Princeton University MAT 202 Spring 2008

Willie W. Wong

Apr 14 - 18

APR 14, 2008

- The topic of the day is diagonalization.
- Again, we motivate our discussion by the study of the linear discrete dynamical systems. Remember that we can let the system be described by a matrix A through the recursive relation

  $$\boldsymbol{x}(t+1) = A\boldsymbol{x}(t)$$

  What we seek, remember, is a closed form solution for the entry

  $$\boldsymbol{x}(t)$$

  that depends only on the initial data and the time t, and not on any of the intermediate steps. Recall now that the linear discrete dynamical system above is equally well described by the equation

  $$\boldsymbol{x}(t) = A^t\boldsymbol{x}(0)$$

  So essentially, the difficulty is finding the matrix of transformation A^t: in general, we will need to compute $A, A^2, A^3 \ldots$ so on until we get to A^t. If there is a way to more easily compute A^t, then we will be definitely able to write down the system in closed form.
- Now, suppose that D is a diagonal matrix: this means that all the entries *not* on the diagonal must be zero. (Though we make no requirements on whether the diagonal entries are zero or not.) Such a matrix, say $M \times M$, looks like

  $$D = \begin{bmatrix} d_{11} & 0 & 0 & \ldots & 0 \\ 0 & d_{22} & 0 & \ldots & 0 \\ 0 & 0 & d_{33} & \ldots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & d_{MM} \end{bmatrix}$$

  Let's do a quick experiment: what is D^2? (Multiply it out). How'bout D^3?

- In fact, for the diagonal matrix D, we can write down

  $$D^t = \begin{bmatrix} d_{11}^t & 0 & 0 & \dots & 0 \\ 0 & d_{22}^t & 0 & \dots & 0 \\ 0 & 0 & d_{33}^t & \dots & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \dots & d_{MM}^t \end{bmatrix}$$

  So taking the power of a diagonal matrix is easy!

- Now what else do we know about powers of matrices? Remember in chapter 3, we learned a useful property of similar matrices: if A and B are similar matrices, in particular, say $A = SBS^{-1}$. Then we know that their powers are also similar by the relation

  $$A^t = SB^tS^{-1}$$

- So this means that should the matrix A be similar to a diagonal matrix, the powers A^t will be very easy to calculate!

- Definition: A $M \times M$ matrix A is called *diagonalizable* if A is similar to a diagonal matrix D: there exists an invertible matrix S such that $S^{-1}AS$ is a diagonal matrix.

- So how do we determine if A is diagonalizable? Let's look back on the condition that $S^{-1}AS = D$: this means that

  $$A = SDS^{-1}$$

  Now since S is invertible, its column vectors, say $\boldsymbol{v}_1, \dots, \boldsymbol{v}_M$ form a basis of $\mathbb{R}^M$. It is a simple calculation to see that if

  $$S = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \dots & \boldsymbol{v}_M \end{bmatrix}$$

  then

  $$S^{-1}\boldsymbol{v}_i = \boldsymbol{e}_i$$

  (this follows from the fact $S^{-1}S = I_M$). So we calculate

  $$A\boldsymbol{v}_i = SDS^{-1}\boldsymbol{v}_i = SD\boldsymbol{e}_i = d_{ii}S\boldsymbol{e}_i = d_{ii}\boldsymbol{v}_i$$

  and see that the vectors $\boldsymbol{i}$ are, in fact, eigenvectors of A. We conclude: if A is a square matrix and S is an invertible matrix such that $S^{-1}AS$ is diagonal, then the column vectors of S are eigenvectors of A. (Note the order of $S^{-1}AS$ and not SAS^{-1}.)

- Now, remembering that the column vectors of S form a basis, we reach the following facts
  - A square matrix A is diagonalizable if and only if we can find an eigenbasis for A.
  - If an $M \times M$ matrix A has M distinct eigenvalues, then A is diagonalizable.

- Based on this, we have an algorithm to diagonalize A (or to determine that A cannot be diagonalized).
  - First, given the square $M \times M$ matrix A, we find its eigenvalues: we solve the characteristic equation $f_A(\lambda) = \det(A - \lambda I_M) = 0$.
  - If the sum of algebraic multiplicities of the real roots comes out to be strictly less than M: stop! The matrix cannot be diagonalized. (Example, rotation by 90 degrees in the plane).
  - Otherwise, for each eigenvalue λ_j, find a basis for its eigenspace $E_\lambda = \ker(A - \lambda_j I_M)$.
  - Add up the geometric multiplicities (dimensions of E_λs) of the eigenvalues: if this sums to strictly less than M, stop! The matrix is not diagonalizable.
  - If the geometric multiplicities add up, then we can write all the M basis vectors of the various eigenspaces in a list, say $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M$. Then defining S as

    $$S = \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_M \end{bmatrix}$$

    we have that A is diagonalizable: $S^{-1}AS = D$, where D is a diagonal matrix with the entries d_{ii} being the corresponding eigenvalue of the eigenvector $\boldsymbol{v}_i$.

- Example: Back to the Ninja-Pirate-Cowboys example from last week. The matrix in question is

  $$A = \begin{bmatrix} 0.78 & 0.06 & 0.16 \\ -0.14 & 1.22 & -0.08 \\ 0 & 0 & 1 \end{bmatrix}$$

  Recall that on Friday we wrote down its characteristic polynomial as

  $$f_A(\lambda) = (1-\lambda)(1.2-\lambda)(0.8-\lambda)$$

  In this particular case, we can skip step 2, since we just found that A, a 3×3 matrix, has 3 distinct eigenvalues, and thus must be diagonalizable. To find its eigenvectors, we solve for the kernels of the respective eigenvalues using Gauss Jordan elimination: for example, for the 0.8 eigenvalue

  $$\left[\begin{array}{ccc|c} -0.02 & 0.06 & 0.16 & 0 \\ -0.14 & 0.42 & -0.08 & 0 \\ 0 & 0 & 0.2 & 0 \end{array}\right] \Rightarrow \left[\begin{array}{ccc|c} 1 & -3 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

  from which we take $P = s$ to be the free parameter and find that the solution is

  $$\boldsymbol{v}_{0.8} = \begin{bmatrix} N \\ P \\ C \end{bmatrix} = s \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$$

Similarly we solve to see that

$$\boldsymbol{v}_1 = t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

and

$$\boldsymbol{v}_{1.2} = r \begin{bmatrix} 1 \\ 7 \\ 0 \end{bmatrix}$$

So in the end, A is diagonalizable with

$$S = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 1 & 7 \\ 0 & 1 & 0 \end{bmatrix}$$

and

$$D = S^{-1}AS = \begin{bmatrix} 0.8 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1.2 \end{bmatrix}$$

- Example: when is the matrix

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

  diagonalizable? We begin by looking at its eigenvalues. Its characteristic equation is

  $$\det(A - \lambda I_3) = (1-\lambda)^3 + c^3 = 0$$

  by directly calculating the determinant of the 3×3 matrix. Remembering that we can factor the expression

  $$a^3 + b^3 = (a+b)(a^2 - ab + b^2)$$

  we have that the characteristic polynomial can be factored as

  $$f_A(\lambda) = [(1-\lambda)+c][(1-\lambda)^2 - (1-\lambda)c + c^2] = (1+c-\lambda)[(1-c+c^2) + (c-2)\lambda + \lambda^2]$$

  So we know that we have 1 eigenvalue of the form $\lambda = 1 + c$. To deal with the second term, we apply the quadratic discriminant and find that the quadratic expression has real solutions only when

  $$(c-2)^2 - 4 \cdot 1 \cdot (1 - c + c^2) \geq 0$$

  we evaluate and get

  $$c^2 - 4c + 4 - 4 + 4c - 4c^2 = -4c^2 \geq^? 0$$

  The only time that we have real solutions is when $c = 0$. So we can say

  - when $c \neq 0$, the only real root of $f_A(\lambda)$ is $\lambda = 1 + c$ with algebraic multiplicity 1, and therefore the matrix is not diagonalizable.

  - when $c = 0$, the characteristic polynomial becomes $f_A(\lambda) = (1-\lambda)^3$ so $\lambda = 1$ is a real root with algebraic multiplicity 3. Normally what we would do at this step is try to find the kernel of $A - I_3$ to see whether the geometric multiplicity is also 3, but in our case, if $c = 0$, A is exactly the identity matrix in three dimensions, and is already diagonalized, so there is nothing to check.

  so we conclude that A is diagonalizable only when $c = 0$, and when $c = 0$, A is already diagonalized.

APR 16, 2008

No notes today: In class quiz on chapters 5 and 6 and sections 7.1 and 7.2.

APR 18, 2008

- Today we study more closely the case where a polynomial does not have "enough" real eigenvalues. (The example to keep in mind is the case of a two dimensional rotation matrix by an angle θ.)
- Remember something that, by the fundamental theorem of algebra, any degree M polynomial will have M *complex* roots counted with multiplicity. That is where the "missing" eigenvalues go.
- Let's look at an example: Let A be the 2×2 matrix

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

  The characteristic polynomial is

  $$f_A(\lambda) = (1-\lambda)(1/3-\lambda) + 10/9 = 13/9 - 4\lambda/3 + \lambda^2$$

  The discriminant $16/9 - 52/9 = -4 < 0$ so this matrix has no real eigenvalues. The characteristic polynomial, however, has two complex roots:

  $$\lambda = 2/3 \pm i$$

- Before examining the complex roots, let's first look at what happens to the phase diagram of a dynamical system with this matrix as its flow. In the figure we can see that the system has a spiraling behaviour. For dynamical systems in 2 dimensions, this spiraling behaviour identifies when the matrix for the flow has only complex eigenvalues.

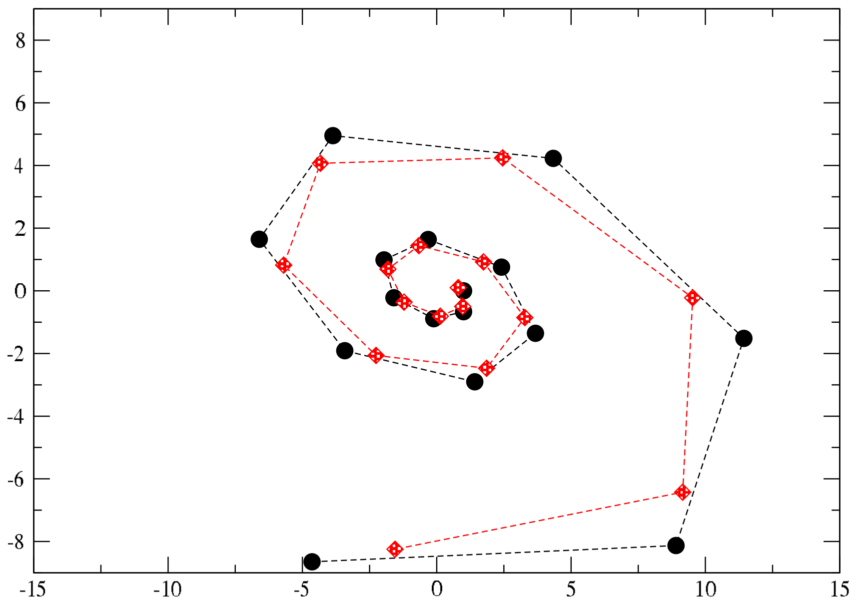

Figure 1: A plot of time evolution of the system $\boldsymbol{x}(t+1) = A\boldsymbol{x}(t)$ using the starting points $(1, 0)$ (black) and $(0.8, 0.1)$ (red). The dashed lines connecting the plotted points is just for illustrative purposes to demonstrate the adjacent times.

- If we project the spiraling picture onto either of its coördinate axes, we see that spiraling in 2 dimensions is the same as oscillation in one dimensions: if we plot the x-coördinate against time, we will see it go up-and-down-and-up-and-down, like a wave. This oscillation is, in general, the hallmark of complex eigenvalues.

- (Hope that I don't need to review complex numbers: point students to the Book for explanation.)

- Another good example: the rotation matrix by angle θ. If we plot the phase diagram, we see that the trajectories sit on concentric circles: this is a special case of a spiral where the radius does not change.

- In fact, qualitatively, a spiraling behaviour looks like one where the transformation between time t and $t + 1$ is some sort of rotation combined with some sort of scaling. This qualitative intuition will be made more precise later.

- Let's go back to the problem of diagonalizing a matrix. Remember from Monday that a matrix is (real) diagonalizable if its geometric multiplicities sum up to the size of the matrix. In the algorithm given on Monday, we see that there are two failure modes:
  - When the matrix does not have enough real eigenvalues; the sum of the algebraic multiplicities of the real eigenvalues do not add up to the size of the matrix.
  - When the matrix has repeated eigenvalues with not enough geometric multiplicity.

Here we are going to solve the first of the failure modes by allowing complex eigenvalues: by the fundamental theorem of algebra, we will have exactly enough algebraic multiplicities from the real and complex eigenvalues to make diagonalization a possibility. Of course, this still doesn't solve the second problem of repeated eigenvalues with not enough geometric multiplicity: this is something for more advanced linear algebra courses (goes under the name of Jordan canonical form if you are interested). To illustrate: the matrix

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

is upper triangular, so its only eigenvalue, 0, is real with multiplicity two. But its kernel ($= E_0$) is clearly not 2-dimensional, so this matrix cannot be diagonalized, even allowing complex numbers.

- Fortunately, given a random matrix A, it is more likely to have M distinct complex eigenvalues than to have repeated eigenvalues, so by working over the complex numbers, we solve a rather large portion of the problem.

- Remember a statement from last week: the trace of a matrix is the same as the sum of all the eigenvalues with (algebraic) multiplicity, and the determinant of a matrix is the product of the eigenvalues. The proof of this statement only depended on the fundamental theorem of algebra. So the same statement is true if we allow ourselves to look at complex eigenvalues.

- Now, diagonalization over $\mathbb{C}$, the complex numbers. The process is exactly the same as when we do it over the real numbers, except now we have to perform calculations using complex numbers. Example: going back to the matrix

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

above, we've already calculated that its eigenvalues are $2/3 \pm i$. Now let's find the eigenvectors for those eigenvalues: for $E_{2/3+i}$, we try to solve the linear equations

$$\left[\begin{array}{cc|c} 1/3-i & 5/3 & 0 \\ -2/3 & -1/3-i & 0 \end{array}\right] \Rightarrow \left[\begin{array}{cc|c} 1 & 5/(1-3i) & 0 \\ -2 & -1-3i & 0 \end{array}\right] \Rightarrow \left[\begin{array}{cc|c} 1 & (1+3i)/2 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

We take $y = s$ to be the free parameter, and the solution is

$$\boldsymbol{v}_+ = \begin{bmatrix} x \\ y \end{bmatrix} = s \begin{bmatrix} -(1+3i)/2 \\ 1 \end{bmatrix}$$

Similarly we can solve for the $E_{2/3-i}$

$$\left[\begin{array}{cc|c} 1/3+i & 5/3 & 0 \\ -2/3 & -1/3+i & 0 \end{array}\right] \Rightarrow \left[\begin{array}{cc|c} 1 & 5/(1+3i) & 0 \\ -2 & -1+3i & 0 \end{array}\right] \Rightarrow \left[\begin{array}{cc|c} 1 & (1-3i)/2 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

and the solution is

$$\boldsymbol{v}_- = \begin{bmatrix} x \\ y \end{bmatrix} = s \begin{bmatrix} -(1-3i)/2 \\ 1 \end{bmatrix}$$

We notice that the two eigenvectors $\boldsymbol{v}_+$ and $\boldsymbol{v}_-$ are *complex conjugates of each other*. So given an arbitrary initial vector $\boldsymbol{w} = \begin{bmatrix} a \\ b \end{bmatrix}$, we want to express it in terms of $\boldsymbol{v}_+$ and $\boldsymbol{v}_-$:

$$\begin{bmatrix} \boldsymbol{v}_+ & \boldsymbol{v}_- \end{bmatrix} \begin{bmatrix} c_+ \\ c_- \end{bmatrix} = \boldsymbol{w} \Rightarrow \begin{bmatrix} 1+3i & 1-3i & | & -2a \\ 1 & 1 & | & b \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & | & b/2 + (2a+b)i/6 \\ 0 & 1 & | & b/2 - (2a+b)i/6 \end{bmatrix}$$

The important point is that c_+ and c_- are, again, complex conjugates.

- (Please be careful when doing complex arithmetic.)
- In general: for a real 2×2 matrix that has no real eigenvalues:
  - By the quadratic formula, the two complex eigenvalues are complex conjugates, so we can write the two complex eigenvalues as λ and $\bar{\lambda}$.
  - Since the two complex eigenvalues are conjugates, they are not the same (if $\lambda = \bar{\lambda}$, then λ is a real number). So each has exactly a 1-dimensional eigenspace. The eigenvectors that span the two eigenspaces are also complex conjugates, meaning that if $\boldsymbol{v}$ is an eigenvector in E_λ, then $\bar{\boldsymbol{v}}$ is an eigenvector in $E_{\bar{\lambda}}$.
    - A quick proof of this fact: suppose $\boldsymbol{v}$ is an eigenvector with complex eigenvalue λ, we can write

      $$\boldsymbol{v} = \boldsymbol{z} + i\boldsymbol{u}$$

      where $\boldsymbol{z}, \boldsymbol{u}$ are vectors with real entries. We can also write $\lambda = r + si$ where r and s are real numbers. And we have that

      $$A\boldsymbol{v} = \lambda\boldsymbol{v} = r\boldsymbol{z} - s\boldsymbol{u} + i(s\boldsymbol{z} + r\boldsymbol{u})$$

      Since A is a real matrix, the above decomposition tells us that

      $$A\boldsymbol{z} = r\boldsymbol{z} - s\boldsymbol{u}$$

      and

      $$A\boldsymbol{u} = s\boldsymbol{z} + r\boldsymbol{u}$$

      so definition $\bar{\boldsymbol{v}} = \boldsymbol{z} - i\boldsymbol{u}$, we have that

      $$A\bar{\boldsymbol{v}} = r\boldsymbol{z} - s\boldsymbol{u} - i(s\boldsymbol{z} + r\boldsymbol{u}) = \bar{\lambda}\bar{\boldsymbol{v}}$$

    - Notice that in the above calculation we reached the conclusion that in the basis given by the real vectors $\boldsymbol{z}$ and $\boldsymbol{u}$, the matrix A can be expressed as

      $$\begin{bmatrix} r & s \\ -s & r \end{bmatrix}$$

which is the general form of a rotation-scaling matrix! So we have the following general statement: *Suppose A is a 2×2 matrix with real entries, and suppose A has no real eigenvalues. Then A has two complex eigenvalues that are complex conjugates. Furthermore, picking one of the eigenvalues, say λ, and writing it as $\lambda = r + si$ with r and s real numbers, and writing the eigenvector $\boldsymbol{v}$ for λ as $\boldsymbol{v} = \boldsymbol{z} + i\boldsymbol{u}$, where $\boldsymbol{z}$ and $\boldsymbol{u}$ are vectors with real entries, then $\boldsymbol{z}$ and $\boldsymbol{u}$ form a basis of $\mathbb{R}^2$ and*

$$A = SRS^{-1}$$

where $S = \begin{bmatrix} \boldsymbol{z} & \boldsymbol{u} \end{bmatrix}$ and

$$R = \begin{bmatrix} r & s \\ -s & r \end{bmatrix}$$

In other words, A is similar to a rotation-scaling matrix.

$*$ (Remark: my matrix looks different from the one given in Bretscher, this is because I used the opposite convention for the coördinate basis; I put the real part first in the matrix S and the imaginary part second. Geometrically, this means that the *frame* rotates backward relative to the *vectors*.)

- Furthermore, given a real vector $\boldsymbol{w}$, its decomposition in the eigenbasis $\boldsymbol{v}, \bar{\boldsymbol{v}}$ we have conjugate coördinate values:

$$\boldsymbol{w} = c\boldsymbol{v} + \bar{c}\bar{\boldsymbol{v}}$$

  where c is a complex number and $\bar{c}$ is its conjugate.

- Together this ensures that if we write the eigenbasis decomposition of the dynamical system corresponding to the matrix

$$\boldsymbol{x}(t) = A^t\boldsymbol{x}(0) = c\lambda^t\boldsymbol{v} + \bar{c}\bar{\lambda}^t\bar{\boldsymbol{v}}$$

  is a sum of a complex vector with its complex conjugate, and therefore will produce a real vector. (So we don't need to worry about the complex diagonalization giving us unphysical answers to the dynamical system.)

HOMEWORK FOR THIS WEEK

7.3: 14, 20, 24, 28, 40

7.4: 16, 22, 28, 34

7.5: 1, 6, 24, 32