Strang

Malcolm

Started 5th June 2025

Contents

1 Vectors and Matrices **4**
- 1.1 Intuition for Dot product, Cosine formula, Schwarz and Triangle inequalities . . . 4
- 1.2 Intuition for column rank being equal to row rank . . . 6
- 1.3 Ways to multiply $\boldsymbol{AB} = \boldsymbol{C}$ . . . 9

2 Solving linear equations $\boldsymbol{Ax} = \boldsymbol{b}$ **10**
- 2.1 Solutions to $\boldsymbol{Ax} = \boldsymbol{b}$ . . . 11
- 2.2 Elimination and Back Substitution . . . 12
- 2.3 Elimination matrices and inverse matrices . . . 15
- 2.4 Gauss-Jordan elimination . . . 17
- 2.5 Proving $\boldsymbol{A} = \boldsymbol{LU}$ . . . 18
- 2.6 Permutation matrices . . . 20
- 2.7 Transposes and symmetric matrices . . . 22
  - 2.7.1 The transpose of $\boldsymbol{AB}$ is $\boldsymbol{B}^T\boldsymbol{A}^T$: Intuition . . . 22
  - 2.7.2 Showing $(\boldsymbol{A}^T)^{-1} = (\boldsymbol{A}^{-1})^T$ . . . 22
  - 2.7.3 $\boldsymbol{A}^T\boldsymbol{A}$ is symmetric . . . 22

3 The Four Fundamental Subspaces **23**
- 3.1 Reduced Row Echelon Form . . . 23
- 3.2 $\boldsymbol{A} = \boldsymbol{CR}$ factorisation . . . 25
- 3.3 Systematic nullspace computation . . . 26
- 3.4 The complete solution to $\boldsymbol{Ax} = \boldsymbol{b}$ . . . 28
  - 3.4.1 Rank and solution . . . 29
- 3.5 Intuition for four subspaces . . . 31
- 3.6 $\boldsymbol{A} = \boldsymbol{CR}$ and Block elimination . . . 32

4 Orthogonality **34**
- 4.1 Orthogonal subspaces . . . 34
  - 4.1.1 Orthogonal complements . . . 35
  - 4.1.2 Intuition for $\boldsymbol{x} = \boldsymbol{x}_{\text{row}} + \boldsymbol{x}_{\text{null}}$ . . . 35
  - 4.1.3 $\boldsymbol{A}$ is invertible from the row space to the column space . . . 36
- 4.2 $\boldsymbol{A}^T\boldsymbol{A}$ is invertible only if $\boldsymbol{A}$ has linearly independent columns . . . 37

4.3 Projections . . . 38
4.3.1 Projection onto a line . . . 38
4.3.2 Projection onto a subspace . . . 39
4.4 Projection matrices are symmetrical . . . 41
4.5 Least Squares Approximations . . . 42
4.5.1 Fitting a straight line . . . 45
4.6 Orthonormal Bases . . . 46
4.6.1 Multiplication by any orthogonal matrix preserves length and angle . . . 47
4.6.2 Projections using orthonormal bases . . . 47
4.7 Gram-Schmidt Process . . . 48
4.8 $\boldsymbol{A} = \boldsymbol{QR}$ factorisation . . . 49
4.9 Pseudoinverses . . . 52

5 Determinants **54**
5.1 Properties of determinants . . . 54
5.1.1 If $\boldsymbol{A}$ has two equal rows, its determinant is 0; row operations don't change the determinant . . . 55
5.1.2 Upper triangular and diagonal matrices, Transpose . . . 56
5.1.3 $\det(\boldsymbol{AB}) = (\det\boldsymbol{A})(\det\boldsymbol{B})$ . . . 56
5.1.4 Orthogonal matrices have determinant 1 or -1 . . . 57
5.1.5 Matrix factorisations and permutation matrix . . . 57
5.2 Intuition for determinant formula . . . 58
5.3 Cofactors, Adjugate, and Inverse . . . 62
5.4 Cramer's Rule to solve $\boldsymbol{Ax} = \boldsymbol{b}$ . . . 64
5.4.1 Inverse formula from Cramer's rule . . . 65
5.5 Sign of a permutation and determinants . . . 66
5.6 Determinants of block-triangular matrices and matrices with identity blocks . . . 68

6 Eigenvalues and Eigenvectors **73**
6.1 Computation . . . 73
6.2 Eigenvalues, Determinant, and Trace . . . 75
6.3 Commuting matrices share eigenvectors . . . 76
6.4 $\boldsymbol{AB}$ and $\boldsymbol{BA}$ have the same eigenvalues . . . 76
6.5 Diagonalisation . . . 77
6.6 Similar Matrices . . . 80
6.7 Matrix Powers . . . 81
6.8 Inutition for spectral theorem . . . 82
6.8.1 Symmetric matrices are always diagonalisable . . . 82
6.8.2 Eigenvectors of distinct eigenvalues of a symmetric matrix are orthogonal . . . 82
6.9 Symmetric Positive Definite Matrices . . . 83
6.9.1 Equivalence of definitions . . . 83
6.9.2 If $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are symmetric positive definite, so is $\boldsymbol{S}_1 + \boldsymbol{S}_2$ 83
6.10 Cholesky decomposition . . . 84

6.11 Other tests for positive definiteness 87
6.11.1 $\boldsymbol{S} = \boldsymbol{A}^T\boldsymbol{A}$ 87
6.11.2 $\boldsymbol{LDL}^T$ factorisation and positive pivots 88
6.12 89

A Misc. topics **90**
A.1 Taylor series, Difference approximations 90
A.2 Conjugate/Hermitian transpose 92
A.3 94

Chapter 1

Vectors and Matrices

1.1 Intuition for Dot product, Cosine formula, Schwarz and Triangle inequalities

Intuition for dot product
The unit vectors $\boldsymbol{v} = (\cos\alpha, \sin\alpha)$ and $\boldsymbol{w} = (\cos\beta, \sin\beta)$ are plotted as follows

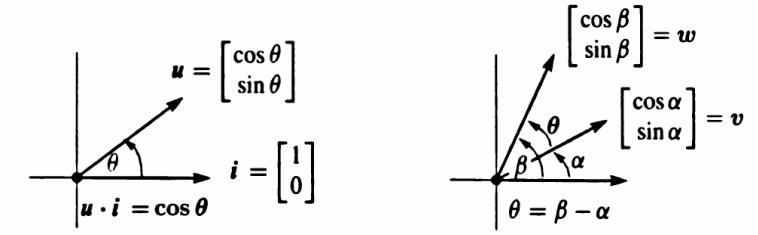


Figure 1.5: Unit vectors: $\boldsymbol{u} \cdot \boldsymbol{i} = \cos\theta$. The angle between the vectors is θ.

See first that when fixed in this form, the magnitude of both vectors is 1, with an angle $\beta - \alpha$ between them. These unit vectors have dot product

$$\boldsymbol{v} \cdot \boldsymbol{w} = \cos\alpha\cos\beta + \sin\alpha\sin\beta = \cos(\beta - \alpha)$$

We have θ as the angle between the two vectors; see that the sign of $\boldsymbol{v} \cdot \boldsymbol{w}$ tells us whether θ is below or above a right angle (due to the cosine function being negative for its argument $> \pi/2$ and positive for $< \pi/2$):

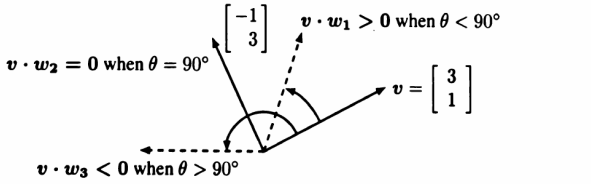


Figure 1.6: Small angle $\boldsymbol{v} \cdot \boldsymbol{w}_1 > 0$. Right angle $\boldsymbol{v} \cdot \boldsymbol{w}_2 = 0$. Large angle $\boldsymbol{v} \cdot \boldsymbol{w}_3 < 0$.

(next page)

Cont.
The idea here is that the dot product reveals the exact angle θ; for unit vectors $\boldsymbol{u}$ and $\boldsymbol{U}$, the dot product $\boldsymbol{u} \cdot \boldsymbol{U}$ is the cosine of θ. Ths remains true in n dimensions (not shown).

See that any $\boldsymbol{u}$ and $\boldsymbol{v}$ can be fixed in the above form by normalising their lengths to get $\boldsymbol{u} = \boldsymbol{v}/||\boldsymbol{v}||$ and $\boldsymbol{U} = \boldsymbol{w}/||\boldsymbol{w}||$. After which their dot product would give $\cos\theta$. This leads us to the *cosine formula*:

$$\textbf{Cosine formula: } \frac{\boldsymbol{v} \cdot \boldsymbol{w}}{||\boldsymbol{v}||\,||\boldsymbol{w}||} = \cos\theta \quad \text{ if } \boldsymbol{v} \text{ and } \boldsymbol{w} \text{ are nonzero vectors}$$

Perpendicular vectors
See that when the angle between $\boldsymbol{v}$ and $\boldsymbol{w}$ is $90°$, its cosine is 0; this gives us a way to test this. Also see that for perpendicular vectors:

$$||\boldsymbol{v} + \boldsymbol{w}||^2 = ||\boldsymbol{v}||^2 + ||\boldsymbol{w}||^2$$

because

$$||\boldsymbol{v} + \boldsymbol{w}||^2 = (\boldsymbol{v} + \boldsymbol{w}) \cdot (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{v} \cdot \boldsymbol{v} + \boldsymbol{v} \cdot \boldsymbol{w} + \boldsymbol{w} \cdot \boldsymbol{v} + \boldsymbol{w} \cdot \boldsymbol{w}$$

where $\boldsymbol{v} \cdot \boldsymbol{w} = 0$.

Schwarz and Triangle inequalities
First, see from the cosine formula that the dot product of $\boldsymbol{v}/||\boldsymbol{v}||$ and $\boldsymbol{w}/||\boldsymbol{w}||$ never exceeds one (since $\cos\theta$ never exceeds one). This is the the *Schwarz inequality*:

$$\textbf{Schwarz inequality: } |\boldsymbol{v} \cdot \boldsymbol{w}| \leq ||\boldsymbol{v}||\,||\boldsymbol{w}||$$

The *Triangle inequality* comes directly from the Schwarz inequality:

$$\textbf{Triangle inequality: } ||\boldsymbol{v} + \boldsymbol{w}|| \leq ||\boldsymbol{v}|| + ||\boldsymbol{w}||$$

This can be seen from

$$||\boldsymbol{v} + \boldsymbol{w}||^2 = \boldsymbol{v} \cdot \boldsymbol{v} + \boldsymbol{v} \cdot \boldsymbol{w} + \boldsymbol{w} \cdot \boldsymbol{v} + \boldsymbol{w} \cdot \boldsymbol{w} \leq ||\boldsymbol{v}||^2 + 2||\boldsymbol{v}||\,||\boldsymbol{w}|| + ||\boldsymbol{w}||^2$$

The square root gives us the triangle equality (side 3 cannot exceed side 1 + side 2).

1.2 Intuition for column rank being equal to row rank

If all columns are in the same direction, why does it happen that all the rows are the same direction?

Consider the matrix, see that column 2 is m times column 1:

$$\boldsymbol{A} = \begin{bmatrix} a & ma \\ b & mb \end{bmatrix}$$

See that the second row is just b/a times the first row—if the column rank is 1, then the row rank is 1. See that transposing the matrix, we have

$$\boldsymbol{A} = \begin{bmatrix} a(1) & b(1) \\ a(m) & b(m) \end{bmatrix}$$

which still has one independent column. Now consider the 3x3 case:

$$\boldsymbol{A} = \begin{bmatrix} a & ma & pa \\ b & mb & pb \\ c & mc & pc \end{bmatrix}$$

See that a similar deduction can also be made in this case, where the row rank of A is equal to its column rank.
(next page)

An informal proof
Consider any matrix $\boldsymbol{A}$, suppose we go from left to right, looking for independent columns of $\boldsymbol{A}$ using the following procedure:

1. If column 1 of $\boldsymbol{A}$ is not zero, put it in matrix $\boldsymbol{C}$
2. If column 2 of $\boldsymbol{A}$ is not a multiple of column 1, put it in into $\boldsymbol{C}$
3. If column 3 of $\boldsymbol{A}$ is not a combination of columns 1 and 2, put it into C. *continue*

See that at the end $\boldsymbol{C}$ will have r columns taken from $\boldsymbol{A}$, where r is the rank of $\boldsymbol{A}$ and $\boldsymbol{C}$. While the n columns of $\boldsymbol{A}$ are dependent, the r columns of $\boldsymbol{C}$ will surely be independent.

For instance consider $\boldsymbol{A}$ with rank 2

$$\boldsymbol{A} = \begin{bmatrix} 2 & 6 & 4 \\ 4 & 12 & 8 \\ 1 & 3 & 5 \end{bmatrix} \quad \text{leads to} \quad \boldsymbol{C} = \begin{bmatrix} 2 & 4 \\ 4 & 8 \\ 1 & 5 \end{bmatrix}$$

Now consider another matrix $\boldsymbol{R}$ to be multiplied by $\boldsymbol{C}$ such that $\boldsymbol{A} = \boldsymbol{CR}$. The first and third columns of $\boldsymbol{A}$ are already in $\boldsymbol{C}$, so those respective columns in $\boldsymbol{R}$ make up a *identity matrix*; the second column of $\boldsymbol{A}$ is a multiple of the first, so we have

$$\boldsymbol{A} = \boldsymbol{CR} \quad \text{is} \quad \begin{bmatrix} 2 & 6 & 4 \\ 4 & 12 & 8 \\ 1 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 4 & 8 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 1 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(See that the ith row of $\boldsymbol{A}$ can be seen as a linear combination of the rows of $\boldsymbol{R}$ specified the ith row of $\boldsymbol{C}$. (or just consider $\boldsymbol{A}^T = \boldsymbol{R}^T\boldsymbol{C}^T$). We know that

1. $\boldsymbol{C}$ contains the full set of r independent columns of $\boldsymbol{A}$.
2. $\boldsymbol{R} = [\boldsymbol{I}\,\boldsymbol{F}]$ contains the identity matrix $\boldsymbol{I}$ in the same r columns that held $\boldsymbol{C}$.
3. The dependent columns of $\boldsymbol{A}$ are combinations of $\boldsymbol{CF}$ of the independent columns in $\boldsymbol{C}$.

Where the matrix $\boldsymbol{F}$ goes into the other $n-r$ columns of $\boldsymbol{R} = [\boldsymbol{I}\,\boldsymbol{F}]$. ($\boldsymbol{A} = \boldsymbol{CR}$ becomes $\boldsymbol{A} = \boldsymbol{C}[\boldsymbol{I}, \boldsymbol{F}] = [\boldsymbol{C}, \boldsymbol{CF}] = [\text{indep cols of } \boldsymbol{A},\ \text{dep cols of } \boldsymbol{A}]$ (in correct order).

See that $\boldsymbol{C}$ has the same column space as $\boldsymbol{A}$, and $\boldsymbol{R}$ *has the same row space* as $\boldsymbol{A}$ (every row of $\boldsymbol{A}$ is a combination of the rows of $\boldsymbol{R}$).
(next page)

Cont.
We had the example

$$\boldsymbol{A} = \boldsymbol{CR} \quad \text{is} \quad \begin{bmatrix} 2 & 6 & 4 \\ 4 & 12 & 8 \\ 1 & 3 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 4 & 8 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 1 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Here is an informal proof that the row rank of $\boldsymbol{A}$ *equals the column rank of* $\boldsymbol{A}$ (based from facts we already know)

1. The r columns of $\boldsymbol{C}$ are independent (chosen that way from $\boldsymbol{A}$)
2. Every column of $\boldsymbol{A}$ is a combination of those r columns of $\boldsymbol{C}$ (since $\boldsymbol{A} = \boldsymbol{CR}$)
3. The r rows of $\boldsymbol{R}$ are independent (they contain the r by r matrix $\boldsymbol{I}$)
4. Every row of $\boldsymbol{A}$ is a combination of the $\boldsymbol{r}$ rows of $\boldsymbol{R}$

See that for every column of $\boldsymbol{A}$ that goes into $\boldsymbol{C}$, a column of $\boldsymbol{I}$ goes into $\boldsymbol{R}$, where each column of $\boldsymbol{I}$ in $\boldsymbol{R}$ adds an independent row.

This means that the column rank of $\boldsymbol{C}$ (column space of $\boldsymbol{A}$) is always equal to the row rank of $\boldsymbol{R}$ (row space of $\boldsymbol{A}$)—the column rank of $\boldsymbol{A}$ is equal to the row rank of $\boldsymbol{A}$.

More examples
Rank 2:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \end{bmatrix}$$

Rank 2:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

Rank 1:

$$\begin{bmatrix} 1 & 2 & 10 & 100 \\ 3 & 6 & 30 & 300 \\ 2 & 4 & 20 & 200 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 10 & 100 \end{bmatrix}$$

1.3 Ways to multiply $AB = C$

Multiplcation by columns of A and rows of B
A lesser known way to multiply AB is through considering the columns of A and the rows of B (contrary to the usual ideas where each entry of the result is a dot product of a row of A and column of B):

$$AB = \begin{bmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{bmatrix} \begin{bmatrix} — & b_1^* & — \\ & \vdots & \\ — & b_n^* & — \end{bmatrix} = a_1 b_1^* + a_2 b_2^* + \cdots + a_n b_n^*.$$

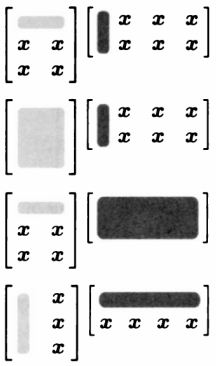
columns a_k **rows b_k^*** **Add columns a_k times rows b_k^***

We multiply each column of A by each row of B; this gives us n rank 1 matrices, which we then sum together; these matrices are called *outer products*

(usually we see the ith column of the result as a linear combination of the columns of A specified by the ith column of B. However in this case the ith outer product is a matrix of the same size as the result, that contains all the contributions of the ith column of A to the final product. By summing this over all n columns of A we get the result.)

Summary of methods

(3 by 2)(2 by 4) = (3 by 4) **Four Ways to Multiply $AB = C$**



(**Row i** of A) · (**Column k** of B) = **Number C_{ik}**
$i = 1$ to 3 $k = 1$ to 4 **12 numbers**

A times (**Column k** of B) = **Column k of C**
$k = 1$ to 4 **4 columns**

(**Row i** of A) times B = **Row i of C**
$i = 1$ to 3 **3 rows**

(**Column j** of A) (**Row j** of B) = **Rank 1 Matrix**
$j = 1$ to 2 **2 matrices**

Dot product way, Column way, Row way, Columns times rows

(A nice way to intuit the the third method is to consider $(AB)^T = B^T A^T$; where the columns of $(AB)^T$ are the rows of AB)

Chapter 2

Solving linear equations $\boldsymbol{Ax = b}$

2.1 Solutions to $\boldsymbol{Ax} = \boldsymbol{b}$

Given a nxn matrix $\boldsymbol{A}$ and an nx1 column vector b, there are three outcomes for the vector $\boldsymbol{x}$ that solves $\boldsymbol{Ax} = \boldsymbol{b}$.

First there may be *no vector* $\boldsymbol{x}$ that solves $\boldsymbol{Ax} = \boldsymbol{b}$, or there may be exactly *one* solution, or there may be *infinitely many* solution vectors $\boldsymbol{x}$. Here are the possibilities:

1. **Exactly one solution** to $\boldsymbol{Ax} = \boldsymbol{b}$ means that $\boldsymbol{A}$ has independent columns (only one particular linear combination of the columns of $\boldsymbol{A}$ leads to $\boldsymbol{b}$. That combination is specified by $\boldsymbol{x}$). $\boldsymbol{A}$ is full rank and the only solution to $\boldsymbol{Ax} = \boldsymbol{0}$ is $\boldsymbol{x} = \boldsymbol{0}$. $\boldsymbol{A}$ *has an inverse matrix* $\boldsymbol{A}^{-1}$ (given $\boldsymbol{b}$, we can work backward to get $\boldsymbol{x}$ since only one $\boldsymbol{x}$ leads to $\boldsymbol{b}$).

2. **No solution** to $\boldsymbol{Ax} = \boldsymbol{b}$ means that $\boldsymbol{b}$ is not in the column space of $\boldsymbol{A}$, so $\boldsymbol{A}$ is not full rank.

3. **Infinitely many solutions**. See that when the columns of $\boldsymbol{A}$ are not independent (not full rank), then there are infinitely many ways to produce the zero vector $\boldsymbol{b} = \boldsymbol{0}$ (this is the meaning of dependent columns), and so there are infinitely many solutions to $\boldsymbol{AX} = \boldsymbol{0}$.

Also see that if $\boldsymbol{A}$ is not full rank it means that its column space is some subspace, where solutions only exist for $\boldsymbol{b}$ within that subspace.

As such, if there so happens to be a solution to $\boldsymbol{Ax} = \boldsymbol{b}$ then we can add any solution to $\boldsymbol{AX} = \boldsymbol{0}$:

$$\boldsymbol{A}(\boldsymbol{x} + \alpha\boldsymbol{X}) = \boldsymbol{Ax} + \alpha\boldsymbol{AX} = \boldsymbol{b} + \boldsymbol{0} = \boldsymbol{b}$$

For some constant α, which gives us $\boldsymbol{b}$ again—we have infinitely many solutions.

2.2 Elimination and Back Substitution

Elimination
We want to produce an *upper triangular* matrix $\boldsymbol{U}$ from a square matrix $\boldsymbol{A}$. This is done through elimination; the procedure (for a 3x3 matrix) is as follows (assuming no row exchanges):

1. Use the first equation(row) to produce zeros in column 1 below the first pivot.
2. Use the new second equation(row) to clear out column 2 below pivot 2 in row 2.
3. *Continue to column 3.* The expected result is an upper triangular matrix $\boldsymbol{U}$.

These steps can be carried out using elimination matrices $\boldsymbol{E}$.

Consider $\boldsymbol{A}$ and $\boldsymbol{b}$

$$\boldsymbol{A} = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 11 & 14 \\ 2 & 8 & 17 \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} 19 \\ 55 \\ 50 \end{bmatrix}$$

$\boldsymbol{E}_{21}$ multiplies equation 1 by 2 and subtracts that from equation 2 to get a zero in the first column below the first pivot:

$$\boldsymbol{E}_{21} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \implies \boldsymbol{E}_{21}\boldsymbol{A} = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 2 & 8 & 17 \end{bmatrix}, \quad \boldsymbol{E}_{21}\boldsymbol{b} = \begin{bmatrix} 19 \\ 17 \\ 50 \end{bmatrix}$$

(For intuition on the elimination matrices, consider the row perspective of matrix multiplication) This produced the desired zero in column 1. It changed equation 2. To make the first pivot column zero, we subtract row 1 from row 3 using $\boldsymbol{E}_{31}$

$$\boldsymbol{E}_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \implies \boldsymbol{E}_{31}\boldsymbol{E}_{21}\boldsymbol{A} = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 5 & 13 \end{bmatrix}, \quad \boldsymbol{E}_{31}\boldsymbol{E}_{21}\boldsymbol{b} = \begin{bmatrix} 19 \\ 17 \\ 31 \end{bmatrix}$$

This completes elimination in column 1. Moving on to column 2 row 2 (the second pivot row). We use $\boldsymbol{E}_{32}$ to subtract equation 2 from equation 3

$$\boldsymbol{E}_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \implies \boldsymbol{U} = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 0 & 7 \end{bmatrix}, \quad \boldsymbol{c} = \begin{bmatrix} 19 \\ 17 \\ 14 \end{bmatrix}$$

$\boldsymbol{E}_{32}\boldsymbol{E}_{31}\boldsymbol{E}_{21}\boldsymbol{A} = \boldsymbol{U}$ is triangular. See that the same steps were applied to the right hand side $\boldsymbol{b}$ to produce a new right hand side $\boldsymbol{c}$.
(next page)

Possible breakdown of elimination
Elimination might fail. This occurs when zero appears in a pivot position—subtracting that zero from lower rows will not clear out the column below that pivot. For example:

$$\boldsymbol{A} = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 6 & 14 \\ 2 & 8 & 17 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 3 & 4 \\ 0 & 0 & 6 \\ 0 & 5 & 13 \end{bmatrix} = \boldsymbol{B}$$

A possible way to get around this would be to *exchange* row 2 (with the zero pivot) for row (with the nonzero in that column), then carry out elimination as per usual. This exchange is carried out using the permuation matrix $\boldsymbol{P}$:

$$\boldsymbol{PB} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 3 & 4 \\ 0 & 0 & 6 \\ 0 & 5 & 13 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 13 \\ 0 & 0 & 6 \end{bmatrix}$$

In this example the exchange produced $\boldsymbol{U}$ with nonzero pivots; normally there may be more columns eliminate before $\boldsymbol{U}$ is reached.

At times this exchange strategy may not work. This occurs when there is no pivot is available. Consider $\boldsymbol{A'}$:

$$\boldsymbol{A'} = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 6 & 14 \\ 2 & 3 & 17 \end{bmatrix} \rightarrow \begin{bmatrix} 2 & 3 & 4 \\ 0 & 0 & 6 \\ 0 & 0 & 13 \end{bmatrix} = \boldsymbol{U'}$$

There is no second pivot in this case. This tells us that the matrix $\boldsymbol{A'}$ *did not have full rank* (intuitively see that the non-pivot row can be expressed as a linear combination of the other pivot rows, so the row space (which is equal to the row space of the original $\boldsymbol{A'}$ since the elimination steps are just linear combinations of the existing rows) is not full rank, and so the column space of the original $\boldsymbol{A'}$ is also not full rank.

In this case there will be nonzero solutions $\boldsymbol{X}$ to $\boldsymbol{A'X} = 0$. The columns of $\boldsymbol{U'}$ (and $\boldsymbol{A'}$) are not independent.

Augmented matrix
During elimination, in order to make sure that the operations on the matrix $\boldsymbol{A}$ are also executed on $\boldsymbol{b}$, one can *include* $\boldsymbol{b}$ *as an extra column of* $\boldsymbol{A}$; this combination $[\boldsymbol{A}, \boldsymbol{b}]$ is called an *augmented matrix*:

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 & 19 \\ 4 & 11 & 14 & 55 \\ 2 & 8 & 17 & 50 \end{bmatrix} \xrightarrow{E} \begin{bmatrix} 2 & 3 & 4 & 19 \\ 0 & 5 & 6 & 17 \\ 0 & 0 & 7 & 14 \end{bmatrix} = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{c} \end{bmatrix}$$

(next page)

Back Substitution to solve $\boldsymbol{Ux} = \boldsymbol{c}$
Elimination (ideally) produces an upper triangular matrix $\boldsymbol{U}$ that has all zeros below the diagonal, with nonzero pivots. For instance we had, for some $\boldsymbol{x}$

$$\boldsymbol{Ax} = \boldsymbol{b}, \quad \boldsymbol{A} = \begin{bmatrix} 2 & 3 & 4 \\ 4 & 11 & 14 \\ 2 & 8 & 17 \end{bmatrix} \quad \boldsymbol{b} = \begin{bmatrix} 19 \\ 55 \\ 50 \end{bmatrix}$$

undergo elimination to to become

$$\boldsymbol{Ux} = \boldsymbol{c}, \quad \boldsymbol{U} = \begin{bmatrix} 2 & 3 & 4 \\ 0 & 5 & 6 \\ 0 & 0 & 7 \end{bmatrix} \quad \boldsymbol{c} = \begin{bmatrix} 19 \\ 17 \\ 14 \end{bmatrix}$$

See that this form allows us to easily solve the equations by going from bottom to top in a procedural manner, finding x_3, then x_2, then x_1:

1. *Back substitution*: The last equation $7x_3 = 14$ gives $x_3 = 2$
2. *Work upwards*: The next equation $5x_2 + 6(2) = 17$ gives $x_2 = 1$
3. *repeat*: The first equation $2x_1 + 3(1) + 4(2) = 19$ gives $x_1 = 4$

giving us the only solution to this example $\boldsymbol{x} = (4, 1, 2)$. Remember the pivots need to be nonzero(full rank) for a single specific solution to be found.

2.3 Elimination matrices and inverse matrices

Elimination matrices
The basic elimination step *subtracts* a multiple ℓ_{ij} of equation j from equation i. We always speak about *subtractions* as elimination proceeds. For instance even if the first pivot $a_{11} = 3$ and below it is $a_{21} = -3$ where we could just add equation 1 to 2, we *subtract* $\ell_{21} = -1$ *times equation 1 from equation 2* (which gives us the same result).

For instance here is the matrix that subtracts 2 times row 1 from row 3:

$$\boldsymbol{E}_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}$$

If no row exchanges are needed, then three elimination matrices $\boldsymbol{E}_{21}, \boldsymbol{E}_{31}, \boldsymbol{E}_{32}$ will produce three zeros below the diagonal to change $\boldsymbol{A}$ to the upper triangular $\boldsymbol{U}$(this just carries out elimination using matrices to represent each step).

Inverse
See that the *inverse* of each matrix $\boldsymbol{E}_{ij}$ just *adds back* $\ell_{ij} \cdot$ (row j) to row i. This leads to the inverse of their product $\boldsymbol{E} = \boldsymbol{E}_{32}\boldsymbol{E}_{31}\boldsymbol{E}_{21}$. We denote the inverse of $\boldsymbol{E}$ by $\boldsymbol{L}$. For instance, say some $\boldsymbol{E}$ subtracts 5 times row 1 from row 2, then $\boldsymbol{E}^{-1}$ adds 5 times row 1 to row 2:

$$\boldsymbol{E} = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{E}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

(See that sequential application of these matrices, in either order, to some vector leads to no net change—as if we had multiplied by the identity.)

Now lets consider $\boldsymbol{F}$ which subtracts 4 times row 2 from row 3 (which might be a next step during elimination), naturally $\boldsymbol{F}^{-1}$ adds it back:

$$\boldsymbol{F} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -4 & 1 \end{bmatrix}, \quad \boldsymbol{F}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}$$

During elimination we would first apply $\boldsymbol{E}$ then $\boldsymbol{F}$, which is the same as applying $\boldsymbol{FE}$. Reversing the elimination would amount to $(\boldsymbol{FE})^{-1}$, which is the same as applying $\boldsymbol{E}^{-1}\boldsymbol{F}^{-1}$:

$$\boldsymbol{FE} = \begin{bmatrix} 1 & 0 & 0 \\ -5 & 1 & 0 \\ 20 & -4 & 1 \end{bmatrix} \quad \text{is inverted by} \quad \boldsymbol{E}^{-1}\boldsymbol{F}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 5 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix}$$

See that the product $\boldsymbol{FE}$ contains '20' but its inverse doesn't. In $\boldsymbol{FE}$, row 3 feels an effect of size 20 from row 1. However in $\boldsymbol{E}^{-1}\boldsymbol{F}^{-1}$ that doesn't happen. (next page)

L is the inverse of E

E is the product of all the elimination matrices E_{ij}, it turns A into its upper triangular form $EA = U$ (assuming no row exchanges). The difficulty with E is multiplying all the separate elimination steps E_{ij} does not produce a good formula; illustrating $E = E_{32}E_{31}E_{21}$:

$$E = \begin{bmatrix} 1 & & \\ 0 & 1 & \\ 0 & -\ell_{32} & 1 \end{bmatrix}\begin{bmatrix} 1 & & \\ 0 & 1 & \\ -\ell_{31} & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & & \\ -\ell_{21} & 1 & \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ -\ell_{21} & 1 & \\ (\ell_{32}\ell_{21} - \ell_{31}) & -\ell_{32} & 1 \end{bmatrix}.$$

See that the bottom left corner is dependent on multiple constants (since mutating the third row requries knowledge of the mutations that already occured to the second row).

Now consider the inverse $E_{21}^{-1}E_{31}^{-1}E_{32}^{-1} = E^{-1} = L$ (see that inverses need to be applied in reverse):

$$\boxed{E^{-1} = \begin{bmatrix} 1 & & \\ \ell_{21} & 1 & \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & & \\ 0 & 1 & \\ \ell_{31} & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & & \\ 0 & 1 & \\ 0 & \ell_{32} & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ \ell_{21} & 1 & \\ \ell_{31} & \ell_{32} & 1 \end{bmatrix} = L}$$

(In the inverse matrix, the mutation of each row doesn't depend on previous mutations, so the multipliers fall into place in the lower triangular L. Also see that the final matrix is only lower triangular because the elimination algorithm specifies that we don't manipulate the first row, and that we don't manipulate the second row with the third row.)

This is why we might want to consider $A = LU$ to go back from triangular U to the original A.

2.4 Gauss-Jordan elimination

How would one compute the inverse of an nxn matrix $\boldsymbol{A}$? Before answering that question, one might want to consider whether it is really necessary to know $\boldsymbol{A}^{-1}$; although it is possible to find the solution to $\boldsymbol{Ax} = \boldsymbol{b}$ using $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$, computing $\boldsymbol{A}^{-1}$ and taking $\boldsymbol{A}^{-1}\boldsymbol{b}$ is a very slow way to find $\boldsymbol{x}$.

Say we want to compute $\boldsymbol{A}^{-1}$. This is equivalent to solving for $\boldsymbol{AX} = \boldsymbol{I}$. In that sense, by performing the manipulations on $\boldsymbol{A}$ to make it look like $\boldsymbol{I}$. We essentially replicate the effect of $\boldsymbol{A}^{-1}$ on $\boldsymbol{A}$; by repeating those steps on the identity, it becomes as if we were taking $\boldsymbol{A}^{-1}\boldsymbol{I} = \boldsymbol{A}^{-1}$, allowing us to obtain the desired matrix.

This whole process can be done with an augmented matrix using *Gauss-Jordan elimination*, where we essentially take steps to reduce $\boldsymbol{A}$ to reduced row echelon form, while repeating said steps on the identity to obtain the inverse:

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{I} \end{bmatrix} = \left[\begin{array}{ccc|ccc} \mathbf{1} & 0 & 0 & 1 & 0 & 0 \\ \mathbf{-1} & \mathbf{1} & 0 & 0 & 1 & 0 \\ \mathbf{0} & \mathbf{-1} & \mathbf{1} & 0 & 0 & 1 \end{array}\right] \rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ \mathbf{0} & 1 & 0 & 1 & 1 & 0 \\ \mathbf{0} & -1 & 1 & 0 & 0 & 1 \end{array}\right]$$

Gauss-Jordan elimination
Solve $AX = I \Rightarrow X = A^{-1}$
Slower than solving $\boldsymbol{Ax} = \boldsymbol{b}$

$$\rightarrow \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \mathbf{1} & 0 & 0 \\ \mathbf{0} & 1 & 0 & \mathbf{1} & \mathbf{1} & 0 \\ \mathbf{0} & \mathbf{0} & 1 & \mathbf{1} & \mathbf{1} & \mathbf{1} \end{array}\right] = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{A^{-1}} \end{bmatrix}$$

The Gauss-Jordan essentially turns $[\boldsymbol{A}, \boldsymbol{I}]$ into $[I, \boldsymbol{A}^{-1}]$, where the elimination steps are essentially equivalent to multiplicaiton by $\boldsymbol{A}^{-1}$.

2.5 Proving $A = LU$

Elimination is expressed by $\boldsymbol{EA} = \boldsymbol{U}$ and inverted by $\boldsymbol{LU} = \boldsymbol{A}$. It starts with $\boldsymbol{A}$ and ends with upper triangular $\boldsymbol{U}$, with each elimination step being carried out by elimination matrices $\boldsymbol{E}_{ij}$. To invert one elimination step we add rows instead of subtracting:

$$E_{31} = \begin{bmatrix} 1 & & \\ 0 & 1 & \\ -\ell_{31} & 0 & 1 \end{bmatrix} \quad \text{and} \quad L_{31} = \text{inverse of } E_{31} = \begin{bmatrix} 1 & & \\ 0 & 1 & \\ \ell_{31} & 0 & 1 \end{bmatrix}.$$

Recall that $\boldsymbol{E} = \boldsymbol{E}_{32}\boldsymbol{E}_{31}\boldsymbol{E}_{21}$ gives us a fairly messy result:

$$\boldsymbol{E} = \begin{bmatrix} 1 & & \\ 0 & 1 & \\ 0 & -\ell_{32} & 1 \end{bmatrix}\begin{bmatrix} 1 & & \\ 0 & 1 & \\ -\ell_{31} & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & & \\ -\ell_{21} & 1 & \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ -\ell_{21} & 1 & \\ (\ell_{32}\ell_{21} - \ell_{31}) & -\ell_{32} & 1 \end{bmatrix}.$$

while the inverse, $\boldsymbol{E}^{-1} = \boldsymbol{E}_{21}^{-1}\boldsymbol{E}_{31}^{-1}\boldsymbol{E}_{32}^{-1} = \boldsymbol{L}$ produces a much simpler result

$$\boldsymbol{E^{-1}} = \begin{bmatrix} 1 & & \\ \ell_{21} & 1 & \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & & \\ 0 & 1 & \\ \ell_{31} & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & & \\ 0 & 1 & \\ 0 & \ell_{32} & 1 \end{bmatrix} = \begin{bmatrix} 1 & & \\ \ell_{21} & 1 & \\ \ell_{31} & \ell_{32} & 1 \end{bmatrix} = \boldsymbol{L}$$

We then have a elegant expression in $\boldsymbol{A} = \boldsymbol{E}^{-1}\boldsymbol{U} = \boldsymbol{LU}$. We now show that this equation holds over larger matrices of size n.

Proof 1
Following each step in elimination, consider the pivot rows that are subtracted from lower rows; see that these rows are not original rows of $\boldsymbol{A}$, since they have been mutated by the previous elimination steps; they are instead rows of $\boldsymbol{U}$.

When computing the, say, third row of $\boldsymbol{U}$, we subtract multiples of earlier rows of $\boldsymbol{U}$:

$$\text{Row 3 of } \boldsymbol{U} = (\text{Row 3 of } \boldsymbol{A}) - \ell_{31}(\text{Row 1 of } \boldsymbol{U}) - \ell_{32}(\text{Row 2 of } \boldsymbol{U})$$

Rewriting, see that

$$\text{Row 3 of } \boldsymbol{A} = \ell_{31}(\text{Row 1 of } \boldsymbol{U}) + \ell_{32}(\text{Row 2 of } \boldsymbol{U}) + 1(\text{Row 3 of } \boldsymbol{U})$$

Row $[\ell_{31}, \ell_{32}, 1]$ is multiplying the matrix $\boldsymbol{U}$. This is *exactly row 3 of* $\boldsymbol{A} = \boldsymbol{LU}$. All rows look like this regardless of the size of $\boldsymbol{A}$. With no row exchanges, we have $\boldsymbol{A} = \boldsymbol{LU}$.
(next page)

Proof 2
Here is another proof. The idea here is to see elimination as removing one rank 1 matrix at a time—one column of $\boldsymbol{L}$ times one row of $\boldsymbol{U}$ from $\boldsymbol{A}$; where the problem becomes one size smaller with each iteration.

Elimination begins with pivot row = row 1 of $\boldsymbol{A}$. We multiply that pivot row by the numbers ℓ_{21}, then ℓ_{31}, and eventually ℓ_{n1}; we subtract the respective products from row 2, row 3, and eventually row n of $\boldsymbol{A}$. By choosing $\ell_{21} = a_{21}/a_{11}$, $\ell_{31} = a_{31}/a_{11}$ and so on until $\ell_{n1} = a_{n1}/a_{11}$; now consider if we also subtracted away the pivot row away from itself—this subtraction leaves zeros in column 1:

$$\textbf{Step 1 removes } \begin{bmatrix} 1\ (\text{row } 1) \\ \ell_{21}(\text{row } 1) \\ \ell_{31}(\text{row } 1) \\ \ell_{41}(\text{row } 1) \end{bmatrix} \textbf{ from } A \textbf{ to leave } \boldsymbol{A_2} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & \times & \times & \times \end{bmatrix}.$$

The idea here is that we *removed a rank 1 matrix* with columns made up of multiples of $\boldsymbol{\ell}_1 = (1, \ell_{21}, \ell_{31}, \ell_{41}, \ldots)$, scaled by each respective entry of the first row of $\boldsymbol{A}$, which is also the first pivot row $\boldsymbol{u}_1$ of the final upper triangular matrix.

We continue with elimination in the second pivot column, using the second row as the pivot row; as before we also subtract the second pivot row from itself. See that this is again equivalent to subtracting a rank 1 matrix with basis $\boldsymbol{\ell}_2 = (0, 1, \ell_{32}, \ell_{42}, \ldots)$. Recall that the second pivot row is the second row of the upper triangular matrix $\boldsymbol{U}$ in our factorisation.

$$\textbf{Step 2 removes } \begin{bmatrix} 0\ (\text{row 2 of } A_2) \\ 1\ (\text{row 2 of } A_2) \\ \ell_{32}(\text{row 2 of } A_2) \\ \ell_{42}(\text{row 2 of } A_2) \end{bmatrix} \textbf{ from } A_2 \textbf{ to leave } \boldsymbol{A_3} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \times & \times \\ 0 & 0 & \times & \times \end{bmatrix}.$$

Notice that both $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}_2$ are columns of $\boldsymbol{L}$. We removed a column $\boldsymbol{\ell}_2$ times the second pivot row. Continuing in the same way, we successively remove columns with each step removing a column $\boldsymbol{\ell}_j$ of $\boldsymbol{L}$ times a pivot row $\boldsymbol{u}_j$ of $\boldsymbol{U}$. See that this entire process can be depicted as

$$\boldsymbol{A = \ell_1 u_1 + \ell_2 u_2 + \cdots + \ell_n u_n} = \begin{bmatrix} \boldsymbol{\ell_1} & \cdots & \boldsymbol{\ell_n} \end{bmatrix} \begin{bmatrix} \boldsymbol{u_1} \\ \vdots \\ \boldsymbol{u_n} \end{bmatrix} = \boldsymbol{LU}$$
$$\textbf{columns times rows}$$

Notice that $\boldsymbol{U}$ is upper triangular; the pivot row u_k begins with $k-1$ zeros. $\boldsymbol{L}$ is lower triangular with 1's on the main diagonal. Column $\boldsymbol{\ell}_k$ also begins with $k-1$ zeros. (See that this proof makes use of the idea that matrix multiplication can be seen as a sum of rank 1 matrices)

2.6 Permutation matrices

Examples
Permutation matrices have a 1 in every row and a 1 in every column. All other entries are 0. When this matrix $\boldsymbol{P}$ multiplies a vector, it changes the order of its components; for instance:

$\boldsymbol{P}$ has the rows of $\boldsymbol{I}$
$\boldsymbol{Px}$ = Circular shift of $\boldsymbol{x}$
1, 2, 3 to 3, 1, 2

$$\boldsymbol{Px} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{x_3} \\ \boldsymbol{x_1} \\ \boldsymbol{x_2} \end{bmatrix}$$

(A nonzero ith entry on row m of $\boldsymbol{P}$ means that row i of $\boldsymbol{x}$ goes on that row m in the result) Other examples include

$\boldsymbol{Px}$ reverses the order to $\boldsymbol{x_4, x_3, x_2, x_1}$
$\boldsymbol{P^2 = I}$

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

$\boldsymbol{x_4}$ stays fixed
Input $\boldsymbol{x_1, x_2, x_3, x_4}$
output $\boldsymbol{x_3, x_1, x_2, x_4}$
$\boldsymbol{P^3 = I}$

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

Circular shift
Input $\boldsymbol{x_1, x_2, x_3, x_4}$
output $\boldsymbol{x_4, x_1, x_2, x_3}$
$\boldsymbol{P^4 = I}$

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

Evens before odds
$\boldsymbol{x_0, x_2}$ before $\boldsymbol{x_1, x_3}$
Fast Fourier Transform
Number from 0 to 3

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

Recall that elimination may require row exchanges. If $\boldsymbol{A}$ is invertible, then there is a permuation $\boldsymbol{P}$ to order its rows in advance, so that elimination on $\boldsymbol{PA}$ meets no zeros in the pivot positions. Then $\boldsymbol{PA = LU}$.

Inverse
We can intuit that the inverse of $\boldsymbol{P}$ is just its transpose:

The rows of any $\boldsymbol{P}$ are the columns of $\boldsymbol{P^{-1} = P^T}$
$\boldsymbol{P^T}$ = "transpose of $\boldsymbol{P}$"

$$\boldsymbol{P^T P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix} = \boldsymbol{I}$$

(For instance, a nonzero entry in row 1 column 2 means 'row 2 of $\boldsymbol{x}$ becomes row 1 of result'. Its transpose corresponds to a nonzero entry in row 2 column 1, so 'row 1 of $\boldsymbol{x}$ becomes row 2 in the result'—reversing the change. This logic can be extrapolated to every other row.)
(next page)

$PA = LU$ factorisation
Recall that elimination may require row exchanges; consider for instance

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & a \\ 2 & 4 & b \\ 3 & 7 & c \end{bmatrix} \overset{\boldsymbol{E}}{\rightarrow} \begin{bmatrix} 1 & 2 & a \\ 0 & \mathbf{0} & b-2a \\ 0 & \mathbf{1} & c-3a \end{bmatrix} \overset{\boldsymbol{P}}{\rightarrow} \begin{bmatrix} 1 & 2 & a \\ 0 & \mathbf{1} & c-3a \\ 0 & \mathbf{0} & b-2a \end{bmatrix} = \boldsymbol{U}$$

To rescue elimination, $\boldsymbol{P}$ exchanced row 2 with 3, bringing 1 to the second pivot so elimination could continue.

See that we could order the rows in advance, first exchanging rows 2 and 3 to get $\boldsymbol{PA}$, then $\boldsymbol{LU}$ factorisation becomes $\boldsymbol{PA} = \boldsymbol{LU}$; the matrix $\boldsymbol{PA}$ sails through elimination without seeing that zero pivot:

$$\underset{\boldsymbol{P}}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}} \underset{\boldsymbol{A}}{\begin{bmatrix} 1 & 2 & a \\ 2 & 4 & b \\ 3 & 7 & c \end{bmatrix}} = \underset{\boldsymbol{PA}}{\begin{bmatrix} 1 & 2 & a \\ 3 & 7 & c \\ 2 & 4 & b \end{bmatrix}} = \underset{\boldsymbol{L}}{\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}} \underset{\boldsymbol{U}}{\begin{bmatrix} 1 & 2 & a \\ 0 & 1 & c-3a \\ 0 & 0 & b-2a \end{bmatrix}}$$

We may require several row exchanges; in that case the an overall permutation $\boldsymbol{P}$ would include them all, still producing $\boldsymbol{PA} = \boldsymbol{LU}$. A useful way to keep track of the permutations might be to add a column of indices at the end of $\boldsymbol{A}$ so that the original indices are displayed:

$$\begin{bmatrix} 1 & 2 & a & \mathbf{1} \\ 2 & 4 & b & \mathbf{2} \\ 3 & 7 & c & \mathbf{3} \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & a & \mathbf{1} \\ 0 & \mathbf{0} & b-2a & \mathbf{2} \\ 0 & \mathbf{1} & c-3a & \mathbf{3} \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 2 & a & \mathbf{1} \\ 0 & \mathbf{1} & c-3a & \mathbf{3} \\ 0 & \mathbf{0} & b-2a & \mathbf{2} \end{bmatrix}. \ \boldsymbol{P_{132}} \text{ is } \begin{bmatrix} \mathbf{1} & 0 & 0 \\ 0 & 0 & \mathbf{1} \\ 0 & \mathbf{1} & 0 \end{bmatrix}$$

Column permutations
We know that we can reorder rows using

$$\boldsymbol{P} = \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} \end{bmatrix} \qquad \boldsymbol{PA} = \begin{bmatrix} a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{11} & a_{12} & a_{13} \end{bmatrix}$$

See that we can also reorder columns by applying a permutation matrix on the right:

$$\boldsymbol{Q} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} \end{bmatrix} \qquad \boldsymbol{PAQ} = \begin{bmatrix} a_{23} & a_{22} & a_{21} \\ a_{33} & a_{32} & a_{31} \\ a_{13} & a_{12} & a_{11} \end{bmatrix}$$

Note that these two operations don't cover all possible permutations of the 9 entries in $\boldsymbol{A}$. The first index is constant on every row and the second index is constant on every column.

2.7 Transposes and symmetric matrices

2.7.1 The transpose of AB is B^TA^T: Intuition

We have

$$(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$$

Consider $\boldsymbol{AB}$. Computing the first *row* of the result (the first column of $(\boldsymbol{AB})^T$); this can be seen as a linear combination of the *rows* of $\boldsymbol{B}$ (the columns of $\boldsymbol{B}^T$) specified by the entries in the first *row* of $\boldsymbol{A}$ (the first column of $\boldsymbol{A}^T$).

2.7.2 Showing $(A^T)^{-1} = (A^{-1})^T$

Consider taking the transpose of $\boldsymbol{A}^{-1}\boldsymbol{A}$,

Transpose of inverse $\boldsymbol{A^{-1}A = I}$ **is transposed to** $\boldsymbol{A^T(A^{-1})^T = I.}$

Similarly, transposing $\boldsymbol{AA}^{-1} = \boldsymbol{I}$ leads to $(\boldsymbol{A}^{-1})^T\boldsymbol{A}^T = \boldsymbol{I}$. Notice especially that $\boldsymbol{A}^T$ *is invertible exactly when* $\boldsymbol{A}$ *is invertible.*

2.7.3 A^TA is symmetric

Choose any matrix $\boldsymbol{A}$, probably rectangular. Multiply $\boldsymbol{A}^T$ times $\boldsymbol{A}$, then the product $\boldsymbol{S} = \boldsymbol{A}^T\boldsymbol{A}$ is automatically a square symmetric matrix:

The transpose of $\boldsymbol{A^TA}$ ***is*** $\boldsymbol{A^T(A^T)^T}$ ***which is*** $\boldsymbol{A^TA}$ ***again.***

The matrix $\boldsymbol{AA}^T$ is also symmetric (see that their shapes permit multiplication in either order), note howerver that $\boldsymbol{AA}^T$ *is a different matrix* from $\boldsymbol{A}^T\boldsymbol{A}$.

Chapter 3

The Four Fundamental Subspaces

3.1 Reduced Row Echelon Form

Motivation
Consider reducing a 2x4 matrix $\boldsymbol{A}$ to its reduced row echelon form $\boldsymbol{R}$:

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{11} & \mathbf{17} \\ \mathbf{3} & \mathbf{7} & \mathbf{37} & \mathbf{57} \end{bmatrix} \rightarrow \begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{11} & \mathbf{17} \\ \mathbf{0} & \mathbf{1} & \mathbf{4} & \mathbf{6} \end{bmatrix} \rightarrow \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{3} & \mathbf{5} \\ \mathbf{0} & \mathbf{1} & \mathbf{4} & \mathbf{6} \end{bmatrix} = \boldsymbol{R}$$

See how the added upward elimination essentially inverted the a portion $\boldsymbol{W}$ of $\boldsymbol{A}$:

$$\boldsymbol{W} = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix}$$

to turn that part of the matrix into the identity:

$$\textbf{Multiply } \boldsymbol{W^{-1}A} = \boldsymbol{W^{-1}} \begin{bmatrix} \boldsymbol{W} & \boldsymbol{H} \end{bmatrix} \textbf{ to produce } \boldsymbol{R} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{W^{-1}H} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{F} \end{bmatrix}.$$

We have $\boldsymbol{H} = \boldsymbol{WF}$, where $\boldsymbol{H}$ refers to the other parts of $\boldsymbol{A}$, and $\boldsymbol{W}$ the portion that was inverted—we can combine the components of $\boldsymbol{W}$ (the first independent columns) to produce the other (dependent) columns. The matrix $\boldsymbol{F}$ specifies the parameters to do this:

$$\begin{matrix} \textbf{Dependent} \\ \textbf{columns} \end{matrix} \quad \boldsymbol{H} = \begin{bmatrix} 11 & 17 \\ 37 & 57 \end{bmatrix} = \boldsymbol{WF} = \begin{bmatrix} 1 & 2 \\ 3 & 7 \end{bmatrix} \text{ times } \begin{bmatrix} 3 & 5 \\ 4 & 6 \end{bmatrix}.$$

See that the first r independent columns of $\boldsymbol{A}$ locate the columns of $\boldsymbol{R}$ containing $\boldsymbol{I}$. Also see that the last $m - r$ rows of $\boldsymbol{R}$ will be rows of zeros (the dependent columns in the reduced form can be expressed in terms of the identity contained within that form, so there can't be any additional nonzero rows).
(next page)

Algorithm
Elimination goes a column at a time from left to right; after k columns, that part of the matrix is in the reduced form, and we move to the $(k+1)$th column; this new column has an upper part $\boldsymbol{u}$ and a lower part $\boldsymbol{\ell}$:

$$\textbf{First } \boldsymbol{k+1} \textbf{ columns} \qquad \begin{bmatrix} I_k & F_k \\ 0 & 0 \end{bmatrix} P_k \text{ followed by } \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{\ell} \end{bmatrix}.$$

The idea here is to decide whether this $(k+1)$th column joins $\boldsymbol{I}_k$ or $\boldsymbol{F}_k$.

See that if $\boldsymbol{\ell}$ is *all zeros*, the new column is *dependent* on the first k columns; then $\boldsymbol{u}$ joins with $\boldsymbol{F}_k$ to produce $\boldsymbol{F}_{k+1}$.

If $\boldsymbol{\ell}$ is *not* all zero, then the column is *independent* of the first k columns. Pick any nonzero in $\boldsymbol{\ell}$ as the pivot, move that row of $\boldsymbol{A}$ up into row $k+1$, then subtract multiples of that pivot row to zero out all the rest of column $k+1$ (eliminate up and down). If necessary, divide the row by its first nonzero entry to have a pivot of 1. Column $k+1$ joins I_k (see that it adds a new nonzero row to the identity) to produce I_{k+1}.

Repeat for the next column and so on.

Row operations
There are three row operations allowed in elimination from $\boldsymbol{A}$ to $\boldsymbol{R}$:

1. Subtract a multiple of one row from another (below or above)
2. Divide a row by its first nonzero entry (to reach pivot 1)
3. Exchange rows (to move all zero rows to the bottom)

A different series of steps could be used reach the same $\boldsymbol{R}$. But the result $\boldsymbol{R}$ can't change.

3.2 $A = CR$ factorisation

We can apply elimination to reduce $\boldsymbol{A}$ to $\boldsymbol{R}_0$ (reduced echelon form with zero rows); then $\boldsymbol{I}$ in $\boldsymbol{R}_0$ locates the matrix $\boldsymbol{C}$ of *independent columns* in $\boldsymbol{A}$. Removing any zero rows from $\boldsymbol{R}_0$ produces the row matrix $\boldsymbol{R}$ such that $\boldsymbol{A} = \boldsymbol{CR}$. For instance,

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

where the independent columns of $\boldsymbol{A}$ are 1 and 3, and

$$\boldsymbol{R} = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 1 & 1 \\ 2 & 5 \\ 3 & 9 \end{bmatrix}$$

so

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 5 \\ 3 & 6 & 9 \end{bmatrix} = \boldsymbol{CR} = \begin{bmatrix} 1 & 1 \\ 2 & 5 \\ 3 & 9 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{matrix} \text{(column basis)} \\ \text{in } C \end{matrix} \quad \begin{matrix} \text{(row basis)} \\ \text{in } R \end{matrix}$$

3.3 Systematic nullspace computation

Elimination gives us a systematic way to find a basis for the nullspace. Say we have

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

Consider attempting to find a basis for the nullspace, meaning a basis for the solutions of $\boldsymbol{Ax}=\boldsymbol{0}$; say that $\boldsymbol{W^{-1}A}=\boldsymbol{R}_0$, then for some $\boldsymbol{x}$,

$$\boldsymbol{Ax}=\boldsymbol{0}\implies\boldsymbol{W^{-1}Ax}=\boldsymbol{W^{-1}0}\implies\boldsymbol{R_0x}=\boldsymbol{0}$$

See that the elimination process doesn't change the space of solutions that satisfy $\boldsymbol{Ax}=\boldsymbol{0}$. Also see that using $\boldsymbol{R}$ instead of $\boldsymbol{R}_0$ only removes the redundant zero rows without removing any solutions.

Given some $\boldsymbol{R}$, we have a systematic way to find a basis for the nullspace: considering all the indices of $\boldsymbol{x}$ corresponding to dependent indices, by letting one of those indices be 1 and the rest 0, we have a simple solution for $\boldsymbol{Rx}=\boldsymbol{0}$:

$$\boldsymbol{Rs_1}=\boldsymbol{0}\qquad\begin{bmatrix}1 & 7 & 0 & 8\\ 0 & 0 & 1 & 9\end{bmatrix}\begin{bmatrix}\boldsymbol{-7}\\ \boldsymbol{1}\\ \boldsymbol{0}\\ \boldsymbol{0}\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}\qquad\begin{matrix}\textbf{Put 1 and 0}\\ \textbf{in positions 2 and 4}\end{matrix}$$

$$\boldsymbol{Rs_2}=\boldsymbol{0}\qquad\begin{bmatrix}1 & 7 & 0 & 8\\ 0 & 0 & 1 & 9\end{bmatrix}\begin{bmatrix}\boldsymbol{-8}\\ \boldsymbol{0}\\ \boldsymbol{-9}\\ \boldsymbol{1}\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}\qquad\begin{matrix}\textbf{Put 0 and 1}\\ \textbf{in positions 2 and 4}\end{matrix}$$

See that if we were to write $\boldsymbol{R}$ as $[\boldsymbol{I},\boldsymbol{F}]$, the solutions can be found in an algorithmic manner:

The two special solutions to $\begin{bmatrix}I & F\end{bmatrix}\boldsymbol{x}=\boldsymbol{0}$ are the columns of $\begin{bmatrix}-F\\ I\end{bmatrix}$ in Example 1
The special solutions to $\begin{bmatrix}I & F\end{bmatrix}P\boldsymbol{x}=\boldsymbol{0}$ are the columns of $P^{\mathrm{T}}\begin{bmatrix}-F\\ I\end{bmatrix}$ in Example 2

The second case occurs should we have to permute $\boldsymbol{R}$ to reorder the independent columns. Recall that $\boldsymbol{PP}^T=\boldsymbol{I}$, and so

$$\boldsymbol{Rx}=\boldsymbol{0}\qquad\begin{bmatrix}I & F\end{bmatrix}P\ \text{times}\ P^{\mathrm{T}}\begin{bmatrix}-F\\ I\end{bmatrix}\ \text{reduces to}\ \begin{bmatrix}I & F\end{bmatrix}\begin{bmatrix}-F\\ I\end{bmatrix}=\begin{bmatrix}0\end{bmatrix}$$

(next page)

Cont.
To put these two ideas together, suppose the mxn matrix $\boldsymbol{A}$ has rank r. To find the $n-r$ special solutions to $\boldsymbol{Ax}=\boldsymbol{0}$, compute the reduced row echelon form $\boldsymbol{R}_0$ of $\boldsymbol{A}$; remove the $m-r$ zero rows of $\boldsymbol{R}_0$ to produce $\boldsymbol{R}=[\boldsymbol{I},\boldsymbol{F}]\boldsymbol{P}$, where $\boldsymbol{A}=\boldsymbol{CR}$.

Then the special solutions to $\boldsymbol{Ax}=0$ are the $n-r$ columns of $\boldsymbol{P}^T\begin{bmatrix}-\boldsymbol{F}\\ \boldsymbol{I}\end{bmatrix}$.

3.4 The complete solution to $Ax = b$

Finding a particular solution
We can reduce $\boldsymbol{Ax} = \boldsymbol{b}$ to a simpler system $\boldsymbol{R_0 x} = \boldsymbol{d}$ with the same solutions (if any). A useful aid here is to augment $\boldsymbol{A}$ with $\boldsymbol{b}$ to produce an augmented matrix $[\boldsymbol{A}, \boldsymbol{b}]$:

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 7 \end{bmatrix} \quad \text{has the augmented matrix} \quad \begin{bmatrix} 1 & 3 & 0 & 2 & 1 \\ 0 & 0 & 1 & 4 & 6 \\ 1 & 3 & 1 & 6 & 7 \end{bmatrix} = \begin{bmatrix} A & b \end{bmatrix}.$$

Applying elimination to reach $\boldsymbol{R_0}$, the augmented part also undergoes the same elimination steps to produce a $\boldsymbol{d}$:

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix} \quad \text{has the augmented matrix} \quad \boxed{\begin{bmatrix} 1 & 3 & 0 & 2 & 1 \\ 0 & 0 & 1 & 4 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} R_0 & d \end{bmatrix}.}$$

For an easy solution x_p we can *choose the free variables to be 0.* In this case $x_2 = x_4 = 0$, then the pivot variables can just be read off from $\boldsymbol{d}$:

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

Pivot variables 1, 6
Free variables 0, 0
$\boldsymbol{x_{\text{particular}} = (1, 0, 6, 0).}$

See that for a solution to exist, the zero rows in $\boldsymbol{R_0}$ must also be zero in $\boldsymbol{d}$.

This procedure allows us to come up with a particular solution. For a complete solution, we consider any sum of the particular solution x_p with any nullspace vectors x_n:

One $\boldsymbol{x_{\text{particular}}}$	*The particular solution solves*	$Ax_p = b$
All $\boldsymbol{x_{\text{nullspace}}}$	*The $n - r$ special solutions solve*	$Ax_n = 0$

Complete solution
one x_p + many x_n
particular x_p
nullspace vectors x_n

$$x = x_p + x_n = \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -2 \\ 0 \\ -4 \\ 1 \end{bmatrix}.$$

The nullspace can be found as per outlined earlier (setting one variable to one and the rest to zero to get each nullspace basis vector). See that the complete solution includes all nullspace basis vectors, and that if the nullspace is just the zero vector then every particular solution $\boldsymbol{x_p}$ is the only solution to that $\boldsymbol{b}$.
(next page)

3.4.1 Rank and solution

Full column rank

In the case where $\boldsymbol{A}$ has *full column rank*, every column has a pivot and the rank is $r = n$. The matrix is tall and thin ($m \geq n$). Row reduction puts $R = I$ at the top when $\boldsymbol{A}$ is reduced to $\boldsymbol{R}_0$ with rank n:

$$\textbf{Full column rank } \boldsymbol{r = n} \quad R_0 = \begin{bmatrix} I \\ 0 \end{bmatrix} = \begin{bmatrix} n \text{ by } n \text{ identity matrix} \\ m - n \text{ rows of zeros} \end{bmatrix}$$

See that in this case the matrix has the following properties:

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

1. **All columns of A are pivot columns (independent). No free variables.**
2. **The nullspace $\mathbf{N}(A)$ contains only the zero vector $\boldsymbol{x} = \mathbf{0}$.**
3. **If $A\boldsymbol{x} = \boldsymbol{b}$ has a solution (it might not) then it has only *one solution*.**

With full column rank, $\boldsymbol{Ax} = \boldsymbol{b}$ will have *one solution or no solution.* (think of a mapping from n dimensional space into m dimensional space where $m \geq n$, there aren't enough vectors in n space to encompass all of m space, so there may be no solution for a particular vector in m space; or see that the columns only span a n dimensional hyperplane in a higher m dimensional space—any m dimensional vector not on that hyperplane has no solution)

Full row rank

The other case is full row rank—now $\boldsymbol{Ax} = \boldsymbol{b}$ has *one or infinitely many solutions.* In this case $\boldsymbol{A}$ must be *short and wide* ($m \leq n$); the matrix has full row rank if $r = m$, where every row has a pivot.

(If $m < n$, then a nullspace consisting of more than just the zero vector exists and there will be multiple solutions to $\boldsymbol{Ax} = \boldsymbol{b}$.) See that a matrix like this has the following properties:

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

1. **All rows have pivots, and R_0 has no zero rows: $R_0 = R$.**
2. **$A\boldsymbol{x} = \boldsymbol{b}$ has a solution for every right side $\boldsymbol{b}$.**
3. **The column space of A is the whole space $\mathbf{R}^m$.**
4. **If $m < n$ the equation $A\boldsymbol{x} = \boldsymbol{b}$ is underdetermined (many solutions).**

(next page)

Cont.

The four possibilities for linear equations depend on the rank r

$r = m$	and	$r = n$	*Square and invertible*	$Ax = b$	has 1 solution
$r = m$	and	$r < n$	*Short and wide*	$Ax = b$	has ∞ solutions
$r < m$	and	$r = n$	*Tall and thin*	$Ax = b$	has 0 or 1 solution
$r < m$	and	$r < n$	*Not full rank*	$Ax = b$	has 0 or ∞ solutions

If the matrix $\boldsymbol{A}$ is square and invertible, it indicates a mapping from $m = n = r$ space into that same space; the basis specified by $\boldsymbol{A}$ spans the whole of r space and there is only one linear combination of that basis that solves for $\boldsymbol{Ax} = \boldsymbol{b}$ (since $\boldsymbol{A}$ has linearly independent columns and therefore a trivial nullspace).

If the matrix is not full rank, then the columns span a hyperplane in the space equal to the row length, meaning some $\boldsymbol{Ax} = \boldsymbol{b}$ may not have a solution, and if some $\boldsymbol{b}$ does lie within that hyperplane there will be infinite solutions to it (since the columns are not independent and therefore a nontrivial nullspace exists).

Their reduced row echelon form will fall into the following categories:

Four types for R_0	$\begin{bmatrix} I \end{bmatrix}$	$\begin{bmatrix} I & F \end{bmatrix}$	$\begin{bmatrix} I \\ 0 \end{bmatrix}$	$\begin{bmatrix} I & F \\ 0 & 0 \end{bmatrix}$
Their ranks	$r = m = n$	$r = m < n$	$r = n < m$	$r < m, r < n$

Cases 1 and 2 have full row rank, case 3 has full column rank but not row rank, case 4 is not full rank. See that in the last two cases, when $\boldsymbol{Ax} = \boldsymbol{b}$ is reduced to $\boldsymbol{R}_0\boldsymbol{x} = \boldsymbol{d}$, $\boldsymbol{d}$ must end in $m - r$ zeros for the equation to be solvable. $\boldsymbol{F}$ refers to the top part of the dependent columns.

3.5 Intuition for four subspaces

We have

	Four Fundamental Subspaces	Dimensions
1.	The *row space* is $\mathbf{C}(A^{\mathrm{T}})$, a subspace of $\mathbf{R}^n$.	r
2.	The *column space* is $\mathbf{C}(A)$, a subspace of $\mathbf{R}^m$.	r
3.	The *nullspace* is $\mathbf{N}(A)$, a subspace of $\mathbf{R}^n$.	$n-r$
4.	The *left nullspace* is $\mathbf{N}(A^{\mathrm{T}})$, a subspace of $\mathbf{R}^m$.	$m-r$

Column space of reduced form not the same as original matrix
Note that although the reduced row echelon form provides the pivot positions to determine the row and column rank of a matrix, while the *rows* of the rref form span the same row space as the original unreduced matrix, the columns of the rref form cannot act as substitutes for the original column space, instead only providing information on dimensionality and on which original columns act as basis vectors for the column space. That is

$$\boldsymbol{C}(\boldsymbol{A}) \neq \boldsymbol{C}(\boldsymbol{R}_0) \quad \text{but} \quad \boldsymbol{C}(\boldsymbol{A}^T) = \boldsymbol{C}(\boldsymbol{R}_0^T)$$

(none of the row operations in elimination change the row space, but they do change the column space)

Nullspace dimension is $n-r$
The intuition for the dimension of $\boldsymbol{N}(A)$ being $n-r$ is clear from elimination, where each dependent column provides a new basis vector for the nullspace (we can find a basis by substituting one free variable as 1 and the rest zero, repeating for all free variables), so the number of basis vectors equals the number of dependent columns equals total columns minus independent columns—so $n-r$.

Left nullspace dimensionality
We know the dimensions for $\boldsymbol{A}$ is the same as that of $\boldsymbol{A}^T$. Transposing $\boldsymbol{A}$ gives us m columns, of which we know r are independent. So there are $m-r$ dependent columns which each provide a basis vector for the nullspace.

3.6 $A = CR$ and Block elimination

The earlier notes on the rref form introduced it as a simple way to express the entire matrix in terms of its dependent rows, where for a matrix $\boldsymbol{A}$, with independent components $\boldsymbol{W}$ and dependent components $\boldsymbol{H}$, reduction to rref looks like:

$$\boldsymbol{A} = \begin{bmatrix} W & H \end{bmatrix} \implies \boldsymbol{W}^{-1}\boldsymbol{A} = \boldsymbol{W}^{-1}\begin{bmatrix} W & H \end{bmatrix} = \begin{bmatrix} I & W^{-1}H \end{bmatrix}$$

See that we can recover the dependent components of $\boldsymbol{H}$ with $\boldsymbol{W}^{-1}\boldsymbol{H}$ by multiplying it by the independent columns $\boldsymbol{W}$, so

$$\boldsymbol{A} = \underbrace{\boldsymbol{W}}_{\boldsymbol{C}}\underbrace{\begin{bmatrix} I & W^{-1}H \end{bmatrix}}_{\boldsymbol{R}}$$

where $\boldsymbol{C}$ means the same thing as $\boldsymbol{W}$, that is, the independent columns of $\boldsymbol{A}$. See that $\boldsymbol{W}$ is gauranteed to be square since row and column rank are equal.

Block elimination
The above doesn't show how the zero rows in $\boldsymbol{R}_0$ (if they exist) are redundant. We can develop a better idea of elimination if we consider working with blocks of the original $\boldsymbol{A}$. This is called block elimination; consider reducing a matrix $\boldsymbol{A}$ with rank r to rref using the following three steps:

Step 1: Exchange columns of $\boldsymbol{A}$ using $\boldsymbol{P}_C$ and exchange rows of $\boldsymbol{A}$ using $\boldsymbol{P}_R$ to put the r independent columns first and r independent rows first. This looks like

$$\boldsymbol{P}_R\boldsymbol{A}\boldsymbol{P}_C = \begin{bmatrix} \boldsymbol{W} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix}$$

$\begin{bmatrix} \boldsymbol{W} \\ \boldsymbol{J} \end{bmatrix}$ has full column rank and $\begin{bmatrix} \boldsymbol{W} & \boldsymbol{H} \end{bmatrix}$ has full row rank. See that $\boldsymbol{W}$ is a square matrix with full rank r and is therefore invertible.

Step 2: Multiply the top rows by $\boldsymbol{W}^{-1}$:

$$\begin{bmatrix} \boldsymbol{W} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} \rightarrow \begin{bmatrix} \boldsymbol{I} & \boldsymbol{W}^{-1}\boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix}$$

Step 3: Subtract $\boldsymbol{J}[\ \boldsymbol{I} \quad \boldsymbol{W}^{-1}\boldsymbol{H}\ ]$ from the $m-r$ rows $[\ \boldsymbol{J} \quad \boldsymbol{K}\ ]$ to produce $[\ \boldsymbol{0} \quad \boldsymbol{0}\ ]$:

$$\begin{bmatrix} \boldsymbol{I} & \boldsymbol{W}^{-1}\boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} \rightarrow \begin{bmatrix} \boldsymbol{I} & \boldsymbol{W}^{-1}\boldsymbol{H} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

To understand why $\boldsymbol{J}\boldsymbol{W}^{-1}\boldsymbol{H} = \boldsymbol{K}$, we know that the first r rows $[\ \boldsymbol{I} \quad \boldsymbol{W}^{-1}\boldsymbol{H}\ ]$ are linearly independent, and since $\boldsymbol{A}$ has rank r, the lower rows $[\ \boldsymbol{J} \quad \boldsymbol{K}\ ]$ must be combinations of the upper rows. These combinations must be given by $\boldsymbol{J}$ to get the first r columns correct: $\boldsymbol{J}\boldsymbol{I} = \boldsymbol{J}$. Then $\boldsymbol{J}$ times $\boldsymbol{W}^{-1}\boldsymbol{H}$ must equal $\boldsymbol{K}$ to make the last columns correct.
(next page)

Cont.
We have block elimination as the following transformations:

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

We can, in a manner similar to the steps taken, factor out parts of $\boldsymbol{A}$ as follows to acquire the reduced row echelon form; for step 2:

$$\boldsymbol{P}_R \boldsymbol{A} \boldsymbol{P}_C = \begin{bmatrix} \boldsymbol{W} & \boldsymbol{H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} = \begin{bmatrix} \boldsymbol{W} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{W^{-1}H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix}$$

For step 3:

$$\begin{bmatrix} \boldsymbol{W} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{W^{-1}H} \\ \boldsymbol{J} & \boldsymbol{K} \end{bmatrix} = \begin{bmatrix} \boldsymbol{W} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1} \end{bmatrix} \begin{bmatrix} \boldsymbol{1} & \boldsymbol{0} \\ \boldsymbol{J} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{W^{-1}H} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

The result is

$$\boldsymbol{P}_R \boldsymbol{A} \boldsymbol{P}_C = \begin{bmatrix} \boldsymbol{W} & \boldsymbol{0} \\ \boldsymbol{J} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{W^{-1}H} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{W} \\ \boldsymbol{J} \end{bmatrix} \begin{bmatrix} \boldsymbol{I} & \boldsymbol{W^{-1}H} \end{bmatrix}$$

See that this is just the $\boldsymbol{A} = \boldsymbol{CR}$ factorisation, but this time showing that the zero rows are redundant rather than just using a reduced form where zero rows don't exist. Another way of writing this result is

$$\boxed{\textbf{The conclusion is that } \boldsymbol{P_R A P_C} = \begin{bmatrix} \boldsymbol{W} \\ \boldsymbol{J} \end{bmatrix} \boldsymbol{W^{-1}} \begin{bmatrix} \boldsymbol{W} & \boldsymbol{H} \end{bmatrix} = \boldsymbol{CW^{-1}B}}$$

Chapter 4

Orthogonality

4.1 Orthogonal subspaces

Recall that two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ are orthogonal if $\boldsymbol{v}^T\boldsymbol{w} = 0$. To say that two subspaces are orthogonal means that every vector in one space is orthogonal to every vector in the other space.

We can make the claim that for any matrix $\boldsymbol{A}$, *the nullspace of* $\boldsymbol{A}$ *is orthogonal to the row space of* $\boldsymbol{A}$. To see this just consider $\boldsymbol{Ax} = \boldsymbol{0}$:

$$\boldsymbol{Ax} = \begin{bmatrix} \textbf{row 1 of } \boldsymbol{A} \\ \vdots \\ \textbf{row } \boldsymbol{m} \textbf{ of } \boldsymbol{A} \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{x} \\ \\ \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \vdots \\ \boldsymbol{0} \end{bmatrix} \quad \begin{matrix} \longleftarrow \ \textbf{(row 1)} \cdot \boldsymbol{x} \textbf{ is zero} \\ \\ \longleftarrow \ \textbf{(row } \boldsymbol{m}\textbf{)} \cdot \boldsymbol{x} \textbf{ is zero} \end{matrix}$$

Every row has a zero dot product with $\boldsymbol{x}$. Then every combination of the rows is perpendicular to $\boldsymbol{x}$. That is, the whole row space $\boldsymbol{C}(\boldsymbol{A}^T)$ is orthogonal to the nullspace $\boldsymbol{N}(\boldsymbol{A})$ in n dimensional space (since $\boldsymbol{A}^T$ has n rows). Another way to show this is

$$\boldsymbol{x} \textbf{ in nullspace orthogonal to } \boldsymbol{A}^{\mathbf{T}}\boldsymbol{y} \textbf{ in row space } \quad \boldsymbol{x}^{\mathbf{T}}(\boldsymbol{A}^{\mathbf{T}}\boldsymbol{y}) = (\boldsymbol{Ax})^{\mathbf{T}}\boldsymbol{y} = \boldsymbol{0}^{\mathbf{T}}\boldsymbol{y} = \boldsymbol{0}.$$

See that we can analagously reason that the *the column space* $\boldsymbol{C}(\boldsymbol{A})$ *and the left nullspace* $\boldsymbol{N}(\boldsymbol{A}^T)$ *are perpendicular inside* m *dimensional space.*
(next page)

Cont.
See that the only common vector between two orthogonal subspaces would be the zero vector (we can't sum the columns of a matrix to produce a vector orthogonal to all the columns). There is a very important restriction on the dimension of any two orthogonal subspaces:

If $\boldsymbol{V}$ and $\boldsymbol{W}$ are orthogonal subspaces in $\mathbf{R}^n$ then $\dim \boldsymbol{V} + \dim \boldsymbol{W} \leq \boldsymbol{n}$.

For instance, the row space of a matrix is a subspace of dimension r in n space, and its orthogonal complement, by construction, will have dimensionality $n-r$.

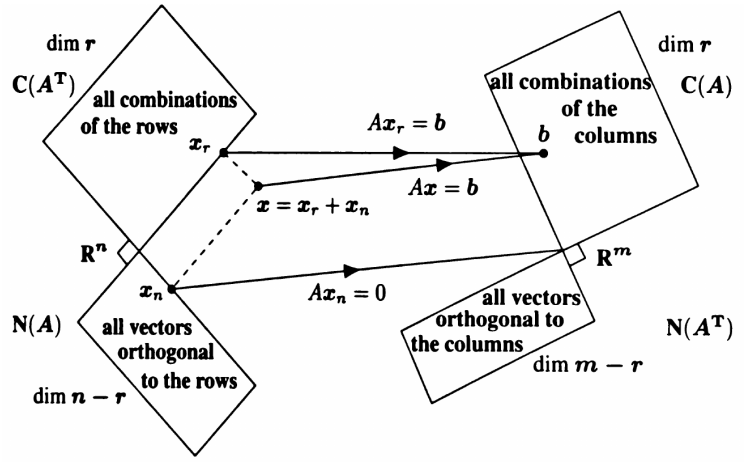
4.1.1 Orthogonal complements

When two orthogonal subspaces account for the whole space, they are called *orthogonal complements*. The orthogonal complement of $\boldsymbol{V}$ is written as $\boldsymbol{V}^{\perp}$. The row space and the nullspace, as well as the column space and left nullspace, are instances of orthogonal complements:

Orthogonal complements	Row space and nullspace	$\boldsymbol{r + (n - r) = n}$
	Column space and left nullspace	$\boldsymbol{r + (m - r) = m}$

4.1.2 Intuition for $\boldsymbol{x = x_{\text{row}} + x_{\text{null}}}$

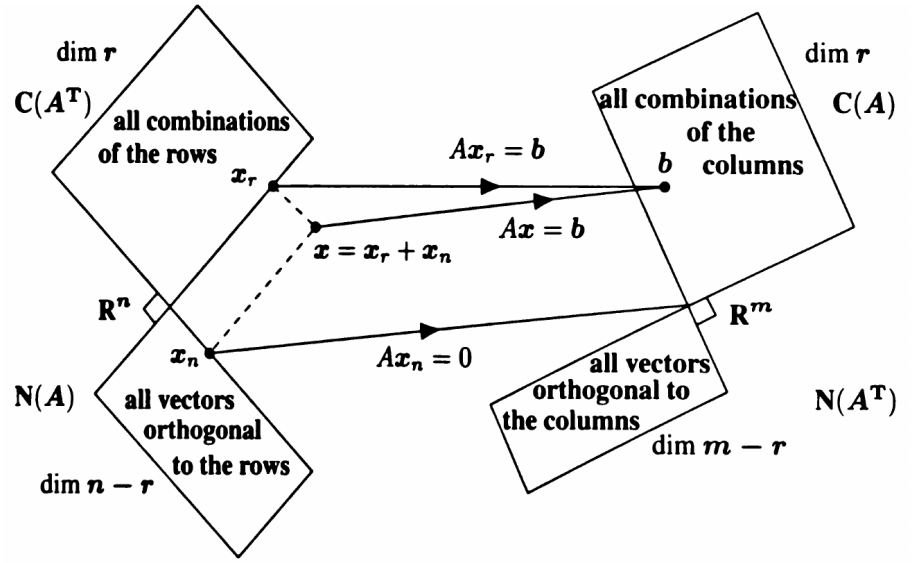
See that because the row space and the nullspace combined span the entire n space, we can say that any vector in $\mathbb{R}^n$ can be expressed as a sum of vectors from both subspaces. That is, any vector $\boldsymbol{x}$ in $\mathbb{R}^n$ is the sum $\boldsymbol{x} = \boldsymbol{x}_{\text{row}} + \boldsymbol{x}_{\text{null}}$ of its row space component and its nullspace component. Any vector $\boldsymbol{y}$ in $\mathbb{R}^m$ is the sum $\boldsymbol{y} = \boldsymbol{y}_{\text{col}} + \boldsymbol{y}_{\text{null}}$ of its column space component and its left nullspace component from $\boldsymbol{N}(\boldsymbol{A}^T)$:



(next page)

4.1.3 A is invertible from the row space to the column space

Consider again the illustration from the previous page:



Every vector in $\boldsymbol{Ax}$ is in the column space; that is obvious. Every $\boldsymbol{x}$ comes from $\mathbb{R}^n$ space, which is made up of the the row space $\boldsymbol{C}(\boldsymbol{A}^T)$ and its orthogonal complement, the nullspace $\boldsymbol{N}(\boldsymbol{A})$. In that context, every candidate $\boldsymbol{x}$ either comes from the nullspace, which would map to $\boldsymbol{0}$, or the row space, which would then map to some $\boldsymbol{b}$ in the column space—every $\boldsymbol{b}$ in the column space can be found by some $\boldsymbol{Ax}$ with $\boldsymbol{x}$ from the row space.

We can also go on to say that *every vector $\boldsymbol{b}$ in the column space comes from exactly one vector $\boldsymbol{x}_r$ in the row space.* Proof: If $\boldsymbol{Ax}_r = \boldsymbol{Ax}'_r$, the difference $\boldsymbol{x}_r - \boldsymbol{x}'_r$ is in the nullspace (since $\boldsymbol{A}(\boldsymbol{x}_r - \boldsymbol{x}'_r) = \boldsymbol{0}$); however, should $\boldsymbol{x}_r$ and $\boldsymbol{x}'_r$ come from the row space, then $\boldsymbol{x}_r - \boldsymbol{x}'_r$ would also have to be in the row space. The only way for a vector to be in both the nullspace and the row space would be if it were the zero vector; so $\boldsymbol{x}_r - \boldsymbol{x}'_r = \boldsymbol{0}$ and $\boldsymbol{x}_r = \boldsymbol{x}'_r$.

So $\boldsymbol{A}$ maps every row space component to a specific column space component; see that there is an r by r invertible matrix hiding inside $\boldsymbol{A}$ if we throw away the two nullspaces. *$\boldsymbol{A}$ is invertible from the row space to the column space.*

4.2 A^TA is invertible only if A has linearly independent columns

$\boldsymbol{A}^T\boldsymbol{A}$ is a square matrix (n by n). We prove the above by showing that $\boldsymbol{A}^T\boldsymbol{A}$ *has the same nullspace as* $\boldsymbol{A}$; so that when the columns of $\boldsymbol{A}$ are linearly independent, and its nullspace only contains the zero vector, then $\boldsymbol{A}^T\boldsymbol{A}$, with this same nullspace, is also invertible.

First we show that a vector in the nullspace of $\boldsymbol{A}$ is also in that of $\boldsymbol{A}^T\boldsymbol{A}$. Let $\boldsymbol{A}$ be any matrix; if $\boldsymbol{x}$ is in the nullspace, then $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$. Multiplying by $\boldsymbol{A}^T$ gives $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, so $\boldsymbol{x}$ is also in the nullspace of $\boldsymbol{A}^T\boldsymbol{A}$.

Now the other way, we show that a vector in the nullspace of $\boldsymbol{A}^T\boldsymbol{A}$ is in that of $\boldsymbol{A}$. Starting with $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, multiplying by $\boldsymbol{x}^T$:

$$(\boldsymbol{x}^\mathbf{T})\boldsymbol{A}^\mathbf{T}\boldsymbol{A}\boldsymbol{x} = 0 \quad \text{or} \quad (\boldsymbol{A}\boldsymbol{x})^\mathbf{T}(\boldsymbol{A}\boldsymbol{x}) = 0 \quad \text{or} \quad \|\boldsymbol{A}\boldsymbol{x}\|^\mathbf{2} = 0.$$

So if $\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ then $\boldsymbol{A}\boldsymbol{x}$ has length 0, so $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$.

Every vector $\boldsymbol{x}$ in one nullspace is in the other nullspace, so if $\boldsymbol{A}^T\boldsymbol{A}$ has dependent columns then so does $\boldsymbol{A}$, and if $\boldsymbol{A}^T\boldsymbol{A}$ has independent columns then so does $\boldsymbol{A}$:

When $\boldsymbol{A}$ has independent columns, $\boldsymbol{A}^\mathbf{T}\boldsymbol{A}$ is square, symmetric, and invertible.

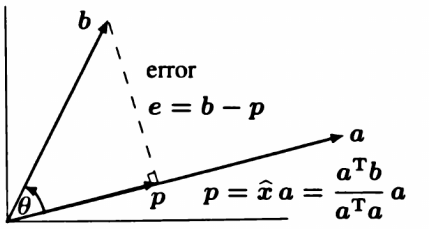
4.3 Projections

4.3.1 Projection onto a line

We want to project any $\boldsymbol{b}$ onto the column space of any m by n matrix $\boldsymbol{A}$. This entails finding the point $\boldsymbol{p}$ on that subspace that is closest to our arbitrary point $\boldsymbol{b}$. We start with a line.

Considering a one dimensional subspace (a line) with basis $\boldsymbol{a} = (a_1, \ldots, a_m)$. Along that line, we want the point $\boldsymbol{p}$ closest to $\boldsymbol{b} = (b_1, \ldots, b_m)$; the key to projection is orthogonality: *the line from* $\boldsymbol{b}$ *to* $\boldsymbol{p}$*, that is* $\boldsymbol{b} - \boldsymbol{p}$*, is perpendicular to the vector* $\boldsymbol{a}$.

We know that the projection $\boldsymbol{p}$ will be some multiple of $\boldsymbol{a}$; call it $\boldsymbol{p} = \hat{x}\boldsymbol{a}$ (so here $\hat{x}$ denotes a constant)—computing this number $\hat{\boldsymbol{x}}$ will give us the vector $\boldsymbol{p}$. We can call the perpendicular line $\boldsymbol{b} - \boldsymbol{p}$ the 'error' $\boldsymbol{e} = \boldsymbol{b} - \hat{x}\boldsymbol{a}$, and use the fact that it is perpendicular to $\boldsymbol{a}$ to derive $\hat{x}$:



Projecting $\boldsymbol{b}$ onto $\boldsymbol{a}$ with error $\boldsymbol{e} = \boldsymbol{b} - \hat{\boldsymbol{x}}\boldsymbol{a}$

$\boldsymbol{a} \cdot (\boldsymbol{b} - \hat{\boldsymbol{x}}\boldsymbol{a}) = 0$ or $\boldsymbol{a} \cdot \boldsymbol{b} - \hat{\boldsymbol{x}}\boldsymbol{a} \cdot \boldsymbol{a} = 0$

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

With that we can compute the projection $\boldsymbol{p} = \hat{x}\boldsymbol{a}$:

The projection of $\boldsymbol{b}$ onto the line through $\boldsymbol{a}$ is the vector $\boldsymbol{p} = \hat{\boldsymbol{x}}\boldsymbol{a} = \dfrac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}\boldsymbol{a}$.

Special case 1: If $\boldsymbol{b} = \boldsymbol{a}$ then $\hat{\boldsymbol{x}} = 1$. The projection of $\boldsymbol{a}$ onto $\boldsymbol{a}$ is itself. $P\boldsymbol{a} = \boldsymbol{a}$.

Special case 2: If $\boldsymbol{b}$ is perpendicular to $\boldsymbol{a}$ then $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b} = 0$. The projection is $\boldsymbol{p} = \boldsymbol{0}$.

(next page)

Projection matrix
We had the projection of $\boldsymbol{b}$ onto the line through $\boldsymbol{a}$ as the vector $\boldsymbol{p}$ given by

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

Now comes the *projection matrix*, where we want a matrix $\boldsymbol{P}$ that can be applied to the arbitrary point $\boldsymbol{b}$ to get the desired projection $\boldsymbol{p}$, so $\boldsymbol{Pb} = \boldsymbol{p}$. This can found easily; since $\hat{x}$ is a constant, we can rearrange the above equation:

Projection matrix P $\quad \boldsymbol{p} = \boldsymbol{a}\hat{\boldsymbol{x}} = \boldsymbol{a}\dfrac{\boldsymbol{a}^{\mathsf{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathsf{T}}\boldsymbol{a}} = \boldsymbol{Pb}$ *when the matrix is* $\boldsymbol{P} = \dfrac{\boldsymbol{a}\boldsymbol{a}^{\mathsf{T}}}{\boldsymbol{a}^{\mathsf{T}}\boldsymbol{a}}.$

The projection matrix $\boldsymbol{P}$ is a m by m rank one matrix. Notice that the line through $\boldsymbol{a}$, the subspace we are projecting onto, is the column space of $\boldsymbol{P}$.

4.3.2 Projection onto a subspace

Now we project onto a n-dimensional subspace of $\mathbb{R}^m$. We start with n vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ in $\mathbb{R}^m$, *assume that they are linearly independent.*

We want to find a projection of $\boldsymbol{b}$ onto the subspace spanned by these vectors, that is, *the combination* $\boldsymbol{p} = \hat{x}_1\boldsymbol{a}_1 + \cdots + \hat{x}_n\boldsymbol{a}_n$ *closest to a given vector* $\boldsymbol{b}$. See that this can be written as $\boldsymbol{A}\hat{\boldsymbol{x}}$, where $\hat{\boldsymbol{x}} = (\hat{x}_1, \ldots, \hat{x}_n)$.

As before, the idea here is that the error vector $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = \boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}$ is perpendicular to the subspace we are projecting onto. Every basis vector is perpendicular to this error vector, so

$$\begin{matrix} \boldsymbol{a}_1^{\mathsf{T}}(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}) = 0 \\ \vdots \\ \boldsymbol{a}_n^{\mathsf{T}}(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}) = 0 \end{matrix} \quad \text{or} \quad \begin{bmatrix} — \, \boldsymbol{a}_1^{\mathsf{T}} \, — \\ \vdots \\ — \, \boldsymbol{a}_n^{\mathsf{T}} \, — \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{0} \\ \\ \end{bmatrix} \quad \text{or} \quad \boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^{\mathsf{T}}\boldsymbol{b}.$$

We have

$$\boldsymbol{A}^T(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}) = \boldsymbol{0} \implies \boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$$

(next page)

Cont.
We had the projection of $\boldsymbol{b}$ onto the column space of $\boldsymbol{A}$ as $\boldsymbol{p} = \boldsymbol{A}\hat{\boldsymbol{x}}$. We showed that

$$\boldsymbol{A}^T(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}) = \boldsymbol{0} \implies \boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$$

With this we can derive the projection matrix:

The combination $\boldsymbol{p} = \hat{x}_1\boldsymbol{a}_1 + \cdots + \hat{x}_n\boldsymbol{a}_n = A\hat{\boldsymbol{x}}$ that is closest to $\boldsymbol{b}$ comes from $\hat{\boldsymbol{x}}$:

$$\textbf{Find } \hat{\boldsymbol{x}}\,(n \times 1) \quad A^{\mathrm{T}}(\boldsymbol{b} - A\hat{\boldsymbol{x}}) = \boldsymbol{0} \quad \text{or} \quad A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}. \tag{5}$$

This symmetric matrix $A^{\mathrm{T}}A$ is n by n. It is invertible if the $\boldsymbol{a}$'s are independent. The solution is $\hat{\boldsymbol{x}} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b}$. The *projection* of $\boldsymbol{b}$ onto the subspace is $\boldsymbol{p}$:

$$\textbf{Find } \boldsymbol{p}\,(m \times 1) \qquad \boldsymbol{p} = A\hat{\boldsymbol{x}} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b}. \tag{6}$$

The *projection matrix* $\boldsymbol{P}$ is multiplying $\boldsymbol{b}$ in (6). It has four $\boldsymbol{A}$'s:

$$\textbf{Find } \boldsymbol{P}\,(m \times m) \qquad \boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}. \tag{7}$$

Again see that this only works if $\boldsymbol{A}^T\boldsymbol{A}$ is invertible, which is only when the column vectors of $\boldsymbol{A}$ are independent. See that the final projection matrix formula is analagous to the formula derived in the one dimensional case.

The subspace we are projecting onto is the the column space of $\boldsymbol{A}$, that is, $\boldsymbol{C}(\boldsymbol{A})$. See that the error $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}$ *belongs to the perpendicular subspace* $\boldsymbol{N}(\boldsymbol{A}^T)$ (since $\boldsymbol{A}^T(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}) = \boldsymbol{0}$).

A note on the projection matrix
The matrix $\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$ is deceptive. One might try to split $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ into $\boldsymbol{A}^{-1}$ times $(\boldsymbol{A}^T)^{-1}$ and find that the entire formula reduces to $\boldsymbol{I}$.

This is wrong because *the matrix $\boldsymbol{A}$ is rectangular and therefore has no inverse.* We cannot split $(\boldsymbol{A}^T\boldsymbol{A})^{-1}$ into $\boldsymbol{A}^{-1}$ times $(\boldsymbol{A}^T)^{-1}$ because there is no $\boldsymbol{A}^{-1}$ in the first place.

■ REVIEW OF THE KEY IDEAS ■

1. The projection of $\boldsymbol{b}$ onto the line through $\boldsymbol{a}$ is $\boldsymbol{p} = \boldsymbol{a}\hat{\boldsymbol{x}} = \boldsymbol{a}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a})$.
2. The rank one projection matrix $P = \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}/\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ multiplies $\boldsymbol{b}$ to produce $\boldsymbol{p}$.
3. Projecting $\boldsymbol{b}$ onto a subspace leaves $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ perpendicular to the subspace.
4. When A has full rank n, the equation $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ leads to $\hat{\boldsymbol{x}}$ and $\boldsymbol{p} = A\hat{\boldsymbol{x}}$.
5. The projection matrix $P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}$ has $P^{\mathrm{T}} = P$ and $P^2 = P$ and $P\boldsymbol{b} = \boldsymbol{p}$.

4.4 Projection matrices are symmetrical

Given a projection matrix $\boldsymbol{P}$, we can decompose any vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ into

$$\begin{aligned}\boldsymbol{v} &= \boldsymbol{v}_p + \boldsymbol{v}_n \\ \boldsymbol{w} &= \boldsymbol{w}_p + \boldsymbol{w}_n\end{aligned}$$

With the p subscript indicating the component of the vector in the subspace of $\boldsymbol{P}$, and the n subscript the component outside (normal to) the subspace of $\boldsymbol{P}$. (the sum of the projection and its orthogonal 'error' make up the original vector before projecion)

The projection of a vector lies in a subspace. The dot produc of anything in this subspace with anything orthogonal to this subspace is zero. So see that

$$\begin{aligned}(\boldsymbol{P}\boldsymbol{v}) \cdot \boldsymbol{w} &= \boldsymbol{v}_p \cdot \boldsymbol{w} \\ &= \boldsymbol{v}_p \cdot (\boldsymbol{w}_p + \boldsymbol{w}_n) \\ &= \boldsymbol{v}_p \cdot \boldsymbol{w}_p + \boldsymbol{v}_p \cdot \boldsymbol{w}_n \\ &= \boldsymbol{v}_p \cdot \boldsymbol{w}_p\end{aligned}$$

and similarly

$$\begin{aligned}\boldsymbol{v} \cdot (\boldsymbol{P}\boldsymbol{w}) &= \boldsymbol{v} \cdot \boldsymbol{w}_p \\ &= (\boldsymbol{v}_p + \boldsymbol{v}_n) \cdot \boldsymbol{w}_p \\ &= \boldsymbol{v}_p \cdot \boldsymbol{w}_p + \boldsymbol{v}_n \cdot \boldsymbol{w}_p \\ &= \boldsymbol{v}_p \cdot \boldsymbol{w}_p\end{aligned}$$

As such we can conclude that

$$(\boldsymbol{P}\boldsymbol{v}) \cdot \boldsymbol{w} = \boldsymbol{v} \cdot (\boldsymbol{P}\boldsymbol{w})$$

and using the definition of the dot product:

$$\begin{aligned}(\boldsymbol{P}\boldsymbol{v})^T \boldsymbol{w} &= \boldsymbol{v}^T \cdot (\boldsymbol{P}\boldsymbol{v}) \\ \boldsymbol{v}^T \boldsymbol{P}^T \boldsymbol{w} &= \boldsymbol{v}^T \boldsymbol{P}\boldsymbol{w}\end{aligned}$$

so

$$\boldsymbol{P}^T = \boldsymbol{P}$$

4.5 Least Squares Approximations

Illustrative example
Consider trying to find the closest line to three points $(0,6),(1,0),(2,0)$. No straight line $\boldsymbol{b} = C + D\boldsymbol{t}$ goes through those three points; we want two numbers C and D that satisfy a system of three equations (this leads us to a 3 by 2 matrix):

$$\begin{array}{lll} \boldsymbol{t=0} & \textbf{The first point is on the line } b = C + Dt \textbf{ if} & \boldsymbol{C + D \cdot 0 = 6} \\ \boldsymbol{t=1} & \textbf{The second point is on the line } b = C + Dt \textbf{ if} & \boldsymbol{C + D \cdot 1 = 0} \\ \boldsymbol{t=2} & \textbf{The third point is on the line } b = C + Dt \textbf{ if} & \boldsymbol{C + D \cdot 2 = 0.} \end{array}$$

(This is essentially sampling a straight line at the three time points and comparing it to the desired data points at those times) This 3 by 2 system has *no solution*: $\boldsymbol{b} = (6,0,0)$ is not a combination of the columns $(1,1,1)$ and $(0,1,2)$ (of the matrix $\boldsymbol{A}$):

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

The idea here is to choose a vector of parameters $\boldsymbol{x}$ that makes $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{Ax}$ as small as possible—so finding a point in the column space of $\boldsymbol{A}$ that is *closest to* $\boldsymbol{b}$—see that this is just the projection. So the best $\boldsymbol{x}$, which we denote as $\hat{\boldsymbol{x}}$, is found by $\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$, where the projection $\boldsymbol{p}$ of $\boldsymbol{b}$ onto the column space of $\boldsymbol{A}$ is denoted by $\boldsymbol{p} = \boldsymbol{A}\hat{\boldsymbol{x}}$:

$$\boldsymbol{Ax = b = p + e} \quad \textbf{is impossible} \qquad \boldsymbol{A\hat{x} = p} \quad \textbf{is solvable} \qquad \boldsymbol{\hat{x}} \quad \textbf{is } (\boldsymbol{A^TA})^{-1}\boldsymbol{A^Tb}.$$

See that the solution to $\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{p}$ results in the least possible squared error:

$$\textbf{Squared error for any } \boldsymbol{x} \qquad \boldsymbol{\|Ax - b\|^2 = \|Ax - p\|^2 + \|e\|^2.}$$

Since any $\boldsymbol{x}$ other than $\hat{\boldsymbol{x}}$ would lead to $||\boldsymbol{Ax} - \boldsymbol{p}||^2 > 0$. (This expression comes from the pythagorean theorem, where the vector from some point $\boldsymbol{Ax}$ to point $\boldsymbol{b}$, that is $\boldsymbol{Ax} - \boldsymbol{b}$, whose length we want to minimise, is the hypotenuse, while the the error $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$, perpendicular to the vector $\boldsymbol{Ax} - \boldsymbol{p}$, are the opposite and adjacent sides respectively)

By choosing $\boldsymbol{x} = \hat{\boldsymbol{x}}$, we leave the smallest possible error $\boldsymbol{e} = (e_1, e_2, e_3)$ that we can't reduce, while reducing $\boldsymbol{Ax} - \boldsymbol{p}$ to zero. There is some ambiguity in the idea of 'smallest', but here it means that we minimise the *squared length of* $\boldsymbol{Ax} - \boldsymbol{b}$:

> ***The least squares solution $\hat{\boldsymbol{x}}$ makes $\boldsymbol{E} = \boldsymbol{\|Ax - b\|^2}$ as small as possible.***

(next page)

Cont.
By taking $\boldsymbol{x} = \hat{\boldsymbol{x}}$, where $\boldsymbol{A}\hat{\boldsymbol{x}}$ corresponds to the projection of $\boldsymbol{b}$ onto the column space of $\boldsymbol{A}$, we have the $\boldsymbol{x}$ that minimises the squared error $||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}||^2$:

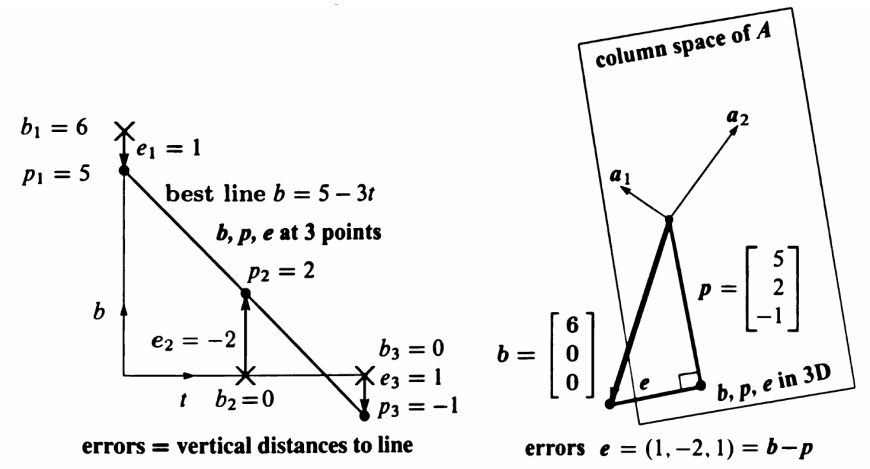


Figure 4.6: **Best line and projection: Two pictures, same problem.** The line has heights $\boldsymbol{p} = (5, 2, -1)$ with errors $\boldsymbol{e} = (1, -2, 1)$. The equations $A^{\mathrm{T}}A\hat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}$ give $\hat{\boldsymbol{x}} = (5, -3)$. Same answer! The best line is $b = 5 - 3t$ and the closest point is $\boldsymbol{p} = 5\boldsymbol{a}_1 - 3\boldsymbol{a}_2$.

Projection allows us to minimise squared error.

Intuition from calculus
Consider minimising the squared error by calculus. The total squared error of the system, with parameters C and D, is

$$\boldsymbol{E} = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 = (C + D \cdot 0 - 6)^2 + (C + D \cdot 1)^2 + (C + D \cdot 2)^2.$$

We take the partial derivative with respect to each parameter and set them to 0:

$$\boldsymbol{\partial E/\partial C} = 2(C + D \cdot 0 - 6) \quad + 2(C + D \cdot 1) \quad + 2(C + D \cdot 2) \quad = 0$$

$$\boldsymbol{\partial E/\partial D} = 2(C + D \cdot 0 - 6)(\mathbf{0}) + 2(C + D \cdot 1)(\mathbf{1}) + 2(C + D \cdot 2)(\mathbf{2}) = 0.$$

Simplifying leaves us with the system

$$\begin{array}{ll} \text{The } C \text{ derivative is zero:} & 3C + 3D = 6 \\ \text{The } D \text{ derivative is zero:} & 3C + 5D = 0 \end{array} \quad \textbf{This matrix } \begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \textbf{ is } \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$$

This system is exactly $\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$, where optimisation leads us to the exact same system we derived by projection:

$$\boxed{\textit{The partial derivatives of } \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|^2 \textit{ are zero when } \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.}$$

(next page)

The big picture for least squares

The intuition for the four subspaces can be applied here. Given some $\boldsymbol{b}$ that isn't in the column space of $\boldsymbol{A}$, we can split it into the sum of its projection $\boldsymbol{p}$ in the column space and the minimum error $\boldsymbol{e}$ in the orthogonal subspace; that is, we have $\boldsymbol{b} = \boldsymbol{p} + \boldsymbol{e}$:

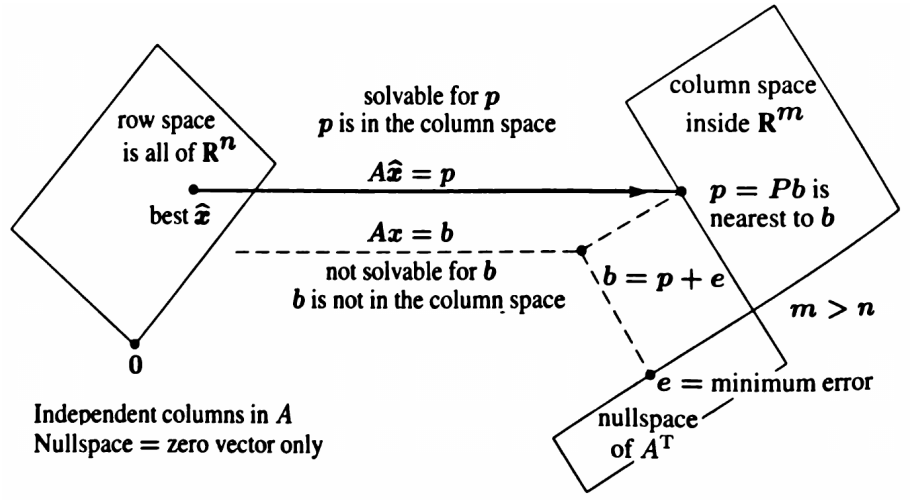


Figure 4.7: The projection $\boldsymbol{p} = A\hat{\boldsymbol{x}}$ is closest to $\boldsymbol{b}$, so $\hat{\boldsymbol{x}}$ minimizes $E = \|\boldsymbol{b} - A\boldsymbol{x}\|^2$.

As discussed earlier, see that the row space has a solution for any vector in the column space of $\boldsymbol{A}$, meaning that it contains $\hat{x}$. Also notice that since we construct $\boldsymbol{A}$ as having full column rank, its nullspace will just be the zero vector. (next page)

4.5.1 Fitting a straight line

Fitting a line is the clearest application of least squares. It starts with $m > 2$ points $b_1, \ldots, b_m$, at times $t_1, \ldots, t_m$. The best fit line $C+Dt$ misses the point by vertical distances $e_1, \ldots, e_m$. The least squares line minimises $E = e_1^2 + \cdots + e_m^2$ (note that this is not the minimum error as in the previous pages).

A line goes through all m points when we can exactly solve $\boldsymbol{Ax} = \boldsymbol{b}$; generally this isn't possible so we want a best fit. Two unknowns C and D determine a line, so $\boldsymbol{A}$ only has $n = 2$ columns. To fit m points, we are trying to solve m equations:

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

The column space is so thin that $\boldsymbol{b}$ is almost certainly outside of it (if $\boldsymbol{b}$ happens to lie in the column space then $\boldsymbol{b} = \boldsymbol{p}$ and the errors are $\boldsymbol{e} = \boldsymbol{0}$). The parameters $\hat{\boldsymbol{x}}$ for the line that minimises the squared error $||\boldsymbol{Ax} - \boldsymbol{b}||^2$ is found by the projection:

The closest line $C + Dt$ has heights $p_1, \ldots, p_m$ with errors $e_1, \ldots, e_m$.
Solve $A^{\mathrm{T}}A\hat{x} = A^{\mathrm{T}}b$ for $\hat{x} = (C, D)$. The errors are $e_i = b_i - C - Dt_i$.

See that the matrix $\boldsymbol{A}^T\boldsymbol{A}$ looks like

$$\textbf{Dot-product matrix} \quad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix} = \begin{bmatrix} m & \sum t_i \\ \sum t_i & \sum t_i^2 \end{bmatrix}.$$

and $\boldsymbol{A}^T\boldsymbol{b}$,

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = \begin{bmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_m \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \sum b_i \\ \sum t_i b_i \end{bmatrix}.$$

The line $C + D\boldsymbol{t}$ minimises the error $\boldsymbol{E}(x) = e_1^2 + \cdots + e_m^2 = ||\boldsymbol{Ax} - \boldsymbol{b}||^2$ when $\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$:

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

$$\boldsymbol{E}(\boldsymbol{x}) = \|\boldsymbol{Ax} - \boldsymbol{b}\|^2 = (C + Dt_1 - b_1)^2 + \cdots + (C + Dt_m - b_m)^2.$$

4.6 Orthonormal Bases

The vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are *orthogonal* when their dot products $\boldsymbol{q}_1 \cdots \boldsymbol{q}_j$ are zero—$\boldsymbol{q}_i^T \boldsymbol{q}_j = 0$ whenever $i \neq j$. By *dividing each vector by its length*, the vectors become orthogonal *unit vectors*—their lengths are all 1—in this case the basis is called *orthonormal*:

DEFINITION The n vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ are ***orthonormal*** if

$$\boldsymbol{q}_i^{\mathrm{T}} \boldsymbol{q}_j = \begin{cases} 0 & \text{when } i \neq j \quad (\textbf{\textit{orthogonal}} \text{ vectors}) \\ 1 & \text{when } i = j \quad (\textbf{\textit{unit}} \text{ vectors: } \|\boldsymbol{q}_i\| = 1) \end{cases}$$

A matrix Q with orthonormal columns has $Q^{\mathrm{T}}Q = I$. Typically $m > n$.

The matrix $\boldsymbol{Q}$ is easy to work with because $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$ (since each column is orthonormal):

$$\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \begin{bmatrix} -\,\boldsymbol{q}_1^{\mathrm{T}}\,- \\ -\,\boldsymbol{q}_2^{\mathrm{T}}\,- \\ \\ -\,\boldsymbol{q}_n^{\mathrm{T}}\,- \end{bmatrix} \begin{bmatrix} | & | & & | \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & & \boldsymbol{q}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \boldsymbol{I}$$

On the diagonal the unit vectors give $\boldsymbol{q}_i^T \boldsymbol{q}_i = ||\boldsymbol{q}_i||^2 = 1$. Note an important implication:

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

Note that generally if $m \neq n$ then $\boldsymbol{Q}\boldsymbol{Q}^T \neq \boldsymbol{I}$; although $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$ even when $\boldsymbol{Q}$ is rectangular, $\boldsymbol{Q}^T$ here is only a 1-sided inverse from the left. Only for square matrices do we also have $\boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I}$ (from the definition of the inverse of square matrices as covered earlier).

(The matrix must have full column rank since if its columns were to be orthonormal; so we have $m \geq n$. A mapping by $\boldsymbol{Q}$ in the case where $m > n$ is invertible since it has full column rank, but the transpose (size nxm) doesn't have full column rank and therefore a specific mapping $\boldsymbol{Q}^T\boldsymbol{x} = \boldsymbol{b}$ would have infinite solutions—a $\boldsymbol{x}$ cannot be exactly recovered given a $\boldsymbol{Q}^T$ and $\boldsymbol{b}$. As such only applying $\boldsymbol{Q}$ first works.)

If we have a *square* $\boldsymbol{Q}$ with orthonormal columns, its transpose is its inverse from both sides; and as such also has orthonormal columns—both the rows and columns of $\boldsymbol{Q}$ are orthonormal. A square matrix with orthonormal columns is called an *orthogonal matrix*.

Also see that if the columns are only orthogonal and not orthonormal, their dot products still give a diagonal matrix.
(next page)

4.6.1 Multiplication by any orthogonal matrix preserves length and angle

Length is preserved during multiplication with an orthogonal matrix:

Proof $\|Q\boldsymbol{x}\|^2$ **equals** $\|\boldsymbol{x}\|^2$ **because** $(Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}}Q^{\mathrm{T}}Q\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}I\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$.

If Q has orthonormal columns $(Q^{\mathrm{T}}Q = I)$***, it leaves lengths unchanged:***

Same length for $Q\boldsymbol{x}$ $\quad \|Q\boldsymbol{x}\| = \|\boldsymbol{x}\|$ **for every vector** $\boldsymbol{x}$. (3)

The dot product, and therefore the angle, is also preserved:

$$\boldsymbol{Q}\textbf{ also preserves dot products: } (Q\boldsymbol{x})^{\mathrm{T}}(Q\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{T}}Q^{\mathrm{T}}Q\boldsymbol{y} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}.$$

4.6.2 Projections using orthonormal bases

Using orthonormal bases for projection greatly simplifies computation. For projections onto subspaces, all formulas involve $\boldsymbol{A}^T\boldsymbol{A}$; suppose the basis vectors are orthonormal, then $\boldsymbol{A}^T\boldsymbol{A}$ simplifies to $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$. Our projection becomes

$$\hat{\boldsymbol{x}} = Q^{\mathrm{T}}\boldsymbol{b} \quad \text{and} \quad \boldsymbol{p} = Q\hat{\boldsymbol{x}} \quad \text{and} \quad P = QQ^{\mathrm{T}}. \tag{4}$$

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

See that no inversion is required—this is the point of an orthonormal basis—the best $\hat{\boldsymbol{x}} = \boldsymbol{Q}^T\boldsymbol{b}$ is just dot products of $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ with $\boldsymbol{b}$:

Projection onto q's

$$\boldsymbol{p} = \begin{bmatrix} | & & | \\ \boldsymbol{q}_1 & \cdots & \boldsymbol{q}_n \\ | & & | \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b} \\ \vdots \\ \boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b} \end{bmatrix} = \boldsymbol{q}_1(\boldsymbol{q}_1^{\mathrm{T}}\boldsymbol{b}) + \cdots + \boldsymbol{q}_n(\boldsymbol{q}_n^{\mathrm{T}}\boldsymbol{b}).$$

Also see that when $\boldsymbol{Q}$ is square (so an orthonormal basis spanning the entire space), $\hat{\boldsymbol{x}} = \boldsymbol{Q}^T\boldsymbol{b}$ leads to $\boldsymbol{p} = \boldsymbol{Q}\hat{\boldsymbol{x}} = \boldsymbol{Q}\boldsymbol{Q}^T\boldsymbol{b} = \boldsymbol{I}\boldsymbol{b} = \boldsymbol{b}$. The projection of $\boldsymbol{b}$ onto the basis that spans the entire space is $\boldsymbol{b}$ itself. (See that $\boldsymbol{Q}\boldsymbol{Q}^T = \boldsymbol{I}$ only works for square matrices)

4.7 Gram-Schmidt Process

Given three *independent vectors* a, b, c, we intend to construct three *orthogonal* vectors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$. Then, by dividing $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ by their lengths, we can produce three orthonormal vectors $\boldsymbol{q}_1 = \boldsymbol{A}/\|\boldsymbol{A}\|$, $\boldsymbol{q}_2 = \boldsymbol{B}/\|\boldsymbol{B}\|$, $\boldsymbol{q}_3 = \boldsymbol{C}/\|\boldsymbol{C}\|$.

Begin by choosing $\boldsymbol{A} = \boldsymbol{a}$ (we make all other vectors orthogonal to this). The next vector $\boldsymbol{B}$ must be perpendicular to $\boldsymbol{A}$; we *start with* $\boldsymbol{b}$ *and subtract its projection along* $\boldsymbol{A}$. This leaves the perpendicular part, which is the orthogonal vector $\boldsymbol{B}$:

First Gram-Schmidt step
$\boldsymbol{B}$ is perpendicular to $\boldsymbol{A}$

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

We use the one dimensional projection of $\boldsymbol{b}$ onto $\boldsymbol{a}$ to produce a vector orthogonal to $\boldsymbol{a}$; then we use it as our next basis vector.

The third direction starts with $\boldsymbol{c}$. We can say that $\boldsymbol{c}$ is not a comination of $\boldsymbol{A}$ and $\boldsymbol{B}$ because $\boldsymbol{c}$ is not a combination of $\boldsymbol{a}$ and $\boldsymbol{b}$ (see that B is just a combination of $\boldsymbol{b}$ and $\boldsymbol{A} = \boldsymbol{a}$ since $\hat{\boldsymbol{x}}$ in the one dimensional case is just a constant). It is also unlikely that $\boldsymbol{c}$ is already perpendicular to $\boldsymbol{A}$ and $\boldsymbol{B}$; the idea here is that we subtract off its components in those two directions to get a perpendicular direction $\boldsymbol{C}$:

Next Gram-Schmidt step
$\boldsymbol{C}$ is perpendicular to $\boldsymbol{A}$ and $\boldsymbol{B}$

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

This is the entire idea behind the Gram-Schmidt process; *Subtract from every vector its projections in the directions already set*, repeating for each new basis vector. At the end, or immediately when each one is found, dividing the orthogonal vectors $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ by their lengths leaves us with *orthonormal* results $\boldsymbol{q}_i$

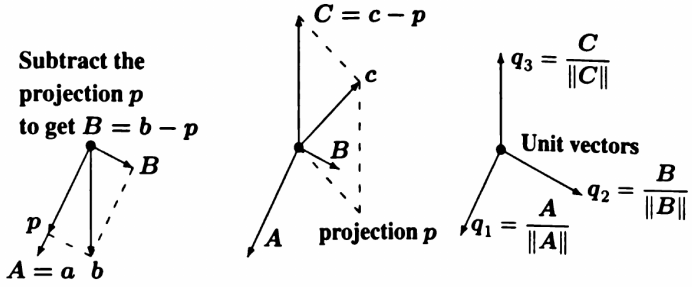


Figure 4.10: First project $\boldsymbol{b}$ onto the line through $\boldsymbol{a}$ and find the orthogonal $\boldsymbol{B} = \boldsymbol{b} - \boldsymbol{p}$. Then project $\boldsymbol{c}$ to $\boldsymbol{p}$ on the $\boldsymbol{AB}$ plane and find $\boldsymbol{C} = \boldsymbol{c} - \boldsymbol{p}$. Divide by $\|\boldsymbol{A}\|, \|\boldsymbol{B}\|, \|\boldsymbol{C}\|$.

4.8 $A = QR$ factorisation

Using the Gram-Schmidt process we can derive an orthonormal matrix $\boldsymbol{Q}$ with columns $\boldsymbol{q}_1, \boldsymbol{q}_2$, and $\boldsymbol{q}_3$ from a matrix $\boldsymbol{A}$ with columns $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$. The vectors $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ come from linear combinations of $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ and vice versa. As such there must be a third matrix that maps $\boldsymbol{Q}$ to $\boldsymbol{A}$.

Matrix $\boldsymbol{A}$ has independent columns, and so any mapping using its column space has a one to one mapping; a mapping to $\boldsymbol{Q}$ in the same dimension suggests a *square invertible matrix* applied to the right of $\boldsymbol{A}$, lets call it $\boldsymbol{R}^{-1}$; so $\boldsymbol{A}\boldsymbol{R}^{-1} = \boldsymbol{Q}$ (the linear combinations of the columns of $\boldsymbol{A}$ to produce $\boldsymbol{Q}$ must themselves be linearly independent). This square invertible matrix has an inverse $\boldsymbol{R}$, which when applied to the right of this equation, gives us

$$\boldsymbol{A}\boldsymbol{R}^{-1}\boldsymbol{R} = \boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$$

so we can say that there is some $\boldsymbol{R}$ connecting the orthonormal $\boldsymbol{Q}$ back to $\boldsymbol{A}$.

To find this $\boldsymbol{R}$, we simply multiply the entire expression by $\boldsymbol{Q}^T$:

$$\boldsymbol{Q}^T\boldsymbol{A} = \boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{R} = \boldsymbol{R}$$

So we have $\boldsymbol{R} = \boldsymbol{Q}^T\boldsymbol{A}$:

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{q}_1^T \\ \boldsymbol{q}_2^T \\ \boldsymbol{q}_3^T \end{bmatrix} \begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1^T\boldsymbol{a} & \boldsymbol{q}_1^T\boldsymbol{b} & \boldsymbol{q}_1^T\boldsymbol{c} \\ \boldsymbol{q}_2^T\boldsymbol{a} & \boldsymbol{q}_2^T\boldsymbol{b} & \boldsymbol{q}_2^T\boldsymbol{c} \\ \boldsymbol{q}_3^T\boldsymbol{a} & \boldsymbol{q}_3^T\boldsymbol{b} & \boldsymbol{q}_3^T\boldsymbol{c} \end{bmatrix}$$

This can be simplified. We know that

- The vectors $\boldsymbol{a}$ and $\boldsymbol{A}$ and $\boldsymbol{q}_1$ are all along a single line
- The vectors $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{q}_1$, $\boldsymbol{q}_2$ are all in the same plane
- The vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ and $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ and $\boldsymbol{q}_1$, $\boldsymbol{q}_2$, $\boldsymbol{q}_3$ are all in the same 3 dimensional subspace

The point here is that q_2 is orthogonal to that entire first line, and q_3 is orthogonal to that entire second plane. Looking back at $\boldsymbol{R}$, see that $\boldsymbol{q}_2^T\boldsymbol{a} = 0$ by orthogonality, as well as $\boldsymbol{q}_3^T\boldsymbol{a}$ and $\boldsymbol{q}_3^T\boldsymbol{b}$. Our simplified matrix looks like

$$\begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^T\boldsymbol{a} & \boldsymbol{q}_1^T\boldsymbol{b} & \boldsymbol{q}_1^T\boldsymbol{c} \\ & \boldsymbol{q}_2^T\boldsymbol{b} & \boldsymbol{q}_2^T\boldsymbol{c} \\ & & \boldsymbol{q}_3^T\boldsymbol{c} \end{bmatrix} \quad \text{or} \quad \boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}.$$

(next page)

Cont.
We had

$$\begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} & \boldsymbol{c} \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^T\boldsymbol{a} & \boldsymbol{q}_1^T\boldsymbol{b} & \boldsymbol{q}_1^T\boldsymbol{c} \\ & \boldsymbol{q}_2^T\boldsymbol{b} & \boldsymbol{q}_2^T\boldsymbol{c} \\ & & \boldsymbol{q}_3^T\boldsymbol{c} \end{bmatrix} \quad \text{or} \quad \boldsymbol{A} = \boldsymbol{QR}.$$

Intuitively, at every step $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ are combinations of $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k$, and so column i of $\boldsymbol{R}$, which specifies $\boldsymbol{a}_i$ as a combination of the columns of $\boldsymbol{Q}$, will only have nonzero entries in the rows corresponding to the columns of $\boldsymbol{Q}$ that contribute to that $\boldsymbol{a}_i$.

See that this leads to $\boldsymbol{R}$ being *triangular*—any $\boldsymbol{a}_j$ only depends on the $\boldsymbol{q}_i$ where $i < j$—so all the entries below the jth row in $\boldsymbol{R}$ will be zero.

(Gram-Schmidt) From independent vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, Gram-Schmidt constructs orthonormal vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$. The matrices with these columns satisfy $\boldsymbol{A} = \boldsymbol{QR}$. Then $\boldsymbol{R} = \boldsymbol{Q}^T\boldsymbol{A}$ *is upper triangular* because later $\boldsymbol{q}$'s are orthogonal to earlier $\boldsymbol{a}$'s.

In the context of Least Squares
See that using our factorisation:

$$\boldsymbol{A}^T\boldsymbol{A} = (\boldsymbol{QR})^T\boldsymbol{QR} = \boldsymbol{R}^T\boldsymbol{Q}^T\boldsymbol{QR} = \boldsymbol{R}^T\boldsymbol{R}$$

the least squares formula

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

simplifies to

$$\begin{aligned} \boldsymbol{R}^T\boldsymbol{R}\hat{\boldsymbol{x}} &= (\boldsymbol{QR})^T\boldsymbol{b} \\ \boldsymbol{R}^T\boldsymbol{R}\hat{\boldsymbol{x}} &= \boldsymbol{R}^T\boldsymbol{Q}^T\boldsymbol{b} \\ \boldsymbol{R}\hat{\boldsymbol{x}} = \boldsymbol{Q}^T\boldsymbol{b} \quad &\text{or} \quad \hat{\boldsymbol{x}} = \boldsymbol{R}^{-1}\boldsymbol{Q}^T\boldsymbol{b} \end{aligned}$$

($\boldsymbol{R}$ is invertible since it has full column rank. It has full row rank too so $\boldsymbol{R}^T$ also has full column rank and is invertible (justifying the last step))

This allows us to solve for the least squares solution $\hat{\boldsymbol{x}}$ using back substitution (which is very fast). The real cost is the computation for Gram-Schmidt, which is needed to construct the orthogonal $\boldsymbol{Q}$ and the triangular $\boldsymbol{R} = \boldsymbol{Q}^T\boldsymbol{A}$.
(next page)

An algorithm
Here is an algorithm for the Gram-Schmidt process:

```
for j = 1:n                    % modified Gram-Schmidt in n loops
   v = A(:,j);                 % v begins as column j of the original A
   for i = 1:j-1               % columns q_1 to q_{j-1} are already settled in Q
      R(i,j) = Q(:,i)' * v;    % compute R_ij = q_i^T a_j which is q_i^T v
      v = v - R(i,j)*Q(:,i);   % subtract the projection (q_i^T v) q_i
   end                         % v is now perpendicular to all of q_1, ..., q_{j-1}
   R(j,j) = norm(v);           % the diagonal entries R_jj in R are lengths
   Q(:,j) = v/R(j,j);          % divide v by its length to get the next q_j
end                            % for j = 1 : n produces all of the q_j
```

Which essentially executes

$$r_{kj} = \sum_{i=1}^{m} q_{ik} v_{ij} \text{ and } v_{ij} = v_{ij} - q_{ik} r_{kj} \text{ and } r_{jj} = \left(\sum_{i=1}^{m} v_{ij}^2 \right)^{1/2} \text{ and } q_{ij} = \frac{v_{ij}}{r_{jj}}.$$

Where r denotes entries of $\boldsymbol{R}$ and $\boldsymbol{v}$ a column in the original $\boldsymbol{A}$.

The diagonal entries of $\boldsymbol{R}$ are just the norms of the unnormalised projections in the Gram-schmidt process. For instance

$$\boldsymbol{q}_2^T \boldsymbol{b} = \boldsymbol{q}_2 \cdot \boldsymbol{b} = \cos(\theta) ||\boldsymbol{q}_2|| \, ||\boldsymbol{b}|| = \cos(\theta) ||\boldsymbol{b}|| = ||\boldsymbol{B}||$$

θ, the angle between $\boldsymbol{b}$ and its projection $\boldsymbol{q}_2$, is the same angle between $\boldsymbol{b}$ and its unnormalised projection $\boldsymbol{B}$.

4.9 Pseudoinverses

Left Inverse
Consider a 'tall and thin' matrix $\boldsymbol{A}$, where $m > n$ with rank $r = n$. A matrix like this only has a *left sided* inverse (consider that $\boldsymbol{x}^T\boldsymbol{A}$ would produce a matrix whose rows are combinations of linearly dependent rows—this cannot be 'undone' to recover $\boldsymbol{x}^T$ given a result. On the other hand $\boldsymbol{A}\boldsymbol{x}$ can be 'undone' since this way the mapping $\boldsymbol{A}$ is unique to each $\boldsymbol{x}$.)

In this case $\boldsymbol{A}$ has independent columns, so $\boldsymbol{A}^T\boldsymbol{A}$ (nxn) is invertible, and

$$(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}$$

We say that

$$\boldsymbol{A}_{\text{left}}^{-1} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$$

is the *left inverse* of $\boldsymbol{A}$ (there may be other left inverses), where $\boldsymbol{A}_{\text{left}}^{-1}\boldsymbol{A} = \boldsymbol{I}$.

Right inverse
In another case consider if $\boldsymbol{A}$ were 'short and fat', where $m < n$ with rank $r = m$. A matrix like this only has a right sided inverse. $\boldsymbol{A}$ has independent rows, so now $\boldsymbol{A}\boldsymbol{A}^T$ is invertible and

$$\boldsymbol{A}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1} = \boldsymbol{I}$$

A nice *right inverse* is therefore:

$$\boldsymbol{A}_{\text{right}}^{-1} = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}$$

where $\boldsymbol{A}\boldsymbol{A}_{\text{right}}^{-1} = \boldsymbol{I}$.

Left and right inverses as projections
If $\boldsymbol{A}$ has full column rank and $\boldsymbol{A}_{\text{left}}^{-1} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$, applying the left inverse on the right leads us to

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

Is just a matrix which projects $\mathbb{R}^m$ onto the *column space* of $\boldsymbol{A}$.

Similarly, if $\boldsymbol{A}$ has full row rank then

$$\boldsymbol{A}_{\text{right}}^{-1}\boldsymbol{A} = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{A} = \boldsymbol{P}$$

Is just a matrix that projects $\mathbb{R}^n$ onto the *row space* of $\boldsymbol{A}$.
(next page)

Intuition for the pseudoinverse

The remaining case to consider is a matrix $\boldsymbol{A}$ for which $r < n$ and $r < m$. The vector $\boldsymbol{Ax}$ is always in the column space of $\boldsymbol{A}$; recall that the correspondence between vectors $\boldsymbol{x}$ in the (r dimensional) row space and vectors $\boldsymbol{Ax}$ in the (r dimensional) column space is one-to-one—meaning that if $\boldsymbol{x} \neq \boldsymbol{y}$ are vectors in the row space of $\boldsymbol{A}$ then $\boldsymbol{Ax} \neq \boldsymbol{Ay}$ in the column space of $\boldsymbol{A}$—we can show this.

Proving that if $\boldsymbol{x} \neq \boldsymbol{y}$ *then* $\boldsymbol{Ax} \neq \boldsymbol{Ay}$: Suppose that the statement is false, then we can find $\boldsymbol{x} \neq \boldsymbol{y}$ in the row space of $\boldsymbol{A}$ for which $\boldsymbol{Ax} = \boldsymbol{Ay}$. But then $\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{y}) = \boldsymbol{0}$ and $\boldsymbol{x} - \boldsymbol{y}$ is in the nullspace of $\boldsymbol{A}$; however since the row space is closed under linear combinations (including subtraction), $\boldsymbol{x} - \boldsymbol{y}$ is also in the row space of $\boldsymbol{A}$. The only vector in both the nullspace and row space is the zero vector, so $\boldsymbol{x} - \boldsymbol{y} = \boldsymbol{0}$, and the two vectors must be equal.

We can therefore conclude that the mapping $\boldsymbol{x} \mapsto \boldsymbol{Ax}$ from the row space to the column space is *invertible*. The inverse of this operation is called the *pseudoinverse*.

Chapter 5

Determinants

5.1 Properties of determinants

We can derive a number of ideas given the following three properties of determinants:

1. The n by n identity matrix has $\det \boldsymbol{I} = 1$.

2. Exchanging two rows of A reverses $\det A$ to $-\det \boldsymbol{A}$.

3. If row 1 of A is a combination $c\boldsymbol{v} + d\boldsymbol{w}$, then add 2 determinants:

$$\det \begin{bmatrix} c\boldsymbol{v} + d\boldsymbol{w} \\ \text{row } 2 \\ \cdot\cdot \\ \text{row } n \end{bmatrix} = c \det \begin{bmatrix} \boldsymbol{v} \\ \text{row } 2 \\ \cdot\cdot \\ \text{row } n \end{bmatrix} + d \det \begin{bmatrix} \boldsymbol{w} \\ \text{row } 2 \\ \cdot\cdot \\ \text{row } n \end{bmatrix}$$

For a nice geometric way to think of the determinant, consider a unit cube in N dimensional space: the set of N vectors of length 1 with coordinates 0 or 1 in each spot. The determinant of the linear transformation (matrix) $\boldsymbol{T}$ is the *signed volume of the region after applying $\boldsymbol{T}$ to the unit cube.*

If we apply the identity to the unit cube, we get back the unit cube, with volume 1 (the first property). If we stretch the cube by a constant factor in one direction only, the new volume is that constant; if we stack two blocks together aligned on the same direction, their combined volume is the sum of their volumes—showing that the signed volume is linear in each coordinate (property 3).
(next page)

Cont.—linearity in all rows separately
The third property essentially states that the determinant is linear with respect to row 1; then property 2 leads us further: $\det(\boldsymbol{A})$ is linear *to every row separately*:

$$\begin{aligned}
\det\begin{bmatrix} \text{row } 1 \\ c\boldsymbol{v}+d\boldsymbol{w} \\ \cdots \\ \text{row } n \end{bmatrix} &= -\det\begin{bmatrix} c\boldsymbol{v}+d\boldsymbol{w} \\ \text{row } 1 \\ \cdots \\ \text{row } n \end{bmatrix} \\
&= -c\det\begin{bmatrix} \boldsymbol{v} \\ \text{row } 1 \\ \cdots \\ \text{row } n \end{bmatrix} - d\det\begin{bmatrix} \boldsymbol{w} \\ \text{row } 1 \\ \cdots \\ \text{row } n \end{bmatrix} \\
&= c\det\begin{bmatrix} \text{row } 1 \\ \boldsymbol{v} \\ \cdots \\ \text{row } n \end{bmatrix} + d\det\begin{bmatrix} \text{row } 1 \\ \boldsymbol{w} \\ \cdots \\ \text{row } n \end{bmatrix}
\end{aligned}$$

5.1.1 If $\boldsymbol{A}$ has two equal rows, its determinant is 0; row operations don't change the determinant

Consider property 2; if $\boldsymbol{A}$ were to have two equal rows, exchanging them would have no effect on $\boldsymbol{A}$. Then $\det(\boldsymbol{A})$=$-\det(\boldsymbol{A})$, and $\det(\boldsymbol{A})$ must be zero.

Now consider this with property 3, and see that *subtracting d times row i from row j leaves det($\boldsymbol{A}$) unchanged*:

$$\det\begin{bmatrix} \text{row } 1 \\ \text{row } 2 - d\,\text{row } 1 \\ \cdots \\ \text{row } n \end{bmatrix} = \det\begin{bmatrix} \text{row } 1 \\ \text{row } 2 \\ \cdots \\ \text{row } n \end{bmatrix} - d\det\begin{bmatrix} \text{row } 1 \\ \text{row } 1 \\ \cdots \\ \text{row } n \end{bmatrix} = \det\begin{bmatrix} \text{row } 1 \\ \text{row } 2 \\ \cdots \\ \text{row } n \end{bmatrix}$$

See that this means that the elimination steps (excluding row exchanges) *does not change the determinant*, and elimination (without row exchanges) is the way to simplify the determinant of a matrix.

5.1.2 Upper triangular and diagonal matrices, Transpose

Finally, see that given any *upper triangular matrix*, row exchanges are not required and upward elimination will reduce the matrix to its diagonal. As such, given an upper triangular matrix, we can just immediately multiply the diagonal entries to find the determinant:

Triangular matrix
Diagonal matrix

$$\det \begin{bmatrix} a & b & c \\ & q & r \\ & & z \end{bmatrix} = \det \begin{bmatrix} a & & \\ & q & \\ & & z \end{bmatrix} = aqz$$

Now see that transposing an invertible matrix won't change the diagonals, and if the matrix is invertible its transposed form is guaranteed to have an upper triangular form (taking into account row exchanges). So by row reductions we can obtain the same diagonal matrix as if we had performed elimination on the orignal matrix:

$$\det(\boldsymbol{A}) = \det(\text{rref form}) = \det(\boldsymbol{A}^T)$$

and so we can conclude that transposing a matrix $\boldsymbol{A}$ doesn't change the determinant:

Transpose the matrix $\qquad \det(\boldsymbol{A}^{\mathbf{T}}) = \det(\boldsymbol{A})$

5.1.3 $\det(\boldsymbol{AB}) = (\det \boldsymbol{A})(\det \boldsymbol{B})$

For a singular matrix $\boldsymbol{A}$, recall that it can be reduced to an identity matrix by elimination matrices (all of which are invertible); then by using the inverse (which is just the reverse of each row operation) of each elimination matrix, we can express $\boldsymbol{A}$ as

$$\boldsymbol{A} = \boldsymbol{E}_k \boldsymbol{E}_{k-1} \cdots \boldsymbol{E}_1$$

These matrices are also called *elementary row matrices*, which are just identity matrices on which exactly *one elementary row operation has been performed*. These elementary row operations are the same as those executed during elimination—scaling a row, adding a scaled row to another row, and exchanging rows.

We can write

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{E}_k \boldsymbol{E}_{k-1} \cdots \boldsymbol{E}_1 \boldsymbol{B})$$

See that any elementary row operation just scales the determinant by a constant. Scaling a single row scales the entire determinant by that constant, exchanging rows negates the determinant, and adding (or subtracting) a scaled row to (or from) another row doesn't change the determinant. As such,

$$\det(\boldsymbol{E}_k \boldsymbol{E}_{k-1} \cdots \boldsymbol{E}_1 \boldsymbol{B}) = c_k c_{k-1} \cdots c_1 \det(\boldsymbol{B}) = \alpha \det(\boldsymbol{B})$$

where c denotes the constants multiplied to the determinant as a result of the row operations.

(next page)

Cont.
We had

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{E}_k\boldsymbol{E}_{k-1}\cdots\boldsymbol{E}_1\boldsymbol{B}) = c_k c_{k-1}\cdots c_1 \det(\boldsymbol{B}) = \alpha\det(\boldsymbol{B})$$

Then by setting $\boldsymbol{B}$ to $\boldsymbol{I}$:

$$\det(\boldsymbol{AI}) = \det(\boldsymbol{E}_k\boldsymbol{E}_{k-1}\cdots\boldsymbol{E}_1\boldsymbol{I}) = c_k c_{k-1}\cdots c_1 \det(\boldsymbol{I}) = \alpha\det(\boldsymbol{I})$$
$$\det(\boldsymbol{A}) = \alpha$$

and

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$$

5.1.4 Orthogonal matrices have determinant 1 or -1

We know that

$$\det(\boldsymbol{A}) = \det(\boldsymbol{A}^T)$$

and that

$$\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$$

Given an orthogonal matrix $\boldsymbol{Q}$, we know

$$\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$$

so

$$\det(\boldsymbol{Q}^T\boldsymbol{Q}) = \det(\boldsymbol{Q}^T)\det(\boldsymbol{Q}) = \det(\boldsymbol{I}) = 1$$

and since the determinant doesn't change under transposition:

$$(\det(\boldsymbol{Q}))^2 = 1$$

and $\det(\boldsymbol{Q})$ is ± 1.

5.1.5 Matrix factorisations and permutation matrix

Recall the matrix factorisation $\boldsymbol{A} = \boldsymbol{LU}$. $\boldsymbol{L}$ is invertible with 1s along the main diagonal, $\boldsymbol{U}$ is upper triangular. In this case

$$\det\boldsymbol{A} = (\det\boldsymbol{L})(\det\boldsymbol{U}) = \det\boldsymbol{U}$$

Where the determinant can be found by just multiplying down the diagonal.

Similarly, we could have $\boldsymbol{PA} = \boldsymbol{LU}$ due to row exchanges being required before elimination. We know that the permutation matrix $\boldsymbol{P}$ is just the identity with row exchanges applied to it, and so we just have to factor in $\det\boldsymbol{P}$, which will be equal to 1 or -1 depending on how many row exchanges occur.

Elimination, followed by multiplying down the diagonal, is how determinants are computed by virtually all computer systems for linear algebra.

5.2 Intuition for determinant formula

Consider a 3 by 3 determinant of the following matrix

$$\det \begin{bmatrix} a & b & c \\ p & q & r \\ x & y & z \end{bmatrix}$$

By linearity,

$$\det \begin{bmatrix} a & b & c \\ p & q & r \\ x & y & z \end{bmatrix} = \det \begin{bmatrix} a & 0 & 0 \\ p & q & r \\ x & y & z \end{bmatrix} + \det \begin{bmatrix} 0 & b & 0 \\ p & q & r \\ x & y & z \end{bmatrix} + \det \begin{bmatrix} 0 & 0 & c \\ p & q & r \\ x & y & z \end{bmatrix}$$

Looking closer at the first matrix on the right side, again by linearity,

$$\begin{aligned} \det \begin{bmatrix} a & 0 & 0 \\ p & q & r \\ x & y & z \end{bmatrix} &= \det \begin{bmatrix} a & 0 & 0 \\ p & 0 & 0 \\ x & y & z \end{bmatrix} + \det \begin{bmatrix} a & 0 & 0 \\ 0 & q & 0 \\ x & y & z \end{bmatrix} + \det \begin{bmatrix} a & 0 & 0 \\ 0 & 0 & r \\ x & y & z \end{bmatrix} \\ &= \det \begin{bmatrix} a & 0 & 0 \\ p & 0 & 0 \\ x & 0 & 0 \end{bmatrix} + \det \begin{bmatrix} a & 0 & 0 \\ p & 0 & 0 \\ 0 & y & 0 \end{bmatrix} + \det \begin{bmatrix} a & 0 & 0 \\ p & 0 & 0 \\ 0 & 0 & z \end{bmatrix} + \\ &\quad \det \begin{bmatrix} a & 0 & 0 \\ 0 & q & 0 \\ x & 0 & 0 \end{bmatrix} + \det \begin{bmatrix} a & 0 & 0 \\ 0 & q & 0 \\ 0 & y & 0 \end{bmatrix} + \det \begin{bmatrix} a & 0 & 0 \\ 0 & q & 0 \\ 0 & 0 & z \end{bmatrix} + \\ &\quad \det \begin{bmatrix} a & 0 & 0 \\ 0 & 0 & r \\ x & 0 & 0 \end{bmatrix} + \det \begin{bmatrix} a & 0 & 0 \\ 0 & 0 & r \\ 0 & y & 0 \end{bmatrix} + \det \begin{bmatrix} a & 0 & 0 \\ 0 & 0 & r \\ 0 & 0 & z \end{bmatrix} \end{aligned}$$

Looking at that final equality, see that the matrices containing a zero column are immediately noninvertible and have determinant 0. This leaves us with the simplification

$$\det \begin{bmatrix} a & 0 & 0 \\ p & q & r \\ x & y & z \end{bmatrix} = \det \begin{bmatrix} a & 0 & 0 \\ 0 & q & 0 \\ 0 & 0 & z \end{bmatrix} + \det \begin{bmatrix} a & 0 & 0 \\ 0 & 0 & r \\ 0 & y & 0 \end{bmatrix}$$

The other two matrices on the right side of that first equation can be expanded in a similar manner,

$$\det \begin{bmatrix} 0 & b & 0 \\ p & q & r \\ x & y & z \end{bmatrix} = \det \begin{bmatrix} 0 & b & 0 \\ p & 0 & 0 \\ 0 & 0 & z \end{bmatrix} + \det \begin{bmatrix} 0 & b & 0 \\ 0 & 0 & r \\ x & 0 & 0 \end{bmatrix}$$

$$\det \begin{bmatrix} 0 & 0 & c \\ p & q & r \\ x & y & z \end{bmatrix} = \det \begin{bmatrix} 0 & 0 & c \\ 0 & q & 0 \\ x & 0 & 0 \end{bmatrix} + \det \begin{bmatrix} 0 & 0 & c \\ p & 0 & 0 \\ 0 & y & 0 \end{bmatrix}$$

(next page)

Cont.
As per our reasoning, we can write the determinant of a 3 by 3 matrix as

$$\det\begin{bmatrix} a & b & c \\ p & q & r \\ x & y & z \end{bmatrix} = \det\begin{bmatrix} a & 0 & 0 \\ 0 & q & 0 \\ 0 & 0 & z \end{bmatrix} + \det\begin{bmatrix} a & 0 & 0 \\ 0 & 0 & r \\ 0 & y & 0 \end{bmatrix} +$$
$$\det\begin{bmatrix} 0 & b & 0 \\ p & 0 & 0 \\ 0 & 0 & z \end{bmatrix} + \det\begin{bmatrix} 0 & b & 0 \\ 0 & 0 & r \\ x & 0 & 0 \end{bmatrix} +$$
$$\det\begin{bmatrix} 0 & 0 & c \\ 0 & q & 0 \\ x & 0 & 0 \end{bmatrix} + \det\begin{bmatrix} 0 & 0 & c \\ p & 0 & 0 \\ 0 & y & 0 \end{bmatrix}$$

This is the sum of the determinants of the following matrices

$$\begin{bmatrix} \boldsymbol{a} & & \\ & \boldsymbol{q} & \\ & & \boldsymbol{z} \end{bmatrix}\begin{bmatrix} & \boldsymbol{b} & \\ \boldsymbol{p} & & \\ & & \boldsymbol{z} \end{bmatrix}\begin{bmatrix} & \boldsymbol{b} & \\ & & \boldsymbol{r} \\ \boldsymbol{x} & & \end{bmatrix}\begin{bmatrix} & & \boldsymbol{c} \\ & \boldsymbol{q} & \\ \boldsymbol{x} & & \end{bmatrix}\begin{bmatrix} & & \boldsymbol{c} \\ \boldsymbol{p} & & \\ & \boldsymbol{y} & \end{bmatrix}\begin{bmatrix} \boldsymbol{a} & & \\ & & \boldsymbol{r} \\ & \boldsymbol{y} & \end{bmatrix}$$

See that this is as if we had taken all the possible row permutations of a 3 by 3 identity, and then for each permutation, we replace the ones with the numbers at those same positions from the matrix of interest. Then we sum up their determinants.

So how do we compute the determinants? Starting with the identity matrix (det($\boldsymbol{A}$)=1), exchanging two rows negates the determinant to -1, exchanging again puts it back to 1 and so on. See that there are six ways to permute the rows of the 3 by 3 identity, each with determinants alternating between 1 and -1:

$$\begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}\begin{bmatrix} & 1 & \\ 1 & & \\ & & 1 \end{bmatrix}\begin{bmatrix} & 1 & \\ & & 1 \\ 1 & & \end{bmatrix}\begin{bmatrix} & & 1 \\ & 1 & \\ 1 & & \end{bmatrix}\begin{bmatrix} & & 1 \\ 1 & & \\ & 1 & \end{bmatrix}\begin{bmatrix} 1 & & \\ & & 1 \\ & 1 & \end{bmatrix}$$
$$\textbf{det} = \textbf{+1} \qquad \textbf{-1} \qquad \textbf{+1} \qquad \textbf{-1} \qquad \textbf{+1} \qquad \textbf{-1}$$

Multiplying a row by a constant multiplies the determinant by that constant. So supposing the three rows are a, b, c and p, q, r and x, y, z,

$$\begin{bmatrix} \boldsymbol{a} & & \\ & \boldsymbol{q} & \\ & & \boldsymbol{z} \end{bmatrix}\begin{bmatrix} & \boldsymbol{b} & \\ \boldsymbol{p} & & \\ & & \boldsymbol{z} \end{bmatrix}\begin{bmatrix} & \boldsymbol{b} & \\ & & \boldsymbol{r} \\ \boldsymbol{x} & & \end{bmatrix}\begin{bmatrix} & & \boldsymbol{c} \\ & \boldsymbol{q} & \\ \boldsymbol{x} & & \end{bmatrix}\begin{bmatrix} & & \boldsymbol{c} \\ \boldsymbol{p} & & \\ & \boldsymbol{y} & \end{bmatrix}\begin{bmatrix} \boldsymbol{a} & & \\ & & \boldsymbol{r} \\ & \boldsymbol{y} & \end{bmatrix}$$
$$\textbf{det} = \boldsymbol{+aqz} \quad \boldsymbol{-bpz} \quad \boldsymbol{+brx} \quad \boldsymbol{-cqx} \quad \boldsymbol{+cpy} \quad \boldsymbol{-ary}$$

Summing up all the individual determinants gives the final desired determinant. See that the 'checkerboard' pattern of negating the determinant in the laplace expansion just comes from row exchanges.
(next page)

Link to recursive computation
Consider the 4 by 4 determinant

$$\det \begin{bmatrix} A & 0 & 0 & 0 \\ 0 & a & b & c \\ 0 & p & q & r \\ 0 & x & y & z \end{bmatrix}$$

and see that by linearity, this just evaluates to

$$\begin{aligned} A \det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & a & b & c \\ 0 & p & q & r \\ 0 & x & y & z \end{bmatrix} = A \Bigg(& \det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & q & 0 \\ 0 & 0 & 0 & z \end{bmatrix} + \det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & 0 & z \end{bmatrix} + \\ & \det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & r \\ 0 & x & 0 & 0 \end{bmatrix} + \det \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & c \\ 0 & 0 & q & 0 \\ 0 & x & 0 & 0 \end{bmatrix} + \cdots \Bigg) \\ = A \det & \begin{bmatrix} a & b & c \\ p & q & r \\ x & y & z \end{bmatrix} \end{aligned}$$

and so from a general 4 by 4 matrix, (now using different alphabets for the entries so don't confuse this with the earlier notation)

$$\begin{aligned} \det \begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix} = & \det \begin{bmatrix} a & 0 & 0 & 0 \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix} + \det \begin{bmatrix} 0 & b & 0 & 0 \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix} \\ & + \det \begin{bmatrix} 0 & 0 & c & 0 \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix} + \det \begin{bmatrix} 0 & 0 & 0 & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix} \end{aligned}$$

considering the first right hand side matrix, see that by linearity

$$\det \begin{bmatrix} a & 0 & 0 & 0 \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix} = \underbrace{\det \begin{bmatrix} a & 0 & 0 & 0 \\ e & 0 & 0 & 0 \\ i & j & k & l \\ m & n & o & p \end{bmatrix}}_{=0} + \det \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix}$$

where the first matrix on the right hand side is clearly noninvertible and therefore has determinant 0.
(next page)

Cont.
Continuing

$$\det\begin{bmatrix} a & 0 & 0 & 0 \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix} = \det\begin{bmatrix} a & 0 & 0 & 0 \\ 0 & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix}$$

$$= \underbrace{\det\begin{bmatrix} a & 0 & 0 & 0 \\ 0 & f & g & h \\ i & 0 & 0 & 0 \\ m & n & o & p \end{bmatrix}}_{=0} + \det\begin{bmatrix} a & 0 & 0 & 0 \\ 0 & f & g & h \\ 0 & j & k & l \\ m & n & o & p \end{bmatrix}$$

$$= \underbrace{\det\begin{bmatrix} a & 0 & 0 & 0 \\ 0 & f & g & h \\ 0 & j & k & l \\ m & 0 & 0 & 0 \end{bmatrix}}_{=0} + \det\begin{bmatrix} a & 0 & 0 & 0 \\ 0 & f & g & h \\ 0 & j & k & l \\ 0 & n & o & p \end{bmatrix}$$

and using our previous idea,

$$\det\begin{bmatrix} a & 0 & 0 & 0 \\ 0 & f & g & h \\ 0 & j & k & l \\ 0 & n & o & p \end{bmatrix} = a\det\begin{bmatrix} f & g & h \\ j & k & l \\ n & o & p \end{bmatrix}$$

deriving the formula for the laplace expansion from here is fairly straightforward. It is just a matter of applying this reasoning to the other matrices while taking into account the negation caused by required row exchanges. We eventually get

$$\det\begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix} = a\det\begin{bmatrix} f & g & h \\ j & k & l \\ n & o & p \end{bmatrix} - b\det\begin{bmatrix} e & g & h \\ i & k & l \\ m & o & p \end{bmatrix} +$$

$$c\det\begin{bmatrix} e & f & h \\ i & j & l \\ m & n & p \end{bmatrix} - d\det\begin{bmatrix} e & f & g \\ i & j & k \\ m & n & o \end{bmatrix}$$

which is the usual 'formula' used when computing matrices by hand.

5.3 Cofactors, Adjugate, and Inverse

The laplace expansion allows us to reduce a n by n determinant to a sum of scaled $(n-1)$ by $(n-1)$ determinants. The *cofactors* in a laplace transform are essentially these smaller determinants, but with alternanting negation to account for row exchanges:

For the i, j cofactor C_{ij}, remove row i and column j from the matrix A.
C_{ij} equals $(-1)^{i+j}$ times the determinant of the remaining matrix (size $n-1$).
The cofactor formula along row i is $\det A = a_{i1}C_{i1} + \cdots + a_{in}C_{in}$

So given a 3 by 3 matrix $\boldsymbol{A}$

$$\boldsymbol{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

Taking the cofactor for every entry in the matrix, we can construct a cofactor matrix $\boldsymbol{C}$ as

$$\boldsymbol{C} = \begin{bmatrix} \det\begin{bmatrix} e & f \\ h & i \end{bmatrix} & -\det\begin{bmatrix} d & f \\ g & i \end{bmatrix} & \det\begin{bmatrix} d & e \\ g & h \end{bmatrix} \\ -\det\begin{bmatrix} b & c \\ h & i \end{bmatrix} & \det\begin{bmatrix} a & c \\ g & i \end{bmatrix} & -\det\begin{bmatrix} a & b \\ g & h \end{bmatrix} \\ \det\begin{bmatrix} b & c \\ e & f \end{bmatrix} & -\det\begin{bmatrix} a & c \\ d & f \end{bmatrix} & \det\begin{bmatrix} a & b \\ d & e \end{bmatrix} \end{bmatrix}$$

We define $\boldsymbol{C}^T$ as the *adjugate matrix*:

$$\boldsymbol{C}^T = \begin{bmatrix} \det\begin{bmatrix} e & f \\ h & i \end{bmatrix} & -\det\begin{bmatrix} b & c \\ h & i \end{bmatrix} & \det\begin{bmatrix} b & c \\ e & f \end{bmatrix} \\ -\det\begin{bmatrix} d & f \\ g & i \end{bmatrix} & \det\begin{bmatrix} a & c \\ g & i \end{bmatrix} & -\det\begin{bmatrix} a & c \\ d & f \end{bmatrix} \\ \det\begin{bmatrix} d & e \\ g & h \end{bmatrix} & -\det\begin{bmatrix} a & b \\ g & h \end{bmatrix} & \det\begin{bmatrix} a & b \\ d & e \end{bmatrix} \end{bmatrix}$$

(next page)

Cont.
We had

$$\boldsymbol{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

and

$$\boldsymbol{C}^T = \begin{bmatrix} \det\begin{bmatrix} e & f \\ h & i \end{bmatrix} & -\det\begin{bmatrix} b & c \\ h & i \end{bmatrix} & \det\begin{bmatrix} b & c \\ e & f \end{bmatrix} \\ -\det\begin{bmatrix} d & f \\ g & i \end{bmatrix} & \det\begin{bmatrix} a & c \\ g & i \end{bmatrix} & -\det\begin{bmatrix} a & c \\ d & f \end{bmatrix} \\ \det\begin{bmatrix} d & e \\ g & h \end{bmatrix} & -\det\begin{bmatrix} a & b \\ g & h \end{bmatrix} & \det\begin{bmatrix} a & b \\ d & e \end{bmatrix} \end{bmatrix}$$

Now consider the expression $\boldsymbol{A}\boldsymbol{C}^T$. See that computing the *diagonal* of the result is the same as taking the laplace expansion along each row of $\boldsymbol{A}$:

$$\boldsymbol{A}\boldsymbol{C}^T = \begin{bmatrix} \det(\boldsymbol{A}) & \cdots & \cdots \\ \cdots & \det(\boldsymbol{A}) & \cdots \\ \cdots & \cdots & \det(\boldsymbol{A}) \end{bmatrix}$$

This expression has a unique property—every entry on the off diagonal is 0. This is because computing those entries would be equivalent to the *laplace expansion for the determinant of a matrix with two linearly dependent or identical rows.* For instance consider the $(1, 2)$ of $\boldsymbol{A}\boldsymbol{C}^T$:

$$(\boldsymbol{A}\boldsymbol{C}^T)_{1,2} = -a\det\begin{bmatrix} b & c \\ h & i \end{bmatrix} + b\det\begin{bmatrix} a & c \\ g & i \end{bmatrix} - c\det\begin{bmatrix} a & b \\ g & h \end{bmatrix}$$

See that this is the same as taking the determinant of

$$\begin{bmatrix} -a & -b & -c \\ a & b & c \\ g & h & i \end{bmatrix}$$

Which is zero because two rows are linearly dependent. This property can be seen in any of the off-diagonal entries. Our final result is

$$\boldsymbol{A}\boldsymbol{C}^T = \begin{bmatrix} \det(\boldsymbol{A}) & 0 & 0 \\ 0 & \det(\boldsymbol{A}) & 0 \\ 0 & 0 & \det(\boldsymbol{A}) \end{bmatrix}$$

This pattern can be extrapolated to higher order matrices. We have the formula

$$\boldsymbol{A}\boldsymbol{C}^T = (\det(\boldsymbol{A}))\boldsymbol{I}$$

This gives us a nice formula for the inverse matrix $\boldsymbol{A}^{-1}$:

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})}\boldsymbol{C}^T$$

See that, as expected, the determinant of an invertible matrix cannot be zero.

5.4 Cramer's Rule to solve $Ax = b$

Cramer's rule allows us to procedurally find the components x_i, of the solution vector in $\boldsymbol{Ax} = \boldsymbol{b}$. The idea here is to construct a matrix starting with $\boldsymbol{I}$; by replacing the first column of $\boldsymbol{I}$ by $\boldsymbol{x}$, *this triangular matrix* $\boldsymbol{M}_1$ *has a determinant equal to* x_1:

$$\det \begin{bmatrix} x_1 & 0 & 0 \\ x_2 & 1 & 0 \\ x_3 & 0 & 1 \end{bmatrix} = x_1$$

(consider eliminating downward using row operations then taking the product down the diagonal). By multiplying by $\boldsymbol{A}$, *the first column of the result becomes* $\boldsymbol{Ax}$ *which is* $\boldsymbol{b}$; the other columns are copied from $\boldsymbol{A}$. We denote this final matrix by $\boldsymbol{B}_1$:

$$\textbf{Key idea} \quad \boldsymbol{AM_1 = B_1} \qquad \begin{bmatrix} & & \\ & \boldsymbol{A} & \\ & & \end{bmatrix} \begin{bmatrix} \boldsymbol{x_1} & 0 & 0 \\ \boldsymbol{x_2} & 1 & 0 \\ \boldsymbol{x_3} & 0 & 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{b_1} & a_{12} & a_{13} \\ \boldsymbol{b_2} & a_{22} & a_{23} \\ \boldsymbol{b_3} & a_{32} & a_{33} \end{bmatrix} = \boldsymbol{B_1}.$$

See that if we had the determinants of $\boldsymbol{B}_1$ and $\boldsymbol{A}$, we can obtain x_1:

$$\textbf{Product rule} \qquad \boxed{(\det \boldsymbol{A})(\boldsymbol{x_1}) = \det \boldsymbol{B_1} \quad \text{or} \quad \boldsymbol{x_1} = \det \boldsymbol{B_1}/\det \boldsymbol{A}.}$$

This allowed us to obtain the first component of $\boldsymbol{x}$. To obtain the other components we apply the same principle again; to find x_2 and $\boldsymbol{B}_2$, we replace the *second column* of $\boldsymbol{I}$ with $\boldsymbol{x}$:

$$\textbf{Same idea} \quad \boldsymbol{AM_2 = B_2} \qquad \begin{bmatrix} \boldsymbol{a_1} & \boldsymbol{a_2} & \boldsymbol{a_3} \\ & & \end{bmatrix} \begin{bmatrix} 1 & \boldsymbol{x_1} & 0 \\ 0 & \boldsymbol{x_2} & 0 \\ 0 & \boldsymbol{x_3} & 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{a_1} & \boldsymbol{b} & \boldsymbol{a_3} \\ & & \end{bmatrix} = B_2.$$

and take determinants to get

$$(\det \boldsymbol{A})(x_2) = \det \boldsymbol{B}_2 \implies x_2 = \frac{\det \boldsymbol{B}_2}{\det \boldsymbol{A}}$$

CRAMER's RULE If $\det A$ is not zero, $\boldsymbol{Ax} = \boldsymbol{b}$ is solved by determinants:

$$\boldsymbol{x_1} = \frac{\det \boldsymbol{B_1}}{\det \boldsymbol{A}} \qquad \boldsymbol{x_2} = \frac{\det \boldsymbol{B_2}}{\det \boldsymbol{A}} \qquad \cdots \qquad \boldsymbol{x_n} = \frac{\det \boldsymbol{B_n}}{\det \boldsymbol{A}}$$

The matrix $\boldsymbol{B_j}$ has the jth column of $\boldsymbol{A}$ replaced by the vector $\boldsymbol{b}$.

To solve an n by n system, Cramer's rule evaluates $n + 1$ determinants (of $\boldsymbol{A}$ and n different $\boldsymbol{B}_i$). Computing each determinant would involve $n!$ terms (row permutations of n by n $\boldsymbol{I}$)—totalling $(n + 1)!$ terms. This is a very inefficient method for solving equations (although it does give an explicit formula).

5.4.1 Inverse formula from Cramer's rule

Consider using Cramer's rule to solve for $\boldsymbol{A}^{-1}$. For the 2 by 2 case, $\boldsymbol{A}^{-1}$ has two columns; denoting them by $\boldsymbol{x}$ and $\boldsymbol{y}$, we solve

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}\begin{bmatrix} \boldsymbol{x} & \boldsymbol{y} \end{bmatrix} = \boldsymbol{I}$$

solving each column at a time,

$$\boldsymbol{A}\boldsymbol{x} = \begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} \\ \boldsymbol{c} & \boldsymbol{d} \end{bmatrix}\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{1} \\ \boldsymbol{0} \end{bmatrix} \qquad \boldsymbol{A}\boldsymbol{y} = \begin{bmatrix} \boldsymbol{a} & \boldsymbol{b} \\ \boldsymbol{c} & \boldsymbol{d} \end{bmatrix}\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{1} \end{bmatrix}$$

Given $|\boldsymbol{A}|$ and the other required determinants

$$|\boldsymbol{A}| = \begin{vmatrix} \boldsymbol{a} & \boldsymbol{b} \\ \boldsymbol{c} & \boldsymbol{d} \end{vmatrix} \text{ and } \begin{vmatrix} \boldsymbol{1} & \boldsymbol{b} \\ \boldsymbol{0} & \boldsymbol{d} \end{vmatrix} \quad \begin{vmatrix} \boldsymbol{a} & \boldsymbol{1} \\ \boldsymbol{c} & \boldsymbol{0} \end{vmatrix} \quad \begin{vmatrix} \boldsymbol{0} & \boldsymbol{b} \\ \boldsymbol{1} & \boldsymbol{d} \end{vmatrix} \quad \begin{vmatrix} \boldsymbol{a} & \boldsymbol{0} \\ \boldsymbol{c} & \boldsymbol{1} \end{vmatrix}$$

with that we compute the components as

$$\boldsymbol{x}_1 = \frac{\boldsymbol{d}}{|\boldsymbol{A}|},\ \boldsymbol{x}_2 = \frac{\boldsymbol{-c}}{|\boldsymbol{A}|},\ \boldsymbol{y}_1 = \frac{\boldsymbol{-b}}{|\boldsymbol{A}|},\ \boldsymbol{y}_2 = \frac{\boldsymbol{a}}{|\boldsymbol{A}|} \text{ and } \boldsymbol{A}^{-1} = \frac{1}{\boldsymbol{ad-bc}}\begin{bmatrix} \boldsymbol{d} & \boldsymbol{-b} \\ \boldsymbol{-c} & \boldsymbol{a} \end{bmatrix}$$

When using Cramer's rule to solve $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$, *the determinant of each* $\boldsymbol{B}_j$ *is a cofactor of* $\boldsymbol{A}$. Consider the 3 by 3 case; we want the inverse of

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

solving for the first *column* of $\boldsymbol{A}^{-1}$ requires the determinants

$$\textbf{Determinants of } \boldsymbol{B}\textbf{'s are Cofactors of } \boldsymbol{A} \quad \begin{vmatrix} \boldsymbol{1} & \boldsymbol{a_{12}} & \boldsymbol{a_{13}} \\ \boldsymbol{0} & \boldsymbol{a_{22}} & \boldsymbol{a_{23}} \\ \boldsymbol{0} & \boldsymbol{a_{32}} & \boldsymbol{a_{33}} \end{vmatrix} \quad \begin{vmatrix} \boldsymbol{a_{11}} & \boldsymbol{1} & \boldsymbol{a_{13}} \\ \boldsymbol{a_{21}} & \boldsymbol{0} & \boldsymbol{a_{23}} \\ \boldsymbol{a_{31}} & \boldsymbol{0} & \boldsymbol{a_{33}} \end{vmatrix} \quad \begin{vmatrix} \boldsymbol{a_{11}} & \boldsymbol{a_{12}} & \boldsymbol{1} \\ \boldsymbol{a_{21}} & \boldsymbol{a_{22}} & \boldsymbol{0} \\ \boldsymbol{a_{31}} & \boldsymbol{a_{32}} & \boldsymbol{0} \end{vmatrix}$$

see that these determinants are the same as those on the first *row* of the cofactor matrix (transposing and eliminating downward using row operations, then transposing back gives the exact cofactor). As such we have

$$\textbf{FORMULA FOR } \boldsymbol{A^{-1}} \qquad (\boldsymbol{A^{-1}})_{\boldsymbol{ij}} = \frac{\boldsymbol{C_{ji}}}{\textbf{det } \boldsymbol{A}} \qquad \boldsymbol{A^{-1}} = \frac{\boldsymbol{C^{T}}}{\textbf{det } \boldsymbol{A}}$$

Which leaves us with the inverse matrix formula as derived earlier.

5.5 Sign of a permutation and determinants

Notation
We introduce the concept of a sign/signature of a permutation of a set of natural numbers. Consider a set of the first n natural numbers

$$1, 2, \ldots, n$$

We will denote a permuation by

$$\pi(1), \ldots, \pi(n)$$

where $\pi(1)$ is the first element of the permutation, $\pi(2)$ is the second, and so on. For instance, a possible permutation of the first 4 natural numbers is 2,1,4,3; its elements can be denoted as

$$\pi(1) = 2, \quad \pi(2) = 1, \quad \pi(3) = 4, \quad \pi(4) = 3$$

Inversions
Two elements of a permutation $\pi(i)$ and $\pi(j)$ is said to be an inversion if and only if

$$i < j \quad \text{but} \quad \pi(i) > \pi(j)$$

in other words, two elements are an inversion if their natural order is inverted. For instance, the permutation 4,1,3,2 has the inversions

$$(4, 1), \quad (4, 3), \quad (4, 2), \quad (3, 2)$$

Parity of a permutation
A permutation is said to be *even* if and only if the total number of inversions it contains is even. Otherwise it is said to be *odd*.

In the previous example, the permutation 4,1,3,2 has 4 inversions. As such it is *even*.

Definition of sign
The sign of a permutation π, denoted by $\text{sgn}(\pi)$ is a function defined as follows:

$$\text{sgn}(\pi) = \begin{cases} 1 & \text{if } \pi \text{ is even} \\ -1 & \text{if } \pi \text{ is odd} \end{cases}$$

For instance, the permutation $\pi = \langle 4, 1, 3, 2 \rangle$ was even. Therefore $\text{sgn}(\pi) = 1$.
(next page)

Transpositions
The operation of *interchanging any two distinct elements of a permutation* is called a *transposition*.

For instance, considering the following permutation of the first five natural numbers:

$$4, 3, 2, 1, 5$$

The operation of interchanging its second and fourth element so as to obtain the new permuation

$$4, 1, 2, 3, 5$$

is a transposition. Note the effect of transpositions on parity:

- If a permuation is odd, a transposition makes it even.
- If a permutation is even, a transposition makes it odd.

Intuitively, a transposition either adds or removes an inversion. Either way it makes an odd number of inversions even or an even number of inversions odd.

Definition of determinant
Let $\boldsymbol{A}$ be a K x K matrix. Let P be the set of all possible permutations of the first K natural numbers.

The determinant of $\boldsymbol{A}$, denoted by $\det(\boldsymbol{A})$ or $|\boldsymbol{A}|$, is

$$\det = \sum_{\pi \in P} \operatorname{sgn}(\pi) \prod_{k=1}^{K} A_{k,\pi(k)}$$

See how this formally describes the intuitive idea of the determinant as a discussed before. The product

$$\prod_{k=1}^{K} A_{k,\pi(k)}$$

is over K entries of $\boldsymbol{A}$. For each row $k = 1, \ldots, K$, we choose the entry located in column $\pi(k)$; there is exactly one chosen entry $A_{k,\pi(k)}$ in each column and row. The sign handles the negation (the row exchanges), and we sum over the set P of all possible permuations π.

5.6 Determinants of block-triangular matrices and matrices with identity blocks

A block matrix is of the form

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}$$

where

- $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same number of rows;
- $\boldsymbol{C}$ and $\boldsymbol{D}$ have the same number of rows;
- $\boldsymbol{A}$ and $\boldsymbol{C}$ have the same number of columns;
- $\boldsymbol{B}$ and $\boldsymbol{D}$ have the same number of columns.

An important fact about block matrices is that their multiplication can be carried out as per standard matrix multiplication

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} E & F \\ G & H \end{bmatrix} = \begin{bmatrix} AE+BG & AF+BH \\ CE+DG & CF+DH \end{bmatrix}$$

With the only caveat being that the shape of the blocks must be conformable (for instance the columns of $\boldsymbol{A}$ and the rows of $\boldsymbol{E}$ must coincide.

Determinant of block-diagonal matrix with identity blocks The first result concerns block matrices of the form

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{bmatrix} \quad \text{or} \quad \boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A} \end{bmatrix}$$

Where $\boldsymbol{I}$ denotes the identity and $\boldsymbol{A}$ a square matrix. Block matrices whose off-diagonal blocks are all zero are called block-diagonal.

We can show that, given one of the above forms,

$$\det(\boldsymbol{\Gamma}) = \det(\boldsymbol{A})$$

(next page)

Cont.
We can show that,

$$\det\left(\begin{bmatrix} \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix}\right) = \det(\boldsymbol{A}) = \det\left(\begin{bmatrix} \boldsymbol{I} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A} \end{bmatrix}\right)$$

First consider the case where

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I} \end{bmatrix}$$

Suppose $\boldsymbol{A}$ is KxK and $\boldsymbol{I}$ is 1x1. Then $\boldsymbol{\Gamma}$ is $(K+1)$x$(K+1)$. We use the definition of the determinant:

$$\begin{aligned} \det(\Gamma) &= \sum_{\pi \in P} \operatorname{sgn}(\pi) \prod_{k=1}^{K+1} \Gamma_{k,\pi(k)} \\ &= \sum_{\pi \in P} \operatorname{sgn}(\pi) \Gamma_{K,\pi(K+1)} \prod_{k=1}^{K} \Gamma_{k,\pi(k)} \end{aligned}$$

where P is the set of all permutations of the first $K+1$ natural numbers. See that the term $\boldsymbol{\Gamma}_{K,\pi(K+1)}$ is only equal to 1 when $\pi(K+1) = K+1$ (since that row is zero everywhere except on the $(K+1)$th entry).

Also notice that the sign of the permutations would then only depend on $\pi(1), \ldots, \pi(K)$ (because $\pi(K+1) = K+1$ does not determine any inversion). As such we have

$$\begin{aligned} \det(\Gamma) &= \sum_{\pi \in Q} \operatorname{sgn}(\pi) \prod_{k=1}^{K} \Gamma_{k,\pi(k)} \\ &= \sum_{\pi \in Q} \operatorname{sgn}(\pi) \prod_{k=1}^{K} A_{k,\pi(k)} \\ &= \det(A) \end{aligned}$$

where Q is the set of permutations of the first K natural numbers.
(next page)

Cont.
The result for the case in which $\boldsymbol{I}$ is not 1x1 is proved recursively. For instance if $\boldsymbol{I}$ is 2x2, we have

$$\begin{aligned}
&\det\left(\begin{bmatrix} A & 0 \\ 0 & I \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} \begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix} & 0 \\ 0 & 1 \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} A & 0 \\ 0 & 1 \end{bmatrix}\right) = \det(A)
\end{aligned}$$

and analagously for larger dimensions. The proof for the other case:

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A} \end{bmatrix}$$

Is similar to the one just proved.

Block-upper-triangular matrix
An block-upper-triangular matrix is of the form

$$\boldsymbol{\Gamma} = \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}$$

We can show that

$$\det(\boldsymbol{\Gamma}) = \det(\boldsymbol{A})\det(\boldsymbol{D})$$

Suppose that $\boldsymbol{A}$ is KxK, $\boldsymbol{D}$ is LxL, $\boldsymbol{B}$ is KxL and $\boldsymbol{0}$ is LxK. Using subscripts to denote shape, we can factor the matrix as

$$\begin{aligned}
&\begin{bmatrix} I_K & 0_{KL} \\ 0_{LK} & D \end{bmatrix}\begin{bmatrix} I_K & B \\ 0_{LK} & I_L \end{bmatrix}\begin{bmatrix} A & 0_{KL} \\ 0_{LK} & I_L \end{bmatrix} \\
&= \begin{bmatrix} I_K I_K + 0_{KL}0_{LK} & I_K B + 0_{KL} I_L \\ 0_{LK} I_K + D0_{LK} & 0_{LK} B + D I_L \end{bmatrix}\begin{bmatrix} A & 0_{KL} \\ 0_{LK} & I_L \end{bmatrix} \\
&= \begin{bmatrix} I_K & B \\ 0_{LK} & D \end{bmatrix}\begin{bmatrix} A & 0_{KL} \\ 0_{LK} & I_L \end{bmatrix} \\
&= \begin{bmatrix} I_K A + B0_{LK} & I_K 0_{KL} + B I_L \\ 0_{LK} A + D0_{LK} & 0_{LK}0_{KL} + D I_L \end{bmatrix} \\
&= \begin{bmatrix} A & B \\ 0 & D \end{bmatrix}
\end{aligned}$$

(next page)

Cont.
Thus,

$$
\begin{aligned}
&\det(\Gamma) \\
&= \det\left(\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} I_K & 0_{KL} \\ 0_{LK} & D \end{bmatrix}\begin{bmatrix} I_K & B \\ 0_{LK} & I_L \end{bmatrix}\begin{bmatrix} A & 0_{KL} \\ 0_{LK} & I_L \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} I_K & 0_{KL} \\ 0_{LK} & D \end{bmatrix}\right)\det\left(\begin{bmatrix} I_K & B \\ 0_{LK} & I_L \end{bmatrix}\right)\det\left(\begin{bmatrix} A & 0_{KL} \\ 0_{LK} & I_L \end{bmatrix}\right) \\
&= \det(D)\det\left(\begin{bmatrix} I_K & B \\ 0_{LK} & I_L \end{bmatrix}\right)\det(A) \\
&= \det(A)\det(D)
\end{aligned}
$$

In the last step we have

$$\det\left(\begin{bmatrix} I_K & B \\ 0_{LK} & I_L \end{bmatrix}\right) = 1$$

(the determinant is just the product of pivots since we can eliminate upward without changing the determinant) The other two determinants come from the earlier derived determinants of matrices with identity blocks.

Block-lower-triangular matrix
A block-lower-triangular matrix is of the form

$$\Gamma = \begin{bmatrix} A & 0 \\ C & D \end{bmatrix}$$

Again we can show that

$$\det(\boldsymbol{\Gamma}) = \det(\boldsymbol{A})\det(\boldsymbol{D})$$

(next page)

Cont.
Suppose that $\boldsymbol{A}$ is KxK, $\boldsymbol{D}$ is LxL, $\boldsymbol{C}$ is LxK and $\mathbf{0}$ is KxL. Using subscripts to denote shape, we can factor the matrix as

$$\begin{aligned}
&\begin{bmatrix} A & 0_{KL} \\ 0_{LK} & I_L \end{bmatrix}\begin{bmatrix} I_K & 0_{KL} \\ C & I_L \end{bmatrix}\begin{bmatrix} I_K & 0_{KL} \\ 0_{LK} & D \end{bmatrix} \\
&= \begin{bmatrix} AI_K + 0_{KL}C & A0_{KL} + 0_{KL}I_L \\ 0_{LK}I_K + I_L C & 0_{LK}0_{KL} + I_L I_L \end{bmatrix}\begin{bmatrix} I_K & 0_{KL} \\ 0_{LK} & D \end{bmatrix} \\
&= \begin{bmatrix} A & 0_{KL} \\ C & I_L \end{bmatrix}\begin{bmatrix} I_K & 0_{KL} \\ 0_{LK} & D \end{bmatrix} \\
&= \begin{bmatrix} AI_K + 0_{KL}0_{LK} & A0_{KL} + 0_{KL}D \\ CI_K + I_L 0_{KL} & C0_{KL} + I_L D \end{bmatrix} \\
&= \begin{bmatrix} A & 0 \\ C & D \end{bmatrix}
\end{aligned}$$

Then, similar to the previous proof,

$$\begin{aligned}
&\det(\Gamma) \\
&= \det\left(\begin{bmatrix} A & 0 \\ C & D \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} A & 0_{KL} \\ 0_{LK} & I_L \end{bmatrix}\begin{bmatrix} I_K & 0_{KL} \\ C & I_L \end{bmatrix}\begin{bmatrix} I_K & 0_{KL} \\ 0_{LK} & D \end{bmatrix}\right) \\
&= \det\left(\begin{bmatrix} A & 0_{KL} \\ 0_{LK} & I_L \end{bmatrix}\right)\det\left(\begin{bmatrix} I_K & 0_{KL} \\ C & I_L \end{bmatrix}\right)\det\left(\begin{bmatrix} I_K & 0_{KL} \\ 0_{LK} & D \end{bmatrix}\right) \\
&= \det(A)\det(I_L)\det(I_K)\det(D) \\
&= \det(A)\det(D)
\end{aligned}$$

Chapter 6

Eigenvalues and Eigenvectors

6.1 Computation

$\boldsymbol{x}$ is an eigenvector if

$$\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$$

We have

$$\boldsymbol{A}\boldsymbol{x} - \lambda\boldsymbol{I}\boldsymbol{x} = \boldsymbol{0}$$
$$(\boldsymbol{A}\boldsymbol{x} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$$

The matrix $(\boldsymbol{A} - \lambda\boldsymbol{I})$ times the eigenvector $\boldsymbol{x}$ is the zero vector. *the eigenvectors make up the nullspace of* $\boldsymbol{A} - \lambda\boldsymbol{I}$.

First we find the eigenvalues. If $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$ has a nonzero solution, then $(\boldsymbol{A} - \lambda\boldsymbol{I})$ is *not invertible*—the determinant of $(\boldsymbol{A} - \lambda\boldsymbol{I})$ must be zero. This gives us a means to find the eigenvalues:

Eigenvalues **The number λ is an eigenvalue of A if and only if $A - \lambda I$ is singular.**

Equation for the n eigenvalues of A $\boxed{\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = 0.}$ (5)

An n by n $(\boldsymbol{A} - \lambda\boldsymbol{I})$ has $A_{ii} - \lambda$ down the main diagonal. This means that the *characterstic polynomial* $\det(\boldsymbol{A} - \lambda\boldsymbol{I})$ has a degree of n—the determinant, after expansion into a polynomial, has n solutions λ_1 to λ_n—and $\boldsymbol{A}$ has n eigenvalues:

An n by n matrix has n eigenvalues (repeated λ's are possible !) Each λ leads to $\boldsymbol{x}$:

For each eigenvalue λ solve $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$ or $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ to find an eigenvector $\boldsymbol{x}$.

(next page)

Cont.
After obtaining the eigenvalues, we can find the eigenvectors by substituting the eigenvalues back into

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0}$$

and solve for $\boldsymbol{x}$.

Invertibility
Consider again the intial equation

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$$

If $\boldsymbol{A}$ were singular, it has a nontrivial nullspace and 0 is an eigenvalue, with the eigenvectors for $\lambda = 0$ filling the nullspace.

If $\boldsymbol{A}$ were invertible, zero cannot be an eigenvalue. (the nullspace is just the zero vector, so any eigenvector cannot have a zero eigenvalue)

Summary To solve the eigenvalue problem for an n by n matrix, follow these steps:

1. ***Compute the determinant of*** $A - \lambda I$. With λ subtracted along the diagonal, this determinant starts with λ^n or $-\lambda^n$. It is a polynomial in λ of degree n.
2. ***Find the roots of this polynomial***, by solving $\det(A - \lambda I) = 0$. The n roots are the n eigenvalues of A. They make $A - \lambda I$ singular.
3. For each eigenvalue λ, ***solve*** $(A - \lambda I)\boldsymbol{x} = \boldsymbol{0}$ ***to find an eigenvector*** $\boldsymbol{x}$.

6.2 Eigenvalues, Determinant, and Trace

Note that if we perform row operations on $\boldsymbol{A}$, the eigenvalues usually change. However, the determinant and trace of a matrix still harbour useful properties.

The determinant of $\boldsymbol{A}$ is the product of its eigenvalues
We can show that

$$\det(\boldsymbol{A}) = \lambda_1 \lambda_2 \cdots \lambda_n$$

Supposing that $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $\boldsymbol{A}$, then the λ are also the roots of the characteristic polynomial:

$$\begin{aligned} \det(\boldsymbol{A} - \lambda \boldsymbol{I}) = p(\lambda) &= (-1)^n (\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_n) \\ &= (-1)(\lambda - \lambda_1)(-1)(\lambda - \lambda_2) \cdots (-1)(\lambda - \lambda_n) \\ &= (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda) \end{aligned}$$

The leading coefficient $(-1)^n$ can be obtained by considering the contribution of the main diagonal to the determinant (the main diagonal is the only permutation that contributes to the coefficient of λ^n). Setting λ to zero gives us our desired result.

The trace of $\boldsymbol{A}$ is the sum of its eigenvalues
Look again at the characteristic polynomial. See that it can be rewritten as

$$p(\lambda) = (-1)^n (\lambda^n - (\lambda_1 + \cdots + \lambda_n)\lambda^{n-1} + \cdots)$$

Looking at the determinant of interest $\det(\boldsymbol{A} - \lambda \boldsymbol{I})$, see that, in a similar manner to the previous part, the main diagonal is the only permutation that contributes to the λ^{n-1} coefficient in the result. Consider the laplace expansion along a row i, we see that any term involving an off diagonal element $[\boldsymbol{A} - \lambda \boldsymbol{I}]_{ij}$ eliminates $a_{ii} - \lambda$ and $a_{jj} - \lambda$ from the main diagonal (also see that any permutation with an off diagonal entry (most of them) would have to be missing at least two elements on the main diagonal).

Therefore the coefficient for λ^{n-1} comes solely from the main diagonal:

$$(a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda) = \cdots + (-1)^{n-1}(a_{11} + \cdots + a_{nn})\lambda^{n-1} + \cdots$$

Comparing coefficients with our characteristic polynomial we get the result

$$a_{11} + \cdots + a_{nn} = \operatorname{tr}(\boldsymbol{A}) = \lambda_1 + \cdots + \lambda_n$$

In short

The product $(\lambda_1) \cdots (\lambda_n)$ of the n eigenvalues equals the determinant of A
$\lambda_1 + \lambda_2 + \cdots + \lambda_n$ equals the sum of the n diagonal entries = (trace of A)

6.3 Commuting matrices share eigenvectors

Say we have two matrices that commute:

$$\boldsymbol{AB} = \boldsymbol{BA}$$

Starting from $\boldsymbol{Ax} = \lambda \boldsymbol{x}$, we have

$$\boldsymbol{ABx} = \boldsymbol{BAx} = \boldsymbol{B}\lambda\boldsymbol{x} = \lambda\boldsymbol{Bx}$$

So $\boldsymbol{x}$ and $\boldsymbol{Bx}$ are both eigenvectors of $\boldsymbol{A}$ with the same eigenvalue.

If we make a further assumption that the eigenvalues of $\boldsymbol{A}$ are distinct—the eigenspaces of $\boldsymbol{A}$ are one dimensional—then $\boldsymbol{Bx}$ must also be a multiple of $\boldsymbol{x}$, making $\boldsymbol{x}$ also an eigenvector of $\boldsymbol{B}$.

6.4 $\boldsymbol{AB}$ and $\boldsymbol{BA}$ have the same eigenvalues

If $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{AB}$ with some nonzero eigenvalue λ, then $\boldsymbol{Bv} \neq 0$ and

$$\lambda\boldsymbol{Bv} = \boldsymbol{B}(\boldsymbol{ABv}) = (\boldsymbol{BA})\boldsymbol{Bv}$$

so $\boldsymbol{Bv}$ is an eigenvector for $\boldsymbol{BA}$ with the same eigenvalue.

For the case of a zero eigenvalue, if 0 is an eigenvalue of $\boldsymbol{AB}$, then the matrix is noninvertible and

$$0 = \det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B}) = \det(\boldsymbol{BA})$$

so 0 will also be an eigenvalue of $\boldsymbol{BA}$.

6.5 Diagonalisation

Suppose the n by n matrix $\boldsymbol{A}$ has $\boldsymbol{n}$ linearly independent eigenvectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. Placing those $\boldsymbol{x}_i$ into the columns of an invertible *eigenvector matrix* $\boldsymbol{X}$, then $\boldsymbol{X}^{-1}\boldsymbol{A}\boldsymbol{X}$ is the diagonal *eigenvalue matrix* $\boldsymbol{\Lambda}$:

Eigenvector matrix X
Eigenvalue matrix Λ

$$\boldsymbol{X^{-1}AX = \Lambda} = \begin{bmatrix} \boldsymbol{\lambda_1} & & \\ & \boldsymbol{\ddots} & \\ & & \boldsymbol{\lambda_n} \end{bmatrix}.$$

To understand this, see that $\boldsymbol{AX} = \boldsymbol{X\Lambda}$:

A times X

$$\boldsymbol{AX} = \boldsymbol{A}\begin{bmatrix} \boldsymbol{x_1} & \cdots & \boldsymbol{x_n} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda_1 x_1} & \cdots & \boldsymbol{\lambda_n x_n} \end{bmatrix}.$$

X times Λ

$$\begin{bmatrix} \boldsymbol{\lambda_1 x_1} & \cdots & \boldsymbol{\lambda_n x_n} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x_1} & \cdots & \boldsymbol{x_n} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda_1} & & \\ & \boldsymbol{\ddots} & \\ & & \boldsymbol{\lambda_n} \end{bmatrix} = \boldsymbol{X\Lambda}.$$

Assuming that n by n $\boldsymbol{X}$ consists of n linearly independent eigenvectors, it then has an inverse, and so

$$\boxed{AX = X\Lambda \quad \text{is} \quad \boldsymbol{X^{-1}AX = \Lambda} \quad \text{or} \quad \boldsymbol{A = X\Lambda X^{-1}}.}$$

One benefit of diagonalisation is that it becomes easy to compute matrix powers, $\boldsymbol{A}^k$:

$$\boldsymbol{A^k = (X\Lambda X^{-1})(X\Lambda X^{-1})\ldots(X\Lambda X^{-1}) = X\Lambda^k X^{-1}}.$$

(next page)

Diagonalisability

We have a condition that guarantees diagonalisability: *Eigenvectors for n different λ's are independent*—if the eigenvectors are distinct then we can diagonalise $\boldsymbol{A}$:

> **Independent $\boldsymbol{x}$ from different λ** Eigenvectors that correspond to distinct (all different) eigenvalues are linearly independent. An n by n matrix that has n different eigenvalues (no repeated λ's) must be diagonalizable.

Proof: Suppose $c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 = \boldsymbol{0}$ (suppose two eigenvectors with different eigenvalues are dependent) Multiply the entire expression by $\boldsymbol{A}$:

$$\boldsymbol{A}(c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2) = c_1\lambda_1\boldsymbol{x}_1 + c_2\lambda_2\boldsymbol{x}_2 = \boldsymbol{0}$$

now consider the first expression multiplied by λ_2

$$\lambda_2(c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2) = c_1\lambda_2\boldsymbol{x}_1 + c_2\lambda_2\boldsymbol{x}_2 = \boldsymbol{0}$$

subtracting the two expressions gives us

$$c_1\lambda_1\boldsymbol{x}_1 - c_1\lambda_2\boldsymbol{x}_1 = (\lambda_1 - \lambda_2)c_1\boldsymbol{x}_1 = \boldsymbol{0}$$

Since $\lambda_1 \neq \lambda_2$ and $\boldsymbol{x}_1 \neq \boldsymbol{0}$, we must have $c_1 = 0$. See that we can repeat the same process but multiply by λ_1 instead to show that we must also have $c_2 = 0$. Only the combination with $c_1 = c_2 = 0$ gives $c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 = \boldsymbol{0}$—the eigenvectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ must be independent.

Generalising to n eigenvectors: Suppose that

$$c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n = \boldsymbol{0}$$

First multiply by $\boldsymbol{A}$:

$$c_1\lambda_1\boldsymbol{x}_1 + \cdots + c_n\lambda_n\boldsymbol{x}_n = \boldsymbol{0}$$

then multiply that same first expression by λ_n:

$$c_1\lambda_n\boldsymbol{x}_1 + \cdots + c_n\lambda_n\boldsymbol{x}_n = \boldsymbol{0}$$

and subtract the two expressions to get

$$\begin{aligned}&(\lambda_1 - \lambda_n)c_1\boldsymbol{x}_1 + \cdots + c_n(\lambda_n - \lambda_n)\boldsymbol{x}_n\\ =&(\lambda_1 - \lambda_n)c_1\boldsymbol{x}_1 + \cdots + c_{n-1}(\lambda_{n-1} - \lambda_n)\boldsymbol{x}_{n-1} = \boldsymbol{0}\end{aligned}$$

Now take the result and multiply by $\boldsymbol{A}$ again. Subtract that same expression, but now multiplied by λ_{n-1}:

$$(\lambda_1 - \lambda_n)(\lambda_1 - \lambda_{n-1})c_1\boldsymbol{x}_1 + \cdots + c_{n-1}(\lambda_{n-1} - \lambda_n)(\lambda_{n-1} - \lambda_{n-1})\boldsymbol{x}_{n-1} = \boldsymbol{0}$$

repeating this process, we eventually end up with

$$(\lambda_1 - \lambda_n)(\lambda_1 - \lambda_{n-1})\cdots(\lambda_1 - \lambda_2)c_1\boldsymbol{x}_1 = \boldsymbol{0}$$

(next page)

Cont.
We have

$$(\lambda_1 - \lambda_n)(\lambda_1 - \lambda_{n-1}) \cdots (\lambda_1 - \lambda_2)c_1\boldsymbol{x}_1 = \mathbf{0}$$

Forcing $c_1 = 0$. We can repeat this same process to show that every $c_i = 0$; when the λ are different, $c_1\boldsymbol{x}_1 + \ldots + c_n\boldsymbol{x}_n = \mathbf{0}$ only when $c_i = 0$, the eigenvectors are independent.

Repeated eigenvalues don't imply non-diagonalisability
Distinct eigenvalues guarantee diagonalisability, however this does not mean that repeated eigenvalues implies non-diagonalisability—if all the eigenvalues of a matrix are distinct, then the matrix is automatically diagonalisable, but there are plenty of cases where a matrix is diagonalisable, but has repeated eigenvalues.

6.6 Similar Matrices

Suppose the eigenvalue matrix $\boldsymbol{\Lambda}$ is fixed. As we change the eigenvector matrix $\boldsymbol{X}$, we get a whole family of different matrices $\boldsymbol{A} = \boldsymbol{X\Lambda X}^{-1}$—*all with the same eigenvalues in* $\boldsymbol{\Lambda}$. All those matrices $\boldsymbol{A}$ (with the same $\boldsymbol{\Lambda}$) are called *similar*.

This idea extends to matrices that *can't be diagonalised*. We choose one constant matrix $\boldsymbol{C}$ (not necessarily $\boldsymbol{\Lambda}$), and we look at the whole family of matrices $\boldsymbol{A} = \boldsymbol{BCB}^{-1}$, allowing all invertible matrices $\boldsymbol{B}$. It turns out that the matrices $\boldsymbol{C}$ and $\boldsymbol{A}$ are similar—they have the same n eigenvalues.

We say $\boldsymbol{C}$ instead of $\boldsymbol{\Lambda}$ because $\boldsymbol{C}$ need not be diagonal. We only require invertible $\boldsymbol{B}$.

All the matrices $\boldsymbol{A} = \boldsymbol{BCB}^{-1}$ are "similar." They all share the eigenvalues of $\boldsymbol{C}$.

To show this, see that if $\boldsymbol{Cx} = \lambda\boldsymbol{x}$, then $\boldsymbol{BCB}^{-1}$ has the same eigenvalue λ with new eigenvector $\boldsymbol{Bx}$:

Similar matrix, same $\boldsymbol{\lambda}$ $\qquad (BCB^{-1})(B\boldsymbol{x}) = BC\boldsymbol{x} = B\lambda\boldsymbol{x} = \lambda(B\boldsymbol{x}).$

A fixed matrix $\boldsymbol{C}$ produces a family of similar matrices $\boldsymbol{BCB}^{-1}$.

6.7 Matrix Powers

The eigenvector matrix $\boldsymbol{X}$ allows us to diagonalise $\boldsymbol{A}$ as

$$\boldsymbol{A} = \boldsymbol{X}\boldsymbol{\Lambda}\boldsymbol{X}^{-1}$$

This factorisation is particularly useful when computing powers:

Powers of A $\qquad A^k u_0 = (X\Lambda X^{-1})\cdots(X\Lambda X^{-1})u_0 = X\Lambda^k X^{-1}u_0$

We can intuitively understand $\boldsymbol{X}\boldsymbol{\Lambda}^k\boldsymbol{X}^{-1}\boldsymbol{u}_0$. The expression can be constructed as follows:

1. Write u_0 as a combination $Xc = c_1x_1 + \cdots + c_nx_n$ of eigenvectors. $\quad c = X^{-1}u_0$.
2. Multiply each eigenvector x_i by $(\lambda_i)^k$. Now we have $\Lambda^k X^{-1}u_0$.
3. Add up the pieces $c_i(\lambda_i)^k x_i$ to find the solution $u_k = A^k u_0$. This is $X\Lambda^k X^{-1}u_0$.

Solution for $u_{k+1} = Au_k$ $\qquad u_k = A^k u_0 = c_1(\lambda_1)^k x_1 + \cdots + c_n(\lambda_n)^k x_n.$ $\qquad$ (11)

We find the initial linear combination of eigenvectors that produces $\boldsymbol{u}_0$, whose coefficients we denote as $\boldsymbol{c}$. Each time we multiply by $\boldsymbol{A}$, we scale each eigenvector by its corresponding eigenvalue, so the result of $\boldsymbol{A}\boldsymbol{u}_0$ is just that linear combination with coefficients $\boldsymbol{\Lambda}\boldsymbol{c}$. We scale each coefficient as many times as we like by its corresponding eigenvalue, $\boldsymbol{\Lambda}^k\boldsymbol{c}$; evaluating the final linear combination is precisely our factorisation. That is,

$$A^k u_0 = X\Lambda^k X^{-1}u_0 = X\Lambda^k c = \begin{bmatrix} x_1 & \dots & x_n \end{bmatrix} \begin{bmatrix} (\lambda_1)^k & & \\ & \ddots & \\ & & (\lambda_n)^k \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

is the same as

$$\boldsymbol{A}^k\boldsymbol{u}_0 = \boldsymbol{u}_k = c_1(\lambda_1)^k\boldsymbol{x}_1 + \cdots + c_n(\lambda_n)^k\boldsymbol{x}_n$$

which is a result of starting with $\boldsymbol{u}_0$

$$u_0 = c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n$$

and applying $\boldsymbol{A}$ repeatedly

$$\boldsymbol{A}\boldsymbol{u}_0 = u_1 = c_1(\lambda_1)\boldsymbol{x}_1 + \cdots + c_n(\lambda_n)\boldsymbol{x}_n$$

6.8 Inutition for spectral theorem

6.8.1 Symmetric matrices are always diagonalisable

If $\boldsymbol{A}$ is an nxn symmetric matrix with eigenvector $\boldsymbol{v}$, and $\boldsymbol{u} \in \mathbb{C}$ is such that $\boldsymbol{u} \perp \boldsymbol{v}$, then $\boldsymbol{Au} \perp \boldsymbol{v}$. We can show this:

$$\boldsymbol{v} \cdot (\boldsymbol{Au}) = \boldsymbol{v}^T \boldsymbol{Au} = \boldsymbol{v}^T \boldsymbol{A}^T \boldsymbol{u} = (\boldsymbol{v}^T \boldsymbol{A}^T)\boldsymbol{u} = (\boldsymbol{Av})^T \boldsymbol{u} = \lambda \boldsymbol{v}^T \boldsymbol{u} = 0$$

where λ is the eigenvalue of $\boldsymbol{v}$.

Let $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k\}$ be any set of k linearly independent eigenvectors of $\boldsymbol{A}$. Let $\boldsymbol{V} \subseteq \mathbb{C}^n$ be the vector space spanned by these vectors, and $\boldsymbol{V}^\perp \subseteq \mathbb{C}^n$ be the orthogonal space, containing all vectors orthogonal to $\boldsymbol{V}$.

Every vector $\boldsymbol{u} \in \boldsymbol{V}^\perp$ is orthogonal to each $\boldsymbol{v}_i$. By ourfirst statement, so is $\boldsymbol{Au}$. Therefore, $\boldsymbol{V}^\perp$ is fixed under $\boldsymbol{A}$; we can make a restricted linear transformation $\boldsymbol{A}|_{\boldsymbol{V}^\perp} : \boldsymbol{V}^\perp \to \boldsymbol{V}^\perp$.

Every linear transformation over a nonzero, finite dimensional space of an algebraically closed field has at least one eigenvector (this can be shown using the fundamental theorem of algebra). If $\boldsymbol{V}^\perp \neq \{\boldsymbol{0}\}$, then there is another eigenvector $\boldsymbol{v}_{k+1} \in \boldsymbol{V}^\perp$ of $\boldsymbol{A}$, *orthogonal to* (and linearly independent from) $\{\boldsymbol{v}_1, \ldots \boldsymbol{v}_k\}$.

By induction, we can continue to extend $\boldsymbol{V}$ by more eigenvectors until $\boldsymbol{V} = \mathbb{C}^n$. At this point we then have n linearly independent eigenvectors of $\boldsymbol{A}$, providing a basis for its diagonalisation.

6.8.2 Eigenvectors of distinct eigenvalues of a symmetric matrix are orthogonal

For any real matrix $\boldsymbol{A}$ and any vectors $\boldsymbol{x}$ and $\boldsymbol{y}$, we have

$$\langle \boldsymbol{Ax}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{A}^T \boldsymbol{y} \rangle$$

Now assuming that $\boldsymbol{A}$ is symmetric, and that $\boldsymbol{x}$ and $\boldsymbol{y}$ are eigenvectors of $\boldsymbol{A}$ corresponding to distinct eigenvalues λ and μ. Then

$$\begin{aligned}\lambda\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \lambda \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{Ax}, \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{A}^T \boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{Ay} \rangle = \langle \boldsymbol{x}, \mu \boldsymbol{y} \rangle \\ = \mu \langle \boldsymbol{x}, \boldsymbol{y} \rangle\end{aligned}$$

Therefore,

$$(\lambda - \mu)\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$$

since $\lambda - \mu \neq 0$, then $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$, and $\boldsymbol{x} \perp \boldsymbol{y}$.

6.9 Symmetric Positive Definite Matrices

6.9.1 Equivalence of definitions

A symmetric matrix is *positive definite* if the energy $\boldsymbol{x}^T\boldsymbol{S}\boldsymbol{x}$ is positive for all vectors $\boldsymbol{x} \neq \boldsymbol{0}$.

Another definition: A symmetric matrix is positive definite if all its eigenvalues are positive.

We can show that the two definitions are equivalent:

$$\boldsymbol{S}\boldsymbol{x} = \lambda\boldsymbol{x} \implies \boldsymbol{x}^T\boldsymbol{S}\boldsymbol{x} = \lambda\boldsymbol{x}^T\boldsymbol{x}$$

So $\lambda > 0$ leads to energy $\boldsymbol{x}^T\boldsymbol{S}\boldsymbol{x} > 0$. If the eigenvalues are all positive, then the energy definition holds when $\boldsymbol{x}$ are eigenvectors.

We can show further that if $\boldsymbol{x}^T\boldsymbol{S}\boldsymbol{x} > 0$ for the eigenvectors of $\boldsymbol{S}$, then $x^T\boldsymbol{S}\boldsymbol{x} > 0$ for every nonzero vector $\boldsymbol{x}$. Recall that symmetric matrices are always diagonalisable, and that their eigenvectors can be chosen orthogonal. As such, every $\boldsymbol{x}$ is a combination $c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n$ of the eigenvectors, and we can show

$$\begin{aligned}\boldsymbol{x}^T\boldsymbol{S}\boldsymbol{x} &= (c_1\boldsymbol{x}_1^T + \cdots + c_n\boldsymbol{x}_n^T)\,\boldsymbol{S}\,(c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n)\\ &= (c_1\boldsymbol{x}_1^T + \cdots + c_n\boldsymbol{x}_n^T)(c_1\lambda_1\boldsymbol{x}_1 + \cdots + c_n\lambda_n\boldsymbol{x}_n)\\ &= (c_1^2\lambda_1\boldsymbol{x}_1^T\boldsymbol{x}_1 + \cdots + c_n^2\lambda_n\boldsymbol{x}_n^T\boldsymbol{x}_n)\end{aligned}$$

(In the last step we used the orthogonality of the eigenvectors of $\boldsymbol{S}$, so $\boldsymbol{x}_i^t\boldsymbol{x}_j = 0$ for $i \neq j$) See that the entire final expression is > 0 if every $\lambda_i > 0$.

6.9.2 If $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are symmetric positive definite, so is $\boldsymbol{S}_1 + \boldsymbol{S}_2$

We can prove this simply adding energies:

$$\boldsymbol{x}^T(\boldsymbol{S}_1 + \boldsymbol{S}_2)\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{S}_1\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{S}_2\boldsymbol{x}$$

If both $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are positive definite, so is their sum.

6.10 Cholesky decomposition

Let $\boldsymbol{A}$ be a K by K matrix. We say that $\boldsymbol{A}$ possesses a cholesky decomposition if and only if there exists a lower triangular K by K matrix $\boldsymbol{L}$ such that its diagonal entries are strictly positive real numbers and

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^*$$

Where $\boldsymbol{L}^*$ denotes the conjugate transpose of $\boldsymbol{L}$. When $\boldsymbol{L}$ is real, then the conjugate transpose $\boldsymbol{L}^*$ coincides with the transpose $\boldsymbol{L}^T$ and the cholesky factorisation is

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}^T$$

When we restrict our attention to real vectors and matrices, then we say that a real matrix $\boldsymbol{A}$ is positive definite if and only if $\boldsymbol{A}$ is symmetric and

$$\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$$

for any non-zero real vector $\boldsymbol{x}$.

We can show that a square matrix possesses a cholesky factorisation if and only if it is positive definite. Since a positive definite matrix $\boldsymbol{A}$ is Hermitian ($\boldsymbol{A} = \boldsymbol{A}^*$), it can be diagonalised as

$$\boldsymbol{A} = \boldsymbol{P}^*\boldsymbol{D}\boldsymbol{P}$$

Where $\boldsymbol{P}$ is a unitary matrix (orthogonal but allowed to be complex) and $\boldsymbol{D}$ is a diagonal matrix with the eigenvalues of $\boldsymbol{A}$ on its main diagonal. Since $\boldsymbol{A}$ is positive definite, its eigenvalues are *strictly positive numbers*. Thus we can write

$$\boldsymbol{D} = \boldsymbol{D}^{1/2}\boldsymbol{D}^{1/2}$$

where $\boldsymbol{D}^{1/2}$ is a diagonal matrix such that its (k,k)-th entry satisfies

$$(\boldsymbol{D}^{1/2})_{kk} = \sqrt{\boldsymbol{D}_{kk}}$$

for $\boldsymbol{k} = 1, \ldots, K$. Therefore,

$$\begin{aligned}\boldsymbol{A} &= \boldsymbol{P}^*\boldsymbol{D}\boldsymbol{P} \\ &= \boldsymbol{P}^*\boldsymbol{D}^{1/2}\boldsymbol{D}^{1/2}\boldsymbol{P} \\ &= (\boldsymbol{D}^{1/2}\boldsymbol{P})^*\boldsymbol{D}^{1/2}\boldsymbol{P}\end{aligned}$$

The matrix $\boldsymbol{P}$ is unitary and therefore full rank. The matrix $\boldsymbol{D}^{1/2}$ is diagonal with strictly positive entries—and is therefore also full-rank. The product of two full rank matrices is full rank, and so $\boldsymbol{D}^{1/2}\boldsymbol{P}$ is full-rank.
(next page)

Cont.
$\boldsymbol{D}^{1/2}\boldsymbol{P}$ is full-rank. It therefore has a $\boldsymbol{QR}$ decomposition (discussed earlier):

$$\boldsymbol{D}^{1/2}\boldsymbol{P} = \boldsymbol{QR}$$

where $\boldsymbol{Q}$ is a unitary matrix and $\boldsymbol{R}$ is upper triangular *with strictly positive entries on its main diagonal*. Thus we have

$$\begin{aligned}\boldsymbol{A} &= (\boldsymbol{D}^{1/2}\boldsymbol{P})^*\boldsymbol{D}^{1/2}\boldsymbol{P} \\ &= (\boldsymbol{QR})^*\boldsymbol{QR} \\ &= \boldsymbol{R}^*\boldsymbol{Q}^*\boldsymbol{QR} \\ &= \boldsymbol{R}^*\boldsymbol{R} \\ &= \boldsymbol{LL}^*\end{aligned}$$

This shows that the all positive definite matrices have cholesky factorisations.

Say that some $\boldsymbol{A}$ has a cholesky decomposition. Since the diagonal entries of $\boldsymbol{L}$ are strictly positive, $\boldsymbol{L}$ and $\boldsymbol{L}^*$ are full-rank. Therefore for any $\boldsymbol{x} \neq \boldsymbol{0}$, we have

$$\boldsymbol{L}^*\boldsymbol{x} \neq \boldsymbol{0}$$

which then satisfies

$$\boldsymbol{x}^*\boldsymbol{Ax} = \boldsymbol{x}^*\boldsymbol{LL}^*\boldsymbol{x} = ||L^*\boldsymbol{x}||^2 \geq 0$$

for any $\boldsymbol{x} \neq \boldsymbol{0}$. This shows that any matrix with a cholesky factorisation is positive definite.

Uniqueness
We can show that the cholesky factorisation for any positive definite $\boldsymbol{A}$ is unique. Given that we have

$$\boldsymbol{A} = \boldsymbol{LL}^*$$

Suppose that there exists another decomposition

$$\boldsymbol{A} = \boldsymbol{MM}^*$$

Then we have

$$\boldsymbol{LL}^* = \boldsymbol{MM}^*$$

and

$$\boldsymbol{M}^{-1}\boldsymbol{L} = \boldsymbol{M}^*(\boldsymbol{L}^*)^{-1}$$

where the existence of the inverses $(\boldsymbol{L}^*)^{-1}$ and $\boldsymbol{M}^{-1}$ are guaranteed by the fact that $\boldsymbol{L}$ and $\boldsymbol{M}$ are triangular with strictly positive diagonal entries.
(next page)

Cont.
We had

$$\boldsymbol{M}^{-1}\boldsymbol{L} = \boldsymbol{M}^{*}(\boldsymbol{L}^{*})^{-1}$$

Since $\boldsymbol{M}$ and $\boldsymbol{L}$ are lower triangular, $\boldsymbol{M}^{-1}\boldsymbol{L}$ is lower triangular. Since $\boldsymbol{M}^{*}$ and $\boldsymbol{L}^{*}$ are upper triangular, $\boldsymbol{M}^{*}(\boldsymbol{L}^{*})^{-1}$ is upper triangular.

The lower triangular $\boldsymbol{M}^{-1}\boldsymbol{L}$ can be equal to the upper triangular $\boldsymbol{M}^{*}(\boldsymbol{L}^{*})^{-1}$ only if both matrices are diagonal. Therefore,

$$\boldsymbol{M}^{-1}\boldsymbol{L} = \boldsymbol{D} = \boldsymbol{M}^{*}(\boldsymbol{L}^{*})^{-1}$$

where $\boldsymbol{D}$ is a diagonal matrix. Now see that

$$\begin{aligned}(\boldsymbol{D}^{*})^{-1} &= ((\boldsymbol{M}^{*}(\boldsymbol{L}^{*})^{-1})^{*})^{-1} \\ &= (\boldsymbol{L}^{-1}\boldsymbol{M})^{-1} \\ &= \boldsymbol{M}^{-1}\boldsymbol{L} = \boldsymbol{D}\end{aligned}$$

and so we can then write

$$\boldsymbol{D}\boldsymbol{D}^{*} = (\boldsymbol{D}^{*})^{-1}\boldsymbol{D}^{*} = \boldsymbol{I}$$

This means that any diagonal entry of $\boldsymbol{D}$, which we can denote as $\boldsymbol{D}_{kk}$, satisfies

$$\boldsymbol{D}_{kk}\overline{\boldsymbol{D}_{kk}} = |\boldsymbol{D}_{kk}|^{2} = 1$$

See that this means that the diagonal entries of $\boldsymbol{D}$ are all located on the unit circle (the one with the real and complex axes).

Moreover, $\boldsymbol{D}$ needs to satisfy the constraint

$$\boldsymbol{L} = \boldsymbol{M}\boldsymbol{D}$$

where the diagonal entries of both $\boldsymbol{M}$ and $\boldsymbol{L}$ are *real and strictly positive*. The only way to satisfy this constraint while remaining on the unit circle is to have

$$\boldsymbol{D}_{kk} = 1$$

for all k. Therefore, $\boldsymbol{D} = \boldsymbol{I}$, and

$$\boldsymbol{L} = \boldsymbol{M}$$

6.11 Other tests for positive definiteness

6.11.1 $\boldsymbol{S} = \boldsymbol{A}^T\boldsymbol{A}$

A symmetric matrix $\boldsymbol{S}$ is positive definite if

$$\boldsymbol{S} = \boldsymbol{A}^T\boldsymbol{A} \text{ for some matrix } \boldsymbol{A} \text{ with independent columns}$$

See that if $\boldsymbol{S} = \boldsymbol{A}^T\boldsymbol{A}$, then

$$\boldsymbol{x}^T\boldsymbol{S}\boldsymbol{x} = \boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{A}\boldsymbol{x}) = ||\boldsymbol{A}\boldsymbol{x}||^2$$

See that the energy is positive (for $\boldsymbol{x} \neq \boldsymbol{0}$) provided that $\boldsymbol{A}\boldsymbol{x}$ is not the zero vector. To assure that $\boldsymbol{A}\boldsymbol{x} \neq \boldsymbol{0}$ when $\boldsymbol{x} \neq \boldsymbol{0}$, the columns of $\boldsymbol{A}$ must be independent.

If the columns are not independent, then there will be a nontrivial nullspace and $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ for some $\boldsymbol{x}$, in which case the energy can then be zero, but *will still never be negative*, meaning

$$\boldsymbol{x}^T\boldsymbol{S}\boldsymbol{x} \geq 0$$

As such, we can conclude that $\boldsymbol{S} = \boldsymbol{A}^T\boldsymbol{A}$ is at least semidefinite: $\boldsymbol{x}^T\boldsymbol{S}\boldsymbol{x} = ||\boldsymbol{A}\boldsymbol{x}||^2$ will be > 0 if the columns of $\boldsymbol{A}$ are independent (positive definite), or it will be ≥ 0 otherwise (positive semidefinite).

6.11.2 $\boldsymbol{LDL^T}$ factorisation and positive pivots

$\boldsymbol{LDL^T}$ factorisation
Assuming now row exchanges are needed, we have Gaussian elimination as $\boldsymbol{A} = \boldsymbol{LU}$ (instead of $\boldsymbol{PA} = \boldsymbol{LU}$). Looking at the upper triangular $\boldsymbol{U}$, see that we can factor out the pivots of each row separately into a diagonal matrix $\boldsymbol{D}$, leaaving 1's along the diagonal of the factored $\boldsymbol{U}_1$:

$$\boldsymbol{LU} = \boldsymbol{LDU}_1$$

If $\boldsymbol{A}$ were symmetric, this implies

$$\boldsymbol{LDU}_1 = \boldsymbol{U}_1^T\boldsymbol{DL}^T$$

assuming that $\boldsymbol{U}$ and $\boldsymbol{L}$ are invertible, then we have

$$(\boldsymbol{U}_1^T)^{-1}\boldsymbol{LD} = \boldsymbol{DL}^T\boldsymbol{U}_1^{-1}$$

Notice that the matrix on the left hand side is lower triangular, while that on the right hand side is upper triangular. The only way they can be equal is if both sides are diagonal.

$\boldsymbol{D}$ is diagonal, both $(\boldsymbol{U}_1^T)^{-1}\boldsymbol{L}$ and $\boldsymbol{L}^T\boldsymbol{U}_1^{-1}$ must be diagonal. Finally, because there are only 1's on the diagonals of both expressions, they must both be the identity. Therefore,

$$\boldsymbol{L} = \boldsymbol{U}_1^T, \quad \boldsymbol{L}^T = \boldsymbol{U}_1$$

and our original factorisation can be written as

$$\boldsymbol{A} = \boldsymbol{LDL}^T$$

Positive pivots as a test for positive definiteness
Every positive definite matrix $\boldsymbol{S}$ has a unique cholesky factorisation

$$\boldsymbol{S} = \boldsymbol{AA}^T$$

Where $\boldsymbol{A}$ has real positive diagonal entries. Working backward, we can factor out the (positive) diagonal entries of $\boldsymbol{A}$, leaving 1's on the diagonals:

$$\boldsymbol{S} = \boldsymbol{LD}'\boldsymbol{D}'\boldsymbol{L}^T = \boldsymbol{LDL}^T$$

Where $\boldsymbol{D} = \boldsymbol{D}'\boldsymbol{D}'$. See that $\boldsymbol{D}$ (the pivots) must have positive entries.

Since the cholesky decomposition is unique, this factorisation is also unique. See that this means that a positive definite matrix will have *positive pivots*, and that elimination of a positive definite matrix *will not require row exchanges*.

6.12

Appendix A

Misc. topics

A.1 Taylor series, Difference approximations

Taylor series
Recall the idea of the taylor series, where $f(x)$ near some point $x = x_0$ can be approximated as

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \frac{f'''(x_0)}{3!}(x - x_0)^3 + \cdots$$

See that this comes from approximating $f(x)$ as a polynomial

$$y(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + a_3(x - x_0)^3 + \cdots$$

where we have $a_0 = y(x_0)$; now consider differentiating:

$$\begin{aligned} y'(x) &= a_1 + 2a_2(x - x_0) + 3a_3(x - x_0)^2 + \cdots \\ y''(x) &= 2a_2 + (3)(2)a_3(x - x_0) + (4)(3)a_4(x - x_0)^2 \cdots \\ y^{(n)}(x) &= n!\, a_n + ((n+1) \cdot (n-1) \cdots 3 \cdot 2)a_{n+1}(x - x_0) + \cdots \end{aligned}$$

so we have

$$a_1 = y'(x_0), \quad a_2 = \frac{y''(x_0)}{2!}, \cdots, a_n = \frac{y^{(n)}(x_0)}{n!}$$

where substituting into our initial approximation gives us

$$y(x) \approx y(x_0) + y'(x_0)(x - x_0) + \frac{y''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{y^{(n)}(x_0)}{n!}(x - x_0)^n + \cdots$$

(next page)

Cont.
We had

$$y(x) \approx y(x_0) + y'(x_0)(x - x_0) + \frac{y''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{y^{(n)}(x_0)}{n!}(x - x_0)^n + \cdots$$

substituting $x - x_0 = h$ gives us

$$y(x_0 + h) \approx y(x_0) + y'(x_0)(h) + \frac{y''(x_0)}{2!}(h)^2 + \cdots + \frac{y^{(n)}(x_0)}{n!}(h)^n + \cdots$$

Which can be rewritten as

$$y(x + h) \approx y(x) + hy'(x) + \frac{1}{2!}h^2y''(x) + \cdots + \frac{1}{n!}h^n y^{(n)}(x) + \cdots$$

Difference formulas
Considering the first few terms of the taylor series approximation, we have

$$y(x + h) - y(x) \approx hy'(x) + \frac{1}{2}h^2y''(x)$$
$$y(x - h) - y(x) \approx -hy'(x) + \frac{1}{2}h^2y''(x)$$

This allows us to have an approximation to dy/dx:

Centered difference	$\boldsymbol{\dfrac{y(x+h) - y(x-h)}{2h} \approx \dfrac{dy}{dx}}$

and also to the second derivative

Second difference	$\boldsymbol{\dfrac{y(x+h) - 2y(x) + y(x-h)}{h^2} \approx \dfrac{d^2y}{dx^2}}$

The individual formulas also yield approximations:

Forward difference	$\boldsymbol{\dfrac{dy}{dx} = \dfrac{y(x+h) - y(x)}{h} + O(h)}$ error
Backward difference	$\boldsymbol{\dfrac{dy}{dx} = \dfrac{y(x) - y(x-h)}{h} + O(h)}$ error

A.2 Conjugate/Hermitian transpose

The conjugate transpose of a matrix $\boldsymbol{A}$ is the matrix $\boldsymbol{A}^*$ defined by

$$\boldsymbol{A}^* = \bar{\boldsymbol{A}}^T = \bar{\boldsymbol{A}^T}$$

Where T denotes transposition and the bar denotes complex conjugation—transpose, and complex conjugation of all entries. The order in which the transposition and conjugation are performed is irrelevant.

Properties
The properties of conjugate transpostition are immediate consequences of the properties of transposition and conjugation. First, recall that for two complex numbers $x = (a + ib)$ and $y = (c + id)$:

$$\bar{(xy)} = \bar{x}\bar{y}$$

which can be shown:

$$xy = (a + ib)(c + id) = ac - bd + i(ad + bc)$$

and

$$\bar{x}\bar{y} = (a - ib)(c - id) = ac - bd - i(ad + bc) = \bar{(xy)}$$

with that we list some properties of conjugate transpositions:

- $(\boldsymbol{A}^*)^* = \boldsymbol{A}$
- $(\boldsymbol{A} + \boldsymbol{B})^* = \boldsymbol{A}^* + \boldsymbol{B}^*$
- $(z\boldsymbol{A})^* = \bar{z}\boldsymbol{A}^*$
- $(\boldsymbol{A}\boldsymbol{B})^* = \boldsymbol{B}^*\boldsymbol{A}^*$
- $(\boldsymbol{A}^*)^{-1} = (\boldsymbol{A}^{-1})^*$ (provided $\boldsymbol{A}$ is square invertible)

The proof of the final point is analagous to the proof of $(\boldsymbol{A})^T)^{-1} = (\boldsymbol{A}^{-1})^T$.
(next page)

Hermitian matrix
A matrix that is equal to its conjugate transpose is called *Hermitian* or *self-adjoint*. In other words, $\boldsymbol{A}$ is Hermitian if and only if

$$\boldsymbol{A}^* = \boldsymbol{A}$$

Denoting $\boldsymbol{A}_{kl}$ as the (k,l)-th entry of $\boldsymbol{A}$ and $(\boldsymbol{A}^*)_{kl}$ the (k,l)-th entry of $\boldsymbol{A}^*$, we have

$$(\boldsymbol{A}^*)_{kl} = \overline{\boldsymbol{A}_{lk}}$$

Therefore, $\boldsymbol{A}$ is Hermitian if and only if

$$\boldsymbol{A}_{kl} = \overline{\boldsymbol{A}_{lk}}$$

for all k and l. See that this implies that the *diagonal entries of* $\boldsymbol{A}$ *must be real* in order to satisfy

$$\boldsymbol{A}_{kk} = \overline{\boldsymbol{A}_{kk}}$$

A.3