1 \\ 0\end{bmatrix} = Xc$, or $\begin{bmatrix}\lambda_+ & \lambda_- \\ 1 & 1\end{bmatrix} c = \begin{bmatrix}1 \\ 0\end{bmatrix} \Rightarrow c = \frac{1}{\sqrt{5}}\begin{bmatrix}1 & -\lambda_- \\ -1 & \lambda_+\end{bmatrix}\begin{bmatrix}1 \\ 0\end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix}1 \\ -1\end{bmatrix}$.

$u_{100} = A^{100}u_0 = X\Lambda^{100}c = \frac{1}{\sqrt{5}}\begin{bmatrix}\lambda_+ & \lambda_- \\ 1 & 1\end{bmatrix}\begin{bmatrix}\lambda_+^{100} & 0 \\ 0 & \lambda_-^{100}\end{bmatrix}\begin{bmatrix}1 \\ -1\end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix}\lambda_+ & \lambda_- \\ 1 & 1\end{bmatrix}\begin{bmatrix}\lambda_+^{100} \\ -\lambda_-^{100}\end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix}\lambda_+^{101} - \lambda_-^{101} \\ \lambda_+^{100} - \lambda_-^{100}\end{bmatrix}$,
$\therefore F_{100} = \frac{\lambda_+^{100} - \lambda_-^{100}}{\sqrt{5}}$.

$\star$ When things are evolving in time by a first-order system starting from an original u_0, the eigenvalues of A tell you where the solution is going, and $u_k = X\Lambda^k c$.

24. Differential Equations and e^{At}

"There are two beautiful Taylor's series in this world."

Outline.

- Differential equations $\frac{du}{dt} = Au$
- Exponential e^{At} of a matrix

Differential equations. When A has a complete set of eigenvectors, the solution to the differential equation $\frac{du}{dt} = Au$ is $u(t) = \sum_{i=1}^n c_i e^{\lambda_i t} x_i$, where the λ's and the x's are the eigenvalues and eigenvectors of A.

Example. Consider a system evolving in time according to $\frac{du}{dt} = \begin{bmatrix}-1 & 2 \\ 1 & -2\end{bmatrix} u$ with an initial condition $u(0) = \begin{bmatrix}1 \\ 0\end{bmatrix}$. Solving for the eigenvalues of $A = \begin{bmatrix}-1 & 2 \\ 1 & -2\end{bmatrix}$ gives $\lambda_1 = 0$ and $\lambda_2 = -3$ with corresponding eigenvectors $x_1 = \begin{bmatrix}2 \\ 1\end{bmatrix}$ and $x_2 = \begin{bmatrix}1 \\ -1\end{bmatrix}$. The solution to the differential equation is:

$u(t) = c_1 e^{0t} x_1 + c_2 e^{-3t} x_2 = c_1 \begin{bmatrix}2 \\ 1\end{bmatrix} + c_2 e^{-3t}\begin{bmatrix}1 \\ -1\end{bmatrix}$.

We use the initial condition to find c_1 and c_2. At $t = 0$, $c_1\begin{bmatrix}2 \\ 1\end{bmatrix} + c_2\begin{bmatrix}1 \\ -1\end{bmatrix} = \begin{bmatrix}1 \\ 0\end{bmatrix} \Rightarrow \begin{bmatrix}2 & 1 \\ 1 & -1\end{bmatrix} c = \begin{bmatrix}1 \\ 0\end{bmatrix}$. Solving gives $c_1 = c_2 = \frac{1}{3}$. $\therefore u(t) = \frac{1}{3}\begin{bmatrix}2 \\ 1\end{bmatrix} + \frac{1}{3}e^{-3t}\begin{bmatrix}1 \\ -1\end{bmatrix}$.

- As $t \to \infty$, the second term disappears, and the flow u attains a steady state: $u(\infty) = \frac{1}{3}\begin{bmatrix}2 \\ 1\end{bmatrix}$.
- For stability ($u(t) \to 0$) we need $e^{\lambda t} \to 0$ or $\lambda < 0$ ($\mathrm{Re}(\lambda) < 0$, actually).

  `Comment.` In case of a 2×2, both eigenvalues should be negative which gives the condition that the trace should be negative and the determinant should be positive.
- For attaining a steady state in this diagonalizable setup, one eigenvalue must be 0 and the real parts of all others must be negative.
- A mode blows up if its eigenvalue has positive real part and that mode is present in the initial condition.
- The matrix $\begin{bmatrix}2 & 1 \\ 1 & -1\end{bmatrix}$ we used to find the coefficients is the eigenvector matrix, and the equation is $Xc = u_0$.

The whole point of the eigenvector matrix is to **uncouple** that is diagonalize the representation as opposed to A being full of nonzeros. Uncouple using X: set $u = Xv \Rightarrow X\frac{dv}{dt} = AXv \Rightarrow \frac{dv}{dt} = X^{-1}AXv = \frac{dv}{dt} = \Lambda v$.

Hence, if the eigenvector matrix is used as the basis, there is no coupling anymore: $\frac{dv_1}{dt} = \lambda_1 v_1$, $\frac{dv_2}{dt} = \lambda_2 v_2$, ... "It's a system of equations but they're not connected," making them easier to solve.

Solution: $v(t) = e^{\Lambda t} v(0) \Rightarrow \boxed{u(t) = Xe^{\Lambda t} X^{-1} u(0).}$

Matrix exponential. What's the exponential of a matrix? Expand using Taylor's series:
$e^{At} = I + At + \frac{(At)^2}{2} + \frac{(At)^3}{6} \ldots + \frac{(At)^n}{n!} + \ldots$ (always converges).

`Aside.` Similarly, $(I - At)^{-1} = I + At + (At)^2 + (At)^3 + \ldots + (At)^n + \ldots$. This series converges and gives the inverse when all eigenvalues of At have absolute value less than 1.

- $e^{At} = Xe^{\Lambda t} X^{-1}$ (assuming A can be diagonalized).

  `Proof sketch.` Rewrite I as XX^{-1} and A as $X\Lambda X^{-1}$ in the Taylor's series expansion to get the powers $(At)^k = X\Lambda^k X^{-1} t^k = X(\Lambda t)^k X^{-1}$. Left-factorize X and right-factorize X^{-1}.
- $[e^{\Lambda t}]_{ii} = e^{\lambda_i t}$.

Question. When does e^{At} go to 0?
Answer. When $e^{\Lambda t}$ goes to 0; i.e. when $e^{\lambda_i t}$ goes to 0 i.e. when $\text{Re}(\lambda_i) < 0$.

	Condition	Region in $\mathbb{C}$
Stability in differential equations	$\text{Re}(\lambda_i) < 0$	Left half plane
$A^k \to O$	$\lvert\lambda_i\rvert < 1$	Open unit disk

Example. Consider the equation $y'' + by' + ky = 0$. Change the one second order equation into a 2×2 first order system: $u = \begin{bmatrix} y' \\ y \end{bmatrix} \Rightarrow u' = \begin{bmatrix} y'' \\ y' \end{bmatrix} = \begin{bmatrix} -b & -k \\ 1 & 0 \end{bmatrix} u$.

Similarly, a fifth order equation can be changed to a 5×5 first order system: $u' = \begin{bmatrix} c_1 & c_2 & c_3 & c_4 & c_5 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} u$.

25. Markov Matrices; Fourier Series

Outline.

- Markov matrices
- Steady state
- Fourier series projections

Markov matrices. A matrix must satisfy three properties to be a Markov matrix:

1. It is square.
2. All entries ≥ 0.
3. All columns add to 1.

$\star$ The powers of a Markov matrix are all Markov matrices.

$\star$ **Theorem.** A matrix has an eigenvalue λ if all columns add to λ.
`Proof sketch.` Show that $A - \lambda I$ is singular, which implies A has an eigenvalue λ.

Key points of Markov matrices:

1. $\lambda = 1$ is an eigenvalue.

2. All eigenvalues satisfy $|\lambda_i| \le 1$. Strict inequality for every eigenvalue other than 1 requires additional hypotheses, such as a regular (primitive) Markov matrix; periodic or reducible examples may have other eigenvalues on the unit circle.

3. When 1 is the only eigenvalue on the unit circle and the remaining modes decay, the steady state of the evolution $u_k = A^k_{\text{Markov}} u_0$ retains only the term corresponding to $\lambda_1 = 1$.

4. The eigenvector x_1 corresponding to $\lambda_1 = 1$ is $\succcurlyeq \mathit{0}$ and hence so is the steady state.

Expansion. We want to expand a vector v in terms of an orthonormal basis $q_1, q_2, \ldots, q_n$. In other words, we want to solve for the coefficients $x_1, x_2, \ldots, x_n$ such that $v = \sum_{i=1}^{n} x_i q_i$. Let $Q = \begin{bmatrix} q_1 & q_2 & \ldots & q_n \end{bmatrix}$ and write the equation in matrix form as $Qx = v \Rightarrow x = Q^{-1}v$. But $Q^{-1} = Q^{\mathsf{T}}$. We immediately get the solution as $x = Q^{\mathsf{T}} v = \begin{bmatrix} {q_1}^{\mathsf{T}} \\ {q_2}^{\mathsf{T}} \\ \vdots \\ {q_n}^{\mathsf{T}} \end{bmatrix} v \Rightarrow \boxed{x_i = {q_i}^{\mathsf{T}} v.}$

Fourier series. Consider the function $f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \ldots$. This is a *Fourier series*. We are now in an infinite dimensional function space with the basis $1, \cos x, \sin x, \cos 2x, \sin 2x, \ldots$. An inner product of real-valued functions $g(x)$ and $h(x)$ is defined as $g^{\mathsf{T}} h = \int_{\ell_1}^{\ell_2} g(x)h(x)\,dx$. Since $f(x)$ is periodic with a period 2π, an inner product involving f will have limits of integration as 0 to 2π. Coming back to the basis in our setup, we can show with this definition of the inner product that it is orthogonal. The problem of Fourier series reduces to expressing $f(x)$ in this orthogonal basis:

$$
\begin{aligned}
&\cos^{\mathsf{T}} f = \cos^{\mathsf{T}}(a_0 + a_1 \cos x + \ldots) \\
\Rightarrow &\int_0^{2\pi} f(x) \cos x\,dx = 0 + a_1 \int_0^{2\pi} \cos^2 x\,dx + 0 \\
\Rightarrow &\boxed{a_1 = \frac{1}{\pi} \int_0^{2\pi} f(x) \cos x\,dx}
\end{aligned}
$$

Similarly, the other Fourier coefficients can be computed.

26. Quiz 2 Review

"Nevertheless, linear algebra was still as great as ever. … And, that wasn't meant to be a joke."

Outline.

1. Orthogonality:
   - $Q = \begin{bmatrix} q_1 & \ldots & q_n \end{bmatrix}$, $Q^{\mathsf{T}} Q = I$.
   - Projections—least squares.
   - Gram-Schmidt.
   - Expansion in orthogonal bases, Fourier series.

2. Determinants:
   - $\det A$ properties 1 - 3, 4 - 10.
   - Big formula (BF): $n!$ terms.
   - Cofactor formula (CF).

3. Eigenvalues and eigenvectors
   - $Ax = \lambda x$.
   - $\det(A - \lambda I) = 0$.
   - Diagonalize $X^{-1} A X = \Lambda$.

- Powers A^k, Markov matrices.
- Exponentials e^{At}, differential equations.

Example 1. Given $a = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$.

- The projection matrix $P = \frac{aa^\intercal}{a^\intercal a} = \frac{1}{9}\begin{bmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{bmatrix}$ for projection onto a.
- The column space is the line through $\begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$. The rank of P is 1 $\Rightarrow$ the nullspace is $n - r = 2$-dimensional $\Rightarrow \lambda = 0$ is an eigenvalue of P with multiplicity 2. The third eigenvalue is $\mathrm{tr}[P] = 1$.
- The eigenvector for $\lambda = 1$ is a. The eigenvectors for $\lambda = 0$ are $\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ -2 \end{bmatrix}$, which form a basis for $\mathrm{N}(P)$. $X = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & 1 \\ 2 & -1 & -2 \end{bmatrix}$, $\Lambda = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.
- Set up a difference equation as $u_0 = \begin{bmatrix} 9 \\ 9 \\ 0 \end{bmatrix}$ and $u_{k+1} = Pu_k$. $u_1 = Pu_0 = a\frac{a^\intercal u_0}{a^\intercal a} = \frac{27}{9}a = 3a = \begin{bmatrix} 6 \\ 3 \\ 6 \end{bmatrix}$. This is the projection of u_0 on a. u_2 will be the projection of u_1 on a, which will be the same as $u_1 = \begin{bmatrix} 6 \\ 3 \\ 6 \end{bmatrix}$. All subsequent $u_k = u_1$.

Example 2. Fit a straight line through the points $(1, 4)$, $(2, 5)$, and $(3, 8)$ that goes through the origin. The equation to solve is $y = Dt$, with one parameter D, "one degree of freedom."

- $A = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $b = \begin{bmatrix} 4 \\ 5 \\ 8 \end{bmatrix}$, $\hat{x} = D$. $A^\intercal AD = A^\intercal b \Rightarrow 14D = 38 \Rightarrow D = \frac{19}{7}$.
- We projected the vector $b = \begin{bmatrix} 4 \\ 5 \\ 8 \end{bmatrix}$ onto the column space of A, which is just the line through $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Example 3. Suppose $a_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $a_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. The goal is to find two orthogonal vectors in the plane. By Gram-Schmidt, $v_1 = a_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $v_2 = a_2 - v_1\frac{{v_1}^\intercal a_2}{{v_1}^\intercal v_1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{6}{14}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{7}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Question. Under what condition on the λ's will a matrix be invertible?
Answer. Invertible if and only if none of them are 0.

"We can guess what the examiner has got in his little mind."

27. Symmetric Matrices and Positive Definiteness

"Forgive me for doing such a thing, but I'll look at the book."

Outline.

- Symmetric matrices: eigenvalues, eigenvectors
- Positive definite matrices

S is a symmetric matrix if $S = S^\mathsf{T}$. Suppose S is a <u>real</u> symmetric matrix, then:

1. The eigenvalues are <u>real</u>.
   - `Proof sketch.` $Sx = \lambda x \Rightarrow \bar{S}\bar{x} = \bar{\lambda}\bar{x} \Rightarrow S\bar{x} = \bar{\lambda}\bar{x} \Rightarrow \bar{x}^\mathsf{T}S^\mathsf{T} = \bar{\lambda}\bar{x}^\mathsf{T} \Rightarrow \bar{x}^\mathsf{T}Sx = \bar{\lambda}\bar{x}^\mathsf{T}x$. Also, $Sx = \lambda x \Rightarrow \bar{x}^\mathsf{T}Sx = \lambda\bar{x}^\mathsf{T}x$. Hence, $\lambda\bar{x}^\mathsf{T}x = \bar{\lambda}\bar{x}^\mathsf{T}x$. But $\bar{x}^\mathsf{T}x = \|x\|^2 > 0 \because x \neq 0 \Rightarrow \lambda = \bar{\lambda}$. $\therefore \lambda \in \mathbb{R}$.
2. The eigenvectors are <u>perpendicular</u> (a complete set can be chosen and be normalized).

$\star$ Symmetric matrices are always diagonalizable.

If we say a "good" matrix is one which has real eigenvalues and perpendicular eigenvectors, then for the real case being good is equivalent to being symmetric ($S = S^\mathsf{T}$). However, this fails in the general complex case (as can be seen in the aforementioned proof sketch).

$\star$ If A is complex, A is "good" if and only if $A = \bar{A}^\mathsf{T}$. Such a matrix is called **Hermitian**.

$\star$ **Spectral Theorem.** For the real symmetric case, $S = Q\Lambda Q^{-1} = Q\Lambda Q^\mathsf{T}$, where Q is an orthonormal matrix and Λ is real.

- The **spectrum** of a matrix is the set of all its eigenvalues.

$\star$ **Spectral Theorem (rephrase).** Every symmetric matrix is a combination of mutually perpendicular projection matrices.

- How are the two statements equivalent?
  $S = Q\Lambda Q^\mathsf{T} = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \begin{bmatrix} {q_1}^\mathsf{T} \\ {q_2}^\mathsf{T} \\ \vdots \\ {q_n}^\mathsf{T} \end{bmatrix} = \sum_{i=1}^n \lambda_i q_i {q_i}^\mathsf{T}$. For each q_i, $q_i{q_i}^\mathsf{T}$ is a projection matrix (recall the $\frac{aa^\mathsf{T}}{a^\mathsf{T}a}$ form with ${q_i}^\mathsf{T}q_i = 1$). Also, each $q_i{q_i}^\mathsf{T}$ is a rank 1 matrix $\Rightarrow$ the spectral theorem can also be rephrased as every symmetric $n\times n$ matrix being decomposed into n rank 1 matrices.

$\star$ **Another great fact for real symmetric matrices.** When symmetric elimination proceeds with nonzero pivots and no row exchanges, the signs of the pivots give the same inertia as the eigenvalues. The number of positive pivots is the same as the number of positive eigenvalues.

- Also, the product of the pivots is equal to the product of the eigenvalues, but "nobody is mixing up and thinking that the pivots *are* the eigenvalues".

Positive definite (symmetric) matrices. "If symmetric matrices are good, then positive definite matrices are a subclass that are excellent." They are symmetric matrices with all eigenvalues <u>positive</u>.

- All pivots are positive. Also, the other way round serves as a test that if all pivots are positive then the matrix is positive definite.
- The trace and determinant are positive. If either of them is not positive then the matrix is definitely not positive definite (no pun intended).
- All **leading principal subdeterminants** are positive. These are the n determinants of the leading $i\times i$ principal matrices. The converse is also true: if all leading principal subdeterminants are positive then the symmetric matrix is positive definite.

28. Complex Matrices; Fast Fourier Transform

"That symbol, H, it stands for a guy named Hermite, who didn't actually pronounce the 'H.' "

Outline.

- Complex vectors, matrices, inner products
- Fast Fourier transform (FFT)
- Fourier matrix F_n

Complex vectors. $z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$ is a vector in $\mathbb{C}^n$ where all $z_i \in \mathbb{C}$.

- To compute the length of z, $z^{\mathsf{T}}z$ "is no good" because the length must be real and positive. We really want $\begin{bmatrix} \bar{z_1} & \bar{z_2} & \dots & \bar{z_n} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} = \sum_{i=1}^{n} |z_i|^2 = \|z\|^2 \Rightarrow \bar{z}^{\mathsf{T}} z$ "is good."

- `Notation.` Transpose of the conjugate is written as one operation: z^{H}, called "z Hermitian." Hence the length of a complex vector z is given by $z^{\mathsf{H}}z$.

- The inner product of two complex vectors x and y is $y^{\mathsf{H}}x \Rightarrow x$ and y are perpendicular if $x^{\mathsf{H}}y = 0$.

Hermitian matrices. $S^{\mathsf{T}} = S$ symmetry is no good if S is complex. The "right complex version of symmetry" is $S^{\mathsf{H}} = S$. Such matrices are called **Hermitian** matrices.

- They have real eigenvalues and perpendicular eigenvectors.

Unitary matrices. A complex orthonormal matrix $Q = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix}$ has orthonormal vectors q_i for columns such that ${q_i}^{\mathsf{H}} q_j = \delta_{ij} \Rightarrow Q^{\mathsf{H}}Q = I$. Complex orthogonal matrices are called **unitary**. Hence, a unitary matrix Q is one that satisfies $Q^{-1} = Q^{\mathsf{H}}$.

Fourier matrices. "The most famous complex matrices."

$F_n = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & w & w^2 & \dots & w^{n-1} \\ 1 & w^2 & w^4 & \dots & w^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & w^{n-1} & w^{2(n-1)} & \dots & w^{(n-1)^2} \end{bmatrix}$, where w is the n^{th} root of unity. In other words, $(F_n)_{ij} = w^{ij}$, where $w = e^{\iota \frac{2\pi}{n}}$.

- Fourier matrices are symmetric $\Rightarrow {F_n}^{\mathsf{T}} = F_n$.

- Fourier matrices are orthogonal but not quite orthonormal since length of each column is $\sqrt{n}$. Let $F'_n = \frac{1}{\sqrt{n}} F_n \Rightarrow F'_n$ is unitary $\Rightarrow {F'_n}^{-1} = {F'_n}^{\mathsf{H}}$.

- $F_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \iota & -1 & -\iota \\ 1 & -1 & 1 & -1 \\ 1 & -\iota & -1 & \iota \end{bmatrix}$. This is the 4×4 matrix that comes into the 4-point Fourier transform of a vector. Its inverse computes the inverse transform.

Fast Fourier transform (FFT). A Fourier transform (multiplication of F_n by a vector) costs n^2 multiplications. The merit of the *fast Fourier transform (FFT)* is that it costs on the order of $n \log n$ multiplications. The idea is to factorize F_{2k} in terms of F_k.

$$F_{2k} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & w & \dots & w^{2k-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & w^{2k-1} & \dots & w^{(2k-1)^2} \end{bmatrix}.$$

Permute the columns of F_{2k} such that the even numbered columns $(0, 2, 4, \dots)$ appear first followed by the remaining k odd numbered columns. This operation can be achieved using a permutation matrix $P'_{2k\times 2k}$ such that the first k columns of P are the even numbered columns of the identity and the next k columns are the odd numbered ones. $P' = \begin{bmatrix} 1 & & & & & & & \\ & & & & 1 & & & \\ & 1 & & & & & & \\ & & & & & 1 & & \\ & & \ddots & & & & & \\ & & & & & & \ddots & \\ & & & 1 & & & & \\ & & & & & & & 1 \end{bmatrix}$

$$\Rightarrow F_{2k}P' = \begin{bmatrix} 1 & 1 & 1 & 1 & \dots & 1 & 1 & 1 \\ 1 & w^2 & w^4 & w^6 & \dots & w^{2k-5} & w^{2k-3} & w^{2k-1} \\ 1 & w^4 & w^8 & w^{12} & \dots & w^{2(2k-5)} & w^{2(2k-3)} & w^{2(2k-1)} \\ 1 & w^6 & w^{12} & w^{18} & \dots & w^{3(2k-5)} & w^{3(2k-3)} & w^{3(2k-1)} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 1 & w^{2(2k-3)} & w^{4(2k-3)} & w^{6(2k-3)} & \dots & w^{(2k-5)(2k-3)} & w^{(2k-3)^2} & w^{(2k-1)(2k-3)} \\ 1 & w^{2(2k-2)} & w^{4(2k-2)} & w^{6(2k-2)} & \dots & w^{(2k-5)(2k-2)} & w^{(2k-3)(2k-2)} & w^{(2k-1)(2k-2)} \\ 1 & w^{2(2k-1)} & w^{4(2k-1)} & w^{6(2k-1)} & \dots & w^{(2k-5)(2k-1)} & w^{(2k-3)(2k-1)} & w^{(2k-1)^2} \end{bmatrix}.$$

Here, $w = e^{\frac{\pi}{k}\iota}$ is the $(2k)^{\text{th}}$ root of unity. Split $F_{2k}P'$ into four quadrants of size $k\times k$ and call them A_I, A_{II}, A_{III}, and A_{IV} starting from the top right and going anti-clockwise:

1. $A_I = \begin{bmatrix} 1 & 1 & \dots & 1 \\ w & w^3 & \dots & w^{2k-1} \\ \vdots & \vdots & \ddots & \vdots \\ w^{k-1} & w^{3(k-1)} & \dots & w^{(2k-1)(k-1)} \end{bmatrix} : \ (A_I)_{ij} = w^{i(2j+1)} = w^{2ij+i} = w^i w^{2ij},$

2. $A_{II} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & w^2 & \dots & w^{2(k-1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & w^{2(k-1)} & \dots & w^{2(k-1)^2} \end{bmatrix} : \ (A_{II})_{ij} = w^{2ij},$

3. $A_{III} = \begin{bmatrix} 1 & w^{2k} & \dots & w^{(2k-2)k} \\ 1 & w^{2(k+1)} & \dots & w^{(2k-2)(k+1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & w^{2(2k-1)} & \dots & w^{(2k-2)(2k-1)} \end{bmatrix} : \ (A_{III})_{ij} = w^{2(k+i)j} = w^{2kj}w^{2ij} = w^{2ij},$

4. $A_{IV} = \begin{bmatrix} w^k & w^{3k} & \dots & w^{(2k-1)k} \\ w^{k+1} & w^{3(k+1)} & \dots & w^{(2k-1)(k+1)} \\ \vdots & \vdots & \ddots & \vdots \\ w^{2k-1} & w^{3(2k-1)} & \dots & w^{(2k-1)^2} \end{bmatrix} : \ (A_{IV})_{ij} = w^{(k+i)(2j+1)} = w^{2kj+k+i}w^{2ij} = -w^i w^{2ij}.$

Since $w^2 = e^{\frac{2\pi}{k}\iota}$ is the k^{th} root of unity that makes up F_k, A_{II} is nothing but F_k. Also, A_I can be factorized as DF_k, where D is a diagonal matrix that pulls out w^i from each row i of A_I leaving us with F_k. Hence, $D = \begin{bmatrix} 1 & & & & \\ & w & & & \\ & & w^2 & & \\ & & & \ddots & \\ & & & & w^{k-1} \end{bmatrix}$ and $A_I = DF_k$. For A_{III}, $(A_{III})_{ij} = w^{2kj}w^{2ij}$, where

$w^{2kj} = (w^{2k})^j = 1 \Rightarrow (A_{III})_{ij} = w^{2ij} \Rightarrow A_{III} = A_{II} = F_k$. For A_{IV}, $(A_{IV})_{ij} = w^{2kj+k+i}w^{2ij}$, where $w^{2kj+k+i} = (w^{2k})^j w^k w^i = -w^i$ since $w^{2k} = 1$ and $w^k = e^{\iota\frac{\pi}{k}\cdot k} = e^{\pi\iota} = -1$. Hence, $(A_{IV})_{ij} = -w^i w^{2ij} \Rightarrow A_{IV} = -A_I = -DF_k$. Finally, the results are: $A_I = DF_k,\ A_{II} = F_k,\ A_{III} = F_k$ and $A_{IV} = -DF_k$.

$$\therefore F_{2k}P' = \begin{bmatrix} A_{II} & A_I \\ A_{III} & A_{IV} \end{bmatrix} = \begin{bmatrix} F_k & DF_k \\ F_k & -DF_k \end{bmatrix} = \begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} F_k & O \\ O & F_k \end{bmatrix}.$$

To complete the factorization, right-multiply both sides by ${P'}^{-1}$. Set $P = {P'}^{-1} \Rightarrow P$ reverses the permutation obtained by P'; i.e. it alternates the columns as $0,\ k,\ 1,\ k+1,\ \ldots,\ i,\ k+i,\ \ldots$ (as opposed to the even-odd alternating permutation obtained by P'). If the identity is permuted in this way, then P is obtained as

$$P = \begin{bmatrix} 1 & & & & & & & \\ & & 1 & & & & & \\ & & & & \ddots & & & \\ & & & & & & 1 & \\ & 1 & & & & & & \\ & & & 1 & & & & \\ & & & & & \ddots & & \\ & & & & & & & 1 \end{bmatrix}.$$

$$\therefore \boxed{F_{2k} = \begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} F_k & O \\ O & F_k \end{bmatrix} P}.$$

`Benefit.` A Fourier transform using F_{2k} costs $(2k)^2 = 4k^2$ multiplications. On the other hand, the cost of multiplying by the factorization is attributed to multiplication by two copies of F_k in $\begin{bmatrix} F_k & O \\ O & F_k \end{bmatrix}$, and by $D \Rightarrow$ a Fourier transform using the factorization costs $2k^2 + k$ multiplications. This is the improvement because of one factorization. However, the factorization idea can be used recursively to factorize F_k in terms of $F_{\frac{k}{2}}$ and so on to get further improvement. For instance, the cost after the second factorization would be $2(2(\frac{k}{2})^2 + \frac{k}{2}) + k$. Ultimately, the squared terms will vanish but the "fix up" factor will remain: $k + 2\cdot\frac{k}{2} + 2\cdot 2\cdot\frac{k}{4} + \ldots = \boxed{k\log k} < 4k^2$. This is the *fast* Fourier transform.

29. Positive Definite Matrices and Minima

> "I'm doing mentally a multiplication of this matrix times a row vector on the left times a column vector on the right and I know that these numbers show up in the answer."

Outline.

- Positive definite matrix (tests)
- Tests for minimum ($x^\mathsf{T} Sx > 0$)
- Ellipsoids in $\mathbb{R}^n$

Tests for positive definiteness. A (symmetric) matrix $S_{n\times n}$ is positive definite iff:

1. $\lambda_1 > 0,\ \lambda_2 > 0, \ldots, \lambda_n > 0$.
2. All leading principal subdeterminants > 0.
3. All pivots > 0 (in elimination without row exchanges).
4. $\star\ x^\mathsf{T} Sx > 0$ for any vector $x \neq 0 \in \mathbb{R}^n$.

Positive semidefiniteness. The eigenvalue and quadratic-form tests remain valid with ≥ 0. For the determinant test, positive semidefiniteness requires all principal minors (not only the leading ones) to be nonnegative; simply replacing the pivot inequalities can fail when a zero pivot occurs.

The 4^{th} test is the definition of positive definiteness (the definition of positive semidefiniteness is $x^{\mathsf{T}}Sx \geq 0$ for all x). This also means that if any one of the first three tests fails, I can definitely find $x_1, x_2, \ldots, x_n \in \mathbb{R}^n$ not all zero such that the vector $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ satisfies $x^{\mathsf{T}}Sx \leq 0$.

2×2 **example.** Let $S = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ be a positive definite matrix. By definition, this means $\begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} > 0$ for all $x_1,\ x_2 \in \mathbb{R}$ not both zero. Expand and write this condition in **quadratic form**: $\forall\, x_1, x_2 \in \mathbb{R},\ (x_1, x_2) \neq (0,0) \Rightarrow ax_1^2 + 2bx_1x_2 + cx_2^2 > 0$. Let $f(x_1, x_2)$ be the function given by the quadratic form.

- The graph of $f(x_1, x_2)$ is a 3D "bowl" that goes upwards from $(0,0)$ in all directions. Its cross-section at any height $f(x_1, x_2) = k > 0$ is an ellipse.
- Since S is positive definite, $f(x_1, x_2)$ will have a **minimum** at $(0,0)$.
- If S were indefinite, $f(x_1, x_2)$ would have a **saddle point** at $(0,0)$ (the graph would go upwards in some directions and downwards in others).
- We can also work the other way round: given the quadratic form, check whether the matrix is positive definite. Idea: if the quadratic form can be expressed as a sum of squares then it is always positive and the matrix is positive definite. Start by completing the square:

$$\begin{aligned} f(x_1, x_2) &= ax_1^2 + 2bx_1x_2 + cx_2^2 \\ &= a(x_1^2 + 2\frac{b}{a}x_1x_2) + cx_2^2 \\ &= a(x_1^2 + 2\frac{b}{a}x_1x_2 + \frac{b^2}{a^2}x_2^2 - \frac{b^2}{a^2}x_2^2) + cx_2^2 \\ &= a(x_1^2 + 2\frac{b}{a}x_1x_2 + \frac{b^2}{a^2}x_2^2) + cx_2^2 - \frac{b^2}{a}x_2^2 \\ &= a(x_1 + \frac{b}{a}x_2)^2 + \frac{ac - b^2}{a}x_2^2. \end{aligned}$$

  This is of the form $f(x_1, x_2) = k(x_1 + \ell x_2)^2 + mx_2^2$. For this to be positive at all x_1, x_2 not both zero, the condition we get is k is positive *and* m is positive, i.e. a and $\frac{ac-b^2}{a}$ are positive. BUT "those numbers are no accident," "those numbers come from elimination!" $k = a$ and $m = \frac{ac-b^2}{a}$ are the pivots obtained on the diagonal after Gaussian elimination and $\ell = \frac{b}{a}$ is the multiplier that goes into L. $\therefore$ The final condition we get is nothing but both pivots are positive.
- The matrix of second derivatives of $f(x_1, x_2)$ is $H_f = \begin{bmatrix} f_{x_1x_1} & f_{x_1x_2} \\ f_{x_2x_1} & f_{x_2x_2} \end{bmatrix}$.

Minima. In calculus, the first derivative being 0 at a point is not a sufficient condition for that point to be a minimum of the function. Assuming the function is single variate, its second derivative at that point being positive is what confirms that the point is a minimum. The corresponding condition for a general multivariate function $f(x_1, x_2, \ldots, x_n)$ is that its matrix of second derivatives is positive definite (in addition to each first derivative being 0).

3×3 **example.** $S = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$.

- Subdeterminants: 2, 3, 4 (compute).

- Pivots: 2, $\frac{2}{3}$, $\frac{3}{4}$ (the subdeterminants should be the products when one pivot is taken, when two pivots are taken, and when all three are taken respectively).
- Eigenvalues: $2-\sqrt{2}$, 2, $2+\sqrt{2}$ (compute).
- `Quadratic form:` $f(x,y,z) = v^\mathsf{T} S v = 2x^2 + 2y^2 + 2z^2 - 2xy - 2yz$. The matrix S is positive definite by the subdeterminants test $\Rightarrow$ the quadratic form can be written as a sum of three squares.
- "What's a long word for a 'bowl?' Maybe *paraboloid*, ... I'll edit the tape and get that word in. 'Bowl,' let's say ..." The graph of $f(x,y,z)$ is a 4D "bowl" that goes upwards in all directions starting at $(0,0,0)$. Its level set at any height $f(x,y,z) = k > 0$ is an ellips**oid**.
- The ellipsoid will be a sphere if all three eigenvalues are equal. It will be like a rugby ball if one eigenvalue is repeated and the other is different. But here, all three eigenvalues are different $\Rightarrow$ it will be like a "lopsided football." It will have a major axis, a middle axis, and a minor axis, and those three axes will be in the directions of the eigenvectors. For the level set $x^\mathsf{T} S x = k$, the half-lengths are $\sqrt{k/\lambda_i}$.
  - `Proof sketch.` Write $S = Q\Lambda Q^\mathsf{T}$. "The **principal axis theorem** we'll call it now."

30. Similar Matrices and Jordan Form

"Excuse me—I think of these matrices as people by this point in 18.06."

Outline.

- $A^\mathsf{T} A$ is positive definite!
- Similar matrices A, B
- $B = M^{-1}AM$, Jordan form

⋆ The inverse of a positive definite matrix is also positive definite.

⋆ If A and B are positive definite, then $A+B$ is positive definite.

⋆ If A is an $m{\times}n$ matrix, then $A^\mathsf{T} A$ is positive semidefinite. If A has rank n (full column rank $\Rightarrow$ no nullspace) then $A^\mathsf{T} A$ is positive definite.

- `Proof sketch.` $x^\mathsf{T} A^\mathsf{T} A x = (Ax)^\mathsf{T} Ax = \|Ax\|^2$.

Similar matrices. Matrices $A_{n\times n}$ and $B_{n\times n}$ are *similar* if there exists a (invertible) matrix $M_{n\times n}$ such that $B = M^{-1}AM$.

- `Example.` A diagonalizable matrix A is similar to Λ.

Example. $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \Rightarrow \Lambda = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$. Let $M = \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}$. Then, $M^{-1}AM = \begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} -2 & -15 \\ 1 & 6 \end{bmatrix} = B$. Check that the eigenvalues of B are also 3 and 1.

Main fact about similar matrices. Similar matrices have the same eigenvalues!

- `Proof sketch.` $Ax = \lambda x \Rightarrow AMM^{-1}x = \lambda x \Rightarrow M^{-1}AMM^{-1}x = \lambda M^{-1}x \Rightarrow BM^{-1}x = \lambda M^{-1}x \Rightarrow By = \lambda y$.
- `Corollary.` The eigenvectors of B are M^{-1} times the eigenvectors of A.
- Converse is not always true: All *nondegenerate* matrices with the same eigenvalues are similar.
- `Corollary.` Similar matrices have the same trace and determinant.

BAD CASE. $\lambda_1 = \lambda_2 = 4$. Then there is not one family of similar matrices with these eigenvalues, there are two!

1. One (small) family has $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ (loner).

2. Big family includes $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}, \ldots$.

$4I$ is in a family all by itself. The only matrix $4I$ is similar to is itself since $M^{-1}(4I)M = 4M^{-1}M = 4I$.

Jordan form. "This guy, Jordan, studied these families of matrices and in each family, he picked out the nicest, most diagonal one." The big family above does not have any diagonal matrices, so there's a 1 up there in the Jordan form: $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$. "In this topic of similar matrices, the climax is the Jordan form."

The Jordan form helps cover the case of nondiagonalizable matrices, but it requires the eigenvalues to be exactly the same. The slightest change in the eigenvalues or the rank completely changes the Jordan form, which makes it not so good numerically.

More members of the big family: $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}, \begin{bmatrix} 5 & 1 \\ -1 & 3 \end{bmatrix}, \begin{bmatrix} 4 & 10 \\ 0 & 4 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 17 & 4 \end{bmatrix}, \begin{bmatrix} a & 8a-a^2-16 \\ 1 & 8-a \end{bmatrix}, \ldots$

$\star$ Similar matrices have the same number of (independent) eigenvectors.

Example. "If you want nightmares, think about matrices like these." $A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

- $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0$. "We're really getting bad matrices now."
- A has two independent eigenvectors because the dimension of the nullspace is 2 (2 zero rows).

Now consider the matrix $B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. B also has four 0 eigenvalues and 2 independent eigenvectors. But B is NOT similar to A.

Jordan blocks. A Jordan block is a square block/matrix with 1 repeated eigenvalue and 1 eigenvector such that the eigenvalue appears on the diagonal and there are 1's above the diagonal and 0's everywhere else. $J_i = \begin{bmatrix} \lambda_i & 1 & 0 & \ldots & 0 \\ 0 & \lambda_i & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & \ldots & \lambda_i \end{bmatrix}$. In the example above, A has two Jordan blocks: $J_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$ and $J_2 = 0$, and B has two Jordan blocks: $J_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ and $J_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$.

Jordan matrix. A Jordan matrix J is one composed of Jordan blocks on the diagonals and 0's everywhere else. $J = \begin{bmatrix} J_1 & O & \ldots & O \\ O & J_2 & \ldots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \ldots & J_d \end{bmatrix}$.

$\star$ **Jordan's theorem.** Every square matrix A over $\mathbb{C}$ is similar to a Jordan matrix J.

$\star$ The number of blocks is the number of (independent) eigenvectors.

$\star$ If A is nondegenerate, $J = \Lambda$ and $J_i = \lambda_i$.

"I'm not that crazy about the Jordan form."

31. Singular Value Decomposition

"This is the lecture on the singular value decomposition, but everybody calls it the SVD."

Outline.

- Singular value decomposition = **SVD** ("The final and best factorization of a matrix")
- $A = U\Sigma V^\mathsf{T}$ (Σ diagonal, U, V orthogonal)

SVD. Any matrix A has a factorization into two orthogonal matrices U and V and a diagonal matrix Σ such that $A = U\Sigma V^\mathsf{T}$.

- For a real symmetric positive semidefinite matrix, the SVD is $A = Q\Lambda Q^\mathsf{T}$, where $U = V = Q$ and $\Sigma = \Lambda$. For a general real symmetric matrix, Σ contains the absolute values of the eigenvalues and the corresponding signs are absorbed into one of the orthogonal factors.

The goal of the SVD for a matrix $A_{m\times n}$ is to find an orthogonal basis $v_1, v_2, \ldots, v_r$ of the row space in $\mathbb{R}^n$ such that it can be knocked over to an orthogonal basis $u_i = Av_i$ of the column space in $\mathbb{R}^m$.

- This is a special setup because if I start with any old orthogonal basis, there is no reason why it should translate to an orthogonal basis in $\mathrm{C}(A)$ when multiplied by A.

Let's now say that $v_1, v_2, \ldots, v_r$ is an ortho*normal* basis of the row space and so is $u_1, u_2, \ldots, u_r$ of the column space. We have now added the condition that each v_i and u_i should be unit vectors. This calls for a scaling factor $\sigma > 0$ since Av_i would be in the direction of u_i but scaled: $Av_i = \sigma_i u_i$. The revised goal is to find such an orthonormal basis $v_1, v_2, \ldots, v_r$.

The goal: $A\begin{bmatrix} v_1 & v_2 & \ldots & v_r\end{bmatrix} = \begin{bmatrix}\sigma_1 u_1 & \sigma_2 u_2 & \ldots & \sigma_r u_r\end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \ldots & u_r\end{bmatrix}\begin{bmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r\end{bmatrix}$

$\Rightarrow A_{m\times n}V_{n\times r} = U_{m\times r}\Sigma_{r\times r}$.

Now let $v_{r+1}, v_{r+2}, \ldots, v_n$ be an orthonormal basis for the nullspace such that $v_1, \ldots, v_n$ is a basis for the entire space $\mathbb{R}^n$. Similarly, complete the orthonormal basis for the column space with $u_{r+1}, u_{r+2}, \ldots, u_m$ such that $u_1, \ldots, u_m$ is an orthonormal basis for the entire space $\mathbb{R}^m$. The new scaling factors will all be 0's because the new v_j's form the nullspace basis with $Av_j = \mathit{0}$. Complete V with the new v_j's, complete U with the new u_j's, and complete Σ with 0's,

$\Rightarrow A_{m\times n}V_{n\times n} = U_{m\times m}\Sigma_{m\times n}$.

Now invert V and right-multiply by the inverse. But since V is orthonormal, $V^{-1} = V^\mathsf{T}$,

$\Rightarrow \boxed{A = U\Sigma V^\mathsf{T}}$

$\Rightarrow A^\mathsf{T}A = V\Sigma^\mathsf{T}U^\mathsf{T}U\Sigma V^\mathsf{T} = V\Sigma^\mathsf{T}\Sigma V^\mathsf{T} = V\begin{bmatrix}\sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \ddots\end{bmatrix}V^\mathsf{T}$. But now this is of the form $A^\mathsf{T}A = Q\Lambda Q^\mathsf{T}$, where $A^\mathsf{T}A$ is symmetric positive semidefinite (positive definite when A has full column rank) $\Rightarrow V = Q$ and $\Lambda = \Sigma^\mathsf{T}\Sigma \Rightarrow$ the v's are the eigenvectors of $A^\mathsf{T}A$.

Similarly, $AA^\mathsf{T} = U\Sigma\Sigma^\mathsf{T}U^\mathsf{T} \Rightarrow$ the u's are the eigenvectors of AA^T. The nonzero eigenvalues of $A^\mathsf{T}A$ and AA^T are the same, and the σ's are their positive square roots. "That's the singular value decomposition."

Question. What are we really doing?
Answer. We're choosing the right basis for the four subspaces of linear algebra:

1. $v_1, \ldots, v_r$: orthonormal basis for the row space,
2. $u_1, \ldots, u_r$: orthonormal basis for the column space,
3. $v_{r+1}, \ldots, v_n$: orthonormal basis for the nullspace,
4. $u_{r+1}, \ldots, u_m$: orthonormal basis for the left nullspace.

and $Av_i = \sigma_i u_i \Rightarrow A = U\Sigma V^{\mathsf{T}}$.

32. Linear Transformations and Their Matrices

"You see I'm moving to coordinates so quickly, I'm not a true physicist here."

Outline.

- Linear transformations T
- Without coordinates: no matrix
- With coordinates $\rightarrow$ MATRIX

Coordinates. *Coordinates* come from a basis $v_1, \ldots, v_n$. The coordinates of v are the numbers $c_1, \ldots, c_n$ such that $v = \sum_{i=1}^{n} c_i v_i$.

Linear transformation. A *linear transformation* T is a function, map, mapping such that:

1. $T(v + w) = T(v) + T(w)$,
2. $T(\mathtt{c}v) = \mathtt{c}T(v)$.

The condition can be expressed compactly as $T(\mathtt{c}v + \mathtt{d}w) = \mathtt{c}T(v) + \mathtt{d}T(w)$.

- `Example 1.` Projections are linear transformations that can be described without matrices as mappings.
- `NONexample 1.` Shifting the whole plane ($T(v) = v + v_0$) is not a linear transformation.
- `NONexample 2.` $T : \mathbb{R}^3 \rightarrow \mathbb{R}^1,\ v \mapsto \|v\|$ is not a linear transformation.
- `Example 2.` Rotations are linear transformations that can be described as mappings.
- $T(0) = 0$.
- We don't necessarily need coordinates to describe linear transformations. "I haven't told you a matrix yet."
- Linear transformations coming from matrices can also be described as mappings. Given matrix A, $T(v) = Av$.
- `Example 3.` Reflections in the X-axis in $\mathbb{R}^2$ can be described using the matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and also by $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2,\ v \mapsto Av$.

Question. How much information is needed to know $T(v)$ for all inputs?
Answer. It is enough to know $T(v_1), T(v_2), \ldots, T(v_n)$ for any basis $v_1, \ldots, v_n$ to know everything.

The goal is to find a matrix for a given linear transformation T such that T corresponds to multiplication by that matrix: $T(v) = Av$.

To construct a matrix A that represents a linear transformation $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$, we need to introduce coordinates. For this, choose basis vectors $v_1, \ldots, v_n$ for $\mathbb{R}^n$ and basis vectors $u_1, \ldots, u_m$ for $\mathbb{R}^m$. For each basis vector v_i, $T(v_i)$ can be expressed using coordinates $a_{1i}, \ldots, a_{mi}$ such that $T(v_i) = \sum_{j=1}^{m} a_{ji} u_j \Rightarrow$ for any input $v = \sum_{i=1}^{n} c_i v_i,\ T(v) = \sum_{i=1}^{n} c_i T(v_i) = \sum_{i=1}^{n} c_i \sum_{j=1}^{m} a_{ji} u_j = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} c_i a_{ji} \right) u_j$. The term in the parentheses is the j^{th} coordinate of $T(v)$, $c'_j \Rightarrow \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} c'_1 \\ \vdots \\ c'_m \end{bmatrix}$. The vector $\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$ is nothing but the coordinates of v in the basis $v_1, v_2, \ldots, v_n$, and the vector $\begin{bmatrix} c'_1 \\ \vdots \\ c'_m \end{bmatrix}$ is nothing but the coordinates of $T(v)$ in the basis $u_1, \dots, u_m$. $\therefore$ We have got our matrix $\boxed{A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}}$ such that $T(v) = Av$.

Rule to find A. Given a linear transformation T and bases $v_1, \dots, v_n$ and $u_1, \dots, u_m$; the i^{th} column of A will be the numbers multiplying $u_1, \dots, u_m$ when $T(v_i)$ is expressed in the u-basis.

Example. The operator (!) that takes the derivative $T = \frac{d}{dx}$ is a linear transformation. Suppose the input space is all functions of the form $c_0 + c_1 x + c_2 x^2$ and the output is its derivative $c_1 + 2c_2 x \Rightarrow$ a basis for the input space is $1, x, x^2$ and a basis for the output space is $1, x$. We are going from a 3D space of inputs to a 2D space of outputs by taking the derivative. The goal is to find the matrix D that multiplies the input coordinates $D \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix}$ and gives the output coordinates $\begin{bmatrix} c_1 \\ 2c_2 \end{bmatrix}$. Follow the rule:

1. $T(1) = 0 = (0)1 + (0)x \Rightarrow D_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$,
2. $T(x) = 1 = (1)1 + (0)x \Rightarrow D_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$,
3. $T(x^2) = 2x = (0)1 + (2)x \Rightarrow D_3 = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$.

$\therefore D = \begin{bmatrix} D_1 & D_2 & D_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$.

$\star$ The inverse of a matrix for a linear transformation gives the inverse transformation.

$\star$ If A and B are matrices for transformations T_1 and T_2, then AB is the matrix for $T_1 \circ T_2$.

33. Change of Basis; Image Compression

Outline.

- Change of basis
- Compression of images
- Transformation $\leftrightarrow$ Matrix

Image compression. A vector x that describes a 512×512 image belongs to $\mathbb{R}^{512^2}$ which is huge. Each entry of 8 bits (0 through 255) represents a pixel. If the image is color than the dimension n increases to $3(512)^2$. Compressing this information is essential to prevent overloading the transmission system.

JPEG. Stands for Joint Photographic Experts Group. It's the standard compression protocol. It's a change of basis.

- Standard basis: $\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$.
- Better basis: $\begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \\ \vdots \\ \pm 1 \end{bmatrix}, \dots, \begin{bmatrix} 1 \\ 1 \\ \vdots \\ -1 \\ -1 \end{bmatrix}$. This allows to encode solid images and alternate light dark

images much more conveniently.

The basis used by JPEG is the discrete cosine transform (DCT) basis, closely related to the Fourier basis. The 512×512 big signal gets broken up into 8×8 blocks, and the basis is changed on every piece. Each block has 64 pixels and can be encoded as vectors in 64D space, with 64 coordinates for 64 basis vectors, originally the standard basis.

Signal $x \xrightarrow[\text{lossless}]{\text{change of basis}}$ coefficients $c \xrightarrow[\text{lossy}]{\text{compression}}$ sparse, compressed coefficients $\hat{c} \xrightarrow{\text{reconstruction}} \hat{x} = \sum_i \hat{c}_i v_i$

The competition to the Fourier basis is called <u>wavelets</u>. The 8×8 wavelet basis is:

$$\begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\1\\1\\-1\\-1\\-1\\-1\end{bmatrix}, \begin{bmatrix}1\\1\\-1\\-1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\1\\1\\-1\\-1\end{bmatrix}, \begin{bmatrix}1\\-1\\0\\0\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\-1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\1\\-1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\0\\0\\0\\0\\1\\-1\end{bmatrix}.$$

The linear algebra step is to find the new (compressed) coefficients. Suppose a 1×8 row of pixels is represented in the standard basis as $p = \begin{bmatrix}p_1\\p_2\\\vdots\\p_8\end{bmatrix}$. The goal is to express this in the wavelets basis $w_!, \dots, w_8$ as $p = \sum_{i=1}^{8} c_i w_i$, or $p = Wc$, where $W = \begin{bmatrix}w_1 & \dots & w_8\end{bmatrix}$. This gives $c = W^{-1}p$. Hence, a good basis would be:

1. *Fast:* Multiplication by W and its inverse should happen *fast*.
   - FFT and FWT (fast wavelet transform) enable exactly that. After normalizing the displayed orthogonal basis, $W^{-1} = W^\mathsf{T}$.
2. *Good compression:* "Few is enough:" a few basis vectors are enough to reproduce the original signal.

Change of basis. Let x be a vector in an old basis, W be the matrix whose columns are the new basis vectors, and c be the corresponding vector in the new basis. Then, $\boxed{c = W^{-1}x.}$

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Suppose it has a matrix A w.r.t the basis $v_1, \dots, v_n$, and a matrix B w.r.t the basis $w_1, \dots, w_n$. Then, A and B are <u>similar</u> with $B = W^{-1}AW$.

- `Proof sketch.` $A[x]_v = [T(x)]_v$ and $B[x]_w = [T(x)]_w$. $[T(x)]_v = W[T(x)]_w \Rightarrow A[x]_v = WB[x]_w$. But $[x]_v = W[x]_w \Rightarrow AW[x]_w = WB[x]_w$. This is true for all x. $\therefore AW = WB \Rightarrow B = W^{-1}AW$.

34. Quiz 3 Review

"I didn't read the book."

Outline.

- λ and x, $A = X\Lambda X^{-1}$
- $S = S^\mathsf{T}$, positive definiteness, $S = Q\Lambda Q^\mathsf{T}$
- Similarity $B = M^{-1}AM$
- SVD $A = U\Sigma V^\mathsf{T}$

$\star$ A real matrix has a complete orthonormal basis of eigenvectors over $\mathbb{C}$ exactly when $AA^\mathsf{T} = A^\mathsf{T}A$.

- Suppose

  $Av = \lambda v, \quad Aw = \mu w$

  with $v, w \in \mathbb{R}^n$ eigenvectors of A for (not-necessarily distinct) eigenvalues λ, μ. Consider the inner products

  $\langle Av,\, w\rangle \quad \text{and} \quad \langle v,\, A^T w\rangle.$

  On one hand,

  $\langle Av,\, w\rangle = v^T A^T w = \langle v,\, A^T w\rangle.$

  On the other hand, since $Av = \lambda v$ and $A^T w = \mu\, w$ (because $Aw = \mu w$ implies $A^T w = \mu w$ when $AA^T = A^T A$), we get

  $\langle Av,\, w\rangle = \lambda\, \langle v, w\rangle, \qquad \langle v,\, A^T w\rangle = \mu\, \langle v, w\rangle.$

  Equating them gives

  $\lambda\, \langle v, w\rangle \;=\; \mu\, \langle v, w\rangle \quad \Longrightarrow \quad (\lambda - \mu)\, \langle v, w\rangle = 0.$

  - If $\lambda \neq \mu$, then $\langle v, w\rangle = 0$, so **eigenvectors for distinct eigenvalues are orthogonal.**
  - If $\lambda = \mu$, one works inside the eigenspace for λ and applies Gram–Schmidt to choose an orthonormal basis there.

  Hence **any real normal matrix is orthogonally diagonalizable**—it admits a full set of mutually orthonormal eigenvectors.

⋆ If $A = U\Sigma V^\intercal$ is the SVD for $A_{m\times n}$, then $A^\intercal A = V(\Sigma^\intercal \Sigma)V^\intercal$ is the SVD for $A^\intercal A$, and $AA^\intercal = U(\Sigma\Sigma^\intercal)U^\intercal$ is the SVD for $AA^\intercal$.

In solving for U and V in the SVD, the sign must be taken care of when choosing the eigenvectors of $A^\intercal A$ and $AA^\intercal$.

⋆ The eigenvalues of an orthogonal matrix ought to have magnitude 1.

- `Proof sketch.` $Qx = \lambda x \Rightarrow \|Qx\| = |\lambda|\|x\| \Rightarrow \|x\| = |\lambda|\|x\| \Rightarrow |\lambda| = 1 \because \|Qx\| = \sqrt{x^\intercal Q^\intercal Q x} = \sqrt{x^\intercal x} = \|x\|$.

35. Left and Right Inverses; Pseudoinverse

> "I shouldn't say anything bad about calculus, but I will. You have formulas for surface area and other awful things ... Linear algebra's about the nice part of calculus where everything's flat."

Outline.

- 4 subspaces
- Left inverses
- Right inverses
- Pseudoinverses

2-sided inverse. This is "what we just call 'inverse.'" $AA^{-1} = I = A^{-1}A$.

- $r = m = n$. Square matrix. Full rank. "Period."

Left inverse. Full column rank ($r = n < m$).

- Independent columns, nullspace $= \{0\}$.
- 0 or 1 solutions to $Ax = b$.
- $A^\mathsf{T}A$ has independent columns, no nullspace. $\therefore A^\mathsf{T}A$ is invertible.
- $\boxed{A_{\text{left}}^{-1} = \left(A^\mathsf{T}A\right)^{-1}A^\mathsf{T}}$ such that $A_{\text{left}}^{-1}A = I_n$.

⋆ A rectangular matrix can't have a 2-sided inverse "cuz there's gotta be some nullspace, right?"

Right inverse. Full row rank ($r = m < n$).

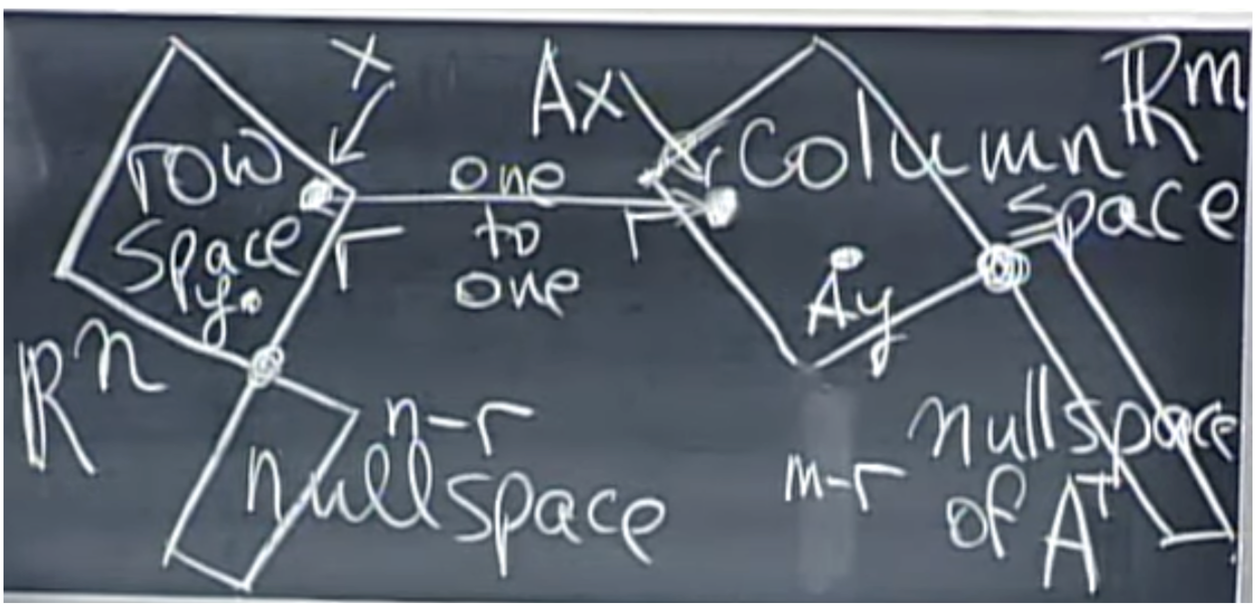
- Independent rows, left nullspace $= \{0\}$.
- ∞ solutions to $Ax = b$.
- AA^T has independent columns, no nullspace. $\therefore AA^\mathsf{T}$ is invertible.
- $\boxed{A_{\text{right}}^{-1} = A^\mathsf{T}\left(AA^\mathsf{T}\right)^{-1}}$ such that $AA_{\text{right}}^{-1} = I_m$.

⋆ $AA_{\text{left}}^{-1} = A\left(A^\mathsf{T}A\right)^{-1}A^\mathsf{T} = P_A$ (projection onto the column space).

⋆ $A_{\text{right}}^{-1}A = A^\mathsf{T}\left(AA^\mathsf{T}\right)^{-1}A = P_{A^\mathsf{T}}$ (projection onto the row space).

Multiplying with the one-sided inverses on the wrong side yields projection matrices. They are *trying* to be the identity but they have been given an impossible job. They are the identity wherever they can be and everywhere else, they are O. "There's only so much the matrix can do."

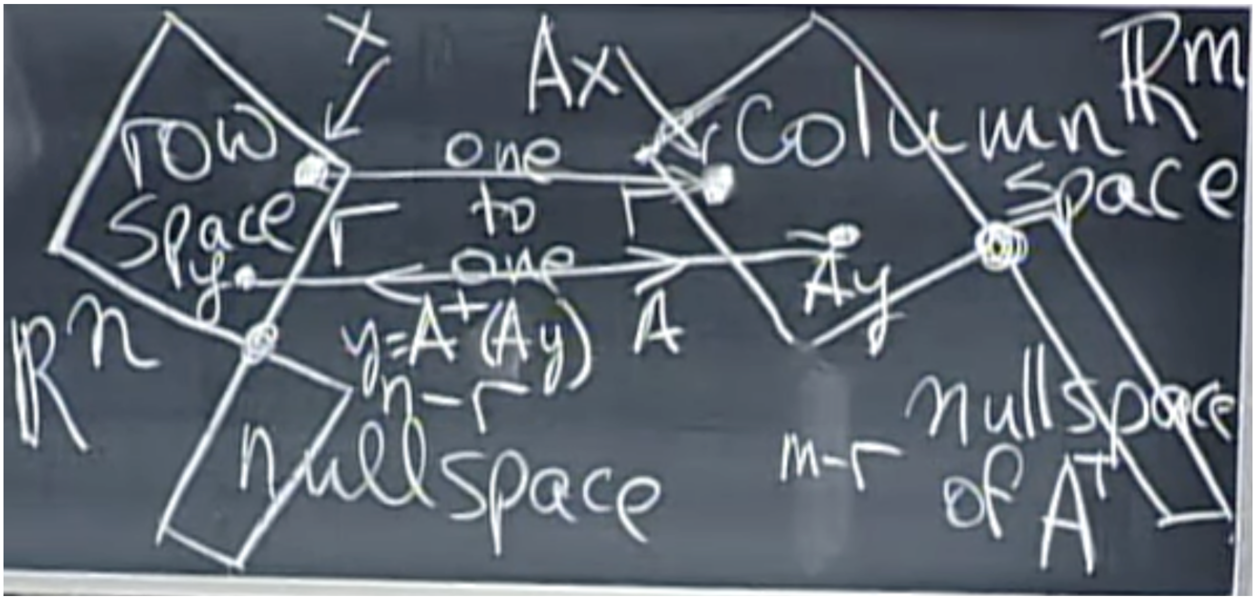
Pseudoinverse. The general case: $r < m,\ r < n$.



$f : \mathrm{C}(A^\mathsf{T}) \to \mathrm{C}(A),\ x \mapsto Ax$ is a bijective mapping between the row space and the column space.

- `Proof sketch.`
    1. `Injective.` If $Ax = Ay$, then $A(x-y) = 0 \Rightarrow x-y \in \mathrm{N}(A)$. Also, $x-y \in \mathrm{C}(A^\mathsf{T}) \because x \in \mathrm{C}(A^\mathsf{T})$ and $y \in \mathrm{C}(A^\mathsf{T})$. But the nullspace and row space are orthogonal complements $\Rightarrow x - y = 0 \Rightarrow x = y$.
    2. `Surjective.` $\forall\, b \in \mathrm{C}(A)$, there exists $x \in \mathbb{R}^n$ such that $Ax = b$. Decompose $x = x_r + x_n$ with $x_r \in \mathrm{C}(A^\mathsf{T})$ and $x_n \in \mathrm{N}(A)$. Then $Ax = Ax_r + Ax_n = Ax_r = b$. $\therefore \forall\, b \in \mathrm{C}(A),\ \exists\, x_r \in \mathrm{C}(A^\mathsf{T})$ such that $Ax_r = b$.

Therefore, f (i.e. A) is a bijection from the row space to the column space $\Rightarrow$ there exists an inverse that takes vectors from the column space to the row space. Call this the **pseudoinverse**. The nullspaces are just hanging around and screwing things up by knocking vectors to zero since an inverse cannot magically bring them back to life. In the cases of the left, right, and 2-sided inverse, one or both nullspaces are absent. The pseudoinverse of A is denoted by A^+. Hence $f^{-1} : \mathrm{C}(A) \to \mathrm{C}(A^\intercal),\ x \mapsto A^+x$ such that $A^+Ax = x \in \mathrm{C}(A^\intercal)$ and $AA^+x = x \in \mathrm{C}(A)$.



Goal: Find the pseudoinverse A^+:

Start with the SVD $A = U\Sigma V^\intercal$, where Σ is an $m \times n$ diagonal matrix of rank r with r singular values on the diagonal and zip everywhere else. Σ^+ is an $n \times m$ diagonal matrix also of rank r with the reciprocals of the r singular values on the diagonal and zilch everywhere else. $\Sigma\Sigma^+$ gives an $m \times m$ diagonal matrix with r 1's on the diagonal and rest 0's. Similarly, $\Sigma^+\Sigma$ gives an $n \times n$ diagonal matrix with r 1's on the diagonal and rest 0's. The pseudoinverse $A^+ = V\Sigma^+U^\intercal$.

36. Final Course Review

"Thanks for taking the course. Thank you."

THANK *YOU*.