Math, Problem Set #2, Inner Product Spaces

Name: Benjamin Lim

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

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

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

3.3 i) Define $\langle f,g\rangle = \int_0^1 f(x)g(x)dx$ as the inner product. Then,
$cos\theta = \frac{\langle x,x^5\rangle}{\|x\|\|y\|}$
$= \frac{\int_0^1 x^6 dx}{\sqrt{\int_0^1 x^2 \int_0^1 x^{10} dx}}$
$= \frac{\frac{1}{7}}{\sqrt{\frac{1}{33}}}$
Therefore, $\theta = arccos\frac{\sqrt{33}}{7}$

3.3 ii) $cos\theta = \frac{\langle x^2,x^4\rangle}{\|x\|\|y\|}$
$= \frac{\int_0^1 x^6 dx}{\sqrt{\int_0^1 x^4 \int_0^1 x^8 dx}}$
$= \frac{\frac{1}{7}}{\sqrt{\frac{1}{45}}}$
Therefore, $\theta = arccos\frac{\sqrt{45}}{7}$

3.8i) Claim: S is an orthonormal set because its columns have norm 1 and they are orthogonal to each other.

They are orthogonal to each other because their inner products are each 0.

$\frac{1}{\pi}\int_{-\pi}^{\pi} cos(t)sin(t)dt = 0$
$\frac{1}{\pi}\int_{-\pi}^{\pi} cos(t)cos(2t)dt = 0$
$\frac{1}{\pi}\int_{-\pi}^{\pi} cos(t)sin(2t)dt = 0$
$\frac{1}{\pi}\int_{-\pi}^{\pi} sin(t)sin(2t)dt = 0$
$\frac{1}{\pi}\int_{-\pi}^{\pi} sin(2t)cos(2t)dt = 0$

Their norms also each equal 1.

$\frac{1}{\pi}\int_{-\pi}^{\pi} cos(t)cos(t)dt = 1$
$\frac{1}{\pi}\int_{-\pi}^{\pi} cos(2t)cos(2t)dt = 1$
$\frac{1}{\pi}\int_{-\pi}^{\pi} sin(t)sin(t)dt = 1$
$\frac{1}{\pi}\int_{-\pi}^{\pi} sin(2t)sin(2t)dt = 1$

3.8ii) $\|t\| = \sqrt{\langle t, t\rangle}$
$= \sqrt{\frac{1}{\pi}\int_{-\pi}^{\pi} t^2 dt} = \sqrt{\frac{2}{3}}\pi$

3.8iii) $proj_x cos(3t)$
$= \langle x, cos(3t)\rangle \frac{x}{\|x\|^2}$
$= \sum_{n=1} \frac{1}{\pi}\int_{-\pi}^{\pi} x_n cos(3t)dt$
$= \frac{1}{\pi}\int_{-\pi}^{\pi} cos(t)cos(3t)dt + \frac{1}{\pi}\int_{-\pi}^{\pi} sin(t)cos(3t)dt + \frac{1}{\pi}\int_{-\pi}^{\pi} cos(2t)cos(3t)dt + \frac{1}{\pi}\int_{-\pi}^{\pi} sin(2t)cos(3t)dt$
$= 0 + 0 + 0 + 0$
$= 0$

3.8iv) $proj_x t$
$= \frac{1}{pi}\int_{-pi}^{pi} tcos(t)dt + \frac{1}{pi}\int_{-pi}^{pi} tsin(t)dt + \frac{1}{pi}\int_{-pi}^{pi} tcos(2t)dt + \frac{1}{pi}\int_{-pi}^{pi} tsin(2t)dt$
$= 2sin(t) - sin(2t)$

3.9) Claim: A rotation in $\mathbb{R}^2$ is an orthonormal transformation with respect to the usual inner product. We want to show that $\langle x, y\rangle = \langle Lx, Ly\rangle$ where L is the rotation transformation. Let x = $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and y = $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$. Then, $\begin{bmatrix} \cos\theta & \text{-}\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos\,\theta x_1 - sin\theta x_2 \\ \sin\,\theta x_1 + cos\theta x_1 \end{bmatrix}$. Similarly, $\begin{bmatrix} \cos\theta & \text{-}\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos\,\theta y_1 - sin\theta y_2 \\ \sin\,\theta y_1 + cos\theta y_1 \end{bmatrix}$. Taking the dot product of the two resulting vectors, we get $(cos\theta x_1 - sin\theta x_2)(cos\theta y_1 - sin\theta y_2) + (sin\theta x_1 + cos\theta x_2)(sin\theta y_1 + cos\theta y_2) = x_1y_1 + x_2y_2 = \langle x, y\rangle$

3.10i) Proof
Let x and y be nonzero elements of $\mathbb{F}$. Because Q is orthonormal, $\langle Qx, Qy\rangle = \langle x, y\rangle$. By the adjoint, we know that this is true if and only if $\langle Q^HQx, y\rangle = \langle x, y\rangle$. Subtracting $\langle x, y\rangle$ on both sides, we get $\langle Q^HQx - x, y\rangle = 0$. Since y is nonzero, this implies that $Q^HQx - x = 0$, which in turn implies that $Q^HQx = x$. Therefore, Q^HQ is the iden-

tity matrix. Similarly, we know that $\langle Qx, Qy\rangle = \langle x, y\rangle$. Multiplying the hermitian conjugate on both sides, we get $\langle Q^H Qx, Q^H Qy\rangle = \langle Q^H x, Q^H y\rangle = \langle x, y\rangle$. Like before, taking the adjoint and subtracting $\langle x, y\rangle$ on both sides, we get $\langle QQ^H x - x, y\rangle = 0$. By a similar argument as before, $QQ^H = 0$.

For the other direction, suppose $Q^H Q = QQ^H = I$, then we know that $\langle x, y\rangle = \langle Q^H Qx, y\rangle = \langle Qx, Qy\rangle$. Therefore, Q is an orthonormal matrix.

3.10ii) Proof
$\|Qx\| = \sqrt{\langle Qx, Qx\rangle} = \sqrt{\langle x, x\rangle} = \|x\|$.

3.10iii) Proof
Suppose Q is orthonormal. Then we know that Q^H is orthonormal. Finally, note that $Q^H = Q^{-1}$ because $Q^H Q = QQ^H = I$. Therefore, Q^{-1} is also orthonormal.

3.10iv) Proof
We know from part i) that the columns of the matrix Q are orthonormal if and and only if $Q^H Q = QQ^H = I$. Let Q be represented by $\begin{pmatrix} a_1 & a_2 & ... & a_n \end{pmatrix}$, where a_i is the ith column vector. Then, $Q^H Q = \begin{pmatrix} a_1 & a_2 & ... & a_n \end{pmatrix}^H \begin{pmatrix} a_1 & a_2 & ... & a_n \end{pmatrix}$, which gives us:
$\begin{bmatrix} a_1^H a_1 & a_1^H a_2 & \dots & a_1^H a_n \\ a_2^H a_1 & a_2^H a_2 & \dots & a_2^H a_n \\ \hdotsfor{4} \\ a_1^H a_n & a_n^H a_2 & \dots & a_n^H a_n \end{bmatrix}$. Then, note that $a_i^H a_j = \langle a_i a_j \rangle$ 1 if $i = j$ and 0 otherwise. Therefore, the columns are orthonormal.

3.10v) Proof
No. As a counterexample, consider the 2x2 matrix of reals, $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$. The determinant is $2 - 1 = 1$, but the columns are not orthonormal because their usual inner product is 3 (instead of 0).

3.10vi) Proof
From part i), we know that Q is an orthonormal matrix if and only if $Q^H Q = QQ^H = I$. Let Q_1 and Q_2 be two distinct orthogonal matrices. Then, $(Q_1 Q_2)^H (Q_1 Q_2) = Q_2^H Q_1^H Q_1 Q_2 = I$. Similarly, $(Q_1 Q_2)(Q_1 Q_2)^H = Q_1 Q_2 Q_2^H Q_1^H = I$. Therefore, $Q_1 Q_2$ are orthonormal.

3.11) If two vectors are linearly dependent, then the projection of one onto the other is just itself. This means that $q_2 = fracx_2 - x2\|x_2 - p_1\| = 0$. The span that the subsequent vectors would be projected on would therefore remain as it was before, ultimately, resulting in a basis of only one vector.

3.16i) Since any matrix $A = QR$, where Q is orthonormal and R is upper triangular, we can let $Q_1 = -Q$ and $R_1 = -R$. Then, since $QR = (-Q)(-R)$, we know

$A = QR = Q_1R_1$ where $Q_1 \neq Q$ and R_1R.

3.16ii) Suppose that A is invertible such that its decomposition matrices Q and R are similarly invertible. Suppose to the contrary that $A = Q_1R_1 = Q_2R_2$ and that $Q_1 \neq Q_2$ and $R_1 \neq R_2$. Then, multiplying Q_2^{-1} followed by R_1^{-1} on both sides, we get $Q_2^{-1}Q_1 = R_2R_1^{-1}$. Since Q_2 is orthonormal, Q_2^{-1} is orthonormal. Because the product of orthonormal matrices is orthonormal, Q_2^{-1} is orthonormal. We also know that the inverse of upper triangular matrices is orthonormal and that the product of upper triangular matrices is upper triangular. Therefore, $R_2R_1^{-1}$ is upper triangular. By the equality of both sides, we know that the matrix is both orthonormal and upper triangular. These two characteristics imply that the resulting matrix is the identity. To see why, note that an upper triangular matrix must have $a_{11} = 1$ and $a_{i1} = 0$ for all $i \neq 1$. Because it is orthonormal, the column's norm must equal 1. Therefore, $a_{11} = 1$. Now looking at the second column, we know that only the first two entries are allowed to be nonzero. Since the diagonal entries are restricted to be positive, this implies that $a_{12} = 0$ and $a_{22} = 1$. Else, the second column would not be orthogonal to the first column as the resulting usual inner product would be positive instead of nonzero. By induction, one can see that the ith column of the matrix must have $a_{ii} = 1$ and $a_{ji} = 0$ for all $j \neq i$. Since $Q_2^{-1}Q_1 = I$ and $R_2R_1^{-1} = I$, we know that $Q_2 = Q_1$ and $R_2 = R_1$, which contradicts our initial assumption that the decomposition matrices were distinct. Therefore, the QR decomposition of A such that R has only positive diagonal elements is unique.

3.17) Given $A^HAx = A^Hb$, we can replace A with the reduced QR decomposition to get $(\hat{Q}\hat{R})^H\hat{Q}\hat{R}x = (\hat{Q}\hat{R})^Hb$. Distributing the Hermitian conjugate, we get $\hat{R}^H\hat{Q}^H\hat{Q}\hat{R}x = \hat{R}^H\hat{Q}^Hb$. Since $\hat{Q}$ and therefore $\hat{Q}^H$ are orthonormal, their product is the identity matrix, giving us $\hat{R}^H\hat{R}x = \hat{R}^H\hat{Q}^Hb$. Finally, because A has rank N, we know that the columns of A are linearly independent. Therefore, the $\hat{R}$ matrix resulting from applying the Gram-Schmidt procedure only has nonzero diagonal entries. Since $\hat{R}$ is also square and upper triangular, it follows that $\hat{R}$ is invertible, and therefore $\hat{R}^H$ is invertible. Therefore, we can multiply the equation by $\hat{R}^{-1}$ throughout, which leaves us with the equation $\hat{R}x = \hat{Q}^Hb$, as desired.

3.23) Since normed linear spaces are always weakly positive, we can square both sides and preserve the inequality. This gives us $|\|x-y\||^2 \leq \|x-y|\hat{2}$. Expanding the left-hand side, $\|x\|^2+\|y\|^2-2\|x\|\|y\| \leq \|x\|^2+\|y\|^2-2\|x\|\|y\|cos\theta = \|x-y\|^2$ because $-1 \leq cos\theta \leq 1$. Taking the squareroot of both sides, we get $|\|x\| - \|y\|| = \|x-y\|$, as desired.

3.24i) $\int_a^b |f(t)|dt \geq 0$ because $|f(t)| \geq 0$ for all real numbers t. $\int_a^b |kf(t)|dt = |k|\int_a^b |f(t)|dt$. Finally, $\int_a^b |f(t)+g(t)|dt \leq \int_a^b |f(t)|dt + \int_a^b |g(t)|dt$ because $|f(t)| + |g(t)| \leq |f(t)+g(t)|$ for all $t \in [a,b]$ by the usual triangle inequality.

3.24ii) $(\int_a^b |f(t)|^2dt)^{1/2} \geq 0$ because $|f(t)|^2 \geq 0$ for all $t \in [a,b]$. Moreover, $(\int_a^b |kf(t)|^2dt)^{1/2} =$

$(|k|^2(1/2)\int_a^b|f(t)|^2dt)^{1/2} = |k|\int_a^b|f(t)|^2dt)^{1/2}$. Lastly, $(\int_a^b|f(t)+g(t)|^2dt)^{1/2} \leq (\int_a^b|f(t)|^2dt)^{1/2} + (\int_a^b|g(t)|^2dt)^{1/2}$. To see why, note that $\int_a^b|f(t)+g(t)|^2dt \leq \int_a^b(|f(t)|+|g(t)|)^2dt$ by the usual triangle inequality, which equals $\int_a^b|f(t)|^2+|g(t)|^2+2|f(t)||g(t)|dt \leq \int_a^b|f(t)|^2+|g(t)|^2+2\|f(t)\|\|g(t)\|dt$ by the Cauchy-Schwarz inequality. However, this simply equals $\int_a^b|f(t)|^2+|g(t)|^2+2\int_a^b|f(t)|dt\int_a^b|g(t)|dt \leq \int_a^b|f(t)|^2+|g(t)|^2+2(\int_a^b|f(t)|^2dt)^{1/2}(\int_a^b|g(t)|^2dt)^{1/2}$ by the Cauchy-Schwarz inequality again. The last inequality holds because the integral of squareroots is smaller than the squareroot of integrals. To see why, consider $\int_a^b|f(t)|^{1/2}dt \leq (\int_a^b|f(t)|dt)^{1/2}(\int_a^b dt)^{1/2} = \sqrt{b-a}(\int_a^b|f(t)|dt)^{1/2}$. Finally, because the integrands and thus integrals are positive, we can take the squareroot on both sides and conclude that $(\int_a^b|f(t)+g(t)|^2dt)^{1/2} \leq (\int_a^b|f(t)|^2dt)^{1/2} + (\int_a^b|g(t)|^2dt)^{1/2}$.

3.24iii) $sup_{x\in[a,b]}|f(x)| \geq 0$ because the absolute value of the range of f over $[a,b]$ is always weakly positive. Also, $sup_{x\in[a,b]}|kf(x)| \geq |k|sup_{x\in[a,b]}|f(x)|$ because $|kf(x)| = |k||f(x)|$ for all scalars k, and because , $sup|k|f(x)|x\in[a,b] = |k|supf(x)|x\in[a,b]$. To see why, let $s = supf(x)|x\in[a,b]$, and thus $s \geq f(x)$ for all $in[a,b]$ and $s \leq y$ for any upper bound y. Since $s \geq f(x)$, it follows that $|k|s \geq |k|f(x)$ since $|k| \geq 0$. Since s is a least upper bound, we know that $s \leq y$ and thus $|k|s \leq |k|y$ for any upper bound y. Finally,
$sup_{x\in[a,b]}|f(x)+g(x)|$
$\leq |sup_{x\in[a,b]}f(x) + sup_{x\in[a,b]}g(x)|$ by lemma
$\leq |sup_{x\in[a,b]}f(x)| + |sup_{x\in[a,b]}g(x)|$ by the triangle inequality.

Lemma: $sup(f+g) \leq sup(f) + sup(g)$

3.26) Claim: Topological equivalence is an equivalence relation.
Reflexive: Let $m = M = 1$. Then, $M\|x\|_a \leq \|x\|_a \leq M\|x\|_a$.
Symmetric: Suppose $m\|x\|_a \leq \|x\|_b \leq M\|x\|_b$. Then, $m\|x\|_b \leq mM\|x\|_a \leq M\|x\|_b$.
Transitive: Suppose $m_1\|x\|_a\|leq\|x\|_b \leq M_1\|x\|_a$ and $m_2\|x\|_b \leq \|x\|_c\|M_2\|x\|_b$. Then we know that $m_2m_1\|x\|_a \leq \|x\|_c \leq M_2M_1\|x\|_a$.

i) Claim: $\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2$
The first inequality follows because $(\|x\|_2)^2 = |x_1|^2+...+|x_n|^2 \leq |x_1|^2+...+|x_n|^2+2|x_1||x_2|+2|x_2||x_3|+... = (|x_1|+...+|x_n|)^2 = (\|x_1\|)^2$. Since both sides are positive, we can take the squareroot and conclude that $\|x\|_2 \leq \|x\|_1$. For the second inequality, we know $\|x\|_1 = \sum_{k=1}^n|x_k| = \sum_{k=1}^n 1\cdot|x_k| \leq \sqrt{\sum_{k=1}^n 1^2}\sqrt{\sum_{k=1}^n|x_k|^2} = \sqrt{n}\|x\|_2$, where the last inequality is the Cauchy-Schwarz inequality.

ii) Claim: $\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty$
$(sup|x_1|,|x_2|,...,|x_n|)^2 = (|x|_{sup})^2 \leq (|x|_{max})^2 + \sum_{k\neq max}^n(|x|_k)^2 \leq ((|x_1|^2+...+|x_n|^2)^{1/2})^2$ because $\sum_{k\neq max}^n(|x|_k)^2 \geq |x|_{sup})^2 - |x|_{max})^2$, as if it were not, then $|x|_{sup}$ would not be the least upper bound since there is an upper bound that is smaller than itself. Taking the squareroot on both sides, we get the first inequality as desired.

The second inequality follows from the fact that $\sum_{k=1}^{n} |x_k|^2 \leq n \cdot |x_{max}|^2 \leq n \cdot |x_{sup}|^2$. Taking the squareroot of both sides, we get $\|x_2\| \leq \|x\|_\infty$, as desired.

3.28) Let A be a n x n matrix. For the subsequent claims, we will denote Ax as v, where v is defined on the vector space V.

Claim: $\frac{1}{\sqrt{n}}\|A\|_2 \leq \|A\|_1 \leq \sqrt{n}\|A\|_2$

Note that $\frac{1}{\sqrt{n}}\|A\|_2 = \frac{1}{\sqrt{n}} sup_{x\neq 0}\frac{\|Ax\|_2}{\|x\|_2} = \frac{1}{\sqrt{n}} sup_{x\neq 0}\frac{\|v\|_2}{\|x\|_2}$ and $\|A\|_1 = sup_{x\neq 0}\frac{\|Ax\|_1}{\|x\|_1} = sup_{x\neq 0}\frac{\|v\|_1}{\|x\|_1}$. The first inequality follows because we know from 26 i) that $\|v\|_2 \leq \|v\|_1$ and $\sqrt{n}\|x|_2 \geq \|x\|_1$ for all vectors x and v. For the second inequality, note that $\sqrt{n}\|A\|_2 = \sqrt{n} sup_{x\neq 0}\frac{\|v\|_2}{\|x\|_2}$. From 26 i), we know that $\sqrt{n}\|v\|_2 \geq \|v\|_1$ (larger numerator) and $\|x\|_2 \leq \|x\|_1$ (smaller denominator), such that the resulting fraction on the right-hand side is larger, giving us the result $\|A\|_1 \leq \sqrt{n}\|A\|_2$.

Claim: $\frac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2 \leq \sqrt{n}\|A\|_\infty$

Note that $\frac{1}{\sqrt{n}}\|A\|_\infty = \frac{1}{\sqrt{n}} sup_{x\neq 0}\frac{\|Ax\|_\infty}{\|x\|_\infty} = \frac{1}{\sqrt{n}} sup_{x\neq 0}\frac{\|v\|_\infty}{\|x\|_\infty}$ and $\|A\|_2 = sup_{x\neq 0}\frac{\|Ax\|_2}{\|x\|_2} = sup_{x\neq 0}\frac{\|v\|_2}{\|x\|_2}$. We know from 26 ii) that $\|v\|_\infty \leq \|v\|_2$ (smaller numerator) and $\sqrt{n}\|x\|_\infty \geq \|x\|_2$ (bigger denominator). Therefore, the left-hand side is a smaller fraction than the right-hand side for the first inequality, as desired. For the second inequality, note that $\sqrt{n}\|A\|_\infty = \sqrt{n} sup_{x\neq 0}\frac{\|v\|_\infty}{\|x\|_\infty}$. From 26 ii), we know that $\|v\|_2 \leq \sqrt{n}\|v\|_\infty$ (smaller numerator) and $\|x\| \leq \|x\|_\infty$ (bigger denominator), causing the fraction on the left to be smaller than that on the right of the second inequality, as desired.

3.29) Let Q be orthonormal. Then $\|Q\| = sup_{x\neq 0}\frac{\|Qx\|}{\|x\|} = sup_{x\neq 0}\frac{\|x\|}{\|x\|} = 1$. $\|R_x\| = sup_{x\neq 0}\frac{\|Ax\|}{\|A\|} = sup_{A\neq 0}\frac{\|Ax\|}{\|A\|} = sup_{A\neq 0}\frac{\|Ax\|}{sup_{x\neq 0}\frac{\|Ax\|}{\|x\|}} \leq sup_{x\neq 0}\|x\| \leq sup_{A\neq 0}\frac{\|Ax\|}{\frac{\|Ax\|}{\|x\|}} = \|x\|$. By Gram-Schmidt, any vector x with norm $\|x\|_2 = 1$ is part of an orthnormal basis, an dhence is the first column of an orthonormal matrix. Therefore, $\|R_x\| \geq \|x\|$ as well. Thus, we know $\|R_x\| = \|x\|$.

3.30) For an arbitrary matrix norm $\|\cdot\|$, $\|A\|_s = \|SAS^{-1}\| \geq 0$ because $\|\cdot\|$ is a matrix norm. Also, for any constant k, $\|kA\|_s = \|SkAS^{-1}\| = |k|\|SAS^{-1}\|$. Finally, $\|S(A+B)S^{-1}\| = \|SAS^{-1}+SBS^{-1}\| \leq \|SAS^{-1}\|+\|SBS^{-1}\|$ because $\|\cdot\|$ is a matrix norm so it fulfils the triangle inequality.

3.37)
Let $p(x) = cx^2 + bx + a$, such that we can express p in terms of the power basis $(a, b, c)^T$. Now let $q = 2x^2 + x$, such that it's matrix representation in the power basis is $q = (0, 1, 2)^T$. Then $\langle q, p\rangle = (a, b, c)^T \cdot (0, 1, 2)^T = b + 2c = P'(1) = b + 2c(1)$, as desired.

3.38) Let $p(x) = cx^2 + bx + a$, so $D[p](x) = p'(x)$ means that $D = (0, b, 2cx)^T$ when written in the power basis. Because the adjoint of D equals to Hermitian conjugate of D, and the Hermitian conjugate of D in the real numbers is just the transpose of D, we know that $D^* = (0, b, 2cx)$.

3.39) Claim : $(S+T)^* = S^* + T^*$
$\langle (S+T)v, w\rangle = \langle v, (S+T)^* w\rangle$
Also, $\langle (S+T)v, w\rangle$
$= \langle Sv, w\rangle + \langle Tv, w\rangle$
$= \langle v, S^* w\rangle + \langle v, T^* w\rangle$
$= \langle v, (S^* + T^*)w\rangle$
Therefore, $\langle v, (S^* + T^*)w\rangle = \langle v, (S+T)^* w\rangle$

Claim: $(aT)^* = \bar{a}T^*$
$\langle (aT)v, w\rangle$
$= \langle v, (aT)^* w\rangle$
$= a\langle Tv, w\rangle$
$= a\langle v, T^* w\rangle$
$= \langle v, \bar{a}T^* w\rangle$
Therefore, $(aT)^* w = \bar{a}T^* w$.

Claim: $(T^*)^* = T$
$\langle T^* w, v\rangle = \langle w, (T^*)^* v\rangle$
Also, since $\langle T^* w, v\rangle = \overline{\langle v, T^*\rangle} = \overline{\langle Tv, w\rangle} = \langle w, Tv\rangle$, we know that $(T^*)^* = T$.

Claim: $(ST)^* = T^* S^*$
$\langle ST(v), w\rangle = \langle v, (ST)^* w$
Also, $\langle ST(v), w\rangle = \langle S(T(v)), w\rangle = \langle T(v), S^* w\rangle = \langle v, T^*(S^* w)\rangle = \langle v, (T^* S^*)w\rangle$.
Therefore, $(ST)^* = T^* S^*$.

Claim: $(T^*)^{-1} = (T^{-1})^*)$
$\langle T^*(T^{-1})^* u, v\rangle = \langle (T^{-1})^* u, Tv\rangle = \langle u, T^{-1}Tv\rangle = \langle u, v\rangle$.
This implies that $T^*(T^{-1})^* = I$ and therefore $(T^{-1})^* = (T^*)^{-1}$.

3.40i) Claim: $A^* = A^H$
$\langle y, Ax\rangle = tr((Y^H)Ax) = tr((Y^H A)x) = \langle A^H y, x\rangle$. By the existence and uniqueness of the adjoint from Theorem 3.7.10, we know that $A^H = A^*$.

3.40ii) Claim: $\langle A_2, A_3 A_1\rangle = \langle A_2 A_1^*, A_3\rangle$
$\langle A_2, A_3 A_1\rangle = tr(A_2^H A_3 A_1) = tr(A_1 A2^H A_3) = \langle A_2 A_1^H, A_3\rangle = \langle A_2 A_1^*, A_3\rangle$.

3.40iii) Claim: $(T_A)^* = T_{A^*}$
$\langle T_{A^*}(Y), X\rangle = \langle A^* Y - Y A^*, X\rangle = \langle A^* Y, X\rangle - \langle Y A^*, X\rangle = \langle Y, AX\rangle - \langle Y, XA\rangle = \langle y, AX - XA\rangle = \langle y, T_A(X)\rangle = \langle T_A^*(Y), x\rangle$. Therefore, $(T_A)^* = T_{A^*}$.

3.44) $Ax = b$ has a solution if and only if $b \in R(A)$. Let $y \in N(A^H)$. By the Fundamental Subspaces Theorem, y is orthogonal to all elements in $R(A)$. Therefore, y is orthogonal to b if and only if $b \in R(A)$, which is true if and only if $\langle y, b\rangle = 0$.

3.45) We first show that the subspace of skew matrices is the complement of the subspace of symmetric matrices. This is because any real square matrix A, can be written as the sum of a symmetric matrix H, and a skew matrix K. To see why, let $K = \frac{1}{2}(A - A^T)$ and $H = \frac{1}{2}(A + A^T)$. Then, $H + K = \frac{1}{2}(A - A^T) + \frac{1}{2}(A + A^T) = A$. It remains to show that the two subspaces are orthogonal to each other. That is, $\langle H, K\rangle = tr(H^T K) = 0$ for all symmetric matrices H and all skew matrices K.

$\langle H, K\rangle = \langle \frac{1}{2}(A + A^T), \frac{1}{2}(A - A^T)\rangle$
$= \frac{1}{4}\langle A + A^T, A - A^T\rangle$
$= \frac{1}{4}(\langle A, A\rangle - \langle A, A^T\rangle + \langle A^T A\rangle - \langle A^T A^T\rangle)$
$= \frac{1}{4}(tr(A^T A) - tr(A^T A^T) + tr(AA) - tr(AA^T))$
$= \frac{1}{4}(tr(A^T A) - tr(A^T A) + tr(AA) - tr(AA))$
$= 0$

3.46i) Suppose $x \in N(A^H A)$. We already know $Ax \in R(A)$ because $R(A) = \{b \in \mathbb{F} : Ax = b\}$. Further, since x is in the null space of $A^H A$, we know $A^H Ax = 0$ and therefore that $Ax \in N(A^H)$.

3.46ii) We aim to show that $N(A^H A) = N(A)$ by showing that they are subsets of each other. Let $x \in N(A^H A)$. Then, by part i), $Ax \in R(A)$ and $x \in N(A^H)$. By the Fundamental Subspaces Theorem, we know that $R(A)$ and $N(A^H)$ are orthogonal to each other, which means that the only element they have in common is 0. This implies that $Ax = 0$, which implies that $x \in N(A)$. Now let $x \in N(A)$, which means that $Ax = 0$. Multiplying both sides by A^H, we get that $A^H Ax = 0$, such that $x \in N(A^H A)$, as desired.

3.46iii) Let A be a m x n matrix. Since $N(A^H A) = N(A)$, we know that $dim(N(A^H A)) = dim(N(A))$. By the rank-nullity theorem, $rank(A^H A) + dim(N(A^H A)) = n$ and that $rank(A) + dim(N(A)) = n$. These together imply that $rank(A^H A) = n - dim(N(A^H A))$ and $rank(A) = n - dim(N(A))$. Therefore, $rank(A^H A) = rank(A)$.

3.46iv) Suppose m x n matrix A has linearly independent columns. Then it has rank n. From part iii) we know that $rank(A) = rank(A^H A) = n$. Since $A^H A$ has rank n and is itself a n x n matrix, we know that $A^H A$ has n linearly independent columns and is thus nonsingular.

3.47i) $P^2 = A(A^H A)^{-1} A^H A(A^H A)^{-1} A^H = A(A^H A)^{-1} A^H$.

3.47ii) $P^H = (A(A^H A)^{-1} A^H)^H$
$= ((A^H A)^{-1} A^H)^H A^H$
$= A((A^H A)^{-1})^H A^H$

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

3.47iii) $rank(P) = rank(A(A^H A)^{-1} A^H)$. We know that $rank(A) = n$ and that $rank(A^H A) = rank(A)$, so $rank(A^H A) = n$. In addition, because $A^H A$ is invertible, we know its inverse also has rank n. By the lemma, we can let $B = (A^H A)^{-1}$ and know that $rank(A(A^H A)^{-1}) = n$. Then we know that $rank(A^H) = rank(A) = n$ and that A^H is a n x m matrix with rank n. Therefore, we can apply the lemma again, letting A^H be B this time, and conclude that $rank(A(A^H A)^{-1} A^H) = n$.

Lemma: If A is m x n and B is n x k with rank n, then $rank(AB) = rank(A)$.

3.48i) $P(kA) = \frac{kA+(kA)^T}{2} = k\frac{A+A^T}{2} = kP(A)$
$P(A+B) = \frac{A+B+(A+B)^T}{2} = \frac{A+A^T}{2} + \frac{B+B^T}{2} = P(A) + P(B)$

3.48ii) $P^2 = \frac{\frac{A+A^T}{2}+(\frac{A+A^T}{2})^T}{2} = 2\frac{(\frac{A+A^T}{2})}{2} = \frac{A+A^T}{2}$

3.48iii) $P^* = P^H = P^T = (\frac{A+A^T}{2})^T = \frac{A+A^T}{2}$

3.48iv) $x \in N(P) \iff \frac{A+A^T}{2}x = 0 \iff \frac{A+A^T}{x} = 0 \iff Ax = -A^T x \iff A = -A^T$.

3.48v) $x \in R(P) \iff \frac{A+A^T}{2}x = Ax \iff A + A^T x = 2Ax \iff A = A^T$.

3.48vi) $\|A - P(A)\|_F$
$= \sqrt{tr[(A-P(A))^T(A-P(A))]}$
$= \sqrt{tr[(A^T - P(A)^T)(A-P(A))]}$
$= \sqrt{tr[A^T A - A^T P(A) - P(A)^T A + P(A)^T P(A)]}$
$= \sqrt{tr[A^T A - A^T(\frac{A+A^T}{2}) - (\frac{A+A^T}{2})^T A + (\frac{A+A^T}{2})^2]}$
$= \sqrt{tr[\frac{4A^T A - 2A^T A - 2(A^T)^2 - 2A^2 - 2A^T A + A^2 + 2AA^T + (A^T)^2}{4}]}$
$= \sqrt{\frac{tr(A^T A) - tr(A^2)}{4}}$

3.50) We can rewrite the equation fo the ellipse as $y^2 = \frac{1}{s} - \frac{r}{s}x^2$. Then, letting $\mathbf{b}$ be the vector of y^2 data points and A be the matrix of x data points, we get $\mathbf{b} = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}$

and A $= \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & \\ . & \\ . & \\ 1 & x_n \end{bmatrix}$, and $\mathbf{x} = \begin{bmatrix} \frac{1}{s} \\ \\ \frac{-r}{s} \end{bmatrix}$. The least squares solution is obtained by solving the normal equation $A^H Ax = A^H b$, or put differently, $\hat{x} = (A^H A)^{-1} A^H b$.