MATH 217 - W24 - LINEAR ALGEBRA
HOMEWORK 4, SOLUTIONS

Part A (15 points)

Solve the following problems from the book:
Section 2.4: 28, 30, 42
Section 3.1: 6, 14

Solution.

2.4.28: We are asked to find the inverse of the matrix $A = \begin{bmatrix} 22 & 13 & 8 & 3 \\ -16 & -3 & -2 & -2 \\ 8 & 9 & 7 & 2 \\ 5 & 4 & 3 & 1 \end{bmatrix}$. If we row-reduce the 4×8 matrix $[A \mid I_4]$ to its reduced row echelon form using Gauss-Jordan elimination, we will obtain the matrix $[I_4 \mid A^{-1}]$. Thus

$$A^{-1} = \begin{bmatrix} 1 & -2 & 9 & -25 \\ -2 & 5 & -22 & 60 \\ 4 & -9 & 41 & -112 \\ -9 & 17 & 80 & 222 \end{bmatrix},$$

and T^{-1} is the transformation from $\mathbb{R}^4$ to $\mathbb{R}^4$ with matrix A^{-1}.

(It is perhaps worth pointing out here that *once* you have mastered the technique of Gauss-Jordan elimination, there is no need to continue carrying out laborious calculations that are better left to a computer; whenever such computations get too complicated, it's best to ask Wolfram Alpha to do them!)

2.4.30 The question is equivalent to: for which constants b, c is $\operatorname{rref}(A) = I_3$? We put the matrix into reduced row echelon form:

$$\begin{bmatrix} 0 & 1 & b \\ -1 & 0 & c \\ -b & -c & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -c \\ 0 & 1 & b \\ -b & -c & 0 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -c \\ 0 & 1 & b \\ 0 & -c & -bc \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -c \\ 0 & 1 & b \\ 0 & 0 & 0 \end{bmatrix}$$

We see that the RREF has a row of zeroes in it, which means that this matrix is *never* invertible for any values of $b, c \in \mathbb{R}$.

2.4.42: Permutation matrices are invertible since they can be row-reduced to an identity matrix by applying a sequence of row-interchange operations. The inverse of a permutation matrix A is also a permutation matrix, since $\operatorname{rref}\big([A \mid I_n]\big) = [I_n \mid A^{-1}]$ is obtained from $[A \mid I_n]$ by a sequence of row swaps.

3.1.6: Solving $A\vec{x} = \vec{0}$ yields the $\ker(A) = \operatorname{span} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$.

3.1.14: By theorem 3.1.3, the image of A is the span of the columns vectors of A:

$$\operatorname{im}(A) = \text{ span } \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} \right\}.$$

Since these three vectors are parallel, we only need one of them to span the image. Thus $\operatorname{im}(A) = \text{span} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Part B (25 points)

Problem 1. Let $\mathbb{R}^n \xrightarrow{T} \mathbb{R}^n$ be a linear transformation. As on HW 3, we define T^k to be the k-fold composition of T with itself,

$$= \underbrace{T \circ T \circ T \circ \cdots \cdots \circ T}_{k \text{ times}}.$$

Let A be the standard matrix of T, by which we mean the unique $n \times n$ matrix such that $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$.

(a) Prove that for all k, the standard matrix for T^k is the matrix A^k. [Hint: induction works nicely.]

(b) We define T to be **nilpotent** if there exists some $k \in \mathbb{N}$ such that T^k is the zero transformation. Prove that if T is nilpotent, then A is not invertible.

(c) Prove that if T is nilpotent, then $A - I_n$ **is** invertible.
[Hint: try multiplying out $(A - I_n)(I_n + A + A^2 + \cdots + A^{k-1})$ and see what you get.]

Solution.

(a) We induct on k.

Base case: if $k = 1$, then $T^1 = T$, so the statement is our starting assumption that A is the standard matrix of T, as $A = A^1$.

Inductive step: Assume that T^k has standard matrix A^k. Since standard matrix of a composition $T_1 \circ T_2$ is the product $A_1 A_2$ of the corresponding matrices (by the book's Definition 2.3.1, or Problem 1f on Worksheet 7), then $T^{k+1} = T \circ T^k$ has standard matrix

$$A\, A^k = A^{k+1}.$$

This completes the proof, by induction.

(b) Assume T^k is the zero transformation. Assume, by way of contradiction, that A is invertible. Note that this implies A^k is invertible for all k. (This was on HW 3, Problem 2d; it can also be proved directly by induction.) On the other hand, by part (a), we know that A^k is the standard matrix T^k, which (by hypothesis) sends every vector, including $\vec{e}_i$, to zero; consequently each column of A^k is zero by the Key theorem. In other words, A^k is the zero matrix, which has rank 0. This contradicts the fact that A is invertible, because an invertible $n \times n$ matrix has rank n, not zero (Theorem 2.4.3).

(c) Assume that A^k is the 0-matrix and use the convention $A^0 = I_n$ (note that this is consistent with our convention that T^0 is the identity function; see HW 3, Problem 5).

Consider the product

$$\begin{aligned}(A - I_n)(I_n + A + A^2 + \cdots + A^{k-1}) &= A(I_n + A + \cdots + A^{k-1}) - I_n(I_n + A + \cdots + A^{k-1}) \\ &= (A + A^2 + \cdots + A^k) - (I_n + A + \cdots + A^{k-1}) \\ &= A^k - I_n \\ &= -I_n\end{aligned}$$

where in the last step we have used the assumption that $A^k = 0$. The equation

$$(A - I_n)(I_n + A + A^2 + \cdots + A^{k-1}) = -I_n$$

is equivalent to

$$(A - I_n)(-I_n - A - A^2 - \cdots - A^{k-1}) = I_n$$

which shows that the matrix $-I_n - A - A^2 - \cdots - A^{k-1}$ is the inverse matrix of $A - I_n$.

Problem 2. Let V be any vector space, and let S be any set. Let $\mathcal{F}(S, V)$ denote the set of all functions from S to V. (Note: we are not assuming $S \subseteq V$ here, just that S is some set. S is not assumed to be a vector space, but it could be. Similarly, the functions in $\mathcal{F}(S, V)$ are not assumed to be linear transformations, although it is possible that some of them might be.)

For any functions $f, g \in \mathcal{F}(S, V)$ we can define their *sum* to be the function $f + g$ given by the formula $(f + g)(s) = f(s) + g(s)$, where s is any element in S. Similarly, for any scalar $c \in \mathbb{R}$ and any function $f \in \mathcal{F}(S, V)$ we define the function cf to be given by the formula $(cf)(s) = c(f(s))$ for all $s \in S$.

(a) Prove that $\mathcal{F}(S, V)$ is a vector space. **Note:** For this problem you must *explicitly prove* that each of the vector space properties VS1-8 from Worksheet 6 is true. (These proofs should be very short but are not skippable.)
(b) Is $0_{\mathcal{F}(S,V)}$ the same element as 0_V? If not, explain how they are different.
(c) We could similarly define $\mathcal{F}(V, S)$ to be the set of all functions from V to S. Would $\mathcal{F}(V, S)$ also a vector space? Why or why not?
(d) The familiar vector spaces $\mathcal{P}, \mathcal{P}_n$ and $\mathcal{C}^\infty$ (all from Worksheet 6) are all subsets of $\mathcal{F}(S, V)$ for some S and V. What are S and V for each of these functions?

Solution.

(a) We first observe that $\mathcal{F}(S, V)$ is closed under addition and scalar multiplication: if $f, g \in \mathcal{F}(S, V)$ then the definition $(f + g)(s) = f(s) + g(s)$ defines a function $f + g \in \mathcal{F}(S, V)$, and likewise if $c \in \mathbb{R}$ then the definition $(cf)(s) = c(f(s))$ defines a function $cf \in \mathcal{F}(S, V)$.

Now we verify all of the properties of a vector space from Worksheet 6:

(VS-1) We must show that for all $f, g, h \in \mathcal{F}(S, V)$, $(f + g) + h = f + (g + h)$. Note that both the left-hand side and the right-hand side of this equation are *functions*. To prove that two functions are equal, we must prove that they agree for all elements of their domain: that is, we must prove that $((f + g) + h)(s) = (f + (g + h))(s)$ for all $s \in S$. By our definition of function addition, this is equivalent to proving that $(f(s) + g(s)) + h(s) = f(s) + (g(s) + h(s))$. Notice that in this equation both the left-hand side and the right-hand side are *elements of* V. We already know that V is a vector space, hence addition of elements in V already satisfies property

(VS-1). Consequently $(f(s)+g(s))+h(s) = f(s)+(g(s)+h(s))$ holds for all $s \in S$, and we conclude that $(f+g)+h = f+(g+h)$.

(VS-2) We follow the same outline as in (VS-1), but in less detail. We want to show that for all $f, g \in \mathcal{F}(S, V)$, $f+g = g+f$. This is the same as proving that $(f+g)(s) = (g+f)(s)$ for all $s \in S$, which is equivalent to proving $f(s)+g(s) = g(s) = f(s)$ for all $s \in S$, which is true because addition in V satisfies (VS-2).

(VS-3) The element $0_{\mathcal{F}(S,V)}$ is the function that sends every element of S to 0_V. We can verify that for any $f \in \mathcal{F}(S, V)$,

$$(f + 0_{\mathcal{F}(S,V)})(s) = f(s) + 0_{\mathcal{F}(S,V)}(s) = f(s) + 0_V = f(s)$$

(VS-4) Let $f \in \mathcal{F}(S, V)$. We define a new function $-f$ by the rule $(-f)(s) = -(f(s))$. We can verify that for all $s \in S$,

$$(f + (-f))(s) = f(s) + (-f)(s) = f(s) + (-f(s)) = 0_V$$

Since $f + (-f)$ sends every element of S to 0_V, we conclude $f + (-f) = 0_{\mathcal{F}(S,V)}$.

(VS-5) Let $c \in \mathbb{R}$ and let $f, g \in \mathcal{F}(S, V)$. We want to show $c(f+g) = cf + cg$. That is, we want to show that for all $s \in S$, $(c(f+g))(s) = (cf)(s) + (cg)(s)$. By our definitions, this is equivalent to proving $c(f(s)+g(s)) = cf(s) + cg(s)$, which is true because V has property (VS-5).

Have you noticed the trend yet? In each case, the set of functions $\mathcal{F}(S, V)$ "inherits" the relevant property of a vector space from V, which has all of those properties by hypothesis.

(VS-6) Let $a, b \in R$ and let $f \in \mathcal{F}(S, V)$. We want to show $(a+b)f = af + bf$, i.e. we want to show that for all $s \in S$, $(a+b)f(s) = af(s) + bf(s)$, which is true by (VS-6) in V.

(VS-7) Let $a, b \in R$ and let $f \in \mathcal{F}(S, V)$. We want to show $a(bf) = (ab)f$, i.e. we want to show that for all $s \in S$, $a(bf(s)) = (ab)f(s)$, which is true by (VS-7) in V.

(VS-8) Let $f \in \mathcal{F}(S, V)$. We want to show that $1f = f$, i.e. we want to show that for all $s \in S$, $1f(s) = f(s)$, which is true by (VS-8) in V.

(b) As was mentioned above, $0_{\mathcal{F}(S,V)}$ is the *function* from S to V that maps each element $s \in S$ to $0_V \in V$. So these two "zero elements" are not the same thing. The difference is analogous to the distinction between the *number* 0, on the one hand, and the *function* $f(x) = 0$, on the other hand. The number 0 can be represented by a single point on a number line; the function $f(x) = 0$ would be represented by a horizontal line coinciding with the x-axis in the xy-plane.

(c) As soon as we try to define addition of functions in $\mathcal{F}(V, S)$ we immediately run into trouble. If $f, g \in \mathcal{F}(V, S)$ we cannot define $f + g \in \mathcal{F}(V, S)$ to be the function defined by $(f+g)(v) = f(v) + g(v)$ for all $v \in V$, because $f(v)$ and $g(v)$ are elements of S, and we have no way of knowing at all if there is a way to add elements of S together, or (if addition is defined) whether the sum belongs to S, or (if the sum does belong to S) whether any of the properties VS1-8 are satisfied. So S does not have any properties that $\mathcal{F}(V, S)$ can "inherit". Consequently, $\mathcal{F}(V, S)$ is not, in general, a vector space.

Of course, $\mathcal{F}(V, S)$ **could be** a vector space, if S happened to be one. But in that case we would really be back in the case of part (a).

(d) All of these vector spaces are subsets of $\mathcal{F}(\mathbb{R},\mathbb{R})$. (Once we learn the definition of "subspace" – coming up soon on Worksheet 9! – we will say: $\mathcal{P}, \mathcal{P}_n$ and C^∞ are all subspaces of $\mathcal{F}(\mathbb{R},\mathbb{R})$.)

Problem 3. Let $\mathcal{P}$ be the vector space of all polynomial functions from $\mathbb{R}$ to $\mathbb{R}$ in the variable t, and for each $n \in \mathbb{N}$, let $\mathcal{P}_n$ be (as usual) the subset of $\mathcal{P}$ consisting of all polynomial functions of degree at most n. (We already know that $\mathcal{P}_n$ is also a vector space.) Also let $T : \mathcal{P} \to \mathcal{P}$ be the map defined by $T(p)(t) = p'(t) + p(0)$ for each $p \in \mathcal{P}$ and for all $t \in \mathbb{R}$.

(a) Show that T is a linear transformation.

Solution. Let $p, q \in \mathcal{P}$ and $c \in \mathbb{R}$. Then, using familiar properties of the derivative from calculus, we see that for all $x \in \mathbb{R}$,

$$T(p+q)(x) = (p+q)'(x) + (p+q)(0) = p'(x) + p(0) + q'(x) + q(0) = T(p)(x) + T(q)(x)$$

and

$$T(cp)(x) = (cp)'(x) + (cp)(0) = cp'(x) + cp(0) = c(p'(x) + p(0)) = cT(p)(x).$$

This shows that T is linear.

(b) Let $n \in \mathbb{N}$, and let $T_n : \mathcal{P}_n \to \mathcal{P}_n$ be defined by $T_n(p)(t) = p'(t) + p(0)$, so that T_n is just T with both domain and codomain restricted to $\mathcal{P}_n$. Is T_n injective? Is T_n surjective?

Solution. If $n = 0$ then $\mathcal{P}_n$ is just the set of scalars (i.e., constant functions), and for any constant function $p(x) = c$ we have $T(p)(x) = c$ as well. So for the case $n = 0$ $T : \mathcal{P}_0 \to \mathcal{P}_0$ is just the identity function, which is both injective and surjective.

For the rest of the problem we assume $n \geq 1$, and consider $p, q \in \mathcal{P}_n$ given by $p(x) = x^2 + 1$ and $q(x) = x^2 + x$. Direct computation shows that $T(p)(x) = 2x + 1$ and $T(q)(x) = 2x + 1$. Since $p \neq q$ but $T(p) = T(q)$ we find that T is not injective.

Next, from calculus we know that if p is a polynomial of degree at most n, then $T(p)$ will be a polynomial of degree at most $n-1$, so there is no polynomial $p \in \mathcal{P}_n$ such that $T(p)(x) = x^n$. So T is not surjective, either.

(c) Is T injective? Is T surjective?

Solution. By the same argument given in part (b), T is not injective. However, in this case T *is* surjective, with $\operatorname{im}(T) = \mathcal{P}$. To see this we argue again as in part (b). Simply observe that for any polynomial $p(x)$, if we let $q(x)$ denote the antiderivative of $p(x)$ that has zero for its constant term (i.e., for which $q(0) = 0$), then by construction $T(q)(x) = p(x)$.

Problem 4. We denote by $\mathbb{R}^{n\times n}$ the vector space of all $n \times n$ matrices. Let A be an $n \times n$ matrix, and define the function $L_A : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ by $L_A(B) = AB$ for all $B \in \mathbb{R}^{n\times n}$. (Note carefully: this is *not* the same function as T_A. While both can be described informally as "multiplication by A", the two functions L_A and T_A have different domains and codomains. Make sure you understand this distinction before beginning to work on this problem!)

(a) Show that L_A is a linear transformation.

Solution. Let $B, C \in \mathbb{R}^{n\times n}$ be any two $n \times n$ matrices. Then

$$L_A(B+C) = A(B+C) = AB + AC = L_A(B) + L_A(C)$$

Also, let $k \in R$ be any scalar and let $B \in \mathbb{R}^{n\times n}$ be any $n \times n$ matrix. Then

$$L_A(kB) = A(kB) = k(AB) = kL_A(B)$$

which completes the proof.

(b) Show that the matrix A is invertible if and only if the linear transformation L_A is invertible.

Solution.

Suppose first that A is invertible (so that a matrix A^{-1} exists, with the property that $AA^{-1} = A^{-1}A = I_n$). We want to prove that L_A is invertible. We claim that $L_{A^{-1}}$ is the inverse of L_A. To prove this, we must show that $L_A \circ L_{A^{-1}} = Id_{\mathbb{R}^{n\times n}}$ and $L_{A^{-1}} \circ L_A = Id_{\mathbb{R}^{n\times n}}$. To verify this, we calculate the action of each of these compositions on an arbitrary matrix $B \in \mathbb{R}^{n\times n}$:

$$(L_A \circ L_{A^{-1}})(B) = L_A(L_{A^{-1}}(B)) = L_A(A^{-1}B) = A(A^{-1}B) = (AA^{-1})B = I_nB = B$$

and

$$(L_{A^{-1}} \circ L_A)(B) = L_{A^{-1}}(L_A(B)) = L_{A^{-1}}(AB) = A^{-1}(AB) = (A^{-1}A)B = I_nB = B$$

Now for the converse: we assume L_A is invertible, so there exists some linear transformation $T : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ such that $L_A \circ T = Id_{\mathbb{R}^{n\times n}}$ and $T \circ L_A = Id_{\mathbb{R}^{n\times n}}$. Let $B = T(I_n)$. We claim that in fact $BA = AB = I_n$ (this would prove that $B = A^{-1}$ and therefore that A is invertible.) The following calculation shows that $AB = I_n$:

$$AB = L_A(B) = L_A(T(I_n)) = (L_A \circ T)(I_n) = I_n$$

Showing directly that $BA = I_n$ turns out to be trickier than it seems at first, so instead of doing that, we will finish the proof by citing Theorem 2.4.8 from the textbook: *if two **square** matrices A, B satisfy $AB = I_n$, then both A and B are invertible, with $B = A^{-1}$ and $A = B^{-1}$.* (The statement of this in the textbook actually interchanges the letters A and B but that doesn't make any substantive difference here.) [**WARNING: this is not true in general for non-square matrices!**] Since we have already proved $AB = I_n$ it follows that A is invertible, which completes the proof.

Now let $\mathcal{F}$ be the set of all functions from $\mathbb{R}^{n\times n}$ to $\mathbb{R}^{n\times n}$, and define the function $L : \mathbb{R}^{n\times n} \to \mathcal{F}$ by $L(A) = L_A$.

(c) Show that L is injective.

Solution. Suppose A, B are two $n \times n$ matrices such that $L_A = L_B$. Then $L_A(I_n) = L_B(I_n)$, which means that $AI_n = BI_n$, or in other words that $A = B$.

(d) Is L surjective? Be sure to justify your claim.

Solution. Definitely not! Let $G \in \mathcal{F}$ be any non-linear function. For example, suppose $G : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ is defined by $G(X) = I_n$ for all $X \in \mathbb{R}^{n\times n}$. This is not linear, because for any two matrices X, Y we would have $G(X + Y) = I_n \neq G(X) + G(Y)$. Since G

is not linear, there does not exist a matrix A such that $G = L_A = L(A)$ (since, by (a), any function of the form L_A would be a linear transformation).

Note: other examples of non-linear functions in $\mathcal{F}$ are $H(X) = X^2$ or $K(X) = X + I_n$.

Problem 5. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined as follows:

$$T = \mathrm{Rot}_{-80^\circ} \circ \mathrm{Proj}_y \circ \mathrm{Rot}_{35^\circ},$$

where Rot_θ is counter-clockwise rotation by θ, and Proj_y is projection onto the y-axis.

(a) Sketch $\mathrm{im}(T)$ in $\mathbb{R}^2$. Indicate the angle between $\mathrm{im}(T)$ and the x-axis.
(b) Sketch $\ker(T)$ in $\mathbb{R}^2$. Indicate the angle between $\ker(T)$ and the x-axis.
(c) Let $T_{\phi,\theta} := \mathrm{Rot}_\phi \circ \mathrm{Proj}_y \circ \mathrm{Rot}_\theta$. For which ϕ and θ is $\mathrm{im}(T_{\phi,\theta}) = \ker(T_{\phi,\theta})$?

Solution.

(a) $\mathrm{im}(T)$ is the line through the origin in $\mathbb{R}^2$ that makes an angle of 10° with the positive x-axis.
(b) $\ker(T)$ is the line through the origin in $\mathbb{R}^2$ that makes an angle of 145° with the positive x-axis.
(c) Multiplying the standard matrices of Rot_ϕ, Proj_y, and Rot_θ, we see that that standard matrix of $T_{\phi,\theta}$ is

$$\begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} -\sin\phi\sin\theta & -\sin\phi\cos\theta \\ \cos\phi\sin\theta & \cos\phi\cos\theta \end{bmatrix}.$$

Thus $\mathrm{im}(T_{\phi,\theta})$ is spanned by $\begin{bmatrix} -\sin\phi \\ \cos\phi \end{bmatrix}$, and $\ker(T_{\phi,\theta})$ is spanned by $\begin{bmatrix} -\cos\theta \\ \sin\theta \end{bmatrix}$, so $\mathrm{im}(T_{\phi,\theta}) = \ker(T_{\phi,\theta})$ precisely when $\sin\phi = \cos\theta$ and $\cos\phi = \sin\theta$. This happens whenever θ and ϕ are complementary angles, or (more generally) when $\theta + \phi = \pi/2 + k\pi$ for some $k \in \mathbb{Z}$.