Linear Algebra

KNU Math 231

Classnotes

Mark Siggers

v. 2019/06/05

These notes are for a second year class in Linear Algebra based on the fourth edition of Gilbert Strang's 'Linear Algebra and its applications'. This is referred to as The Text. (I will try to keep updating these notes, but I might not be able to for the whole term, so keep taking your own notes, if you do that.)

1 Matrices and Gaussian Elimination

1.1 Introduction

The basic object of study of this course is a system, or set, of linear equations. Here are some examples. Can you solve them? That is, can you find the values x and y, (and z in the last example) that make the equations true?

Example 1.1.

$$\begin{array}{rcrcl} 2x & - & 2y & = & 2 \\ 2x & - & y & = & 5 \end{array}$$

Example 1.2.

$$\begin{array}{rcrcl} 2x & - & 2y & = & 2 \\ x & - & y & = & 4 \end{array}$$

Example 1.3.

$$\begin{array}{rcrcl} 2x & - & 2y & = & 2 \\ x & - & y & = & 1 \end{array}$$

Example 1.4.

$$\begin{array}{rcrcrcl} x & + & 2y & - & 3z & = & 3 \\ x & - & 2y & + & 5z & = & 7 \\ x & + & y & + & 7z & = & 9 \end{array}$$

In this course we look at

- deciding if a system of equations has a solution,
- solving systems of equations,
- solving them quickly,
- approximating solutions (more quickly), and
- interpreting solutions geometrically.

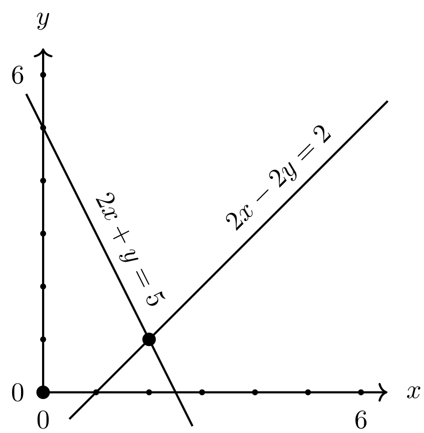
In all of these tasks, it will be convenient to represent systems of linear equations with matrices.

1.2 The Geometry of Linear Equations

There are two basis ways to interpret a system of linear equations geometrically.

The row picture

Each row is an equation that we can graph. The rows of the system in Example 1.1 can be graphed as follows. The solution $(x, y) = (2, 1)$ is the intersection of the two lines.



Practice

Draw similar pictures for Examples 1.2 and 1.3. What do 'no solutions' or 'many solutions' correspond to under this point of view? How is the picture different for Example 1.4?

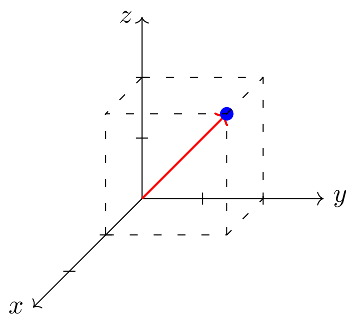
The column picture

For the column view we need some notation. An *n-dimensional vector* (over $\mathbb{R}$ is an ordered set of n real numbers: $(1, 2, 1/2)$ and $(1/2, 2, 1)$ are different 3-dimensional vectors. They are *row vectors* but we could write them as column vectors:

$$\begin{bmatrix} 1 \\ 2 \\ \frac{1}{2} \end{bmatrix} \text{ and } \begin{bmatrix} \frac{1}{2} \\ 2 \\ 1 \end{bmatrix}.$$

Any n-dimensional vector can be viewed as a point in $\mathbb{R}^n$, or as the arrow from the origin to that point.

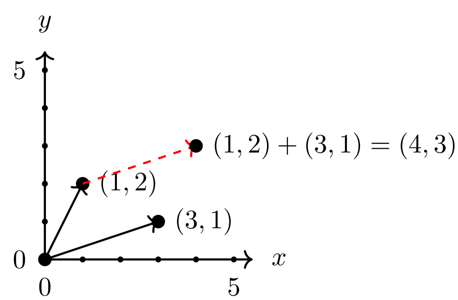
Example 1.5. The vector $(1, 2, 2)$ can be viewed either as the point or the arrow in the following picture.



We add vectors componentwise:

$$\begin{bmatrix}1\\2\end{bmatrix} + \begin{bmatrix}3\\1\end{bmatrix} = \begin{bmatrix}4\\3\end{bmatrix}$$

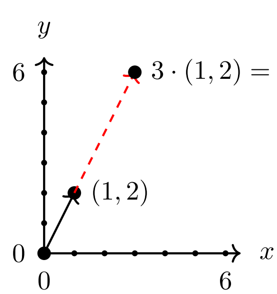
which corresponds to concatenating arrows 'head-to-tail'



and we can multiply them by a scalar:

$$3\begin{bmatrix}1\\2\end{bmatrix} = \begin{bmatrix}3\\6\end{bmatrix}$$

which corredsponds to stretching or shrinking or flipping them.



The system of equations from Example 1.1

$$\begin{array}{ccccc} 2x & - & 2y & = & 2 \\ 2x & - & y & = & 5 \end{array}$$

can be represented as an equation of column vectors

$$x\begin{bmatrix}2\\2\end{bmatrix} + y\begin{bmatrix}-2\\1\end{bmatrix} = \begin{bmatrix}2\\5\end{bmatrix}.$$

Practice

Try to intepret a solution to this equation of column vectors in terms of the column picture.

Problems from the text

Sect 1.2: 2,5,9,16 (Non International Student Edition: 4,3,6,16)

1.3 Gaussian Elimination

Let's now find the solutions to the system of equations in Example 1.4:

$$\begin{array}{ccccccc} x & + & 2y & - & 3z & = & 3 \\ x & - & 2y & + & 5z & = & 7 \\ x & + & y & + & 7z & = & 9. \end{array}$$

We do this by a series of *elementary row operations:*

- $R_i + cR_j$ – Add c times the j^{th} row to the i^{th} row.
- cR_i – Multiply the i^{th} row by $c \neq 0$.

It is clear that by doing either of these row operations, the solutions of the system of equation do not change. Our goal is to get to a system that looks like the following.

$$\begin{array}{ccccc} x & & & = & 4\frac{3}{8} \\ & y & & = & \frac{2}{8} \\ & & z & = & \frac{5}{8} \end{array}$$

Where the coefficient of the i^{th} variable in the i^{th} equation is called the i^{th} *pivot*, *Gaussian Elimination* is using row operations to do this as follows.

i. For $i = 1$ to $n-1$, change the i^{th} pivot to 1 and then clear all entries below it.

ii. For $i = n$ to 2, clear all entries above the i^{th} pivot.

The first step is called *elimination*, it *triangulises* the system:

$$\begin{bmatrix} x + 2y - 3z = 3 \\ x - 2y + 5z = 7 \\ x + y + 7z = 9 \end{bmatrix} \xrightarrow{R_2 - R_1} \begin{bmatrix} x + 2y - 3z = 3 \\ - 4y + 8z = 4 \\ x + y + 7z = 9 \end{bmatrix} \xrightarrow{-R_2/4} \begin{bmatrix} x + 2y - 3z = 3 \\ y - 2z = -1 \\ x + y + 7z = 9 \end{bmatrix}$$
$$\to \cdots \to \begin{bmatrix} x + 2y - 3z = 3 \\ y - 2z = -1 \\ z = 5/8 \end{bmatrix}$$

The second step is called *back substitiution*, it *diagonalises* the system.

$$\begin{bmatrix} x + 2y - 3z = 3 \\ y - 2z = -1 \\ z = \frac{5}{8} \end{bmatrix} \to \begin{bmatrix} x + 2y - 3z = 3 \\ y = \frac{2}{8} \\ z = \frac{5}{8} \end{bmatrix} \to \begin{bmatrix} x + 2y = 4\frac{7}{8} \\ y = \frac{2}{8} \\ z = \frac{5}{8} \end{bmatrix} \to \begin{bmatrix} x = 4\frac{3}{8} \\ y = \frac{2}{8} \\ z = \frac{5}{8} \end{bmatrix}$$

Our solution is thus $(x, y, z) = \frac{1}{8}(35, 2, 5)$.

This is a heck of a lot of writing. Let's use some shorthand. We almost did it above. Lets just remove the variables and replace the equals signs with a big line. Writing

$$\begin{array}{ccccccc} x & + & 2y & - & 3z & = & 3 \\ x & - & 2y & + & 5z & = & 7 \\ x & + & y & + & 7z & = & 9 \end{array} \quad \text{as} \quad \left[\begin{array}{ccc|c} 1 & 2 & 3 & 3 \\ 1 & -2 & 5 & 7 \\ 1 & 1 & 7 & 9 \end{array}\right]$$

elimination becomes

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 3 \\ 1 & -2 & 5 & 7 \\ 1 & 1 & 7 & 9 \end{array}\right] \xrightarrow{R_2 - R_1} \left[\begin{array}{ccc|c} 1 & 2 & 3 & 3 \\ 0 & -4 & 8 & 4 \\ 1 & 1 & 7 & 9 \end{array}\right] \to \ldots.$$

Practice

Solve the following systems. You might encounter some problems. How should you deal with them.

Example 1.6.

$$\begin{array}{ccccccc} 2u & + & v & + & w & = & 5 \\ 4u & - & 6v & & & = & -2 \\ -2u & + & 7v & + & 2w & = & 9 \end{array}$$

Example 1.7.

$$\begin{array}{ccccccc} x & + & 2y & + & 2z & = & 4 \\ 2x & + & 4y & + & z & = & 3 \\ x & + & 3y & + & z & = & 1 \end{array}$$

Example 1.8.

$$\begin{array}{rcrcrcl} x & + & 2y & + & 2z & = & 4 \\ x & + & 2y & + & 3z & = & 4 \\ 2x & + & 4y & + & z & = & 2 \end{array}$$

Example 1.9.

$$\begin{array}{rcrcrcl} x & + & 2y & + & 2z & = & 4 \\ x & + & 3y & + & 3z & = & 4 \\ & & y & + & z & = & 0 \end{array}$$

Problems from the text

1.3: 1, 11, 12, 15, 22 (5, 12, 10, 18, 20)

1.4 Matrices

We 'kind of' used a matrix in the last section, but now we define it properly.

An m by n matrix is a list of $m \times n$ numbers arranged in m rows of n numbers each. For example,

$$\begin{bmatrix} 1 & 2 & -2 & 4 \\ 1 & 1/3 & 3 & \pi \\ \sin(1/2) & 1 & 1 & 0 \end{bmatrix}$$

is a 3×4 matrix.

Notation

Given an $m \times n$ matrix A we write $A = [a_{ij}]$ to mean that a_{ij} is the entry in the i^{th} row and the j^{th} column. So $a_{12} = 3$ if

$$[a_{ij}] = \begin{bmatrix} 0 & 3 & 1 \\ 4 & 2 & 5 \end{bmatrix}.$$

Some Special Kinds of Matrices

The pair 'm by n' is the *dimension* of the matrix. A $1 \times n$ matrix is a *column matrix*, and an $m \times 1$ matrix is a *row matrix*. An $n \times n$ matrix is a *square*. If a square matrix $A = [a_{ij}]$ has $a_{ij} = 0$ whenever $i < j$ then it is *upper-triangular* and if $a_{ij} = 0$ whenever $i > j$, it is lower-triangular. It it is upper and lower triangular, then it is diagonal.

upper-triangular: $\begin{bmatrix} 1 & 2 & 4 \\ 0 & 3 & 1 \\ 0 & 0 & 2 \end{bmatrix}$ lower-triangular: $\begin{bmatrix} 1 & 0 & 0 \\ 3 & 3 & 0 \\ 4 & 7 & 2 \end{bmatrix}$ diagonal: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$

The diagonal matrix with ones all along the diagonal is the *identity:*

$$I = I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Notationally, matrices will be denoted by capital letters such as A, B and M except that the row and column vectors are often denoted by small bold letters such as $\mathbf{b}$, $\mathbf{c}$ and $\mathbf{x}$.

Addition of Matrices

We add matrices of the same dimensions by adding entrywise:

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

Scalar Multiplication

Multiplying a matrix A by a scalar r is in the obvious way: $r[a_{ij}] = [ra_{ij}]$. For example

$$2 \cdot \begin{bmatrix} 1 & 0 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 4 & 6 \end{bmatrix}.$$

Matrix Multiplication

This is the more complicated operation. Given an $\ell \times m$ matrix $A = [a_{ij}]$ and an $m \times n$ matrix $B = [b_{ij}]$, the *product matrix* AB is the $\ell \times n$ matrix $C = [c_{ij}]$ where

$$c_{ij} = \sum_{\alpha=1}^{m} a_{i\alpha} \cdot b_{\alpha j}.$$

For example we can take the product of a row and a column vector of the same length:

$$\begin{bmatrix} 1 & 4 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 1/2 \end{bmatrix} = [1 \cdot 3 + 4 \cdot 1 + 2 \cdot 1/2] = [8].$$

Note:

This is also known as the *dot product* of the two vectors.

Practice

Evaluate

$$\begin{bmatrix}1 & 7\\2 & 8\end{bmatrix}\begin{bmatrix}1 & 4\\2 & 3\end{bmatrix} \text{ and } \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}1 & 4 & 3\\2 & 3 & 1\end{bmatrix}$$

Notice that $AB = BA$ is not generally true, and that the identity is the multiplicative identity.

There are three useful ways to view the product. The first is the *dot product view*. The ij^{th} entry of the product AB is the i^{th} row of A dotted with the j^{th} column of B.

$$\begin{bmatrix} \cdot & \cdot & \cdot \\ \boxed{1 \quad 4 \quad 2} \\ \cdot & \cdot & \cdot \end{bmatrix}\begin{bmatrix} \cdot & \boxed{\begin{matrix}3\\1\\1/2\end{matrix}} & \begin{matrix}\cdot & \cdot\\ \cdot & \cdot\\ \cdot & \cdot\end{matrix} \end{bmatrix} = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & 8 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

The second is the *row view* of AB. We view the rows of A as providing the co-efficeints in linear combinations of the rows of B:

$$\begin{bmatrix}1 & 4\\2 & 3\end{bmatrix}\begin{bmatrix}1 & 7\\2 & 8\end{bmatrix} = \begin{bmatrix}1\cdot\begin{bmatrix}3 & 7\end{bmatrix} + 4\cdot\begin{bmatrix}2 & 8\end{bmatrix}\\ 2\cdot\begin{bmatrix}3 & 7\end{bmatrix} + 3\cdot\begin{bmatrix}2 & 8\end{bmatrix}\end{bmatrix} = \begin{bmatrix}11 & 39\\12 & 38\end{bmatrix}$$

This seems silly perhaps. But we will use it this section in 'encoding' row operations for Gaussian elimination as matrix multiplications.

The third is the *column view* of AB where we veiw the columns of B a providing co-efficients in linear combinations of the columns of A.

$$\begin{bmatrix}1 & 4\\2 & 3\end{bmatrix}\begin{bmatrix}1 & 7\\2 & 8\end{bmatrix} = \begin{bmatrix} 3 & 7 \\ \times & \times \\ \begin{bmatrix}1\\2\end{bmatrix} & \begin{bmatrix}1\\2\end{bmatrix} \\ + & + \\ 2 & 8 \\ \times & \times \\ \begin{bmatrix}4\\3\end{bmatrix} & \begin{bmatrix}4\\3\end{bmatrix} \end{bmatrix} = \begin{bmatrix}11 & 39\\12 & 38\end{bmatrix}$$

Basic properties of matrix multiplication

We have seen that the identity matrix I is the multiplicative identity, and that matrix multiplication is not generally commuative (ie. $AB = BA$ does not generally hold.) Here are a couple more basic properties.

Fact 1.10. *For any matrices A, B and C of the proper dimensions, the following hold.*

- ***Associativity:*** $A(BC) = A(BC)$
- ***Distributivity:*** $A(B + C) = AB + AC$ *and* $(A + B)C = AC + BC$

As we have an identity matrix, one might ask if there is an inverse? Given a matrix A is there a matrix B such that $AB = I$ or $BA = I$.

Systems of Equations as Matrix equations

Consider the following system of equation.

$$\begin{array}{rcrcr} x & + & 4y & = & 7 \\ 2x & - & y & = & 3 \end{array}$$

To solve it, our first step in Gaussian elimination is

$$\left[\begin{array}{cc|c} 1 & 4 & 7 \\ 2 & -1 & 3 \end{array}\right] \xrightarrow{R_2 - 2R_1} \left[\begin{array}{cc|c} 1 & 4 & 7 \\ 0 & -9 & -11 \end{array}\right]$$

taking 2 times row one from row two. This was shorthand. One can properly write the original system now as

$$\begin{bmatrix} 1 & 4 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \end{bmatrix}.$$

Thinking about the row view of multiplication, the row operation $R_2 - 2R_1$ can by viewed as multiplying on the left by the matrix $\left[\begin{smallmatrix} 1 & 0 \\ -2 & 1 \end{smallmatrix}\right]$:

$$\begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 7 \\ 3 \end{bmatrix}.$$

Multiplying this out yields

$$\begin{bmatrix} 1 & 4 \\ 0 & -9 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 7 \\ -11 \end{bmatrix}$$

which is the system of equations we have after the row operation.

Our next row operations would be $R_2/-9$ and $R_1 - 4R_2$.

Practice

What matrices to we have to multiply the equation by (on the left) to do these row operations?

The matrices corresponding to elementary row operations are called *elementary matrices*. Except for one entry, they will be 1 on the diagonal, and 0 everywhere else. Those used for elimination will have 0 on the upper triangle, and those used for back-substitution will have 0 on the lower triangle.

The full Gaussian Elimination

$$\left[\begin{array}{cc|c} 1 & 4 & 7 \\ 2 & -1 & 3 \end{array}\right] \to \cdots \to \left[\begin{array}{cc|c} 1 & 0 & 19/9 \\ 0 & 1 & 11/9 \end{array}\right]$$

can by achieved by multiplying on the left by the matrix:

$$B = \underbrace{\begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix}}_{R_1 - 4R_2} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & -1/9 \end{bmatrix}}_{-R_2/9} \underbrace{\begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}}_{R_2 - 2R_1} = \frac{1}{9}\begin{bmatrix} 5 & 4 \\ 1 & -1 \end{bmatrix}.$$

That is:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = B\begin{bmatrix} 1 & 4 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = B\begin{bmatrix} 7 \\ 3 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 19 \\ 11 \end{bmatrix}.$$

The matrix B was constructed from elementary matrices by Gaussian Elimination so that it has the property that $B\left[\begin{smallmatrix} 1 & 4 \\ 2 & -1 \end{smallmatrix}\right] = I$. It is an inverse of $\left[\begin{smallmatrix} 1 & 4 \\ 2 & -1 \end{smallmatrix}\right]$. This will be a useful idea.

Problems from the text

1.4: 2, 4, 7, 11, 13, 19, 26, 43, 44, 45 (3, 7, 2, 10, 17, 21, 22, 40, 41, 46)

1.6 Inverses and Transposes

Recall that the identity I had the nice property that for any square matrix A of the same dimension $IA = A = AI$. We asked, given A about a matrix B such that $AB = I$ or $BA = I$. Well, if there are B and B' such that $BA = I$ and $AB' = I$ then they must be the same:

$$B = BI = BAB' = IB' = B'.$$

Definition 1.11. For an $n \times n$ matrix A, an *inverse* A^{-1} is a matrix such that

$$A^{-1}A = I = AA^{-1}.$$

If A has an inverse, it is *invertible* or *non-singular*.

Practice

Show that not all square matrices A need have a inverse. Show also that if a square matrix A has an inverse, then it is unique.

Observe that if two $n \times n$ matrices A and B both have inverses, then so does their product, and $(AB)^{-1} = B^{-1}A^{-1}$. Indeed

$$(AB)B^{-1}A^{-1} = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I.$$

Practice

Thinking of how to reverse the row operations that it represents, invert the following upper triangular matrix $\begin{bmatrix} 1 & 3 & 5 \\ 0 & 2 & -3 \\ 0 & 0 & 1 \end{bmatrix}$.

Note

For an $n \times n$ matrix A, we will see that if there is any B such that $BA = I$, then we also have that $AB = I$, and so B is the inverse A^{-1}. In Section 1.4 we saw that if Gaussian elimination can be used to reduce A to I, then it is invertible, and we found its inverse. We will see if A is invertible then we will always be able to find it this way.

We now find the inverse of another matrix, but in a much neater way, by a process known as Gauss-Jordan elimination.

Where $A = \begin{bmatrix} 1 & 4 \\ 2 & -1 \end{bmatrix}$ we constuct the augmented matrix

$$[A \mid I] = \left[\begin{array}{cc|cc} 1 & 4 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{array}\right].$$

This is really just an $n \times 2n$ matrix, but we've added a cosmetic bar down the middle. The idea is that by multiplying by A^{-1} we get $A^{-1}[A \mid I] = [I \mid A^{-1}]$. Multiplying by A^{-1} is really just doing row operations, so if we do these row operations to the augmented matrix, we get A^{-1} in the second part.

Let's try:

$$\left[\begin{array}{cc|cc} 1 & 4 & 1 & 0 \\ 2 & -1 & 0 & 1 \end{array}\right] \to \left[\begin{array}{cc|cc} 1 & 4 & 1 & 0 \\ 0 & -9 & -2 & 1 \end{array}\right] \to \left[\begin{array}{cc|cc} 1 & 4 & 1 & 0 \\ 0 & 1 & 2/9 & -1/9 \end{array}\right] \to \left[\begin{array}{cc|cc} 1 & 0 & 1/9 & 4/9 \\ 0 & 1 & 2/9 & -1/9 \end{array}\right].$$

So $A^{-1} = \frac{1}{9}\begin{bmatrix} 1 & 4 \\ 2 & -1 \end{bmatrix}$.

Practice

Find the inverse of $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 2 & 2 & 3 \end{bmatrix}$.

Now, we could go almost exactly the same reduction on the augmented matrix

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 2 & 3 & 4 & 3 \\ 2 & 2 & 3 & 4 \end{array}\right]$$

to solve the system of linear equations

$$\begin{array}{ccccccc} x & + & y & + & z & = & 2 \\ 2x & + & 3y & + & 4z & = & 3, \\ 2x & + & 2y & + & 3z & = & 4 \end{array}$$

or equivalently the matrix equation

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 2 & 2 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}.$$

But we don't have to now that we have A^{-1}. We can use it to 'divide' and so solve the system:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = I \begin{bmatrix} x \\ y \\ z \end{bmatrix} = A^{-1}A \begin{bmatrix} x \\ y \\ z \end{bmatrix} = A^{-1} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}.$$

That is:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = A^{-1} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -3 & -1 & 1 \\ 2 & 1 & -2 \\ -2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -5 \\ -1 \\ 0 \end{bmatrix}.$$

What's really nice about this is that if we have to solve $A\mathbf{x} = \mathbf{a}$ for several different **a**, we only have to do the elimination once. For each **a** we get away with multiplying a matrix by a vector.

Let's discuss the time this takes in a bit more detail.

The computational cost of matrix operations

When working with bigger matrices, the work required to do various operations becomes important. To add two $n \times n$ matrices we have n^2 pairs of numbers that we must add. For the product of a scalar and an $n \times n$ matrix, we must multiply n^2 pairs of numbers. Though for big number multiplication is harder, we will call each of 'add' or 'multiply' or 'add and multiply' a basic operation. We implicilty assume that we are working with $n \times n$ matrices and say that matrix addition and scalar multiplication both take n^2 operations, or n^2 time.

The dot product of two vectors of length n takes n multiplications and n additions, so we say it takes n operations. Multiplying an matrix by a vector requires n dot products, so n^2 operations. Multiplying two $n \times n$ matrics requires n matrix-vector multiplications, so costs n^3 operations.

Note

We can multiply two matrices in n^3 operations. This means that if your computer can multiply two 3×3 matrices in one second, then it takes about 8 seconds to multiply two 6×6 matrices, and in an hour, or 3600 seconds, the best it can do is multiply two 45×45 matrices, (because $\sqrt[3]{3600} \approx 15$).

If we could reduce this to n^2 operations, then as $\sqrt{3600} = 60$, your computer could mutliply two 180×180 matrices in an hour. These differences can be important in real situations.

What does Gaussian Elimination cost? Let's count this for solving an $n \times n$ system.

- In elimination, to clear below the first pivot we must clear $n-1$ rows, and each row add a scalar multple of a vector to another vector, so takes $n+1$ operations. Together takes $(n+1)(n-1) = n^2 - 1$, let's say n^2, operations to clear under the first pivot.
- Clearing under the second pivot takes $(n-1)^2$, under the third $(n-2)^2$ and all together, elimination takes

$$2^2 + \cdots + n^2 = \sum_{i=2}^{n} i^2 \approx n^3/3 \approx n^3$$

operations.
- Back substitution is a bit faster. Clearing over the n^{th} pivot only takes one operation per row, so $n-1$ operations, clearling over the second pivot takes $n-2$ operations, etc, and altogether back substitution takes

$$\sum_{i=1}^{n-1} i \approx n^2.$$

- The whole process takes about $n^3 + n^2$ operations.

Note

Using Gaussian Elimination to find an inverse takes about twice as long, as you have to un-eliminate the augmented matrice I as you are eliminating A.

The Transpose Matrix of A

For a matrix A, the *transpose* A^T is the matrix we get by switching rows and columns. For example:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 2 & 3 \end{bmatrix}^T = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 3 \end{bmatrix}$$

It is not hard to see that

- $(AB)^T = B^T A^T$ and
- $(A^{-1})^T = (A^T)^{-1}$.

A matrix A is *symmetric* if $A^T = A$.

For any matrix A, AA^T is (square and) symmetric. Indeed,

$$(AA^T)^T = (A^T)^T A^T = AA^T.$$

Similarily $A^T A$ is symmetric.

Practice

Find an example to show that $A^T A = AA^T$ does not generally hold.

Problems from the text

1.6: 1, 2, 4, 6, 16, 21, 25, 27, 39, 49, 56 (9, 5, 10, 4, 13, 25, 28, 24, 38, 51, 60)

1.5 Triangular Factors and Row Exchanges

Note:

We always use $\mathbf{x}$ for the column vector $\begin{bmatrix} x \\ y \end{bmatrix}$ or $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$, etc. of n variables.

We observed that if we compute A^{-1} at a cost of $2n^3$ operations, then we can solve $A\mathbf{x} = \mathbf{b}$ at a cost of n^2 operations for each vector $\mathbf{b}$ that we want to solve it for. But actually we don't really have to completely find A^{-1} to make these savings, we can stop halfway.

Doing elimination to $[A \mid I]$ we get to $[I \mid A^{-1}]$. But stopping before back-substitution, we get, in half the time, a matrix $[U \mid L^{-1}]$ where U is upper-triangular and L is lower triangular.

This means that $L^{-1}A = U$, and applied to the system $A\mathbf{x} = \mathbf{B}$ gives

$$U\mathbf{x} = L^{-1}A\mathbf{x} = L^{-1}\mathbf{b} =: \mathbf{c}.$$

So we can solve for $\mathbf{x}$ by (i) letting $\mathbf{c} = L^{-1}\mathbf{b}$, and (ii) solving $U\mathbf{x} = \mathbf{c}$. Why is this good? Well, computing $L^{-1}\mathbf{b}$ is a matrix-vector multiplication, so takes n^2 operations (and actually only half of this as L^{-1} is triangular), and U is already eliminated, so solving $U\mathbf{x} = \mathbf{c}$ is just back-substitiution, so takes $n^2/2$ operations. This is the same as computing $A^{-1}\mathbf{b}$, but we do so without having to compute A^{-1}. We save half the time on our most expensive step.

Example 1.12. We eliminate the augmented matrix $[A \mid I]$ below, and save time by not worrying about making the pivots 1:

> Note:
>
> It is important that we only use elementary row operations here, not row swaps, if we use row sways, then U may not be upper triangular.

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 5 & 8 & 0 & 1 & 0 \\ 2 & 1 & 2 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & -3 & -4 & -2 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & -2 & 1 & 0 \\ 0 & 0 & 2 & -8 & 3 & 1 \end{array}\right] =: [U \mid L^{-1}]$$

Now to solve, say, $A\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$ we compute

$$\mathbf{c} = L^{-1}\mathbf{b} = \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -8 & 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix}$$

and then solve

$$\left[\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 2 & -3 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 2 & 0 & 11/2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -3/2 \end{array}\right] \to \left[\begin{array}{ccc|c} 1 & 0 & 0 & -1/2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -3/2 \end{array}\right]$$

to get $\mathbf{x} = (-1, 6, -3)/2$.

The LU-factorisation

Apart from the use we just say, this partial elimination $[A \mid I] \to [U \mid L^{-1}]$ will have other uses. As $L^{-1}A = U$ we can write A as

$$A = LU$$

where L is lower-triangular and U is upper-triangular.

This LU-factorisation of a matrix A, if it exists, need not be unique:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 2 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 2/3 & 8 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 2/3 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 3 & 6 \\ 0 & 0 & 2 \end{bmatrix}$$

This suggests an LDU-factorisation, in which D is diagonal, and L and U have only 1s on the diagonal. This is unique, if it exists, and is also used later.

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 4 & 2/3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

Practice

Show that if A is symmetric and has an LDU-factorisation, then it is of the form $A = LDL^T$.

Permutation Matrices

We saw that Gaussian Elimination can fail for some matrices A, we also saw that we might need to use 'row swaps', and we said that there are not 'elementary row operations'. The matrice that we can use to multiply on the left to give a row swap is called a permutation matrix. A *permutation matrix* is an matrix that we can get from the identity by doing row swaps.

Practice

Which of the following matrices are permutation matrices?

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

Practice

How many $n \times n$ permutation matrices are there?

When we are doing elimination, if we have to do a row swap, we can pretend that we did it before we started the elimination. (The other row operations might change, but we will still be able to do them.) So if we can eliminate A to I using row swaps, then we can eliminate PA to I without using row swaps, for some permutation matrix P.

Thus if A is invertable, then there are U, L and P such that

$$PA = LU.$$

Practice

Show, for a permutation matrix P, that $P^{-1} = P^T$.

Problems from the text

1.5: 1, 2, 7, 10, 15, 18 (3, 1, 5, 6, 12, 11)

2 Vector Spaces

2.1 Vector Spaces and Subspaces

Recall that we called an n-tuple $(r_1, r_2, \dots, r_n)$ of elements of $\mathbb{R}^4$ a ‘vector’. The reason we use the word vector is that the set $\mathbb{R}^4$, with scalar multiplication and componentwise addition, satisfies the following definition of a vactor space.

Definition 2.1. A set X with addition '+' and scalar multiplication $\cdot : \mathbb{R} \times X \to X$ is a *(real) vector space* if it satisfies the following properties for all $x, y, z \in X$ and $q, r \in \mathbb{R}$.

a) $x + y = y + x$ for all $x, y \in X$

b) $x + (y + z) = (x + y) + z$ for all $x, y, z \in X$

c) There is an element $0 \in X$ such that $0 + x = x$ for all $x \in X$. This element is called zero.

d) For every $x \in X$ there is a element $-x \in X$ such that $x + (-x) = 0$.

e) $1 \cdot x = x$ for all $x \in X$.

f) $(qr)x = q(rx)$ for $x \in X$ and $q, r \in \mathbb{R}$.

g) $r(x + y) = rx + ry$ for all $x, y \in X$ and $r \in \mathbb{R}$.

h) $(q + r)x = qx + rx$ for all $x \in X$ and $q, r \in \mathbb{R}$.

Elements of a vector space X are called *vectors*.

Practice

Verify $\mathbb{R}^n$ is a vector space.

As we said, $\mathbb{R}^n$ is a vector space, and it is the one we talk about the most. But there are other real vector spaces:

Example 2.2. The following are vector spaces.

- $\emptyset$
- $\mathbb{R}^\infty$, the space of all infinite tuples: $(1, 2, 1, 1, 3, \dots,)$. (Addition and scalar multiplication are again componentwise.)
- The space of 2×2 real matrices. (This is actually isomorphic as a space to $\mathbb{R}^4$. Matrix multiplication is not part of the space, it will define operations on the space.)

- The space of functions (like $f : x \mapsto 2x$, say) on a set $D \subset \mathbb{R}$. Addition is defined by $[f+g](x) = f(x)+g(x)$ and scalar multiplication by $[r \cdot f](x) = r \cdot f(x)$.

Note

An *isomorphism* of vector spaces U and V is a bijective function $f : U \to V$ such that

- $f(a+b) = f(a) + f(b)$, and
- $f(ra) = rf(a)$.

If there is an isomorphism between two spaces, they are *isomorphic.*

Subspaces

A *subspace* of a vector space V is a subset U of V that is *closed under addition and scalar multiplication:*

- $x, y \in U \Rightarrow x + y \in U$, and
- $x \in U$, $r \in \mathbb{R} \Rightarrow rx \in U$.

A subspace of a vector space is itself a vector space.

Practice

In $\mathbb{R}^2$, is the subset $U = \{(1, 1), (1, -1)\}$ a subspace?

Example 2.3. Some subspaces of $\mathbb{R}^2$ are

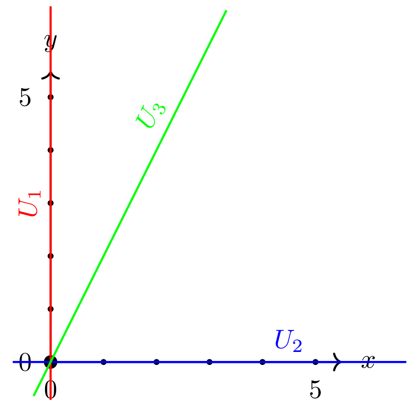
- $\emptyset$, $\mathbb{R}^2$.
- $U_1 = \{(r, 0) \mid r \in \mathbb{R}\}$
- $U_2 = \{(0, r) \mid r \in \mathbb{R}\}$
- $U_3 = \{(r, 2r) \mid r \in \mathbb{R}\}$

Practice

Show that $U_3 = \{(r, 2r) \mid r \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$.

These subspaces are sets, so we can graph them.

ver. 2019/06/05



In general in $\mathbb{R}^2$ if a subspace contains a vector (a, b) then it must contain all scalar multiples $r(a, b)$ of (a, b) so must contain the line through $0(a, b) = (0, 0)$ and (a, b).

As soon as it also contains some (a', b') not on this line, what does it contain?

The only subspace of $\mathbb{R}^2$ are $\emptyset, \mathbb{R}^2$ and lines through $(0, 0)$. A line $\{r(a, b) \mid r \in \mathbb{R}\}$ through $\mathbf{0}$ is isomorphic to $\mathbb{R}$ via the isomorphism

$$r \mapsto r(a, b).$$

So though $\mathbb{R}^1$ is not a subspace of R^2 it is isomorphic to one. The subspaces of $\mathbb{R}^2$ are, **up to isomorphism**, $\mathbb{R}^0, \mathbb{R}^1$, or $\mathbb{R}^2$.

Practice

Up to isomorphism, what are the subspaces of $\mathbb{R}^n$?

2.1.1 The subspace generated by a set

We saw that if (a, b) is in a subspace U of $\mathbb{R}^2$ then the whole line $\{r(a, b) \mid r \in \mathbb{R}\}$ must be in U.

For a subset S of vectors in a vector space V, the *subspace generated by* S is the smallest subspace of V that contains S.

Practice

Show that if U and U' are two subspaces of V then so is their intesection $U \cap U'$. Use this to assert that for any subsets S of V, the subspace generated by S is uniquely defined.

Example 2.4. The subspace of $\mathbb{R}^4$ generated by $\{(0, 1, 2, 1)\}$ is the line

$$\{r(0, 1, 2, 3) \mid r \in \mathbb{R}\}.$$

The subspace of $\mathbb{R}^3$ generated by $\{(0,1,2),(1,1,0)\}$ is the plane

$$\{a(0,1,2)+b(1,1,0) \mid a,b\in\mathbb{R}\};$$

it is isomorphic to $\mathbb{R}^2$. (You can show that $(a,b)\mapsto a(0,1,2)+b(1,1,0)$ is an isomorphism.)

We will see that every subspace of $\mathbb{R}^n$ is isomorphic to $\mathbb{R}^m$ for some non-negative $m\le n$. This m is the *dimension* of the subspace.

The Column Space of a Matrix

As the columns of an $m\times n$ matrix M are vectors in $\mathbb{R}^m$, the generate a subspace of $\mathbb{R}^m$. This is called the the *column space* $C(M)$ of M.

Example 2.5. The column space of

$$\begin{bmatrix}0&4\\1&2\\5&3\end{bmatrix} \text{ is } \left\{u\begin{bmatrix}0\\1\\5\end{bmatrix}+v\begin{bmatrix}4\\2\\3\end{bmatrix} \mid u,v\in\mathbb{R}\right\}.$$

The matrix equation

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

has a solution if and only if $\begin{bmatrix}7\\2\\1\end{bmatrix}$ is in this space.

> **Note**
>
> Given M, the set of vectors $\mathbf{b}$ for which $M\mathbf{x}=\mathbf{b}$ has a solution, is a vector space.

The Nullspace of a Matrix

As the solutions of

$$\begin{bmatrix}1&2\\1&3\\1&7\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}$$

are vectors in $\mathbb{R}^2$, the solutions of $M\mathbf{x}=\mathbf{0}$ for an $m\times n$ matrix are vectors in $\mathbb{R}^n$. Observing that if

$$M\mathbf{x}=\mathbf{0} \text{ and } M\mathbf{x}'=\mathbf{0}$$

then

$$M\mathbf{x}+\mathbf{x}'=M\mathbf{x}+M\mathbf{x}'=\mathbf{0}+\mathbf{0}=\mathbf{0},$$

> Note:
>
> We let $\mathbf{0}$ denote the appropriate vector or matrix of all '0's.

and

$$M(r\mathbf{x}) = (M\mathbf{x})r = \mathbf{0}r = \mathbf{0},$$

we see that the set of solutions to $M\mathbf{x} = \mathbf{0}$ is a subspace of $\mathbb{R}^n$. This space is called the *nullspace* $N(M)$ of M.

Note:

Think of the nullspace of M as the set of vectors that M nullifies, or 'takes to zero'.

Problems from the text

2.1: 2, 3, 5, 6, 8, 13, 17, 25, 30 (2, 4, 3, 5, 8, 12, 14, 26, 29)

2.2 Solving $M\mathbf{x} = \mathbf{0}$ and $M\mathbf{x} = \mathbf{b}$

When a matrix M has an inverse there is a unique solution $M^{-1}\mathbf{b}$ to $M\mathbf{x} = \mathbf{b}$, and we know how to solve this.

But matricies are not always invertible. A matrix that has no inverse is *singular* (so invertible matrices are also called non-singular). For a singular matrices M, the equation $M\mathbf{x} = \mathbf{b}$ may have no solutions, or many solutions. We want to find all of them.

Though the set of solutions of $M\mathbf{x} = \mathbf{0}$ is a subspace of $\mathbb{R}^n$, the set of solutions of $M\mathbf{x} = \mathbf{b}$ generally is not. As expected, we find it using Gaussian elimination.

Echelon Form of a matrix

Consider the matrix $M = \left[\begin{smallmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 7 \\ -1 & -3 & 3 & 4 \end{smallmatrix}\right]$. We reduce it using Gaussian Elimination:

$$\begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 7 \\ -1 & -3 & 3 & 4 \end{bmatrix} \rightsquigarrow \underbrace{\begin{bmatrix} 1 & 3 & 3 & 2 \\ 0 & 0 & 3 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\text{Echelon Form}} \rightsquigarrow \underbrace{\begin{bmatrix} 1 & 3 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\text{Row Reduced Echelon Form}}$$

Note:

The symbol $\rightsquigarrow$ hides a lot of row operations.

Echelon Form is as 'close' as we can get to triangular: we try to make the system triangular but we may get unavoidable 0's on the diagonal. The Row Reduced Echelon Form (RREF) is a 'close' as we can get to diagonal. The first non-zero entry in a row is a *pivot*, the column it is in is a *pivot column*. The other columns are *non-pivot* or *free* columns. To find the set of solutions to $M\mathbf{x} = \mathbf{b}$ using the RREF of M the pivot columns and free columns treated differently.

Solving $M\mathbf{x} = \mathbf{0}$

To solve $M\mathbf{x} = \mathbf{0}$ we have reduced to

$$\begin{bmatrix} 1 & 3 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The variables in $\begin{bmatrix} w & x & y & z \end{bmatrix}$ corresponding to pivot and free columns are called *dependent* and *independent (or free)* respectively. So w and y are dependent and x and z are free.

The free variables can be anything, so we leave them as variables: $x = x$ and $z = z$. Once z is fixed, y is determined: we solve $y = -z$. Once x, y and z are fixed, we get $w = z - 3x$. So our solution is

$$\begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = x \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + z \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix},$$

for any choice of $x, z \in \mathbb{R}$.

We have found the nullspace of M. Observe that it is also the column space of the matrix $\begin{bmatrix} -3 & 1 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{bmatrix}$. We call this the *nullspace matrix* of M. The number of pivots in M is an important number.

The *rank* $r \leq n$ of an $m \times n$ matrix M is the number of pivots when it has been reduced to Echelon Form. The number $n - r$ of columns of the nullspace matrix is called the *nullity* of M.

Solving $M\mathbf{x} = \mathbf{b}$

We observed that unlike the set of solutions of $M\mathbf{x} = \mathbf{0}$, the set of solutions of $M\mathbf{x} = \mathbf{b}$ need not be a space. However, if we find one solution $\mathbf{x_0}$ to $M\mathbf{x} = \mathbf{b}$ then $\mathbf{x_0} + \mathbf{z}$ is another for any solution $\mathbf{z}$ of $M\mathbf{x} = \mathbf{0}$. Indeed

$$M(\mathbf{x_0} + \mathbf{z}) = M\mathbf{x_0} + M\mathbf{z} = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

The set of solutions of $M\mathbf{x} = \mathbf{b}$ is called its *complete solution.* We will see that it is

$$S = \{\mathbf{x_0} + z \mid z \in \text{nullspace(M)}\}$$

for any one *particular solution* $\mathbf{x_0}$ of $M\mathbf{x} = \mathbf{b}$.

Example 2.6. We find the complete solution of

$$\begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 7 \\ -1 & -3 & 3 & 4 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 5 \end{bmatrix} =: \mathbf{b}$$

by eliminating the augmented matrix to RREF:

$$\left[\begin{array}{cccc|c} 1 & 3 & 3 & 2 & 1 \\ 2 & 6 & 9 & 7 & 5 \\ -1 & -3 & 3 & 4 & 5 \end{array}\right] \to \left[\begin{array}{cccc|c} 1 & 3 & 0 & -1 & -2 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Again we solve: $z = z$, $y = 1 - z$, $x = x$, and $w = -2 - 3x + z$. So we have complete solution

$$\begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \underbrace{\begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix}}_{\text{particular soln}} + x \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + z \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix}.$$

Problems from the text

2.2: 4, 5, 7, 10, 18, 29, 38, 41, 53, 65 (5, 2, 6, 12, 20, 30, 38, 39, 54, 60)

2.3 Linear Independence

The subspace of $\mathbb{R}^4$ generated by $\{(1,1,1,0),(1,1,0,0),(-1,-1,1,0)\}$ is also generated by $\{(0,0,1,0),(1,1,0,0)\}$. This smaller generating set is more desirable than the big one for obvious reasons. How do we find the smallest generating set of a space?

Definition 2.7. A *linear combination* of vectors $v_1, \dots v_k$ is a sum of the vectors, each multiplied by some scalar:

$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k.$$

A set $\{v_1, \dots, v_k\}$ of vectors is *linearly independent* if no one of them can be written as a linear combination of the others.

Practice

Show that a set $\{v_1, \dots, v_k\}$ of vectors is linearly independent if and only if

$$(c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = 0) \Rightarrow (c_1 = c_2 = \cdots = c_k = 0).$$

Example 2.8. Is $\{(1,1,1),(2,3,4),(0,1,2)\}$ independent?

No! $(2,3,4)-(1,1,1)=(0,1,2)$.

Is $\{(1,1,1),(2,3,4)\}$ independent?

Yes! If $a(1,1,1)+b(2,3,4)=0$ then

$$\begin{bmatrix}1 & 2\\ 1 & 3\\ 1 & 4\end{bmatrix}\begin{bmatrix}a\\ b\end{bmatrix}=\begin{bmatrix}0\\ 0\end{bmatrix}$$

which we can quickly show yields $(a,b)=(0,0)$.

We can generalise this to the following assertion:

Fact 2.9. *A set of vectors is linearly independent if and only if the matrix M of which they are columns yields a unique solution to $M\mathbf{x}=\mathbf{0}$; that is, if and only if M is non-singular.*

Example 2.10. Are the vectors $(1,1,2),(3,2,2)$, and $(2,-1,4)$ linearly independent?

Well, eliminating, we get

$$\begin{bmatrix}1 & 3 & 2\\ 1 & 2 & -1\\ 2 & 2 & 4\end{bmatrix}\rightsquigarrow\begin{bmatrix}1 & 3 & 2\\ 0 & 1 & 3\\ 0 & 0 & 1\end{bmatrix},$$

so yes.

Example 2.11. What if we include $(7,0,-8)$ also?

No! A 3×4 matrix cannot by non-singular, it necessarily has a free column.

Example 2.12. OK. Then we should be able to express $\mathbf{b}=(7,0,-8)$ as a linear combination of the other three vectors. How do we do this?

One method: We could solve $M\mathbf{x}=\mathbf{b}$.

$$\left[\begin{array}{ccc|c}1 & 3 & 2 & 7\\ 1 & 2 & -1 & 0\\ 2 & 2 & 4 & -8\end{array}\right]\rightsquigarrow\left[\begin{array}{ccc|c}1 & 0 & 0 & -10.5\\ 0 & 1 & 0 & 5.5\\ 0 & 0 & 1 & .5\end{array}\right]$$

So $\left[\begin{smallmatrix}7\\ 0\\ -8\end{smallmatrix}\right]=-10.5\left[\begin{smallmatrix}1\\ 1\\ 2\end{smallmatrix}\right]+5.5\left[\begin{smallmatrix}3\\ 2\\ 2\end{smallmatrix}\right]+.5\left[\begin{smallmatrix}2\\ -1\\ 4\end{smallmatrix}\right]$.

Another method: Put the vectors as rows and eliminate the following augmented matrix.

$$\left[\begin{array}{ccc|cccc} 1 & 1 & 2 & 1 & 0 & 0 & 0 \\ 3 & 2 & 2 & 0 & 1 & 0 & 0 \\ 2 & -1 & 4 & 0 & 0 & 1 & 0 \\ 7 & 0 & -8 & 0 & 0 & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|cccc} 1 & 1 & 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 4 & 3 & -1 & 0 & 0 \\ 0 & 0 & 12 & 7 & -3 & 1 & 0 \\ 0 & 0 & 0 & 10.5 & -5.5 & -.5 & 1 \end{array}\right]$$

Hmm.

Practice

How does this give us the answer?

Note

A nice thing to notice about this second method is that all the reducef rows are in space generated by the original rows. They seem to give us a nicer generating set of the same space.

Definition 2.13. A set of vectors S *span* a vector space V if V is the space they generate. The set S is a *basis* for V if it spans V and is linearly independent.

Example 2.14. The set $\{e_1 = (1,0,0), e_2 = (0,1,0), e_3 = (0,0,1)\}$ is a basis for $\mathbb{R}^3$. Indeed, it is called the *standard basis.*

Practice

Find another basis for $\mathbb{R}^3$.

Practice

Find one containing $(10, 10, -2)$.

Fact 2.15. *Given a basis B of a vector space V, a vector v in V has a unique expression as a linear combination of vectors in B.*

Practice

Why is this true?

Note:

'Basis' is singular, 'bases' is plural.

Though a space can have different bases, they must have the same size.

Theorem 2.16. *Different bases of the same vector space have the same number of vectors.*

Proof. Assume that $\{b_1, \ldots b_m\}$ and $\{a_1, \ldots, a_n\}$ are two different bases of a vector space $V \subset \mathbb{R}^d$. As every a_i is a linear combination of the b_i there is an $m \times n$ matrix M such that

$$\underbrace{[b_1|b_2|\ldots|b_m]}_{B} M = \underbrace{[a_1|a_2|\ldots|a_n]}_{A}.$$

If $n > m$ then $M\mathbf{x} = \mathbf{0}$ has a non-zero solution $\mathbf{x_0}$ and so

$$A\mathbf{x_0} = BM\mathbf{x_0} = B\mathbf{0} = \mathbf{0}$$

is a linear combination of the a_i that equals 0. This is impossible as $\{a_1, \ldots, a_n\}$ is a basis, and so $n \leq m$. Similarily we can show that $m \leq n$. ☐

Definition 2.17. The number of vectors in a basis of a vector space is its *dimension*.

Problems from the text

2.3: 3, 4, 7, 9, 10, 18, 21, 30, 34, 36 (5, 1, 2, 10, 8, 18, 26, 33, 30, 29)

2.4 The Four Fundamental Subspaces

For a matrix M we have seen the column space $C(M)$ and the nullspace $N(C)$. We can find bases for these spaces by elimination. Lets eliminate some M to some upper triangular U:

$$M = \begin{bmatrix} 1 & 3 & 3 & 2 \\ 2 & 6 & 9 & 7 \\ -1 & -3 & 3 & 4 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 3 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} =: U.$$

Basis of the nullspace

As $M\mathbf{x} = \mathbf{0} \iff U\mathbf{x} = \mathbf{0}$, M and U have the same nullspace

$$N(M) = N(U) = \left\{ z \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} + x \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} \mid x, z \in \mathbb{R} \right\}$$

having dimension $n - r$, and basis $B_{N(M)} = \{(1, 0, -1, 1), (-3, 1, 0, 0)\}$.

Basis of the column space

The column spaces $C(M)$ and $C(U)$ are different. (This is easy to see for, say, the matrix $M = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ which would eliminate to $C = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$.) The column space $C(U)$ of U is easy to read off: it is generated by the pivot columns, so has dimension $r = 2$ and basis $B_{C(U)} = \{(1,0,0),(0,1,0)\}$.

Now because $M\mathbf{x} = \mathbf{0} \iff U\mathbf{x} = \mathbf{0}$, columns of M are independent if and only if the same columns of C are independent. As the pivot columns of U were the first and third, and these gave a basis of $C(U)$ the first and third columns of M are a basis of $C(M)$. It has dimension $r = 2$ and basis $B_{C(A)} = \{(1,2,-1),(3,9,3)\}$.

The row space of M

The *row space* of an $m \times n$ matrix M is the subspace of $\mathbb{R}^n$ generated by the rows of M. These are the columns of M^T so we denote it by $C(M^T)$.

Clearly M and U have the same row space; and it is generated by the pivot (non-zero) rows of U. SO $B_{C(M^T)} = \{(1,3,0,-1),(0,0,1,1)\}$. It has dimension r.

The left nullspace of M

The *left nullspace* of M is the space of row vectors $\mathbf{x} \in \mathbb{R}^m$ such that $\mathbf{x}M = \mathbf{0}$. This is the nullspace of M^T so we denote it $N(M^T)$.

As $N(M^T)$ is the nullspace of M^T if has dimension $m - r$. We can again get the basis by elimination. Computing

$$\left[\begin{array}{cccc|ccc} 1 & 3 & 3 & 2 & 1 & 0 & 0 \\ 2 & 6 & 9 & 7 & 0 & 1 & 0 \\ -1 & -3 & 3 & 4 & 0 & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc|ccc} 1 & 3 & 3 & 2 & 1 & 0 & 0 \\ 0 & 0 & 3 & 3 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 5 & -2 & 1 \end{array}\right] =: [U \mid L],$$

We see that $\begin{bmatrix} 5 & -2 & 1 \end{bmatrix} A = \mathbf{0}$, so this row is in the left nullspace. More generally, rows of L corresponding to the non-pivot rows of M are in the nullspace, and there are exactly $m - r$ of them, so they are a basis for $N(M^T)$. So $B_{N(M^T)} = \{(5,-2,1)\}$.

Left and right inverses

A *left inverse* of an $m \times n$ matrix M is an $n \times m$ matrix L such that

$$LM = I_{n \times n}.$$

A *right inverse* is an $n \times m$ matrix R such that

$$MR = I_{m \times m}.$$

If such an L exists, then the rows of $I_{n \times n}$ are in the $R(L) \leq R^n$. As they are the standard basis of R^n we have then that $R(L) = R^n$ and so

$$n = r \leq m.$$

In particular, the matrix is *tall*: $n \leq m$. Similarily, we can show that if such an R exists then

$$m = r \leq n.$$

In particular, the matrix is *wide*: $n \geq m$. If both exist then $m = n$ and (as we saw in Problem #21 of Section 2.2), $L = R$ is the inverse of M.

Observe that if M has a left inverse L then L^T is the right inverse of A^T:

$$(A^T L^T) = (LA)^T = I^T = I.$$

Finding the left inverse

We saw that the left inverse of $m \times n$ matrix M exists only if $n = r \leq m$. Indeed, this is 'if and only if'. The following example shows how one can find a left inverse for a tall matrix M: one where $n \leq m$. Observe how if this fails, we will have that $r < n$.

Example 2.18. By Gauss Jordan, reduce $[M \mid I]$:

$$\left[\begin{array}{ccc|cccc} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 3 & 2 & 1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 2 & 0 & 0 & 1 & 0 \\ 3 & 3 & 2 & 0 & 0 & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{ccc|cccc} 1 & 0 & 0 & -4 & 2 & 1 & 0 \\ 0 & 1 & 0 & 7 & -3 & -2 & 0 \\ 0 & 0 & 1 & -2 & 1 & 1 & 0 \\ 0 & 0 & 0 & -5 & 1 & 1 & 1 \end{array}\right] =: [U \mid B]$$

As

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

the first three rows of B make a left inverse L of M. So $L = \begin{bmatrix} -4 & 2 & 1 & 0 \\ 7 & -3 & -2 & 0 \\ -2 & 1 & 1 & 0 \end{bmatrix}$.

Notice also that $(-5, 1, 1, 1)$ is in $N(M^T)$. So adding it to any row of L doesn't change the fact that it is a left inverse of M. So there can be may left inverses of M.

Finding the right inverse

The right inverse exists only if the matrix is wide. The following example shows that it exists if $m = r$.

Example 2.19. Reduce M by elimination

$$\left[\begin{array}{cccc|ccc} 1 & 1 & 2 & 3 & 1 & 0 & 0 \\ 1 & 2 & 0 & 3 & 0 & 1 & 0 \\ 1 & 3 & 1 & 2 & 0 & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{cccc|ccc} 1 & 0 & 0 & 7 & -2 & 7 & -4 \\ 0 & 1 & 0 & -2 & 1 & -3 & 2 \\ 0 & 0 & 1 & -1 & 1 & -2 & 1 \end{array}\right] =: [U \mid B].$$

As $m = r$ some choice of m columns of U contains $I_{m\times m}$. Choosing the corresponding columns M' of M we get

$$\begin{bmatrix} -2 & 7 & -4 \\ 1 & -3 & 2 \\ 1 & -2 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 0 \\ 1 & 3 & 1 \end{bmatrix}}_{M'} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

So $B = M'^{-1}$ is the inverse of the square matrix M'. So it is also the right inverse of M':

$$\begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 0 \\ 1 & 3 & 1 \end{bmatrix} \begin{bmatrix} -2 & 7 & -4 \\ 1 & -3 & 2 \\ 1 & -2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We want a right inverse of M though. So when we add the column back to M' we have to add a row of 0s back to B to get R:

$$\begin{bmatrix} 1 & 1 & 2 & 3 \\ 1 & 2 & 0 & 3 \\ 1 & 3 & 1 & 2 \end{bmatrix} \begin{bmatrix} -2 & 7 & -4 \\ 1 & -3 & 2 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Here we have found a right inverse R of the wide matrix M, and the only thing we needed was that there were m pivots. This works whenever $m = r$. Observe also that as M is wide, it has non-trivial nullspace. Adding vectors from the $N(M)$ to the columns of R does not change the product, so gives other right inverses.

Problems from the text

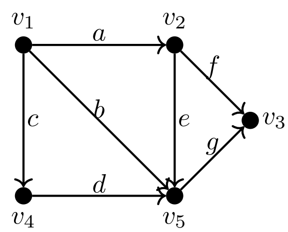
2.4: 2, 4, 6, 9, 11, 15, 23, 28, 30, 31, 33 (3, 5, 12, 14, 17, 16, 27, 29, 33, 34, 30)

2.5 Graphs applications

A graph consists of a set of *vertices* (or *nodes*) $V = \{v_1, \ldots, v_n\}$, and a set E two elements subsets of V, called *edges*.

For the applications in this we consider *oriented graphs*. The edges will be will be ordered pairs rather than sets. So we usually call them *arcs* instead of *edges*. We cannot have both the edge (u, v) and the edge (v, u).

Example 2.20. We can draw a graph G as follows.



and can alternately represent it by its incidence matrix M_G:

$$
\text{Arcs}\quad
\begin{array}{c}
\\ a \\ b \\ c \\ d \\ e \\ f \\ g
\end{array}
\overset{\text{Nodes}}{
\begin{array}{ccccc}
v_1 & v_2 & v_3 & v_4 & v_5 \\
\hline
-1 & 1 & 0 & 0 & 0 \\
-1 & 0 & 0 & 0 & 1 \\
-1 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1 & 1 \\
0 & -1 & 0 & 0 & 1 \\
0 & -1 & 1 & 0 & 0 \\
0 & 0 & -1 & 0 & 1
\end{array}}
$$

We interpret the spaces associated to the incidence matrix M_g of a graph G.

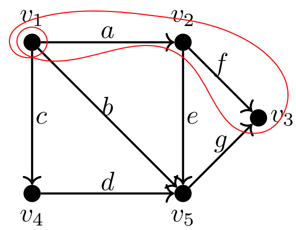
Nullspace of M_G

Consider the product

$$
M_G \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \\ -2 \\ 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} \begin{matrix} a \\ b \\ c \\ d \\ e \\ f \\ g \end{matrix}
$$

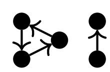
assigning a number to each arc: -2 to a etc.

We can interpret this as follows. The column matrix $(2,1,0,0,1)$ picks out some nodes: node v_1 twice, node v_2 once, etc. View this as defining a curve around these nodes.



The resulting column counts the arcs going from ourside the curve to inside. Arc g goes from outside to inside. Arc e goes from inside to outside. Arc a goes from inside 'two levels' to 'inside one level'.

Thus an integer solution to $M_G\mathbf{x} = \mathbf{0}$ would correspond to a curve that doesn't cross any of the arcs. $M_G(1,1,1,1,1)^T = \mathbf{0}$. The dimension of the nullspace of M_G is the number of components of G. That is if G is,



then M_G has nullity 2.

Rowspace and Left Nullspace

A *cycle* in a graph is an alternating sequence of distinct vertices and edges $v_1, e_1, v_2, e_2, \ldots, v_g, e_g$ such that $e_i = \{v_i, v_{i+1}\}$ for all $i < g$ and $e_g = \{v_g, v_1\}$. When our edges are oriented, the orientation doesn't matter. We often describe a cycle just by its edges.

One will notice the following.

Fact 2.21. *Adding up the rows of M_G corresponding to a cycle of G gives $\mathbf{0}$.*

Example 2.22. The cycle $a+e-b$ that goes from v_1 to v_2 to v_5 and back to v_1 corresponds to the product

$$\begin{bmatrix}1 & -1 & 0 & 0 & 1 & 0 & 0\end{bmatrix} M_G = \begin{matrix} 1\begin{bmatrix}1 & -1 & 0 & 0 & 0\end{bmatrix} + \\ -1\begin{bmatrix}0 & -1 & 0 & 0 & 1\end{bmatrix} + \\ 1\begin{bmatrix}-1 & 0 & 0 & 0 & 1\end{bmatrix}\end{matrix} = \mathbf{0}$$

Fact 2.21 tells us that **cycles of G are vectors in the left nullspace of** M_G. As the properties of a space are useful, we often generalise the definition of cycles as any set of edges corresponding to a vector in the nullspace of M_G. Thus we can add cycles, and so get a *cycle space of G*.

Example 2.23. Where C_1 is the cycle $a + e - b$ and C_2 is $b - d - c$ the sum $C_1 + C_2 = a + e - d - c$.

Practice

Exemplify this sum of cycles with a picture.

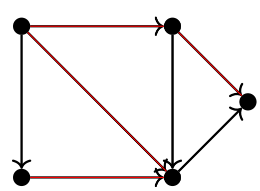
Practice

Show that under this definition, cycles need not be connected.

We also see from Fact 2.21 that the rows of M_G corresponding to a cycle of G are dependent. One can prove the following.

Fact 2.24. *The arcs corresponding to any dependent set of rows of M_G contain a cycle in G. If the co-efficients in the dependence are all ± 1, then M_G is a (union of disjoint) cycles.*

The dimension r of the rowspace is the size of the biggest subgraph with no cycles. In a connected graph this is a spanning tree,



which has $r = n - 1$ edges. In general $r = n - c(G)$ where $c(G)$ is the number of components of G. It follows that the dimension of the column space is $n - r = c(G)$, as we observed above.

Practice

Show that the dimension of the cycles space of G is $m - n + c(G)$?

An application: Euler's Formula

We (non-rigorously) prove a (relatively easy) classical result of topology / graph theory.

A graph is *planar* if it can be drawn on the plane without an crossing edges. Take a planar drawing of a planar graph. Cutting the plane up by cutting along the edges you partition the plane into the *faces* of the the graph. The edges separating a face from the rest of the plane are a cycle, and it is not hard to see that these 'face cycles' generate the cycle space over $\mathbb{Z}$ (though maybe we haven't really defined that.)

Thus the the number of faces is exactly the dimension of the cycle space, and so we have

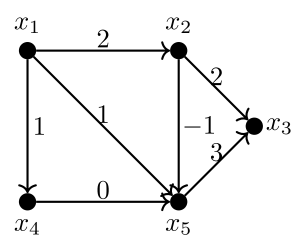
$$f = m - n + g(G)$$

where f denotes the number of faces for the plane graph G. This is Euler's Formula.

The column space

Recall that the column space of a matrix M is the set of $\mathbf{b}$ such that $M\mathbf{x} = \mathbf{b}$ has a solution. What does this mean when M is the adjacency matrix M_G of a graph G?

With our previously pictured G, let $\mathbf{b} = (2, 1, 1, 0, -1, 2, 3)$ and assume that $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{bmatrix}^T$ is a solution to $M_G\mathbf{x} = \mathbf{b}$. The solution x_i to each node v_i, and $\mathbf{b}$ assigns a value to each arc.

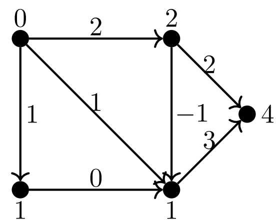


As $(1, 1, 1 \ldots, 1)$ is in the nullspace, if there is a solution, then there is one with $x_1 = 0$, so lets assume that $x_1 = 0$.

The first row of $M_G\mathbf{x} = \mathbf{b}$, that is, of

$$\begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & -1 & 0 & 0 & 1 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 0 \\ -1 \\ 2 \\ 3 \end{bmatrix}$$

says that $-x_1 + x_2 = 2$. Thus the 2 corresponding on the arc $v_1 \to v_2$ means that $x_2 = x_1 + 2$. So a solution looks like this:



Were were lucky that this all worked out. When we follow the arcs around a circle, adding values to the vertices according to the values on the edges, things might not work out. We need that the sum of the arc values around any cycle (paying attention to orientation) is 0.

Fact 2.25. *$M_G\mathbf{x} = \mathbf{b}$ has a solution if and only if the sum of the values in $\mathbf{b}$ corresponding to any cycle c of G, add up to 0.*

Viewing the values x_i on each vertex as a voltage, and the values on the arcs as voltage drops (or as current when the wire between any two vertices is assumed to have resistance 1) then 'there is a solution to $M_G\mathbf{x} = \mathbf{b}$ if and only if the current around each closed walk adds up to 0.' You might recognise this as Kirchoff's law of conservation of energy.

Problems from the text

2.5: 2 (1)

2.6 Linear Transformations

A function $f : V \to W$ of vector spaces is a *linear transformation* if it satisfies the following for all $a, b \in \mathbb{R}$ and $\mathbf{x}, \mathbf{y} \in V$:

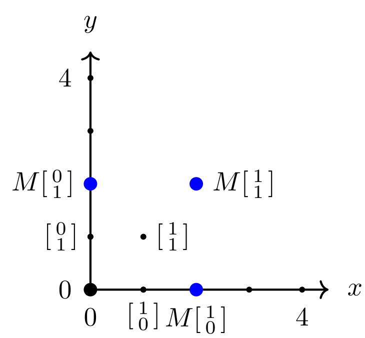
$$T(a\mathbf{x} + b\mathbf{y}) = aT(\mathbf{x}) + bT(\mathbf{y}).$$

Observe that matrix multiplication satisfies this:

$$M(a\mathbf{x} + b\mathbf{y}) = aM\mathbf{x} + bM\mathbf{y},$$

so multiplication (on the left) by a matrix M is a linear transformation.

Example 2.26. The matrix $\left[\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right]$ takes the point $(0, 1)$ to the point $\left[\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right]\left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right]$. It takes the point $(1, 1)$ to $(2, 2)$ and the point $(1, 0)$ to $(2, 0)$.



It seems to be 'blowing $\mathbb{R}^2$ up by a factor of 2', or 'zooming in on $\mathbb{R}^2$.

Practice

What does multiplication by 1) $\left[\begin{smallmatrix}0 & -1\\1 & 0\end{smallmatrix}\right]$ 2) $\left[\begin{smallmatrix}0 & 1\\1 & 0\end{smallmatrix}\right]$ 3) $\left[\begin{smallmatrix}1 & 0\\0 & 0\end{smallmatrix}\right]$ do to these same points? Describe what multiplication by these matrices seems to be doing to $\mathbb{R}^2$ as a whole.

Practice

Which of the following are linear transformations of R^2?

i. Reflection in the x-axis.

ii. Mapping everything to 0.

iii. Shifting left 1.

For those that are, are there matrices that give this transformation?

A good way to find the matrix that gives a transformation, is to check first what the transformation does to a basis of the vector space.

Example 2.27. Consider the transformation T of $\mathbb{R}^2$ in which every vector is reflected in the x-axis. On the standard basis we have $T((0,1)) = (0,-1)$ and $T((1,0)) = (1,0)$, so T is represented by the matrix M such that

$$M\begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\-1\end{bmatrix} \text{ and } M\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}1\\0\end{bmatrix}.$$

So $M = \left[\begin{smallmatrix}1 & 0\\0 & -1\end{smallmatrix}\right]$.

Rotation by $\pi/2$ degree is represented by the matrix $\left[\begin{smallmatrix}0 & -1\\1 & 0\end{smallmatrix}\right]$; what about rotation by an angle θ?

Practice

Represent the transformation R_θ which rotates everything in $\mathbb{R}^2$ by θ radians, by a matrix.

Now to rotate by 2θ we could multiply by R_θ twice, or by $(R_\theta)^2$ once. So rotation by 2θ is (writing c for $\cos\theta$ and s for $\sin\theta$.

$$\begin{bmatrix}\cos 2\theta & -\sin 2\theta \\ \sin 2\theta & \cos 2\theta\end{bmatrix} = (R_\theta)^2 = \begin{bmatrix} c & -s \\ s & c\end{bmatrix}\begin{bmatrix} c & -s \\ s & c\end{bmatrix} = \begin{bmatrix} c^2 - s^2 & -2cs \\ 2cs & c^2 - s^2\end{bmatrix}.$$

This gives us our double angle formulae:

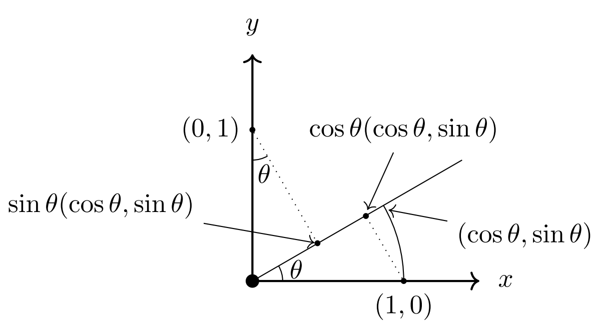
- $\cos 2\theta = \cos^2\theta - \sin^2\theta$
- $\sin 2\theta = 2\cos\theta\sin\theta$

Practice

Prove the angle sum identities for $\sin(\phi+\theta)$ and $\cos(\phi+\theta)$ in the same way.

Let's do some harder transformations.

Example 2.28. How do we represent the projection onto the line L_θ that makes a positive angle of θ with the x-axis? Well, we could figure out where the vector $(1, 0)$ projects:



And read off that this is represented by the matrix

$$\begin{bmatrix}\cos^2\theta & \cos\theta\sin\theta \\ cos\theta\sin\theta & \sin^2\theta\end{bmatrix}.$$

Or we it another way, being clever instead of having to think. It is easy to find the matrix P_x that projects onto the x-axis: $P_x = \left[\begin{smallmatrix}1 & 0 \\ 0 & 0\end{smallmatrix}\right]$. To project onto the line L_θ we can rotate a vector by $-\theta$, projecting onto the x-axis, and then rotating back by θ. So projection onto L_θ is

$$\begin{aligned}
R_\theta P_x R_{-\theta} &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \cos -\theta & -\sin -\theta \\ \sin -\theta & \cos -\theta \end{bmatrix} \\
&= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos -\theta & -\sin -\theta \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ 0 & 0 \end{bmatrix} \\
&= \begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{bmatrix}
\end{aligned}$$

Multiplying by R_θ and R_θ^{-1} is called a change of basis. This is a useful technique that we well use later.

For the homework you will need to know that the *kernal* of a transformation T is the set of vectors v such that $T(v) = 0$. So it is the nullspace of the matrix M that represents the transformation T.

> Problems from the text
>
> **2.6:** 5, 9, 12, 16, 20, 23, 24, 26, 28, 29, 49 (6, 1, 3, 23, 22, 19, 17, 29, 26, 27, 50

3 Orthogonality

3.1 Orthogonal Vectors and Subspaces

We saw the standard basis for R^n:

$$\{\underbrace{(1,0,0,\dots,0)}_{e_1}, \underbrace{(0,1,0,\dots,0)}_{e_2}, \dots, \underbrace{(0,0,0,\dots,1)}_{e_n}\}.$$

It is a really nice basis to work with. When we are looking for a basis of some subspace U of R^n, we can't always get such a nice basis. What's nice about them? They all have length (distance from the origin) 1, and they are all at right angles with each other. We try to get close to this for U.

Definition 3.1. Two non-zero vectors u and v in $\mathbb{R}^n$ are *orthogonal*, written $u \perp v$, if the (smallest) angle between them is $\pi/2$ (i.e., 90°). (It is useful to assume that the zero vector is orthogonal to all vectors.) A vector is *unit* if it has length 1. A basis is *normal* if it consists of unit vectors, *orthogonal* if it consists of vectors are pairwise orthogonal, and *orthanormal* if it is normal and orthogonal. (Some books use othanogonal to mean orthanormal.)

The goal in this chapter is to take a basis, and find an equivalent orthanormal basis. We investigate many nice properties of orthogonality along the way.

Length of a vector

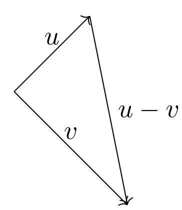
As the length of (x, y) in $\mathbb{R}^2$ is $\sqrt{x^2+y^2}$ by pythagoras, the length of (x, y, z) in $\mathbb{R}^3$ is $\sqrt{x^2+y^2+z^2}$. The length of $v = (v_1, v_2, \dots, v_n)$ is

$$\|v\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

So $\|v\| = \left[\begin{smallmatrix} v_1 & v_2 & \dots & v_n \end{smallmatrix}\right] \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = v^T v$, (viewing vectors as columns.) To get a basis of unit vectors from a basis, we can just replace each vector v by $v/\|v\|$. This is called *normalising* the vector space.

Orthogonal vectors

Now, vectors u and v are orthogonal



if and only if $\|u\|^2 + \|v\|^2 = \|u - v\|^2$. But

$$\begin{aligned} \|u - v\|^2 &= (u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2 \\ &= (u_1^2 - 2u_1v_1 + v_1^2) + (u_2^2 - 2u_2v_2 + v_2^2) + \dots \\ &= (u_1^2 + \cdots + u_n^2) + (v_1^2 + \dots, v_n^2) - 2(u_1v_1 + \cdots + u_nv_n) \\ &= \|u\|^2 + \|v\|^2 - 2u^Tv \end{aligned}$$

So $u \perp v$ if and only if $u^Tv = 0$.

Example 3.2.

- $(0, 0, 1)(0, 1, 0)^T = 0 + 0 + 0 = 0$: orthogonal.

- $(1,0)(1,1)^T = 1 + 0 = 1$: not orthogonal.
- $(1,1)(1,-1)^T = 1 - 1 = 0$: orthogonal.

> **Note**
>
> Hey! is it $u^T v$ or uv^T? Sorry, yeah. When I write it as v we think of it as a column vector, so we write $u^T v$. But when I'm writing it out. It takes less space to write it as a row vector, so it is $(u_1, u_2)(v_1, v_2)^T$. However we write it, it often denoted $u \cdot v$ and called the *dot product*. I like this notation, as then we don't have to worry about whether we are writing u as a row or a column. I'll try to use this more. The book doesn't though.

Theorem 3.3. *A set of non-zero vectors that are pairwise othogonal, are independent.*

Proof. Let $v_1, \ldots, v_n$ be pairwise orthogonal, non-zero vectors. To see they are independent, let $0 = c_1v_1 + c_2v_2 + \cdots + c_nv_n$. Taking inner products with v_i we get that

$$\begin{aligned} 0 = v_i \cdot 0 &= v_i \cdot (c_1v_1 + c_2v_2 + \cdots + c_nv_n) \\ &= c_1v_i \cdot v_1 + c_2v_i \cdot v_2 + \cdots + c_nv_i^tv_n \\ &= c_iv_i \cdot v_i = c_i \end{aligned}$$

This shows that $c_i = 0$, and works for all i, so the vectors are linearly independent. ☐

Orthogonal Subspaces

Definition 3.4. Two subspaces U and V of $\mathbb{R}^n$ are orthgonal, written $U \perp V$, if for all $u \in U$ and $v \in V$, $u \perp v$. For a subspace V of $\mathbb{R}^n$, the orthogonal complement $V^\perp$ (read 'V perp') of V, is the space of all vectors orthogonal to all vectors of V.

> **Note**
>
> To show that two subspaces are orthogonal, it is enough to show that every vector in the basis of one space is orthogonal to every vector in a basis of the other.

Observe that each row r in M and each vector x of the nullspace, $r \cdot x = 0$. As the rows of a matrix contain a basis of its rowspace we get that the row space and nullspace of any matrix are othogonal. In fact, as the nullspace is the set of all vectors x such that $r \cdot x = 0$, the nullspace $N(M)$ of a matrix M is exactly the orthogonal complement $C(M^T)^\perp$ of its row space.

Example 3.5. The matrix $M = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 5 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -1 \end{bmatrix}$ has rowspace $\{a(1,0,3) + b(0,1,-1) \mid a, b \in \mathbb{R}\}$ and nullspace $\{z(-3,1,1) \mid z \in \mathbb{R}\}$. We have

$$(1,0,3) \cdot (-3,1,1) = 0 \qquad \text{and} \qquad (0,1,-1) \cdot (-3,1,1) = 0,$$

as expected.

Now consider a subspace V of $\mathbb{R}^n$ with basis $b_1, \ldots, b_r$, and let M be the matrix whose rows are the b_i. Then $V^\perp = N(M)$, which gives us $\dim(V^\perp) = n - r$, and so $\dim((V^\perp)^\perp) = n - (n-r) = r$. As we clearly have $V \subseteq (V^\perp)^\perp$, we get that $V = (V^\perp)^\perp$.

Fact 3.6. *For any subspace* $V \subseteq \mathbb{R}^n$, $(V^\perp)^\perp = V$ *and* $\dim(V^\perp) = n - \dim(V)$.

Applying this to a matrix M, and its transpose, we get the following.

Theorem 3.7. *[Fundamental Theorem of Linear Algebra] For an* $m \times n$ *matrix* M*,*

- $C(M^T)^\perp = N(M)$ *and* $N(M)^\perp = C(M^T)$*, and*
- $C(M)^\perp = N(M^T)$ *and* $N(M^T)^\perp = C(M)$.

Practice

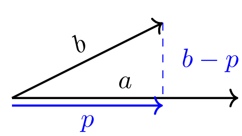
What is the dimension of the orthogonal complement $V^\perp$ of a space V spanned by the vectors $(1,2,4)$ and $(2,1,1)$? Find $V^\perp$?

Problems from the text

3.1: 3, 6, 7, 8, 12, 15, 17, 23, 35 (6, 4, 1, 19, 14, 20, 18, 26, 42)

3.2 Cosines and Projections

We talked about normalising vector spaces– changing the basis so that it consisted of unit vectors. This was easy. We now look at changing the basis to a basis of pairwise orthogonal vectors. The idea is simple, for for two basis vectors a and b we split b up into an a-part p and a part $b - p$ that is orthogonal to a.



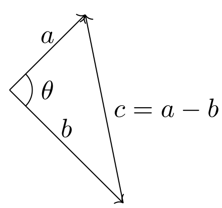
We can then write b as $p + (b - p)$ so $\{a, b - p\}$ spans the same space as $\{a, b\}$. This p such that $b - p$ is orthogonal to a is called the *projection* of b onto a, and denoted $p = \text{proj}_a(b)$. It is clearly the vector of length $\|b\| \cos\theta$ in the direction of a so we can write it as

$$\text{proj}_a(b) = \|b\| \cos\theta \frac{a}{\|a\|}. \tag{1}$$

We get a nicer formula for this using the inner product.

Formulae for $a \cdot b$ and for $\text{proj}_a(b)$

We saw that $a \perp b$ if and only if $a^T b = 0$. Let's start with an expression for $a^T b$ when $a \not\perp b$. Recall the pythagorean theorem for non-right angles (i.e., the cosine law):



$$c^2 = a^2 + b^2 - 2ab\cos\theta$$

On the one hand, we have

$$\begin{aligned} \|a - b\|^2 &= (a - b) \cdot (a - b) = a \cdot a + b \cdot b - b \cdot a - a \cdot b \\ &= a \cdot a + b \cdot b - 2a \cdot b \end{aligned}$$

while on the other hand we have by the cosine rule

$$\begin{aligned} \|a - b\|^2 &= \|a\|^2 + \|b\|^2 - 2\|a\| \cdot \|b\| \cos\theta \\ &= a \cdot a + b \cdot b - 2\|a\| \cdot \|b\| \cos\theta \end{aligned}$$

From this we see that $a \cdot b = \|a\| \cdot \|b\| \cos\theta$ where θ is the angle between a and b. Solving this for $\|b\| \cos\theta$ and plugging it into (1) we get

$$\text{proj}_a(b) = \frac{a \cdot b}{\|a\|} \frac{a}{\|a\|} = \frac{a \cdot b}{a \cdot a} a.$$

Note:

Memory Aid:

$$\cos\theta = \frac{u}{\|u\|} \cdot \frac{v}{\|v\|}$$

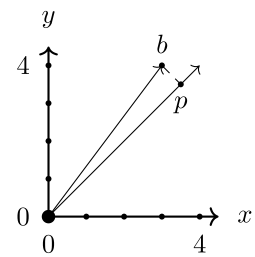
or if u and v are unit, just

$$\cos\theta = u \cdot v$$

We use the projection formula to:

- find the distance from a point to a line,
- write a vector in terms of a basis, or
- find an orthanormal basis of a space.

Example 3.8. To find the projection p of the point $b = (3, 4)$ onto the line $y = x$



we project $b = (3,4)$ as a vector onto the vector $a = (1,1)$ which defines the line. So p is $\frac{b \cdot a}{a \cdot a} a = \frac{(3,4)\cdot(1,1)}{2} a = \frac{7}{2}(1,1)$.

We often refer to the difference $e = b - \text{proj}_a(b) = (3,4) - (3.5, 3.5) = \frac{1}{2}(-1,1)$ as the *error.* Its length $\|e\| = \sqrt{2}/2$ is the distance from p to $y = x$.

Example 3.9. We find an orthanormal basis of the subspace U of $\mathbb{R}^3$ that is generated by $\{a = (1,1,0), b = (1,2,3)\}$.

First, we orthogonalise by removing the a-component of b:

$$\begin{aligned} b' &= b - \text{proj}_a(b) = (1,2,3) - \frac{(1,2,3)\cdot(1,1,0)}{(1,1,0)\cdot(1,1,0)}(1,1,0) \\ &= (1,2,3) - \frac{3}{2}(1,1,0) = (-1/2, 1/2, 3). \end{aligned}$$

So $\{a, b'\}$ is an orthogonal basis of U. Now, normalising, we get the orthonormal basis $\{\sqrt{1/2}(1,1,0), \sqrt{1/38}(-1,1,6)\}$.

Finding an orthanormal basis of a higher dimensional subspace is the same idea: from the third basis vector we have to remove the projections onto the previous two. We will come back to this in a couple of sections, and look at an orderly way to do it.

We have one last thing to point out here. The projection onto a vector a is a linear transformation, so can be represented as (multiplication on the left by) a matrix P. Indeed, as $a^T b$ is a scalar, we have $(a^T b)a = a(a^T b)$ so the projection formula becomes

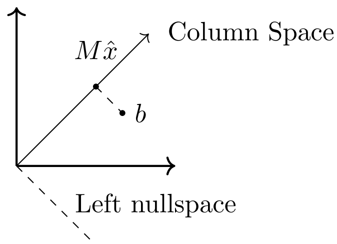
$$\text{proj}_a(b) = \frac{a^T b}{a^T a} a = \frac{a a^T}{a^T a} b =: Pb,$$

where $\frac{1}{a^T a}$ is a scalar and aa^T is a square matrix, so $P := \frac{aa^T}{a^T a}$ is a square matrix.

Problems from the text

3.2: 5, 9, 11, 12, 15, 19, 20, 22 (11, 12, 3, 10, 16, 17, 19, 24)

3.3 Projections and least squares



Sometimes the matrix equation $Mx = b$ we are trying to solve has no solution, but, entitled as we are, we want one anyways. That there is no solution means that b is not in the column space. We settle for the vector in the column space that is closest to b. That is, the projection of b onto the column space C. We find $\hat{x}$ such that $M\hat{x} = \text{proj}_C(b)$. The error $e = b - M\hat{x}$, is orthogonal to the column space, so is in the left nullspace of M.

So we would like to find $M\hat{x}$ and $\hat{x}$. How do we do this? Observe first that as $M\hat{x} = b - e$ we have

$$M^T M\hat{x} = M^T(b - e) = M^T b - 0 = M^T b.$$

Now while M is not invertible (because $Mx = b$ has no solution), $M^T M$ is. Multipling on the left by $(M^T M)^{-1}$ we get:

$$\hat{x} = (M^T M)^{-1} M^T b \qquad \text{and so} \qquad M\hat{x} = M(M^T M)^{-1} M^T b.$$

Note:

We don't divide by matrices, but if we did, we might write the projection matrix at $\frac{MM^T}{M^T M}$. Compare this to the formula for the projection matrix projecting onto a vector a.

This $\hat{x}$ is called the *least squares estimate* for $Mx = b$. We will explain this is a second. As $M\hat{x}$ is the projection of x onto the column space of M, this matrix $M(M^T M)^{-1} M^T$ is called the *projection matrix*. The matrix $M^T M$ is called the *cross product matrix*.

We will look at some properties of these matrices. But first lets practice computing with them.

Example 3.10. It is clear by looking at the first and third row that when $M = \begin{bmatrix} -1 & -1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$ and $b = \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix}$, the equation $Mx = b$ has no solution. We find $\hat{x}$ and $M\hat{x}$.

Indeed,

- $M^T M = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}$
- $(M^T M)^{-1} : \left[\begin{array}{cc|cc} 2 & 2 & 1 & 0 \\ 2 & 3 & 0 & 1 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|cc} 1 & 0 & 3/2 & -1 \\ 0 & 1 & -1 & 1 \end{array}\right]$
- $\hat{x} = (M^T M)^{-1} M^T b = \begin{bmatrix} 3/2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 3/2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} -7/2 \\ 3 \end{bmatrix}$
- $M\hat{x} = \begin{bmatrix} -1 & -1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -7/2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 3 \\ -1/2 \end{bmatrix}$

Properties of the cross product and projection matrices

For an $m \times n$ matrix M the cross product matrix M^TM is a symmetric $n \times n$ matrix which is invertible if and only if the rank of M is n. Indeed, it has rank n if and only if M does. It cannot have greater rank than M as each row is a linear combination of rows of M. On the other hand, it cannot have rank less than that of M as M and M^TM because of the following.

Practice

Show that M and M^TM have the same nullspace.

The projection matrix $P = M(M^TM)^{-1}M^T$ projects any vector b onto the column space $C(M)$ of M. That is to say, for all $b \in \mathbb{R}^n$:

- $Pb \in C(M)$, and
- $e = b - Pb$ is orthogonal to $C(M)$.

Practice

Show that P is symmetric and idempotent: $P^2 = P$.

Now, any matrix P with these properties projects vectors onto the column space of some matrix M. Indeed, P itself is such a matrix: P certainly maps b to $C(P)$. To see that $b - Pb$ is orthogonal to $C(P)$ we show that it is orthogonal to any $Pc \in C(P)$:

$$(b - Pb)^TPc = b^T(I - P)^TPc = b^T(P - P^2)c = 0.$$

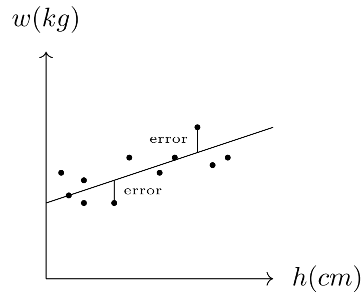
An application: the least squares line

A nice application of finding the approximate solution $\hat{x}$ of a system is using it to find the least squares regression line of a set of data.

We believe that a persons weight w is approximately a linear function of their height h. That is, we believe that

$$w = mh + b$$

for some constants m and b. We sample n people in order to find m and b.



The points seem to best fit some line $w = mh + b$. But what line? What are m and b? Ideally we want m and b such that for each point (w_i, h_i) we have $w_i = mh_i + b$. We want to solve $Mx = w$, i.e.,

$$\begin{bmatrix} 1 & h_1 \\ 1 & h_2 \\ \vdots & \\ 1 & h_n \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix},$$

but it has no solution. How about replacing w with the closest vector $\hat{w}$ in the column space, and finding the solution $\hat{x}$ for that. What this does is minimises $\|MX - w\|$ and so

$$\|Mx - w\|^2 = \sum_i (b + mh_i - w_i)^2 = \sum_i e_i^2.$$

So this is called the least squares line.

Example 3.11. What is the the least squares line for the points $(1, 3), (1, 4)$ and $(2, 6)$. We get the approximate solution to:

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix}}_{M} x = \underbrace{\begin{bmatrix} 3 \\ 4 \\ 6 \end{bmatrix}}_{b}.$$

- $M^T M = \left[\begin{smallmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \end{smallmatrix}\right] \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 2 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 3 & 4 \\ 4 & 6 \end{smallmatrix}\right]$
- $(M^T M)^{-1} = \frac{1}{2} \left[\begin{smallmatrix} 6 & -4 \\ -4 & 3 \end{smallmatrix}\right]$
- $M^T b = \left[\begin{smallmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \end{smallmatrix}\right] \left[\begin{smallmatrix} 3 \\ 4 \\ 6 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 13 \\ 19 \end{smallmatrix}\right]$
- $\hat{x} = -\frac{1}{2} \left[\begin{smallmatrix} 6 & -4 \\ -4 & 3 \end{smallmatrix}\right] \left[\begin{smallmatrix} 13 \\ 19 \end{smallmatrix}\right] = \left[\begin{smallmatrix} 1 \\ 5/2 \end{smallmatrix}\right]$

So $w = \frac{5}{2}h + 1$ is our least squares line.

Problems from the text

3.3: 1, 3, 6, 7, 8, 9, 11, 15, 17, 18, 22 (5, 2, 6, 10, 8, 13, 14, 12, 11, 18, 20)

3.4 Orthogonal Basis and Gram Schmidt

Given a basis $\{b_1, \ldots, b_m\}$ of a subspace V of $\mathbb{R}^n$ we now find an equivalent orthanormal basis $\{q_1, \ldots, q_m\}$ using the following 'Gram-Schmidt process'. First, let $q_1 = b_1/\|b_1\|$ and for $i = 2, \ldots, m$ let $q_i = q_i'/\|q_i'\|$ where

$$q_i' = b_i - \sum_{j=1}^{i-1} \operatorname{proj}_{q_j} b_i.$$

Example 3.12. Let $V \subseteq \mathbb{R}^3$ be generated by

$$B = \{b_1 = (1,0,1), b_2 = (1,0,0), b_3 = (2,1,0)\}.$$

Clearly the standard normal basis is an equivalent basis, but ignoring this for the moment, we use the Graham-Schmidt process to get another orthanormal basis.

- $q_1 = b_1/\|b_1\| = \frac{1}{\sqrt{2}}(1,0,1)$
- $q_2' = b_2 - (q_1^T b_2)q_1 = (1,0,0) - \frac{1}{\sqrt{2}}(1)\frac{1}{\sqrt{2}}(1,0,1) = (1/2, 0, -1/2)$
- $q_2 = \frac{1}{\sqrt{2}}(1,0,-1)$
- $q_3' = b_3 - (q_1^T b_3)q_1 - (q_2^T b_3)q_2 = \cdots = (0,1,0)$
- $q_3 = (0,1,0)$.

Decomposing a vector with an orthonormal basis

Okay. That was fun, but what is really fun is expressing a vector as a linear combination of basis vectors. Once we have an orthanormal basis, this is easy. Indeed, any vector $v \in V$ can be written as

$$v = r_1 q_1 + r_2 q_2 + \cdots + r_m q_m,$$

for some scalars r_i. But as $q_i^T v = \sum_{j=1}^m x_j q_i^T q_j = x_i$ this can be written

$$v = \sum_{i=1}^m (q_i^T v) q_i = \sum_{i=1}^m q_i (q_j^T v),$$

or more compactly as

$$v = QQ^T v$$

where Q is the matrix whose i^{th} column is q_i so

$$Q^t v = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}$$

Properties of Orthogonal Matrices

The matrix Q whose columns are a set of orthanormal vectors has some other nice properties that arise from the fact that $Q^T Q = I$.

Such a matrix Q **preserves inner products**:

$$(Qu)^T(Qv) = u^T Q^T Q v^T = u^T v.$$

It follows immediately that it **preserves length**:

$$\|Qv\| = \|v\|$$

and **preserves the angle between vectors**:

$$\cos^{-1}\left(\frac{(Qu)^T(Qv)}{\|Qu\| \cdot \|Qv\|}\right) = \cos^{-1}\left(\frac{u^T v}{\|u\| \cdot \|v\|}\right).$$

Though $Q^T Q = I$, Q need not be invertible, as it need not be square. If it is square though, we give it a special name, and call it *orthogonal.* For an orthogonal matrix Q then $Q^T = Q^{-1}$., which is convenient: as the inverse is immediate it is easy to solve $Qx = b$.

Even when Q is not square, so is a rectangular matrix with orthogonal columns, Q^T is its left inverse. The matrix QQ^T is square and also has some nice properties. Indeed

$$Q(Q^T Q)Q^T = QIQ^T = QQ^T$$

so $QQ^T =: P_Q$ is the matrix that projects onto the column space of Q.

Thus the projection of b onto $C(Q)$ is $QQ^T b$, and the least squares approximation for

$$Qx = b$$

is just $\hat{x} = Q^T b$.

The QR-factorisation of M (Skipped this bit)

Writing out our original basis $\{b_1, \ldots, b_m\}$ in terms of our orthonormal basis $\{q_1, \ldots, q_m\}$:

$$\begin{array}{lllllllll} b_1 & = & q_1^T b_1 q_1 & +q_2^T b_1 q_2 & +q_3^T b_1 q_3 & = & q_1^T b_1 q_1 & +0 & +0 \\ b_2 & = & q_1^T b_2 q_1 & +q_2^T b_2 q_2 & +q_3^T b_2 q_3 & = & q_1^T b_2 q_1 & +q_2^T b_2 q_2 & +0 \\ b_3 & = & q_1^T b_3 q_1 & +q_2^T b_3 q_2 & +q_3^T b_3 q_3 & = & q_1^T b_3 q_1 & +q_2^T b_3 q_2 & +q_3^T b_3 q_3 \end{array}$$

we see that this can be written as

$$\underbrace{\begin{bmatrix} b_1 \mid b_2 \mid b_3 \end{bmatrix}}_{M} = \underbrace{\begin{bmatrix} q_1 \mid q_2 \mid q_3 \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} q_1^T b_1 & q_1^T b_2 & q_1^T b_3 \\ 0 & q_2^T b_2 & q_2^T b_3 \\ 0 & 0 & q_2^T b_3 \end{bmatrix}}_{R}$$

where R is upper triangular.

Instead of computing $\hat{x} = (M^T M)M^T b$ for the least squares problem we have that

$$M^T M = R^T Q^T Q R = R^T R$$

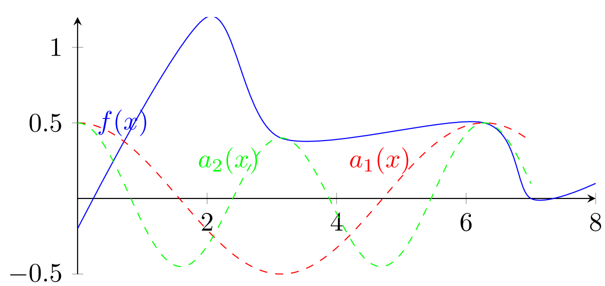
so we can compute $R^T R(M^T b)$. This is easier as R is triangular.

Problems from the text

3.4: 1, 6, 7, 8, 10 (2, 14, 11, 10, 6)

3.5 Fast Fourier Transforms

Take a continuous function $f(x)$. One can approximate it by periodic functions on some interval.



Indeed, $a_1(x) = f(2\pi)\cos x$ agrees with it at $x = 2\pi$. And $a_2(x) = \frac{f(2\pi)-f(\pi)}{2}\cos x + \frac{f(2\pi)+f(\pi)}{2}\cos 2x$ agrees with it at π and 2π. How can we make an approximation that agrees with it on n points?

Well for a_2 we wrote $a_2(x) = c_1 \cos x + c_2 \cos 2x$, where c_1 and c_2 were such that this agreed with $f(x)$ when $x \in \{\pi, 2\pi\}$. That is, we found c_1 and c_2 such that

$$\begin{bmatrix}\cos(1\cdot\pi) & \cos(2\cdot\pi)\\ \cos(1\cdot 2\pi) & \cos(2\cdot 2\pi)\end{bmatrix}\begin{bmatrix}c_1\\ c_2\end{bmatrix} = \begin{bmatrix}f(\pi)\\ f(2\pi)\end{bmatrix}.$$

For this to have a solution it was important that the rows of

$$F_2 = \begin{bmatrix}\cos(1\cdot\pi) & \cos(2\cdot\pi)\\ \cos(1\cdot 2\pi) & \cos(2\cdot 2\pi)\end{bmatrix} = \begin{bmatrix}-1 & 1\\ 1 & 1\end{bmatrix} \tag{2}$$

were independent.

Practice

Solve

$$\begin{bmatrix}-1 & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}c_1\\ c_2\end{bmatrix} = \begin{bmatrix}f(\pi)\\ f(2\pi)\end{bmatrix}.$$

Do you get the same c_1 and c_2 that I did?

To generalise this- to approximate f by periodic functions agreeing at n points- we should find functions r_i such as $\cos x$, $\cos kx$, $\sin x$, $\sin^3(kx) + \cos(2kx)$, with period dividing 2π such that the vectors

$$\left(r_i(1\frac{2\pi}{n}), r_i(2\frac{2\pi}{n}), \ldots, r_i(n\frac{2\pi}{n})\right)$$

are independent.

It seems it could be a nasty bit of business to find such rows, but it is not so bad when we use complex numbers.

Complex numbers

Recall that in the complex plan $\mathbb{C} = \mathbb{R} \times i\mathbb{R}$ a point $a + ib$ can be expressed as

$$re^{i\theta} = r(\cos\theta + i\sin\theta)$$

where $e^{i\theta} = \cos\theta + i\sin\theta$ is the complex exponential function. This notation is exceptionally useful, and is appropriate, as one can show that

$$e^{i\theta}e^{i\theta'} = e^{i(\theta+\theta')}.$$

The number $\omega_n = e^{i\frac{2\pi}{n}}$ is the *primitive n^{th} root of unity*, as one has that

$$\omega_n^n = (e^{i\frac{2\pi}{n}})^n = e^{i2\pi} = \cos 2\pi + i\sin 2\pi = 1.$$

These points are on the unit circle in the complex plane, and two multiply two we simply add their angles from the x axis:

PICTURE

So, for example $\omega_2 = \cos\pi + i\sin\pi = -1$ and $\omega_4 = \cos\pi/2 + i\sin\pi/2 = -1$.

Practice

Show that $\omega_n^d = \omega_n^m$ where $m = d \mod n$.

The Fourier Matrix

The row functions r_α we want are now $r_\alpha(x) = e^{i\alpha x}$; that is, evaluting $e^{i\alpha x}$ at the points $\overline{x} = (1\frac{2\pi}{n}, 2\frac{2\pi}{n}, \ldots n\frac{2\pi}{n})$ we get

$$r_\alpha(\overline{x}) = (e^{i1\alpha\frac{2\pi}{n}}, e^{i2\alpha\frac{2\pi}{n}}, \ldots, e^{in\alpha\frac{2\pi}{n}}) = (\omega_n^{1\alpha}, \omega_n^{2\alpha}, \ldots, \omega_n^{n\alpha}).$$

We claim that these rows are orthogonal, as long as n is even. Indeed,

$$(\omega^i, \omega^{2i}, \ldots, \omega^{ni}) \cdot (\omega^j, \omega^{2j}, \ldots, \omega^{nj}) = \omega^{i+j} + \omega^{2(i+j)} + \cdots + \omega^{n(i+j)}.$$

This sum is a sum of even n vectors even spaced around the unit circle in the complex plane. The i^{th} vector and the $(i+n)^{th}$ cancel out, so this sum is 0.

We make the fourier matrix from these rows. That is, for even n we let F_n be defined by

$$[F_n]_{ij} = \omega_n^{ij}.$$

For example,

$$F_2 = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \qquad F_4 = \begin{bmatrix} \omega_4 & \omega_4^2 & \omega_4^3 & \omega_4^4 \\ \omega_4^2 & \omega_4^4 & \omega_4^6 & \omega_4^8 \\ \omega_4^3 & \omega_4^6 & \omega_4^9 & \omega_4^{12} \\ \omega_4^4 & \omega_4^8 & \omega_4^{12} & \omega_4^{16} \end{bmatrix} = \begin{bmatrix} i & -1 & -i & 1 \\ -1 & 1 & -1 & 1 \\ -i & -1 & i & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Now observe that

$$(\omega^i, \omega^{2i}, \ldots, \omega^{ni}) \cdot (\omega^{-i}, \omega^{-2i}, \ldots, \omega^{-ni}) = 1 + 1 + \cdots + 1 = n.$$

So where $\overline{F}_n$ is defined by $[\overline{F}_n]_{ij} = \omega_n^{-ij}$, we have $F_n\overline{F}_n = nI$. So $F_n^{-1} = \frac{1}{n}\overline{F}_n$.

For example

$$F_4^{-1} = \frac{1}{4}\begin{bmatrix} \omega_4^{-1} & \omega_4^{-2} & \omega_4^{-3} & \omega_4^{-4} \\ \omega_4^{-2} & \omega_4^{-4} & \omega_4^{-6} & \omega_4^{-8} \\ \omega_4^{-3} & \omega_4^{-6} & \omega_4^{-9} & \omega_4^{-12} \\ \omega_4^{-4} & \omega_4^{-8} & \omega_4^{-12} & \omega_4^{-16} \end{bmatrix} = \frac{1}{4}\begin{bmatrix} -i & -1 & i & 1 \\ -1 & 1 & -1 & 1 \\ i & -1 & -i & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Thus for $f(x) = x^2$ say, we can approximate it by computing

$$\frac{1}{4}\begin{bmatrix} -i & -1 & i & 1 \\ -1 & 1 & -1 & 1 \\ i & -1 & -i & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} (\pi/2)^2 \\ (2\pi/2)^2 \\ (3\pi/2)^2 \\ (4\pi/2)^2 \end{bmatrix} \approx \frac{\pi^2}{4} \begin{bmatrix} 3+2i \\ 5/2 \\ 3-2i \\ 15/2 \end{bmatrix}.$$

The function

$$\frac{\pi^2}{4}(3+2i)e^{i\frac{\pi}{2}} + 5/2e^{i2\frac{\pi}{2}} + (3-2i)e^{i3\frac{\pi}{2}} + 15/2e^{i4\frac{\pi}{2}}$$

is 2π periodic and agrees with x^2 at $\pi/2, \pi, 3\pi/2$ and 2π.

Problems from the text

3.5: This section isn't on the test.

4 Determinants

4.1 Introduction

The determinant of a 2×2 matrix $M = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ is

$$\det(M) = |M| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

It is a number with a lot of nice properties. For example, $|M|$ is the volume of the box whose edges are the rows of M. Also $|M| = 0$ if and only if M is singular. If $M \neq 0$ then

$$M^{-1} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} / |M|.$$

In this chapter, we define the determinant for all square matrices and investigate its properties.

Problems from the text

4.1:

4.2 Properties of the Determinant

We have not defined the determinant $\det(M)$ of a square matrix M, except for when M is 2×2. In general, the determinant is a function from the set of $n \times n$ matrices to $\mathbb{R}$ such that the following three properties hold.

1) $\det(I) = \left|\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right| = 1.$

2) A row exchange of M changes the determinant of M by a factor of -1.

3) The determinant is linear in the first row:

$$\begin{vmatrix} ta_1 + a_2 & tb_1 + b_2 & tc_1 + c_2 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = t \begin{vmatrix} a_1 & b_1 & c_1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} + \begin{vmatrix} a_2 & b_2 & c_2 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

Notice that the determinant is **not** linear in the whole matrix, just in the first row.

Practice

Using property 2) show that the determinant is linear in any row of the matrix. Show in particular that if we get M' by multiplying a row of M by r, then $|M'| = r|M|$.

Practice

Show that the above three properties hold for the determinant we defined for 2×2 matrices.

The above three properties are enough to uniquely define the determinant. We now look at some properties that will help us compute it.

4) If two rows of M are equal, then $\det(M) = 0$.

5) The row operation of adding a multiple of one row to another doesn't change the determinat.

6) If M has a row of zeros, then $\det(M) = 0$.

Practice

We now know how the determinant acts under row operations, so we can compute the determinant using Gaussian elimination. Try this with the matrix $M = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 2 \\ 0 & 0 & 1 \end{bmatrix}$.

7) If M is triangular then $\det(M)$ is the product of the diagonal entries.

8) M is singualr if and only if $|M| = 0$.

Practice

Prove properties 4) - 8) using properties 1) - 3). Hint: if 7) seems hard, make sure do the previous practice problem first.

The last two properties we give are a bit more difficult to prove. We will prove them after developing some formulae for the determinant in the next section.

9) The determinant is multiplicative: $|AB| = |A||B|$.

10) $|M| = |M^T|$.

Practice

Prove properties 9) and 10) for 2×2 matrices.

Problems from the text

4.2: 2, 4, 5, 7, 8, 9, 17, 24, 25, 26 (1, 4, 7, 5, 10, 6, 19, 27, 22, 28)

4.3 Formulae for Determinants

We saw how to calculate the determinant by elimination, keeping track of row swaps:

$$\begin{vmatrix} 1 & 2 & 3 \\ 2 & 4 & 4 \\ 1 & 0 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 0 & 0 & -2 \\ 0 & -2 & -1 \end{vmatrix} = -\begin{vmatrix} 1 & 2 & 3 \\ 0 & -2 & -1 \\ 0 & 0 & -2 \end{vmatrix} = -(1 \cdot -2 \cdot -2) = -4.$$

This always works, there are other methods that are useful at times. For a 2×2 matrix $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ we can use the elimination method to get

$$\left|\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right| = \left|\begin{smallmatrix} a & b \\ 0 & d-b\frac{c}{a} \end{smallmatrix}\right| = ad - cb.$$

This is an easy formula to memorise, and quick to use.

Practice

Compute the determinant of $\left[\begin{smallmatrix} 1 & 4 \\ 3 & 1 \end{smallmatrix}\right]$.

We can derive this same formula by linearity:

$$\begin{aligned}
\left|\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right| &= \left|\begin{smallmatrix} a & 0 \\ c & d \end{smallmatrix}\right| + \left|\begin{smallmatrix} 0 & b \\ c & d \end{smallmatrix}\right| \\
&= \left|\begin{smallmatrix} a & 0 \\ c & 0 \end{smallmatrix}\right| + \left|\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix}\right| + \left|\begin{smallmatrix} 0 & b \\ c & 0 \end{smallmatrix}\right| + \left|\begin{smallmatrix} 0 & b \\ 0 & d \end{smallmatrix}\right| \\
&= \left|\begin{smallmatrix} a & 0 \\ c & 0 \end{smallmatrix}\right| + \left|\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix}\right| - \left|\begin{smallmatrix} c & 0 \\ 0 & b \end{smallmatrix}\right| + \left|\begin{smallmatrix} 0 & b \\ 0 & d \end{smallmatrix}\right| \\
&= 0 + ad - bc + 0 \\
&= ad - bc
\end{aligned}$$

What have we done? We have decomposed the matrix into 'summand matrics' with one non-zero entry in each row. Unless there was one non-zero entry in each column, the determinant of the summand was zero. If there was one non-zero entry per column, there determinant was the product of the entries, times a factor of -1 if the columns were not in the right order.

We can do the same for an $n \times n$ matrix. In this decompositon, we get a non-zero summand by choosing a permuation of the columns, and then multiplying the entries down the diagonal. We sum these summands, multiplied by -1 depending on the number of column switches we used.

Observe that for a permutation matrix

$$\begin{vmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{vmatrix} = - \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1.$$

That is, the determinant is $(-1)^s$ if it takes s switches to get to the identity. This yields the *permutation formula for the determinant*:

$$|M| = \sum_P \text{Product of diagonals of } PM \cdot |P|$$

where the sum is over all $n!$ permutation matrices P.

This notation is a bit clumsy. For a permuation matrix P let the *permutation of* P be the vector $(\sigma_P(1), \sigma_P(2), \ldots \sigma_P(n))$ be such that the i^{th} column of P is the $\sigma_P(i)^{th}$ column of I.

Example 4.1. The permuation of $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ is $(1, 3, 2)$.

Where S_m is the set of $n \times n$ permutation matrices, the permuation formula for $M = [m_{i,j}]$ then becomes

$$\det(M) = \sum_{P \in S_n} (\det(P) \prod_{i=1}^{n} m_{i\sigma_P(i)}).$$

Example 4.2.

$$\begin{aligned} \begin{vmatrix} 1 & 2 & 3 \\ 2 & 4 & 4 \\ 1 & 0 & 2 \end{vmatrix} &= \begin{matrix} 1 & 2 & 3 \\ 2 & 4 & 4 \\ 1 & 0 & 2 \end{matrix} \quad - \quad \begin{matrix} 1 & 3 & 2 \\ 2 & 4 & 4 \\ 1 & 2 & 0 \end{matrix} \quad - \quad \begin{matrix} 2 & 1 & 3 \\ 4 & 2 & 4 \\ 0 & 1 & 2 \end{matrix} \quad + \\ & \begin{matrix} 2 & 3 & 1 \\ 4 & 4 & 2 \\ 0 & 2 & 1 \end{matrix} \quad - \quad \begin{matrix} 3 & 2 & 1 \\ 4 & 4 & 2 \\ 2 & 1 & 0 \end{matrix} \quad + \quad \begin{matrix} 3 & 1 & 2 \\ 4 & 2 & 4 \\ 2 & 1 & 0 \end{matrix} \\ &= 8 - 0 - 8 + 8 - 12 + 0 = -4 \end{aligned}$$

This permutation formula can be described in another way, it leads to what is called co-factor expansion formula.

The $(i, j)^{th}$ *minor* of an $n \times n$ matrix M is the $(n-1) \times (n-1)$ matrix M_{ij} that we get from M by removing the i^{th} row and the j^{th} column. The $(i, j)^{th}$ *cofactor* of M is $c_{ij} = (-1)^{i+j} |M_{i,j}|$.

The cofactor formula says that fixing i we get

$$|M| = \sum_{j=1}^{n} c_{ij} m_{ij}$$

or fixing j we get

$$|M| = \sum_{i=1}^{n} c_{ij} m_{ij}.$$

The first equation is called the 'cofactor expansion along the i^{th} row' and the second is the 'cofactor expansion along the j^{th} column'.

Example 4.3. Expanding along the first row we get:

$$\begin{aligned} \begin{vmatrix} 2 & 3 & 4 \\ 1 & 0 & 3 \\ -1 & 3 & 2 \end{vmatrix} &= 2\begin{vmatrix} 0 & 3 \\ 3 & 2 \end{vmatrix} - 3\begin{vmatrix} 1 & 3 \\ -1 & 2 \end{vmatrix} + 4\begin{vmatrix} 1 & 0 \\ -1 & 3 \end{vmatrix} \\ &= 2(-7) - 3(5) + 4(3) = -17 \end{aligned}$$

Practice

Try expanding along the second column.

Practice

True or false:

- The determinant of $S^{-1}MS$ is the same as that of M.
- If $\det(M) = 0$ then at least one of the cofactors of M is 0.
- A matrix whose entries are 0 and 1 has determinant $-1, 0$ or 1.

Practice

Show that

$$\begin{vmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & 0 & * & * \end{vmatrix} = 0$$

where each $*$ is a different real number.

Problems from the text

4.3: 1,3,10,11,13(A,B),15,23,27 (3, 1, 8, 7, 14(A,B), 18, 17, 25)

4.4 Applications of Determinants

A formula for M^{-1}

Observe that for a 2×2 matrix $M = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ we have that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \det(M) I$$

and so we get the formula

$$M^{-1} = \frac{1}{\det M} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Practice

Compute $\begin{vmatrix} 3 & 8 \\ 3 & -1 \end{vmatrix}$.

We can generalise this formula. Recalling that the cofactor c_{ij} of M is $(-1)^{i+j}|M_{ij}|$, the cofactor matrix C of M is $C = [c_{ij}]$.

The cofactor matirx of $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ is

$$C = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}.$$

So this equation above becomes

$$M^{-1} = C^T / \det M. \tag{3}$$

This holds not only for 2×2 matrices, but for all square matrices. Indeed, consider the product

$$MC^T = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} c_{11} & c_{21} & \dots & c_{n1} \\ \vdots & & & \vdots \\ c_{1n} & c_{2n} & \dots & c_{nn} \end{bmatrix}.$$

The diagonal entries of the product are $\det(M)$ by the co-factor expansion formula. The non-diagonal entries are 0 as, say, the entry in the first row and second column is

$$a_{11}c_{21} + a_{12}c_{22} + a_{13}c_{23} + \cdots + a_{1n}c_{2n}.$$

Now what is this? This is the expansion along the first row of a matrix with the same first row of as M, but we are replacing the cofactor c_{1n} with the cofactor c_{2n}. This is the $(i,n)^{th}$ cofactor of the matrix M' that we get from M by replacing its second row with its first. So the whole expression is the determinant of the matrix M', the first two rows of which are the first row of M. Thus the expression is 0.

We thus have

$$MC^T = \det(M)I$$

which proves the formula.

Example 4.4.

$$\begin{bmatrix} 1 & 3 & 2 \\ 0 & 1 & 4 \\ 0 & 0 & 3 \end{bmatrix}^{-1} = \frac{1}{3}\begin{bmatrix} 3 & 0 & 0 \\ -9 & 3 & 0 \\ 10 & -4 & 1 \end{bmatrix}^T = \frac{1}{3}\begin{bmatrix} 3 & -9 & 10 \\ 0 & 3 & -4 \\ 0 & 0 & 1 \end{bmatrix}^T$$

Using this formula isn't always quicker than simply eliminating to a traingular matrix, but sometimes it is.

Cramer's Rule

If $Mx = b$, then $x = M^{-1}b = \frac{1}{\det M}C^T b$. Reading just the i^{th} row of this, we get that

$$x_i = \frac{b_1 c_{i1} + b_2 c_{i2} + \cdots + b_n c_{in}}{\det M}.$$

This is the cofactor expancion along the i^{th} column of the matrix B_i which we get from M by replacing the i^{th} column with the vector b. So we get

$$x_i = \det B_i / detM.$$

This is called Cramer's Rule.

Example 4.5. To solve $\left[\begin{smallmatrix} 1 & 3 \\ 2 & 4 \end{smallmatrix}\right] x = \left[\begin{smallmatrix} 0 \\ 6 \end{smallmatrix}\right]$ by Cramer's rule, we can compute that

$$x_1 = \frac{\left|\begin{smallmatrix} 0 & 3 \\ 6 & 4 \end{smallmatrix}\right|}{\left[\begin{smallmatrix} 1 & 3 \\ 2 & 4 \end{smallmatrix}\right]} = \frac{-18}{-2} = 9$$

$$x_2 = \frac{\left|\begin{smallmatrix} 1 & 0 \\ 2 & 6 \end{smallmatrix}\right|}{\left[\begin{smallmatrix} 1 & 3 \\ 2 & 4 \end{smallmatrix}\right]} = \frac{6}{-2} = -3$$

so $x = (9, -3)$.

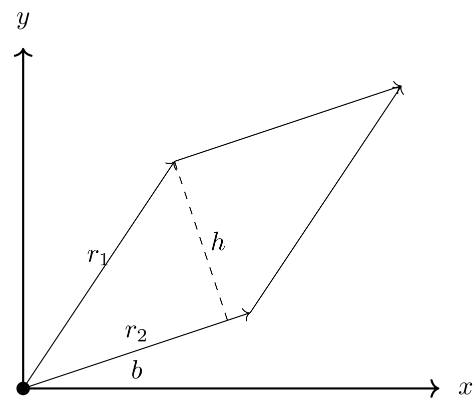
Volume of a box

Given a matrix M with orthogonal rows $b_1, b_2, \dots, b_n$ we have that

$$MM^T = \begin{bmatrix} \|b_1\|^2 & 0 & \dots & 0 \\ 0 & \|b_2\|^2 & \dots & 0 \\ \vdots & & & \vdots \\ 0 & \dots & 0 & \|b_n\|^2 \end{bmatrix},$$

so $\det(MM^T) = \prod \|b_i\|^2 = \prod l_i^2$ where l_i is the length of b_i.

But $\det(MM^T) = \det(M)\det(M^T) = \det(M)^2$ and so $|\det(M)| = \prod l_i$ is the volume of the box whose edges are the vectors b_i. The most wonderful thing you have seen in days, is that this continues to hold when the sides of the box are not orthogonal. Indeed, consider the parallelopiped shown below with edges r_1 and r_2. It has volume $b \cdot h$, where $b = \|r_2\|$ and $h = r_1 - \frac{r_1^T r_2}{r_1^T r_1} r_1$.



We saw that $\det(M)$ doesn't change when we remove a multiple of one row from another, so replacing r_1 with the vector h which is perpendicular to r_2, the determinant is $b \cdot h$ as needed. The same argument works in higher dimensions, so we get the following.

Fact 4.6. *Where M is the matrix whose row vectors are the edges of a parallelopiped P, $|\det(M)|$ is the volume of P.*

Practice

A box has edges from $(0,0,0)$ to $(3,1,1), (1,3,1)$ and $(1,1,3)$. Find its volume.

As the determinant of M is preserved by taking transpose, we can alternately view it as the volume of the figure spanned by the columns of M. Viewing M as a transformation, it transforms the unit cube into this parallelopiped having volume $|\det(M)|$. More generally though, as a transformation, it transforms any solid in $\mathbb{R}^n$.

Let B be the set of points that make up the ball of volume $\frac{4}{3}\pi$:

$$B = \{x \in \mathbb{R}^n \mid \|x\| \leq 1\}.$$

Transforming by M maps this set to the surface of some other solid (an ellipsoid) that has volume $\frac{4}{3}\pi \cdot |\det M|$.

Finding Pivots

Consider eliminating a non-singular matrix M to triangular matrix T without scaling any rows or doing row switches. (So we only do row operations under which the determinant is unchanged.) If we can succesfully eliminate M to triangular T, let the i^{th} diagonal entry of the eliminated matrix be called the i^{th} pivot t_i. Computing determinants, we can find these pivots without eliminating.

Let M_m refer to the square matrix made up of the intersection of the first m rows and m columns of M. Observe that as we eliminate M to T, we are also eliminating M_m to T_m by the same determinant preserving operations, so for each m, we have $\det(T_m) = \det(M_m)$.

So clearly $M_1 = T_1$ has the single entry t_1. The second pivot t_2 is $|T_2|/|T_1| = |M_2|/|M_1|$, and inductively we see that

$$t_i = |M_i|/|M_{i-1}|$$

for all i.

Example 4.7. The pivots of $\begin{bmatrix} 3 & 1 & 4 \\ 1 & 2 & 3 \\ 1 & 3 & 5 \end{bmatrix}$ are

$$\begin{aligned} t_1 &= 3 \\ t_2 &= \frac{\left|\begin{smallmatrix} 3 & 1 \\ 1 & 2 \end{smallmatrix}\right|}{3} = \frac{5}{3} \\ t_3 &= \frac{\left|\begin{smallmatrix} 3 & 1 & 4 \\ 1 & 2 & 3 \\ 1 & 3 & 5 \end{smallmatrix}\right|}{5} = \frac{5}{5} = 1 \end{aligned}$$

Wonderful. As we have to compute the full determinant in the last step, and this essentially requires triangulating the matrix, this seems like a painstaking way to compute these pivots. But is does give us some insight. We see the following fact that will come in handy.

Fact 4.8. *The matrix M can be eliminated to triangular without row exchanges if and only if the leading submatrices $M_1, M_2, \ldots, M_n$ are non-singular.*

Problems from the text

4.4: 1, 3(A), 4, 7, 8, 10, 14, 19, 23, 29, 32 (1, 7(A), 8, 3, 6, 11, 13, 20, 21, 28, 29)

References

[1] Gilbert Strang *Linear Algebra and its applications* Fourth Edition, International Student Edition. 2006 Thompson Learning.