Math 1b	TA: Padraic Bartlett

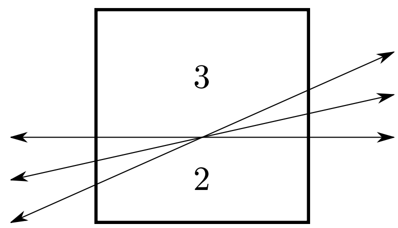
Recitation 7: Diagonalization!

Week 7 — *Caltech 2011*

1 Random Question

Suppose you have a square S and nine lines $l_1, \dots l_9$, such that each line divides S into a pair of quadrilaterals, so that the ratio formed by the areas of these quadrilaterals is 2:3.

Show that there must be three of these lines that meet at a common point.



2 Homework comments

- Section average: 70/80, or about 87.5%. This is roughly identical/slightly higher than the course average.
- People did really really well! Pretty much the main source of points lost here wasn't people failing to understand anything; rather, it was just people not attaching their (Mathematica/Wolfram Alpha/Matlab/Maple/Hex) code[1]! So, I'm pretty pleased.

3 Diagonalization: Theorems, Definitions, and Motivations

To start off, we restate the two theorems we use in diagonalizing matrices, and review all of our relevant definitions (including, say, just what diagonalization <u>is</u>.)

Theorem 1 *If A is a $n \times n$ matrix with distinct eigenvalues $\lambda_1, \dots \lambda_k$, such that*

$$\sum_{i=1}^{k} \textit{geometric multiplicity}(\lambda_i) = n,$$

*then we can find a basis for $\mathbb{R}^n$ made entirely out of vectors which are eigenvectors for A. (As an aside: the **geometric multiplicity** of an eigenvalue is the dimension of the eigenspace associated to λ_i. Equivalently, it is the largest number of linearly independent vectors you can find that are all eigenvectors for A, with λ_i as their eigenvalue.)*

[1] Relatedly: ATTACH YOUR CODE.

Furthermore, if $\{\mathbf{v}_{i,j}\}_{j=1}^{n_i}$ *is a basis for the eigenspace of* λ_i*, for every* i*, we can explicitly write out our basis for* $\mathbb{R}^n$ *as the following union:*

$$\textit{eigenvector basis for } \mathbb{R}^n := \bigcup_{i=1}^{k} \left(\bigcup_{j=1}^{n_i} \mathbf{v}_{i,j} \right).$$

The above theorem is kind of an odd thing: when would we want to find such a basis? What would we do with it?

The answer to the above questions, in a word, is **diagonalization**! and in a theorem, is our next result:

Theorem 2 *Suppose that* A *is a* $n \times n$ *matrix such that we can find a basis* $\{\mathbf{e}_1, \ldots \mathbf{e}_n\}$ *for* $\mathbb{R}^n$ *made out of* A*'s eigenvectors. (In other words, suppose that* A *is a matrix to which we can apply our above theorem!)*

Then A *is* ***diagonalizable****! Specifically, there is an invertible matrix*

$$E = \begin{pmatrix} | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \\ | & | & & | \end{pmatrix}$$

made out of the eigenvectors $\mathbf{e}_1, \ldots \mathbf{e}_n$*, such that such that*

$$E^{-1}AE = \begin{pmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ 0 & 0 & \lambda_3 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \lambda_n \end{pmatrix},$$

where the numbers $\lambda_1, \ldots \lambda_n$ *are the eigenvalues corresponding to the eigenvectors* $\mathbf{e}_1, \ldots \mathbf{e}_n$*.*

So: why would we ever want to do this? Well, one excellent application is for taking **powers** of matrices! For example, suppose we have a $n \times n$ matrix A, and we want to find what happens to A when we raise it to some power – say, $3 \cdot 10^8$. How can we calculate this?

Well, one (naïve) way to try this is just to perform matrix multiplication. How many operations will this take?

Well: whenever we multiply two matrices, to find the entry in the (i, j)-th spot we have to multiply the i-th row with the j-th column. This will require us to perform $2n - 1$ operations: n operations to multiply the relevant matrix entries together, and $n - 1$ more operations to add them all up. Our matrix is a $n \times n$ grid, so we'll have to perform the above process n^2 many times per pair of matrices multipled; and we're performing $3 \cdot 10^8 - 1$ many such instances of matrix multiplication.

So, in total, we're performing about

$$6n^3 \cdot 10^8$$

many operations to find $A^{3\cdot 10^8}$, which for many reasonable values of n will quickly become massive and fairly intractible.

However, suppose furthermore that we know A is diagonalizable: i.e. there is a matrix E such that $E^{-1}AE = \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix}$. Then, by solving for A in the equality above, we can write

$$\begin{aligned}
(A)^{3\cdot 10^8} &= \left(E \cdot \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix} \cdot E^{-1} \right)^{3\cdot 10^8} \\
&= E \cdot \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix} \cdot \cancel{E^{-1} \cdot E} \cdot \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix} \cdot \cancel{E^{-1} \cdot E} \dots \\
&= E \cdot \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix}^{3\cdot 10^8} \cdot E^{-1} \\
&= E \cdot \begin{pmatrix} \lambda_1^{3\cdot 10^8} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^{3\cdot 10^8} \end{pmatrix} \cdot E^{-1},
\end{aligned}$$

which we can find by performing simply n exponentiations and then multiplying together three matrices (each of which takes about $2n^2$ many operations, as noted before.) So, we've reduced $6n^3 \cdot 10^8$ many operations to about $6n^3 + n$-many operations; an improvement by a factor of almost 10^8! This is a serious algorithmic improvement; for example, suppose you took a program whose run time was about three years, and improved its run time by a factor of 10^8. Your program will now run in *under a second.*

Furthermore, we can use diagonalization to find *roots* of matrices, something that otherwise we don't really have any well-established methods to *do*. For example, suppose that we had a matrix A, and we wanted to find $A^{1/2}$: i.e. a matrix such that when you multiply it by itself, you get A.

In general, this is really hard! However, if A is diagonalizable as $E \cdot \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n \end{pmatrix} \cdot E^{-1}$, we can easily express $A^{1/2}$ as

$$E \cdot \begin{pmatrix} \lambda_1^{1/2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^{1/2}, \end{pmatrix} \cdot E^{-1}.$$

(It bears noting that this will sometimes give you a matrix with *complex* entries, in the cases that you're finding even roots and your matrix has negative eigenvalues.)

4 Diagonalization: A Carefully Worked Example

To illustrate how we use these methods, let's work an example!

Example. Find A^{301}, A^{302} and $A^{1/301}$, for

$$A = \begin{pmatrix} 3 & 4 & -2 \\ -4 & -5 & 2 \\ -4 & -4 & 1 \end{pmatrix}.$$

Solution. As suggested by our discussion above, multiplying A by itself 301 times in a row might be a bit tedious. So let's try to diagonalize A! By our theorems, this means we want to do the following:

1. First, we want to find all of A's eigenvalues.
2. Then, for each of these eigenvalues, we want to find a basis for their corresponding eigenspaces.
3. How many vectors did we find in the above step? If the answer is the dimension of A (in this case 3, as A in our example is a 3×3 matrix), then we can apply our theorems to write $A = E \cdot \begin{pmatrix} \lambda_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n, \end{pmatrix} \cdot E^{-1}$, where $\lambda_1 \dots \lambda_n$ are A's eigenvalues and $E = \begin{pmatrix} | & | & & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \dots & \mathbf{e}_n \\ | & | & & | \end{pmatrix}$ is the matrix made out of the eigenvectors we found in step (2).
4. Finally, we can calculate A^{301} by simply finding $A = E \cdot \begin{pmatrix} \lambda_1^{301} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \lambda_n^{301}, \end{pmatrix} \cdot E^{-1}$, and do a similar calculation to find $A^{1/301}$.

So: let's try this! First, to find A's eigenvalues, we calculate its characteristic polynomial:

$$\begin{aligned} p_A(\lambda) &= \det(\lambda I - A) \\ &= (\lambda - 3)(\lambda^2 + 4\lambda + 3) + 4(4\lambda + 4) + 2(-4\lambda - 4) \\ &= \lambda^3 + \lambda^2 - \lambda - 1 \\ &= (\lambda + 1)^2(\lambda - 1). \end{aligned}$$

So A's two eigenvalues are -1 and 1. What are their associated eigenspaces?
For -1: we know that

$$\begin{aligned} E_{-1} &= \text{nullspace}\,(A - (-1)I) \\ &= \text{nullspace} \begin{pmatrix} 4 & 4 & -2 \\ -4 & -4 & 2 \\ -4 & -4 & 2 \end{pmatrix}. \end{aligned}$$

Because all of the rows of this matrix are constant multiples of the first row, we know that its rowspace has dimension 1. Therefore, its nullspace has to have dimension $3 - 1 = 2$; so it suffices to find two linearly independent vectors $\mathbf{e}_1, \mathbf{e}_2$ such that $A\mathbf{e}_i = 0$.

To do this: you can either use the methods in our handout on (row/null)spaces, or simply notice that in the matrix above,

- the first column minus the second column is the zero vector, as is
- the first column plus two times the third column.

In other words, $(A+I)\cdot(1,-1,0)^T = \mathbf{0}$ and $(A+I)\cdot(1,0,2)^T = \mathbf{0}$; therefore, both of these vectors are in the nullspace of $A+I$. These two vectors are linearly independent; therefore, we know that they form a basis for this nullspace, and thus E_{-1}.

So we just need to find a basis for E_1, which is

$$\begin{aligned} E_{-1} &= \text{nullspace}\,(A - (1)I) \\ &= \text{nullspace}\begin{pmatrix} 2 & 4 & -2 \\ -4 & -6 & 2 \\ -4 & -4 & 0 \end{pmatrix}. \end{aligned}$$

By inspection, one way to combine the columns in the matrix above to get zero is $(-1, 1, 1)$. Are there any others?

In fact, up to constant multiples, there aren't! One way to see this is to calculate the rank of the matrix $A - I$,and see that is is 2; another way to see this is to notice that we've already found a subspace of R^3 of dimension 2 that's made out of eigenvectors for -1, and therefore there's at most one more dimension left for the 1-eigenvectors.

After either bit of reasoning, we're free to conclude that E_1 has $(-1, 1, 1)$ as a basis.

So: we've found three vectors! Therefore, by our first theorem, these vectors form a basis for $\mathbb{R}^3$, and thus by our second theorem we can write

$$A = \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix}^{-1}.$$

Consequently, we can write

$$
\begin{aligned}
A^{301} &= \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1^{301} & 0 & 0 \\ 0 & (-1)^{301} & 0 \\ 0 & 0 & (-1)^{301} \end{pmatrix} \cdot \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix}^{-1} \\
&= \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix}^{-1} \\
&= A, \\
A^{1/301} &= \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1^{1/301} & 0 & 0 \\ 0 & (-1)^{1/301} & 0 \\ 0 & 0 & (-1)^{1/301} \end{pmatrix} \cdot \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix}^{-1} \\
&= \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \cdot \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix}^{-1} \\
&= A, \text{ and} \\
A^{302} &= \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1^{302} & 0 & 0 \\ 0 & (-1)^{302} & 0 \\ 0 & 0 & (-1)^{302} \end{pmatrix} \cdot \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix}^{-1} \\
&= \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & 2 \end{pmatrix}^{-1} \\
&= EIE^{-1} = EE^{-1} = I.
\end{aligned}
$$

As you may have guessed, the properties above aren't specific to the numbers 301 and 302: A raised to any odd power is itself, and A raised to any even power is the identity matrix! Furthermore, the only property we used about A to conclude this was that A had n eigenvalues that were all ± 1; for any such matrix, we've just proven that it is in fact orthogonal! Cool, right?

5 Diagonalization: Can It Always Be Done?

One cautionary remark to make is that not every matrix can be diagonalized. For example, the $n \times n$ matrix

$$
\begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix}
$$

has 1 as its only eigenvalue and $(0, \ldots 0, x)$ as its only eigenvectors. As we clearly cannot span $\mathbb{R}^n$ with such vectors whenever $n > 1$, we cannot apply either of our theorems to diagonalize this matrix. (As it turns out, it's not just that our methods are lacking; there is in fact no way to diagonalize this matrix at all! We call such matrices **defective** matrices.)