Math, Problem Set #2, Inner Product Spaces

OSM Lab, Zach Boyd

Due Monday, July 5 at 8:00am

3.1

(i) $\langle x, y \rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2)$

Proof $\frac{1}{4}(\|x+y\|^2 - \|x-y\|^2)$

$$= \frac{1}{4}(\langle x+y, x+y \rangle - \langle x-y, x-y \rangle) = \frac{1}{4}(\langle x+y, x \rangle + \langle x+y, y \rangle - (\langle x-y, x \rangle - \langle x-y, y \rangle))$$

$$= \frac{1}{4}(\langle x, x \rangle + \langle y, y \rangle + \langle y, x \rangle + \langle y, y \rangle - (\langle x, x \rangle - \langle x, y \rangle - \langle y, x \rangle + \langle y, y \rangle))$$

$$\frac{1}{4}(4 \cdot \langle x, y \rangle) = \langle x, y \rangle \qquad \square$$

(ii) $\|x\|^2 + \|y\|^2 = \frac{1}{2}(\|x+y\|^2 + \|x-y\|^2)$

Proof Replace the middle minus with a plus above, and observe:

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

$$= \frac{1}{2}(2\langle x, x \rangle + 2\langle y, y \rangle) = \langle x, x \rangle + \langle y, y \rangle = \|x\|^2 + \|y\|^2$$

$\square$

3.2 $\langle x, y \rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2)$

Proof Using 1(a), we observe that we already showed that $\langle x, y \rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2 + 2\langle x, y \rangle - 2\langle y, x \rangle)$ (we could simplify thanks to the identity $\|x+iy\|^2 - \|x-iy\|^2 = 0$.

$$\|x+iy\|^2 - \|x-iy\|^2 = \langle x+iy, x+iy \rangle - \langle x-iy, x-iy \rangle$$

$$= \frac{1}{4}(\langle x+iy, x \rangle + i\langle x+iy, y \rangle - (\langle x-iy, x \rangle - i\langle x-iy, y \rangle))$$

$$= \frac{1}{4}(\langle x, x \rangle - i\langle y, x \rangle + i\langle x, y \rangle + \langle y, y \rangle - (\langle x, x \rangle + i\langle y, x \rangle - i\langle x, y \rangle + \langle y, y \rangle))$$

$$= \frac{1}{4}(-2i\langle y, x \rangle + 2i\langle x, y \rangle)$$

Thus, we are able to conclude that

$$\frac{1}{4}(\|x+y\|^2 - \|x-y\|^2 + i\|x+iy\|^2 - i\|x-iy\|^2)$$

$$= \langle x, y \rangle + \frac{1}{4}(2\langle x, y \rangle - 2\langle y, x \rangle + i(-2i\langle y, x \rangle + 2i\langle x, y \rangle)) = \langle x, y \rangle$$

□

3.3 Let $\mathbb{R}[x]$ have the inner product $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$. Find the angle between:

(i) x and x^5:

Solution From $cos(\theta) = \frac{\langle f, g \rangle}{\|f\|\|g\|}$, we observe that

$$\langle f, g \rangle = \langle x^k, x^l \rangle = \int_0^1 x^{k+l}dx = \frac{1}{k+l+1}$$

and

$$\|f\|^2 = \langle f, f \rangle = \langle x^k, x^k \rangle = \int_0^1 x^{2k}dx = \frac{1}{2k+1}$$

Hence for (i),

$$\frac{\langle f, g \rangle}{\|f\|\|g\|} = \frac{\langle x, x^5 \rangle}{\|x\|\|x^5\|} = \frac{\frac{1}{7}}{\sqrt{\frac{1}{3}\frac{1}{11}}} = \frac{\sqrt{33}}{7} \approx 0.82065 \Leftrightarrow \theta = 0.608 = 34.85°$$

□

(ii) x^2 and x^4:

Solution

$$\frac{\langle f, g \rangle}{\|f\|\|g\|} = \frac{\langle x^2, x^4 \rangle}{\|x^2\|\|x^4\|} = \frac{\frac{1}{7}}{\sqrt{\frac{1}{5}\frac{1}{9}}} = \frac{3\sqrt{5}}{7} \approx 0.9583 \Leftrightarrow \theta = 0.290 = 16.6°$$

□

3.8 Let V be the inner product space $C([-\pi, \pi]; \mathbb{R})$ with inner product

$$\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t)g(t)dt$$

Let $X = span(S) \subset V$, where $S = cos(t), sin(t), cos(2t), sin(2t)$.

(i) Prove that S is an orthonormal set.

Proof

$$\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} sin(kt)cos(lt)dt = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{e^{ikt} - e^{-ikt}}{2i} \frac{e^{ilt} + e^{-ilt}}{2} dt$$

$$= \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{e^{i(k+l)t} - e^{i(l-k)t} + e^{i(k-l)t} - e^{-i(k+l)t}}{4i} dt$$

$$= \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{e^{i(k+l)t} - e^{-i(k+l)t} + e^{i(k-l)t} - e^{-i(k-l)t}}{4i} dt = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{sin((k+l)t) + sin((k-l)t)}{2} dt$$

$$= 0 \qquad \text{since sin is an odd function}$$

$$\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} sin(kt)sin(lt)dt = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{e^{ikt} - e^{-ikt}}{2i} \frac{e^{ilt} - e^{-ilt}}{2i} dt$$

$$= \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{e^{i(k+l)t} - e^{-i(k-l)t} - e^{i(k-l)t} + e^{-i(k+l)t}}{-4} dt = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{cos((k+l)t) - cos((k-l)t)}{-2}$$

$$= \begin{cases} 1 & |k| = |l| \neq 0 \\ 0 & |k| \neq |l| \quad \text{and} \quad |k| = |l| = 0 \end{cases} \qquad \text{since } sin(m\pi) = 0 \text{ unless } m = 0$$

$$\langle f, g \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} cos(kt)cos(lt)dt = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{e^{ikt} + e^{-ikt}}{2} \frac{e^{ilt} + e^{-ilt}}{2} dt$$

$$= \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{e^{i(k+l)t} + e^{-i(k-l)t} + e^{i(k-l)t} + e^{-i(k+l)t}}{4} dt = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{cos((k+l)t) + cos((k-l)t)}{2}$$

$$= \begin{cases} 1 & |k| = |l| \neq 0 \\ 0 & |k| \neq |l| \\ 2 & |k| = |l| = 0 \end{cases} \qquad \text{since } sin(m\pi) = 0 \text{ unless } m = 0$$

Lastly,

$$\frac{1}{\pi} \int_{-\pi}^{\pi} (\frac{e^{ikt} \pm e^{-ikt}}{2a_{-i,1}i})^2 = \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{e^{2ikt} \pm 2 + e^{-2ikt}}{\pm 4}$$

$$= \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{1}{2} + \frac{cos(2kt)}{2} = 1$$

Hence, S is an orthonormal set. □

(ii) Compute $\|t\|$.

Computation $\|t\|^2 = \langle t, t \rangle = \frac{1}{\pi} \int_{-\pi}^{\pi} t^2 dt = \frac{1}{\pi}(\frac{\pi^3}{3} - \frac{(-\pi)^3}{3}) = \frac{2\pi^2}{3}$

Hence $\|t\| = \sqrt{\frac{2}{3}}\pi$ □

(iii) Compute the projection $proj_X(cos(3t))$.

Computation $proj_X(h) = \sum_{i=1}^{m}\langle f_i, h\rangle f_i$ with $f_i \in S$. From (i), we observe that $\langle sin(kt), cos(3t)\rangle = 0$, and $\langle cos(kt), cos(3t)\rangle = 0$. Hence, cos(3t) is orthogonal to the entire set. $\square$

(iv) Compute the projection $proj_X(t)$.

Computation $proj_X(h) = \sum_{i=1}^{m}\langle f_i, h\rangle f_i$ with $f_i \in S$.

$$\langle cos(kt), t\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} cos(kt)tdt = [(\frac{1}{k\pi}tsin(kt))|_{-\pi}^{\pi}] - \frac{1}{k\pi}\int_{-\pi}^{\pi} sin(kt)dt = 0$$

and

$$\langle sin(kt), t\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} sin(kt)tdt = [(\frac{-1}{k\pi}tcos(kt))|_{-\pi}^{\pi}] - \frac{1}{k\pi}\int_{-\pi}^{\pi} -cos(kt)dt$$

$$= \frac{-2cos(k\pi)}{k} = \frac{2(-1)^{k+1}}{k}$$

Thus $proj_X(t) = \sum_{i=1}^{m}\langle f_i, t\rangle f_i = 2sin(t) - sin(2t)$. $\square$

3.9 Prove that a rotation in $\mathbb{R}^2$ is an orthonormal transformation with respect to the usual inner product.

Proof We recall that any orthonormal transformation in $\mathbb{R}^2$ is of the form

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

Now observe that $\langle R_\theta x, R_\theta y\rangle = \langle x, (R_\theta)^H R_\theta y\rangle$ and $R_\theta^H R_\theta = \begin{bmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{bmatrix}$.
$\begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ by recognizing that $cos^2(\theta) + sin^2(\theta) = 1$ $\square$

3.10 Prove the following:

(i) The matrix $Q \in M_n(\mathbb{F})$ is an orthonormal matrix if and only if $QQ^H = Q^H Q = I$:

Proof From our definition of orthonormal transformations, we know that for all x,

$$\langle x, x \rangle = \langle Qx, Qx \rangle = (Qx)^H Qx = x^H Q^H Qx = \langle QQ^H x, x \rangle = x^H (Q^H Q)x = \langle x, Q^H Qx \rangle$$

Since

$$\langle x, x \rangle = \langle QQ^H x, x \rangle = \langle x, Q^H Qx \rangle \quad \forall \quad x,$$

$$QQ^H x = Q^H Qx = x \quad , \text{i.e.} \quad QQ^H = Q^H Q = I.$$

By this chain of equalities, Q is an orthonormal matrix if and only if $QQ^H = Q^H Q = I$ □

(ii) If $Q \in M_n(\mathbb{F})$ is an orthonormal matrix, then $\|Qx\| = \|x\|$ for all $x \in \mathbb{F}^n$.

Proof By orthonormality, we have $\langle x, x \rangle = \langle Qx, Qx \rangle$, and thus $\|Qx\|^2 = \langle x, x \rangle = \langle Qx, Qx \rangle = \|x\|^2$, i.e. $\|Qx\| = \sqrt{\langle x, x \rangle} = \sqrt{\langle Qx, Qx \rangle} = \|x\|$ □

(iii) If $Q \in M_n(\mathbb{F})$ is an orthonormal matrix, then so is Q^{-1}.

Proof From (i), we see that $Q^{-1} = Q^H$. We can easily see Q^H is orthonormal by observing that

$$\langle Q^{-1}x, Q^{-1}x \rangle = \langle Q^H x, Q^H x \rangle = (Q^H x)^H Q^H x = x^H Q Q^H x = x^H x = \langle x, x \rangle □$$

(iv) The columns of an orthonormal matrix $Q \in M_n(\mathbb{F})$ are orthonormal.

Proof Note $Q = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \dots & v_n \\ | & | & & | \end{bmatrix}$, where $\{v_1, v_2, \cdots, v_n\}$ is an orthonormal set of vectors. Then

$$I = Q^H Q = \begin{bmatrix} - & \overline{v_1^T} & - \\ - & \overline{v_2^T} & - \\ & \vdots & \\ - & \overline{v_n^T} & - \end{bmatrix} \cdot \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \dots & v_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \overline{v_1^T}v_1 & \overline{v_1^T}v_2 & \dots & \overline{v_1^T}v_n \\ \overline{v_2^T}v_1 & \overline{v_2^T}v_2 & \dots & \overline{v_2^T}v_n \\ \vdots & \vdots & \ddots & \vdots \\ \overline{v_n^T}v_1 & \overline{v_n^T}v_2 & \dots & \overline{v_n^T}v_n \end{bmatrix},$$

thus $\overline{v_k^T}v_l = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$, i.e. the columns are orthonormal.

(v) If $Q \in M_n(\mathbb{F})$ is an orthonormal matrix, then $\|det(Q)\| = 1$ so is Q^{-1}. Is the converse true?

Proof Observe that $1 = det(I) = det(QQ^{-1}) = det(Q)det(Q^H)$
$= det(Q)det(conj(Q)) = det(Q)^2 = |det(Q)|^2$, where we used that the determinant of a matrix and its transpose are the same (Thm 2.8.21), and that taking the complex conjugate of all entries does not change the determinant. Hence $|det(Q)| = |det(Q^{-1})| = 1$.

The converse is not true: take $Q = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Then $det(Q) = 1$ but Q is clearly not orthonormal.

(vi) If $Q_1, Q_2 \in M_n(\mathbb{F})$ are orthonormal matrices, then the product Q_1Q_2 is also an orthonormal matrix.

Proof Observe $\langle Q_1Q_2x, Q_1Q_2x \rangle = \langle Q_1(Q_2x), Q_1(Q_2x) \rangle = \langle Q_2x, Q_2x \rangle = \langle x, x \rangle$ by orthonormality. $\square$

3.11 Describe what happens when we apply the Gram-Schmidt orthonormalization process to a collection of linearly *dependent* vectors.

Answer Once we get to a vector that is in the span of the previous vectors, its projection onto the previous vector's span is just itself. Hence $v - proj(v) = 0$, hence it is not contributing to the orthonormal basis, (and we cannot normalize), so we just move on to the next vector. If we did not skip the vector, the algorithm would break down since we would try to divide by zero when normalizing.

3.16 Prove the following results about the QR decomposition:

(i) The QR decomposition is not unique.

Example As the hint suggests, take $D = D^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Suppose $A = QR$ in a QR-decomposition. Then $A = QR = QDD^{-1}R = (QD)(QR) = Q'R'$, which is still a QR-decomposition, since multiplying the second column by -1 does not affect orthonormality of the columns of Q, and R remains upper triangular, with the only change that $R'_{22} = -R_{22}$ $\square$

(ii) If A is invertible, then there is a unique QR decomposition of A such that R has only positive diagonal elements.

Proof Suppose $QR = Q'R' = A$ are two QR decompositions such that all diagonal entries of R, R' are positive. Then since A is invertible, and Q, Q' are orthonormal, R, R' must be invertible, and thus $D = R'^{-1}R = Q^{-1}Q'$. Note D must be upper-triangular, with positive entries along the diagonal. At

the same time, by Exercise 3.11(vi), it must be orthonormal, and this (upper-triangular and orthonormal) can only be the case if $D = I$. From Exercise 3.11, it follows that $Q'^{-1} = Q^{-1}$, and by uniqueness of inverses, $Q = Q'$. Thus $R' = Q'^{-1}A = Q^{-1}A = R$, i.e. the decomposition is unique. $\square$

3.17 Let $A \in M_{m \times n}$ have rank $n \leq m$, and let $A = \hat{Q}\hat{R}$ be a reduced QR decomposition. Prove that solving the system $A^H A x = A^H b$ is equivalent to solving the system $\hat{R}x = \hat{Q}^H b$.

Proof Note $A^H A x = (\hat{Q}\hat{R})^H \hat{Q}\hat{R}x = \hat{R}^H \hat{Q}^H \hat{Q}\hat{R}x = \hat{R}^H \hat{R}x$ by orthonormality, and $A^H b = (\hat{Q}\hat{R})^H b = \hat{R}^H \hat{Q}^H b$. In the reduced QR-decomposition, $\hat{R}$ is invertible, hence $\hat{R}x = (\hat{R}^H)^{-1}\hat{R}^H \hat{R}x = (\hat{R}^H)^{-1}\hat{Q}^H b = (\hat{R}^H)^{-1}\hat{R}^H \hat{Q}^H b = \hat{Q}^H b$ $\square$

3.23 Let $(V, \|\cdot\|)$ be a normed linear space. Prove that $|\|x\| - \|y\|| \leq \|x - y\|$ for all $x, y \in V$

Proof The result follows by triangle inequality in both cases:

$$|\|x\| - \|y\|| = \begin{cases} \|x\| - \|y\| \leq \|x - y\| + \|y\| - \|y\| = \|x - y\| & \|x\| \geq \|y\| \\ \|y\| - \|x\| \leq \|y - x\| + \|x\| - \|x\| = \|y - x\| = \|x - y\| & \|x\| \geq \|y\| \end{cases}$$

$\square$

3.24 Let $C([a, b]; \mathbb{F})$ be the vector space of all continuous functions from $[a, b] \subset \mathbb{R}$ to $\mathbb{F}$. Prove that each of the following is a norm on $C([a, b]; \mathbb{F})$.

Remark A norm has to satisfy the following:

1. Positivity: $\|f\| \geq 0$ with equality iff $f = 0$

2. Scale Preservation: $\|\alpha f\| = |\alpha|\|f\|$

3. Triangle inequality: $\|f + g\| \leq \|f\| + \|g\|$

(i) $\|f\|_{L^1} = \int_a^b |f(t)| dt$

   Proof

   (a) $|f(t)| \geq 0$ by definition, and the integral of a nonnegative function is nonnegative. $\int_a^b |0| dt = 0$. To show the 'only if', suppose a function is not zero anywhere on the interval, say at x_0 it takes the value with $y_0 = |f(x_0)|$. Then by continuity, for $\frac{y_0}{2}$, there exists $\delta > 0$ such that $||f(x)| - y_0| < \frac{y_0}{2}$. Then

   $$\|f\|_{L^1} = \int_a^b |f(t)| \geq \int_{max(a,x_0-\delta)}^{min(b,x_0+\delta)} \frac{y_0}{2} dt \geq (min(b, x_0+\delta) - max(a, x_0-\delta))\frac{y_0}{2} > 0$$

(b) $\|\alpha f\| = \int_a^b |\alpha f(t)| dt = |\alpha| \int_a^b |f(t)| dt = |\alpha| \|f\|$

(c) $\|f + g\| = \int_a^b |f(t) + g(t)| dt \le \int_a^b |f(t)| + |g(t)| dt = \|f\| + \|g\|$

Thus, $\|f\|_{L^1}$ is a valid norm. $\square$

(ii) $\|f\|_{L^2} = (\int_a^b |f(t)|^2 dt)^{\frac{1}{2}}$

Proof

(a) $|f(t)| \ge 0$ by definition, and the integral of a nonnegative function is nonnegative. $\int_a^b |0^2| dt = 0$. To show the 'only if', suppose a function is not zero anywhere on the interval, say at x_0 it takes the value with $y_0 = |f^2(x_0)|$. Then by continuity, for $\frac{y_0}{2}$, there exists $\delta > 0$ such that $||f^2(x)| - y_0| < \frac{y_0}{2}$. Then

$$\|f\|_{L^2}^2 = \int_a^b |f^2(t)| dt \ge \int_{max(a,x_0-\delta)}^{min(b,x_0+\delta)} \frac{y_0}{2} dt$$

$$\ge (min(b, x_0 + \delta) - max(a, x_0 - \delta)) \frac{y_0}{2} > 0$$

(b) $\|\alpha f\| = (\int_a^b |\alpha f(t)|^2 dt)^{\frac{1}{2}} = (\int_a^b |\alpha|^2 |f(t)|^2 dt)^{\frac{1}{2}} = |\alpha| (\int_a^b |f(t)|^2 dt)^{\frac{1}{2}} = |\alpha| \|f\|$

(c) $\|f + g\|^2 = \int_a^b |f(t) + g(t)|^2 dt \le \int_a^b |f(t)|^2 + |g(t)|^2 dt = \|f\| + \|g\|$

Thus, $\|f\|_{L^2}$ is a valid norm. $\square$

(iii) $\|f\|_{L^\infty} = sup_{x \in [a,b]} |f(x)|$

Proof

(a) $|f(t)| \ge 0$ by definition, and the sup detects the highest value, such that, unless f $= 0$ on [a, b], the inequality is strict.

(b) $\|\alpha f\| = sup_{x \in [a,b]} |\alpha f(x)| = |\alpha| sup_{x \in [a,b]} |f(x)| = |\alpha| \|f\|$

(c) $\|f + g\| = sup_{x \in [a,b]} |f(x) + g(x)|$

$$\le sup_{x \in [a,b]} |f(x)| + sup_{x \in [a,b]} |g(x)| = \|f\| + \|g\|$$

Thus, $\|f\|_{L^\infty}$ is a valid norm. $\square$

3.26 Two norms $\| \cdot \|_a$ and $\| \cdot \|_b$ on the vector space X are *topologically equivalent* if there exist constants $0 < m \le M$ such that

$$m \| \cdot \|_a \le \| \cdot \|_b \le M \| \cdot \|_a, \quad \forall x \in X$$

Prove that topological equivalence is an equivalence relation. Then prove that the p-norms for $p = 1, 2, \infty$ on $\mathbb{F}^n$ are topologically equivalent by establishing the following inequalities:

Proof: Equivalence Relation

1. Topological equivalence is reflexive: $1 \cdot \|x\|_a \leq \|x\|_b \leq 1 \cdot \|x\|_a$ (m = M = 1), for all x.

2. Topological equivalence is symmetric:

   Suppose $\|\cdot\|_a \sim \|\cdot\|_b$, then $m\|\cdot\|_a \leq \|\cdot\|_b \leq M\|\cdot\|_a, \quad \forall x \in X$, but then clearly also $\frac{1}{M}\|\cdot\|_b \leq \|\cdot\|_a \leq \frac{1}{m}\|\cdot\|_b \leq, \quad \forall x \in X$.

3. Topological equivalence is transitive: Suppose $\|\cdot\|_a \sim \|\cdot\|_b$ and $\|\cdot\|_b \sim \|\cdot\|_c$, then there are m, M such that $m\|\cdot\|_a \leq \|\cdot\|_b \leq M\|\cdot\|_a, \quad \forall x \in X$, and m', M' such that $m\|\cdot\|_b \leq \|\cdot\|_c \leq M\|\cdot\|_b, \quad \forall x \in X$, and thus mm', MM' such that $mm'\|\cdot\|_a \leq \|\cdot\|_c \leq MM'\|\cdot\|_a, \quad \forall x \in X$. Hence topological equivalence is an equivalence relation. $\square$

(i) $\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2$

   Proof Note $\|x\|_2^2 = \sum_{i=1}^n |x_i|^2 \leq (\sum_{i=1}^n |x_i|)^2 = \|x\|_1^2$ Further, Cauchy-Schwarz shows that $\|x\|_1 = \sum_{i=1}^n |x_i| = \sum_{i=1}^n 1 \cdot |x_i| \leq (\sum_{i=1}^n 1^2)^{\frac{1}{2}}(\sum_{i=1}^n |x_i|^2)^{\frac{1}{2}} = \sqrt{n}\|x\|_2$
   $\square$

(ii) $\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty$

   Proof $\|x\|_\infty = sup_i\{|x_i|\} = ((sup_i\{|x_i|\})^2)^{\frac{1}{2}} \leq (\sum_{i=1}^n |x_i|^2)^{\frac{1}{2}} = \|x\|_2$

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

   $\square$

3.28 Let A be an $n \times n$ matrix. Prove that the operator p-norms are topologically equivalent for $p = 1, 2, \infty$ by establishing the following inequalities:

(i) $\frac{1}{\sqrt{n}}\|A\|_2 \leq \|A\|_1 \leq \sqrt{n}\|A\|_2$

   Proof $\frac{1}{\sqrt{n}}\|A\|_2 = \frac{1}{\sqrt{n}}sup_{x\neq 0}(\frac{\|Ax\|_2}{\|x\|_2}) \leq sup_{x\neq 0}(\frac{\|Ax\|_1}{\sqrt{n}\|x\|_2}) \leq sup_{x\neq 0}(\frac{\|Ax\|_1}{\|x\|_1}) = \|A\|_1$

   $$\|A\|_1 = sup_{x\neq 0}(\frac{\|Ax\|_1}{\|x\|_1}) \leq sup_{x\neq 0}(\sqrt{n}\frac{\|Ax\|_2}{\|x\|_1}) \leq \sqrt{n}sup_{x\neq 0}(\frac{\|Ax\|_2}{\|x\|_2}) = \sqrt{n}\|A\|_2$$

   where we used 3.26(i) to justify all the inequalities. $\square$

(ii) $\frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2 \leq \sqrt{n}\|A\|_\infty$

Proof $\frac{1}{\sqrt{n}}\|A\|_\infty = \frac{1}{\sqrt{n}}sup_{x\neq 0}(\frac{\|Ax\|_\infty}{\|x\|_\infty}) \leq sup_{x\neq 0}(\frac{\|Ax\|_2}{\sqrt{n}\|x\|_\infty}) \leq sup_{x\neq 0}(\frac{\|Ax\|_2}{\|x\|_2}) = \|A\|_2$

$$\|A\|_2 = sup_{x\neq 0}(\frac{\|Ax\|_2}{\|x\|_2}) \leq sup_{x\neq 0}(\sqrt{n}\frac{\|Ax\|_\infty}{\|x\|_2}) \leq \sqrt{n}sup_{x\neq 0}(\frac{\|Ax\|_\infty}{\|x\|_\infty}) = \sqrt{n}\|A\|_\infty$$

where we used 3.26(ii) to justify all the inequalities. $\square$

3.29 Take $\mathbb{F}^n$ with the 2-norm, and let the norm on $M_n(\mathbb{F})$ be the corresponding induced norm. Prove that any orthonormal matrix $Q \in M_n(\mathbb{F})$ has $\|Q\| = 1$.

Proof $\|Q\| = sup_{x\neq 0}(\frac{\|Qx\|_2}{\|x\|_2}) = sup_{x=1}(\|Qx\|_2) = sup_{x=1}(\|Q\sum_{i=1}^n \langle q_i, x\rangle q_i\|_2) = 1$, by orthonormality of Q and diadic expansion.

3.29 (ctd.) For any $x \in \mathbb{F}^n$, let $R_x : M_n(\mathbb{F}) \to \mathbb{F}^n$ be the linear transformation $A \mapsto Ax$. Prove that the induced norm of the transformation R_x is equal to $\|x\|_2$.

Proof $\|R_x\| = sup_{x\neq 0}(\frac{\|Ax\|}{\|A\|}) = sup_{x\neq 0}(\frac{\|Ax\|\|x\|}{\|A\|\|x\|}) \leq (\frac{\|Ax\|\|x\|}{\|Ax\|}) = \|x\|$, where the inequality comes from 3.28 (ii). $\square$

3.30 Let $S \in M_n(\mathbb{F})$ be an invertible matrix. Given any matrix norm $\|\cdot\|$ on M_n, define $\|\cdot\|_S$ by $\|A\|_S = \|SAS^{-1}\|$. Prove that $\|\cdot\|_S$ is a matrix norm on M_n.

Proof As above, we show the definition of norm applies (see 3.24)

1. $\|A\|_S = \|SAS^{-1}\| \geq 0$ by definition of matrix norm, and $\|A\|_S = \|SAS^{-1}\| = 0$ iff $SAS^{-1} = 0$, and since S is invertible, iff $A = 0$.

2. $\|\alpha A\|_S = \|S(\alpha A)S^{-1}\| = |\alpha|\|SAS^{-1}\| = |\alpha|\|A\|_S$

3. $\|A + B\|_S = \|S(A + B)S^{-1}\| = \|SAS^{-1} + SBS^{-1}\| \leq \|SAS^{-1}\| + \|SBS^{-1}\| = \|A\|_S + \|B\|_S$, where the inequality stems from the triangle inequality in the well-defined matrix norm.

Thus, the norm is well-defined. $\square$

3.37 Let $V = \mathbb{R}[x; 2]$ be the space of polynomials of at most two, which is a subspace of the inner product space $L^2([0, 1]; \mathbb{R})$. Let $L : V \to \mathbb{R}$ be the linear functional given by $L[p] = p'(1)$. Find the unique $q \in V$ such that $L[p] = \langle q, p\rangle$, as guaranteed by the Riesz representation theorem.

Unique q Since the space is isomorphic to $\mathbb{R}^3$, so we represent all polynomials $p(x) = a_2 x^2 + a_1 x + a_0$ as $[a_2 \quad a_1 \quad a_0]^T$. Then the derivative of such a polynomial is $p'(x) = 2a_2 x + a_1$, i.e. $[0 \quad 2a_2 \quad a_1]^T$, such that $p'(1) = 2a_2 + a_1$. Thus, to capture the desired q, it needs to satisfy

$$2a_2 + a_1 = L[p] = \langle q, p \rangle = q^T p = q^T [a_2 \quad a_1 \quad a_0]^T$$

i.e. $q = [2 \quad 1 \quad 0]^T$ is the desired q. $\square$

3.38 Let $V = \mathbb{F}[x; 2]$, which is a subspace of the inner product space $L^2([0, 1]; \mathbb{R})$. Let D be the derivative operator $D : V \to V$; that is $D[p](x) = p'(x)$. Write the matrix representation of D with respect to the power basis $[1, x, x^2]$ of $\mathbb{F}[x; 2]$. Write the matrix representation of the adjoint of D with respect to this basis.

Proof In the words of 3.37, we are looking for the D that satisfies

$$[0 \quad 2a_2 \quad a_1]^T = D \cdot [a_2 \quad a_1 \quad a_0]^T$$

i.e. $D = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ $\square$

The adjoint is the D^* such that $\langle D^* q, p \rangle = \langle q, Dp \rangle$ matrix representation. It follows that $D^* = D^H = \begin{bmatrix} 0 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

Note that if we order the basis in reversed order, such that a polynomial p is expressed as $[a_0 \quad a_1 \quad a_2]^T$, D and D^* will be reversed.

3.39 Prove the following:

Proposition 3.7.12 Let V and W be finite-dimensional inner product spaces. The adjoint has the following properties:

Note I will use the results from 3.40 to prove the following.

(i) If $S, T \in \mathscr{L}(V; W)$, then $(S + T)^* = S^* + T^*$ and $(\alpha T)^* = \overline{\alpha} T^*$, $\alpha \in \mathbb{F}$.

    Proof By linearity of the inner product, we have

$$\langle x, (S + T)^* y \rangle = \langle (S + T)x, y \rangle = \langle Sx + Tx, y \rangle = \langle Sx, y \rangle + \langle Tx, y \rangle = S^* + T^*$$

(ii) If $S \in \mathscr{L}(V; W)$, then $(S^*)^* = S$.

Proof   This follows from the definition of the adjoint, since $(S^*)^*$ is the operator such that $\langle (S^*)^* x, y \rangle = \langle x, (S^*)y \rangle$. Now observe that for all x, y, $\langle x, (S^*)y \rangle = x^H S^* y = (Sx)^H y = \langle Sx, y \rangle$. This holds for all x, y, so $(S^*)^* = S$ $\square$

(iii) If $S, T \in \mathscr{L}(V)$, then $(ST)^* = T^* S^*$.

Proof   We have that for all x, y,

$$\langle x, (ST)^* y \rangle = \langle STx, y \rangle = \langle Tx, S^* y \rangle = \langle x, T^* S^* y \rangle$$

, so $(ST)^* = T^* S^*$ $\square$

(iv) If $T \in \mathscr{L}(V)$ and T is invertible, then $(T^*)^{-1} = (T^{-1})^*$.

Proof   Observe simply that $T^*(T^{-1})^* = (T^{-1}T)* = I^* = I$ (with the first equality following from (iii)).

3.40   Let $M_n(\mathbb{F})$ be endowed with the Frobenius inner product. Any $A \in M_n(\mathbb{F})$ defines a linear operator on $M_n(\mathbb{F})$ by left multiplication: $B \mapsto AB$.

(i) Show that $A^* = A^H$.

Proof   By definition of the Frobenius inner product, for all B, C, we have

$$\langle A^H B, C \rangle = tr((A^H B)^H C) = tr(B^H AC) = \langle B, AC \rangle = \langle A^* B, C \rangle   \square$$

(ii) Show that for any $A_1, A_2, A_3 \in M_n(\mathbb{F})$, we have $\langle A_2, A_3 A_1 \rangle = \langle A_2 A_1^*, A_3 \rangle$.

Proof   By (i), and the invariance of the trace in even permutation, we have:

$$\langle A_2, A_3 A_1 \rangle = tr(A_2^H A_3 A_1) = tr(A_1 A_2^H A_3)$$

$$= tr((A_2 A_1^H)^H A_3) = \langle A_2 A_1^H, A_3 \rangle = \langle A_2 A_1^*, A_3 \rangle   \square$$

(iii) Let $A \in M_n(\mathbb{F})$. Define the linear operator $T_A : M_n(\mathbb{F}) \to M_n(\mathbb{F})$ by $T_A(X) = AX - XA$, and show that $(T_A)^* = T_{A^*}$.

Proof $\langle T_{A^*} B, C \rangle = \langle A^* B - BA^*, C \rangle = \langle A^* B, C \rangle - \langle BA^*, C \rangle$

$$= tr((A^* B)^H C) - tr((BA^*)^H C) = tr(B^H AC) - tr(AB^H C) = tr(B^H AC) - tr(B^H CA)$$

$$= \langle B, AC \rangle - \langle B, CA \rangle = \langle B, T_A C \rangle = \langle (T_A)^* B, C \rangle   \square$$

3.44   Given $A \in M_{m \times n}(\mathbb{F})$ and $\mathbf{b} \in \mathbb{F}^{\mathbf{m}}$, prove the *Fredholm alternative*: Either $Ax = b$ has a solution $x \in \mathbb{F}^n$ or there exists $y \in \mathscr{N}(A^H)$ such that $\langle y, b \rangle \neq 0$.

Proof Reformulating, we need to show that :

$Ax = b$ has a solution $x \in \mathbb{F}^n$ if and only if $\langle y, b \rangle = 0$ for all $y \in \mathscr{N}(A^H)$.

Note that if $Ax = b$, then for all $y \in \mathscr{N}(A^H)$, we have $\langle y, b \rangle = \langle y, Ax \rangle = \langle A^H y, x \rangle = \langle 0, x \rangle = 0$

Conversely, suppose $\langle y, b \rangle = 0$ for all $y \in \mathscr{N}(A^H)$. Now if there were no solution, $b \notin \mathscr{R}(A)$, hence $b \in \mathscr{N}(A^H)$, but then, by taking $y = b$, $\langle b, b \rangle > 0$, unless $b = 0$. But even, the equation $Ax = b = 0$ would have the solution $x = 0$. Hence, $Ax = b$ must have a solution. $\square$

3.45 Consider the vector space $M_n(\mathbb{R})$ with the Frobenius inner product. Show that $Sym_n(\mathbb{R})^\perp = Skew_n(\mathbb{R})$.

Proof Suppose $A \in Skew_n(\mathbb{R})$, $B \in Sym_n(\mathbb{R})$. Then $\langle A, B \rangle = tr(AB) = tr((AB)^H) = tr(B^H A^H) = -tr(BA) = -tr(AB)$, such that we need to conclude that $tr(AB) = 0$, i.e. $Skew_n(\mathbb{R}) \subset Sym_n(\mathbb{R})^\perp$.

Now suppose $A \in Sym_n(\mathbb{R})^\perp$. Note that for any A, $A + A^H \in Sym_n(\mathbb{R})$, since $(A + A^H)^H = A + A^H$. Observe that for any $B \in Sym_n(\mathbb{R})$,

$$\langle A + A^H, B \rangle = \langle A, B \rangle + \langle A^H, B \rangle = 0 + tr(AB) = tr(AB^H)$$

$$= tr(B^H A) = tr((B^H A)^H) = tr(A^H B) = \langle A, B \rangle = 0$$

Note however, by choosing $B = A + A^H$, and definition of the norm, we know that $\langle B, B \rangle \geq 0$, with equality iff $B = 0$. Hence $A + A^H = 0$, i.e. $A = -A^H$ and $A \in Skew_n(\mathbb{R})$

Combining the subset statements, we see that $Sym_n(\mathbb{R})^\perp = Skew_n(\mathbb{R})$ $\square$

3.46 Prove the following for an $m \times n$ matrix A:

(i) If $x \in \mathscr{N}(A^H A)$, then Ax is in both $\mathscr{R}$ and $\mathscr{N}(A^H)$.

Proof We have $x \in \mathscr{N}(A^H A)$, i.e. $A^H(Ax) = 0$, so $Ax \in \mathscr{N}(A^H)$. Further, Ax is clearly in the range of A, by definition.

(ii) $\mathscr{N}(A^H A) = \mathscr{N}(A)$

Proof Let $x \in \mathscr{N}(A^H A)$, then $A^H(Ax) = 0$, so also $\|Ax\| = x^H A^H(Ax) = 0$, hence $Ax = 0$, i.e. $\mathscr{N}(A^H A) \subset \mathscr{N}(A)$. That $\mathscr{N}(A^H A) \supset \mathscr{N}(A)$ is clear, since if $Ax = 0$, clearly $A^H(Ax) = A^H(0) = 0$. $\square$

(iii) A and $A^H A$ have the same rank.

Proof This follows directly from (ii) and dimension theorem. $\square$

(iv) If A has linearly independent columns, then $A^H A$ is nonsingular.

Proof This follows by (iii), since we linearly independent columns means full column rank n, and $A^H A$ is dimension $n \times n$. $\square$

3.47 Assume A is an $m \times n$ matrix of rank n. Let $P = A(A^H A)^{-1} A^H$. Prove the following:

(i) $P^2 = P$.

Proof Note simply that $P^2 = A(A^H A)^{-1} A^H A (A^H A)^{-1} A^H = A(A^H A)^{-1} I A^H = A(A^H A)^{-1} A^H = P$ $\square$

(ii) $P^H = P$.

Proof Note simply that $P^H = (A(A^H A)^{-1} A^H)^H = (A^H)^H ((A^H A)^{-1})^H A^H$
$= A((A^H A)^H)^{-1} A^H = A(A^H A)^{-1} A^H = P$ where we used that we can interchange inverse and Hermitian operator for a symmetric matrix. $\square$

(iii) $rank(P) = n$.

Proof Note that $rank(A) = n$, so the rank of a composition of A, is at most n. Now, observe that if $y \in \mathscr{R}(A)$, then $y = Ax = AIx = A(A^H A)^{-1} A^H A x = A(A^H A)^{-1} A^H y$ is also in the range of P. Thus, the rank of P, must be lower or equal, and hence equal to n. $\square$

3.48 Consider the vector space $M_n(\mathbb{R})$ with the Frobenius inner product. Let $P(A) = \frac{A+A^T}{2}$ be the map $P : M_n(\mathbb{R}) \mapsto M_n(\mathbb{R})$. Prove that

(i) P is linear.

Proof $P(\alpha A + B) = \frac{\alpha A + B + (\alpha A + B)^T}{2} = \frac{\alpha A + B + \alpha A^T + B^T}{2} = \frac{\alpha(A+A^T) + B + B^T}{2} = \alpha P(A) + P(B)$ $\square$

(ii) $P^2 = P$.

Proof Observe $P^2(A) = P(\frac{A+A^T}{2}) = \frac{(\frac{A+A^T}{2}) + (\frac{A+A^T}{2})^T}{2} = (\frac{A+A^T}{4}) + \frac{A+A^T}{4} = \frac{A+A^T}{2} = P(A)$ $\square$

(iii) $P^* = P$.

Proof Note first of all that for $A \in M_n(\mathbb{R})$, $A^T = A^H$. Using 3.40, we observe that $\langle P^*(A), B \rangle = \langle A, P(B) \rangle = \langle A, \frac{B+B^T}{2} \rangle = \langle A, \frac{B}{2} \rangle + \langle A, \frac{B^T}{2} \rangle$

$$= tr(\frac{A^H B}{2}) + tr(\frac{A^H B^H}{2}) = tr(\frac{A^H B}{2}) + tr(\frac{(A^H B^H)^H}{2}) = tr(\frac{A^H B}{2}) + tr(\frac{BA}{2}) = tr(\frac{A^H B}{2}) + tr(\frac{AB}{2})$$

$$= \langle \frac{A}{2}, B \rangle + \langle \frac{A^T}{2}, B \rangle = \langle \frac{A+A^T}{2}, B \rangle = \langle P(A), B \rangle$$

Hence $P^* = P$. $\square$

(iv) $\mathscr{N}(P) = Skew_n(\mathbb{R})$.

Proof For $A \in \mathscr{N}(P)$, we have $\frac{A+A^T}{2} = 0$, i.e. $A = -A^T = -A^H$ since we work in $\mathbb{R}$, i.e. $A \in Skew_n(\mathbb{R})$ and $\mathscr{N}(P) \subset Skew_n(\mathbb{R})$.

Now if $A \in Skew_n(\mathbb{R})$, then $A = -A^H = -A^T$, and thus $P(A) = 0$, i.e. $\mathscr{N}(P) \supset Skew_n(\mathbb{R})$.

Thus, we have shown that $\mathscr{N}(P) = Skew_n(\mathbb{R})$ $\square$

(v) $\mathscr{R}(P) = Sym_n(\mathbb{R})$.

Proof We have shown in a previous exercise that $\mathscr{R}(0.5P) \subset Sym_n(\mathbb{R})$, so $\mathscr{R}(P) = \mathscr{R}(0.5P) \subset Sym_n(\mathbb{R})$.

Suppose on the other hand, that $A \in Sym_n(\mathbb{R})$, then observe that $P(A) = \frac{A+A^T}{2} = \frac{2A}{2} = A$, so clearly $\mathscr{R}(P) \supset Sym_n(\mathbb{R})$.

Thus, we have shown that $\mathscr{R}(P) = Sym_n(\mathbb{R})$ $\square$

(vi) $\|A - P(A)\|_F = \sqrt{\frac{tr(A^T A) - tr(A^2)}{2}}$.

Proof $\|A - P(A)\|_F^2 = \langle A - P(A), A - P(A) \rangle = \langle A - \frac{A+A^T}{2}, A - \frac{A+A^T}{2} \rangle$

$$= \langle \frac{A-A^T}{2}, \frac{A-A^T}{2} \rangle = tr((\frac{A-A^T}{2})^T \frac{A-A^T}{2}) = tr(\frac{A^T - A}{2} \frac{A-A^T}{2})$$

$$= tr(\frac{A^T A - A^2 - (A^T)^2 + AA^T}{4}) = tr(\frac{A^T A - A^2 - A^2 + A^T A}{4})$$

$$= tr(\frac{A^T A - A^2}{2}) = \frac{tr(A^T A) - tr(A^2)}{2}$$

So $\|A - P(A)\|_F = \sqrt{\frac{tr(A^T A) - tr(A^2)}{2}}$ where we again used that the trace is linear and invariant under even permutations. $\square$

3.50 Let $(x_i, y_i)_{i=1}^n$ be a collection of data points that we have reason to believe should lie (roughly) on an ellipse of the form $rx^2 + sy^2 = 1$. We wish to find the least-squares approximation for r and s. Write A, x, and b for the corresponding normal equation in terms of the data x_i and y_i and the unknowns r and s.

Notation We want to minimize $\sum_{i=1}^n |1 - rx_i^2 - sy_i^2|^2$, or equivalently,

$$\sum_{i=1}^n |y_i^2 - (\frac{1}{s} - \frac{r}{s}x_i^2)|^2$$

That is, we want to find the closest 'solution' $\hat{x}$ to the following equation:

$$b = \begin{bmatrix} y_1^2 \\ y_2^2 \\ \vdots \\ y_n^2 \end{bmatrix} = \begin{bmatrix} 1 & x_1^2 \\ 1 & x_2^2 \\ \vdots & \vdots \\ 1 & x_n^2 \end{bmatrix} \cdot \begin{bmatrix} (\frac{\hat{1}}{\hat{s}}) \\ (\frac{-r}{s}) \end{bmatrix} = A\hat{x}$$

The resulting normal equation(s) is as always

$$A^H A\hat{x} = \begin{bmatrix} \sum_{i=1}^n 1 & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^4 \end{bmatrix} \hat{x} = \begin{bmatrix} n & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^4 \end{bmatrix} \begin{bmatrix} (\frac{\hat{1}}{\hat{s}}) \\ (\frac{-r}{s}) \end{bmatrix} = \begin{bmatrix} \frac{n}{\hat{s}} - \frac{\hat{r}}{\hat{s}} \sum_{i=1}^n x_i^2 \\ \frac{\hat{1}}{\hat{s}} \sum_{i=1}^n x_i^2 - \frac{\hat{r}}{\hat{s}} \sum_{i=1}^n x_i^4 \end{bmatrix}$$

$$= A^H b = \begin{bmatrix} \sum_{i=1}^n y_i^2 \\ \sum_{i=1}^n x_i^2 y_i^2 \end{bmatrix}$$