MA 314: Linear Algebra

Dr Padraig Ó Catháin

April 4, 2012

National University of Ireland, Galway

Contents

Introduction

These notes are based on a course I taught in NUI Galway, from January to March 2012. The students were predominantly final year arts students. This was their third course in Linear Algebra, so familiarity with basic concepts is taken for granted. The course is largely a methods course, so while some results are proved, many are given as facts. The emphasis is on problem solving and on computation.

The notes are based on material prepared by David O'Keeffe and Ted Hurley, who taught the course in previous years. All of the material has been rewritten from scratch. The major changes are a new approach to the section on symmetric and orthogonal matrices, the coverage of the assignment problem as an introduction to optimization, and a comprehensive selection of exercises.

Topics covered

- Inner products, lengths and angles
- Orthogonality, orthonormal bases and the Gram-Schmidt process, applications
- Symmetric and orthogonal matrices, diagonalisation, applications
- An introduction to linear programming and the simplex method

Assessment

- Three homework assignments, each worth 10% of the final mark.
- One two hour exam, four questions, one on each topic. Full marks for solutions to 3 questions, worth 70% of the final mark.

Recommended texts

- These lecture notes.
- *Linear Algebra* by Fraleigh and Beauregard
- *Elementary Linear Algebra: Applications version* by Anton and Rorres

1 Inner products

1.1 Inner products

Definition 1. An *inner product* on a vector space over $\mathbb{R}$ is a mapping $V \times V \mapsto \mathbb{R}$ with the following properties:

1. $\langle v, w\rangle = \langle w, v\rangle$ for all $v, w \in V$.
2. $\langle v, u + w\rangle = \langle v, u\rangle + \langle v, w\rangle$ for all $u, v, w \in V$.
3. $\langle v + u, w\rangle = \langle v, w\rangle + \langle u, w\rangle$ for all $u, v, w \in V$.
4. $\langle \alpha v, w\rangle = \langle v, \alpha w\rangle = \alpha\langle v, w\rangle$ for all $v, w \in V$ and all $\alpha \in \mathbb{R}$.
5. $\langle v, v\rangle \geq 0$ for all $v \in V$, and $\langle v, v\rangle = 0$ if and only if v is the zero vector.

A vector space together with an inner product is an *inner product space.*

Example 2. The familiar dot product on $\mathbb{R}^2$ is an inner product. We use the standard $\bullet$ notation.

$$(u_1, u_2) \bullet (v_1, v_2) = u_1v_1 + u_2v_2.$$

Example 3. A less familiar example is the set of all polynomial functions with real coefficients defined on the interval $[0, 1]$. The inner product is given by

$$\langle f(x), g(x)\rangle = \int_0^1 f(x)g(x)\,\mathrm{dx}\,.$$

Exercise 1. 1. Verify that both of the examples satisfy the axioms of an inner product.

2. Show that the dot product on $\mathbb{R}^n$ given by $u \bullet v = \sum_{i=1}^{n} u_iv_i$ is an inner product.

3. Show that $\langle f(x), g(x)\rangle = \int_0^1 f(x)g(x)\,\mathrm{dx}$ is an inner product on the space of all integrable functions for which $\int_0^1 f(x)\,\mathrm{dx}$ is finite.

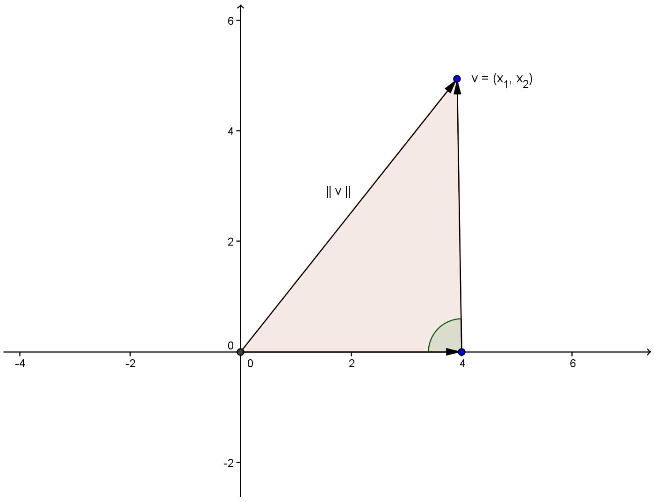
1.2 Length and angle

The dot product allows us to measure distances and angles in $\mathbb{R}^2$. We show how this is done, and then show how it can be generalised to any inner product space.

Definition 4. Let u, v be vectors in $\mathbb{R}^2$.

- the length of v is $\|v\| = \sqrt{v \bullet v}$
- the angle θ between u and v is given by $u \bullet v = \|u\| \, \|v\| \operatorname{Cos}(\theta)$

Geometrically, the formula for length is a consequence of Pythagoras' Theorem.



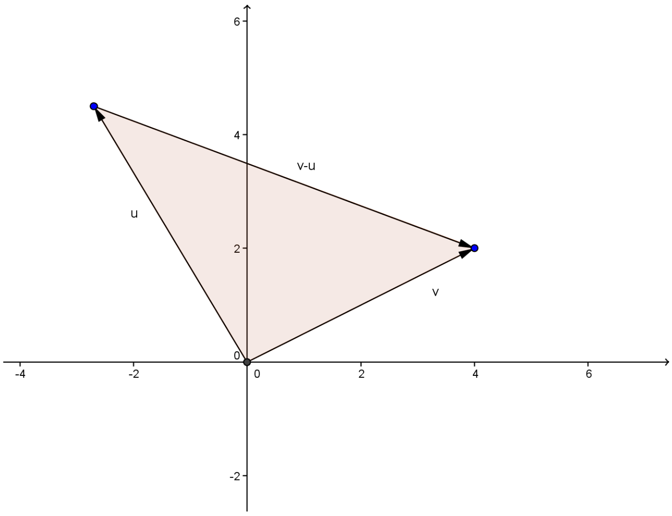
$$\begin{aligned} \|v\|^2 &= x_1^2 + x_2^2 \\ &= v \bullet v \end{aligned}$$

Let a, b, c be the sides of a triangle in $\mathbb{R}^2$, and let θ be the angle which is opposite a. Recall that the Cosine Rule states that:

$$a^2 = b^2 + c^2 - 2bc \operatorname{Cos}(\theta)$$

Or equivalently, when b and c are non-zero:

$$\operatorname{Cos}(\theta) = \frac{b^2 + c^2 - a^2}{2bc}.$$



Lemma 5. *For any non-zero vectors u and v, the angle θ between them is given by*

$$\mathrm{Cos}(\theta) = \frac{u \bullet v}{\|u\| \, \|v\|}$$

Proof. From the Cosine rule, the angle between u and v is

$$\mathrm{Cos}(\theta) = \frac{\|u\|^2 + \|v\|^2 - \|v-u\|^2}{2\,\|u\| \, \|v\|}.$$

First, we rearrange $\|v-u\|^2$.

$$\begin{aligned} \|v-u\|^2 &= (v-u) \bullet (v-u) \\ &= v \bullet (v-u) - u \bullet (v-u) \\ &= v \bullet v - v \bullet u - u \bullet v + u \bullet u \\ &= \|v\|^2 + \|u\|^2 - 2u \bullet v \end{aligned}$$

We substitute this into the Cosine rule, to obtain

$$\begin{aligned} \mathrm{Cos}(\theta) &= \frac{\|u\|^2 + \|v\|^2 - \|v-u\|^2}{2\,\|u\| \, \|v\|} \\ &= \frac{\|u\|^2 + \|v\|^2 - \Big(\|v\|^2 + \|u\|^2 - 2u \bullet v\Big)}{2\,\|u\| \, \|v\|} \\ &= \frac{u \bullet v}{\|u\| \, \|v\|} \end{aligned}$$

□

We give an example in the inner product space of polynomial functions.

Example 6. We find the angle between $f(x) = x + 1$ and $g(x) = x^2$.

- From the formula for the inner product, $\mathrm{Cos}(\theta) = \frac{\langle f(x), g(x)\rangle}{\|f(x)\|\|g(x)\|}$.
- So we calculate each of these inner products.

$$\begin{aligned}\langle f(x), g(x)\rangle &= \int_0^1 (x+1)x^2\,\mathrm{dx}\\ &= \int_0^1 x^3 + x^2\,\mathrm{dx}\\ &= \frac{1}{4}x^4 + \frac{1}{3}x^3 \mid_0^1\\ &= \frac{7}{12}\end{aligned}$$

- Similarly, we calculate $\|f(x)\| = \langle f(x), f(x)\rangle^{\frac{1}{2}}$ and $\|g(x)\|$, finding these to be $\sqrt{\frac{7}{3}}$ and $\sqrt{\frac{1}{5}}$ respectively. So we find that

$$\mathrm{Cos}(\theta) = \frac{7}{12\sqrt{\frac{7}{3}}\sqrt{\frac{1}{5}}}.$$

Using log tables or otherwise, we find θ to be approximately 31°.

Definition 7. The vectors u, v are *orthogonal* if and only if $u \bullet v = 0$.

1.3 Applications

As an application, we show how to use inner products to find the minimum distance between a line and the origin.

Example 8. We wish to find the shortest distance between the line $3x + 2y = 6$ and the point $(0, 0)$.

- We place the line in parametric form: $y = 3 - \frac{3}{2}x$.
- So every point on the line is of the form $(t, 3 - \frac{3}{2}t)$.
- The distance of $(t, 3 - \frac{3}{2}t)$ from the origin is given by $\left\|(t, 3 - \frac{3}{2}t)\right\|$.
- Multiplying out, this is $\sqrt{t^2 + (3 - \frac{3}{2}t)^2}$.
- We minimize the distance by taking the first derivative of this function, and solving for $f'(t) = 0$.
- This will give the local minmium of the function as $(\frac{18}{13}, \frac{12}{13})$.

As a more complicated application, we show how to find a vector perpendicular to a plane in $\mathbb{R}^3$.

Example 9. We wish to find a vector of length 1 perpendicular to the plane P generated by $x = (1, 2, 3)$ and $y = (-1, 0, 1)$.

- A vector v is perpendicular to P if and only if $v \bullet x = 0$ and $v \bullet y = 0$.
- So $v_1 + 2v_2 + 3v_3 = 0$, and $-v_1 + v_3 = 0$.
- The general solution to this pair of equations is $v = (t, -2t, t)$ for $t \in \mathbb{R}$.
- We want a solution of length 1: we solve $\|v\| = 1$, to find that $t = \pm\frac{1}{\sqrt{6}}$.
- So the solution is $v = \pm(\frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}}, \frac{1}{\sqrt{6}})$.

The vector v is called the *normal* vector to P.

Now, we define length and angle in any inner product space.

Definition 10. Let u, v be vectors in V.

- the length of v is $\|v\| = \sqrt{\langle v, v\rangle}$
- the angle θ between u and v is given by $\langle u, v\rangle = \|u\| \, \|v\| \operatorname{Cos}(\theta)$

1.4 Revision

You should be able to do the following.

1. State the axioms of an inner product.
2. Decide with proof whether a given function is an inner product.
3. Derive the formula for angle from the Cosine rule.
4. Calculate lengths and angles with a given inner product.
5. Find the minimal distance from a point to a line.
6. Find the normal vector to a plane in $\mathbb{R}^3$.

2 Orthonormal bases

2.1 Orthogonality, orthonormality

Definition 11. Let X be a set of vectors in an inner product space. We say that X is an *orthogonal set* if every pair of distinct elements of X is orthogonal. We say that X is an *orthonormal set* if it is orthogonal **and** every vector in X has length 1.

An orthogonal set is always linearly independent.

Theorem 12. *If $X = \{x_1, x_2, \ldots, x_n\}$ is orthogonal, then X is a linearly independent set.*

Proof. Consider the vector

$$y = a_1 x_1 + a_2 x_2 + \ldots + a_n x_n.$$

Since X is orthogonal, we have

$$\begin{aligned} \langle y, x_i \rangle &= \langle a_1 x_1, x_i \rangle + \langle a_2 x_2, x_i \rangle + \ldots + \langle a_n x_n, x_i \rangle \\ &= a_1 \langle x_1, x_i \rangle + a_2 \langle x_2, x_i \rangle + \ldots + a_n \langle x_n, x_i \rangle \\ &= a_i \langle x_i, x_i \rangle \end{aligned}$$

Now, suppose that $y = 0$: this implies that $a_i = 0$ for all a_i. Thus X is linearly independent. □

Given linearly independent vectors u and v we often want to write $u = u' + v'$, where u' is orthogonal to v, and v' is parallel to v. We do this using the following lemma.

Lemma 13. *Let u and v be linearly independent vectors in a vector space V. Then*

$$u' = u - \frac{\langle u, v \rangle}{\langle v, v \rangle} v$$

is orthogonal to v.

Proof. We use the axioms of an inner product. Recall that $\frac{\langle u,v\rangle}{\langle v,v\rangle}$ is a **scalar**.

$$\begin{aligned}
\langle u', v\rangle &= \left\langle u - \frac{\langle u,v\rangle}{\langle v,v\rangle} v, v \right\rangle \\
&= \langle u,v\rangle - \left\langle \frac{\langle u,v\rangle}{\langle v,v\rangle} v, v \right\rangle \\
&= \langle u,v\rangle - \frac{\langle u,v\rangle}{\langle v,v\rangle} \langle v,v\rangle \\
&= \langle u,v\rangle - \langle u,v\rangle \\
&= 0
\end{aligned}$$

□

Example 14. Take $u = (1,3,4)$ and $v = (2,0,-1)$. We find the components of u parallel and orthogonal to v. First we calculate

$$\frac{u \bullet v}{v \bullet v},$$

which we find to be $\frac{-2}{5}$. So the component of u parallel to v is $v' = \frac{-2}{5} v = (\frac{-4}{5}, 0, \frac{2}{5})$. The component of u orthogonal to v is $u - v' = (1,3,4) - (\frac{-4}{5}, 0, \frac{2}{5}) = (\frac{9}{5}, 3, \frac{18}{5})$. (Check that $u - v'$ is orthogonal to v. Find the parallel and orthogonal components of other pairs of vectors.)

2.2 Orthonormal bases, the Gram-Schmidt process

Consider the sets $S = \{(1,2),(3,4)\}$ and $T = \{(1,0),(0,1)\}$ of $\mathbb{R}^2$. Both are bases for $\mathbb{R}^2$.

T has the property that

$$t_i \bullet t_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

This makes T more useful for calculation than S.

Fact 15. *Every finite dimensional vector space has an orthonormal basis.*

Moreso: given any spanning set T for V we can find an orthonormal basis. The procedure for doing so is called the *Gram-Schmidt* process. We break it into two parts:

- Given a spanning set T, we calculate an orthogonal set of vectors S.
- Given an orthogonal set of vectors S, we normalise to obtain an orthonormal set of vectors B.

Given T we construct our orthogonal set of vectors S step by step, using the following procedure.

1. We set $s_1 = t_1$.
2. We set $s_2 = t_2 - \frac{\langle t_2, s_1\rangle}{\langle s_1, s_1\rangle} s_1$.
3. We set $s_3 = t_3 - \frac{\langle t_3, s_2\rangle}{\langle s_2, s_2\rangle} s_2 - \frac{\langle t_3, s_1\rangle}{\langle s_1, s_1\rangle} s_1$.
4. We set $s_4 = t_4 - \frac{\langle t_4, s_3\rangle}{\langle s_3, s_3\rangle} s_3 - \frac{\langle t_4, s_2\rangle}{\langle s_2, s_2\rangle} s_2 - \frac{\langle t_4, s_1\rangle}{\langle s_1, s_1\rangle} s_1$.
5. And so on. The general term in the sequence is

$$s_i = t_i - \sum_{k=1}^{i-1} \frac{\langle t_i, s_k\rangle}{\langle s_k, s_k\rangle} s_k$$

6. If at any step one of the t_i is linearly dependent on elements that we have already seen, then we will get $s_i = 0$. Our last step in producing an orthogonal set of vectors is to discard any such s_i. (So our set S might be strictly smaller than T.)

We have produced a set of orthogonal, linearly independent vectors, S. Now, we obtain an orthonormal set, B. The second step of the process is much easier than the first.

- For each s_i, we set $b_i = \frac{s_i}{\|s_i\|}$

And that is the entire Gram-Schmidt process.

Example 16. We take $T = \{(1,1,0), (0,2,-1), (3,-1,-1), (2,2,2)\}$, which is a spanning set for $\mathbb{R}^3$ (equipped with the dot product). We carry out the Gram-Schmidt process.

- $s_1 = t_1 = (1,1,0)$.
- $s_2 = t_2 - \frac{\langle t_2, s_1\rangle}{\langle s_1, s_1\rangle} s_1$. First we calculate $\langle t_2, s_1\rangle = 2$ and $\langle s_1, s_1\rangle = 2$. So $s_2 = t_2 - s_1 = (-1,1,-1)$.
- We calculate $s_3 = t_3 - \frac{-3}{3} s_2 - \frac{2}{2} s_1 = (3,-1,-1) + (-1,1,-1) - (1,1,0) = (1,-1,-2)$.
- Finally, $s_4 = (0,0,0)$ (verify this!)

Now, we discard s_4, to obtain $S = \{(1,1,0), (-1,1,-1), (1,-1,-2)\}$, which is a basis for $\mathbb{R}^3$. We normalise each of the vectors to obtain the orthonormal basis

$$B = \left\{ \frac{1}{\sqrt{2}}(1,1,0), \frac{1}{\sqrt{3}}(-1,1,-1), \frac{1}{\sqrt{6}}(1,-1,-2) \right\}$$

2.3 Orthogonal decomposition

Orthonormal sets are very useful for computation with vectors, as the following theorem shows.

Theorem 17. *Let $B = \{b_1, b_2, \dots, b_n\}$ be an orthonormal basis for an inner product space V. Then for any vector $v \in V$,*

$$v = \sum_{i=1}^{n} \langle v, b_i \rangle \, b_i.$$

Proof. Two vectors u and v in V are equal if and only if $u - v = 0$. Furthermore, $u - v = 0$ if and only $\langle u - v, b_i \rangle = 0$ for all b_i in a basis B of V.

We show that $v - \sum_{i=1}^{n} \langle v, b_i \rangle \, b_i = 0$.

$$\begin{aligned} \left\langle v - \sum_{i=1}^{n} \langle v, b_i \rangle \, b_i, b_j \right\rangle &= \langle v, b_j \rangle - \sum_{i=1}^{n} \langle v, b_i \rangle \langle b_i, b_j \rangle \\ &= \langle v, b_j \rangle - \langle v, b_j \rangle \langle b_j, b_j \rangle \\ &= 0 \end{aligned}$$

But b_j can be chosen to be any element of B. The result follows. □

We give an example of an application of this theorem.

Example 18. We find the polynomial of degree ≤ 2 which is closest to $\operatorname{Sin} x$ on the interval $[-1, 1]$.

- We find an orthonormal basis for $P_2(x) = \{a_0 + a_1 x + a_2 x^2 \mid a_i \in \mathbb{R}\}$,

$$B = \left\{ \frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}} x, \sqrt{\frac{45}{8}} (x^2 - \frac{1}{3}) \right\}$$

- We work out $\langle \operatorname{Sin}(x), b_1 \rangle \, b_1 + \langle \operatorname{Sin}(x), b_2 \rangle \, b_2 + \langle \operatorname{Sin}(x), b_3 \rangle \, b_3$.
- We find that the closest polynomial is $p(x) = .7377x$.

(Do the calculations!)

We can also generalise this idea of orthogonal decomposition to subspaces.

Definition 19. Let $U = \{u_1, u_2, \dots, u_m\}$ be a linearly independent set in $\mathbb{R}^n$, and let $v \in \mathbb{R}^n$. The *orthogonal decomposition of v with respect to U* is an expression $v = p + q$ with the following properties:

- $p = a_1 u_1 + a_2 u_2 + \dots a_m u_m$
- $q \bullet u_i = 0$ for all $1 \leq i \leq m$.

Example 20. The orthogonal decomposition has many applications: e.g. we can define the distance of a vector from a subspace. Examples were done in class on 27/01/2012.

2.4 Application: Fourier series

We describe a special orthonormal set of functions defined on $[-\pi, \pi]$ which are used to approximate functions on this domain.

For any positive integer n we define the functions

$$p_n = \frac{1}{\sqrt{\pi}} \operatorname{Cos} nt, \quad q_n = \frac{1}{\sqrt{\pi}} \operatorname{Sin}(nt)$$

Furthermore, we define $p_0(t) = \frac{1}{\sqrt{2\pi}}$.

Theorem 21. *The set $F = \{p_0, q_1, p_1, q_2, p_2, q_3, \ldots\}$ is orthonormal.*

Proof. We demonstrate that $\langle p_m, p_n \rangle = 1$ if $m = n$ and 0 otherwise. We will use standard identities for Sin and Cos without comment.

First, we show that $\|p_n\| = 1$ for all $n \geq 1$.

$$\begin{aligned}
\langle p_n, p_n \rangle &= \int_{-\pi}^{\pi} \frac{1}{\pi} \operatorname{Cos}^2(nt)\, \mathrm{dt} \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} 1 + \operatorname{Cos}(2nt)\, \mathrm{dt} \\
&= \frac{1}{2\pi} \left(t + \frac{1}{2n} \operatorname{Sin}(2nt) \right) \Big|_{-\pi}^{\pi} \\
&= \frac{1}{2\pi} \left(\pi + \frac{1}{2n} \operatorname{Sin} 2\pi t \right) - \frac{1}{2\pi} \left(-\pi + \frac{1}{2n} \operatorname{Sin}(-2\pi n) \right) \\
&= 1
\end{aligned}$$

Now, we show that $\langle p_m, p_n \rangle = 0$ for $m \neq n \neq 0$.

$$\begin{aligned}
\langle p_m, p_n \rangle &= \frac{1}{\pi} \int_{-\pi}^{\pi} \operatorname{Cos}(mt) \operatorname{Cos}(nt)\, \mathrm{dt} \\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Cos}(m+n)t + \operatorname{Cos}(m-n)t\, \mathrm{dt} \\
&= \frac{1}{2\pi} \left(\frac{1}{m+n} \operatorname{Sin}(m+n)t + \frac{1}{m-n} \operatorname{Sin}(m-n)t \right) \Big|_{-\pi}^{\pi} \\
&= 0
\end{aligned}$$

Thus the set $\{p_1, p_2, p_3, \ldots\}$ is orthonormal. The other parts of the proof are similar, and are left as exercises. □

In a technical sense (which we will not make precise), F is a basis for the set of all continuous functions $f : [\pi, \pi] \to \mathbb{R}$.

Fact 22. *Let $f : [\pi, \pi] \to \mathbb{R}$ be a continuous function. Then there exist two sequences of real numbers $a_0, a_1, a_2, \ldots$ and $b_1, b_2, b_3, \ldots$ such that*

$$\begin{aligned} f(t) &= a_0 p_0(t) + \sum_{n=1}^{\infty} a_n p_n(t) + b_n q_n(t) \\ &= \frac{a_0}{\sqrt{2\pi}} + \sum_{n=1}^{\infty} \frac{a_n}{\sqrt{\pi}} \operatorname{Cos}(nt) + \frac{b_n}{\sqrt{\pi}} \operatorname{Sin}(nt) \end{aligned}$$

This is the Fourier series for f.

Fourier analysis is a well studied area of mathematics with many applications. In particular, any wave-type function which arises in engineering or physics (sound, shock waves, light etc.) can be decomposed into a sum of trigonometric functions. (This is done by normalising the period of the wave to be 2π units, and then setting $f(x + k2\pi) = f(x)$ for any $k \in \mathbb{Z}$.) This decomposition often reveals important information about the system being studied. We illustrate how to find the Fourier series for some simple periodic functions.

Theorem 23. *Let $f : [\pi, \pi] \to \mathbb{R}$ be a continuous function. Then the Fourier coefficients of $f(t)$ may be found as follows:*

$$\begin{aligned} a_0 &= \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(t)\,\mathrm{dt} \\ a_n &= \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(t) \operatorname{Cos}(nt)\,\mathrm{dt} \\ b_n &= \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(t) \operatorname{Sin}(nt)\,\mathrm{dt} \end{aligned}$$

Proof. By Fact 22, we know that the Fourier series of $f(t)$ exists and is unique. By Theorem 21, we know that all of the functions in the Fourier sequence are orthogonal. So by linearity of the inner product:

$$\begin{aligned} \langle f(t), p_n(t) \rangle &= \langle a_0 p_0(t) + a_1 p_1 t + b_1 q_1(t) + a_2 p_2(t) + \ldots, p_n(t) \rangle \\ &= \langle a_0 p_0(t), p_n(t) \rangle + \langle a_1 p_1(t), p_n(t) \rangle + \langle b_1 q_1(t), p_n(t) \rangle + \ldots \\ &= \langle a_n p_n(t), p_n(t) \rangle \\ &= a_n \end{aligned}$$

The proof for the b_n terms is identical. □

Example 24. Let $f(t) = t$ for $t \in [-\pi, \pi]$, and set $f(t + 2\pi) = f(t)$. (So $f(t)$ can be visualised as a 'saw-tooth' wave.) We calculate the Fourier series.

First of all, we observe that $f(t)$ is an odd function: $f(-t) = -f(t)$. This means that

$$a_0 = \frac{1}{\sqrt{2\pi}} \int_{\pi}^{\pi} f(t)\,\mathrm{dt} = 0.$$

Now, for $n \geq 1$, we need to find

$$a_n = \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(t) \operatorname{Cos}(nt) \, \mathrm{dt} \, .$$

We use integration by parts, setting $u = t$ and $\mathrm{dv} = \operatorname{Cos}(nt)$. We find that $a_n = 0$ for all n.

Finally,

$$b_n = \frac{1}{\sqrt{\pi}} \int_{-\pi}^{\pi} f(t) \operatorname{Sin}(nt) \, \mathrm{dt}$$

Integration by parts yields that

$$b_n = \frac{1}{\sqrt{\pi}} \left(\frac{-t}{n} \operatorname{Cos}(nt) + \frac{1}{n^2} \operatorname{Sin}(nt) \right) \Bigg|_{-\pi}^{\pi}$$

Evaluating, we find that $b_n = \frac{-2\sqrt{\pi}}{n} \operatorname{Cos}(n\pi)$. We observe that $\operatorname{Cos}(n\pi) = 1$ if n is even, and -1 otherwise. So $b_n = \frac{2\sqrt{\pi}}{n}(-1)^{n+1}$.

Thus the Fourier series of $f(t) = t$ is

$$f(t) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{2}{n} \operatorname{Sin} nt$$

(Fill in the missing steps in the above integrations.)

2.5 Revision

You should be able to do the following:

1. State and prove Lemma 3.
2. Apply the Gram-Schmidt process to a set of vectors from an arbitrary inner product space.
3. Find the orthogonal decomposition of a vector v with respect to a subspace S, and hence find the distance from v to S, and the point in S closest to v.
4. Give the definition of a Fourier Series.
5. Show that a given subset of the Fourier basis is orthonormal (i.e. prove part of Theorem 11).
6. Compute the first n Fourier coefficients of a given function. (See Question 15 for some examples.)

3 Symmetric and orthogonal matrices

3.1 Change of basis

Let T and U be two bases for V, an n-dimensional vector space.

By definition every $v \in V$ can be written in the form

$$v = \alpha_1 t_1 + \alpha_2 t_2 + \ldots + \alpha_n t_n$$

in a unique way.

Definition 25. The vector $v_T = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ is the *coefficient vector of v with respect to T.*

Since T is a basis, every element of U can be uniquely expressed as a linear combination of elements of T. Often we are given a vector v_U, and we want to find v_T.

This is done via a change of basis.

1. For each $u \in U$, calculate u_T.
2. Plug these expressions for the vectors u into v_U to obtain an expression in terms of the t_i.

This is can be viewed as a linear transformation of V, and so can be represented as a matrix with respect to the basis T. The following steps produce a matrix which takes v_T and produces v_U.

- Write the basis elements of T as a matrix: $M_T = [t_1|t_2|\ldots|t_n]$, and similarly for U: $M_U = [u_1|u_2|\ldots|u_n]$.
- Then we are looking for a matrix X such that the image under X of t_i is u_i.
- That is $X = M_U^{-1} M_T$.

Example 26. Take $T = \{(1,2), (-1,0)\}$ and $U = \{(1,0), (-3,1)\}$. We wish to find the change of basis matrix which takes v_T to v_U for any vector v.

$$M_T = \begin{pmatrix} 1 & -1 \\ 2 & 0 \end{pmatrix} \quad M_U = \begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix}$$

So we find that

$$M_U^{-1} M_T = \begin{pmatrix} 7 & -1 \\ 2 & 0 \end{pmatrix}$$

Now, we observe that the vector $(1,0)_T$ corresponds to the vector $(1,2)$ in the standard basis. We want its expression with respect to the basis U. We calculate $M_U^{-1}M_T(1,0)^\top$, and find this to be $(7,2)_U$. This means that $7(1,0)+2(-3,1) = (1,2)$.

Note that X takes a vector with co-ordinates given in terms of T, and returns its co-ordinates relative to U. Now suppose that M_T is a matrix relative to the basis T, and we want to know how it acts relative to the basis U.

1. We calculate X with respect to the basis T, which maps t_i to u_i for all $1 \leq i \leq n$.
2. Then X^{-1} maps the basis U to the basis T.
3. So XMX^{-1} (in that order!) takes a vector expressed in terms of U to its image in terms of T, applies M and then expresses it in terms of U again.
4. Thus $M_U = XM_TX^{-1}$ gives us the action of M relative to the basis U.

Typically, M_T is expressed in terms of the standard basis, in which case $A = I$ and $X = B$.

Example 27. We express the linear transformation $f(t_1,t_2) = (2t_1+t_2, 3t_1-t_2)$ in terms of the basis $T = \{(1,0),(0,1)\}$ and then in terms of $U = \{(1,2),(2,3)\}$.

The matrix with respect to the standard basis T is

$$M_T = \begin{pmatrix} 2 & 1 \\ 3 & -1 \end{pmatrix}$$

Now, the change of basis matrix from T to U is

$$X = \begin{pmatrix} -3 & 2 \\ 2 & -1 \end{pmatrix}$$

So the matrix of the linear transformation f, relative to the basis U is:

$$M_U = XM_TX^{-1} = \begin{pmatrix} -10 & 15 \\ 7 & 11 \end{pmatrix}$$

We show that this really does represent the same linear transformation. Let $v = (1,0)_U$, so $v = (1,2)_T$. We show that $M_Uv_U = M_Tv_T$.

$$M_Uv_U = (-10,7)_U \quad M_Tv_T = (4,1)_T$$

Now, the linear transformation from U to T is given by X, so we check that $Xv_U = v_T$. We observe that this does indeed hold: $(4,1)_T = (-10,7)_U$.

3.2 Eigenvalues and eigenvectors

In this section we describe a method for finding a basis relative to which a linear transformation is represented by a diagonal matrix. This is via eigenvalues and eigenvectors.

Definition 28. Let M be a $n \times n$ matrix. An *eigenvector* is a vector v such that

$$Mv = \lambda v.$$

An *eigenvalue* for v is the value of λ corresponding to v.

Definition 29. The *characteristic polynomial* of a matrix is the determinant of $M - \lambda I$ (treating λ as an unknown).

Fact 30. *The roots of the characteristic polynomial of M are the eigenvalues of M.*

So: to find the eigenvalues and eigenvectors of a matrix M, we do the following.

- Find the characteristic polynomial of M.
- Find the roots of the characteristic polynomial (i.e. the eigenvalues).
- For each eigenvalue, solve the system of linear equations $Mv = \lambda v$, where the co-ordinates of v are unknowns.
- A basis for the set of solutions, for a fixed value of λ is a set of eigenvectors *corresponding to* λ.

It is not always possible to find a set of n eigenvectors for an $n \times n$ matrix over $\mathbb{R}$. Consider for example the matrices

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

The first has no real eigenvalues (though it is diagonalisable over $\mathbb{C}$), the second has an eigenvalue of 1 with *algebraic multiplicity* 2, and *geometric multiplicity* 1 (i.e. the factor $\lambda - 1$ occurs twice in the characteristic polynomial, but has only one linearly independent eigenvector).

Fact 31. *If the eigenvectors of M are linearly independent, then the matrix $P = [v_1 \mid v_2 \mid \ldots \mid v_n]$ is invertible. The matrix $P^{-1}MP$ is a diagonal matrix, whose non-zero entries are the eigenvalues of M. This calculation is called diagonalising the matrix M.*

3.3 Bilinear forms

Bilinear forms are a generalization of inner products.

Definition 32. A *bilinear form* is a map $V \times V \to \mathbb{R}$ with the following properties.

1. $\langle u + u', v\rangle = \langle u, v\rangle + \langle u', v\rangle$
2. $\langle u, v + v'\rangle = \langle u, v\rangle + \langle u, v'\rangle$
3. $\langle \lambda u, v\rangle = \langle u, \lambda v\rangle = \lambda \langle u, v\rangle$

Definition 33. Let V be an n-dimensional vector space, and B a basis for V. Then each $v \in V$ has a unique expression as a linear combination of elements of B.

$$v = \sum_{i=1}^{n} \alpha_i b_i$$

We call the vector $v_B = (\alpha_1, \alpha_2, \ldots, \alpha_n)^\top$ the *image of v with respect to B*. (Note: v_B is a **column** vector.)

The notion of a bilinear form is useful for the following reason.

Definition 34. Let $\langle -, -\rangle$ be a bilinear form on an n-dimensional vector space V. Fix a basis B, and form the matrix

$$M_B = [\langle b_i, b_j\rangle]_{1 \le i,j \le n}\,.$$

We say that M_B is the *matrix of M relative to B*.

Theorem 35. *Let $\langle -, -\rangle$ be a bilinear form on an n-dimensional vector space V, and let B be a basis for V. Then for any $u, v \in V$,*

$$\langle u, v\rangle = u_B^\top M_B v_B.$$

Proof. The i^{th} basis element is represented by the vector e_i, which has a 1 in the i^{th} co-ordinate and 0s elsewhere. First we show that $\langle e_i, e_j\rangle = m_{i,j}$ for any $1 \le i, j \le n$. The result then follows from the properties of a bilinear form. We consider the matrix product

$$\begin{aligned}(e_i^\top M)e_j &= \left(\sum_{t=1}^{n} e_{it} m_{tk}\right)_{1 \le k \le n} e_j \\ &= (m_{i,k})_{1 \le k \le n}\, e_j \\ &= m_{i,j}\end{aligned}$$

But by definition, the entry $m_{i,j}$ of m is $\langle b_i, b_j\rangle$.

Now, any vectors $u, v \in V$ can be expressed in the form $u = \sum_{t=1}^{n} \alpha_t b_t$ and $v = \sum_{k=1}^{n} \beta_k b_k$. By bilinearity of the inner product,

$$\begin{aligned}
\langle u, v\rangle &= \sum_{t=1}^{n} \langle \alpha_t b_t, v\rangle \\
&= \sum_{t=1}^{n} \alpha_t \langle b_t, v\rangle = \sum_{t=1}^{n} \alpha_t \sum_{k=1}^{n} \langle b_t, \beta_k b_k\rangle \\
&= \sum_{t=1}^{n} \alpha_t \sum_{k=1}^{n} \beta_k \langle b_t, b_k\rangle \\
&= \sum_{t=1}^{n} \alpha_t \sum_{k=1}^{n} \beta_k m_{tk} = \sum_{t=1}^{n} \sum_{k=1}^{n} \alpha_t \beta_k m_{tk}
\end{aligned}$$

We show that evaluating the matrix product gives the same result. This proves the theorem.

$$u_B^\top M_B v_B = \left(\sum_{t=1}^{n} \alpha_t e_t^\top\right) M_B \left(\sum_{k=1}^{n} \beta_k e_k\right)$$

But matrix multiplication is bilinear, so we obtain

$$\sum_{t=1}^{n} \sum_{k=1}^{n} \alpha_t \beta_k m_{tk}$$

as required. □

Thus, once we choose a basis, every bilinear form is represented by a matrix, and every matrix gives a bilinear form on our vector space. Orthonormal bases correspond to the identity matrix.

Lemma 36. *Let B be a basis for the n-dimensional vector space V. Then B is orthonormal with respect to the inner product $\langle -, -\rangle$ if and only if $M_B = I_n$.*

Proof. If the basis B is orthonormal, then by definition $\langle b_i, b_j\rangle = 1$ if $i = j$ and 0 otherwise.

So every diagonal entry of M_B is 1, and every off-diagonal entry is 0. □

Note that even though the identity matrix is unique, the choice of orthonormal basis is **not**: recall the choices that are made during the Gram-Schmidt process. Many different bases can produce the same matrix for a given bilinear form. On the other hand, the same bilinear form can give rise to many different matrices.

Example 37. The standard basis of $\mathbb{R}^n$ is orthonormal with respect to the dot product. So the dot product is represented by the identity matrix.

Example 38. Consider the bilinear form $\langle (x_1, y_1), (x_2, y_2)\rangle = x_1x_2 + x_1y_2 + y_1x_1 + y_2x_2$. Relative to the basis $\{(0, 1), (1, 0)\}$ this form has matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

But relative to the basis $\{(1, 1), (1, -1)\}$, the matrix is

$$\begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix}.$$

3.4 Isometries of bilinear forms

Definition 39. Let $\langle -, -\rangle$ be a bilinear form on a vector space V. An *isometry* of $\langle -, -\rangle$ is a linear transformation L such that $\langle u, v\rangle = \langle Lu, Lv\rangle$ for all $u, v \in V$.

Describing a bilinear form in terms of its matrix we obtain the following result.

Lemma 40. *The linear transformation L is an isometry of $\langle -, -\rangle$ if and only if $L_B^\top M_B L_B = M_B$.*

Proof. The following calculation proves the result.

$$\begin{aligned} \langle Lu, Lv\rangle &= L_B u_B \cdot L_B v_B \\ &= (L_B u_B)^\top M_B L_B v_B \\ &= u_B^\top L_B^\top M_B L_B v_B \\ &= u_B^\top (L_B^\top M_B L_B) v_B \end{aligned}$$

But by hypothesis $\langle Lu, Lv\rangle = \langle u, v\rangle$ for all $u, v \in V$, so $M_B = L_B^\top M_B L_B$. □

For the remainder of this section, all vectors and all linear transformations are given with respect to the standard basis of $\mathbb{R}^n$. We study the isometries of the dot product. These are linear transformations L with the property that for all $u, v \in \mathbb{R}^n$,

$$\begin{aligned} u \cdot v &= Lu \cdot Lv \\ &= (Lu)^\top I_n Lv \\ &= u^\top L^\top I_n Lv \\ &= u^\top (L^\top L) v \end{aligned}$$

We observe that L satisfies this condition if and only if $L^\top L = I_n$.

Definition 41. An $n \times n$ matrix L is *orthogonal* if $L^\top L = I_n$.

This means that $L^{-1} = L^\top$.

Theorem 42. *The $n \times n$ matrix L is orthogonal if and only if the columns of L form an orthonormal basis of $\mathbb{R}^n$ (with respect to the dot product).*

Proof. We observe that the entry in position (i, j) of $L^\top L$ is the inner product $\langle c_i, c_j \rangle$ of column i with column j.

Suppose that $L^\top L = I_n$. We find that $\langle c_i, c_j \rangle = 0$ if $i \neq j$, and $\langle c_i, c_i \rangle = 1$. Thus the columns of L form an orthonormal basis for $\mathbb{R}^n$ when L is an orthogonal matrix.

In the other direction, suppose that the columns of L form an orthonormal basis for $\mathbb{R}^n$. Then $L^\top L = I_n$, and L is orthogonal. □

We can form many examples of orthogonal matrices: choose a basis for $\mathbb{R}^n$, use the Gram-Schmidt process to orthogonalise, and use these basis elements as the columns of L.

Example 43. We describe all 2 × 2 orthogonal matrices. Suppose that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is an orthogonal matrix. Then by Theorem 42, $\{(a, c), (b, d)\}$ is an orthonormal basis of $\mathbb{R}^2$ with respect to the dot product. So we obtain the following identities:

$$\begin{aligned} (a, c) \bullet (a, c) = a^2 + c^2 &= 1 \\ (b, d) \bullet (b, d) = b^2 + d^2 &= 1 \\ (a, c) \bullet (b, d) = ab + cd &= 0 \end{aligned}$$

Thus (a, c) and (b, d) are points on the unit circle. We express (a, c) as $(\mathrm{Cos}(\theta), \mathrm{Sin}(\theta))$, and $(b, d) = (\mathrm{Cos}(\nu), \mathrm{Sin}(\nu))$. The third equation tells us that $\mathrm{Cos}(\theta - \nu) = 0$. (Recall the definition of angle in an inner product space!) Thus $\nu = \theta \pm \frac{\pi}{2}$. We normally express (b, d) as $(-\mathrm{Sin}(\theta), \mathrm{Cos}(\theta))$.

Thus M is a 2×2 orthogonal matrix if and only if it is equal to one of the following, for some value of θ:

$$M_\theta = \begin{pmatrix} \mathrm{Cos}(\theta) & -\mathrm{Sin}(\theta) \\ \mathrm{Sin}(\theta) & \mathrm{Cos}(\theta) \end{pmatrix}, \qquad M'_\theta \begin{pmatrix} \mathrm{Cos}(\theta) & \mathrm{Sin}(\theta) \\ \mathrm{Sin}(\theta) & -\mathrm{Cos}(\theta) \end{pmatrix}$$

Observe that the matrices M_θ have determinant 1: these are called the *special orthogonal matrices*. The matrix M_θ acts on $\mathbb{R}^2$ by rotating points anti-clockwise by angle of θ around the origin.

Secondly, we observe that $M'_\theta = M_\theta T$, where T is the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Now, T acts as reflection in the x-axis, so the action of M'_θ corresponds to reflection in the x-axis followed by rotation about the origin by an angle of θ.

3.5 Symmetric bilinear forms

In this section we examine a special type of bilinear form, closely related to inner products.

Definition 44. A bilinear form is *symmetric* if $\langle u, v\rangle = \langle v, u\rangle$ for all vectors $u, v \in V$. A bilinear form is *positive definite* if $\langle v, v\rangle > 0$ for all non-zero vectors $v \in V$.

An inner product is a positive definite bilinear form.

Definition 45. The matrix M is *symmetric* if $M^\top = M$.

Lemma 46. *The bilinear form $\langle -, -\rangle$ is symmetric if and only if the matrix M_B is symmetric for any choice of basis B.*

Proof. We have that $m_{i,j} = \langle b_i, b_j\rangle = \langle b_j, b_i\rangle = m_{j,i}$ for any basis vectors b_i, b_j in B. □

Definition 47. A matrix is *positive definite* if for any $v \in V$,

$$v^\top M v \geq 0.$$

So M is the matrix of an inner product if and only if M is symmetric and positive definite. While symmetry is a condition easily verified from the entries of the matrix, there is no known condition which characterizes completely positive matrices.

Recall that Fact 5 of the handout *Orthonormal bases* states that every finite dimensional inner product space has an orthonormal basis. This means that, for any inner product space, there exists a basis B such that M_B is the identity matrix. This is not true for bilinear forms in general.

We show that a symmetric matrix is always diagonalisable. This shows that every symmetric inner product can be *polarised.*

Theorem 48. *Suppose that M is a symmetric matrix, and that v_1, v_2 are two eigenvectors of M corresponding to the eigenvalues λ_1 and λ_2 respectively. Then $v_1 \bullet v_2 = 0$.*

Proof.

$$\begin{aligned} \lambda_1 v_1^\top &= (Av_1)^\top v_2 \\ &= v_1^\top A^\top v_2 \\ &= v_1^\top A v_2 \\ &= \lambda_2 v_1^\top v_2 \end{aligned}$$

But by assumption, $\lambda_1 \neq \lambda_2$: so we obtain a contradiction unless $v_1 \bullet v_2 = 0$. This proves the result. □

Corollary 49. *Let M be an $n \times n$ symmetric matrix with all eigenvalues distinct. Then the eigenvectors of M form an orthonormal basis for $\mathbb{R}^n$.*

Proof. By Theorem 48, the eigenvectors of M are orthogonal. Hence they are linearly independent, so it remains only to normalise these vectors to obtain an orthonormal basis for $\mathbb{R}^n$. $\square$

Actually, we do not need the eigenvalues of M to be distinct, the proof is more complicated in the general case, however. So we omit it.

Fact 50. *Let M be an $n \times n$ real symmetric matrix. Then the eigenvectors of M form an orthonormal basis for $\mathbb{R}^n$.*

Thus by Fact 31, for any symmetric matrix M, there exists an orthogonal matrix O such that $O^{-1}MO = D$, where D is a diagonal matrix.

Example 51. We diagonalise the following matrix:

$$M = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{pmatrix}.$$

- The characteristic polynomial of M is $\lambda^3 - 4\lambda^2 - \lambda + 4$.
- This has roots at $-1, 1$ and 4.
- We find eigenvectors for each of the eigenvalues (solving $Mv = \lambda v$ in each case).
- We obtain the eigenvectors $(0, 1, -1)$, $(-2, 1, 1)$ and $(1, 1, 1)$ (in that order).
- We note that these vectors are orthogonal.
- Normalizing these vectors, we obtain the matrix:

$$O = \begin{pmatrix} 0 & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}.$$

- We find that

$$O^{-1}MO = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix}.$$

- Note that the diagonal entries of $O^\top MO$ are the eigenvectors of M.

Now, we combine the above facts and our results on symmetric matrices.

Theorem 52. *Let $\langle u, v\rangle$ be a symmetric bilinear form on $\mathbb{R}^n$. Then there exists an orthogonal matrix O which polarises $\langle u, v\rangle$.*

Proof. Choose a basis B of $\mathbb{R}^n$. Then M_B is a symmetric matrix. By Fact 50, M_B is orthogonally diagonalisable, hence there exists an orthogonal linear transformation L such that $L^\top M_{B'} L$ is a diagonal matrix. ☐

In practice, to calculate an inner product, we find the matrix O that diagonalizes M_B, then $\langle u, v\rangle = \langle OO^\top u, OO^\top v\rangle$, but $O^\top M_B O$ is diagonal, so we can just evaluate the inner product of $O^\top u$ and $O^\top v$ with respect to this polarized form to find $\langle u, v\rangle$.

3.6 Revision

You should be able to do the following.

1. Be able to find a change of basis matrix from given bases B_1 and B_2 of V, be able to calculate the coefficient vector of a given $v \in V$ for a given basis, be able to find M_{B_1} and M_{B_2} for a given linear transformation M of V, and explain how they are related by a change of basis matrix.
2. Find the eigenvalues and eigenvectors of a 2×2 or 3×3 matrix with real entries.
3. Give the definition of a bilinear form, write the matrix of a form relative to a basis, and recover a bilinear form from a given matrix and basis.
4. Define an isometry of a bilinear form and prove Lemma 16.
5. Define an orthogonal matrix, prove Theorem 18 and be able to manipulate orthogonal matrices (as in the questions on Problem Sheet 2).
6. Explain the relation between symmetric matrices and symmetric bilinear forms, prove Theorem 24 and orthogonally diagonalize a symmetric matrix.

4 The assignment problem

We describe an example of an *optimization problem*, and give a solution using linear algebra.

Example 53. A university intends to renovate 3 lecture theaters over the Summer. It has received three tenders from three different contractors, as follows. The entry in row i and column j of the following matrix is the bid of the i^{th} company for the j^{th} contract.

$$C = \begin{pmatrix} 53 & 96 & 37 \\ 47 & 87 & 41 \\ 60 & 92 & 36 \end{pmatrix}$$

The university must assign one contract to each contractor and wishes to minimize the total cost.

Definition 54. A *cost matrix* C is a matrix all of whose entries are non-negative integers. We interpret c_{ij} as the cost of having contractor i carry out task j.

Definition 55. A *transversal* of an $n \times n$ matrix M is a choice of n entries of M no two of which lie in the same row or the same column. The *weight* of a transversal is the sum of its entries.

Thus the problem in the above example is to find the transversal which has minimal weight. In this case, there are 3! possibilities, which can be checked by hand. We find a general method for when there are n contracts and n contractors.

Definition 56. Suppose that there are n tasks that must be completed, and that there are n contractors, each of which must be assigned one task. The cost of having task i completed by contractor j is c_{ij}. The *assignment problem* is to find an assignment of tasks which minimizes the total cost. In terms of linear algebra, given a cost matrix, we wish to find a transversal of minimal weight.

This problem occurs in many guises. E.g. we may want to find an optimal assignment of workers to jobs, or of players in a team to field positions, or of equipment to construction sites, etc.

The Hungarian method is a way of finding such a minimal solution. It is based on the following two observations, the proofs of which are straightforward.

Lemma 57. *If a cost matrix contains a transversal of weight zero, then this is an optimal solution to the assignment problem.*

Proof. The entries of the cost matrix are non-negative integers. So 0 is obviously optimal! □

Lemma 58. *If a number is added to or subtracted from all of the entries of a fixed row or column of a cost matrix, then an optimal assignment for the resulting matrix is also an optimal assignment for the original matrix.*

Proof. Suppose we add x to every entry in row i of our cost matrix. Any transversal must contain exactly one element from row i, so the weight of every transversal increases by x. So optimality is preserved. The argument for columns is analogous. □

Definition 59. A Hungarian operation is to add or subtract a fixed value from every entry in a fixed row or column of a cost matrix.

Essentially, we apply these Hungarian operations until we obtain a matrix for which the assignment problem is trivial.

1. Subtract the smallest entry in each row from every entry in the row.
2. Subtract the smallest entry in every column from every entry in that column.
3. Draw lines through the rows and columns to cover all the zeros so that the minimum number of lines is used.
4. If this minimum number is n, an optimal assignment is possible.
5. Otherwise, such an assignment is not yet possible, find the smallest entry not covered by any line, subtract this entry from all uncovered entries and add it to all entries covered by both a horizontal and a vertical line. Return to Step 3.

Example 60. We solve the matrix from the example above using the Hungarian method.

$$A = \begin{pmatrix} 53 & 96 & 37 \\ 47 & 87 & 41 \\ 60 & 92 & 36 \end{pmatrix}$$

After steps 1 and 2 we obtain the matrices

$$\begin{pmatrix} 16 & 59 & 0 \\ 6 & 46 & 0 \\ 24 & 56 & 0 \end{pmatrix}, \quad \begin{pmatrix} 10 & 13 & 0 \\ 0 & 0 & 0 \\ 18 & 10 & 0 \end{pmatrix}$$

respectively. Now, two lines cover all the zeros in this array, so we need to move to step 5. The minimal uncovered value is 10.

$$\begin{pmatrix} 0 & 3 & 0 \\ 0 & 0 & 10 \\ 8 & 0 & 0 \end{pmatrix}$$

We observe that there are multiple choices of transversal here. Any choice is optimal. We take $c_{11}+c_{22}+c_{33}$. With reference to the original matrix, we find that the optimal solution for the assignment problem is $53 + 87 + 36 = 176$.

4.1 Analysis of algorithm

The following argument is slightly complicated as written, but clear when accompanied by an example.

Lemma 61. *Step 5 of the algorithm given is a combination of Hungarian operations.*

Proof. Let x be the value of the smallest uncovered entry in the cost matrix C.

First we subtract x from every entry in every column which contains an uncovered entry. Then we add x to every row which contain a negative entry.

Observe that all negative entries must have been 0 before we subtracted x, so we are adding x to a row which is covered. We observe that this combination of Hungarian operations gives us Step 5. In particular, we have subtracted x from every uncovered element, done nothing to any element which is covered precisely once (i.e. we have subtracted and added x), and we have added x to elements which are covered by a row and a column. □

We also observe that we must find a solution after a finite number of steps. It suffices to show that the sum of all the entries in C decreases each time step 5 is executed. Since the sum is finite to begin with, we arrive at a transversal of weight 0 within a finite number of steps. Thus the following observation suffices.

Lemma 62. *Suppose the cost matrix C is $n \times n$. If there are less than n lines covering the zeros in the cost matrix, then there are at least n uncovered entries and at most n line intersections.*

Proof. Every pair of lines intersect in at most one point, and say that there are $b \leq n-1$ lines. Let k be the number of line intersections.

Then the number of points covered by the b lines is $bn - k$. So the number of points not covered by any line is $n^2 - bn + k$. We need to show that $n^2 - bn + k > k$, subtracting k from both sides, it suffices to show that

$$n^2 - bn > 0$$

but this is obvious since $b < n$. □

Hence, each time we apply step 5 of the Hungarian algorithm, we reduce the sum of the entries in the matrix. So within a finite number of steps we reach the zero matrix, at which point we certainly have a solution. It follows that the algorithm always terminates. The solution obtained is optimal by Lemma 58.

4.2 The Marriage problem

This is a maximization problem similar to the Assignment problem.

Example 63. A matchmaker has on her books n brides and n grooms, and for each pair has a rating c_{ij} between 0 and 10 which rates their compatibility. Again, we obtain a cost matrix with rows labelled by brides and columns labelled by grooms. We require a transversal for the cost matrix wit **maximum** weight.

This is known as the marriage problem, and is equivalent to many (more likely) real world scenarios. The method of solution is a modification of the Hungarian algorithm. The key step is to observe that the problem of maximizing the weight function on transversals of the cost matrix C is **exactly equivalent** to minimizing the weight function on transversals of $-C$. And the Hungarian solves this problem. Essentially the only additional information required is that, in step one, we consider the **smallest number** to be the **most negative** number.

Example 64. Maximize the weight function on transversals of the matrix

$$C = \begin{pmatrix} 3 & 6 & 9 & 2 \\ 4 & 5 & 1 & 5 \\ 6 & 2 & 7 & 6 \\ 8 & 5 & 6 & 6 \end{pmatrix}$$

The first step is to negate this matrix, then we subtract the most negative entry to obtain a positive matrix.

$$-C = \begin{pmatrix} -3 & -6 & -9 & -2 \\ -4 & -5 & -1 & -5 \\ -6 & -2 & -7 & -6 \\ -8 & -5 & -6 & -6 \end{pmatrix} \quad C' = \begin{pmatrix} 6 & 3 & 0 & 7 \\ 1 & 0 & 4 & 0 \\ 1 & 4 & 0 & 1 \\ 0 & 3 & 2 & 2 \end{pmatrix}$$

Step 2 is trivial in this case, there is already a zero in every column. Three lines suffice to cover all of the zeros (columns 1, 3 with row 2), however, so we proceed to step 5.

$$C'' = \begin{pmatrix} 6 & 2 & 0 & 6 \\ 2 & 0 & 5 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 2 & 2 & 1 \end{pmatrix}$$

Now we do require four lines to cover the zeros. We see that one possible transversal takes $c_{4,1}, c_{2,2}, c_{1,3}, c_{3,4}$. Thus the maximum weight is 26.

(Verify this by checking other alternatives!)

4.3 Generalizations

The first generalization of this problem is for assignments when we have more contractors than jobs. In this case we introduce 'dummy jobs' (rows of zeros) to form a square cost matrix, and then implement the Hungarian method as before.

Example 65. An antique dealer has four coins to sell by auction. Bids have been placed in advance by five collectors, with instructions that at most one one bid is to be honoured. The following cost matrix records the bids (with columns indexed by coins and rows by bidders). The seller wants to maximize his profit.

$$C = \begin{pmatrix} 150 & 65 & 210 & 135 \\ 175 & 75 & 230 & 155 \\ 135 & 85 & 200 & 140 \\ 140 & 70 & 190 & 130 \\ 170 & 50 & 200 & 160 \end{pmatrix}$$

To solve this problem, we first introduce an extra column of 0s which represents a dummy coin. The bidder matched to this coin will pay 0 and receive nothing. Since this is a maximization problem, we deal with it as the marriage problem. We give the first 3 steps of the process below (since every row contains a zero, Step 1 is trivial).

$$C' = \begin{pmatrix} -150 & -65 & -210 & -135 & 0 \\ -175 & -75 & -230 & -155 & 0 \\ -135 & -85 & -200 & -140 & 0 \\ -140 & -70 & -190 & -130 & 0 \\ -170 & -50 & -200 & -160 & 0 \end{pmatrix}, \quad C'' = \begin{pmatrix} 25 & 20 & 20 & 25 & 0 \\ 0 & 10 & 0 & 5 & 0 \\ 40 & 0 & 30 & 20 & 0 \\ 35 & 15 & 40 & 30 & 0 \\ 5 & 35 & 30 & 0 & 0 \end{pmatrix}$$

At this point we require four lines to cover all zeros, (we choose rows 2 and 5 and columns 2 and 5). The smallest uncovered entry is 20. So we apply step 5.

$$C''' = \begin{pmatrix} 5 & 20 & 0 & 5 & 0 \\ 0 & 30 & 0 & 5 & 20 \\ 20 & 0 & 10 & 0 & 0 \\ 15 & 15 & 20 & 10 & 0 \\ 5 & 55 & 30 & 0 & 20 \end{pmatrix}$$

We observe that the zeros now require 5 lines to cover them, and we obtain the transversal $c_{21}, c_{32}, c_{13}, c_{45}, c_{54}$, which has minimum value $175+85+210+160+0 = 650$. Note that bidder 4 obtains nothing.

A final generalization: suppose we wish to allow a single contractor to do more than one job. In this case, we can duplicate the row of bids by that contractor in the cost matrix, and introduce an additional dummy column to compensate. Either it is not more efficient to award that contractor a second contract and he will be assigned the dummy column, or a more efficient solution will be found (the order in which the two jobs are assigned to the contractor is irrelevant).

5 Linear Programming

5.1 Linear programming problems

We represent by $\overline{x}$ the vector $(x_1, x_2, \ldots, x_n)^\top$.

Definition 66. The *standard maximum problem* is as follows:

- We are given one *objective function*,

  $$P(\overline{x}) = \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n,$$

  and a set of m *constraint functions*

  $$\begin{array}{rcl} a_{1,1}x_1 + a_{1,2}x_2 + \ldots a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \ldots a_{2,n}x_n & = & b_2 \\ \vdots & = & \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \ldots a_{m,n}x_n & = & b_m, \end{array}$$

  where all of the $b_i \geq 0$ for $1 \leq i \leq m$ and $x_i \geq 0$ for $1 \leq i \leq n$.
- The space of all solutions to the constraint functions is called the *feasible set.*
- We wish to find a vector $\overline{x}$ in the feasible set which **maximises** the value of the objective function.

A typical example of this type of problem is the following.

Example 67. Maximise $P = 5x_1 + 6x_2$ subject to the constraints

$$\begin{array}{rcl} 3x_1 + 2x_2 & \leq & 120 \\ 4x_1 + 6x_2 & \leq & 260 \end{array}$$

Such problems arise quite often in business contexts, where we interpret the constraints as physical limitations, and the objective function as profit. An example of a 'real world' problem involving linear programming is as follows.

Example 68. You own 200 cattle. There are three feeds available, f_1, f_2, f_3 (or silage, calf-nuts and grain). These contain nutrients n_1, n_2, n_3 (vitamins, say) essential to the well-being of your cattle. You have studied Agricultural Science, and know that each cow requires b_i units of n_i to stay healthy. You also know that in each kilogram of f_i there are $a_{i,j}$ units of n_j. Now, you can tabulate all the information you know as follows:

$$\begin{aligned}
a_{1,1}x_i + a_{1,2}x_2 + a_{1,3}x_3 &= b_1 \\
a_{2,1}x_1 + a_{2,2}x_2 + a_{2,3}x_3 &= b_2 \\
a_{3,1}x_1 + a_{3,2}x_2 + a_{3,3}x_3 &= b_3
\end{aligned}$$

Where the x_is are the quantity in kilograms of each foodstuff which must be consumed by the cattle. A diet for the cattle is *feasible* if it meets all of the nutritional requirements. Unfortunately, you have to buy the feed for your cattle, and foodstuff f_i costs α_i euro per kilo. Given a space of feasible diets, we want to find the one of minimal cost, i.e. which minimises the function

$$P(x_1, x_2, x_3) = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3.$$

Equivalently, we wish to find a vector (x_1, x_2, x_3) which **maximizes** the function

$$-P(x_1, x_2, x_3) = -\alpha_1 x_1 - \alpha_2 x_2 - \alpha_3 x_3.$$

So this is a standard maximum problem. (The definition of a standard maximum problem does **not** require that the α_i are positive.)

We can write our linear programming problem very compactly using the language of matrices: the standard maximum problem is to maximize the function

$$P = \overline{\alpha}^\top \overline{x}$$

subject to the constraints

$$A\overline{x} \leq \overline{b}$$

and $\overline{x} \geq 0$, $\overline{b} \geq 0$.

5.2 Geometry of linear programming problems

We examine the geometry of the feasible set of a linear programming problem.

Definition 69. Consider the linear function $f(x,y) = ax + by + c$. It divides $\mathbb{R}^2$ into 3 parts depending on whether $f(x,y) < 0$, $f(x,y) = 0$ or $f(x,y) > 0$. The second case is the familiar equation of a line. We call the other two parts *open half spaces*. The subsets $f(x,y) \leq 0$ and $f(x,y) \geq 0$ are *closed half spaces*.

Remark 70. The space $\mathbb{R}^3$ is divided into 3 parts by a plane (2-dimensional subspace). More generally, $\mathbb{R}^n$ is divided into half-spaces by an $(n-1)$-dimensional subspace.

In general, the solution space of a linear programming problem is the intersection of a number of closed half spaces (one for each constraint).

Example 71. In Example 67, the solution space is the intersection of 4 half spaces: $x_1 \geq 0$, $x_2 \geq 0$, $3x_1 + 2x_2 - 120 \leq 0$, $4x_1 + 6x_2 - 260 \leq 0$. (Sketch this region.)

Example 72. The unit cube is the intersection of the following 6 half-spaces in $\mathbb{R}^3$:

$$x \geq 0, \;\; x \leq 1, \;\; y \geq 0 \;\; y \leq 1 \;\; z \geq 0, \;\; z \leq 1$$

We obtain a tetrahedron in $\mathbb{R}^3$ by taking the intersection of the the following 4 half spaces:

$$x \geq 0, \quad y \geq 0 \quad z \geq 0 \quad x + y + z \leq \alpha$$

for any $\alpha \in \mathbb{R}$.

Because we require all of the $x_i \geq 0$, and we are only interested in the case that all of the variables are finite, the space of solutions to the constraints will always form a convex polytope.

Definition 73. A polytope is a generalisation a polygon in $\mathbb{R}^2$ to higher dimensional spaces. An n-dimensional polytope is the (bounded) intersection of a number of half-spaces in $\mathbb{R}^n$.

Though all of our examples will be in 2- and 3-dimensions, we then generalise to higher dimensions algebraically.

Fact 74. *If a standard linear programming problem has an optimal solution, this optimal solutions will be achieved at a corner point of the solution space.*

Fact 75. *A corner of a polytope in* $\mathbb{R}^n$ *occurs at the intersection of* n *half-spaces.*

Thus in $\mathbb{R}^2$, a corner occurs at the intersection of two lines (as is familiar for polygons). In $\mathbb{R}^3$, a corner occurs at the intersection of three planes (visualise the cube or the tetrahedron).

Now, our linear programming problem involves $n + m$ inequalities in n variables. This corresponds to a polytope with at most $n+m$ faces which lives in n-dimensional space. We observe that some of the inequalities could be redundant (e.g. $x > 0$ and $x > 1$) in which case we can have strictly less faces than inequalities. We get a corner in the feasible set by replacing some n of the inequalities by equality.

Example 76. Consider the unit cube, described above. Its eight corners are at the points (x, y, z) where $x, y, z \in \{0, 1\}$. We observe that each of these is obtained by replacing inequality with equality in three of the equations. (Notice that some half-spaces do not intersect, and that their conditions are mutually exclusive.) We also note that the numbers of half-spaces and corners to consider quickly makes this approach impractical for solving linear programming problems.

5.3 The simplex method

The simplex method is a variation on Gaussian elimination used to solve linear programming problems. We choose pivot columns and then apply row operations just as in Gaussian elimination. The modification is an extra step in which the data from the problem determines the row and column where pivoting occurs. We describe the algorithm and give an example of its use in this section. We give an explanation of how it works in the next section.

- Turn all inequalities into equalities by adding 'slack' variables.
- Write the objective function as $P - \alpha_1 x_1 \dots \alpha_n x_n = 0$.
- Write out the initial conditions in a 'tableau':

	x_1	x_2	$\dots$	x_n	b_n
	$a_{1,1}$	$a_{1,2}$	$\dots$	$a_{1,n}$	b_1
	$a_{2,1}$	$a_{2,2}$	$\dots$	$a_{2,n}$	b_2
	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
	$a_{n,1}$	$a_{n,2}$	$\dots$	$a_{n,n}$	b_n
P	$-\alpha_1$	$-\alpha_2$	$\dots$	$-\alpha_n$	0

- While at least one entry in the last row of the tableau is negative:
  - Choose the negative entry of largest absolute value in the last row, say it is in column i.
  - For every positive entry in column i, calculate $b_k/a_{k,i}$. Choose the row which minimizes this value, say this is row j.
  - Perform Gaussian elimination on column i, using row j as a pivot (i.e. perform row operations so that column i contains a 1 in row j and 0s elsewhere.)
- When all entries in the last row are positive, the solution can be read directly from the tableau: each column labelled by a non-slack variable x_i will contain a unique non-zero entry. The corresponding value in the final column is the optimum value of x_i.

Remark 77. Note that after each iteration all of the entries in the final column are positive. It can be proved that the sum of these entries decreases at each step in the algorithm. This is somewhat analogous to the Hungarian algorithm for the assignment problem.

Example 78. We solve Example 67 using the simplex method.

- We introduce the slack variables y_1 and y_2, and rewrite our inequalities as equalities.

$$\begin{aligned} 3x_1 + 2x_2 + y_1 &= 120 \\ 4x_1 + 6x_2 + y_2 &= 260 \end{aligned}$$

where $x_1, x_2, y_1, y_2 \geq 0$.

- We rewrite our objective function: $P - 5x_1 - 6x_2 = 0$.
- Thus we obtain the initial tableau

	x_1	x_2	y_1	y_2	
	3	2	1	0	120
	4	6	0	1	260
P	-5	-6	0	0	0

- Now the most negative term in the bottom row is -6, in the second column. $\frac{260}{6} < \frac{120}{2}$, so the second row is our pivot row. (We scale the second row so that it contains a 1 in the second column. Then we use row operations to kill off the other entries in the second column: this is really just Gaussian elimination.)

	x_1	x_2	y_1	y_2	
	$\frac{5}{3}$	0	1	$\frac{-1}{3}$	$120 - \frac{260}{3}$
	$\frac{4}{6}$	1	0	$\frac{1}{6}$	$\frac{260}{6}$
P	-1	0	0	1	260

- Now the 'most negative' term in the bottom row is -1 in the first column. The smaller ratio is in the first row, so this becomes the pivot row. (We scale the first row so that its first entry is 1, and kill off the first column entries of the other rows.)

	x_1	x_2	y_1	y_2	
	1	0	$\frac{3}{5}$	$\frac{-1}{5}$	20
	0	1	$\frac{-6}{15}$	$\frac{3}{10}$	30
P	0	0	$\frac{3}{5}$	$\frac{4}{5}$	280

- Now there are no negative entries in the bottom row. So the algorithm terminates. The value of x_1 is 20 and the value of x_2 is 30. The maximal value of the objective function is 280.

5.4 Interpretation of the simplex method

As noted in Fact 74, an optimal solution to a linear programming problem necessarily occurs at a corner point of the feasible set.

We explain how the steps in the algorithm correspond to evaluating the objective function at each of the corners in turn.

Definition 79. The variables that are set equal to 0 at a corner point are *non-basic*, and the others are *basic*.

We begin our algorithm at the origin, where all of our x_is are set equal to zero. That is they are the non-basic variables. Our goal is to make the x_is basic one-by-one and to make the slack variables non-basic. This corresponds to the Gaussian elimination steps.

The choice of column corresponds to finding the direction in which $P(\overline{x})$ is increasing fastest, and moving to a corner in that direction. At that point, the algorithm starts again. We illustrate this with another example.

Example 80. Maximise $P = 5x_1 + 4x_2 + 6x_3$ subject to the constraints

$$\begin{aligned} x_1 + x_2 + x_3 &\leq 25 \\ 2x_1 + x_2 + 3x_3 &\leq 51 \end{aligned}$$

First we introduce the slack variables y_1 and y_2, to obtain

$$\begin{aligned} x_1 + x_2 + x_3 + y_1 &= 25 \\ 2x_1 + x_2 + 3x_3 + y_2 &= 51 \end{aligned}$$

Remark 81. So the feasible set is a subset of $\mathbb{R}^5$ given by the equations above together with $x_i \geq 0$ for $i = 1, 2, 3$ and $y_j \geq 0$ for $j = 1, 2$. A corner point is given by the intersection of 5 of these 7 half-spaces. This problem would be infeasible to solve using the 'corner-point' method.

So we end up with a tableau as follows:

	x_1	x_2	x_3	y_1	y_2	
	1	1	1	1	0	25
	2	1	3	0	1	51
P	-5	-4	-6	0	0	0

Our initial solution is $x_1 = x_2 = x_3 = 0$, $y_1 = 25$, $y_2 = 51$. Our non-basic variables are x_1, x_2 and x_3 and the point we start at is $(0, 0, 0)$. Notice that the columns labelled by y_1 and y_2 correspond to the identity matrix.

Remark 82. Now, we wish to move to another corner in a way which increases the objective function as quickly as possible. Choosing the negative number of largest absolute value will cause the objective function tog row fastest: we choose x_2 in this case. Now we choose the smallest positive ratio of the last column with the third column. (This ensures that all of the entries in the last column remain positive; so we still have a standard maximum problem at the next step.)

And we perform Gaussian elimination:

	x_1	x_2	x_3	y_1	y_2	
	$\frac{1}{3}$	$\frac{2}{3}$	0	1	$\frac{-1}{3}$	8
	$\frac{2}{3}$	$\frac{1}{3}$	1	0	$\frac{1}{3}$	17
P	-1	-2	0	0	2	102

Now the non-basic variables are x_1, x_2 and y_2. The basic variables are $y_1 = 8$ and $x_3 = 17$. The value of the objective function is now 102.

We now iterate the process: since there are still negative entries in the last row, our solution is not optimal. We now want to make x_2 a basic variable.

	x_1	x_2	x_3	y_1	y_2	
	$\frac{1}{2}$	1	0	$\frac{3}{2}$	$\frac{-1}{2}$	12
	$\frac{1}{2}$	0	1	$\frac{-1}{2}$	$\frac{1}{2}$	13
P	0	0	0	3	2	126

Now the non-basic variables are x_1, y_1 and y_2. The basic variables are $x_2 = 12$ and $x_3 = 13$, which gives $P = 126$. As there are no negative variables left in the bottom row, there is no possibility to increase this solution further, and the algorithm finishes.

5.5 The general linear programming problem

The standard maximum problem is a restricted version of the General Linear Programming Problem.

Definition 83. The *general linear programming problem* (GLPP) is to maximize $P = \sum_{i=1}^{n} \alpha_i x_i$ given constraints

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + \dots a_{1,n}x_n &= b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \dots a_{2,n}x_n &= b_2 \\ \vdots &= \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \dots a_{m,n}x_n &= b_m, \end{aligned}$$

such that $x_i \geq 0$ for $1 \leq i \leq n$.

The difference from the standard maximum problem is that we no longer require the b_i to be positive. This allows us to solve systems involving new types of inequalities: recall that $a \geq b$ if and only if $-a \leq -b$.

Example 84. Very often, we want to bound our variables from below as well as from above. E.g. in real world problems, for a company to produce 0 product and make 0 profit may not be a feasible solution. Thus, a large class of problems will have constraints of the form

$$b_1 \leq x_1 \leq b_2.$$

We interpret this as a pair of inequalities:

$$x_1 \leq b_2 \qquad -x_1 \leq -b_2$$

Thus any such problem may be stated as a GLPP.

Any problem involving a linear objective function and linear constraints can be put into the form of a GLPP, thus justifying the name. To solve the GLPP, we reduce it to a standard maximum problem, in which all the b_i are positive, and then use the simplex method.

Example 85. Maximise $P = 3x_1 - 2x_2$ subject to

$$\begin{aligned} 4x_1 + x_2 &\leq 16 \\ x_1 + x_2 &\geq 7 \\ x_2 - 2x_1 &\geq 4 \end{aligned}$$

First, we turn this into a GLPP.

$$\begin{aligned} 4x_1 + x_2 &\leq 16 \\ -x_1 - x_2 &\leq -7 \\ 2x_1 - x_2 &\leq -4 \end{aligned}$$

Then, as usual, we introduce slack variables. This time, we also introduce a new term, x_0, which we add to each of the b_i.

$$\begin{aligned} 4x_1 + x_2 + y_1 &= 16 + x_0 \\ -x_1 - x_2 + y_2 &= -7 + x_0 \\ 2x_1 - x_2 + y_3 &= -4 + x_0 \end{aligned}$$

This is called the *augmented problem.*

Remark 86. Basically, we do here exactly what we did in the marriage problem: we multiply by -1 to obtain a problem that we know how to solve. Then we add a suitable x_0 to every row to make all of the entries in the tableau positive. This reduces the problem to a standard maximum problem, to which we apply the simplex algorithm.

We give an example of how GLPP may be reduced to a standard maximum problem. We introduce a new row $\hat{P}$. It's only purpose is to keep track of the value we use for x_0.

	x_0	x_1	x_2	y_1	y_2	y_3	
	-1	4	1	1	0	0	16
	-1	-1	-1	0	1	0	-7
	-1	2	-1	0	0	1	-4
P	0	-3	2	0	0	0	0
$\hat{P}$	1	0	0	0	0	0	0

Now, we want to eliminate x_0 so as to make all the b_i positive. We choose the row which contains the negative term of largest absolute value (i.e. the 'most negative' term), and do row reduction on this. (Here it is the second row.)

	x_0	x_1	x_2	y_1	y_2	y_3	
	0	5	2	1	-1	0	23
	1	1	1	0	-1	0	7
	0	3	0	0	-1	1	3
P	0	-3	2	0	0	0	0
$\hat{P}$	0	-1	-1	0	1	0	-7

We now maximize $\hat{P}$ in the usual way. (We pivot the last row and third column. Note that choosing the second column would negate some of the b_i and leave us back where we started.)

	x_0	x_1	x_2	y_1	y_2	y_3	
	-2	3	0	1	1	0	9
	1	1	1	0	-1	0	7
	0	3	0	0	-1	1	3
P	-2	-5	0	0	2	0	-14
$\hat{P}$	1	0	0	0	0	0	0

Now x_0 is non-basic (i.e. $= 0$), and we can safely remove both it and $\hat{P}$ from the tableau.

	x_1	x_2	y_1	y_2	y_3	
	3	0	1	1	0	9
	1	1	0	-1	0	7
	3	0	0	-1	1	3
P	-5	0	0	2	0	-14

What remains is a standard maximum problem. We solve this using the simplex method to find $x_1 = 1$, $x_2 = 6$.

To recap, the steps to solve a GLPP are as follows:

1. Write out the augmented tableau (i.e. introducing x_0 and $\hat{P}$).
2. Choose the row with the negative entry of largest absolute value, and row reduce the x_0 column.
3. Maximise $\hat{P}$ using the simplex method (i.e. choose the columns with negative entry of largest absolute value, etc.).
4. The system has a solution **if and only if** $\hat{P}$ has maximum value 0.
5. If $\hat{P}$ has maximum value 0, then $\hat{P}$ and x_0 can be removed, leaving a standard maximum problem, which can be solved using the simplex method.

Remark 87. The **definition** of a GLPP is on the syllabus for this course, the **solution** of a GLPP is not.

5.6 The dual linear programming problem

This is most easily stated in matrix form.

Definition 88. Suppose that we have a GLPP: i.e. we wish to maximize $\overline{cx}$ subject to $A\overline{x} \leq \overline{b}$, $\overline{x} \geq 0$. Then the *dual problem* is to **minimize** $\overline{by}$ subject to $A^\top \overline{y} \geq \overline{c}$, $\overline{y} \geq 0$.

Essentially, we replace every constraint in the original problem with a variable, and every variable in the original problem with a constraint. While this is quite a natural thing to do in linear algebra, it does in fact have value in solving physical problems.

Fact 89 (The duality theorem)**.** *A linear programming problem has an optimal solution* ***if and only if*** *its dual problem has an optimal solution. Furthermore, the optimal value attained by both problems is the same.*

Definition 90. When speaking about duality, we refer to the original problem posed as the *primal problem*, and the dual problem as its *dual*.

We given an example of how to dualise a problem.

Example 91. We take as our primal problem the same problem as in the previous example: maximise $P = 5x_1 + 4x_2 + 6x_3$ subject to the constraints

$$\begin{aligned} x_1 + x_2 + x_3 &\leq 25 \\ 2x_1 + x_2 + 3x_3 &\leq 51 \end{aligned}$$

We formulate the dual problem.

In particular, we currently have $\overline{c} = (5,4,6)$, $\overline{b} = (25,51)^\top$ and

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 3 \end{pmatrix}.$$

So our dual problem is to minimize $25y_1 + 51y_2$ subject to the constraints

$$\begin{aligned} y_1 + 2y_2 &\geq 5 \\ y_1 + y_2 &\geq 4 \\ y_1 + 3y_2 &\geq 6 \end{aligned}$$

Subject to $y_1 \geq 0$, $y_2 \geq 0$. Note that the duality theorem and our solution of this problem in Example 80 tell us that the optimal solution to this problem is 126.

Quite often it is easier to solve a problem with few variables and many constraints than vice versa. Alternatively, we may find that a GLPP primal problem has a dual which is already in standard maximum form, and so is easier to solve.

5.7 Revision

Note: you will **not** be expected to solve the general linear programming problem in the exam. You may be asked to do any of the following.

- State the standard maximum problem, the general linear programming problem and the duality theorem.
- Solve standard maximum problems using the simplex method.
- Put an arbitrary problem into the form of the general linear programming problem.
- Dualise a general linear programming problem.

6 Exercises

The correspondence between exercises and chapters is as follows.
The first homework assignment consisted of problems $2(c, d),\ 4(b, c),\ 6(a),\ 9(c, d),\ 10$.

Chapter	Problems
1	1-7
2	8-15
3	16-27
4	28-32
5	33-40

The second homework assignment consisted of problems $18(b, c),\ 19,\ 23,\ 25,\ 26(b)$.
The third homework assignment consisted of problems $36,\ 38,\ 40$ (for questions $36,\ 38$ only).
Solutions for these problems is given in the next chapter.

1. Denote by $\mathcal{M}_3(\mathbb{R})$ the space of 3×3 matrices with entries in $\mathbb{R}$. Which of the the following are subspaces of $\mathcal{M}_3(\mathbb{R})_3(\mathbb{R})$? If X_i is a subspace, then provide a basis. If X_i is not a subspace, provide a counter-example.
    a) $X_1 = \{A \in \mathcal{M}_3(\mathbb{R}) \mid \mathrm{tr}(A) = 0\}$ (The trace of A is $a_{1,1} + a_{2,2} + a_{3,3}$.)
    b) $X_2 = \{A \in \mathcal{M}_3(\mathbb{R}) \mid A \text{ symmetric}\}$ (i.e. $A^\top = A$)
    c) $X_3 = \{A \in \mathcal{M}_3(\mathbb{R}) \mid \det(A) = 0\}$
    d) $X_4 = \{A \in \mathcal{M}_3(\mathbb{R}) \mid A \text{ skew-symmetric}\}$ (i.e. $A^\top = -A$)

2. Which of the following are inner products on $\mathbb{R}^2$?
    a) $\langle (x_1, x_2), (y_1, y_2) \rangle = 4x_1y_1 + 3x_2y_2$
    b) $\langle (x_1, x_2), (y_1, y_2) \rangle = x_1x_2 + y_1y_2$
    c) $\langle (x_1, x_2), (y_1, y_2) \rangle = 2x_1y_1 - x_2y_2$
    d) $\langle (x_1, x_2), (y_1, y_2) \rangle = x_1y_2 + x_2y_1$
    e) $\langle (x_1, x_2), (y_1, y_2) \rangle = x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2$

3. Let $u = (1, 2, 3)$, $v = (2, -2, 1)$ and $w = (0, 6, 5)$ be vectors in $\mathbb{R}^3$. Using the dot product, calculate the following:
    a) $\|x\|$, for $x \in \{u, v, w\}$
    b) $\langle x, y \rangle$ for $x, y \in \{u, v, w\}, x \neq y$
    c) The angle between x and y for $x, y \in \{u, v, w\}, x \neq y$

4. Find the point on each of the following curves which minimises the distance to $(0,0)$ (in $\mathbb{R}^2$, using the dot product).
   a) $5x + 2y - 1 = 0$
   b) $-2x + 4y + 8 = 0$
   c) $y = x^2 + x + 1$ (You may assume that $2x^3 + 3x^2 + 4x + 1$ has a unique real root at approximately -0.3.)
   d) $y = \frac{1}{x}$

5. Find a normal vector to the plane in $\mathbb{R}^3$ generated by each of the following pairs of vectors.
   a) $u = (1,2,3)$, $v = (-1,0,1)$
   b) $u = (2,4,5)$, $v = (-1,2,4)$

6. Find a normal vector to the following planes in $\mathbb{R}^3$.
   a) $x + y - z = 0$
   b) $3x - y + 2z = 0$

7. Denote by $\mathcal{F}$ the set of functions $f(x)$ such that $\int_{-\pi}^{\pi} f(x)\,\mathrm{dx}$ exists and is finite.
   a) Prove that $\mathcal{F}$ is a vector space.
   b) Show that $\langle f(x), g(x)\rangle = \int_{-\pi}^{\pi} f(x)g(x)\,\mathrm{dx}$ is an inner product on $\mathcal{F}$.
   c) Calculate $\|\mathrm{Cos}(\theta)\|$, $\|\mathrm{Sin}(\theta)\|$
   d) Hence calculate the angle between $\mathrm{Cos}(\theta)$ and $\mathrm{Sin}(\theta)$.

8. Let $u = (1,2,3,4)$ and $v = (1,1,1,-1)$ be vectors in $\mathbb{R}^4$. Find the component of u which is orthogonal to v. Hence find the component of u which is parallel to v.

9. Apply the Gram-Schmidt process to the following sets of vectors. $\mathbb{R}^n$ is equipped with the dot product, and $\mathcal{F}$ with the integral on $[-\pi, \pi]$.
   a) $(1,1,0)$, $(1,0,1)$ and $(0,1,1)$ in $\mathbb{R}^3$
   b) $(1,2,3)$, $(2,3,4)$, $(3,4,5)$ and $(4,5,6)$ in $\mathbb{R}^3$
   c) $(0,1,-1,0)$, $(2,1,1,1)$ and $(3,1,-1,0)$ in $\mathbb{R}^4$
   d) $\mathrm{Cos}(\theta)$, $\mathrm{Cos}(2\theta)$ and $\mathrm{Sin}(\theta)$ in $\mathbb{F}$

10. Find an orthonormal basis for the solution space in $\mathbb{R}^5$ of the equations

$$\begin{aligned} x_1 + x_2 + x_3 - x_4 - x_5 &= 0 \\ 2x_1 + x_2 - x_3 + 3x_4 + x_5 &= 0 \end{aligned}$$

11. Find a basis in $\mathbb{R}^3$ which is orthonormal with respect to the inner product

$$\langle (x_1,x_2,x_3),(y_1,y_2,y_3)\rangle = 2x_1y_1 + x_2y_2 + x_3y_2 + x_2y_3 + 2x_3y_3$$

12. Find the component of $(1,0,-1,2)$ which lies in the subspace spanned by the vectors $(1,2,3,0)$ and $(-1,0,2,0)$ in $\mathbb{R}^4$.
13. Let $\mathbb{R}_2[x] = \{a_0 + a_1x + a_2x^2 \mid a_i \in \mathbb{R}\}$ be the space of polynomials of degree at most 2. Let $\mathbb{R}_3[x]$ be the space of polynomials of degree at most 3. Find an orthonormal basis of each space with respect to the following inner products
    a) $\langle f,g\rangle = \int_0^1 f(x)g(x)\,\mathrm{d}x$
    b) $\langle f,g\rangle = \int_0^2 f(x)g(x)\,\mathrm{d}x$
    c) $\langle f,g\rangle = \int_{-1}^1 f(x)g(x)\,\mathrm{d}x$
14. Complete the proof of Theorem 13.
15. Find the first three non-zero Fourier coefficients of the following functions, with respect to the inner product

$$\langle f,g\rangle = \int_{-\pi}^{\pi} f(x)g(x)\,\mathrm{d}x$$

    a) $f(x) = 1$ for $-\pi \le x \le \pi$
    b) $f(x) = -1$ for $\pi \le x \le \pi$
    c) $f(x) = |x|$
    d) $f(x) = \frac{\pi}{2}$ for $-\pi \le x < 0$, and $\frac{-\pi}{2}$ for $0 \le x \le \pi$
    e) $f(x) = 0$ for $-\pi \le x < 0$, and 1 for $0 \le x \le \pi$
16. For each of the following pairs of bases, find a matrix which maps vectors expressed relative to B_1 to vectors expressed relative to B_2.
    a) $B_1 = (1,0),(0,1)$ and $B_2 = (1,1),(1,-1)$.
    b) $B_1 = (2,4),(-1,1)$ and $B_2 = (2,-1),(-1,-1)$.
    c) $B_1 = (3,3),(1,2)$ and $B_2 = (-1,1),(2,2)$.
17. Find the eigenvalues and eigenvectors of each of the following matrices. Hence diagonalize them.

$$\text{(i)} \begin{pmatrix} 2 & 9 \\ 1 & 2 \end{pmatrix} \quad \text{(ii)} \begin{pmatrix} 3 & 2 \\ 7 & -2 \end{pmatrix} \quad \text{(iii)} \begin{pmatrix} 0 & 4 \\ 4 & 6 \end{pmatrix}$$

18. Write down the matrix of the bilinear form $\langle (u_1,u_2),(v_1,v_2)\rangle = u_1v_1 + u_1v_2 - u_2v_1 + 2u_2v_2$ with respect to the following bases for $\mathbb{R}^2$.
    a) $B_1 = \{(1,0),(0,1)\}$
    b) $B_2 = \{(1,1),(1,-1)\}$
    c) $B_3 = \{(1,1),(-1,2)\}$

19. Find the change of basis matrix X from B_3 to B_2, and verify that $X^\top M_{B_2} X = M_{B_3}$ where M_{B_2} and M_{B_3} are the matrices obtained in parts (b) and (c) of Question 3 respectively.

20. Write down a bilinear form for each of the following matrices relative to the bases $\{(1,0),(0,1)\}$ and $\{(1,1),(1,-1)\}$.

$$\text{(i)} \begin{pmatrix} 2 & 3 \\ 1 & 2 \end{pmatrix} \quad \text{(ii)} \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix} \quad \text{(iii)} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$$

21. Show that the set of $\mathbb{S}$ of 3×3 symmetric matrices over $\mathbb{R}$ is a vector space over $\mathbb{R}$.
    a) What is its dimension?
    b) Write down a basis for $\mathbb{S}$.
    c) Use the Gram-Schmidt process to find an orthonormal basis for $\mathbb{S}$ relative to the inner product $\langle A, B\rangle = \operatorname{tr} AB$.

22. Suppose that M is an $n \times n$ orthogonal matrix. Show that $M^\top$ is orthogonal. Hence show that the rows of an orthogonal matrix form a basis for $\mathbb{R}^n$.

23. Let M be a real orthogonal matrix. Show that all of the entries satisfy the inequality $-1 \leq m_{i,j} \leq 1$.

24. Suppose that M and N are $n \times n$ orthogonal matrices. Show that MN is orthogonal.

25. Show that an $n \times n$ matrix M which is both symmetric and orthogonal satisfies $M^2 = I_n$. Hence find all 2×2 symmetric orthogonal matrices.

26. Orthogonally diagonalize the following matrices.

$$\text{(a)} \begin{pmatrix} 1 & 3 & 2 \\ 3 & 4 & 1 \\ 2 & 1 & 9 \end{pmatrix}, \quad \text{(b)} \begin{pmatrix} 1 & 3 & 1 \\ 3 & 1 & 1 \\ 1 & 1 & 3 \end{pmatrix}$$

    You may assume that the (a) has an eigenvalue at -1.

27. Find a basis of $\mathbb{R}^3$ which polarises each of the bilinear forms below.
    a) $\langle (x_1, x_2, x_3), (y_1, y_2, y_3)\rangle = x_1y_1 + x_2y_2 + x_3y_3 + x_1y_2 + x_1y_3 + x_2y_3$
    b) $\langle (x_1, x_2, x_3), (y_1, y_2, y_3)\rangle = x_1y_1 + 2x_3y_3 + x_1y_2 - x_1y_3 + 3x_2y_3$

28. A building contractor owns 5 cranes, located in five different locations. These must be moved to five new building sites. The cost of moving crane i to site j is given in the following matrix. Find an optimal assignment of cranes to building sites.

$$C = \begin{pmatrix} 6 & 12 & 3 & 11 & 15 \\ 4 & 2 & 7 & 1 & 10 \\ 8 & 11 & 10 & 7 & 11 \\ 16 & 19 & 12 & 23 & 21 \\ 9 & 5 & 7 & 6 & 10 \end{pmatrix}$$

29. A matchmaker has five grooms and four brides on her books. Their compatibility is given by the following matrix. Find an optimal matching.

$$C = \begin{pmatrix} 7 & 4 & 7 & 3 & 10 \\ 5 & 9 & 3 & 8 & 7 \\ 3 & 5 & 6 & 2 & 9 \\ 6 & 5 & 0 & 4 & 8 \end{pmatrix}$$

30. An antiques dealer is holding a postal auction on 4 paintings, and has received bids in advance from 5 bidders. These are arranged in the following matrix. Maximize the sum of the successful bids, if each bidder is to receive at most 1 painting.

$$C = \begin{pmatrix} 13 & 15 & 9 & 7 \\ 14 & 12 & 10 & 7 \\ 11 & 16 & 11 & 9 \\ 10 & 12 & 11 & 10 \\ 13 & 16 & 8 & 8 \end{pmatrix}$$

31. a) In the auction of the previous question, the second bidder sends a message that he is willing to buy up to two paintings at the prices previously agreed. Can the dealer increase his profit?

    b) The dealer forges a copy of the fourth painting, and sells it for the same price as the original. What is the maximum profit if he sells a single painting to each bidder?

32. Graph the feasible set of each of the following sets of constraints. In all cases you may assume that $x_1 \geq 0$, $x_2 \geq 0$.

    a) $x_1 \leq 3$, $x_2 \leq 4$

    b) $3x_1 + 5x_2 \leq 15$

    c) $4x_1 + 6x_2 \leq 24$, $x_2 - x_1 \leq 1$, $x_1 \leq 4$

    d) $5x_1 + 5x_2 \leq 25$, $x_2 - x_1 \leq 1$, $x_1 - 4x_2 \leq 3$

33. Recall that the maximum value of an objective function occurs at a corner point of the feasible set. List the corner points of each of feasible set in Question 5. Hence maximize the following objective functions on each feasible set.

    a) $P = x_1 + x_2$

b) $P = 5x_1 + 3x_2$

c) $P = 2x_2 - 3x_1$

34. Now use the simplex method to maximize the objective functions given in Question 6 (*b*) and (*c*) subject to the constraints in Question 5 (*c*) and (*d*). Verify that the method produces the same answer as Question 6.

35. Use the simplex method to maximize $P = 5x_1 + 2x_2 + x_3$ subject to the constraints

$$\begin{aligned} x_1 + 3x_2 - x_3 &\le 6 \\ x_2 + x_3 &\le 4 \\ 3x_1 + x_2 &\le 7 \end{aligned}$$

36. Use the simplex method to maximise $P = 2x_1 + 2x_2 - x_3 + 3x_4$ subject to the constraints

$$\begin{aligned} 4x_1 + 6x_2 + 2x_3 - 2x_4 &\le 7 \\ 3x_1 + x_2 - 4x_3 + 5x_4 &\le 8 \end{aligned}$$

37. You run a pension fund. Three types of government bonds are available: German bonds paying 5% per annum, Spanish bonds paying 6% and bonds Irish bonds paying 9%. Maximise the return subject to the following constraints:
    - The total amount invested cannot exceed one million euro.
    - No more that €200,000 can be invested in Irish bonds.
    - The amount invested Irish bonds cannot exceed the amount invested in German bonds.

38. A power plant burns coal, oil and gas to generate electricity.
    - A tonne of coal generates 600 kilowatt hours, emits 20 units of sulphur dioxide and 15 units of particulate matter, and costs €200.
    - A tonne of oil generates 550 kilowatt hours, emits 18 units of sulphur dioxide and 12 units of particulate matter, and costs €220.
    - A tonne of gas generates 500 kilowatt hours, emits 15 units of sulphur dioxide and 10 units of particulate matter, and costs €250.

    European Union guidelines restrict the daily emission of sulphur dioxide to no more than 60 units, and that of particulate matter to no more than 75 units. If the daily fuel budget of the power plant is €2000, what is the maximum amount of electricity that can be generated while staying within the emission limits?

39. Express the following linear programming problem as a standard maximum problem: Maximise $P = x_1 - 3x_2 - 8x_3$ subject to the constraints

$$\begin{aligned} 2x_2 - 20 &\leq 2x_1 - x_3 \\ -x_1 + 2x_2 - 2x_3 &\geq -4 \\ 16 - 4x_1 - 8x_2 + x_3 &\geq 0 \end{aligned}$$

40. State the dual of each of the linear programming problems in Questions $8, 9, 10, 11$.

7 Solutions to homework

7.1 Homework 1

Question $2(c)$**:**

$$\langle (x_1, x_2), (y_1, y_2) \rangle = 2x_1y_1 - x_2y_2$$

is not an inner product. Consider the vector $(0, 1)$:

$$\langle (0, 1), (0, 1) \rangle = 2(0)^2 - (1)^2 = -1.$$

This is negative, which violates property (5) of an inner product.

Question $2(d)$**:**

$$\langle (x_1, x_2), (y_1, y_2) \rangle = x_1y_2 + x_2y_1$$

is not an inner product. Consider the vector $(1, -1)$:

$$\langle (1, -1), (1, -1) \rangle = 2(1)(-1) + (-1)(1) = -3.$$

This is negative, which violates property (5) of an inner product.

Question $4(b)$**:**
We minimize the distance from $-2x + 4y + 8 = 0$ to the origin.

- We put the line in parametric form:
$$y = \frac{x}{2} - 2.$$
- So every point on the line is of the form $(t, \frac{t}{2} - 2)$.
- The distance to the origin is $f(t) = \sqrt{t^2 + (\frac{t}{2} - 2)^2}$.
- We express $f(t) = \sqrt{\frac{5t^2}{4} - 2t + 4}$, and minimize.
- The derivative of $f(t)$ is
$$f'(t) = \frac{1}{\sqrt{\frac{5t^2}{4} - 2t + 4}} \left(\frac{5t}{2} - 2 \right)$$

- To solve $f'(t) = 0$, we can multiply on both sides by $\frac{5t^2}{4} - 2t + 4$, since the roots of this polynomial are complex.
- So $f'(t) = 0$ if and only if $\frac{5t}{2} - 2 = 0$, i.e. when $t = \frac{4}{5}$.
- We substitute this value into the formula for a point; the solution is $(\frac{4}{5}, \frac{-8}{5})$.

Question $4(c)$**:**
We minimize the distance from $y = x^2 + x + 1$ to the origin.

- Every point on the curve is of the form $(t, t^2 + t + 1)$.
- The distance to the origin is $f(t) = \sqrt{t^2 + (t^2 + t + 1)^2}$.
- The derivative of this function is

$$f'(t) = \frac{1}{\sqrt{t^2 + (t^2 + t + 1)^2}}(2t + 2(t^2 + t + 1)(2t + 1)).$$

- The roots of the term under the line are complex, so the distance is defined, we can multiply on both sides by this term.
- So we need to solve $4t^3 + 6t^2 + 8t + 2 = 0$, but dividing both sides by 2 this is $2t^3 + 3t^2 + 4t + 1 = 0$, and the unique real root of this polynomial was given as -0.3.
- So the point on the curve which minimizes the distance to $(0, 0)$ is (very close to) the point $(-0.3, 0.79)$.

Question $6(a)$**:**

- We find a normal vector to $x + y - z = 0$
- Every point in this plane is of the form $(x, y, x + y)$.
- So a vector (a, b, c) is orthogonal to the plane if and only if

$$(a, b, c) \bullet (x, y, x + y) = ax + by + c(x + y) = 0.$$

- This means that $(a + c)x = 0$, and that $(b + c)y = 0$ for all choices of x and y.
- So $a + c = 0$, and $b + c = 0$.
- So a vector orthogonal to $x + y - z = 0$ is of the form $(t, t, -t)$.
- We want a vector of length 1 on this line: $\pm\frac{1}{\sqrt{3}}(1, 1, -1)$.

Question $9(c)$**:**

- We apply the Gram-Schmidt process to $(0,1,-1,0)$, $(2,1,1,1)$ and $(3,1,-1,0)$ in $\mathbb{R}^4$
- We set $s_1 = (0,1,-1,0)$.
- We calculate:

  $$s_2 = (2,1,1,1) - \frac{(2,1,1,1) \bullet (0,1,-1,0)}{(0,1,-1,0) \bullet (0,1,-1,0)}(0,1,-1,0) = (2,1,1,1)$$

  (This means that s_1 and s_2 were already orthogonal.)
- We calculate:

  $$\begin{aligned} s_3 &= (3,1,-1,0) - \frac{(3,1,-1,0) \bullet (2,1,1,1)}{(2,1,1,1) \bullet (2,1,1,1)}(2,1,1,1) - \frac{(3,1,-1,0) \bullet (0,1,-1,0)}{(0,1,-1,0) \bullet (0,1,-1,0)}(0,1,-1,0) \\ s_3 &= (3,1,-1,0) - \frac{6}{7}(2,1,1,1) - \frac{2}{2}(0,1,-1,0) \\ &= \frac{1}{7}(9,-1,-1,-1) \end{aligned}$$

- Finally, we normalise: our solution is $b_1 = \frac{1}{\sqrt{2}}(0,1,-1,0)$, $b_2 = \frac{1}{\sqrt{7}}(2,1,1,1)$ and $b_3 = \frac{2\sqrt{21}}{7}(9,-1,-1,-1)$.

Question $9(d)$**:**

- We apply the Gram-Schmidt process to $\mathrm{Cos}(\theta)$, $\mathrm{Cos}(2\theta)$ and $\mathrm{Sin}(\theta)$ in $\mathbb{F}$.
- In fact the three functions given are already orthogonal.
- So we only have to normalise: an orthonormal basis is $\frac{1}{\sqrt{\pi}}\mathrm{Cos}(\theta)$, $\frac{1}{\sqrt{\pi}}\mathrm{Cos}(2\theta)$ and $\frac{1}{\sqrt{\pi}}\mathrm{Sin}(\theta)$

Question 10**:**

- We find an orthonormal basis for the space of solutions in $\mathbb{R}^5$ of the simultaneous equations

  $$\begin{aligned} x_1 + x_2 + x_3 - x_4 - x_5 &= 0 \\ 2x_1 + x_2 - x_3 + 3x_4 + x_5 &= 0 \end{aligned}$$

- First, these are two simultaneous equations in 5 unknowns, so the space of solutions is three dimensional. We begin by finding a basis for this space.

- We find the general form of a point satisfying the first equation (solving for x_5):
$$p = (x_1, x_2, x_3, x_4, x_1 + x_2 + x_3 - x_4)$$
- Now we substitute these values into the second equation to find that
$$3x_1 + 2x_2 + 2x_4 = 0.$$
- So we can express x_4 as $-\frac{3}{2}x_1 - x_2$.
- We substitute this value for x_4 into our equation for p, so a point in the space of solutions has form
$$p = (x_1, x_2, x_3, -\frac{3}{2}x_1 - x_2, \frac{5}{2}x_1 + 2x_2 + x_3).$$
- The set of all points of the form p form a three dimensinal subspace of $\mathbb{R}^5$.

From this point, there are many possible choices of basis. An obvious one is obtained by setting $x_1 = 1$ with $x_2 = x_3 = 0$, then $x_2 = 1$ and $x_1 = x_3 = 0$ and finally $x_3 = 1$, $x_1 = x_2 = 0$. This gives

$$T_1 = \left\{(1, 0, 0, \frac{-3}{2}, \frac{5}{2}), (0, 1, 0, -1, 2), (0, 0, 1, 0, 1)\right\}.$$

Applying the Gram-Schmidt process, one obtains the orthonormal basis:

$$B_1 = \left\{\sqrt{\frac{2}{19}}(1, 0, 0, \frac{-3}{2}, \frac{5}{2}), \sqrt{\frac{1}{2242}}(-26, 38, 0, 1, 11), \sqrt{\frac{1}{1618724}}(-152, -209, 1121, 437, 323)\right\}.$$

This is a valid solution, but is unpleasant to work with. A better choice of basis is as follows:

$$T_2 = \{(0, -1, 2, 1, 0), (0, 0, 1, 0, 1), (2, 0, 0, -3, 5)\}.$$

We found this by solving for $x_4 = 1, x_5 = 0$ and $x_5 = 1, x_4 = 0$ in the formula for p. (Setting as a many coordinates equal to zero as possible.) Then we observed that our first two points had $x_1 = 0$, so anything with $x_1 \neq 0$ will be linearly independent from the first two vectors. We set $x_1 = 2$ to eliminate fractions.

We then apply the Gram-Schmidt process, setting $s_1 = (0, 0, 1, 0, 1)$ because this vector is easiest to work with. We obtain the orthonormal basis

$$B_2 = \left\{(\frac{1}{\sqrt{2}}(0, 0, 1, 0, 1), \frac{1}{2}(0, -1, 1, 1, -1), \sqrt{\frac{2}{51}}(2, -4, \frac{3}{2}, 1, \frac{3}{2})\right\}$$

A better choice of the third vector for T_2 could probably have given us a nicer final element for B_2. But this still a much nicer basis than B_1.

7.2 Homework 2

Question 18(b, c)**:**
Recall that the matrix of a bilinear form relative to a basis B is formed by placing $\langle b_i, b_j\rangle$ in row i and column j of the matrix M_B.

$$M_{B_2} = \begin{pmatrix} 3 & -3 \\ 1 & 3 \end{pmatrix} \quad M_{B_3} = \begin{pmatrix} 3 & 6 \\ 0 & 9 \end{pmatrix}$$

Question 19**:**
As in Example 2 in the notes on *Symmetric and orthogonal matrices*, we compute the change of basis matrix from B_3 to B_2 as

$$X = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 2 & 1 \\ 0 & -3 \end{pmatrix}.$$

And then we compute that

$$\frac{1}{4}\begin{pmatrix} 2 & 0 \\ 1 & -3 \end{pmatrix}\begin{pmatrix} 3 & -3 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 0 & -3 \end{pmatrix} = \begin{pmatrix} 3 & 6 \\ 0 & 9 \end{pmatrix}$$

as required.

Question 23**:**
Let M be an $n \times n$ real orthogonal matrix. Denote by m_{ij} the entry of M in row i and column j. Since the columns of M form an orthonormal basis for $\mathbb{R}^n$ with respect to the dot product, we have the identity:

$$\sum_{j=1}^{n} m_{ij}^2 = 1$$

But the entries of M are real, so $m_{ij}^2 \geq 0$ for all $1 \leq j \leq n$, we conclude that $m_{ij}^2 \leq 1$ for any choice of i and j. Hence $-1 \leq m_{ij} \leq 1$ as required.

Question 25**:**
An orthogonal matrix satisfies $M^\top M = I_n$, and a symmetric matrix satisfies $M^\top = M$. Hence a symmetric orthogonal matrix satisfies $MM = I_n$, or $M^2 = I_n$.

As seen in class, every 2×2 orthogonal matrix is of one of the following types:

$$M_\theta = \begin{pmatrix} \operatorname{Cos}(\theta) & -\operatorname{Sin}(\theta) \\ \operatorname{Sin}(\theta) & \operatorname{Cos}(\theta) \end{pmatrix}, \qquad M'_\theta = \begin{pmatrix} \operatorname{Cos}(\theta) & \operatorname{Sin}(\theta) \\ \operatorname{Sin}(\theta) & -\operatorname{Cos}(\theta) \end{pmatrix}$$

We find all symmetric matrices of type M_θ first: $\operatorname{Sin}(\theta) = -\operatorname{Sin}(\theta)$ implies that $\operatorname{Sin}(\theta) = 0$. Thus, $\theta = 0$ or $\theta = \pi$, so the only possibilities here are I_2 and $-I_2$.

Those of type M'_θ are all symmetric. Thus the set of 2×2 symmetric orthogonal matrices is as follows:

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} \mathrm{Cos}(\theta) & \mathrm{Sin}(\theta) \\ \mathrm{Sin}(\theta) & -\mathrm{Cos}(\theta) \end{pmatrix} \mid 0 \leq \theta < 2\pi \right\}.$$

Question $26(b)$**:**

- The characteristic polynomial of the matrix M is $\chi(\lambda) = \lambda^3 - 5\lambda^2 - 4\lambda + 20$.
- By inspection, this has roots at -2, 2 and 5. (We evaluated $\chi(1), \chi(-1), \chi(2)$ etc. to find a root.)
- The corresponding eigenvectors are $(-1, 1, 0), (-1, -1, 2), (1, 1, 1)$ (in that order).
- We observe that these eigenvectors are orthogonal. We normalize them to obtain the basis

$$\frac{1}{\sqrt{2}}(-1, 1, 0),\ \frac{1}{\sqrt{6}}(-1, -1, 2),\ \frac{1}{\sqrt{3}}(1, 1, 1).$$

- So the matrix

$$O = \begin{pmatrix} \frac{-1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}$$

is orthogonal.

- Finally, we observe that $O^\top M O$ is indeed a diagonal matrix.

7.3 Homework 3

Question 36**:**

We arrange the problem in a tableau.

	x_1	x_2	x_3	x_4	y_1	y_2	
	4	6	2	-2	1	0	7
	3	1	-4	5	0	1	8
P	-2	-2	1	-3	0	0	0

The most negative entry is -3 so we pivot in the column labelled by x_4. Now, the only positive entry in this column is in the second row, so we pivot about this.

	x_1	x_2	x_3	x_4	y_1	y_2	
	$\frac{26}{5}$	$\frac{32}{5}$	$\frac{2}{5}$	0	1	$\frac{2}{5}$	$\frac{51}{5}$
	$\frac{3}{5}$	$\frac{1}{5}$	$\frac{-4}{5}$	1	0	$\frac{1}{5}$	$\frac{8}{5}$
P	$\frac{-1}{5}$	$\frac{-7}{5}$	$\frac{-7}{5}$	0	0	$\frac{3}{5}$	$\frac{24}{5}$

For the next step, we can choose either the column labelled by x_2 or the column labelled by x_3. We choose x_3 for now (in either case we pivot around the first row).

	x_1	x_2	x_3	x_4	y_1	y_2	
	13	16	1	0	$\frac{5}{2}$	1	$\frac{51}{2}$
	11	13	0	1	2	1	22
P	18	21	0	0	$\frac{7}{2}$	2	$\frac{81}{2}$

We observe that there are no longer any negative entries in the last row, so we are done. We read the solution from the tableau as $x_1 = x_2 = 0$, $x_3 = 25.5$, $x_4 = 22$. We observe that this does in fact meet both conditions (so corresponds to a corner point), and is optimal, with $P = 40.5$.

Question 38**:**

We first express the problem as a standard maximum problem. Denote the number of tonnes of coal bought by x_1, the number of tonnes of oil by x_2 and the number of tonnes of gas by x_3. Then we wish to maximize $P = 600x_1 + 550x_2 + 500x_3$ subject to the constraints:

$$20x_1 + 18x_2 + 15x_3 \leq 60, \quad 15x_1 + 12x_2 + 10x_3 \leq 75, \quad 200x_1 + 220x_2 + 250x_3 \leq 2000.$$

In tableau form:

	x_1	x_2	x_3	y_1	y_2	y_3	
	20	18	15	1	0	0	60
	15	12	10	0	1	0	75
	200	220	250	0	0	1	2000
P	-600	-550	-500	0	0	0	0

The most negative entry is in the first column, and we see that $\frac{60}{20} < \frac{75}{15} < \frac{2000}{200}$, so we pivot on the first row to obtain the following tableau.

	x_1	x_2	x_3	y_1	y_2	y_3	
	1	$\frac{9}{10}$	$\frac{3}{4}$	$\frac{1}{20}$	0	0	3
	0	$\frac{-3}{2}$	$\frac{-5}{4}$	$\frac{-3}{4}$	1	0	30
	0	40	100	-10	0	1	1400
P	0	-10	-50	30	0	0	1800

Now the most negative term is in the column labelled by x_3. We ignore the second row, as the entry there is negative. Since $3(\frac{4}{3}) < \frac{1400}{100}$, we pivot in the first row again.

	x_1	x_2	x_3	y_1	y_2	y_3	
	$\frac{4}{3}$	$\frac{6}{5}$	1	$\frac{1}{15}$	0	0	4
	$\frac{5}{4}$	0	0	$\frac{-2}{3}$	1	0	35
	$\frac{-400}{3}$	-80	0	$\frac{-50}{3}$	0	1	1000
P	$\frac{200}{3}$	50	0	$\frac{100}{3}$	0	0	2000

All entries in the final row are positive. Thus we are finished. The pivot columns are x_3, y_2 and y_3. The variables x_1 and x_2 are non-basic here, so we set them to be zero. The slack variables y_2 and y_3 do note appear in P, so the optimal solution is $x_1 = x_2 = 0$, $x_3 = 4$, which generates 2000 kilowatt hours of electricity. We observe that the main constraint on the power plant's production is the sulfur dioxide allowance.

Question 40**:**

For question 36, we obtain the problem: Minimize the function $P^*(y_1, y_2) = 7y_1 + 8y_2$ subject to the constraints

$$\begin{aligned} 4y_1 + 3y_2 &\geq 2 \\ 6y_1 + y_2 &\geq 2 \\ 2y_1 - 4y_2 &\geq -1 \\ -2y_1 + 5y_2 &\geq 3 \end{aligned}$$

where $y_1, y_2 \geq 0$.

For question 38, we obtain the problem Minimize the function $P^*(\overline{y}) = 60y_1 + 75y_2 + 2000y_3$ subject to the constraints

$$\begin{aligned} 20y_1 + 15y_2 + 200y_3 &\geq 600 \\ 18y_1 + 12y_2 + 220y_3 &\geq 550 \\ 15y_1 + 10y_2 + 250y_3 &\geq 500 \end{aligned}$$

where $y_1, y_2, y_3 \geq 0$.