Matrix Analysis Homework 8

Michael Nelson

Problem a

Problem a.1

The matrix A has eigenvalue 0 with multiplicity n if and only if X^n divides $\chi_A(X)$. Since $\chi_A(X)$ has degree n, we have $aX^n = \chi_A(X)$ where $a \in \mathbb{C}$. Since the lead coefficient of $\chi_A(X)$ is $(-1)^n$, we must have $a = (-1)^n$. Thus $\chi_A(X) = (-1)^n X^n$. In particular, the Cayley-Hamilton Theorem implies

$$\begin{aligned} 0 &= \chi_A(A) \\ &= (-1)^n A^n. \end{aligned}$$

which implies $A^n = 0$. The matrix

$$E_{12} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

is nonzero and satisfies $E_{12}^n = 0$.

Problem a.2

Let $n \in \mathbb{N}$. Let

$$K[T]_{<n} = \{p(T) \in K[T] \mid \deg(p(T)) < n\}.$$

Let $D\colon K[T]_{<n} \to K[T]_{<n}$ be the differentiation operator, given by

$$D(p(T)) = p'(T)$$

for all $p(T) \in K[T]_{<n}$. Observe that $D^n = 0$ (over any field). Indeed, the differentiation operator drops the degree of a polynomial by at least 1. Since any $p(T) \in K[T]_{>n}$ has degree at most $n-1$, applying the differentiation operator n times to $p(T)$ will drop the degree by at least n, which can only mean $D^n(p(T)) = 0$. We now describe all nonzero D-invariant subspaces of $K[T]_{<n}$ (obviously the zero subspace of $K[T]_{<n}$ is D-invariant). To do so, we must consider two cases.

Case 1: K has characteristic 0 or K has characteristic p with $p \geq n$. Let $m \in \mathbb{N}$ such that $1 \leq m \leq n$. Let

$$K[T]_{<m} = \{p(T) \in K[T] \mid \deg(p(T)) < m\}.$$

Then $K[T]_{<m}$ is a nonzero D-invariant subspace of $K[T]_{<n}$. Indeed, it is nonzero since the constant polynomial 1 belongs to $K[T]_{<m}$. Also, if $p(T) \in K[T]_{<m}$, then $\deg(p'(T)) < m-1$ implies $p'(T) \in K[T]_{<m}$. We claim that these are all of the nonzero D-invariant subspaces of $K[T]_{<n}$. To see this, we first make the following observation. For all k such that $1 \leq k \leq n$, we have

$$\ker(D^k) = K[T]_{<k}. \tag{1}$$

Indeed, it is clear that $K[T]_{<k} \subseteq \ker(D^k)$ since D^k drops the degree of any polynomial by at least k. For the reverse direction, assume for a contradiction that $K[T]_{<k} \not\supseteq \ker(D^k)$. Choose a $p(T) \in \ker(D^k)$ of degree m where $m \geq k$. Let $a_m T^m$ be the lead term of $p(T)$ (so $a_m \neq 0$). Then $m(m-1)\cdots(m-k+1)a_m T^{m-k}$ is the lead term of $p^{(k)}(T)$. Since K has characteristic 0 or K has characteristic p with $p \geq n$, we must have $m(m-1)\cdots(m-k+1)a_m \neq 0$. Thus the lead term of $p^{(k)}(T)$ is nonzero. This contradicts the fact that $p(T) \in \ker(D^k)$.

Now let W be any nonzero D-invariant subspace of $K[T]_{<n}$. Choose a nonzero polynomial in W with maximum degree, say $p(T) \in W$ with $\deg(p(T)) = k$ where $0 \leq k < n$. Thus if $q(T) \in W$, then $\deg(q(T)) \leq k$. In particular, we have $W \subseteq K[T]_{<k+1}$. We claim that $\{p(T), p'(T), \dots, p^{(k)}(T)\}$ is a linearly independent set. If $k = 0$, then $\{p(T)\}$ is linearly independent since $p(T)$ is nonzero, thus assume $k > 0$. Suppose

$$c_0 p(T) + c_1 p'(T) + \cdots + c_k p^{(k)}(T) = 0 \tag{2}$$

for some $c_0, c_1, \dots, c_k \in K$. Applying D^k to both sides of (2) gives us

$$c_0 p^{(k)}(T) = 0 \tag{3}$$

by (1). The equation (3) implies $c_0 = 0$. Thus we may rewrite (2) as

$$c_1 p'(T) + \cdots + c_k p^{(k)}(T) = 0. \tag{4}$$

Applying D^{k-1} to both sides of (4) gives us

$$c_1 p^{(k)}(T) = 0 \tag{5}$$

by (1). The equation (5) implies $c_1 = 0$. Proceeding with this argument inductively, we obtain $c_0 = c_1 = \cdots = c_k = 0$. Thus $\{p(T), p'(T), \dots, p^{(k)}(T)\}$ is a linearly independent subset of W. Since $W \subseteq K[T]_{<k+1}$, $\dim K[T]_{<k+1} = k+1$, and W contains a linearly independent set of size $k+1$, we must have $W = K[T]_{<k+1}$.

Thus $K[T]_{<n}$ is not decomposable in this case since we cannot write $K[T]_{<n}$ as a direct sum of subspaces of the form $K[T]_{<m}$ where $1 \leq m < n$ (every such subspace has nonzero intersection, namely $K[T]_{<1}$ belongs to the intersection of any two subspaces of the form $K[T]_{<m}$ where $1 \leq m < n$).

Case 2: Assume has characteristic p where $p < n$. Write $n = pk + r$ where $k \geq 1$ and $0 \leq r < p$. For each $p(T) \in K[T]_{<n}$, write

$$p(T) = p_0(T) + T^p p_1(T) + \cdots + T^{(k-1)p} p_{k-1}(T) + T^{kp} p_k(T) \tag{6}$$

where $p_i(T) \in K[T]_{<p}$ for all $0 \leq i \leq k-1$ and $p_k(T) \in K[T]_{<r}$. Using the expression (6), we obtain an isomorphism of vector spaces

$$\Psi \colon K[T]_{<n} \to \bigoplus_{j=0}^{k-1} K[T]_{<p} \oplus K[T]_{<r},$$

given by

$$\Psi(p(T)) = (p_0(T), \dots, p_{k-1}(T), p_k(T))$$

for all $p(T) \in K[T]_{<n}$. In particular, the D-invariant subspaces of $K[T]_{<n}$ are in one-to-one correspondence with the $(\Psi \circ D \circ \Psi^{-1})$-invariant subspaces of $\bigoplus_{j=0}^{k-1} K[T]_{<p} \oplus K[T]_{<r}$ (by Proposition (0.1) given in the Appendix). Thus it suffices to describe all $(\Psi \circ D \circ \Psi^{-1})$-invariant subspaces of $\bigoplus_{j=0}^{k-1} K[T]_{<p} \oplus K[T]_{<r}$. A calculation gives

$$\begin{aligned}
(\Psi \circ D \circ \Psi^{-1})(p_0(T), \dots, p_{k-1}(T), p_k(T)) &= \Psi(D(\Psi^{-1}(p_0(T), \dots, p_{k-1}(T), p_k(T))) \\
&= \Psi(D(p_0(T) + \cdots + T^{(k-1)p} p_{k-1}(T) + T^{kp} p_k(T)) \\
&= \Psi(p_0'(T) + \cdots + T^{(k-1)p} p_{k-1}'(T) + T^{kp} p_k'(T)) \\
&= (p_0'(T), \dots, p_{k-1}'(T), p_k'(T)).
\end{aligned}$$

for all $(p_0(T), \dots p_{k-1}(T), p_k(T)) \in \bigoplus_{j=0}^{k-1} K[T]_{<p} \oplus K[T]_{<r}$. Thus the map $\Psi \circ D \circ \Psi^{-1}$ is simply the derivative operator taken component-wise. Thus it suffices to describe all D-invariant subspaces of $K[T]_{<p}$ and $K[T]_{<r}$ since every $(\Psi \circ D \circ \Psi^{-1})$-invariant subspace is a direct sum of D-invariant subspaces of each of the $K[T]_{<p}$ and of $K[T]_{<r}$. Both of these cases are described in case 1.

Contrary to case 1, $K[T]_{<n}$ is decomposable in case 2. Indeed, $K[T]_{<n}$ being decomposable under D is equivalent to $\bigoplus_{j=0}^{k-1} K[T]_{<p} \oplus K[T]_{<r}$ being decomposable under $\Psi \circ D \circ \Psi^{-1}$, and since $\Psi \circ D \circ \Psi^{-1}$ is just componentwise differentiation, $\bigoplus_{j=0}^{k-1} K[T]_{<p} \oplus K[T]_{<r}$ is obviously decomposable.

Problem b

Problem b.1

Let $\beta = \{v_1, \dots, v_n\}$. The matrix representation of f with respect to β is given by

$$[f]_\beta = \begin{pmatrix} A & 0 \\ 0 & \mathrm{I}_{n-m} \end{pmatrix}$$

where I_{n-m} is the $(n-m) \times (n-m)$ identity matrix and A is the $m \times m$ matrix given by

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

The characteristic polynomial of f is given by

$$\begin{aligned}
\chi_f(X) &= \det([f]_\beta - X\mathrm{I}_n) \\
&= \det\left(\begin{pmatrix} A & 0 \\ 0 & \mathrm{I}_{n-m} \end{pmatrix} - X\mathrm{I}_n\right) \\
&= \det\begin{pmatrix} A - X\mathrm{I}_m & 0 \\ 0 & (1-X)\mathrm{I}_{n-m} \end{pmatrix} \\
&= \det(A - X\mathrm{I}_m)\det((1-X)\mathrm{I}_{n-m}) \\
&= \det\begin{pmatrix} 1-X & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1-X & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1-X & \ddots & \cdots & 0 & 0 \\ 0 & 0 & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & \ddots & 1 \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 1-X \end{pmatrix} \det\begin{pmatrix} 1-X & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1-X & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1-X & \ddots & \cdots & 0 & 0 \\ 0 & 0 & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 1-X \end{pmatrix} \\
&= (1-X)^m(1-X)^{n-m} \\
&= (1-X)^n,
\end{aligned}$$

where we used the fact that the determinant of an upper triangular matrix is given by the product of its diagonal entries.

Problem b.2

A basis for the eigenspace E_1 (corresponding to $\lambda = 1$) is given by $\{v_1, v_{m+1}, v_{m+2}, \ldots, v_n\}$. Indeed, we have $f(v_1) = v_1$ and $f(v_k) = v_k$ for all $m < k \leq n$. Also the set $\{v_1, v_{m+1}, v_{m+2}, \ldots, v_n\}$ is clearly linearly independent. Thus it suffices to check that the dimension of E_1 is equal to the cardinality of the set $\{v_1, v_{m+1}, v_{m+2}, \ldots, v_n\}$. We have

$$\begin{aligned}
\operatorname{rank}([f]_\beta - \mathrm{I}_n) &= \operatorname{rank}\begin{pmatrix} A - \mathrm{I}_m & 0 \\ 0 & 0 \end{pmatrix} \\
&= \operatorname{rank}(A - \mathrm{I}_m) \\
&= \operatorname{rank}\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \\
&= m - 1.
\end{aligned}$$

Therefore

$$\begin{aligned}
\dim(E_1) &= \dim(\ker([f]_\beta - \mathrm{I}_n)) \\
&= n - \operatorname{rank}([f]_\beta - \mathrm{I}_n) \\
&= n - (m-1) \\
&= n - m + 1 \\
&= \#\{v_1, v_{m+1}, v_{m+2}, \ldots, v_n\}.
\end{aligned}$$

Problem b.3

Write $m_f(X)$ for the minimal polynomial of f. First note that the minimal polynomial has the same roots as the characteristic polynomial (see Proposition (0.3) in Appendix for proof). Moreover the minimal polynomial divides the characteristic polynomial. It follows that

$$m_f(X) = (X-1)^k$$

for some $1 \leq k \leq n$. A calculation shows

$$\begin{aligned}
([f]_\beta - \mathrm{I}_n)^k &= \left(\sum_{i=1}^{m-1} E_{i,i+1}\right)^k \\
&= \sum_{i=1}^{m-k} E_{i,i+k} \\
&\neq 0
\end{aligned}$$

for all $1 \leq k < m$ and

$$([f]_\beta - \mathrm{I}_n)^k = 0$$

for all $m \leq k \leq n$. In particular, this implies

$$\begin{aligned}
m_f(X) &= m_{[f]_\beta}(X) \\
&= (X-1)^m.
\end{aligned}$$

Problem c

Problem c.1

Let $x, y \in E$ and let $\beta = \{e_1, \dots, e_n\}$ be a basis of E. Then there exists unique $c_1, \dots, c_n \in K$ and unique $d_1, \dots, d_n \in K$ such that

$$x = \sum_{i=1}^{n} c_i e_i \quad \text{and} \quad y = \sum_{i=1}^{n} d_i e_i.$$

The column representation of x and y with respect to the basis β is written

$$[x]_\beta := \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \quad \text{and} \quad [y]_\beta := \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix}.$$

Then

$$\begin{aligned}
[x]_\beta^\top A[y]_\beta &= \begin{pmatrix} c_1 & \cdots & c_n \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix} \\
&= \begin{pmatrix} c_1 & \cdots & c_i & \cdots & c_n \end{pmatrix} \begin{pmatrix} a_{11}d_1 + \cdots + a_{1n}d_n \\ \vdots \\ a_{i1}d_1 + \cdots + a_{in}d_n \\ \vdots \\ a_{n1}d_1 + \cdots + a_{nn}d_n \end{pmatrix} \\
&= c_1(a_{11}d_1 + \cdots + a_{1n}d_n) + \cdots + c_i(a_{i1}d_1 + \cdots + a_{in}d_n) + \cdots + c_n(a_{n1}d_1 + \cdots + a_{nn}d_n) \\
&= \sum_{1 \leq i,j \leq n} a_{ij} c_i d_j \\
&= \sum_{1 \leq i,j \leq n} c_i d_j a_{ij} \\
&= \sum_{1 \leq i,j \leq n} c_i d_j \varphi(e_i, e_j) \\
&= \varphi\left(\sum_{i=1}^{n} c_i e_i, \sum_{j=1}^{n} d_j e_j\right) \\
&= \varphi(x, y).
\end{aligned}$$

Problem c.2

Assume that A is a symmetric matrix. Then $a_{ij} = a_{ji}$ for all $1 \leq i, j \leq n$. Let $x, y \in E$ where

$$x = \sum_{i=1}^{n} c_i e_i \quad \text{and} \quad y = \sum_{j=1}^{n} d_j e_j.$$

Then

$$\begin{aligned}
\varphi(x, y) &= \sum_{1 \leq i, j \leq n} a_{ij} c_i d_j \\
&= \sum_{1 \leq i, j \leq n} a_{ji} c_i d_j \\
&= \sum_{1 \leq i, j \leq n} a_{ji} d_j c_i \\
&= \varphi(y, x).
\end{aligned}$$

Conversely, if φ is symetric, then we have

$$\begin{aligned}
a_{ij} &= \varphi(e_i, e_j) \\
&= \varphi(e_j, e_i) \\
&= a_{ji}
\end{aligned}$$

for all $1 \leq i, j \leq n$. It follows that A is a symmetric matrix.

Problem c.3

Let $\gamma = \{f_1, \ldots, f_n\}$ be another basis of V and let $P = [1_V]_\gamma^\beta$ be the change of basis matrix from β to γ. Then

$$\begin{aligned}
\varphi(x, y) &= [x]_\beta^\top A [y]_\beta \\
&= (P[x]_\gamma)^\top A \, (P[y]_\gamma) \\
&= [x]_\gamma P^\top A P [y]_\gamma.
\end{aligned}$$

It follows that the matrix representation with respect to γ is $P^\top AP$.

Appendix

Problem a.2

Proposition 0.1. *Let $\Psi \colon V_1 \to V_2$ be an isomorphism from the vector space V_1 to the vector space V_2 and let $T \colon V_1 \to V_1$ be a linear map. Then the T-invariant subspaces of V_1 are in one-to-one correspondence with the $(\Psi \circ T \circ \Psi^{-1})$-invariant subspaces of V_2.*

Proof. Let $\mathrm{Inv}_T(V_1)$ denote the set of T-invariant subspaces of V_1 and let $\mathrm{Inv}_{\Psi \circ T \circ \Psi^{-1}}(V_2)$ denote the set of $(\Psi \circ T \circ \Psi^{-1})$-invariant subspaces of V_2. The isomorphism $\Psi \colon V_1 \to V_2$ induces a bijection $\Psi \colon \mathrm{Inv}_T(V_1) \to \mathrm{Inv}_{\Psi \circ T \circ \Psi^{-1}}(V_2)$ given by $W_1 \mapsto \Psi(W_1)$. Observe that this map lands in the target space. Indeed, if $W_1 \in \mathrm{Inv}_T(V_1)$, then

$$\begin{aligned}
(\Psi \circ T \circ \Psi^{-1})(\Psi(W_1)) &= (\Psi \circ T)(\Psi \circ \Psi^{-1})(W_1) \\
&= (\Psi \circ T)(W_1) \\
&= \Psi(T(W_1)) \\
&\subset \Psi(W_1).
\end{aligned}$$

The inverse map is given by $\Psi^{-1} \colon \mathrm{Inv}_{\Psi \circ T \circ \Psi^{-1}}(V_2) \to \mathrm{Inv}_T(V_1)$. $\square$

Proposition 0.2. *Let $V = V_1 \oplus \cdots \oplus V_n$ be a direct sum of vectors spaces $V_1, \ldots, V_n$. Let $T \colon V \to V$ be given by $T = \oplus_i T_i$ where $T_i \colon V_i \to V_i$ are linear maps for each $1 \leq i \leq n$. Then the T-invariant subspaces of V consist of subspaces of the form*

$$W = W_1 \oplus \cdots \oplus W_n \tag{7}$$

where W_i is a T_i-invariant subspace for each $1 \leq i \leq n$.

Proof. Let $W = W_1 \oplus \cdots \oplus W_n$ be a subspace of V such that W_i is T_i-invariant for all $1 \leq i \leq n$. Let $w \in W$ and write $w = w_1 + \cdots + w_n$ where $w_i \in W_i$ for all $1 \leq i \leq n$. Then

$$\begin{aligned} T(w) &= T(w_1 + \cdots + w_n) \\ &= T(w_1) + \cdots + T(w_n) \\ &= T_1(w_1) + \cdots + T_n(w_n) \\ &\in W_1 \oplus \cdots \oplus W_n \\ &= W. \end{aligned}$$

Thus W is T-invariant. Conversely, let $W = W_1 \oplus \cdots \oplus W_n$ be any T-invariant subspace of V. Then for any $1 \leq i \leq n$ and for any $w \in W_i$, we have

$$\begin{aligned} T_i(w) &= T(w) \\ &\subseteq W. \end{aligned}$$

Since $\operatorname{im}(T_i) \subseteq V_i$, this implies $T_i(w) \in W \cap V_i = W_i$. Thus W_i is T_i-invariant for all $1 \leq i \leq n$. □

Problem b.3

Proposition 0.3. *Let $T \colon V \to V$ be a linear map from a finite dimensional K-vector space V to itself. Suppose λ is an eigenvalue of T. Then λ is a root of the minimal polynomial $m_T(X)$. Conversely, if α is a root of $m_T(X)$, then α is an eigenvalue of T.*

Proof. Choose an eigenvector $v \in V$ corresponding to the eigenvalue λ. Since $T^k v = \lambda^k v$ for all $k \geq 1$, we have $p(X)(v) = p(\lambda)(v)$ for all $p(X) \in K[X]$. Thus since $m_T(X)$ kills all of V, we have

$$\begin{aligned} 0 &= m_T(X) \cdot v \\ &= m_T(T)(v) \\ &= m_T(Tv) \\ &= m_T(\lambda)v, \end{aligned}$$

which implies $m_T(\lambda) = 0$. For the converse direction, we have

$$\begin{aligned} \alpha \text{ is a root of } m_T(X) &\implies (X - \alpha) \text{ divides } m_T(X) \\ &\implies (X - \alpha) \text{ divides } \chi_T(X) \\ &\implies \alpha \text{ is a root of } \chi_T(X) \\ &\implies \alpha \text{ is an eigenvalue of } T. \end{aligned}$$

□