Teaching Plan, week 7

Plan for today: 1. Quiz (15 mins)
2. Orthogonal Transformations & Matrices
3. Least Squares
4. Determinants
5. Problems

1. Quiz

2. Orthogonal Transformations & Matrices

The linear transform \( T: \mathbb{R}^n \rightarrow \mathbb{R}^n \) is called orthogonal with orthogonal matrix \( A \) (where \( T(x) = Ax \)), if \( T \) preserves lengths, i.e., \( \|T(x)\| = \|x\| \) for all \( x \in \mathbb{R}^n \). Note: if \( T \) is orthogonal, then \( x \in \ker(T) \iff T(x) = 0 \iff \|T(x)\| = \|0\| = 0 \) since the only vector with length 0 is the zero vector. So \( T \) is bijective (since \( T \) maps from \( \mathbb{R}^n \) to \( \mathbb{R}^n \)).

If \( T \) is orthogonal and so are vectors \( \mathbf{v}, \mathbf{w} \in \mathbb{R}^n \), then

\[
\|T(\mathbf{v}) + T(\mathbf{w})\|^2 = \|T(\mathbf{v} + \mathbf{w})\|^2 = \|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 = \|T(\mathbf{v})\|^2 + \|T(\mathbf{w})\|^2
\]

Linearly, \( T \) preserves lengths by the Pythagorean Theorem.

Proving (by the Pythagorean theorem) that \( \langle \mathbf{v}, \mathbf{w} \rangle = 0 \). In fact,

\[
\forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^n, \|\mathbf{v} - \mathbf{w}\|^2 = (\mathbf{v} - \mathbf{w}) \cdot (\mathbf{v} - \mathbf{w}) = \mathbf{v} \cdot \mathbf{v} - 2 \mathbf{v} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{w} = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - 2 \mathbf{v} \cdot \mathbf{w}, \text{ so } \mathbf{v} \cdot \mathbf{w} = \frac{1}{2}(\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - \|\mathbf{v} - \mathbf{w}\|^2).
\]

Letting \( T(\mathbf{v}) \) and \( T(\mathbf{w}) \) assume the roles of \( \mathbf{v} \) and \( \mathbf{w} \) for the moment, we get

\[
T(\mathbf{v}) \cdot T(\mathbf{w}) = \frac{1}{2}(\|T(\mathbf{v})\|^2 + \|T(\mathbf{w})\|^2 - \|T(\mathbf{v}) - T(\mathbf{w})\|^2) = \frac{1}{2}(\|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 - \|\mathbf{v} - \mathbf{w}\|^2) = \mathbf{v} \cdot \mathbf{w}, \text{ so } T \text{ preserves dot products too.} \]
Since the standard unit basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of $\mathbb{R}^n$ are orthonormal, and $T$ preserves lengths (and thus also inner products), $T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)$ is also orthonormal (this is also a basis for $\mathbb{R}^n$). $T(\mathbf{e}_i) = A\mathbf{e}_i$ is the $i^{th}$ column of $A$, so we know that $A$'s columns form an orthonormal basis of $\mathbb{R}^n$. Conversely, if $A$'s columns form an orthonormal basis for $\mathbb{R}^n$, or (equivalently) if $T(\mathbf{e}_1), \ldots, T(\mathbf{e}_n)$ is an orthonormal basis for $\mathbb{R}^n$, then for all $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 \mathbf{e}_1 + \cdots + x_n \mathbf{e}_n$, we get $\|T(\mathbf{x})\|^2 = \|x_1 T(\mathbf{e}_1) + \cdots + x_n T(\mathbf{e}_n)\|^2 = x_1^2 + \cdots + x_n^2 = \|\mathbf{x}\|^2$, so $\|T(\mathbf{x})\| = \|\mathbf{x}\|$.

So $T$ preserves lengths, so $T$ is orthogonal.

If $A, B$ are orthogonal $n \times n$ matrices, then for all $\mathbf{x} \in \mathbb{R}^n$, $\|AB(\mathbf{x})\| = \|A(B\mathbf{x})\| = \|B\mathbf{x}\| = \|\mathbf{x}\|$, so $AB$ preserves lengths, so $AB$ is orthogonal. Furthermore, $A^{-1}$ exists (by an earlier remark, since $\ker(A) = \{0\}$), and for all $\mathbf{x} \in \mathbb{R}^n$, $\|A^{-1} \mathbf{x}\| = \|AA^{-1} \mathbf{x}\| = \|\mathbf{x}\|$, so $A^{-1}$ preserves lengths, so $A^{-1}$ is orthogonal as well. We've already discussed transposes. $A$ is symmetric if $A^T = A$, and skew-symmetric if $A^T = -A$. If $A$ is orthogonal, $A = [\mathbf{v}_1 \ldots \mathbf{v}_n]$, then

$$A^T A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \\ \vdots & \ddots & \ddots & \ddots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = I_n.$$ This calculation also shows that if $A$ is $n \times n$ and $A^T A = I_n$, then $A$'s columns are orthonormal, so $A$ is orthogonal. So if $A$ is orthogonal, then $A^{-1} = A^T$, and since $A A^{-1} = I_n$,
\(A^T\) was already shown to be orthogonal, we know \(A^T\) must be orthogonal as well. This means if \(A\)'s columns are an orthonormal basis of \(\mathbb{R}^n\), then so are its rows. Now consider orthogonal projections. If \(V\) is a subspace of \(\mathbb{R}^n\) with some orthonormal basis \(\vec{u}_1, \ldots, \vec{u}_m\) (\(m \leq n\)), then \(\forall \vec{x} \in \mathbb{R}^n\), \(\text{proj}_V(\vec{x}) = (\vec{x} \cdot \vec{u}_1)\vec{u}_1 + \cdots + (\vec{x} \cdot \vec{u}_m)\vec{u}_m = [\vec{u}_1, \ldots, \vec{u}_m] \begin{bmatrix} \vec{x} \cdot \vec{u}_1 \\ \vdots \\ \vec{x} \cdot \vec{u}_m \end{bmatrix} = [\vec{u}_1, \ldots, \vec{u}_m] \begin{bmatrix} \vec{u}_1^T \\ \vdots \\ \vec{u}_m^T \end{bmatrix} \vec{x} = QQT^{-1} \vec{x}\), where 

\(Q\) is the \(n \times m\) matrix defined as \(Q = [\vec{u}_1, \ldots, \vec{u}_m]\).

The projection matrix \(P = QQ^T\) is symmetric, since \(P^T = (QQ^T)^T = Q^TQ = QQ^T = P\). Finally, if \(A\) is any \(n \times k\) matrix, then \(\text{rank}(AT) = \text{rank}(A)\). The reason for this is as follows: say \(\text{rref}(A) = E_{m-1} \cdots E_1 \cdot A\). Suppose \(\text{rref}(A)\) has \(p\) pivots.

Put \(E = E_{m-1} \cdots E_1\), so \(E\) is invertible. Then \(p \leq n\) and \(\tilde{e}_i^T \cdot EA = 0\) for \(i = 1, 2, \ldots, n\). Also, \(\forall \vec{x} \in \mathbb{R}^n\), \(\tilde{e}_i^T \cdot EA = 0\) implies \(E_{(i)} \cdot \text{span}(\vec{e}_{p+1}, \ldots, \vec{e}_n) = 0\) since the \(j\)th component of \(\vec{x}\) must be 0 for \(j \leq p\), by looking at where the pivots are. So \(\{\vec{e}_{p+1}, \ldots, \vec{e}_n\}\) is a basis for \(\ker((EA)^T) = \ker(A^TE^T)\), which we get by taking transposes of \(\tilde{e}_i^T \cdot EA = 0\). Then since \(E^T\) is invertible, \(c_{p+1}ET\vec{e}_{p+1} + \cdots + c_nET\vec{e}_n = \vec{0}\) implies \(c_{p+1} \tilde{e}_{p+1} + \cdots + c_n \tilde{e}_n = (E^T)^{-1} \vec{0} = \vec{0} \Rightarrow c_{p+1} = \cdots = c_n = 0\) by LI-ness of the \(\tilde{e}_i\)s, \(i = p+1, \ldots, n\). So \(\{ET\vec{e}_{p+1}, \ldots, ET\vec{e}_n\}\) is LI. If \(A^T \vec{x} = \vec{0}\), then \(\vec{x} = ETT^{-1} \vec{x}\), so \(\ker(AT) = \text{span}(ET\vec{e}_{p+1}, \ldots, ET\vec{e}_n)\), which means...
\[
\text{dim (ker}(A^T)) = n - p, \text{ so by rank-nullity, rank}(A^T) = \\
\text{dim (Im}(A^T)) = n - \text{dim (ker}(A^T)) = n - (n - p) = p = \text{\# of pivots in matrix } A
\]
\[= \text{dim (Im (A))} = \text{rank (A)}. \text{ So } \text{rank}(A^T) = \text{rank}(A).
\]
If \( V \) is a subspace of \( \mathbb{R}^n \) with orthonormal basis \( \mathbf{u}_1, \ldots, \mathbf{u}_m \), then put \( Q = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix} \). Then \( \mathbf{V} \cong \mathbb{R}^m \), the orthogonal projection of \( \mathbf{x} \) onto \( V \) is given by:
\[
\text{proj}_V(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{u}_1) \mathbf{u}_1 + \cdots + (\mathbf{x} \cdot \mathbf{u}_m) \mathbf{u}_m
\]
\[
= \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \mathbf{x} \\ \vdots \\ \mathbf{u}_m^T \mathbf{x} \end{bmatrix}
\]
\[
= \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_m^T \end{bmatrix} \mathbf{x}
\]
\[
= QQ^T \mathbf{x}. \text{ This means the projection matrix can be easily found once we have an orthonormal basis for the subspace we're projecting (orthogonally) onto. Let's do the case } V = \text{span} \{ \mathbf{u} \} \text{ in } \mathbb{R}^2, \text{ where } \|\mathbf{u}\| = 1. \text{ The set } \{ \mathbf{u} \} \text{ is of course orthonormal (no other vectors to dot with)}.
\]
\[
\text{and so } \text{proj}_V(\mathbf{x}) = [\mathbf{u}] [\mathbf{u}^T] \mathbf{x} = [u_1] [u_1 u_2] \mathbf{x} = [u_1^2 \ u_1 u_2 \ u_2^2] \mathbf{x}, \\
\]
\[
(\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix})
\]
which we are familiar with from section 2.2.
3. Least Squares

First, for any n x k matrix A, \((\text{Im} A)^\perp = \ker(A^T)\)
if \(\tilde{x} \in (\text{Im} A)^\perp\), then \((A\tilde{x}) \cdot \tilde{x} = 0\) for all \(y \in \mathbb{R}^k\), so
\((A\tilde{x}) \cdot \tilde{x} = 0\). Thus if \(A = [\tilde{v}_1, \ldots, \tilde{v}_k]\), then \(A\tilde{e}_i = \tilde{v}_i\), so
\(\tilde{v}_i \cdot \tilde{x} = 0\) for \(i = 1, \ldots, k\), which means \(\tilde{v}_i^T \tilde{x} = 0\), and so
\[A^T \tilde{x} = \begin{bmatrix} \tilde{v}_1^T \\ \vdots \\ \tilde{v}_k^T \end{bmatrix} \tilde{x} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \tilde{0},\] so \(\tilde{x} \in \ker(A^T)\). Conversely, if
\(\tilde{x} \in \ker(A^T)\), then \(A^T \tilde{x} = \tilde{0}\), so \(y \cdot (A^T \tilde{x}) = y^T A^T \tilde{x} = 0\), so
\((A\tilde{y})^T \tilde{x} = 0\), which means \(\tilde{x} \in (\text{Im} (A))^\perp\), since \(\tilde{y} \in \mathbb{R}^k\) was arbitrary, so \((\text{Im} (A))^\perp = \ker(A^T)\).

Next, \(\ker(A) = \ker(A^T A)\): if \(\tilde{x} \in \ker(A)\), then \(A\tilde{x} = \tilde{0}\), so
\[A^T A \tilde{x} = A^T (A\tilde{x}) = A^T \tilde{0} = \tilde{0},\] so \(\tilde{x} \in \ker(A^T A)\). Conversely, if \(\tilde{x} \in \ker(A^T A)\), then \(A^T A \tilde{x} = \tilde{0}\), so \(A\tilde{x} \in \ker(A^T)\) and \(A\tilde{x} \in \text{Im}(A)\) and since \((\text{Im}(A))^\perp = \ker(A^T)\), and the only element a subspace has in common with its orthogonal complement is \(\tilde{0}\), we know \(A\tilde{x} = \tilde{0}\), so \(\tilde{x} \in \ker(A)\). Thus \(\ker(A) = \ker(A^T A)\) (works for any size matrix). Since \(A\) and \(A^T A\) have the same number of columns, by rank-nullity \(A\) and \(A^T A\) must have the same rank.
If $A$ is $m \times k$ and $\mathbf{b} \in \mathbb{R}^m$, then $A\mathbf{x} = \mathbf{b}$ might have no solutions $\mathbf{x}$. So instead of looking for exact solutions, we look for close solutions, with $A\mathbf{x}$ as close to $\mathbf{b}$ as possible. $A\mathbf{x}$ lies in $\text{im}(A)$, so the closest $A\mathbf{x}$ will ever get to $\mathbf{b}$ is $\text{proj}_V(\mathbf{b})$, where $V = \text{im}(A)$. That $\text{proj}_V(\mathbf{b})$ is really closer to $\mathbf{b}$ (in distance) than every other vector in $V$ was the first problem in week 6's discussion notes.

$\mathbf{b}'' = \text{proj}_V(\mathbf{b})$ is the component of $\mathbf{b}$ in $V$, and $\mathbf{b}^\perp$ is the component of $\mathbf{b}$ orthogonal to $V$. So $\mathbf{b} = \mathbf{b}'' + \mathbf{b}^\perp$, and the equation we want to solve is $A\mathbf{x} = \mathbf{b}''$, which will always have a solution, since $\mathbf{b}''$ lies in $\text{im}(A)$ already. Left-multiplying by $A^\top$ kills off $\mathbf{b}^\perp$, and that's why we do it: $A^\top(\mathbf{b}) = A^\top(\mathbf{b}'' + \mathbf{b}^\perp) = A^\top \mathbf{b}'' + A^\top \mathbf{b}^\perp = A^\top \mathbf{b}''$.

Also, if $A\mathbf{x} = \mathbf{b}''$, then $A^\top A\mathbf{x} = A^\top \mathbf{b}'' = A^\top \mathbf{b}$, and if $A^\top A\mathbf{x} = A^\top \mathbf{b}$, then $A^\top A\mathbf{x} = A^\top \mathbf{b}''$, so $A^\top(A\mathbf{x} - \mathbf{b}'') = \mathbf{0}$. Since $\mathbf{b}'' \in \text{im}(A)$, we have $\mathbf{b}'' = A\mathbf{y}$ for some $\mathbf{y} \in \mathbb{R}^k$, so $A^\top(A\mathbf{x} - \mathbf{y}) = \mathbf{0}$, which means $A\mathbf{x} - \mathbf{y} \in \ker(A^\top A) = \ker(A^\top)$, so $A(A\mathbf{x} - \mathbf{y}) = \mathbf{0}$. Then $A\mathbf{x} = A\mathbf{y} = \mathbf{b}''$. This shows that $A\mathbf{x} = \mathbf{b}''$ and $A^\top A\mathbf{x} = A^\top \mathbf{b}$ have the same set of solutions. So to solve the least-squares problem, just solve $A^\top A\mathbf{x} = A^\top \mathbf{b}$ for $\mathbf{x}$. This last equation is known as the "normal equation." We know $\text{rank}(A^\top A) = \text{rank}(A)$, so if $\text{rank}(A) = k$, then $\text{rank}(A^\top A) = k$, so $A^\top A$ is invertible, and thus the normal equation becomes $\mathbf{x} = (A^\top A)^{-1} A^\top \mathbf{b}$. Since $A$ might not be square, $A^\top$ might not exist, though...
This also shows that $V \subseteq \mathbb{R}^n$, if $V$ is a subspace of $\mathbb{R}^n$ with basis $\mathbf{v}_1, \ldots, \mathbf{v}_k$, then form $A = [\mathbf{v}_1, \ldots, \mathbf{v}_k]$. Since the $\mathbf{v}_i$'s are LI, $\ker(A) = \{ \mathbf{0} \}$, which means by rank-nullity, $\text{rank}(A) = k$. Then $A^TA$ is invertible (as just discussed), so $\mathbf{x} = (A^TA)^{-1}A^T\mathbf{b}$ is the unique solution to $A\mathbf{x} = b^n = \text{proj}_V(\mathbf{b})$ (note $\text{im}(A) = V$), which implies $\text{proj}_V(\mathbf{b}) = A\mathbf{x} = A(A^TA)^{-1}A^T\mathbf{b}$. Thus, the matrix of orthogonal projection onto $V$ is $A(A^TA)^{-1}A^T$. This works even if $A$'s columns, aka the basis vectors of $V$, are not orthonormal! If they were orthonormal, then $A^TA$ would be the identity $I_k$, so we would recover our formula $A(A^TA)^{-1}A^T = A(I_k)^{-1}A^T = AA^T$ ($P=QQ^T$...).

4. Determinants

Defined only for square matrices, difficult to compute numerically, contains some deep secrets of the matrix. Let $A$ be an $n \times n$ matrix.

A pattern is a way to choose $n$ entries of $A$ that get their own row & column. We define $\text{prod}(P)$ to be the product of the $n$ entries in the pattern $P$ in matrix $A$. Two entries in a pattern are inverted if one of them is located to the right and above the other in the matrix. The signature of a pattern $P$ is defined as $\text{sgn}(P) = (-1)^{\text{# of inversions in } P}$.

The determinant of $A$ is defined as $\det(A) = \sum_{\text{P a pattern in } A} \text{sgn}(P) \text{prod}(P)$.

Each $n\times n$ matrix has $n! = n(n-1)(n-2)\cdots(3)(2)(1)$ patterns in it, since there are $n$ choices for the element in the first column, then $(n-1)$ choices left in the second column (after removing the row our first element was in), $(n-2)$ choices for the element in the third column, etc.
By this definition, it follows that the determinant of any (upper or lower) triangular matrix is just the product of its diagonal entries. For 2×2 matrices \( A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \), \( \det(A) = ad - bc \) by the determinant formula (there are only two patterns) and we know from long ago that \( A \) is invertible iff \( \det(A) \neq 0 \). For 3×3 matrices \( A = [\vec{u} \ \vec{v} \ \vec{w}] \), where \( \vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^3 \), \( \det(A) = \vec{u} \cdot (\vec{v} \times \vec{w}) \) ends up being true, which can be verified by expanding this out using the entries of \( \vec{u}, \vec{v}, \) and \( \vec{w} \) and checking that the resulting expression is equal to the definition of \( \det(A) \) using \( A \)'s six patterns.

\((A \text{ not invertible}) \iff (\vec{u} \text{ in plane spanned by } \vec{v}, \vec{w} \iff \vec{v}, \vec{w} \text{ are LI, or } \vec{v}, \vec{w} \text{ are collinear})\)

\(\iff (\vec{u} \text{ is orthogonal to the vector } \vec{v} \times \vec{w} \text{ normal to the plane spanned by } \vec{v}, \vec{w} \iff \vec{v}, \vec{w} \text{ are LI, or } \vec{v} \times \vec{w} = \vec{0})\)

\(\iff (\det(A) = 0)\). So in the 3×3 case, \( A \) is invertible iff \( \det(A) \neq 0 \), as well.

5. Problems

a) If \( \vec{u}_1, ..., \vec{u}_m \in \mathbb{R}^n \) are orthonormal, and \( \vec{x} \in \mathbb{R}^n \) is arbitrary, show that \( p = (\vec{x} \cdot \vec{u}_1)^2 + ... + (\vec{x} \cdot \vec{u}_m)^2 \leq \|\vec{x}\|^2 \), and we get equality iff \( \vec{x} \in \text{span}(\vec{u}_1, ..., \vec{u}_m) \).

Sol'n: Put \( V = \text{span}(\vec{u}_1, ..., \vec{u}_m) \). Then \( \text{proj}_V(\vec{x}) = (\vec{x} \cdot \vec{u}_1)\vec{u}_1 + ... + (\vec{x} \cdot \vec{u}_m)\vec{u}_m \), orthonormal basis for \( V \) and \( p = \|\text{proj}_V(\vec{x})\|^2 \leq \|\vec{x}\|^2 \). Since \( \vec{x} = \vec{x} - \vec{x}' \) and by the Pythagorean theorem we have \( \|\vec{x}\|^2 = \|\vec{x} - \vec{x}'\|^2 + \|\vec{x}'\|^2 = \|\vec{x} - \vec{x}'\|^2 + \|\text{proj}_V(\vec{x})\|^2 = \|\vec{x} - \vec{x}'\|^2 + p \), we get that \( p = \|\vec{x}\|^2 \iff \|\vec{x}'\| = 0 \iff \vec{x} = \vec{x}' \iff \vec{x} \in \text{span}(\vec{u}_1, ..., \vec{u}_m) \).
b) Perform Gram-Schmidt on the following basis for $\mathbb{R}^3$:

$\vec{v}_1 = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} b \\ c \\ 0 \end{bmatrix}$, $\vec{v}_3 = \begin{bmatrix} d \\ e \\ f \end{bmatrix}$, where $a, c, \text{ and } f$ are positive and the rest of the variables are arbitrary.

Soln: $\vec{u}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ (since $a > 0$, otherwise the top component would be $-1$)

$\vec{u}_2 = \frac{\vec{v}_2 - (\vec{v}_2 \cdot \vec{u}_1)\vec{u}_1}{\|\vec{v}_2\|}$ = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ (since we are just removing the component of $\vec{u}_2$ parallel to $\vec{u}_1$)

$\vec{u}_3 = \frac{\vec{v}_3 - (\vec{v}_3 \cdot \vec{u}_1)\vec{u}_1 - (\vec{v}_3 \cdot \vec{u}_2)\vec{u}_2}{\|\vec{v}_3\|}$ = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ (since we are removing the components of $\vec{v}_3$ parallel to $\vec{u}_1$ and $\vec{u}_2$, which amounts to zeroing the first two coordinates of $\vec{v}_3$, then scaling, and $c > 0$)

So, Gram-Schmidt yields $[\vec{u}_1, \vec{u}_2, \vec{u}_3]$ as an orthonormal set of vectors with the same span as $\vec{v}_1, \vec{v}_2, \vec{v}_3$ (and some other nice properties).

c) Consider vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in \mathbb{R}^4$. If entry $a_{ij}$ of $A = \begin{bmatrix} 3 & 5 & 11 \\ 5 & 9 & 20 \\ 0 & 12 & 49 \end{bmatrix}$ is $\vec{v}_i \cdot \vec{v}_j$, and $\text{proj}_V(\vec{v}_i)$, where $V = \text{span}(\vec{v}_2, \vec{v}_3)$.

Soln: One way to do this is to do Gram-Schmidt on $\vec{v}_1, \vec{v}_2, \vec{v}_3$ using the entries in $A$ in our calculations and keeping each $\vec{u}_i$ as a linear combo of the $\vec{v}_k$'s, then using the projection formula. This is the "brute-force" way, but a more elegant solution exists, even though it's harder to come up with. Form $B = [\vec{v}_2 \ \vec{v}_3]$. Then the projection formula when our subspace basis vectors $\vec{v}_2, \vec{v}_3$ are not orthonormal is:

$\text{proj}_V(\vec{v}_i) = B(B^TB)^{-1}B^T\vec{v}_i = [\vec{v}_2 \ \vec{v}_3][(\vec{v}_2^T \vec{v}_2)^{-1}\vec{v}_2^T][\vec{v}_2 \ \vec{v}_3]^T \vec{v}_i$

$= [\vec{v}_2 \ \vec{v}_3]\begin{bmatrix} \vec{v}_2^T \vec{v}_2 & \vec{v}_2^T \vec{v}_3 \\ \vec{v}_3^T \vec{v}_2 & \vec{v}_3^T \vec{v}_3 \end{bmatrix}^{-1}\begin{bmatrix} \vec{v}_2^T \vec{v}_i \\ \vec{v}_3^T \vec{v}_i \end{bmatrix}$

$= \frac{25}{41} \vec{v}_2 - \frac{41}{41} \vec{v}_3$. 

\[\text{proj}_V(\vec{v}_i) = B(B^TB)^{-1}B^T\vec{v}_i = [\vec{v}_2 \ \vec{v}_3][(\vec{v}_2^T \vec{v}_2)^{-1}\vec{v}_2^T][\vec{v}_2 \ \vec{v}_3]^T \vec{v}_i\]
d) Consider an invertible \( n \times n \) matrix \( A \) whose columns are orthogonal but not necessarily orthonormal. What does the QR factorization of \( A \) look like?

**Sln:** \( A = [ \overline{v}_1 \cdots \overline{v}_n ] \), where \( \overline{v}_i \cdot \overline{v}_j = 0 \) if \( i \neq j \). Since \( A \) is invertible, each \( \overline{v}_i \neq \overline{0} \), so \( \| \overline{v}_i \| > 0 \). \( \overline{u}_1 = \frac{\overline{v}_1}{\| \overline{v}_1 \|}, \overline{u}_2 = \frac{\overline{v}_2 - (\overline{v}_2 \cdot \overline{u}_1) \overline{u}_1}{\| \overline{v}_2 - (\overline{v}_2 \cdot \overline{u}_1) \overline{u}_1 \|} = \frac{\overline{v}_2}{\| \overline{v}_2 \|} \), since \( \overline{v}_2 \cdot \overline{u}_1 = \overline{v}_2 \cdot \left( \frac{\overline{v}_1}{\| \overline{v}_1 \|} \right) = \frac{\overline{v}_1 \cdot \overline{v}_2}{\| \overline{v}_1 \|} = 0 \) (\( \overline{u}_1 \) and \( \overline{v}_2 \) are collinear). Keep going in this manner to get that \( \overline{u}_i = \frac{\overline{v}_i}{\| \overline{v}_i \|} \) for \( i = 1, \ldots, n \). So \( Q = [ \overline{u}_1 \cdots \overline{u}_n ] \)

\[
\begin{bmatrix}
\frac{\overline{v}_1}{\| \overline{v}_1 \|} & \cdots & \frac{\overline{v}_n}{\| \overline{v}_n \|}
\end{bmatrix}, \text{ and } R = Q^T A = \begin{bmatrix}
\overline{u}_1^T / \| \overline{v}_1 \| & 0 & \cdots & 0 \\
0 & \overline{u}_2^T / \| \overline{v}_2 \| & \cdots & 0 \\
0 & 0 & \cdots & \overline{u}_n^T / \| \overline{v}_n \|
\end{bmatrix}
\]

Thus, \( [ \overline{v}_1 \cdots \overline{v}_n ] = \begin{bmatrix}
\begin{bmatrix}
\| \overline{v}_1 \| & 0 & \cdots & 0
\end{bmatrix} \\
\begin{bmatrix}
0 & \| \overline{v}_2 \| & \cdots & 0
\end{bmatrix} \\
\begin{bmatrix}
0 & 0 & \cdots & \| \overline{v}_n \|
\end{bmatrix}
\end{bmatrix} \), and \( R = \begin{bmatrix}
\begin{bmatrix}
\| \overline{v}_1 \| & 0 & \cdots & 0
\end{bmatrix} \\
\begin{bmatrix}
0 & \| \overline{v}_2 \| & \cdots & 0
\end{bmatrix} \\
\begin{bmatrix}
0 & 0 & \cdots & \| \overline{v}_n \|
\end{bmatrix}
\end{bmatrix} \).

e) If \( n \times n \) matrices \( A \) and \( B \) are orthogonal, which of the following must be orthogonal as well?

- \( 3A \), \( -B \), \( AB \), \( A + B \), \( B^{-1} \), \( B^T AB \), \( A^T \)

**Sln:** not \( 3A \), since \( A = (1 \ 0) \) is orthogonal but \( 3A = (3 \ 0) \) isn’t (counterexample).

- yes \( -B \), since \( A \times e_1^T \in \mathbb{R}^n \), \( \| -B e_1 \| = \| B e_1 \| = \| e_1 \| \), so \( B \) preserves lengths and is thus orthogonal.

- yes \( AB \), since \( A \times e_1^T \in \mathbb{R}^n \), \( \| A B e_1 \| = \| A (B e_1) \| = \| B e_1 \| = \| e_1 \| \), since \( B \) preserves lengths.

- not \( A + B \), since \( A = (0 \ 1) \), \( B = (1 \ 0) \) are orthogonal, but \( A + B = (1 \ 1) \) isn’t.

- yes \( B^{-1} \), since \( A \times e_1^T \in \mathbb{R}^n \), \( \| B^{-1} e_1 \| = \| B e_1 \| = \| e_1 \| \).

- yes \( B^T A B \), by combining products and inverses. \( (A B) \times e_1^T \in \mathbb{R}^n \), \( \| B^{-1} (A B) e_1 \| = \| A B e_1 \| = \| e_1 \| \).

- yes \( A^T \), since \( A \) orthogonal \( \Rightarrow A \)'s columns are orthogonal \( \Rightarrow A^T A = I_n \)

\( \Rightarrow A^T = A^{-1} \) \( \Rightarrow A^T \) is orthogonal, since \( A^{-1} \) is (by a previous part).
1) If \( n \times n \) matrices \( A \& B \) are symmetric, which of the following must be symmetric as well? \( A, B, AB, A+B, B^{-1}, A^{10}, 2I_n + 3A - 4A^2A^2 \)  

Solution: all except for \( AB \). Matrices are symmetric iff they equal their transposes. \((AB)^T = B^TA^T = BA\), which hints to us to find a counterexample. 

If \( A = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix}, B = \begin{pmatrix} 6 & 4 \\ 0 & 5 \end{pmatrix}\), then \( A, B \) are symmetric, but \( AB = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 6 & 4 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 22 & 18 \\ 24 & 20 \end{pmatrix} \) is not. 

\((3A)^T = 3A^T = 3A, \quad (-B)^T = -B^T = -B, \quad (A+B)^T = A^T + B^T = A + B, \quad (B^{-1})^T = (B^T)^{-1} = B^{-1}\) 

\((A^{10})^T = (A^{10})^T = A^{T10} = A^{10}, \quad (2I_n + 3A - 4A^2)^T = 2I_n^T + 3A^T - 4A^2A^T\) 

\( = 2I_n + 3A - 4A^2, \) and \((AB^2)^T = (ABBA)^T = A^TB^TB^TA = ABB^A = A^2B^2A^T\) 

Side note: if \( A, B, C, D, E \) are any matrices such that the matrix product \( ABCDE \) is defined, then \((ABCDE)^T = E^TD^CT^BT^AT^A\).

2) If \( m \times n \) matrices \( A \& B \) are arbitrary, which of the following must be symmetric? \( A^TA, B^TB, A - A^T, A^TBA, A^TB^TB^TA, B(A + A^T)B^T\) 

Solution: yes \( A^TA : (A^TA)^T = A^T^TA^T = A^TA\) 

yes \( B^TB : (B^TB)^T = B^T^TB^T = B^TB\) 

\( \Rightarrow \) but it is skew-symmetric. 

no \( A - A^T : (A - A^T)^T = A^T - A = -(A - A^T) \) (take \( A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \) for a counterexample). 

no \( A^TBA : (A^TBA)^T = A^T^TB^A^T = A^TB^A \) (take \( A = I_n, B \) any nonsymmetric matrix). 

\( \Rightarrow \) yes \( A^TBA : (A^TBA)^T = A^TB^TBA^T = A^TB^TBA \) 

yes \( B(A + A^T)B^T : (B(A + A^T)B^T)^T = B^T^T(A^T + A^TT)^T = B( A + A^T)B^T\)

3) Are the rows of an orthonormal matrix \( A \) necessarily orthonormal? 

Solution: yes, since \( A^T \) was proven (above) to be orthonormal, so \( A \)'s columns, i.e. \( A \)'s rows, form an orthonormal basis of \( \mathbb{R}^n \).
i) Find all orthogonal $2 \times 2$ matrices.
Soln: Set $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. If $A$ is to be orthogonal, then we must have

$$a^2 + c^2 = 1 = b^2 + d^2 \quad \text{and} \quad ab + cd = 0,$$

and conversely if these conditions are met then $A$ is orthogonal. If $c = 0$, then $a = \pm 1$, so $b = 0$. This forces $a = \pm 1$, $d = \pm 1$. If $a = 0$, then $c \neq 0$, so $d = 0$ and $A = \begin{pmatrix} a & 0 \\ c & d \end{pmatrix}$. If neither $a$ nor $c$ is zero, then $\frac{b}{c} = -\frac{d}{a}$, so $b^2 = 1 - d^2 = 1 - \left(\frac{b}{c}\right)^2 = 1 - \frac{a^2 c^2}{a^2 + c^2}$

so $b^2 (1 + \frac{a^2 c^2}{a^2 + c^2}) = 1$, so $b^2 = \frac{1}{1 + \frac{a^2 c^2}{a^2 + c^2}} = \frac{c^2}{a^2 + c^2} = c^2$, so $b = \pm c$, so

$$-\frac{d}{a} = \pm 1,$$

so $d = \mp a$. Thus if $a^2 + c^2 = 1$, then $(b = c, d = -a)$, or $(b = c, d = a)$.

Combining with the other cases, all $2 \times 2$ orthogonal matrices take the form

$$\begin{pmatrix} a & c \\ -c & a \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} a & -c \\ c & a \end{pmatrix},$$

where $a^2 + c^2 = 1$, which are either reflections or rotations.

j) Find all $3 \times 3$ matrices of the form $\begin{pmatrix} a & b & 0 \\ c & d & 0 \\ e & f & 0 \end{pmatrix}$.
Soln: $\begin{pmatrix} a \\ c \\ e \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0$ and $\begin{pmatrix} b \\ d \\ f \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = 0$ must be 0. Then our matrix must look like $\begin{pmatrix} a & b & 0 \\ 0 & 0 & 0 \\ e & f & 0 \end{pmatrix}$, at which point we use the previous problem to say that $a^2 + e^2 = 1$ and

$$\begin{pmatrix} a & -e & 0 \\ 0 & 0 & 0 \\ e & a & 0 \end{pmatrix}, \begin{pmatrix} a & e & 0 \\ 0 & 0 & 0 \\ e & a & 0 \end{pmatrix}$$

captures the full set of such orthogonal matrices.

k) Find an orthogonal transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ st. $T \begin{pmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix}$.
Soln: We can reflect about the line $l = \text{span} \left( \begin{pmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \right)$. Set $V = \text{span} \left( \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right)$.

$$T(x) = 2 \text{proj}_V(x) - x = 2 \begin{pmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \left( \begin{pmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \right)^T \begin{pmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} - x = \frac{1}{2} \begin{pmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} x^T \begin{pmatrix} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\sqrt{3}}{2} \frac{\sqrt{3}}{2} \frac{\sqrt{3}}{2} \\ \frac{1}{2} \frac{1}{2} \frac{1}{2} \\ \frac{1}{2} \frac{1}{2} \frac{1}{2} \end{pmatrix} x$$

l) Is there an orthogonal transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ st. $T \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ and $T \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}$?
Soln: no, because $T$ must preserve dot products (proven earlier), and

$$\begin{pmatrix} \frac{3}{2} \\ -\frac{3}{2} \\ 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{3}{2} \\ -\frac{3}{2} \\ 0 \end{pmatrix} = 0,$$

and yet $\begin{pmatrix} \frac{3}{2} \\ 2 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} \frac{3}{2} \\ 2 \\ 0 \end{pmatrix} = 6 \neq 0.$
w) If an n x n matrix $A$ is skew-symmetric, is $A^2$ necessarily symmetric?

Soln: Yes, because then $A^T = -A$, and so $(A^2)^T = (AA)^T = A^TA^T = (A)(-A) = AA = A^2$.

w) $W = \text{span}(\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -5 \end{bmatrix})$. Find the matrix of orthogonal projection onto $W$.

Soln: Put $A = \begin{bmatrix} 1 \\ 1 \\ -5 \end{bmatrix}$. Then \( \text{proj}_W(\vec{x}) = A(ATA)^{-1}A^T \vec{x} \)

\[
\begin{bmatrix} 1 \\ 1 \\ -5 \end{bmatrix} \left( \begin{bmatrix} 1 & 1 & -5 \\ 1 & 1 & -5 \\ 1 & 1 & -5 \end{bmatrix} \right)^{-1} \left( \begin{bmatrix} 1 \\ 1 \\ -5 \end{bmatrix} \right) \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}
\]

\[
= \frac{1}{400} \begin{bmatrix} 110 & -8 & 11 \\ -8 & 11 & -8 \\ 11 & -8 & 11 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{400} \begin{bmatrix} 110 & -8 & 11 \\ -8 & 11 & -8 \\ 11 & -8 & 11 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}
\]

\[
= \frac{1}{400} \begin{bmatrix} 26 & 12 & 24 \\ 12 & 74 & 32 \\ 24 & 32 & 108 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}
\]

So, the matrix of orthogonal projection onto $W$ is

\[
\begin{bmatrix} .26 & .12 & .24 \\ .12 & .74 & .32 \\ .24 & .32 & .108 \end{bmatrix}
\]

o) Let $A$ be the matrix of an orthogonal projection. Find $A^2$.

Soln: Geometrically, projecting onto the same subspace twice is the same as doing it once, since once the projection lies in the subspace, projecting that won't budge it. So $A^2 = A$. If $Q$'s columns are an orthonormal basis for $\text{im}(A)$, the space being projected onto, then $A = QQ^T$, so $A^2 = (QQ^T)(QQ^T) = QQ^T = Q(\text{Q}^TQ) = A$.

p) If $A$ is $n \times k$, find $\text{dim}(\text{im}(A)) + \text{dim}(\text{ker}(A^T))$.

Soln: $\text{im}(A)^\perp = \text{ker}(A^T)$, so $\text{dim}(\text{im}(A)) + \text{dim}(\text{ker}(A^T)) = \text{dim}(\text{im}(A)) + \text{dim}(\text{im}(A)^\perp) = n$, since $\text{im}(A)$ and $\text{im}(A)^\perp$ are complementary subspaces in $\mathbb{R}^n$ (so any basis for $\text{im}(A)$ union any basis for $\text{im}(A)^\perp$ is a basis for $\mathbb{R}^n$).
9) For which $n \times k$ matrices $A$ does the equation $\dim(\ker(A)) = \dim(\ker(A^T))$ hold?
Solution: $\ker(A^T) = (\im(A))^\perp$ and $\dim(\im(A)^\perp) = n - \dim(\im(A))$, so $\dim(\ker(A)) = \dim(\ker(A^T)) \iff \dim(\ker(A)) = \dim(\im(A)^\perp) \iff \dim(\ker(A)) = n - \dim(\im(A)) \iff \dim(\im(A)) + \dim(\ker(A)) = n$ $\iff n = k$, by rank-nullity.

r) If $A = QR$ is a QR factorization, what is the relationship between $A^T A$ and $R^T R$?
Solution: $A^T A = (QR)^T (QR) = R^T Q^T Q R = R^T R = I_k$.

s) If $Q_1$ and $Q_2$ are $n \times k$ matrices with orthonormal columns, and $S$ is a (necessarily $k \times k$) matrix s.t. $Q_1 = Q_2 S$, show that $S$ is orthogonal.
Hint: $Q_1^T Q_1 = I_k = Q_2^T Q_2^T$, and compute $(Q_2 S)^T (Q_2 S)$.
Solution: $S^T S = S^T (Q_2^T Q_2) S = (Q_2 S)^T (Q_2 S) = Q_1^T Q_1 = I_k$, which implies that $S$'s columns are orthonormal, and since $S$ is $k \times k$, $S$ is orthogonal.

t) Find a basis of the space $V$ of all symmetric $3 \times 3$ matrices, and find $\dim(V)$.
Solution: every symmetric $3 \times 3$ matrix is determined by its lower diagonal entries, so a basis for $V$ is $\{[1, 0, 0], [0, 1, 0], [0, 0, 1], [0, 0, 0], [0, 0, 0], [0, 0, 0]\}$.
Thus $\dim(V) = 6$.

u) Find the dimension of the space of all symmetric $n \times n$ matrices.
Solution: it's equal to the number of lower triangular elements, call it $L$.
Then $2L = (\# \text{of diagonal elements}) = n^2$ $\Rightarrow L = \frac{n^2}{2}$, total $\# \text{of elements in an $n \times n$ matrix,}$
because diagonals in $L$ get double-counted.
$\Rightarrow L = \frac{n^2 + n}{2} = \frac{n(n+1)}{2}$, Note that this will be an integer because one of $n, n+1$ must be even.
V) Find all orthogonal 2x2 matrices $A$ s.t. all the entries of $10A$ are integers and s.t. both entries in the first column are positive.

Soln: $A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ or $\begin{pmatrix} a & b \\ b & -a \end{pmatrix}$ for some $a, b > 0$ s.t. $a^2 + b^2 = 1$

and $10a, 10b$ are integers. So $(10a)^2 + (10b)^2 = 100$. Say $c = 10a, d = 10b$. Then $c, d \in \mathbb{N} = \{1, 2, 3, \ldots\}$ and $c^2 + d^2 = 100$. Through trial and error, much strife and personal sacrifice, we get that $c = 8 \& d = 6$ or $c = 6 \& d = 8$ (these are the only positive integer solutions). So either $(a = 8/10, b = 6/10)$ or $(a = 6/10, b = 8/10)$. Thus the possibilities for $A$ are

$\begin{pmatrix} 8/10 & -6/10 \\ 6/10 & 8/10 \end{pmatrix}$, $\begin{pmatrix} 8/10 & 6/10 \\ -6/10 & 8/10 \end{pmatrix}$, and $\begin{pmatrix} 6/10 & -8/10 \\ 8/10 & 6/10 \end{pmatrix}$.

w) Let $V$ be the solutions (in $\mathbb{R}^4$) to \[
\begin{align*}
\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix} + \begin{bmatrix} 1 & 2 & 5 & 4 \end{bmatrix} = 0 \\
\begin{bmatrix} x_1 & 2x_2 + 5x_3 + 4x_4 = 0 
\end{align*}
\]

Find a basis for $V^\perp$.

Soln: $V = \ker\left(\begin{bmatrix} 1 & 2 & 5 & 4 \end{bmatrix}\right)$, so $V^\perp = \ker\left(\begin{bmatrix} 1 & 2 & 5 & 4 \end{bmatrix}^\top\right)$

= $\text{im}\left(\begin{bmatrix} 1 & 1 & 2 & 5 \\\n5 & 1 & 2 & 5 \end{bmatrix}\right)$, for which $\left\{\begin{bmatrix} 1 \\
1 \end{bmatrix}, \begin{bmatrix} 2 \\
5 \end{bmatrix}\right\}$ is a basis.

X) If $A$ is non symmetric, what is the relationship between $\text{im}(A)$ and $\ker(A)$?

Soln: $A^T = A$, so $\text{im}(A) = (\ker(A^T))^\perp = (\ker(A))^\perp$, so $\text{im}(A)$ and $\ker(A)$ are orthogonally complemented subspaces of $\mathbb{R}^n$ if $A$ is non symmetric. (Note that $B^\top = B$ and $(V^\perp)^\perp = V$ for any matrix $B$ and subspace of $\mathbb{R}^n$).
1) If $A$ is $n \times k$ with $\ker(A) = \{0\}$, show that there exists a $k \times n$ matrix $B$ such that $BA = I_k$. Hint: use $\text{rank}(A^TA) = \text{rank}(A)$. Solution: $\text{rank}(A) = k - 0 = \text{rank}(A^TA)$, so $A^TA$, which is $k \times k$, has full rank, and is thus invertible. So \( (A^TA)^{-1} A^TA = I_k \).

Set $B = (A^TA)^{-1} A^T$. Then $B$ is $k \times n$ and $BA = I_k$.

2) Use the formula $(\text{im} A)^\perp = \ker(A^T)$ to prove that $\text{rank}(A) = \text{rank}(A^T)$.

Solution: if $A$ is $n \times k$, then $\text{rank}(A) = k - \dim(\ker(A)) = k - \dim(\text{im}(A^T))$.

This proof is much slicker than the one given earlier in these notes. But it's also less intuitive.

aa) Let $A$ be an $n \times k$ matrix. Is $\text{rank}(A) = \text{rank}(A^TA)$?

Solution: Yes, because $\text{rank}(A^TA) = k - \dim(\ker(A^TA)) = k - \dim(\ker(A)) = k - [k - \text{rank}(A)] = \text{rank}(A)$, as $\ker(A) = \ker(A^TA)$.

ab) Using the last two problems, show that if $A$ is $n \times k$, then $\text{rank}(A^TA) = \text{rank}(AA^T)$.

Solution: $\text{rank}(A^TA) = \text{rank}(A) = \text{rank}(AT) = \text{rank}(A^TA) = \text{rank}(AA^T)$. 
ac) Find the least-squares solution of the system $A\mathbf{x} = \overline{b}$, where

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \overline{b} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}.$$\

Soh: Solve the normal equation $A^TA\mathbf{x} = A^T\overline{b}$. So

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 3 \\ 3 \end{bmatrix},$$

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 3 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

is the least-squares solution.

ad) If $A$ is $3 \times 2$ and the least squares solution to $A\mathbf{x} = \overline{b}$ is

$$\mathbf{x}^* = \begin{bmatrix} 7 \\ 11 \end{bmatrix},$$

find the least-squares solution to $SA\mathbf{x} = S\overline{b}$, where $S$ is some $3 \times 3$ orthogonal matrix.

Soh: The least-squares solution to our system $SA\mathbf{x} = S\overline{b}$ is given by the normal equation $(SA)^TSA\mathbf{x} = (SA)^TS\overline{b}$, so

$$A^TS^TS\mathbf{x} = A^TS^TS\overline{b},$$

is the normal equation to the least-squares problem $A\mathbf{x} = \overline{b}$, which we know has solution $\mathbf{x}^* = \begin{bmatrix} 7 \\ 11 \end{bmatrix}$, so this is our desired solution to both least-squares problems.

ae) Fit a quadratic polynomial to the data points $(0,27), (1,0), (2,0), (3,0)$, using least squares.

Soh: Let $f(t) = at + bt + ct^2$ be the quadratic polynomial. We'd like to solve

$$\begin{cases} 27 = a \\ 0 = a + b + c \\ 0 = a + 2b + 4c \\ 0 = a + 3b + 9c \end{cases},$$

but this system is overdetermined (more constraints than variables).

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \\ 1 & 3 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 27 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$\

The normal equation is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 3 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 27 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$\

so the least-squares solution is $f(t) = \frac{513}{20} - \frac{567}{20} t + \frac{27}{2} t^2$. 

If \( x_1, \ldots, x_n \) are any real numbers, then
\[
\left( \sum_{k=1}^{n} x_k \right)^2 \leq n \sum_{k=1}^{n} (x_k^2)
\]
holds. True or False?

Sohln: 
\[
\sum_{k=1}^{n} x_k = \left( \begin{array}{c} x_1 \\ \vdots \\ x_n \end{array} \right) \cdot \left( \begin{array}{c} 1 \\ \vdots \\ 1 \end{array} \right) = \left\| \left( \begin{array}{c} x_1 \\ \vdots \\ x_n \end{array} \right) \right\| \left\| \left( \begin{array}{c} 1 \\ \vdots \\ 1 \end{array} \right) \right\| = \sqrt{\sum_{k=1}^{n} x_k^2} \cdot \sum_{k=1}^{n} 1^2 \]
\[
= n \sqrt{\sum_{k=1}^{n} x_k^2}
\]
Square both sides to get the result. So, this is true.