Univ. of Washington

Lecture : Eigenvalues and eigenvectors

Winter 2021

Lecturer: Dan Calderone

Traces and Determinants

Two useful numbers associated with square matrices are the *trace* and the *determinant*. The trace is the sum of the diagonals

$$\text{Tr}(A) = \sum_i A_{ii} \tag{1}$$

Traces are very well behaved algebraic. One can check immediately the following identities.

$$\text{Tr}(A) = \text{Tr}(A^T), \qquad \text{Tr}(A+B) = \text{Tr}(A) + \text{Tr}(B), \qquad \text{Tr}(AB) = \text{Tr}(BA) \tag{2}$$

Formulas for the determinant are generally complicated but they compute how the volume of the unit cube changes under the transformation A.

$$\text{det}(A) = \text{signed volume of the unit cube transformed by } A \tag{3}$$

The sign of the determinant flips if the unit cube is reflected across some axis.

Determinants have the properties

$$\text{det}(A) = \text{det}(A^T), \qquad \text{det}(A^{-1}) = \text{det}(A)^{-1}, \qquad \text{det}(AB) = \text{det}(BA) = \text{det}(A)\text{det}(B) \tag{4}$$

Both the trace and determinant have special relationships with the eigenvalues of A (see below for discussion of eigenvalues). If the eigenvalues of A, $\lambda_1, \ldots, \lambda_n$ then we have that

$$\text{Tr}(A) = \sum_i \lambda_i, \qquad \text{det}(A) = \prod_i \lambda_i \tag{5}$$

Eigenvectors, Eigenvalues, and Diagonalization

In general, multiplying a column vector $x \in \mathbb{R}^n$ by a square matrix $A \in \mathbb{R}^{n\times n}$ causes that vector to stretch and to rotate. However, some vectors in specific subspaces are *only stretched, not rotated*. Another way to say this is that those subspaces are *invariant* with respect to A. These invariant subspaces are called *right eigenspaces* and vectors within them are called *right eigenvectors*. The amount each eigenvector is stretched is called it's *eigenvalue*. We can also consider a similar situation where left multiplying A by specific row vectors only causes them to stretch. These row vectors are called *left eigenvectors* and they live in *left eigenspaces*. (The eigenvalues for left and right eigenvectors turn out to be the same, ie. left and right eigenspaces come in pairs.) Finding a linearly independent sets of eigenvectors (either left or right) for a square matrix A is one of the fundamental problems of linear algebra. **If we represent vectors as coordinates with respect to a basis of eigenvectors, then the action of the matrix simply becomes scaling each individual coordinate by the appropriate eigenvalue.** If a matrix has a linearly independent basis of eigenvectors then we say it is *diagonalizable*. Not all matrices are diagonalizable, but if we choose a matrix at random then it will be (with probability 1), ie. we have to specifically work to construct a matrix that is not diagonalizable. The reason for this is that non-diagonalizable matrices are a low dimensional subset of the space of all matrices. Many arguments in linear algebra are best understood by understanding them for diagonalizable matrices and then generalizing them to the non-diagonalizable case.

The right and left eigenvector equations are given by

$$\lambda v = Av, \qquad\qquad \lambda w^T = w^T A \tag{6}$$

respectively. Suppose the columns of $P \in \mathbb{R}^{n\times n}$ are a linearly independent set of right eigenvectors of A and with eigenvalues $\lambda_1, \ldots, \lambda_n$. Let $D \in \mathbb{R}^n$ be a diagonal matrix with the eigenvalues on the diagonal, ie. $D = \mathrm{diag}([\lambda_1, \ldots, \lambda_n])$. The columns of P being right eigenvectors is equivalent to the equation

$$AP = PD \tag{7}$$

$$\Rightarrow \qquad A = PDP^{-1} \tag{8}$$

We say that the matrix of eigenvectors P diagonalizes A because it relates A to a diagonal matrix D via a similarity transform. In other words if $x = Pz$, $z' = Px'$ and $x' = Ax$, then $z' = Dz$. Note that in the z-coordinates, D simply scales each coordinate by the appropriate eigenvalue.

Left multiplying (8) by P^{-1} gives $P^{-1}A = DP^{-1}$. Note that this means that the rows of P^{-1} are a set of linearly independent left-eigenvectors of A. Note that this also shows why the left and right eigenvectors come in pairs and share eigenvalues. To summarize, let

$$P = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}, \qquad D = \begin{bmatrix} \lambda_1 & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & \lambda_n \end{bmatrix}, \qquad P^{-1} = \begin{bmatrix} - & w_1^* & - \\ & \vdots & \\ - & w_n^* & - \end{bmatrix}, \tag{9}$$

with v_i and w_j being right and left eigenvectors. A can be decomposed as

$$A = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} - & w_1^* & - \\ & \vdots & \\ - & w_n^* & - \end{bmatrix} = \sum_i \lambda_i v_i w_i^* \tag{10}$$

Note that real eigenvalues denote how much each eigenvectors get stretched when they are multiplied by the matrix.

Computing Eigenvalues and Eigenvectors

As stated above the determinant of a matrix is equal to the product of its eigenvalues. This means that if a matrix has a zero eigenvalue than its determinant is zero. Any vector in the nullspace of a matrix is an eigenvector with an eigenvalue of 0. Note that if $\lambda v = Av$ then $(\lambda I - A)v = 0$. In other words, if v is eigenvector of A with eigenvalue λ, then v is also an eigenvector of $\lambda I - A$ with eigenvalue 0. We can find eigenvalues of A by finding values of λ such that $(\lambda I - A)$ has a 0 eigenvalue. This leads us to characterize eigenvalues as solutions to the equation

$$\chi_A(s) = \det\big(sI - A\big) = 0 \tag{11}$$

$\chi_A(s)$ is called the *characteristic polynomial* of A.

$$\chi_A(s) = \det(sI - A) = s^n + \alpha_{n-1}s^{n-1} + \cdots + \alpha_1 s + \alpha_0$$

Based on properties of determinants, $\chi_A(s)$ will always have order n and the first term will always be s^n.

Once we find roots of $\chi_A(s)$, λ_i, we find the corresponding right and left eigenvectors by finding vectors in the right and left nullspace of $\lambda_i I - A$ respectively.

Spectral Mapping Theorem

Polynomial Functions

As stated above computing eigenvectors and eigenvalues simplifies matrix computations. In particular, note that given a diagonalization of $A = PDP^{-1}$, we can compute powers of A as

$$A^k = \underbrace{A \times \cdots \times A}_{\times k} = PD^k \underbrace{P^{-1} \times P}_{I} D^k P^{-1} \times \cdots \times PDP^{-1} = PD^kP^{-1} \tag{12}$$

This implies that if a function $f : \mathbb{C}^{n\times n} \to \mathbb{C}^{n\times n}$ is a polynomial (or more generally analytic function) of A, then

$$f(A) = Pf(D)P^{-1} = P \begin{bmatrix} f(\lambda_1) & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & f(\lambda_n) \end{bmatrix} P^{-1} \tag{13}$$

In other words, we can compute polynomial functions of A simply by applying that function to the eigenvalues of A and leaving the eigenvectors unchanged. This is known as the *spectral mapping theorem*. Note that this analysis applies to polynomials with an infinite number of terms such as Taylor expansions of functions such as $e^{(\cdot)}$, $\cos(\cdot)$, and $\sin(\cdot)$ as well.

Matrix Exponential

One important function of A that we want to compute is the *matrix exponential* e^A where which can be defined by its Taylor expansion.

$$e^A := I + A + \tfrac{1}{2!}A^2 + \tfrac{1}{3!}A^3 + \cdots = \sum_{k=0}^{\infty} \tfrac{1}{k!}A^k \tag{14}$$

Note that by the spectral mapping theorem we have that

$$e^A = Pe^D P^{-1} = P \begin{bmatrix} e^{\lambda_1} & & 0 \\ \vdots & \ddots & \vdots \\ 0 & & e^{\lambda_n} \end{bmatrix} P^{-1} \tag{15}$$

Exponential functions are interesting because they are functions who are equal to their own derivative (times some scaling), ie. $\frac{d}{dt}e^{\lambda t} = \lambda e^{\lambda t}$. (Note that $e^{\lambda t}$ is actually an *eigenfunction* of the derivative operator $\frac{d}{dt}$ with eigenvalue λ.)

Cayley-Hamilton Theorem

The Cayley-Hamilton theorem says that a matrix satisfies its own characteristic polynomial, ie. ${}_A(A) = 0$. For diagonalizable matrices, this is a direct application of the spectral mapping theorem.

$$\chi_A(A) = P\chi_A(D)P^{-1} = P \begin{bmatrix} \chi_A(\lambda_1) & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & \chi_A(\lambda_n) \end{bmatrix} P^{-1} = 0$$

Consequently,

$$A^n = -\alpha_{n-1}A^{n-1} - \cdots - \alpha_1 A - \alpha_0 I$$

As a result of this, any polynomial function of A could be expressed in terms of powers of A only up through $n-1$. Higher powers of A can be reduced by iteratively plugging in the above equation.

Another application of Cayley-Hamilton gives a polynomial expression for a matrix inverse.

$$\begin{aligned} 0 &= \Big(A^n + \alpha_{n-1}A^{n-1} + \cdots + \alpha_1 A + \alpha_0 I\Big)A^{-1} \\ A^{-1} &= -\tfrac{1}{\alpha_0}A^{n-1} - \tfrac{\alpha_{n-1}}{\alpha_0}A^{n-2} - \cdots - \tfrac{\alpha_1}{\alpha_0}I \end{aligned}$$

Jordan Form

To motivate a study of Jordan form, we consider the following matrix

$$J_i = \lambda_i I + N_i = \begin{bmatrix} \lambda_i & 1 & \cdots & \cdots & 0 \\ 0 & \lambda_i & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \lambda_i & 1 \\ 0 & \cdots & \cdots & 0 & \lambda_i \end{bmatrix}$$

where N_i is a matrix with 1's on the first super diagonal. This matrix N_i is an example of a *nilpotent matrix* since raising it to some power gives a matrix of 0's, ie. for example

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

Note that any matrix similar to a nilpotent matrix is also nilpotent. If $N_i^k = 0$, then $(PN_iP^{-1})^k = PN_i^kP^{-1} = 0$. If $J_i = \lambda_i I + N_i$, then clearly, $J_i - \lambda_i I$ is nilpotent, ie. $J_i - \lambda_i I = N_i$. Since the eigenvalues of a triangular matrix are just the diagonal values, we have that the only eigenvalue of N_i is simply 0. However, N_i clearly has $n-1$ linearly independent columns, ie. rank $n-1$. Thus it only has a one dimensional nullspace. One can check that the characteristic polynomial of N_i is $\chi_{N_i}(s) = s^n$ and the charactestic polynomial of $J_i = \lambda_i I + N_i$ is $\chi_{J_i}(s) = (s - \lambda_i)^n$.

A matrix is not diagonalizable when a full basis of eigenvectors does not exist. For a matrix $A \in \mathbb{R}^{n \times n}$ with n distinct eigenvalues, there must be a basis of n linearly independent eigenvectors since each eigenvalue λ_i is associated with the nullspace of $\lambda_i I - A$. We know these eigenvectors are linearly independent since if not

$$\begin{aligned} v_i &= \sum_{j \neq i} \alpha_j v_j \\ Av_i &= A \left(\sum_{j \neq i} \alpha_j v_j \right) \\ 0 &= \sum_{j \neq i} \alpha_j \lambda_j v_j - \lambda_i v_i \\ 0 &= \sum_{j \neq i} \alpha_j (\lambda_j - \lambda_i) v_j \end{aligned}$$

An inductive argument shows that $\lambda_i = \lambda_j$ for some i and j which is a contradiction.

In this case, the characteristic polynomial is

$$\chi_A(s) = (s - \lambda_1)(s - \lambda_2) \cdots (s - \lambda_n)$$

In the general case with repeated eigenvaleus, the characteristic polynomial is given by

$$\chi_A(s) = \prod_{i=1}^{k} (s - \lambda_i)^{k_i}$$

where k is the number of distinct eigenvalues and k_i is the number of times each eigenvalue is repeated. If $\dim\big(\mathcal{N}(\lambda_i I - A)\big) = k_i$ for all i, then the matrix is diagonalizable. In this case,

$$\mathcal{N}(\lambda_i I - A) = \mathcal{N}\big((\lambda_i I - A)^2\big) = \mathcal{N}\big((\lambda_i I - A)^3\big) = \dots$$

and

$$\dim\big(\mathcal{N}(\lambda_i I - A)\big) = \dim\big(\mathcal{N}\big((\lambda_i I - A)^2\big)\big) = \dim\big(\mathcal{N}\big((\lambda_i I - A)^3\big)\big) = \cdots = k_i$$

This happens when $\mathcal{N}(\lambda_i I - A) \cap \mathcal{R}(\lambda_i I - A) = 0$ for all i.

It is also possible that $\dim\big(\mathcal{N}(\lambda_i I - A)\big) < k_i$ In this case

$$\mathcal{N}(\lambda_i I - A) \subset \mathcal{N}\big((\lambda_i I - A)^2\big) \subset \mathcal{N}\big((\lambda_i I - A)^3\big) \subset \dots$$

and

$$\dim\big(\mathcal{N}(\lambda_i I - A)\big) < \dim\big(\mathcal{N}\big((\lambda_i I - A)^2\big)\big) < \dim\big(\mathcal{N}\big((\lambda_i I - A)^3\big)\big) < \cdots < k_i \tag{16}$$

ie., $\mathcal{N}(\lambda_i I - A) \cap \mathcal{R}(\lambda_i I - A) \neq 0$. A regular eigenvector satisfies

$$(\lambda_i I - A) v_i = 0$$

If $\dim\big(\mathcal{N}(\lambda_i I - A)\big) < \dim\big(\mathcal{N}(\lambda_i I - A)^2\big)$, then we should be able to find generalized eigenvectors that satisfy

$$(\lambda_i I - A) w_i^2 \in \mathcal{N}(\lambda_i I - A), \qquad (\lambda_i I - A) w_i^3 \in \mathcal{N}(\lambda_i I - A)^2, \qquad \text{etc}$$

$w_i^2 \in \mathbb{C}^n$ is a 2nd order eigenvector, $w_i^3 \in \mathbb{C}^n$ is a 3rd order eigenvector, etc.

Note that

$$(\lambda_i I - A)^2 w_i^2 = 0, \qquad (\lambda_i I - A)^3 w_i^3 = 0, \qquad \text{etc}$$

If we are careful in picking, v_i, w_i^2, w_i^3, $\dots$ we can choose them so that

$$0 = (\lambda_i I - A) v_i, \qquad v_i = (\lambda_i I - A) w_i^2, \qquad w_i^2 = (\lambda_i I - A) w_i^3, \qquad \text{etc} \tag{17}$$

A general organization of these equations is given by

$$AP = PJ = \underbrace{\begin{bmatrix} V_1 & \cdots & V_q \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} J_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & J_q \end{bmatrix}}_{J}$$

where

$$V_i = \begin{bmatrix} | & | & | & \\ v_1 & w_1^2 & w_1^3 & \cdots \\ | & | & | & \end{bmatrix}, \qquad J_i = \lambda_i I + N_i = \begin{bmatrix} \lambda_i & 1 & \cdots & \cdots & 0 \\ 0 & \lambda_i & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \lambda_i & 1 \\ 0 & \cdots & \cdots & 0 & \lambda_i \end{bmatrix}$$

$$N_i = \begin{bmatrix} 0 & 1 & \cdots & \cdots & 0 \\ 0 & 0 & & & \vdots \\ \vdots & & \ddots & & \vdots \\ \vdots & & & 0 & 1 \\ 0 & \cdots & \cdots & 0 & 0 \end{bmatrix}$$

J_i is called a Jordan block and q is the number of Jordan blocks. Each Jordan block corresponds to one true eigenvector and a chain of generalized eigenvectors as in (17). Note that if each distinct eigenvalue has only one Jordan block (and only one true eigenvector), then $q = k$, the number of distinct eigenvalues. It is possible that a distinct eigenvalue has more than one Jordan block. In this case, $q > k$. Most matrices are diagonalizable, but every matrix can be put in *Jordan form*. Note that

$$\begin{aligned} A - \lambda_1 I &= PJP^{-1} - \lambda_1 PP^{-1} \\ &= P(J - \lambda_1 I)P^{-1} \\ &= P \begin{bmatrix} N_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & J_q \end{bmatrix} \end{aligned}$$

and that

$$(A - \lambda_1 I)^\ell = P \begin{bmatrix} N_1^\ell & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & J_q^\ell \end{bmatrix}$$

Since N_1 is nilpotent, as ℓ increases the nullspace of $(A - \lambda_1 I)^\ell$ grows as in (16).

We now perform several manipulations with a simple non-diagonalizable matrix to illustrate some simple properties of Jordan form. Consider

$$
\begin{aligned}
A &= \begin{bmatrix} | & | & | \\ v & w^2 & w^3 \\ | & | & | \end{bmatrix} \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} | & | & | \\ v & w^2 & w^3 \\ | & | & | \end{bmatrix}^{-1} \\
&= \begin{bmatrix} | & | & | \\ v & w^2 & w^3 \\ | & | & | \end{bmatrix} \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} - & (q^3)^T & - \\ - & (q^2)^T & - \\ - & p^T & - \end{bmatrix} \\
&= \lambda v(q^3)^T + (v + \lambda w^2)(q^2)^T + (w^2 + \lambda w^3)p^T \\
&= \lambda v(q^3)^T + \lambda w^2(q^2)^T + \lambda w^3 p^T + v(q^2)^T + w^2 p^T
\end{aligned}
$$

Note that

- The first order right eigenvector v matches up with the third order left generalized eigenvector $(q^3)^T$
- The second order right eigenvector w^2 matches up with the second order left generalized eigenvector $(q^2)^T$
- The third order right eigenvector w^3 matches up with the first order left eigenvector p^T

We note that we could also write A in other ways related to Jordan form (These are just a sample of how the Jordan block and eigenvectors could be shuffled.)

$$
\begin{aligned}
A &= \begin{bmatrix} | & | & | \\ w^2 & v & w^3 \\ | & | & | \end{bmatrix} \begin{bmatrix} \lambda & 0 & 1 \\ 1 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} - & (q^2)^T & - \\ - & (q^3)^T & - \\ - & p^T & - \end{bmatrix} \\
&= \begin{bmatrix} | & | & | \\ w^3 & w^2 & v \\ | & | & | \end{bmatrix} \begin{bmatrix} \lambda & 0 & 0 \\ 1 & \lambda & 0 \\ 0 & 1 & \lambda \end{bmatrix} \begin{bmatrix} - & p^T & - \\ - & (q^2)^T & - \\ - & (q^3)^T & - \end{bmatrix} \\
&= \text{etc...}
\end{aligned}
$$