Problem Set 2 - Solutions

Problem 1. *Carefully applying axioms*

It is important to make sure that all of our work in applications is properly founded in theory. One of the more common vector spaces we will work in is the space $(\mathcal{M}_{m\times n}(\mathbb{F}), \mathbb{F})$ of $m \times n$ matrices with entries in the field $\mathbb{F}$, where addition of matrices and $\mathbb{F}$-scalar multiples of matrices are defined in the usual, componentwise sense. We will now abbreviate $(\mathcal{M}_{m\times n}(\mathbb{F}), \mathbb{F})$ simply by its set of vectors $\mathbb{M}_{m\times n}(\mathbb{F})$.

(a) Carefully prove that $\mathcal{M}_{m\times n}(\mathbb{F})$ satisfies all 10(!) of the axioms of a vector space. [*Hint*: Much of the work will be taken care of by appealing to the field axioms that $\mathbb{F}$ satisfies.]

For simplicity, take $V = \mathcal{M}_{m\times n}(\mathbb{F})$. Let $A, B, C \in \mathcal{M}_{m\times n}$ with $(A)_{ij} = a_{ij}$, $(B)_{ij} = b_{ij}$, and $(C)_{ij} = c_{ij}$. Let $r, s \in \mathbb{F}$. I will use the same labels as the book (on page 76):

(i) By definition, $(A+B)_{ij} = a_{ij} + b_{ij}$. Because $a_{ij}, b_{ij} \in \mathbb{F}$, so is their sum. Hence, $A + B \in V$.

(ii) By definition, $(rA)_{ij} = ra_{ij}$. Because $r, a_{ij} \in \mathbb{F}$, so is their product. Hence, $rA \in V$.

(a) Consider that $(B + A)_{ij} = b_{ij} + a_{ij} = a_{ij} + b_{ij} = (A + B)_{ij}$, where the middle equality uses the commutativity of addition in $\mathbb{F}$. Hence, $B + A = A + B$.

(b) Consider that $(A + [B + C])_{ij} = a_{ij} + (b_{ij} + c_{ij}) = (a_{ij} + b_{ij}) + c_{ij} = ([A+B]+C)_{ij}$, where the middle equality uses the associativity of addition in $\mathbb{F}$.

(c) The element $\mathbb{O}$ defined as $(\mathbb{O})_{ij} = 0 \in \mathbb{F}$ defines the zero vector in V. Indeed, $(A + \mathbb{O})_{ij} = a_{ij} + 0 = a_{ij} = 0 + a_{ij} = (\mathbb{O} + A)_{ij}$, where we have used the properties of the zero element of $\mathbb{F}$ in the middle two equalities.

(d) The element $-A$ defined by $(-A)_{ij} = -a_{ij}$ defines the additive inverse of A. Indeed, $(A+[-A])_{ij} = a_{ij}+(-a_{ij}) = 0 = (-a_{ij})+a_{ij} = ([-A]+A)_{ij}$, where we have used the existence and properties of the additive inverse $-a_{ij}$ of a_{ij} in $\mathbb{F}$.

(e) Consider that $([r+s]A)_{ij} = (r+s)a_{ij} = ra_{ij}+sa_{ij} = (rA+sA)_{ij}$, where we have used the distributivity of scalars in $\mathbb{F}$ in the middle equality.

(f) Consider that $(c[dA])_{ij} = c(da_{ij}) = (cd)a_{ij} = ([cd]A)_{ij}$, where we have used the associativity of scalars of $\mathbb{F}$ in the middle equality.

(g) The scalar $1 \in \mathbb{F}$ indeed acts as a unit: $(1A)_{ij} = 1a_{ij} = a_{ij} = (A)_{ij}$, where we have used the unit properties of 1 in $\mathbb{F}$ in the middle equality.

(b) Define $\mathcal{M}_n(\mathbb{F}) \coloneqq \mathcal{M}_{n\times n}(\mathbb{F})$. Show that set $\mathcal{S}_n(\mathbb{F}) = \{A \in \mathcal{M}_n(\mathbb{F}) \mid A = A^T\}$ of symmetric matrices is a subspace of $\mathcal{M}_n(\mathbb{F})$.

Because $\mathcal{M}_n(\mathbb{F})$ is a vector space over $\mathbb{F}$, we can simply show that $\mathcal{S}_n(\mathbb{F})$ is closed under addition and scalar multiplication. Let $A, B \in \mathcal{S}_n(\mathbb{F})$, with $(A)_{ij} = a_{ij}$ and $(B)_{ij} = b_{ij}$, and $r \in \mathbb{F}$.

Consider that $([A+B]^T)_{ij} = (A+B)_{ji} = a_{ji} + b_{ji} = a_{ij} + b_{ij} = (A+B)ij$, by symmetry of A and B in the middle equality. Therefore, $A+B \in \mathcal{S}_n(\mathbb{F})$ as well.

Also consider that $([rA]^T)_{ij} = (rA)_{ji} = ra_{ji} = ra_{ij} = (rA)_{ij}$, by symmetry of A in the middle equality. Therefore, $rA \in \mathcal{S}_n(\mathbb{F})$ as well.

Therefore, $\mathcal{S}_n(\mathbb{F})$ is a subspace of $\mathcal{M}_n(\mathbb{F})$.

(c) Determine the dimension of $\mathcal{M}_{m\times n}(\mathbb{F})$ by exhibiting (with proof!) a set of matrices that constitutes a basis for this vector space.

Define $M_{ij} = [0\cdots 0\ \vec{\mathbf{e}}_i\ 0\cdots 0]$, where the standard basis vectors $\vec{\mathbf{e}}_i \in \mathbb{R}^m$ occurs in the j^{th} column of M_{ij}; that is, M_{ij} is an $m \times n$ matrix comprised of zeroes in all entries except the $(i,j)^{th}$, which is a 1. We claim that $\beta = \{M_{ij}\}_{i=1,j=1}^{i=m,j=n}$ is a basis for $\mathcal{M}_{m\times n}(\mathbb{F})$.

Indeed, any matrix $A \in \mathcal{M}_{m\times n}(\mathbb{F})$ given by $(A)_{ij} = a_{ij}$ can be written as

$$A = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}M_{ij}.$$

Hence, β spans $\mathcal{M}_{m\times n}(\mathbb{F})$.

If

$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}M_{ij} = \mathbb{O}$$

for some collection of scalars $c_{ij} \in \mathbb{F}$, then consider that

$$c_{ij} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}M_{ij}\right)_{ab} = (\mathbb{O})_{ab} = 0$$

for all $c_{ij} \in \mathbb{F}$. Hence, β is linearly independent.

Therefore, β is a basis for $\mathcal{M}_{m\times n}(\mathbb{F})$. In total there are mn many of these matrices in β, so $\dim(\mathcal{M}_{m\times n}(\mathbb{F}) = mn$.

(d) In lecture, I mistakenly stated that the set $\mathbb{L}_m(\mathbb{R})$ of $m \times m$ lower unitriangular matrices formed a field. Briefly explain why I was wrong, and offer a way to change $\mathbb{L}_m(\mathbb{R})$ to the field $\mathcal{F}$ of scalars so that we can think of $(\mathbb{M}_{m\times n}(\mathbb{R}), \mathcal{F})$ as a vector space as well. Also comment on how your change might affect the original interpretation of elementary matrices being identified with elementary row operations.

First observe that $\mathbb{L}_m(\mathbb{R})$ doesn't contain the zero matrix $\mathbb{O}$. Therefore, it cannot be a field.

Suppose we add $\mathbb{O}$ to this set and set the union $\mathbb{L}_m(\mathbb{R} \cup \{\mathbb{O}\} = \mathcal{F}$. Then $\mathcal{F}$ would need to be closed under addition. This is not the case, since the sum of any two unitriangular matrices contains twos along its diagonal instead of ones. Hence, we must include in $\mathcal{F}$ the lower triangular matrices L such that $(L)_{ii} = c \in \mathbb{R}$ for all $1 \leq i \leq m$.

We have now included non-trivial lower triangular matrices with zeros along their diagonals. These matrices have no inverses. Therefore, we must remove all elements with non-zero elements below their diagonal. Hence, $\mathcal{F} = \{cI_m\}_{c\in\mathbb{R}}$ can consist only of all scalar multiples of the identity matrix I_m, and $\mathcal{F}$ is now a field.

Given this field $\mathcal{F}$, the only interpretation that is preserved with regard to elementary row operations is to scale the entire system of equations by the same number. Therefore, $\mathcal{F}$ is now useless in studying elementary row operations.

Note: The space $(\mathbb{M}_{m\times n}(\mathbb{R}), \mathbb{L}_m(\mathbb{R}))$ is indeed a legitimate, algebraic object; but, alas, it is not a field [EDIT - we must change the definition of $\mathbb{L}_m(\mathbb{F})$ here to include all lower triangular matrices, not only the lower unitriangular matrices]. It is considered a *left* $\mathbb{L}_m(\mathbb{R})$*-module*, and it is a generalization of vector spaces. Instead of using fields of scalars, modules use *rings* of scalars; but they largely behave the same as vector spaces. Basically, rings ignore the commutativity of multiplication and existence of multiplicative inverses in the set of scalars (they don't even have to have a unit 1 element!). If you're interested in pursuing this further, feel free to talk to me or take MATH 143 in the Fall and MATH 144 sometime thereafter.

Problem 2. *Sample spaces, huh? Tell me more...*

Because continuous functions $f\colon \mathbb{R}\to\mathbb{R}$ have an infinity of inputs, we can only ever approximate them in the computer. This suggests that we should develop a theory about *sample vectors* and how they relate to the functions they approximate. Ideally we would like to characterize all of the important information about a function using a "small" sample vector - that way we can instead use the sample vector to study the function.

Let $S = \{x_1, x_2, \ldots, x_n\}$ be a set of distinct sample points. Let $F = \{f_1, \ldots, f_n\} \subset C(\mathbb{R})$ be a collection of continuous functions on $\mathbb{R}$, and define

$$\vec{\mathbf{f}}_i = \begin{bmatrix} f_i(x_1) \\ f_i(x_2) \\ \vdots \\ f_i(x_n) \end{bmatrix}$$

to be the i^{th} *sample vector on* S for all $1 \le i \le n$.

(a) Show that F is linearly independent in $C(\mathbb{R})$ if the sample vectors $\vec{\mathbf{F}} = \{\vec{\mathbf{f}}_1, \vec{\mathbf{f}}_\mathbf{2}, \ldots, \vec{\mathbf{f}}_n\}$ are linearly independent vectors in $\mathbb{R}^n$.

Suppose indeed that $\vec{\mathbf{F}}$ is linearly independent. Consider if

$$\sum_{j=1}^{n} c_j f_j \equiv 0$$

for some scalars $c_j \in \mathbb{R}$. Then we could evaluate this equation at each of the sample points:

$$\sum_{j=1}^{n} c_j f_j(x_i) = 0$$

Each equation in a linear equation in the variables c_j. This can be represented by the matrix equation

$$A\vec{\mathbf{c}} = [\vec{\mathbf{f}}_1\ \vec{\mathbf{f}}_2\ \cdots\ \vec{\mathbf{f}}_n]\vec{\mathbf{c}} = \vec{\mathbf{0}},$$

where $(\vec{\mathbf{c}})_j = c_j$. Because the columns of the matrix A are linearly independent (they are the sample vectors!), this matrix equation has only the trivial solution. Therefore, each $c_j = 0$. Hence, F is linearly independent as well.

(b) Show that it is not necessarily the case that the converse holds; that is, find a set of linearly independent functions whose sample vectors are linearly dependent.

Approach 1: Notice that $F = \{x, x^2\}$ is linearly independent. If we choose $x_1 = 0$ and $x_2 = 1$, then $\vec{\mathbf{F}} = \{(0,1)^T, (0,1)^T\}$ is linearly dependent.

Approach 2: Notice that if $f(x) = \sin(\pi x) \in F$ and $x_i = \pi i$, then $\vec{\mathbf{f}} = \vec{\mathbf{0}} \in \vec{\mathbf{F}}$ so that $\vec{\mathbf{F}}$ is linearly dependent.

(c) Fix an integer $K \geq 1$, and define

$$F_K = \{1, \sin x, \cos x, \sin(2x), \cos(2x), \ldots, \sin(Kx), \cos(Kx)\}.$$

Show that F_K is a set of linearly independent vectors in $C(\mathbb{R})$.

The idea is to use part (a) to show this statement. The hard part is to select the correct number of sample points and, more importantly, a *general* set of sample points $\{x_i\}_{i=1}^{2K+1}$.

I expected solutions to differ dramatically at this point. I mostly wanted to see how you would approach this problem of choosing your sample points and, perhaps more intriguingly, how you proved that your sample vectors were linearly independent.

It turns out that the proof I had originally thought of fails (quite extraordinarily, in fact). I experimented a bit with MATLAB and chose $2K + 1$ equally spaced sample points on the interval $[0, \pi]$. Initial observations for $K \leq 5000$ show that this set of sample points is almost perfect for the job, but they get hung up on $\sin(Kx)$ because its sample vector is $\vec{\mathbf{0}}$. If we simply increase the number of sample points to $2K + 2$ instead, then $[\vec{\mathbf{f}}_1, \ldots, \vec{\mathbf{f}}_{2K+1}]$ is full-rank and we indeed get the desired result. A little bit of digging turned up a non-trivial result that explains why adding just one more point matters so much:

Theorem 1. (Shannon-Nyquist, 1928) If a signal contains frequencies no higher than K, then the signal can be completely recovered if the sample rate is higher than $2K$.

Shannon's proof of this theorem makes use of the methods of *Fourier Analysis.*

The main take-away here is that although we have the theoretical means to show that a set of functions is linearly independent, it might be still be difficult to show that their associated sample vectors are linearly independent.

Note: We will address this sticking point more efficiently (and with a proper solution!) when we talk about orthogonal sets.

Problem 3. *Every vector space has a basis*

These words are a mantra for most students of theoretical linear algebra, and they favor a rather dramatic claim. We have effectively shown this result for $\mathbb{R}^n$, but we would like to make sure this holds for abstract vector spaces over arbitrary fields. We will show that this is indeed true for the finite-dimensional case and then examine some of the challenges we face in the infinite-dimensional case.

Let $(V, \mathbb{F})$ be an n-dimensional vector space over the field $\mathbb{F}$, and let $(W, \mathbb{F})$ be an infinite-dimensional vector space.

(a) Let $\mathcal{L} \subset V$ be a linearly independent subset of vectors from V. Suppose $\vec{\mathbf{v}} \in V$ doesn't lie in $\mathcal{L}$. Show that $\mathcal{L} \cup \{\vec{\mathbf{v}}\}$ is linearly independent if and only if $\vec{\mathbf{v}}$ is not in $\operatorname{span}(\mathcal{L})$.

Assume that $\mathcal{L} = \{\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_n\}$ is linearly independent. We show the contrapositive statements: "$\mathcal{L} \cup \{\vec{\mathbf{v}}\}$ is linearly dependent if and only if $\vec{\mathbf{v}}$ is not in span($\mathcal{L}$."

Suppose $\mathcal{L} \cup \{\vec{\mathbf{v}}\}$ is linearly dependent. Then

$$c\vec{\mathbf{v}} + \sum_{i=1}^{n} \vec{\mathbf{v}}_i = \vec{\mathbf{0}}$$

for some scalars $c, c_i \in \mathbb{F}$, not all of which are zero. If $c = 0$, the vectors from $\mathcal{L}$ form a non-trivial linear combination of the zero vector. Since this would contradict the fact that $\mathcal{L}$ is linearly independent, we can assume that $c \neq 0$. In this case,

$$v = \sum_{i=1}^{n} -\frac{c_i}{c}\vec{\mathbf{v}}_{\mathbf{i}}$$

so that $\vec{\mathbf{v}} \in \operatorname{span}(\mathcal{L})$.

Now suppose that $\vec{\mathbf{v}} \in \operatorname{span}(\mathcal{L})$. Then

$$\vec{\mathbf{v}} = \sum_{i=1}^{n} d_i \vec{\mathbf{v}}_i$$

so that

$$\vec{\mathbf{v}} + \sum_{i=1}^{n} -d_i \vec{\mathbf{v}}_i = \vec{\mathbf{0}}.$$

is a non-trivial linear combination producing the zero vector. Therefore, $\mathcal{L} \cup \{v\}$ is linearly dependent.

(b) Does the statement in (a) hold for W?

The proof in part (a) shows that this is true whenever $\mathcal{L}$ is *finite*. In the case of $\mathcal{L} \subset W$, it is possible that $\mathcal{L}$ is infinite. If $\mathcal{L}$ is *infinite*, the definition of linear independence for $\mathcal{L}$ is that "any finite collection $L \subset \mathcal{L}$ is linearly independent."

If L doesn't include $\vec{\mathbf{v}}$, there is nothing to show. Then $\vec{\mathbf{v}} \in L$, and the same proof above holds for L. Indeed, this statement is true for infinite-dimensional vector spaces as well.

(c) Let $\mathcal{S} \subset V$. Show that if $\text{span}(\mathcal{S}) = V$, then there are at least n vectors in $\mathcal{S}$.

We show that the negation is false: "$\mathcal{S}$ spans V and $\mathcal{S}$ has fewer than n vectors."

Indeed, assume that $\text{span}(\mathcal{S}) = V$. Suppose also that $\mathcal{S} = \{\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_k\}$, where $0 \leq k \leq n-1$. Because V is n-dimensional, $\mathcal{S}$ cannot be a basis for V. Specifically, any basis for V spans V. This produces a contradiction: $\mathcal{S}$ spans V, and does not span V. This contradiction shows that the negation is indeed false. We conclude that the original statement must be true.

(d) Does the statement in (c) hold for W?

The statement for W is "if $\text{span}(\mathcal{S}) = W$, then $\mathcal{S}$ must be infinite." Recall that $\text{span}(\mathcal{S}) = W$ means that "every vector of W can be written as a *finite* linear combination of a *finite* subcollection of vectors in $\mathcal{S}$." Similarly to part (c), we can contradict the negation: "$\text{span}(\mathcal{S}) = W$ and $\mathcal{S}$ is finite."

Indeed, assume that $\text{span}(\mathcal{S} = W$. Suppose also that $\mathcal{S}$ is finite. We show below in part (g) that $\mathcal{S}$ can be reduced to be a basis for W. This cannot be since W is infinite-dimensional and, by definition, has no basis. This contradiction shows that the negation is indeed false. We conclude that the original statement must be true.

(e) Show that any linearly independent set $\mathcal{L}$ of vectors in V can be extended to a basis of V; that is, if $\mathcal{L}$ isn't already a basis for V, we can add vectors to it so that it is.

Suppose that $\mathcal{L}$ is linearly independent but does not span V. Then there is some vector $\vec{\mathbf{v}} \in V$ that is not a linear combination of $\mathcal{S}$. By part (a), we see that $\mathcal{L}' = \mathcal{L} \cup \{\vec{\mathbf{v}}\}$ is also linearly independent. If $\mathcal{L}'$ doesn't span V either, then we continue this process.

Because V has finite dimension, this process must eventually terminate once $\mathcal{L}'$ consists of n vectors.

(f) Does your argument in (e) work for W?

No. The process of extending $\mathcal{L}$ to $\mathcal{L}'$ is not guaranteed to terminate. A counterexample for $W = \mathcal{P}(\mathbb{R})$, the vector space of formal power series with real coefficients, begins with $\mathcal{L} = \{1\}$. We can add x, and then x^2, and so on. Each time, $\mathcal{L}'$ is still linearly independent, but this process continues without end.

(g) Show that any spanning set $\mathcal{S}$ of V can be reduced to a basis of V; that is, if $\mathcal{S}$ isn't a basis for V, we can remove vectors from it until it is.

If $\mathcal{S}$ is a *finite* spanning set, then we proceed by induction on the size $k \geq 1$ of $\mathcal{S}$.

For $k = 1$, $\mathcal{S} = \{\vec{\mathbf{v}}\}$ for some non-zero $\vec{\mathbf{v}} \in V$. If $\mathcal{S}$ is a basis, then we are done. Suppose that $\mathcal{S}$ spans V but is not a basis for V. Then $\mathcal{S}$ is linearly dependent so that $c\vec{\mathbf{v}} = \vec{\mathbf{0}}$ for some non-zero scalar $c \in \mathbb{F}$. Therefore, $\vec{\mathbf{v}} = \vec{\mathbf{0}}$

so that V is the zero vector space. Removing $\vec{\mathbf{v}} = \vec{\mathbf{0}}$ from $\mathcal{S}$ produces the empty set $\mathcal{S}' = \emptyset$, which indeed is a basis for V.

Suppose now that the statement holds for any $\mathcal{S}$ of size k. If $\mathcal{S} = \{\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_{k+1}\}$ spans V but is not a basis already, then $\mathcal{S}$ is linearly dependent. Therefore,

$$\sum_{i=1}^{k+1} c_i \vec{\mathbf{v}}_i = \vec{\mathbf{0}}$$

for some scalars $c_i \in \mathbb{F}$, not all of which are zero. Possibly by renumbering the vectors in $\mathcal{S}$, suppose that $c_{k+1} \neq 0$. Then we can write

$$v_{k+1} = \sum_{i=1}^{k} -\frac{c_i}{c_{k+1}} \vec{\mathbf{v}}_i$$

so that $v_{k+1} \in \operatorname{span}\{\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_k\}$. Moreover, $\mathcal{S}' = \{\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_k\}$ spans V as well. Therefore, by the induction hypothesis, $\mathcal{S}'$ can be reduced to a basis for V. This completes the proof when $\mathcal{S}$ is finite.

If $\mathcal{S}$ is infinite but still spans V, then $\mathcal{S}$ is linearly dependent because V is finite-dimensional. In particular, part (c) states that $\mathcal{S}$ must have at least n vectors in it. Choose n vectors from $\mathcal{S}$ and create the set $\mathcal{S}' = \{\vec{\mathbf{v}}_1, \ldots, \vec{\mathbf{v}}_n\}$. If $\mathcal{S}'$ is linearly independent, we are done. If not, then the induction argument for the finite case shows that we can remove vectors from $\mathcal{S}'$ until the resulting subset $\mathcal{S}''$ is linearly independent. We can then apply part (e) to create a basis of V from $\mathcal{S}''$. This completes the proof for the case when $\mathcal{S}$ is infinite.

(h) Does your argument in (g) work for W?

No. The set $\mathcal{S}'$ is infinite by part (c). The process that creates $\mathcal{S}''$ is not guaranteed to terminate in this case. Therefore, there is no guarantee to produce a *finite* linearly independent set $\mathcal{S}''$ to which we can apply part (e).

(i) Use your observations above to explain some things that make it difficult to show that every *infinite-dimensional* vector space has a basis.

It appears that it is not hard to add vectors to a linearly independent set and keep the resulting set linearly independent; so making sets bigger is not an issue. It also appears that is not hard to show that spanning sets need to have a certain minimal number of elements in them; so putting an infinite number of vectors into a spanning set is not an issue.

The problems seem to occur when we try to produce an infinite set from a finite set or to produce a finite set from an infinite set. We cannot guarantee that adding vectors to a finite set one at a time will produce an infinite set; and we cannot guarantee that removing vectors from an infinite set will produce a finite set. The processes that work for the finite-dimensional cases break down when we *transcend* the finite into the infinite.

Note: Brute force is one solution to these issues - make the things you can't do into axioms! The axiom of most use to us in this instance is *Zorn's Lemma*:

> Suppose a partially ordered set P has the property that every chain in P has an upper bound in P. Then P contains at least one maximal element.

In this context, $P = W$ and the partial ordering is $\subseteq$. We can conclude that the process of adding things to the linearly independent set $\mathcal{L}$ must have a maximal linearly independent upper bound; that is, a basis for W.

Other note: This may well have been the most difficult problem that I have ever assigned. Problems on exams will not be this involved. I include them on homework because you have the time to think properly about them. Not to mention this might be the only time you get a chance to learn about something so deep and remarkable.