Strang

Malcolm

Started 5th June 2025

Contents

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

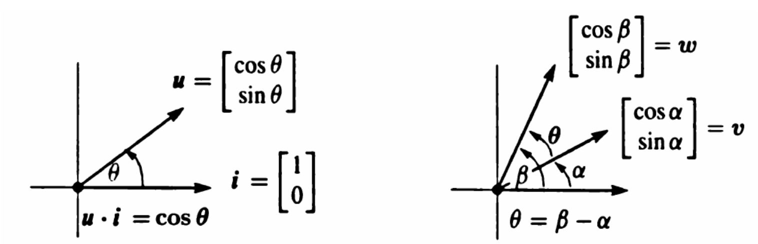


Figure 1.5: Unit vectors: $\boldsymbol{u} \cdot \boldsymbol{i} = \cos\theta$. The angle between the vectors is θ.

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 $< \pi/2$):

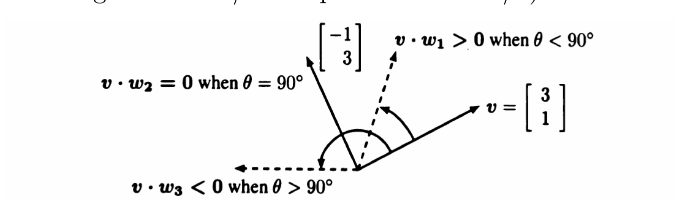


Figure 1.6: Small angle $\boldsymbol{v} \cdot \boldsymbol{w_1} > 0$. Right angle $\boldsymbol{v} \cdot \boldsymbol{w_2} = 0$. Large angle $\boldsymbol{v} \cdot \boldsymbol{w_3} < 0$.

(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 $\boldsymbol{A} = \boldsymbol{C}\boldsymbol{R}$. 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

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

(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 $\boldsymbol{C}\boldsymbol{F}$ of the independent columns in $\boldsymbol{C}$.

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

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

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 $\boldsymbol{A} = \boldsymbol{C}\boldsymbol{R}$)
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

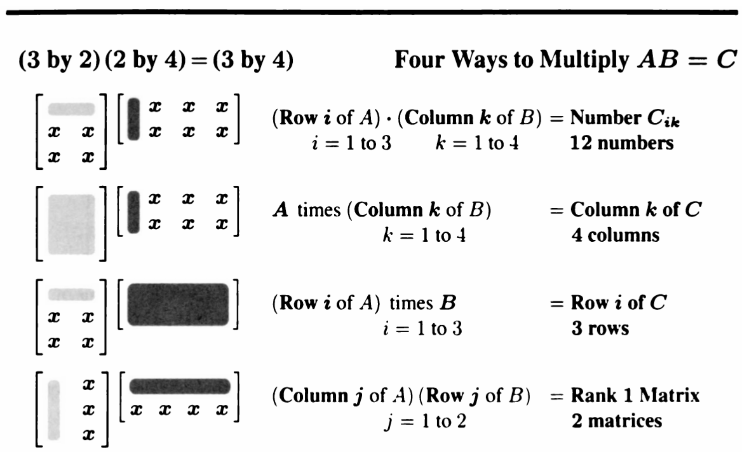
A lesser known way to multiply $\boldsymbol{AB}$ is through considering the columns of $\boldsymbol{A}$ and the rows of $\boldsymbol{B}$ (contrary to the usual ideas where each entry of the result is a dot product of a row of $\boldsymbol{A}$ and column of $\boldsymbol{B}$):

$$\boxed{\boldsymbol{AB} = \underbrace{\begin{bmatrix} | & & | \\ \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_n \\ | & & | \end{bmatrix}}_{\textbf{columns } \boldsymbol{a}_k} \underbrace{\begin{bmatrix} - & \boldsymbol{b}_1^* & - \\ & \vdots & \\ - & \boldsymbol{b}_n^* & - \end{bmatrix}}_{\textbf{rows } \boldsymbol{b}_k^*} = \underbrace{\boldsymbol{a}_1\boldsymbol{b}_1^* + \boldsymbol{a}_2\boldsymbol{b}_2^* + \cdots + \boldsymbol{a}_n\boldsymbol{b}_n^*.}_{\textbf{Add columns } \boldsymbol{a}_k \textbf{ times rows } \boldsymbol{b}_k^*}}$$

We multiply each column of $\boldsymbol{A}$ by each row of $\boldsymbol{B}$; 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 $\boldsymbol{A}$ specified by the ith column of $\boldsymbol{B}$. 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 $\boldsymbol{A}$ to the final product. By summing this over all n columns of $\boldsymbol{A}$ we get the result.)

Summary of methods



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

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

Chapter 2

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

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 $\boldsymbol{Ax} = \mathbf{0}$ is $\boldsymbol{x} = \mathbf{0}$. $\boldsymbol{A}$ *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 $\boldsymbol{b} = \mathbf{0}$ (this is the meaning of dependent columns), and so there are infinitely many solutions to $\boldsymbol{AX} = \mathbf{0}$.

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 $\boldsymbol{AX} = \mathbf{0}$:

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

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)

$\boldsymbol{L}$ is the inverse of $\boldsymbol{E}$

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

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

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 $\boldsymbol{E}_{21}^{-1}\boldsymbol{E}_{31}^{-1}\boldsymbol{E}_{32}^{-1} = \boldsymbol{E}^{-1} = \boldsymbol{L}$ (see that inverses need to be applied in reverse):

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

(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 $\boldsymbol{L}$. 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 $\boldsymbol{A} = \boldsymbol{LU}$ to go back from triangular $\boldsymbol{U}$ to the original $\boldsymbol{A}$.

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:

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

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

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

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:

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

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

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

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:

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

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.

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

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

$$\boxed{\underset{\textbf{columns times rows}}{\boldsymbol{A} = \boldsymbol{\ell}_1\boldsymbol{u}_1 + \boldsymbol{\ell}_2\boldsymbol{u}_2 + \cdots + \boldsymbol{\ell}_n\boldsymbol{u}_n} = \begin{bmatrix} \boldsymbol{\ell}_1 & \cdots & \boldsymbol{\ell}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1 \\ \vdots \\ \boldsymbol{u}_n \end{bmatrix} = \boldsymbol{LU}}$$

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:

P has the rows of I
Px = Circular shift of x
1, 2, 3 to 3, 1, 2

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

(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

Px reverses the order to $x_4 . x_3 . x_2, x_1$; $P^2 = I$

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

x_4 stays fixed; Input x_1, x_2, x_3, x_4; output x_3, x_1, x_2, x_4; $P^3 = I$

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

Circular shift; Input x_1, x_2, x_3, x_4; output x_4, x_1, x_2, x_3; $P^4 = I$

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

Evens before odds; x_0, x_2 before 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 $\boldsymbol{PA} = \boldsymbol{LU}$.

Inverse

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

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

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

(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

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

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:

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

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:

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

Column permutations
We know that we can reorder rows using

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

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

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

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 $\boldsymbol{AB}$ is $\boldsymbol{B}^T\boldsymbol{A}^T$: 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 $(\boldsymbol{A}^T)^{-1} = (\boldsymbol{A}^{-1})^T$

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

$$\textbf{Transpose of inverse} \qquad A^{-1}A = I \quad \textbf{is transposed to} \quad A^{\mathrm{T}}(A^{-1})^{\mathrm{T}} = I.$$

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 $\boldsymbol{A}^T\boldsymbol{A}$ 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:

$$\textit{The transpose of} \quad \mathbf{A^T A} \quad \textit{is} \quad \mathbf{A^T (A^T)^T} \quad \textit{which is} \quad \mathbf{A^T A} \ \textit{again.}$$

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 $\boldsymbol{R}$:

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

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:

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

We have $\boldsymbol{H} = \boldsymbol{W}\boldsymbol{F}$, 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:

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

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 $\boldsymbol{\ell}$:

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

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 $\boldsymbol{A} = \boldsymbol{CR}$ 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,

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

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

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

3.3 Systematic nullspace computation

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

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

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

$$\boldsymbol{Ax} = \mathbf{0} \implies \boldsymbol{W^{-1}Ax} = \boldsymbol{W^{-1}}\mathbf{0} \implies \boldsymbol{R}_0\boldsymbol{x} = \mathbf{0}$$

See that the elimination process doesn't change the space of solutions that satisfy $\boldsymbol{Ax} = \mathbf{0}$. 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 $\boldsymbol{Rx} = \mathbf{0}$:

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

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

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

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

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

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

(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 $\boldsymbol{Ax}=\mathbf{0}$, 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 $\boldsymbol{Ax} = \boldsymbol{b}$

Finding a particular solution

We can reduce $\boldsymbol{Ax} = \boldsymbol{b}$ to a simpler system $\boldsymbol{R}_0\boldsymbol{x} = \boldsymbol{d}$ with the same solutions (if any). A useful aid here is to augment $\boldsymbol{A}$ with $\boldsymbol{b}$ to produce an augmented matrix $[\boldsymbol{A}, \boldsymbol{b}]$:

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

Applying elimination to reach $\boldsymbol{R}_0$, the augmented part also undergoes the same elimination steps to produce a $\boldsymbol{d}$:

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

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 $\boldsymbol{d}$:

$$R_0\boldsymbol{x}_p = \begin{bmatrix} \mathbf{1} & 3 & \mathbf{0} & 2 \\ \mathbf{0} & 0 & \mathbf{1} & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{1} \\ 0 \\ \mathbf{6} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{1} \\ \mathbf{6} \\ 0 \end{bmatrix} = \boldsymbol{d} \qquad \begin{array}{l} \textbf{Pivot variables 1, 6} \\ \textbf{Free variables 0, 0} \\ \boldsymbol{x}_{\textbf{particular}} = \mathbf{(1, 0, 6, 0)}. \end{array}$$

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***	$A\boldsymbol{x}_p = \boldsymbol{b}$
All $\boldsymbol{x}_{\text{nullspace}}$	***The $n - r$ special solutions solve***	$A\boldsymbol{x}_n = \mathbf{0}$

Complete solution
one $\boldsymbol{x}_p$+many $\boldsymbol{x}_n$
particular $\boldsymbol{x}_p$
nullspace vectors $\boldsymbol{x}_n$

$$\boldsymbol{x} = \boldsymbol{x}_p + \boldsymbol{x}_n = \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix} + \boldsymbol{x}_2 \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \boldsymbol{x}_4 \begin{bmatrix} -2 \\ 0 \\ -4 \\ 1 \end{bmatrix}.$$

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** $(\boldsymbol{r = n})$ 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*** $(\boldsymbol{r{=}m})$ has all these properties:

1. **All rows have pivots, and $\boldsymbol{R}_0$ has no zero rows: $\boldsymbol{R}_0 = \boldsymbol{R}$.**
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 $\mathbf{R}^n$.	$\boldsymbol{r}$
2.	The ***column space*** is $\mathbf{C}(A)$, a subspace of $\mathbf{R}^m$.	$\boldsymbol{r}$
3.	The ***nullspace*** is $\mathbf{N}(A)$, a subspace of $\mathbf{R}^n$.	$\boldsymbol{n-r}$
4.	The ***left nullspace*** is $\mathbf{N}(A^{\mathrm{T}})$, a subspace of $\mathbf{R}^m$.	$\boldsymbol{m-r}$

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 Block elimination

3.7 Every rank r matrix is a sum of r rank one matrices

Rank one matrices have the form $\boldsymbol{uv^T}$. Recall the $\boldsymbol{A} = \boldsymbol{CR}$ factorisation, where we reduce the matrix by row operations to $\boldsymbol{R}_0$, which has the same row space as $\boldsymbol{A}$, and remove the zero rows to find $\boldsymbol{R}$, which also has the same row space. For instance:

$$\textbf{Rank two} \quad \boldsymbol{A} = \begin{bmatrix} 1 & 0 & 3 \\ 1 & 1 & 7 \\ 4 & 2 & 20 \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} \\ \mathbf{4} & \mathbf{2} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{3} \\ \mathbf{0} & \mathbf{1} & \mathbf{4} \end{bmatrix} = \boldsymbol{CR}$$

For each independent column in $\boldsymbol{A}$, a new independent column is added to $\boldsymbol{C}$ and a new indentity component added to $\boldsymbol{R}$. On the other hand for each dependent column in $\boldsymbol{A}$, only a new column is added to $\boldsymbol{R}$ and $\boldsymbol{C}$ is left unchanged. Each independent column in $\boldsymbol{C}$ contributes a rank 1 matrix:

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

$$y(x + h) \approx y(x) + hy'(x) + \frac{1}{2!}h^2 y''(x) + \cdots + \frac{1}{n!}h^n y^{(n)}(x) + \cdots$$

Difference formulas
Considering the first few terms of the taylor series approximation, we have

$$y(x + h) - y(x) \approx hy'(x) + \frac{1}{2}h^2 y''(x)$$
$$y(x - h) - y(x) \approx -hy'(x) + \frac{1}{2}h^2 y''(x)$$

This allows us to have an approximation to dy/dx:

Centered difference $$\frac{y(x+h) - y(x-h)}{2h} \approx \frac{dy}{dx}$$

and also to the second derivative

Second difference $$\frac{y(x+h) - 2y(x) + y(x-h)}{h^2} \approx \frac{d^2y}{dx^2}$$

The individual formulas also yield approximations:

Forward difference $$\frac{dy}{dx} = \frac{y(x+h) - y(x)}{h} + O(h) \text{ error}$$

Backward difference $$\frac{dy}{dx} = \frac{y(x) - y(x-h)}{h} + O(h) \text{ error}$$