Math 108a Professor: Padraic Bartlett

Lecture 1: Motivation: Eigenthings and Orthogonality

Week 1 *UCSB 2014*

One particularly fun[1] aspect of planning this year's run of Math 108B has been dealing with the variations in possible background material. Between the two distinct runs of Math 108a from the fall, students who took 108a in previous years, and transfer students, the following subjects have all been seen by some but not a majority of the class:

- Eigenvalues and eigenvectors.
- Orthogonality and the Gram-Schimidt process.
- The determinant.
- The fundamental theorem of algebra.
- Various decomposition results.

Initially, this seemed like a fairly serious problem; how can you create a class that would simultaneously not bore students who've already seen these concepts, without going too fast for the students who are new to these ideas? However, in practice we've found out that this isn't actually a serious issue, because the focus of Math 108B is radically different than that of Math 108A. While 108A was best thought of as an overgrown version of Math 8 — i.e. a class that's meant to keep you working on your proofs — 108B is a class that is now assuming that you **know** proofs[2] and instead want to focus on **using** proofs to understand things!

So, while many of you will run into terms over the next week that you "know," pay attention: we're going to be discussing these things from a different perspective than you may be used to, and using them alongside terms you've not seen to get to some beautiful mathematics.

1 Eigenvalues and Eigenvectors: Basic Definitions and Examples

We first start off by stating a few of these "previously-covered" concepts, along with some basic examples to refresh the memory:

Definition. Let A be a $n \times n$ matrix with entries from some field F. (In practice, in examples we will assume that F is the real numbers $\mathbb{R}$ unless otherwise stated. It is worthwhile

[1] For certain values of "fun."

[2] Which is not to say that you should not keep working on your proofs. Developing a mathematical style is a lifelong project, and is something that I and other professors are still refining. Just like there will never be a day when LeBron is "done" with his jumper, there similarly is never a day when you're "done" with developing your mathematical style. You just keep refining it.

to occasionally think about this field being $\mathbb{C}$, when you're working on problems on your own.) We say that $\vec{v} \in F^n$ is an **eigenvector** and $\lambda \in F$ is an **eigenvalue** if

$$A\vec{v} = \lambda\vec{v}.$$

We usually ask that $\vec{v} \neq \vec{0}$, to avoid silly trivial cases.

Example. Does the matrix

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

have any real or complex-valued eigenvalues or eigenvectors? If so, find them.

Proof. We find these eigenvalues and eigenvectors via brute force (i.e. just using the definition and solving the system of linear equations.) To do this, notice that if $(x, y) \neq (0, 0)$ is an eigenvector of A and λ is a corresponding eigenvalue to (x, y), then we have

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}.$$

In other words, this is the pair of linear equations

$$\begin{aligned} x + 2y &= \lambda x, \\ 2x + y &= \lambda y. \end{aligned}$$

We want to find what values of λ, x, y are possible solutions to this equation. To find these solutions, notice that if we add these two equations together, we get

$$3x + 3y = \lambda(x + y).$$

If $x + y$ is nonzero, we can divide through by $x + y$ to get $\lambda = 3$, which is one potential value of λ. For this value of λ, we have

$$\begin{aligned} x + 2y &= 3x, \\ 2x + y &= 3y \\ \Rightarrow 2y &= 2x, \\ 2x &= 2y. \end{aligned}$$

In other words, it seems like 3 is an eigenvalue for any eigenvector of the form (x, x). This is easily verified: simply notice that

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ x \end{bmatrix} = \begin{bmatrix} 3x \\ 3x \end{bmatrix} = 3 \begin{bmatrix} x \\ -x \end{bmatrix},$$

which precisely means that 3 is an eigenvalue corresponding to vectors of the form (x, x).

Otherwise, if $x+y=0$, we can use this observation in our earlier pair of linear equations to get

$$y = \lambda x,$$
$$x = \lambda y.$$

If one of $x, y = 0$, then $x + y = 0$ forces the other to be zero, and puts us in the trivial case where $(x, y) = (0, 0)$, which we're not interested in. Otherwise, we have both $x, y \neq 0$, and therefore that $\lambda = \frac{x}{y} = \frac{y}{x}$. In particular, this forces $x = \lambda y$ and $\lambda = \pm 1$.

If we return to the two equations

$$x + 2y = \lambda x,$$
$$2x + y = \lambda y,$$

we can see that $\lambda = 1$ is not possible, as it would force $x + 2y = x$, i.e. $y = 0$. Therefore we must have $\lambda = -1$, and therefore $x = -y$. In other words, it seems like -1 is an eigenvalue with corresponding eigenvectors $(x, -x)$. Again, this is easily verified: simply notice that

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ -x \end{bmatrix} = \begin{bmatrix} -x \\ x \end{bmatrix} = -1 \begin{bmatrix} x \\ -x \end{bmatrix}$$

which precisely means that -1 is an eigenvalue corresponding to vectors of the form $(x, -x)$. □

In this case, we had a 2×2 matrix with 2 distinct eigenvalues, each one of which corresponded to a one-dimensional family of eigenvectors. This was rather nice! It bears noting that things are often not this nice, as the following example illustrates:

Example.

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

have any real or complex-valued eigenvalues or eigenvectors? If so, find them.

Proof. We proceed using the same brute-force method as before. If there was an eigenvector $(x, y) \neq (0, 0)$ and corresponding eigenvalue λ, we would have

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}.$$

In other words, we would have a solution to the system of equations

$$-y = \lambda x$$
$$x = \lambda y.$$

If one of x, y are equal to zero, then the two linear equations above force the other to be zero; this puts us in the trivial case where $(x, y) = (0, 0)$, which we're not interested in. Otherwise, we can solve each of the linear equations above for λ, and get

$$-\frac{y}{x} = \lambda$$
$$\frac{x}{y} = \lambda.$$

In other words, we have $-\frac{y}{x} = \frac{x}{y}$. Because both x and y are nonzero, this equation is equivalent to (by multiplying both sides by xy)

$$-y^2 = x^2.$$

This equation has no real-valued solutions, because any nonzero real number squared is positive. If we extend our results to complex-valued solutions, we can take square roots to get $iy = \pm x$. This gives us two possible values of λ : either i with corresponding eigenvectors $(x, -ix)$, or $-i$ with corresponding eigenvectors (x, ix). We check that these are indeed eigenvalues and eigenvectors here:

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ -ix \end{bmatrix} = \begin{bmatrix} ix \\ x \end{bmatrix} = i \begin{bmatrix} x \\ -ix \end{bmatrix},$$
$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ ix \end{bmatrix} = \begin{bmatrix} -ix \\ x \end{bmatrix} = -i \begin{bmatrix} x \\ ix \end{bmatrix}.$$

□

This was an example of a 2×2 matrix that has no real eigenvalues or eigenvectors, but **did** have two distinct complex-valued eigenvalues, each with corresponding one-dimensional families of eigenvectors. You might hope here that this is always true; i.e. that working in the complex plane is enough to always give you lots of eigenvectors and eigenvalues!

This is not true, as the following example indicates:

Example. Consider the $n \times n$ matrix

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix},$$

formed by filling the main diagonal and the stripe directly above the main diagonal with 1's, and filling the rest with zeroes. Does this matrix have any real or complex-valued eigenvalues or eigenvectors? If so, find them.

Proof. We proceed as before. Again, let $(x_1, \ldots x_n) \neq \vec{0}$ denote a hypothetical eigenvector and λ a corresponding eigenvalue; then we have

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 & 1 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \ldots \\ x_{n-1} \\ x_n \end{bmatrix} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \ldots \\ x_{n-1} \\ x_n \end{bmatrix}.$$

This gives the following set of n linear equations:

$$\begin{aligned} x_1 + x_2 &= \lambda x_1, \\ x_2 + x_3 &= \lambda x_2, \\ &\vdots \\ x_{n-1} + x_n &= \lambda x_{n-1}, \\ x_n &= \lambda x_n. \end{aligned}$$

There are at most two possiblities:

1. $\lambda = 1$. In this case, we can use our first equation $x_1 + x_2 = x_1$ to deduce that $x_2 = 0$, the second equation $x_2 + x_3 = x_2$ to deduce that $x_3 = 0$, and in general use the k-th linear equation $x_k + x_{k+1} = x_k$ to deduce that $x_{k+1} = 0$. In other words, if $\lambda = 1$, all of the entries $x_2, \ldots x_n$ are all 0. In this case, we have only the vectors $(x_1, 0, \ldots 0)$ remaining as possible candidates. We claim that these are indeed eigenvectors for the eigenvalue 1: this is easily checked, as

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 & 1 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ 0 \\ 0 \\ \ldots \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 + 0 \\ 0 \\ 0 \\ \ldots \\ 0 \\ 0 \end{bmatrix} = 1 \begin{bmatrix} x_1 \\ 0 \\ 0 \\ \ldots \\ 0 \\ 0 \end{bmatrix}.$$

2. We are trying to find any value of $\lambda \neq 1$. If this is true, then our last equation $x_n = \lambda x_n$ can only hold if $x_n = 0$. Plugging this observation into our second-to-last equation $x_{n-1} + 0 = \lambda x_{n-1}$ tellsus that x_{n-1} is also zero. In general, using induction (where our base case was proving that $x_n = 0$, and our inductive step is saying that if $x_k = 0$, we have x_{k-1} also equal to zero) we get that every x_k must be equal to zero. But this is the trivial case where $(x_1, \ldots x_n) = \vec{0}$, which we're not interested in if we're looking for eigenvectors. Therefore, there are no eigenvectors corresponding to non-1 eigenvalues.

□

So; in this case, we found a $n \times n$ matrix with only **one** eigenvalue, corresponding only to a one-dimensional space of eigenvectors! In other words, sometimes there are very very few eigenvectors or eigenvalues to be found.

2 Eigenvalues and Eigenvectors: Why We Care

So: eigenvalues and eigenvectors! We have an ad-hoc method for finding them (lots of linear equations) and have seen through examples that there are sometimes very few of them for a given matrix. We have not yet talked about **why** we care about them, though — why look for these things?

The short answer is that **they're useful**. Honestly? They're probably the most useful thing in linear algebra, and arguably in mathematics as a whole. Understanding eigenvalues and eigenvectors is fundamental to thousands of problems, ranging from the most practical of applications in physics and economics to the airiest of theoretical constructions in higher mathematics. To give a bit of an idea for what these applications look like, we do two examples here:

2.1 The Googles

Perhaps one of the most commonly used applications of eigenvalues and eigenvectors is Google's PageRank algorithm. Basically, before Google came along, web search engines were **atrocious**; many search results were not very-sophisticated massive keyword-bashes plus some well-meaning but dumb attempts to improve these results by hand. Then Brin and Page came onto the scene, with the following simple idea:

Important websites are the websites other important websites link to.

This seems kinda circular, so let's try framing this in more of a graph-theoretic framework: Take the internet. Think of it as a collection of webpages, which we'll think of as "vertices," along with a collection of hyperlinks, which we'll think of as directed lines[3] going between webpages. Call these webpages $\{v_1, \ldots v_n\}$ for shorthand, and denote the collection of webpages linking to some v_i as the set $\text{LinksTo}(v_i)$.

In this sense, if we have some quantity of "importance" $rank(v_i)$ that we're associating to each webpage i, we still want it to obey the entire "important websites are the websites other important websites link to" idea. However, we can refine what we mean by this a little bit. For example, suppose that we know a website is linked to by Google. On one hand, this might seem important — Google is an important website, after all! — but on the other hand, this isn't actually that relevant, because Google basically links to **everything.** So we don't want to simply say something is important if it's linked to by something important — we want to **weight** that importance by how many **other** things that important website links to! In other words, if you're somehow important and also only link to a few things, we want to take those links very seriously — i.e. if something is linked to by the front page of the New York Times or the Guardian, that's probably pretty important!

If we write this down with symbols and formulae, we get the following equation:

$$\text{rank}(v_i) = \sum_{v_j \in \text{LinksTo}(v_i)} \frac{\text{rank}(v_j)}{\text{number of links leaving}(v_j)}.$$

In other words, to find your rank, we add up all of the ranks of the webpages that link to you, scaling each of those links by the number of other links leaving those webpages. This

[3] A "series of tubes," if you will.

is ... still circular. But it looks mathier! Also, it's more promising from a linear-algebra point of view. Suppose that we don't think of each ranking individually, but rather take them all together as some large rank vector $\vec{r} = (\text{rank}(v_1), \ldots \text{rank}(v_n))$.

As well, instead of thinking of the links one-by-one, consider the following $n \times n$ "link-matrix" A, formed by doing the following:

- If there is a link to v_i from v_j, put a $\frac{1}{\text{number of links leaving}(v_j)}$ in the entry (i, j).
- Otherwise, put a 0.

This contains all of the information about the internet's links, in one handy $n \times n$ matrix!

Now, notice that if we multiply this matrix A by our rank vector $\vec{r}$, we get

$$A \cdot \vec{r} = \begin{bmatrix} \sum_{v_j \in \text{LinksTo}(v_1)} \frac{\text{rank}(v_j)}{\text{number of links leaving}(v_j)} \\ \sum_{v_j \in \text{LinksTo}(v_2)} \frac{\text{rank}(v_j)}{\text{number of links leaving}(v_j)} \\ \vdots \\ \sum_{v_j \in \text{LinksTo}(v_n)} \frac{\text{rank}(v_j)}{\text{number of links leaving}(v_j)} \end{bmatrix}.$$

In other words, if we the "mathy" version of the importance rule we derived earlier, we have

$$A \cdot \vec{r} = \vec{r}.$$

In other words, the vector $\vec{r}$ that we're looking for is an **eigenvector** for A, corresponding to the eigenvalue 1! The entries in this eigenvector then correspond to the "importance" ranks we were looking for. In particular, the coordinate in the vector $\vec{r}$ with the highest value corresponds to the "most important" website, and should be the first page suggested by the search engine.

Up to tweaks and small modifications, this is precisely how search works nowadays; people come up with quick and efficient ways to find eigenvectors for subgraphs of the internet that correspond to the eigenvalue 1. (Actually finding this eigenvector in an efficient manner is a problem people are still working on — there are lots of interesting techniques, some of which I hope we'll get to later in the course!)

2.2 Fibonacci

It bears noting that eigenvalues aren't only useful for applications: they have lots of theoretical and mathy uses as well! Consider the Fibonacci sequence, defined below:

Definition. The **Fibonacci sequence** $\{f_n\}_{n=1}^{\infty}$ is the sequence of numbers defined recursively as follows:

- $f_0 = 0$,
- $f_1 = 1$,
- $f_{n+1} = f_n + f_{n-1}$.

The first sixteen Fibonacci numbers are listed here:

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610 \ldots$$

Here's a question you might want to ask, at some point in time: what's f_{1001}? On one hand, you could certainly calculate this directly, by just finding all of the numbers in the sequence from 1 up to 1001. But what if you needed to calculate this **quickly**? Could you find a **closed form**?

The answer is yes, and the solution comes through using eigenvectors and eigenvalues!

Specifically, notice the following:

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} f_n \\ f_{n-1} \end{bmatrix} = \begin{bmatrix} f_n + f_{n-1} \\ f_n \end{bmatrix} = \begin{bmatrix} f_{n+1} \\ f_n \end{bmatrix}.$$

In other words, if we take a vector formed by two consecutive Fibonacci sequence elements, and multiply it by the matrix $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, we shift this sequence one step forward along the Fibonacci sequence!

Therefore, if we want to find f_{1001}, we can just calculate

$$\begin{aligned}
\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^k \cdot \begin{bmatrix} f_1 \\ f_0 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{999} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} f_1 \\ f_0 \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{999} \cdot \begin{bmatrix} f_2 \\ f_1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{998} \cdot \begin{bmatrix} f_3 \\ f_2 \end{bmatrix} \\
&\vdots \\
&= \begin{bmatrix} f_{1001} \\ f_k \end{bmatrix}.
\end{aligned}$$

So: we just need to find $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^k$! This is not an obviously easy task: multiplying the matrix by itself a thousand times seems about as difficult as adding the Fibonacci numbers to themselves that many times. However, with the help of eigenvalues, eigenvectors, and the concept of **orthogonality**, this actually can be made rather trivial!

First, let's find the eigenvalues and eigenvectors for this matrix. As before, we just use brute force: we seek $(x, y) \neq (0, 0)$ and λ such that

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}.$$

In terms of linear equations, this is just asking for x, y, λ such that

$$\begin{aligned} x + y &= \lambda x \\ x &= \lambda y. \end{aligned}$$

First, notice that if either x or y are zero, then the other is zero by the second equation, which puts us in the trivial case $(x, y) = (0, 0)$, which we don't care about.

Now, note that if we substitute the second equation into the first, we get

$$\lambda^2 y - \lambda y - y = 0.$$

If we divide through by y (which we can do, because it is nonzero,) we get

$$\lambda^2 - \lambda - 1 = 0.$$

We can use the quadratic formula to see that this has the roots

$$\frac{1 \pm \sqrt{5}}{2}.$$

These are very famous values! In particular, the quantity

$$\frac{1 + \sqrt{5}}{2}$$

is something that people have been studying for millenia — it's the famous golden ratio, denoted by the symbol φ. It has tons of weird and useful properties, but the main one I want us to note here is that

$$\frac{1 - \sqrt{5}}{2} = \frac{(1 - \sqrt{5})(1 + \sqrt{5})}{2(1 + \sqrt{5})} = \frac{1 - 5}{2(1 + \sqrt{5})} = -\frac{2}{1 + \sqrt{5}} = -\frac{1}{\varphi}.$$

In other words, the two possible values of λ are $\varphi, -\frac{1}{\varphi}$.

For each of these, we can solve for x, y: if we have $\lambda = \varphi$, then the pair of equations

$$\begin{aligned} x + y &= \varphi x \\ x &= \varphi y \end{aligned}$$

has solutions given by $(\varphi y, y)$. Similarly, if we have $\lambda = -\frac{1}{\varphi}$, then the pair of equations

$$\begin{aligned} x + y &= -\frac{1}{\varphi} x \\ x &= -\frac{1}{\varphi} y \end{aligned}$$

has solutions given by $(-\frac{y}{\varphi}, y)$.

So: we have the eigenvectors and eigenvalues! Now, notice the following very clever trick we can do with these eigenvalues and eigenvectors. First, notice that we have

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix} = \begin{bmatrix} \varphi^2 x & \frac{y}{\varphi^2} \\ \varphi x & -\frac{y}{\varphi} \end{bmatrix}.$$

This is not hard to see: if you think of the $\begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix}$ matrix as just two columns, each of which are eigenvectors, then the right-hand-side is just a result of that eigenvector property.

Now: notice that the right-hand-side can be written

$$\begin{bmatrix} \varphi^2 x & \frac{y}{\varphi^2} \\ \varphi x & -\frac{y}{\varphi} \end{bmatrix} = \begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix} \cdot \begin{bmatrix} \varphi & 0 \\ 0 & -\frac{1}{\varphi} \end{bmatrix}.$$

As a result of this, we have

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix} = \begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix} \cdot \begin{bmatrix} \varphi & 0 \\ 0 & -\frac{1}{\varphi} \end{bmatrix},$$

which implies

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix} \cdot \begin{bmatrix} \varphi & 0 \\ 0 & -\frac{1}{\varphi} \end{bmatrix} \cdot \begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix}^{-1}$$

And this is **fantastic!** Why? Well, notice that if we're calculating something like $\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^k$, we have

$$\overbrace{\begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix} \cdot \begin{bmatrix} \varphi & 0 \\ 0 & -\frac{1}{\varphi} \end{bmatrix} \cdot \cancel{\begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix}^{-1}} \cdot \cancel{\begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix}} \cdot \begin{bmatrix} \varphi & 0 \\ 0 & -\frac{1}{\varphi} \end{bmatrix} \cdot \cancel{\begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix}^{-1}} \cdot \ldots \cdot \cancel{\begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix}} \cdot \begin{bmatrix} \varphi & 0 \\ 0 & -\frac{1}{\varphi} \end{bmatrix} \cdot \begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix}^{-1}}^{\text{k copies}}.$$

$$= \begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix} \cdot \begin{bmatrix} \varphi & 0 \\ 0 & -\frac{1}{\varphi} \end{bmatrix}^k \cdot \begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix}^{-1}.$$

And this is easy to calculate — if we take a diagonal matrix and raise it to a large power, we just get the matrix formed by raising those diagonal entries to that power[4]! In other words, we get

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^k = \begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix} \cdot \begin{bmatrix} \varphi^k & 0 \\ 0 & (-\frac{1}{\varphi})^k \end{bmatrix} \cdot \begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix}^{-1}.$$

Great! If we can just find $\begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix}^{-1}$, then this is a very easy calculation: we just have to multiply **three** matrices, instead of a thousand. Much less work!

To find this inverse matrix, notice the following special property about these eigenvectors: if we take one eigenvector for φ and another for $-\frac{1}{\varphi}$, those two vectors are **orthogonal**! Specifically, recall the following definitions:

[4] This property emphatically does **not** hold for normal matrices. I.e. **NEVER** ever write something like $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^3 = \begin{bmatrix} a^3 & b^3 \\ c^3 & d^3 \end{bmatrix}$ on a test or quiz, because it is made of lies and will result in you getting no points and a lot of red ink.

Definition. Take two vectors $(x_1, \dots x_n), (y_1, \dots y_n) \in \mathbb{R}^n$. Their **dot product** is simply the sum

$$x_1y_1 + x_2y_2 + \dots x_ny_n.$$

Alternately, you can prove that the quantity above is also equal to the product

$$||\vec{x}|| \cdot ||\vec{y}|| \cos(\theta),$$

where θ is the angle between $\vec{x}$ and $\vec{y}$.

Definition. Given two vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$, we say that these two vectors are **orthogonal** if their dot product $\vec{v} \cdot \vec{w} = 0$. Note that geometrically, if both of these vectors have nonzero length, this can only happen if the cosine of the angle between these two vectors is zero: i.e. if these two vectors meet at a right angle!

With these definitions restated, it is not hard to check that an eigenvector $(\varphi x, x)$ for φ and an eigenvector $(-\frac{y}{\varphi}, y)$ for $-\frac{1}{\varphi}$ are orthogonal: we just calculate

$$(\varphi x, x) \cdot (-\frac{y}{\varphi}, y) = -xy + xy = 0.$$

Why do we care? Well: notice that if we look at the product

$$\begin{bmatrix} \varphi x & x \\ -\frac{y}{\varphi} & y \end{bmatrix} \cdot \begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix} = \begin{bmatrix} (\varphi x, x) \cdot (\varphi x, x) & (\varphi x, x) \cdot (-\frac{y}{\varphi}, y) \\ (-\frac{y}{\varphi}, y) \cdot (\varphi x, x) & (-\frac{y}{\varphi}, y) \cdot (-\frac{y}{\varphi}, y) \end{bmatrix},$$

we get that the upper-right and bottom-left entries are 0, because those vectors are orthogonal! Therefore, we have that this product is

$$\begin{bmatrix} \varphi x & x \\ -\frac{y}{\varphi} & y \end{bmatrix} \cdot \begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix} = \begin{bmatrix} \varphi^2 x^2 + x^2 & 0 \\ 0 & \frac{y^2}{\varphi^2} + y^2 \end{bmatrix}.$$

So, in particular, if we wanted to make this product the identity matrix, we could just pick x, y such that

$$x^2(1+\varphi^2) = 1 \quad \Leftarrow \quad x = \frac{1}{\sqrt{1+\varphi^2}}$$
$$y^2(1+\frac{1}{\varphi^2}) = 1 \quad \Leftarrow \quad y = \frac{1}{\sqrt{1+\frac{1}{\varphi^2}}} = \frac{\varphi}{\sqrt{1+\varphi^2}}.$$

In other words: we have just calculated $\begin{bmatrix} \varphi x & -\frac{y}{\varphi} \\ x & y \end{bmatrix}^{-1}$ for free! In the case where we set x, y as above, it's just the transpose of this matrix: i.e. $\begin{bmatrix} \varphi x & x \\ -\frac{y}{\varphi} & y \end{bmatrix}$!

So: we've just proven the following formula:

$$\begin{aligned}\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}^k &= \begin{bmatrix}\frac{\varphi}{\sqrt{1+\varphi^2}} & -\frac{1}{\sqrt{1+\varphi^2}}\\ \frac{1}{\sqrt{1+\varphi^2}} & \frac{\varphi}{\sqrt{1+\varphi^2}}\end{bmatrix} \cdot \begin{bmatrix}\varphi^k & 0\\ 0 & (-\frac{1}{\varphi})^k\end{bmatrix} \cdot \begin{bmatrix}\frac{\varphi}{\sqrt{1+\varphi^2}} & \frac{1}{\sqrt{1+\varphi^2}}\\ -\frac{1}{\sqrt{1+\varphi^2}} & \frac{\varphi}{\sqrt{1+\varphi^2}}\end{bmatrix}\\ &= \begin{bmatrix}\frac{\varphi}{\sqrt{1+\varphi^2}} & -\frac{1}{\sqrt{1+\varphi^2}}\\ \frac{1}{\sqrt{1+\varphi^2}} & \frac{\varphi}{\sqrt{1+\varphi^2}}\end{bmatrix} \cdot \begin{bmatrix}\frac{\varphi^{k+1}}{\sqrt{1+\varphi^2}} & \frac{\varphi^k}{\sqrt{1+\varphi^2}}\\ -\frac{(-\frac{1}{\varphi})^k}{\sqrt{1+\varphi^2}} & \frac{\varphi(-\frac{1}{\varphi})^k}{\sqrt{1+\varphi^2}}\end{bmatrix}\\ &= \begin{bmatrix}\frac{\varphi^{k+2}+(-\frac{1}{\varphi})^k}{1+\varphi^2} & \frac{\varphi^{k+1}-\varphi(-\frac{1}{\varphi})^k}{1+\varphi^2}\\ \frac{\varphi^{k+1}-\varphi(-\frac{1}{\varphi})^k}{1+\varphi^2} & \frac{\varphi^k+\varphi^2(-\frac{1}{\varphi})^k}{1+\varphi^2}\end{bmatrix}.\end{aligned}$$

If we multiply the numerator and denominator in each fraction by $\frac{1}{\varphi}$, we get

$$\begin{bmatrix}\frac{\varphi^{k+1}-(-\frac{1}{\varphi})^{k+1}}{\frac{1}{\varphi}+\varphi} & \frac{\varphi^k-(-\frac{1}{\varphi})^k}{\frac{1}{\varphi}+\varphi}\\ \frac{\varphi^k-(-\frac{1}{\varphi})^k}{\frac{1}{\varphi}+\varphi} & \frac{\varphi^{k-1}-(-\frac{1}{\varphi})^{k-1}}{\frac{1}{\varphi}+\varphi}\end{bmatrix}.$$

We do this because

$$\varphi + \frac{1}{\varphi} = \frac{1+\sqrt{5}}{2} + \frac{2}{1+\sqrt{5}} = \frac{1+\sqrt{5}}{2} - \frac{1-\sqrt{5}}{2} = \sqrt{5},$$

which allows us to simplify the above into

$$\frac{1}{\sqrt{5}}\begin{bmatrix}\varphi^{k+1}-(-\frac{1}{\varphi})^{k+1} & \varphi^k-(-\frac{1}{\varphi})^k\\ \varphi^k-(-\frac{1}{\varphi})^k & \varphi^{k-1}-(-\frac{1}{\varphi})^{k-1}\end{bmatrix}.$$

Whew! Ok. In the end, we've finally proven the following theorem:

Theorem.

$$\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}^k = \frac{1}{\sqrt{5}}\begin{bmatrix}\varphi^{k+1}-(-\frac{1}{\varphi})^{k+1} & \varphi^k-(-\frac{1}{\varphi})^k\\ \varphi^k-(-\frac{1}{\varphi})^k & \varphi^{k-1}-(-\frac{1}{\varphi})^{k-1}\end{bmatrix}.$$

As a particular consequence, we have

$$\begin{aligned}\begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}^k \cdot \begin{bmatrix}f_1\\ f_0\end{bmatrix} &= \begin{bmatrix}1 & 1\\ 1 & 0\end{bmatrix}^k \cdot \begin{bmatrix}1\\ 0\end{bmatrix}\\ &= \frac{1}{\sqrt{5}}\begin{bmatrix}\varphi^{k+1}-(-\frac{1}{\varphi})^{k+1}\\ \varphi^k-(-\frac{1}{\varphi})^k\end{bmatrix}.\end{aligned}$$

In other words, we have

$$f_{k+1} = \frac{\varphi^{k+1}-(-\frac{1}{\varphi})^{k+1}}{\sqrt{5}}.$$

This is exactly what we wanted! A way to calculate the Fibonacci numbers quickly, without having to calculate everything else on the way. Success!