Math 313 Final KEY

Fall 2012

sections 008 and 011

Instructor: Scott Glasgow

Serious Instructions: Write your name very clearly on this exam. In this booklet, write your mathematics clearly, legibly, in big fonts, and, most important, "have a point", i.e. make your work logically and even pedagogically acceptable. (Other human beings not already understanding 313 should be able to learn from your exam.) To avoid excessive erasing, first put your ideas together on scratch paper, then commit the logically acceptable fraction of your scratchings to this exam booklet. More is not necessarily better: say what you mean and mean what you say.

Instructions for those who want their psychology to be optimal for an assessment[1]: a) you should communicate in complete sentences, 2) you should write on your own paper and d) you should be neat as possible.

NOTE: Almost none of the problems below are worth 25 points. That's funny.

---

[1] In case of an overview of this document by an administrator, my students have learned of research indicating that humorous instruction may increase capacity on exams. This claim is similar to the following: "Three grams of soluble fiber daily from whole grain oat foods, like Honey Nut Cheerios, in a diet low in saturated fat and cholesterol, may reduce the risk of heart disease." So there.

1. Put the following matrix in reduced row-echelon form via elementary row operations:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 0 & 2 & 0 \\ 6 & -2 & 3 & 4 \end{bmatrix} \tag{1}$$

20pts

Solution

The row reduction might proceed as follows:

$$\begin{aligned}
&\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 0 & 2 & 0 \\ 6 & -2 & 3 & 4 \end{bmatrix}^{\substack{R2-2R1 \\ R3-6R1}} \sim \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -2 & 0 & -2 \\ 0 & -8 & -3 & -2 \end{bmatrix}^{\substack{R2/-2 \\ R3-4R2}} \sim \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -3 & 6 \end{bmatrix}^{R3/-3} \\
&\sim \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix}^{R1-R3} \sim \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix}^{R1-R2} \sim \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix}
\end{aligned} \tag{2}$$

However the row reduction proceeds, the row echelon form is unique—the last matrix indicated in (2) is ***the*** answer.

2. Give an expression for the general solution to the system of equations

$$\begin{aligned} w + x + y + z &= 0 \\ 2w \quad + 2y \quad &= 0 \\ 6w - 2x + 3y + 4z &= 0. \end{aligned} \tag{3}$$

   For the final 10 points of this problem, insert your expression for the solution into (3) to make sure that it works. That is, check your answer by plugging it back into (3).

10+10pts

Solution

The augmented matrix, less the augmentation, is the matrix in (1). Thus, according to the last matrix in (2), (3) is equivalent to

$$\begin{aligned} w \qquad\quad &+2z=0 \\ x \quad\; &+z=0 \\ y &-2z=0, \end{aligned} \tag{4}$$

whose general solution is clearly

$$\begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -2t \\ -t \\ 2t \\ t \end{bmatrix} = t\begin{bmatrix} -2 \\ -1 \\ 2 \\ 1 \end{bmatrix} =: t\mathbf{x}_0. \tag{5}$$

Since (3) is homogeneous, in order to check that (5) is an expression of at least some of its solutions, it is enough to there plug in

$$\mathbf{x}_0 := \begin{bmatrix} -2 \\ -1 \\ 2 \\ 1 \end{bmatrix} =: \begin{bmatrix} w_0 \\ x_0 \\ y_0 \\ z_0 \end{bmatrix}. \tag{6}$$

We find

$$\begin{bmatrix}1 & 1 & 1 & 1\\ 2 & 0 & 2 & 0\\ 6 & -2 & 3 & 4\end{bmatrix}\mathbf{x}_0 = \begin{bmatrix}1 & 1 & 1 & 1\\ 2 & 0 & 2 & 0\\ 6 & -2 & 3 & 4\end{bmatrix}\begin{bmatrix}w_0\\ x_0\\ y_0\\ z_0\end{bmatrix}$$

$$= \begin{bmatrix}w_0 + x_0 + y_0 + z_0\\ 2w_0 + 2y_0\\ 6w_0 - 2x_0 + 3y_0 + 4z_0\end{bmatrix} \tag{7}$$

$$= \begin{bmatrix}-2-1+2+1\\ 2(-2)+2(2)\\ 6(-2)-2(-1)+3(2)+4(1)\end{bmatrix}$$

$$= \begin{bmatrix}-3+3\\ -4+4\\ -12+2+6+4\end{bmatrix} = \begin{bmatrix}0\\ 0\\ -10+10\end{bmatrix} = \begin{bmatrix}0\\ 0\\ 0\end{bmatrix}.$$

It works. This does not of course check that there aren't more solutions.

3. Give an expression for the general solution to the system of equations

$$\begin{aligned} x+y+z &= 1\\ 2x+2z &= 0\\ 6x-2y+3z &= 4. \end{aligned} \tag{8}$$

For the final 10 points of this problem, insert your expression for the solution into (8) to make sure that it works. That is, check your answer by plugging it back into(8).

<u>10+10pts</u>

<u>Solution</u>

The augmented matrix for (8) is the matrix in (1), hence by the last matrix in (2) the unique solution of (8) is

$$\mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}. \tag{9}$$

We check that

$$\begin{aligned} x+y+z &= 2+1+(-2) = 3-2 = 1 \\ 2x+2z &= 2(2)+2(-2) = 4-4 = 0 \\ 6x-2y+3z &= 6(2)-2(1)+3(-2) = 12-2-6 = 10-6 = 4. \end{aligned} \tag{10}$$

It works.

4. Let A be a square matrix, and consider the linear system

$$A\mathbf{x} = \mathbf{b}. \tag{11}$$

The usual theorem is that if A is invertible—which means ‘left and right invertible’—then (11) has exactly one solution, namely

$$\mathbf{x} = A^{-1}\mathbf{b}\ . \tag{12}$$

a) Imagine a universe in which A has a left inverse B, but is not guaranteed to have a right one. So we have

$$BA = I \tag{13}$$

but we are not at all guaranteed that there is a matrix C (be it B or any other matrix) such that $AC = I$. Now prove one of the following, whichever one is possible: (It is not possible to prove both with only (13).)

i) Prove that (11) has at least one solution $\mathbf{x}$, namely $\mathbf{x} = B\mathbf{b}$.
ii) Prove that (11) has at most one solution $\mathbf{x}$, namely $\mathbf{x} = B\mathbf{b}$.

b) Imagine a universe in which A has a right inverse C, but is not guaranteed to have a left one. So we have

$$AC = I \tag{14}$$

but we are not at all guaranteed that there is a matrix B (be it C or any other matrix) such that $BA = I$. Now prove one of the following, whichever one is possible: (It is not possible to prove both with only (14).)

i) Prove that (11) has at least one solution $\mathbf{x}$, namely $\mathbf{x} = C\mathbf{b}$.
ii) Prove that (11) hat at most one solution $\mathbf{x}$, namely $\mathbf{x} = C\mathbf{b}$.

20 points

Solution

a) **If** the system has a solution $\mathbf{x}$, then, for any such $\mathbf{x}$,we may write

$$A\mathbf{x} = \mathbf{b} \tag{15}$$

(without implicitly lying), and then, by left application of B to (15), as well as by the associative property of matrix multiplication, obtain that

$$B\mathbf{b} = B(A\mathbf{x}) = (BA)\mathbf{x} = I\mathbf{x} = \mathbf{x}, \tag{16}$$

i.e., (15) and (11) imply

$$\mathbf{x} = B\mathbf{b}. \tag{17}$$

(In (16) we also used that B is a left inverse of A, i.e., we used (13), as well as the fact that the so-called identity matrix I is in fact a "multiplicative identity".) Here then we have just showed that **if** (15)/(11) has a solution, it's got to be $\mathbf{x} = B\mathbf{b}$. Thus we have showed that (15)/(11) has at most one solution, namely $\mathbf{x} = B\mathbf{b}$. This is the second option ii) above. We have not at all proved that (17) actually solves (15). The latter is impossible to show without knowing B is a right inverse of A.

b) With right inverse C satisfying (14) we easily show that (15)/(11) has at least one solution, namely $\mathbf{x} = C\mathbf{b}$: we simply note that with $\mathbf{x} = C\mathbf{b}$ we certainly get

$$A\mathbf{x} = A(C\mathbf{b}) = (AC)\mathbf{b} = I\mathbf{b} = \mathbf{b}. \tag{18}$$

This is the first option i) above. This does not at all prove that there aren't other solutions. The latter is impossible to show without knowing that C is a left inverse of A.

5. Suppose $S := \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis for a vector space V. Show that

$$S' := \{\mathbf{v}_1', \mathbf{v}_2', \mathbf{v}_3'\} := \{\mathbf{v}_1 + 2\mathbf{v}_2 + 3\mathbf{v}_3, 4\mathbf{v}_1 + 5\mathbf{v}_2 + 6\mathbf{v}_3, 7\mathbf{v}_1 + 8\mathbf{v}_2 + 10\mathbf{v}_3\} \tag{19}$$

is also a basis for V. Tools: After **sufficient** explanation of any possible relevance, you may use **either** that the reduced row-echelon form of

$$A := \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{bmatrix} \tag{20}$$

is the identity

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{21}$$

or that the reduced row-echelon form of

$$B := \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix} \tag{22}$$

is the identity (as in (21)). One of these facts is directly relevant, the other irrelevant (or only relevant very indirectly). You will lose points for citing the not-directly-relevant fact.

20pts

Solution

Since S is a basis for V, and since $|S| := |\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}| = 3$ [2], V is 3 dimensional. S' is a set of three vectors in V (using $S := \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ spans V), so S' is a basis provided it is linearly independent or spans V, the former easier to prove: Consider whether the equation

$$k_1' \cdot \mathbf{v}_1' + k_2' \cdot \mathbf{v}_2' + k_3' \cdot \mathbf{v}_3' = \mathbf{z} \tag{23}$$

has only the trivial solution $(k_1', k_2', k_3') = (0,0,0)$ (which solution it certainly has). From (19) we see (23) is equivalent to both

$$k_1' \cdot (\mathbf{v}_1 + 2\mathbf{v}_2 + 3\mathbf{v}_3) + k_2' \cdot (4\mathbf{v}_1 + 5\mathbf{v}_2 + 6\mathbf{v}_3) + k_3' \cdot (7\mathbf{v}_1 + 8\mathbf{v}_2 + 10\mathbf{v}_3) = \mathbf{z}, \tag{24}$$

[2] Since the $\mathbf{v}$'s are independent, there are no copies in the set, and the cardinality is actually 3, not less.

and then, via vector space algebra (made possible by the vector space axioms),

$$\left(k_1' + 4k_2' + 7k_3'\right)\cdot \mathbf{v}_1 + \left(2k_1' + 5k_2' + 8k_3'\right)\cdot \mathbf{v}_2 + \left(3k_1' + 6k_2' + 10k_3'\right)\cdot \mathbf{v}_3 = \mathbf{z}, \tag{25}$$

which is

$$k_1 \cdot \mathbf{v}_1 + k_2 \cdot \mathbf{v}_2 + k_3 \cdot \mathbf{v}_3 = \mathbf{z} \tag{26}$$

where

$$k_1 := k_1' + 4k_2' + 7k_3', \qquad k_2 := 2k_1' + 5k_2' + 8k_3', \qquad k_3 := 3k_1' + 6k_2' + 10k_3'. \tag{27}$$

But since $S := \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is a basis, (26) implies (and is in fact equivalent to) $(k_1, k_2, k_3) = (0,0,0)$, which, with (27) is the system of equations

$$\begin{aligned} k_1' + 4k_2' + 7k_3' &= 0, \\ 2k_1' + 5k_2' + 8k_3' &= 0, \\ 3k_1' + 6k_2' + 10k_3' &= 0 \end{aligned} \tag{28}$$

$$\Leftrightarrow$$

$$\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 10 \end{bmatrix} \begin{bmatrix} k_1' \\ k_2' \\ k_3' \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

which, by(20)/(21)—***NOT*** (22)/(21)—does indeed have exactly the trivial solution.

6. Let $S_W = \{\mathbf{w}_1, \cdots, \mathbf{w}_r\}$ and $S_{W^\perp} = \{\mathbf{w}_{r+1}, \cdots, \mathbf{w}_n\}$ be bases for nonzero subspace W and nonzero orthogonal compliment subspace $W^\perp$ of finite dimensional (dimension n) inner product space V. So what we've effectively said is that

$$\begin{aligned} W &= span(S_W) = span\{\mathbf{w}_1, \cdots, \mathbf{w}_r\} \subset V, \\ W^\perp &= span(S_{W^\perp}) = span\{\mathbf{w}_{r+1}, \cdots, \mathbf{w}_n\} \subset V, \end{aligned} \tag{29}$$

and that

$$k_1\mathbf{w}_1 + \cdots + k_r\mathbf{w}_r = \mathbf{z} \Rightarrow k_1, \ldots, k_r = 0,$$
$$k_{r+1}\mathbf{w}_{r+1} + \cdots + k_n\mathbf{w}_n = \mathbf{z} \Rightarrow k_{r+1}, \ldots, k_n = 0, \tag{30}$$

and finally that

$$\langle \mathbf{w}_j, \mathbf{w}_k \rangle = 0, \quad j \in \{1, \ldots, r\}, k \in \{r+1, \ldots, n\}. \tag{31}$$

((29) is the "spanning part" of basis, (30) is the "independence part" of basis, and (31) is the "orthogonal compliment" relationship between W and $W^{\perp}$.)

a) Prove first that

$$S_W \cap S_{W^{\perp}} = \{\ \}. \tag{32}$$

($\{\ \}$ denotes the empty set. So (32) just says these two sets have nothing in common.) **Hint:** Do a proof by contradiction. Suppose they did have even just one object $\mathbf{w}$ in common. Use (31), and a relevant property of the innerproduct, to deduce that $\mathbf{w}$ would be a vector that neither ***BASIS*** S_W nor ***BASIS*** $S_{W^{\perp}}$ could possibly possess!!

b) Use (32) to prove second that

$$\left| S_W \cup S_{W^{\perp}} \right| = \left| \{\mathbf{w}_1, \cdots, \mathbf{w}_r, \mathbf{w}_{r+1}, \cdots, \mathbf{w}_n\} \right| = n. \tag{33}$$

Recall that $|S|$ denotes the cardinality of set S, i.e., the number of ***distinct*** objects in the set.

(Why is (32) important for (33)? Because, letting a, b, b', c denote distinct objects, we always have

$$\left| \{a,b\} \cup \{b,c\} \right| = \left| \{a,b,b,c\} \right| = \left| \{a,b,c\} \right| = 3 \neq 4 = \left| \{a,b,b',c\} \right|. \tag{34}$$

)

c) Use (33) to prove that ***IF*** $S_W \cup S_{W^{\perp}}$ is linearly independent, then it's a basis for V. (Nothing much to do here but state the theorem.)

d) Prove that $S_W \bigcup S_{W^\perp}$ ***IS*** linearly independent (hence a basis for V). Hint: You may use the fact that $W \bigcap W^\perp = \{\mathbf{z}\}$ (which, FYI —but otherwise not really important now—follows from (31) rather directly, just like as (32) does). If you use the hint that $W \bigcap W^\perp = \{\mathbf{z}\}$, you should also consider rewriting the linear-independence-checking equation

$$k_1\mathbf{w}_1 + \cdots + k_r\mathbf{w}_r + k_{r+1}\mathbf{w}_{r+1} + \cdots + k_n\mathbf{w}_n = \mathbf{z} \tag{35}$$

as

$$k_1\mathbf{w}_1 + \cdots + k_r\mathbf{w}_r = (-k_{r+1})\mathbf{w}_{r+1} + \cdots + (-k_n)\mathbf{w}_n. \tag{36}$$

Ultimately then you'll also want to use both parts of (30).

30pts

Solution

For a) we just note that if they had even just one vector $\mathbf{w}$ in common, then from (31) we'd get

$$\langle \mathbf{w}, \mathbf{w} \rangle = 0 \Rightarrow \mathbf{w} = \mathbf{z} \tag{37}$$

and we'd have bases containing the additive identity, which is impossible since any set containing the additive identity is linearly dependent, hence, by definition, not a basis!

For b) , we note that since they've nothing in common the union doesn't "collapse" at all (such as it did on the left side of (34)), hence the cardinality is just the total number of distinctly-labeled objects; i.e., it's n.

For c), we just state the theorem that since we've got a set of n (distinct) vectors in an n-dimensional space V, it's a basis provided it's linearly independent.

For d) we could do as directed by the hint, get (36), then use that $W \bigcap W^\perp = \{\mathbf{z}\}$ implies that both (hence either) side of (36) is the additive identity $\mathbf{z}$: by closure under linear combination of both $W(\supset S_W)$ and $W^\perp(\supset S_{W^\perp})$, we get both sides of (36) are in both W and $W^\perp$, hence by $W \bigcap W^\perp = \{\mathbf{z}\}$, both sides are $\mathbf{z}$.But then we get both equations and implications in (30), giving all the k's in (35) needed to be 0. By definition then, that means that the set $\{\mathbf{w}_1,\cdots,\mathbf{w}_r,\mathbf{w}_{r+1},\cdots,\mathbf{w}_n\}$ is linearly independent (hence, with the other facts, a basis for V).

7. Let $S_W := \{\mathbf{w}_1,\ldots,\mathbf{w}_m\}$ be a basis for a (obviously nonzero) subspace W of finite-dimensional innerproduct space V. (Hence $S_W := \{\mathbf{w}_1,\ldots,\mathbf{w}_m\}$ is linearly independent, and spans W.) Further, assume $W \neq V$, even while we certainly have $W \subset V$. That is, assume that there's at least one element $\mathbf{v} \in V$ yet $\mathbf{v} \notin W$. By theorem, since $\mathbf{v} \notin W$ we have then that $S_W{}^+ := \{\mathbf{w}_1,\ldots,\mathbf{w}_m,\mathbf{v}\}$ is linearly independent hence a basis for its span

$$\begin{aligned} W^+ &:= span(S_W{}^+) \\ &:= span\{\mathbf{w}_1,\ldots,\mathbf{w}_m,\mathbf{v}\} \underset{\text{space}}{\subset} V. \end{aligned} \tag{38}$$

Now we can always choose this $\mathbf{v}$ (that's in V but not in W) orthogonal to every element in $S_W := \{\mathbf{w}_1,\ldots,\mathbf{w}_m\}$, this even if $S_W := \{\mathbf{w}_1,\ldots,\mathbf{w}_m\}$ is not itself an orthogonal set. Let's prove this together. (For your confidence, this is a much-simplified version of problem 7 from midterm III. And the reason this theorem is important is to show that the beginning assumption in the previous problem is a good one, namely that we always have $\dim W + \dim W^\perp = \dim V$.)

Well, if $\mathbf{v}$ (which is ***not*** in W) is not "already" orthogonal to every element in $S_W := \{\mathbf{w}_1,\ldots,\mathbf{w}_m\}$, then attempt to make it so ***without making it*** in W by subtracting from it an appropriate linear combination of the elements of the basis $S_W := \{\mathbf{w}_1,\ldots,\mathbf{w}_m\}$ for W: replace $\mathbf{v}$ in $S_W{}^+ := \{\mathbf{w}_1,\ldots,\mathbf{w}_m,\mathbf{v}\}$ by

$$\mathbf{v}^\perp := \mathbf{v} - \left(c_1\mathbf{w}_1 + \cdots + c_m\mathbf{w}_m\right), \tag{39}$$

where (it's not too hard to show) demanding the c's satisfy

$$\langle \mathbf{w}_i,\mathbf{w}_1\rangle c_1 + \cdots + \langle \mathbf{w}_i,\mathbf{w}_m\rangle c_m = \langle \mathbf{w}_i,\mathbf{v}\rangle, \quad i = 1,\ldots,m \tag{40}$$

will make $\mathbf{v}^\perp$ orthogonal to every element of $S_W := \{\mathbf{w}_1,\ldots,\mathbf{w}_m\}$ as required. And note from (39) that any m-tuple $(c_1,\ldots,c_m) \in \mathbb{R}^m$ satisfying (40) will ***not*** make $\mathbf{v}^\perp \in W$ simply because $\mathbf{v} \notin W = span\{\mathbf{w}_1,\ldots,\mathbf{w}_m\}$. Finally note also then from (39) that we certainly get

$$span\{\mathbf{w}_1,\ldots,\mathbf{w}_m,\mathbf{v}\} = span\{\mathbf{w}_1,\ldots,\mathbf{w}_m,\mathbf{v}^{\perp}\}. \tag{41}$$

So all we need to show is that (40) has a solution $(c_1,\ldots,c_m)\in\mathbb{R}^m$. In the process we'll also show that it has only one solution. To that end note that (40) is just a linear system of equations

$$A\mathbf{x} = \mathbf{b} \tag{42}$$

with initially unknown column vector $\mathbf{x}$ given by

$$\mathbf{x} = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} \tag{43}$$

and known right-hand side

$$\mathbf{b} = \begin{bmatrix} \langle \mathbf{w}_1,\mathbf{v}\rangle \\ \vdots \\ \langle \mathbf{w}_m,\mathbf{v}\rangle \end{bmatrix} \tag{44}$$

and square matrix

$$A = \begin{bmatrix} \langle \mathbf{w}_1,\mathbf{w}_1\rangle & \cdots & \langle \mathbf{w}_1,\mathbf{w}_m\rangle \\ & \vdots & \\ \langle \mathbf{w}_m,\mathbf{w}_1\rangle & \cdots & \langle \mathbf{w}_m,\mathbf{w}_m\rangle \end{bmatrix}. \tag{45}$$

So (40) will certainly have one (and only one) solution if A in (45) is invertible. Well, from equivalent statements (which, unfortunately, depend on row reduction, which dependence I don't like—for reasons you should know by now), we have A is invertible if the homogeneous equation

$$A\mathbf{x} = \mathbf{0} \tag{46}$$

has only the trivial solution. Translating back to the original notation we find that (40) will have one (and only one) solution $(c_1,\ldots,c_m)\in\mathbb{R}^m$ provided

$$\langle \mathbf{w}_i,\mathbf{w}_1\rangle c_1 + \cdots + \langle \mathbf{w}_i,\mathbf{w}_m\rangle c_m = 0, \quad i = 1,\ldots,m \tag{47}$$

has only the trivial solution

$$(c_1,\ldots,c_m) = (0,\ldots,0). \tag{48}$$

Ok, well, let's think about whether (47)has only a trivial solution. To that end rewrite (47) as

$$\langle \mathbf{w}_i, c_1\mathbf{w}_1 + \cdots + c_m\mathbf{w}_m \rangle = 0, \quad i = 1, \ldots, m \tag{49}$$

which is possible by linearity of the inner product. Now note that (49) implies that

$$\begin{aligned}\langle c_1\mathbf{w}_1 + \cdots + c_m\mathbf{w}_m, c_1\mathbf{w}_1 + \cdots + c_m\mathbf{w}_m \rangle &= \left\langle \sum_{i=1}^{m} c_i\mathbf{w}_i, c_1\mathbf{w}_1 + \cdots + c_m\mathbf{w}_m \right\rangle \\ &= \sum_{i=1}^{m} c_i \langle \mathbf{w}_i, c_1\mathbf{w}_1 + \cdots + c_m\mathbf{w}_m \rangle \\ &= \sum_{i=1}^{m} c_i 0 = \sum_{i=1}^{m} 0 = 0,\end{aligned} \tag{50}$$

i.e., (49) (hence (47)) implies that

$$\langle c_1\mathbf{w}_1 + \cdots + c_m\mathbf{w}_m, c_1\mathbf{w}_1 + \cdots + c_m\mathbf{w}_m \rangle = 0. \tag{51}$$

Well, I'm thinking you can take it from here: use the given assumptions to prove that (51) implies that $c_1 = \ldots = c_m = 0$.

20pts

Solution

By nondegeneracy of the innerproduct, (51) implies that

$$c_1\mathbf{w}_1 + \cdots + c_m\mathbf{w}_m = \mathbf{z}, \tag{52}$$

and by linear independence of $S_W := \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ we get (52) implies $c_1 = \ldots = c_m = 0.$

8. In this problem, assume $b \neq a$.

   Find the characteristic polynomial $P_A(\lambda)$ of

$$A = \begin{bmatrix} a & 1 \\ 0 & b \end{bmatrix}. \tag{53}$$

Now find the eigenvalues λ of A. Then find the eiqenspaces of A. (“Finding” a vector space ALWAYS means writing it as the span of a basis.) Now write

$$A = P\Lambda P^{-1} \tag{54}$$

for some diagonal matrix Λ and some invertible matrix P. Now find a formula for all four entries of $f(A)$, where f is (the appropriate extension of) any sufficiently well-behaved function of a real variable.

Partial hint: your answer for $f(A)$ should imply that for every positive integer n,

$$A^n = \begin{bmatrix} a^n & \dfrac{b^n - a^n}{b-a} \\ 0 & b^n \end{bmatrix}. \tag{55}$$

<u>20pts</u>

<u>Solution</u>

By definition

$$\begin{aligned} P_A(\lambda) &:= \det(\lambda I - A) = \det\left(\lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} a & 1 \\ 0 & b \end{bmatrix}\right) \\ &= \det \begin{bmatrix} \lambda - a & -1 \\ 0 & \lambda - b \end{bmatrix} = (\lambda - a)(\lambda - b). \end{aligned} \tag{56}$$

By theorem λ is an eigenvalue of A iff $P_A(\lambda) = 0$, so from (56) the eigenvalues of A are $\lambda_1 = a$ and $\lambda_2 = b$.

By theorem we have

$$\begin{aligned}
E_A(a) &= Nul\left[aI - A\right] = Nul\left[a\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} - \begin{bmatrix}a & 1\\ 0 & b\end{bmatrix}\right]\\
&= Nul\begin{bmatrix}a-a & -1\\ 0 & a-b\end{bmatrix} = Nul\begin{bmatrix}0 & -1\\ 0 & a-b\end{bmatrix}\\
&= Nul\begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix} = Nul\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix} = span\left\{\begin{pmatrix}1\\ 0\end{pmatrix}\right\},\\
E_A(b) &= Nul\left[bI - A\right] = Nul\left[b\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} - \begin{bmatrix}a & 1\\ 0 & b\end{bmatrix}\right]\\
&= Nul\begin{bmatrix}b-a & -1\\ 0 & b-b\end{bmatrix} = Nul\begin{bmatrix}b-a & -1\\ 0 & 0\end{bmatrix}\\
&= span\left\{\begin{pmatrix}1\\ b-a\end{pmatrix}\right\}.
\end{aligned} \tag{57}$$

Since the eigenvalues are distinct, by theorem we can take

$$\Lambda = \begin{bmatrix}a & 0\\ 0 & b\end{bmatrix}, P = \left[\begin{array}{c|c}1 & 1\\ 0 & b-a\end{array}\right] = \begin{bmatrix}1 & 1\\ 0 & b-a\end{bmatrix}, \tag{58}$$

and find then that

$$P^{-1} = \frac{1}{\det P}\begin{bmatrix}b-a & -1\\ 0 & 1\end{bmatrix} = \frac{1}{b-a}\begin{bmatrix}b-a & -1\\ 0 & 1\end{bmatrix}. \tag{59}$$

So we have

$$\begin{bmatrix}a & 1\\ 0 & b\end{bmatrix} = A = P\Lambda P^{-1} = \begin{bmatrix}1 & 1\\ 0 & b-a\end{bmatrix}\begin{bmatrix}a & 0\\ 0 & b\end{bmatrix}\frac{1}{b-a}\begin{bmatrix}b-a & -1\\ 0 & 1\end{bmatrix}. \tag{60}$$

(And we can check that

$$\begin{aligned}
\begin{bmatrix}1 & 1\\ 0 & b-a\end{bmatrix}\begin{bmatrix}a & 0\\ 0 & b\end{bmatrix}\frac{1}{b-a}\begin{bmatrix}b-a & -1\\ 0 & 1\end{bmatrix} &= \frac{1}{b-a}\begin{bmatrix}a & b\\ 0 & b(b-a)\end{bmatrix}\begin{bmatrix}b-a & -1\\ 0 & 1\end{bmatrix}\\
&= \frac{1}{b-a}\begin{bmatrix}a(b-a) & -a+b\\ 0 & b(b-a)\end{bmatrix}\\
&= \begin{bmatrix}a & 1\\ 0 & b\end{bmatrix} = A.
\end{aligned} \tag{61}$$

)

By theorem we have

$$A = P\Lambda P^{-1} \Rightarrow f(A) = Pf(\Lambda)P^{-1}, \tag{62}$$

and moreover $f(\Lambda)$ is just the diagonal matrix arising from f applied to the diagonal entries of Λ when Λ is diagonal. So in the present case we get

$$\begin{aligned}
f(\begin{bmatrix} a & 1 \\ 0 & b \end{bmatrix}) &= f(A) = Pf(\Lambda)P^{-1} \\
&= \begin{bmatrix} 1 & 1 \\ 0 & b-a \end{bmatrix} \begin{bmatrix} f(a) & 0 \\ 0 & f(b) \end{bmatrix} \frac{1}{b-a} \begin{bmatrix} b-a & -1 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} f(a) & f(b) \\ 0 & f(b)(b-a) \end{bmatrix} \frac{1}{b-a} \begin{bmatrix} b-a & -1 \\ 0 & 1 \end{bmatrix} \\
&= \frac{1}{b-a} \begin{bmatrix} f(a)(b-a) & f(b)-f(a) \\ 0 & f(b)(b-a) \end{bmatrix} \\
&= \begin{bmatrix} f(a) & \dfrac{f(b)-f(a)}{b-a} \\ 0 & f(b) \end{bmatrix}.
\end{aligned} \tag{63}$$

9. The Singular Value Decomposition theorem says we can write every matrix $A \in \mathbb{R}^{mxn}$ as

$$A = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^{\ T} + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^{\ T} \tag{64}$$

where $k = rank(A)$ $(\leq m,n)$, $\sigma_1, \ldots, \sigma_k > 0$, $S_{Col(A)} := \{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \subset \mathbb{R}^m$ is an orthonormal basis (of column vectors) for the column space of A, and $S_{Row(A)} := \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset \mathbb{R}^n$ is an orthonormal basis (of column vectors) for the row space of A—giving $\{\mathbf{v}_1^{\ T}, \ldots, \mathbf{v}_k^{\ T}\} \subset \mathbb{R}^n$ is an orthonormal basis of row vectors for the row space of A. (If (64) holds with $\sigma_1, \ldots, \sigma_k > 0$, and $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset \mathbb{R}^n$ orthonormal then

$$
\begin{aligned}
Col(A) &= \left\{ A\mathbf{x} \middle| \mathbf{x} \in \mathbb{R}^n \right\} \\
&= \left\{ \left(\sigma_1 \mathbf{u}_1 \mathbf{v}_1{}^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k{}^T \right) \mathbf{x} \middle| \mathbf{x} \in \mathbb{R}^n \right\} \\
&= \left\{ \sigma_1 \mathbf{u}_1 \mathbf{v}_1{}^T \mathbf{x} + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k{}^T \mathbf{x} \middle| \mathbf{x} \in \mathbb{R}^n \right\} \\
&= \left\{ \sigma_1 \left(\mathbf{v}_1 \cdot \mathbf{x} \right) \mathbf{u}_1 + \cdots + \sigma_k \left(\mathbf{v}_k \cdot \mathbf{x} \right) \mathbf{u}_k \middle| \mathbf{x} \in \mathbb{R}^n \right\} \\
&= \left\{ c_1 \mathbf{u}_1 + \cdots + c_k \mathbf{u}_k \middle| c_1, \ldots, c_k \in \mathbb{R} \right\} \\
&=: span \left\{ \mathbf{u}_1, \ldots, \mathbf{u}_k \right\} =: span\ S_{Col(A)}.
\end{aligned} \tag{65}
$$

And with $\left\{ \mathbf{u}_1, \ldots, \mathbf{u}_k \right\} \subset \mathbb{R}^m$ orthonormal then $\left\{ \mathbf{u}_1, \ldots, \mathbf{u}_k \right\}$ is linearly independent and, so, a basis for the column space of A. In (65) we used that choosing $\mathbf{x} = c_1 \mathbf{v}_1 / \sigma_1 + \cdots + c_k \mathbf{v}_k / \sigma_k \in \mathbb{R}^n$ gives

$$
\mathbf{v}_j \cdot \mathbf{x} = c_j / \sigma_j, \quad j = 1, \ldots, k \tag{66}
$$

and then

$$
\begin{aligned}
\sigma_1 \left(\mathbf{v}_1 \cdot \mathbf{x} \right) \mathbf{u}_1 + \cdots + \sigma_k \left(\mathbf{v}_k \cdot \mathbf{x} \right) \mathbf{u}_k &= \sigma_1 c_1 / \sigma_1 \mathbf{u}_1 + \cdots + \sigma_k c_k / \sigma_k \mathbf{u}_k \\
&= c_1 \mathbf{u}_1 + \cdots + c_k \mathbf{u}_k .
\end{aligned} \tag{67}
$$

A similar argument holds to show that orthonormal set $\left\{ \mathbf{v}_1, \ldots, \mathbf{v}_k \right\} \subset \mathbb{R}^n$ will be a basis for the row space of A given (64), etc.)

Note that (64) immediately implies that

$$
A^T = \sigma_1 \mathbf{v}_1 \mathbf{u}_1{}^T + \cdots + \sigma_k \mathbf{v}_k \mathbf{u}_k{}^T , \tag{68}
$$

whence with $\left\{ \mathbf{u}_1, \ldots, \mathbf{u}_k \right\}$ orthonormal we get

$$
\begin{aligned}
A^T A &= \left(\sigma_1 \mathbf{v}_1 \mathbf{u}_1^T + \cdots + \sigma_k \mathbf{v}_k \mathbf{u}_k^T\right)\left(\sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T\right) \\
&= \sigma_1 \mathbf{v}_1 \mathbf{u}_1^T \left(\sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T\right) + \cdots + \sigma_k \mathbf{v}_k \mathbf{u}_k^T \left(\sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{u}_k \mathbf{v}_k^T\right) \\
&= \left(\sigma_1 \mathbf{v}_1 \mathbf{u}_1^T \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_1 \mathbf{v}_1 \mathbf{u}_1^T \sigma_k \mathbf{u}_k \mathbf{v}_k^T\right) + \cdots \\
&\quad + \left(\sigma_k \mathbf{v}_k \mathbf{u}_k^T \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T + \cdots + \sigma_k \mathbf{v}_k \mathbf{u}_k^T \sigma_k \mathbf{u}_k \mathbf{v}_k^T\right) \\
&= \left(\sigma_1^2 \mathbf{v}_1 \left(\mathbf{u}_1^T \mathbf{u}_1\right) \mathbf{v}_1^T + \cdots + \sigma_1 \sigma_k \mathbf{v}_1 \left(\mathbf{u}_1^T \mathbf{u}_k\right) \mathbf{v}_k^T\right) + \cdots + \\
&\quad \left(\sigma_k \sigma_1 \mathbf{v}_k \left(\mathbf{u}_k^T \mathbf{u}_1\right) \mathbf{v}_1^T + \cdots + \sigma_k^2 \mathbf{v}_k \left(\mathbf{u}_k^T \mathbf{u}_k\right) \mathbf{v}_k^T\right) \qquad (69) \\
&= \left(\sigma_1^2 \mathbf{v}_1 \left(\mathbf{u}_1 \cdot \mathbf{u}_1\right) \mathbf{v}_1^T + \cdots + \sigma_1 \sigma_k \mathbf{v}_1 \left(\mathbf{u}_1 \cdot \mathbf{u}_k\right) \mathbf{v}_k^T\right) + \cdots + \\
&\quad \left(\sigma_k \sigma_1 \mathbf{v}_k \left(\mathbf{u}_k \cdot \mathbf{u}_1\right) \mathbf{v}_1^T + \cdots + \sigma_k^2 \mathbf{v}_k \left(\mathbf{u}_k \cdot \mathbf{u}_k\right) \mathbf{v}_k^T\right) \\
&= \left(\sigma_1^2 \mathbf{v}_1 (1) \mathbf{v}_1^T + \cdots + \sigma_1 \sigma_k \mathbf{v}_1 (0) \mathbf{v}_k^T\right) + \cdots + \\
&\quad \left(\sigma_k \sigma_1 \mathbf{v}_k (0) \mathbf{v}_1^T + \cdots + \sigma_k^2 \mathbf{v}_k (1) \mathbf{v}_k^T\right) \\
&= \sigma_1^2 \mathbf{v}_1 \mathbf{v}_1^T + \cdots + \sigma_k^2 \mathbf{v}_k \mathbf{v}_k^T,
\end{aligned}
$$

i.e., we get

$$A^T A = \sigma_1^2 \mathbf{v}_1 \mathbf{v}_1^T + \cdots + \sigma_k^2 \mathbf{v}_k \mathbf{v}_k^T \qquad (70)$$

(from which it's not too hard to show that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset \mathbb{R}^n$ is just normalized eigenvectors of $A^T A$ corresponding to positive eigenvalues $\sigma_1^2, \ldots, \sigma_k^2$ of $A^T A$.) So then the (always solvable) normal equation

$$A^T A\mathbf{x} = A^T \mathbf{b} \qquad (71)$$

is

$$\left(\sigma_1^2 \mathbf{v}_1 \mathbf{v}_1^T + \cdots + \sigma_k^2 \mathbf{v}_k \mathbf{v}_k^T\right)\mathbf{x} = \left(\sigma_1 \mathbf{v}_1 \mathbf{u}_1^T + \cdots + \sigma_k \mathbf{v}_k \mathbf{u}_k^T\right)\mathbf{b}. \qquad (72)$$

Show that

$$\mathbf{x} = \left(\frac{\mathbf{v}_1 \mathbf{u}_1^T}{\sigma_1} + \cdots + \frac{\mathbf{v}_k \mathbf{u}_k^T}{\sigma_k}\right)\mathbf{b} =: A_{MP}^{-1}\mathbf{b} \qquad (73)$$

solves the normal equation (72). (Here, FYI, A_{MP}^{-1} is called the "Moore-Penrose pseudo inverse" of $A \in \mathbb{R}^{m \times n}$.) (Remember NOT to do the high school thing of immediately manipulating both sides of (72) given (73) until you get "truth", which doesn't prove anything, but rather manipulate one side or the other of (72) using (73) until you get the un-manipulated side, which DOES prove the assertion.)

.

20pts

Solution

With (73) and the orthonormality of $\{\mathbf{v}_1,\ldots,\mathbf{v}_k\}$ we get

$$\begin{aligned}
\left(\sigma_1^2\mathbf{v}_1\mathbf{v}_1^T+\cdots+\sigma_k^2\mathbf{v}_k\mathbf{v}_k^T\right)\mathbf{x} &= \left(\sigma_1^2\mathbf{v}_1\mathbf{v}_1^T+\cdots+\sigma_k^2\mathbf{v}_k\mathbf{v}_k^T\right)\left(\frac{\mathbf{v}_1\mathbf{u}_1^T}{\sigma_1}+\cdots+\frac{\mathbf{v}_k\mathbf{u}_k^T}{\sigma_k}\right)\mathbf{b} \\
&= \begin{pmatrix} \left(\sigma_1^2\mathbf{v}_1\mathbf{v}_1^T+\cdots+\sigma_k^2\mathbf{v}_k\mathbf{v}_k^T\right)\dfrac{\mathbf{v}_1\mathbf{u}_1^T}{\sigma_1}+\cdots \\ +\left(\sigma_1^2\mathbf{v}_1\mathbf{v}_1^T+\cdots+\sigma_k^2\mathbf{v}_k\mathbf{v}_k^T\right)\dfrac{\mathbf{v}_k\mathbf{u}_k^T}{\sigma_k} \end{pmatrix}\mathbf{b} \\
&= \left(\left(\sigma_1^2\mathbf{v}_1\mathbf{v}_1^T\right)\frac{\mathbf{v}_1\mathbf{u}_1^T}{\sigma_1}+\cdots+\left(\sigma_k^2\mathbf{v}_k\mathbf{v}_k^T\right)\frac{\mathbf{v}_k\mathbf{u}_k^T}{\sigma_k}\right)\mathbf{b} \\
&= \left(\sigma_1\mathbf{v}_1\left(\mathbf{v}_1^T\mathbf{v}_1\right)\mathbf{u}_1^T+\cdots+\sigma_k\mathbf{v}_k\left(\mathbf{v}_k^T\mathbf{v}_k\right)\mathbf{u}_k^T\right)\mathbf{b} \\
&= \left(\sigma_1\mathbf{v}_1(1)\mathbf{u}_1^T+\cdots+\sigma_k\mathbf{v}_k(1)\mathbf{u}_k^T\right)\mathbf{b} \\
&= \left(\sigma_1\mathbf{v}_1\mathbf{u}_1^T+\cdots+\sigma_k\mathbf{v}_k\mathbf{u}_k^T\right)\mathbf{b}
\end{aligned} \tag{74}$$

whence (72) holds.

10. In this problem, assume $b \neq a$.

    Let $T : P_1 \to P_1$ be defined by

$$T[f](x) := \int_0^1 \left(a+3\left(1+b(2x-1)\right)(2y-1)\right) f(y)dy. \tag{75}$$

    Note that T is linear, and note that defining $f_0(x) := 1$, $f_1(x) := 2x-1$ renders $S := \{f_0, f_1\}$ a basis for P_1, and gives

$$\begin{aligned} T[f_0](x) &:= \int_0^1 \left(a+3\left(1+b(2x-1)\right)\left(2y-1\right)\right) f_0(y)dy \\ &= \int_0^1 \left(a+3\left(1+b(2x-1)\right)\left(2y-1\right)\right) dy \\ &= \left[ay+3\left(1+b(2x-1)\right)\left(y^2-y\right)\right]_{y=0}^{y=1} \\ &= a+3\left(1+b(2x-1)\right)\left(1^2-1\right) = a = a\cdot 1 = af_0(x), \end{aligned} \tag{76}$$

and

$$\begin{aligned} T[f_1](x) &:= \int_0^1 \left(a+3\left(1+b(2x-1)\right)\left(2y-1\right)\right) f_1(y)dy \\ &= \int_0^1 \left(a+3\left(1+b(2x-1)\right)\left(2y-1\right)\right)\left(2y-1\right) dy \\ &= \int_0^1 \left(a\left(2y-1\right)+3\left(1+b(2x-1)\right)\left(2y-1\right)^2\right) dy \\ &= \left[a\left(y^2-y\right)+3\left(1+b(2x-1)\right)\frac{\left(2y-1\right)^3}{6}\right]_{y=0}^{y=1} \\ &= a\cdot 0+\left(1+b(2x-1)\right)\left(\frac{\left(2\cdot 1-1\right)^3}{2}-\frac{\left(2\cdot 0-1\right)^3}{2}\right) \\ &= \left(1+b(2x-1)\right)\left(\frac{1}{2}-\frac{-1}{2}\right) = 1+b(2x-1) = 1\cdot 1+b(2x-1) \\ &= 1f_0(x)+bf_1(x), \end{aligned} \tag{77}$$

i.e., we get

$$T[f_0] = af_0 = af_0 + 0f_1, \qquad T[f_1] = 1f_0 + bf_1. \tag{78}$$

What is the matrix $[T]_{SS}$ for T relative to basis $S := \{f_0, f_1\}$? What are the eigenvalues of $[T]_{SS}$? What are the eigenvalues of T? What are $[T]_{SS}$'s eigenspaces? What are T's eigenspaces? Given that n is a positive integer , and using the theorem that

$$\left[T^n\right]_{SS} = [T]_{SS}{}^n, \tag{79}$$

argue that

$$T^n\left[\alpha f_0 + \beta f_1\right] = \left(\alpha a^n + \beta\frac{b^n - a^n}{b-a}\right) f_0 + \beta b^n f_1. \tag{80}$$

Hint: See problem 8. Also, if it's easier, instead of arguing (80), just argue that

$$T^n\left[f_0\right] = a^n f_0 = a^n f_0 + 0\, f_1, \tag{81}$$

and that

$$T^n\left[f_1\right] = \left(\frac{b^n - a^n}{b-a}\right) f_0 + b^n f_1. \tag{82}$$

20pts

Solution

By easy theorem we have

$$\begin{aligned}
[T]_{SS} &= \left[\left(T\left[f_0\right]\right)_S \middle| \left(T\left[f_1\right]\right)_S\right] \\
&= \left[\left(af_0\right)_S \middle| \left(1 f_0 + bf_1\right)_S\right] \\
&= \left[\left(af_0 + 0\, f_1\right)_{\{f_0, f_1\}} \middle| \left(1 f_0 + bf_1\right)_{\{f_0, f_1\}}\right] \\
&= \left[\begin{array}{c|c} a & 1 \\ 0 & b \end{array}\right] = \begin{bmatrix} a & 1 \\ 0 & b \end{bmatrix} = A,
\end{aligned} \tag{83}$$

where A is the matrix of problem 8. Hence we have from just below (56) that the eigenvalues are $\lambda_1 = a$ and $\lambda_2 = b$. By theorem we have that the eigenvalues of T are the same. By (57) we have that

$$\begin{aligned}
E_{[T]_{SS}}(a) &= E_A(a) = span\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\}, \\
E_{[T]_{SS}}(b) &= E_A(b) = span\left\{\begin{pmatrix} 1 \\ b-a \end{pmatrix}\right\},
\end{aligned} \tag{84}$$

and then by theorem we have that

$$\begin{aligned}
E_T(a) &= span\left\{1 f_0 + 0\, f_1\right\} = span\left\{f_0\right\}, \\
E_T(b) &= span\left\{1 f_0 + \left(b-a\right) f_1\right\} = span\left\{f_0 + \left(b-a\right) f_1\right\}.
\end{aligned} \tag{85}$$

So now since from(79), (55) and (83) we have

$$\left[T^n\right]_{SS} = \left[T\right]_{SS}^{\;n} = A^n = \begin{bmatrix} a^n & \dfrac{b^n - a^n}{b-a} \\ 0 & b^n \end{bmatrix}, \tag{86}$$

and since

$$T^n\left[\alpha f_0 + \beta f_1\right] = \alpha T^n\left[f_0\right] + \beta T^n\left[f_1\right] \tag{87}$$

then

$$\begin{aligned} \left(T^n\left[\alpha f_0 + \beta f_1\right]\right)_S &= \alpha\left(T^n\left[f_0\right]\right)_S + \beta\left(T^n\left[f_1\right]\right)_S \\ &= \alpha\left[T^n\right]_{SS}\left(f_0\right)_S + \beta\left[T^n\right]_{SS}\left(f_1\right)_S \\ &= \alpha\left[T^n\right]_{SS}\mathbf{e}_1 + \beta\left[T^n\right]_{SS}\mathbf{e}_2 \\ &= \alpha\begin{bmatrix} a^n \\ 0 \end{bmatrix} + \beta\begin{bmatrix} \dfrac{b^n - a^n}{b-a} \\ b^n \end{bmatrix} \\ &= \begin{bmatrix} \alpha a^n + \beta\dfrac{b^n - a^n}{b-a} \\ \beta b^n \end{bmatrix}, \end{aligned} \tag{88}$$

and finally then, by definition of coordinate vector with respect to basis $S := \left\{f_0, f_1\right\}$, we must have

$$T^n\left[\alpha f_0 + \beta f_1\right] = \left(\alpha a^n + \beta\frac{b^n - a^n}{b-a}\right) f_0 + \beta b^n f_1 \tag{89}$$

which is (80).

11. Consider the following, which is one of the most important theorems in linear algebra:

Theorem 1 (Decomposition)

Let W denote a subspace of finite-dimensional innerproduct space V, and let $W^\perp$ denote the orthogonal complement of W with respect to V. Then, for every $\mathbf{v} \in V$ there is exactly one pair $\left(\mathbf{w}^\parallel, \mathbf{w}^\perp\right) \in W \times W^\perp$ such that

$$\mathbf{v} = \mathbf{w}^\parallel + \mathbf{w}^\perp. \tag{90}$$

Use this theorem to prove that $\mathbf{w}^{\parallel}$ in (90) is the closest vector in W to $\mathbf{v}$. That is, prove that for every ***other*** $\mathbf{w} \in W$, ("***other*** "means $\mathbf{w} \neq \mathbf{w}^{\parallel} \in W$), we have

$$d(\mathbf{w},\mathbf{v}) > d(\mathbf{w}^{\parallel},\mathbf{v}). \tag{91}$$

Actually, it's enough (and easier) to prove that

$$\left(d(\mathbf{w},\mathbf{v})\right)^2 > \left(d(\mathbf{w}^{\parallel},\mathbf{v})\right)^2, \tag{92}$$

which is what then you should try to prove. Related to the idea that (92) is the basic/fundamental thing to prove (from which (91) will follow) is the fact that (92) will be quite impossible to prove if you don't know that in an innerproduct space the notion of distance is as follows:

$$\left(d(\mathbf{w},\mathbf{v})\right)^2 := \left\|\mathbf{w}-\mathbf{v}\right\|^2 := \left\langle \mathbf{w}-\mathbf{v},\mathbf{w}-\mathbf{v}\right\rangle. \tag{93}$$

Also it will be quite impossible to prove (92) if you don't know what orthogonal complement means. We have this definition:

$$W^{\perp} := \left\{\mathbf{w}^{\perp} \in V \middle| \left\langle \mathbf{w},\mathbf{w}^{\perp}\right\rangle = 0 \text{ for every } \mathbf{w} \in W\right\}. \tag{94}$$

(Recall that by symmetry of the innerproduct, in (94) we could have just as well written $\left\langle \mathbf{w}^{\perp},\mathbf{w}\right\rangle$ as $\left\langle \mathbf{w},\mathbf{w}^{\perp}\right\rangle$.) Thus, for (very relevant) example (but not so relevant as to give you the answer immediately), we have that decomposition (90) and definition (94) (together with $\mathbf{w}^{\parallel} \in W$) give the following "Pythagorean theorem":

$$\begin{aligned}
\left\|\mathbf{v}\right\|^2 &= \left\|\mathbf{w}^{\parallel}+\mathbf{w}^{\perp}\right\|^2 := \left\langle \mathbf{w}^{\parallel}+\mathbf{w}^{\perp},\mathbf{w}^{\parallel}+\mathbf{w}^{\perp}\right\rangle \\
&= \left\langle \mathbf{w}^{\parallel},\mathbf{w}^{\parallel}+\mathbf{w}^{\perp}\right\rangle + \left\langle \mathbf{w}^{\perp},\mathbf{w}^{\parallel}+\mathbf{w}^{\perp}\right\rangle \\
&= \left\langle \mathbf{w}^{\parallel},\mathbf{w}^{\parallel}\right\rangle + \left\langle \mathbf{w}^{\parallel},\mathbf{w}^{\perp}\right\rangle + \left\langle \mathbf{w}^{\perp},\mathbf{w}^{\parallel}\right\rangle + \left\langle \mathbf{w}^{\perp},\mathbf{w}^{\perp}\right\rangle \\
&= \left\langle \mathbf{w}^{\parallel},\mathbf{w}^{\parallel}\right\rangle + 0 + 0 + \left\langle \mathbf{w}^{\perp},\mathbf{w}^{\perp}\right\rangle \\
&= \left\langle \mathbf{w}^{\parallel},\mathbf{w}^{\parallel}\right\rangle + \left\langle \mathbf{w}^{\perp},\mathbf{w}^{\perp}\right\rangle \\
&=: \left\|\mathbf{w}^{\parallel}\right\|^2 + \left\|\mathbf{w}^{\perp}\right\|^2,
\end{aligned} \tag{95}$$

i.e., $\left(\mathbf{w}^{\parallel},\mathbf{w}^{\perp}\right) \in W \times W^{\perp}$ implies that

$$\left\|\mathbf{w}^{\parallel}+\mathbf{w}^{\perp}\right\|^2 = \left\|\mathbf{w}^{\parallel}\right\|^2 + \left\|\mathbf{w}^{\perp}\right\|^2. \tag{96}$$

Well, you should have enough tools now to show that the left side of (92) is indeed strictly greater than the right side of (92) when $\mathbf{w}, \mathbf{w}^{\parallel} \in W$ yet $\mathbf{w} \neq \mathbf{w}^{\parallel}$. To get strictly greater as in (92), instead of just getting "greater than or equal to" ($\geq$), it will be important to know that

$$\mathbf{w} \neq \mathbf{w}^{\parallel} \Rightarrow d\left(\mathbf{w}, \mathbf{w}^{\parallel}\right) := \left\|\mathbf{w} - \mathbf{w}^{\parallel}\right\| > 0, \tag{97}$$

which is itself another hint on how to get started in (92).

20pts

Solution

There are slicker ways to prove this, but how about the following "can't help but get the right answer" method: On the one hand we have

$$\begin{aligned}
\left(d(\mathbf{w}, \mathbf{v})\right)^2 := \left\|\mathbf{w} - \mathbf{v}\right\|^2 &= \left\|\mathbf{w} - \left(\mathbf{w}^{\parallel} + \mathbf{w}^{\perp}\right)\right\|^2 \\
&= \left\|\mathbf{w} - \mathbf{w}^{\parallel} + \left(-\mathbf{w}^{\perp}\right)\right\|^2 \\
&= \left\|\mathbf{w} - \mathbf{w}^{\parallel}\right\|^2 + \left\|\left(-\mathbf{w}^{\perp}\right)\right\|^2 \\
&= \left\|\mathbf{w} - \mathbf{w}^{\parallel}\right\|^2 + \left\|\mathbf{w}^{\perp}\right\|^2 \\
&> \left\|\mathbf{w}^{\perp}\right\|^2,
\end{aligned} \tag{98}$$

where we used a version of Pythagorean theorem (96) (made possibly by $\mathbf{w}, \mathbf{w}^{\parallel} \in W$ implies $\mathbf{w} - \mathbf{w}^{\parallel} \in W$, etc.) and that $\left\|\mathbf{w} - \mathbf{w}^{\parallel}\right\| > 0$ as in (97). On the other hand we have

$$\begin{aligned}
\left(d(\mathbf{w}^{\parallel}, \mathbf{v})\right)^2 := \left\|\mathbf{w}^{\parallel} - \mathbf{v}\right\|^2 &= \left\|\mathbf{w}^{\parallel} - \left(\mathbf{w}^{\parallel} + \mathbf{w}^{\perp}\right)\right\|^2 \\
&= \left\|-\mathbf{w}^{\perp}\right\|^2 = \left\|\mathbf{w}^{\perp}\right\|^2,
\end{aligned} \tag{99}$$

and between (98) and (99) we get

$$\left(d(\mathbf{w}, \mathbf{v})\right)^2 > \left\|\mathbf{w}^{\perp}\right\|^2 = \left(d(\mathbf{w}^{\parallel}, \mathbf{v})\right)^2, \tag{100}$$

i.e.,

$$\left(d(\mathbf{w},\mathbf{v})\right)^2 > \left(d(\mathbf{w}^{\|},\mathbf{v})\right)^2. \tag{101}$$

12. Another one of the “big ones” from Linear algebra:

Theorem 2

The equation

$$A\mathbf{x} = \mathbf{b} \tag{102}$$

has a solution $\mathbf{x}$ iff $\mathbf{b}$ is orthogonal to every vector $\mathbf{y} \in Nul(A^T)$.

For the normal equation

$$A^T A\mathbf{x} = A^T\mathbf{b} \tag{103}$$

this theorem becomes as follows: (103) has a solution $\mathbf{x}$ iff $A^T\mathbf{b}$ is orthogonal to every vector $\mathbf{y} \in Nul\left(\left(A^T A\right)^T\right) = Nul(A^T A)$. So consider that

$$\left\langle \mathbf{y}, A^T\mathbf{b} \right\rangle = \mathbf{y} \cdot A^T\mathbf{b} = \mathbf{y}^T A^T\mathbf{b} = \left(A\mathbf{y}\right)^T \mathbf{b} = A\mathbf{y} \cdot \mathbf{b} = \left\langle A\mathbf{y}, \mathbf{b} \right\rangle \tag{104}$$

and that

$$\begin{gathered} \mathbf{y} \in Nul(A^T A) \\ \Leftrightarrow \\ A^T A\mathbf{y} = \mathbf{0} \\ \Leftrightarrow \\ 0 \equiv \left\langle \mathbf{w}, A^T A\mathbf{y} \right\rangle, \end{gathered} \tag{105}$$

the latter meaning that $\left\langle \mathbf{w}, A^T A\mathbf{y} \right\rangle = 0$ holds for every $\mathbf{w}$. ($A^T A\mathbf{y} = \mathbf{0} \Rightarrow 0 \equiv \left\langle \mathbf{w}, A^T A\mathbf{y} \right\rangle$ easily, and then $0 \equiv \left\langle \mathbf{w}, A^T A\mathbf{y} \right\rangle \;\Rightarrow\; 0 = \left\langle A^T A\mathbf{y}, A^T A\mathbf{y} \right\rangle$ $= \left\| A^T A\mathbf{y} \right\|^2 \;\Leftrightarrow\; A^T A\mathbf{y} = \mathbf{0}$.) Thus

$$
\begin{aligned}
&\mathbf{y} \in Nul(A^T A) \\
&\Leftrightarrow \\
&A^T A\mathbf{y} = \mathbf{0} \\
&\Leftrightarrow \\
&0 \equiv \langle \mathbf{w}, A^T A\mathbf{y} \rangle \\
&\Rightarrow \\
&0 = \langle \mathbf{y}, A^T A\mathbf{y} \rangle = \mathbf{y} \cdot A^T A\mathbf{y} \\
&= \mathbf{y}^T A^T A\mathbf{y} = (A\mathbf{y})^T A\mathbf{y} \\
&= A\mathbf{y} \cdot A\mathbf{y} = \|A\mathbf{y}\|^2 \\
&\Leftrightarrow \\
&A\mathbf{y} = \mathbf{0} \\
&\Leftrightarrow \\
&\mathbf{y} \in Nul(A).
\end{aligned}
\tag{106}
$$

So in summary we have that a) for any $\mathbf{y} \in Nul(A^T A)$,

$$\langle \mathbf{y}, A^T\mathbf{b} \rangle = \langle A\mathbf{y}, \mathbf{b} \rangle \tag{107}$$

and b)

$$A^T A\mathbf{y} = \mathbf{0} \Leftrightarrow \mathbf{y} \in Nul(A^T A) \Leftrightarrow \mathbf{y} \in Nul(A) \Leftrightarrow A\mathbf{y} = \mathbf{0}. \tag{108}$$

(In (108) we included that easily $A\mathbf{y} = \mathbf{0} \Rightarrow A^T A\mathbf{y} = \mathbf{0}$.) So given **Theorem 2**, prove that the normal equation (103) is consistent for every $\mathbf{b}$.

20pts

Solution

By the new version of the theorem and (107) and (108) we find that (103) is consistent iff

$$0 = \langle A\mathbf{y}, \mathbf{b} \rangle \tag{109}$$

for every $\mathbf{y}$ such that $A\mathbf{y} = \mathbf{0}$, i.e., is consistent iff

$$0 = \langle \mathbf{0}, \mathbf{b} \rangle, \tag{110}$$

which holds whatever $\mathbf{b}$ is.