MATH2040A Week 8 Tutorial Notes

In this tutorial, we will consider only finite dimensional vector spaces.

1 Diagonalizability

A linear map $T \in L(V)$ on a vector space V is *diagonalizable* if there exists a basis of V that consists only of eigenvectors of T. If T is diagonalizable with eigenbasis β, $[T]_\beta$ is then a diagonal matrix. As noted in lecture, diagonalizability depends on the behavior of the characteristic polynomial $f_T(t) = \det([T]_\alpha - tI)$:

Theorem 1.1. *T is diagonalizable if and only if f_T splits and the algebraic multiplicity of each eigenvalue is the same as their geometric multiplicity*[1].

Here,

- a polynomial *splits* if it can be factorized into linear factors (of form $X - a$) (optionally multiplied by a constant): $p(X) = c(X - a_1) \dots (X - a_k)$
- the *algebraic multiplicity* of a root λ of a polynomial $p(X)$ is the largest integer $m_\lambda \geq 1$ such that $(X-\lambda)^{m_\lambda}$ is a factor of $p(X)$
- the *geometric multiplicity* of an eigenvalue λ of a linear map T is $\text{nullity}(T - \lambda I) = \dim \mathsf{N}(T - \lambda I)$

As noted in lecture,

- a *complex* polynomial always splits
- for eigenvalue λ of T, $1 \leq \text{nullity}(T - \lambda I) \leq m_\lambda \leq n$
- if $p(X)$ splits and all (unique) roots are $\lambda_1, \dots, \lambda_k$, then $\sum m_{\lambda_i} = n$

So, to check if a linear map T is diagonalizable, typically you would need to

1. compute the characteristic polynomial, usually by computing the determinant
2. factorize the characteristic polynomial and check if it splits
3. for each eigenvalue, compute the dimension of $\mathsf{N}(T - \lambda I)$, usually by finding a basis
4. check for each eigenvalue if the two multiplicities match

Once diagonalizability is verified, it is simple to find a diagonalizing basis:

Theorem 1.2. *If T is diagonalizable with (unique) eigenvalues $\lambda_1, \dots, \lambda_k$, and β_i is a basis of $\mathsf{N}(T - \lambda I)$, then $\beta = \beta_1 \cup \dots \cup \beta_k$ is an eigenbasis of V*

So, to construct an eigenbasis, you just need to merge all bases you find in step 3 above into one basis.

By definition of diagonalizability, $[T]_\beta = \text{diag}(\lambda_1, \dots, \lambda_n)$ is diagonal if $\beta = \{v_1, \dots, v_n\}$ is an eigenbasis with associated eigenvalues $\lambda_1, \dots, \lambda_n$, and so on a basis α of V we have $[T]_\alpha = Q \,\text{diag}(\lambda_1, \dots, \lambda_n)\, Q^{-1}$ with $Q = [\text{Id}]_\beta^\alpha$. In particular, if $V = F^n$, $T = L_A$ with $A \in F^{n \times n}$, and α is the standard ordered basis, we recover the familiar eigendecomposition from MATH1030:

$$A = \begin{pmatrix} | & & | \\ v_1 & \dots & v_n \\ | & & | \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \begin{pmatrix} | & & | \\ v_1 & \dots & v_n \\ | & & | \end{pmatrix}^{-1}$$

[1] Such eigenvalue is said to be *semisimple*.

1.1 (Optional) Interpretation of Diagonalizability

Suppose $T \in L(V)$ is diagonalizable with (complete set of distinct) eigenvalues $\lambda_1, \ldots, \lambda_k$ and associated eigenspace $E_1 = \mathsf{N}(\, T - \lambda_1 \mathrm{Id}\,), \ldots, E_k$. From the matrix representation in eigenbasis, we can see the following decompositions:

- $V = E_1 \oplus \ldots \oplus E_k$[2]
- If $P_i \in L(V)$ is the projection map onto E_i along $\sum_{j \neq i} E_j$, then $P_i P_j = 0$ for $i \neq j$, $\sum P_i = \mathrm{Id}$ and $\sum \lambda_i P_i = T$

Recall from textbook Sec. 2.3 Q17 (Homework 5 Optional part) that a linear map $P \in L(V)$ is a projection if and only if $P^2 = P$.

Conversely, if V is a direct sum of eigenspaces of T, then T is diagonalizable: on basis β_i of eigenspace E_i of T, it is easy to verify that $\beta = \bigcup \beta_i$ is an eigenbasis of V. So, we have the following theorem:

Theorem 1.3. *T is diagonalizable if and only if V is a direct sum of eigenspaces of T.*

This decomposition is *sometimes* useful when working with diagonalizable operators. We will see more about this decomposition later when we are talking about a similar theorem on inner product spaces.

2 Cayley–Hamilton Theorem

In the lecture the following theorem is proven:

Theorem 2.1 (Cayley–Hamilton theorem)**.** *If V is a finite dimensional vector and $T \in L(V)$ with characteristic polynomial f_T, then $f_T(T) = 0$*[3].

The proof of this theorem is done by the following two concepts and one theorem:

Definition 2.1. The *T-cyclic subspace* generated by $v \in V$ / *Krylov subspace* generated by T and v is $K(T, v) = \mathrm{Span}(\, \{\, T^i v : i \geq 0 \,\}\,)$, which is the smallest T-invariant subspace that contains v.

Definition 2.2. For polynomial $p(X) = (-1)^n (X^n + c_{n-1} X^{n-1} + \ldots + c_0) \in P(F)$, the corresponding *companion matrix* is

$$A = \begin{pmatrix} 0 & 0 & \ldots & 0 & -c_0 \\ 1 & & & & -c_1 \\ & 1 & & & -c_2 \\ & & \ddots & & \vdots \\ & & & 1 & -c_{n-1} \end{pmatrix} \in F^{n \times n}$$

which, as you can verify, has characteristic polynomial $\det(A - tI) = p(t)$

Theorem 2.2. *If W is a T-invariant subspace, then the characteristic polynomial f_{T_W} of the restriction T_W of T on W is a factor of f_T: there exists a polynomial g such that $f_T(t) = g(t)\, f_{T_W}(t)$*

Occasionally this last theorem is quite powerful.

If $\dim(V) = n$, we have $\dim(L(V)) = n^2$ and so for a general linear map $T \in L(V)$ we may only expect $p(T) = 0$ for some polynomial with degree up to $n^2 - 1$, with little information on what this polynomial can be. Cayley–Hamilton theorem claims that this can always be done with a polynomial of degree n by choosing the characteristic polynomial of T.

One use of Cayley–Hamilton theorem is to quickly compute T^m with large m (or in general $p(T)$ for some polynomial p with high degree), with typical approaches like

---

[2] Here, $V = W_1 \oplus \ldots \oplus W_k$ means that for each $v \in V$ there exists unique $w_i \in W_i$ for each i such that $v = \sum w_i$. You can show that this is equivalent to $V = W_i \oplus (\sum_{j \neq i} W_j)$ for each i.

[3] Note that the characteristic polynomial f_T is just one of many polynomials that makes $p(T) = 0$ (annihilates T). With some knowledge from e.g. MATH3030, you can show that (a) there exists a unique nonzero polynomial $p_{\min}$ (the *minimal polynomial*) with leading coefficient 1 and has minimal degree that annihilates T, and (b) every (nonzero) polynomial that annihilates T is a multiple of $p_{\min}$. Usually $p_{\min}$ is not the characteristic polynomial.

- on $f_T(X) = (-1)^n X^n + \sum_{i=0}^{n-1} c_i X^i$, reduce each high order term X^k with $k \geq n$ to a lower degree polynomial via the identity $T^n = (-1)^n \sum_{i=0}^{n-1} c_i T^i$, then evaluate the low degree polynomial directly; or (equivalently)
- find the remainder polynomial r with $\deg(r) < n$ such that $X^m = f_T(X)q(X) + r(X)$, then evaluate $r(T)$

There are many ways to use Cayley–Hamilton theorem to simplify such computations.

3 Exercises

1. Let V be a finite dimensional vector space. Find all diagonalizable linear maps $T \in L(V)$ such that $(T - \mathrm{Id})^k = 0$ for some $k \geq 1$.

Solution: Let $T \in L(V)$ be one such map.
Let $\beta = \{ v_1, \ldots, v_n \}$ be an eigenbasis of T, and $\lambda_1, \ldots, \lambda_n$ be the associated eigenvalues.
Then for each i, $(T - \mathrm{Id})v_i = (\lambda_i - 1)v_i$, so $0 = (T - \mathrm{Id})^k v_i = (\lambda_i - 1)^k v_i$.
This implies that for each i, $(\lambda_i - 1)^k = 0$ and so $\lambda_i = 1$.
Thus $T = \mathrm{Id}$.

It is easy to verify that $T = \mathrm{Id}$ satisfies the requirement, so the only linear map that satisfies $(T - \mathrm{Id})^k = 0$ for some $k \geq 1$ is the identity map Id.

2. For $A = \begin{pmatrix} 1 & 4 \\ -1 & -3 \end{pmatrix} \in \mathbb{R}^{2\times 2}$, compute $A^{14}(A + 2I)^{13}$.

Solution: The characteristic polynomial of A is $f_A(t) = \det(A - tI_2) = \det \begin{pmatrix} 1-t & 4 \\ -1 & -3-t \end{pmatrix} = t^2 + 2t + 1$.
By Cayley–Hamilton theorem $A^2 + 2A + I = 0$, so $A(A + 2) = -I$.
This implies that $A^{14}(A + 2I)^{13} = (A(A + 2))^{13} A = (-I)^{13} A = -A = \begin{pmatrix} -1 & -4 \\ 1 & 3 \end{pmatrix}$

Note

Typically, there seems to be (at least) 3 ways to compute a matrix polynomial:

- compute the matrix exponential $A^0 = I, A, A^2, \ldots$ iteratively. With some tricks, you can compute compute $A^{14}(A + 2I)^{13}$ with about 7 matrix multiplications
- diagonalize $A = Q^{-1}DQ$ with diagonal D, then evaluate $A^{14}(A + 2I)^{13} = Q^{-1}\left(D^{14}(D + 2I)^{13}\right)Q$ where the middle term can be evaluated quickly since it is just operations on diagonal matrices. However, diagonalizing a matrix requires quite some computations, and A is not diagonalizable even when we consider the complex field.
(Instead of diagonalization, you can consider converting A into *Jordan normal form*, but that appears to be beyond the syllabus)
- find the remainder polynomial $r(t)$ such that $p(t) = q(t)f_A(t) + r(t)$, then evaluate $r(A)$ by other means (e.g. compute directly, which is usually easy enough)

The approach here in this solution, on the other hand, is to exploit the structure of the target polynomial and simplify it with the identity given by Cayley–Hamilton theorem.
As you may imagine, this approach is critically affected by the form of the characteristic polynomial. For example, consider $B = A + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ -1 & -3 \end{pmatrix}$ which has characteristic polynomial $t^2 + t - 2$. Cayley–Hamilton theorem does not (seem to) give a quick way to simplify $B^{14}(B + 2I)^{13}$, and in this case other approaches would preferable. (For example, B is diagonalizable.)

Solution: Here is another (more) brute-force approach with Cayley–Hamilton theorem.

The characteristic polynomial of A is $f_A(t) = t^2 + 2t + 1 = (t+1)^2$. On $p(t) = t^{14}(t+2)^{13}$, the degree of the remainder polynomial is at most 1.
More explicitly, $p(t) = q(t)(t+1)^2 + at + b$ for some scalars $a, b \in \mathbb{R}$ and some polynomial $q(t) \in \mathsf{P}(\mathbb{R})$.
Evaluating p and its (formal) derivative p' at $t = -1$, we obtain $1 = p(-1) = -a + b$ and $-14 + 13 = p'(-1) = a$, which implies $a = -1$, $b = 0$, and so by Cayley–Hamilton theorem $p(A) = -A = \begin{pmatrix} -1 & -4 \\ 1 & 3 \end{pmatrix}$.

Note

Here, the (formal) derivative $D : p \mapsto p'$ is the linear map on $\mathsf{P}(F) = F[X]$ that maps each monomial X^n to nX^{n-1}. You can verify that this (formal) derivative on polynomials, just like the (usual) derivative from real analysis, satisfies the product rule $(pq)' = p'q + pq'$.

Even though $B = \begin{pmatrix} 2 & 4 \\ -1 & -3 \end{pmatrix}$ does not (seem to) have a good characteristic polynomial to simplify $p(B)$, the same approach can also yield a quick result: $p(t) = q(t)f_B(t) + 3^{12}t + 2 \cdot 3^{12}$.

3. Let $T \in L(V)$ be a linear map on a finite dimensional vector space V over scalar field F, and $r = \operatorname{rank}(T)$. Show that there exists a polynomial $p \in \mathsf{P}_{r+1}(F)$ such that $p(T) = 0$.

Solution: Let $R = \mathsf{R}(\,T\,) \subseteq V$. It is already shown that R is T-invariant, so we may consider the restriction T_R of T on R.
By Cayley–Hamilton theorem, there exists a polynomial $p_R \in \mathsf{P}(F)$ of degree $\deg(p_R) = r$ such that $p_R(T_R) = 0$.
Consider the polynomial $q(X) = p_R(X)X \in \mathsf{P}(F)$ which is of degree $\deg(q) = \deg(p_R) + 1 = r + 1$. Since for each $v \in V$, $Tv \in R$, we have $T^k Tv = T_R^k Tv$ for all $k \geq 0$, and so by induction $p(T)Tv = p(T_R)Tv$ for all polynomials $p \in \mathsf{P}(F)$.
This implies that $q(T)v = p_R(T)Tv = p_R(T_R)Tv = 0$ for all $v \in V$, so $q(T) = 0$.

4. Let V be a nontrivial finite dimensional vector space over *complex number*, and $T, U \in L(V)$ be such that $TU = UT$. Show that there exists a nonzero vector in V that is an eigenvector to both T, U.

Solution: Since V is nontrivial and finite dimensional, by the fundamental theorem of algebra, T has an eigenvalue $\lambda \in F$. Let $v \in V$ be an associated eigenvector, so $v \neq 0$ and $Tv = \lambda v$.
Then $TUv = UTv = \lambda Uv$, so $Uv \in E_\lambda(T)$. By induction, $U^n v \in E_\lambda(T)$ for all $n \geq 0$, so the U-cyclic subspace $K = \operatorname{Span}(\ \{ U^n v : n \geq 0 \}\ )$ is a subspace of $E_\lambda(T)$.
By definition, K is U-invariant, so we may consider the restriction U_K of U on K.
Since $v \in K$, K is nontrivial. By the fundamental theorem of algebra, U_K has an eigenvalue μ.
This implies that there exists a nonzero $w \in K \subseteq E_\lambda$ that is an eigenvector of U_K, so an eigenvalue of U.
As $w \in E_\lambda(T)$ and is nonzero, w is also an eigenvector of T (with eigenvalue λ).
So T, U has a common eigenvector.

Note

The idea of this approach is to exploit the fact that U must have an eigenvector in a nontrivial U-invariant subspace, and construct such a subspace in an eigenspace of T.

We can consider $T_1, \ldots, T_n \in L(V)$ for some $n \geq 2$ that are pairwise commuting $T_iT_j = T_jT_i$. By repeating this argument, we can show that they must share some common eigenvector. The complex field requirement can also be relaxed if we assume that the characteristic polynomials of T and U split.