3. Linear Algebra

3.1. Vectors as arrows.

Write vectors as arrows but the “real vector” is the location of the tip of the arrow. Important that in visual applications, we often draw vectors that don’t have their base at the origin. E.g., the gradient vector at $\mathbf{x}$ is always drawn with its base at the point $\mathbf{x}$. Strictly speaking, you have to translate it back to the origin to interpret it as a vector.
3.2. Vector operations

Row and column vectors: doesn’t make any difference whether the vector is written as a row or a column vector. Purely a matter of convenience.

The norm of a vector is its euclidean length: measure the arrow with a ruler. Written \( ||x|| = \sqrt{\sum_{k=1}^{n} x_k^2} \). Note that \( ||x|| = d_2(x, 0) \).

Adding and subtracting vectors. Intuitive what the sum of two vectors looks like. A little trickier to figure out what the difference between two vectors looks like, but you should try to figure out the picture.

How to visualize \( a - b \): do \( a + (-b) \).

Take the positive weighted sum of two vectors: \( \alpha v^1 + (1 - \alpha)v^2 \). Draw it.

Scalar multiples: do it.

The inner product of two vectors \( x, y \in \mathbb{R}^n \) is the sum of the products of the components. That is, \( x \cdot y = \sum_{k=1}^{n} x_k y_k \). When I think of inner products, I think of a row vector and a column vector; purely a convention.

It is hard to visualize what \( x \cdot y \) looks like. Look at a picture of \( x \) and \( y \) and say whether \( x \cdot y \) is positive, negative, zero. Answer is given by the angle between the two vectors.

- acute angle means \( x \cdot y \) is positive.
• obtuse angle means $\mathbf{x} \cdot \mathbf{y}$ is negative.

• ninety degree angle means $\mathbf{x} \cdot \mathbf{y}$ is zero.

**Theorem:** $\mathbf{a} \cdot \mathbf{u} = ||\mathbf{a}|| ||\mathbf{u}|| \cos(\theta)$, where $\theta$ is the angle between $\mathbf{a}$ and $\mathbf{u}$. (We’ll prove this later in the lecture.) Note the beauty of cos: doesn’t matter whether you look at the big angle between the vectors or the little one, get the same answer!

In Fig. 1, rank the inner products $\mathbf{x} \cdot \mathbf{a}$, $\mathbf{x} \cdot \mathbf{b}$ and $\mathbf{x} \cdot \mathbf{c}$. Answer: all the vectors are the same length, so that the only thing that determines the inner product is the angle between them. Hence $\mathbf{x} \cdot \mathbf{a} > \mathbf{x} \cdot \mathbf{b} > \mathbf{x} \cdot \mathbf{c}$.

Application: a fact that we’ll learn soon is that for small vectors $\mathbf{dx}$, $f(\mathbf{x} + \mathbf{dx}) \approx f(\mathbf{x}) + \nabla f(\mathbf{x}) \cdot \mathbf{dx}$. Just believe this for the moment.

• Draw $\mathbf{x}$ in the domain and $\mathbf{dx}$, then add them to get $\mathbf{x} + \mathbf{dx}$. Now think about $f(\mathbf{x} + \mathbf{dx})$: is it bigger or smaller than $f(\mathbf{x})$?
• First answer is graphical. Assume the domain is $\mathbb{R}^2$, draw $x$ and a level set through $x$.\footnote{We’ll look at level sets in more detail later. For now, the following definitions will suffice}

Now draw 3 small vectors $\mathbf{dx}$ starting from $x$.

- if $\mathbf{dx}$ points into the upper contour set, then $x + \mathbf{dx}$ is in the upper contour set. That is, by definition of the upper contour set, $f(x + \mathbf{dx}) > f(x)$. Which $\mathbf{dx}$’s point into the upper contour sets? The ones that make an acute angle with the gradient vector.

- if $\mathbf{dx}$ points along the level set, then $f(x + \mathbf{dx}) \approx f(x)$ i.e., $f$ is flat in this direction. Which $\mathbf{dx}$’s point into the level sets? The ones that make an right angle (90 degrees) with the gradient vector.

- if $\mathbf{dx}$ points into the lower contour set, then $x + \mathbf{dx}$ is in the lower contour set. That is, by definition of the lower contour set, $f(x + \mathbf{dx}) < f(x)$. Which $\mathbf{dx}$’s point into the lower contour sets? The ones that make an obtuse angle with the gradient vector.

• Now observe that you get the same answer when you use the fact that $f(x + \mathbf{dx}) \approx f(x) + \nabla f(x) \cdot \mathbf{dx}$. and apply the cos formula to $\nabla f(x) \cdot \mathbf{dx}$. Answer depends on the angle between $\nabla f(x)$ and $\mathbf{dx}$.

  - if angle is acute, then $f(x + \mathbf{dx}) > f(x)$.
  - if angle is 90°, then $f(x + \mathbf{dx}) = f(x)$. (Well, not exactly, but then we only said that $f(x + \mathbf{dx})$ was \textit{approximately} equal to $f(x) + \nabla f(x) \cdot \mathbf{dx}$.
  - if angle is obtuse, then $f(x + \mathbf{dx}) < f(x)$.

This verifies that the gradient of a function at $x$ points into the upper contour set of the function at $x$, and that the gradient is perpendicular to the level set.

\footnote{We’ll look at level sets in more detail later. For now, the following definitions will suffice}

1. **Level set:** A level set of a function $f$ consists of all of the points in the domain of $f$ at which the function takes a certain value. In other words, take any two points that belong to the same level set of a function $f$: this means that $f$ assigns the same value to both points.

2. **Upper contour set:** An upper contour set of a function $f$ consists of all of the points in the domain of $f$ at which the value of the function is at least a certain value. We talk about “the upper contour set of a function $f$ corresponding to $\alpha$”, referring to the set of points to which $f$ assigns the value at least $\alpha$.

3. **Lower contour set:** A lower contour set of a function $f$ consists of all of the points in the domain of $f$ at which the value of the function is no more than a certain value.
3.3. Projections

Given a vector $\mathbf{u} \in \mathbb{R}^n$ and another vector $\mathbf{a} \in \mathbb{R}^n$, we often want to project $\mathbf{u}$ “along” $\mathbf{a}$. That is we find two vectors $\mathbf{w}^1$ and $\mathbf{w}^2$ that are perpendicular to each other (i.e., the inner product $\mathbf{w}^1 \cdot \mathbf{w}^2$ is zero) such that $\mathbf{w}^1$ is a scalar multiple of $\mathbf{a}$. That is, we are looking for a scalar $\alpha$ such that

\begin{align*}
(1) & \quad \mathbf{w}^1 = \alpha \mathbf{a}, \\
(2) & \quad \mathbf{w}^2 = \mathbf{u} - \mathbf{w}^1 \\
(3) & \quad \mathbf{a} \cdot (\mathbf{u} - \alpha \mathbf{a}) = 0.
\end{align*}

Terminology:

- $\mathbf{w}^1$ is called the projection of $\mathbf{u}$ along $\mathbf{a}$ or the vector component of $\mathbf{u}$ along $\mathbf{a}$
- $\mathbf{w}^2$ is called vector component of $\mathbf{u}$ orthogonal to $\mathbf{a}$

Figure 2. The projection of $\mathbf{u}$ along $\mathbf{a}$
Notice that \( w^1 \) is the closest vector to \( u \) of all vectors that are colinear with \( a \). \( \quad (1) \)

That is, \( \|w^2\| = \min \{\|u - v\| : v \text{ is colinear with } a\} \). To see this, pick \( v \) such that \( v = \alpha a \) and note that by Pythagoras (see Fig. 3):

\[
\|u - v\|^2 = \|u - w^1\|^2 + \|w^1 - v\|^2
\]

squared length of hypotenuse squared length of opposite squared length of adjacent

\[
= \|w^2\|^2 + \|w^1 - v\|^2
\]

Therefore

\[
\|w^2\|^2 = \|u - v\|^2 - \|w^1 - v\|^2
\]
Figure 4. The cosine formula

which, since the latter term is positive except when \( v = w^1 \)

\[
\leq ||u - v||^2
\]

proving statement (1).

You’ll see a lot more of this when you do regression analysis: \( \alpha \) will be the regression coefficient of a single variable regression, when \( u \) is the vector of observations of the dependent variable and \( a \) is the vector of observations of the independent variable. But this is a digression for us.

3.4. Proof of the cosine formula theorem

We need to prove that \( a \cdot u = ||a|| ||u|| \cos(\theta) \). Recall from Simon-Blume (Figures 10.18 and 10.19) that \( \cos(\theta) = \begin{cases} 
\text{length(adjacent)/length(hypotenuse)} & \text{if } \theta \text{ is acute} \\
-\text{length(adjacent)/length(hypotenuse)} & \text{if } \theta \text{ is obtuse}
\end{cases} \) (see Fig. 4). Now let \( u \) be the hypotenuse. What’s the adjacent line? Recall that opposite and adjacent are perpendicular to each other. In other words, adjacent is the projection of \( u \) (the hypotenuse) onto \( a \), i.e., adjacent = \( \alpha a \), where \( \alpha \) is defined by conditions 1-3 above. From Fig. 4 \( \alpha < 0 \) iff \( \theta \) is an obtuse angle.
Therefore $|\alpha a| = \begin{cases} 
\alpha |a| & \text{if } \theta \text{ is acute} \\
-\alpha |a| & \text{if } \theta \text{ is obtuse}
\end{cases}$.

So we have

$$
\cos(\theta) = \begin{cases} 
\frac{|\text{length(adjacent)}|}{|\text{length(hypotenuse)}|} & \text{if } \theta \text{ is acute} \\
-\frac{|\text{length(adjacent)}|}{|\text{length(hypotenuse)}|} & \text{if } \theta \text{ is obtuse}
\end{cases}
= \begin{cases} 
\alpha \frac{|a|}{|u|} & \text{if } \theta \text{ is acute} \\
-\left(\alpha \frac{|a|}{|u|}\right) & \text{if } \theta \text{ is obtuse}
\end{cases}
$$

Now plug the expression for $\cos(\theta)$ into the right hand side of our expression to obtain

$$
|a||u||\cos(\theta)| = |a||u|\alpha \frac{|a|}{|u|} = \alpha |a||a| = |a|^2 \alpha
$$

But by definition of $\alpha$

$$
0 = w^1 \cdot w^2 \equiv \alpha a \cdot (u - \alpha a) = a \cdot (u - \alpha a) = a \cdot u - \alpha a \cdot a
$$

so that

$$
\alpha = \frac{a \cdot u}{a \cdot a} = \frac{a \cdot u}{|a|^2}
$$

Hence, if $|a| > 0$

$$
|a||u||\cos(\theta)| = |a|^2 \alpha \frac{a \cdot u}{|a|^2} = a \cdot u
$$

proving the cosine formula theorem.
3.5. Linear Combinations, Linear Independence, Linear Dependence and Cones.

**Defn:** \( x \in \mathbb{R}^n \) is a *linear combination* of a set of \( m \) vectors \( \{v^1, \ldots, v^k, \ldots, v^m\} \) in \( \mathbb{R}^n \) if there exists a vector \( t \in \mathbb{R}^m \) such that \( x = \sum_{k=1}^{m} t_k v^k \); in words, if \( x \) can be written as the sum of scalar multiples of the original vectors \( v^k \)'s.

**Example:** in \( \mathbb{R}^2 \), \( \{v^1, v^2\} \) point in the same (or opposite) direction, the linear combinations of these vectors all lie on the same line. If not, then any point in \( \mathbb{R}^2 \) can be written as a linear combination of these vectors.

**Defn:** \( x \in \mathbb{R}^n \) is a *nonnegative linear combination* of a set of \( m \) vectors \( \{v^1, \ldots, v^k, \ldots, v^m\} \) in \( \mathbb{R}^n \) if there exists a vector \( t \in \mathbb{R}^m_+ \) such that \( x = \sum_{k=1}^{n} t_k v^k \). I.e., the coefficients all have to be nonnegative.

**Defn:** The *nonnegative (positive) cone* defined by a set of vectors \( \{v^1, \ldots, v^k, \ldots, v^m\} \) is the set of all nonnegative (positive) linear combinations of these vectors.

Note that

1. if you have two vectors in \( \mathbb{R}^2 \) that aren’t colinear, the difference between the nonnegative cone and the positive cone is that the “edges” of the cone aren’t included in the positive cone but are included in the nonnegative cone.
2. if you have two vectors in \( \mathbb{R}^2 \) that are colinear, with a *negative* coefficient i.e., you have the vectors \( x \) and \( \alpha x \), with \( \alpha < 0 \), then the positive and the nonnegative cones are identical and consist of the entire line through these vectors.
Convex combinations: what’s the difference between one of these and a linear combination? Ans: the set of convex combinations of two vectors is the line segment between them, which is a subset of the nonnegative cone defined by these two vectors. The set of convex combinations of three vectors is a plane.

Defn: $x \in \mathbb{R}^n$ is a convex combination of a set of vectors $\{v^1, ... v^k, ... v^m\}$ if there exists a vector $t \in \mathbb{R}^m_+$ such that $\sum_{k=1}^m t_k = 1$ and $x = \sum_{k=1}^m t_k v^k$. I.e., the coefficients have to be nonnegative and sum to one.

Informal and not quite correct defn: a set of vectors $\{v^1, ... v^k, ... v^m\}$ is a linear independent set if no one of them can be written as a linear combination of all the others.

Correct defn: a set of vectors $\{v^1, ... v^k, ... v^m\}$ is a linear independent set if for all $t \in \mathbb{R}^m$, $\sum_{k=1}^m t_k v^k = 0$ implies $t = 0$.

Defn: a set of vectors $\{v^1, ..., v^m\}$ is a linear dependent set if it is not a linear independent set.

The only difference between my friendly defn and the unfriendly formal definition is that the set $\{0\}$ is linear independent by my definition and linear dependent by the formal definition. To see that it is linear independent according to my definition, note that if 0 is the only element of the set, then, trivially, you can’t write 0 as a lin comb of other vectors in the set, because there aren’t any. To see that it is linear dependent by the formal definition, let $t = 1$, and note that $t \times 0 = 0$, but $t \neq 0$. So the test for linear independence fails. It’s a useful exercise to check that the two definitions are equivalent for any set of vectors other than the singleton set zero.
Examples:

- Can you construct a linear independent set of vectors in which one of the vectors is zero?
- What’s the largest set of linearly independent 2-vectors you can have?
- What’s the largest set of linearly independent 3-vectors you can have?