640:350: LINEAR ALGEBRA HIGHLIGHTS OF LECTURES FOR WEEK 9

Sections to be covered: 5.1, 5.2, 5.3.

(1) First, let's recall what we learned in elementary linear algebra about eigenvalues, eigenvectors, and diagonalizability of an $n \times n$ matrix A.

  (a) λ is called an eigenvalue of A if there is a non-zero vector $\mathbf{v}$ such that $A\mathbf{v} = \lambda\mathbf{v}$. $\mathbf{v}$ will then be called an eigenvector of A associated with the eigenvalue λ.

  (b) The condition for λ to be an eigenvalue and $\mathbf{v}$ to be an eigenvector of A can be reformulated as

$$(A - \lambda I)\mathbf{v} = \mathbf{0}$$

  having a non-zero solution $\mathbf{v}$. This happens iff

$$\det(A - \lambda I) = 0.$$

  So we have an algorithm to find the eigenvalues and eigenvectors of A: first solve $\det(A - \lambda I) = 0$ to find all eigenvalues of A; then for each eigenvalue λ of A, solve $(A - \lambda I)\mathbf{v} = \mathbf{0}$ to find the associated eigenvectors.

  (c) What are potential complications in solving $\det(A - \lambda I) = 0$?

  - $\det(A - \lambda I)$ is a degree n polynomial in the unknown λ, and is called the characteristic polynomial of A.
  - When $n \geq 3$, it may not be easy to find the roots of the equation $\det(A - \lambda I) = 0$ by closed formula; more importantly, some of the roots may not be in the field of scalars that we used in constructing A. For instance, the entries of A may consist of rational numbers, but the roots may end up to be irrational numbers, or complex numbers; this may also happend when the entries of A consist of real numbers. However, for any $n \times n$ matrix A with entries in the complex numbers, all of its eigenvalues stay within the field of complex numbers. This property of the the field of complex numbers is called the fundamental theorem of algebra.
  - Any field with such a property, namely, the roots of polynomial equations with coefficients in the field $\mathbb{F}$ stay within $\mathbb{F}$, is called algebraically closed.
  - For a given polynomial $f(t)$ of degree n with coefficients in the field $\mathbb{F}$, if all of its roots are in F, then there must be n such roots, counting multiplicity. We can list them as $a_1, \cdots, a_n$, and can prove that $f(t)$ can then be factorized as

$$f(t) = c(t - a_1) \cdots (t - a_n),$$

    for some non-zero scalar $c \in \mathbb{F}$. Such a polynomial f is called split over $\mathbb{F}$.

    Example. $f(t) = t^2 + t - 2$ *is split over the field of rational numbers* $\mathbb{Q}$, *as* $f(t) = (t-1)(t+2)$. $g(t) = t^2 - 2$ *is not split over the field of rational numbers* $\mathbb{Q}$, *as there are roots of* $g(t) = 0$ *which are not rational numbers; however,* $g(t)$ *is split over the field of real numbers, as* $g(t) = (t-\sqrt{2})(t+\sqrt{2})$, *and both* $\pm\sqrt{2}$ *are real numbers.* $h(t) = t^2 + 2$ *is not split over the field of rational numbers, as there are roots of* $h(t) = 0$ *which are not rational numbers.*

So we have to resolve this issue when solving for the eigenvalues of an $n \times n$ matrix A. One approach is to work with whatever eigenvalue one can find within the starting field $\mathbb{F}$; another approach is to extend the field $\mathbb{F}$ to a larger field $\bar{\mathbb{F}}$, and work with possible eigenvalues in the extended field $\bar{\mathbb{F}}$, just as the example above shows.

(d) matrix A is called diagonalizable if there exists an $n \times n$ invertible matrix Q and a diagonal matrix Λ such that $A = Q\Lambda Q^{-1}$. When this happens, we can rewrite the relation $A = Q\Lambda Q^{-1}$ as $AQ = Q\Lambda$, which implies that each column of Q is an eigenvector of A with the corresponding eigenvalue as a diagonal entry in the diagonal matrix Λ. Based on this discussion, we have the following diagonalizability criterion: *An $n \times n$ matrix A is diagonalizable iff there exists a basis of $\mathbb{F}^n$ consisting of eigenvectors of A.*

(e) For each eigenvalue λ of a matrix A, the maximum number of linearly independent eigenvectors of A associated with the eigenvalue λ is the dimension of $Null(A - \lambda I)$. If $\lambda_1, \cdots, \lambda_r$ denote the set of all distinct eigenvalues of A, then one needs to find a maximum number of linearly independent eigenvectors of A associated with each λ_i, and pools these eigenvectors together. One needs to make sure that the union of these eigenvectors is still linearly independent—this is true, see **Theorem 5.8.** But this union may not form a basis of $\mathbb{F}^n$. The union will be a basis of $\mathbb{F}^n$ by **Corollary 2** on p. 47, provided the union contains n vectors. But the union contains $\sum_i \dim(Null(A - \lambda_i I))$ vectors, $\dim(Null(A - \lambda_i I)) \le m_i$—**Theorem 5.7**, where m_i denotes the algebraic multiplicity of λ_i, and $\sum_i m_i = n$, so the union will be a basis of $\mathbb{F}^n$ iff $\dim(Null(A - \lambda_i I)) = m_i$ for each i—**Theorem 5.9**.

(2) The concepts of eigenvalues and eigenvectors generalize to a linear operator $T : V \mapsto V$ directly. The concept of characteristic polynomial can also be defined for a linear operator $T : V \mapsto V$ as the characteristic polynomial of its matrix representation $[T]_\beta$ in any basis β. This definition is independent of the choice of the basis used, because for any two bases β and γ of V, $[T]_\beta = P[T]_\gamma P^{-1}$ for some invertible P, and $[T]_\beta$ has the same characteristic polynomial as $[T]_\gamma$. For the same reason, the algebraic multiplicity of an eigenvalue of the operator T can also be defined. Then we can use our discussion above on matrix diagonalizability to discuss the diagonalizability of a linear operator.

(3) A subspace W of V is called a T-**invariant subspace** of V if $T(W) \subset W$.

(4) An invariant subspace W of T is useful in that one can restrict $T : V \mapsto V$ to W and regard $T : W \mapsto W$ as a linear operator T_W. When V can be split into the direct sum of several proper subspaces W_i of V: $V = W_1 \oplus \cdots \oplus W_r$, one can reduce problems about T to ones about T_{W_i}. Since each W_i has smaller dimension than V, hopefully T_{W_i} is easier to analyse than T.

(5) Let W be an invariant subspace of T, and T_W be the restriction operator. Then the characteristic polynomial of T_W is related to that of T: the former divides the latter, see **Theorem 5.21**. When W is a proper invariant subspace of T, the degree of the characteristic polynomial of T_W is smaller than that of the characteristic polynomial of T.

(6) For any vector $\mathbf{v} \in V$, one can generate an invariant subspace of V by $\text{span}\{\mathbf{v}, T(\mathbf{v}), T^2(\mathbf{v}), \cdots\}$. It is called the T-cyclic subspace of V generated by $\mathbf{v}$. When V is finite dimensional,

this subspace must be finite dimensional as well. Let k denote the first $T^k(\mathbf{v})$ which can be represented as a linear combination of the previous terms in this sequence:

$$T^k(\mathbf{v}) = c_0\mathbf{v} + c_1T(\mathbf{v}) + \cdots + c_{k-1}T^{k-1}(\mathbf{v}).$$

This implies that $\{\mathbf{v}, T(\mathbf{v}), \cdots, T^{k-1}(\mathbf{v})\}$ is linearly independent, as otherwise, there would be some $T^i(\mathbf{v})$, $i \leq k-1$, which can be written as a linear combination of its previous terms. One can now see that any higher power $T^l(\mathbf{v})$, $l \geq k$, can be written as a linear combination of $\mathbf{v}, T(\mathbf{v}), \cdots, T^{k-1}(\mathbf{v})$. For instance,

$$\begin{aligned}T^{k+1}(\mathbf{v}) &= T(T^k(\mathbf{v}))\\ &= T(c_0\mathbf{v} + c_1T(\mathbf{v}) + \cdots + c_{k-1}T^{k-1}(\mathbf{v}))\\ &= c_0T(\mathbf{v}) + c_1T^2(\mathbf{v}) + \cdots + c_{k-1}T^k(\mathbf{v}))\\ &= c_0T(\mathbf{v}) + c_1T^2(\mathbf{v}) + \cdots + c_{k-1}\left[c_0\mathbf{v} + c_1T(\mathbf{v}) + \cdots + c_{k-1}T^{k-1}(\mathbf{v})\right]\end{aligned}$$

which is a linear combination of $\mathbf{v}, T(\mathbf{v}), \cdots, T^{k-1}(\mathbf{v})$. Thus k is the dimension of this cyclic invariant subspace, see **Theorem 5.22**. **Theorem 5.22** further says that the characteristic polynomial of T_W is $(-1)^k(t^k - c_{k-1}t^{k-1} - \cdots - c_0)$.

(7) Other invariant subspaces of T also arise naturally: $N(T)$, $R(T)$, $N(T^k)$, $R(T^k)$, $E_\lambda = N(T - \lambda I)$, $N((T - \lambda I)^k)$, $R((T - \lambda I)^k)$, etc.

(8) Here is a proof for # 21 in 5.4: If, for some $\mathbf{v} \in V$, the T-cyclic subspace W generated by $\mathbf{v}$ is 2 dimensional, then V must be equal to W, thus itself a T-cyclic subspace. It remains to consider the situation that for any non-zero $\mathbf{v} \in V$, the T-cyclic subspace W generated by $\mathbf{v}$ is 1-dimensional. This implies that for any $\mathbf{v} \in V$, $T(\mathbf{v}) = c\mathbf{v}$ for some scalar c, which at this point may depend on $\mathbf{v}$. We prove now that this c is independent of $\mathbf{v}$, thus $T(\mathbf{v}) = c\mathbf{v}$ for all $\mathbf{v} \in V$, proving that $T = cI$. Let $\{\mathbf{v}_1, \mathbf{v}_2\}$ be a basis for V. In our situation, $T(\mathbf{v}_1) = c_1\mathbf{v}_1$ for some scalar c_1, $T(\mathbf{v}_2) = c_2\mathbf{v}_2$, for some scalar c_2, and $T(\mathbf{v}_1 + \mathbf{v}_2) = c_3(\mathbf{v}_1 + \mathbf{v}_2)$, for some scalar c_3. It then follows that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = c_3(\mathbf{v}_1 + \mathbf{v}_2)$. Since $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a a basis for V, we conclude that $c_1 = c_2 = c_3$. Now, any vector $\mathbf{v} \in V$ can be written as $a\mathbf{v}_1 + b\mathbf{v}_2$ for some scalars a and b. It then follows that

$$T(\mathbf{v}) = T(a\mathbf{v}_1 + b\mathbf{v}_2) = aT(\mathbf{v}_1) + bT(\mathbf{v}_2) = a \cdot c_3\mathbf{v}_1 + b \cdot c_3\mathbf{v}_2 = c_3(a\mathbf{v}_1 + b\mathbf{v}_2) = c_3\mathbf{v},$$

namely, for all vector $\mathbf{v} \in V$, we have $T(\mathbf{v}) = c_3\mathbf{v}$.