Linear Algebra and Curve Fitting
Math 135: Applied Calculus

Andrew Beveridge, Dan Flath, Tom Halverson,
Danny Kaplan, Karen Saxe

February 2, 2010

Contents

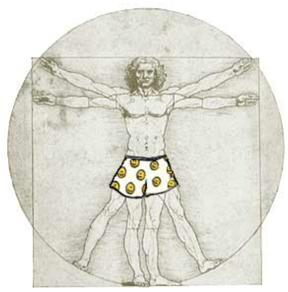
1 Introduction

1.1 Motivation

In applied mathematics many of the systems of equations that you would like to solve do not have any solutions. The reason is that there are very often far more equations than variables. This is especially so when doing statistics, because the goal is to explain a large quantity of data with a small number of parameters.

When there are more equations than variables you can not expect to find a solution, though this situation does not automatically rule out the existence of a solution. We can get an idea of the problem by considering motivating example.

Example 1.1. *Isometry of the human arm.* We often marvel at the patterns and symmetry of nature. For example, Leonardo da Vinci's famous drawing of the *Vitruvian Man* correlates proportions of the human body with ideal geometries. We include a more, er, modest version of the drawing here:



One of the most famous numbers in mathematics is the *golden ratio* $\phi = \frac{1+\sqrt{5}}{2} \approx 1.618$. Leonardo suggested that various body parts obey a proportional relationship governed by the golden ratio. One example is the combined length of an adult's forearm and hand compared to the length of his hand. Leonardo hypothesized that

$$\frac{\text{length of forearm and hand}}{\text{length of hand}} = \frac{1+\sqrt{5}}{2}.$$

Let y denote the combined length of the forearm and hand, and let x denote the length of the hand. We can rewrite this hypothesis as

$$y = \frac{1+\sqrt{5}}{2} x \approx 1.61\, x.$$

Such a proportional relationship is called an *isometry.* Let us compare this hypothesis to some real data, contained in the following table.

forearm (cm)	forearm + hand (cm)	proportion
26.0	44.5	1.712
24.4	41.4	1.697
26.4	46.1	1.746
26.0	43.6	1.658
27.2	46.8	1.721

First off, we can see that Leonardo was not quite correct for this small data set (and, in fact, for larger data sets as well). All of our calculated proportions are above his hypothesized proportion of 1.618. So what is the correct constant of proportionality? Ideally, we would like to find a constant a such that $y = ax$ for all of our data points. In other words, we want to solve the system of equations

$$\begin{aligned} 44.5 &= 26.0\,a \\ 41.4 &= 24.4\,a \\ 46.1 &= 26.4\,a \\ 43.6 &= 26.0\,a \\ 46.8 &= 27.2\,a. \end{aligned}$$

This is a *system of 5 equations in 1 unknown variable.* We have far more constraining equations than we have variables! We can't even find a value for a that works for *two* of these equations, let alone all six. So we cannot find an *exact* solution to this system of equations. But we can find an approximate solution. A naive way to find an approximate solution is to take the average of these six values, which would give $a = 1.707$. But is this the best approximation for a that we can find?

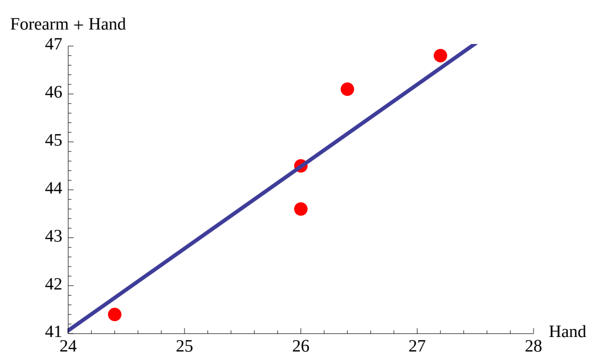
To answer this question, we need to explicitly define what we mean by "best approximation." Intuitively, we want our function $y = ax$ to be close to all of the points simultaneously. We define the *line of best fit* as the line that minimizes the sum of the squares of the vertical distances between the data points and the line. In other words, we find the value of a that minimizes

$$(44.5 - 26a)^2 + (41.4 - 24.4a)^2 + (46.1 - 26.4a)^2 + (43.6 - 26a)^2 + (46.8 - 27.2a)^2.$$

Here we square before adding them because these differences may be positive or negative. Squaring ensures that all the terms are positive. Using techniques that we will learn in Section 5, we find that the value that minimizes this equation is $a = 1.711$. The line of best fit

$$y = 1.711\,x$$

is also called the *least squares solution.* The "squares" that we are minimizing are the sum of the squares of the differences between our approximation and the actual data. Here is a graphical representation of the data and our least squares approximation.



1.2 Goals of this Reader

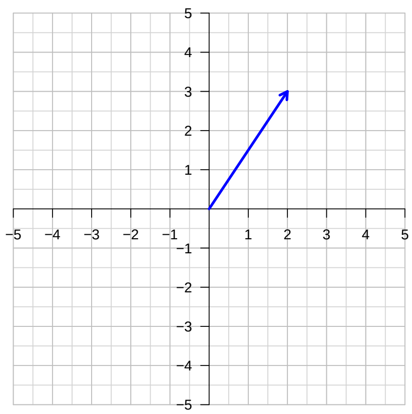
The goal of this reader is to introduce the basics of *linear algebra* and use them to explain the method of *least squares approximation.*

Simply put, linear algebra is a mathematical framework for describing a multidimensional world. Linear algebra has applications in nearly every scientific discipline. In particular, much of statistics is based on linear algebra, and we will cover enough of the material for you to use in Math 155, Introduction to Statistical Modeling. We introduce the fundamental ideas of linear algebra: vectors, linear combinations and vector spaces. We then discuss the dot product, which is a simple function that captures the geometry of vector spaces.

The latter part of this reader is devoted to curve fitting via least squares approximation. The key is to transform a curve fitting problem into a (relatively simple) linear algebra problem. We use the inherent geometry of vector spaces to identify the "best" function to model a set of data. Finally, we should how to transform our data so that we can use the method of least squares to fit our data to a variety of models (polynomials, logarithmic, exponential, etc).

2 Vectors

The pair of coordinates $(2,3)$ indicates the point in the xy-plane where $x = 2$ and $y = 3$. This point can also be thought of as the vector, which is an arrow in the xy-plane with tail at the origin $(0,0)$ and head at the point $(2,3)$. We often write the coordinates of a vector vertically, $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$, to indicate that we are thinking of the vector rather than the point $(2,3)$, but we will not always do so. In the figure below, you find a drawing of this vector.

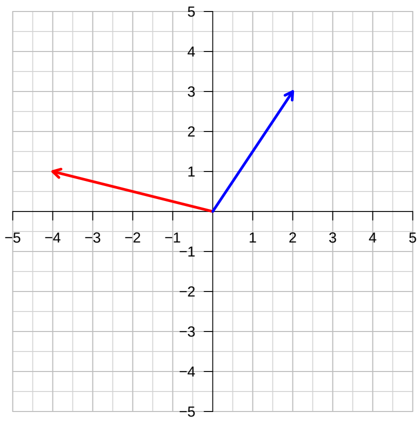


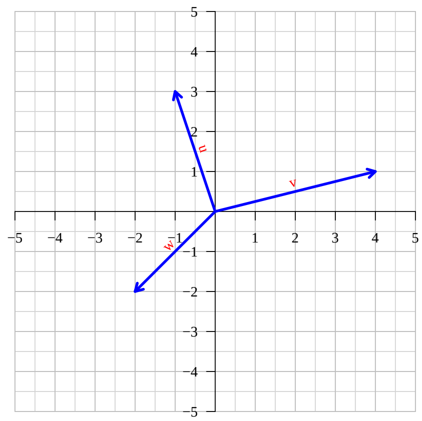
The zero vector $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is just a point since its head and tail are both at the origin.

Example 2.1.

1. The vectors $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ and $\begin{pmatrix} -4 \\ 1 \end{pmatrix}$ are drawn in the plot on the left below. Include the vectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$, $\begin{pmatrix} -2 \\ -3 \end{pmatrix}$ in the plot.

2. In the right plot below, there are three vectors u, v and w drawn. Give their coordinate names.



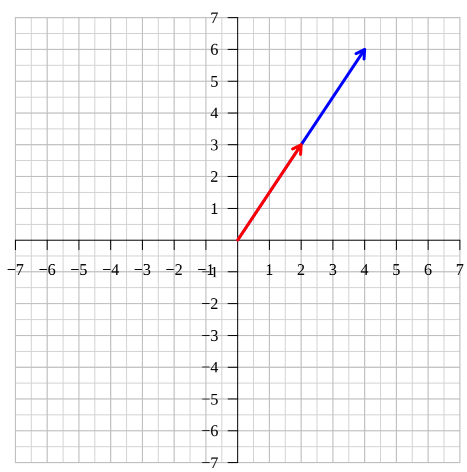


2.1 Scalar Multiplication

With vectors, we do two operations that we do not do with points in the plane. The first is called *scalar multipliation*. We multiply a vector by a constant to get another vector whose length is rescaled by that constant. This works both algebraically and geometrically. For example, using algebra we have

$$2\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix}.$$

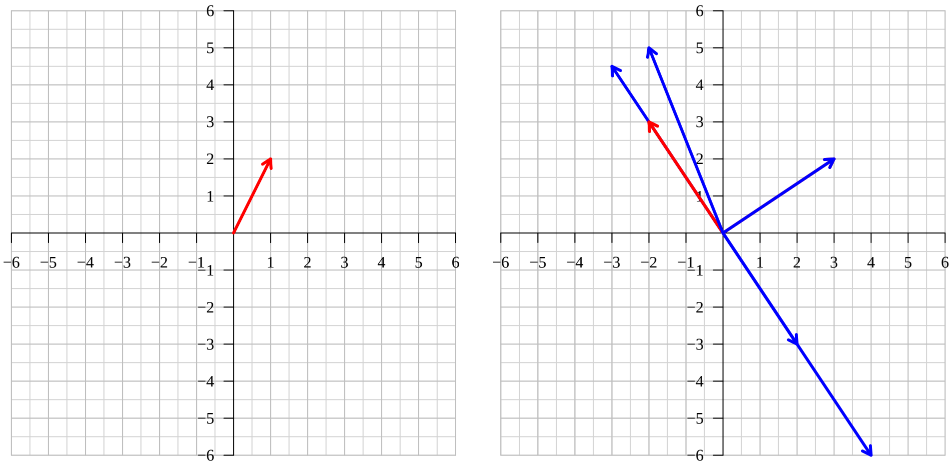
We have multiplied each coordinate of the vector by the constant 2. From the Figure below you can see that multiplying the vector by 2 doubles its length but does not change its direction.



Example 2.2.

1. The figure below left shows the vector $\vec{\mathbf{u}} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Give coordinates (by computing them algebraically) and sketch a picture of $2\vec{\mathbf{u}}, 3\vec{\mathbf{u}}, 1.5\vec{\mathbf{u}}, (-1)\vec{\mathbf{u}}$. Notice that that multiplying a vector by a negative constant causes it to point in the opposite direction.

2. The figure below right shows the vector $\vec{\mathbf{w}} = \begin{pmatrix} -2 \\ 3 \end{pmatrix}$ and three other vectors. Decide which are constant multiples of $\vec{\mathbf{w}}$ and for those that are find the multiple.

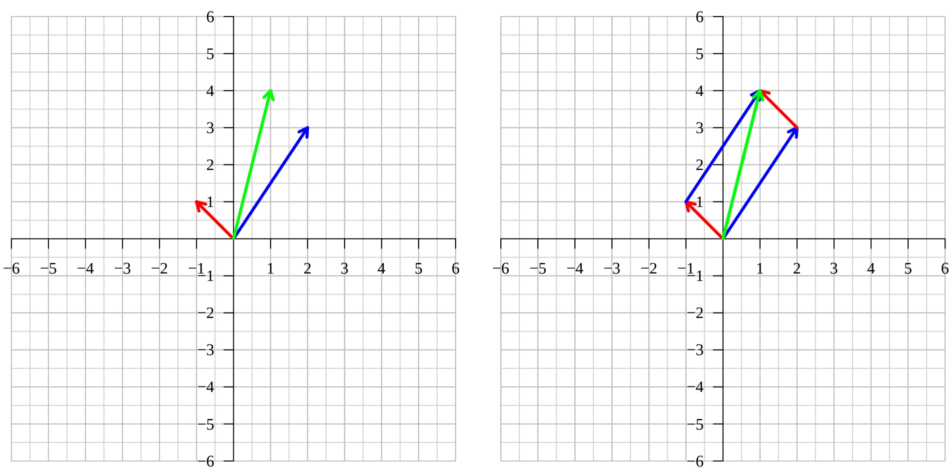


2.2 Vector Addition

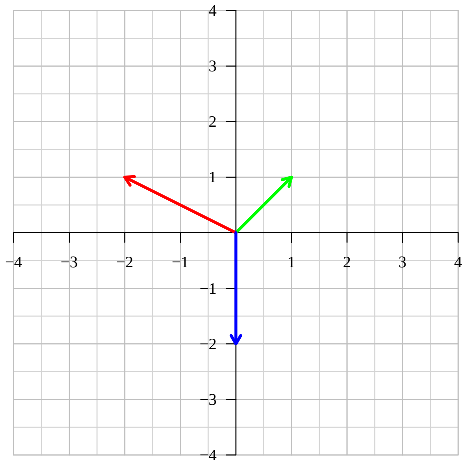
We also can add two vectors to get a vector. Algebraically, we do it just like you would think

$$\begin{pmatrix} 2 \\ 3 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \end{pmatrix},$$

This is called coordinate-wise addition, since you add corresponding coordinates. The Figure below shows the sum as the diagonal of the parallelogram with the original two vectors on two sides. Also interpreted geometrically as head to tail concatenation of arrows.



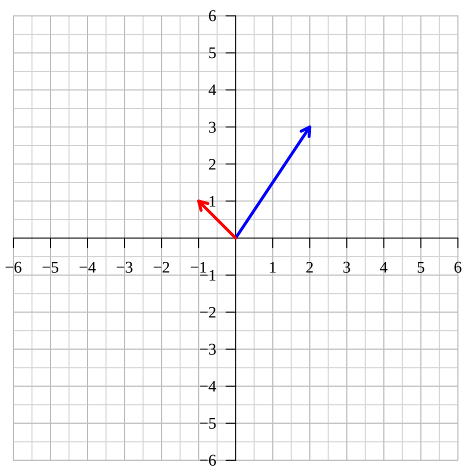
Example 2.3. The figure below shows three vectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 0 \\ -2 \end{pmatrix}$. Give coordinates and sketch the sum (showing the parallelogram in each case) for each pair of these vectors.



2.3 Linear Combinations

We can combine the two operations, adding multiples of two vectors. A sum of multiples is a *linear combination.*

Example 2.4. The graph below shows the vectors $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Compute and show the three linear combinations $\begin{pmatrix} 2 \\ 3 \end{pmatrix} + 3\begin{pmatrix} -1 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 2 \\ 3 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, and $2\begin{pmatrix} 2 \\ 3 \end{pmatrix} - \begin{pmatrix} -1 \\ 1 \end{pmatrix}$.



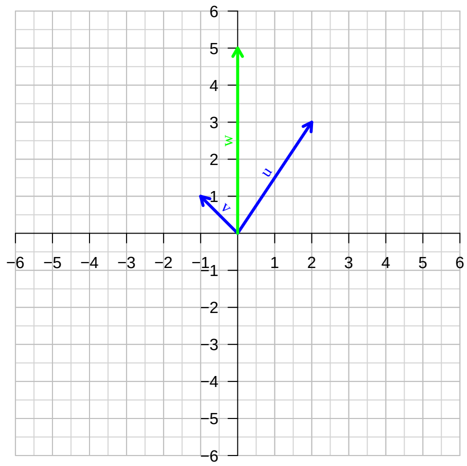
3 Linear Equations: two interpretations

Consider the following three vectors

$$\vec{\mathbf{u}} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \qquad \vec{\mathbf{v}} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \qquad \vec{\mathbf{w}} = \begin{pmatrix} 0 \\ 5 \end{pmatrix}.$$

In this section, we explore the following kind of question: What multiples of $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ can you add to get to $\vec{\mathbf{w}}$. In fancier language, we are asking, which linear combinations of $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ equal $\vec{\mathbf{w}}$?

First we answer the question graphically in the plot below. Is it possible to appropriately re-scale $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ so that they sum to $\vec{\mathbf{w}}$? Do it in the plot and keep track of what the scaling factors x and y are.



Now we solve the problem algebraically. Since the multiples are unknown, we denote them by x and y. We are asking what numbers x and y are such that

$$x\begin{pmatrix} 2 \\ 3 \end{pmatrix} + y\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \end{pmatrix}.$$

This is an equation in two variables x and y. Is $x = 3$, $y = 2$ a solution? Check by substituting the proposed solution values for x and y into the equation: $3\begin{pmatrix} 2 \\ 3 \end{pmatrix}+2\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 11 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 5 \end{pmatrix}$. No, it is not a solution.

To find a solution, let us proceed to transform the equation a little bit.

$$x\begin{pmatrix} 2 \\ 3 \end{pmatrix} + y\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \end{pmatrix}$$

is equivalent to

$$\begin{pmatrix} 2x \\ 3x \end{pmatrix} + \begin{pmatrix} -y \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \end{pmatrix}$$

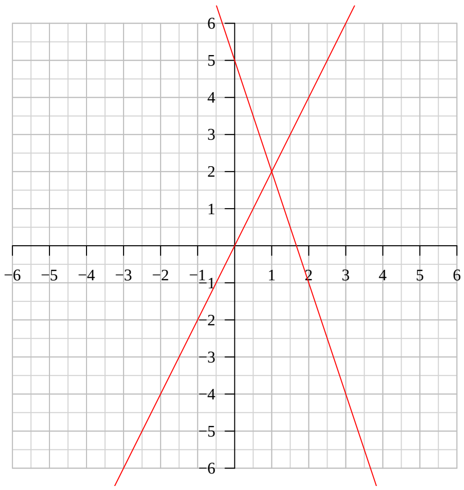
and to

$$\begin{pmatrix} 2x - y \\ 3x + y \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \end{pmatrix}.$$

Thus, a solution to our problem is a pair (x, y) which satisfies the simultaneous system of two equations and two unknowns

$$\begin{array}{ccccc} 2x & - & y & = & 0 \\ 3x & + & y & = & 5 \end{array}.$$

Each of these equations represents a line in the plane, as seen in the plot below, and the solution to the system of equations is the point at the intersection of these two lines.



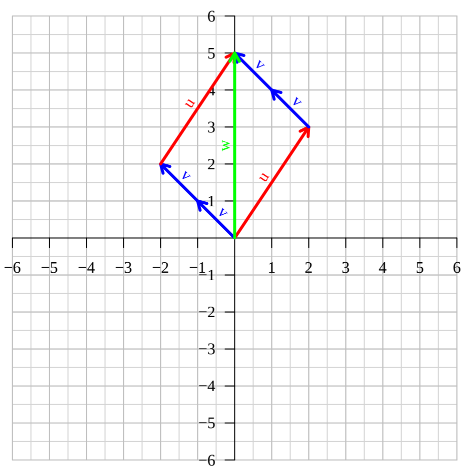
Solving such systems of linear equations is familiar from high school, and it is (fairly) easily solved to get

$$x = 1 \text{ and } y = 2.$$

This confirms what the figure above shows: that

$$1\begin{pmatrix} 2 \\ 3 \end{pmatrix} + 2\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \end{pmatrix}.$$

In terms of vectors, the solution $\vec{\mathbf{u}} + 2\vec{\mathbf{v}} = \vec{\mathbf{w}}$ is drawn in the plot below



Example 3.1. *Question*: Write the statement that the sum of x times $\begin{pmatrix} 1 \\ 3 \end{pmatrix}$ and y times $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ gives $\begin{pmatrix} 4 \\ 8 \end{pmatrix}$ in two ways: (a) as an equation in vectors; (b) as a system of linear equations.

Solution:

(a) $x\begin{pmatrix} 1 \\ 3 \end{pmatrix} + y\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 8 \end{pmatrix}$.

(b) $\begin{array}{rcrcr} x & + & y & = & 4 \\ 3x & + & y & = & 8 \end{array}$

Example 3.2. *Question*: Translate the system of equations into the language of vectors.

$$\begin{aligned} x - 3y &= 8 \\ 2x + y &= 3. \end{aligned}$$

Solution:

$$\begin{pmatrix} x - 3y \\ 2x + y \end{pmatrix} = \begin{pmatrix} 8 \\ 3 \end{pmatrix}.$$

Then

$$\begin{pmatrix} x \\ 2x \end{pmatrix} + \begin{pmatrix} -3y \\ y \end{pmatrix} = \begin{pmatrix} 8 \\ 3 \end{pmatrix}.$$

Then

$$x \begin{pmatrix} 1 \\ 2 \end{pmatrix} + y \begin{pmatrix} -3 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 3 \end{pmatrix}.$$

The system of equations can be interpreted to mean: What linear combination of the vectors $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} -3 \\ 1 \end{pmatrix}$ equals $\begin{pmatrix} 8 \\ 3 \end{pmatrix}$?

3.1 The geometry of linear equations

We have seen that a system of equations for unknowns x and y

$$\begin{aligned} ax + by &= e \\ cx + dy &= f \end{aligned}$$

can also be expressed as

$$x \begin{pmatrix} a \\ c \end{pmatrix} + y \begin{pmatrix} b \\ d \end{pmatrix} = \begin{pmatrix} e \\ f \end{pmatrix}.$$

In other words, this problem has two fundamental but different geometric interpretations:

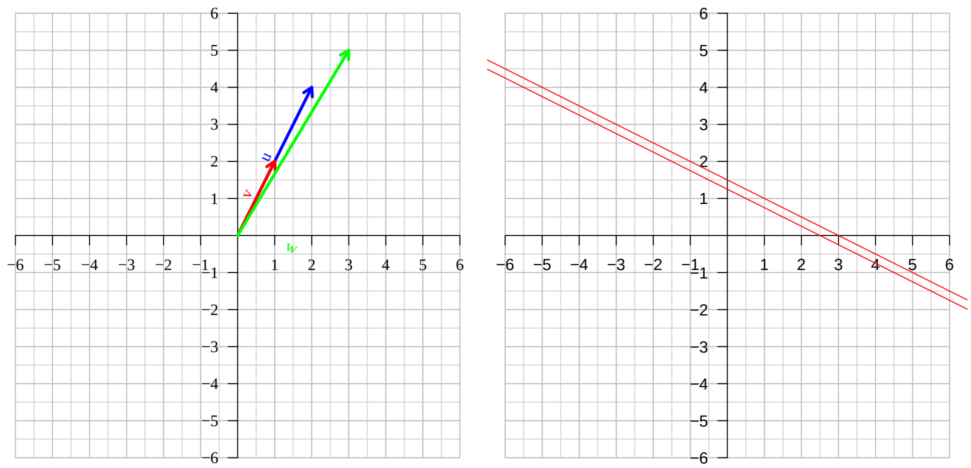
1. (Points interpretation) What point (x, y) is on the intersection of two lines $ax + by = e$ and $cx + dy = f$?

2. (Vector interpretation) What multiples x of $\vec{\mathbf{u}} = \begin{pmatrix} a \\ c \end{pmatrix}$ and y of $\vec{\mathbf{v}} = \begin{pmatrix} b \\ d \end{pmatrix}$ add to give $\vec{\mathbf{w}} = \begin{pmatrix} e \\ f \end{pmatrix}$?

Example 3.3. What multiples of $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ can you add to get $\begin{pmatrix} 3 \\ 5 \end{pmatrix}$? The corresponding system of equations is

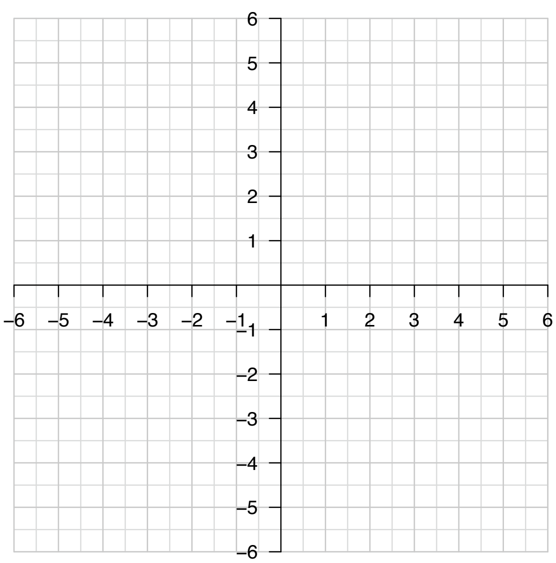
$$\begin{aligned} x + 2y &= 3 \\ 2x + 4y &= 5 \end{aligned}$$

The points and vector interpretations are shown below.



There is a problem here. The trouble is that the vectors $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ are in line and no linear combination gets you off the line where $\begin{pmatrix} 3 \\ 5 \end{pmatrix}$ is. Similarly, the two lines are parallel and do not intersect.

Example 3.4. *Question*: What multiples of $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ can you add to get $\begin{pmatrix} 4 \\ 8 \end{pmatrix}$? Draw a vector graph of the situation on the coordinate axes below:



Solution: The trouble is that the vectors $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 4 \end{pmatrix}$ are both in line with (pointing in the same direction as) the vectors $\begin{pmatrix} 4 \\ 8 \end{pmatrix}$, so there are many ways to get to $\begin{pmatrix} 4 \\ 8 \end{pmatrix}$. For example

$$\begin{aligned}
\begin{pmatrix} 4 \\ 8 \end{pmatrix} &= 4\begin{pmatrix} 1 \\ 2 \end{pmatrix} + 0\begin{pmatrix} 2 \\ 4 \end{pmatrix} \\
&= 0\begin{pmatrix} 1 \\ 2 \end{pmatrix} + 2\begin{pmatrix} 2 \\ 4 \end{pmatrix} \\
&= 2\begin{pmatrix} 1 \\ 2 \end{pmatrix} + 1\begin{pmatrix} 2 \\ 4 \end{pmatrix} \\
&= 6\begin{pmatrix} 1 \\ 2 \end{pmatrix} + (-1)\begin{pmatrix} 2 \\ 4 \end{pmatrix}
\end{aligned}$$

Notice that in the last case, we overshoot with the first vector and then compensate with the last.

As you can see from the intersection of lines in the examples above, a system of equations can have

(0) **No solution**: Here the two lines are parallel (and not the same), and the two vectors are parallel and the vector we are trying to get to is not on that line.

(1) **Exactly One Solution**: Here the two lines intersect in a point, and the two vectors can be combined in exactly one way to get to that vector.

(∞) **Infinitely Many Solutions**: Here the two lines are the same and any point on this line works. The corresponding vectors are parallel and so is the vector we are trying to get to.

And these are the only possible solutions since these are the only ways that lines can intersect.

3.2 Higher Dimensions

Everything we have done so far has been in the 2-dimensional xy-plane, which we denote by $\mathbb{R}^2$ (to stand for the 2-dimensional plane of real numbers). But this all works in $3, 4, 5$ or more dimensions. For example, the vectors

$$\vec{\mathbf{u}} = \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, \quad \vec{\mathbf{v}} = \begin{pmatrix} 5 \\ -2 \\ 0 \end{pmatrix}, \quad \vec{\mathbf{w}} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

live in 3-dimensional space, denoted $\mathbb{R}^3$. It is possible to draw these on 2-dimensional paper using perspective (thought we have not done it here). We will have to try to “graph” these in class. You can write down linear equations in the same way. For example, find scaling factors x and y such that the following vector equation is satisfied.

$$x\begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} + y\begin{pmatrix} 5 \\ -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

We can then convert this two a points equation of the form

$$\begin{array}{rcrcl} x & + & 5y & = & 1 \\ 2x & + & -2y & = & 1 \\ 4x & + & 0y & = & 1 \end{array}$$

Similarly we could have equations of the form:

$$x\begin{pmatrix}1\\2\\4\end{pmatrix} + y\begin{pmatrix}5\\-2\\0\end{pmatrix} + z\begin{pmatrix}1\\2\\3\end{pmatrix} = \begin{pmatrix}1\\1\\1\end{pmatrix} \qquad \Leftrightarrow \qquad \begin{array}{rcrcrcl} x & + & 5y & + & z & = & 1 \\ 2x & + & -2y & + & 2z & = & 1 \\ 4x & + & 0y & + & 3z & = & 1 \end{array}$$

Example 3.5. *Isometry of the human arm, continued.* Now that we have developed some linear algebra, we return to Leonardo's isometry of the human arm. We can rewrite the system of equations in Example 1.1 using vector notation. We get the vector equation:

$$\begin{pmatrix} 44.5 \\ 41.4 \\ 46.1 \\ 43.6 \\ 46.8 \end{pmatrix} = a \begin{pmatrix} 26.0 \\ 24.4 \\ 26.4 \\ 26.0 \\ 27.2 \end{pmatrix}.$$

Each of these vectors live in a 5-dimensional vector space. We may not be able to visualize them directly, but our intuition from 2-dimensions and 3-dimensions carries over without a hitch.

We already know that this vector equation has no solutions. Instead, we would like to find an approximate solution. In other words, we must find the scalar multiple of the right hand vector that is closest to the left hand vector. The coefficient a of that nearest vector will be the coefficient in our least squared solution $y = ax$. All that we have done is to rephrase our problem using the language of linear algebra: we think about linear combinations of vectors instead of systems of linear equations.

We will return to this problem once more in Section 5, where we will actually see how to find the best approximate value for a.

3.3 Matrix notation: $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$

Linear combinations of vectors occur so frequently that a special notation has been developed using matrices.

Definition. An $m \times n$ *matrix* is a rectangular array of numbers, arranged in m rows and n columns.

For example, here is a 2×2 matrix, a 3×2 matrix, and a 3×4 matrix respectively,

$$\begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 5 \\ 2 & -2 \\ 4 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 5 & 1 & 2 \\ 2 & -2 & 0 & 3 \\ 4 & 0 & 1 & 1 \end{pmatrix}.$$

We think of matrices as lists of vectors, one for each column. Thus

$$\begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix} \text{ represents the list } \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \text{ and}$$

$$\begin{pmatrix} 1 & 5 \\ 2 & -2 \\ 4 & 0 \end{pmatrix} \text{ represents the list } \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 5 \\ -2 \\ 0 \end{pmatrix}.$$

$$\begin{pmatrix} 1 & 5 & 1 & 2 \\ 2 & -2 & 0 & 3 \\ 4 & 0 & 1 & 1 \end{pmatrix} \text{ represents the list } \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix}, \begin{pmatrix} 5 \\ -2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}.$$

A linear combination of the n columns of an $m \times n$ matrix A is indicated by multiplication of A by a vector of n coefficients. For example

$$\begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ -5 \end{pmatrix} = 2 \begin{pmatrix} 2 \\ 3 \end{pmatrix} - 5 \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 9 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 5 \\ 2 & -2 \\ 4 & 0 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} + 2 \begin{pmatrix} 5 \\ -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 13 \\ 2 \\ 12 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 5 & 1 & 2 \\ 2 & -2 & 0 & 3 \\ 4 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ -1 \\ 2 \\ 4 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} - \begin{pmatrix} 5 \\ -2 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + 4 \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 20 \\ 18 \end{pmatrix}$$

Now, consider our original system of equations which we wrote both in vector notation and as a system of equations:

$$x \begin{pmatrix} 2 \\ 3 \end{pmatrix} + y \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \end{pmatrix} \qquad \Leftrightarrow \qquad \begin{array}{rcrcl} 2x & - & y & = & 0 \\ 3x & + & y & = & 5 \end{array}$$

Since

$$\begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x \begin{pmatrix} 2 \\ 3 \end{pmatrix} + y \begin{pmatrix} -1 \\ 1 \end{pmatrix},$$

we can write this same system of equations in matrix notation as

$$\begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \end{pmatrix}.$$

We found that a solution to this system is $x = 1$ and $y = 2$. This is represented by the matrix-vector product

$$\begin{pmatrix} 2 & -1 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix} + 2 \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 5 \end{pmatrix}.$$

Similarly,

$$\begin{pmatrix} 1 & 5 \\ 2 & -2 \\ 4 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 13 \\ 2 \\ 12 \end{pmatrix}$$

indicates the system

$$x_1 \begin{pmatrix} 1 \\ 2 \\ 4 \end{pmatrix} + x_2 \begin{pmatrix} 5 \\ -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 12 \\ 2 \\ 12 \end{pmatrix} \qquad \Leftrightarrow \qquad \begin{array}{rcrcr} x_1 & + & 5x_2 & = & 13 \\ 2x_1 & + & -2x_2 & = & 2 \\ 4x_1 & + & 0x_2 & = & 12 \end{array}$$

Letting

$$A = \begin{pmatrix} 1 & 5 \\ 2 & -2 \\ 4 & 0 \end{pmatrix}, \quad \vec{\mathbf{x}} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \text{and} \quad \vec{\mathbf{b}} = \begin{pmatrix} 13 \\ 2 \\ 12 \end{pmatrix}.$$

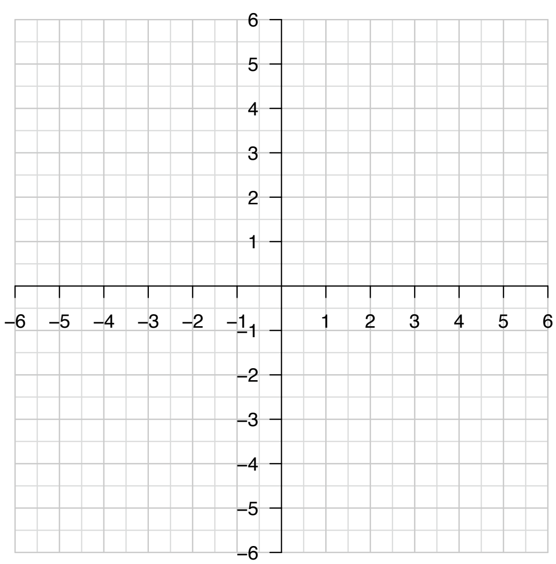
the same system can we written in shorthand as

$$A\vec{\mathbf{x}} = \vec{\mathbf{b}}.$$

3.4 Exercises

Here is some graph paper on which to do problems.

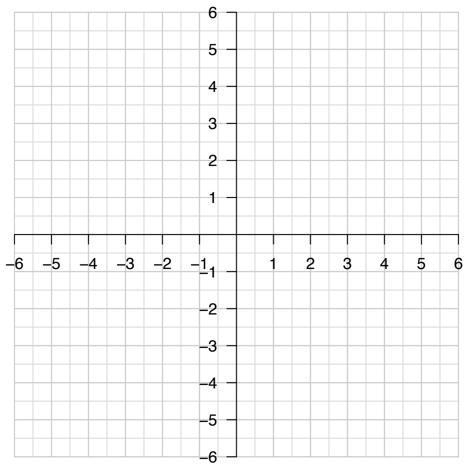
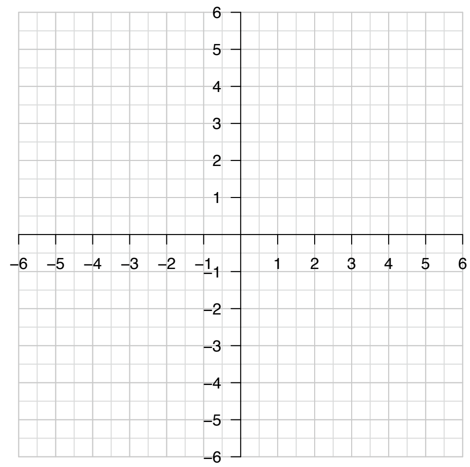
1. Draw the two vectors $\mathbf{u} = (4,1)$ and $\mathbf{v} = (2,3)$. Add to the graph the 3 vectors $\mathbf{u}+\mathbf{v}$, $-\mathbf{u}$, $2\mathbf{u} - 0.5\mathbf{v}$. You are to do these drawings without first finding the vectors numerically!



2. This problem emphasizes the connection between the vector and algebraic interpretations of systems of equations.

(a) Make up a system of two linear equations in two unknowns that has no solution. On the graph paper plot the two lines that your equations represent, illustrating that they do not intersect.

(b) Write the vector interpretation of your system. Draw, on graph paper, the two coefficient vectors, and use these plotted vectors to explain why the 'target' vector is not a linear combination of those two vectors.

3. Translate the system of equations into the language of vectors.

$$\begin{aligned} x - 3y &= 8 \\ 2x + y &= 3. \end{aligned}$$

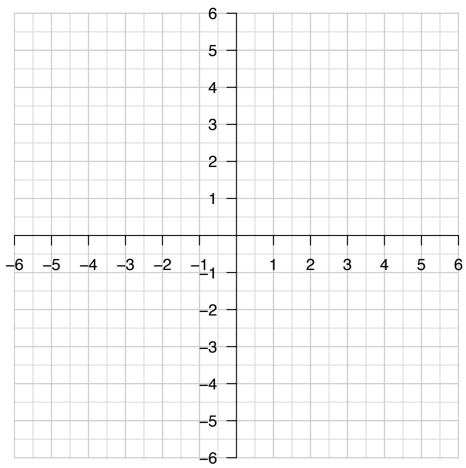
4. Consider two vectors, $\mathbf{p} = (1, 2)$ and $\mathbf{q} = (3, -2)$.

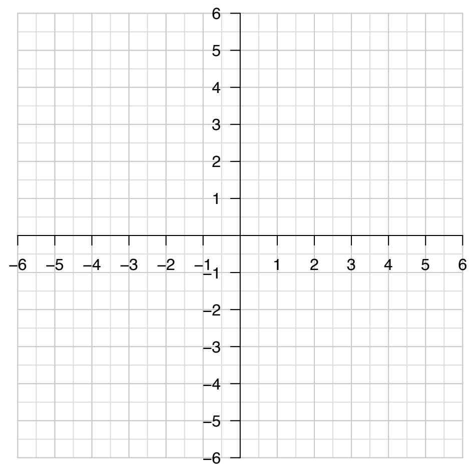
(a) For several values of the scalars α and β (that you pick), plot out the position of $\alpha\mathbf{p} + \beta\mathbf{q}$. Explain in words what points in the space can be reached by appropriate choice of α and β.

(b) Repeat the above, but holding $\alpha = 1$ and setting β however you want. What points in the space can be reached this way?

(c) Now hold $\beta = 1$ and setting α however you want. What points in the space can be reached this way?

(d) Now pick β as you want, but set $\alpha = 1 - \beta$, so that $\alpha + \beta$ always equals 1. What points in the space can be reached this way?





5. On graph paper, show how $\mathbf{w} = (1, -8)$ is a linear combination of $\mathbf{u} = (4, 1)$ and $\mathbf{v} = (-3, 2)$.

6. Given that
$$A = \begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}, \quad \vec{\mathbf{x}} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad \text{and} \quad \vec{\mathbf{b}} = \begin{pmatrix} 0 \\ -4 \end{pmatrix}$$
Write the system of equations $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$ in vector form and as a system of equations.

7. If the equation below is written in $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$ form, what are A, $\vec{\mathbf{x}}$, and $\vec{\mathbf{b}}$?
$$\begin{aligned} 2x - 5y &= 2 \\ x + y &= 3. \end{aligned}$$

8. If the system of equations below is written in $A\vec{\mathbf{x}} = \vec{\mathbf{b}}$ form, what are A, $\vec{\mathbf{x}}$, and $\vec{\mathbf{b}}$?
$$x_1 \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} + x_2 \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

9. Compute the following products:

   (a) $\begin{pmatrix} 1 & 3 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} -1 \\ 5 \end{pmatrix} =$

   (b) $\begin{pmatrix} 1 & 3 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} =$

   (c) $\begin{pmatrix} 1 & 3 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} =$

   (d) $\begin{pmatrix} 1 & 3 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \end{pmatrix} =$

   (e) $\begin{pmatrix} 1 & 3 & 4 & 5 \\ -2 & 1 & -1 & 5 \\ 1 & 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 0 \\ -1 \end{pmatrix} =$

(f) $\begin{pmatrix} 1 & 3 & 4 & 5 \\ -2 & 1 & -1 & 5 \\ 1 & 1 & -1 & 2 \\ 0 & 2 & -1 & 2 \\ 1 & 3 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} =$

(g) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} =$

10. What do you think of these matrix-vector products?

$$\begin{pmatrix} 2 & 2 & 0 & 2 \\ 1 & 1 & -1 & 2 \\ 1 & 1 & -1 & 0 \\ 0 & 1 & -1 & 3 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \qquad \begin{pmatrix} 2 & 2 & 0 \\ 1 & 1 & -1 \\ 1 & 1 & -1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

4 Vector Spaces

In this section, we further develop our understanding of vector spaces. First, we introduce the *dot product*. This remarkable function captures the geometry of a vector space. We show how to use the dot product to find the length of a vector and to determine the angle between two vectors.

Next, we cover three big ideas involving vectors: the *span* of a set of vectors, whether a set of vectors is *linearly dependent* or *linearly independent*, and the *dimension* of the span of a set of vectors.

4.1 The dot product and vector space geometry

The *dot product* $\vec{\mathbf{u}} \cdot \vec{\mathbf{v}}$ of two vectors $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ is the real number obtained by multiplying corresponding coordinates of the vectors and adding. For example,

$$\begin{pmatrix} 2 \\ -1 \\ 3 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 5 \\ 2 \\ 3 \\ 0 \end{pmatrix} = 2 \cdot 5 + (-1) \cdot 2 + 3 \cdot 3 + 1 \cdot 0 = 17.$$

One nice thing about a dot produce is that it is *easy* to compute. Be sure, however, to do the addition of the products. Many dot product beginners will give the answer to $\vec{\mathbf{v}} \cdot \vec{\mathbf{w}}$ to be the vector that consists of the numbers $(10, -2, 9, 0)$. In other words, they will forget to add them up and get the scalar 17. Sometimes the dot product is called a scalar product.

The amazing thing about the dot product is that this simple algebraic function actually captures the geometry of the vector space! In other words, we can use the dot product to find lengths of vectors and to find angles between vectors. This first property comes from the Pythagorean Theorem (which holds in all dimensions). The second property comes from the Law of Cosines (which you probably learned in high school trigonometry).

Definition. (1) The *length* of a vector $\mathbf{u} \in \mathbb{R}^n$ is defined to be

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}.$$

(2) The *angle* θ between two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ can be computed using the formula

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \, \|\mathbf{v}\| \cos\theta.$$

Example 4.1. *Question*: Find the lengths of the two vectors $\mathbf{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} a \\ b \end{pmatrix}$ and the angle between them.

Solution: The lengths are

$$\|\mathbf{u}\| = \sqrt{1^2 + 0^2} = 1 \qquad \|\mathbf{v}\| = \sqrt{a^2 + b^2}.$$

The angle θ between the two vectors is computed from its cosine:

$$\cos\theta = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} = \frac{a}{1 \cdot \sqrt{a^2 + b^2}} = \frac{a}{\sqrt{a^2 + b^2}}.$$

Thus

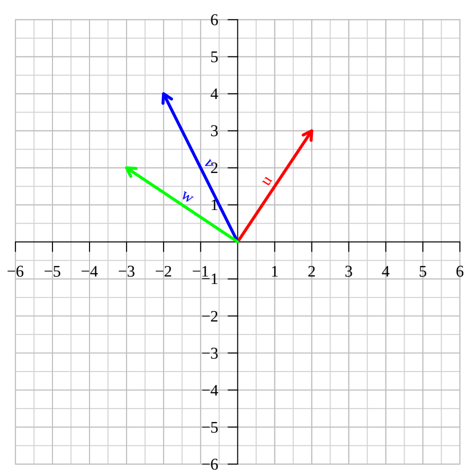
$$\theta = \cos^{-1}\left(\frac{a}{\sqrt{a^2 + b^2}}\right).$$

One of the most useful applications of the dot product is that

- *the dot product of two vectors is zero if the vectors are* perpendicular *(or* orthogonal*), and*
- *the dot product is nonzero if the two vectors are* not perpendicular.

In other words, if $\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} = 0$ then $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ are perpendicular. If $\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} \neq 0$, then $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ are not perpendicular. A proof can be found in linear algebra texts.

Example 4.2. It is easy to see from the picture below that $\vec{\mathbf{u}} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ and $\vec{\mathbf{w}} = \begin{pmatrix} -3 \\ 2 \end{pmatrix}$ are orthogonal, but neither are orthogonal to $\vec{\mathbf{v}} = \begin{pmatrix} -2 \\ 4 \end{pmatrix}$. Furthermore $\vec{\mathbf{u}} \cdot \vec{\mathbf{w}} = 0$, $\vec{\mathbf{u}} \cdot \vec{\mathbf{v}} = 8$, and $\vec{\mathbf{w}} \cdot \vec{\mathbf{v}} = 14$.



Example 4.3. *Question*: Use the dot product to decide which pairs of the given three vectors are perpendicular.

$$\mathbf{u} = \begin{pmatrix} 2 \\ -1 \\ 3 \\ 1 \end{pmatrix} \qquad \mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad \mathbf{w} = \begin{pmatrix} 2 \\ 2 \\ 1 \\ -5 \end{pmatrix}.$$

Solution: Computing dot products, we have $\mathbf{u} \cdot \mathbf{v} = 5$, $\mathbf{u} \cdot \mathbf{w} = 0$, and $\mathbf{v} \cdot \mathbf{w} = 0$. Thus $\mathbf{u}$ and $\mathbf{w}$ are perpendicular; $\mathbf{v}$and $\mathbf{w}$ are perpendicular; but $\mathbf{u}$ and $\mathbf{v}$ are not perpendicular.

Example 4.4. The dot product criterion for perpendicularity is a vast generalization to higher dimensions of the high school test for perpendicularity of two lines in the plane. The vectors $\begin{pmatrix} 1 \\ m \end{pmatrix}$ and $\begin{pmatrix} 1 \\ n \end{pmatrix}$ in the plane, which have slopes m and n (try drawing a figure), are perpendicular if and only if $\begin{pmatrix} 1 \\ m \end{pmatrix} \cdot \begin{pmatrix} 1 \\ n \end{pmatrix} = 1 + mn = 0$; that is, if and only if the product of their slopes is -1.

4.2 Spanning: which vectors can we get to?

We start with the key definition.

Definition. The *span* of a list of vectors is the set of all linear combinations that can be made with those vectors.

Example 4.5. The vector $\begin{pmatrix} 6 \\ 2 \end{pmatrix}$ is in the span of the list $\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ as is shown by the equation

$$\begin{pmatrix} 6 \\ 2 \end{pmatrix} = 4 \begin{pmatrix} 2 \\ 1 \end{pmatrix} - 2 \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Example 4.6. The vector $\begin{pmatrix} 6 \\ 2 \\ 5 \end{pmatrix}$ is not in the span of the vectors $\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ because no linear combination

$$x \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2x + y \\ x + y \\ 0 \end{pmatrix}$$

has third coordinate 5.

Example 4.7. *Question*: Let $\vec{\mathbf{u}} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and $\vec{\mathbf{v}} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$.

(a) Describe the span of $\vec{\mathbf{u}}$,

(b) Describe the span of $\vec{\mathbf{u}}, \vec{\mathbf{v}}$.

Solution: (a) The span of $\vec{\mathbf{u}}$ is the set of all linear combinations of the single vector $\vec{\mathbf{u}}$, which means all multiples $x\vec{\mathbf{u}}$ of $\vec{\mathbf{u}}$. The multiples all lie on a line through the origin in 3-space. The span consists of all vectors on that line.

(b) The span of $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$ is the set of all linear combinations $x\vec{\mathbf{u}} + y\vec{\mathbf{v}}$. This includes vectors on the line $x\vec{\mathbf{u}} + 0\vec{\mathbf{v}} = x\vec{\mathbf{u}}$ of multiples of $\vec{\mathbf{u}}$ and also vectors on the different line $0\vec{\mathbf{u}} + y\vec{\mathbf{v}} = y\vec{\mathbf{v}}$ of multiples of $\vec{\mathbf{v}}$. But it includes other vectors such as $\vec{\mathbf{u}} + \vec{\mathbf{v}}$. The span consists of all vectors in the plane through the origin that contains both $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$

In terms of matrices, if we put our vectors in the column of a matrix A, then $A\vec{\mathbf{x}}$ is the set of all vectors we can get to using A. For example, in Example 4.6, we see that

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x+y \\ x+y \\ 0 \end{pmatrix},$$

is the set of all possible vectors that we can get to using the columns of A, as we let x and y vary over all possible real numbers.

Similarly,

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x+y \\ x+2y \end{pmatrix}, \qquad \text{and} \qquad \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ x+2y \\ 1 \end{pmatrix},$$

are all of the vectors we can get to in Examples 4.5 and 4.7, respectively.

4.3 Linear independence: are there redundant vectors?

Definition. A list of vectors is *linearly independent* if no vector on the list is a linear combination of the other vectors on the list. A list of vectors is *linearly dependent* if at least one of the vectors on the list is a linear combination of the other vectors on the list.

Example 4.8. The list

$$\vec{\mathbf{i}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \vec{\mathbf{j}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \vec{\mathbf{k}} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

is linearly independent because there are no constants x and y such that $\vec{\mathbf{i}} = x\vec{\mathbf{j}} + y\vec{\mathbf{k}}$ or $\vec{\mathbf{j}} = x\vec{\mathbf{i}} + y\mathbf{k}$ or $\vec{\mathbf{k}} = x\vec{\mathbf{i}} + y\vec{\mathbf{j}}$.

Example 4.9. The list

$$\vec{\mathbf{i}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \vec{\mathbf{j}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \vec{\mathbf{w}} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

$\vec{\mathbf{i}}$, $\vec{\mathbf{j}}$, is linearly dependent because $\vec{\mathbf{w}} = \vec{\mathbf{i}} - \vec{\mathbf{j}}$ is a linear combination of $\vec{\mathbf{i}}$ and $\vec{\mathbf{j}}$.

Example 4.10. The list

$$\vec{\mathbf{u}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \vec{\mathbf{v}} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \vec{\mathbf{w}} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

is linearly independent. The vector $\vec{\mathbf{u}}$ cannot be written as a combination $\vec{\mathbf{u}} = x_2\vec{\mathbf{v}} + x_3\vec{\mathbf{w}}$, because any vector of that form will have non-zero scalars in the second or third coordinate (unless x_2 and x_3 are both 0). The vector $\vec{\mathbf{v}}$ cannot be written as a combination $\vec{\mathbf{v}} = x_1\vec{\mathbf{u}} + x_3\vec{\mathbf{w}}$ (this is a homework problem). Finally, the vector $\vec{\mathbf{w}}$ cannot be written as a combination $\vec{\mathbf{w}} = x_1\vec{\mathbf{u}} + x_2\vec{\mathbf{v}}$, since all the vectors of this form will have a 0 in the third coordinate.

Example 4.11. The list

$$\vec{\mathbf{u}} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \vec{\mathbf{v}} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \vec{\mathbf{w}} = \begin{pmatrix} 3 \\ 7 \end{pmatrix}$$

is linearly dependent, since $\vec{\mathbf{w}} = -3\vec{\mathbf{u}} + 2\vec{\mathbf{v}}$.

You can think of a linearly dependent list of vectors as a list with some built-in *redundancy.* In Example 4.11 the vector $\vec{\mathbf{w}}$ is not needed, since anything that it can get to, could be gotten to using only $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$. In Example 4.9, the span of $\vec{\mathbf{i}}$ is the line of all multiples $x\vec{\mathbf{i}}$, and the span of $\vec{\mathbf{i}}$ and $\vec{\mathbf{j}}$ is the plane of all linear combinations of those two vectors, i.e.,

$$x\vec{\mathbf{i}} = \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} \qquad \text{and} \qquad x\vec{\mathbf{i}} + y\vec{\mathbf{j}} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix},$$

The second span is larger than the first span and it contains the first span. But the span of $\vec{\mathbf{i}}$, $\vec{\mathbf{j}}$, and $\vec{\mathbf{w}}$ is just the same as the span of $\vec{\mathbf{i}}$, $\vec{\mathbf{j}}$ because the new vector $\mathbf{w}$ on the list does not enable you to get out of the xy-plane already reached by $\vec{\mathbf{i}}$ and $\vec{\mathbf{j}}$. The problem is that $\vec{\mathbf{i}}$, $\vec{\mathbf{j}}$, $\mathbf{w}$ is a linearly dependent list. The vector $\vec{\mathbf{w}}$ adds nothing to the reach of the vectors $\vec{\mathbf{i}}$ and $\vec{\mathbf{j}}$ because $\vec{\mathbf{w}}$ itself is already reachable by a linear combination of $\vec{\mathbf{i}}$ and $\vec{\mathbf{j}}$. In terms of matrices, you could say that

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

reaches just as many vectors as does

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

and the third column of this matrix is redundant.

Similarly, the third column of (see Example 4.11)

$$\begin{pmatrix} 1 & 3 & 3 \\ -1 & 2 & 7 \end{pmatrix}$$

is redundant, since we can reach just as many vectors using the first two vectors.

Generally speaking, you can expect that lengthening a list of vectors increases the span of the list, that with more vectors to start with you can reach more vectors by taking linear combinations. This is true if the longer list is linearly independent, but if lengthening the list creates linear dependence then the span may not be as large as you might guess.

Example 4.12. *Question*: Is the span of the vectors $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$, $\vec{\mathbf{w}}$ shown below a line, a plane, or all of 3-space?

$$\vec{\mathbf{u}} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \vec{\mathbf{v}} = \begin{pmatrix} 3 \\ -2 \\ -1 \end{pmatrix}, \vec{\mathbf{w}} = \begin{pmatrix} 5 \\ 0 \\ 1 \end{pmatrix}.$$

Solution: The span of $\vec{\mathbf{u}}$ is a line, the set of all multiples of the single vector $\vec{\mathbf{u}}$. Since the vector $\vec{\mathbf{v}}$ is not a multiple of $\vec{\mathbf{u}}$ it is not on the same line. The vector $\vec{\mathbf{v}}$ gives a second independent direction, so $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$ spans a plane. However, $\vec{\mathbf{w}} = 2\vec{\mathbf{u}} + \vec{\mathbf{v}}$ so $\mathbf{w}$ is already in the plane of $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$ and does not provide an independent direction. The span of $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$, $\vec{\mathbf{w}}$ is the same as the span of $\vec{\mathbf{u}}$ and $\vec{\mathbf{v}}$, which is a plane. The list $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$, $\mathbf{w}$ is linearly dependent.

Example 4.13. The concepts of spanning and linear independence/dependence explain the trichotomy in the number of solutions of linear equations we have seen before. For example, given specific vectors $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$, $\vec{\mathbf{w}}$, $\vec{\mathbf{b}}$, the *number* of solutions (x, y, z) of the equation

$$x\vec{\mathbf{u}} + y\vec{\mathbf{v}} + z\vec{\mathbf{w}} = \vec{\mathbf{b}}$$

is

(0) 0 if $\vec{\mathbf{b}}$ is not in the span of $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$, $\vec{\mathbf{w}}$.

(1) 1 if $\vec{\mathbf{b}}$ is in the span of $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$, $\mathbf{w}$ and $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$, $\vec{\mathbf{w}}$ is a linearly independent list.

(∞) infinite if $\vec{\mathbf{b}}$ is in the span of $\vec{\mathbf{u}}$, $\mathbf{v}$, $\vec{\mathbf{w}}$ and $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$, $\vec{\mathbf{w}}$ is a linearly dependent list.

Example 4.14. *Question*: You are interested in the equation $x\vec{\mathbf{u}} + y\vec{\mathbf{v}} + z\vec{\mathbf{w}} = \vec{\mathbf{b}}$ where $\vec{\mathbf{u}}$, $\vec{\mathbf{v}}$, $\vec{\mathbf{w}}$, $\vec{\mathbf{b}}$ are specific vectors. You happen to know that $(x, y, z) = (2, 3, 5)$ is a solution, and that $\vec{\mathbf{w}} = \vec{\mathbf{u}} - \vec{\mathbf{v}}$. Find two more solutions.

Solution: We have been given two equations relating the vectors $\vec{\mathbf{u}}$, $\mathbf{v}$, and $\vec{\mathbf{w}}$:

$$\begin{aligned} 2\vec{\mathbf{u}} + 3\vec{\mathbf{v}} + 5\vec{\mathbf{w}} &= \mathbf{b} \\ \vec{\mathbf{u}} - \vec{\mathbf{v}} - \vec{\mathbf{w}} &= \mathbf{0}. \end{aligned}$$

Addition of the two equations gives

$$3\vec{\mathbf{u}} + 2\vec{\mathbf{v}} + 4\vec{\mathbf{w}} = \vec{\mathbf{b}}$$

which shows that $(x, y, z) = (3, 2, 4)$ is a solution of the original equation. To find more solutions, add any multiple $n\vec{\mathbf{u}} - n\vec{\mathbf{v}} - n\vec{\mathbf{w}} = \mathbf{0}$ to $2\vec{\mathbf{u}} + 3\vec{\mathbf{v}} + 5\vec{\mathbf{w}} = \vec{\mathbf{b}}$ to get

$$(2 + n)\vec{\mathbf{u}} + (3 - n)\vec{\mathbf{v}} + (5 - n)\vec{\mathbf{w}} = \vec{\mathbf{b}}$$

which shows that $(x, y, z) = (2 + n, 3 - n, 5 - n)$ solves the original equation for any n. For example, with $n = 2$ we have the solution $(x, y, z) = (4, 1, 3)$. If you choose n to be your student id number, then everyone in class will have a different solution to this problem.

4.4 Subspaces and dimension

Definition. A *subspace* is a set of vectors that is the span of some list of vectors. The *dimension* of a subspace is the minimum number of vectors required to span the subspace.

As suggested by the preceding examples, a 1-dimensional vector space is a line through the origin and a 2-dimensional subspace is a plane through the origin. A subspace is a very special collection of vectors.

It can be shown that a list of d linearly independent vectors always spans a d-dimensional subspace. A list of d linearly dependent vectors spans a subspace of dimension less than d because you can always leave out one of the vectors in the list and still span the same subspace.

4.5 Exercises

1. Find the length of the vector $(-1, 2, 0, 4)$.

2. Find the angles between the following pairs of vectors, drawing the vectors on the graph paper to confirm that your answer is right.

   (a) $(1, 0)$ and $(0, 1)$
   (b) $(1, 1)$ and $(1, 0)$
   (c) $(1, 1)$ and $(-1, 1)$
   (d) $(3, 2)$ and $(-3, -2)$

3. Find three non-parallel vectors each orthogonal to $(3, 2, 5, 4)$.

4. Is the vector $(3, 1, 2)$ orthogonal to the plane spanned by $(0, 2, -1)$ and $(-1, 3, 0)$?

5. Find the angles between the following pairs of vectors:

   (a) $(5, 2, 3)$ and $(-3, 2, -3)$
   (b) $(7, 4, 2, 3)$ and $(4, 2, -3, 7)$

6. Let
$$\vec{\mathbf{u}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \vec{\mathbf{v}} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \vec{\mathbf{w}} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$
and show that it is not possible to find vectors x_1 and x_3 so that $\vec{\mathbf{v}} = x_1\vec{\mathbf{u}} + x_3\vec{\mathbf{w}}$ (this is part of Example 4.10).

7. The following vectors are linearly dependent. Write one of them as a linear combination of the others
$$\begin{pmatrix} 10 \\ 6 \end{pmatrix}, \quad \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

8. How many solutions does the following system have?
$$\begin{pmatrix} 10 & 1 & 3 \\ 6 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 10 \\ 6 \end{pmatrix}$$

If it has one solution, find it. If it has more than one solution, find two of them. If it has no solutions, explain how you know that.

9. How many solutions does the following system have?

$$\begin{pmatrix} 10 & 1 \\ 6 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 10 \\ 6 \end{pmatrix}$$

If it has one solution, find it. If it has more than one solution, find two of them. If it has no solutions, explain how you know that.

10. How many solutions does the following system have?

$$\begin{pmatrix} 10 & 5 \\ 6 & 3 \\ 4 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 10 \\ 6 \\ 1 \end{pmatrix}$$

If it has one solution, find it. If it has more than one solution, find two of them. If it has no solutions, explain how you know that.

11. For each of the matrices that appear in problem 3, 4, and 5, describe the dimension of the space spanned by the columns of the matrix.

5 Projections and Least Squares Approximation

5.1 Motivation

In applied mathematics many of the equations that you would like to solve do not have any solutions. The reason is that there are very often more equations than variables, often far more equations. This is especially so when doing statistics, because the goal is to explain a large quantity of data with a small number of parameters.

When there are more equations than variables you can not expect to find a solution, though this situation does not automatically rule out the existence of a solution. We can get an idea of the problem by considering motivating example using three equations in two variables.

Example 5.1. *Points interpretation.* Consider the following system of three linear equations in two unknowns:

$$\begin{aligned} 3x + 2y &= 7 \\ x - y &= 6 \\ x + y &= 14 \end{aligned}$$

Each of the three equations describes a line in the plane and we are looking for a single point on all three. The first two lines intersect in a single point. It would be amazing if a randomly chosen third line happened to pass through that point, and in this case, it does not.

Vector interpretation. The vector interpretation of the system of equations in the preceding example is:

$$x \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} + y \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 6 \\ 14 \end{pmatrix}.$$

The linear combinations of the two vectors $\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}$ all lie in a plane through the origin in 3-dimensional space. It would be amazing if a randomly chosen third vector happened to lie on the same plane, and in this case, the the target vector $\begin{pmatrix} 7 \\ 6 \\ 14 \end{pmatrix}$ does not.

When confronted with equations with no solutions, we do not give up. We look for *approximate* solutions. In the preceding examples, there is no choice of x and y for which the left side of the equations equals the right side. We change the question and ask: For which values of x and y is the left side of the equation *closest* to the right side? It turns out that this question is easiest to understand in the vector interpretation, which is one of the main reasons why vectors are introduced in this course.

To solve this problem, we need to further study the geometry of vectors, mainly lengths of vectors and angles between them. The fundamental tool that we will use is the dot product.

5.2 Projection of a vector on a line

Example. Let $\mathbf{u} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $\mathbf{a} = \begin{pmatrix} 6 \\ 8 \end{pmatrix}$. The equation $x\mathbf{u} = \mathbf{a}$ has no solution because the vector $\mathbf{a}$ is not a multiple of the vector $\mathbf{u}$, as can be seen from the Figure.

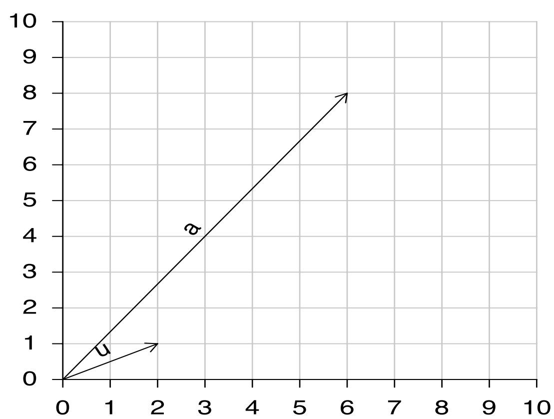


Figure: The vector $\mathbf{a}$ is not a multiple of $\mathbf{u}$.

Since we can not make $x\mathbf{u}$ equal to $\mathbf{a}$, let's try to make it close. The question is, what multiple $x\mathbf{u}$ of $\mathbf{u}$ comes closest to $\mathbf{a}$? The next figure shows how to find it geometrically. The segment from the tip of the desired vector $x\mathbf{u}$ to the tip of $\mathbf{a}$ is perpendicular to the line through $\mathbf{u}$. Denoting the vector along this segment by $\mathbf{r}$, called the *residual vector*, we can express the geometry in equations:

$$\begin{aligned} x\mathbf{u} + \mathbf{r} &= \mathbf{a} \\ \mathbf{u} \cdot \mathbf{r} &= 0. \end{aligned}$$

The goal is to make $\mathbf{r}$ as short as possible. That is achieved by the second equation, which says that $\mathbf{r}$ is perpendicular to the line through $\mathbf{u}$.

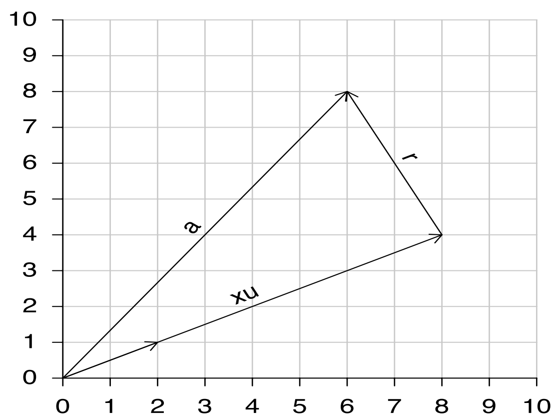


Figure: The vector $x\mathbf{u}$ is the projection of $\mathbf{a}$ on the line spanned by $\mathbf{u}$.

To solve for the value of m, take the dot product of the first equation with the vector $\mathbf{u}$.

$$\begin{aligned} \mathbf{u} \cdot (x\mathbf{u} + \mathbf{r}) &= \mathbf{u} \cdot \mathbf{a} \\ x\mathbf{u} \cdot \mathbf{u} + \mathbf{u} \cdot \mathbf{r} &= \mathbf{u} \cdot \mathbf{a} \\ x\mathbf{u} \cdot \mathbf{u} + 0 &= \mathbf{u} \cdot \mathbf{a} \\ x\mathbf{u} \cdot \mathbf{u} &= \mathbf{u} \cdot \mathbf{a} \\ x &= \frac{\mathbf{u} \cdot \mathbf{a}}{\mathbf{u} \cdot \mathbf{u}} \end{aligned}$$

Plugging in our values for $\mathbf{a}$ and $\mathbf{u}$, we find that

$$x = \frac{12 + 8}{4 + 1} = \frac{20}{5} = 4.$$

Thus

$$x\mathbf{u} = 4\mathbf{u} = \begin{pmatrix} 8 \\ 4 \end{pmatrix}$$

and

$$\mathbf{r} = \mathbf{a} - x\mathbf{u} = \begin{pmatrix} -2 \\ 4 \end{pmatrix}.$$

The vector $x\mathbf{u}$ is the *projection of* $\mathbf{a}$ *on the line spanned by* $\mathbf{u}$. In our calculation

Our argument leading to the formula for the value of x is a general one: it applied to any vectors $\mathbf{a}$ and $\mathbf{u}$ (in *any* vector space R^n):

The formula for the projection of $\mathbf{a}$ on the line spanned by $\mathbf{u}$ is $x\mathbf{u}$ where

$$x = \frac{\mathbf{u} \cdot \mathbf{a}}{\mathbf{u} \cdot \mathbf{u}}.$$

5.3 Least squares approximation with one parameter

In this section, we show that finding a least squares approximation for a one parameter model is solved by projecting a vector onto a line through the origin.

Suppose you believe that two variables x and y are proportional, so that $y = mx$ for some proportionality constant m that you want to determine. You measure a specimen and find $x_1 = 5$, $y_1 = 10$. To be on the safe side, you measure a second specimen and find $x_2 = 8$, $y_2 = 15$. Now you have a problem! The single constant m must satisfy both equations in the system

$$\begin{aligned} 5m &= 10 \\ 8m &= 15 \end{aligned}$$

which has no solution. The geometric explanation is that the two data points $(5, 10)$ and $(8, 15)$ do not lie on a line through the origin in the xy-plane. The physical explanation may be that there is some measurement error that causes measurements not to be exactly proportional even though the true quantities are proportional. We want to use the data we have to do the best we can at figuring out what the correct proportionality constant m is.

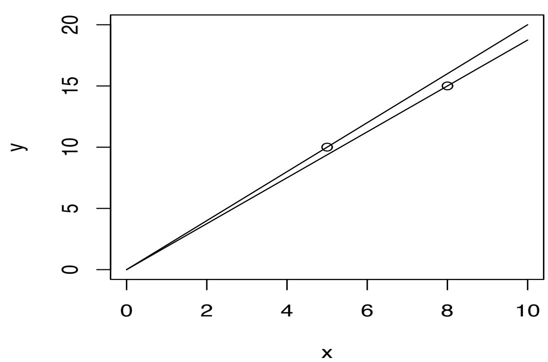


Figure: The two data points are on different lines through the origin.

Let's look at the vector interpretation of the system of equations.

$$m \begin{pmatrix} 5 \\ 8 \end{pmatrix} = \begin{pmatrix} 10 \\ 15 \end{pmatrix}.$$

Again, there is no solution. The geometric explanation is that the vector $\mathbf{y} = \begin{pmatrix} 10 \\ 15 \end{pmatrix}$ can not be reached by multiples of the vector $\mathbf{x} = \begin{pmatrix} 5 \\ 8 \end{pmatrix}$.

The figure shows that there is a multiple $\hat{m}\mathbf{x}$ of $\mathbf{x}$ that reaches closest to $\mathbf{y}$. That multiple is called the *projection of* $\mathbf{y}$ *on the line spanned by* $\mathbf{x}$. The step from the tip of the projection $\hat{m}\mathbf{x}$ to $\mathbf{y}$ is the *residual vector* $\mathbf{r}$. We will never know the value of m; but we can find $\hat{m}$ which is our best estimate of m. We can evaluate the quality of our estimate by the length of the residual vector $\mathbf{r}$, because $\|\mathbf{r}\|$ is the distance between $\hat{m}\mathbf{x}$ and $\mathbf{y}$, the two sides of what we had hoped would be an equality between $m\mathbf{x}$ and $\mathbf{y}$.

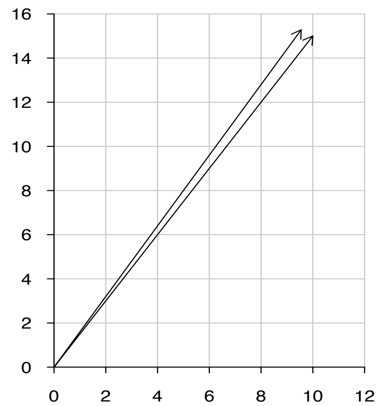


Figure: The vector $\hat{m}\mathbf{x}$ is close to $\mathbf{y}$.

How to find $\hat{m}$ and $\mathbf{r}$? We use the fact, clear from the Figure, that $\hat{m}\mathbf{x}$ and $\mathbf{r}$ are perpendicular, namely that $\mathbf{x} \cdot \mathbf{r} = 0$. We must solve for $\hat{m}$ and $\mathbf{r}$ in the two equations that describe the geometry:

$$\begin{aligned} \hat{m}\mathbf{x} + \mathbf{r} &= \mathbf{y} \\ \mathbf{x} \cdot \mathbf{r} &= 0. \end{aligned}$$

To solve, take the dot product of the first equation with the vector $\mathbf{x}$, then do some simple algebra.

$$\begin{aligned} \mathbf{x} \cdot (\hat{m}\mathbf{x} + \mathbf{r}) &= \mathbf{x} \cdot \mathbf{y} \\ \hat{m}\mathbf{x} \cdot \mathbf{x} + \mathbf{x} \cdot \mathbf{r} &= \mathbf{x} \cdot \mathbf{y} \\ \hat{m}\mathbf{x} \cdot \mathbf{x} + 0 &= \mathbf{x} \cdot \mathbf{y} \\ \hat{m} &= \frac{\mathbf{x} \cdot \mathbf{y}}{\mathbf{x} \cdot \mathbf{x}} \end{aligned}$$

In our example,

$$\hat{m} = 170/89 = 1.91$$

The residual vector is

$$\mathbf{r} = \mathbf{y} - \hat{m}\mathbf{x} = \begin{pmatrix} 0.45 \\ -0.28 \end{pmatrix}$$

which has length

$$\|\mathbf{r}\| = \sqrt{0.45^2 + (-0.28)^2} = 1.6$$

The projection of the vector $\mathbf{y}$ on the line generated by $\mathbf{x}$ is the vector

$$\hat{m}\mathbf{x} = \begin{pmatrix} 9.55 \\ 15.28 \end{pmatrix}$$

The projection is the closest you can get to $\mathbf{y}$ using only multiples of $\mathbf{x}$. See all these on Figure.

In summary, we have $\hat{m}\mathbf{x} + \mathbf{r} = \mathbf{y}$ where the shorter the residual vector $\mathbf{r}$, the closer $\hat{m}$ comes to solving the original unsolvable equation $m\mathbf{x} = \mathbf{y}$.

Returning to the initial problem of modeling the relationship between x and y with a proportion, we find that $y = \hat{m}x$ is a good fit to the data. The residual error is represented on the graph by the vertical segments from the modeling line to the two data points. The data point $(5, 10)$ is 0.45 above the line $y = \hat{m}x$, and the data point $(8, 15)$ is 0.28 below the line. These residual errors give the coordinates of the residual vector $\mathbf{r} = \begin{pmatrix} 0.45 \\ -0.28 \end{pmatrix}$, but $\mathbf{r}$ itself is not in this figure. Notice that minimizing the length of $\mathbf{r}$ is equivalent to minimizing the sum of the squares of the residual errors. The line $y = \hat{m}x$ is called the *least squares fit* to the data.

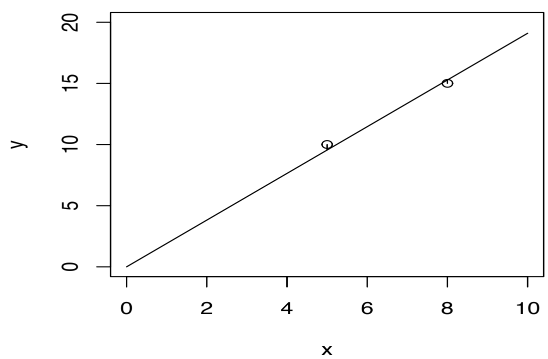


Figure: The line $y = \hat{m}x$ is a good fit to the data, with $\hat{m} = 1.91$.

Example 5.2. *Isometry of the human arm, concluded.* Now that we know how to solve projections onto a 1-dimensional subspace, we can complete our analysis of the isometry of the human arm, which was introduced in Example 1.1. Let $\hat{a}$ be the coefficient of our least squares solution. We have

$$\vec{\mathbf{y}} = \hat{a}\vec{\mathbf{x}} + \vec{\mathbf{r}}$$

and as in the previous example, we can take the dot product with $\vec{\mathbf{x}}$ and solve for $\hat{a}$ to obtain

$$\hat{a} = \frac{\vec{\mathbf{x}} \cdot \vec{\mathbf{y}}}{\vec{\mathbf{x}} \cdot \vec{\mathbf{x}}} = \frac{5790.76}{3384.16} = 1.711.$$

Our residual vector is

$$\vec{\mathbf{r}} = \begin{pmatrix} 44.5 \\ 41.4 \\ 46.1 \\ 43.6 \\ 46.8 \end{pmatrix} - 1.711 \begin{pmatrix} 26.0 \\ 24.4 \\ 26.4 \\ 26.0 \\ 27.2 \end{pmatrix} = \begin{pmatrix} 0.01 \\ -0.35 \\ 0.93 \\ -0.89 \\ 0.26 \end{pmatrix}$$

which has length $\|\vec{\mathbf{r}}\| = 1.36$.

5.4 Least squares approximation with two parameters

In this section, we show that finding a least squares approximation for a two parameter model is solved by projecting a vector onto a 2-dimensional subspace (which is a plane through the origin).

Suppose that we have N data points (x, y). Furthermore, we believe that these data points are modeled by a line $y = mx + b$. We have to approximate *two* parameters: m and b. We achieve this goal by generalizing our method in the previous example. Instead of projecting onto a 1-dimensional subspace of $\mathbb{R}^N$, we will project onto a 2-dimensional subspace of $\mathbb{R}^N$. A 2-dimensional subspace looks like a plane sitting inside the larger vector space $\mathbb{R}^N$.

Example 5.3. Suppose that a business wants to know the effect of advertising dollars on sales. They gather the following data

a(advertising in \$1000s)	3	4	5	6
s(sales in \$1000s)	105	117	141	152

and plot it to observe that there is roughly a linear relationship between sales and advertising dollars:

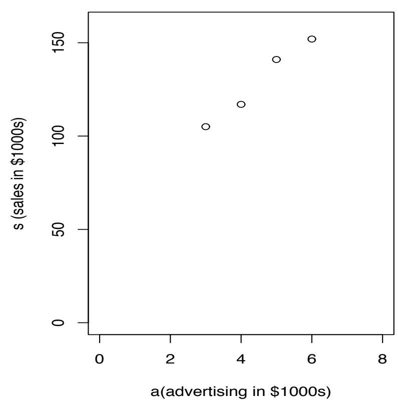


Figure: The data shows a linear tendency.

The scatterplot of s vs a reveals an approximately linear relationship of the form $s = ma + b$, where slope m and intercept b are to be determined from the data. For such an equation to hold, we must have

$$\begin{aligned} 3m + b &= 105 \\ 4m + b &= 117 \\ 5m + b &= 141 \\ 6m + b &= 152. \end{aligned}$$

There can not be an exact solution: from the picture it is clear that the four data points (a, s) do not lie in a single line. Instead, we want to find values for m and b so that the line $y = mx + b$ is close to all four data points.

In order to find our best approximation, we put this system of equations into vector form:

$$m \begin{pmatrix} 3 \\ 4 \\ 5 \\ 6 \end{pmatrix} + b \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 105 \\ 117 \\ 141 \\ 152 \end{pmatrix}.$$

We can also rewrite this system in matrix form:

$$\begin{pmatrix} 3 & 1 \\ 4 & 1 \\ 5 & 1 \\ 6 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 105 \\ 117 \\ 141 \\ 152 \end{pmatrix}.$$

This matrix form is particularly useful when solving problems using mathematical software.

We have now turned our data fitting problem (in $\mathbb{R}^2$) into a linear combination problem (in $\mathbb{R}^4$). Indeed, we are looking for a linear combination of two vectors $\mathbf{u}$ and $\mathbf{v}$ that equals a given vector $\mathbf{s}$, where

$$\mathbf{u} = \begin{pmatrix} 3 \\ 4 \\ 5 \\ 6 \end{pmatrix} \qquad \mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \qquad \mathbf{s} = \begin{pmatrix} 105 \\ 117 \\ 141 \\ 152 \end{pmatrix}.$$

There is no such linear combination. Linear combinations of two vectors $\mathbf{u}$ and $\mathbf{v}$ lie in only a small 2-dimensional subspace of the full 4-dimensional space of vectors, and the vector $\mathbf{s}$ does not happen to lie in that subspace.

We do not give up. We change the question to: *Find the linear combination* $\hat{m}\mathbf{u} + \hat{b}\mathbf{v}$ *that comes closest to* $\mathbf{s}$. We will have to add an additional vector $\mathbf{r}$ to that linear combination to get all the way to $\mathbf{s}$. The key observation is that the additional vector $\mathbf{r}$ is perpendicular to all the vectors in the plane of linear combinations of $\mathbf{u}$ and $\mathbf{v}$, including both $\mathbf{u}$ and $\mathbf{v}$. You can understand the geometry physically by thinking about the shortest line segment from a point in your room to the plane of the floor; that segment is perpendicular to the floor. Expressing the geometry in equations, we have:

$$\begin{aligned} \hat{m}\mathbf{u} + \hat{b}\mathbf{v} + \mathbf{r} &= \mathbf{s} \\ \mathbf{u} \cdot \mathbf{r} &= 0 \\ \mathbf{v} \cdot \mathbf{r} &= 0. \end{aligned}$$

These equations are usually solved by computer. To find a solution algebraically, take the dot product of the first equation with $\mathbf{u}$ and then with $\mathbf{v}$, getting the system

$$\begin{aligned} \hat{m}\mathbf{u} \cdot \mathbf{u} + \hat{b}\mathbf{u} \cdot \mathbf{v} &= \mathbf{u} \cdot \mathbf{s} \\ \hat{m}\mathbf{v} \cdot \mathbf{u} + \hat{b}\mathbf{v} \cdot \mathbf{v} &= \mathbf{v} \cdot \mathbf{s} \end{aligned}$$

which after computing the dot products is

$$\begin{aligned} 86\hat{m} + 18\hat{b} &= 2400 \\ 18\hat{m} + 4\hat{b} &= 515. \end{aligned}$$

The system of two equations in two unknowns has solution $\hat{m} = 16.5$ and $\hat{b} = 54.5$.

The projection of the vector $\mathbf{s}$ on the subspace spanned by the vectors $\mathbf{u}$ and $\mathbf{v}$ is the vector

$$\hat{m}\mathbf{u} + \hat{b}\mathbf{v} = \begin{pmatrix} 104.0 \\ 120.5 \\ 137.0 \\ 153.5 \end{pmatrix}$$

which is as close to $\mathbf{s}$ as you can get with linear combinations of $\mathbf{u}$ and $\mathbf{v}$.

The residual vector

$$\mathbf{r} = \mathbf{s} - \text{projection} = \begin{pmatrix} 1.0 \\ -3.5 \\ 4.0 \\ -1.5 \end{pmatrix}$$

has length

$$\|\mathbf{r}\| = 5.6$$

The graph shows that the line $s = \hat{m}a + \hat{b}$ gives a good fit to the data. The four residual errors, represented on the figure by short vertical segments from the data points to the line, give the four coordinates of the residual vector $\mathbf{r}$. Since $\mathbf{r}$ is as short as possible, the sum of the squares of the residual errors is minimized. The line is the least squares fit to the data.

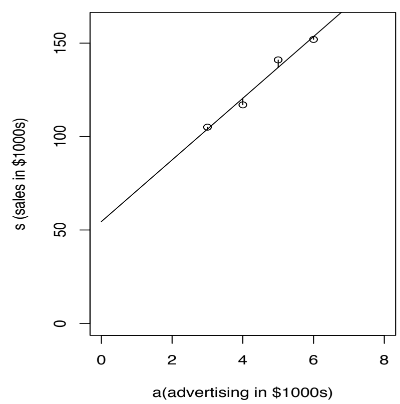


Figure: The line $s = \hat{m}a + \hat{b}$ is a good fit with $\hat{m} = 16.5$ and $\hat{b} = 54.5$.

5.5 The fundamental problem of linear modeling

We can generalize even further. We solve any least square curve fitting by projecting onto a subspace of a larger vector space. If we have N data points and our model requires k parameters, then the least squares method becomes a projection onto a subspace spanned by k vectors. This subspace has dimension at most k, since it is spanned by k vectors. In summary:

	Curve fitting problem	**Corresponding vector problem**
N	number of data points	dimension of the vector space
k	number of parameters in the model	upper bound on the dimension of the subspace

The preceding sections illustrate the **Fundamental Problem of Linear Modeling:** Given $k+1$ vectors in $\mathbb{R}^N$

$$\mathbf{u_1}, \mathbf{u_2}, \ldots, \mathbf{u_k} \text{ and } \mathbf{b},$$

find a linear combination

$$x_1\mathbf{u_1} + x_2\mathbf{u_2} + \cdots + x_k\mathbf{u_k}$$

that comes as close to $\mathbf{b}$ as possible.

Equivalently, find a linear combination

$$x_1\mathbf{u_1} + x_2\mathbf{u_2} + \cdots + x_k\mathbf{u_k}$$

and a vector $\mathbf{r}$ such that

$$x_1\mathbf{u_1} + x_2\mathbf{u_2} + \cdots + x_k\mathbf{u_k} + \mathbf{r} = \mathbf{b}$$

and the length $\|\mathbf{r}\|$ of $\mathbf{r}$ is as small as possible.

We can rewrite this equation in matrix form. Let

$$U = \begin{pmatrix} \mathbf{u_1} & \mathbf{u_2} & \cdots & \mathbf{u_k} \end{pmatrix}$$

be the $N \times k$ matrix whose jth column is $\mathbf{u_j}$. Let

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}.$$

Our equation becomes

$$U\mathbf{x} + \mathbf{r} = \mathbf{b}.$$

Here U is an $N \times k$ matrix and $\mathbf{x} \in R^k$, so $U\mathbf{x} \in R^N$. We want to find the vector $\mathbf{x} \in \mathbb{R}^k$ that makes the vector $\mathbf{r} \in \mathbb{R}^N$ as small as possible.

The key to a solution is the fact that the *residual vector* $\mathbf{r}$ is perpendicular to all the vectors $\mathbf{u_j}$, so

$$\mathbf{u_1} \cdot \mathbf{r} = \mathbf{u_2} \cdot \mathbf{r} = \cdots = \mathbf{u_k} \cdot \mathbf{r} = 0.$$

We take advantage of this perpendicular relationship to create k equations for the k parameters $x_1, x_2, \ldots, x_k$. We take the dot product of this equation with each of the $\mathbf{u_j}$ to get:

$$\begin{aligned}
\mathbf{u_1} \cdot (U\mathbf{x}) &= \mathbf{u_1} \cdot \mathbf{b} \\
\mathbf{u_2} \cdot (U\mathbf{x}) &= \mathbf{u_2} \cdot \mathbf{b} \\
&\vdots \\
\mathbf{u_k} \cdot (U\mathbf{x}) &= \mathbf{u_k} \cdot \mathbf{b}.
\end{aligned}$$

This system is solvable, and is handled by a computer. The optimal linear combination is called the *projection of the vector* $\mathbf{b}$ *on the subspace spanned by the vectors* $\mathbf{u_1}$, ..., $\mathbf{u_k}$.

5.6 Curve fitting for non-linear models

Even when our models are not linear, we can still use least squares to solve them, provided that we have some idea about what the model should look like.

Example 5.4. Fit a quadratic function $f(t) = c_0 + c_1 t + c_2 t^2$ to the four data points $(0, 6)$, $(1, 5)$, $(2, 2)$, $(3, 12)$.

We would like to find a quadratic function $f(t)$ such that

$$\begin{aligned} f(0) &= 6 \\ f(1) &= 5 \\ f(2) &= 2 \\ f(3) &= 12. \end{aligned}$$

Using the desired function $f(t) = c_0 + c_1 t + c_2 t^2$, this means that we want

$$\begin{array}{llll} c_0 & & & = 6 \\ c_0 & +\,c_1 & +\,c_2 & = 5 \\ c_0 & +2c_1 & +4c_2 & = 2 \\ c_0 & +3c_2 & +9c_2 & = 12 \end{array}$$

and we can rewrite this system of equations in matrix form:

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 6 \\ 5 \\ 2 \\ 12 \end{pmatrix}.$$

To solve for c_0, c_1, c_2, we want to find the least squares approximation for the vector on the right hand side into the subspace generated by the columns of the matrix on the left hand side.

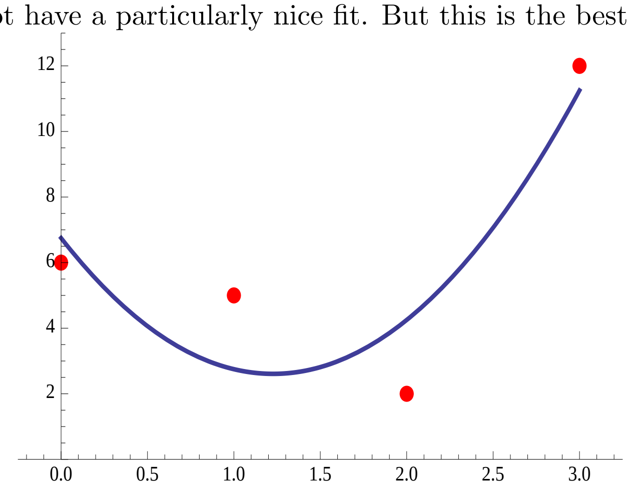
Using mathematical software (like R), we can solve to find

$$\begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 6.75 \\ -6.75 \\ 2.75 \end{pmatrix}$$

with residual vector

$$\mathbf{r} = \begin{pmatrix} -0.75 \\ 2.25 \\ -2.25 \\ 0.75 \end{pmatrix},$$

giving $\|\mathbf{r}\| = 3.35$. When we compare our curve to the original data, we see that we indeed do not have a particularly nice fit. But this is the best possible quadratic fit for the data.



Example 5.5. Let $S(t)$ be the number of daylight hours on the t day of the year. We are given the following data for $S(t)$:

Day	t	$S(t)$
February 1	32	10
March 17	77	12
April 30	121	14
May 31	152	15

Find the best fit trigonometric function of the form

$$f(t) = a + b\sin\left(\frac{2\pi}{366}t\right) + c\cos\left(\frac{2\pi}{366}t\right).$$

We have 4 data points. So we start with the system of equations

$$\begin{pmatrix} f(32) \\ f(77) \\ f(121) \\ f(152) \end{pmatrix} = \begin{pmatrix} a + 0.522b + 0.853c \\ a + 0.969b + 0.246c \\ a + 0.874b - 0.485c \\ a + 0.507b - 0.862c \end{pmatrix} = \begin{pmatrix} 1 & 0.522 & 0.853 \\ 1 & 0.969 & 0.246 \\ 1 & 0.874 & -0.485 \\ 1 & 0.507 & -0.862 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 10 \\ 12 \\ 14 \\ 15 \end{pmatrix}$$

We are ready to perform a least squares approximation. This time we are finding

the projection of $\begin{pmatrix} 10 \\ 12 \\ 14 \\ 15 \end{pmatrix}$ onto the subspace generated by $\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0.522 \\ 0.969 \\ 0.874 \\ 0.507 \end{pmatrix}, \begin{pmatrix} 0.853 \\ 0.246 \\ -0.485 \\ -0.862 \end{pmatrix}$.

We use our trusty mathematical software to find that

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 12.26 \\ 0.43 \\ -2.90 \end{pmatrix},$$

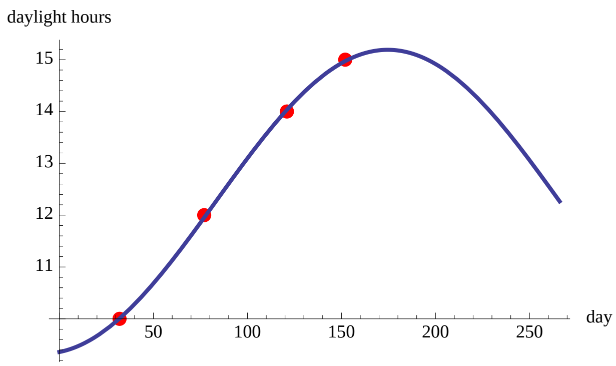
so our best fit trigonometric function is

$$f(t) = 12.26 + 0.43\sin\left(\frac{2\pi}{366}t\right) - 2.90\cos\left(\frac{2\pi}{366}t\right).$$

This time, our residual vector is

$$\mathbf{r} = \begin{pmatrix} -0.013 \\ 0.035 \\ -0.043 \\ -0.022 \end{pmatrix}.$$

with $\|\mathbf{r}\| = 0.061$, giving a nice fit that can be seen in the corresponding graph.



5.7 Exercises

1. Using graph paper, project the vector $(3, 2)$ onto the vector $(1, 0)$.

   (a) Give, numerically, both the projected vector and the residual vector.

   (b) Show, numerically, that the residual vector is orthogonal to the vector.

2. Suppose that we have the following table of data points:

x_i	1	2	3	4	5
y_i	5	6	8	9	9

   We want to find the line $y = mx + b$ that best fits this data.

   (a) Write down the system of equations (5 equations) and unknowns (2 unknowns) that need to be "solved" in this problem. Explain why this system of equations does not have an exact solution.

   (b) Convert this system of equations into a vector equation.

   (c) Convert this vector equation into a matrix equation of the form $A\vec{\mathbf{x}} = \vec{\mathbf{s}}$.

   (d) Using a computer, the least squares solution to this matrix equation is found to be $\vec{\mathbf{x}}^* = \begin{pmatrix} 1.1 \\ 4.1 \end{pmatrix}$.

   (e) The closest vector to $\vec{\mathbf{s}}$ in the span of A is the vector $\hat{s} = A\vec{\mathbf{x}}^*$. Compute this vector by multiplying $A\vec{\mathbf{x}}^*$.

   (f) Compute the residual vector $\vec{\mathbf{r}} = \vec{\mathbf{s}} - \hat{\vec{\mathbf{s}}}$. Show that $\vec{\mathbf{r}}$ and $\hat{s}$ are orthogonal.

   (g) Find $||\vec{\mathbf{r}}||$, this is the distance between the actual data $\vec{\mathbf{s}}$ and the corresponding data on the line $\hat{\vec{\mathbf{s}}}$.

3. We often collect data on variables that are *categorical* rather than numerical. For example, when studying trees, we might record the species of the tree as one of several possibilities: oak, maple, birch, etc. Sexes are recorded as either male or female. In this problem, we'll imagine that the possible categories are A, B, C, or D. The possible values are called the *levels* of the variable.

   Suppose that we measure 10 cases, giving us the categories (A,A,B,D,D,A,D,B,B,C). An indicator variable for a given level is a numerical vector that has the value 1 for those cases that match the level, and 0 for the other cases. For example, the indicator variable for level A would be (1,1,0,0,0,1,0,0,0,0).

When there are 4 levels, there will be 4 distinct indicator variables. Show that all 4 of these indicator variables are mutually orthogonal.

(a) Explain why, for any categorical variable, the set of indicator variables must always be mutually orthogonal.

(b) Show that the vector $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ can be written as a linear combination of the indicator variables for A, B, C, and D.

(c) Explain why, for any categorical variable, the vector of all 1s can always be written as a linear combination of the indicator variables drawn from that variable.

4. In statistics, the *correlation coefficient* is a way of describing the relationship between two variables. When the correlation coefficient is 1, any change in one variable is perfectly reflected in the other. When the correlation coefficient is 0, a change in one variable is not (on average) reflected in the other.

   For a vector $\mathbf{u}$, the vector $\bar{\mathbf{u}}$ is the vector all of whose components equal the average of the components of $\mathbf{u}$. For example , for $\mathbf{u} = (3, 5, 6, 2)$ we have $\bar{\mathbf{u}} = (4, 4, 4, 4)$.

   As it happens, the correlation coefficient between two vectors $\mathbf{u}$ and $\mathbf{v}$ is the cosine of the angle between the two vectors $\mathbf{u} - \bar{\mathbf{u}}$ and $\mathbf{v} - \bar{\mathbf{v}}$.

   Calculate the correlation coefficient between these vectors: $\mathbf{v} = (3, -4, 5, 2, 3, -6, 8)$ and $\mathbf{w} = (9, 2, 3, 5, 1, -2, 4)$.

5. Calculating the mean (average) of the components in a vector is equivalent to projecting the vector onto the vector of all ones: $(1, 1, \cdots, 1)$.

   Project each of the following vectors onto the vector of ones:

   (a) $(4, 2)$

   (b) $(-3, 3)$

   (c) $(5, 5)$

   (d) $(7, 4, -3, 4)$

6. Let P be the plane spanned by the vectors

$$\mathbf{u} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad \mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}.$$

   In this problem we will consider the vector $\mathbf{b} = \begin{pmatrix} -4 \\ 2 \\ 6 \end{pmatrix}$, and its relation to the plane P.

   You will see that this vector is not in this plane, and will see how close to $\mathbf{b}$ we can get using linear combinations of $\mathbf{u}$ and $\mathbf{v}$.

(a) Verify that $\mathbf{w} = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ is orthogonal to $\mathbf{u}$ and $\mathbf{v}$ (so it is on the line $P^{\perp}$ that is perpendicular to P). Try to picture the geometry here. Note that this implies that *any* vector in 3-space can be written as a linear combination of $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$. In the next part, you will show how $\mathbf{b}$ is a linear combination of these three vectors.

(b) Suppose that we solve the equation $\begin{pmatrix} 1 & 2 & -1 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -4 \\ 2 \\ 6 \end{pmatrix}$. This can be done using R, and we will learn how to do so in the lab. The solution is $x_1 = 3, x_2 = -1, x_3 = 5$. Verify that this is a solution by computing the linear combination on the left-hand side.

(c) Show that $\mathbf{b}$ is not in the plane P. That is, $\mathbf{b}$ is not a linear combination of the vectors $\mathbf{u}$ and $\mathbf{v}$.

(d) What is the least-squares projection of $\mathbf{b} = \begin{pmatrix} -4 \\ 2 \\ 6 \end{pmatrix}$ onto the plane P?

(e) Write $\mathbf{b} = \mathbf{p} + \mathbf{r}$ with $\mathbf{p}$ on the plane P and $\mathbf{r}$ on the line $P^{\perp}$.

(f) Verify that $\mathbf{p}$ and $\mathbf{r}$ are orthogonal.

(g) How far is $\mathbf{b}$ from P? Note that the distance from $\mathbf{b}$ from P is the distance from $\mathbf{b}$ from $\mathbf{p}$, and that this equals the length of $\mathbf{r}$.

7. This problem continues the ideas of the last problem. Consider the three points $(1,2)$ $(3,4)$ and $(4,4)$. Plot them, and see that they are not on a line. In practice, we often have *many* points that are not on a line but lay close to being on a line. Or, we just don't know how they lie but decide to find a 'best fit line' in any case. We will develop ideas here with the three points mentioned above. We want the equation of the line they lie on: $y = mx + b$. What is m? What is b? We have to have:

$$\begin{aligned} m + b &= 2 \\ 3m + b &= 4 \\ 4m + b &= 4. \end{aligned}$$

In other words,

$$m \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix} + b \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 4 \end{pmatrix},$$

or

$$\begin{pmatrix} 1 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} m \\ b \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \\ 4 \end{pmatrix}.$$

(a) Verify that no such m and b exist. You can think about this in two ways:
   - The three points do not lie on a line.

- The vector $(2,4,4)$ is not in the span of the vectors $(1,3,4)$ and $(1,1,1)$.

(b) We will now proceed to find the least squares solution. That is, we will find $\begin{pmatrix} \hat{m} \\ \hat{b} \end{pmatrix}$ so that

$$\begin{pmatrix} 1 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} \hat{m} \\ \hat{b} \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix},$$

where $\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$ is the projection of $(2,4,4)$ on to the subspace spanned by the vectors $(1,3,4)$ and $(1,1,1)$. Do this by finding $\hat{m}$ and $\hat{b}$ as on page 14 of the linear algebra reading notes, and then using these values to find the projection vector and the residual vector.

(c) Make a plot of the three original points and the best fit line $y = \hat{m}x + \hat{b}$. Draw vertical line segments from each of the three points to this line. What is the length of each? You should see that it is exactly the value of the corresponding entry in the residual vector; the sign of the entry just tells you if the point is above or below the best fit line.

8. The following table lists the height h, gender g and weight w of some young adults.

height h (inches above 5 feet)	gender g 1='female', 0='male'	weight w (in pounds)
2	1	110
12	0	180
5	1	120
11	1	160
6	0	160

We will find the best fit function of the form

$$w(h,g) = a + bh + cg$$

to this data set.

(a) Before you perform the computations, think about the signs of the constants b and c. What signs would you expect if these data were representative of the general population? Why?

(b) Create the matrix equation that you will use in your least squares analysis.

(c) Calculate the least squares solution using mathematical software.

(d) Create a plot containing the data set and your best fit function.

(e) What is the sign of a? What is the practical significance of a?

9. The following table lists the estimated number g of genes and the estimated number z of cell types for various organisms.

Organism	Number of genes g	Number of cell types z
humans	$600,000$	250
Annelid worms	$200,000$	60
Jellyfish	$60,000$	25
Sponges	$10,000$	12
Yeasts	$2,500$	5

(a) Create the matrix equation you will use to fit a function of the form $\log(z) = a + b\log(g)$ to the data points $(\log(g), \log(z))$.

(b) Use mathematical software to find the least squares solution to part (a).

(c) Create a plot containing the data set $(\log(g), \log(z))$ and your best fit function $\log(z) = a + b\log(g)$.

(d) Use your answer in part (b) to fit a power function $z = kg^r$ to the data points (g, z).

(e) Create a plot containing the data set (g, z) and your best fit function $z = kg^r$.

(f) Using the theory of self-regulator systems, scientists developed a model that predicts z is a square root function of g (that is $z = k\sqrt{g}$ for some constant k). Is your answer in part (c) reasonable close to this form?