Math Problem Set 2

Matthew Brown
OSM Boot Camp 2018

July 1, 2018

Problem 3.1. There are two parts:

- (i)

$$\begin{aligned}
||x+y||^2 - ||x-y||^2 &= \langle x+y, x+y\rangle - \langle x-y, x-y\rangle \\
&= \langle x+y, x\rangle + \langle x+y, y\rangle - (\langle x-y, x\rangle + \langle x-y, -y\rangle) \\
&= \langle x, x+y\rangle + \langle y, x+y\rangle - (\langle x, x-y\rangle + \langle -y, x-y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle - \langle x, x\rangle - \langle x, -y\rangle - \langle -y, x\rangle - \langle -y, -y\rangle \\
&= 4\langle x, y\rangle \text{ after some easy manipulations}
\end{aligned}$$

  So it is clear that $\frac{1}{4}(||x+y||^2 - ||x-y||^2) = \langle x, y\rangle$

- (ii) As above,

$$\begin{aligned}
||x+y||^2 + ||x-y||^2 &= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle \\
&= \langle x+y, x\rangle + \langle x+y, y\rangle + (\langle x-y, x\rangle + \langle x-y, -y\rangle) \\
&= \langle x, x+y\rangle + \langle y, x+y\rangle + (\langle x, x-y\rangle + \langle -y, x-y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle + \langle x, x\rangle + \langle x, -y\rangle + \langle -y, x\rangle + \langle -y, -y\rangle \\
&= 2\langle x, x\rangle + 2\langle y, y, \rangle \text{ after some easy manipulations} \\
&= 2(||x||^2 + ||y||^2)
\end{aligned}$$

  So it is clear that $\frac{1}{2}(||x+y||^2 + ||x-y||^2) = ||x||^2 + ||y||^2$

Problem 3.2. Latex code from kendra: $\frac{1}{4}(||x+y||^2 - ||x-y||^2 + i||x-iy||^2 - i||x+iy||^2)$
$= \frac{1}{4}(||x+y||^2 - ||x-y||^2) + \frac{1}{4}(i||x-iy||^2 - i||x+iy||^2)$
$= < x, y > + \frac{1}{4}(i\sqrt{< x-iy, x-iy >}^2 - i\sqrt{< x+iy, x+iy >}^2)$ by 3.1.i
$= < x, y > + \frac{1}{4}(i < x-iy, x-iy > - i < x+iy, x+iy >)$
$= < x, y > + \frac{1}{4}(i(< x, x-iy > + < -iy, x-iy >) - i(< x, x+iy > + < iy, x+iy >))$
$= < x, y > + \frac{1}{4}(i(< x, x > + < x, -iy > + < -iy, x > + < -iy, -iy >)$
$\quad - i(< x, x > + < x, iy > + < iy, x > + < iy, iy >))$
$= < x, y > + \frac{1}{4}(i(< x, x > + i < x, y > + i < y, x > - < y, y >)$
$\quad - i(< x, x > - i < x, y > - i < y, x > - < y, y >))$

$= < x, y > +\frac{1}{4}(i < x, x > - < x, y > - < y, x > -i < y, y >$
$\quad - i < x, x > + < x, y > + < y, x > +i < y, y >)$
$= < x, y > +\frac{1}{4}(- < x, y > + < x, y >)$
$= < x, y >$

Problem 3.3. Let θ be the angle in question. Recall that

$$cos(\theta) - \frac{\langle f, g\rangle}{||f|| \cdot ||g||}$$

- (i)

$$\langle f, g\rangle = \int_0^1 fgdx = \int_0^1 x^6 dx = \left(\frac{x^7}{7}\right)\Big|_0^1 = \frac{1}{7}$$
$$||f||^2 = \int_0^1 f^2 dx = \int_0^1 x^2 dx = \left(\frac{x^3}{3}\right)\Big|_0^1 = \frac{1}{3}$$
$$||g||^2 = \int_0^1 g^2 dx = \int_0^1 x^1 0 dx = \left(\frac{x^{11}}{11}\right)\Big|_0^1 = \frac{1}{11}$$

  Thus, we see that

$$cos(\theta) = \frac{\frac{1}{7}}{(\frac{1}{3} \cdot \frac{1}{11})^{\frac{1}{2}}} = \frac{33^{\frac{1}{2}}}{7} \tag{1}$$

  And (1) implies that $\theta \approx 35°$.

- (ii)

$$\langle f, g\rangle = \int_0^1 fgdx = \int_0^1 x^6 dx = \left(\frac{x^7}{7}\right)\Big|_0^1 = \frac{1}{7}$$
$$||f||^2 = \int_0^1 f^2 dx = \int_0^1 x^4 dx = \left(\frac{x^5}{8}\right)\Big|_0^1 = \frac{1}{5}$$
$$||g||^2 = \int_0^1 g^2 dx = \int_0^1 x^8 dx = \left(\frac{x^9}{9}\right)\Big|_0^1 = \frac{1}{9}$$

  Thus, we see that

$$cos(\theta) = \frac{\frac{1}{7}}{(\frac{1}{5} \cdot \frac{1}{9})^{\frac{1}{2}}} = \frac{45^{\frac{1}{2}}}{7} \tag{2}$$

  And (1) implies that $\theta \approx 17°$

Problem 3.8. There are four parts

- (i)

Proof. This is just a matter of checking all the relevant details.
Norms = 1:

$$||cos(t)||^2 = \frac{1}{\pi}\int_{-\pi}^{\pi} cos^2(t)dt = \frac{1}{\pi}\frac{cos(x)sin(x)+x}{2}\Big|_{-\pi}^{\pi} = 1$$
$$||sin(t)||^2 = \frac{1}{\pi}\int_{-\pi}^{\pi} sin^2(t)dt = \frac{1}{\pi}\frac{-sin(2x)+2x}{4}\Big|_{-\pi}^{\pi} = 1$$
$$||cos(2t)||^2 = \frac{1}{\pi}\int_{-\pi}^{\pi} cos^2(2t)dt = \frac{1}{\pi}\frac{sin(4x)+4x}{8}\Big|_{-\pi}^{\pi} = 1$$
$$||sin(2t)||^2 = \frac{1}{\pi}\int_{-\pi}^{\pi} sin^2(2t)dt = \frac{1}{\pi}\frac{-sin(4x)+4x}{8}\Big|_{-\pi}^{\pi} = 1$$

Inner Products = 1:

$$\langle cos(t), sin(t)\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} cos(t)sin(t)dt = \frac{sin^2(x)}{2}\Big|_{-\pi}^{\pi} = sin^2(\pi) - sin^2(-\pi) = 0$$

The proof is completed by checking all the other inner products similarly. □

- (ii)

$$||t||^2 = \int_{-\pi}^{\pi} t^2 dt = \frac{t^3}{3}\Big|_{-\pi}^{\pi} = \frac{\pi^3}{3} - \frac{(-\pi)^3}{3} = \frac{2\pi^3}{3}$$
$$||t|| = \left(\frac{2\pi^3}{3}\right)^{\frac{1}{2}}$$

- (iii)

$$\begin{aligned} proj_X(cos(3t)) &= \langle sin(t), cos(3t)\rangle sin(t) + \langle cos(t), cos(3t)\rangle cos(t) \\ &+ \langle sin(2t), cos(3t)\rangle sin(2t) + \langle cos(2t), cos(3t)\rangle cos(2t) \\ &= 0+0+0+0 = 0 \end{aligned}$$

(We see that $cos(3t)$ is orthogonal to X)

- (iv)

$$proj_X(t) = 0 + 2sin(t) + 0 - sin(2t)$$

Problem 3.9. *Proof.* In $\mathbb{R}^2$, a rotation about the origin by arbitrary angle θ can be described by the matrix

$$M = \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix}$$

So that $M\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \left(\begin{smallmatrix} xcos(\theta)-ysin(\theta) \\ xsin(\theta)+ycos(\theta) \end{smallmatrix}\right)$. Let $a = \left(\begin{smallmatrix} a_1 \\ a_2 \end{smallmatrix}\right), b = \left(\begin{smallmatrix} b_1 \\ b_2 \end{smallmatrix}\right) \in \mathbb{R}^2$. Then

$$\begin{aligned} \langle M(a), M(b)\rangle &= \left\langle \begin{pmatrix} a_1cos(\theta) - a_2sin(\theta) \\ a_1sin(\theta) + a_2cos(\theta) \end{pmatrix}, \begin{pmatrix} b_1cos(\theta) - b_2sin(\theta) \\ b_1sin(\theta) + b_2cos(\theta) \end{pmatrix} \right\rangle \\ &= a_1b_1cos^2(\theta) + a_2b_2sin^2(\theta) + a_1b_1sin^2(\theta) + a_2b_2cos^2(\theta \\ &= a_1b_1 + a_2b_2 = \langle a, b\rangle \end{aligned}$$

So we see that M is an orthonormal operator. (N.b that this is a terribly inefficient way to prove this - I should have just shown that the columns of M were orthonormal! $\square$

Problem 3.10. Recall that taking the Hermitian is flipping rows and columns and taking the conjugate.

- (i)

  Proof. We need to show both directions.
  $\Rightarrow$: Let Q be an orthonormal matrix. Then $\langle m, n\rangle = \langle Qm, Qn\rangle \implies m^H n = (Qm)^H Qn = m^H Q^H Qn$. And because m and n were arbitrarily chosen, the only way that this equality holds is if $Q^H Q = I$, and this gives us that $QQ^H = Q$ since left inverse $\implies$ right inverse (for square matrices).
  $\Leftarrow$: Let Q be a matrix so that $Q^H Q = I$. Then consider $\langle Qm, Qn\rangle = (Qm)^H Qn = m^H Q^H Qn = m^H n = \langle m, n\rangle$. $\square$

- (ii)

  Proof. This is pretty easy:

  $$||x|| = \sqrt[2]{\langle x, x\rangle} = \sqrt[2]{\langle x, x\rangle} = ||Qx||$$

  $\square$

- (iii)

  Proof. Assume Q is orthonormal. Then $QQ^H = Q^H Q = I \implies Q^H = Q^{-1}$. I'll prove the following short lemma.

  Lemma 0.1. For Q orthonormal, Q^H is orthonormal.

  Recall that $(Q^H)^H = Q$, and see that

  $$(Q^H)^H Q^H = Q^H (Q^H)^H = I$$

  which proves the lemma.
  And since $Q^{-1} = Q^H$, Q^{-1} is orthonormal. $\square$

- (iv)

  Proof. We'll examine the elements of the identity matrix element by element. First note that:

  $$I_{ij} = \delta_{ij} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

Then we'll compare this to what we get when we multiply QQ^H, which we know is equal to I in all its coordinates. First, though, for any matrix A, define A^i to be the "ith row" of A and A_j to be the jth column. Then

$$\delta_{ij} = (Q^H Q)_{ij} = (Q^H)^i Q_j =$$

Recall now that $(Q^H)^i = \bar{Q}_i$, by definition of the Hermitian. But now we see that

$$\langle \bar{Q}_i, q_j \rangle = \delta_{ij}$$

and the columns of Q are orthonormal. □

- (v)

  Proof. Consider an orthonormal matrix $Q \in M_n(\mathbb{R})$.

  $$\det(Q)\ \det(Q^H) = \ \det(QQ^H) = \ \det(I) = 1$$

  and $\det Q) = \det(Q^H) \implies \det(Q) = 1$ or -1.
  A counterexample to the converse would be the matrix $M = \left(\begin{smallmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{smallmatrix}\right)$. det $B = 1$, but $Be_1 = 2e_1$ and $||e_1|| \neq ||2e_1|| = ||Be_1||$ which violates what we proved in (ii). □

- (vi)

  Proof. This is also quite short:

  $$(Q_1Q_2)(Q_1Q_2)^H = Q_1Q_2Q_2^HQ_1^H = Q_1Q_1^H = I$$

  And we also get the left inverse by properties of inverses. □

Problem 3.11. Suppose that $x_1, ..., x_n$ is as et of linearly *dependent* vectors. Let's apply Gram-Schmidt. Eventually, we will arrive at a vector x_k which is linearly dependent upon $x_1, ..., x_{k-1}$. But then, if $X = span(x_1, ..., x_{k-1})$, then $x_k \in X$ and $p_{k-1} = proj_X(x_k) = x_k$, which forces $q_k = 0$. In the end, if we throw out all these zeroes, we still get an orthonormal basis $q_1, ...q_m$ of X where $m = \dim X$.

Problem 3.16. Solved this with Albi!

- (i) Let $A \in \mathbb{M}_{mxn}$ where rank$(A) = n \leq m$. Then there exist orthonormal $Q \in \mathbb{M}_{mxm}$ and upper triangular $R \in \mathbb{M}_{mxn}$ such that $A = QR$. Since $\tilde{Q} = -Q$ is still orthonormal $(-Q(-Q)^H = -Q(-Q^H) = QQ^H = I$ and similarly one shows $(-Q)^H(-Q) = I)$ and $\tilde{R} = -R$ is still upper triangular, $A = QR = \tilde{Q}\tilde{R}$. Therefore QR-decomposition is not unique.

- (ii) Suppose now that A is invertible and can be decomposed into two different QR decompositions: QR and $\tilde{Q}\tilde{R}$, where the diagonal entries of R and $\tilde{R}$ are strictly positive. Then both R and $\tilde{R}$ are invertible and we conclude that $\tilde{R}^{-1}R = Q^H\tilde{Q}$. Since R and $\tilde{R}$ are upper triangular, so is the LHS of the previous equation. On the other hand, since Q and $\tilde{Q}$ are orthonormal, so is the RHS. Therefore $\tilde{R}^{-1}R = I$ and, by unicity of the inverse, we conclude that $R = \tilde{R}$, and so $Q = \tilde{Q}$.

Problem 3.17. *Proof.*

$$\begin{aligned} A^H Ax &= Ab \\ \iff (\hat{Q}\hat{R})^H \hat{Q}\hat{R}x &= (\hat{Q}\hat{R})^H b \\ \iff (\hat{Q}\hat{R})^H \hat{Q}\hat{R}x & \\ \iff \hat{R}^H \hat{Q}^H \hat{Q}\hat{R}x &= \hat{R}^H \hat{Q}^H b \end{aligned}$$

Multiply both sides by $\hat{R}^{H^{-1}}$ and we see that it is equivalent to $\hat{R}x = \hat{Q}^H b$ ☐

Problem 3.23. *Proof.*

$$||x|| - ||y|| = ||x|| + || - y|| \leq ||x - y||$$

and

$$||y|| - ||x|| = ||y|| + || - x|| \leq ||y - x|| = ||x - y||$$

together imply $\big| ||x|| - ||y|| \big| \leq ||x - y||$ ☐

Problem 3.24. This is procedural, so I used Albi's latex code

(i) Since $|f(t)| \geq 0$ for every t, so is $\int_a^b |f(t)|dt$. In addition, if $f = 0$, then $\int_a^b |f(t)|dt = 0$. On the other hand, if $\int_a^b |f(t)|dt = 0$ and $|f(t)| \geq 0$, it must be that $|f(t)| = 0$ for all t, implying that $f = 0$. Now take a constant $c \in \mathbb{F}$, then $\int_a^b |cf(t)|dt = \int_a^b |c||f(t)|dt = |c| \int_a^b |f(t)|dt$, since c does not depend on t. Finally, take $g \in C([a, b]; \mathbb{F})$. Since $|f(t)+g(t)| \leq |f(t)|+|g(t)|$ for all t and the integral is a linear operator, we have that $\int_a^b |f(t) + g(t)|dt \leq \int_a^b |f(t)|dt + \int_a^b |g(t)|dt$.

(ii) Since $|f(t)|^2 \geq 0$ for every t, so is $\int_a^b |f(t)|^2 dt$ and its square root. In addition, if $f = 0$, then $|f(t)|^2 = 0$ for all t and $\sqrt{\int_a^b |f(t)|^2 dt} = 0$. On the other hand, if $\sqrt{\int_a^b |f(t)|^2 dt} = 0$, then $\int_a^b |f(t)|^2 dt = 0$ and since $|f(t)|^2 \geq 0$ for all t, it must be that $|f(t)|^2 = 0$ for all t, implying that $f = 0$. Now take a constant $c \in \mathbb{F}$, then $\sqrt{\int_a^b |cf(t)|^2 dt} = \sqrt{\int_a^b |c|^2 |f(t)|^2 dt} = |c|\sqrt{\int_a^b |f(t)|^2 dt}$, since c does not depend on t. Finally, take $g \in C([a, b]; \mathbb{F})$. Since $|f(t) + g(t)| \leq |f(t)| + |g(t)|$ for all t, $x \mapsto x^2$ and $x \mapsto \sqrt{x}$ are monotonically increasing for nonnegative x and the integral is a linear operator, we have that $\sqrt{\int_a^b |f(t) + g(t)|^2 dt} \leq \sqrt{\int_a^b |f(t)|^2 dt + \int_a^b |g(t)|^2 dt} \leq ||f||_{L2} + ||g||_{L2}$.

(iii) Since $|f(x)| \geq 0$ for all x, so is the $\sup_{x\in[a,b]} |f(x)|$. In addition, if $f = 0$, then $\sup_{x\in[a,b]} |f(x)|$ is also zero. On the other hand, since $|f(x)| \geq 0$ for all x, $0 \leq \sup_{x\in[a,b]} |f(x)| = 0$ implies that we must have $f = 0$. Now take a constant $c \in \mathbb{F}$, then $\sup_{x\in[a,b]} |cf(x)| = \sup_{x\in[a,b]} |c||f(x)| = |c| \sup_{x\in[a,b]} |f(x)|$. Finally, take $g \in C([a, b]; \mathbb{F})$. Since $|f(x) + g(x)| \leq |f(x)| + |g(x)|$ for all x, we have that $\sup_{x\in[a,b]} |f(x) + g(x)| \leq \sup_{x\in[a,b]}\{|f(x)| + |g(x)|\} \leq \sup_{x\in[a,b]} |f(x)| + \sup_{x\in[a,b]} |g(x)|$.

Problem 3.26. First I must show that topological equivalence is an equivalence relation.

Proof. I must show three things: (i) $x \sim x$. (ii) $x \sim y \implies y \sim x$, (iii) $x \sim y$ and $y \sim z \implies x \sim z$.
(i) $||\cdot||_1 \sim ||\cdot||_1$ trivially. Let $M \geq m$, then $m||x||_1 \leq ||x||_1 \leq M||x||_1$ for all x.
(ii) Also trivial: Suppose $||\cdot||_1 \sim ||\cdot||_2$. Then $m||x||_1 \leq ||x||_2 \leq M||x||_1$ for all x, which implies that $M^{-1}||x||_2 \leq ||x||_1 \leq m^{-1}||x||_2$.
(iii) Suppose $||\cdot||_1 \sim ||\cdot||_2$, and $||\cdot||_2 \sim ||\cdot||_3$. Then $m||x||_1 \leq ||x||_2 \leq M||x||_1$ and $n||x||_2 \leq ||x||_3 \leq N||x||_2$. But we get from this that $mn||x||_1 \leq ||x||_3 \leq MN||x||_1$. □

Now I'll show that the $1, 2$, and ∞ norms are topologically equivalent.

Proof. (i) $||\cdot||_1 \sim ||\cdot||_2$:
If we think about the inner product as the standard dot-product, then we have

$$(||x||_1)^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}|x_i||x_j| \geq \sum_{i=1}^{n} x_i^2 = \langle x, x, \rangle = (||x||_2)^2$$

(the inequality comes because we simply threw out some positive terms on the left side). This implies that $||x||_1 \geq ||x||_2$. Moreover,

$$\sum_{i=1}^{n}|x_i|\cdot 1 \leq \left(\sum_{i=1}^{n}|x_i|^2\right)^{1/2}\left(\sum_{i=1}^{n}1^2\right)^{1/2} = \sqrt{n}\left(\sum_{i=1}^{n}|x_i|^2\right)^{1/2}$$

so $||x||_2 \leq ||x||_1 \leq \sqrt{n}||x||_2$.
(ii) $||\cdot||_\infty \sim ||\cdot||_2$

$$||x||_\infty = max_{1\leq i\leq n}\{x_i\} = \sqrt[2]{(max_{1\leq i\leq n}\{x_i\})^2} \leq \sqrt[2]{\sum_{i=1}^{n} x_i} = ||x||_2$$

and

$$||x||_2^2 = \sum_{i=1}^{n}|x_i|^2 \leq nmax_i\{x_i\} = (\sqrt{n}||x|_\infty|)^2 \implies ||x||_2 = \sqrt{n}||x||_\infty$$

so $||x||\infty \leq ||x||_2 \leq \sqrt{n}||x||_\infty$ □

Problem 3.28. (Albi's latex code)
(i) Notice that (applying the results of the previous exercise)

$$\sup_{x\neq 0}\frac{||Ax||_1}{||x||_1} \leq \sup_{x\neq 0}\frac{||Ax||_1}{||x||_1} \leq \sqrt{n}\sup_{x\neq 0}\frac{||Ax||_2}{||x||_2},$$

and

$$\sup_{x\neq 0}\frac{||Ax||_1}{||x||_1} \geq \sup_{x\neq 0}\frac{||Ax||_2}{||x||_1} \geq \frac{1}{\sqrt{n}}\sup_{x\neq 0}\frac{||Ax||_2}{||x||_2}$$

imply that $\frac{1}{\sqrt{n}}||A||_2 \leq ||A||_1 \leq ||A||_2$.

(ii) Notice that

$$\sup_{x\neq 0}\frac{||Ax||_2}{||x||_2}\leq \sup_{x\neq 0}\frac{\sqrt{n}||Ax||_\infty}{||x||_\infty},$$

and

$$\sup_{x\neq 0}\frac{||Ax||_2}{||x||_2}\geq \sup_{x\neq 0}\frac{||Ax||_\infty}{\sqrt{n}||x||_\infty}.$$

Problem 3.29. I will prove two statements.

The norm of an orthonormal matrix is 1:

Proof. Let Q be an orthonormal matrix. Then

$$||Qx|| = ||x|| \implies sup_{x\neq 0}\frac{||Qx||}{||x||} = ||Q|| = 1$$

□

If $R_x : M_n(\mathbb{F}) \to \mathbb{F}, R_x(A) = Ax$, then $||R_x|| = ||x||$:

Proof. The first step is to show $||R_x|| < ||x||$.

$$||R_x|| = sup_{A\neq 0}\frac{||R_x(A)||}{||A||} = sup_{A\neq 0}\frac{||Ax||}{||A||} = sup_{A\neq 0}\frac{||Ax||\cdot||x||}{||A||\cdot||x||}$$

By Remark 3.5.12, $||Ax|| \leq ||A||\cdot||x||\forall x\in\mathbb{F}^n$, so

$$||R_x|| = sup_{A\neq 0}\frac{||Ax||\cdot||x||}{||A||\cdot||x||} \leq sup_{A\neq 0}\frac{||Ax||\cdot||x||}{||Ax||} = ||x||$$

Now I'll show equality. For the $\leq$ above to be strict, we must have $||Ax|| < ||A||\cdot||A||$ for all operators A (because we're taking the supremum). $||x|| > 0$, so I can rearrange for the condition:

$$\frac{||Ax||}{||x||} < ||A||, \text{ for all operators } A, \text{ vectors } x$$

In other words, no x achieves the supremum which is encoded in the definition of $||A||$. I will use the previous result to show that this will never hold.
Let $q_1 = e_1$ (or some other vector with norm 1). I can use the gram-schmidt algorithm to construct an orthonormal basis $q_1, ...q_n$ for $\mathbb{F}^n$. Let Q be the matrix with these basis vectors as its columns. Then Q is an orthonormal matrix. Specifically, $||Q|| = 1$ and it achieves $\frac{||Qx||}{||x||} = ||Q|| = 1$ at all nonzero x.
This shows that the inequality can never be strict, so we have $||R_x|| = ||x||$ □

Problem 3.30. *Proof.* To show something is a norm, I must show three properties:

- *Positivity:* This follows immediately from the positivity of the underlying matrix norm: $||A||_S = ||SAS^{-1}|| \geq 0$, with equality iff $SAS^{-1} = 0$, and there are no elements that are conjugate to 0 other than itself so $A = 0$ in this case.

- *Scalar Preservation:* I use linearity of S and the corresponding property for the matrix norm.
$$||kA||_S = ||SkAS^{-1}|| = ||kSAS^{-1}|| = k||SAS^{-1}|| = k||A||_S$$

- *Triangle Inequality:* This also follows from the linearity of S.
$$||(A+B)||_S = ||S(A+B)S^{-1}|| = ||SAS^{-1}+SBS^{-1}|| \leq ||||SAS^{-1}|+||SBS^{-1}|| = ||A||_S+||B||_S$$

Finally, to see that it is a matrix norm, I must show that it is submultiplicative. And indeed

$$||AB||_S = ||SABS^{-1}|| = ||SAS^{-1}SBS^{-1}|| \leq ||SAS^{-1}|| \cdot ||SBS^{-1}|| = ||A||_S \cdot ||B||_S$$

□

Problem 3.37. The first thing is to define the standard basis, which is $\mathcal{B} = \{1, x, x^2\}$ Evaluate L on the basis vectors:

$$L(1) = 0, L(x) = 1, L(x^2) = 2$$

Now, for $p \in V$, p can be written as a linear combination of these basis vectors. So

$$L(p) = L(a_1 1+a_2 x+a_3 x^2) = a_1 L(1)+a_2 L(x)+a_3 L(x^2) = \langle (L(1)\cdot 1, L(x), L(x^2)), (a_1, a_2, a_3)\rangle$$

which is the idea behind the Riesz Representation theorem. So we see that in this case, $q = (0, 1, 2)$... which squares with what we know about derivatives.

Problem 3.38. Let $\mathcal{B}$ as above.

$$D = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

By following the same method as exercise 3.7.9, I see that

$$D^* = -D = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -2 \\ 0 & 0 & 0 \end{bmatrix}$$

Problem 3.39. There are 4 things to show.

- (i)

  Proof.
$$\langle (S+T)v, w\rangle = \langle Sv, w\rangle + \langle Tv, w\rangle = \langle v, S^*w\rangle + \langle v, T^*w\rangle = \langle v, (S^* + T^*)w\rangle$$
$$\langle \alpha T^* v, w\rangle = \alpha\langle Tv, w\rangle = \alpha\langle v, T^*w\rangle = \langle v, \overline{\alpha}T^*w\rangle$$

  □

- (ii)

Proof.

$$\langle S^*v, w\rangle = \overline{\langle w, S^*v\rangle} = \overline{\langle Sw, v\rangle} = \langle v, Sw\rangle$$

□

- (iii)

Proof.

$$\langle STv, w\rangle = \langle Tv, S^*w\rangle = \langle v, T^*S^*w\rangle$$

□

- (iv)

Proof. Consider the composition $T^*(T^{-1})^*$.

$$\langle T^*(T^{-1})^*x, y\rangle = \langle (T^{-1})^*x, Ty\rangle = \langle x, (T^{-1})Ty\rangle = \langle x, y\rangle$$

Since the above is true for all x, y, we must have $T^*(T^{-1})^* = I$ □

Problem 3.40.

- (i) (Considering A as the operator)

$$\langle AB, C\rangle = \text{ tr } (AB)^H C = \text{ tr } B^H A^H C = \langle B, A^H C\rangle$$

- (ii)

$$\langle A_2, A_3A_1\rangle = \text{ tr}(A_2^H A_3 A_1) = \text{ tr}(A_1 A_2^H A_3) = \text{ tr}(A_2 A_1^H A_3) = \langle A_2 A_1^*, A_3\rangle$$

- (iii) (Albi) Given $B, C \in \mathbb{M}_n(\mathbb{F})$, we have $< B, AC - CA > = < B, AC > - < B, CA >$. Applying (ii) to the second term we get $< B, CA > = < BA^*, C >$. On the other hand,

$$< B, AC > = \text{tr}(B^H AC) = \text{tr}((A^H B)^H C) = < A^H B, C > = < A^* B, C > .$$

Putting all together we obtain that $T_A^* = T_{A^*}$

Problem 3.44. *Proof.* By the fundamental subspaces theorem, $\text{Ker}(A^H) = \text{Range}(A)$. So we can reformulate the second possibility to: there exists $y \in \text{ Range}(A)^\perp : \langle y, b\rangle \neq 0$. Consider now $p = proj_{\text{Range}(A)} b$. If $p = b$, then $b \in \text{Range}(A)$ and we have the first case. Otherwise, the procedure creates a residual vector r, $r = b - p$. $r \in \text{ Range}(A)^\perp$, and

$$\langle r, b\rangle = \overline{\langle p + r, r\rangle} = \overline{\langle p, r\rangle} + \overline{\langle r, r\rangle} = \langle p, b\rangle + \langle r, r\rangle = \langle r, r\rangle \neq 0$$

which is the second case. □

Problem 3.45. Props to Albi for this idea - I think this is really slick (especially the last part)!

Proof. Double inclusion. First show that $\text{Skew}_n(\mathbb{R}) \subset \text{Sym}_n^\perp(\mathbb{R})$.
Let $A \in Skew_n(\mathbb{R})$ Then, recalling the definitions of the spaces and the properties of trace,

$$\forall B \in \text{Sym}_n(\mathbb{R}),\ \langle A, B\rangle = \text{tr}(A^H B) = \text{tr}(-AB) = \text{tr}(-AB^H) = -\overline{\langle A, B\rangle}$$

But since we are over $\mathbb{R}$, $\langle A, B\rangle = -\overline{\langle A, B\rangle} \implies \langle A, B\rangle = 0$ for all $B \in \text{Sym}_n(\mathbb{R})$, so we see that $A \in \text{Sym}_n(\mathbb{R})^\perp$
Now I'll show that $\text{Sym}_n^\perp(\mathbb{R}) \subset \text{Skew}_n(\mathbb{R})$.
Let $B \in \text{Sym}_n(\mathbb{R})^\perp$. Then for any $A \in \text{Sym}_n(\mathbb{R})$, We will examine $\langle B + B^T, A\rangle$.

$$\langle B + B^T, A\rangle = \langle B, A\rangle + \langle B^T, A\rangle = 0 + \langle B^T, A\rangle$$

and

$$\langle B^T, A\rangle = \text{tr}(BA) = \text{tr}(BA^T) = \text{tr}(A^T B) = \text{tr}(B^T A) = \langle B, A\rangle = 0$$

so $\langle B + B^T, A\rangle = 0$ for all $A \in \text{Sym}_n(\mathbb{R})$. But $B + B^T \in \text{Sym}_n(\mathbb{R})$, so $||B + B^T|| = 0 \implies B + B^T = 0 \implies B^T = -B$ $\square$

Problem 3.46. Four Parts

Proof.

- (i) $Ax \in \text{Range}(A)$ by definition, and $x \in \text{Ker}(A^H A) \implies A^H(Ax) = 0 \implies Ax \in \text{Ker}A^H$.
- (ii) Clearly $\text{Ker}(A) \subset \text{Ker}(A^H A)$ since $A^H(0) = 0$. It remains to show that

  $$A^H Ax = 0 \implies Ax = 0$$

  Suppose $A^H Ax = 0$. Then if we apply the operator x^T as a row vector, we see that

  $$x^H A^H Ax = (Ax)^H Ax = \langle Ax, Ax\rangle = 0$$

  and $\langle Ax, Ax\rangle = 0 \implies Ax = 0$ which is what we wanted to show.
- (iii) A and $A^H A$ both map to the n-dimensional spaces and $\text{Ker}(A) = \text{Ker}(A^H A)$ by the above, so by Rank-Nullity,

  $$n - \dim \text{Ker}(A) = \dim \text{Range}(A)$$
  $$n - \dim \text{Ker}(A^H A) = \dim \text{Range}(A^H A)$$

  The left sides of the two equations are equal so the right sides must also be equal!
- (iv) A has linearly independent columns $\implies$
  A has rank $n \implies$
  $A^H A$ has (full) rank $n \implies$
  $A^H A$ is non singular.

$\square$

Problem 3.47. *Proof.*

- (i)

$$P^2 = (A(A^H A^{-1})A^H)^2 = A(A^H A^{-1})A^H A(A^H A^{-1})A^H = AA^H AA^H = A(A^H A^{-1})A$$

- (ii)

$$\begin{aligned} P^H = (A(A^H A^{-1})A^H)^H &= ((A(A^H))((A^{-1})A^H))^H \\ &= ((A^{-1})A^H))^H((A(A^H))^H \\ &= A(A^{-1})^H AA^H \end{aligned}$$

- (iii) Er... I'm not sure this is even true. Suppose $m < n$, then P is m by m and cannot possibly have rank n? What am I missing here?

$\square$

Problem 3.48. SO MANY PARTS AAAH!

Proof.
- (i) I'll do it all at once... let $k \in \mathbb{R}, A, B \in M_n(\mathbb{R})$

$$\begin{aligned} P(k(A+B)) &= \frac{(k(A+B)) + (k(A+B))^T}{2} \\ &= \frac{k(A+B) + (k(A^T + B^T)}{2} \\ &= \frac{k(A + A^T + B + B^T)}{2} \\ &= k(P(A) + P(B)) \end{aligned}$$

- (ii)

$$P^2(A) = \frac{P(A) + P(A)^T}{2} = \frac{\frac{A+A^T}{2} + \frac{A+A^T}{2}}{2} = \frac{A + A^T}{2} = P(A)$$

- (iii) First see that $P(A) = P(A^T)$ (this is trivial). Then

$$\begin{aligned} \langle P(A), B\rangle = \ \operatorname{tr}(P(A)^T B) = \ \operatorname{tr}(\frac{A+A^T}{2} \cdot B) = \frac{\operatorname{tr}(A^T B + AB)}{2} = \operatorname{tr}(AB) \\ = \frac{\operatorname{tr}(AB + AB^T)}{2} = \ \operatorname{tr}(A \cdot \frac{B + B^T}{2}) = \ \operatorname{tr}(AP(B)) = \langle A, P(B)\rangle \end{aligned}$$

- (iv)

$$A \in \ \operatorname{Ker}(P) \iff P(A) = 0 \iff A + A^T = 0 \iff A = -A^T \iff A \in \ \operatorname{Skew}_n(\mathbb{R})$$

- (v)

$$\begin{aligned} A \in \ \operatorname{Range}(P) &\iff \exists B : A = P(B) \\ &\iff \exists B : B + B^T = 2A \\ &\iff 2A \in \ \operatorname{Sym}_n(\mathbb{R}) \\ &\iff A \in \ \operatorname{Sym}_n(\mathbb{R}) \end{aligned}$$

- (vi) Copied latex code from Albi as this is just long and mechanical.

$$||A - P(A)||_F^2 =< A - P(A), A - P(A) >=< A - \frac{A + A^T}{2}, A - \frac{A + A^T}{2} >=$$
$$< \frac{A - A^T}{2}, \frac{A - A^T}{2} >= \text{Tr}\left(\left(\frac{A - A^T}{2}\right)^T \frac{A - A^T}{2}\right) =$$
$$\text{Tr}\left(\frac{A^T - A}{2}\frac{A - A^T}{2}\right) = \text{Tr}\left(\frac{A^TA - A^2 - (A^T)^2 + AA^T}{4}\right) =$$
$$\text{Tr}\left(\frac{A^TA - A^2 - A^2 + A^TA}{4}\right) = \text{Tr}\left(\frac{A^TA - A^2}{2}\right) = \frac{\text{Tr}(A^TA) - \text{Tr}(A^2)}{2}.$$

$\square$

Problem 3.50. I want to estimate the least squares solution for $Ax = b$ where:

$$A = \begin{bmatrix} x_1^2 & y_1^2 \\ . & . \\ . & . \\ . & . \\ x_n^2 & y_n^2 \end{bmatrix}, x = \begin{bmatrix} r \\ s \end{bmatrix}, b = \begin{bmatrix} 1 \\ . \\ . \\ . \\ 1 \end{bmatrix}$$

The normal equation to solve is:

$$\begin{bmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i y_i & \sum_{i=1}^n y_i^2 \end{bmatrix} \begin{bmatrix} r \\ s \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_i \\ \sum_{i=1}^n y_i \end{bmatrix}$$