Introduction to linear algebra

Course notes written by Samuel Coskey

Based on material from “Introduction to linear algebra”

by Gilbert Strang

Contents

PART I

Systems of linear equations, vectors, and matrices

§1. Systems of linear equations and vectors

READING. Strang, ILA 5th, §2.1.

What is the equation of a line? The most commonly taught version is $y = mx + b$, and there are many other correct answers to the question. However the *standard form* of a line is $ax + by = c$. This is considered standard because it treats the two variables symmetrically. In this course we will look at generalizations of this simple type of equation.

- We will consider equations with more than two variables x, y. For example if we want to have three variables x, y, z, then the analogous equation is $ax + by + cz = d$. Of course, this is not the equation of a line (what is it?). But due to its similarity we still call it a *linear equation*.
- We will consider situations when we have more than one equation. For example we can have the two equations true simultaneously: $\{ax + by = c;\ dx + ey = f\}$. We call this a *system of linear equations*.

Let's look at a few examples of simple problems which can be modeled by a system of linear equations.

1.1. EXAMPLE (Economics). The New York Times reported that the film "The Interview" had two million online sales, split between rentals and downloads. A rental costs \$6 and a download costs \$15. The total revenue was \$15 million. The NYT could not determine how many transactions were rentals and how many were sales. With a little work, we can set up a system of linear equations.

$$\begin{cases} d + r = 2 \text{ million} \\ 15d + 6r = 15 \text{ million} \end{cases}$$

1.2. EXAMPLE (Modeling flow). Vehicles driving in each segment of a one-way grid are counted during a period of one hour. The results are shown in the figure.

$$
\begin{tikzcd}
 & {} \arrow[d, "450"] & {} & \\
{} & {} \arrow[l, "610"'] \arrow[d, "z"] & {} \arrow[l, "y"'] \arrow[u, "310"'] & {} \arrow[l, "640"'] \\
{} \arrow[r, "520"] & {} \arrow[r, "w"] \arrow[d, "480"] & {} \arrow[u, "x"'] \arrow[r, "600"] & {} \\
 & {} & {} \arrow[u, "390"'] &
\end{tikzcd}
$$

Assume that for each intersection, the number of cars entering it is the same as the number of cars leaving it. What are the values of the unknowns? Again with a little work we can approach this using a system of linear equations.

$$
\begin{cases}
x + 640 = y + 310 \\
y + 450 = z + 610 \\
z + 520 = w + 480 \\
w + 390 = x + 600
\end{cases}
$$

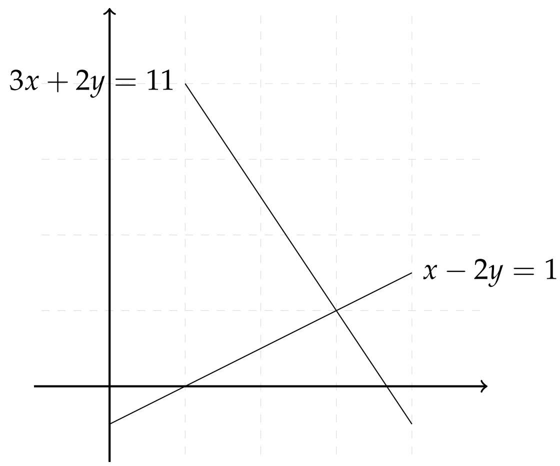
1.3. EXAMPLE (Best fit line). Given sampled data points, what is the equation of a trend line that best describes the overall behavior of the data? For example suppose you are measuring the change in a quantity over time. At time $t = 2$ you measure 3, at time $t = 3$ you measure 5, and at time $t = 6$ you measure 6. We want to find the coefficients m, b of a line $Y = mt + b$ that best fits the data. After a lttle calculus, we arrive at the following.

$$
\begin{cases}
11b + 49m - 57 = 0 \\
\;\;3b + 11m - 14 = 0
\end{cases}
$$

Our short term goal will be to learn how to solve systems like these. But before we begin doing so, let's visualize systems and their solutions geometrically. Consider the following example system:

$$
\begin{cases}
\;\;x - 2y = 1 \\
3x + 2y = 11
\end{cases}
$$

The classical way to visualize the system and its solution is to graph the two equations and highlight the point or points where they intersect.

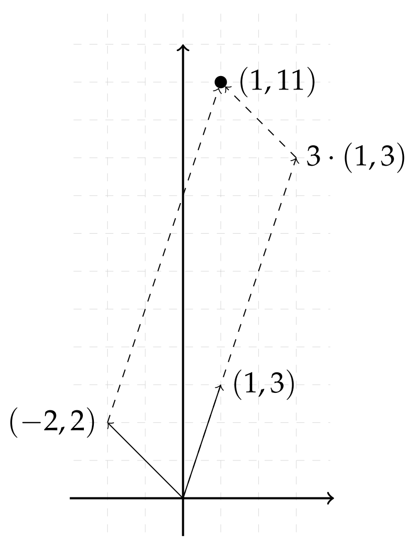


But since the system is linear, there is another very revealing way to visualize the system and its solutions. We will consider the system as a *single equation* with *vector coefficients*.

$$\begin{bmatrix} 1 \\ 3 \end{bmatrix} x + \begin{bmatrix} -2 \\ 2 \end{bmatrix} y = \begin{bmatrix} 1 \\ 11 \end{bmatrix}$$

We should say a little bit about vectors. First, if you are coming from calculus, you may be used to writing vectors horizontally. Now we write them vertically in order to fit our two equations together. Next, recall that we can scale vectors by lenghening and shortening them by a scalar factor, and that we can add vectors "tail to tip".

Viewed this way, the system is asking, how many of the vector $(1, 3)$ and how many of the vector $(-2, 2)$ do we need in order to obtain the vector $(1, 11)$?



In both the classical and vector views, the solution is $x = 3$, $y = 1$.

Activity for §1.

1. Solve the three systems of linear equations from the first lecture. Use any method.

   (a) (costs) $\begin{cases} d + r = 2 \text{ million} \\ 15d + 6r = 15 \text{ million} \end{cases}$

   (b) (traffic flow) $\begin{cases} x + 640 = y + 310 \\ y + 450 = z + 610 \\ z + 520 = w + 480 \\ w + 390 = x + 600 \end{cases}$

   (c) (least squares) $\begin{cases} 11b + 49m - 57 = 0 \\ 3b + 11m - 14 = 0 \end{cases}$

2. In your own words, describe a good method for systematically solving systems of linear equations (such as those in the previous problem).

3. For the following system of linear equations, draw the "row picture (classical)" and the "column picture (vectors)". Then find the solution.

$$\begin{cases} 3x + 2y = 6 \\ 2x + 3y = 6 \end{cases}$$

§2. Elimination

Based on ILA 5th, §2.2.

In the previous class we argued using three examples that it is important to solve systems of linear equations. In this section we will address how to solve systems of linear equations. Of course, there are dozens of valid methods in existence, and no particular method is necessarily right or wrong. For example, many students solved for one variable in one equation, and then plugged the result into the second equation.

In this class we will provide you with an efficient, deterministic method for solving systems of linear equations, called *elimination*. You will be expected to be able to solve systems using this one prescribed method. There are many benefits to using the elimination method; it is efficient and it reveals valuable information about the system. Moreover if all humans and computers agree to use the same exact steps, we can do linear algebra harmoniously with one another and our machines.

We will now illustrate the method with the simple example from the last class.

2.1. EXAMPLE. Consider the system

$$\begin{cases} x - 2y = 1 \\ 3x + 2y = 11 \end{cases}$$

We begin by "eliminating" the $3x$ term. To do this we take the first equation minus three times the second equation, and replace the original second equation with the result. We notate this operation or step by $R_2 - 3R_1 \to R_2$.

$$\begin{cases} x - 2y = 1 \\ \quad\;\; 8y = 8 \end{cases}$$

The elimination is now done in one step. The system is *triangular*, meaning that the only nonzero coefficients are on or above the diagonal. Moreover the new system is equivalent to the old system, in the sense that they have the same solutions.

Having finished elimination, the system can now be easily solved by *back-substitution*. Here we use the last equation to solve for the last variable, the second-to-last to solve for the second-to-last, and so on if there are more equations. The solution is $x = 3$, $y = 1$.

In the previous example, there was only one step of elimination. For larger systems there will be several steps.

2.2. EXAMPLE. Consider the system

$$\begin{cases} 2x + 4y - 2z = 2 \\ 4x + 9y - 3z = 8 \\ -2x - 3y + 7z = 10 \end{cases}$$

Begin by using the upper-left term $2x$ as a *pivot* to eliminate the two terms below it. To do this, we will perform the two operations $R_2 - 2R_1 \to R_2$ and $R_3 + 1R_1 \to R_3$.

$$\begin{cases} 2x + 4y - 2z = 2 \\ \qquad\ \ y + \ \ z = 4 \\ \qquad\ \ y + 5z = 12 \end{cases}$$

Next we move to the second column and use the $1y$ (in the second row) as a pivot to eliminate the term below it. This time we perform the operation $R_3 - 1R_2 \to R_3$.

$$\begin{cases} 2x + 4y - 2z = 2 \\ \qquad\ \ y + \ \ z = 4 \\ \qquad\qquad 4z = 8 \end{cases}$$

The system is now upper triangular, and we can back-solve starting with the last equation and the last variable. The solution is $-1, 2, 2$.

There is one important case when the elimination steps described above cannot be used. This occurs when the term we want to use as a pivot is 0. To get around this issue, we exchange the order of the equations.

2.3. EXAMPLE. In the following system, the first equation has $0x$ in its pivot location, so we will need to exchange it with the second row. We denote this operation $R_1 \leftrightarrow R_2$.

$$\begin{cases} \qquad\ 5y - 7z = 2 \\ 2x + \ \ y + \ \ z = 1 \\ \ \ x - \ \ y + \ \ z = 3 \end{cases}$$

2.4. EXAMPLE. In the following system, there are not as many pivots as there are variables. In this case we can still perform elimination, but we cannot back-solve to get a unique solution. In fact the system has infinitely many solutions.

$$\begin{cases} \ \ x - \ \ y + 2z = 1 \\ 2x + \ \ y - \ \ z = -1 \\ 5x - 2y + 5z = 2 \end{cases}$$

2.5. EXAMPLE. In the following system, there are once again not as many pivots as there are variables. In this case the system has no solution.

$$\begin{cases} \ \ x - \ \ y + 2z = 1 \\ 2x - 2y + \ \ z = 3 \\ -x + \ \ y - \ \ z = -1 \end{cases}$$

Activity for §2.

1. Use the elimination method to reduce each system to an upper triangular one. For each step of the algorithm, write the notation for the step you are performing (*e.g.*, $R_2 + 3R_1 \to R_2$). Then, use back-solving to find a unique solution, or to show the system has infinitely many solutions or no solutions.

   (a) $\begin{cases} 2x - 3y = 5 \\ -4x + 6y = 8 \end{cases}$

   (b) $\begin{cases} 3x - 4y + z = -7 \\ 6x - 8y - z = -16 \\ x + 2y + 2z = 6 \end{cases}$

2. Consider the following system of equations with unknown coefficients a, b, c, and d.

$$\begin{cases} 2x + y + 3z = a \\ 5y + 7z = b \\ dz = c \end{cases}$$

   (a) Choose values of a, b, c, d so that there is no solution. Explain why there is no solution.

   (b) Choose values of a, b, c, d so that there are infinitely many solutions. Explain why there are infinitely many solutions.

3. Look at the matrix A and the vector $\mathbf{b}$. Is it possible to combine the columns of A, using scalar multiples and addition, to get b as a result?

$$A = \begin{bmatrix} 1 & 2 \\ -1 & -1 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

§3. Matrix operations

Based on ILA 5th, §2.4.

When writing systems of linear equations, we rewrite the same variable over and over again. Starting in the next class we will simplify our notation by writing the coefficients as a matrix and the variables as a vector.

However before we dive headlong into new notation, we need to recall the concepts of matrix, vector, and the matrix-times-vector operation. For example, consider the following system.

$$\begin{cases} 3x - 2y + z = 4 \\ 2x + 2y - z = 0 \\ -x + \ \ y + z = 1 \end{cases}$$

The matrix of coefficients of the left-hand-side is

$$\begin{bmatrix} 3 & -2 & 1 \\ 2 & 2 & -1 \\ -1 & 1 & 1 \end{bmatrix}$$

The vector of variables is

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

The vector of scalars on the right-hand-side is

$$\begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}$$

This system can be written in matrix vector form as follows:

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

Here we are relying on the matrix-times-vector operation. Here if A is a matrix and $\mathbf{v}$ is a vector (where the number of columns of A is equal to the number of entries of $\mathbf{v}$), then $A\mathbf{v}$ means to multiply each row of A by $\mathbf{v}$ in the dot product fashion to obtain a new vector $A\mathbf{v}$. For instance

$$\begin{bmatrix} 2 & 3 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 5 \\ 7 \end{bmatrix} = \begin{bmatrix} (2)(5) + (3)(7) \\ (-1)(5) + (2)(7) \end{bmatrix} = \begin{bmatrix} 31 \\ 9 \end{bmatrix}$$

Matrices will be an important feature throughout this course, they are not only a convenient way to represent the coefficients of a system of linear equations. For example, recalling that vectors scaled or added together, matrices can be scaled or added together also. The operations are done component by component. For example:

$$3\begin{bmatrix}2 & 1\\3 & 4\end{bmatrix} = \begin{bmatrix}6 & 3\\9 & 12\end{bmatrix}$$

$$\begin{bmatrix}2 & 1\\3 & 4\end{bmatrix} + \begin{bmatrix}1 & -1\\-2 & 5\end{bmatrix} = \begin{bmatrix}3 & 0\\1 & 9\end{bmatrix}$$

Unlike vectors, matrices can also be multiplied together. Here each row of the first matrix is multiplied by each column of the second matrix, sum-of-products style.

$$\begin{bmatrix}2 & 3\\1 & 5\end{bmatrix}\begin{bmatrix}1 & 4\\2 & -1\end{bmatrix} = \begin{bmatrix}8 & 5\\11 & -1\end{bmatrix}$$

Matrices do not need to be square. We say a matrix is $a \times b$ if it has a rows and b columns. (We *always* write rows $\times$ columns, not the other way around!) When multiplying matrices, the dot products have to match in length. This means the number of columns of the first matrix must match the number of rows of the second matrix.

> **The rule.** if A is $a \times b$ and B is $c \times d$ then to multiply AB we have to have $b = c$, and the result will be $a \times d$.

Here is an example.

$$\begin{bmatrix}2 & 3 & -1\\1 & 1 & 0\end{bmatrix}\begin{bmatrix}1 & 2\\3 & 4\\5 & 6\end{bmatrix} = \begin{bmatrix}6 & 10\\4 & 6\end{bmatrix}$$

Returning to square matrices, the presence of the additon and multiplication operations means that we can do *matrix algebra*. It is natural to ask whether matrxi algebra obeys the same laws as traditional algebra: associativity, commutativity, and distributivity.

Activity for §3. The following problems refer to the 2×2 matrices:

$$A = \begin{bmatrix} 2 & -3 \\ -1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 6 \\ 2 & 2 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

1. Explore whether the associativity law holds by calculating both $(AB)C$ and $A(BC)$. If it worked, was it just a coincidence or would it work for any three matrices?

2. Explore whether the commutativity law holds by calculating both AB and BA, and then both BC and CB. What results to you get?

3. Explore whether the distributivity law holds by calculating both $A(B+C)$ and $AB + AC$. If it worked, was it just a coincidence or would it work for any three matrices?

4. Make up three 3×3 matrices and repeat problems (1)–(3) above to test it out!

5. What 2×2 matrix A satisfies $AB = B$ for all other 2×2 matrices B? What 3×3 matrix A satisfies $AB = B$ for all other 3×3 matrices B?

§4. Elimination using matrices

Based on ILA 5th, §2.3.

We have previously seen that a system of linear equations can be viewed as a single matrix equation $A\mathbf{x} = \mathbf{b}$. For example, consider the following system

$$\begin{cases} 2x + 4y - 2z = 2 \\ 4x + 9y - 3z = 8 \\ -2x - 3y + 7z = 10 \end{cases}$$

We can rewrite it as

$$\begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix}$$

In this section we will see that the row elimination process is just multiplying both sides by a matrix.

4.1. EXAMPLE. Continuing the example from above, the first step of elimination will be $R_2 - 2R_1 \to R_2$. We can carry this out by multiplying both sides (on the left) by an "elimination matrix":

$$\begin{bmatrix} 1 & & \\ -2 & 1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & -2 \\ 4 & 9 & -3 \\ -2 & -3 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 & & \\ -2 & 1 & \\ & & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 8 \\ 10 \end{bmatrix}$$

Resulting in:

$$\begin{bmatrix} 2 & 4 & -2 \\ 0 & 1 & 1 \\ -2 & -3 & 7 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 10 \end{bmatrix}$$

Thus carrying out $R_2 - 2R_1 \to R_2$.

As we have seen, every step of row elimination has a corresponding elimination matrix. Here are the rules

- The step $R_i + \ell R_j \to R_i$ has the matrix with 1's down the diagonal and an ℓ in position (i, j).
- The step $R_i \leftrightarrow R_j$ has a matrix with 1's down the diagonal except row i has a 1 in position j and row j has a 1 in position i.
- The (so far unused) step $kR_i \to R_i$ has a matrix with 1's down the diagonal except for a k in position (i, i).

This means elimination can be viewed as successive multiplication by elimination matrices:

$$E_k \cdots E_2 E_1 A\mathbf{x} = E_k \cdots E_2 E_1 \mathbf{b}$$

Until the left hand side is in its echelon (or upper triangular) form:

$$U\mathbf{x} = \mathbf{b}_{\text{new}}$$

4.2. EXAMPLE. Continuing the previous example, we next apply the elimination matrices

$$E_1 = \begin{bmatrix} 1 & & \\ & 1 & \\ 1 & & 1 \end{bmatrix}, \; E_2 = \begin{bmatrix} 1 & & \\ & 1 & \\ & -1 & 1 \end{bmatrix}$$

Note that since we do the same operation to the left and right sides, we often write them both in one combined *augmented matrix*. This is the final and most compact way to write a system of linear equations.

4.3. EXAMPLE. Consider the system:

$$\begin{cases} x - 2y + 2z = 1 \\ -3x + \ y + \ z = -1 \\ -2x + 2y + \ z = 1 \end{cases}$$

The augmented form of the system is:

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

To solve this we can apply the following elimination matrices.

$$E_1 = \begin{bmatrix} 1 & & \\ 3 & 1 & \\ & & 1 \end{bmatrix}, \; E_2 = \begin{bmatrix} 1 & & \\ & 1 & \\ 2 & & 1 \end{bmatrix}, \; E_3 = \begin{bmatrix} 1 & & \\ & 1 & \\ & -\frac{2}{5} & 1 \end{bmatrix}$$

Since the augmented matrix hides the variables, it is easy to forget that it represents a system of linear equations. If you are confused, try labeling each column with the corresponding variable name!

So far we have only used the first type of elimination step, and the first type of elimination matrix. The following example has a row exchange.

4.4. EXAMPLE. Consider the following augmented matrix.

$$\left[\begin{array}{ccc|c} 2 & 3 & 1 & 3 \\ 2 & 3 & 2 & 5 \\ 4 & 1 & 3 & 3 \end{array}\right]$$

To solve this we can apply the following elemination matrices.

$$E_1 = \begin{bmatrix} 1 & & \\ -1 & 1 & \\ & & 1 \end{bmatrix}, \; E_2 = \begin{bmatrix} 1 & & \\ & 1 & \\ -2 & & 1 \end{bmatrix}, \; E_3 = \begin{bmatrix} 1 & & \\ & & 1 \\ & 1 & \end{bmatrix}$$

Activity for §4.

1. Write down the 4×4 matrices corresponding to the row operations:
   (a) $R_2 - 5R_1 \to R_2$
   (b) $R_4 + 7R_2 \to R_4$
   (c) $R_2 \leftrightarrow R_4$

2. For each system of equations, write the augmented matrix version of the system and use elimination to solve it. In each step, write the elimination matrix you used.
   (a) $\begin{cases} 2x - 3y = 5 \\ \ \ x - \ \ y = 2 \end{cases}$

   (b) $\begin{cases} \qquad v + 2w = 0 \\ u - v - \ \ w = -1 \\ 2u + v \qquad = 4 \end{cases}$

3. Consider the 3×3 row operations $R_2 - 2R_1 \to R_2$ (op1) and $R_3 + R_1 \to R_3$ (op2). Write down an elimination matrix for each operation. Can you find a single matrix which performs both operations together?

   Now add the third operation $R_3 + 2R_2 \to R_3$ (op3). Can you find a single matrix which performs all three operations in order?

§5. LU factorization

Based on ILA 5th, §2.6.

We begin by recalling the method of elimination matrices from the previous section. Given a system of linear equations, $A\mathbf{x} = \mathbf{b}$ we left-multiply the equation by a sequence of elmination matrices E_i as follows.

$$E_k \cdots E_2 E_1 A\mathbf{x} = E_k \cdots E_2 E_1 \mathbf{b}$$

At the end of the process, the left hand side is in its echelon (or upper triangular) form. In other words, we write the above equation as

$$U\mathbf{x} = E_k \cdots E_2 E_1 \mathbf{b}$$

In this section we will use matrix algebra to "divide" the $E_i's$ back over to the left-hand side. While division doesn't necessarily make sense in matrix algebra, it turns out that the elementary matrices E_i always have an "inverse matrix" E_i^{-1} which *undoes* the effect of E_i. For example, the following matrices undo one another.

$$E = \begin{bmatrix} 1 & \\ 3 & 1 \end{bmatrix}, \quad E^{-1} = \begin{bmatrix} 1 & \\ -3 & 1 \end{bmatrix}$$

This allows us to rewrite the previous equation as

$$E_1^{-1} E_2^{-1} \cdots E_k^{-1} U\mathbf{x} = \mathbf{b}$$

For the rest of this section, let's assume there are *no row exchanges* needed in the elimination. Then all of the matrices E_i^{-1} are *lower triangular*, meaning all entries above the main diagonal are 0. We thus arrive at the following conclusion. We can *factor* A as a product LU and write the original system as

$$LU\mathbf{x} = \mathbf{b}$$

where $L = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$ and U is the eliminated matrix.

There are two reasons to do this. First, just as in ordinary arithmetic, it is often useful to factor something as a product of simpler things. Second, it shows the system is a composition of two triangular systems, which are much easier to solve than general systems.

5.1. EXAMPLE. Factor the matrix $A = LU$.

$$A = \begin{bmatrix} 2 & 4 & 2 \\ 1 & 5 & 2 \\ 4 & -1 & 9 \end{bmatrix}$$

The elimination matrices are

$$E_1 = \begin{bmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ & & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 1 & & \\ & 1 & \\ -2 & & 1 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & & \\ & 1 & \\ & 3 & 1 \end{bmatrix}$$

The inverse matrices are

$$E_1^{-1} = \begin{bmatrix} 1 & & \\ \frac{1}{2} & 1 & \\ & & 1 \end{bmatrix} \quad E_2^{-1} = \begin{bmatrix} 1 & & \\ & 1 & \\ 2 & & 1 \end{bmatrix} \quad E_3^{-1} = \begin{bmatrix} 1 & & \\ & 1 & \\ & -3 & 1 \end{bmatrix}$$

We can combine these together to get L. The final answer is

$$A = \begin{bmatrix} 1 & & \\ \frac{1}{2} & 1 & \\ 2 & -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ & 3 & 1 \\ & & 8 \end{bmatrix}$$

Finally we can confirm that $A = LU$.

Next if we have a factored system $LU\mathbf{x} = \mathbf{b}$ we can use the substitution $\mathbf{c} = U\mathbf{x}$ to break it into two triangular systems: $L\mathbf{c} = \mathbf{b}$ and then $U\mathbf{x} = \mathbf{c}$. We can solve the first using "forth-solving" and the the second using back-solving.

5.2. EXAMPLE. Consider the system

$$\begin{cases} 2x + 4y + 2z = 10 \\ \;\; x + 5y + 2z = 11 \\ 4x - \;\; y + 9z = 2 \end{cases}$$

By the work we have done previously, the system can be broken into two:

$$\begin{cases} c \qquad\qquad\quad = 10 \\ \frac{1}{2}c + \;\; d \qquad = 11 \\ 2c - 3d + e = 2 \end{cases} \qquad \text{and} \qquad \begin{cases} 2x + 4y + 2z = c \\ \qquad\;\; 3y + \;\; z = d \\ \qquad\qquad\quad 8z = e \end{cases}$$

The first can easily be solved forwards with solution $c = 10$, $d = 6$, $e = 0$. We then plug these values into the second to solve backwards with solution $x = 1$, $y = 2$, $z = 0$.

We close this section with the really cool part. Suppose the matrix A models a given phenomenon, and you want to solve $A\mathbf{x} = \mathbf{b}$ over and over again for many different $\mathbf{b}$'s. If we had to perform elimination each time, we would be spending $O(n^3)$ operations each time. If on the other hand we factor $A = LU$ once, then we can reduce every future instance to backsolving twice, and this takes only $O(n^2)$ operations!

Activity for §5.

1. Let A be the following matrix.

$$\begin{bmatrix} 2 & -3 \\ 4 & 3 \end{bmatrix}$$

Find the elimination matrix E that puts A in upper triangular form U. Let L be the inverse matrix E^{-1} that "undoes" E. Write $A = LU$ and confirm that it is true.

2. Let A be the following matrix.

$$\begin{bmatrix} 1 & 1 & 1 \\ 2 & 4 & 5 \\ 0 & 4 & 0 \end{bmatrix}$$

(a) Use elimination to put A into upper triangular form U. Keep track of the elimination matrices E_1, E_2, E_3. (Note in this case E_2 is just the identity.)

(b) Write the inverse matrices $E_1^{-1}, E_2^{-1}, E_3^{-1}$ that undo the elimination matrices.

(c) Find the combined matrix $L = E_1^{-1}E_2^{-1}E_3^{-1}$.

(d) Write $A = LU$ and confirm that it is true.

(e) Consider the system

$$\begin{cases} x + \ y + \ z = 1 \\ 2x + 4y + 5z = 5/2 \\ \quad 4y \qquad = -2 \end{cases}$$

Use your factorization $A = LU$ from part (d) to write two systems, $L\mathbf{c} = \mathbf{b}$ and $U\mathbf{x} = \mathbf{c}$.

(f) Solve the system $L\mathbf{c} = \mathbf{b}$ by forth-solving. Solve the system $U\mathbf{x} = \mathbf{c}$ by back-solving. Check that your solution for $\mathbf{x}$ is correct by plugging it into the original system $A\mathbf{x} = \mathbf{b}$.

§6. Inverse matrices

Based on ILA 5th, §2.5.

In the previous section we have seen that while we cannot necessarily divide matrices, elimination matrices E do have an "inverse" E^{-1} which undo the effect of E. In this section we will see that most square matrices A have an inverse A^{-1} which undoes multiplication by A.

The official definition of A^{-1} is that it is a multiplicative inverse for A. First recall that I denotes the identity matrix: the matrix with the same size as A, but with 1's down the diagonal and 0's everywhere else. The identity matrix is the multiplicative identity which means that $AI = IA = A$.

6.1. DEFINITION. If A is any square matrix, its *inverse* is the matrix A^{-1} such that $AA^{-1} = A^{-1}A = I$.

6.2. EXAMPLE. We have already seen that elimination matrices A have an inverse. Let A be the following matrix

$$A = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$$

Then we have

$$A^{-1} = \begin{bmatrix} 1 & 0 \\ -3 & 1 \end{bmatrix}$$

We can verify that $AA^{-1} = I$.

6.3. EXAMPLE. Other types of matrices have inverses too.

$$A = \begin{bmatrix} 2 & 5 \\ 1 & 3 \end{bmatrix}, \qquad A^{-1} = \begin{bmatrix} 3 & -5 \\ -1 & 2 \end{bmatrix}$$

Check that the two matrices are inverses.

Not every matrix has an inverse. In the rest of the section we will talk about how to find out if a matrix has an inverse, and if it does, how to calculate what the inverse is.

Recall that we can use elimination matrices to make a given matrix upper triangular:

$$E_k \cdots E_1 A = U$$

This zeros out the entries below the pivots. The idea is to *continue elimination* to zero out the entries above the pivots as well. We can also ensure the pivots themselves are all 1 leaving only the identity matrix.

$$E_m \cdots E_k \cdots E_1 A = I$$

The inverse of A is now observed to be the product of all the elimination matrices.

In order to keep a running track of the product of the elimination matrices, we begin with a (super) augmented matrix containing A and the identity I. We then perform a side-by-side elimination.

$$\begin{gathered}
[A \mid I] \\
[E_1 A \mid E_1] \\
\vdots \\
[E_m \cdots E_1 A \mid E_m \cdots E_1] \\
[I \mid A^{-1}]
\end{gathered}$$

When the left side is finally transformed to I, the inverse appears in the right side. In order to perform this reduction, we need to eliminate entries below *and above* the pivots, and finally clear the pivots to 1.

6.4. EXAMPLE. Let A be the following matrix.

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

Begin with the matrix

$$\left[\begin{array}{ccc|ccc} 1 & 4 & 3 & 1 & & \\ -1 & -2 & 0 & & 1 & \\ 2 & 2 & 3 & & & 1 \end{array}\right]$$

The elimination results in

$$\left[\begin{array}{ccc|ccc} 1 & & & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ & 1 & & \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ & & 1 & \frac{1}{6} & \frac{1}{2} & \frac{1}{6} \end{array}\right]$$

We may confirm this is A^{-1} by multiplying AA^{-1} and getting I.

It is important to note that not every matrix A has an inverse matrix! First of all the matrix has to be square. Next, in order for the method to work, we have to have a (nonzero) pivots in every diagonal position. A square matrix with no inverse is called *singular*.

6.5. EXAMPLE. The following matrix is singular.

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 4 \\ 2 & 4 & 7 \end{bmatrix}$$

Activity for §6.

1. Find the inverse of each matrix, or else show that it is singular.
   (a) $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$
   (b) $\begin{bmatrix} \frac{1}{2} & 3 \\ 2 & 12 \end{bmatrix}$
   (c) $\begin{bmatrix} 1 & 0 & 3 \\ 2 & 3 & 4 \\ 1 & 0 & 2 \end{bmatrix}$

2. Find the inverse of the matrix (treat the letters a, b as variables):
$$\begin{bmatrix} 1 & a & 0 \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix}$$

3. Assume A is an invertible $n \times n$ matrix and consider the system $A\mathbf{x} = \mathbf{b}$. Is it possible for the system to have no solution? Is it possible for the system to have infinitely many solutions?

§7. Column space

Based on ILA 5th, §3.1.

In previous sections we have seen how to use elimination to decide whether a system of linear equations has zero, one, or infinitely many solutions. In this section we will explore the theory behind whether and when a system of linear equations has any solutions or not.

We will still be viewing a system of linear equations as a single matrix equation

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

If we view A as a function mapping inputs $\mathbf{x}$ to outputs $\mathbf{b}$, then observe that $A\mathbf{x} = \mathbf{b}$ has solutions if and only if $\mathbf{b}$ is in the *range* of A. This means we have to study the concept of range for matrices. Before we do that, we will need to introduce vector spaces. Recall that the key fact about vectors is that they can be scaled and added.

7.1. DEFINITION. A *vector space* is any collection of mathematical objects that may be scaled and added.

For example, let $\mathbb{R}^n$ denote the space of real vectors with n many components. Thus $\mathbb{R}^1$ is the real line, $\mathbb{R}^2$ is the Euclidean plane, $\mathbb{R}^3$ is real 3-space, and so on. Since we have seen that real vectors may be scaled and added, $\mathbb{R}^n$ is a vector space. There are many other examples of vector spaces includeing polynomials, continuous functions, differentiable functions, matrices, and more. But for this class we care exclusively about the vector spaces $\mathbb{R}^n$.

Well that is not exactly true. The spaces $\mathbb{R}^n$ also contain interesting subspaces, and these are what we will need to study to better understand the range of a matrix, and when matrix equations have solutions.

In order to explain what a subspace is, we first make the following definition: if $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are any vectors then a *combination* of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is any vector that can be obtained by scaling and adding them.

7.2. DEFINITION. A *subspace* of $\mathbb{R}^n$ is a set of vectors S with the property that all combinations of vectors in S also lie in S.

Though this definition may seem technical, if we think about subspaces geometrically, it turns out that all subspaces of $\mathbb{R}^n$ have the shape of a point, line, plane, etc, passing through the origin.

7.3. EXAMPLE. Consider the subset of $\mathbb{R}^3$:

$$S = \{\, (x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0 \,\}$$

We may verify that S is a subspace of $\mathbb{R}^3$ by checking that scaling and adding vectors in S_1 results in vectors that are also in S_1. Note that this subspace of $\mathbb{R}^3$ has the shape of a plane.

7.4. EXAMPLE. Consider the set

$$S = \{ (x, y) \in \mathbb{R}^2 \mid 2x + y = 1 \}$$

We can see that this set S is *not* a subspace of $\mathbb{R}^2$ by finding two vectors in S whose sum is not in S..

The most important example of a subspace is the range of a matrix, called its column space.

7.5. DEFINITION. If A is any matrix with column vectors $\mathbf{v}_1, \ldots \mathbf{v}_n$, then the *column space* of A consists of all combinations $\mathbf{v}_1 x_1 + \cdots + \mathbf{v}_n x_n$ of the column vectors of A.

Thus the column space of A is precisely the set of of all vectors $\mathbf{b}$ such that $A\mathbf{x} = \mathbf{b}$ has a solution. The column space of A is the range of A.

7.6. EXAMPLE. Consider the matrix:

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

The column space of A is a plane consisting of all combintations of the two column vectors. We can test whether the vector $(-1, -1, 1)$ lies in the column space of A by deciding whether the system is consistent:

$$\left[\begin{array}{cc|c} 1 & 0 & -1 \\ 4 & 3 & -1 \\ 2 & 3 & 1 \end{array}\right]$$

Elimination yields

$$\left[\begin{array}{cc|c} 1 & 0 & -1 \\ 0 & 3 & 3 \\ 0 & 0 & 0 \end{array}\right]$$

Thus the vector $(-1, -1, 1)$ lies in the column space of A, as witnessed by the values $(x, y) = (-1, 1)$.

In the above example we checked whether a single vector $\mathbf{b}$ is in the column space of A. More generally we can use algebra to decide which vectors $\mathbf{b}$'s are in the column space of A.

7.7. EXAMPLE. Let A be the matrix below. Find the column space of A. The answer consists of one or more equations in the variables b_1, b_2, b_3 that decide when $A\mathbf{x} = \mathbf{b}$ has a solution.

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

Using elimination, we find that the system is consistent precisely when $b_3 - 2b_1 = 0$.

Activity for §7.

1. For each matrix A, find the column space of A. The answer consists of one or more equations in the variables b_1, b_2, b_3 that decide when $A\mathbf{x} = \mathbf{b}$ has a solution.
   (a) $A = \begin{bmatrix} 1 & 4 & 2 \\ 2 & 8 & 4 \\ -1 & -4 & -2 \end{bmatrix}$

   (b) $A = \begin{bmatrix} 1 & 4 \\ 2 & 9 \\ -1 & -4 \end{bmatrix}$

   (c) $A = \begin{bmatrix} -2 & 1 & 4 \\ 0 & -5 & -4 \\ 1 & -3 & -4 \end{bmatrix}$

2. Decide whether each of the following is true or false. Explain your answers!
   (a) The matrix A and $2A$ have the same column space
   (b) Given a matrix A, the set of vectors $\mathbf{b}$ such that $A\mathbf{x} = \mathbf{b}$ has a solution forms a subspace
   (c) Given a matrix A and a vector $\mathbf{b}$, the set of vectors $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$ forms a subspace

§8. Null space: the complete solution to $A\mathbf{x} = \mathbf{0}$

Based on ILA 5th, §3.2.

In the last section we introduced vector spaces, along with our primary example, the column space of a matrix. As it turns out, there are several more important subspaces that are associated with a given matrix A (see the cover of the text!). In this session we introduce the null space.

8.1. DEFINITION. Given a matrix A, the *null space* of A is the set of all $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$.

Thinking of A as a function, the column space of A is the range, and the null space is the set of zeros or roots.

Note that if A is $m \times n$, then the null space of A is a subset of $\mathbb{R}^n$. It is important for us to check that the null space is really a subspace of $\mathbb{R}^n$, that is, closed under scaling and addition: if $A\mathbf{x} = 0$ then $A\alpha\mathbf{x} = \alpha A\mathbf{x} = \mathbf{0}$ too; if $A\mathbf{x} = \mathbf{0}$ and $A\mathbf{w} = \mathbf{0}$, then we have $A(\mathbf{x} + \mathbf{w}) = A\mathbf{x} + A\mathbf{w} = \mathbf{0}$ too.

We want to find the null space of a given matrix. We begin with an example.

8.2. EXAMPLE. Consider the following matrix.

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 4 & 8 \\ 3 & 6 & 10 \end{bmatrix}$$

As usual the first step is to perform elimination.

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix}$$

Next identify the variables whose column has no pivot, called *free variables*, and the variables whose column has a pivot, called *basic variables*. Then solve for the basic variables *in terms of* the free variables.

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -2y \\ y \\ 0 \end{bmatrix}$$

When we find the null space, we can achieve some additional clarity if we eliminate both below and above the pivots, and also clear the pivots to 1s. This is called the *reduced row echelon form* (RREF) of the matrix.

8.3. EXAMPLE. Consider the following matrix.

$$A = \begin{bmatrix} 1 & 2 & 2 & 4 \\ 3 & 8 & 6 & 16 \end{bmatrix}$$

The first step is to eliminate all the way to RREF.

$$A = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix}$$

Then solve for the basic variables in terms of the free variables.

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} -2z \\ -2w \\ z \\ w \end{bmatrix}$$

Even better, by factoring out the free variables, we can write the solution in *parametric vector form*. The resulting numeric vectors are called the *special solutions*, and the free variables are the parameters.

$$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = \begin{bmatrix} -2 \\ 0 \\ 1 \\ 0 \end{bmatrix} z + \begin{bmatrix} 0 \\ -2 \\ 0 \\ 1 \end{bmatrix} w$$

Since the parameters are in some sense unnecessary information, we often write just the *null space matrix* which is the matrix whose columns are the special solutions.

$$N = \begin{bmatrix} -2 & 0 \\ 0 & -2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

It is called the null space matrix because $AN = 0 =$ the zero matrix.

The special solutions $A\mathbf{x} = \mathbf{0}$ each provide a recipe for a combination of the columns of A that gives $\mathbf{0}$.

8.4. EXAMPLE. Calculate the null space matrix of A.

$$A = \begin{bmatrix} 1 & 2 & 2 & 3 \\ 2 & 4 & 8 & 10 \\ 3 & 6 & 10 & 13 \end{bmatrix}$$

The solution is

$$N = \begin{bmatrix} -2 & -1 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{bmatrix}$$

Activity for §8.

1. For each system, set up an augemented matrix and eliminate to RREF. Then write the solution as a combination of special solutions.
   (a) $\begin{cases} 2x + 3y - z = 0 \\ x \qquad + z = 0 \end{cases}$
   (b) $\begin{cases} x + 3y + 3z = 0 \\ 2x + 6y + 9z = 0 \\ -x - 3y + 3z = 0 \end{cases}$

2. Eliminate each matrix to RREF, and find the null space matrix N. (The columns of N are the special solutions.) Check that $AN = 0$.
   (a) $A = \begin{bmatrix} 1 & 0 & 2 & -3 \\ -2 & 1 & 0 & -7 \end{bmatrix}$
   (b) $A = \begin{bmatrix} 1 & 2.5 & 1 & 0 \\ 2 & 5 & 3 & -2 \end{bmatrix}$
   (c) $A = \begin{bmatrix} 2 & 4 & 2 \\ 0 & 4 & 4 \\ -2 & 4 & 6 \end{bmatrix}$

3. Consider the modification to the system in 1(b):
$$\begin{cases} x + 3y + 3z = 2 \\ 2x + 6y + 9z = 7 \\ -x - 3y + 3z = 4 \end{cases}$$
   (a) Using ordinary elimination and back-solving, find *one* solution to this system.
   (b) Add your solution for 3(a) to one of the special solutions from 1(b). Check that the result is a solution to this system too. How can you explain this?

§9. The complete solution to $A\mathbf{x} = \mathbf{b}$

Based on ILA 5th, §3.3.

In the previous section we showed how to find the complete solution to a system of linear equations $A\mathbf{x} = \mathbf{0}$. A system with $\mathbf{0}$ on the right-hand side is called a *homogeneous* system, just as in differential equations. We will now use this as a stepping stone to solving a general system of linear equations, $A\mathbf{x} = \mathbf{b}$.

Recall that the solutions to $A\mathbf{x} = \mathbf{0}$ all lie in the *null space*, which is a vector subspace of $\mathbb{R}^n$. It turns out that solutions to the system $A\mathbf{x} = \mathbf{0}$ all lie in a *translate* of the null space.

To get the idea, let's look at the case of one equation in two variables. For example, first consider the *homogeneous* equation:

$$2x + 3y = 0$$

The solutions to this equation form a line through the origin, a subspace of $\mathbb{R}^2$. We can plot it: the line through the origin with slope $-2/3$.

Next consider an *inhomogeneous* equation with the same left-hand side:

$$2x + 3y = 3$$

The solutions to this equation form a line, but not through the origin. It is the same line as above but shifted to the right by $3/2$ (or up by 1).

We can redo this problem in matrix notation. We are using the matrix $A = \begin{bmatrix} 2 & 3 \end{bmatrix}$. First we find the general solution to $A\mathbf{x} = \mathbf{0}$, that is, the null space of A:

$$\mathbf{x}_{\text{null}} = \begin{bmatrix} -3/2 \\ 1 \end{bmatrix} y$$

The null space is the line through the origin that we found above. The general solution to $A\mathbf{x} = [3]$ is then:

$$\mathbf{x}_{\text{general}} = \begin{bmatrix} 3/2 \\ 0 \end{bmatrix} + \begin{bmatrix} -3/2 \\ 1 \end{bmatrix} y$$

which is the line translated away from the origin that we found above.

In general, we will find that the general solution to a system of linear equations $A\mathbf{x} = \mathbf{b}$ will be equal to a single vector (the particular solution) added to the vectors in the null space (the combinations of the special solutions). This strongly mirrors the method from systems of linear differential equations.

9.1. EXAMPLE. Consider the system $A\mathbf{x} = \mathbf{b}$:

$$\begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 4 \\ 1 & 3 & 1 & 6 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 6 \\ 7 \end{bmatrix}$$

- Step one. Eliminate the full system to RREF.

$$\left[\begin{array}{cccc|c} 1 & 3 & 0 & 2 & 1 \\ & & 1 & 4 & 6 \\ & & & & \end{array}\right]$$

- Step two. Find the null space. Forget about the right-hand side briefly, pretend it is zeros, and find the null space of A in *parametric vector form*.
- Step three. Find the particular solution to the original system by setting the free variables to zero. Note there is also a "trick" here using the right-hand side of RREF.
- Step four. Write the solution; it is just the sum of the particular and general solutions.

$$\mathbf{x}_{\text{general}} = \begin{bmatrix} 1 \\ 0 \\ 6 \\ 0 \end{bmatrix} + \begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} x_2 + \begin{bmatrix} -2 \\ 0 \\ -4 \\ 1 \end{bmatrix} x_4$$

Activity for §9.

1. For each system, solve or state that it is inconsistent. If the system is consistent, find the null space in parametric vector form, the particular solution, and finally write the general solution.

   (a) $\begin{cases} 2x + 3y - z = 1 \\ x \qquad + z = 1 \end{cases}$

   (b) $\begin{cases} x + 3y + 3z = 1 \\ 2x + 6y + 9z = 5 \\ -x - 3y + 3z = 5 \end{cases}$

   (c) $\begin{bmatrix} 1 & 3 & 1 & 2 \\ 2 & 6 & 4 & 8 \\ 0 & 0 & 2 & 4 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$

2. Suppose that A is $m \times n$ and has r many pivots. Assume you know the system $A\mathbf{x} = \mathbf{b}$ is consistent. What is the dimension of the set of solutions?

PART II

Subspaces, orthogonality, and applications

§10. Independence and basis

Based on ILA 5th, §3.4.

Recall that the column space of a matrix A consists of all combinations of the columns of A. Sometimes we say the columns are a *spanning set* for the column space. The idea is if I want to tell you about a plane, I only need to tell you two vectors in the plane (a spanning set) and not every vector in the plane. You can deduce the rest using combinations.

We have seen that sometimes not all the columns are necessary to form a spanning set. For example, consider the matrix:

$$A = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 0 & 2 \\ 3 & 1 & 5 \end{bmatrix}$$

Notice that the third column of A is a combination of the first two. The first two columns already form a spanning set, making the third column redundant. Geometrically, the three columns of A point in three distinct directions, but never manage to break outside of a plane.

10.1. DEFINITION. We say that a set of vectors is *linearly dependent* if one of the vectors is a combination of the others. Otherwise we say the set is *linearly independent*. If V is a vector space, then a linearly independent spanning set for V is called a *basis* for V.

A basis for a vector space has "just the right number" of vectors. A basis for a plane will have two vectors, for a three-dimensional vector space will have three vectors, etc. In fact we can take the definition of *dimension* as the number of vectors in a basis.

Turning to practical matters, suppose we are given a matrix A, and asked to find bases for its column space and null space. Beginning with the column space, the goal is to remove the redundant columns until we are left with the right number, the basis. There is a simple rule to determine which columns should be in the basis and which columns are redundant: The basic (non-free) variable columns belong in the basis, and the free variable columns are redundant.

10.2. EXAMPLE. Consider the following matrix:

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

The RREF is as follows:

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

The RREF shows exactly why the free variable columns are redundant: the entries reveal how they are combinations of earlier columns. The basis for the column space consists of the basic variable columns from *the original matrix*.

$$\text{basis for } C(A)\text{:} \quad \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ -1 \end{bmatrix}$$

We also wish to find a basis for the null space of A. In fact we have already done this: the special solutions are the basis.

10.3. EXAMPLE. The basis for the null space of the matrix in the previous problem consists of the *special solutions*:

$$\text{basis for } N(A)\text{:} \quad \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ -4 \\ 1 \end{bmatrix}$$

To summarize: To find a basis for the column space of A, we eliminate to REF, identify the pivot columns, and select those columns from the original matrix A. To find a basis for the null space of A, we find the special solutions or null space matrix.

We can apply the same reasoning to remove the redundant elements from any set of vectors, and find a linearly independent subset with the same span.

10.4. EXAMPLE. Find a basis for the subspace of $\mathbb{R}^3$ spanned by the given vectors. What is the dimension of the subspace?

$$\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

We can squash these vectors into the columns of a matrix and find the RREF:

$$R = \begin{bmatrix} 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 1 \\ & & & \end{bmatrix}$$

The basic columns are $1, 4$, so the basis consists of the first and fourth vector:

$$\text{basis:} \quad \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

Activity for §10.

1. Find a basis for the null space and column space of each matrix.
   (a) $A = \begin{bmatrix} 1 & -1 & 3 \\ -2 & 2 & -6 \end{bmatrix}$
   (b) $A = \begin{bmatrix} 1 & 2 & 4 \\ 0 & 1 & -2 \end{bmatrix}$
   (c) $A = \begin{bmatrix} 1 & 3 & 0 & -1 \\ 2 & 6 & 1 & -4 \\ 1 & 3 & -2 & 3 \end{bmatrix}$
   (d) $A = \begin{bmatrix} 2 & 0 & -1 & 4 & 2 \\ 4 & 1 & 0 & 2 & -1 \\ 2 & 1 & 1 & -2 & -3 \end{bmatrix}$

2. Determine whether the set of vectors is independent or dependent. If it is dependent, eliminate the (and only the) redundant vectors to find an independent subset.
   (a) $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \quad \begin{bmatrix} 2 \\ 3 \\ 6 \end{bmatrix}$
   (b) $\begin{bmatrix} 2 \\ 0 \\ 0 \\ 4 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ 6 \\ 0 \\ 6 \end{bmatrix}, \quad \begin{bmatrix} 4 \\ 7 \\ 0 \\ 8 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 0 \\ 9 \\ 2 \end{bmatrix}$

3. (a) Give an example of three vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ such that (1) the vectors are linearly dependent, and (2) if you remove any one of them the remaining two are linearly independent.
   (b) Give an example of three vectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ such that (1) the vectors are linearly dependent; (2) if you remove $\mathbf{w}_1$ then the remaining two are linearly independent, and; (3) if you remove $\mathbf{w}_2$ then the remaining two are linearly dependent.

§11. The fundamental subspaces

Based on ILA 5th, §3.5.

We have previously discussed two "fundamental" subspaces associated with a given matrix A: the column space and the null space. The column space is the span of the columns, the null space is the combinations of the columns that give zero. It should not surprise us that there are two more fundamental subspaces corresponding to the rows.

We can view the rows of a matrix as columns of the "transpose" matrix.

11.1. DEFINITION. Given a matrix A we form the *transpose* matrix A^T as follows: The rows of A^T are the columns of A and the columns of A^T are the rows of A.

11.2. EXAMPLE.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

The four fundamental subspaces corresponding to a matrix A are the following.

- $C(A)$, the *column space*, the span (combinations) of the columns of A
- $N(A)$, the *null space*, the solutions to $A\mathbf{x} = \mathbf{0}$ (the combinations of the columns that give zero, also the equations of the row space)
- $R(A)$, the *row space*, the span (combinations) of the rows of A
- $LN(A)$, the *left null space*, the solutions to $\mathbf{y}^T A = \mathbf{0}$ (the combinations of the rows that give zero, also the equations of the column space)

The four subspaces are depicted on the cover of our book.

We wish to find bases and dimensions of the four fundamental subspaces corresponding to A. In all cases we first need to calculate the RREF.

- $C(A)$: The basis consists of the columns of the original matrix corresponding to basic columns of the RREF.
- $N(A)$: The basis consists of the special solutions.
- $R(A)$: The basis consists of the nonzero rows of the RREF matrix.
- $LN(A)$: The basis consists of the coefficients of the equations that determine the column space. (Alternatively, start with A^T and find the null space.)

11.3. EXAMPLE. Consider the matrix A below. Find bases and dimensions of the four fundamental subspaces.

$$A = \begin{bmatrix} 1 & 3 & 0 & 5 \\ 2 & 6 & 1 & 16 \\ 5 & 15 & 0 & 25 \end{bmatrix}$$

We first augment the matrix with an unknown right-hand side $\mathbf{b}$, and find RREF.

$$R = \left[\begin{array}{cccc|c} 1 & 3 & 0 & 5 & b_1 \\ & & 1 & 6 & b_2 - 2b_1 \\ & & & & b_3 - 5b_1 \end{array}\right]$$

The four bases are thus:

$$\text{C(A):}\ \begin{bmatrix}1\\2\\5\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix},\quad \text{N(A):}\ \begin{bmatrix}-3\\1\\0\\0\end{bmatrix},\begin{bmatrix}-5\\0\\-6\\1\end{bmatrix},\quad \text{R(A):}\ \begin{bmatrix}1\\3\\0\\5\end{bmatrix},\begin{bmatrix}0\\0\\1\\6\end{bmatrix},\quad \text{LN(A):}\ \begin{bmatrix}-5\\0\\1\end{bmatrix}$$

We can also see that $C(A)$ is a 2-dimensional subspace of $\mathbb{R}^3$, $N(A)$ is a 2-dimensional subspace of $\mathbb{R}^4$, $R(A)$ is a 2-dimensional subspace of $\mathbb{R}^4$, and $LN(A)$ is a 1-dimensional subspace of $\mathbb{R}^3$.

If A is $m \times n$, the dimensions of the four fundamental subspaces of A can always be calculated m, n, and the number of pivots r.

- $C(A)$ is an r-dimensional subspace of $\mathbb{R}^m$
- $N(A)$ is an $n - r$-dimensional subspace of $\mathbb{R}^n$
- $R(A)$ is an r-dimensional subspace of $\mathbb{R}^n$
- $LN(A)$ is an $m - r$-dimensional subspace of $\mathbb{R}^m$

The number of pivots r is called the *rank* of the matrix A and measures the size of the matrix in terms of the dimension of its row or column space.

Activity for §11.

1. For each of the following matrices A, find bases and dimensions of all four fundamental subspaces corresponding to A.
   (a) $A = \begin{bmatrix} 2 & 4 & 0 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix}$
   (b) $A = \begin{bmatrix} -2 & 1 & 0 & 4 & 1 \\ 4 & -2 & 1 & 3 & 1 \\ 2 & -1 & 1 & 7 & 2 \end{bmatrix}$

2. Given each set of info, find the dimensions of all four fundamental subspaces corresponding to A.
   (a) A is a 3×5 matrix with two pivots.
   (b) The null space is a three-dimensional subspace of $\mathbb{R}^5$ and the left-null space is one-dimensional.

3. Let P be the plane with equation $\begin{bmatrix} 2 & 3 & -1 \end{bmatrix} \mathbf{x} = \mathbf{0}$. Find all solutions to the equation in parametric vector form. Which of the four subspaces is this? Recall that the *normal line* is perpendicular to the plane and passes through $(2, 3, -1)$. Which of the four subspaces is this?

§12. Orthogonality and fundamental subspaces

Based on ILA 5th, §4.1.

In previous sections we saw that solving systems of linear equations is related to two "fundamental" subspaces of a matrix: the column space (which determines when it is consistent), and the null space (which are the solutions up to a translation). We then defined two more fundamental subspaces, the row space and the left null space. We saw that the four spaces are related to each other. In this section we reveal one more special property of the four subspaces.

12.1. DEFINITION. Vectors $\mathbf{v}, \mathbf{w}$ are *orthogonal*, written $\mathbf{v} \perp \mathbf{w}$, if they are perpendicular (at 90 degree angles to one another). Subspaces V, W of $\mathbb{R}^n$ are *orthogonal*, written $V \perp W$, if every vector of V is orthogonal to every vector of W.

You may recall from previous courses that we can test whether vectors are orthogonal using the dot product.

12.2. LEMMA. *Vectors $\mathbf{v}, \mathbf{w}$ are orthogonal if and only if $\mathbf{v}^T\mathbf{w} = 0$.*

PROOF. The reason this statement is true ultimately comes down to the Pythagorean theorem.

Observe that $\mathbf{v}, \mathbf{w}$ are orthogonal if and only if the triangle with corners $\mathbf{0}, \mathbf{v}, \mathbf{w}$ is a right triangle. By the Pythagorean theorem, this is the case if and only if

$$|\mathbf{v}|^2 + |\mathbf{w}|^2 = |\mathbf{v} - \mathbf{w}|^2$$

We now use the distance formula, which states that the length of a vector is $|\mathbf{v}| = \sqrt{\sum v_i^2}$ (the distance formula is yet another version of the pythagorean theorem). Thus the previous equation is equivalent to

$$\sum v_i^2 + \sum w_i^2 = \sum (v_i - w_i)^2$$

Distributing each summand on the right hand side, we arrive at

$$\sum v_i^2 + \sum w_i^2 = \sum v_i^2 - \sum 2v_i w_i + \sum w_i^2$$

After cancelling terms, this last equation is equivalent to the desired $\mathbf{v}^T\mathbf{w} = 0$. □

12.3. EXAMPLE. The vector $(1, 5, -7, -1)$ is orthogonal to the vector $(2, 1, 1, 0)$. It is also orthogonal to the vector $(1, 2, 1, 4)$. By the same reasoning, we can also say that the line V through $(1, 5, -7, -1)$ is orthogonal to the plane W with basis $(2, 1, 1, 0), (1, 2, 1, 4)$.

For any subspace V, there is always a largest subspace W such that V, W are orthogonal.

12.4. DEFINITION. If V is a subspace of some Euclidean space $\mathbb{R}^n$, its *orthogonal complement* $V^\perp$ is the subspace consisting of all vectors $\mathbf{w}$ that are orthogonal to all vectors in V.

12.5. EXAMPLE. If V is the line in $\mathbb{R}^3$ that passes through the vector $(1, -1, 2)$, then the orthogonal complement of V would be the vectors (x, y, z) such that

$$\begin{bmatrix} 1 & -1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 0$$

In other words, the orthogonal compleement is the set of solutions to $x - y + 2z = 0$, or in other words, the plane parameterized by

$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} y + \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} z$$

It makes sense that the orthogonal complement of a line in $\mathbb{R}^3$ would be a plane—the dimensions of the two subspaces should add to 3.

We saw from the example that if $A = \begin{bmatrix} 1 & -1 & 2 \end{bmatrix}$ then the null space of A is the orthogonal compelement of the row space of A. Strang states this is the fundamental theorem of linear algebra.

12.6. THEOREM (Fundamental Theorem of Linear Algebra). *Given any matrix A, the row space and null space are orthogonal complements of each other, and the column space and left-null space are orthogonal compelements of each other.*

This theorem completes our understanding of the four fundamental subspaces. It also gives us a general method to calculate orthogonal complements.

12.7. EXAMPLE. Let V be the subspace of $\mathbb{R}^4$ with the basis below. Find a basis for $V^\perp$.

$$\text{basis for } V\text{:} \quad \begin{bmatrix} 1 \\ 3 \\ 0 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}$$

We solve this using the FTLA as follows: put the given vectors into the rows of a matrix, and then find the null space of the matrix. Here are the matrix and the final RREF.

$$A = \begin{bmatrix} 1 & 3 & 0 & 1 \\ 1 & 1 & 2 & 1 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 0 & 3 & 1 \\ 0 & 1 & -1 & 0 \end{bmatrix}$$

Using the null space trick, we can say:

$$\text{basis for } V^\perp\text{:} \quad \begin{bmatrix} -3 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Activity for §12.

1. In each problem, let V be the subspace with the given basis. Find a basis for $V^{\perp}$.
   - Basis for V: $\begin{bmatrix} 2 \\ -1 \\ \frac{1}{2} \end{bmatrix}$
   - Basis for V: $\begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}$
   - Basis for V: $\begin{bmatrix} 1 \\ 2 \\ -5 \\ 8 \end{bmatrix}, \quad \begin{bmatrix} -1 \\ -1 \\ 3 \\ -5 \end{bmatrix}$

2. - Find a vector orthogonal to the plane spanned by the vectors $(1, 3, 4)$ and $(5, 2, 7)$.
   - Find a vector orthogonal to the plane with the equation $3x - y + 2z = 5$.

3. Consider the vector $(2, 3)$ in $\mathbb{R}^2$. What is the closest vector to $(2, 3)$ on the x-axis? Draw a right triangle that would convince someone your answer is correct.

§13. Orthogonal projections

Based on ILA 5th, §4.2.

Given a system of linear equations $A\mathbf{x} = \mathbf{b}$, if it is consistent we know how to solve it. If it is inconsistent, $\mathbf{b}$ isn't in the column space of A and we can't find any solution. In this case, we can still ask for a *best approximate solution* $\mathbf{x}$ that makes $A\mathbf{x}$ and $\mathbf{b}$ as close as possible. The method is to replace $\mathbf{b}$ with the *nearest vector* $\mathbf{p}$ that is in the column space of A and to solve $A\hat{\mathbf{x}} = \mathbf{p}$ instead.

Geometrically, if $A\mathbf{x} = \mathbf{b}$ is inconsistent, and $\mathbf{b}$ doesn't lie in the column space of A, we want to "project" $\mathbf{b}$ to the column space, or in other words find its shadow. [Picture] It is not difficult to see that an optimal solution is achieved exactly when the error vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the column space of A.

The theory for finding $\hat{\mathbf{x}}$ and $\mathbf{p}$ is known as *orthogonal projection* and runs as follows. Given an inconsistent system $A\mathbf{x} = \mathbf{b}$, we set the error term orthogonal to the columns of A:

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0$$

This simplifies to:

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$$

We then solve this modified system in the usual way to find the best approximate solution $\hat{\mathbf{x}}$. Finally we can find the *projection* vector $\mathbf{p}$ by multiplying $A\hat{\mathbf{x}}$.

13.1. EXAMPLE. Let $A\mathbf{x} = \mathbf{b}$ be the inconsistent system below. Find the best approximate solution $\hat{\mathbf{x}}$ and the projection $\mathbf{p}$ of $\mathbf{b}$ to $C(A)$.

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

The first step is to make the modified system $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$. It turns out to be:

$$\begin{bmatrix} 3 & 3 \\ 3 & 5 \end{bmatrix} \hat{\mathbf{x}} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

The second step is to solve using old-fashioned elimination and back-solving. The third step is to find the projection vector $\mathbf{p} = A\hat{\mathbf{x}}$. The solutions are:

$$\hat{\mathbf{x}} = \begin{bmatrix} 5 \\ -3 \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} 5 \\ 2 \\ -1 \end{bmatrix}$$

The situation is somewhat simpler when A is just a single vector $\mathbf{a}$. In this case we are projecting $\mathbf{b}$ onto the line through $\mathbf{a}$. Once again, we set the error term $\mathbf{b} - \mathbf{a}\hat{x}$ orthogonal to $\mathbf{a}$:

$$\mathbf{a}^T(\mathbf{b} - \mathbf{a}\hat{x}) = 0$$

This time we solve explicitly to obtain

$$\hat{x} = \mathbf{a}^T\mathbf{b}/\mathbf{a}^T\mathbf{a}$$

We can also explicitly find the projection

$$\mathbf{p} = \frac{\mathbf{a}^T\mathbf{b}}{\mathbf{a}^T\mathbf{a}}\mathbf{a}$$

13.2. EXAMPLE. Find the projection of **b** onto the line through **a**.

$$\mathbf{a} = (1, 2, 3), \quad \mathbf{b} = (4, 6, 4)$$

The solution is: ...

We close by mentioning that given any matrix A, there is a single *projection matrix P* which carries vectors **b** to their projections **p** to the column space of A. To find P, we simply observe that $\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}$ and $\mathbf{p} = A\hat{\mathbf{x}} = A(A^TA)^{-1}A^T\mathbf{b}$. In other words:

$$P = A(A^TA)^{-1}A^T$$

13.3. EXAMPLE. Returning to the first example, we can calculate P.

$$P = \dots$$

Activity for §13.

1. Let $A\mathbf{x} = \mathbf{b}$ be the inconsistent system below. Find the best approximate solution $\hat{\mathbf{x}}$ and the projection $\mathbf{p}$ of $\mathbf{b}$ to $C(A)$.
   (a) $\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix}$
   (b) $\begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 4 \\ 4 \\ 6 \end{bmatrix}$

2. Find the projection of $\mathbf{b}$ onto the line through $\mathbf{a}$.
   (a) $\mathbf{a} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$
   (b) $\mathbf{a} = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$

3. For the matrix A in 1(a), find the projection matrix P.

§14. Linear modeling

Based on ILA 5th, §4.3.

In the previous section, we described how to find the orthogonal projection of a given vector **b** onto a vector subspace. The projection vector **p** is optimal in the sense that it is as close as possible to **b**.

This concept has an enormous application in statistics, where it is called "least squares optimization". We often see data sets which, when plotted, do not form a line, but instead have a visible "trend". Given an arbitrary data set, we wish to find the line that most closely fits the data. We will use the fact that the possible data sets form a vector space, and the linear data sets form a subspace. We will take a given data set and *project* it onto the subspace consisting of linear data sets! This is a beautiful and typical example of the power of abstract mathematics!

Here is how it works. Given a sequence of data points $(t_1, b_1), \dots, (t_n, b_n)$, the data lie on a line of the form $C + Dt = y$ if and only if the following system is consistent:

$$\begin{cases} 1C + t_1 D = b_1 \\ \vdots \\ 1C + t_n D = b_n \end{cases}$$

If the system is not consistent we use the method of orthogonal projection, replacing **b** with **p** and finding the best approximate solutions $\hat{C}, \hat{D}$ for C, D.

14.1. EXAMPLE. Recall the example that we gave on the very first day of class. Given the following measurements, find the line that best fits the data.

$$(0,6), (1,0), (2,0)$$

The data lie on a line if and only if the system is consistent:

$$\begin{cases} C + 0D = 6 \\ C + 1D = 0 \\ C + 2D = 0 \end{cases}$$

Write it as a matrix equation:

$$\begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \mathbf{C} = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$$

The data do not lie on a line, so the system is inconsistent. Instead we find the best approximate solution. We actually did this example in the previous section and got $\hat{C} = 5$, $\hat{D} = -3$. Thus the best fit line is $5 - 3t = y$. Graphing the line with the data, we see this is reasonable!

We didn't really need the projection vector $\mathbf{p}$. In this case $\mathbf{p} = (5, 2, -1)$, which is simply the vector of y-coordinates on the best fit line. The error vector is $(1, -2, 1)$ and corresponds to the "residuals".

For a larger data set (t_i, b_i), the A matrix and $\mathbf{b}$ vector will always look like the following:

$$A = \begin{bmatrix} 1 & t_1 \\ 2 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

Here is the really cool part: The method can be used to find the best fit for any linear model, not just a line! Here a linear model means any linear combination of functions. For example we can find a best fit *parabola* using the model $C + Dt + Et^2 = y$.

14.2. EXAMPLE. Find the best fit parabola for the given data set.

$$\begin{bmatrix} 0 \\ 6 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

The data lie on a parabolif if and only if the system is consistent:

$$\begin{cases} 1C + 0D + 0E = 6 \\ 1C + 1D + 1E = 0 \\ 1C + 2D + 4E = 0 \\ 1C + 3D + 9E = 2 \end{cases}$$

In other words, the A matrix and $\mathbf{b}$ vector are:

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_n & t_n^2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

The best approximate solution turns out to be $\hat{C} = 5.8, \hat{D} = -7.2, \hat{E} = 2$. Thus the best fit parabola $5.8 - 7.2t + 2t^2 = y$.

Activity for §14.

1. Use the method of orthogonal projections to find the best fit line for the data set.
   (a) $(0,9),(1,4),(2,1)$
   (b) $(0,0),(1,8),(3,8),(4,20)$

2. Repeat problem 1(b), but this time find the best fit parabola.

PART III

Determinants, eigenvalues, eigenvectors, and applications

§15. The determinant of a matrix

Based on ILA 5th, §5.1.

In this section we introduce the concept of the determinant of a matrix. Given any square matrix A, we will associate with it a determinant $\det(A)$, which is an important real number that tells us some information about A.

As motivation, consider the initial step in the process of inverting a general 2×2 matrix:

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right] \to \left[\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & d - \frac{c}{a}b & -\frac{c}{a} & 1 \end{array}\right]$$
$$\to \left[\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & ad - bc & -c & a \end{array}\right]$$
$$\to \left[\begin{array}{cc|cc} a & 0 & 1 + \frac{ad-bc}{b}c & -\frac{ad-bc}{b}a \\ 0 & ad - bc & -c & a \end{array}\right]$$
$$\to \left[\begin{array}{cc|cc} 1 & 0 & \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ 0 & 1 & \frac{-c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right]$$

Notice that after we cleared denominators, the last pivot is $ad - bc$, and that this quantity shows up in the denominator of the entries of A^{-1}. Notice also that if this quantity is zero, then the method fails and A is not invertible.

The special quantity $ad - bc$ is called the *determinant* of A because it determines whether A is invertible or not. For $A =$ the general 2×2 matrix, we write $\det(A) = ad - bc$ or $|A| = ad - bc$.

Of course we want to go beyond the simple case of a 2×2 matrix and find determinants for larger matrices. So let's step up to the case of a general 3×3 matrix. This time we won't bother calculating the inverse, but just the last pivot.

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \to \begin{bmatrix} a & b & c \\ 0 & e - \frac{d}{a}b & f - \frac{d}{a}c \\ 0 & h - \frac{g}{a}b & i - \frac{g}{a}c \end{bmatrix}$$
$$\to \dots$$
$$\to \begin{bmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & aei + bfg + cdh - afh - bdi - ceg \end{bmatrix}$$

Thus if $A =$ the general 3×3 matrix, we write $\det(A) = aei + bfg + cdh - afh - bdi - ceg$.

These small cases are useful, but still leave us with a lot of questions about the determinant. What is it really, and what makes determinant the determinant? We now proceed to define the determinant by its basic properties.

15.1. DEFINITION. The determinant $\det(A)$ is a mapping from matrices to real numbers and is defined by the following four properties.

- (Scaling) Scaling a row of by a scalar t also scales the determinant by t.
$$\det \begin{bmatrix} ta & tb \\ c & d \end{bmatrix} = tad - tbc = t(ad - bc)$$
- (Row exchange) Exchanging two rows negates the determinant.
$$\det \begin{bmatrix} c & d \\ a & b \end{bmatrix} = cb - da = -(ad - bc)$$
- (Elimination) Adding a of one row to another row has no effect on the determinant.
$$\det \begin{bmatrix} a + tc & b + td \\ c & d \end{bmatrix} = ad - bc$$
- (Identity) The determinant of the identity matrix is 1. This is clear from small examples.

The above definition isn't a formula, it is just four properties. Still, these four rules can be used to find the determinant of any matrix! To see this, we reason as follows.

- The determinant of a diagonal matrix is the product of its diagonal entries. To see this we use the Scaling rule repeatedly plus the Identity rule.
$$\det \begin{bmatrix} d_1 & & \\ & d_2 & \\ & & \ddots \end{bmatrix} = d_1 \det \begin{bmatrix} 1 & & \\ & d_2 & \\ & & \ddots \end{bmatrix} = \cdots = d_1 d_2 \cdots d_n$$
- The determinant of an upper triangular matrix is the product of its diagonal entries too. To see this note that if the determinant is not zero, it has all its pivots

and you can use the Elimination rule until it is diagonal. Then apply the previous item.

- The determinant of any matrix is the product of its pivots. Once again, we can use the Elimination rule to make any matrix upper triangular, and then apply the previous item.

In conclusion, we can always find the determinant by eliminating it to upper triangular form, and then multiplying the pivots.

Warning: if you use any Scaling or Row Exchange steps while eliminating, you must keep track of this along the way.

15.2. EXAMPLE.

$$\det \begin{bmatrix} 1 & 2 & 3 \\ 3 & 0 & 1 \\ 1 & 2 & 1 \end{bmatrix}$$

15.3. EXAMPLE.

$$\det \begin{bmatrix} 2 & 4 & 6 \\ 2 & 4 & 1 \\ 1 & 4 & 2 \end{bmatrix}$$

Activity for §15.

1. Use elimination to find the determinant of each of the following matrices.

$$\text{(a)} \begin{bmatrix} 2 & -1 & 0 \\ 1 & 2 & -2 \\ -1 & 1 & 3 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} a & 1 & 1 \\ 0 & a & 1 \\ 0 & 0 & a \end{bmatrix}$$

$$\text{(c)} \begin{bmatrix} a & 1 & & \\ 1 & a & 1 & \\ & 1 & a & 1 \\ & & 1 & a \end{bmatrix} \qquad \text{(d)} \begin{bmatrix} a & b & & \\ c & d & & \\ & & a & b \\ & & c & d \end{bmatrix}$$

2. For each 2×2 matrix A below, calculate the determinant using $ad - bc$. Then find the inverse (if it exists) using the formula $A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. Finally check your work by verifying that $AA^{-1} = I$.

$$\text{(a)} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} \frac{1}{2} & -3 \\ 5 & \frac{1}{3} \end{bmatrix} \qquad \text{(c)} \begin{bmatrix} -.2 & .8 \\ .3 & .7 \end{bmatrix}$$

3. Take a look at the six-term formula for the determinant of the general 3×3 matrix

$$\det(A) = aei + bfg + cdh - afh - bdi - ceg$$

What do each of the six terms have in common? Each one has a factor from each _______ and each _______. What additionally do the three positive terms have in common? What do the three negative terms have in common?

§16. Formulas for the determinant

Based on ILA 5th, §5.2.

Last week we defined the determinant by its four key properties, and described an algorithm to calculate the determinant using elimination. In this section we additionally give a general formula for the determinant, and a recursive formula for the determinant.

We begin with the general formula for the determinant, called the *big formula*. To motivate it, recall the six-term formula for the determinant of a 3×3 matrix:

$$\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = aei + bfg + cdh - afh - bdi - ceg$$

Observe that each term contains one entry from every row and one entry from every column:

$$\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \det \begin{bmatrix} a & & \\ & e & \\ & & i \end{bmatrix} + \det \begin{bmatrix} & b & \\ & & f \\ g & & \end{bmatrix} + \det \begin{bmatrix} & & c \\ d & & \\ & h & \end{bmatrix} - \det \begin{bmatrix} a & & \\ & & f \\ & h & \end{bmatrix} - \det \begin{bmatrix} & b & \\ d & & \\ & & i \end{bmatrix} - \det \begin{bmatrix} & & \\ & & \\ g & & \end{bmatrix}$$

A selection of one entry from every row and column is called a *permutaion*. A permutation is a bijective function $\sigma \colon n \to n$. The set of all permutations of $1, \dots, n$ is denoted S_n. Every permutation has a *sign*, which is ± 1 depending on whether an even or odd number of swaps are needed to reduce it to the identity.

In sum, the determinant of a matrix is given by the following "big" formula:

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \prod_{1}^{n} a_{i,\sigma(i)}$$

The formula is "big" because it has $n!$ many terms, a huge quantity. A formula with $n!$ many terms is rarely of practical value, but the big formula can occasionally be useful for sparse matrices.

16.1. EXAMPLE. Use the general formula to find the determinant of the matrix.

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

We now introduce the recursive formula for the determinant, called the *cofactor expansion*. We once again begin with the six term formula for the determinant of a 3×3 matrix. This time we factor some of the terms:

$$\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = a(ei - fh) + b(fg - di) + c(dh - eg)$$

The last equation is a combination of three sub-determinants or "cofactors":

$$\det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = a \det \begin{bmatrix} e & f \\ h & i \end{bmatrix} - b \det \begin{bmatrix} d & f \\ g & i \end{bmatrix} + c \det \begin{bmatrix} d & e \\ g & h \end{bmatrix}$$

More generally, given a matrix A consisting of entries a_{ij}, we define the *cofactor* c_{ij} is equal to $(-1)^{i+j}$ times the determinant of the matrix obtained from A by deleting the ith row and jth column. For example in the matrix

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

the cofactor of the 4 is $-\det \begin{bmatrix} 2 & 3 \\ 8 & 9 \end{bmatrix} = 6$.

The *cofactor formula* for the determinant is:

$$\det(A) = a_{11}c_{11} + \cdots + a_{1n}c_{1n}$$

To remember the signs of the cofactors $(-1)^{i+j}$ we use the following memory device:

$$\begin{bmatrix} + & - & + & \\ - & + & + & \cdots \\ + & - & + & \\ & \vdots & & \ddots \end{bmatrix}$$

16.2. EXAMPLE. Use the cofactor expansion to find the determinant of the matrix.

$$\det \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & -1 \\ & & -1 & 2 \end{bmatrix}$$

We close by noting that you don't necessarily have to use the top row to make a cofactor expansion. You can use any row or column, provided you are careful to get the signs of the cofactors just right.

§17. Two applications of determinants

Based on ILA 5th, §5.3.

We have already said that the determinant of a matrix tells you whether it is invertible (nonzero determinant) or singular (zero determinant). We will use this fact repeatedly in the next chapter. But before moving on, we give two further applications of the determinant.

The first application involves the close connection between determinants and the inverse matrix. The first hint of this was the general formula for the inverse of a 2×2 matrix, in which $\det(A)$ appears in the denominator.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

For an even bigger hint, we will reveal the general formula for the inverse of a 3×3 matrix.

$$\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}^{-1} = \frac{1}{\det(A)} \begin{bmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{bmatrix}$$

Once again the determinant appears in the denominator. But note also that the entries of the matrix are determinants as well, though smaller in size. These determinants are obtained by crossing out one row and column of A.

17.1. DEFINITION. Given a matrix A, the corresponding *cofactor matrix* of A consists of entries $c_{ij} = (-1)^{i+j}$ times the determinant of the submatrix of A with the ith row and jth column crossed out.

To remember the signs of the cofactors $(-1)^{i+j}$ we use the memory device

$$\begin{bmatrix} + & - & + & \\ - & + & + & \cdots \\ + & - & + & \\ & \vdots & & \ddots \end{bmatrix}$$

17.2. EXAMPLE. The matrix A has cofactor matrix C:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 2 & 2 \\ 1 & 1 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 0 & -2 & 2 \\ 1 & -2 & 1 \\ -2 & 10 & -6 \end{bmatrix}$$

In the above example, look at what happens when we multiply AC^T:

$$AC^T = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 2 & 2 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & -2 \\ -2 & -2 & 10 \\ 2 & 1 & -6 \end{bmatrix} = \begin{bmatrix} 2 & & \\ & 2 & \\ & & 2 \end{bmatrix}$$

In general the product AC^T is equal to $\det(A)I$, the diagonal matrix with the determinant of A every diagonal entry.

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nn} \end{bmatrix} = \begin{bmatrix} \det(A) & & \\ & \ddots & \\ & & \det(A) \end{bmatrix}$$

17.3. THEOREM (Cramer's formula for the inverse).

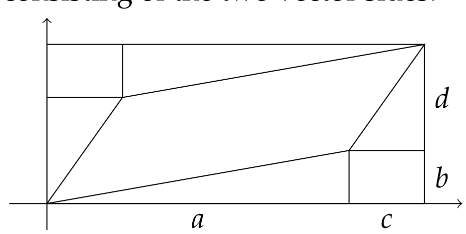
$$A^{-1} = \frac{1}{\det(A)} C^T$$

17.4. EXAMPLE.

$$A = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ 1 & 1 & 1 & \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad A^{-1} = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & -1 & 1 & \\ & & -1 & 1 \end{bmatrix}$$

The second application is about the calculation of the area of a parallelepiped. Recall that a parallelepiped is the higher dimensional analog of a paralellogram and is the region bounded by several vectors and their sums.

Let's first look at the 2-dimensional parallelogram with vector sides (a, b) and (c, d). Using simple geometry, we can find that the area of the paralellogram is $ad - bc$, the determinant of the matrix consisting of the two vector sides.



Perhaps surprisingly, this turns out to be true in higher dimensions! The n-dimensional volume of the paralellepiped with vector edges $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is given by

$$V = \left| \det \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ | & & | \end{bmatrix} \right|$$

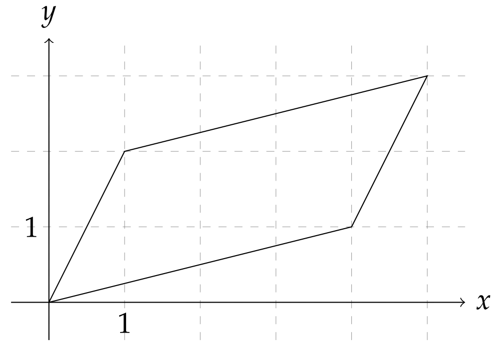
To understand briefly how this can be true, we can see that the volume of a parallelepiped satisfies the same four properties that we used to define the determinant.

- (Identity) If $\mathbf{v}_i = \mathbf{e}_i$ then the parallelepiped is the unit hypercube which clearly has volume 1.
- (Scaling) If you scale one of the sides of a paralellepiped by t this clearly scales the volume by t.
- (Elimination) Adding a multiple of one vector to another slides the parallelepiped along one of its sides. This has no effect on the volume.

- (Row exchange) This has no effect on the volume, but if we consider the volume to be signed, we can assume this is true.

Activity for §17.

1. Use **geometry** to find the area of the parallelogram with the vector sides shown below. Then use a **determinant** to confirm your answer.



2. Find the volume of the parallelepiped with the vector sides.

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix}$$

3. Use Cramer's rule to find the inverse of each matrix (if it exists).

(a) $\begin{bmatrix} 0 & 2 & -1 \\ 2 & 1 & 5 \\ -1 & 0 & 4 \end{bmatrix}$

(b) $\begin{bmatrix} a & & \\ b & c & \\ d & e & f \end{bmatrix}$

§18. Eigenvalues and eigenvectors

Based on ILA 5th, §6.1.

We are now prepared to introduce eigenvalues and eigenvectors of a matrix. This is probably the most powerful and broadly used concept in linear algebra. Before presenting the technical definitions, let's begin with a motivating example.

Imagine a large island with two cities, C and D. The two cities start with equal population of 50 thousand. Every year 20% of the population of city C moves to city D, and 30% of city D moves to city C.

$$\begin{tikzcd} C \arrow[r, bend right, "20\%"'] & D \arrow[l, bend right, "30\%"'] \end{tikzcd}$$

What will happen to the two populations over time? It isn't easy to tell just by guessing! We can look at the first few years:

	0yr	1yr	2yr	3yr
C	50	55	57.5	58.75
D	50	45	42.5	41.25

It looks like in the limit, the population distribution will tend towards 60% for city C and 40% for city D. How could we find the limit population distribution without just estimating? Linear algebra can come to our aid. It turns out that the dynamics of this scenario are controled by a matrix, and the limit population distribution is an *eigenvector* associated with that matrix.

18.1. DEFINITION. Let A be a square matrix. Let $\mathbf{x}$ be a vector and λ be a number such that $A\mathbf{x} = \lambda\mathbf{x}$. Then the vector $\mathbf{x}$ is called an *eigenvector* of A and the number λ is called an *eigenvalue* of A.

If $\mathbf{x}$ is an eigenvectors of A, then $\mathbf{x}$ is special because A doesn't change what direction it points in. It simply scales $\mathbf{x}$ by a factor of λ.

We now turn to the problem of finding eigenvectors and eigenvalues for a given matrix A. First we will lay out the theory, and later we will show how it works with an example. Given a matrix A, we follow the following steps.

- Step 0. We are looking for $\mathbf{x}$, λ which make $A\mathbf{x} = \lambda\mathbf{x}$ true. We rewrite this as $A\mathbf{x} - \lambda\mathbf{x} = 0$, and then again as $(A - \lambda I)\mathbf{x} = \mathbf{0}$.
- Step 1. The last equation tells us we want to find out when $A - \lambda I$ has a nontrivial null space. We know this happens when $A - \lambda I$ is singular, or in other words when $\det(A - \lambda I) = 0$. So we set $\det(A - \lambda I) = 0$ and solve for λ.
- Step 2. For each solution λ obtained in step 1, plug it into $A - \lambda I$ and find the basis for the null space. This vector is the eigenvector $\mathbf{x}$ corresponding to λ. (Sometimes there will be several vectors $\mathbf{x}$ corresponding to λ.)

18.2. EXAMPLE. Consider the matrix A:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

Beginning with Step 1, we have

$$\begin{aligned} \det \begin{bmatrix} 1-\lambda & 2 \\ 2 & 4-\lambda \end{bmatrix} = 0 &\iff (1-\lambda)(4-\lambda) - (2)(2) = 0 \\ &\iff 4 - 5\lambda + \lambda^2 - 4 = 0 \\ &\iff \lambda^2 - 5\lambda = 0 \\ &\iff (\lambda)(\lambda - 5) = 0 \end{aligned}$$

Thus the eigenvalues are 0 and 5. To find the eigenvectors we move on to Step 2. For the eigenvalue $\lambda = 0$ we find the null space of

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

Thus the eigenvector corresponding to $\lambda = 0$ is $\mathbf{x} = (-2, 1)$. For the eigenvalue $\lambda = 5$ we find the null space of

$$\begin{bmatrix} 1-5 & 2 \\ 2 & 4-5 \end{bmatrix}$$

Thus the eigenvector corresponding to $\lambda = 5$ is $\mathbf{x} = (1, 2)$. We can check our solutions by confirming that $A(-2,1)^T = 0(-2,1)^T$ and $A(1,2)^T = 5(1,2)^T$.

18.3. EXAMPLE. Here we return to our motivating example with cities C and D. To go from the population distribution for year n to the population distribution for year $n+1$, we multiply by the matrix A:

$$A = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$$

Now we will find the eigenvalues and eigenvectors of A. The equation in Step 1 will be:

$$(0.8 - \lambda)(0.7 - \lambda) - (0.2)(0.3) = 0$$

The solutions are $\lambda = 1, 0$. The corresponding eigenvectors are $(1.5, 1)$ and $(1, -1)$. The eigenvector corresponding to $\lambda = 1$ is the one we are after (we will see why this is true later). Note that if we renormalize $(1.5, 1)$ by dividing by the sum 2.5, it becomes $(0.6, 0.4)$, which is the limit population distribution!

18.4. EXAMPLE. Consider the matrix A:

$$A = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

We find that it has eigenvalues $0, 1, 3$ with corresponding eigenvectors $(1, 1, 1)$, $(-1, 0, 1)$, $(1, -2, 1)$. Check!

Activity for §18.

1. For each matrix A, find the eigenvalues λ and corresponding eigenvectors $\mathbf{x}$. In each case, check that $A\mathbf{x} = \lambda\mathbf{x}$.
   (a) $A = \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix}$
   (b) $A = \begin{bmatrix} 3 & 4 & 2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix}$
   (c) $A = \begin{bmatrix} 0 & 2 & 0 & 0 \\ 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{bmatrix}$

2. Suppose A and B are cities, and that each year 10% of city A movies to city B, and 20% of city B movies to city A. Assuming there are no other population changes, find the matrix A which models the population distribution over time. Then find the eigenvalues and eigenvectors of A. What is the limit population distribution?

3. Let A be the matrix from problem 1(b). Put the three eigenvectors you found into the columns of a new matrix X. What is AX and why?

§19. Diagonalization

Based on ILA 5th, §6.2.

In the previous section we said that for a square matrix A, if the special equation $A\mathbf{x} = \lambda\mathbf{x}$ is true then we say that λ is an eigenvalue and $\mathbf{x}$ is a corresponding eigenvector of A. In most cases, if A is $n \times n$ then there will be exactly n distinct eigenvectors (not counting scalar multiples). Moreover the eigenvectors are linearly independent, which means they make a *basis* for $\mathbb{R}^n$. The basis of eigenvectors is very special, because A acts really nicely (diagonally) with respect to this basis.

To begin exploring this idea, assume that A is $n \times n$ and has n distinct eigenvectors. We let X be the $n \times n$ matrix whose columns are the eigenvectors. Then the matrices A and X interact in a very special way, which we will explore by example first.

19.1. EXAMPLE. Let A be the matrix:

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

The eigenvalues are $\lambda = 3, 1, 0$, and the corresponding eigenvectors are $(1,0,0)^T$, $(-2,1,0)^T$, and $(2,-2,1)^T$. When we multiply AX we get:

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

We find that $AX =$ the matrix X again but with each column multiplied by the corresponding eigenvalue. To see why this will always be true, we observe:

$$AX = A \begin{bmatrix} \vdots & & \vdots \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n \\ \vdots & & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & & \vdots \\ \lambda_1\mathbf{x}_1 & \cdots & \lambda_n\mathbf{x}_n \\ \vdots & & \vdots \end{bmatrix} = \begin{bmatrix} \vdots & & \vdots \\ \mathbf{x}_1 & \cdots & \mathbf{x}_n \\ \vdots & & \vdots \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

Let Λ be the diagonal matrix with eigenvalues in the diagonal entries. We can summarize the above by saying that $AX = X\Lambda$. We can further rewrite this as

$$A = X\Lambda X^{-1}$$

This equation is said to *diagonalize* the matrix A. Of course most matrices A are not diagonal, but this equation shows that in most cases it is possible to transform A to make it diagonal!

19.2. EXAMPLE. Let A be the matrix below. Diagonalize A by writing it as $A = X\Lambda X^{-1}$.

$$A = \begin{bmatrix} 1 & 5 \\ & 6 \end{bmatrix}$$

The eigenvalues are $\lambda = 1, 6$ and corresponding eigenvectors are $(1, 0)^T$ and $(1, -1)^T$. We use these to populate the matrices Λ and X. We also have to find the inverse of X. The solution is:

$$A = \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix} \begin{bmatrix} 1 & \\ & 6 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ & 1 \end{bmatrix}$$

We can check our answer by confirming that the right-hand side multiplies back to A.

There are many applications of diagonalization. One application is the calculation of matrix powers A^k. If you can diagonalize A as $A = S\Lambda S^{-1}$, then you can also diagonalize its powers as $A^n = S\Lambda^n S^{-1}$. (Show.)

We will see this come in handy when we want to study the city population problem after n years or in the limit.

19.3. EXAMPLE. For A in the previous example, find A^{100}. The answer is

$$A^{100} = \begin{bmatrix} 1 & 1 \\ & 1 \end{bmatrix} \begin{bmatrix} 1 & \\ & 6 \end{bmatrix}^{100} \begin{bmatrix} 1 & -1 \\ & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 + 6^{100} \\ & 6^{100} \end{bmatrix}$$

We conclude this section with one of the coolest applications of diagonalization. Recall that the Fibonacci sequence begins with $F_0 = 0$, $F_1 = 1$ and is defined recursively by $F_{n+1} = F_{n+1} + F_n$. Find an absolute formula for F_n.

- Step 1. Using the variables F_{n+1} and F_n, the process is governed by the equations

$$\begin{cases} F_{n+1} = F_n + F_{n-1} \\ \quad F_n = F_n \end{cases}$$

  Or in matrix terms we have

$$\begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix}$$

  Denoting this square matrix by A, we want to find

$$\begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = A^n \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

- Step 2. We diagonalize the matrix A above. We find that the two eigenvalues are $\lambda_1 = (1 + \sqrt{5})/2$ and $\lambda_2 = (1 - \sqrt{5})/2$. The corresponding eigenvectors are $(\lambda_1, 1)$ and $(\lambda_2, 1)$. Thus we can diagonalize A as

$$A = \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix} \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 1/\lambda_1 \\ -1 & \lambda_1 \end{bmatrix}$$

- Step 3. We are really interested in

$$\begin{bmatrix} F_{n+1} \\ F_n \end{bmatrix} = A^n \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
$$= \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & \\ & \lambda_2 \end{bmatrix}^n \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -\lambda_2 \\ -1 & \lambda_1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
$$= \frac{1}{\sqrt{5}} \begin{bmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} (\lambda_1)^n & \\ & (\lambda_2)^n \end{bmatrix} \begin{bmatrix} 1 & -\lambda_2 \\ -1 & \lambda_1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
$$= \frac{1}{\sqrt{5}} \begin{bmatrix} \lambda_1^{n+1} - (\lambda_2)^{n+1} \\ \lambda_1^n - (\lambda_2)^n \end{bmatrix}$$

  We can finally conclude the formula

$$F_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right)$$

Activity for §19.

1. In the Fibonacci numbers we said F_{n+1} was the sum of the previous two numbers. To construct the *Gibonacci* numbers, we let G_{n+1} = the *average* of the previous two numbers.
   (a) Using the initial conditions $G_0 = 0$ and $G_1 = 1$, write the first 7 terms of the sequence G_n.
   (b) Write the system of equations that models the Gibonacci process.
   $$\begin{cases} G_{n+1} = \\ \quad G_n = \end{cases}$$
   (c) Write the 2×2 matrix A that models the Gibonacci process.
   (d) Diagonalize the matrix A.
   (e) Use your diagonalization to find a formula for the power A^n.
   (f) Find the formula for G_n.
   (g) Use the above information to find $\lim_{n\to\infty} G_n$.

§20. Markov matrices

Based on ILA 5th, §10.3.

We introduced eigenvectors with a problem about population flow between cities. We observed that the limit population distribution seems to be given by the eigenvector corresponding to $\lambda = 1$. Now that we have the diagonalization avialable, we can actually prove this is true.

As a reminder, the setup is that each year 20% of the population of city C moves to city D, and 30% of city D moves to city C. We said that the population flow is controlled by the matrix

$$A = \begin{bmatrix} .8 & .3 \\ .2 & .7 \end{bmatrix}$$

Thus the population distribution in the nth year is given by

$$\begin{bmatrix} C_n \\ D_n \end{bmatrix} = A^n \begin{bmatrix} C_0 \\ D_0 \end{bmatrix}$$

To find this we will need to diagonalize A. We already found the eigenvalues are $\lambda = 1, .5$ and the corresponding eigenvectors are $(1.5, 1)^T$ and $(-1, 1)^T$. Thus we can calculate:

$$A^n \begin{bmatrix} C_0 \\ D_0 \end{bmatrix} = \begin{bmatrix} 1.5 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & \\ & .5^n \end{bmatrix} \frac{1}{2.5} \begin{bmatrix} 1 & 1 \\ -1 & 1.5 \end{bmatrix}$$

Taking a limit we botain

$$\begin{aligned} A^\infty \begin{bmatrix} C_0 \\ D_0 \end{bmatrix} &= \frac{1}{2.5} \begin{bmatrix} 1.5 & 1.5 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} C_0 \\ D_0 \end{bmatrix} \\ &= \frac{1}{2.5} \begin{bmatrix} 1.5 \\ 1 \end{bmatrix} (C_0 + D_0) \end{aligned}$$

This calculation in this example works for a very general class of problems. A vector that sums to 1 is called a *probability vector*. A matrix whose columns are all probability vectors is called a *Markov matrix*.

20.1. THEOREM (Markov matrix thoerem). *Suppose A is a Markov matrix whose entries are nonzero and $\mathbf{x_0}$ is a probability vector. Then $\lim_{n\to\infty} A^n \mathbf{x_0}$ is equal to the eigenvector corresponding to $\lambda = 1$ (normalized to be a probability vector).*

This theorem is actually a special case of the much more powerful Perron–Frobenius theorem: If A is a (reasonable) matrix, then its dominant (greatest) eigenvalue λ_m is real and positive. Then the growth rate of the powers $A^n\mathbf{x}$ is given by λ_m and the limiting distribution is given by the eigenvector corresponding to $\mathbf{x}_m$.

As an example of this theorem, we present the Leslie matrix, which describes how a population changes through the generations.

20.2. EXAMPLE. Suppose you observe the fertility rates and survival rates of three generations: Group 1 is age 0 to 20, Group 2 is age 20 to 40, and Group 3 is age 40 to 60. The fertility rates F_n are the number of children each female in Group n has in 20 years, the survival rates S_n are the proportion of Group n that survives to group $n+1$.

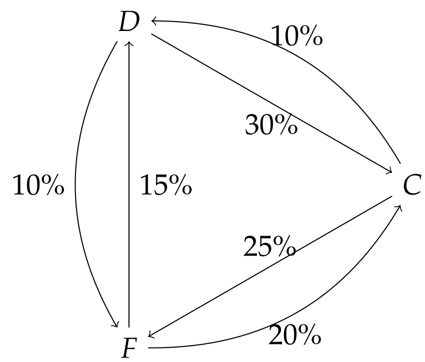
Suppose you measure $F_1 = .04$, $F_2 = 1.1$, $F_3 = .01$, $S_1 = .98$, and $S_2 = .92$. Then the process is controlled by the matrix

$$A = \begin{bmatrix} .04 & 1.1 & .01 \\ .98 & & \\ & .92 & \end{bmatrix}$$

This isn't a Markov matrix, but it is still subject to the Perron–Frobenius theorem. Using a calculator I found that the largest real eigenvalue was $\lambda = 1.06$. Thus the population grows at a rate of 6% per generation. The corresponding eigenvector, renormalized to sum to 1, is $(.37, .34, .29)^T$. This tells the asymptotic distribution of the generations.

Activity for §20.

1. Consider three towns with yearly population movements described by the figure. Assume there are no other population changes.



   (a) Write the Markov matrix for these movements
   (b) Find the eigenvector corresponding to $\lambda = 1$.
   (c) Find the limit population distribution of the three towns.

2. Suppose salmon have three equal growth stages: smolt, ocean dweller, and spawner. The fertility rate of smolts and ocean dwellers is 0, but the fertility rate of spawners is 12. The survival rate of smolts to ocean dwellers is $1/2$, and the survival rate of ocean dwellers to spawners is $1/3$.
   (a) Write the Leslie matrix for this model.
   (b) Find is the dominant eigenvalue λ_m.
   (c) Find the eigenvector corresponding to λ_m.
   (d) Find the asymptotic growth rate of the salmon.

§21. Systems of differential equations

Based on ILA 5th, §6.3.

In our study of discrete processes such as Markov processes, we always have the recursive equation $\mathbf{x_{n+1}} = A\mathbf{x_n}$ and an initial state $\mathbf{x_0}$ given. Here the variables at time $n+1$ depend in a linear way on the variables at time n.

The continuous analog of this situation is a differential equation $\mathbf{u}' = A\mathbf{u}$ and an initial condition $\mathbf{u}(0) = \mathbf{u_0}$. Here the growth rate of the variables depends in a linear way on the variables themselves.

For a simple one-variable example, unrestricted population growth is the situation when $y' = ky$, $y(0) = y_0$. We know that the solution to this differential equation is $y = y_0e^{kt}$, that is, exponential growth.

In the multivariable version, we simply replace a single scalar k with a matrix A. We obtain a linear system of first-order ordinary differential equations.

$$\begin{cases} \mathbf{u}' = A\mathbf{u} \\ \mathbf{u}(0) = \mathbf{u_0} \end{cases}$$

It turns out that we can write the solution to such an equation as

$$\mathbf{u}(t) = e^{At}\mathbf{u_0}$$

But this leaves us with three burning questions. First, what exactly is e^{At}? Second, why does it solve the linear system of differential equations? And third, how can we calculate it?

For the first question (what is e^{At}), recall that e^x may be defined using the series $e^x = 1 + x + x^2/2 + x^3/3! + \cdots$. Thus we define

$$e^{At} = I + At + \frac{(At)^2}{2} + \frac{(At)^3}{3!} + \cdots$$

It is a fact that this sum always converges.

For the second question (why does e^{At} solve $\mathbf{u}' = A\mathbf{u}$), we simply take the derivative of $\mathbf{u} = e^{At}\mathbf{u_0}$:

$$\begin{aligned} \frac{d}{dt}e^{At}\mathbf{u_0} &= \frac{d}{dt}\left[I + At + \frac{(At)^2}{2} + \frac{(At)^3}{3!} + \cdots\right]\mathbf{u_0} \\ &= \left[0 + A + A^2t + A\frac{(At)^2}{2} + \cdots\right]\mathbf{u_0} \\ &= A\left[I + At + \frac{(At)^2}{2} + \cdots\right]\mathbf{u}_0 \\ &= Ae^{At}\mathbf{u_0} \end{aligned}$$

For the last question (how to calculate e^{At}), if A is diagonalizable then one can compute e^{At} using diagonalization. To see this, if $A = X\Lambda X^{-1}$ then

$$\begin{aligned} e^{At} &= e^{X\Lambda t X^{-1}} \\ &= I + X\Lambda X^{-1}t + \frac{X(\Lambda t)^2 X^{-1}}{2} + \cdots \\ &= X\left[I + \Lambda t + \frac{(\Lambda t)^2}{2} + \cdots\right] X^{-1} \\ &= Xe^{\Lambda t}X^{-1} \end{aligned}$$

Since Λ is diagonal it is easy to calculate $e^{\Lambda t}$ by just applying e^{at} to each of its diagonal entries.

Ok, that was the theory, now comes the action. The diffusion of a gas between two adjacent chambers satisfies the equations:

$$\begin{cases} u' = -u + v & u(0) = 15 \\ v' = u - v & v(0) = 5 \end{cases}$$

To solve the system we use the matrix

$$A = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

Which has the diagonalization

$$A = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 0 & \\ & -2 \end{bmatrix} \frac{1}{-2} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}$$

It follows that

$$e^{At} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} e^{0t} & \\ & e^{-2t} \end{bmatrix} \frac{1}{-2} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}$$

And finally

$$\mathbf{u} = e^{At}\mathbf{u}_0 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & \\ & e^{-2t} \end{bmatrix} \frac{1}{-2} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 15 \\ 5 \end{bmatrix}$$

Calculating we arrive at the solution

$$\mathbf{u} = \begin{bmatrix} 10 + 5e^{-2t} \\ 10 - 5e^{-2t} \end{bmatrix}$$

We observe that in the limit, the two chambers will have an equal proportion of gas.

Activity for §21.

1. Consider the linear system of differential equations:

$$\begin{cases} u' = u + v & u(0) = 2 \\ v' = \quad\; 2v & v(0) = 6 \end{cases}$$

   (a) Let A be the matrix below and find e^{At}.

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix}$$

   (b) Use the previuos part to solve the linear system.

2. The rabbit and wolf populations interact according to $r' = 6r - 2w$, and $w' = 2r + w$. In other words, the rabbits breed very fast but lose some to wolf predation, and the wolves breed only modestly but better when there are rabbits.

   Suppose there are initially 30 rabbits and 30 wolves. Find formulas for the populations $r(t)$ and $w(t)$ of each species over time. After a long time, what is the ratio of rabbits to wolves?

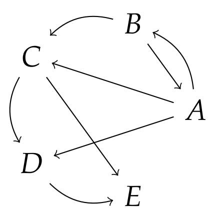
§22. Search rankings

Based on ??

We conclude our introduction to eigenvalues and eigenvectors with one more application. This time we will look at the idea of ranking search results by their "importance".

When you perform an internet search, several things happen. First, the system filters just the sites that match to your search terms. Second, the system ranks the sites according to how important they are. One portion of this second part involves an algorithm that uses eigenvectors. (Google calls it PageRank.)

To see how it works, imagine a very small internet with just five pages. We depict the pages as nods and the links between them as arrows.



We model browsing with a matrix as follows. We postulate a generic user who clicks links on the pages. Each time she opens a page, she chooses one of the links at random, with equal probability. If a page has no links, she chooses one of the other pages at random. Thus for this internet we have

$$A = \begin{bmatrix} & 1/2 & & & 1/5 \\ 1/3 & & & & 1/5 \\ 1/3 & 1/2 & & & 1/5 \\ 1/3 & & 1/2 & & 1/5 \\ & & 1/2 & 1 & 1/5 \end{bmatrix}$$

If we initially assume that the pages are visited with some probability vector $\mathbf{u}_0$, then $A^n\mathbf{u}_0$ gives the likelihood of being on each page after n clicks. The big idea is that the limit $A^\infty\mathbf{u}_0$ gives us a reasonable measurement of the relative importance of each web page!

Since A is a Markov matrix, we can calculate the limit simply by finding the eigenvector corresponding to $\lambda = 1$. Using a computer, we find that $A - I$ has the RREF form

$$A - I \sim \begin{bmatrix} 1 & & & & -9/25 \\ & 1 & & & -8/25 \\ & & 1 & & -12/25 \\ & & & 1 & -14/25 \\ & & & & \end{bmatrix}$$

We conclude that the eigenvector is

$$(9/25, 8/25, 12/25, 14/25, 1)$$

Which we can renormalize to

$$(0.13, 0.12, 0.18, 0.21, 0.37)$$

This tells us that E is the most popular, followed by D.

Activity for §22.

1. Consider the following (very very small) internet.

$$
\begin{tikzcd}
A & B \arrow[l] \arrow[d, bend left] \\
C \arrow[ur] & D \arrow[l] \arrow[u]
\end{tikzcd}
$$

   (a) Write down a 4×4 matrix A that represents this internet.
   (b) Find the eigenvector corresponding to $\lambda = 1$ and rank the pages according to importance.

2. Repeat the above activity with your own internet that you create. Make sure it has at least six pages. Try to make it so it is unclear which page is most important.

PART IV

Orthogonal bases and the SVD

§23. Orthogonal bases

Based on ILA 5th, §6.4.

We have previously discussed the concept of basis for $\mathbb{R}^n$: a family of n vectors $\mathbf{v_1}, \ldots, \mathbf{v_n}$ that are linearly independent. A basis is special because any vector in $\mathbb{R}^n$ can be written as a combination of basis vectors.

In this section we introduce the concept of *orthogonal basis*: a family of n vectors $\mathbf{v_1}, \ldots, \mathbf{v_n}$ such that $\mathbf{v_i} \perp \mathbf{v_j}$ whenever $i \neq j$. An orthogonal basis is very special because it has the shape of the standard basis (the usual axes), except possibly rotated, reflected, and scaled.

An example of an orthogonal basis for $\mathbb{R}^2$ would be $(3,4), (-4,3)$. Note that you can always normalize the vectors (make the norm 1) to get what's called an *orthonormal basis*. In this example just divide by the length of the vector: $(3/5, 4/5), (-4/5, 3/5)$.

An orthonormal basis is special algebraically. For example if Q is a matrix whose columns are an orthonormal basis, then $Q^{-1} = Q^T$.

23.1. Example. Let Q be the matrix

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

Then we can see that $QQ^T = I$.

To see why this always works, note that the columns of Q are the same as the rows of Q. When we take the dot product of a column with itself we get 1 by normality, and when we take the dot product of a column with any other column we get 0 by orthogonality.

Since we have seen the eigenvectors of a diagonalizable matrix always form a basis, it is natural to ask when the eigenvectors form an orthogonal basis.

23.2. Example. Let A be the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

Then A has eigenvalues $\lambda = 0, 5$ and corresponding eigenvectors $(1,2), (-2,1)$. These two eigenvectors are orthogonal.

The matrix in the previous example is symmetric, meaning $A = A^T$. It turns out this always guarantees the eigenvectors are orthogonal, and more.

23.3. THEOREM (Spectral theorem). *Let A be a real symmetric matrix. Then the eigenvalues of A are all real, and the eigenvectors of A are pairwise orthogonal. Thus A can be diagonalized $A = Q\Lambda Q^T$, where the columns of Q are orthonormal.*

We will briefly explain why part of this theorem is true. First we show why a symmetric matrix will have real eigenvalues (recall it is fully possible for a general matrix to have complex eigenvalues). Suppose A is symmetric and that $A\mathbf{x} = \lambda\mathbf{x}$. Then we have:

$$\lambda\|\mathbf{x}\| = \lambda\bar{\mathbf{x}}^T\mathbf{x} = \bar{\mathbf{x}}^T A\mathbf{x} = \bar{\mathbf{x}}^T A^T\mathbf{x} = (\overline{A\mathbf{x}})^T\mathbf{x} = \bar{\lambda}\bar{\mathbf{x}}^T\mathbf{x} = \bar{\lambda}\|\mathbf{x}\|$$

It follows that $\lambda = \bar{\lambda}$.

Next we show why a symmetric matrix will have orthogonal eigenvectors. Suppose A is symmtric, $\lambda_1 \neq \lambda_2$, and $A\mathbf{x_1} = \lambda_1\mathbf{x_1}$ and $A\mathbf{x_2} = \lambda_2\mathbf{x_2}$. Then we have:

$$\lambda_2\mathbf{x_1}^T\mathbf{x_2} = \mathbf{x_1}^T A\mathbf{x_2} = (A\mathbf{x_1})^T\mathbf{x_2} = \lambda_1\mathbf{x_1}^T\mathbf{x_2}$$

It follows that $\mathbf{x_1}^T\mathbf{x_2} = \mathbf{0}$.

It remains to prove that a symmetric matrix A must always have n distinct eigenvectors. We leave this to the textbook.

23.4. EXAMPLE. Let A be the matrix

$$A = \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix}$$

Find the eigenvalues and eigenvectors, then diagonalize it $Q\Lambda Q^T$.

Activity for §23.

1. Find the eigenvalues and eigenvectors of the symmetric matrices. Check that the eigenvectors are orthogonal to each other.

   (a) $\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$

   (b) $\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$

   (c) $\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$

§24. The SVD

Based on ILA 5th, §7.2.

In the previous section we saw how to diagonalize a symmetric matrix $A = Q\Lambda Q^T$. Often in applications a matrix A represents a data structure such as an image. In this case A will very large and its contents will be non-random. This usually means A will have a few eigenvalues that are large (the signal) and many that are small (the noise). If that is true, we can create a compressed version of A that is still a good approximation to A by zeroing out the smaller eigenvalues (and their corresponding eigenvectors).

Of course if A really does represent an image then it isn't very likely to be symmetric, or even square. In this section we generalize the diagonalization to work on an arbitrary matrix A. The cost is that the eigenvectors on the left and right sides of A won't be the same anymore. This is called the singular value decompositoin $A = U\Sigma V^T$.

The key idea to achieve this decomposition is that for any matrix A, the two matrices A^TA and AA^T are square and symmetric. The eigenvalues of A^TA and AA^T are the same, their square roots are called singular values, and they run down the diagonal of the Σ in the SVD. The eigenvectors of A^TA and AA^T are two orthogonal bases and they form the columns of U and V in the SVD.

Here is how it works for an $m \times n$ matrix A.

- Find find A^TA and diagonalize it in an orthonormal way. Let λ_i be the eigenvalues of A^TA and $\mathbf{v_i}$ a corresponding orthonormal family of eigenvectors. The right singular vectors of A are just the vectors $\mathbf{v_i}$. We can put them into the columns of a matrix V.
- Next define the singular values of A by $\sigma_i = \sqrt{\lambda_i}$, and put them in the diagonal of an $m \times n$ matrix Σ. (To see why the λ_i are always nonnegative, observe $\|A\mathbf{v}\|^2 = \mathbf{v}^TA^TA\mathbf{v} = \mathbf{v}^T\lambda\mathbf{v} = \lambda\|\mathbf{v}\|^2 = \lambda$.)
- Now for each nonzero singular value σ_i we define the left singular vector $\mathbf{u_i} = A\mathbf{v_i}/\sigma_i$. (To see why they are orthogonal, observe that $AA^T\mathbf{u_i} = AA^TA\mathbf{v_i}/\sigma_i = A\lambda_i\mathbf{v_i}/\sigma_i = \lambda_i\mathbf{u_i}$ so the $\mathbf{u_i}$ are eigenvectors of AA^T.)
- If necessary, extend the $\mathbf{u_i}$'s to a family of m many orthonormal vectors, and put them in the columns of the matrix U. (For simplicity we will avoid problems that use this step.)

24.1. EXAMPLE. Let A be the following matrix. Find the singular decomposition $A = U\Sigma V^T$.

$$A = \begin{bmatrix} 3 & 0 \\ 4 & 5 \end{bmatrix}$$

We first look at A^TA:

$$A^TA = \begin{bmatrix} 25 & 20 \\ 20 & 25 \end{bmatrix}$$

As in the past, we can calculate that eigenvalues of $A^T A$ are $\lambda = 45, 5$ and the corresponding eigenvectors are $(1,1)^T$ and $(-1,1)^T$.

Now the singular values are $\sigma = \sqrt{45}, \sqrt{5}$, and the right singular vectors are $(1,1)^T/\sqrt{2}$ and $(-1,1)^T/\sqrt{2}$. Finally the left singular vectors are $(1,3)^T/\sqrt{10}$ and $(-3,1)^T/\sqrt{10}$. The final SVD is thus:

$$A = \frac{1}{\sqrt{10}}\begin{bmatrix} 1 & -3 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} \sqrt{45} & \\ & \sqrt{5} \end{bmatrix}\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

Note that in the above example, the eigenvalues of the original matrix are $5, 3$. The singular decomposition has more information in its dominant singular value, and less information in its smaller singular value.

24.2. EXAMPLE. Let A be the following nonsquare matrix. Find the singular decomposition $A = U\Sigma V^T$.

$$A = \begin{bmatrix} 0 & 1 & & \\ & & 2 & \\ & & & 3 \end{bmatrix}$$

We first look at $A^T A$:

$$A^T A = \begin{bmatrix} 0 & & & \\ & 1 & & \\ & & 4 & \\ & & & 9 \end{bmatrix}$$

We can see that the eigenvalues of $A^T A$ are $\lambda = 9, 4, 1, 0$ and the corresponding eigenvectors are $(0,0,0,1)^T$, $(0,0,1,0)^T$, $(0,1,0,0)^T$, and $(0,0,0,1)^T$.

Now the singular values are $\sigma = 3, 2, 1, 0$ and the corresponding right singular vectors are just the same as above. To find the right singular vectors, for each $\sigma_i = 3, 2, 1$ we calculate $A\mathbf{v_i}/\sigma$. We get the vectors $(0,0,1)^T$, $(0,1,0)^T$, and $(1,0,0)^T$. The final SVD is thus:

$$A = \begin{bmatrix} & & 1 \\ & 1 & \\ 1 & & \end{bmatrix}\begin{bmatrix} 3 & & & \\ & 2 & & \\ & & 1 & 0 \end{bmatrix}\begin{bmatrix} & & & 1 \\ & & 1 & \\ & 1 & & \\ 1 & & & \end{bmatrix}$$

Activity for §24.

1. For each matrix A find the singular value decomposition $A = U\Sigma V^T$.
   (a) $A = \begin{bmatrix} 0 & 4 \\ 0 & 0 \end{bmatrix}$
   (b) $A = \begin{bmatrix} 0 & 4 \\ 1 & 0 \end{bmatrix}$
   (c) $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$