Discussion 2

1. Proof of the Fundamental Theorem of Linear Algebra

In this question, we will prove the fundamental theorem of linear algebra. For any $A \in \mathbb{R}^{m\times n}$, let $\mathcal{N}(A)$, $\mathcal{R}(A)$, and $\mathrm{rank}(A)$ denote the null space, range and rank of A respectively. For any subspace, $\mathcal{S}$ with dimension, $\dim(\mathcal{S})$, let $\mathcal{S}^\perp$ denote its the subspace orthogonal to $\mathcal{S}$. The fundamental theorem of linear algebra states that,

$$\mathcal{N}(A) \oplus \mathcal{R}\left(A^\top\right) = \mathbb{R}^n. \tag{1}$$

The proof technique we employ will first show that,

$$\mathcal{N}(A) = \mathcal{R}\left(A^\top\right)^\perp. \tag{2}$$

Then we will prove that we can find orthonormal vectors $\vec{e}_1, \vec{e}_2, \ldots \vec{e}_n$ such that $\mathcal{N}(A) = \mathrm{span}(\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_\ell)$ and $\mathcal{R}\left(A^\top\right) = \mathrm{span}(\vec{e}_{\ell+1}, \vec{e}_{\ell+2}, \ldots, \vec{e}_n)$. As a corollary we get the rank-nullity theorem:

$$\dim(\mathcal{N}(A)) + \mathrm{rank}(A) = n. \tag{3}$$

(a) First, show that $\mathcal{N}(A) \subseteq \mathcal{R}\left(A^\top\right)^\perp$.

HINT: Consider $\vec{u}$ in $\mathcal{N}(A)$, $\vec{v} \in \mathcal{R}\left(A^\top\right)$ and show that $\vec{u}^\top \vec{v} = 0$.

Solution: Take a vector $\vec{u}$ from the nullspace: $A\vec{u} = 0, \vec{u} \neq 0$, and a vector $\vec{w}$ from the rowspace: $A^\top \vec{v} = \vec{w}$. Now we can show orthogonality:

$$\langle \vec{w},\ \vec{u}\rangle = \left\langle A^\top \vec{v},\ \vec{u}\right\rangle = \vec{v}^\top A\vec{u} = \vec{v}^\top \vec{0} = 0 \tag{4}$$

Therefore, $\vec{u} \in \mathcal{R}\left(A^\top\right)^\perp$. Since this holds for every u in $\mathcal{N}(A)$ we have $\mathcal{N}(A) \subseteq \mathcal{R}\left(A^\top\right)^\perp$. Note that this is not an equality since we only proved that every vector in the nullspace is orthogonal , not that these vectors are the **only** orthogonal ones.

(b) Now show that: $\mathcal{R}\left(A^\top\right)^\perp \subseteq \mathcal{N}(A)$.

HINT: Show that any vector $\vec{v}$ that is orthogonal to all vectors in the range of $A^\top$ satisfies $A\vec{v} = 0$. To do this, consider $\vec{v} \in R(A^\top)^\perp$ and what it implies for $\vec{v}^\top A^\top$.

Solution: Consider $\vec{v} \in \mathcal{R}\left(A^\top\right)^\perp$ and $\vec{u} \in \mathcal{R}\left(A^\top\right)$. They are related by $\langle \vec{v},\ \vec{u}\rangle = \vec{v}^\top \vec{u} = 0$. Each column of $A^\top$ is in $\mathcal{R}\left(A^\top\right)$. So we have $\vec{v}^\top \vec{u} = 0$ if $\vec{u}$ is a column of $A^\top$. This implies $\vec{v}^\top A^\top = \vec{0}^\top$. Taking transposes gives $(A\vec{v})^\top = \vec{0}^\top$ which implies that $A\vec{v} = \vec{0}$ and so $\vec{v} \in \mathcal{N}(A)$. Since this holds for any $\vec{v} \in \mathcal{R}\left(A^\top\right)^\perp$ we have $\mathcal{R}\left(A^\top\right)^\perp \subseteq \mathcal{N}(A)$.

(c) Let $\dim(\mathcal{N}(A)) = \ell$ and let $\vec{e}_1, \ldots, \vec{e}_\ell$ be an orthonormal basis for $\mathcal{N}(A)$. Consider an extension of the basis to an orthonormal basis, $\vec{e}_1, \ldots, \vec{e}_n$ for $\mathbb{R}^n$. We will prove that $\vec{e}_{\ell+1}, \ldots, \vec{e}_n$ form a basis for $\mathcal{R}\left(A^\top\right)$ and as a consequence, the dimension of $\mathcal{R}\left(A^\top\right)$ is $n - \ell$.

i. Show that $\mathcal{R}\left(A^\top\right)$ lies in the span of $\vec{e}_{\ell+1}, \ldots, \vec{e}_n$.

HINT: Express any vector $\vec{u} \in \mathcal{R}\left(A^\top\right)$ as $\vec{u} = \sum_{i=1}^n \alpha_i \vec{e}_i$. What are the values of α_i?

Solution: We can obtain the coefficient attached to the basis vector $\vec{e}_i$ by finding the scalar projection (dot product) of $\vec{u}$ onto $\vec{e}_i$. Therefore, by projecting a vector onto an orthonormal one, we see that $\alpha_i = 0$. Now, take any $\vec{u} \in \mathcal{R}\left(A^\top\right)$. We can express it as a linear combination of A's basis vectors:

$$\vec{u} = \sum_{i=1}^n \alpha_i \vec{e}_i. \tag{5}$$

But we also know from parts (a) and (b) that u is orthogonal to any vectors in the nullspace, which are spanned by/include the first ℓ basis vectors:

$$\vec{u}^\top \vec{e}_i = 0 \tag{6}$$

for all $i \in \{1, 2, \ldots, \ell\}$. Therefore, $\vec{u}^\top \vec{e}_i = \alpha_i = 0$ for all $i \in \{1, 2, \ldots, \ell\}$ and subsequently

$$\vec{u} = \sum_{i=\ell+1}^{n} \alpha_i \vec{e}_i. \tag{7}$$

Therefore, any vector $\vec{u} \in \mathcal{R}(A^\top)$ can be spanned by $\vec{e}_{\ell+1}, \ldots, \vec{e}_n$, making $\mathcal{R}(A^\top)$ a subset of the span of $\vec{e}_{\ell+1}, \ldots, \vec{e}_n$.

ii. From part (i) we know that $\mathcal{R}(A^\top) \subseteq \mathrm{span}(\vec{e}_{\ell+1}, \ldots, \vec{e}_n)$, but we want something stronger. Show that in fact $\mathcal{R}(A^\top) = \mathrm{span}(\vec{e}_{\ell+1}, \ldots, \vec{e}_n)$.

HINT: First, prove $\dim(\mathcal{R}(A^\top)) = \dim(\mathrm{span}(\vec{e}_{\ell+1}, \ldots, \vec{e}_n)) = n - \ell$ *by contradiction. Assume* $\dim(\mathcal{R}(A^\top)) = k < n - \ell$*, show that a vector* $\vec{u} \in \mathrm{span}(\vec{e}_{\ell+1}, \ldots, \vec{e}_n)$ *and* $\vec{u} \notin \mathcal{R}(A^\top)$ *cannot exist.*

Specifically, let $\vec{f}_1, \vec{f}_2, \ldots, \vec{f}_k$ *be an orthonormal basis for* $\mathcal{R}(A^\top)$*, we can find non-zero* $\vec{u}_\perp = \vec{u} - \sum_{i=1}^{k}(\vec{f}_i^\top \vec{u})\vec{f}_i$ *that is orthogonal to* $\mathcal{R}(A^\top)$*. Does* $\vec{u}_\perp$ *lie in* $\mathcal{N}(A)$*? Does* $\vec{u}_\perp$ *also lie in* $\mathrm{span}(\vec{e}_{\ell+1}, \ldots, \vec{e}_n)$*? Does this lead to a contradiction? Think of* $n - \ell = 3$ *and* $k = 2$ *for visualization.*

HINT: Second, you can use without proof the fact that for two subspaces, $\mathcal{S}_1$ *and* $\mathcal{S}_2$*, if* $\mathcal{S}_1 \subseteq \mathcal{S}_2$ *and* $\dim(\mathcal{S}_1) = \dim(\mathcal{S}_2)$ *then* $\mathcal{S}_1 = \mathcal{S}_2$*.*

Solution: Assume the contrary and let $\vec{f}_1, \ldots, \vec{f}_k$ be an orthonormal basis for $\mathcal{R}(A^\top)$. Then, there exists $\vec{u}$ in the span of $\vec{e}_{\ell+1}, \ldots, \vec{e}_n$ such that $\vec{u} \notin \mathcal{R}(A^\top)$. From this, we get $\vec{u}_\perp = \vec{u} - \sum_{i=1}^{k}(\vec{f}_i^\top \vec{u})\vec{f}_i$ is non-zero and orthogonal to $\mathcal{R}(A^\top)$. Visually, you can think of the range of $A^\top$ as the x, y plane, and $\vec{u}$ extends into 3D space outside this plane. From there, we can define a $\vec{u}_\perp$ that is perpendicular to the x, y plane simply by using Gram-Schmidt on the basis vectors, for example. Thus, $\vec{u}_\perp \in \mathcal{R}(A^\top)^\perp = \mathcal{N}(A)$. However, we also have $\vec{u}_\perp \in \mathrm{span}(\vec{e}_{\ell+1}, \ldots, \vec{e}_n)$ which is a contradiction as $\mathcal{N}(A)$ and $\mathrm{span}(\vec{e}_{\ell+1}, \ldots, \vec{e}_n)$ are orthogonal to each other. Therefore, the dimension of $\mathcal{R}(A^\top)$ is at least $n - \ell$ and is exactly $n - \ell$ as it is contained in an $n - \ell$ dimensional space. Since, we have $\mathcal{R}(A^\top) \subseteq \mathrm{span}(\vec{e}_{\ell+1} \ldots \vec{e}_n)$, and $\dim(\mathcal{R}(A^\top)) = \dim(\mathrm{span}(\vec{e}_{\ell+1} \ldots \vec{e}_n)) = n - \ell$, we can conclude that $\mathcal{R}(A^\top) = \mathrm{span}(\vec{e}_{\ell+1}, \ldots, \vec{e}_n)$.

(d) Using part (c) argue why $\mathcal{N}(A) \oplus \mathcal{R}(A^\top) = \mathbb{R}^n$ and why the rank nullity theorem holds.

Solution: We have found a basis $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$ such that the first ℓ vectors $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_\ell$ form a basis for $\mathcal{N}(A)$ and the last $n - \ell$ vectors form a basis for $\mathcal{R}(A^\top)$. Thus the claim $\mathcal{N}(A) \oplus \mathcal{R}(A^\top) = \mathbb{R}^n$ follows. Also,

$$\begin{aligned}
\dim(\mathcal{N}(A)) + \mathrm{rank}(A) &= \dim(\mathcal{N}(A)) + \mathrm{rank}(A^\top) && (8)\\
&= \dim(\mathcal{N}(A)) + \dim(\mathcal{R}(A^\top)) && (9)\\
&= \ell + n - \ell = n. && (10)
\end{aligned}$$

2. Symmetric Matrices

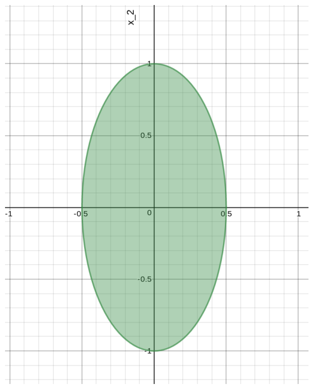

Figure 1: Region corresponding to $4x_1^2 + x_2^2 \leq 1$.

(a) Show that any symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive semidefinite if and only if there exists a symmetric matrix $C \in \mathbb{R}^{n \times n}$ such that $A = C^\top C$.

Solution: Forward direction:

$$\vec{x}^\top C^\top C \vec{x} = \|C\vec{x}\|_2^2 \geq 0 \tag{11}$$

and hence $C^\top C$ is PSD. Reverse direction: If A is PSD, then $A = UDU^\top$ by the spectral theorem with all the diagonal elements of D non-zero. Let $D^{1/2}$ be the square root of D (in this case just the square root of all the diagonal elements) and set $C = UD^{1/2}U^\top$. Observe that C is symmetric. Note that there are other non-symmetric C which also satisfy $C^\top C = A$. For example, $C = D^{1/2}U^\top$.

(b) Draw the region $\left\{ \vec{x} \in \mathbb{R}^2 \middle| \vec{x}^\top \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \vec{x} \leq 1 \right\}$.

Solution: Simplifying we get

$$4x_1^2 + x_2^2 \leq 1. \tag{12}$$

The corresponding region is an ellipse with minor axis 1 and major-axis 2 as shown in Fig. 1:

(c) Draw the region $\left\{ \vec{x} \in \mathbb{R}^2 \middle| \vec{x}^\top \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \vec{x} \leq 1 \right\}$.

Solution: Call the given matrix as A. The eigenvalues of A can be found by solving,

$$\det(A - \lambda I) = 0 \tag{13}$$

$$\implies (1-\lambda)^2 - 1 = 0 \tag{14}$$

$$\implies \lambda = 0, \text{ or } \lambda = 2. \tag{15}$$

The eigenvector associated with eigenvalue 0, say $\vec{v}_1$, is $\left[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right]^\top$. The eigenvector associated with eigenvalue 2, say $\vec{v}_2$, is $\left[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right]$. Observe that $\vec{v}_1, \vec{v}_2$ are orthonormal vectors, and any $\vec{x} \in \mathbb{R}^2$ can be written as $\vec{x} = a\vec{v}_1 + b\vec{v}_2$, for $a, b \in \mathbb{R}$. Then,

$$\vec{x}^\top A \vec{x} = (a\vec{v}_1 + b\vec{v}_2)^\top A (a\vec{v}_1 + b\vec{v}_2) \tag{16}$$

$$= a^2 \vec{v}_1^\top A \vec{v}_1 + b^2 \vec{v}_2^\top A \vec{v}_2 + ab\vec{v}_1^\top A \vec{v}_2 + ba\vec{v}_2^\top A \vec{v}_1 \tag{17}$$

$$= 2b^2. \tag{18}$$

Thus we have the condition $|b| \leq \frac{1}{\sqrt{2}}$. Note that a is a free parameter and can take any value in $\mathbb{R}$. Thus the region is a strip parallel to the line $y = x$ of width $\sqrt{2}$ as shown in Fig. 2. Note that the strip is not bounded and any $\vec{x} = a\vec{v}_1$ satisfies the condition for all $a \in \mathbb{R}$.

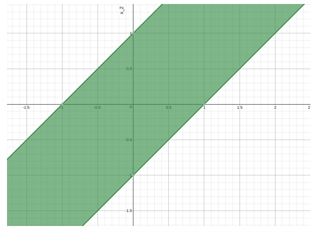

Figure 2: Region corresponding to $x_1^2 + x_2^2 - 2x_1x_2 \leq 1$.

(d) Why is the region in part (b) bounded, whereas the region in part (c) is unbounded?

Solution: This is because A in part (b) is positive definite while the A in part (c) is positive semi-definite and has a non-trivial nullspace. Suppose $\vec{v}$ is a unit vector in the nullspace of A. Then $\vec{x} = t\vec{v}$ will satisfy the equation for all $t \in \mathbb{R}$, however large.