Linear mathematics for applications MAS201

Lecture 1

Introduction

- The lecturer is Professor Neil Strickland.
- The best way to contact me is by email: `N.P.Strickland@sheffield.ac.uk`
- The timetable is as follows:
  - Lecture on Tuesday at 12.00 in the Student Union Auditorium
  - Lecture on Wednesday at 13.00 in Dainton Building Lecture Theatre 1
  - Tutorial in weeks 1, 3, 5, 8, 10: Tuesday 15.00, Wednesday 11.00 or Friday at 10.00.
- The course web page is `http://www.shef.ac.uk/nps/courses/MAS201`.
- I do not plan to use MOLE.
- There will be online tests as for MAS100, and also some homework assignments to be done on paper. Arrangements will be announced when they are finalised.
- However, your mark for the course will be based solely on the final exam.

Background to the course

- This course is mainly about the theory of matrices.
- The (i,j)'th entry in a matrix could represent
  - The brightness of the (i,j)'th pixel in a digitised image (relevant to image processing).
  - The probability that the i'th word in the dictionary will be followed by the j'th word, in typical english text (relevant to machine translation).
  - The number of links from the i'th website to the j'th website, in some list of websites (relevant to search engine design).
  - The response of the i'th patient to the j'th drug in a clinical trial.
  - Many other things.
- We will learn how to calculate many things using matrices. Row reduction is a key ingredient in many methods of calculation. We will either use matrices for which row reduction is easy, or get Maple to do the work. Our main task is to learn how to convert other kinds of questions to row-reduction questions, and to interpret the results.
- Eigenvalues and eigenvectors will be another important ingredient.
- A few applications will be treated in more detail: solution of difference equations; solution of differential equations; long-term behaviour of random systems known as Markov chains.

Notation

- $\mathbb{R}$ is the set of all real numbers ("scalars") so 17, π, $\frac{123}{456} \in \mathbb{R}$ but $1+i \notin \mathbb{R}$.
- $\mathbb{R}^n$ is the set of column vectors with n entries, so

$$\begin{bmatrix}10\\20\\30\end{bmatrix} \in \mathbb{R}^3 \qquad \begin{bmatrix}\pi\\\pi^2\\\pi^3\\\pi^4\end{bmatrix} \in \mathbb{R}^4 \qquad \begin{bmatrix}12.38\\-9.14\end{bmatrix} \in \mathbb{R}^2.$$

- $M_{m\times n}(\mathbb{R})$ is the set of all $m \times n$ matrices (with m rows and n columns, ie height m and width n)

$$\begin{bmatrix}1&2&3\\4&5&6\end{bmatrix} \in M_{2\times 3}(\mathbb{R}) \qquad \begin{bmatrix}1&2\\3&4\\5&6\end{bmatrix} \in M_{3\times 2}(\mathbb{R})$$

a 2×3 matrix   a 3×2 matrix

- $M_n(\mathbb{R}) = M_{n\times n}(\mathbb{R})$ is the set of all $n \times n$ square matrices. I_n is the $n \times n$ identity matrix.

$$\begin{bmatrix}1&2&2&1\\2&3&3&2\\2&3&3&2\\1&2&2&1\end{bmatrix} \in M_4(\mathbb{R}) \qquad I_4 = \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} \in M_4(\mathbb{R})$$

Notation

- The *transpose* of an $m \times n$ matrix A is the $n \times n$ matrix A^T obtained by flipping A over, so the (i,j)'th entry in A^T is the same as the (j,i)'th entry in A. For example, we have

$$\begin{bmatrix}a_1&a_2&a_3&a_4\\b_1&b_2&b_3&b_4\end{bmatrix}^T = \begin{bmatrix}a_1&b_1\\a_2&b_2\\a_3&b_3\\a_4&b_4\end{bmatrix}.$$

- Note also that the transpose of a row vector is a column vector, for example

$$\begin{bmatrix}5&6&7&8\end{bmatrix}^T = \begin{bmatrix}5\\6\\7\\8\end{bmatrix}.$$

- We will typically write column vectors in this way when it is convenient to lay things out horizontally.

Reminder about dot products

For column vectors $u, v \in \mathbb{R}^n$, the dot product is

$$u.v = u_1v_1 + \cdots + u_nv_n = \sum_{i=1}^n u_iv_i.$$

For example: $$\begin{bmatrix}1\\2\\3\\4\end{bmatrix} . \begin{bmatrix}1000\\100\\10\\1\end{bmatrix} = 1000 + 200 + 30 + 4 = 1234.$$

Product of a matrix and a vector

We can multiply an $m \times n$ matrix by a vector in $\mathbb{R}^n$ to get a vector in $\mathbb{R}^m$, for example

$$\begin{bmatrix}a&b&c\\d&e&f\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}ax+by+cz\\dx+ey+fz\end{bmatrix}$$

$$(2 \times 3 \text{ matrix})(\text{vector in } \mathbb{R}^3) = (\text{vector in } \mathbb{R}^2)$$

General rule: divide A into n columns u_i (each u_i in $\mathbb{R}^m$) or into m rows v_j^T (each v_j in $\mathbb{R}^n$)

$$A = \left[\begin{array}{c|c|c} u_1 & \cdots & u_n \end{array}\right] = \left[\begin{array}{c} v_1^T \\ \hline \vdots \\ \hline v_m^T \end{array}\right].$$

Now let $t = \begin{bmatrix}t_1 & \cdots & t_n\end{bmatrix}^T$ be a vector in $\mathbb{R}^n$. The rule is then

$$At = \left[\begin{array}{c} v_1^T \\ \hline \vdots \\ \hline v_m^T \end{array}\right] t = \begin{bmatrix}v_1.t\\\vdots\\v_m.t\end{bmatrix} = t_1u_1 + \cdots + t_nu_n.$$

Product of a matrix and a vector

In the example

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} ax+by+cz \\ dx+ey+fz \end{bmatrix}$$

$$(2 \times 3 \text{ matrix})(\text{vector in } \mathbb{R}^3) = (\text{vector in } \mathbb{R}^2)$$

we have

$$v_1 = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \qquad v_2 = \begin{bmatrix} d \\ e \\ f \end{bmatrix} \qquad t = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad At = \begin{bmatrix} v_1.t \\ v_2.t \end{bmatrix} = \begin{bmatrix} ax+by+cz \\ dx+ey+fz \end{bmatrix}$$

Also

$$u_1 = \begin{bmatrix} a \\ d \end{bmatrix} \qquad u_2 = \begin{bmatrix} b \\ e \end{bmatrix} \qquad u_3 = \begin{bmatrix} c \\ f \end{bmatrix} \qquad t_1 = x \qquad t_2 = y \qquad t_3 = z$$

so

$$t_1u_1 + t_2u_2 + t_3u_3 = x\begin{bmatrix} a \\ d \end{bmatrix} + y\begin{bmatrix} b \\ e \end{bmatrix} + z\begin{bmatrix} c \\ f \end{bmatrix} = \begin{bmatrix} ax+by+cz \\ dx+ey+fz \end{bmatrix} = At$$

as expected.

Product of two matrices

We can multiply an $m \times n$ matrix A by an $n \times p$ matrix B to get an $m \times p$ matrix AB:

$$A = \left[\begin{array}{c} v_1^T \\ \hline \vdots \\ \hline v_m^T \end{array}\right] \qquad B = \left[\begin{array}{c|c|c} w_1 & \cdots & w_p \end{array}\right]$$

$$AB = \left[\begin{array}{c} v_1^T \\ \hline \vdots \\ \hline v_m^T \end{array}\right] \left[\begin{array}{c|c|c} w_1 & \cdots & w_p \end{array}\right] = \begin{bmatrix} v_1.w_1 & \cdots & v_1.w_p \\ \vdots & \ddots & \vdots \\ v_m.w_1 & \cdots & v_m.w_p \end{bmatrix}$$

$AB \neq BA$

If A and B are numbers then of course $AB = BA$, but this does not work in general for matrices. Suppose that A is an $m \times n$ matrix and B is an $n \times p$ matrix, so we can define AB as before.

(a) Firstly, BA may not even be defined. It is only defined if the number of columns of B is the same as the number of rows of A, or in other words $p = m$.

(b) Suppose that $p = m$, so A is an $m \times n$ matrix, and B is an $n \times m$ matrix, and both AB and BA are defined. We find that AB is an $m \times m$ matrix and BA is an $n \times n$ matrix. Thus, it is not meaningful to ask whether $AB = BA$ unless $m = n$.

(c) Suppose that $m = n = p$, so both A and B are square matrices of shape $n \times n$. This means that AB and BA are also $n \times n$ matrices. However, they are usually not equal. For example, we have

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 10 & 10 & 10 \\ 100 & 100 & 100 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 20 & 20 & 20 \\ 300 & 300 & 300 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 10 & 10 & 10 \\ 100 & 100 & 100 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 10 & 20 & 30 \\ 100 & 200 & 300 \end{bmatrix}.$$

$(AB)^T = B^TA^T$

$$AB = \left[\begin{array}{c} u_1^T \\ \hline \vdots \\ \hline u_m^T \end{array}\right] \left[\begin{array}{c|c|c} v_1 & \cdots & v_p \end{array}\right] = \begin{bmatrix} u_1.v_1 & \cdots & u_1.v_p \\ \vdots & \ddots & \vdots \\ u_m.v_1 & \cdots & u_m.v_p \end{bmatrix}$$

$$(AB)^T = \begin{bmatrix} u_1.v_1 & \cdots & u_m.v_1 \\ \vdots & \ddots & \vdots \\ u_1.v_p & \cdots & u_m.v_p \end{bmatrix} = \begin{bmatrix} v_1.u_1 & \cdots & v_1.u_m \\ \vdots & \ddots & \vdots \\ v_p.u_1 & \cdots & v_p.u_m \end{bmatrix}$$

$$B^TA^T = \left[\begin{array}{c} v_1^T \\ \hline \vdots \\ \hline v_p^T \end{array}\right] \left[\begin{array}{c|c|c} u_1 & \cdots & u_m \end{array}\right] = \begin{bmatrix} v_1.u_1 & \cdots & v_1.u_m \\ \vdots & \ddots & \vdots \\ v_p.u_1 & \cdots & v_p.u_m \end{bmatrix} = (AB)^T$$

$(AB)^T = B^T A^T$ for 2×2 matrices

For $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $B = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$ we have

$$AB = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} p & q \\ r & s \end{bmatrix} = \begin{bmatrix} ap+br & aq+bs \\ cp+dr & cq+ds \end{bmatrix}$$

$$(AB)^T = \begin{bmatrix} ap+br & aq+bs \\ cp+dr & cq+ds \end{bmatrix}^T = \begin{bmatrix} ap+br & cp+dr \\ aq+bs & cq+ds \end{bmatrix}$$

$$B^T A^T = \begin{bmatrix} p & r \\ q & s \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} pa+rb & pc+rd \\ qa+sb & qc+sd \end{bmatrix} = (AB)^T.$$

Matrices and linear equations

Systems of linear equations can be rewritten as matrix equations. Consider the equations

$$\begin{aligned} w + 2x + 3y + 4z &= 1 \\ 5w + 6x + 7y + 8z &= 10 \\ 9w + 10x + 11y + 12z &= 100 \end{aligned}$$

Note that

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} w + 2x + 3y + 4z \\ 5w + 6x + 7y + 8z \\ 9w + 10x + 11y + 12z \end{bmatrix}$$

So our system of equations is equivalent to the single matrix equation

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 10 \\ 100 \end{bmatrix}.$$

Matrices and linear equations

Systems of linear equations can be rewritten as matrix equations.

$$\begin{aligned} a + b + c &= 10 \\ a + 2b + 4c &= 20 \\ a + 3b + 9c &= 30 \\ a + 4b + 16c &= 40 \\ a + 5b + 25c &= 50 \end{aligned} \qquad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 10 \\ 20 \\ 30 \\ 40 \\ 50 \end{bmatrix}.$$

The *augmented matrix* for an equation $Au = v$ is $[A|v]$:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 10 \\ 1 & 2 & 4 & 20 \\ 1 & 3 & 9 & 30 \\ 1 & 4 & 16 & 40 \\ 1 & 5 & 25 & 50 \end{array}\right]$$

Tidying

Sometimes we need to tidy up first.

$$\begin{aligned} p + 7s &= q + 1 \\ 5r + 1 &= 7q - p \\ r + s &= p + q \end{aligned} \qquad \begin{array}{cccccc} p & -q & +0r & +7s & = & 1 \\ p & -7q & +5r & +0s & = & -1 \\ p & +q & -r & -s & = & 0 \end{array}$$

$$\begin{bmatrix} 1 & -1 & 0 & 7 \\ 1 & -7 & 5 & 0 \\ 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} p \\ q \\ r \\ s \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

The augmented matrix is

$$\left[\begin{array}{cccc|c} 1 & -1 & 0 & 7 & 1 \\ 1 & -7 & 5 & 0 & -1 \\ 1 & 1 & -1 & -1 & 0 \end{array}\right]$$

Lecture 2

(Reduced) row-echelon form

Definition 5.1: Let A be a matrix of real numbers. Recall that A is said to be in *reduced row-echelon form* (RREF) if the following hold:

RREF0: Any rows of zeros come at the bottom of the matrix, after all the nonzero rows.

RREF1: In any nonzero row, the first nonzero entry is equal to one. These entries are called *pivots*.

RREF2: In any nonzero row, the pivot is further to the right than the pivots in all previous rows.

RREF3: If a column contains a pivot, then all other entries in that column are zero.

We will also say that a system of linear equations (in a specified list of variables) is in RREF if the corresponding augmented matrix is in RREF.

If RREF0, RREF1 and RREF2 are satisfied but not RREF3 then we say that A is in (unreduced) row-echelon form.

(Reduced) row-echelon form

Example 5.2:

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

$$C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ \mathbf{1} & 0 & 0 \end{bmatrix} \qquad D = \begin{bmatrix} 1 & 0 & \mathbf{2} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

A is not in RREF because the middle row is zero and the bottom row is not (RREF0 fails).The matrix B is also not in RREF because the first nonzero entry in the top row is 2 rather than 1 (RREF1 fails). The matrix C is not in RREF because the pivot in the bottom row is to the left of the pivots in the previous rows (RREF2 fails). The matrix D is not in RREF because the last column contains a pivot and also another nonzero entry (RREF3 fails). On the other hand, the matrix

$$E = \begin{bmatrix} \mathbf{1} & 2 & 0 & 3 & 0 & 4 \\ 0 & 0 & \mathbf{1} & 5 & 0 & 6 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

is in RREF.

RREF for systems of equations

Example 5.3: The system of equations

$$\begin{aligned} x - z &= 1 \\ y &= 2 \end{aligned}$$

is in RREF because its augmented matrix is in RREF:

$$A = \left[\begin{array}{ccc|c} 1 & 0 & -1 & 1 \\ 0 & 1 & 0 & 2 \end{array}\right]$$

The system of equations

$$\begin{aligned} x + y + z &= 1 \\ y + z &= 2 \\ z &= 3 \end{aligned}$$

is not in RREF because its augmented matrix is not in RREF:

$$B = \left[\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

Solving RREF systems

If a system of equations is in RREF, it can be solved very easily.

$$\begin{aligned} w + 2x + 3z &= 10 \\ y + 4z &= 20. \end{aligned} \qquad \left[\begin{array}{cccc|c} 1 & 2 & 0 & 3 & 10 \\ 0 & 0 & 1 & 4 & 20 \end{array}\right]$$

Variables in non-pivot columns are *independent*; they can take any value, and we move them to the right hand side. Variables in pivot columns are *dependent*; they stay on the left. The equations now express the dependent variables in terms of the independent ones.

$$\begin{aligned} w &= 10 - 2x - 3z \\ y &= 20 - 4z \end{aligned}$$

Sometimes it is convenient to introduce new letters for the independent variables, say λ and μ. Then the solution is

$$\begin{aligned} w &= 10 - 2\lambda - 3\mu \\ x &= \lambda \\ y &= 20 - 4\mu \\ z &= \mu \end{aligned}$$

where λ and μ can take arbitrary values.

Solving RREF systems — degenerate cases

The augmented matrix

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 10 \\ 0 & 1 & 0 & 0 & 11 \\ 0 & 0 & 1 & 0 & 12 \\ 0 & 0 & 0 & 1 & 13 \end{array}\right]$$

has a pivot in every column to the left of the bar, so there are no independent variables. It corresponds to the system

$$w = 10 \qquad x = 11 \qquad y = 12 \qquad z = 13$$

which is its own (unique) solution.
The augmented matrix

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

has a pivot in the last column, to the right of the bar. It corresponds to the system

$$\begin{aligned} w + z &= 0 \\ 0 &= 1 \end{aligned} \qquad \begin{aligned} x + y &= 0 \\ 0 &= 0 \end{aligned}$$

so there is clearly no solution.

Row operations

Let A be a matrix. The following operations on A are called *elementary row operations*:

ERO1: Exchange two rows.

ERO2: Multiply a row by a nonzero constant.

ERO3: Add a multiple of one row to another row.

We write $A \to B$ if A can be converted to B by a sequence of EROs. As all EROs are reversible, we see that if $A \to B$ then also $B \to A$.

Theorem
Let A be a matrix.

(a) *By applying a sequence of row operations to A, one can obtain a matrix B that is in RREF.*

(b) *Although there are various different sequences that reduce A to RREF, they all give the same matrix B at the end of the process.*

In a moment we will recall the standard procedure for row-reduction. It is not hard to prove (by induction on the number of rows) that this procedure always works as advertised, so (a) is true. Statement (b) is an important fact but we will not prove it in this course.

Row reduction

To reduce a matrix A to RREF, we do the following.

(a) If all rows are zero, then A is already in RREF, so we are done.

(b) Otherwise, we find a row that has a nonzero entry as far to the left as possible. Let this entry be u, in the k'th column of the j'th row say. Because we went as far to the left as possible, all entries in columns 1 to $k-1$ of the matrix are zero.

(c) We now exchange the first row with the j'th row (which does nothing if j happens to be equal to one).

(d) Next, we multiply the first row by u^{-1}. We now have a 1 in the k'th column of the first row.

(e) We now subtract multiples of the first row from all the other rows to ensure that the k'th column contains nothing except for the pivot in the first row.

(f) We now ignore the first row and apply row operations to the remaining rows to put them in RREF.

(g) If we put the first row back in, we have a matrix that is nearly in RREF, except that the first row may have nonzero entries above the pivots in the lower rows. This can easily be fixed by subtracting multiples of those lower rows.

Row reduction example

Consider the following sequence of reductions:

$$\begin{bmatrix} 0 & 0 & -2 & -1 & -13 \\ -1 & -2 & -1 & 1 & -2 \\ -1 & -2 & 0 & -1 & -8 \end{bmatrix} \xrightarrow{1} \begin{bmatrix} -1 & -2 & -1 & 1 & -2 \\ 0 & 0 & -2 & -1 & -13 \\ -1 & -2 & 0 & -1 & -8 \end{bmatrix} \xrightarrow{2} \begin{bmatrix} 1 & 2 & 1 & -1 & 2 \\ 0 & 0 & -2 & -1 & -13 \\ -1 & -2 & 0 & -1 & -8 \end{bmatrix} \xrightarrow{3}$$

$$\begin{bmatrix} 1 & 2 & 1 & -1 & 2 \\ 0 & 0 & -2 & -1 & -13 \\ 0 & 0 & 1 & -2 & -6 \end{bmatrix} \xrightarrow{4} \begin{bmatrix} 1 & 2 & 1 & -1 & 2 \\ 0 & 0 & 1 & 1/2 & 13/2 \\ 0 & 0 & 1 & -2 & -6 \end{bmatrix} \xrightarrow{5} \begin{bmatrix} 1 & 2 & 1 & -1 & 2 \\ 0 & 0 & 1 & 1/2 & 13/2 \\ 0 & 0 & 0 & -5/2 & -25/2 \end{bmatrix} \xrightarrow{6}$$

$$\begin{bmatrix} 1 & 2 & 1 & -1 & 2 \\ 0 & 0 & 1 & 1/2 & 13/2 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix} \xrightarrow{7} \begin{bmatrix} 1 & 2 & 1 & -1 & 2 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix} \xrightarrow{8} \begin{bmatrix} 1 & 2 & 0 & -1 & -2 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix} \xrightarrow{9} \begin{bmatrix} 1 & 2 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix}$$

Exchange the first two rows; Multiply the first row by -1; Add the first row to the third row; Divide the second row by -2; Subtract the second row from the third; Multiplying the third row by $-2/5$; Subtract half the bottom row from the middle row; Subtract the middle row from the top row; Add the bottom row to the top row.

Row reduction example

Consider the following sequence of reductions:

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

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

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

Add row 1 to rows 2 and 4, and subtract it from rows 3 and 5; Subtract row

2 from row 4; Exchange rows 3 and 5; Add 3 times row 3 to row 4, and subtract 3 times row 3 from row 5; Divide row 4 by 7, then add 5 times row 4 to row 5; Subtract 2 times row 2 from row 1; Add 3 times row 3 to row 1; Subtract 3 times row 4 from row 2, and subtract row 4 from row 3.

Deleting columns

We previously saw the following row-reduction:

$$\begin{bmatrix} 0 & 0 & -2 & -1 & -13 \\ -1 & -2 & -1 & 1 & -2 \\ -1 & -2 & 0 & -1 & -8 \end{bmatrix} \to \begin{bmatrix} -1 & -2 & -1 & 1 & -2 \\ 0 & 0 & -2 & -1 & -13 \\ -1 & -2 & 0 & -1 & -8 \end{bmatrix} \to \cdots \to \begin{bmatrix} 1 & 2 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix}$$

We can delete the middle column and it still works the same way:

$$\begin{bmatrix} 0 & 0 & & -1 & -13 \\ -1 & -2 & & 1 & -2 \\ -1 & -2 & & -1 & -8 \end{bmatrix} \to \begin{bmatrix} -1 & -2 & & 1 & -2 \\ 0 & 0 & & -1 & -13 \\ -1 & -2 & & -1 & -8 \end{bmatrix} \to \cdots \to \begin{bmatrix} 1 & 2 & & 0 & 3 \\ 0 & 0 & & 0 & 4 \\ 0 & 0 & & 1 & 5 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 & -1 & -13 \\ -1 & -2 & 1 & -2 \\ -1 & -2 & -1 & -8 \end{bmatrix} \to \begin{bmatrix} -1 & -2 & 1 & -2 \\ 0 & 0 & -1 & -13 \\ -1 & -2 & -1 & -8 \end{bmatrix} \to \cdots \to \begin{bmatrix} 1 & 2 & 0 & 3 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 1 & 5 \end{bmatrix}$$

(However, the final result is no longer in RREF; we need further row operations to fix that.)

In general: suppose that $A \to A'$, and that B is obtained by deleting some columns from A, and that B' is obtained by deleting the corresponding columns from A'. Then $B \to B'$.

Solution by row-reduction

Theorem 6.8: Let A be an augmented matrix, and let A' be obtained from A by a sequence of row operations. Then the system of equations corresponding to A has the same solutions (if any) as the system of equations corresponding to A'.

This should be fairly clear. The three types of elementary row operations correspond to reordering our system of equations, multiplying both sides of one equation by a nonzero constant, and adding one equation to another one. None of these operations changes the solution set. We thus have the following method:

Method 6.9: To solve a system of linear equations:

(a) Write down the corresponding augmented matrix.

(b) Row-reduce it to RREF

(c) Convert it back to a new system of equations, which will have exactly the same solutions as the old ones.

(d) Read off the solutions (which is easy for a system in RREF).

Example solution by row-reduction

We will try to solve the following system:

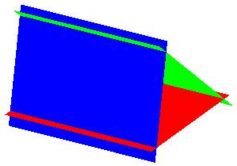
$$\begin{array}{rcrcrcr} 2x & + & y & + & z & = & 1 \\ 4x & + & 2y & + & 3z & = & -1 \\ 6x & + & 3y & - & z & = & 11 \end{array}$$

We construct and then row-reduce the augmented matrix:

$$\left[\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 4 & 2 & 3 & -1 \\ 6 & 3 & -1 & 11 \end{array}\right] \xrightarrow{1} \left[\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & -4 & 8 \end{array}\right] \xrightarrow{2} \left[\begin{array}{ccc|c} 2 & 1 & 1 & 1 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & -4 \end{array}\right] \xrightarrow{3} \left[\begin{array}{ccc|c} 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

There is a pivot in the rightmost column, which means that there are no solutions for the original system.

Each of the equations defines a plane. These are arranged like the three faces of a Toblerone packet, so there is no point where they all meet.



Example solution by row-reduction

We will solve the equations

$$\begin{aligned} a+b+c+d &= 4 \\ a+b-c-d &= 0 \\ a-b+c-d &= 0 \\ a-b-c+d &= 0. \end{aligned}$$

The corresponding augmented matrix can be row-reduced as follows:

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4 \\ 1 & 1 & -1 & -1 & 0 \\ 1 & -1 & 1 & -1 & 0 \\ 1 & -1 & -1 & 1 & 0 \end{array}\right] \xrightarrow{1} \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4 \\ 0 & 0 & -2 & -2 & -4 \\ 1 & -1 & 1 & -1 & 0 \\ 0 & 0 & -2 & 2 & 0 \end{array}\right] \xrightarrow{2} \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4 \\ 0 & 0 & 1 & 1 & 2 \\ 1 & -1 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{array}\right] \xrightarrow{3}$$

$$\left[\begin{array}{cccc|c} 1 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 1 & 2 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{array}\right] \xrightarrow{4} \left[\begin{array}{cccc|c} 1 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & -2 & 0 & 0 & -2 \\ 0 & 0 & 0 & -2 & -2 \end{array}\right] \xrightarrow{5} \left[\begin{array}{cccc|c} 1 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right] \xrightarrow{6}$$

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right] \xrightarrow{7} \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right]$$

Subtract row 1 from row 2, and row 3 from row 4; Multiply rows 2 and 4 by $-1/2$; Subtract row 2 from row 1, and row 4 from row 3; Subtract row 1 from row 3, and row 2 from row 4; Multiply rows 3 and 4 by $-1/2$; Subtract row 3 from row 1, and row 4 from row 2; Exchange rows 2 and 3.
The final matrix corresponds to the equations $a = 1$, $b = 1$, $c = 1$ and $d = 1$, which give the unique solution to the original system of equations.

Homogeneous systems

A system of linear equations is *homogeneous* if the values on the right hand side are all zero. Example:

$$\begin{aligned} a+b+c+d+e+f &= 0 \\ 2a+2b+2c+2d-e-f &= 0 \\ 3a+3b-c-d-e-f &= 0 \end{aligned}$$

The last column of the augmented matrix is zero all through the row reduction, so we need not write it in; we can work with the unaugmented matrix.

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & -1 & -1 \\ 3 & 3 & -1 & -1 & -1 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$a + b = 0 \qquad c + d = 0 \qquad e + f = 0.$$

Move the independent variables (from non-pivot columns) to the RHS:

$$a = -b \qquad c = -d \qquad e = -f.$$

If we prefer we can introduce new variables λ, μ and ν, and say that the general solution is

$$\begin{array}{lll} a = -\lambda & c = -\mu & e = -\nu \\ b = \lambda & d = \mu & f = \nu \end{array}$$

for arbitrary values of λ, μ and ν.

Lecture 3

Linear combinations

Definition 7.1: Let $v_1, \ldots, v_k$ and w be vectors in $\mathbb{R}^n$. We say that w is a *linear combination* of $v_1, \ldots, v_k$ if there exist scalars $\lambda_1, \ldots, \lambda_k$ such that

$$w = \lambda_1 v_1 + \cdots + \lambda_k v_k.$$

Example 7.2: Consider the following vectors in $\mathbb{R}^4$:

$$v_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} \qquad v_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} \qquad v_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \qquad w = \begin{bmatrix} 1 \\ 10 \\ 100 \\ -111 \end{bmatrix}$$

If we take $\lambda_1 = 1$ and $\lambda_2 = 11$ and $\lambda_3 = 111$ we get

$$\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 11 \\ -11 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 111 \\ -111 \end{bmatrix} = \begin{bmatrix} 1 \\ 10 \\ 100 \\ -111 \end{bmatrix} = w,$$

which shows that w is a linear combination of v_1, v_2 and v_3.

Linear combinations example

w is a *linear combination* of $v_1, \ldots, v_k$ if there exist scalars $\lambda_1, \ldots, \lambda_k$ such that

$$w = \lambda_1 v_1 + \cdots + \lambda_k v_k.$$

Consider the following vectors in $\mathbb{R}^4$:

$$v_1 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix} \qquad v_2 = \begin{bmatrix} 0 \\ 1 \\ 4 \\ 9 \end{bmatrix} \qquad v_3 = \begin{bmatrix} 0 \\ 1 \\ 8 \\ 27 \end{bmatrix} \qquad v_4 = \begin{bmatrix} 0 \\ 1 \\ 16 \\ 81 \end{bmatrix} \qquad w = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Any linear combination of $v_1, \ldots, v_4$ has the form

$$\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 + \lambda_4 v_4 = \begin{bmatrix} 0 \\ \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 \\ 2\lambda_1 + 4\lambda_2 + 8\lambda_3 + 16\lambda_4 \\ 3\lambda_1 + 9\lambda_2 + 27\lambda_3 + 81\lambda_4 \end{bmatrix}.$$

Thus, the first component of any such linear combination is zero. (You should be able to see this without writing out the whole formula.) As the first component of w is not zero, we see that w is *not* a linear combination of $v_1, \ldots, v_4$.

Linear combinations example

w is a *linear combination* of $v_1, \ldots, v_k$ if there exist scalars $\lambda_1, \ldots, \lambda_k$ such that

$$w = \lambda_1 v_1 + \cdots + \lambda_k v_k.$$

Consider the following vectors in $\mathbb{R}^3$:

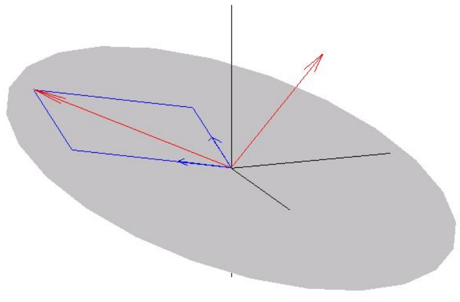
$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \qquad v_2 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \qquad v_3 = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \qquad v_4 = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} \qquad v_5 = \begin{bmatrix} 1 \\ 5 \\ 1 \end{bmatrix} \qquad w = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

Any linear combination of $v_1, \ldots, v_5$ has the form

$$\lambda_1 v_1 + \cdots + \lambda_5 v_5 = \begin{bmatrix} \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 \\ \lambda_1 + 2\lambda_2 + 3\lambda_3 + 4\lambda_4 + 5\lambda_5 \\ \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 + \lambda_5 \end{bmatrix}.$$

The first and last components of any such linear combination are the same. Again, you should be able to see this without writing the full formula. As the first and last components of w are different, we see that w is not a linear combination of $v_1, \ldots, v_5$.

Two vectors in $\mathbb{R}^3$ span a plane



Any vector that lies in the grey plane can be expressed as a linear combination of the two blue vectors.
Any vector that does not lie in the grey plane cannot be expressed as a linear combination of the two blue vectors.

Method for finding linear combinations

Suppose we have vectors $v_1, \dots, v_k \in \mathbb{R}^n$ and another vector $w \in \mathbb{R}^n$, and we want to express w as a linear combination of the v_i (or show that this is not possible).
Let A be the matrix whose columns are the vectors v_i:

$$A = \left[\begin{array}{c|c|c} v_1 & \cdots & v_k \end{array}\right] \in M_{n\times k}(\mathbb{R}).$$

For any k-vector $\lambda = \begin{bmatrix}\lambda_1 & \cdots & \lambda_k\end{bmatrix}^T$ we have

$$A\lambda = \left[\begin{array}{c|c|c} v_1 & \cdots & v_k \end{array}\right] \begin{bmatrix}\lambda_1 \\ \vdots \\ \lambda_k\end{bmatrix} = \lambda_1 v_1 + \cdots + \lambda_k v_k$$

Thus, to express w as a linear combination of the v_i is the same as to solve the vector equation $A\lambda = w$, which we can do by row-reducing the augmented matrix

$$B = \left[\begin{array}{c|c} A & w \end{array}\right] = \left[\begin{array}{c|c|c|c} v_1 & \cdots & v_k & w \end{array}\right]$$

Example of finding a linear combination

Is w a linear combination of v_1, v_2 and v_3?

$$v_1 = \begin{bmatrix}11\\11\\1\\1\end{bmatrix} \qquad v_2 = \begin{bmatrix}1\\11\\11\\1\end{bmatrix} \qquad v_3 = \begin{bmatrix}1\\1\\11\\11\end{bmatrix} \qquad w = \begin{bmatrix}121\\221\\1211\\1111\end{bmatrix}.$$

We write down the relevant augmented matrix and row-reduce it:

$$\left[\begin{array}{ccc|c} 11 & 1 & 1 & 121 \\ 11 & 11 & 1 & 221 \\ 1 & 11 & 11 & 1211 \\ 1 & 1 & 11 & 1111 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 11 & 1111 \\ 11 & 1 & 1 & 121 \\ 11 & 11 & 1 & 221 \\ 1 & 11 & 11 & 1211 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 11 & 1111 \\ 0 & -10 & -120 & -12100 \\ 0 & 0 & -120 & -12000 \\ 0 & 10 & 0 & 100 \end{array}\right]$$

$$\to \left[\begin{array}{ccc|c} 1 & 1 & 11 & 1111 \\ 0 & 1 & 12 & 1210 \\ 0 & 0 & 1 & 100 \\ 0 & 1 & 0 & 10 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 0 & 11 \\ 0 & 1 & 0 & 10 \\ 0 & 0 & 1 & 100 \\ 0 & 1 & 0 & 10 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 10 \\ 0 & 0 & 1 & 100 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

Move the bottom row to the top; Subtract multiples of row 1 from the other rows; Divide rows 2,3 and 4 by -10, -120 and 10; Subtract multiples of row 3 from the other rows; Subtract multiples of row 2 from the other rows.

Example of finding a linear combination

$$v_1 = \begin{bmatrix}11\\11\\1\\1\end{bmatrix} \qquad v_2 = \begin{bmatrix}1\\11\\11\\1\end{bmatrix} \qquad v_3 = \begin{bmatrix}1\\1\\11\\11\end{bmatrix} \qquad w = \begin{bmatrix}121\\221\\1211\\1111\end{bmatrix}.$$

$$\left[\begin{array}{ccc|c} 11 & 1 & 1 & 121 \\ 11 & 11 & 1 & 221 \\ 1 & 11 & 11 & 1211 \\ 1 & 1 & 11 & 1111 \end{array}\right] \to \cdots \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 10 \\ 0 & 0 & 1 & 100 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

The final matrix corresponds to the system of equations

$$\lambda_1 = 1 \qquad \lambda_2 = 10 \qquad \lambda_3 = 100 \qquad 0 = 0$$

so we conclude that $w = v_1 + 10v_2 + 100v_3$.
In particular, w can be expressed as a linear combination of v_1, v_2 and v_3.
We can check the above equation directly:

$$v_1 + 10v_2 + 100v_3 = \begin{bmatrix}11\\11\\1\\1\end{bmatrix} + \begin{bmatrix}10\\110\\110\\10\end{bmatrix} + \begin{bmatrix}100\\100\\1100\\1100\end{bmatrix} = \begin{bmatrix}121\\221\\1211\\1111\end{bmatrix} = w.$$

Example of not finding a linear combination

Is b a linear combination of a_1, a_2 and a_3?

$$a_1 = \begin{bmatrix}2\\-1\\0\end{bmatrix} \qquad a_2 = \begin{bmatrix}3\\0\\-1\end{bmatrix} \qquad a_3 = \begin{bmatrix}0\\3\\-2\end{bmatrix} \qquad b = \begin{bmatrix}1\\2\\3\end{bmatrix}$$

We write down the relevant augmented matrix and row-reduce it:

$$\left[\begin{array}{ccc|c} 2 & 3 & 0 & 1 \\ -1 & 0 & 3 & 2 \\ 0 & -1 & -2 & 3 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & -3 & -2 \\ 0 & 1 & 2 & -3 \\ 2 & 3 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & -3 & -2 \\ 0 & 1 & 2 & -3 \\ 0 & 3 & 6 & 5 \end{array}\right] \to$$

$$\left[\begin{array}{ccc|c} 1 & 0 & -3 & -2 \\ 0 & 1 & 2 & -3 \\ 0 & 0 & 0 & 14 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & -3 & -2 \\ 0 & 1 & 2 & -3 \\ 0 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & -3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

Move the top row to the bottom, and multiply the other two rows by -1; Subtract 2 times row 1 from row 3; Subtract 3 times row 2 from row 3; Divide row 3 by 14; Subtract multiples of row 3 from rows 1 and 2.

Example of not finding a linear combination

$$a_1 = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix} \qquad a_2 = \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix} \qquad a_3 = \begin{bmatrix} 0 \\ 3 \\ -2 \end{bmatrix} \qquad b = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

$$\left[\begin{array}{ccc|c} 2 & 3 & 0 & 1 \\ -1 & 0 & 3 & 2 \\ 0 & -1 & -2 & 3 \end{array}\right] \to \dots \to \left[\begin{array}{ccc|c} 1 & 0 & -3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

The final matrix has a pivot in the rightmost column, corresponding to the equation $0 = 1$. This means that the equation $\lambda_1 a_1 + \lambda_2 a_2 + \lambda_3 a_3 = b$ cannot be solved for λ_1, λ_2 and λ_3, or in other words that b is not a linear combination of a_1, a_2 and a_3.
We can also see this in a more direct but less systematic way, as follows. It is easy to check that $b.a_1 = b.a_2 = b.a_3 = 0$, which means that $b.(\lambda_1 a_1 + \lambda_2 a_2 + \lambda_3 a_3) = 0$ for all possible choices of λ_1, λ_2 and λ_3. However, $b.b = 14 > 0$, so b cannot be equal to $\lambda_1 a_1 + \lambda_2 a_2 + \lambda_3 a_3$.

Linear independence

Definition 8.1: Let $\mathcal{V} = v_1, \dots, v_k$ be a list of vectors in $\mathbb{R}^n$.
A *linear relation* between the v_i is a relation of the form $\lambda_1 v_1 + \cdots + \lambda_k v_k = 0$, where $\lambda_1, \dots, \lambda_k$ are scalars.

For any list we have the trivial linear relation $0v_1 + 0v_2 + \cdots + 0v_k = 0$.
There may or may not be any nontrivial linear relations.

If $\mathcal{V}$ has a nontrivial linear relation, we say that it is *(linearly) dependent*.
If the only linear relation on $\mathcal{V}$ is the trivial one, we instead say that $\mathcal{V}$ is *(linearly) independent*.

Example 8.2: Consider the list $\mathcal{V}$ given by

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad v_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \qquad v_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad v_4 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}.$$

There is a nontrivial linear relation $v_1 + v_2 - v_3 - v_4 = 0$, so the list $\mathcal{V}$ is dependent.

Linear dependence example

The list $v_1, \dots, v_k$ is *dependent* if there is a relation $\lambda_1 v_1 + \cdots + \lambda_k v_k = 0$ with not all λ_i being zero. Otherwise, it is *independent*.

Example 8.3: Consider the list $\mathcal{A}$ given by

$$a_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad a_2 = \begin{bmatrix} 12 \\ 1 \end{bmatrix} \qquad a_3 = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \qquad a_4 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

Just by writing it out, you can check that $3a_1 + a_2 + 3a_3 - 4a_4 = 0$.
This is a nontrivial linear relation on the list $\mathcal{A}$, so $\mathcal{A}$ is dependent.

Example 8.4: Claim: the following list $\mathcal{U}$ is independent.

$$u_1 = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}^T \qquad u_2 = \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix}^T \qquad u_3 = \begin{bmatrix} 0 & 0 & 1 & 1 \end{bmatrix}^T$$

Indeed, consider a linear relation $\lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 = 0$. This gives

$$\begin{bmatrix} \lambda_1 \\ \lambda_1 + \lambda_2 \\ \lambda_2 + \lambda_3 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}; \qquad \lambda_1 = 0; \qquad \lambda_3 = 0; \qquad \lambda_1 + \lambda_2 = 0; \qquad \lambda_2 = 0.$$

As the only linear relation is the trivial one, we see that $\mathcal{U}$ is independent.

Pivots in every column

Definition 8.6: Let B be a $p \times q$ matrix.
We say that B is *wide* if $p < q$, or *square* if $p = q$ or *tall* if $p > q$.

$$\underset{\text{wide}}{\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}} \qquad \underset{\text{square}}{\begin{bmatrix} 1 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 1 \end{bmatrix}} \qquad \underset{\text{tall}}{\begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix}}$$

Lemma 8.7: Let B be a $p \times q$ matrix in RREF.

(a) If B is wide then it is impossible for every column to contain a pivot.
(b) If B is square and every column contains a pivot then $B = I_q$.
(c) If B is tall then the only way for every column to contain a pivot is if B consists of I_q with $(p - q)$ rows of zeros added at the bottom.

$$B = \left[\begin{array}{c} I_q \\ \hline 0_{(p-q)\times q} \end{array}\right]$$

Checking dependence by row-reduction

Method 8.8: Let $\mathcal{V} = v_1, \dots, v_m$ be a list of vectors in $\mathbb{R}^n$.
We can check whether this list is dependent as follows.

(a) Form the $n \times m$ matrix $A = \left[\begin{array}{c|c|c} v_1 & \dots & v_m \end{array}\right]$ whose columns are the vectors v_i.

(b) Row reduce A to get another $n \times m$ matrix B in RREF.

(c) If every column of B contains a pivot (as on the previous slide) then $\mathcal{V}$ is independent.

(d) If some column of B has no pivot, then the list $\mathcal{V}$ is dependent. Moreover, we can find the coefficients λ_i in a nontrivial linear relation by solving the vector equation $B\lambda = 0$ (which is easy because B is in RREF).

Remark 8.9: If $m > n$ then $\mathcal{V}$ is automatically dependent and need not do any more.
(Any list of 5 vectors in $\mathbb{R}^3$ is dependent, any list of 10 in $\mathbb{R}^9$ is dependent, $\dots$.)
Indeed, in this case B is wide, so it cannot have a pivot in every column. This only tells us that there **exists** a nontrivial relation $\lambda_1 v_1 + \dots + \lambda_m v_m = 0$, it does not tell us the coefficients λ_i. To find them we do need to go through the whole method as explained above.

Example of checking for (in)dependence

We previously considered the list

$$v_1 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix} \qquad v_2 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix} \qquad v_3 = \begin{bmatrix}1\\0\\0\\1\end{bmatrix} \qquad v_4 = \begin{bmatrix}0\\1\\1\\0\end{bmatrix}.$$

We can write down the corresponding matrix and row-reduce it as follows:

$$\begin{bmatrix}1&0&1&0\\1&0&0&1\\0&1&0&1\\0&1&1&0\end{bmatrix} \xrightarrow{1} \begin{bmatrix}1&0&1&0\\0&0&-1&1\\0&1&0&1\\0&0&1&-1\end{bmatrix} \xrightarrow{2} \begin{bmatrix}1&0&1&0\\0&1&0&1\\0&0&1&-1\\0&0&-1&1\end{bmatrix} \xrightarrow{3} \begin{bmatrix}1&0&0&1\\0&1&0&1\\0&0&1&-1\\0&0&0&0\end{bmatrix}$$

The end result has no pivot in the last column, so the original list is dependent. To find a specific linear relation, we solve the equation

$$\begin{bmatrix}1&0&0&1\\0&1&0&1\\0&0&1&-1\\0&0&0&0\end{bmatrix}\begin{bmatrix}\lambda_1\\\lambda_2\\\lambda_3\\\lambda_4\end{bmatrix} = \begin{bmatrix}0\\0\\0\\0\end{bmatrix} \qquad \begin{aligned}\lambda_1 &= -\lambda_4\\ \lambda_2 &= -\lambda_4\\ \lambda_3 &= \lambda_4\\ \lambda_4 &\text{ arbitrary}\end{aligned}$$

Taking $\lambda_4 = 1$ gives $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (-1, -1, 1, 1)$, corresponding to the relation $-v_1 - v_2 + v_3 + v_4 = 0$.

Example of checking for (in)dependence

We previously considered the list

$$a_1 = \begin{bmatrix}1\\2\end{bmatrix} \qquad a_2 = \begin{bmatrix}12\\1\end{bmatrix} \qquad a_3 = \begin{bmatrix}-1\\-1\end{bmatrix} \qquad a_4 = \begin{bmatrix}3\\1\end{bmatrix}.$$

Here we have 4 vectors in $\mathbb{R}^2$, so they must be dependent. Thus, there exist nontrivial linear relations $\lambda_1 a_1 + \lambda_2 a_2 + \lambda_3 a_3 + \lambda_4 a_4 = 0$.
To actually find such a relation, we write down the corresponding matrix and row-reduce it as follows:

$$\begin{bmatrix}1&12&-1&3\\2&1&-1&1\end{bmatrix} \to \begin{bmatrix}1&12&-1&3\\0&1&-1/23&5/23\end{bmatrix} \to \begin{bmatrix}1&0&-11/23&9/23\\0&1&-1/23&5/23\end{bmatrix}$$

We now need to solve the matrix equation

$$\begin{bmatrix}1&0&-11/23&9/23\\0&1&-1/23&5/23\end{bmatrix}\begin{bmatrix}\lambda_1\\\lambda_2\\\lambda_3\\\lambda_4\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix}$$

This gives $\lambda_1 = \frac{11}{23}\lambda_3 - \frac{9}{23}\lambda_4$ and $\lambda_2 = \frac{1}{23}\lambda_3 - \frac{5}{23}\lambda_4$ with λ_3 and λ_4 arbitrary.
If we choose $\lambda_3 = 23$ and $\lambda_4 = 0$ we get $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (11, 1, 23, 0)$ so we have a relation $11a_1 + a_2 + 23a_3 + 0a_4 = 0$.

Example of checking for (in)dependence

We previously considered the list $\mathcal{U}$ given by

$$u_1 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix} \qquad u_2 = \begin{bmatrix}0\\1\\1\\0\end{bmatrix} \qquad u_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}.$$

We can write down the corresponding matrix and row-reduce it as follows:

$$\begin{bmatrix}1&0&0\\1&1&0\\0&1&1\\0&0&1\end{bmatrix} \to \begin{bmatrix}1&0&0\\0&1&0\\0&1&1\\0&0&1\end{bmatrix} \to \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&1\end{bmatrix} \to \left[\begin{array}{ccc}1&0&0\\0&1&0\\0&0&1\\\hline 0&0&0\end{array}\right]$$

The final matrix has a pivot in every column. It follows that the list $\mathcal{U}$ is independent.

Proof of correctness of the method

Put $A = \left[\begin{array}{c|c|c} v_1 & \cdots & v_m \end{array}\right]$ as in step (a), and let B be the RREF form of A.
Note that for any vector $\lambda = \begin{bmatrix}\lambda_1 & \dots & \lambda_m\end{bmatrix}^T \in \mathbb{R}^m$, we have

$$A\lambda = \left[\begin{array}{c|c|c} v_1 & \cdots & v_m \end{array}\right] \begin{bmatrix}\lambda_1 \\ \vdots \\ \lambda_m\end{bmatrix} = \lambda_1 v_1 + \cdots + \lambda_m v_m.$$

Thus, linear relations on our list are just the same as solutions to the homogeneous equation $A\lambda = 0$. We saw earlier that these are the same as solutions to the equation $B\lambda = 0$, which can be found by the standard method for RREF equations. If there is a pivot in every column then none of the variables λ_i is independent, so the only solution is $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 0$. Thus, the only linear relation on $\mathcal{V}$ is the trivial one, which means that the list $\mathcal{V}$ is linearly independent.
Suppose instead that some column (the k'th one, say) does not contain a pivot. Then the variable λ_k will be independent, so we can choose $\lambda_k = 1$. This will give us a nonzero to solution to $B\lambda = 0$, or equivalently $A\lambda = 0$, corresponding to a nontrivial linear relation on $\mathcal{V}$. This shows that $\mathcal{V}$ is linearly dependent.

Lecture 4

Spanning

Definition 9.1: Suppose we have a list $\mathcal{V} = v_1, \dots, v_m$ of vectors in $\mathbb{R}^n$. Then $\mathcal{V}$ *spans* $\mathbb{R}^n$ if **every** vector in $\mathbb{R}^n$ can be expressed as a linear combination of $v_1, \dots, v_m$.

Example 9.2: Consider the list $\mathcal{V} = v_1, v_2, v_3, v_4$, where

$$v_1 = \begin{bmatrix}0\\1\\2\\3\end{bmatrix} \qquad v_2 = \begin{bmatrix}0\\1\\4\\9\end{bmatrix} \qquad v_3 = \begin{bmatrix}0\\1\\8\\27\end{bmatrix} \qquad v_4 = \begin{bmatrix}0\\1\\16\\81\end{bmatrix}$$

Previously we saw that the vector $w = \begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}^T$ is not a linear combination of this list, so the list $\mathcal{V}$ does not span $\mathbb{R}^4$.

Example 9.3: Consider the list $\mathcal{V} = v_1, v_2, v_3, v_4, v_5$, where

$$v_1 = \begin{bmatrix}1\\1\\1\end{bmatrix} \qquad v_2 = \begin{bmatrix}1\\2\\1\end{bmatrix} \qquad v_3 = \begin{bmatrix}1\\3\\1\end{bmatrix} \qquad v_4 = \begin{bmatrix}1\\4\\1\end{bmatrix} \qquad v_5 = \begin{bmatrix}1\\5\\1\end{bmatrix}.$$

Previously we saw that the vector $w = \begin{bmatrix}-1 & 0 & 1\end{bmatrix}^T$ is not a linear combination of this list, so the list $\mathcal{V}$ does not span $\mathbb{R}^3$.

Spanning example

Consider the list $\mathcal{U} = u_1, u_2, u_3$, where

$$u_1 = \begin{bmatrix}1\\1\\0\end{bmatrix} \qquad u_2 = \begin{bmatrix}1\\0\\1\end{bmatrix} \qquad u_3 = \begin{bmatrix}0\\1\\1\end{bmatrix}.$$

We will show that these span $\mathbb{R}^3$. Indeed, for any vector $v = \begin{bmatrix}x & y & z\end{bmatrix}^T \in \mathbb{R}^3$ we can put

$$\lambda_1 = \frac{x+y-z}{2} \qquad \lambda_2 = \frac{x-y+z}{2} \qquad \lambda_3 = \frac{-x+y+z}{2}$$

and we find that

$$\lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 = \begin{bmatrix}(x+y-z)/2\\(x+y-z)/2\\0\end{bmatrix} + \begin{bmatrix}(x-y+z)/2\\0\\(x-y+z)/2\end{bmatrix} + \begin{bmatrix}0\\(-x+y+z)/2\\(-x+y+z)/2\end{bmatrix}$$

$$= \begin{bmatrix}(x+y-z+x-y+z)/2\\(x+y-z-x+y+z)/2\\(x-y+z-x+y+z)/2\end{bmatrix} = \begin{bmatrix}x\\y\\z\end{bmatrix} = v.$$

This expresses v as a linear combination of the list $\mathcal{U}$, as required.

Spanning example

Consider the list $\mathcal{A} = a_1, a_2, a_3$ where

$$a_1 = \begin{bmatrix}1\\2\end{bmatrix} \qquad a_2 = \begin{bmatrix}2\\3\end{bmatrix} \qquad a_3 = \begin{bmatrix}3\\5\end{bmatrix}.$$

Let $v = \begin{bmatrix}x\\y\end{bmatrix}$ be an arbitrary vector in $\mathbb{R}^2$. Note that

$$(2y-4x)\begin{bmatrix}1\\2\end{bmatrix} + (x-y)\begin{bmatrix}2\\3\end{bmatrix} + x\begin{bmatrix}3\\5\end{bmatrix} = \begin{bmatrix}2y-4x\\4y-8x\end{bmatrix} + \begin{bmatrix}2x-2y\\3x-3y\end{bmatrix} + \begin{bmatrix}3x\\5x\end{bmatrix} = \begin{bmatrix}x\\y\end{bmatrix}$$

or in other words

$$v = (2y-4x)a_1 + (x-y)a_2 + xa_3.$$

This expresses an arbitrary $v \in \mathbb{R}^2$ as a linear combination of a_1, a_2 and a_3, proving that the list $\mathcal{A}$ spans $\mathbb{R}^2$.

In this case there are actually many different ways in which we can express v as a linear combination of a_1, a_2 and a_3. Another one is

$$v = (y-3x)a_1 + (2x-2y)a_2 + ya_3.$$

Checking spanning by row-reduction

Method 9.7: Let $\mathcal{V} = v_1, \dots, v_m$ be a list of vectors in $\mathbb{R}^n$. We can check whether this list spans $\mathbb{R}^n$ as follows.

(a) Form the $m \times n$ matrix $C = \left[\begin{array}{c} v_1^T \\ \hline \vdots \\ \hline v_m^T \end{array}\right]$ whose rows are the v_i^T.

(b) Row reduce C to get another $m \times n$ matrix D in RREF.

(c) If every column of D contains a pivot (so $D = \left[\begin{array}{c} I_n \\ \hline 0_{(m-n)\times n} \end{array}\right]$) then $\mathcal{V}$ spans $\mathbb{R}^n$.

(d) If some column of D has no pivot, then the list $\mathcal{V}$ does not span $\mathbb{R}^n$.

Remark 9.8: This is almost exactly the same as the method for checking independence, except that here we start by building a matrix C whose rows are v_i^T, instead of building a matrix A whose columns are v_i. These are transposes of each other: $A = C^T$ and $C = A^T$.

Warning: transposing does not interact well with row-reduction, so the matrix D is **not** the transpose of B.

Example of spanning check

Consider the list

$$v_1 = \begin{bmatrix}0\\1\\2\\3\end{bmatrix} \qquad v_2 = \begin{bmatrix}0\\1\\4\\9\end{bmatrix} \qquad v_3 = \begin{bmatrix}0\\1\\8\\27\end{bmatrix} \qquad v_4 = \begin{bmatrix}0\\1\\16\\81\end{bmatrix}$$

The relevant matrix is $C = \begin{bmatrix}0&1&2&3\\0&1&4&9\\0&1&8&27\\0&1&16&81\end{bmatrix}$

The first column is zero, and will remain zero no matter what row operations we perform. Thus C cannot reduce to the identity matrix, so $\mathcal{V}$ does not span (as we already saw by a different method). In fact the row-reduction is

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

but it is not really necessary to go through the whole calculation.

Example of spanning check

Consider the list $\mathcal{V}$ given by

$$v_1 = \begin{bmatrix}1\\1\\1\end{bmatrix} \qquad v_2 = \begin{bmatrix}1\\2\\1\end{bmatrix} \qquad v_3 = \begin{bmatrix}1\\3\\1\end{bmatrix} \qquad v_4 = \begin{bmatrix}1\\4\\1\end{bmatrix} \qquad v_5 = \begin{bmatrix}1\\5\\1\end{bmatrix}.$$

The relevant row-reduction is

$$\begin{bmatrix}1&1&1\\1&2&1\\1&3&1\\1&4&1\\1&5&1\end{bmatrix} \to \begin{bmatrix}1&1&1\\0&1&0\\0&2&0\\0&3&0\\0&4&0\end{bmatrix} \to \left[\begin{array}{ccc}1&0&1\\0&1&0\\0&0&0\\ \hline 0&0&0\\0&0&0\end{array}\right]$$

At the end of the process the last column does not contain a pivot (so the top 3×3 block is not the identity), so $\mathcal{V}$ does not span $\mathbb{R}^3$. Again, we saw this earlier by a different method.

Example of spanning check

For the list

$$\mathcal{A} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ 0 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 3 \\ -2 \end{bmatrix}$$

the relevant row-reduction is

$$\begin{bmatrix} 2 & -1 & 0 \\ 3 & 0 & -1 \\ 0 & 3 & -2 \end{bmatrix} \to \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 3 & 0 & -1 \\ 0 & 3 & -2 \end{bmatrix} \to \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & \frac{3}{2} & -1 \\ 0 & 3 & -2 \end{bmatrix} \to$$

$$\begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -\frac{1}{3} \\ 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 0 \end{bmatrix}.$$

In the last matrix the third column has no pivot, so the list does not span.

Example of spanning check

Consider the list $\mathcal{U} = u_1, u_2, u_3$

$$u_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \qquad u_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad u_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

The relevant row-reduction is

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

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

The end result is the identity matrix, so the list $\mathcal{U}$ spans $\mathbb{R}^3$.

Example of spanning check

Consider the list $\mathcal{A} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \end{bmatrix}$. The relevant row-reduction is

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ 0 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & 2 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \to \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ \hline 0 & 0 \end{array}\right]$$

In the last matrix, the top 2×2 block is the identity. This means that the list $\mathcal{A}$ spans $\mathbb{R}^2$.

Proof of correctness of the method

Lemma 9.15: Let C be an $m \times n$ matrix, and let C' be obtained from C by a single elementary row operation. Let s be a row vector of length n. Then s can be expressed as a linear combination of the rows of C if and only if it can be expressed as a linear combination of the rows of C'.

Proof: Let the rows of C be $r_1, \dots, r_m$. Suppose that s is a linear combination of these rows, say

$$s = \lambda_1 r_1 + \lambda_2 r_2 + \lambda_3 r_3 + \cdots + \lambda_m r_m.$$

(a) Suppose that C' is obtained from C by swapping the first two rows, so the rows of C' are $r_2, r_1, r_3, \dots, r_m$. The sequence of numbers $\lambda_2, \lambda_1, \lambda_3, \dots, \lambda_m$ satisfies

$$s = \lambda_2 r_2 + \lambda_1 r_1 + \lambda_3 r_3 + \cdots + \lambda_m r_m,$$

which expresses s as a linear combination of the rows of C'. The argument is essentially the same if we exchange any other pair of rows.

Proof of correctness of the method

$C \in M_{m\times n}(\mathbb{R})$; C' obtained from C by a single row operation; s a row vector of length n. Claim: s is a linear combination of rows of C iff it is a linear combination of rows of C'.

(b) Suppose instead that C' is obtained from C by multiplying the first row by a nonzero scalar u, so the rows of C' are $ur_1, r_2, \dots, r_m$. The sequence of numbers $u^{-1}\lambda_1, \lambda_2, \dots, \lambda_m$ then satisfies

$$s = (u^{-1}\lambda_1)(ur_1) + \lambda_2 u_2 + \dots + \lambda_m r_m,$$

which expresses s as a linear combination of the rows of C'.The argument is essentially the same if we multiply any other row by a constant.

(c) Suppose instead that C' is obtained from C by adding u times the second row to the first row, so the rows of C' are $r_1 + ur_2, r_2, r_3, \dots, r_m$. The sequence of numbers $\lambda_1, \lambda_2 - u\lambda_1, \lambda_3, \dots, \lambda_n$ then satisfies

$$\lambda_1(r_1+ur_2)+(\lambda_2-u\lambda_1)r_2+\lambda_3 r_3+\dots+\lambda_m r_m = \lambda_1 r_1+\lambda_2 r_2+\dots+\lambda_m r_m = s,$$

which expresses s as a linear combination of the rows of C'. The argument is essentially the same if add a multiple of any row to any other row.

Proof of correctness of the method

$C \in M_{m\times n}(\mathbb{R})$; C' obtained from C by a single row operation; s a row vector of length n. Claim: s is a linear combination of rows of C iff it is a linear combination of rows of C'.

We have now proved half of the lemma: if s is a linear combination of the rows of C, then it is also a linear combination of the rows of C'. We also need to prove the converse: if s is a linear combination of the rows of C', then it is also a linear combination of the rows of C. We will only treat case (c), and leave the other two cases as an exercise. The rows of C' are then $r_1 + ur_2, r_2, r_3, \dots, r_m$. As s is a linear combination of these rows, we have $s = \mu_1(r_1 + ur_2) + \mu_2 r_2 + \dots + \mu_m r_m$ for some numbers $\mu_1, \dots, \mu_m$. Now the sequence of numbers $\mu_1, (\mu_2 + u\mu_1), \mu_3, \dots, \mu_m$ satisfies

$$s = \mu_1 r_1 + (\mu_2 + u\mu_1) r_2 + \mu_3 r_3 + \dots + \mu_m r_m,$$

which expresses s as a linear combination of the rows of C.

Proof of correctness of the method

$C \in M_{m\times n}(\mathbb{R})$; C' obtained from C by a single row operation; s a row vector of length n. Claim: s is a linear combination of rows of C iff it is a linear combination of rows of C'.

Corollary 9.16: Let C be an $m \times n$ matrix, and let D be obtained from C by a sequence of elementary row operation. Let s be a row vector of length n. Then s can be expressed as a linear combination of the rows of C if and only if it can be expressed as a linear combination of the rows of D.

Proof.
Just apply the lemma to each step in the row-reduction sequence. □

Proof of correctness of the method

Lemma 9.17: Let D be an $m \times n$ matrix in RREF.

(a) Suppose that every column of D contains a pivot, so $D = \left[\begin{array}{c} I_n \\ \hline 0_{(m-n)\times n} \end{array}\right]$. Then every row vector of length n can be expressed as a linear combination of the rows of D.

(b) Suppose instead that the k'th column of D does not contain a pivot. Then the k'th standard basis vector e_k **cannot** be expressed as a linear combination of the rows of D.

Proof of (a): In this case the first n rows are the standard basis vectors

$$r_1 = e_1^T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \end{bmatrix}$$
$$r_2 = e_2^T = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \end{bmatrix} \cdots$$
$$r_n = e_n^T = \begin{bmatrix} 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

and $r_i = 0$ for $i > n$. This means that any row vector $v = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$ can be expressed as

$$\begin{aligned} v = &\begin{bmatrix} v_1 & 0 & 0 & \cdots & 0 \end{bmatrix} + \\ &\begin{bmatrix} 0 & v_2 & 0 & \cdots & 0 \end{bmatrix} + \dots + \\ &\begin{bmatrix} 0 & 0 & 0 & \cdots & v_n \end{bmatrix} \\ = &v_1 r_1 + v_2 r_2 + v_3 r_3 + \dots + v_n r_n, \end{aligned}$$

which is a linear combination of the rows of D.

Proof of correctness of the method

Lemma: Let D be an $m \times n$ matrix in RREF.

(a) Suppose that every column of D contains a pivot, so $D = \left[\begin{array}{c} I_n \\ \hline 0_{(m-n)\times n} \end{array}\right]$.
Then every row vector of length n can be expressed as a linear combination of the rows of D.

(b) Suppose instead that the k'th column of D does not contain a pivot. Then the k'th standard basis vector e_k **cannot** be expressed as a linear combination of the rows of D.

Example for proof of (b): Consider the matrix

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

This is in RREF, with pivots in columns 2, 5 and 8. Let r_i be the i'th row, and consider a linear combination

$\lambda_1 r_1 + \lambda_2 r_2 + \lambda_3 r_3 = \begin{bmatrix} 0 & \lambda_1 & 2\lambda_1 & 3\lambda_1 & \lambda_2 & 4\lambda_1+6\lambda_2 & 5\lambda_1+7\lambda_2 & \lambda_3 \end{bmatrix}$.

The entries in the pivot columns 2, 5 and 8 of s are just the coefficients λ_1, λ_2 and λ_3. This is not a special feature of this example: it simply reflects the fact that pivot columns contain nothing except the pivot. Now consider

$$e_6 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

For this to be $\lambda_1 r_1 + \lambda_2 r_2 + \lambda_3 r_3$ we need $\lambda_1 = 0$ and $\lambda_2 = 0$ and $\lambda_3 = 0$ (by looking in the pivot columns). But that means $\lambda_1 r_1 + \lambda_2 r_2 + \lambda_3 r_3 = 0 \neq e_6$.

Proof of correctness of the method

Lemma: Let D be an $m \times n$ matrix in RREF.

(b) Suppose instead that the k'th column of D does not contain a pivot. Then the k'th standard basis vector e_k **cannot** be expressed as a linear combination of the rows of D.

This line of argument works more generally.
Suppose that D is an RREF matrix and that the k'th column has no pivot.
We claim that e_k is not a linear combination of the rows of D.
We can remove any rows of zeros from D without affecting the question, so we may assume that every row is nonzero, so every row contains a pivot.
Suppose that $e_k = \lambda_1 r_1 + \cdots + \lambda_m r_m$ say.
By looking in the column that contains the first pivot, we see that $\lambda_1 = 0$.
By looking in the column that contains the second pivot, we see that $\lambda_2 = 0$.
Continuing in this way, we see that all the coefficients λ_i are zero, so $\sum_i \lambda_i r_i = 0$, which contradicts the assumption that $e_k = \lambda_1 r_1 + \cdots + \lambda_m r_m$.

We conclude that in fact it is impossible to write e_k as $\lambda_1 r_1 + \cdots + \lambda_m r_m$, so e_k is not a linear combination of the rows of D.

Duality

Consider an $n \times m$ matrix

$$P = \left[\begin{array}{c|c|c} v_1 & \cdots & v_m \end{array}\right] = \left[\begin{array}{c} w_1^T \\ \hline \vdots \\ \hline w_n^T \end{array}\right] \in M_{n\times m}(\mathbb{R})$$

$$P^T = \left[\begin{array}{c|c|c} w_1 & \cdots & w_n \end{array}\right] = \left[\begin{array}{c} v_1^T \\ \hline \vdots \\ \hline v_m^T \end{array}\right] \in M_{m\times n}(\mathbb{R})$$

- The vectors v_i are linearly independent in $\mathbb{R}^n$ if and only if $P \to \left[\begin{array}{c} I_m \\ \hline 0 \end{array}\right]$, if and only if the vectors w_j span $\mathbb{R}^m$.
- The vectors v_i span $\mathbb{R}^n$ if and only if $P^T \to \left[\begin{array}{c} I_n \\ \hline 0 \end{array}\right]$, if and only if the vectors w_j are linearly independent in $\mathbb{R}^m$.

In other words:

- The columns of P are independent if and only if the columns of P^T span.
- The columns of P span if and only if the columns of P^T are independent.

Lecture 5

Bases

Definition 10.1: A *basis* for $\mathbb{R}^n$ is a linearly independent list of vectors in $\mathbb{R}^n$ that also spans $\mathbb{R}^n$.

Remark 10.2: Any basis for $\mathbb{R}^n$ must contain precisely n vectors. Indeed, we saw before that a linearly independent list can contain at most n vectors, that a spanning list must contain at least n vectors. As a basis has both these properties, it must contain precisely n vectors.

Basis example

Consider the list $\mathcal{U} = (u_1, u_2, u_3)$, where

$$u_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad u_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \qquad u_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

For an arbitrary vector $v = \begin{bmatrix} x & y & z \end{bmatrix}^T$ we have

$$(a-b)u_1 + (b-c)u_2 + cu_3 = \begin{bmatrix} a-b \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} b-c \\ b-c \\ 0 \end{bmatrix} + \begin{bmatrix} c \\ c \\ c \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} = v,$$

which expresses v as a linear combination of u_1, u_2 and u_3. This shows that $\mathcal{U}$ spans $\mathbb{R}^3$. Now suppose we have a linear relation $\lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 = 0$. This means that

$$\begin{bmatrix} \lambda_1 + \lambda_2 + \lambda_3 \\ \lambda_2 + \lambda_3 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

from which we read off that $\lambda_3 = 0$, then that $\lambda_2 = 0$, then that $\lambda_1 = 0$. This means that the only linear relation on $\mathcal{U}$ is the trivial one, so $\mathcal{U}$ is linearly independent. As it also spans, we conclude that $\mathcal{U}$ is a basis.

Basis criterion

Proposition 10.4: Given $\mathcal{V} = (v_1, \dots, v_n)$ in $\mathbb{R}^n$, put

$$A = \left[\begin{array}{c|c|c} v_1 & \dots & v_n \end{array} \right] \in M_{n\times n}(\mathbb{R})$$

Then $\mathcal{V}$ is a basis iff $A\lambda = x$ has a **unique** solution for every $x \in \mathbb{R}^n$. **Proof:** Suppose that $\mathcal{V}$ is a basis. In particular, this means that any vector $x \in \mathbb{R}^n$ can be expressed as a linear combination $x = \lambda_1 v_1 + \dots + \lambda_n v_n$.
Thus, if we form the vector $\lambda = \begin{bmatrix} \lambda_1 & \cdots & \lambda_n \end{bmatrix}^T$, we have

$$A\lambda = \left[\begin{array}{c|c|c} v_1 & \cdots & v_n \end{array} \right] \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = \lambda_1 v_1 + \dots + \lambda_n v_n = x,$$

so λ is a solution to $A\lambda = x$. Suppose that μ is also a solution, so

$$\mu_1 v_1 + \dots + \mu_n v_n = x.$$

By subtracting this from the earlier equation, we get

$$(\lambda_1 - \mu_1)v_1 + \dots + (\lambda_n - \mu_n)v_n = 0.$$

This is a linear relation on the independent list $\mathcal{V}$, so it must be the trivial one, so the coefficients $\lambda_i - \mu_i$ are zero, so $\lambda = \mu$. In other words, λ is the **unique** solution to $A\lambda = x$, as required.

Basis criterion

Proposition 10.4: Given $\mathcal{V} = (v_1, \dots, v_n)$ in $\mathbb{R}^n$, put

$$A = \left[\begin{array}{c|c|c} v_1 & \dots & v_n \end{array} \right] \in M_{n\times n}(\mathbb{R})$$

Then $\mathcal{V}$ is a basis iff $A\lambda = x$ has a **unique** solution for every $x \in \mathbb{R}^n$.

We now need to prove the converse. Suppose that for every $x \in \mathbb{R}^n$, the equation $A\lambda = x$ has a unique solution. Equivalently, for every $x \in \mathbb{R}^n$, there is a unique sequence of coefficients $\lambda_1, \dots, \lambda_n$ such that $\lambda_1 v_1 + \dots + \lambda_n v_n = x$. Firstly, we can temporarily ignore the uniqueness, and just note that every element $x \in \mathbb{R}^n$ can be expressed as a linear combination of $v_1, \dots, v_n$. This means that the list $\mathcal{V}$ spans $\mathbb{R}^n$. Next, consider the case $x = 0$. The equation $A\lambda = 0$ has $\lambda = 0$ as one solution. By assumption, the equation $A\lambda = 0$ has a unique solution, so $\lambda = 0$ is the only solution. Using the standard equation for $A\lambda$, we can restate this as follows: the only sequence $(\lambda_1, \dots, \lambda_n)$ for which $\lambda_1 v_1 + \dots + \lambda_n v_n = 0$ is the sequence $(0, \dots, 0)$. In other words, the only linear relation on $\mathcal{V}$ is the trivial one. This means that $\mathcal{V}$ is linearly independent, and it also spans $\mathbb{R}^n$, so it is a basis.

Method to check for a basis

Let $\mathcal{V} = (v_1, \dots, v_m)$ be a list of vectors in $\mathbb{R}^n$.

(a) If $m \neq n$ then $\mathcal{V}$ is not a basis.

(b) If $m = n$ then we form the matrix

$$A = \left[\begin{array}{c|c|c} v_1 & \dots & v_m \end{array} \right]$$

and row-reduce it to get a matrix B.

(c) If $B = I_n$ then $\mathcal{V}$ is a basis; otherwise, it is not.

Proof:

(a) Has been discussed already: any basis of $\mathbb{R}^n$ has n vectors.

(b) If $A \to I_n$ then the same steps give $[A|x] \to [I_n|x']$, then $\lambda = x'$ is the unique solution to $A\lambda = x$. Thus $\mathcal{V}$ is a basis.

(c) If $A \to B \neq I_n$ then B cannot have a pivot in every column. By our method for checking independence, the list $\mathcal{V}$ is dependent and so is not a basis.

Basis example

Consider the vectors

$$v_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 2 \\ 1 \end{bmatrix} \quad v_2 = \begin{bmatrix} 3 \\ 2 \\ 1 \\ 2 \\ 3 \end{bmatrix} \quad v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad v_4 = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 3 \\ 1 \end{bmatrix} \quad v_5 = \begin{bmatrix} 5 \\ 3 \\ 1 \\ 3 \\ 5 \end{bmatrix}$$

To decide whether they form a basis, we construct the corresponding matrix A and start row-reducing it:

$$\begin{bmatrix} 1 & 3 & 1 & 1 & 5 \\ 2 & 2 & 1 & 3 & 3 \\ 3 & 1 & 1 & 5 & 1 \\ 2 & 2 & 1 & 3 & 3 \\ 1 & 3 & 1 & 1 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & 3 & 1 & 1 & 5 \\ 0 & -4 & -1 & 1 & -7 \\ 0 & -8 & -2 & 2 & -14 \\ 0 & -4 & -1 & 1 & -7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 3 & 1 & 1 & 5 \\ 0 & -4 & -1 & 1 & -7 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Already after the first step we have a row of zeros, and it is clear that we will still have a row of zeros after we complete the row-reduction, so A does not reduce to the identity matrix, so the vectors v_i do not form a basis.

Basis example

Consider the vectors

$$p_1 = \begin{bmatrix} 1 \\ 1 \\ 11 \\ 1 \end{bmatrix} \quad p_2 = \begin{bmatrix} 1 \\ 11 \\ 1 \\ 11 \end{bmatrix} \quad p_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 11 \end{bmatrix} \quad p_4 = \begin{bmatrix} 1 \\ 11 \\ 11 \\ 11 \end{bmatrix}$$

To decide whether they form a basis, we construct the corresponding matrix A and row reduce it:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 11 & 1 & 11 \\ 11 & 1 & 1 & 11 \\ 1 & 11 & 11 & 11 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 10 & 0 & 10 \\ 0 & -10 & -10 & 0 \\ 0 & 10 & 10 & 10 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix} \to$$

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

After a few more steps, we obtain the identity matrix. It follows that the list p_1, p_2, p_3, p_4 is a basis.

Coefficients in terms of a basis

Suppose that the list $\mathcal{V} = v_1, \dots, v_n$ is a basis for $\mathbb{R}^n$, and that w is another vector in $\mathbb{R}^n$. By the very definition of a basis, it must be possible to express w (in a unique way) as a linear combination $w = \lambda_1 v_1 + \cdots + \lambda_n v_n$. If we want to find the coefficients λ_i, we can use the following:

Method 10.8: Let $\mathcal{V} = v_1, \dots, v_n$ be a basis for $\mathbb{R}^n$, and let w be another vector in $\mathbb{R}^n$.

(a) Let B be the matrix

$$B = \left[\begin{array}{c|c|c|c} v_1 & \cdots & v_n & w \end{array} \right] \in M_{n\times(n+1)}(\mathbb{R}).$$

(b) Let B' be the RREF form of B. Then B' will have the form $[I_n|\lambda]$ for some column vector

$$\lambda = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix}.$$

(c) Now $w = \lambda_1 v_1 + \cdots + \lambda_n v_n$.

It is clear from our recent discussion that this is valid.

Example of coefficients in terms of a basis

We will express $q = \begin{bmatrix} 0.9 & 0.9 & 0 & 10.9 \end{bmatrix}^T$ in terms of the basis p_1, p_2, p_3, p_4 in the previous example. We form the relevant augmented matrix, and apply the same row-reduction steps as before, except that we now have an extra column.

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0.9 \\ 1 & 11 & 1 & 11 & 0.9 \\ 11 & 1 & 1 & 11 & 0 \\ 1 & 11 & 11 & 11 & 10.9 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0.9 \\ 0 & 10 & 0 & 10 & 0 \\ 0 & -10 & -10 & 0 & -9.9 \\ 0 & 10 & 10 & 10 & 10 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0.9 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0.99 \\ 0 & 1 & 1 & 1 & 1 \end{array}\right] \to$$

$$\left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0.9 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0.99 \\ 0 & 0 & 0 & 1 & 0.01 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 0.9 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0.99 \\ 0 & 0 & 0 & 1 & 0.01 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 0.9 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0.99 \\ 0 & 0 & 0 & 1 & 0.01 \end{array}\right] \to$$

$$\left[\begin{array}{cccc|c} 1 & 0 & 1 & 0 & 0.9 \\ 0 & 1 & 0 & 0 & -0.01 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0.01 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & -0.1 \\ 0 & 1 & 0 & 0 & -0.01 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0.01 \end{array}\right]$$

The final result is $[I_4|\lambda]$, where $\lambda = \begin{bmatrix} -0.1 & -0.01 & 1 & 0.01 \end{bmatrix}^T$. This means that q can be expressed in terms of the vectors p_i as follows:

$$q = -0.1 p_1 - 0.01 p_2 + p_3 + 0.01 p_4.$$

Example of coefficients in terms of a basis

One can check that the vectors u_1, u_2, u_3 and u_4 below form a basis for $\mathbb{R}^4$.

$$u_1 = \begin{bmatrix} 1 \\ 1/2 \\ 1/3 \\ 1/4 \end{bmatrix} \quad u_2 = \begin{bmatrix} 1/2 \\ 1/3 \\ 1/4 \\ 1/5 \end{bmatrix} \quad u_3 = \begin{bmatrix} 1/3 \\ 1/4 \\ 1/5 \\ 1/6 \end{bmatrix} \quad u_4 = \begin{bmatrix} 1/4 \\ 1/5 \\ 1/6 \\ 1/7 \end{bmatrix} \quad v = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

We would like to express v in terms of this basis. The matrix formed by the vectors u_i is called the *Hilbert matrix*; it is notoriously hard to row-reduce. We will therefore use Maple.

Example of coefficients in terms of a basis

```
with(LinearAlgebra):
RREF := ReducedRowEchelonForm;
u[1] := <1,1/2,1/3,1/4>;
u[2] := <1/2,1/3,1/4,1/5>;
u[3] := <1/3,1/4,1/5,1/6>;
u[4] := <1/4,1/5,1/6,1/7>;
v    := <1,1,1,1>;
B    := <u[1]|u[2]|u[3]|u[4]|v>;
RREF(B);
```

$$\left[\begin{array}{cccc|c} 1 & 1/2 & 1/3 & 1/4 & 1 \\ 1/2 & 1/3 & 1/4 & 1/5 & 1 \\ 1/3 & 1/4 & 1/5 & 1/6 & 1 \\ 1/4 & 1/5 & 1/6 & 1/7 & 1 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & -4 \\ 0 & 1 & 0 & 0 & 60 \\ 0 & 0 & 1 & 0 & -180 \\ 0 & 0 & 0 & 1 & 140 \end{array}\right].$$

We conclude that

$$v = -4u_1 + 60u_2 - 180u_3 + 140u_4.$$

Duality for bases

Proposition 10.11: Let A be an $n \times n$ matrix. Then the columns of A form a basis for $\mathbb{R}^n$ if and only if the columns of A^T form a basis for $\mathbb{R}^n$.

Proof.
Recall:

- The colums of A span iff the columns of A^T are independent.
- The columns of A are independent iff the columns of A^T span.
- A list is a basis iff it is independent and also spans.

The claim is clear from this. □

Numerical criteria

Proposition 10.12: Let $\mathcal{V}$ be a list of n vectors in $\mathbb{R}^n$ (so the number of vectors is the same as the number of entries in each vector).

(a) If the list is linearly independent then it also spans, and so is a basis.

(b) If the list spans then it is also linearly independent, and so is a basis.

Proof.

Let A be the matrix whose columns are the vectors in $\mathcal{V}$.

(a) Suppose that $\mathcal{V}$ is linearly independent. Let B be the matrix obtained by row-reducing A. By the standard method for checking (in)dependence, B must have a pivot in every column. As B is also square, we must have $B = I_n$. It follows that $\mathcal{V}$ is a basis.

(b) Suppose instead that $\mathcal{V}$ (which is the list of columns of A) spans $\mathbb{R}^n$. By duality, we conclude that the columns of A^T are linearly independent. Now A^T has n columns, so we can apply part (a) to deduce that the columns of A^T form a basis. By duality again, the columns of A must form a basis as well.

$\square$

Lecture 6

Elementary matrices

Definition 11.1: Fix an integer $n > 0$. We define $n \times n$ matrices as follows.

(a) Suppose that $1 \leq p \leq n$ and that λ is a nonzero real number. We then let $D_p(\lambda)$ be the matrix that is the same as I_n except that $(D_p(\lambda))_{pp} = \lambda$.

(b) Supose that $1 \leq p, q \leq n$ with $p \neq q$, and that μ is an arbitrary real number. We then let $E_{pq}(\mu)$ be the matrix that is the same as I_n except that $(E_{pq}(\lambda))_{pq} = \mu$.

(c) Supose again that $1 \leq p, q \leq n$ with $p \neq q$. We let F_{pq} be the matrix that is the same as I_n except that $(F_{pq})_{pp} = (F_{pq})_{qq} = 0$ and $(F_{pq})_{pq} = (F_{pq})_{qp} = 1$.

An *elementary matrix* is a matrix of one of these types.

Example 11.2: In the case $n = 4$, we have

$$D_2(\lambda) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad E_{24}(\mu) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \mu \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad F_{24} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Elementary matrices and row operations

Proposition 11.3: Let A be an $n \times n$ matrix, and let A' be obtained from A by a single row operation. Then $A' = UA$ for some elementary matrix U. In more detail:

(a) Let A' be obtained from A by multiplying the p'th row by λ. Then $A' = D_p(\lambda)A$.

(b) Let A' be obtained from A by adding μ times the q'th row to the p'th row. Then $A' = E_{pq}(\mu)A$.

(c) Let A' be obtained from A by exchanging the p'th row and the q'th row. Then $A' = F_{pq}A$.

Elementary matrices and row operations

Corollary 11.4: Let A and B be $n \times n$ matrices, and suppose that A can be converted to B by a sequence of row operations. Then $B = UA$ for some matrix U that can be expressed as a product of elementary matrices.

Proof.
The assumption is that there is a sequence of matrices $A_0, A_1, \dots, A_r$ starting with $A_0 = A$ and ending with $A_r = B$ such that A_i is obtained from A_{i-1} by a single row operation. By the Proposition, this means that there is an elementary matrix U_i such that $A_i = U_i A_{i-1}$. This gives

$$A_1 = U_1 A_0 = U_1 A$$
$$A_2 = U_2 A_1 = U_2 U_1 A$$
$$A_3 = U_3 A_2 = U_3 U_2 U_1 A$$

and so on. Eventually we get $B = A_r = U_r U_{r-1} \cdots U_1 A$. We can thus take $U = U_r U_{r-1} \cdots U_1$ and we have $B = UA$ as required. □

Invertibility

Theorem 11.5: Let A be an $n \times n$ matrix. Then the following statements are equivalent: if any one of them is true then they are all true, and if any one of them is false then they are all false.

(a) A can be row-reduced to I_n.
(b) The columns of A are linearly independent.
(c) The columns of A span $\mathbb{R}^n$.
(d) The columns of A form a basis for $\mathbb{R}^n$.
(e) A^T can be row-reduced to I_n.
(f) The columns of A^T are linearly independent.
(g) The columns of A^T span $\mathbb{R}^n$.
(h) The columns of A^T form a basis for $\mathbb{R}^n$.
(i) There is a matrix U such that $UA = I_n$.
(j) There is a matrix V such that $AV = I_n$.

Moreover, if these statements are all true then there is a unique matrix U that satisfies $UA = I_n$, and this is also the unique matrix that satisfies $AU = I_n$ (so the matrix V in (j) is necessarily the same as the matrix U in (i)).

Invertibility — what we already know

(a) A can be row-reduced to I_n.
(b) The columns of A are linearly independent.
(c) The columns of A span $\mathbb{R}^n$.
(d) The columns of A form a basis for $\mathbb{R}^n$.
(e),(f),(g),(h): same for A^T
(i) There is a matrix U such that $UA = I_n$.
(j) There is a matrix V such that $AV = I_n$.

Statements (a) to (d) are equivalent to each other by the "numerical criteria" (Proposition 10.12).

Similarly statements (e) to (h) are equivalent to each other.

Moreover, (a) to (d) are equivalent to (e) to (h) by "duality for bases" (Proposition 10.11).

The real issue is to prove that (a) to (h) are equivalent to (i) and (j).

Invertibility

(a) A can be row-reduced to I_n.
(b) The columns of A are linearly independent.
(c) The columns of A span $\mathbb{R}^n$.
(d) The columns of A form a basis for $\mathbb{R}^n$.
(e),(f),(g),(h): same for A^T
(i) There is a matrix U such that $UA = I_n$.
(j) There is a matrix V such that $AV = I_n$.

- If (a) holds then each row operation corresponds to an elementary matrix, and the product of those is a matrix U with $UA = I_n$; so (i) holds.
- Similarly, if (e) holds then there exists W with $WA^T = I_n$, so $AW^T = I_n$, so can take $V = W^T$ to see that (j) holds.
- Conversely, suppose that (i) holds. Let $v_1, \dots, v_r$ be the columns of A. A linear relation $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$ gives a a vector λ with $A\lambda = 0$. As $UA = I_n$ this gives $\lambda = UA\lambda = U0 = 0$, so our linear relation is the trivial one. Thus the columns v_i are linearly independent, so (b) holds.
- Similarly, (j) implies (f).
- Now (a)$\Leftrightarrow \cdots \Leftrightarrow$(h) and (a)$\Rightarrow$(i)$\Rightarrow$(b) and (e)$\Rightarrow$(j)$\Rightarrow$(f); so (a) to (j) are all equivalent. □

Invertibility

(a) A can be row-reduced to I_n.
(b) The columns of A are linearly independent.
(c) The columns of A span $\mathbb{R}^n$.
(d) The columns of A form a basis for $\mathbb{R}^n$.
(e),(f),(g),(h): same for A^T
(i) There is a matrix U such that $UA = I_n$.
(j) There is a matrix V such that $AV = I_n$.

Definition 11.6:
We say that A is *invertible* if (any one of) the conditions (a) to (j) hold.
If so, we write A^{-1} for the unique matrix satisfying $A^{-1}A = I_n = AA^{-1}$ (which exists by the Theorem).

Remark 11.7: It is clear that A is invertible if and only if A^T is invertible.

Elementary matrices are invertible

All elementary matrices are invertible. More precisely:

(a) $D_p(\lambda^{-1})D_p(\lambda) = I_n$, so $D_p(\lambda)$ is invertible with inverse $D_p(\lambda^{-1})$.
For example, when $n = 4$ and $p = 2$ we have

$$D_2(\lambda)D_2(\lambda^{-1}) = \begin{bmatrix}1&0&0&0\\0&\lambda&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&\lambda^{-1}&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} = \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} = I_4.$$

(b) $E_{pq}(\mu)E_{pq}(-\mu) = I_n$, so $E_{pq}(\mu)$ is invertible with inverse $E_{pq}(-\mu)$.
For example, when $n = 4$ and $p = 2$ and $q = 4$ we have

$$E_{24}(\mu)E_{24}(-\mu) = \begin{bmatrix}1&0&0&0\\0&1&0&\mu\\0&0&1&0\\0&0&0&1\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&1&0&-\mu\\0&0&1&0\\0&0&0&1\end{bmatrix} = \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} = I_4.$$

(c) $F_{pq}^2 = I_n$, so F_{pq} is invertible and is its own inverse. For example, when $n = 4$ and $p = 2$ and $q = 4$ we have

$$F_{24}^2 = \begin{bmatrix}1&0&0&0\\0&0&0&1\\0&0&1&0\\0&1&0&0\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&0&0&1\\0&0&1&0\\0&1&0&0\end{bmatrix} = \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{bmatrix} = I_4.$$

Products of invertible matrices are invertible

Proposition 11.9:
If A and B are invertible $n \times n$ matrices, then AB is also invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.

Proof.
Put $C = AB$ and $D = B^{-1}A^{-1}$.

$$DC = B^{-1}A^{-1}AB = B^{-1}I_nB = B^{-1}B = I_n$$

$$CD = ABB^{-1}A = AI_nA^{-1} = AA^{-1} = I_n$$

This shows that D is an inverse for C, so C is invertible with $C^{-1} = D$ as claimed. □

More generally, if $A_1, A_2, \ldots, A_r$ are invertible $n \times n$ matrices, then the product $A_1A_2 \cdots A_r$ is also invertible, with

$$(A_1A_2 \cdots A_r)^{-1} = A_r^{-1} \cdots A_2^{-1}A_1^{-1}.$$

The proof is similar.

Row reduction and invertible matrices

Corollary 11.10: Let A and B be $n \times n$ matrices, and suppose that A can be converted to B by a sequence of row operations. Then $B = UA$ for some invertible matrix U.

Proof.

- Corollary 11.4 tells us that $B = UA$ for some matrix U that is a product of elementary matrices.
- Example 11.8 tells us that elementary matrices are invertible.
- Proposition 11.9 tells us that products of invertible matrices are invertible.
- Thus, U is invertible.

□

Finding inverses by row-reduction

To check whether A is invertible, row-reduce it and see whether you get the identity. We can find the inverse by a closely related procedure.

Method 11.11: Let A be an $n \times n$ matrix.

(a) Form the augmented matrix $[A|I_n]$ and row-reduce it.

(b) If the result has the form $[I_n|B]$, then A is invertible with $A^{-1} = B$.

(c) If the result has any other form then A is not invertible.

Proof of correctness.
Let $[T|B]$ be the row-reduction of $[A|I_n]$.
Then T is the row-reduction of A, so A is invertible if and only if $T = I_n$.
Suppose that this holds, so $[A|I_n]$ reduces to $[I_n|B]$. As in Corollary 11.4 we see that there is a matrix U such that $[I_n|B] = U[A|I_n] = [UA|U]$. This gives $B = U$ and $UA = I_n$ so $BA = I_n$, so $B = A^{-1}$. □

Example of finding an inverse

Consider the matrix $A = \begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}$. We have the following row-reduction:

$$[A|I_3] = \left[\begin{array}{ccc|ccc} 1 & a & b & 1 & 0 & 0 \\ 0 & 1 & c & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & b-ac & 1 & -a & 0 \\ 0 & 1 & c & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -a & ac-b \\ 0 & 1 & 0 & 0 & 1 & -c \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right]$$

We conclude that $A^{-1} = \begin{bmatrix} 1 & -a & ac-b \\ 0 & 1 & -c \\ 0 & 0 & 1 \end{bmatrix}$.
It is a straightforward exercise to check this directly:

$$\begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -a & ac-b \\ 0 & 1 & -c \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -a & ac-b \\ 0 & 1 & -c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}.$$

Example of finding an inverse

Consider the matrix $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}$. We have the following row-reduction:

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 4 & 0 & 1 & 0 \\ 1 & 3 & 9 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & 2 & 8 & -1 & 0 & 1 \end{array}\right] \to$$
$$\left[\begin{array}{ccc|ccc} 1 & 0 & -2 & 2 & -1 & 0 \\ 0 & 1 & 3 & -1 & 1 & 0 \\ 0 & 0 & 2 & 1 & -2 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 3 & -3 & 1 \\ 0 & 1 & 0 & -5/2 & 4 & -3/2 \\ 0 & 0 & 1 & 1/2 & -1 & 1/2 \end{array}\right]$$

We conclude that
$$A^{-1} = \begin{bmatrix} 3 & -3 & 1 \\ -5/2 & 4 & -3/2 \\ 1/2 & -1 & 1/2 \end{bmatrix}.$$

Lecture 7

Determinants

Definition : For a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the determinant is defined as

$$\det(A) = ad - bc.$$

For a 3×3 matrix $A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ the determinant is defined by

$$\begin{aligned} \det(A) &= a \det \begin{bmatrix} e & f \\ h & i \end{bmatrix} - b \det \begin{bmatrix} d & f \\ g & i \end{bmatrix} + c \det \begin{bmatrix} d & e \\ g & h \end{bmatrix} \\ &= aei + bfg + cdh - afh - bdi - ceg. \end{aligned}$$

We will now discuss determinants for square matrices of any size. There are more details in an appendix to the printed notes, which will not be examined.

Determinants

Definition 12.1: Let A be an $n \times n$ matrix, and let a_{ij} denote the entry in the i'th row of the j'th column. We define

$$\det(A) = \sum_{\sigma} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)},$$

where the sum runs over all permutations σ of the set $\{1, \dots, n\}$. Here $\operatorname{sgn}(\sigma)$ is the signature of σ. This means that $\operatorname{sgn}(\sigma) = +1$ if σ can be written as the product of an even number of transpositions, and $\operatorname{sgn}(\sigma) = -1$ otherwise.

One can check that this agrees with the standard formulae on the previous slide, if $n = 2$ or $n = 3$.

Determinants of triangular matrices

Example 12.4: Let A be an $n \times n$ matrix.

(a) If all the entries below the diagonal are zero, then the determinant is just the product of the diagonal entries: $\det(A) = a_{11} a_{22} \cdots a_{nn} = \prod_{i=1}^{n} a_{ii}$. For example, we have

$$\det \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 8 & 9 \\ 0 & 0 & 0 & 10 \end{bmatrix} = 1 \times 5 \times 8 \times 10 = 400.$$

(b) Similarly, if all the entries above the diagonal are zero, then the determinant is just the product of the diagonal entries.

(c) In particular, if A is a diagonal matrix (so all entries off the diagonal are zero) then both (a) and (b) apply and we have $\det(A) = \prod_{i=1}^{n} a_{ii}$.

(d) In particular, we have $\det(I_n) = 1$.

Basic facts about determinants

Example 12.5: If any row or column of A is zero, then $\det(A) = 0$.

Proposition 12.6: The determinants of elementary matrices are $\det(D_p(\lambda)) = \lambda$ and $\det(E_{pq}(\mu)) = 1$ and $\det(F_{pq}) = -1$.

$$D_2(\lambda) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad E_{24}(\mu) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & \mu \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad F_{24} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

Proposition 12.7: For any square matrix A, we have $\det(A^T) = \det(A)$.

Theorem 12.8: If A and B are $n \times n$ matrices, then $\det(AB) = \det(A)\det(B)$.

Determinants and row operations

Method 12.9: Let A be an $n \times n$ matrix. We can calculate $\det(A)$ by applying row operations to A until we reach a matrix B for which we know $\det(B)$, keeping track of some factors as we go along.

(a) Every time we multiply a row by a number λ, we record the factor λ.

(b) Every time we exchange two rows, we record the factor -1.

Let μ be the product of these factors: then $\det(A) = \det(B)/\mu$.

Most obvious approach: continue until we reach B in RREF.

- If $B = I_n$ then $\det(B) = 1$ and $\det(A) = 1/\mu$.
- If $B \neq I_n$ then B must have a row of zeros so $\det(B) = 0$ and $\det(A) = 0$.

It will often be more efficient to stop the row-reduction at an earlier stage.

Example determinant by row-reduction

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = B.$$

- Subtract row 1 from each of the other rows: no factor.
- Subtract multiples of row 2 from rows 3 and 4: no factor.

As B has two rows of zeros, we see that $\det(B) = 0$.
The method therefore tells us that $\det(A) = \det(B)/\mu = 0$ as well.

Example determinant by row-reduction

$$A = \begin{bmatrix} 3 & 5 & 5 & 5 \\ 1 & 3 & 5 & 5 \\ 1 & 1 & 3 & 5 \\ 1 & 1 & 1 & 3 \end{bmatrix} \to \begin{bmatrix} 0 & 2 & 2 & -4 \\ 0 & 2 & 4 & 2 \\ 0 & 0 & 2 & 2 \\ 1 & 1 & 1 & 3 \end{bmatrix} \xrightarrow{\frac{1}{8}} \begin{bmatrix} 0 & 1 & 1 & -2 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 3 \end{bmatrix} \to$$

$$\begin{bmatrix} 0 & 1 & 1 & -2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 3 \end{bmatrix} \to \begin{bmatrix} 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 3 \end{bmatrix} \xrightarrow{-1} \begin{bmatrix} 0 & 1 & 1 & -2 \\ 1 & 1 & 1 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix} \xrightarrow{-1} \begin{bmatrix} 1 & 1 & 1 & 3 \\ 0 & 1 & 1 & -2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

- Add multiples of row 4 to the other rows: no factor.
- Multiply each of the first three rows by $\frac{1}{2}$: overall factor of $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$.
- Subtract row 1 from row 2: no factor.
- Subtract row 3 from row 2: no factor.
- Exchange rows 2 and 4: factor of -1.
- Exchange rows 1 and 2: another factor of -1.

The final matrix B is upper-triangular, so the determinant is just the product of the diagonal entries, which is $\det(B) = 2$. The product of the factors is $\mu = 1/8$, so $\det(A) = \det(B)/\mu = 16$.

Warning

Warning:

Most slides for this lecture have many transitions overlaying each other, so they cannot be printed in a useful way. Those slides have been omitted from this file. You should look at the printed notes and/or the version of the slides designed for online display instead.

Inverse of a Jordan block

Consider the matrix $P = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. The minor matrices are:

$$M_{11} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad M_{12} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad M_{13} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad M_{14} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$M_{21} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad M_{22} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad M_{23} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad M_{24} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$M_{31} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad M_{32} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad M_{33} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad M_{34} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$M_{41} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad M_{42} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad M_{43} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad M_{44} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Each of these matrices is either upper triangular or lower triangular, so the determinant is the product of the diagonal entries.

Inverse of a Jordan block

Consider the matrix

$$P = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

The corresponding minor determinants are as follows:

$$\begin{array}{llll} m_{11} = 1 & m_{12} = 0 & m_{13} = 0 & m_{14} = 0 \\ m_{21} = 1 & m_{22} = 1 & m_{23} = 0 & m_{24} = 0 \\ m_{31} = 1 & m_{32} = 1 & m_{33} = 1 & m_{34} = 0 \\ m_{41} = 1 & m_{42} = 1 & m_{43} = 1 & m_{44} = 1 \end{array}$$

and thus

$$\operatorname{adj}(P) = \begin{bmatrix} +m_{11} & -m_{21} & +m_{31} & -m_{41} \\ -m_{12} & +m_{22} & -m_{32} & +m_{42} \\ +m_{13} & -m_{23} & +m_{33} & -m_{43} \\ -m_{14} & +m_{24} & -m_{34} & +m_{44} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

As P is upper triangular it is easy to see that $\det(P) = 1$ and so P^{-1} is the same as $\operatorname{adj}(P)$.

Lecture 8

Eigenvalues and eigenvectors

Definition 13.1: Let A be an $n \times n$ matrix, and let λ be a real number. A λ-*eigenvector* for A is a **nonzero** n-vector v with the property that $Av = \lambda v$. We say that λ is an *eigenvalue* of A if there exists a λ-eigenvector for A.

- This is for *square* matrices only.
- If v is a λ-eigenvector, then Av points in the same direction as v (if $\lambda > 0$) or the opposite direction (if $\lambda < 0$) or $Av = 0$ (if $\lambda = 0$).
- Some things would work better if we considered complex eigenvalues, and eigenvectors in $\mathbb{C}^n$, even if the entries in A are real. However, we will stick with the real case for the moment.
- The equation $Av = \lambda v$ is equivalent to the homogeneous equation $(A - \lambda I_n)v = 0$. We can solve this by row-reducing $A - \lambda I_n$ to get a matrix B say. If B has a pivot in every column then (because it is square) it must be the identity, so the reduced equation $Bv = 0$ says $v = 0$, so there are no λ-eigenvectors. If B does not have a pivot in every column then there will be at least one independent variable, so the equation $Bv = 0$ will have some nonzero solutions, which are the λ-eigenvectors for A.

Eigenvector example

Consider the case

$$A = \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} \qquad a = \begin{bmatrix}1\\1\end{bmatrix} \qquad b = \begin{bmatrix}1\\-1\end{bmatrix}.$$

We have

$$Aa = \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}2\\2\end{bmatrix} = 2a \qquad Ab = \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}1\\-1\end{bmatrix} = \begin{bmatrix}0\\0\end{bmatrix} = 0b$$

so a is a 2-eigenvector and b is a 0-eigenvector, so 2 and 0 are eigenvalues. We claim that these are the only eigenvalues, or equivalently that when $\lambda \notin \{0,2\}$ the only solution to $(A-\lambda I_2)v = 0$ is $v = 0$, or equivalently that the matrix $A - \lambda I_2 = \begin{bmatrix}1-\lambda & 1\\ 1 & 1-\lambda\end{bmatrix}$ row-reduces to I_2.

Subtract $1-\lambda$ times row 2 from row 1 to get $\begin{bmatrix}0 & 1-(1-\lambda)^2\\ 1 & 1-\lambda\end{bmatrix}$.

Here $1-(1-\lambda)^2 = 2\lambda - \lambda^2 = \lambda(2-\lambda)$, which is nonzero because $\lambda \notin \{0,2\}$.

Divide the row 1 by this to get $\begin{bmatrix}0 & 1\\ 1 & 1-\lambda\end{bmatrix}$; more steps then give $\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} = I_2$.

Eigenvector example

Consider

$$A = \begin{bmatrix}1&1&1&1\\0&2&2&2\\0&0&3&3\\0&0&0&4\end{bmatrix} \qquad a = \begin{bmatrix}1\\0\\0\\0\end{bmatrix} \qquad b = \begin{bmatrix}1\\1\\0\\0\end{bmatrix} \qquad c = \begin{bmatrix}3\\4\\2\\0\end{bmatrix} \qquad d = \begin{bmatrix}8\\12\\9\\3\end{bmatrix}.$$

We have

$$Ad = \begin{bmatrix}1&1&1&1\\0&2&2&2\\0&0&3&3\\0&0&0&4\end{bmatrix}\begin{bmatrix}8\\12\\9\\3\end{bmatrix} = \begin{bmatrix}32\\48\\36\\12\end{bmatrix} = 4d,$$

which means that d is a 4-eigenvector for A, and 4 is an eigenvalue of A. Equally direct calculation shows that $Aa = a$ and $Ab = 2b$ and $Ac = 3c$, so a, b and c are also eigenvectors, and 1, 2 and 3 are also eigenvalues of A. Using the general theory that we will discuss below, we can show that

(a) The only 1-eigenvectors are the nonzero multiples of a.
(b) The only 2-eigenvectors are the nonzero multiples of b.
(c) The only 3-eigenvectors are the nonzero multiples of c.
(d) The only 4-eigenvectors are the nonzero multiples of d.
(e) There are no more eigenvalues: if λ is a real number other than 1, 2, 3 and 4, then the equation $Av = \lambda v$ has $v = 0$ as the only solution, so there are no λ-eigenvectors.

The characteristic polynomial

Definition 13.8: Let A be an $n \times n$ matrix. We define $\chi_A(t) = \det(A - t\,I_n)$ (where I_n is the identity matrix). This is the *characteristic polynomial* of A.

Example 13.9: For $A = \begin{bmatrix}a & b\\ c & d\end{bmatrix}$ we have $A - tI_2 = \begin{bmatrix}a-t & b\\ c & d-t\end{bmatrix}$ so

$$\chi_A(t) = \det\begin{bmatrix}a-t & b\\ c & d-t\end{bmatrix} = (a-t)(d-t) - bc = t^2 - (a+d)t + (ad-bc).$$

When $A = \begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}$ we have

$$\chi_A(t) = t^2 - (1+4)t + (1\times 4 - 2\times 3) = t^2 - 5t - 2.$$

Theorem 13.11: The eigenvalues of A are the roots of the characteristic polynomial.

Characteristic polynomial example

Consider $A = \begin{bmatrix}2 & -1 & 2\\ -1 & 3 & -1\\ 2 & -1 & 2\end{bmatrix}$, so $\chi_A(t) = \det\begin{bmatrix}2-t & -1 & 2\\ -1 & 3-t & -1\\ 2 & -1 & 2-t\end{bmatrix}$

$$= (2-t)\det\begin{bmatrix}3-t & -1\\ -1 & 2-t\end{bmatrix} - (-1)\det\begin{bmatrix}-1 & -1\\ 2 & 2-t\end{bmatrix} + 2\det\begin{bmatrix}-1 & 3-t\\ 2 & -1\end{bmatrix}$$

$$\det\begin{bmatrix}3-t & -1\\ -1 & 2-t\end{bmatrix} = (3-t)(2-t) - (-1)(-1) = t^2 - 5t + 5$$

$$\det\begin{bmatrix}-1 & -1\\ 2 & 2-t\end{bmatrix} = (-1)(2-t) - (-1)(2) = t$$

$$\det\begin{bmatrix}-1 & 3-t\\ 2 & -1\end{bmatrix} = (-1)(-1) - (3-t)(2) = 2t - 5$$

$$\begin{aligned}\chi_A(t) &= (2-t)(t^2-5t+5) + t + 2(2t-5) = -t^3 + 7t^2 - 10t\\ &= -t(t-2)(t-5).\end{aligned}$$

The eigenvalues of A are the roots of $\chi_A(t)$, namely 0, 2 and 5.

Eigenvalue example

Consider $A = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ -1 & 0 & 0 \end{bmatrix}$, so $\chi_A(t) = \det \begin{bmatrix} -1-t & 1 & 0 \\ -1 & -t & 1 \\ -1 & 0 & -t \end{bmatrix}$

$$= (-1-t)\det\begin{bmatrix} -t & 1 \\ 0 & -t \end{bmatrix} - \det\begin{bmatrix} -1 & 1 \\ -1 & -t \end{bmatrix} + 0\det\begin{bmatrix} -1 & -t \\ -1 & 0 \end{bmatrix}$$

$$= -t^2(1+t) - (t+1) + 0 = -(1+t^2)(1+t).$$

Eigenvector example

$$A = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \\ -1 & 0 & 0 \end{bmatrix} \qquad \chi_A(t) = -(1+t^2)(1+t)$$

As $1+t^2$ is always positive, the only way $-(1+t^2)(1+t)$ can be zero is if $t=-1$. Thus, the only real eigenvalue of A is -1. When $\lambda=-1$ we have

$$A - \lambda I_3 = A + I_3 = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix}.$$

To find an eigenvector of eigenvalue -1, solve $(A+I_3)u=0$, or

$$\begin{bmatrix} 0 & 1 & 0 \\ -1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

or ($y=0$ and $-x+y+z=0$ and $-x+z=0$). These equations reduce to $x=z$ with $y=0$, so $\begin{bmatrix} x & y & z \end{bmatrix} = z\begin{bmatrix} 1 & 0 & 1 \end{bmatrix}$. This means that the (-1)-eigenvectors are just the nonzero multiples of the vector $\begin{bmatrix} 1 & 0 & 1 \end{bmatrix}^T$.

General method for eigenvectors

Method 13.14: Suppose we have an $n\times n$ matrix A, and we want to find the eigenvalues and eigenvectors.

(a) Calculate the characteristic polynomial $\chi_A(t) = \det(A - tI_n)$.

(b) Find all the real roots of $\chi_A(t)$, and list them as $\lambda_1,\dots,\lambda_k$. These are the eigenvalues of A.

(c) For each eigenvalue λ_i, row reduce the matrix $A - \lambda_i I_n$ to get a matrix B.

(d) Read off solutions to the equation $Bu=0$ (which is easy because B is in RREF). These are the λ_i-eigenvectors of the matrix A.

Eigenvector example

Consider the matrix

$$A = \begin{bmatrix} 16 & 2 & 1 & 1 \\ 2 & 16 & 1 & 1 \\ 1 & 1 & 16 & 2 \\ 1 & 1 & 2 & 16 \end{bmatrix}$$

We will take it as given here that $\chi_A(t) = (t-14)^2(t-16)(t-20)$. Thus, the eigenvalues of A are 14, 16 and 20. To find the eigenvectors of eigenvalue 14, we write down the matrix $A - 14I_4$ and row-reduce it to get a matrix B as follows:

$$\begin{bmatrix} 2 & 2 & 1 & 1 \\ 2 & 2 & 1 & 1 \\ 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix} \to \begin{bmatrix} 0 & 0 & -3 & -3 \\ 0 & 0 & -3 & -3 \\ 1 & 1 & 2 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

If we write $u = \begin{bmatrix} a & b & c & d \end{bmatrix}^T$, then the equation $Bu=0$ just gives $a+b=c+d=0$, so $a=-b$ and $c=-d$ (with b and d arbitrary), so

$$u = \begin{bmatrix} -b & b & -d & d \end{bmatrix}^T$$

for some $b,d\in\mathbb{R}$. The eigenvectors of eigenvalue 14 are precisely the nonzero vectors of the above form. (Recall that eigenvectors are nonzero, by definition.)

Nasty eigenvalues

Using Maple, we find that one eigenvalue of the matrix

$$A = \begin{bmatrix} -1 & 0 & 0 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & 0 \\ -1 & -1 & -1 & -1 \end{bmatrix} \qquad \text{is}$$

$$-1/2+1/12\,\sqrt{6}\sqrt{\frac{10\,\sqrt[3]{892+36\,\sqrt{597}}+\left(892+36\,\sqrt{597}\right)^{2/3}+28}{\sqrt[3]{892+36\,\sqrt{597}}}}+$$

$$1/12\,\sqrt{6}\sqrt{\frac{20\,\sqrt[3]{892+36\,\sqrt{597}}\sqrt{\frac{10\,\sqrt[3]{892+36\,\sqrt{597}}+\left(892+36\,\sqrt{597}\right)^{2/3}+28}{\sqrt[3]{892+36\,\sqrt{597}}}}-\sqrt{\frac{10\,\sqrt[3]{892+36\,\sqrt{597}}+\left(892+36\,\sqrt{597}\right)^{2/3}+28}{\sqrt[3]{892+36\,\sqrt{597}}}}\left(892+36\,\sqrt{597}\right)^{2/3}-28\sqrt{10\,\sqrt[3]{892\cdots}}}{\sqrt[3]{892+36\,\sqrt{597}}\sqrt{\frac{10\,\sqrt[3]{892+36\,\sqrt{597}}+\left(892+36\,\sqrt{597}\right)^{2/3}+28}{\sqrt[3]{892+36\,\sqrt{597}}}}}}$$

This level of complexity is quite normal, even for matrices whose entries are all 0 or ±1. Most examples in this course are carefully constructed to have simple eigenvalues and eigenvectors, but you should be aware that this is not typical. The methods that we discuss will work perfectly well for all matrices, but in practice we need to use computers to do the calculations. Also, it is rarely useful to work with exact expressions for the eigenvalues when they are as complicated as those above. Instead we should use the numerical approximation $\lambda \simeq 1.496698205$.

Eigenvector example

Consider $A = \begin{bmatrix} 3 & 0 & 0 & 2 \\ 0 & 0 & 2 & 0 \\ 0 & 2 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix}$. We will take it as given that $\chi_A(t) = (t+1)(t+2)(t-2)(t-4)$, so the eigenvalues are -1, -2, 2 and 4. To find the eigenvectors of eigenvalue 2, we write down the matrix $A - 2I_4$ and row-reduce it to get a matrix B in RREF:

$$\begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & -2 & 2 & 0 \\ 0 & 2 & -2 & 0 \\ 2 & 0 & 0 & -2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & -2 & 2 & 0 \\ 0 & 2 & -2 & 0 \\ 0 & 0 & 0 & -4 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 2 & -2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} = B.$$

If we write $u = \begin{bmatrix} a & b & c & d \end{bmatrix}^T$, then the equation $Bu = 0$ just gives $a = b - c = d = 0$, so

$$u = \begin{bmatrix} 0 & c & c & 0 \end{bmatrix}^T = c\begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix}^T.$$

for some $c \in \mathbb{R}$. The eigenvectors of eigenvalue 2 are precisely the nonzero vectors of the above form. In particular, the vector $\begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix}^T$ is an eigenvector of eigenvalue 2.

Eigenvector example

We will find the eigenvalues and eigenvectors for $A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 3 & 0 \\ 4 & 0 & 0 \end{bmatrix}$.

$$\chi_A(t) = \det\begin{bmatrix} -t & 0 & 1 \\ 0 & 3-t & 0 \\ 4 & 0 & -t \end{bmatrix} = -t\det\begin{bmatrix} 3-t & 0 \\ 0 & -t \end{bmatrix} + \det\begin{bmatrix} 0 & 3-t \\ 4 & 0 \end{bmatrix}$$
$$= -t^3 + 3t^2 + 4t - 12 = (4-t^2)(t-3) = -(t-2)(t+2)(t-3)$$

Thus, the eigenvalues are -2, 2 and 3.
For the eigenvectors $\begin{bmatrix} a & b & c \end{bmatrix}^T$ of eigenvalue -2:

$$A + 2I_3 = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 5 & 0 \\ 4 & 0 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The eigenvectors of eigenvalue -2 are solutions to the equation

$$\begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

or $a + c/2 = 0$ and $b = 0$. Take $c = 2$ to get the eigenvector $\begin{bmatrix} 1 & 0 & 2 \end{bmatrix}^T$.

Eigenvector example

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 3 & 0 \\ 4 & 0 & 0 \end{bmatrix} \qquad \text{eigenvalues } -2,\ 2 \text{ and } 3$$

For the eigenvectors of eigenvalue 2:

$$A - 2I_3 = \begin{bmatrix} -2 & 0 & 1 \\ 0 & 1 & 0 \\ 4 & 0 & -2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1/2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

This gives $a - c/2 = 0$ and $b = 0$.
Take $c = 2$ to get the eigenvector $\begin{bmatrix} 1 & 0 & 2 \end{bmatrix}^T$.

For the eigenvectors of eigenvalue 3:

$$A - 3I_3 = \begin{bmatrix} -3 & 0 & 1 \\ 0 & 0 & 0 \\ 4 & 0 & -3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1/3 \\ 0 & 0 & 0 \\ 4 & 0 & -3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1/3 \\ 0 & 0 & 0 \\ 0 & 0 & -5/3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

This gives $a = c = 0$. Take $b = 1$ to get the eigenvector $\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$.

Independence of eigenvectors

Proposition 13.19: Let A be a $d \times d$ matrix, and let $v_1, \dots, v_n$ be eigenvectors of A. Suppose that the corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ are all different.
Then the list $v_1, \dots, v_n$ is linearly independent.

Proof for $n = 2$.
Suppose we have a linear relation $\quad \alpha_1 v_1 + \alpha_2 v_2 = 0. \quad$ (P)
We now multiply both sides of this vector equation by the matrix $A - \lambda_2 I$.

$$\alpha_1(\lambda_1 - \lambda_2)v_1 + \alpha_2(\lambda_2 - \lambda_2)v_2 = 0$$

As the number $\lambda_1 - \lambda_2$ and the vector v_1 are nonzero, we can conclude that $\alpha_1 = 0$. If we instead multiply equation (P) by $A - \lambda_1 I$ we get

$$\alpha_2(\lambda_2 - \lambda_1)v_2 = 0.$$

As the number $\lambda_2 - \lambda_1$ and the vector v_2 are nonzero, we can conclude that $\alpha_2 = 0$. We have now seen that $\alpha_1 = \alpha_2 = 0$, so the relation (P) is the trivial relation. As this works for any linear relation between v_1 and v_2, we see that these vectors are linearly independent. $\square$

Independence of eigenvectors

Proposition 13.19: Let A be a $d \times d$ matrix, and let $v_1, \dots, v_n$ be eigenvectors of A. Suppose that the corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ are all different. Then the list $v_1, \dots, v_n$ is linearly independent.

Proof for $n = 3$.
Suppose we have a linear relation $\quad \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = 0. \quad$ (P)
Multiply both sides by $A - \lambda_3 I$, then by $A - \lambda_2 I$

$$\alpha_1(\lambda_1 - \lambda_3)v_1 + \alpha_2(\lambda_2 - \lambda_3)v_2 + \alpha_3(\lambda_3 - \lambda_3)v_3 = 0$$

$$\alpha_1(\lambda_1 - \lambda_3)(\lambda_1 - \lambda_2)v_1 + \alpha_2(\lambda_2 - \lambda_3)(\lambda_2 - \lambda_2)v_2 + \alpha_3(\lambda_3 - \lambda_3)(\lambda_2 - \lambda_2)v_3 = 0$$

As the eigenvalues are all different $(\lambda_1 - \lambda_3)(\lambda_1 - \lambda_2) \neq 0$. As v_1 is an eigenvector it is nonzero. It follows that $\alpha_1 = 0$. Similarly, multiplying (P) by $(A - \lambda_1 I)(A - \lambda_3 I)$ makes the first and third terms go away leaving $\alpha_2(\lambda_2 - \lambda_1)(\lambda_2 - \lambda_3)v_2 = 0$ and so $\alpha_2 = 0$. Similarly, multiplying (P) by $(A - \lambda_1 I)(A - \lambda_2 I)$ gives $\alpha_3(\lambda_3 - \lambda_1)(\lambda_3 - \lambda_2)v_3 = 0$ and $\alpha_3 = 0$.
We now see that $\alpha_1 = \alpha_2 = \alpha_3 = 0$, so relation (P) is the trivial relation. This means that the list v_1, v_2, v_3 is linearly independent. $\square$

Independence of eigenvectors

Proposition 13.19: Let A be a $d \times d$ matrix, and let $v_1, \dots, v_n$ be eigenvectors of A. Suppose that the corresponding eigenvalues $\lambda_1, \dots, \lambda_n$ are all different. Then the list $v_1, \dots, v_n$ is linearly independent.

Proof for general n.
Suppose we have a linear relation $\quad \alpha_1 v_1 + \cdots + \alpha_n v_n = 0. \quad$ (P)
For any k, we can multiply (P) by the product of all the matrices $A - \lambda_i I$ for $i \neq k$. This makes all the terms go away except for the k'th term. All that is left is

$$\alpha_k \left(\prod_{i \neq k} (\lambda_k - \lambda_i) \right) v_k = 0.$$

As all the eigenvalues are assumed to be different, the product in brackets is nonzero, so we can divide to get $\alpha_k v_k = 0$. As $v_k \neq 0$ this gives $\alpha_k = 0$. This holds for all k, so relation (P) is the trivial relation. This means that the list $v_1, \dots, v_n$ is linearly independent. $\square$

A generalisation

Suppose we have:

- A $d \times d$ matrix A
- A list $\lambda_1, \dots, \lambda_r$ of distinct eigenvalues
- A linearly independent list $\mathcal{V}_1 = (v_{1,1}, \dots, v_{1,h_1})$ of eigenvectors, all with eigenvalue λ_1
- A linearly independent list $\mathcal{V}_2 = (v_{2,1}, \dots, v_{2,h_2})$ of eigenvectors, all with eigenvalue λ_2
- ....
- A linearly independent list $\mathcal{V}_r = (v_{r,1}, \dots, v_{r,h_r})$ of eigenvectors, all with eigenvalue λ_r

We can then combine the lists $\mathcal{V}_1, \dots, \mathcal{V}_r$ into a single list

$$\mathcal{W} = (v_{1,1}, \cdots, v_{1,h_1}, v_{2,1}, \cdots, v_{2,h_2}, \cdots, v_{r,1}, \cdots, v_{r,h_r}).$$

One can show that the combined list $\mathcal{W}$ is linearly independent.
The problem sheet asks you to prove this.

Lecture 9

Eigenvector bases

Let A be an $n \times n$ matrix. Recall:

(a) If $u_1, \dots, u_k$ are eigenvectors, with eigenvalues $\lambda_1, \dots, \lambda_k$, and these eigenvalues are all different, then the vectors $u_1, \dots, u_k$ are independent.

(b) The eigenvalues are the roots of $\chi_A(t)$, which is a polynomial of degree n. Thus, there are at most n different eigenvalues.

(c) Suppose there are exactly n distinct eigenvalues, say $\lambda_1, \dots, \lambda_n$. We can then choose an eigenvector u_i for each eigenvalue λ_i, and part (a) says that the list $\mathcal{U} = u_1, \dots, u_n$ is independent. As $\mathcal{U}$ is an independent list of n vectors in $\mathbb{R}^n$, it is in fact a basis.

Eigenvector basis example

Consider $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$, so $\chi_A(t) = \det(A - tI) = (1-t)(2-t)(3-t)$, so the eigenvalues are 1, 2 and 3. Suppose we have eigenvectors u_1, u_2 and u_3, where u_k has eigenvalue k. By the previous slide: the list u_1, u_2, u_3 is a basis for $\mathbb{R}^3$. We can find the eigenvectors explicitly by row-reduction:

$$\begin{aligned}
A - I &= \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix} \to \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} & u_1 &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\
A - 2I &= \begin{bmatrix} -1 & 1 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} & u_2 &= \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \\
A - 3I &= \begin{bmatrix} -2 & 1 & 1 \\ 0 & -1 & 2 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -3/2 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix} & u_3 &= \begin{bmatrix} 3/2 \\ 2 \\ 1 \end{bmatrix}.
\end{aligned}$$

We can check more directly that the u_i form a basis:

$$[u_1 | u_2 | u_3] = \begin{bmatrix} 1 & 1 & 3/2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3.$$

Eigenvector basis example

Consider $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, so $\chi_A(t) = \det \begin{bmatrix} -t & 1 \\ -1 & -t \end{bmatrix} = t^2 + 1$.

For all $t \in \mathbb{R}$ we have $t^2 + 1 \geq 1 > 0$, so the characteristic polynomial has no real roots, so there are no real eigenvalues or eigenvectors.

However, there are complex eigenvalues i and $-i$, with corresponding eigenvectors $u_1 = \begin{bmatrix} 1 \\ i \end{bmatrix}$ and $u_2 = \begin{bmatrix} 1 \\ -i \end{bmatrix}$, which form a basis for $\mathbb{C}^2$.

This example and the previous one are typical. If we pick an $n \times n$ matrix at random, it will usually have n different eigenvalues (some of which will usually be complex), and so the corresponding eigenvectors will form a basis for $\mathbb{C}^n$. However, there are some exceptions, as we will see soon. Such exceptions usually arise because of some symmetry or other interesting feature of the problem that gives rise to the matrix.

Eigenvector basis example

Consider $A = \begin{bmatrix} 5 & 5 & 0 \\ 0 & 5 & 5 \\ 0 & 0 & 5 \end{bmatrix}$, so $\chi_A(t) = (5-t)^3$, so the only eigenvalue is 5.

The eigenvectors are the solutions of $\begin{bmatrix} 0 & 5 & 0 \\ 0 & 0 & 5 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, which reduces

to $5y = 5z = 0$ so $y = z = 0$, so the eigenvectors are the multiples of $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.

This means that any two eigenvectors are multiples of each other, and so are linearly dependent. Thus, we cannot find a basis consisting of eigenvectors.

Eigenvector basis example

Consider $A = \begin{bmatrix} 0 & 0 & 5 \\ 0 & 5 & 0 \\ 5 & 0 & 0 \end{bmatrix}$. The characteristic polynomial is

$$\chi_A(t) = \det \begin{bmatrix} -t & 0 & 5 \\ 0 & 5-t & 0 \\ 5 & 0 & -t \end{bmatrix} = -t \det \begin{bmatrix} 5-t & 0 \\ 0 & -t \end{bmatrix} + 5 \det \begin{bmatrix} 0 & 5-t \\ 5 & 0 \end{bmatrix}$$
$$= t^2(5-t) - 25(5-t) = -(t-5)(t^2-25) = -(t-5)(t-5)(t+5)$$
$$= -(t-5)^2(t+5).$$

The only eigenvalues (even in $\mathbb{C}$) are 5 and -5. As there are only two distinct eigenvalues, we do not *automatically* have a basis of eigenvectors. However, it turns out that there is a basis of eigenvectors anyway. Indeed, we can take

$$u_1 = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}^T \qquad u_2 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T \qquad u_3 = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}^T.$$

We can check that $Au_1 = 5u_1$ and $Au_2 = 5u_2$ and $Au_3 = -5u_3$, so the u_i are eigenvectors with eigenvalues 5, 5 and -5 respectively. We can also check that the u_i form a basis, either by row-reducing $[u_1|u_2|u_3]$ or using the formula

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} (x+z)/2 \\ 0 \\ (x+z)/2 \end{bmatrix} + \begin{bmatrix} 0 \\ y \\ 0 \end{bmatrix} + \begin{bmatrix} (x-z)/2 \\ 0 \\ (z-x)/2 \end{bmatrix} = \frac{x+z}{2} u_1 + y u_2 + \frac{x-z}{2} u_3.$$

Diagonalisation

Definition 14.1: We write $\mathrm{diag}(\lambda_1, \dots, \lambda_n)$ for the $n \times n$ matrix such that the entries on the diagonal are $\lambda_1, \dots, \lambda_n$ and the entries off the diagonal are zero.

Example 14.2: $\mathrm{diag}(5,6,7,8) = \begin{bmatrix} 5 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 8 \end{bmatrix}$

Definition 14.3: Let A be an $n \times n$ matrix.

- To *diagonalise* A means to give an invertible matrix U and a diagonal matrix D such that $U^{-1}AU = D$ (or equivalently $A = UDU^{-1}$).
- We say that A is *diagonalisable* if there exist matrices U and D with these properties.

Diagonalisation and eigenvectors

Proposition 14.4: Suppose we have a basis $u_1, \dots, u_n$ for $\mathbb{R}^n$ such that each vector u_i is an eigenvector for A, with eigenvalue λ_i say.

Put $U = [u_1|\cdots|u_n]$ and $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$.

Then $U^{-1}AU = D$, so we have a diagonalisation of A.

Moreover, every diagonalisation of A occurs in this way.
The proof will be given after a lemma.

A matrix multiplication lemma

Lemma 14.5: Let A and U be $n \times n$ matrices, let $\lambda_1, \dots, \lambda_n$ be real numbers, and put $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$. Let $u_1, \dots, u_n$ be the columns of U. Then

$$AU = \left[\begin{array}{c|c|c} Au_1 & \cdots & Au_n \end{array}\right] \qquad UD = \left[\begin{array}{c|c|c} \lambda_1 u_1 & \cdots & \lambda_n u_n \end{array}\right].$$

Proof: Let the rows of A be $a_1^T, \dots, a_n^T$. By the definition of matrix multiplication, we have

$$AU = \begin{bmatrix} a_1.u_1 & \cdots & a_1.u_n \\ \cdots & \cdots & \cdots \\ a_n.u_1 & \cdots & a_n.u_n \end{bmatrix}$$

The p'th column is $\begin{bmatrix} a_1.u_p \\ \vdots \\ a_n.u_p \end{bmatrix}$, and this is just the definition of Au_p. In other words, we have

$$AU = \left[\begin{array}{c|c|c} Au_1 & \cdots & Au_n \end{array}\right]$$

A matrix multiplication lemma

Lemma 14.5: Let A and U be $n \times n$ matrices, let $\lambda_1, \dots, \lambda_n$ be real numbers, and put $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$. Let $u_1, \dots, u_n$ be the columns of U. Then

$$AU = \left[\begin{array}{c|c|c} Au_1 & \cdots & Au_n \end{array}\right] \qquad UD = \left[\begin{array}{c|c|c} \lambda_1 u_1 & \cdots & \lambda_n u_n \end{array}\right].$$

Proof continued: For the second claim, we just do the 3×3 case:

$$UD = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = \begin{bmatrix} \lambda_1 a & \lambda_2 b & \lambda_3 c \\ \lambda_1 d & \lambda_2 e & \lambda_3 f \\ \lambda_1 g & \lambda_2 h & \lambda_3 i \end{bmatrix}$$

Everything in the first column gets multiplied by λ_1, everything in the second column gets multiplied by λ_2 and everything in the third column gets multiplied by λ_3. In other words, we have

$$\left[\begin{array}{c|c|c} u_1 & u_2 & u_3 \end{array}\right] \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = \left[\begin{array}{c|c|c} \lambda_1 u_1 & \lambda_2 u_2 & \lambda_3 u_3 \end{array}\right]$$

as claimed.

Diagonalisation and eigenvectors

Proposition 14.4: Suppose we have a basis $u_1, \dots, u_n$ for $\mathbb{R}^n$ such that each vector u_i is an eigenvector for A, with eigenvalue λ_i say. Put $U = [u_1 | \cdots | u_n]$ and $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$. Then $U^{-1}AU = D$, so we have a diagonalisation of A. Moreover, every diagonalisation arises in this way.

Proof.

- The columns u_i of U form a basis for $\mathbb{R}^n$, so U is invertible.
- First half of the lemma: $AU = [Au_1 | \cdots | Au_n]$. But u_i is an eigenvector of eigenvalue λ_i, so $Au_i = \lambda_i u_i$, so $AU = [\lambda_1 u_1 | \cdots | \lambda_n u_n]$.
- Second half of the lemma: $UD = [\lambda_1 u_1 | \cdots | \lambda_n u_n]$. So $AU = UD$.
- It follows that $U^{-1}AU = U^{-1}UD = D$ and $UDU^{-1} = AUU^{-1} = A$.

Conversely: suppose we have an invertible matrix U and a diagonal matrix $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ such that $U^{-1}AU = D$. Let $u_1, \dots, u_n$ be the columns of U. By reversing the above steps: u_i is an eigenvector of eigenvalue λ_i, and $u_1, \dots, u_n$ is a basis for $\mathbb{R}^n$. $\square$

Diagonalisation example

Example 13.23: the matrix $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$ has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 2$ and $\lambda_3 = 3$; and eigenvectors

$$u_1 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T \qquad u_2 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T \qquad u_3 = \begin{bmatrix} 3/2 & 2 & 1 \end{bmatrix}^T.$$

It follows that $A = UDU^{-1}$, where

$$U = \left[\begin{array}{c|c|c} u_1 & u_2 & u_3 \end{array}\right] = \begin{bmatrix} 1 & 1 & 3/2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \qquad D = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

$$\begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -a & ac-b \\ 0 & 1 & -c \\ 0 & 0 & 1 \end{bmatrix}; \qquad U^{-1} = \begin{bmatrix} 1 & -1 & 1/2 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}.$$

We thus have

$$DU^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1/2 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1/2 \\ 0 & 2 & -4 \\ 0 & 0 & 3 \end{bmatrix}$$

$$UDU^{-1} = \begin{bmatrix} 1 & 1 & 3/2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 1/2 \\ 0 & 2 & -4 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix} = A.$$

Diagonalisation example

In Example 13.24 we showed that the matrix $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ does not have any real eigenvalues or eigenvectors, but that over the complex numbers we have eigenvectors $u_1 = \begin{bmatrix} 1 \\ i \end{bmatrix}$ and $u_2 = \begin{bmatrix} 1 \\ -i \end{bmatrix}$ with eigenvalues $\lambda_1 = i$ and $\lambda_2 = -i$.
We thus have a diagonalisation $A = UDU^{-1}$ with

$$U = [u_1|u_2] = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix} \qquad D = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \qquad U^{-1} = \frac{1}{-2i}\begin{bmatrix} -i & -1 \\ -i & 1 \end{bmatrix} = \begin{bmatrix} 1/2 & -i/2 \\ 1/2 & i/2 \end{bmatrix}.$$

This gives

$$UDU^{-1} = \begin{bmatrix} 1 & 1 \\ i & -i \end{bmatrix}\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}\begin{bmatrix} 1/2 & -i/2 \\ 1/2 & i/2 \end{bmatrix} = \begin{bmatrix} i & -i \\ -1 & -1 \end{bmatrix}\begin{bmatrix} 1/2 & -i/2 \\ 1/2 & i/2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

As expected, this is the same as A.

Lecture 10

Non-diagonalisation example

Consider the matrix $A = \begin{bmatrix} 5 & 5 & 0 \\ 0 & 5 & 5 \\ 0 & 0 & 5 \end{bmatrix}$.

The characteristic poly is $(t-5)^3$, so the only eigenvalue is $\lambda = 5$.

Any eigenvector $u = \begin{bmatrix} x & y & z \end{bmatrix}^T$ must satisfy $(A - 5I_3)u = 0$ so

$$\begin{bmatrix} 0 & 5 & 0 \\ 0 & 0 & 5 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \qquad \begin{aligned} 5y &= 0 \\ 5z &= 0 \\ 0 &= 0 \end{aligned}$$

so $u = \begin{bmatrix} x & 0 & 0 \end{bmatrix}^T$.

It follows that there is no basis of eigenvectors, so A is not diagonalisable.

It is possible to understand non-diagonalisable matrices using the theory of "Jordan blocks". However, we will not cover Jordan blocks in this course.

Powers and eigenvectors

Let A be an $n \times n$ matrix. We can form the powers $A^2 = AA$, $A^3 = AAA$ and so on, and these are again $n \times n$ matrices. It is conventional to take $A^0 = I_n$ and $A^1 = A$.

Now let u be an eigenvector of eigenvalue λ.

$$\begin{aligned} A^0u &= I_nu = u \\ A^1u &= Au = \lambda u \\ A^2u &= A.Au = A.\lambda u = \lambda Au = \lambda^2 u \\ A^3u &= A.A^2u = A.\lambda^2 u = \lambda^2 Au = \lambda^3 u \\ A^4u &= A.A^3u = A.\lambda^3 u = \lambda^3 Au = \lambda^4 u \end{aligned}$$

and in general $A^k u = \lambda^k u$ for all $k \geq 0$.

This is a key point in many applications of eigenvalues and eigenvectors.

Powers of diagonalised matrices

Proposition 14.9: Suppose we have a diagonalisation $A = UDU^{-1}$, where $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ say. Then for all $k \geq 0$ we have $D^k = \text{diag}(\lambda_1^k, \dots, \lambda_n^k)$ and

$$A^k = UD^kU^{-1} = U\ \text{diag}(\lambda_1^k, \dots, \lambda_n^k)\ U^{-1}.$$

Proof: For example:

$$\begin{aligned} A^4 &= (UDU^{-1})^4 = (UDU^{-1})(UDU^{-1})(UDU^{-1})(UDU^{-1}) \\ &= UD(U^{-1}U)D(U^{-1}U)D(U^{-1}U)DU^{-1} = UDDDDU^{-1} = UD^4U^{-1} \end{aligned}$$

It is clear that the general case works the same way, so $A^k = UD^kU^{-1}$ for all k. (More formal proof by induction.) Next:

$$\text{diag}(\lambda_1, \dots, \lambda_n)\,\text{diag}(\mu_1, \dots, \mu_n) = \text{diag}(\lambda_1\mu_1, \dots, \lambda_n\mu_n).$$

It follows that

$$D^k = \text{diag}(\lambda_1, \dots, \lambda_n)^k = \text{diag}(\lambda_1^k, \dots, \lambda_n^k).$$

(Again, a formal proof would go by induction on k.)

Diagonalisation example

We will diagonalise the matrix $A = \begin{bmatrix} 3 & 2 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & -3 \end{bmatrix}$ and thus find A^k.

As $A - tI_4$ is upper-triangular we see that the determinant is just the product of the diagonal terms. This gives

$$\chi_A(t) = \det(A - tI_4) = t^2(t-3)(t+3),$$

and it follows that the eigenvalues are $\lambda_1 = \lambda_2 = 0$ and $\lambda_3 = 3$ and $\lambda_4 = -3$. Consider the vectors

$$u_1 = \begin{bmatrix} 1 \\ 0 \\ -3 \\ 0 \end{bmatrix} \qquad u_2 = \begin{bmatrix} 2 \\ -3 \\ 0 \\ 0 \end{bmatrix} \qquad u_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad u_4 = \begin{bmatrix} 2 \\ -3 \\ -6 \\ -9 \end{bmatrix}$$

It is straightforward to check that $Au_1 = Au_2 = 0$ and $Au_3 = 3u_3$ and $Au_4 = -3u_4$, so the vectors u_i are eigenvectors for A, with eigenvalues 0, 0, 3 and -3 respectively.
(These vectors were found by row-reducing the matrices $A - \lambda_i I_4$.)

Diagonalisation example

$\lambda_1 = \lambda_2 = 0$ and $\lambda_3 = 3$ and $\lambda_4 = -3$; $Au_i = \lambda_i u_i$ where

$$u_1 = \begin{bmatrix} 1 \\ 0 \\ -3 \\ 0 \end{bmatrix} \qquad u_2 = \begin{bmatrix} 2 \\ -3 \\ 0 \\ 0 \end{bmatrix} \qquad u_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad u_4 = \begin{bmatrix} 2 \\ -3 \\ -6 \\ -9 \end{bmatrix}$$

Now put

$$U = [u_1|u_2|u_3|u_4] = \begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & -3 & 0 & -3 \\ -3 & 0 & 0 & -6 \\ 0 & 0 & 0 & -9 \end{bmatrix} \qquad V = \frac{1}{9}\begin{bmatrix} 0 & 0 & -3 & 2 \\ 0 & -3 & 0 & 1 \\ 9 & 6 & 3 & -2 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

One can check that $UV = I_4$, so $U^{-1} = V$.
(V was found by row-reduction $[I_4|U] \to [V|I_4]$.)
We now have $A = UDU^{-1} = UDV$, where

$$D = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}.$$

Diagonalisation example

$$A = UDU^{-1} = \frac{1}{9}\begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & -3 & 0 & -3 \\ -3 & 0 & 0 & -6 \\ 0 & 0 & 0 & -9 \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}\begin{bmatrix} 0 & 0 & -3 & 2 \\ 0 & -3 & 0 & 1 \\ 9 & 6 & 3 & -2 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

$$\begin{aligned} A^k = UD^kU^{-1} &= \frac{1}{9}\begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & -3 & 0 & -3 \\ -3 & 0 & 0 & -6 \\ 0 & 0 & 0 & -9 \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 3^k & 0 \\ 0 & 0 & 0 & (-3)^k \end{bmatrix}\begin{bmatrix} 0 & 0 & -3 & 2 \\ 0 & -3 & 0 & 1 \\ 9 & 6 & 3 & -2 \\ 0 & 0 & 0 & -1 \end{bmatrix} \\ &= 3^{k-2}\begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & -3 & 0 & -3 \\ -3 & 0 & 0 & -6 \\ 0 & 0 & 0 & -9 \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 9 & 6 & 3 & -2 \\ 0 & 0 & 0 & (-1)^{k+1} \end{bmatrix} \\ &= 3^{k-2}\begin{bmatrix} 9 & 6 & 3 & -2(1+(-1)^k) \\ 0 & 0 & 0 & 3(-1)^k \\ 0 & 0 & 0 & 6(-1)^k \\ 0 & 0 & 0 & 9(-1)^k \end{bmatrix} \end{aligned}$$

Diagonalisation example

We will diagonalise the matrix $A = \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 5 & 5 & 2 \\ 2 & 5 & 5 & 2 \\ 2 & 2 & 2 & 2 \end{bmatrix}$. Recall $\chi_A(t) = \det(B)$,

where

$$B = A - tI_4 = \begin{bmatrix} 2-t & 2 & 2 & 2 \\ 2 & 5-t & 5 & 2 \\ 2 & 5 & 5-t & 2 \\ 2 & 2 & 2 & 2-t \end{bmatrix}.$$

Method 12.9: row-reduce B and keep track of row operation factors.

$$\begin{bmatrix} 2-t & 2 & 2 & 2 \\ 2 & 5-t & 5 & 2 \\ 2 & 5 & 5-t & 2 \\ 2 & 2 & 2 & 2-t \end{bmatrix} \to \begin{bmatrix} 2-t & 2 & 2 & 2 \\ 2 & 5-t & 5 & 2 \\ 0 & t & -t & 0 \\ t & 0 & 0 & -t \end{bmatrix} \xrightarrow{1/t^2} \begin{bmatrix} 2-t & 2 & 2 & 2 \\ 2 & 5-t & 5 & 2 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \to$$

$$\begin{bmatrix} 0 & 0 & 4 & 4-t \\ 0 & 0 & 10-t & 4 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & -1 \end{bmatrix} \xrightarrow{-1} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 10-t & 4 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 4 & 4-t \end{bmatrix} \xrightarrow{-1} \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 10-t & 4 \\ 0 & 0 & 4 & 4-t \end{bmatrix}$$

- Subtract row 1 from row 4, and row 2 from row 3.
- Multiply rows 3 and 4 by $1/t$ (factor $1/t^2$)
- Subtract multiples of rows 3 and 4 from rows 1 and 2.
- Swap rows 1 and 4 (factor -1); Swap rows 2 and 3 (factor -1).

Diagonalisation example

$$B = A - tI_4 \to C = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 10-t & 4 \\ 0 & 0 & 4 & 4-t \end{bmatrix}; \text{ product of factors } \mu = 1/t^2$$

Expand down the columns to get

$$\det(C) = \det \begin{bmatrix} 10-t & 4 \\ 4 & 4-t \end{bmatrix} = (10-t)(4-t)-16 = t^2-14t+24 = (t-2)(t-12).$$

Thus $\chi_A(t) = \det(B) = \det(C)/\mu = (t-2)(t-12)t^2$.
This means that the eigenvalues of A are 2, 12 and 0.

Diagonalisation example

$$A = \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 5 & 5 & 2 \\ 2 & 5 & 5 & 2 \\ 2 & 2 & 2 & 2 \end{bmatrix}$$

$$B = A - tI_4 \to C = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 10-t & 4 \\ 0 & 0 & 4 & 4-t \end{bmatrix}; \text{ product of factors } \mu = 1/t^2$$

$$\chi_A(t) = \det(B) = \det(C)/\mu = (t-2)(t-12)t^2$$

Eigenvalues are 0, 2 and 12.

Diagonalisation example

$$B = A - tI_4 \to C = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 10-t & 4 \\ 0 & 0 & 4 & 4-t \end{bmatrix}; \text{ eigenvalues } 2, 12, 0.$$

To find an eigenvector of eigenvalue 2 we need to row-reduce the matrix $A - 2I_4$, which is just the matrix B with $t = 2$. We can therefore substitute $t = 2$ in C and then perform a few more steps to complete the row-reduction.

$$A - 2I_4 \to \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 8 & 4 \\ 0 & 0 & 4 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 1/2 \\ 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The eigenvector $u_1 = \begin{bmatrix} w & x & y & z \end{bmatrix}^T$ of eigenvalue 2 must therefore satisfy $w - z = x + z/2 = y + z/2 = 0$, so $u_1 = z \begin{bmatrix} 1 & -1/2 & -1/2 & 1 \end{bmatrix}^T$, with z arbitrary. It will be convenient to take $z = 2$ so $u_1 = \begin{bmatrix} 2 & -1 & -1 & 2 \end{bmatrix}^T$.

Diagonalisation example

$$B = A - tI_4 \to C = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 10-t & 4 \\ 0 & 0 & 4 & 4-t \end{bmatrix}; \text{ eigenvalues } 2, 12, 0.$$

To find an eigenvector of eigenvalue 12 we need to row-reduce the matrix $A - 12I_4$, which is just the matrix B with $t = 12$. We can therefore substitute $t = 12$ in C and then perform a few more steps to complete the row-reduction.

$$A - 12I_4 \to \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & -2 & 4 \\ 0 & 0 & 4 & -8 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The eigenvector $u_2 = \begin{bmatrix} w & x & y & z \end{bmatrix}^T$ of eigenvalue 12 must therefore satisfy $w - z = x - 2z = y - 2z = 0$, so $u_2 = z \begin{bmatrix} 1 & 2 & 2 & 1 \end{bmatrix}^T$, with z arbitrary. It will be convenient to take $z = 1$ so $u_2 = \begin{bmatrix} 1 & 2 & 2 & 1 \end{bmatrix}^T$.

Diagonalisation example

$$B = A - tI_4 \to C = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 10-t & 4 \\ 0 & 0 & 4 & 4-t \end{bmatrix}; \text{ eigenvalues } 2, 12, 0.$$

Finally, we need to find the eigenvectors of eigenvalue 0. Our reduction $B \to C$ involved division by t, so it is not valid in this case where $t = 0$. We must therefore start again and row-reduce the matrix $A - 0I_4 = A$ directly, but that is easy:

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

We conclude that the eigenvectors of eigenvalue 0 are the vectors $\begin{bmatrix} w & x & y & z \end{bmatrix}^T$ with $w + z = x + y = 0$. These form a two-dimensional space, and the vectors

$$u_3 = \begin{bmatrix} 1 & 0 & 0 & -1 \end{bmatrix}^T \qquad u_4 = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix}^T$$

form a basis.

Diagonalisation example

$$\begin{aligned} \lambda_1 &= 2 \\ \lambda_2 &= 12 \\ \lambda_3 &= 0 \\ \lambda_4 &= 0 \end{aligned} \qquad u_1 = \begin{bmatrix} 2 \\ -1 \\ -1 \\ 2 \end{bmatrix} \qquad u_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix} \qquad u_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix} \qquad u_4 = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}$$

Now put

$$U = [u_1|u_2|u_3|u_4] = \begin{bmatrix} 2 & 1 & 1 & 0 \\ -1 & 2 & 0 & 1 \\ -1 & 2 & 0 & -1 \\ 2 & 1 & -1 & 0 \end{bmatrix} \qquad D = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 12 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Row reduce $[U|I_4] \to [I_4|U^{-1}]$. Answer is

$$U^{-1} = \frac{1}{10} \begin{bmatrix} 2 & -1 & -1 & 2 \\ 1 & 2 & 2 & 1 \\ 5 & 0 & 0 & -5 \\ 0 & 5 & -5 & 0 \end{bmatrix}.$$

We now have a diagonalisation $A = UDU^{-1}$.

Lecture 11

Systems of differential equations

- If $\dot{x} = ax$ with $x = c$ at $t = 0$, then $x = c\,e^{at}$.
- If $\dot{x}_i = a_i x_i$ with $x_i = c_i$ at $t = 0$ (for $i = 1, 2, 3$), then $x_i = c_i\,e^{a_i t}$
- Put

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad c = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} \qquad D = \begin{bmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{bmatrix} \qquad e^{Dt} = \begin{bmatrix} e^{a_1 t} & 0 & 0 \\ 0 & e^{a_2 t} & 0 \\ 0 & 0 & e^{a_3 t} \end{bmatrix}$$

  The equations are $\dot{x} = Dx$ with $x = c$ at $t = 0$; the solution is $x = e^{Dt}c$.
- Suppose instead $x = c$ at $t = 0$ with

$$\begin{aligned} \dot{x}_1 &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ \dot{x}_2 &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ \dot{x}_3 &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{aligned} \qquad \text{so } \dot{x} = Ax \text{ where } A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

- To solve this, diagonalise $A = UDU^{-1}$ with $D = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$ say, so $\dot{x} = UDU^{-1}x$. Put $y = U^{-1}x$ and $d = U^{-1}c$ so $\dot{y} = U^{-1}\dot{x} = DU^{-1}x = Dy$, with $y = d$ at $t = 0$. This gives $y = e^{Dt}d$, where

$$e^{Dt} = \mathrm{diag}(e^{\lambda_1 t}, e^{\lambda_2 t}, e^{\lambda_3 t});$$

  so $x = Uy = Ue^{Dt}d = Ue^{Dt}U^{-1}c$.

Differential equations example

If $\dot{x} = Ax$, $x = c$ at $t = 0$, $A = UDU^{-1}$, then $x = Ue^{Dt}U^{-1}c$ where $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ and $e^{Dt} = \mathrm{diag}(e^{\lambda_1 t}, \dots, e^{\lambda_n t})$.

Example 15.1: Suppose $\dot{x}_1 = x_1 + x_2 + x_3$; $\dot{x}_2 = 2x_2 + 2x_3$; $\dot{x}_3 = 3x_3$ with $x_1 = x_2 = 0$ and $x_3 = 1$ at $t = 0$. This can be written as $\dot{x} = Ax$, where $A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$. By Example 14.6: $A = UDU^{-1}$, where

$$U = \begin{bmatrix} 1 & 1 & 3/2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \qquad D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \qquad U^{-1} = \begin{bmatrix} 1 & -1 & 1/2 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}.$$

So $x = Ue^{Dt}U^{-1}c$, where $c = \text{ initial value } = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$. Thus

$$\begin{aligned} x &= \begin{bmatrix} 1 & 1 & 3/2 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} e^t & 0 & 0 \\ 0 & e^{2t} & 0 \\ 0 & 0 & e^{3t} \end{bmatrix} \begin{bmatrix} 1 & -1 & 1/2 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} e^t & e^{2t} & \frac{3}{2}e^{3t} \\ 0 & e^{2t} & 2e^{3t} \\ 0 & 0 & e^{3t} \end{bmatrix} \begin{bmatrix} 1/2 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}e^t - 2e^{2t} + \frac{3}{2}e^{3t} \\ -2e^{2t} + 2e^{3t} \\ e^{3t} \end{bmatrix}. \end{aligned}$$

Differential equations example

Suppose $\dot{x} = \dot{y} = \dot{z} = x + y + z$ with $x = z = 0$ and $y = 1$ at $t = 0$.

Thus $\dot{v} = Av$, where $v = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ and $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$; $v = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ at $t = 0$

The characteristic polynomial is

$$\chi_A(t) = \det \begin{bmatrix} 1-t & 1 & 1 \\ 1 & 1-t & 1 \\ 1 & 1 & 1-t \end{bmatrix} = 3t^2 - t^3 = t^2(3-t).$$

Eigenvalues are $\lambda_1 = 0$, $\lambda_2 = 0$ and $\lambda_3 = 3$. If we put

$$u_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \qquad u_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \qquad u_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

we find that $Au_1 = Au_2 = 0$ and $Au_3 = 3u_3$. Thus, the vectors u_i form a basis for $\mathbb{R}^3$ consisting of eigenvectors for A. This means that we have a diagonalisation $A = UDU^{-1}$, where

$$U = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix} \qquad D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Differential equations example

$\dot{v} = UDU^{-1}v$ and $v = c$ at $t = 0$ where

$$U = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix} \qquad D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix} \qquad c = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

We can find U^{-1} by the following row-reduction:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ -1 & 1 & 1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 3 & 1 & 1 & 1 \\ 0 & -1 & 1 & 0 & 0 & 0 \end{array}\right] \to$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2/3 & -1/3 & -1/3 \\ 0 & 0 & 1 & 1/3 & 1/3 & 1/3 \\ 0 & -1 & 0 & -1/3 & -1/3 & 2/3 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2/3 & -1/3 & -1/3 \\ 0 & 1 & 0 & 1/3 & 1/3 & -2/3 \\ 0 & 0 & 1 & 1/3 & 1/3 & 1/3 \end{array}\right].$$

The conclusion is that

$$U^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ 1 & 1 & -2 \\ 1 & 1 & 1 \end{bmatrix}.$$

Differential equations example

$\dot{v} = UDU^{-1}v$ and $v = c$ at $t = 0$ where

$$U = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix} \quad D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{bmatrix} \quad U^{-1} = \frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ 1 & 1 & -2 \\ 1 & 1 & 1 \end{bmatrix} \quad c = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

The solution to our differential equation is now $v = Ue^{Dt}U^{-1}c$:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 1 \\ 0 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{3t} \end{bmatrix}\begin{bmatrix} 2 & -1 & -1 \\ 1 & 1 & -2 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$= \frac{1}{3}\begin{bmatrix} 1 & 0 & e^{3t} \\ -1 & 1 & e^{3t} \\ 0 & -1 & e^{3t} \end{bmatrix}\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} (e^{3t}-1)/3 \\ (e^{3t}+2)/3 \\ (e^{3t}-1)/3 \end{bmatrix}.$$

Solving difference equations

Problem: find a formula for the sequence where $a_0 = -1$, $a_1 = 0$, and $a_{i+2} = 6a_{i+1} - 8a_i$ for all $i \geq 0$.

$$a_2 = 6a_1 - 8a_0 = 6 \times 0 - 8 \times (-1) = 8$$
$$a_3 = 6a_2 - 8a_1 = 6 \times 8 - 8 \times 0 = 48$$
$$a_4 = 6a_3 - 8a_2 = 6 \times 48 - 8 \times 8 = 224 \qquad \text{etc.}$$

Vector formulation: put $v_i = \begin{bmatrix} a_i \\ a_{i+1} \end{bmatrix} \in \mathbb{R}^2$, so $v_0 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$ and

$$v_{n+1} = \begin{bmatrix} a_{n+1} \\ a_{n+2} \end{bmatrix} = \begin{bmatrix} a_{n+1} \\ 6a_{n+1} - 8a_n \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -8 & 6 \end{bmatrix}\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -8 & 6 \end{bmatrix} v_n.$$

We write $A = \begin{bmatrix} 0 & 1 \\ -8 & 6 \end{bmatrix}$, so the above reads $v_{n+1} = Av_n$. Thus $v_1 = Av_0$, $v_2 = Av_1 = A^2v_0$, $v_3 = Av_2 = A^3v_0$, $v_n = A^nv_0$.

We can be more explicit by finding the eigenvalues and eigenvectors of A.

Solving difference equations

$$v_n = \begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = A^nv_0 \qquad A = \begin{bmatrix} 0 & 1 \\ -8 & 6 \end{bmatrix} \qquad v_0 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

The characteristic polynomial is

$$\chi_A(t) = \det\begin{bmatrix} -t & 1 \\ -8 & 6-t \end{bmatrix} = t^2 - 6t + 8 = (t-2)(t-4),$$

so the eigenvectors are 2 and 4. Using the row-reductions

$$A - 2I = \begin{bmatrix} -2 & 1 \\ -8 & 4 \end{bmatrix} \to \begin{bmatrix} 1 & -1/2 \\ 0 & 0 \end{bmatrix} \qquad A - 4I = \begin{bmatrix} -4 & 1 \\ -8 & 2 \end{bmatrix} \to \begin{bmatrix} 1 & -1/4 \\ 0 & 0 \end{bmatrix}$$

we see that $u_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $u_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$ are eigenvectors of eigenvalues 2 and 4 (forming a basis for $\mathbb{R}^2$). Recall that $A^nu_1 = 2^nu_1$ and $A^nu_2 = 4^nu_2$. We can express $v_0 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$ in terms of this basis by row-reducing $[u_1|u_2|v_0]$:

$$\left[\begin{array}{cc|c} 1 & 1 & -1 \\ 2 & 4 & 0 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 1 & -1 \\ 0 & 2 & 2 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 1 & -1 \\ 0 & 1 & 1 \end{array}\right] \to \left[\begin{array}{cc|c} 1 & 0 & -2 \\ 0 & 1 & 1 \end{array}\right].$$

By reading off the last column, we deduce that $v_0 = -2u_1 + u_2$ (which could also have been obtained by inspection).

Solving difference equations

$$v_n = \begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = A^nv_0 \quad A = \begin{bmatrix} 0 & 1 \\ -8 & 6 \end{bmatrix} \quad v_0 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}. \quad u_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}. \quad u_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}.$$

$$v_0 = -2u_1 + u_2 \qquad v_n = A^nv_0 \qquad A^nu_1 = 2^nu_1, \quad A^nu_2 = 4^nu_2$$

It follows that

$$v_n = A^nv_0 = A^nu_2 - 2A^nu_1 = 4^nu_2 - 2 \times 2^nu_1$$
$$= 2^{2n}\begin{bmatrix} 1 \\ 4 \end{bmatrix} - 2^{n+1}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 2^{2n} - 2^{n+1} \\ 2^{2n+2} - 2^{n+2} \end{bmatrix}.$$

Moreover, v_n was defined to be $\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix}$, so a_n is the top entry in v_n, so we conclude that

$$a_n = 2^{2n} - 2^{n+1}.$$

We will check that this formula does indeed give the required properties:

$$a_0 = 2^0 - 2^1 = 1 - 2 = -1$$
$$a_1 = 2^2 - 2^2 = 0$$
$$6a_{i+1} - 8a_i = 6(2^{2i+2} - 2^{i+2}) - 8(2^{2i} - 2^{i+1}) = 24 \times 2^{2i} - 24 \times 2^i - 8 \times 2^{2i} + 16 \times 2^i$$
$$= 16 \times 2^{2i} - 8 \times 2^i = 2^{2i+4} - 2^{i+3} = 2^{2(i+2)} - 2^{(i+2)+1} = a_{i+2}.$$

Another difference equation

We will find the sequence satisfying $b_0 = 3$ and $b_1 = 6$ and $b_2 = 14$ and

$$b_{i+3} = 6b_i - 11b_{i+1} + 6b_{i+2}.$$

The vectors $v_i = \begin{bmatrix} b_i & b_{i+1} & b_{i+2} \end{bmatrix}^T$ satisfy $v_0 = \begin{bmatrix} 3 & 6 & 14 \end{bmatrix}^T$ and

$$v_{i+1} = \begin{bmatrix} b_{i+1} \\ b_{i+2} \\ b_{i+3} \end{bmatrix} = \begin{bmatrix} b_{i+1} \\ b_{i+2} \\ 6b_i - 11b_{i+1} + 6b_{i+2} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -11 & 6 \end{bmatrix} \begin{bmatrix} b_i \\ b_{i+1} \\ b_{i+2} \end{bmatrix} = Bv_i.$$

It follows that $v_n = B^n v_0$ for all n, and b_n is the top entry in the vector v_n. Now write v_0 in terms of the eigenvectors of B. The characteristic polynomial is

$$\chi_B(t) = \det \begin{bmatrix} -t & 1 & 0 \\ 0 & -t & 1 \\ 6 & -11 & 6-t \end{bmatrix} = -t \det \begin{bmatrix} -t & 1 \\ -11 & 6-t \end{bmatrix} - \det \begin{bmatrix} 0 & 1 \\ 6 & 6-t \end{bmatrix}$$
$$= -t(t^2 - 6t + 11) - (-6) = 6 - 11t + 6t^2 - t^3 = (1-t)(2-t)(3-t),$$

so the eigenvalues are 1, 2 and 3.

Another difference equation

$$v_n = \begin{bmatrix} b_n \\ b_{n+1} \\ b_{n+2} \end{bmatrix} = B^n \begin{bmatrix} 3 \\ 6 \\ 14 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 6 & -11 & 6 \end{bmatrix} \text{ has eigenvalues } 1, 2, 3.$$

Now find the eigenvectors:

$$B - I = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 6 & -11 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \qquad u_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$B - 2I = \begin{bmatrix} -2 & 1 & 0 \\ 0 & -2 & 1 \\ 6 & -11 & 4 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1/4 \\ 0 & 1 & -1/2 \\ 0 & 0 & 0 \end{bmatrix} \qquad u_2 = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

$$B - 3I = \begin{bmatrix} -3 & 1 & 0 \\ 0 & -3 & 1 \\ 6 & -11 & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1/9 \\ 0 & 1 & -1/3 \\ 0 & 0 & 0 \end{bmatrix} \qquad u_3 = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix}.$$

Another difference equation

$$v_n = B^n \begin{bmatrix} 3 \\ 6 \\ 14 \end{bmatrix} \qquad u_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad u_2 = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} \quad u_3 = \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix} \qquad Bu_k = ku_k.$$

By inspection: $v_0 = \begin{bmatrix} 3 \\ 6 \\ 14 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} + \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix} = u_1 + u_2 + u_3.$

This could also have been obtained by row-reducing $[u_1|u_2|u_3|v_0]$:

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 1 & 2 & 3 & 6 \\ 1 & 4 & 9 & 14 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 1 & 1 & 3 \\ 0 & 1 & 2 & 3 \\ 0 & 3 & 8 & 11 \end{array}\right] \to \cdots \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right].$$

As u_k is an eigenvector of eigenvalue k, we have $B^n u_k = k^n u_k$, so

$$v_n = B^n v_0 = B^n u_1 + B^n u_2 + B^n u_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + 2^n \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} + 3^n \begin{bmatrix} 1 \\ 3 \\ 9 \end{bmatrix} = \begin{bmatrix} 1 + 2^n + 3^n \\ 1 + 2^{n+1} + 3^{n+1} \\ 1 + 2^{n+2} + 3^{n+2} \end{bmatrix}.$$

Moreover, b_n is the top entry in v_n, so we conclude that

$$b_n = 1 + 2^n + 3^n.$$

Fibonacci numbers

The Fibonacci numbers are given by $F_0 = 0$ and $F_1 = 1$ and $F_{n+2} = F_n + F_{n+1}$.
The vectors $v_i = \begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}$ therefore satisfy $v_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and

$$v_{n+1} = \begin{bmatrix} F_{n+1} \\ F_{n+2} \end{bmatrix} = \begin{bmatrix} F_{n+1} \\ F_n + F_{n+1} \end{bmatrix} = Av_n, \qquad \text{where} \qquad A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

It follows that $v_n = A^n v_0$. We have $\chi_A(t) = t^2 - t - 1$, which has roots $\lambda_1 = (1 + \sqrt{5})/2$ and $\lambda_2 = (1 - \sqrt{5})/2$. To find an eigenvector of eigenvalue λ_1, we must solve

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda_1 \begin{bmatrix} x \\ y \end{bmatrix} \qquad \text{or} \qquad \begin{aligned} y &= \lambda_1 x \\ x + y &= \lambda_1 y \end{aligned}$$

Substituting $y = \lambda_1 x$ in $x + y = \lambda_1 y$ gives $x + \lambda_1 x = \lambda_1^2 x$, or $(\lambda_1^2 - \lambda_1 - 1)x = 0$, which is automatic, because λ_1 is a root of $t^2 - t - 1 = 0$.
Take $x = 1$ to get an eigenvector $u_1 = \begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix}$ of eigenvalue λ_1.

Similarly, we have an eigenvector $u_2 = \begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix}$ of eigenvalue λ_2.

Fibonacci numbers

$$v_n = A^n \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad u_k = \begin{bmatrix} 1 \\ \lambda_k \end{bmatrix} \qquad Au_k = \lambda_k u_k \qquad \begin{aligned} \lambda_1 &= (1+\sqrt{5})/2 \\ \lambda_2 &= (1-\sqrt{5})/2 \end{aligned}$$

We now need to find α and β such that $\alpha u_1 + \beta u_2 = v_0$, or equivalently

$$\alpha \begin{bmatrix} 1 \\ \lambda_1 \end{bmatrix} + \beta \begin{bmatrix} 1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \qquad \text{or} \qquad \begin{aligned} \beta &= -\alpha \\ \alpha(\lambda_1 - \lambda_2) &= 1. \end{aligned}$$

Now $\lambda_1 - \lambda_2 = \sqrt{5}$ so $\alpha = 1/\sqrt{5}$ and $\beta = -1/\sqrt{5}$ and $v_0 = (u_1 - u_2)/\sqrt{5}$.

$$v_n = A^n v_0 = \frac{A^n u_1 - A^n u_2}{\sqrt{5}} = \frac{\lambda_1^n u_1 - \lambda_2^n u_2}{\sqrt{5}} = \frac{1}{\sqrt{5}} \begin{bmatrix} \lambda_1^n - \lambda_2^n \\ \lambda_1^{n+1} - \lambda_2^{n+1} \end{bmatrix}.$$

Moreover, F_n is the top entry in v_n, so we obtain the formula

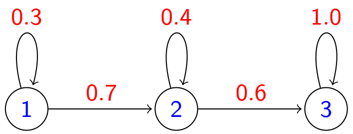
$$F_n = \frac{\lambda_1^n - \lambda_2^n}{\sqrt{5}} = \frac{(1+\sqrt{5})^n - (1-\sqrt{5})^n}{2^n\,\sqrt{5}}.$$

It is also useful to note here that $\lambda_1 \simeq 1.618033988$ and $\lambda_2 \simeq -0.6180339880$. As $|\lambda_1| > 1$ and $|\lambda_2| < 1$ we see that $|\lambda_1^n| \to \infty$ and $|\lambda_2^n| \to 0$ as $n \to \infty$. When n is large we can neglect the λ_2 term and we have $F_n \simeq \lambda_1^n/\sqrt{5}$.

Lecture 12

Markov chains

Consider a system that can be in three different states.



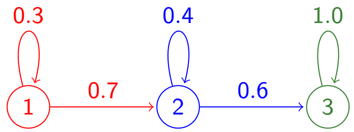
Once per second, it can change state in a random way. If it is in state 1, it jumps to state 2 with probability 0.7 and stays in state 1 with probability 0.3. If it is in state 2, it jumps to state 3 with probability 0.6 and stays in state 1 with probability 0.4. If it is in state 3, it stays there (with probability 1).

This is an example of a *Markov chain*. These are widely used to model (pseudo)-random processes in economics, population biology, information technology and other areas. Some questions about a Markov chain:

- How much time to we spend in state i, on average?
- If we start in state i, what is the average wait before reaching j?
- If we start in state i, what is the probability of reaching j before k?

We will take the first steps towards answering such questions.

Markov chains



Notation: $p_{j\leftarrow i}$ is the probability of jumping from state i to state j.
The *transition matrix* has $p_{j\leftarrow i}$ in the i'th column of the j'th row.

$$P = \begin{bmatrix} p_{1\leftarrow 1} & p_{1\leftarrow 2} & p_{1\leftarrow 3} \\ p_{2\leftarrow 1} & p_{2\leftarrow 2} & p_{2\leftarrow 3} \\ p_{3\leftarrow 1} & p_{3\leftarrow 2} & p_{3\leftarrow 3} \end{bmatrix} = \begin{bmatrix} 0.3 & 0.0 & 0.0 \\ 0.7 & 0.4 & 0.0 \\ 0.0 & 0.6 & 1.0 \end{bmatrix}.$$

All entries are probabilities so they lie between 0 and 1.
The entries in column 1 are the probabilities of all possible steps when we start in state 1, so they must add up to 1.
Similarly, each column has nonnegative entries adding up to 1, in other words it is a *probability vector*. By definition, this means that P is a *stochastic matrix*.

Distribution vectors

Suppose that the probability of being in state i (at a certain time) is q_i. Let q'_j be the probability of being in state j one second later. Then $q'_j = \sum_i p_{j\leftarrow i} q_i$.

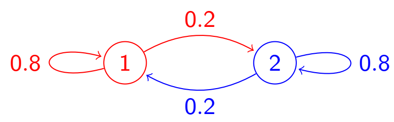
In terms of distribution vectors $q = \begin{bmatrix} q_1 & \cdots & q_n \end{bmatrix}^T$ and $q' = \begin{bmatrix} q'_1 & \cdots & q'_n \end{bmatrix}^T$ this says that $q' = Pq$. For example, when there are three states we have

$$q' = \begin{bmatrix} q'_1 \\ q'_2 \\ q'_3 \end{bmatrix} = \begin{bmatrix} p_{1\leftarrow 1}q_1 + p_{1\leftarrow 2}q_2 + p_{1\leftarrow 3}q_3 \\ p_{2\leftarrow 1}q_1 + p_{2\leftarrow 2}q_2 + p_{2\leftarrow 3}q_3 \\ p_{3\leftarrow 1}q_1 + p_{3\leftarrow 2}q_2 + p_{3\leftarrow 3}q_3 \end{bmatrix} = \begin{bmatrix} p_{1\leftarrow 1} & p_{1\leftarrow 2} & p_{1\leftarrow 3} \\ p_{2\leftarrow 1} & p_{2\leftarrow 2} & p_{2\leftarrow 3} \\ p_{3\leftarrow 1} & p_{3\leftarrow 2} & p_{3\leftarrow 3} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = Pq.$$

Thus, if r_t is the distribution vector at time t we have $r_t = P^t r_0$. This can be calculated using the eigenvalues and eigenvectors of P.

Markov chain example

Consider a two-state Markov chain which stays in the same state with probability 0.8, and flips to the other state with probability 0.2.



$$P = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$$

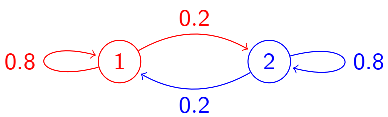
The characteristic polynomial is $\chi_P(t) = t^2 - 1.6t + 0.6$ so the eigenvalues are $(1.6 \pm \sqrt{2.56 - 4 \times 0.6})/2$, which works out as $\lambda_1 = 0.6$ and $\lambda_2 = 1$.

Corresponding eigenvectors: $u_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $u_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Now $P = UDU^{-1}$ and so $P^n = UD^nU^{-1}$, where

$$U = [u_1|u_2] = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \qquad D = \begin{bmatrix} 0.6 & 0 \\ 0 & 1 \end{bmatrix} \qquad U^{-1} = \begin{bmatrix} 0.5 & -0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

$$P^n = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} (0.6)^n & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.5 & -0.5 \\ 0.5 & 0.5 \end{bmatrix} = \begin{bmatrix} 0.5(1+0.6^n) & 0.5(1-0.6^n) \\ 0.5(1-0.6^n) & 0.5(1-0.6^n) \end{bmatrix}$$

Markov chain example



$$P^n = \begin{bmatrix} 0.5(1+0.6^n) & 0.5(1-0.6^n) \\ 0.5(1-0.6^n) & 0.5(1-0.6^n) \end{bmatrix}$$

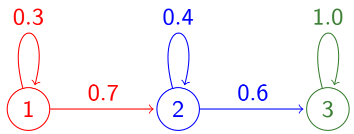
Suppose we are given that the system starts at $t = 0$ in state 1, so $r_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. It follows that

$$r_n = P^n r_0 = \begin{bmatrix} 0.5(1+0.6^n) & 0.5(1-0.6^n) \\ 0.5(1-0.6^n) & 0.5(1-0.6^n) \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.5(1+0.6^n) \\ 0.5(1-0.6^n) \end{bmatrix}.$$

Thus, at time n the probability of being in state 1 is $0.5(1+0.6^n)$, and the probability of being in state 2 is $0.5(1-0.6^n)$.

When n is large, we observe that $(0.6)^n$ will be very small, so $r_n \simeq \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$, so it is almost equally probable that X will be in either of the two states. This should be intuitively plausible, given the symmetry of the situation.

Markov chain example



$$P = \begin{bmatrix} 0.3 & 0.0 & 0.0 \\ 0.7 & 0.4 & 0.0 \\ 0.0 & 0.6 & 1.0 \end{bmatrix}.$$

We start in state 1 at $t = 0$. What is the probability that we are in state 3 at $t = 5$? We are given $r_0 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$ and we need to find $r_5 = P^5 r_0$.

$$\chi_P(t) = \det \begin{bmatrix} 0.3-t & 0.0 & 0.0 \\ 0.7 & 0.4-t & 0.0 \\ 0.0 & 0.6 & 1.0-t \end{bmatrix} = (0.3-t)(0.4-t)(1-t),$$

so the eigenvalues are 0.3, 0.4 and 1.
To find an eigenvector of eigenvalue 0.3, we row-reduce the matrix $P - 0.3I$:

$$\begin{bmatrix} 0 & 0 & 0 \\ 7/10 & 1/10 & 0 \\ 0 & 6/10 & 7/10 \end{bmatrix} \to \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1/7 & 0 \\ 0 & 1 & 7/6 \end{bmatrix} \to \begin{bmatrix} 1 & 1/7 & 0 \\ 0 & 1 & 7/6 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1/6 \\ 0 & 1 & 7/6 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus take $u_1 = \begin{bmatrix} 1 & -7 & 6 \end{bmatrix}^T$ as an eigenvector of eigenvalue 0.3.
Eigenvectors u_2 and u_3 can be found similarly.

Markov chain example

$$P = \begin{bmatrix} 0.3 & 0.0 & 0.0 \\ 0.7 & 0.4 & 0.0 \\ 0.0 & 0.6 & 1.0 \end{bmatrix} \qquad u_1 = \begin{bmatrix} 1 \\ -7 \\ 6 \end{bmatrix} \qquad u_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \qquad u_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\lambda_1 = 0.3 \qquad \lambda_2 = 0.4 \qquad \lambda_3 = 1$$

We have $P = UDU^{-1}$ where

$$D = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3) = \begin{bmatrix} 0.3 & 0 & 0 \\ 0 & 0.4 & 0 \\ 0 & 0 & 1.0 \end{bmatrix} \qquad U = [u_1|u_2|u_3] = \begin{bmatrix} 1 & 0 & 0 \\ -7 & 1 & 0 \\ 6 & -1 & 1 \end{bmatrix}$$

Now find U^{-1} by row-reducing $[U|I_3]$:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ -7 & 1 & 0 & 0 & 1 & 0 \\ 6 & -1 & 1 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 7 & 1 & 0 \\ 0 & -1 & 1 & -6 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 7 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{array}\right]$$

$$\begin{aligned} P^k = UD^kU^{-1} &= \begin{bmatrix} 1 & 0 & 0 \\ -7 & 1 & 0 \\ 6 & -1 & 1 \end{bmatrix} \begin{bmatrix} (0.3)^k & 0 & 0 \\ 0 & (0.4)^k & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 7 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} (0.3)^k & 0 & 0 \\ 7(0.4)^k - 7(0.3)^k & (0.4)^k & 0 \\ 1 + 6(0.3)^k - 7(0.4)^k & 1 - (0.4)^k & 1 \end{bmatrix}. \end{aligned}$$

Markov chain example

$$P^k = \begin{bmatrix} (0.3)^k & 0 & 0 \\ 7(0.4)^k - 7(0.3)^k & (0.4)^k & 0 \\ 1 + 6(0.3)^k - 7(0.4)^k & 1 - (0.4)^k & 1 \end{bmatrix}$$

We are definitely in state 1 at $t = 0$, so $r_0 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$. It follows that

$$\begin{aligned} r_k = P^k r_0 &= \begin{bmatrix} (0.3)^k & 0 & 0 \\ 7(0.4)^k - 7(0.3)^k & (0.4)^k & 0 \\ 1 + 6(0.3)^k - 7(0.4)^k & 1 - (0.4)^k & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} (0.3)^k \\ 7(0.4)^k - 7(0.3)^k \\ 1 + 6(0.3)^k - 7(0.4)^k \end{bmatrix}. \end{aligned}$$

For the probability p that X is in state 3 at time $t = 5$, we need to take $k = 5$ and look at the third component, giving

$$p = 6(0.3)^5 - 7(0.4)^5 + 1 \simeq 0.94290.$$

Stochastic matrices have eigenvalue 1

In both of the last two examples, one of the eigenvalues of the transition matrix P was equal to one. This was not a coincidence.

Proposition 17.10: If P is an $n \times n$ stochastic matrix, then 1 is an eigenvalue of P.

We will prove this after two lemmas.

A and A^T have the same eigenvalues

Lemma: Let B be an $n \times n$ matrix.
Then 0 is an eigenvalue of B iff 0 is an eigenvalue of B^T.

Proof: We can divide B and B^T into columns, say

$$B = \left[\begin{array}{c|c|c} v_1 & \cdots & v_n \end{array}\right] \qquad B^T = \left[\begin{array}{c|c|c} w_1 & \cdots & w_n \end{array}\right]$$

Now 0 is an eigenvalue of B
iff $\exists \alpha \neq 0$ with $B\alpha = 0$ or $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$
iff the columns v_i are linearly dependent
iff the v_i are not a basis (using the fact that there are n columns)
iff the w_j are not a basis (by duality)
iff the w_j are linearly dependent (using the fact that there are n columns)
iff $\exists \beta \neq 0$ with $\beta_1 w_1 + \cdots + \beta_n w_n = 0$ or $B^T\beta = 0$
iff 0 is an eigenvalue of B^T. □

A and A^T have the same eigenvalues

Corollary: For any $n \times n$ matrix A, the eigenvalues of A are the same as the eigenvalues of A^T.

Proof.
Let λ be an eigenvalue of A, so there is a nonzero vector u with $Au = \lambda u$.
This means that $(A - \lambda I_n)u = 0$, so 0 is an eigenvalue of $A - \lambda I_n$.
The lemma then tells us that 0 is also an eigenvalue of $(A - \lambda I_n)^T = A^T - \lambda I_n$.
This means that there is a nonzero vector v with $(A^T - \lambda I_n)v = 0$, or equivalently $A^T v = \lambda v$.
This proves that λ is also an eigenvalue of A^T.

The whole argument can be reversed to prove the converse as well: if λ is an eigenvalue of A^T, then it is also an eigenvalue of A. □

Stochastic matrices have eigenvalue 1

Corollary: For any $n \times n$ matrix A, the eigenvalues of A are the same as the eigenvalues of A^T.

Proposition 17.10: If P is an $n \times n$ stochastic matrix, then 1 is an eigenvalue of P.

Proof.
Let the columns of P be $v_1, \dots, v_n$.
Put $d = \begin{bmatrix} 1 & 1 & \cdots & 1 & 1 \end{bmatrix}^T \in \mathbb{R}^n$.
Because P is stochastic we know that the sum of the entries in v_i is one, or in other words that $v_i.d = 1$. This means that

$$P^T d = \begin{bmatrix} v_1^T \\ \hline \vdots \\ \hline v_n^T \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} v_1.d \\ \vdots \\ v_n.d \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} = d.$$

Thus, d is an eigenvector of P^T with eigenvalue 1.
It follows by the Corollary that 1 is also an eigenvalue of P, as required. □

Stationary distribution

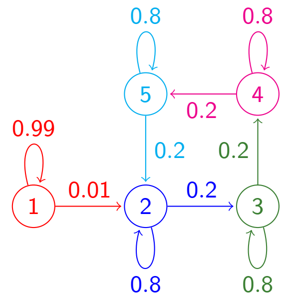
Definition 17.11: A *stationary distribution* for a Markov chain is a probability vector q that satisfies $Pq = q$ (so q is an eigenvector of eigenvalue 1).
Remark 17.12: It often happens that the distribution vectors r_n converge (as $n \to \infty$) to a distribution r_∞, whose i'th component is the long term average probability of the system being in state i. Because $Pr_n = r_{n+1}$ we then have

$$Pr_\infty = P \lim_{n\to\infty} r_n = \lim_{n\to\infty} Pr_n = \lim_{n\to\infty} r_{n+1} = r_\infty,$$

so r_∞ is a stationary distribution. Moreover, it often happens that there is only one stationary distribution. There are many theorems about this sort of thing, but we will not explore them in this course.

Stationary distribution example

We will use a heuristic argument to guess what the stationary distribution should be, and then give a rigorous proof that it is correct.
At each time there is a (small but) nonzero probability of leaving state 1 and entering the square, so if we wait long enough we can expect this to happen.
After we have entered the square there is no way we can ever return to state 1, so the long-run average probability of being in state 1 is zero.



Once we have entered the square things are symmetric so we spend $\frac{1}{4}$ of the time in each of states $2, \dots, 5$. Thus $q = \begin{bmatrix} 0 & 0.25 & 0.25 & 0.25 & 0.25 \end{bmatrix}^T$ should be a stationary distribution. It is a probability vector and

$$Pq = \begin{bmatrix} 0.99 & 0 & 0 & 0 & 0 \\ 0.01 & 0.8 & 0 & 0 & 0.2 \\ 0 & 0.2 & 0.8 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 & 0 \\ 0 & 0 & 0 & 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} = q \text{ as required.}$$

Lecture 13

Page rank

PageRank

Google assigns to each web page a number called the PageRank, calculated using eigenvectors; pages with higher rank come higher in search results. We will describe a simplified version.

- Imagine pages $S_1, \dots, S_n$, with some links between them.
- Say S_j links to N_j different pages, and assume $N_j > 0$.
- We want rankings $r_i \geq 0$, normalised so that $\sum_i r_i = 1$; so r is a probability vector in $\mathbb{R}^n$.
- A link from S_j to S_i is a vote by S_j that S_i is important.
- Links from important pages should count for more; links from pages with many links should count for less.
- We use this rule: a link from S_j to S_i contributes r_j/N_j to r_i.
- Thus, the following consistency condition should be satisfied:

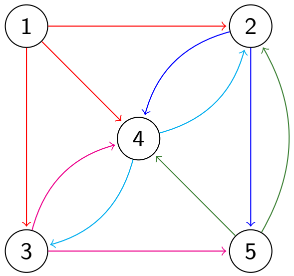
$$r_i = \sum_{\text{pages } S_j \text{ that link to } S_i} r_j/N_j.$$

PageRank as an eigenvector

Pages $S_1, \dots, S_n$; rankings $r_i \geq 0$ with $\sum_i r_i = 1$; S_j links to N_j pages;
Consistency condition $r_i = \sum_{\text{pages } S_j \text{ that link to } S_i} r_j/N_j$.

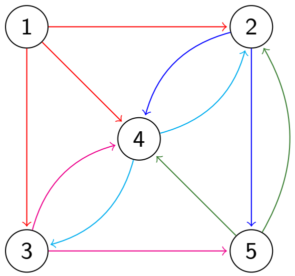
Define matrix P by $P_{ij} = \begin{cases} 1/N_j & \text{if there is a link from } S_j \text{ to } S_i \\ 0 & \text{otherwise.} \end{cases}$

Consistency condition is $r_i = \sum_j P_{ij} r_j$, so $r = Pr$, so r is an eigenvector for P with eigenvalue 1. Column j has N_j entries of $1/N_j$ so P is stochastic.



$$P = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 & 1/2 \\ 1/3 & 0 & 0 & 1/2 & 0 \\ 1/3 & 1/2 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 & 0 \end{bmatrix}.$$

PageRank as a Markov chain



$$P = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 & 1/2 \\ 1/3 & 0 & 0 & 1/2 & 0 \\ 1/3 & 1/2 & 1/2 & 0 & 1/2 \\ 0 & 1/2 & 1/2 & 0 & 0 \end{bmatrix}.$$

Imagine a surfer who clicks a randomly chosen link on the current page once per minute. This gives a Markov chain X with transition matrix P.
The PageRank vector r must satisfy $r_i \geq 0$ and $\sum_i r_i = 1$ and $Pr = r$, so it is a stationary distribution for X.
Take $q = \begin{bmatrix} 1/n & \cdots & 1/n \end{bmatrix}^T$ (distribution for a uniformly random page).
Typically there is a unique stationary distribution r, and $P^k q$ converges quickly to r as $k \to \infty$. When n is millions or billions, this is the best way to find r.
Conceptually: r_i is the long run average proportion of time that a random surfer spends on page i.

Calculating PageRank in Maple

```
with(LinearAlgebra):
n := 5;
P := << 0 | 0 | 0 | 0 | 0 >,
 <1/3 | 0 | 0 | 1/2 | 1/2 >,
 <1/3 | 0 | 0 | 1/2 | 0 >,
 <1/3 | 1/2 | 1/2 | 0 | 1/2 >,
 < 0 | 1/2 | 1/2 | 0 | 0 >>;
NS := NullSpace(P - IdentityMatrix(n));
r := NS[1];
r := r / add(r[i],i=1..n);
r := evalf(r);
```

Result: $r = \begin{bmatrix} 0.0 \\ 0.2777777778 \\ 0.1666666667 \\ 0.3333333333 \\ 0.2222222222 \end{bmatrix}$; so

page 1 has rank 0.0
page 2 has rank 0.2777777778
page 3 has rank 0.1666666667
page 4 has rank 0.3333333333
page 5 has rank 0.2222222222

Calculating PageRank as a limit

```
with(LinearAlgebra):
n := 5;
P := << 0 | 0 | 0 | 0 | 0 >,
 <1/3 | 0 | 0 | 1/2 | 1/2 >,
 <1/3 | 0 | 0 | 1/2 | 0 >,
 <1/3 | 1/2 | 1/2 | 0 | 1/2 >,
 < 0 | 1/2 | 1/2 | 0 | 0 >>;
q := Vector(n,[1/n $ n]);
r := evalf(P^10 .  q);
```

Result: $r = \begin{bmatrix} 0.0 \\ 0.2783203125 \\ 0.1667317708 \\ 0.3332682292 \\ 0.2216796875 \end{bmatrix}$, close to the exact value of $\begin{bmatrix} 0.0 \\ 0.2777777778 \\ 0.1666666667 \\ 0.3333333333 \\ 0.2222222222 \end{bmatrix}$

q is a vector of length n, whose entries are $1/n$, repeated n times.
We have seen that $r = \lim_{k\to\infty} P^k q$, so $r = P^{10} q$ should be approximately right.

Damping

Google found it useful to modify the PageRank algorithm with a *damping factor* d, where $0 < d < 1$. Consider a surfer who clicks a random link on the current page with probability d, but with probability $1 - d$ chooses a uniformly random page (whether or not there is a link to it).
This gives a new transition matrix:

$$Q_{ij} = \begin{cases} \frac{d}{N_j} + \frac{1-d}{n} & \text{if there is a link from } S_j \text{ to } S_i \\ \frac{1-d}{n} & \text{otherwise.} \end{cases}$$

Equivalently: let R be the stochastic matrix with $R_{ij} = 1/n$ for all i and j; then $Q = dP + (1 - d)R$. Now the PageRank vector r should satisfy $(Q - I_n)r = 0$. We can approximate r by finding $Q^k q$ for large q.

```
d := 0.85;
R := Matrix(n,n,[1/n $ n^2]);
Q := d * P + (1-d) * R;
NS := NullSpace(Q - IdentityMatrix(n));
r := NS[1];
r := r / add(r[i],i=1..n);
```

or

```
r := Q^10 .  q;
```

Lecture 14

Subspaces

In $\mathbb{R}^2$ and $\mathbb{R}^3$, lines and planes are important, especially through the origin. We now discuss analogous structures in $\mathbb{R}^n$, where n may be bigger than 3.

Definition 19.1: A subset $V \subseteq \mathbb{R}^n$ is a *subspace* if

(a) The zero vector is an element of V.

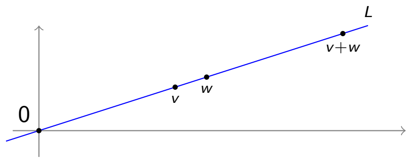
(b) Whenever v and w are two elements of V, the sum $v + w$ is also an element of V. (In other words, V is closed under addition.)

(c) Whenever v is an element of V and t is a real number, the vector tv is again an element of V. (In other words, V is closed under scalar multiplication.)

Subspace example

A subspace must contain 0, and be closed under addition and scalar multiplication.

Let L be the line in $\mathbb{R}^2$ with equation $y = x/\pi$.



- The point $0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^T$ lies on L.
- Suppose we have $v, w \in L$, so $v = \begin{bmatrix} a & a/\pi \end{bmatrix}^T$ and $w = \begin{bmatrix} b & b/\pi \end{bmatrix}^T$ for some numbers a and b. Then $v + w = \begin{bmatrix} a+b & (a+b)/\pi \end{bmatrix}^T$, which again lies on L. Thus, L is closed under addition.
- Suppose again that $v \in L$, so $v = \begin{bmatrix} a & a/\pi \end{bmatrix}^T$ for some a. Suppose also that $t \in \mathbb{R}$. Then $tv = \begin{bmatrix} ta & ta/\pi \end{bmatrix}^T$, which again lies on L, so L is closed under scalar multiplication.

So L is a subspace.

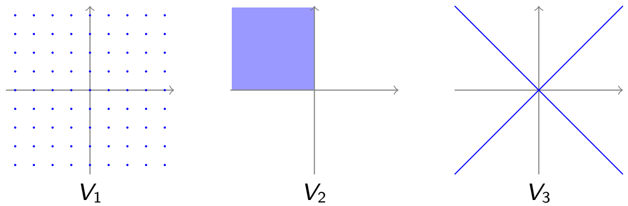
Subspace non-examples

Consider the following subsets of $\mathbb{R}^2$:

$$V_1 = \mathbb{Z}^2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \mid x \text{ and } y \text{ are integers} \right\}$$

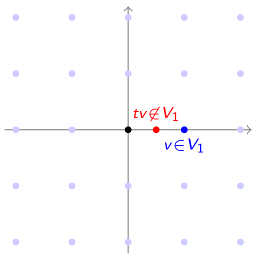
$$V_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \mid x \leq 0 \leq y \right\}$$

$$V_3 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \mid x^2 = y^2 \right\} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \mid x = \pm y \right\}.$$



None of these are subspaces.

V_1 is not a subspace

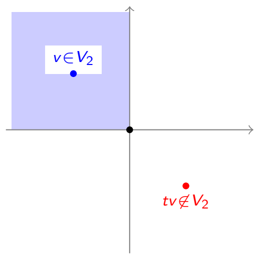


$$V_1 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \mid x \text{ and } y \text{ are integers} \right\}$$

It is clear that the zero vector has integer coordinates and so lies in V_1. Next, if v and w both have integer coordinates then so does $v + w$. In other words, if $v, w \in V_1$ then also $v + w \in V_1$, so V_1 is closed under addition. However, it is not closed under scalar multiplication. Indeed, if we take $v = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $t = 0.5$ then $v \in V_1$ and $t \in \mathbb{R}$ but the vector $tv = \begin{bmatrix} 0.5 \\ 0 \end{bmatrix}$ does not lie in V_1.

(This is generally the best way to prove that a set is not a subspace: provide a completely specific and explicit example where one of the conditions is not satisfied.)

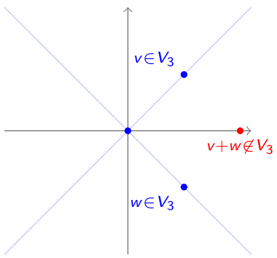
V_2 is not a subspace



$$V_2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \mid x \leq 0 \leq y \right\}$$

As $0 \leq 0 \leq 0$ we see that $0 \in V_2$.
Suppose we have vectors $v = \begin{bmatrix} x & y \end{bmatrix}^T$ and $v' = \begin{bmatrix} x' & y' \end{bmatrix}^T$ in V_2,
so $x \leq 0 \leq y$ and $x' \leq 0 \leq y'$. As $x, x' \leq 0$ it follows that $x + x' \leq 0$.
As $y, y' \geq 0$ it follows that $y + y' \geq 0$. This means that the sum
$v + v' = \begin{bmatrix} x + x' & y + y' \end{bmatrix}^T$ is again in V_2, so V_2 is closed under addition.
However, it is not closed under scalar multiplication.
Indeed, if we take $v = \begin{bmatrix} -1 & 1 \end{bmatrix}^T$ and $t = -1$ then $v \in V_2$ and $t \in \mathbb{R}$ but the vector $tv = \begin{bmatrix} 1 & -1 \end{bmatrix}^T$ does not lie in V_2.

V_3 is not a subspace



$$V_3 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \mid x^2 = y^2 \right\} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \mid x = \pm y \right\}.$$

It is again clear that $0 \in V_3$.
Now suppose we have $v = \begin{bmatrix} x & y \end{bmatrix}^T \in V_3$ (so $x^2 = y^2$) and $t \in \mathbb{R}$.
It follows that $(tx)^2 = t^2 x^2 = t^2 y^2 = (ty)^2$,
so the vector $tv = \begin{bmatrix} tx & ty \end{bmatrix}^T$ again lies in V_3.
This means that V_3 is closed under scalar multiplication.
However, it is not closed under addition,
because the vectors $v = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $w = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ lie in V_3,
but $v + w$ does not.

Two extreme cases

(a) The set $\{0\}$ (just consisting of the zero vector) is a subspace of $\mathbb{R}^n$.
(b) The whole set $\mathbb{R}^n$ is a subspace of itself.

Linear combinations in subspaces

Proposition 19.6: Let V be a subspace of $\mathbb{R}^n$. Then any linear combination of elements of V is again in V.

Proof.
Suppose we have elements $v_1, \dots, v_k \in V$, and suppose that w is a linear combination of the v_i, say $w = \sum_i \lambda_i v_i$ for some $\lambda_1, \dots, \lambda_k \in \mathbb{R}$. As $v_i \in V$ and $\lambda_i \in \mathbb{R}$ and V is closed under scalar multiplication we have $\lambda_i v_i \in V$. Now $\lambda_1 v_1$ and $\lambda_2 v_2$ are elements of V, and V is closed under addition, so $\lambda_1 v_1 + \lambda_2 v_2 \in V$. Next, as $\lambda_1 v_1 + \lambda_2 v_2 \in V$ and $\lambda_3 v_3 \in V$ and V is closed under addition we have $\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 \in V$. By extending this in the obvious way, we eventually conclude that the vector $w = \lambda_1 v_1 + \dots + \lambda_k v_k$ lies in V as claimed. $\square$

Dependent lists of length two

Lemma 8.5: Let v and w be vectors in $\mathbb{R}^n$, and suppose that $v \neq 0$ and that the list (v, w) is linearly dependent. Then there is a number α such that $w = \alpha v$.

Proof.
Because the list is dependent, there is a linear relation $\lambda v + \mu w = 0$ where λ and μ are not both zero. There are apparently three possibilities: (a) $\lambda \neq 0$ and $\mu \neq 0$; (b) $\lambda = 0$ and $\mu \neq 0$; (c) $\lambda \neq 0$ and $\mu = 0$. However, case (c) is not really possible. Indeed, in case (c) the equation $\lambda v + \mu w = 0$ would reduce to $\lambda v = 0$, and we could multiply by λ^{-1} to get $v = 0$; but $v \neq 0$ by assumption. In case (a) or (b) we can take $\alpha = -\lambda/\mu$ and we have $w = \alpha v$. $\square$

Subspaces of $\mathbb{R}^2$

Proposition 19.7: Let V be a subspace of $\mathbb{R}^2$. Then V is either $\{0\}$ or all of $\mathbb{R}^2$ or a straight line through the origin.

The proof will rely on two lemmas from last week.

Proposition 19.6: Let V be a subspace of $\mathbb{R}^n$. Then any linear combination of elements of V is again in V.

Lemma 8.5: Let v and w be vectors in $\mathbb{R}^n$, and suppose that $v \neq 0$ and that the list (v, w) is linearly dependent. Then there is a number α such that $w = \alpha v$.

Subspaces of $\mathbb{R}^2$

Proposition 19.7: Let V be a subspace of $\mathbb{R}^2$. Then V is either $\{0\}$ or all of $\mathbb{R}^2$ or a straight line through the origin.

Proof.

(a) If $V = \{0\}$ then there is nothing more to say.

(b) Suppose that V contains two vectors v and w such that the list (v, w) is linearly independent. As this is a linearly independent list of two vectors in $\mathbb{R}^2$, it must be a basis. Thus, every vector $x \in \mathbb{R}^2$ is a linear combination of v and w, and therefore lies in V by the last Proposition. Thus, we have $V = \mathbb{R}^2$.

(c) Suppose instead that neither (a) nor (b) holds. As (a) does not hold, we can choose a nonzero vector $v \in V$. Let L be the set of all scalar multiples of v, which is a straight line through the origin. As V is a subspace and $v \in V$ we know that every multiple of v lies in V, or in other words that $L \subseteq V$. Now let w be any vector in V. As (b) does not hold, the list (v, w) is linearly dependent, so the last Lemma tells us that w is a multiple of v and so lies in L. This shows that $V \subseteq L$, so $V = L$.

$\square$

Lecture 15

Spans and annihilators

Definition 19.8: Let $\mathcal{W} = (w_1, \dots, w_r)$ be a list of vectors in $\mathbb{R}^n$.

(a) $\mathrm{span}(\mathcal{W})$ is the set of all vectors $v \in \mathbb{R}^n$ that can be expressed as a linear combination of the list $\mathcal{W}$.

(b) $\mathrm{ann}(\mathcal{W})$ is the set of all vectors $u \in \mathbb{R}^n$ such that $u.w_1 = \cdots = u.w_r = 0$.

Remark 19.9: The terminology in (a) is related in an obvious way to the terminology used earlier: the list $\mathcal{W}$ spans $\mathbb{R}^n$ if and only if every vector in $\mathbb{R}^n$ is a linear combination of $\mathcal{W}$, or in other words $\mathrm{span}(\mathcal{W}) = \mathbb{R}^n$.

Span and annihilator example

$\mathrm{span}(w_1, \dots, w_r) = \{$ linear combinations of $w_1, \dots, w_r\}$;
$\mathrm{ann}(w_1, \dots, w_r) = \{v \mid v.w_1 = \cdots = v.w_r = 0\}$

Consider the plane P in $\mathbb{R}^3$ with equation $x + y + z = 0$. More formally:

$$P = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 \mid x + y + z = 0 \right\}.$$

If we put $v = \begin{bmatrix} x & y & z \end{bmatrix}^T$ and $t = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^T$, then we have $v.t = x + y + z$. It follows that

$$P = \{v \in \mathbb{R}^3 \mid v.t = 0\} = \mathrm{ann}(t).$$

On the other hand, if $x + y + z = 0$ then $z = -x - y$ so

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ -x-y \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

Thus, if we put $u_1 = \begin{bmatrix} 1 & 0 & -1 \end{bmatrix}^T$ and $u_2 = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix}^T$ then

$P = \{x\,u_1 + y\,u_2 \mid x, y \in \mathbb{R}\} = \{$ linear combinations of u_1 and $u_2\} = \mathrm{span}(u_1, u_2)$.

Span and annihilator example

Put

$$V = \{\begin{bmatrix} w & x & y & z \end{bmatrix}^T \in \mathbb{R}^4 \mid w + 2x + 3y + 4z = 4w + 3x + 2y + z = 0\}.$$

If we put $a = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}^T$ and $b = \begin{bmatrix} 4 & 3 & 2 & 1 \end{bmatrix}^T$ then

$$w+2x+3y+4z = a.\begin{bmatrix} w & x & y & z \end{bmatrix}^T \qquad 4w+3x+2y+z = b.\begin{bmatrix} w & x & y & z \end{bmatrix}^T$$

so we can describe V as $\mathrm{ann}(a, b)$.
On the other hand, suppose we have a vector $v = \begin{bmatrix} w & x & y & z \end{bmatrix}^T$ in V, so that

$$w + 2x + 3y + 4z = 0 \qquad \text{(A)}$$
$$4w + 3x + 2y + z = 0 \qquad \text{(B)}$$

If we subtract 4 times (A) from (B) and then divide by -15 we get equation (C) below. Similarly, if we subtract 4 times (B) from (A) and divide by -15 we get (D).

$$\tfrac{1}{3}x + \tfrac{2}{3}y + z = 0 \qquad \text{(C)}$$
$$w + \tfrac{2}{3}x + \tfrac{1}{3}y = 0 \qquad \text{(D)}$$

Span and annihilator example

$$
\begin{aligned}
V &= \{\begin{bmatrix} w & x & y & z \end{bmatrix}^T \in \mathbb{R}^4 \mid w + 2x + 3y + 4z = 4w + 3x + 2y + z = 0\} \\
&= \{\begin{bmatrix} w & x & y & z \end{bmatrix}^T \mid w = -\tfrac{2}{3}x - \tfrac{1}{3}y, \qquad z = -\tfrac{1}{3}x - \tfrac{2}{3}y\} \\
&= \left\{ \begin{bmatrix} -\frac{2}{3}x - \frac{1}{3}y \\ x \\ y \\ -\frac{1}{3}x - \frac{2}{3}y \end{bmatrix} \mid x, y \in \mathbb{R} \right\} = \left\{ x \begin{bmatrix} -2/3 \\ 1 \\ 0 \\ -1/3 \end{bmatrix} + y \begin{bmatrix} -1/3 \\ 0 \\ 1 \\ -2/3 \end{bmatrix} \mid x, y \in \mathbb{R} \right\}
\end{aligned}
$$

Thus, if we put

$$
c = \begin{bmatrix} -\frac{2}{3} \\ 1 \\ 0 \\ -\frac{1}{3} \end{bmatrix} \qquad d = \begin{bmatrix} -\frac{1}{3} \\ 0 \\ 1 \\ -\frac{2}{3} \end{bmatrix}
$$

then

$$
V = \{xc + yd \mid x, y \in \mathbb{R}\} = \operatorname{span}(c, d).
$$

Annihilators are subspaces

A subspace must contain 0, and be closed under addition and scalar multiplication.

Proposition 19.23: For any list $\mathcal{W} = (w_1, \dots, w_r)$ of vectors in $\mathbb{R}^n$, the set

$$
\operatorname{ann}(\mathcal{W}) = \{x \in \mathbb{R}^n \mid x.w_1 = \dots = x.w_r = 0\}
$$

is a subspace of $\mathbb{R}^n$.

Proof.

(a) The zero vector clearly has $0.w_i = 0$ for all i, so $0 \in \operatorname{ann}(\mathcal{W})$.

(b) Suppose that $u, v \in \operatorname{ann}(\mathcal{W})$. This means that $u.w_i = 0$ for all i, and that $v.w_i = 0$ for all i. It follows that $(u + v).w_i = u.w_i + v.w_i = 0 + 0 = 0$ for all i, so $u + v \in \operatorname{ann}(\mathcal{W})$. Thus, $\operatorname{ann}(\mathcal{W})$ is closed under addition.

(c) Suppose instead that $u \in \operatorname{ann}(\mathcal{W})$ and $t \in \mathbb{R}$. As before, we have $u.w_i = 0$ for all i. It follows that $(tu).w_i = t(u.w_i) = 0$ for all i, so $tu \in \operatorname{ann}(\mathcal{W})$. Thus, $\operatorname{ann}(\mathcal{W})$ is closed under scalar multiplication.

$\square$

Spans are subspaces

A subspace must contain 0, and be closed under addition and scalar multiplication.

Proposition 19.24: For any list $\mathcal{W} = (w_1, \dots, w_r)$ of vectors in $\mathbb{R}^n$, the set $\operatorname{span}(\mathcal{W})$
(of linear combinations of $\mathcal{W}$) is a subspace of $\mathbb{R}^n$.

Proof.

(a) The zero vector can be written as a linear combination $0 = 0w_1 + \dots + 0w_r$, so $0 \in \operatorname{span}(\mathcal{W})$.

(b) Suppose that $u, v \in \operatorname{span}(\mathcal{W})$. This means that for some sequence of coefficients $\lambda_i \in \mathbb{R}$ we have $u = \sum_i \lambda_i w_i$, and for some sequence of coefficients μ_i we have $v = \sum_i \mu_i w_i$. If we put $\nu_i = \lambda_i + \mu_i$ we then have $u + v = \sum_i \nu_i w_i$. This expresses $u + v$ as a linear combination of $\mathcal{W}$, so $u + v \in \operatorname{span}(\mathcal{W})$. Thus, $\operatorname{span}(\mathcal{W})$ is closed under addition.

(c) Suppose instead that $u \in \operatorname{span}(\mathcal{W})$ and $t \in \mathbb{R}$. As before, we have $u = \sum_i \lambda_i w_i$ for some sequence of coefficients λ_i. If we put $\kappa_i = t\lambda_i$ we find that $tu = \sum_i \kappa_i w_i$, which expresses tu as a linear combination of $\mathcal{W}$, so $tu \in \operatorname{span}(\mathcal{W})$. Thus, $\operatorname{span}(\mathcal{W})$ is closed under scalar multiplication.

$\square$

Bases for subspaces

Definition 20.1: Let V be a subspace of $\mathbb{R}^n$. A *basis* for V is a linearly independent list $\mathcal{V} = (v_1, \dots, v_r)$ of vectors in $\mathbb{R}^n$ such that $\operatorname{span}(\mathcal{V}) = V$.
Definition 20.2: Let V be a subspace of $\mathbb{R}^n$. The *dimension* of V (written $\dim(V)$) is the maximum possible length of any linearly independent list in $\mathcal{V}$.
The empty list always counts as linearly independent, so $\dim(V) \geq 0$.
Any linearly independent list in $\mathbb{R}^n$ has length at most n, so $\dim(V) \leq n$.
Proposition 20.3: Let V be a subspace of $\mathbb{R}^n$, and put $d = \dim(V)$.
Then any linearly independent list of length d in V is automatically a basis.
In particular, V has a basis.

Independent lists of the right length are bases

Proposition: Let V be a subspace of $\mathbb{R}^n$, and put $d = \dim(V)$. Then any linearly independent list $\mathcal{V} = (v_1, \cdots, v_d)$ of length d in V is a basis.

Proof.
Let u be an arbitrary vector in V. Consider the list $\mathcal{V}' = (v_1, \dots, v_d, u)$. This is a list in V of length $d + 1$, but d is the maximum possible length for any linearly independent list in V, so the list $\mathcal{V}'$ must be dependent. This means that there is a nontrivial relation

$$\lambda_1 v_1 + \cdots + \lambda_d v_d + \mu u = 0.$$

We claim that μ cannot be zero. Indeed, if $\mu = 0$ then the relation would become $\lambda_1 v_1 + \cdots + \lambda_d v_d = 0$, but $\mathcal{V}$ is linearly independent so this would give $\lambda_1 = \cdots = \lambda_d = 0$ as well as $\mu = 0$, so the original relation would be trivial, contrary to our assumption. Thus $\mu \neq 0$, so the relation can be rearranged as

$$u = -\frac{\lambda_1}{\mu} v_1 - \cdots - \frac{\lambda_d}{\mu} v_d,$$

which expresses u as a linear combination of $\mathcal{V}$. This shows that an arbitrary vector $u \in V$ can be expressed as a linear combination of $\mathcal{V}$, or in other words $V = \text{span}(\mathcal{V})$. As $\mathcal{V}$ is also linearly independent, it is a basis for V. □

Any d-dimensional subspace is $\mathbb{R}^d$ in disguise

Proposition 20.4: Let V be a subspace of $\mathbb{R}^n$, and let $\mathcal{V} = (v_1, \dots, v_d)$ be a basis for V.
Define a function $\phi\, \mathbb{R}^d \to V$ by $\phi(\lambda) = \lambda_1 v_1 + \cdots + \lambda_d v_d$.
Then there is an inverse function $\psi\, V \to \mathbb{R}^d$ with $\phi(\psi(v)) = v$ for all $v \in V$, and $\psi(\phi(\lambda)) = \lambda$ for all $\lambda \in \mathbb{R}^d$. Moreover, both ϕ and ψ respect addition and scalar multiplication:

$$\phi(\lambda + \mu) = \phi(\lambda) + \phi(\mu) \qquad \phi(t\lambda) = t\phi(\lambda)$$
$$\psi(v + w) = \psi(v) + \psi(w) \qquad \psi(tv) = t\psi(v).$$

Proof.
By assumption the list $\mathcal{V}$ is linearly independent and $\text{span}(\mathcal{V}) = V$. Consider an arbitrary vector $u \in V$. As $u \in \text{span}(\mathcal{V})$ we can write u as a linear combination $u = \sum_i \lambda_i v_i$ say, which means that $u = \phi(\lambda)$ for some λ. We claim that this λ is unique. Indeed, if we also have $u = \phi(\mu) = \sum_i \mu_i v_i$ then we can subtract to get $\sum_i (\lambda_i - \mu_i) v_i = 0$. This is a linear relation on the list $\mathcal{V}$, but $\mathcal{V}$ is assumed to be independent, so it must be the trivial relation. This means that all the coefficients $\lambda_i - \mu_i$ are zero, so $\lambda = \mu$ as required. It is now meaningful to define $\psi(u)$ to be the unique vector λ with $\psi(\lambda) = u$. Properties are left as an exercise. □

Numerical criteria

Corollary: Let V be a d-dimensional subspace of $\mathbb{R}^n$.
- (a) Any linearly independent list in V has at most d elements.
- (b) Any list that spans V has at least d elements.
- (c) Any basis of V has exactly d elements.
- (d) Any linearly independent list of length d in V is a basis.
- (e) Any list of length d that spans V is a basis.

Proof:
- (a) This is just the definition of $\dim(V)$.
- (b) Recall: we have inverse functions $\mathbb{R}^d \xrightarrow{\phi} V \xrightarrow{\psi} \mathbb{R}^d$ with $\phi(\lambda) = \sum_i \lambda_i v_i$.
  Let $\mathcal{W} = (w_1, \dots, w_r)$ be a list that spans V. We claim that the list $(\psi(w_1), \dots, \psi(w_r))$ spans $\mathbb{R}^d$. Indeed, for any $x \in \mathbb{R}^d$ we have $\phi(x) \in V$, and $\mathcal{W}$ spans V so $\phi(x) = \sum_j \mu_j w_j$ say. We can apply ψ to this to get

  $$x = \psi(\phi(x)) = \psi(\sum_j \mu_j w_j) = \sum_j \mu_j \psi(w_j),$$

  which expresses x as a linear combination of the vectors $\psi(w_j)$, as required. We saw earlier that any list that spans $\mathbb{R}^d$ must have length at least d, so $r \geq d$ as claimed.
- (c) This holds by combining (a) and (b).
- (d) This was proved two slides ago.

Numerical criteria

Corollary: Let V be a d-dimensional subspace of $\mathbb{R}^n$.
- (a) Any linearly independent list in V has at most d elements.
- (b) Any list that spans V has at least d elements.
- (c) Any basis of V has exactly d elements.
- (d) Any linearly independent list of length d in V is a basis.
- (e) Any list of length d that spans V is a basis.

Proof:
- (e) Recall: we have inverse functions $\mathbb{R}^d \xrightarrow{\phi} V \xrightarrow{\psi} \mathbb{R}^d$ with $\phi(\lambda) = \sum_i \lambda_i v_i$.
  Let $\mathcal{W} = (w_1, \dots, w_d)$ be a list of length d that spans V.
  As in (b) we use ϕ and ψ to see that the list $(\psi(w_1), \dots, \psi(w_d))$ spans $\mathbb{R}^d$.
  This is a list of length d that spans $\mathbb{R}^d$, so it must be a basis.
  In particular, it is linearly independent.
  Claim: the original list $\mathcal{W}$ is also linearly independent.
  To see this, consider a linear relation $\sum_j \lambda_j w_j = 0$.
  By applying ψ to both sides, we get $\sum_j \lambda_j \psi(w_j) = 0$.
  As the vectors $\psi(w_j)$ are independent we see that $\lambda_j = 0$ for all j.
  This means that the original relation is the trivial one, as required.
  As $\mathcal{W}$ is linearly independent and spans V, it is a basis for V.

Lecture 16

Canonical bases

Proposition 20.6: Let V be a subspace of $\mathbb{R}^n$. Then there is a unique RREF matrix B such that the columns of B^T form a basis for V.
(We call this basis the *canonical basis* for V.)

Proof of existence.
Let $\mathcal{U} = (u_1, \dots, u_d)$ be any basis for V, and let A be the matrix with rows $u_1^T, \dots, u_d^T$.

$$A = \begin{bmatrix} u_1^T \\ \hline \vdots \\ \hline u_d^T \end{bmatrix} \to B = \begin{bmatrix} v_1^T \\ \hline \vdots \\ \hline v_d^T \end{bmatrix} \qquad B^T = \left[\begin{array}{c|c|c} v_1 & \cdots & v_d \end{array}\right]$$

Let B be the row-reduction of A, let $v_1^T, \dots, v_d^T$ be the rows of B, and put $\mathcal{V} = (v_1, \dots, v_d) =$ the list of columns of B^T. We saw earlier that a row vector can be expressed as a linear combination of the rows of A if and only if it can be expressed as a linear combination of the rows of B. This implies that $\mathrm{span}(\mathcal{V}) = \mathrm{span}(\mathcal{U}) = V$. As $\mathcal{V}$ is a list of length d that spans the d-dimensional space V, we see that $\mathcal{V}$ is actually a basis for V. $\square$

Canonical bases — towards uniqueness

Definition 20.9: Let $x = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^T$ be a nonzero vector in $\mathbb{R}^n$.
We say that x *starts in slot* k if $x_1, \dots, x_{k-1}$ are zero, but x_k is not.

Given a subspace $V \subseteq \mathbb{R}^n$, we say that k is a *jump* for V if there is a nonzero vector $x \in V$ that starts in slot k. We write $J(V)$ for the set of all jumps for V.

Example
20.10

- The vector $\begin{bmatrix} 0 & 0 & \mathbf{1} & 11 & 111 \end{bmatrix}^T$ starts in slot 3;
- The vector $\begin{bmatrix} \mathbf{1} & 2 & 3 & 4 & 5 \end{bmatrix}^T$ starts in slot 1;
- The vector $\begin{bmatrix} 0 & 0 & 0 & 0 & \mathbf{0.1234} \end{bmatrix}^T$ starts in slot 5.

Examples of jumps

Example: Consider $V = \{\begin{bmatrix} s & -s & t+s & t-s \end{bmatrix}^T \mid s, t \in \mathbb{R}\} \subseteq \mathbb{R}^4$.
If $s \neq 0$ then the vector $x = \begin{bmatrix} s & -s & t+s & t-s \end{bmatrix}^T$ starts in slot 1.
If $s = 0$ but $t \neq 0$ then $x = \begin{bmatrix} 0 & 0 & t & t \end{bmatrix}^T$ and this starts in slot 3.
If $s = t = 0$ then $x = 0$ and x does not start anywhere.
Thus, the possible starting slots for x are 1 and 3, which means that $J(V) = \{1, 3\}$.

Example: Consider the subspace
$W = \{\begin{bmatrix} a & b & c & d & e & f \end{bmatrix}^T \in \mathbb{R}^6 \mid a = b + c = d + e + f = 0\}$.
Any vector $w = \begin{bmatrix} a & b & c & d & e & f \end{bmatrix}^T$ in W can be written as
$w = \begin{bmatrix} 0 & b & -b & d & e & -d-e \end{bmatrix}^T$, where b, d and e are arbitrary.
If $b \neq 0$ then w starts in slot 2.
If $b = 0$ but $d \neq 0$ then $w = \begin{bmatrix} 0 & 0 & 0 & d & e & -d-e \end{bmatrix}^T$ starts in slot 4.
If $b = d = 0$ but $e \neq 0$ then $w = \begin{bmatrix} 0 & 0 & 0 & 0 & e & -e \end{bmatrix}^T$ starts in slot 5.
If $b = d = e = 0$ then $w = 0$ and w does not start anywhere.
Thus, the possible starting slots for w are 2, 4 and 5, so $J(W) = \{2, 4, 5\}$.

Jumps and pivots

Lemma: Let B be an RREF matrix, and suppose that the columns of B^T form a basis for a subspace $V \subseteq \mathbb{R}^n$. Then $J(V) = \{\text{cols of } B \text{ that contain pivots}\}$.

Example proof: Consider $B = \begin{bmatrix} v_1^T \\ \hline v_2^T \\ \hline v_3^T \end{bmatrix} = \begin{bmatrix} 0 & 1 & \alpha & 0 & \beta & 0 & \gamma \\ 0 & 0 & 0 & 1 & \delta & 0 & \epsilon \\ 0 & 0 & 0 & 0 & 0 & 1 & \zeta \end{bmatrix}$.

Put $V = \text{span}(v_1, v_2, v_3) \subseteq \mathbb{R}^7$, so the v_i (= cols of B^T) form a basis for V. There are pivots in columns 2, 4 and 6, so we must show that $J(V) = \{2, 4, 6\}$. Any $x \in V$ has the form $x = \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3$

$$= \begin{bmatrix} 0 & \lambda_1 & \lambda_1\alpha_1 & \lambda_2 & \lambda_1\beta + \lambda_2\delta & \lambda_3 & \lambda_1\gamma + \lambda_2\epsilon + \lambda_3\zeta \end{bmatrix}^T.$$

Note that λ_k occurs on its own in the k'th pivot column, and all entries to the left of that involve only $\lambda_1, \dots, \lambda_{k-1}$. Thus, if $\lambda_1, \dots, \lambda_{k-1}$ are all zero but $\lambda_k \neq 0$ then x starts in the k'th pivot column. In more detail:

- If $\lambda_1 \neq 0$ then $x = \begin{bmatrix} 0 & \lambda_1 & * & * & * & * & * \end{bmatrix}^T$ and so x starts in slot 2 (the first pivot column).
- If $\lambda_1 = 0 \neq \lambda_2$ then $x = \begin{bmatrix} 0 & 0 & 0 & \lambda_2 & * & * & * \end{bmatrix}^T$ and so x starts in slot 4 (the second pivot column).
- If $\lambda_1 = \lambda_2 = 0 \neq \lambda_3$ then $x = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \lambda_3 & * \end{bmatrix}^T$ and so x starts in slot 6 (the third pivot column).
- If $\lambda_1 = \lambda_2 = \lambda_3 = 0$ then $x = 0$ and so x does not start anywhere.

Canonical bases — proof of uniqueness

Proposition 20.6: Let V be a subspace of $\mathbb{R}^n$. Then there is a **unique** RREF matrix B such that the columns of B^T form a basis for V.

Sketch proof of uniqueness.

Suppose we have a subspace $V \subseteq \mathbb{R}^n$ and two RREF matrices B and C such that the columns of B^T form a basis for V, and the columns of C^T also form a basis for V. Both B and C must be $d \times n$ matrices, where $d = \dim(V)$. Let $v_1, \dots, v_d$ be the columns of B and let $w_1, \dots, w_d$ be the columns of C. Both B and C have all rows nonzero, and so have d pivots each. The pivot columns are the jumps for V and so are the same for B and C: say columns $p_1, \dots, p_d$.

Now consider one of the vectors v_i. As $v_i \in V$ and $V = \text{span}(w_1, \dots, w_d)$ we can write v_i as a linear combination of the vectors w_j, say $v_i = \lambda_1 w_1 + \dots + \lambda_d w_d$. By looking in slot p_i we see that $1 = \lambda_i$. By looking in slot p_j (where $j \neq i$) we see that $\lambda_j = 0$. Thus, the sum on the right is just w_i and we get $v_i = w_i$. This holds for all i, so we have $B = C$ as claimed. □

Finding the canonical basis for a span

Method: To find the canonical basis for a subspace $V = \text{span}(v_1, \dots, v_r)$, form the matrix

$$A = \begin{bmatrix} v_1^T \\ \hline \vdots \\ \hline v_r^T \end{bmatrix}$$

Then row-reduce to get an RREF matrix B, and discard any rows of zeros to get another RREF matrix C. The columns of C^T are the canonical basis for V.

Proof of correctness.

We showed earlier that row operations do not change the span of the rows, and it is clear that discarding rows of zeros does not change the span of the rows either, so the rows of C have the same span as the rows of A. Equivalently, the span of the columns of C^T is the same as the span of the columns of A^T, namely V. Moreover, as each pivot column of C contains a single one, it is easy to see that the rows of C are linearly independent or equivalently the columns of C^T are linearly independent. As they are linearly independent and span V, they form a basis for V. As C is in RREF, this must be the canonical basis. □

Example of finding the canonical basis for a span

Consider again the plane

$$P = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 \mid x + y + z = 0 \right\}.$$

We showed before that $P = \text{span}(u_1, u_2)$, where

$$u_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \qquad u_2 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$$

As the matrix

$$A = \begin{bmatrix} u_1^T \\ \hline u_2^T \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$$

is already in RREF, we see that the list $\mathcal{U} = (u_1, u_2)$ is the canonical basis for P.

Example of finding the canonical basis for a span

Consider again the subspace

$$V = \{\begin{bmatrix} w & x & y & z \end{bmatrix}^T \in \mathbb{R}^4 \mid w + 2x + 3y + 4z = 4w + 3x + 2y + z = 0\}.$$

We showed previously that the vectors

$$c = \begin{bmatrix} -\frac{2}{3} & 1 & 0 & -\frac{1}{3} \end{bmatrix}^T \qquad \text{and} \qquad d = \begin{bmatrix} -\frac{1}{3} & 0 & 1 & -\frac{2}{3} \end{bmatrix}^T.$$

give a (non-canonical) basis for V. To find the canonical basis, we perform the following row-reduction:

$$\left[\begin{array}{c} c^T \\ \hline d^T \end{array}\right] = \begin{bmatrix} -\frac{2}{3} & 1 & 0 & -\frac{1}{3} \\ -\frac{1}{3} & 0 & 1 & -\frac{2}{3} \end{bmatrix} \to \begin{bmatrix} 1 & -\frac{3}{2} & 0 & \frac{1}{2} \\ -\frac{1}{3} & 0 & 1 & -\frac{2}{3} \end{bmatrix} \to$$

$$\begin{bmatrix} 1 & -\frac{3}{2} & 0 & \frac{1}{2} \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} \end{bmatrix} \to \begin{bmatrix} 1 & -\frac{3}{2} & 0 & \frac{1}{2} \\ 0 & 1 & -2 & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -3 & 2 \\ 0 & 1 & -2 & 1 \end{bmatrix}$$

We conclude that the vectors $u_1 = \begin{bmatrix} 1 & 0 & -3 & 2 \end{bmatrix}^T$ and $u_2 = \begin{bmatrix} 0 & 1 & -2 & 1 \end{bmatrix}^T$ form the canonical basis for V.

Finding the canonical basis for an annihilator

Method: Suppose $V = \operatorname{ann}(u_1, \dots, u_r) = \{x \in \mathbb{R}^n \mid x.u_1 = \dots = x.u_r = 0\}$. To find the canonical basis for V:

- Write out the equations $x.u_1 = 0, \dots, x.u_r = 0$, listing the variables in backwards order (x_r down to x_1); then solve by row-reduction.
- Write the general solution as a sum of terms, each of which is an independent variable times a constant vector.
- These constant vectors form the canonical basis for V.

Finding the canonical basis for an annihilator

Example: Put $V = \operatorname{ann}(u_1, u_2, u_3)$, where

$$u_1 = \begin{bmatrix} 9 & 13 & 5 & 3 \end{bmatrix}^T \qquad u_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^T \qquad u_3 = \begin{bmatrix} 7 & 11 & 3 & 1 \end{bmatrix}^T.$$

The equations $x.u_3 = x.u_2 = x.u_1 = 0$ can be written as follows:

$$x_4 + 3x_3 + 11x_2 + 7x_1 = 0 \qquad x_4 + x_3 + x_2 + x_1 = 0 \qquad 3x_4 + 5x_3 + 13x_2 + 9x_1 = 0$$

We can row-reduce the matrix of coefficients as follows:

$$\begin{bmatrix} 1 & 3 & 11 & 7 \\ 1 & 1 & 1 & 1 \\ 3 & 5 & 13 & 9 \end{bmatrix} \to \begin{bmatrix} 0 & 2 & 10 & 6 \\ 1 & 1 & 1 & 1 \\ 0 & 2 & 10 & 6 \end{bmatrix} \to \begin{bmatrix} 0 & 1 & 5 & 3 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -4 & -2 \\ 0 & 1 & 5 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

This gives $x_4 - 4x_2 - 2x_1 = x_3 + 5x_2 + 3x_1 = 0$ so $x_4 = 4x_2 + 2x_1$ and $x_3 = -5x_2 - 3x_1$. We thus have

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ -5x_2 - 3x_1 \\ 4x_2 + 2x_1 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \\ -3 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ -5 \\ 4 \end{bmatrix}.$$

so $\begin{bmatrix} 1 & 0 & -3 & 2 \end{bmatrix}^T$ and $\begin{bmatrix} 0 & 1 & -5 & 4 \end{bmatrix}^T$ form the canonical basis for V.

Finding the canonical basis for an annihilator

Example: Put $V = \operatorname{ann}(u_1, u_2, u_3)$, where

$$u_1 = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \end{bmatrix}^T \qquad u_2 = \begin{bmatrix} 1 & 2 & 3 & 3 & 3 \end{bmatrix}^T \qquad u_3 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T.$$

To find the canonical basis, write the equations $x.u_3 = x.u_2 = x.u_1 = 0$ as:

$$\begin{aligned} x_5 + x_4 + x_3 + x_2 + x_1 &= 0 \\ 3x_5 + 3x_4 + 3x_3 + 2x_2 + x_1 &= 0 \\ 5x_5 + 4x_4 + 3x_3 + 2x_2 + x_1 &= 0. \end{aligned}$$

We now row-reduce the matrix of coefficients:

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

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

This gives $x_5 - x_3 + x_1 = 0$ and $x_4 + 2x_3 - 2x_1 = 0$ and $x_2 + 2x_1 = 0$ so $x_5 = x_3 - x_1$ and $x_4 = -2x_3 + 2x_1$ and $x_2 = -2x_1$, (x_1, x_3 independent)

Finding the canonical basis for an annihilator

$V = \mathrm{ann}(u_1, u_2, u_3)$, where

$$u_1 = \begin{bmatrix}1 & 2 & 3 & 4 & 5\end{bmatrix}^T \qquad u_2 = \begin{bmatrix}1 & 2 & 3 & 3 & 3\end{bmatrix}^T \qquad u_3 = \begin{bmatrix}1 & 1 & 1 & 1 & 1\end{bmatrix}^T.$$

Equations $x.u_3 = x.u_2 = x.u_1 = 0$ give
$x_5 = x_3 - x_1$ and $x_4 = -2x_3 + 2x_1$ and $x_2 = -2x_1$
(with x_1 and x_3 independent).

Thus

$$x = \begin{bmatrix}x_1\\x_2\\x_3\\x_4\\x_5\end{bmatrix} = \begin{bmatrix}x_1\\-2x_1\\x_3\\-2x_3+2x_1\\x_3-x_1\end{bmatrix} = x_1\begin{bmatrix}1\\-2\\0\\2\\-1\end{bmatrix} + x_3\begin{bmatrix}0\\0\\1\\-2\\1\end{bmatrix} = x_1v_1 + x_3v_2 \text{ say.}$$

It follows that the vectors

$$v_1 = \begin{bmatrix}1 & -2 & 0 & 2 & 1\end{bmatrix}^T \qquad \text{and} \qquad v_2 = \begin{bmatrix}0 & 0 & 1 & -2 & 1\end{bmatrix}^T$$

form the canonical basis for V.

Pure matrix method for annihilators

Method: Let A be a $k \times n$ matrix, and let $V \subseteq \mathbb{R}^n$ be the annihilator of the columns of A^T. We can find the canonical basis for V as follows:

(a) Rotate A through 180° to get a matrix A^*.

(b) Row-reduce A^* and discard any rows of zeros to obtain a matrix B^* in RREF. This will have shape $m \times n$ for some m with $m \leq \min(k, n)$.

(c) The matrix B^* will have m pivots (one in each row). Let columns $p_1, \dots, p_m$ be the ones with pivots, and let columns $q_1, \dots, q_{n-m}$ be the ones without pivots.

(d) Delete the pivot columns from B^* to leave an $m \times (n-m)$ matrix, which we call C^*. Let the i'th row of C^* be c_i^T (so $c_i \in \mathbb{R}^{n-m}$ for $1 \leq i \leq m$).

(e) Now construct a new matrix D^* of shape $(n-m) \times n$ as follows: the p_i'th column is $-c_i$, and the q_j'th column is the standard basis vector e_j.

(f) Rotate D through 180° to get a matrix D.

(g) The columns of D^T then form the canonical basis for V.

Pure matrix method for annihilators

Example: Again consider $V = \mathrm{ann}(u_1, u_2, u_3)$, where

$$u_1 = \begin{bmatrix}9 & 13 & 5 & 3\end{bmatrix}^T \qquad u_2 = \begin{bmatrix}1 & 1 & 1 & 1\end{bmatrix}^T \qquad u_3 = \begin{bmatrix}7 & 11 & 3 & 1\end{bmatrix}^T.$$

$$A = \begin{bmatrix}u_1^T\\ \hline u_2^T \\ \hline u_3^T\end{bmatrix} = \begin{bmatrix}9 & 13 & 5 & 3\\1 & 1 & 1 & 1\\7 & 11 & 3 & 1\end{bmatrix} \qquad A^* = \begin{bmatrix}1 & 3 & 11 & 7\\1 & 1 & 1 & 1\\3 & 5 & 13 & 9\end{bmatrix}.$$

The matrix A^* is the the matrix of coefficients appearing in our previous approach; as we saw we can row-reduce and delete zeros as follows:

$$A^* = \begin{bmatrix}1 & 3 & 11 & 7\\1 & 1 & 1 & 1\\3 & 5 & 13 & 9\end{bmatrix} \to \begin{bmatrix}1 & 0 & -4 & -2\\0 & 1 & 5 & 3\\0 & 0 & 0 & 0\end{bmatrix} \to \begin{bmatrix}1 & 0 & -4 & -2\\0 & 1 & 5 & 3\end{bmatrix} = B^*.$$

The pivot columns are $p_1 = 1$ and $p_2 = 2$, whereas the non-pivot columns are $q_1 = 3$ and $q_2 = 4$. We now delete the pivot columns to get

$$C^* = \begin{bmatrix}c_1^T\\ \hline c_2^T\end{bmatrix} = \begin{bmatrix}-4 & -2\\5 & 3\end{bmatrix}.$$

$$D^* = \left[\begin{array}{c|c|c|c}-c_1 & -c_2 & e_1 & e_2\end{array}\right] = \begin{bmatrix}4 & -5 & 1 & 0\\2 & -3 & 0 & 1\end{bmatrix}; \; D = \begin{bmatrix}1 & 0 & -3 & 2\\0 & 1 & -5 & 4\end{bmatrix}.$$

Canonical basis for V: $\begin{bmatrix}1 & 0 & -3 & 2\end{bmatrix}^T$ and $\begin{bmatrix}0 & 1 & -5 & 4\end{bmatrix}^T$.

Pure matrix method for annihilators

Example: Again consider $V = \mathrm{ann}(u_1, u_2, u_3)$, where

$$u_1 = \begin{bmatrix}1 & 2 & 3 & 4 & 5\end{bmatrix}^T \qquad u_2 = \begin{bmatrix}1 & 2 & 3 & 3 & 3\end{bmatrix}^T \qquad u_3 = \begin{bmatrix}1 & 1 & 1 & 1 & 1\end{bmatrix}^T.$$

$$A = \begin{bmatrix}1 & 2 & 3 & 4 & 5\\1 & 2 & 3 & 3 & 3\\1 & 1 & 1 & 1 & 1\end{bmatrix} \qquad A^* = \begin{bmatrix}1 & 1 & 1 & 1 & 1\\3 & 3 & 3 & 2 & 1\\5 & 4 & 3 & 2 & 1\end{bmatrix}$$

A^* = matrix of coefficients in previous approach. As before:

$$A^* = \begin{bmatrix}1 & 1 & 1 & 1 & 1\\3 & 3 & 3 & 2 & 1\\5 & 4 & 3 & 2 & 1\end{bmatrix} \to \begin{bmatrix}1 & 0 & -1 & 0 & 1\\0 & 1 & 2 & 0 & -2\\0 & 0 & 0 & 1 & 2\end{bmatrix} = B^*.$$

Pivot cols $p_1 = 1$, $p_2 = 2$ and $p_3 = 4$; non-pivot cols $q_1 = 3$ and $q_2 = 5$.

Deleting pivot columns leaves $C^* = \begin{bmatrix}c_1^T\\ \hline c_2^T\\ \hline c_3^T\end{bmatrix} = \begin{bmatrix}-1 & -1\\2 & -2\\0 & 2\end{bmatrix}$

$$D^* = \left[\begin{array}{c|c|c|c|c}-c_1 & -c_2 & e_1 & -c_3 & e_2\end{array}\right] = \begin{bmatrix}-1 & 2 & 1 & 0 & 0\\-1 & -2 & 0 & -2 & 1\end{bmatrix}.$$

Rotate: $D = \begin{bmatrix}1 & -2 & 0 & -2 & -1\\0 & 0 & 1 & 2 & -1\end{bmatrix}$. Rows of D give canonical basis for V.

Describing spans as annihilators

We have just discussed a method that finds a basis for an annihilator, and so describes the annihilator as a span.
Opposite problem: describe a span as an annihilator.
In more detail: given $v_1, \dots, v_r$ find $u_1, \dots, u_s$ such that $\mathrm{span}(v_1, \dots, v_r) = \mathrm{ann}(u_1, \dots, u_s)$.

Method:

(a) Write out the equations $x.v_r = 0, \dots, x.v_1 = 0$, listing the variables in backwards order (x_r down to x_1).

(b) Solve by row-reduction in the usual way.

(c) Write the general solution as a sum of terms, each of which is an independent variable times a constant vector.

(d) Call these constant vectors $u_1, \dots, u_s$. Then $V = \mathrm{ann}(u_1, \dots, u_s)$.

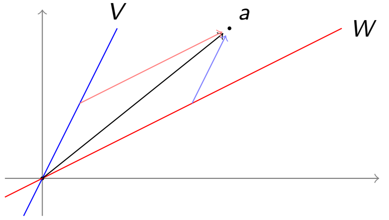
Lecture 17

Sums and intersections of subspaces

Definition: Let V and W be subspaces of $\mathbb{R}^n$. We define

$$V + W = \{x \in \mathbb{R}^n \mid x \text{ can be expressed as } v + w \text{ for some } v \in V \text{ and } w \in W\}$$
$$V \cap W = \{x \in \mathbb{R}^n \mid x \in V \text{ and also } x \in W\}.$$

Example: Put $V = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \mid y = 2x \right\}$ $\quad W = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \mid 2y = x \right\}$



Then $V \cap W$ is the set of points lying on both lines, but the lines only meet at the origin, so $V \cap W = \{0\}$.
Every point $a \in \mathbb{R}^2$ can be expressed as the sum of a point on V with a point on W, so $V + W = \mathbb{R}^2$.

Sums and intersections of subspaces

Definition: Let V and W be subspaces of $\mathbb{R}^n$. We define

$$V + W = \{x \in \mathbb{R}^n \mid x \text{ can be expressed as } v + w \text{ for some } v \in V \text{ and } w \in W\}$$
$$V \cap W = \{x \in \mathbb{R}^n \mid x \in V \text{ and also } x \in W\}.$$

Example: Put $V = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \mid y = 2x \right\}$ $\quad W = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 \mid 2y = x \right\}$

Algebraically:

- If $\begin{bmatrix} x \\ y \end{bmatrix} \in V \cap W$ then $y = 2x$ and also $x = 2y$, so $x = y = 0$, so $\begin{bmatrix} x \\ y \end{bmatrix} = 0$. Thus $V \cap W = \{0\}$.
- Consider an arbitrary point $a = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$. If we put

$$v = \frac{2y - x}{3} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad w = \frac{2x - y}{3} \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

we find that $v \in V$ and $w \in W$ and $a = v + w$, which shows that $a \in V + W$. Thus $V + W = \mathbb{R}^2$.

Sum and intersection example

Put
$$V = \{\begin{bmatrix} w & x & y & z\end{bmatrix}^T \in \mathbb{R}^4 \mid w = y \text{ and } x = z\}$$
$$W = \{\begin{bmatrix} w & x & y & z\end{bmatrix}^T \in \mathbb{R}^4 \mid w + z = x + y = 0\}.$$

For a vector $u = \begin{bmatrix} w & x & y & z\end{bmatrix}^T$ to lie in $V \cap W$ we must have $w = y$ and $x = z$ and $w = -z$ and $x = -y$,
so $u = \begin{bmatrix} w & -w & w & -w\end{bmatrix}^T$, so $V \cap W$ is just the set of multiples of $\begin{bmatrix} 1 & -1 & 1 & -1\end{bmatrix}$. Now put

$$U = \{\begin{bmatrix} w & x & y & z\end{bmatrix}^T \mid w - x - y + z = 0\} = \operatorname{ann}(\begin{bmatrix} 1 & -1 & -1 & 1\end{bmatrix}^T).$$

We claim that $V + W = U$. Proof: consider a $u = \begin{bmatrix} w & x & y & z\end{bmatrix}^T$.

- Suppose $u \in V + W$. Then $u = v + w$ for some $v \in V$ and $w \in W$, say $v = \begin{bmatrix} p & q & p & q\end{bmatrix}$ and $w = \begin{bmatrix} -r & -s & s & r\end{bmatrix}^T$. This gives $u = v + w = \begin{bmatrix} p-r & q-s & p+s & q+r\end{bmatrix}$, so
$$w-x-y+z = (p-r)-(q-s)-(p+s)+(q+r) = p-r-q+s-p-s+q+r = 0,$$
proving that $u \in U$ as required.
- Suppose $u \in U$, so $z = x + y - w$. Put $v = \begin{bmatrix} y & x & y & x\end{bmatrix}^T$ and $w = \begin{bmatrix} w-y & 0 & 0 & y-w\end{bmatrix}^T$. We find that $v \in V$ and $w \in W$ and $v + w = u$, which proves that $u \in V + W$ as required.

Sum of spans, intersection of annihilators

Proposition: For lists $v_1, \dots, v_r$ and $w_1, \dots, w_s$ of vectors in $\mathbb{R}^n$, we have

(a) $\operatorname{span}(v_1, \dots, v_r) + \operatorname{span}(w_1, \dots, w_s) = \operatorname{span}(v_1, \dots, v_r, w_1, \dots, w_s)$.

(b) $\operatorname{ann}(v_1, \dots, v_r) \cap \operatorname{ann}(w_1, \dots, w_s) = \operatorname{ann}(v_1, \dots, v_r, w_1, \dots, w_s)$.

Proof.

(a) An arbitrary element $x \in \operatorname{span}(v_1, \dots, v_r) + \operatorname{span}(w_1, \dots, w_s)$ has the form $x = v + w$, where v is an arbitrary element of $\operatorname{span}(v_1, \dots, v_r)$ and w is an arbitrary element of $\operatorname{span}(w_1, \dots, w_s)$. This means that $v = \sum_{i=1}^r \lambda_i v_i$ and $w = \sum_{j=1}^s \mu_j w_j$ for some coefficients $\lambda_1, \dots, \lambda_r$ and $\mu_1, \dots, \mu_s$, so
$$x = \lambda_1 v_1 + \cdots + \lambda_r v_r + \mu_1 w_1 + \cdots + \mu_s w_s.$$
This is also the general form for an element of $\operatorname{span}(v_1, \dots, v_r, w_1, \dots, w_s)$.

(b) A vector $x \in \mathbb{R}^n$ lies in $\operatorname{ann}(v_1, \dots, v_r)$ if and only if $x.v_1 = \cdots = x.v_r = 0$. Similarly, x lies in $\operatorname{ann}(w_1, \dots, w_s)$ iff $x.w_1 = \cdots = x.w_s$. Thus, x lies in $\operatorname{ann}(v_1, \dots, v_r) \cap \operatorname{ann}(w_1, \cdots, w_s)$ iff both sets of equations are satisfied, or in other words $x.v_1 = \cdots = x.v_r = x.w_1 = \cdots = x.w_s = 0$.
This is precisely the condition for x to lie in $\operatorname{ann}(v_1, \dots, v_r, w_1, \dots, w_s)$.

$\square$

Finding sums and intersections

Method 21.5: To find the sum of two subspaces $V, W \subseteq \mathbb{R}^n$:

(a) Find a list $\mathcal{V}$ such that $V = \operatorname{span}(\mathcal{V})$. (If V is given as an annihilator, use earlier method to find the canonical basis $\mathcal{V}$ for V; then $V = \operatorname{span}(\mathcal{V})$.)

(b) Find a list $\mathcal{W}$ such that $W = \operatorname{span}(\mathcal{W})$ (in the same way).

(c) Now $V + W$ is the span of the combined list $\mathcal{V}, \mathcal{W}$. If desired, we can make this list the rows of a matrix then row-reduce and discard zeros to get the canonical basis for $V + W$.

Method 21.6: To find the intersection of two subspaces $V, W \subseteq \mathbb{R}^n$:

(a) Find a list $\mathcal{V}'$ such that $V = \operatorname{ann}(\mathcal{V}')$. It may be that V is given to us as the annihilator of some list, in which case there is nothing to do. Alternatively, if V is given to as as the span of some list, then gave a method earlier to find a list $\mathcal{V}'$ such that $\operatorname{ann}(\mathcal{V}') = V$.

(b) Find a list $\mathcal{W}'$ such that $W = \operatorname{ann}(\mathcal{W}')$ (in the same way).

(c) Now $V \cap W$ is the annihilator of the combined list $\mathcal{V}', \mathcal{W}'$. Earlier we described how to find the canonical basis for an annihilator, so we can use that to get the canonical basis for $V \cap W$.

The dimension formula

Dimensions of V, W, $V \cap W$ and $V + W$ are linked by the following formula:
$$\dim(V \cap W) + \dim(V + W) = \dim(V) + \dim(W).$$

Example:
$$\begin{aligned}
V &= \operatorname{span}(e_1, \dots, e_p, e_{p+1}, \dots, e_{p+q}) \\
W &= \operatorname{span}(e_1, \dots, e_p, e_{p+q+1}, \dots, e_{p+q+r}) \\
V \cap W &= \operatorname{span}(e_1, \dots, e_p) \\
V + W &= \operatorname{span}(e_1, \dots, e_{p+q+r}) \\
\dim(V \cap W) + \dim(V + W) &= p + (p + q + r) = 2p + q + r \\
\dim(V) + \dim(W) &= (p + q) + (p + r) = 2p + q + r.
\end{aligned}$$

- If we know three of the numbers $\dim(V \cap W)$, $\dim(V + W)$, $\dim(V)$ and $\dim(W)$, we can rearrange the formula to find the fourth.
- If you believe that you have found bases for V, W, $V \cap W$ and $V + W$, you can use the formula as a check that your bases are correct.

We will not prove the formula, except to say that one can choose bases to make the proof like the above example. Details would be a digression.

Sum and intersection example

Put $V = \text{span}(v_1, v_2, v_3)$ and $W = \text{span}(w_1, w_2, w_3)$ where

$$v_1 = \begin{bmatrix}1\\0\\0\\0\end{bmatrix} \quad v_2 = \begin{bmatrix}1\\1\\1\\0\end{bmatrix} \quad v_3 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} \quad w_1 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix} \quad w_2 = \begin{bmatrix}0\\1\\1\\0\end{bmatrix} \quad w_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$$

Claim: $V + W = \mathbb{R}^4$. Systematic proof: recall
$V + W = \text{span}(v_1, v_2, v_3, w_1, w_2, w_3))$ and row-reduce:

$$\left[\begin{array}{c} v_1^T \\ \hline v_2^T \\ \hline v_3^T \\ \hline w_1^T \\ \hline w_2^T \\ \hline w_3^T \end{array}\right] = \begin{bmatrix}1&0&0&0\\1&1&1&0\\1&1&1&1\\1&1&0&0\\0&1&1&0\\0&0&1&1\end{bmatrix} \to \left[\begin{array}{cccc}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\\ \hline 0&0&0&0\\0&0&0&0\end{array}\right] = \left[\begin{array}{c} e_1^T \\ \hline e_2^T \\ \hline e_3^T \\ \hline e_4^T \\ \hline 0 \\ \hline 0 \end{array}\right]$$

Conclusion: (e_1, e_2, e_3, e_4) is the canonical basis for $V + W$, so $V + W = \mathbb{R}^4$.
More efficiently:

$$e_1 = v_1 \qquad e_2 = w_1 - v_1 \qquad e_3 = v_2 - w_1 \qquad e_4 = v_3 - v_2.$$

It follows that e_1, e_2, e_3 and e_4 are all in $V + W$, so $V + W = \mathbb{R}^4$.

Sum and intersection example

$$v_1 = \begin{bmatrix}1\\0\\0\\0\end{bmatrix} \quad v_2 = \begin{bmatrix}1\\1\\1\\0\end{bmatrix} \quad v_3 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} \quad w_1 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix} \quad w_2 = \begin{bmatrix}0\\1\\1\\0\end{bmatrix} \quad w_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$$

Now find $V \cap W$. First step: describe V as an annihilator. Write equations $x.v_3 = 0$, $x.v_2 = 0$ and $x.v_1 = 0$, with the variables x_i in descending order:

$$\begin{aligned} x_4 + x_3 + x_2 + x_1 &= 0 \\ x_3 + x_2 + x_1 &= 0 \\ x_1 &= 0. \end{aligned}$$

Clearly $x_1 = x_4 = 0$ and $x_3 = -x_2$, with x_2 arbitrary. Thus:

$$x = \begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix} = \begin{bmatrix}0\\x_2\\-x_2\\0\end{bmatrix} = x_2 \begin{bmatrix}0\\1\\-1\\0\end{bmatrix}$$

We conclude that $V = \text{ann}(a)$, where $a = \begin{bmatrix}0 & 1 & -1 & 0\end{bmatrix}^T$.

Sum and intersection example

$$v_1 = \begin{bmatrix}1\\0\\0\\0\end{bmatrix} \quad v_2 = \begin{bmatrix}1\\1\\1\\0\end{bmatrix} \quad v_3 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} \quad w_1 = \begin{bmatrix}1\\1\\0\\0\end{bmatrix} \quad w_2 = \begin{bmatrix}0\\1\\1\\0\end{bmatrix} \quad w_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$$

Next step: describe W as an annihilator. Write down the equations $x.w_3 = 0$, $x.w_2 = 0$ and $x.w_1 = 0$, with the variables x_i in descending order:

$$\begin{aligned} x_4 + x_3 &= 0 \\ x_3 + x_2 &= 0 \\ x_2 + x_1 &= 0. \end{aligned}$$

This easily gives $x_4 = -x_3 = x_2 = -x_1$, so

$$x = \begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix} = \begin{bmatrix}x_1\\-x_1\\x_1\\-x_1\end{bmatrix} = x_1 \begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}$$

We conclude that $W = \text{ann}(b)$, where $b = \begin{bmatrix}1 & -1 & 1 & -1\end{bmatrix}^T$.

Sum and intersection example

$$a = \begin{bmatrix}0 & 1 & -1 & 0\end{bmatrix}^T \qquad b = \begin{bmatrix}1 & -1 & 1 & -1\end{bmatrix}^T$$

We now have $V = \text{ann}(a)$ and $W = \text{ann}(b)$ so $V \cap W = \text{ann}(a, b)$. To find the canonical basis for this, we write the equations $x.b = 0$ and $x.a = 0$, again with the variables in decreasing order:

$$\begin{aligned} -x_4 + x_3 - x_2 + x_1 &= 0 \\ -x_3 + x_2 &= 0 \end{aligned}$$

After row-reduction we get $x_4 = x_1$ and $x_3 = x_2$ with x_1 and x_2 arbitrary. This gives

$$x = \begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix} = \begin{bmatrix}x_1\\x_2\\x_2\\x_1\end{bmatrix} = x_1 \begin{bmatrix}1\\0\\0\\1\end{bmatrix} + x_2 \begin{bmatrix}0\\1\\1\\0\end{bmatrix}.$$

We conclude that the vectors $u_1 = \begin{bmatrix}1 & 0 & 0 & 1\end{bmatrix}^T$ and $u_2 = \begin{bmatrix}0 & 1 & 1 & 0\end{bmatrix}^T$ form the canonical basis for $V \cap W$. As a sanity check we can note that

$$\begin{aligned} u_1 &= v_1 - v_2 + v_3 \in V & u_2 &= v_2 - v_1 \in V \\ u_1 &= w_1 - w_2 + w_3 \in W & u_2 &= w_2 \in W. \end{aligned}$$

These equations show directly that u_1 and u_2 lie in $V \cap W$.

Dimension check

We will use the dimension formula to check our calculation.

- $V = \mathrm{span}(v_1, v_2, v_3)$ and one can check that this list is independent so $\dim(V) = 3$.
- $W = \mathrm{span}(w_1, w_2, w_3)$ and one can check that this list is independent so $\dim(W) = 3$.
- We showed that $V + W = \mathbb{R}^4$ so $\dim(V + W) = 4$.
- We showed that u_1, u_2 is a basis for $V \cap W$ so $\dim(V \cap W) = 2$.
- Now $\dim(V + W) + \dim(V \cap W) = 4 + 2 = 6$ and $\dim(V) + \dim(W) = 3 + 3 = 6$. As expected, these are the same.

Sum and intersection example

Put $V = \mathrm{span}(v_1, v_2)$ and $W = \mathrm{span}(w_1, w_2)$ where

$$v_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} \qquad v_2 = \begin{bmatrix}1\\2\\3\\4\end{bmatrix} \qquad w_1 = \begin{bmatrix}-3\\-1\\1\\3\end{bmatrix} \qquad w_2 = \begin{bmatrix}0\\1\\-1\\0\end{bmatrix}$$

We will find the canonical bases for V, W, $V + W$ and $V \cap W$. For V:

$$\left[\frac{v_1^T}{v_2^T}\right] = \begin{bmatrix}1&1&1&1\\1&2&3&4\end{bmatrix} \to \begin{bmatrix}1&1&1&1\\0&1&2&3\end{bmatrix} \to \begin{bmatrix}1&0&-1&-2\\0&1&2&3\end{bmatrix}$$

Thus: the vectors $v_1' = \begin{bmatrix}1&0&-1&-2\end{bmatrix}^T$ and $v_2' = \begin{bmatrix}0&1&2&3\end{bmatrix}^T$ form the canonical basis for V.
Similarly, the row-reduction

$$\left[\frac{w_1^T}{w_2^T}\right] = \begin{bmatrix}-3&-1&1&3\\0&1&-1&0\end{bmatrix} \to \begin{bmatrix}-3&0&0&3\\0&1&-1&0\end{bmatrix} \to \begin{bmatrix}1&0&0&-1\\0&1&-1&0\end{bmatrix}$$

shows that the vectors $w_1' = \begin{bmatrix}1&0&0&-1\end{bmatrix}^T$ and $w_2' = \begin{bmatrix}0&1&-1&0\end{bmatrix}^T$ form the canonical basis for W.

Sum and intersection example

$V = \mathrm{span}(v_1', v_2')$ and $W = \mathrm{span}(w_1', w_2')$ where

$$v_1' = \begin{bmatrix}1\\0\\-1\\-2\end{bmatrix} \qquad v_2' = \begin{bmatrix}0\\1\\2\\3\end{bmatrix} \qquad w_1' = \begin{bmatrix}1\\0\\0\\-1\end{bmatrix} \qquad w_2' = \begin{bmatrix}0\\1\\-1\\0\end{bmatrix}$$

Next find the canonical basis for $V + W = \mathrm{span}(v_1', v_2', w_1', w_2')$, by row-reducing either the matrix $[v_1'|v_2'|w_1'|w_2']^T$:

$$\begin{bmatrix}1&0&-1&-2\\0&1&2&3\\1&0&0&-1\\0&1&-1&0\end{bmatrix} \to \begin{bmatrix}1&0&-1&-2\\0&1&2&3\\0&0&1&1\\0&0&-3&-3\end{bmatrix} \to \begin{bmatrix}1&0&0&-1\\0&1&0&1\\0&0&1&1\\0&0&0&0\end{bmatrix}.$$

We conclude that the following vectors form the canonical basis for $V + W$:

$$u_1 = \begin{bmatrix}1\\0\\0\\-1\end{bmatrix} \qquad u_2 = \begin{bmatrix}0\\1\\0\\1\end{bmatrix} \qquad u_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}.$$

In particular, we have $\dim(V + W) = 3$.

Sum and intersection example

$$V = \mathrm{span}(v_1', v_2') \qquad v_1' = \begin{bmatrix}1&0&-1&-2\end{bmatrix}^T \qquad v_2' = \begin{bmatrix}0&1&2&3\end{bmatrix}^T$$

$$W = \mathrm{span}(w_1', w_2') \qquad w_1' = \begin{bmatrix}1&0&0&-1\end{bmatrix}^T \qquad w_2' = \begin{bmatrix}0&1&-1&0\end{bmatrix}^T$$

Next, to understand $V \cap W$, we need to write V and W as annihilators.
For W: put $b_1 = \begin{bmatrix}1&0&0&1\end{bmatrix}^T$ and $b_2 = \begin{bmatrix}0&1&1&0\end{bmatrix}^T$.
After considering the form of the vectors w_1' and w_2' we see that

$$W = \left\{ \begin{bmatrix}x_1\\x_2\\-x_2\\-x_1\end{bmatrix} \;\middle|\; x_1, x_2 \in \mathbb{R} \right\} = \left\{ \begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix} \;\middle|\; x_1 + x_4 = x_2 + x_3 = 0 \right\} = \mathrm{ann}(b_1, b_2).$$

For V: the equations $x.v_1' = 0$ and $x.v_2' = 0$ are $-2x_4 - x_3 + x_1 = 0$ and $3x_4 + 2x_3 + x_2 = 0$. Solution:

$$x_3 = -2x_2 - 3x_1 \qquad x_4 = x_2 + 2x_1 \qquad (\,x_2 \text{ and } x_1 \text{ arbitrary})$$

$$x = \begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix} = \begin{bmatrix}x_1\\x_2\\-2x_2 - 3x_1\\x_2 + 2x_1\end{bmatrix} = x_1 \begin{bmatrix}1\\0\\-3\\2\end{bmatrix} + x_2 \begin{bmatrix}0\\1\\-2\\1\end{bmatrix} = x_1 a_1 + x_2 a_2 \text{ say.}$$

Thus $V = \mathrm{ann}(a_1, a_2)$, where $a_1 = \begin{bmatrix}1&0&-3&2\end{bmatrix}^T$, $a_2 = \begin{bmatrix}0&1&-2&1\end{bmatrix}^T$.

Sum and intersection example

$$a_1 = \begin{bmatrix} 1 & 0 & -3 & 2 \end{bmatrix}^T \qquad a_2 = \begin{bmatrix} 0 & 1 & -2 & 1 \end{bmatrix}^T$$
$$b_1 = \begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix}^T \qquad b_2 = \begin{bmatrix} 0 & 1 & 1 & 0 \end{bmatrix}^T$$

We now have

$$V \cap W = \text{ann}(a_1, a_2) \cap \text{ann}(b_1, b_2) = \text{ann}(a_1, a_2, b_1, b_2).$$

To find the canonical basis, solve $x.b_2 = x.b_1 = x.a_2 = x.a_1 = 0$:

$$\begin{aligned} x_3 + x_2 &= 0 & x_4 + x_1 &= 0 \\ x_4 - 2x_3 + x_2 &= 0 & 2x_4 - 3x_3 + x_1 &= 0. \end{aligned}$$

The first two equations give $x_3 = -x_2$ and $x_4 = -x_1$, which we can substitute into the remaining equations to get $x_2 = x_1/3$. This leads to $x = x_1 \begin{bmatrix} 1 & 1/3 & -1/3 & -1 \end{bmatrix}^T$, so the vector $c = \begin{bmatrix} 1 & 1/3 & -1/3 & -1 \end{bmatrix}^T$ is (by itself) the canonical basis for $V \cap W$. In particular, we have $\dim(V \cap W) = 1$.
As a check, we note that

$$\dim(V + W) + \dim(V \cap W) = 3 + 1 = 2 + 2 = \dim(V) + \dim(W),$$

as expected.

Lecture 18

Rank

Definition 22.1: For any matrix A, put

$$\begin{aligned} \text{rank}(A) &= \dim(\text{ span of the columns of } A) \\ &= \dim(\text{ span of the rows of } A^T). \end{aligned}$$

Column operations

Definition 22.2: A matrix A is in *reduced column echelon form* (RCEF) if A^T is in RREF, or equivalently:

- **RCEF0**: Any column of zeros come at the right hand end of the matrix, after all the nonzero columns.
- **RCEF1**: In any nonzero column, the first nonzero entry is equal to one. These entries are called *copivots*.
- **RCEF2**: In any nonzero column, the copivot is further down than the copivots in all previous rows.
- **RCEF3**: If a row contains a copivot, then all other entries in that row are zero.

Definition 22.3: Let A be a matrix. The following operations on A are called *elementary column operations*:

- **ECO1**: Exchange two columns.
- **ECO2**: Multiply a column by a nonzero constant.
- **ECO3**: Add a multiple of one column to another column.

Rank of an RCEF matrix

Proposition 22.4: If a matrix A is in RCEF, then the rank of A is just the number of nonzero columns.

Proof.
Let the nonzero columns be $u_1, \dots, u_r$, and put $U = \mathrm{span}(u_1, \dots, u_r)$.
This is the same as the span of *all* the columns, because columns of zeros do not contribute anything to the span.
We claim that the vectors u_i are linearly independent.
To see this, note that each u_i contains a copivot, say in the q_i'th row. As the matrix is in RCEF we have $q_1 < \dots < q_r$, and the q_i'th row is all zero apart from the copivot in u_i. In other words, for $j \neq i$ the q_i'th entry in u_j is zero.
Now suppose we have a linear relation $\lambda_1 u_1 + \dots + \lambda_r u_r = 0$.
By looking at the q_i'th entry, we see that λ_i is zero.
This holds for all i, so we have the trivial linear relation.
This proves that the list $u_1, \dots, u_r$ is linearly independent, so it forms a basis for U, so $\dim(U) = r$. We thus have $\mathrm{rank}(A) = r$ as claimed. $\square$

Basic facts about column operations

Proposition 22.5: Any matrix A can be converted to RCEF by a sequence of elementary column operations.
Proof: Analogous to Method 6.3 for row operations. $\square$

Proposition 22.6: Suppose that A can be converted to B by a sequence of elementary column operations. Then $B = AV$ for some invertible matrix V.

Proof.
A^T can be converted to B^T by a sequence of row operations corresponding to the column operations that were used to convert A to B.
Thus, Corollary 11.10 tells us that $B^T = UA^T$ for some invertible matrix U.
We thus have $B = B^{TT} = (UA^T)^T = A^{TT}U^T = AU^T$.
Here U^T is also invertible, so we can take $V = U^T$. $\square$

Proposition 22.7: Suppose that A can be converted to B by a sequence of elementary column operations. Then the span of the columns of A is the same as the span of the columns of B (and so $\mathrm{rank}(A) = \mathrm{rank}(B)$).
Proof: Analogous to Corollary 9.16 for row operations. $\square$

Invariance under row operations

Proposition 22.8: Suppose that A can be converted to B by a sequence of elementary *row* operations. Then $\mathrm{rank}(A) = \mathrm{rank}(B)$.

Proof: Let the columns of A be $v_1, \dots, v_n$ and put $V = \mathrm{span}(v_1, \dots, v_n)$ so $\mathrm{rank}(A) = \dim(V)$. There is an invertible matrix P such that

$$B = PA = P\left[\begin{array}{c|c|c} v_1 & \cdots & v_n \end{array}\right] = \left[\begin{array}{c|c|c} Pv_1 & \cdots & Pv_n \end{array}\right],$$

so the vectors Pv_i are the columns of B. Thus, if we put $W = \mathrm{span}(Pv_1, \dots, Pv_n)$, then $\mathrm{rank}(B) = \dim(W)$.

Claim: if $x \in V$ then $Px \in W$. Indeed, if $x \in V$ then $x = \sum_{i=1}^n \lambda_i v_i$ for some sequence of coefficients $\lambda_1, \dots, \lambda_n$. This means that $Px = \sum_{i=1}^n \lambda_i Pv_i$, which is a linear combination of $Pv_1, \dots, Pv_n$, so $Px \in W$.

Claim: if $y \in W$ then $P^{-1}y \in V$. Indeed, if $y \in W$ then $y = \sum_{i=1}^n \lambda_i Pv_i$ for some sequence of coefficients $\lambda_1, \dots, \lambda_n$. This means that $P^{-1}y = \sum_{i=1}^n \lambda_i P^{-1}Pv_i = \sum_{i=1}^n \lambda_i v_i$, which is a linear combination of $v_1, \dots, v_n$, so $P^{-1}y \in V$.

Invariance under row operations

$V = \mathrm{span}(v_1, \dots, v_n)$; $W = \mathrm{span}(Pv_1, \dots, Pv_n)$;
if $x \in V$ then $Px \in W$; if $y \in W$ then $P^{-1}y \in V$.

Now choose a basis $a_1, \dots, a_r$ for V (so $\mathrm{rank}(A) = \dim(V) = r$).
Claim: the vectors $Pa_1, \dots, Pa_r$ form a basis for W.
We just showed that $Px \in W$ whenever $x \in V$, so at least $Pa_i \in W$.

Consider an arbitrary element $y \in W$. We then have $P^{-1}y \in V$, but the vectors a_i form a basis for V, so we have $P^{-1}y = \sum_{i=1}^r \mu_i a_i$ for some sequence of coefficients μ_i. This means that $y = PP^{-1}y = \sum_i \mu_i Pa_i$, which expresses y as a linear combination of the vectors Pa_i. It follows that the list $Pa_1, \dots, Pa_r$ spans W.

We need to check that it is also linearly independent.
Suppose we have a linear relation $\sum_i \lambda_i Pa_i = 0$. After multiplying by P^{-1}, we get a linear relation $\sum_i \lambda_i a_i = 0$. The list $a_1, \dots, a_r$ is assumed to be a basis for V, so this must be the trivial relation, so $\lambda_1 = \dots = \lambda_r = 0$, or in other words the original relation $\sum_i \lambda_i Pa_i = 0$ was the trivial one.

We have now shown that $Pa_1, \dots, Pa_r$ is a basis for W, so $\dim(W) = r$. In conclusion, we have $\mathrm{rank}(A) = r = \mathrm{rank}(B)$ as required. $\square$

Normal form

Definition 22.9: An $n \times m$ matrix A is in *normal form* if it has the form

$$A = \left[\begin{array}{c|c} I_r & 0_{r\times(m-r)} \\ \hline 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{array}\right]$$

for some r. ($r = 0$ is allowed, in which case A is just the zero matrix.)

If A is in normal form as above, then $\operatorname{rank}(A) = r =$ the number of ones in A.

Example 22.10: There are precisely four different 3×5 matrices that are in normal form, one of each rank from 0 to 3 inclusive.

$$A_0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \qquad A_1 = \left[\begin{array}{c|cccc} 1 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$A_2 = \left[\begin{array}{cc|ccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right] \qquad A_3 = \left[\begin{array}{ccc|cc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{array}\right]$$

Reduction to normal form

Proposition 22.11: Any $n \times m$ matrix A can be converted to a matrix C in normal form by a sequence of row and column operations. Moreover:

(a) There is an invertible $n \times n$ matrix U and an invertible $m \times m$ matrix V such that $C = UAV$.

(b) $\operatorname{rank}(A) = \operatorname{rank}(C) =$ the number of ones in C.

Proof: Perform row operations to get a matrix B in RREF.
By Corollary 11.10 there is an invertible matrix U such that $B = UA$.
(This has to be an $n \times n$ matrix for the product UA to make sense.)
Now subtract multiples of pivot columns from columns further to the right.
As each pivot column contains nothing but the pivot, the only effect of these column operations is to set everything to the right of a pivot equal to zero.
However, every nonzero entry in B is either a pivot or to the right of a pivot, so after these ops we just have the pivots from B and everything else is zero.
Now just move all columns of zeros to the right hand end, which leaves a matrix C in normal form. As C was obtained from B by a sequence of elementary column operations, we have $C = BV$ for some invertible $m \times m$ matrix V. As $B = UA$, it follows that $C = UAV$. Propositions 22.7 and 22.8 tell us that neither row nor column operations affect the rank, so $\operatorname{rank}(A) = \operatorname{rank}(C)$, and because C is in normal form, $\operatorname{rank}(C)$ is just the number of ones in C.

Example of reduction to normal form

Consider the matrix

$$A = \begin{bmatrix} 1 & 3 & 0 & 1 \\ 2 & 6 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 2 & 9 \end{bmatrix}.$$

This can be row-reduced as follows:

$$\begin{bmatrix} 1 & 3 & 0 & 1 \\ 2 & 6 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 2 & 9 \end{bmatrix} \to \begin{bmatrix} 1 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 2 & 8 \end{bmatrix} \to \begin{bmatrix} 1 & 3 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 3 & 0 & 1 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We now perform column operations:

$$\begin{bmatrix} 1 & 3 & 0 & 1 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \to \left[\begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

(Subtract column 1 from column 4, and 3 times column 1 from column 2; subtract 4 times column 3 from column 4; exchange columns 2 and 3.)
We are left with a matrix of rank 2 in normal form, so $\operatorname{rank}(A) = 2$.

Example of reduction to normal form

Consider the matrix

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \\ 4 & 5 & 6 \end{bmatrix}$$

This can be reduced to normal form as follows: $A \to$

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ 0 & -2 & -4 \\ 0 & -3 & -6 \end{bmatrix} \to \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & -2 & -4 \\ 0 & -3 & -6 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \to \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

- Subtract multiples of row 1 from the other rows
- Multiply row 2 by -1
- Subtract multiples of row 2 from the other rows
- Add column 1 to column 3
- Subtract 2 times column 2 from column 3.

The final matrix has rank 2, so we must also have $\operatorname{rank}(A) = 2$.

$\mathrm{rank}(A) = \mathrm{rank}(A^T)$

Proposition 22.14: For any matrix A we have $\mathrm{rank}(A) = \mathrm{rank}(A^T)$.

Proof: We can convert A by row and column ops to a matrix C in normal form, and $\mathrm{rank}(A)$ is the number of ones in C. If we transpose everything then the row ops become column ops and *vice-versa*, so A^T can be converted to C^T by row and column ops, and C^T is also in normal form, so $\mathrm{rank}(A^T) =$ the number of ones in $C^T =$ the number of ones in $C = \mathrm{rank}(A)$. □

Alternative terminology:

column rank of $A = \dim(\mathrm{span}(\text{ columns of } A)) = \mathrm{rank}(A)$

row rank of $A = \dim(\mathrm{span}(\text{ rows of } A)) = \dim(\mathrm{span}(\text{ cols of } A^T)) = \mathrm{rank}(A^T)$

With this terminology, the proposition says row rank=column rank.

Corollary 22.16: If A is an $n \times m$ matrix. Then $\mathrm{rank}(A) \leq \min(n, m)$.

Proof: Let V be the span of the columns of A, and let W be the span of the columns of A^T. Now V is a subspace of $\mathbb{R}^n$, so $\dim(V) \leq n$, but W is a subspace of $\mathbb{R}^m$, so $\dim(W) \leq m$. On the other hand, Proposition 22.14 tells us that $\dim(V) = \dim(W) = \mathrm{rank}(A)$, so we have $\mathrm{rank}(A) \leq n$ and also $\mathrm{rank}(A) \leq m$, so $\mathrm{rank}(A) \leq \min(n, m)$.

Lecture 19

Orthogonal matrices and orthonormal lists

Definition 23.1: Let A be an $n \times n$ matrix.
We say that A is an *orthogonal matrix* it is invertible and $A^{-1} = A^T$.

Definition 23.2: Let $v_1, \dots, v_r$ be a list of r vectors in $\mathbb{R}^n$.
We say that this list is *orthonormal* if $v_i.v_i = 1$ for all i,
and $v_i.v_j = 0$ whenever i and j are different.

Proposition 23.4: Any orthonormal list of length n in $\mathbb{R}^n$ is a basis.

Proof: Let $v_1, \dots, v_n$ be an orthonormal list of length n.
Suppose we have a linear relation $\sum_{i=1}^n \lambda_i v_i = 0$.
We can take the dot product of both sides with v_p to get $\sum_{i=1}^n \lambda_i (v_i.v_p) = 0$.
Most of the terms $v_i.v_p$ are zero, because $v_i.v_j = 0$ whenever $i \neq j$.
After dropping the terms where $i \neq p$, we are left with $\lambda_p (v_p.v_p) = 0$.
Here $v_p.v_p = 1$ (by the definition of orthonormality) so $\lambda_p = 0$.
This works for all p, so our linear relation is the trivial one.
This proves that the list $v_1, \dots, v_n$ is linearly independent.
A linearly independent list of n vectors in $\mathbb{R}^n$ is automatically a basis by Proposition 10.12. □

Orthogonal matrices and orthonormal lists

Proposition 23.5: Let A be an $n \times n$ matrix. Then A is an orthogonal matrix if and only if the columns of A form an orthonormal list.

Proof.
By definition, A is orthogonal if and only if A^T is an inverse for A, or in other words $A^T A = I_n$. Let the columns of A be $v_1, \dots, v_n$. Then

$$A^T A = \left[\begin{array}{c} v_1^T \\ \hline \vdots \\ \hline v_n^T \end{array}\right] \left[\begin{array}{c|c|c} v_1 & \cdots & v_n \end{array}\right] = \begin{bmatrix} v_1.v_1 & \cdots & v_1.v_n \\ \vdots & \ddots & \vdots \\ v_n.v_1 & \cdots & v_n.v_n \end{bmatrix}$$

In other words, the entry in the (i, j) position in $A^T A$ is just the dot product $v_i.v_j$. For $A^T A$ to be the identity we need the diagonal entries $v_i.v_i$ to be one, and the off-diagonal entries $v_i.v_j$ (with $i \neq j$) to be zero. This means precisely that the list $v_1, \dots, v_n$ is orthonormal. □

Symmetric matrices

Definition 23.6: Let A be an $n \times n$ matrix, with entries a_{ij}.
We say that A is *symmetric* if $A^T = A$,
or equivalently $a_{ij} = a_{ji}$ for all i and j.

Example: A 4×4 matrix is symmetric if and only if it has the form

$$\begin{bmatrix} a & b & c & d \\ b & e & f & g \\ c & f & h & i \\ d & g & i & j \end{bmatrix}.$$

Example: The matrices A and B are symmetric, but C and D are not.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 3 \\ 3 & 3 & 3 \end{bmatrix} \qquad B = \begin{bmatrix} 111 & 11 & 1 \\ 11 & 111 & 11 \\ 1 & 11 & 111 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \qquad D = \begin{bmatrix} 1 & 10 & 1000 \\ 1 & 10 & 1000 \\ 1 & 10 & 1000 \end{bmatrix}$$

Dot products and symmetric matrices

Lemma 23.9: Let A be an $n \times n$ matrix, and let u and v be vectors in $\mathbb{R}^n$.
Then $u.(Av) = (A^T u).v$. Thus, if A is symmetric then $u.(Av) = (Au).v$.

Proof.
Put $p = A^T u$ and $q = Av$, so the claim is that $u.q = p.v$.
By the definition of matrix multiplication, we have $q_i = \sum_j A_{ij} v_j$,
so $u.q = \sum_i u_i q_i = \sum_{i,j} u_i A_{ij} v_j$.
Similarly, we have $p_j = \sum_i (A^T)_{ji} u_i$, but $(A^T)_{ji} = A_{ij}$ so $p_j = \sum_i u_i A_{ij}$.
It follows that $p.v = \sum_j p_j v_j = \sum_{i,j} u_i A_{ij} v_j$,
which is the same as $u.q$, as claimed.

Alternatively: for $x, y \in \mathbb{R}^n$ the dot product $x.y$ is the matrix product $x^T y$.
Thus $(Au).v = (Au)^T v$, but $(Au)^T = u^T A^T$ (by Proposition 3.4)
so $(Au).v = u^T (A^T v) = u.(A^T v)$. $\square$

Eigenvalues of symmetric matrices

Proposition 23.10: Let A be an $n \times n$ symmetric matrix (with real entries).

(a) All eigenvalues of A are real numbers.

(b) If u and v are (real) eigenvectors for A with distinct eigenvalues, then u and v are orthogonal.

Proof of (a): Let $\lambda = \alpha + i\beta$ be a complex eigenvalue of A ($\alpha, \beta \in \mathbb{R}$).
We must show that $\beta = 0$, so that λ is actually a real number.
As λ is an eigenvalue, there is a nonzero vector u with $Au = \lambda u$.
Let $v, w \in \mathbb{R}^n$ be the real and imaginary parts of u, so $u = v + iw$.

$$Av + iAw = A(v + iw) = Au = \lambda u = (\alpha + i\beta)(v + iw) = (\alpha v - \beta w) + i(\beta v + \alpha w).$$

As the entries in A are real, we see that the vectors Av and Aw are real.
Compare real and imaginary parts to get

$$Av = \alpha v - \beta w \qquad Aw = \beta v + \alpha w$$

$$(Av).w = \alpha v.w - \beta w.w \qquad v.(Aw) = \beta v.v + \alpha v.w.$$

However, A is symmetric, so $(Av).w = v.(Aw)$ by Lemma 23.9.
Rearrange to get $\beta(v.v + w.w) = 0$ or $\beta(\|v\|^2 + \|w\|^2) = 0$.
By assumption $u \neq 0$ so ($v \neq 0$ or $w \neq 0$) so $\|v\|^2 + \|w\|^2 > 0$.
Divide by this to get $\beta = 0$ and $\lambda = \alpha \in \mathbb{R}$ as claimed.

Eigenvalues of symmetric matrices

Proposition 23.10: Let A be an $n \times n$ symmetric matrix (with real entries).

(a) All eigenvalues of A are real numbers.

(b) If u and v are (real) eigenvectors for A with distinct eigenvalues, then u and v are orthogonal.

Proof of (b): Suppose that u and v are eigenvectors of A with distinct eigenvalues, say λ and μ. This means that

$$Au = \lambda u \qquad Av = \mu v \qquad \lambda \neq \mu.$$

As A is symmetric we have $(Au).v = u.(Av)$.
As $Au = \lambda u$ and $Av = \mu v$ this becomes $\lambda\, u.v = \mu\, u.v$.
Rearrange to get $(\lambda - \mu) u.v = 0$.
As $\lambda \neq \mu$ we can divide by $\lambda - \mu$ to get $u.v = 0$,
which means that u and v are orthogonal. $\square$

Alternative proof for 2 × 2 matrices

A 2×2 symmetric matrix has the form

$$A = \begin{bmatrix} a & b \\ b & d \end{bmatrix} \qquad \text{so} \qquad A - tI_2 = \begin{bmatrix} a-t & b \\ b & d-t \end{bmatrix}$$

so

$$\chi_A(t) = (a-t)(d-t) - b^2 = t^2 - (a+d)t + (ad - b^2).$$

The eigenvalues are

$$\lambda = \frac{a + d \pm \sqrt{(a+d)^2 - 4(ad - b^2)}}{2}.$$

The expression under the square root is

$$\begin{aligned} (a+d)^2 - 4(ad - b^2) &= a^2 + 2ad + d^2 - 4ad + 4b^2 \\ &= a^2 - 2ad + d^2 + 4b^2 \\ &= (a-d)^2 + (2b)^2. \end{aligned}$$

This is the sum of two squares, so it is nonnegative, so the square root is real, so the two eigenvalues are both real.

Orthonormal basis of eigenvectors

Proposition 23.12: Let A be an $n \times n$ symmetric matrix. Then there is an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors for A.

Partial proof.
We will show that the Theorem holds whenever A has n distinct eigenvalues. In fact it is true even without that assumption, but the proof is harder.

Let the eigenvalues of A be $\lambda_1, \dots, \lambda_n$ (so $\lambda_i \in \mathbb{R}$).
For each i we choose a (real) eigenvector u_i of eigenvalue λ_i.
As u_i is an eigenvector we have $u_i \neq 0$ and so $u_i.u_i > 0$
so we can define $v_i = u_i/\sqrt{u_i.u_i}$. This is just a real number times u_i, so it is again an eigenvector of eigenvalue λ_i.
It satisfies $v_i.v_i = \dfrac{u_i.u_i}{\sqrt{u_i.u_i}\,\sqrt{u_i.u_i}} = 1$ (so it is a unit vector).
Proposition 23.10(b): eigenvectors of a symmetric matrix with distinct eigenvalues are orthogonal. Thus $v_i.v_j = 0$ for $i \neq j$.
This shows that the sequence $v_1, \dots, v_n$ is orthonormal.
Proposition 23.4: any orthonormal list of length n in $\mathbb{R}^n$ is a basis.
Proposition 13.22: any n eigenvectors in $\mathbb{R}^n$ with distinct eigenvalues form a basis.
Either of these results implies that $v_1, \dots, v_n$ is a basis. □

Our special case is the usual case

Let A be an $n \times n$ symmetric matrix again.
The characteristic polynomial $\chi_A(t)$ has degree n,
so by well-known properties of polynomials it can be factored as

$$\chi_A(t) = \prod_{i=1}^{n} (\lambda_i - t)$$

for some complex numbers $\lambda_1, \dots, \lambda_n$.

By Proposition 23.10(a) these eigenvalues λ_i are in fact all real.

Some of them might be the same, but that would be a concidence which could only happen if the matrix A was very simple or had some kind of hidden symmetry.

Thus, our proof of Proposition 23.12 covers almost all cases
(but some of the cases that are not covered are the most interesting ones).

Orthonormal eigenvector example

Consider the symmetric matrix $A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$
(which appeared on one of the problem sheets) and the vectors

$$u_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad u_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} \quad u_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix} \quad u_4 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ -1 \end{bmatrix} \quad u_5 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

These satisfy $Au_1 = Au_2 = Au_3 = Au_4 = 0$ and $Au_5 = 5u_5$,
so they are eigenvectors of eigenvalues $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 0$ and $\lambda_5 = 5$.
Because λ_5 is different from $\lambda_1, \dots, \lambda_4$, Proposition 23.10(b) tells us that u_5 must be orthogonal to $u_1, \dots, u_4$, and indeed it is easy to see directly that $u_1.u_5 = \cdots = u_4.u_5 = 0$. However, the eigenvectors $u_1, \dots, u_4$ all share the same eigenvalue so there is no reason for them to be orthogonal and in fact they are not.

Orthonormal eigenvector example

$$u_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad u_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} \quad u_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \\ 0 \end{bmatrix} \quad u_4 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ -1 \end{bmatrix} \quad u_5 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

These satisfy $Au_1 = Au_2 = Au_3 = Au_4 = 0$ and $Au_5 = 5u_5$.

The vectors $u_1, \dots, u_4$ are not orthogonal:

$$u_1.u_2 = u_1.u_3 = u_1.u_4 = u_2.u_3 = u_2.u_4 = u_3.u_4 = 1.$$

However, it is possible to choose a different basis of eigenvectors where all the eigenvectors are orthogonal to each other. One such choice is as follows:

$$v_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad v_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \\ 0 \\ 0 \end{bmatrix} \quad v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -3 \\ 0 \end{bmatrix} \quad v_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -4 \end{bmatrix} \quad v_5 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

It is easy to check directly that $Av_1 = Av_2 = Av_3 = Av_4 = 0 \qquad Av_5 = 5v_5$

$$v_1.v_2 = v_1.v_3 = v_1.v_4 = v_1.v_5 = v_2.v_3 = v_2.v_4 = v_2.v_5 = v_3.v_4 = v_3.v_5 = v_4.v_5 = 0,$$

so the v_i are eigenvectors and are orthogonal to each other.

Orthonormal eigenvector example

$$v_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad v_2 = \begin{bmatrix} 1 \\ 1 \\ -2 \\ 0 \\ 0 \end{bmatrix} \quad v_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -3 \\ 0 \end{bmatrix} \quad v_4 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -4 \end{bmatrix} \quad v_5 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

These satisfy $Av_1 = Av_2 = Av_3 = Av_4 = 0$ and $Av_5 = 5v_5$, and they are orthogonal to each other.

However, the list $v_1, \dots, v_5$ is not orthonormal, because

$$v_1.v_1 = 2 \quad v_2.v_2 = 6 \quad v_3.v_3 = 12 \quad v_4.v_4 = 20 \quad v_5.v_5 = 5.$$

This is easily fixed: if we put

$$w_1 = \frac{v_1}{\sqrt{2}} \quad w_2 = \frac{v_2}{\sqrt{6}} \quad w_3 = \frac{v_3}{\sqrt{12}} \quad w_4 = \frac{v_4}{\sqrt{20}} \quad w_5 = \frac{v_5}{\sqrt{5}}$$

then $w_1, \dots, w_5$ is an orthonormal basis for $\mathbb{R}^5$ consisting of eigenvectors for A.

Lecture 20

Orthogonal diagonalisation of symmetric matrices

Corollary 23.15: Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal matrix U and a diagonal matrix D such that $A = UDU^T = UDU^{-1}$.

Proof.
Choose an orthonormal basis of eigenvectors $u_1, \dots, u_n$, and let λ_i be the eigenvalue of u_i.

Put $U = [u_1 | \cdots | u_n]$ and $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$.

Proposition 14.4 tells us that $U^{-1}AU = D$ and so $A = UDU^{-1}$.

Proposition 23.5 tells us that U is an orthogonal matrix, so $U^{-1} = U^T$. $\square$

Example of orthogonal diagonalisation

Let A be the 5×5 matrix in which every entry is one, as in Example 23.14. Following the prescription in the above proof, we put

$$U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{12} & 1/\sqrt{20} & 1/\sqrt{5} \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{12} & 1/\sqrt{20} & 1/\sqrt{5} \\ 0 & -2/\sqrt{6} & 1/\sqrt{12} & 1/\sqrt{20} & 1/\sqrt{5} \\ 0 & 0 & -3/\sqrt{12} & 1/\sqrt{20} & 1/\sqrt{5} \\ 0 & 0 & 0 & -4/\sqrt{20} & 1/\sqrt{5} \end{bmatrix} \qquad D = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix}$$

The general theory tells us that $A = UDU^T$. We can check this directly:

$$UD = \begin{bmatrix} * & * & * & * & 1/\sqrt{5} \\ * & * & * & * & 1/\sqrt{5} \\ * & * & * & * & 1/\sqrt{5} \\ * & * & * & * & 1/\sqrt{5} \\ * & * & * & * & 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & \sqrt{5} \\ 0 & 0 & 0 & 0 & \sqrt{5} \\ 0 & 0 & 0 & 0 & \sqrt{5} \\ 0 & 0 & 0 & 0 & \sqrt{5} \\ 0 & 0 & 0 & 0 & \sqrt{5} \end{bmatrix}$$

$$UDU^T = \begin{bmatrix} 0 & 0 & 0 & 0 & \sqrt{5} \\ 0 & 0 & 0 & 0 & \sqrt{5} \\ 0 & 0 & 0 & 0 & \sqrt{5} \\ 0 & 0 & 0 & 0 & \sqrt{5} \\ 0 & 0 & 0 & 0 & \sqrt{5} \end{bmatrix} \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} = A.$$

Example of orthogonal diagonalisation

Write $\rho = \sqrt{3}$ for brevity (so $\rho^2 = 3$), and consider the symmetric matrix

$$A = \begin{bmatrix} 0 & 1 & \rho \\ 1 & 0 & -\rho \\ \rho & -\rho & 0 \end{bmatrix}.$$

The characteristic polynomial is

$$\begin{aligned} \chi_A(t) &= \det \begin{bmatrix} -t & 1 & \rho \\ 1 & -t & -\rho \\ \rho & -\rho & -t \end{bmatrix} \\ &= -t \det \begin{bmatrix} -t & -\rho \\ -\rho & -t \end{bmatrix} - \det \begin{bmatrix} 1 & -\rho \\ \rho & -t \end{bmatrix} + \rho \det \begin{bmatrix} 1 & -t \\ \rho & -\rho \end{bmatrix} \\ &= -t(t^2 - \rho^2) - (-t + \rho^2) + \rho(-\rho + t\rho) = -t^3 + 3t + t - 3 - 3 + 3t \\ &= -t^3 + 7t - 6 = -(t-1)(t-2)(t+3). \end{aligned}$$

It follows that the eigenvalues are $\lambda_1 = 1$, $\lambda_2 = 2$ and $\lambda_3 = -3$.

Example of orthogonal diagonalisation

$$\rho = \sqrt{3} \qquad A = \begin{bmatrix} 0 & 1 & \rho \\ 1 & 0 & -\rho \\ \rho & -\rho & 0 \end{bmatrix} \qquad \begin{aligned} \lambda_1 &= 1 \\ \lambda_2 &= 2 \\ \lambda_3 &= -3. \end{aligned}$$

Eigenvectors can be found by row-reduction:

$$A - I = \begin{bmatrix} -1 & 1 & \rho \\ 1 & -1 & -\rho \\ \rho & -\rho & -1 \end{bmatrix} \to \begin{bmatrix} 1 & -1 & -\rho \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$A - 2I = \begin{bmatrix} -2 & 1 & \rho \\ 1 & -2 & -\rho \\ \rho & -\rho & -2 \end{bmatrix} \to \begin{bmatrix} 1 & -2 & -\rho \\ 0 & -3 & -\rho \\ 0 & \rho & 1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -\rho/3 \\ 0 & 1 & \rho/3 \\ 0 & 0 & 0 \end{bmatrix}$$

$$A + 3I = \begin{bmatrix} 3 & 1 & \rho \\ 1 & 3 & -\rho \\ \rho & -\rho & 3 \end{bmatrix} \to \begin{bmatrix} 1 & 3 & -\rho \\ 0 & -8 & 4\rho \\ 0 & -4\rho & 6 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & \rho/2 \\ 0 & 1 & -\rho/2 \\ 0 & 0 & 0 \end{bmatrix}$$

From this we can read off the following eigenvectors:

$$u_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \qquad u_2 = \begin{bmatrix} \rho/3 \\ -\rho/3 \\ 1 \end{bmatrix} \qquad u_3 = \begin{bmatrix} -\rho/2 \\ \rho/2 \\ 1 \end{bmatrix}.$$

Example of orthogonal diagonalisation

$$\begin{aligned} \lambda_1 &= 1 \\ \lambda_2 &= 2 \\ \lambda_3 &= -3 \end{aligned} \qquad u_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \qquad u_2 = \begin{bmatrix} \rho/3 \\ -\rho/3 \\ 1 \end{bmatrix} \qquad u_3 = \begin{bmatrix} -\rho/2 \\ \rho/2 \\ 1 \end{bmatrix}$$

Because the matrix A is symmetric and the eigenvalues are distinct, it is automatic that the eigenvectors u_i are orthogonal to each other. However, they are not normalised: instead we have

$$\begin{aligned} u_1.u_1 &= 1^2 + 1^2 = 2 \\ u_2.u_2 &= (\rho/3)^2 + (-\rho/3)^2 + 1^2 = 1/3 + 1/3 + 1 = 5/3 \\ u_3.u_3 &= (-\rho/2)^2 + (\rho/2)^2 + 1^2 = 3/4 + 3/4 + 1 = 5/2. \end{aligned}$$

The vectors $v_i = u_i/\sqrt{u_i.u_i}$ form an orthonormal basis of eigenvectors. Explicitly, this works out as follows:

$$v_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \qquad v_2 = \begin{bmatrix} 1/\sqrt{5} \\ -1/\sqrt{5} \\ \sqrt{3/5} \end{bmatrix} \qquad v_3 = \begin{bmatrix} -\sqrt{3/10} \\ \sqrt{3/10} \\ \sqrt{2/5} \end{bmatrix}.$$

Example of orthogonal diagonalisation

$$\begin{aligned}\lambda_1 &= 1\\ \lambda_2 &= 2\\ \lambda_3 &= -3\end{aligned} \qquad v_1 = \begin{bmatrix}1/\sqrt{2}\\1/\sqrt{2}\\0\end{bmatrix} \qquad v_2 = \begin{bmatrix}1/\sqrt{5}\\-1/\sqrt{5}\\\sqrt{3/5}\end{bmatrix} \qquad v_3 = \begin{bmatrix}-\sqrt{3/10}\\\sqrt{3/10}\\\sqrt{2/5}\end{bmatrix}.$$

The eigenvectors v_i form orthonormal basis for $\mathbb{R}^3$.

It follows that if we put

$$U = \begin{bmatrix}1/\sqrt{2} & 1/\sqrt{5} & -\sqrt{3/10}\\ 1/\sqrt{2} & -1/\sqrt{5} & \sqrt{3/10}\\ 0 & \sqrt{3/5} & \sqrt{2/5}\end{bmatrix} \qquad D = \begin{bmatrix}1 & 0 & 0\\0 & 2 & 0\\0 & 0 & -3\end{bmatrix}$$

then U is an orthogonal matrix and $A = UDU^T$.

Square roots of positive matrices

Corollary 23.18: Let A be an $n \times n$ real symmetric matrix, and suppose that all the eigenvalues of A are positive.
Then there is a real symmetric matrix B such that $A = B^2$.

Proof.
Choose an orthonormal basis of eigenvectors $u_1, \dots, u_n$, and let λ_i be the eigenvalue of u_i.
Put $U = [u_1|\cdots|u_n]$ and $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$.
We saw in Corollary 23.15 that U is orthogonal (so $U^TU = I = UU^T$) and that $A = UDU^T$.

By assumption the eigenvectors λ_i are positive, so we have a real diagonal matrix $E = \mathrm{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n})$. Put $B = UEU^T$. It is clear that $E^T = E$, and it follows that

$$B^T = (UEU^T)^T = U^{TT}E^TU^T = UEU^T = B.$$

We also have

$$B^2 = UEU^TUEU^T = UEEU^T = UDU^T = A.$$

□

Linear and quadratic forms

Definition 23.19:

(a) A *linear form* on $\mathbb{R}^n$ is a function of the form $L(x) = \sum_{i=1}^n a_ix_i$ (for some constants $a_1, \dots, a_n$).

(b) A *quadratic form* on $\mathbb{R}^n$ is a function of the form $Q(x) = \sum_{i=1}^n \sum_{j=1}^n b_{ij}x_ix_j$ (for some constants b_{ij}).

Example 23.20:

(a) We can define a linear form on $\mathbb{R}^3$ by $L(x) = 7x_1 - 8x_2 + 9x_3$.

(b) We can define a quadratic form on $\mathbb{R}^4$ by $Q(x) = 10x_1x_2 + 12x_3x_4 - 14x_1x_4 - 16x_2x_3$.

Given a linear form $L(x) = \sum_i a_ix_i$, we can form the vector $a = \begin{bmatrix}a_1 & \cdots & a_n\end{bmatrix}^T$, and clearly $L(x) = a.x = a^Tx$.

Symmetric expressions for quadratic forms

Consider a quadratic form $Q(x) = \sum_{i,j} b_{ij}x_ix_j$.
Form the matrix B with entries b_{ij}: we find that $Q(x) = x^TBx$.
For example, if $n = 2$ and $Q(x) = 2x_1^2 + 4x_1x_2 + 7x_2^2$ then $B = \begin{bmatrix}2 & 4\\0 & 7\end{bmatrix}$ and

$$x^TBx = \begin{bmatrix}x_1 & x_2\end{bmatrix}\begin{bmatrix}2 & 4\\0 & 7\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}x_1 & x_2\end{bmatrix}\begin{bmatrix}2x_1 + 4x_2\\7x_2\end{bmatrix} = 2x_1^2 + 4x_1x_2 + 7x_2^2 = Q(x).$$

Alternatively, note that $x_1x_2 = x_2x_1$:

(a) Rewriting the same $Q(x)$ as $2x_1^2 + 3x_1x_2 + 1x_2x_1 + 7x_2^2$ gives $B = \begin{bmatrix}2 & 3\\1 & 7\end{bmatrix}$.

(b) Rewriting the same $Q(x)$ as $2x_1^2 + 2x_1x_2 + 2x_2x_1 + 7x_2^2$ gives $B = \begin{bmatrix}2 & 2\\2 & 7\end{bmatrix}$.

(c) Rewriting the same $Q(x)$ as $2x_1^2 + 1x_1x_2 + 3x_2x_1 + 7x_2^2$ gives $B = \begin{bmatrix}2 & 1\\3 & 7\end{bmatrix}$.

In option (b) we "share the coefficient equally" between x_1x_2 and x_2x_1, so the matrix B is symmetric. This is the preferred option.
We can do the same for any quadratic form.

Symmetric expressions for quadratic forms

For example, we considered above the quadratic form

$$Q(x) = 10x_1x_2 + 12x_3x_4 - 14x_1x_4 - 16x_2x_3.$$

This can be rewritten symmetrically as

$$Q(x) = 5x_1x_2 + 5x_2x_1 + 6x_3x_4 + 6x_4x_3 - 7x_1x_4 - 7x_4x_1 - 8x_2x_3 - 8x_3x_2,$$

which corresponds to the symmetric matrix

$$B = \begin{bmatrix} 0 & 5 & 0 & -7 \\ 5 & 0 & -8 & 0 \\ 0 & -8 & 0 & 6 \\ -7 & 0 & 6 & 0 \end{bmatrix}$$

$$x^TBx = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix} \begin{bmatrix} 0 & 5 & 0 & -7 \\ 5 & 0 & -8 & 0 \\ 0 & -8 & 0 & 6 \\ -7 & 0 & 6 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

$$= \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix} \begin{bmatrix} 5x_2 - 7x_4 \\ 5x_1 - 8x_3 \\ -8x_2 + 6x_4 \\ -7x_1 + 6x_3 \end{bmatrix} = Q(x).$$

Diagonalisation of quadratic forms

Proposition 23.23: Let $Q(x)$ be a quadratic form on $\mathbb{R}^n$. Then there are integers $r, s \geq 0$ and nonzero vectors $v_1, \dots, v_r, w_1, \dots, w_s$ such that all the v's and w's are orthogonal to each other, and

$$Q(x) = (x.v_1)^2 + \cdots + (x.v_r)^2 - (x.w_1)^2 - \cdots - (x.w_s)^2.$$

Or in terms of linear forms $L_i(x) = x.v_i$ and $M_j(x) = x.w_j$:

$$Q = L_1{}^2 + \cdots + L_r{}^2 - M_1{}^2 - \cdots - M_s{}^2.$$

The *rank* of Q is defined to be $r + s$, and the *signature* is defined to be $r - s$.

Proof: There is a symmetric matrix B such that $Q(x) = x^TBx$.
By Proposition 23.12, we can find an orthonormal basis $u_1, \dots, u_n$ for $\mathbb{R}^n$ such that each u_i is an eigenvector for B, with eigenvalue $\lambda_i \in \mathbb{R}$ say.
Let r be the number of indices i for which $\lambda_i > 0$,
and let s be the number of indices i for which $\lambda_i < 0$.
We can assume that things have been ordered such that $\lambda_1, \dots, \lambda_r > 0$ and $\lambda_{r+1}, \dots, \lambda_{r+s} < 0$ and any eigenvalues after λ_{r+s} are zero.
Now put $U = [u_1|\cdots|u_n]$ and $D = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$.
We have seen that $B = UDU^T$, so

$$Q(x) = x^TBx = x^TUDU^Tx = (U^Tx)^T(DU^Tx) = (U^Tx).(DU^Tx).$$

Diagonalisation of quadratic forms

$$U = [u_1|\cdots|u_n] \qquad D = \mathrm{diag}(\lambda_1, \dots, \lambda_n) \qquad Q(x) = (U^Tx).(DU^Tx)$$

$$U^Tx = \begin{bmatrix} u_1^T \\ \hline \vdots \\ \hline u_n^T \end{bmatrix} x = \begin{bmatrix} u_1.x \\ \vdots \\ u_n.x \end{bmatrix}$$

$$DU^Tx = \mathrm{diag}(\lambda_1, \dots, \lambda_n) \begin{bmatrix} u_1.x \\ \vdots \\ u_n.x \end{bmatrix} = \begin{bmatrix} \lambda_1 u_1.x \\ \vdots \\ \lambda_n u_n.x \end{bmatrix}$$

$$Q(x) = (U^Tx).(DU^Tx) = \begin{bmatrix} u_1.x \\ \vdots \\ u_n.x \end{bmatrix} . \begin{bmatrix} \lambda_1 u_1.x \\ \vdots \\ \lambda_n u_n.x \end{bmatrix} = \lambda_1(u_1.x)^2 + \cdots + \lambda_n(u_n.x)^2.$$

Diagonalisation of quadratic forms

$$U = [u_1|\cdots|u_n] \qquad D = \mathrm{diag}(\lambda_1, \dots, \lambda_n) \qquad Q(x) = \lambda_1(u_1.x)^2 + \cdots + \lambda_n(u_n.x)^2.$$

$$\lambda_1, \dots, \lambda_r > 0 \qquad \lambda_{r+1}, \dots, \lambda_{r+s} < 0 \qquad \lambda_{r+s+1}, \dots, \lambda_n = 0$$

- For $1 \leq i \leq r$ we have $\lambda_i > 0$ and we put $v_i = \sqrt{\lambda_i}u_i$ so $\lambda_i(u_i.x)^2 = (v_i.x)^2$.
- For $r + 1 \leq i \leq r + s$ we have $\lambda_i < 0$ and we put $w_{i-r} = \sqrt{|\lambda_i|}u_i$ so $\lambda_i(u_i.x)^2 = -(w_{i-r}.x)^2$.
- For $i > r + s$ we have $\lambda_i = 0$ and $\lambda_i(u_i.x)^2 = 0$.

We thus have

$$Q(x) = (x.v_1)^2 + \cdots + (x.v_r)^2 - (x.w_1)^2 - \cdots - (x.w_s)^2$$

as required.

Example of diagonalising a quadratic form

Consider the quadratic form $Q(x) = x_1x_2 - x_3x_4$ on $\mathbb{R}^4$.
It is elementary that for all $a, b \in \mathbb{R}$ we have

$$\left(\frac{a+b}{2}\right)^2 - \left(\frac{a-b}{2}\right)^2 = \frac{a^2+2ab+b^2-a^2+2ab-b^2}{4} = ab.$$

Using this, we can rewrite $Q(x)$ as

$$Q(x) = \left(\frac{x_1+x_2}{2}\right)^2 - \left(\frac{x_1-x_2}{2}\right)^2 - \left(\frac{x_3+x_4}{2}\right)^2 + \left(\frac{x_3-x_4}{2}\right)^2.$$

Now put

$$v_1 = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \\ 0 \end{bmatrix} \qquad v_2 = \begin{bmatrix} 0 \\ 0 \\ 1/2 \\ -1/2 \end{bmatrix} \qquad w_1 = \begin{bmatrix} 1/2 \\ -1/2 \\ 0 \\ 0 \end{bmatrix} \qquad w_2 = \begin{bmatrix} 0 \\ 0 \\ 1/2 \\ 1/2 \end{bmatrix}.$$

We then have

$$Q(x) = (x.v_1)^2 + (x.v_2)^2 - (x.w_1)^2 - (x.w_2)^2$$

and it is easy to see that the v's and w's are all orthogonal.

Example of diagonalising a quadratic form

Consider the quadratic form $Q(x) = 4x_1x_2 + 6x_2x_3 + 4x_3x_4$ on $\mathbb{R}^4$.
Rewritten symmetrically: $Q(x) = 2x_1x_2 + 2x_2x_1 + 3x_2x_3 + 3x_3x_2 + 2x_3x_4 + 2x_4x_3$.

Corresponding matrix: $B = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 2 & 0 & 3 & 0 \\ 0 & 3 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{bmatrix}$. Characteristic polynomial:

$$\det \begin{bmatrix} -t & 2 & 0 & 0 \\ 2 & -t & 3 & 0 \\ 0 & 3 & -t & 2 \\ 0 & 0 & 2 & -t \end{bmatrix} = -t \det \begin{bmatrix} -t & 3 & 0 \\ 3 & -t & 2 \\ 0 & 2 & -t \end{bmatrix} - 2 \det \begin{bmatrix} 2 & 3 & 0 \\ 0 & -t & 2 \\ 0 & 2 & -t \end{bmatrix}$$

$$\det \begin{bmatrix} -t & 3 & 0 \\ 3 & -t & 2 \\ 0 & 2 & -t \end{bmatrix} = -t(t^2-4) - 3(-3t) = 13t - t^3$$

$$\det \begin{bmatrix} 2 & 3 & 0 \\ 0 & -t & 2 \\ 0 & t & -2 \end{bmatrix} = 2(t^2-4) - 3(0-0) = 2t^2 - 8$$

$$\begin{aligned} \chi_B(t) &= -t(13t - t^3) - 2(2t^2 - 8) = t^4 - 17t^2 + 16 \\ &= (t^2-1)(t^2-16) = (t-1)(t+1)(t-4)(t+4). \end{aligned}$$

Example of diagonalising a quadratic form

$$Q(x) = x^T Bx \qquad B = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 2 & 0 & 3 & 0 \\ 0 & 3 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{bmatrix} \qquad \chi_B(t) = (t-1)(t+1)(t-4)(t+4)$$

Eigenvalues: $\lambda_1 = 1$, $\lambda_2 = 4$, $\lambda_3 = -1$ and $\lambda_4 = -4$.
Row-reduce $B - \lambda_i I$ to find the eigenvectors:

$$t_1 = \begin{bmatrix} 2 \\ 1 \\ -1 \\ -2 \end{bmatrix} \qquad t_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix} \qquad t_3 = \begin{bmatrix} 2 \\ -1 \\ -1 \\ 2 \end{bmatrix} \qquad t_4 = \begin{bmatrix} 1 \\ -2 \\ 2 \\ -1 \end{bmatrix}$$

In each case we see that $t_i.t_i = 10$ so the corresponding orthonormal basis consists of the vectors $u_i = t_i/\sqrt{10}$. Following Proposition 23.23:

$$\begin{aligned} v_1 &= \sqrt{\lambda_1}u_1 &&= t_1/\sqrt{10} &&= \sqrt{1/10}\,[2 \quad 1 \quad -1 \quad -2]^T \\ v_2 &= \sqrt{\lambda_2}u_2 &&= \sqrt{4}t_2/\sqrt{10} &&= \sqrt{2/5}\,[1 \quad 2 \quad 2 \quad 1]^T \\ w_1 &= \sqrt{|\lambda_3|}u_3 &&= t_3/\sqrt{10} &&= \sqrt{1/10}\,[2 \quad -1 \quad -1 \quad 2]^T \\ w_2 &= \sqrt{|\lambda_4|}u_4 &&= \sqrt{4}t_4/\sqrt{10} &&= \sqrt{2/5}\,[1 \quad -2 \quad 2 \quad -1]^T \end{aligned}$$

Example of diagonalising a quadratic form

$$Q(x) = 4x_1x_2 + 6x_2x_3 + 4x_3x_4$$

$$v_1 = \sqrt{\frac{1}{10}} \begin{bmatrix} 2 \\ 1 \\ -1 \\ -2 \end{bmatrix} \qquad v_2 = \sqrt{\frac{2}{5}} \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \end{bmatrix} \qquad w_1 = \sqrt{\frac{1}{10}} \begin{bmatrix} 2 \\ -1 \\ -1 \\ 2 \end{bmatrix} \qquad w_2 = \sqrt{\frac{2}{5}} \begin{bmatrix} 1 \\ -2 \\ 2 \\ -1 \end{bmatrix}$$

Conclusion: $Q(x) = (x.v_1)^2 + (x.v_2)^2 - (x.w_1)^2 - (x.w_2)^2$.