Summary of definitions about bases and coordinates

Before starting the problem set, I'll summarize some definitions here that we have introduced in class or in the previous problem set.

Suppose V is is a vector space over a field k. A *basis* B for V is a set of vectors in V which is linearly independent and which spans V. If V has a finite basis, it is called *finite-dimensional*; in that case, we will typically write something like

$$B = \{\vec{e}_1, \dots \vec{e}_n\} = \{\vec{e}_i\}_{i=1}^n.$$

If there does not exist a finite basis for V, then V is called *infinite-dimensional*, and in that case we will write something like

$$B = \{\vec{e}_\alpha\}_{\alpha \in I},$$

where I is called the *index set*. (A finite basis is a special case of this notation, if we set $I = \{1, \dots, n\}$.) Recall that a *linear combination* of a set of vectors $\{\vec{e}_\alpha\}_{\alpha \in I}$ is a *finite* sum

$$x_{\alpha_1}\vec{e}_{\alpha_1} + \cdots + x_{\alpha_m}\vec{e}_{\alpha_m} = \sum_{j=1}^{m} x_{\alpha_j}\vec{e}_{\alpha_j},$$

where $\alpha_1, \dots, \alpha_m \in I$. The *span* of a set B of vectors is the set of all possible linear combinations of vectors from that set; we say B *spans* V is the span of B equals the whole space V. The set of vectors B is called *linearly dependent* if there exists some linear combination of vectors in B equal to the zero vector, with not all the coefficients being zero. The set B is called *linearly independent* if it is not linearly dependent. We will prove that, if B is a basis for V, then every vector $\vec{v} \in V$ can be written uniquely (in only one way) as a linear combination of elements of B. If B is finite, this looks like

$$\vec{v} = x_1\vec{e}_1 + \cdots + x_n\vec{e}_n$$

where the $x_1, \dots, x_n$ are elements of k (scalars) that are uniquely determined by the choice of $\vec{v}$. In this case, we call x_i the *coordinate* of $\vec{v}$ corresponding to the basis vector $\vec{e}_i$. We write the coordinates, by definition, as a column matrix:

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_B := x_1\vec{e}_1 + \cdots + x_n\vec{e}_n.$$

Note that as a special case,

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_B = 1\vec{e}_1 + 0\vec{e}_2 + \cdots + 0\vec{e}_n = \vec{e}_1,$$

and similarly for the other basis vectors.

(We will discuss what happens for infinite-dimensional spaces later.)

Given two vector spaces U and V over the same field k, a *linear function*, or *linear map*, or *linear transformation* (all synonymous) $A : U \to V$ is a function from U to V as sets, with the additional conditions that

(i) For any vectors $\vec{u}, \vec{u}' \in U$, we have $A(\vec{u} + \vec{u}') = A\vec{u} + A\vec{u}'$; and
(ii) for any vector $\vec{u} \in U$ and any scalar $a \in k$, we have $A(k\vec{u}) = kA\vec{u}$.

The matrix of a linear transformation

You probably remember that any linear transformation $A : \mathbb{R}^m \to \mathbb{R}^n$ can be written as multiplication by an $n \times m$ matrix. In these problems, I want to work through why this is true, and how it works in any arbitrary finite-dimensional vector spaces (with the matrices you remember as a special case). (I will come back to infinite-dimensional vector spaces later.) I also want to introduce the "dual space", and to show how the space of all linear transformations can be spanned by a basis made up of "tensor products" of vectors and dual vectors (I'll explain all those terms below!). This formalism will be very valuable in writing projections (which we'll do very soon), and also in writing spectral decompositions, singular value decompositions, and other tools which come up in data compression, page rank, principal components, and other applications. I also want to introduce some new notations, such as "Einstein summation notation", which will be helpful when we get to tensors.

1. (Core) (Coordinate representation of a linear transformation) In this problem, I want to run through how any linear transformation between finite-dimensional vector spaces can be written in a coordinate system, that is, relative to some bases. I'm going to have you write this in several different ways, each meaning something slightly different conceptually.

   Let's assume that U and V are vector spaces. Let's also assume that U has a basis $E = \{\vec{e}_1, \vec{e}_2\}$, and that V has a basis $F = \{\vec{f}_1, \vec{f}_2, \vec{f}_3\}$. Suppose $A : U \to V$ is a linear transformation. (The linear transformation $A : \mathbb{R}^2 \to \mathbb{R}^3$ that occurred in the projection problem, PS 3 #1, is a special case of this.) (Everything we do here will work equally well if U has a basis $E = \{\vec{e}_1, \ldots, \vec{e}_m\}$ and V has a basis $F = \{\vec{f}_1, \ldots, \vec{f}_n\}$; I am picking two specific numbers to make the notation easier.)
   (a) Suppose $\vec{u}$ is any vector in U. Show that $A\vec{u}$ can be written as a linear combination of the vectors $A\vec{e}_1$ and $A\vec{e}_2$.
   (b) Now, the vector $A\vec{e}_1$ can be written as a linear combination of the vectors $\{\vec{f}_1, \vec{f}_2, \vec{f}_3\}$. (Why?) Write that out, and do the same for $A\vec{e}_2$. What is a good way of labelling these coefficients, that will let you keep track, and that will work in any dimension?
   (c) How many number (i. e. elements of k) does it take to specify (i. e. completely know) a linear transformation $A : U \to V$ (with the spaces U and V having bases as we said above)? What would be the answer if U and V had bases with m and n elements, respectively?
   (d) Now, substitute your results of # 1b into your answer for # 1a. Collect all the terms in each $\vec{f}_i$, to obtain $A\vec{u}$ as a linear combination of the vectors $\{\vec{f}_1, \vec{f}_2, \vec{f}_3\}$.
   (e) Note that in # 1d, you have, by definition, determined the coordinates of $A\vec{u}$, with respect to the F basis of V, in terms of the coordinates of $\vec{u}$ in the E basis, and the parameters for A you identified in # 1b. This means that we could symbolically write

$$[A\vec{u}]_F = \begin{bmatrix} (A\vec{u})_1 \\ (A\vec{u})_2 \\ (A\vec{u})_3 \end{bmatrix}_F = [A]_{E,F} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}_E,$$

where $\left[A\right]_{E,F}$ is a matrix (i. e. grid with two indices) of coefficients that you found for A in # 1b. Note that I have put a subscript E, F, because these coefficients were calculated with reference to the bases E and F of U and V respectively. Use your answer to # 1d to define what matrix multiplication should be. Does it agree with what you remember for matrix multiplication? (Look it up if you don't recall.) If it doesn't agree, it's probably because you made a different convention about rows and columns. If that is the case, go back and adjust your notation until your matrix multiplication agrees with the standard one.

(f) Based on your work in the previous part, convince yourself that if the basis E had had m elements, and if F had had n elements, then the matrix for A in E, F would have been $n \times m$, that is, n rows and m columns.

(g) There are two ways to look at this matrix multiplication; each is very important in different situations.

(i) First way is to look at the matrix as made up of columns. This is what you worked out in # 1a: giving a basis $\{\vec{e}_1, \vec{e}_2\}$ for U means that every output vector $\vec{v} = A\vec{u}$ is a linear combination of $A\vec{e}_1$ and $A\vec{e}_2$. The coefficients are the coordinates of $\vec{u}$ in the E basis, that is, u_1 and u_2, where $\vec{u} = u_1\vec{e}_1 + u_2\vec{e}_2$. Write this out in matrix form. Check (at least mentally) that this new way of writing the matrix multiplication agrees with what you wrote above.

(ii) Second way is to look at the matrix as made up of rows. This is what you worked out in # 1b: if the output $\vec{v} = A\vec{u}$ has coordinates with respect to the F basis given by $\vec{v} = v_1\vec{f}_1 + v_2\vec{f}_2 + v_3\vec{f}_3$, then each v_i is a function of u_1 and u_2, and that function can be written as a row vector times the column vector $\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}_E$. Write this out in matrix form, and check that it agrees with what you've already established about matrix multiplication. (It's not quite as easy to write the row interpretation in a nice way as it is for the column interpretation; I'll return to this in a later problem.)

2. (Core/Theory Everyone should do the first part; theory people may want to do the proofs). Suppose that U and V are any vector spaces over a common field k. Let $T(U, V)$ be the set of *all* linear transformations from U to V:

$$T(U, V) := \{A : U \to V \Big| A \text{ is linear}\}.$$

- Define a natural addition and scalar multiplication operation on $T(U, V)$. (Note that I am NOT assuming that U and V are finite-dimensional, so I am NOT assuming that linear transformations are given by matrices.)
- Prove that $T(U, V)$ forms a vector space over k.
- If U and V are finite-dimensional, with bases E and F, prove that the addition and scalar multiplication you defined agree with addition and scalar multiplication of matrices.

3. (Core/Theory Everyone should do answer the problems, but can skip formal proofs if wanted; theory people should do the proofs) (Dual spaces and coordinates) Suppose that U is a vector space over a field k. We define the *dual space* to U, written U^*, as the set of all linear transformation from U to k. In the

notation of the previous problem, $U^* = T(U,k)$, where we are viewing k as a vector space over k.

(a) Note that the previous problem proves that U^* is itself a vector space. If you did not do the previous problem, prove this now!

(b) Suppose that U has a basis $E = \{\vec{e}_1, \ldots, \vec{e}_m\}$ (so U is assumed finite-dimensional). We define the *dual basis to* E for U to be the set of dual vectors $E^* = \{\overleftarrow{\phi}_1, \ldots, \overleftarrow{\phi}_n\}$, where, for each i, $\overleftarrow{phi}_i : U \to k$ is defined by the condition

$$\overleftarrow{\phi}_i(\vec{e}_j) = \begin{cases} 1 & : i = j \\ 0 & : i \neq j \end{cases}$$

and by the assumption that $\overleftarrow{\phi}_i$ is linear. Prove that, if E is a finite basis for U, then, true to its name, the dual basis E^* is, in fact, a basis for the dual vector space U^*. (Note that the backwards arrow for a dual vector is not a standard notation.)

(c) Show that, if U has a finite basis $E = \{\vec{e}_1, \ldots, \vec{e}_m\}$, and if $\xi \in U*$, then the matrix for ξ, $\left[\xi\right]_E$ has only one row (called a *row vector*). (Why was I able to have only one basis E in the subscript?)

(d) Assuming U has a finite basis $E = \{\vec{e}_1, \ldots, \vec{e}_m\}$, what are the matrices for the dual basis vectors $\overleftarrow{\phi}_i$?

4. (Core)(Basis for $T(U,V)$ and tensor products) Suppose that U and V vector spaces over a common field k. We have seen that the set $T(U,V)$ of all linear transformations from U to V is a vector space. If U and V are finite-dimensional vector spaces, with bases $E = \{\vec{e}_1, \ldots, \vec{e}_m\}$ and $F = \{\vec{f}_1, \ldots, \vec{f}_n\}$ respectively, I would like to give a corresponding basis for the vector space $T(U,V)$. This will give ways of decomposing linear transformations (or matrices) that will be useful in many applications.

This depends on a definition of a "tensor product". (The definition I give here is a special case of a more general definition we'll see later.) Suppose that $\vec{v}$ is any vector in V, and that $\overleftarrow{\phi}$ is any dual vector in U^*. We define the *tensor product of* $\vec{v}$ *and* $\overleftarrow{\phi}$, written $\vec{v} \otimes \overleftarrow{\phi}$, to be the linear function $\vec{v} \otimes \overleftarrow{\phi} : U \to V$ given by, for any $\vec{u} \in U$,

$$\vec{v} \otimes \overleftarrow{\phi}(\vec{u}) := \phi(\vec{u})\vec{v}.$$

That is, $\vec{v} \otimes \overleftarrow{\phi}$ takes as input vector $\vec{u}$ in U; you apply $\overleftarrow{\phi}$ to it to get a scalar; then you multiply that scalar by $\vec{v}$.

(a) Suppose that U and V are finite-dimensional, with bases E and F as written above, and suppose that $A : U \to V$ is a linear transformation. To keep things simple, let's assume that $m = 2$ and $n = 3$, so that $E = \{\vec{e}_1, \vec{e}_2\}$ and $F = \{\vec{f}_1, \vec{f}_2, \vec{f}_3\}$. Let $E^* = \{\overleftarrow{\phi}_1, \overleftarrow{\phi}_2\}$ be the dual basis to E. Suppose $\vec{u} \in U$ is any vector in U. Referring back to your answers for question # 1, write $A\vec{u}$ as a linear combination of the $\vec{f}_j \otimes \overleftarrow{\phi}_i$, where $i = 1, 2$ and $j = 1, 2, 3$. (Your answer should come out to look pretty simple.)

(b) Use your answer to the previous part, to write the linear transformation A as a linear combination of the $\vec{f}_j \otimes \overleftarrow{\phi}_i$, where $i = 1, 2$ and $j = 1, 2, 3$.

(c) Based on this, prove that the set

$$\{\vec{f}_j \otimes \overleftarrow{\phi}_i \big| i = 1, 2 \; j = 1, 2, 3\}$$

forms a basis for the space of linear transformations $T(U, V)$.

(d) Write the matrices of the $\vec{f}_j \otimes \overleftarrow{\phi}_i$, where $i = 1, 2$ and $j = 1, 2, 3$, in the E, F basis (as defined in # 1). Make sure your answers make sense with both interpretations of matrix multiplication. (Your answers should come out to look pretty simple.)

(e) According to the interpretation of matrices of linear transformations, a 3×1 matrix should represent a linear transformation from a 1-dimensional space to a 3-dimensional space. On the other hand, separately from this, we have been writing vector coordinates in a basis as 3×1 matrices as well. Can you find a way of thinking about things which makes the two notations coincide? And does this way of thinking about things then make the matrices for the $\vec{f}_j \otimes \overleftarrow{\phi}_i$ make sense as matrix multiplications?

5. (Core) (Example: spaces of polynomials) Here, I want to give a practical example of vector spaces which are not just the standard $\mathbb{R}^n$, and where there isn't one best basis to work in. For this problem, let V be the space of real polynomials of degree less than or equal to 4, that is,

$$V = \mathbb{R}_4[x] := \{a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4\}.$$

Let $\frac{\mathrm{d}}{\mathrm{d}x}$ be the operation of taking the derivative of a polynomial, considered as a function from V to V.

(a) Check that $\frac{\mathrm{d}}{\mathrm{d}x}$ is a linear transformation. (This requires you to remember a couple of rules from calculus; you don't need to prove them.)

(b) Let $B = \{f_0, f_1, f_2, f_3\}$, where

$$\begin{aligned} f_0(x) &= 1 \\ f_1(x) &= x \\ f_2(x) &= x^2 \\ f_3(x) &= x^3 \\ f_4(x) &= x^4. \end{aligned}$$

I claim that B is a basis of V. Write the matrix of the operator $\frac{\mathrm{d}}{\mathrm{d}x}$ in this basis.

(c) Any guesses what the matrix for $\frac{\mathrm{d}}{\mathrm{d}x}$ would look like on the infinite-dimensional space $\mathbb{R}[x]$ of all real polynomials (if we extend basis B in the expected way)?

(d) Let $L = \{P_0, P_1, P_2, P_3\}$, where

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_2(x) &= \frac{1}{2}(3x^2 - 1) \\ P_3(x) &= \frac{1}{2}(5x^3 - 3x) \\ P_4(x) &= \frac{1}{8}(35x^4 - 30x^2 + 3). \end{aligned}$$

These are the first five *Legendre polynomials* which I mentioned earlier. I claim that L is a basis of V. Give the matrix for the operator $\frac{\mathrm{d}}{\mathrm{d}x}$ in the basis L. (Keep in mind that you must write both the input AND the output in basis L!)

(e) Any guesses what the matrix for $\frac{\mathrm{d}}{\mathrm{d}x}$ would look like on the infinite-dimensional space $\mathbb{R}[x]$ of all real polynomials in the basis of Legendre polynomials? (Note that I haven't told you what the other Legendre polynomials are; if the pattern holds, could you use this to find them?)

6. (Optional, but interesting and good practice with bases) (Logic as linear algebra) This question is for those who took the Logic and Proofs class at some point. Suppose that we are looking at logical operations, like OR="$\vee$" or AND="$\wedge$". If we look at them as truth tables, then we have two inputs (say p and q), and one output (for example, $p \vee q$ or $p \wedge q$). Let $\mathbb{GF}(2)$ be the field with two elements $\{0, 1\}$ and modulo 2 addition and multiplication. Every possible set of values for p and q is an ordered pair of such numbers, that is, an element of the vector space $\mathbb{GF}(2)^2$. Then every possible truth table is a logical function of two arguments; that is, it is a function $f : \mathbb{GF}(2)^2 \to \mathbb{GF}(2)$.
   (a) I have earlier claimed that, given any set S and any field k, the set of all functions $f : S \to k$ is a vector space over k. You might want to prove this now, or at least convince yourself it is true. This applies here, to show that the set of all logical functions $f : \mathbb{GF}(2)^2 \to \mathbb{GF}(2)$ is a vector space. Let's call it V.
   (b) How many elements are there in V? (That is, how many possible truth tables exist?)
   (c) List all possible truth tables (elements of V). Try to give each one the simplest logical name you can (you can use symbols $\neg$, $\vee$, $\wedge$, $\to$, or any others you happen to know).
   (d) Find a basis for the space V. How many elements does your basis have?
   (e) How would you interpret the operation of addition in V logically?
   (f) Write the other elements of V as linear combinations of your basis. See if you can interpret at least a few of them logically. (It might be boring to write them all.)
   (g) Now, I'm betting that the basis you chose probably wasn't the most logically natural one. Can you find a basis which seems natural, thinking of the functions as logical functions? (There isn't just one good answer to this question.)
   (h) Try writing a few of the other elements of V in terms of this alternate basis, and interpreting logically.
   (i) The elements of V are all functions $f : \mathbb{GF}(2)^2 \to \mathbb{GF}(2)$. Are these all linear functions? If not, which ones are linear functions? The linear functions should be expressible as matrices in your basis: do so. (If you're feeling ambitious, you could do this both in the first simple basis you found, and in the other basis.)