MATH 320 NOTES, WEEK 14

Diagonalization

Recall that if V is finite dimensional and $T : V \to V$ is a linear transformation, then the matrix representation of T, $[T]_\alpha$ for some basis α captures all the properties of T. For example, T is invertible iff $[T]_\alpha$ is invertible, and so on.

And when it comes to matrices, some of the nicest ones to deal with are diagonal matrices – the ones whose nonzero entries appear only on the diagonal.

This chapter is motivated by investigating when we can view a linear transformation T as a diagonal matrix,

5.1 Eigenvalues and eigenvectors

Definition 1. *Let V be finite dimensional and $T : V \to V$ be a linear transformation. T is* **diagonalizable** *if there is a basis α for V, such that $[T]_\alpha$ is a diagonal matrix.*

Similarly, a matrix $A \in M_{n\times n}(F)$ is diagonalizable if there is an invertible matrix Q, such that QAQ^{-1} is diagonal.

Note that any diagonal matrix is (trivially) diagonalizable, for example, the zero matrix, I_n, aI_n for any scalar a.

Exercise: A matrix A is diagonalizable iff there is a basis α, such that $[L_A]_\alpha$ is diagonal.

Suppose now that $\dim(V) = n$ and $T : V \to V$ is diagonalizable, i.e. for some basis $\alpha = \{v_1, ..., v_n\}$, $D := [T]_\alpha$ is diagonal. Say,

$$D := \begin{pmatrix} \lambda_1 & 0... & 0 \\ 0 & \lambda_2... & 0 \\ \vdots & & \\ 0 & ... & \lambda_n \end{pmatrix}$$

Let's look at the first column of D: it is $[T(v_1)]_\alpha$, so we must have that $T(v_1) = \lambda_1 v_1$. Similarly, it must be that $T(v_2) = \lambda_2 v_2$, and so on. More generally, for all $i \leq n$, $T(v_i) = \lambda_i v_i$. This motivates the next definition.

Definition 2. *Let $T : V \to V$ be a linear transformation, V finite dimensional. An* **eigenvector** *for T is a vector $v \neq \vec{0}$, such that for some scalar λ,*

$$T(v) = \lambda v.$$

Then we say that λ is an **eigenvalue**.

Similarly, if $A \in M_{n\times n}(F)$*, an eigenvector for* A *is* $v \in F^n$*,* $v \neq \vec{0}$*, such that* $Av = \lambda v$*for some scalar* $\lambda \in F$*; such a* λ *is an eigenvalue for* A*.*

Theorem 3. *Let* V *be finite dimensional and* $T : V \to V$ *be a linear transformation.* T *is diagonalizable iff* V *has a basis of eigenvectors for* T*.*

Proof. For one direction, suppose that T is diagonalizable. Let $\alpha = \{v_1, ..., v_n\}$ be a basis for V, such that $D = [T]_\alpha$ is diagonal. But then by the above analysis, for each $i \leq n$, $T(v_i) = \lambda v_i$, where $\lambda = D_{ii}$, the (i,i)-th entry of D. So, each vector in α is an eigenvector.

For the other direction, suppose that $\alpha = \{v_1, ..., v_n\}$ is a basis of eigenvectors for T. For each $i \leq n$, let λ_i be a scalar such that $T(v_i) = \lambda_i v_i$. Then,

$$[T]_\alpha = \begin{pmatrix} \lambda_1 & 0... & 0 \\ 0 & \lambda_2... & 0 \\ \vdots & & \\ 0 & ... & \lambda_n \end{pmatrix}.$$

So, T is diagonalizable. $\square$

Next we want to see how, given a linear transformation T, we can find eigenvalues and eigenvectors for T. First, note the following lemma.

Lemma 4. λ *is an eigenvalue for* $A \in M_{n\times n}$ *iff* $\det(A - \lambda I_n) = 0$*. Also, if* v *is an eigenvector corresponding to the eigenvalue* λ*, then* $v \in ker(A - \lambda I_n)$*.*

Proof. λ is an eigenvalue for A iff for some $v \neq \vec{0}$, $Av = \lambda v$ iff $Av - \lambda v = \vec{0}$ iff $(A - \lambda I_n)v = \vec{0}$ for some nonzero v iff $(A - \lambda I_n)$ is not one-to-one iff $(A - \lambda I_n)$ is not invertible iff $\det(A - \lambda I_n) = 0$.

For the second statement, suppose that v is an eigenvector with eigenvalue λ. Then $Av = \lambda v$, so $(A - \lambda I_n)v = Av - \lambda v = \vec{0}$, so $v \in ker(A - \lambda I_n)$. $\square$

Definition 5. *Let* $A \in M_{n\times n}(F)$*. The* **characteristic polynomial of** A *is*

$$f_A(t) = \det(A - tI_n).$$

What we have so far is that λ is an eigenvalue of A iff λ is a root of the characteristic polynomial $f_A(t)$.

Next we want to generalize the above lemma and definition of the characteristic polynomial to a linear transformation $T : V \to V$, V finite dimensional. To define the characteristic polynomial of T, we have to take a matrix representation, with respect to some basis. So we should make sure the choice of basis does not matter i.e. any basis will yield the same characteristic polynomial. To do this we make use of the following fact about similar matrices.

Lemma 6. *Let* $A, B \in M_{n\times n}(F)$ *be similar matrices. Then they have the same characteristic polynomial. I.e.* $\det(A - tI_n) = \det(B - tI_n)$*.*

Proof. Let Q be an invertible matrix, such that $A = Q^{-1}BQ$. Then,

$$\det(A - tI_n) = \det(Q^{-1}BQ - tI_n) = \det(Q^{-1}(B - tI_n)Q) =$$
$$\det(Q^{-1})\det(B - tI_n)\det(Q) = \det(Q^{-1}Q)\det(B - tI_n) = \det(B - tI_n).$$

□

Corollary 7. *Let α, β be two bases for a finite dimensional V and $T : V \to V$ be a linear transformation. Then $\det([T]_\alpha - tI_n) = \det([T]_\beta - tI_n)$.*

Proof. Since $[T]_\alpha$ and $[T]_\beta$ are similar matrices, the result follows by the above lemma.

□

Definition 8. *Let $T : V \to V$ be a linear transformation, V finite dimensional. The characteristic polynomial of T is*

$$f_T(t) = \det([T]_\alpha - tI_n),$$

where α is any basis for V.

By the above lemma the characteristic polynomial of T is well defined i.e. it is independent of the choice of the basis α.

Theorem 9. *Let $T : V \to V$ be a linear transformation, V finite dimensional. Then,*

1. *λ is an eigenvalue iff λ is a root of the characteristic polynomial f_T i.e. $f_T(\lambda) = 0$;*
2. *v is an eigenvector corresponding to λ iff $v \neq \vec{0}$ and $v \in \ker(T - \lambda I)$. Here $I : V \to V$ is the identity linear transformation $I(x) = x$, and $\lambda I(x) = \lambda x$.*

Proof. Let's show the second item first: v is an eigenvector corresponding to λ iff $v \neq \vec{0}$ and $T(v) = \lambda v$ iff $v \neq \vec{0}$ and $(T - \lambda I)(v) = T(v) - \lambda v = \vec{0}$ iff $v \neq \vec{0}$ and $v \in \ker(T - \lambda I)$.

For the first item, λ is an eigenvalue iff there is $v \neq \vec{0}$ and $T(v) = \lambda v$ iff there is there is $v \neq \vec{0}$, $v \in \ker(T - \lambda I)$ iff $\ker(T - \lambda I) \neq \{\vec{0}\}$ iff $\ker([T]_\alpha - \lambda I) \neq \{\vec{0}\}$ for some basis α iff $f_T(\lambda) = \det([T]_\alpha - \lambda I) = 0$. □

Steps to find eigenvalues and eigenvectors of T:

1. Solve $f(t) = 0$, where f is the characteristic polynomial of T. Say the roots are $\{\lambda_1, ..., \lambda_k\}$.
2. For each $i \leq k$, solve for v in $T(v) = \lambda_i v$. The (nonzero) solutions are the eigenvectors corresponding to λ_i.
3. Check if we have enough linearly independent eigenvectors to form a basis for V. If yes, then T is diagonalizable.

Below we give some examples.

Example 1. Find the eigenvalues and eigenvectors of $A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$

Compute the characteristic polynomial of A, $f_A(t) = \det(A - tI_3) = \det\begin{pmatrix} 1-t & 2 & 0 \\ 0 & 1-t & 0 \\ 1 & 0 & -t \end{pmatrix} = (1-t)(1-t)(-t) = -t(t-1)^2 = 0$. The solutions are $\lambda_1 = 0, \lambda_2 = 1$.

For $\lambda_1 = 0$, solving $Av = A\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \vec{0}$, we get

- $v_1 + 2v_2 = 0$,
- $v_2 = 0$,
- $v_1 = 0$.

So the solutions set is $Span(\langle 0,0,1\rangle)$. Then we have the eigenvector $\langle 0,0,1\rangle$ (every other solution is a multiple of it).

For $\lambda_2 = 1$, solving $Av = v$, $v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$, we get

- $v_1 + 2v_2 = v_1$,
- $v_2 = v_2$,
- $v_1 = v_3$.

So the solutions set is $Span(\langle 1,0,1\rangle)$. Then we have the eigenvector $\langle 1,0,1\rangle$ (every other solution is a multiple of it).

To sum up we have two eigenvalues: $\{0,1\}$ and corresponding eigenvectors $\langle 0,0,1\rangle, \langle 1,0,1\rangle$. It follows that A is not diagonalizable because there are only two linearly independent eigenvectors, so there is no basis of eigenvectors.

Example 2. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be

$$T(\langle v_1, v_2, v_3\rangle) = \langle v_1 + v_3, -v_2 + v_3, 2v_3\rangle$$

Taking $\alpha = \{e_1, e_2, e_3\}$ to be the standard basis, we get that

$$[T]_\alpha = \begin{pmatrix} 1 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

The characteristic polynomial of T is $f(t) = \det([T]_\alpha - tI_3) = \det\begin{pmatrix} 1-t & 0 & 1 \\ 0 & -1-t & 1 \\ 0 & 0 & 2-t \end{pmatrix} =$ $(1-t)(-1-t)(2-t) = -(t+1)(t-1)(t-2)$.
The solutions are $1, -1, 2$.

For $\lambda = 1$, solve for $T(v) = v$ and get $v_1 + v_3 = v_1, -v_2 + v_3 = v_2, 2v_3 = v_3$, so $v_3 = 0 = v_2$. Eigenvector: $\langle 1,0,0\rangle$.

For $\lambda = -1$, solve for $T(v) = -v$ and get $v_1 + v_3 = -v_1, -v_2 + v_3 = -v_2, 2v_3 = -v_3$, so $v_3 = 0 = v_1$. Eigenvector: $\langle 0,1,0\rangle$.

For $\lambda = 2$, solve for $T(v) = v$ and get $v_1 + v_3 = 2v_1, -v_2 + v_3 = 2v_2, 2v_3 = 2v_3$, so $v_1 = v_3, v_3 = 3v_2$. Eigenvector: $\langle 3,1,3\rangle$.

In this case we have a basis of eigenvectors: $\beta = \{\langle 1,0,0\rangle, \langle 0,1,0\rangle, \langle 3,1,3\rangle\}$, and so T is diagonalizable. In particular, we have that

$$[T]_\beta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

5.2 Diagonalizability In this next section, we analyze what type of matrices/linear transformations are diagonalizable.

Theorem 10. *Let $T : V \to V$ be a linear transformation. Suppose $\{\lambda_1, ..., \lambda_k\}$ are distinct eigenvalues, with corresponding eigenvectors $\{v_1, ..., v_k\}$. Then $\{v_1, ..., v_k\}$ are linearly independent.*

Proof. By induction on k. If $k = 1$, this is since eigenvectors are nonzero.

Suppose that $k > 1$, and $\{v_1, ..., v_{k-1}\}$ are linearly independent. Suppose that $a_1v_1 + ... + a_kv_k = \vec{0}$. Since for each i, $T(v_i) = \lambda_i v_i$, we have that

$$T(a_1v_1 + ... + a_kv_k) = a_1\lambda_1v_1 + ... + a_k\lambda_kv_k = \vec{0}.$$

The last equality is because $T(\vec{0}) = \vec{0}$. We also have that

$$\lambda_k(a_1v_1 + ... + a_kv_k) = \lambda_ka_1v_1 + ... + \lambda_ka_kv_k = \vec{0}.$$

By solving for $a_k\lambda_kv_k$ in both equations, we get

$$\lambda_ka_1v_1 + ... + \lambda_ka_{k-1}v_{k-1} = \lambda_1a_1v_1 + ... + \lambda_{k-1}a_{k-1}v_{k-1}.$$

It follows that,

$$(\lambda_k - \lambda_1)a_1v_1 + (\lambda_k - \lambda_2)a_2v_2 + ... + (\lambda_k - \lambda_{k-1})a_{k-1}v_{k-1} = \vec{0}.$$

Since the eigenvalues are distinct, for all $i < k$, $(\lambda_k - \lambda_i) \neq 0$. Then by the inductive hypothesis , that $\{v_1, ..., v_{k-1}\}$ are linearly independent, we get that for all $i < k$, $a_i = 0$. But then also $a_k = 0$, since $v_k \neq \vec{0}$.

□

As an immediate corollary we have:

Corollary 11. *Let $T : V \to V$ be a linear transformation, with $\dim(V) = n$. Suppose that T has n many distinct eigenvalues. Then there are n many linearly independent eigenvectors and so T is diagonalizable.*

What about the case when the number of eigenvalues is less than $\dim(V)$? As we have seen earlier in some of these cases, the matrix is diagonalizable, such as the example above with $A = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.

On the other hand, $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ has two eigenvalues $1, 2$ and it is diagonal.

So next, we analyze, given an eigenvalue, how many linearly independent eigenvectors correspond to it. Let λ be an eigenvalue for $T : V \to V$ (or a matrix A). Define

$$E_\lambda := \{v \mid T(v) = \lambda v\}.$$

E_λ is called the **eigenspace** corresponding to λ.

Lemma 12. *E_λ is a subspace.*

Proof. Exercise. □

Now, note that the number of linearly independent eigenvectors for λ is exactly $\dim(E_\lambda)$. So, in order of have enough eigenvectors for a basis for V, we would need that for each eigenvalue λ, $\dim(E_\lambda)$ matches the power with which λ appears as a root in the characteristic polynomial. That brings us to the following definition.

Definition 13. *A polynomial $f(t)$ with coefficients in F* **splits** *over F if we can factor $f(t) = c(t - a_1)(t - a_2)...(t - a_n)$, where c and each $a_i \in F$. Here that a_i's do not need to be distinct.*

For example,

- $(x - 1)^3$ splits over $\mathbb{Q}$ (and so over $\mathbb{R}$).
- $x^3 - x = x(x - 1)(x + 1)$ splits over $\mathbb{Q}$.
- $x^2 + 1$ does not split over $\mathbb{R}$, but it does split over $\mathbb{C}$.
- $x^2 - 2$ does not split over $\mathbb{Q}$, but it does split over $\mathbb{R}$.

Definition 14. *Let λ be an eigenvalue for $T : V \to V$ (or a matrix A), with a characteristic polynomial $f(t)$.*

(1) *The* **algebraic multiplicity** *of λ is the largest power k, such that $f(t) = (t - \lambda)^k g(t)$.*
(2) *The* **geometric multiplicity** *of λ is* $\dim(E_\lambda)$.

Remark 15. The book uses the word "multiplicity" to denote "algebraic multiplicity".

Lemma 16. *Let V, T be as above, and let λ be an eigenvalue with algebraic multiplicity a. Then $1 \leq \dim(E_\lambda) \leq a$. I.e. the geometric multiplicity is no more that the algebraic multiplicity.*

Proof. Let $\{v_i, ..., v_k\}$ be a basis for E_λ. Extend it to a basis α for V, and let $A = [T]_\alpha$. Then

$$A = \begin{pmatrix} \lambda I_k & B \\ O & C \end{pmatrix},$$

and so the characteristic polynomial is $f(t) = (t - \lambda)^k g(t)$, which means that k is no more than the algebraic multiplicity of λ, □

And now for the main theorem of the section:

Theorem 17. *Suppose that V is a finite dimensional vector space over F, $\dim(V) = n$, and $T : V \to V$ is linear. Then T is diagonalizable iff both of the following hold:*

(1) *The characteristic polynomial $f(t)$ splits over F.*
(2) *For every eigenvalue λ, the algebraic multiplicity of λ equals the geometric multiplicity of λ (i.e. $\dim(E_\lambda)$).*

Proof. First recall that the characteristic polynomial has degree n.

For the first direction, suppose that T is diagonalizable. Let $\alpha = \{v_1, ..., v_n\}$ be a basis of eigenvectors. Then $D := [T]_\alpha$ is diagonal, say

$$D = \begin{pmatrix} d_{11} & 0 & ... & 0 \\ 0 & d_{22} & ... & 0 \\ \vdots & & & \\ 0 & 0 & ... & d_{nn} \end{pmatrix}.$$

Then $f(t) = (t - d_{11})(t - d_{22})...(t - d_{nn})$, and so it splits.

Now let the eigenvalues for T be $\{\lambda_1, ..., \lambda_k\}$, $k \leq n$. Note that above, each d_{ii} is one of those eigenvalues. For each $i \leq k$, let a_i be the algebraic multiplicity of λ_i and let b_i be its geometric multiplicity i.e. $\dim(E_{\lambda_i}) = b_i$. Then $f(t) = c(t - \lambda_1)^{a_1}...(t - \lambda_k)^{a_k}$, and

$$b_1 + b_2 + ... + b_k = n = a_1 + a_2 + ... + a_k.$$

And since each $b_i \leq a_i$, we have to have that $a_i = b_i$.

For the other direction, suppose that items (1) and (2) hold. Again, let $\{\lambda_1, ..., \lambda_k\}$ be the eigenvalues for T, with algebraic multiplicity a_i and geometric multiplicity b_i for λ_i. Since the characteristic polynomial splits, we must have that $n = a_1 + a_2 + ... + a_k$. By item (2) for each i, $a_i = b_i$, so $b_1 + b_2 + ... + b_k = n$.

For each i, let $\alpha_i = \{v_{i1}, ..., v_{ib_i}\}$ be a basis for E_{λ_i}, and let $\alpha = \cup_i \alpha_i$. We will show that α is a basis for V. Since $|\alpha| = b_1 + b_2 + ... + b_k = n$, it is enough to show that α is linearly independent.

So, suppose that

$$\Sigma_{i \leq i \leq k, 1 \leq j \leq b_i} a_{ij} v_{ij} = \vec{0}.$$

For $i \leq i \leq k$, let $x_i = \Sigma_{1 \leq j \leq b_i} a_{ij} v_{ij}$. Then each $x_i \in E_{\lambda_i}$, and

$$x_1 + .. + x_k = \vec{0}.$$

Claim 18. *For all $i \leq k$, $x_i = \vec{0}$.*

Proof. Note that if $x_i \neq \vec{0}$, then it is an eigenvector for the eigenvalue λ_i. So if some of them are nonzero, we have a linear combination of eigenvectors for distinct eigenvalues equal to $\vec{0}$. We already showed that is impossible. □

Then for all $i \leq k$, $x_i = \Sigma_{1 \leq j \leq b_i} a_{ij} v_{ij} = \vec{0}$, and since α_i is linearly independent, the coefficients must be zero i.e. $a_{i1} = a_{i2} = ... = a_{ib_i} = 0$. So for all i, j, $a_{ij} = 0$. This concludes the proof that α is linearly independent.

Then α is a basis of eigenvectors, and so T is diagonalizable.

$\square$