Engsci 721 – IPLD Supplement I

Oliver Maclaren (oliver.maclaren@auckland.ac.nz)

Overview

Some helpful background facts on linear spaces and linear algebra.

A lot is beyond the scope of this course, but by the end of the lecture material you should be able to answer the **'Test your understanding'** questions in each section. I will give you similar additional handouts as we go.

Spaces, norms, inner products etc

Mathematical spaces

To define continuity, distance, size etc in general mathematical spaces we need to abstract these concepts beyond their usual setting (e.g. beyond the real number line $\mathbb{R}$). This leads to ideas like **topological spaces**, **metric spaces**, **vector spaces**, **normed vector spaces**, **inner product vector spaces** etc. Intuitively, we can think of these spaces as **sets (collections of things) containing objects (elements) which all 'behave'/can be manipulated according to 'common rules'**.

A **core concept of linear algebra is the vector space**; another common name for vector spaces is **linear spaces**. According to the above idea of spaces, a 'vector' is *defined* as an element of a vector space, which in turn is defined by the rules describing how the elements can be manipulated. **Any objects that follow these rules can be represented as vectors**. Briefly, these state that we can add vectors and multiply them by numbers in the ways expected. More formally…

Vector space

A **vector space** is a set V of objects, that we will call vectors, that can be manipulated according to the operations of **vector addition** and **scalar multiplication**, where the **scalars** are elements

of a set $\mathbb{F}$ of numbers, such as $\mathbb{R}$ or $\mathbb{C}$, and the operations satisfy (note: the symbol $\forall$ means 'for all'):

(1) $u+v=v+u \quad \forall u,v \in V$
(2) $(u+v)+w=u+(v+w) \quad \forall u,v \in V$
(3) there is a zero element (vector) $0 \in V$ such that $0+v=v \ \forall v \in V$
(4) for all $v \in V$, there is an element (vector) $-v \in V$ such that $v+(-v)=0$
(5) $1v=v \ \forall v \in V, 1 \in \mathbb{F}$
(6) $\alpha(\beta v)=(\alpha\beta)v \ \forall v \in V, \alpha, \beta \in \mathbb{F}$
(7) $\alpha(u+v)=\alpha u+\alpha v$ and $(\alpha+\beta)v=\alpha v+\beta v \ \forall u,v \in V, \alpha,\beta \in \mathbb{F}$

We call V a vector space *over* $\mathbb{F}$. (In general, the field $\mathbb{F}$ of numbers is also a vector space over itself, as we can add scalars and multiply them by scalars etc!)

While we won't make much use of the fine distinctions between different possible spaces and mathematical structures, keeping in mind the available structure can be important. For example, in contrast to vector spaces, neither topological nor metric spaces come with any algebraic structure on the elements in general: it doesn't necessarily make sense to 'add' two elements of a general metric space together (though we can if the space is *also* a vector space). On the other hand, not all vector spaces come with a notion of 'distance between' or 'size of' vectors (though they do if they are *also* e.g. metric or normed spaces). For more detail on these spaces and their relationships, see e.g. courses on **functional analysis**.

In this course we work in the fairly familiar $\mathbb{R}^n$, which is a **normed vector space** (also an inner product space): we can add vectors together, multiply them by scalars and measure their 'size'. This latter notion is defined via the concept of a **norm**. (Side note: we will also allow any standard operations that follow from treating them as $n \times 1$ arrays/matrices, which goes beyond just a normed vector space.)

Norms - general definition

Given a vector space X, a **norm** is a function defined on elements $x \in X$ such that

(1) $||x|| = 0$ iff $x = 0$
(2) $||ax|| = |a| \ ||x|| \ \forall x \in X, a \in \mathbb{R}$
(3) $||x+y|| \leq ||x|| + ||y|| \ \forall x, y \in X$.

Intuitively, a norm **measures the size of a vector**, and generalises the usual concept of **magnitude** $|a|$ for $a \in \mathbb{R}$ to $\mathbb{R}^n$.

Norms - examples

We consider the p-norms, also called L_p-norms, in $\mathbb{R}^n$:

$$||x||_p = (|x_1|^p + ... + |x_n|^p)^{1/p}$$

where $p \geq 1$. Particular cases include

- $p = 2$, the **Euclidean/$\mathbf{L_2}$ norm**.
- $p = 1$, the **taxicab/$\mathbf{L_1}$ norm**
- $p \to \infty$, the **max/$\mathbf{L_\infty}$ norm** where $||x||_\infty = \max_i |x_i|$

Norms - relationships

As discussed in class, there is a sense in which all norms in finite dimensional $\mathbb{R}^n$ are 'equivalent' and each can approximate the others. As n grows large, however, these approximations become much looser. Many difficulties of the infinite-dimensional case are inherited by the large finite n case due to e.g. numerical rounding.

Furthermore, the **solutions** to problems involving optimisation with respect to a norm are often **qualitatively different in character** even if they have **quantitatively similar sizes**. For example, minimising with respect to the L_1 norm tends to produce solutions with many **exactly zero** entries, while the L_2 norm produces entries with many **close-to-but-not-exactly-zero** entries.

- E.g. in an $\mathbb{R}^3$ problem we might get L_1 solutions of the form $(0,0,1)^T$ and L_2 solutions of the form $\frac{1}{\sqrt{3}}(1,1,1)^T$. Both have norm of 1 in their respective norms, but e.g. $||\frac{1}{\sqrt{3}}(1,1,1)^T||_1 = \sqrt{3} > ||(0,0,1)^T||_1 = 1$.

Thus if, for example, **true zeros** or **sparse** solutions are desired, the L_1 norm would be a more natural choice than the L_2 norm.

Inner products

The space $\mathbb{R}^n$ is also an **inner product** space. (If you're keeping track, inner product spaces are special cases of normed spaces, which are special cases of metric spaces, which are special cases of topological spaces…).

This means that we can take inner products $\langle x, y\rangle$ between vectors x, y, where an inner product $\langle\cdot,\cdot\rangle$ satisfies, for all x, y, z in an inner produce space X and $a, b \in \mathbb{R}$:

$$
\begin{aligned}
&(1)\quad \langle x, x\rangle \geq 0 \text{ and } \langle x, x\rangle = 0 \text{ iff } x = 0\\
&(2)\quad \langle x, y\rangle = \langle y, x\rangle\\
&(3)\quad \langle ax + by, z\rangle = a\langle x, z\rangle + b\langle y, z\rangle
\end{aligned}
$$

Note that (2) and (3) imply that $\langle\cdot,\cdot\rangle$ is a **linear** function in both arguments (i.e. is bilinear).

Roughly, you can think of the inner product as first **projecting** one vector onto the other, giving something like two elements in $\mathbb{R}$ (i.e. two points lying along a common line), and then multiplying these two elements in the usual $\mathbb{R}$ way.

Inner products and dot products

In $\mathbb{R}^n$ we also have the **dot product** of two vectors x, y defined via

$$x \cdot y = x^T y = x_1 y_1 + \ldots + x_n y_n$$

where we take vectors to be **column vectors** by default. This leads to the **dot product version of the inner product**

$$\langle x, y\rangle = x \cdot y = x^T y$$

which we will assume by default. Note though that other inner products can be used.

Inner products and norms

Given an inner produce one can define the **inner product norm** via

$$||x|| = \langle x, x\rangle^{1/2}$$

We use this relationship often, especially in the form

$$||x||^2 = \langle x, x\rangle$$

e.g. when using the squared L_2 norm. Along with the previous note on dot products we have

$$||x||^2 = \langle x, x\rangle = x^T x$$

We will **usually convert expressions involving norms to their expressions in terms of standard matrix/vector operations**, which we will also learn to manipulate in different ways, e.g. differentiate, later.

Test your understanding of norms and inner products

- Use the usual norms and inner products on $\mathbb{R}^n$ to show that our least-squares objective function $||Ax - y||_2^2$ can be written as $x^T A^T A x - 2y^T A x - y^T y$.
- What do the L_0, L_1, L_2 and L_∞ norms give for the vectors $(1,1,1)^T$, $(0,1,1)^T$, and $(0,1,0)^T$?

Matrices

Like vectors, matrices can be added together, multiplied by scalars etc, and hence **can themselves be considered elements of the vector space of matrices, and hence as 'vectors' themselves** in the algebraic sense. We will return to this idea later. **For now**, though, we will just consider matrices of dimensions $m \times n$ as representing **linear functions that map vectors of length n to vectors of length m**, i.e. as representing mappings between the vectors spaces $\mathbb{R}^n$ and $\mathbb{R}^m$.

Linear independence, image, kernel

A **linear combination** of vectors $v_1, v_2, ..., v_n$ has the form

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = 0$$

for scalars $\alpha_1, \dots, \alpha_n$. The vectors $v_1, v_2, ..., v_n$ are called **linearly independent** if the only solution to

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = 0$$

is the trivial solution $\alpha_1, \dots, \alpha_n = 0$.

Since **multiplying a vector by a matrix can be thought of as taking a linear combination of the columns of the matrix with weights given by the vector** (see later supplement for refresher), the above is equivalent to

$$Ax = 0 \implies x = 0$$

where the A matrix has n columns, each corresponding to one of the vectors $v_1, \dots, v_n$, and the vector x is given by $x = (\alpha_1, \dots, \alpha_n)$.

A **subspace** of a linear space is a subset of the original space that is also **itself a linear space**, i.e. also satifies the vector space axioms. In particular, it must be **closed under vector addition and scalar multiplication** and include the **zero vector**.

A **basis** for a linear space is a set of vectors that are **linearly independent** and such that every element of the space is a **linear combination** of the vectors in the basis. The **dimension** of a space V, written $\dim V$ or $\dim(V)$, is, for finite dimensions, the number of vectors in the basis.

The **image** (or column space/range) of an $m \times n$ matrix A, written $\mathrm{Im}A$ or $\mathrm{im}A$, is the set of all linear combinations of the the columns of A. That is,

$$\mathrm{Im}A = \{Ax \in \mathbb{R}^m \mid x \in \mathbb{R}^n\} \subseteq \mathbb{R}^m.$$

The **kernel** (or null space) of an $m \times n$ matrix A, written $\mathrm{Ker}A$ or $\ker A$ is the set of all vectors x such that $Ax = 0$, i.e.

$$\ker A = \{x \in \mathbb{R}^n \mid Ax = 0\} \subseteq \mathbb{R}^n.$$

The **rank-nullity theorem** states that, for an $m \times n$ matrix A, $\dim(\mathrm{Im}A) + \dim(\mathrm{Ker}A) =$ 'rank' + 'nullity' $= n =$ number of columns of A.

Tall and wide matrices

- We call an $m \times n$ matrix A **tall** if $m > n$, i.e. it has more rows than columns.
- We call an $m \times n$ matrix A **wide** if $m < n$, i.e. it has more columns than rows.

Invertibility of tall and wide matrices

- When the **columns** of a **tall** (or square) matrix are **linearly independent** then it has a **left** inverse (a *retraction*) $LA = I$. The converse is also true - the existence of a left inverse implies the columns of a tall (or square) matrix are linearly independent.
- When the **rows** of a **wide** (or square) matrix are **linearly independent** then it has a **right** inverse (a *section*). The converse is also true - the existence of a right inverse implies the rows of a wide (or square) matrix are linearly independent.
- It follows that a matrix has **both** a left inverse and a right inverse iff it is **square** with **linearly independent** rows and columns (iff it is **square** with **all eigenvalues non-zero** iff is **square** with a **non-zero determinant**). In this case the left and right inverses are the same matrix, and just called the **inverse**.

Test your understanding of tall/wide systems

- A typical least squares data approximation problem requires one to 'solve' $Ax = y$ when A is $m \times n$ with $m > n$ and (say) linearly independent columns. In what sense is this 'solvable'? What sort of inverse, if any, 'solves' this problem?
- A typical least squares model reduction problem requires one to 'solve' $Ax = y$ when A is $m \times n$ with $m < n$ and (say) linearly independent rows. In what sense is this 'solvable'? What sort of inverse, if any, 'solves' this problem?