ENGG 5781: Matrix Analysis and Computations **2015–16 Second Term**

Additional Materials for Lecture 1: Basic Concepts (More)

Instructor: Wing-Kin Ma

The aim of this note is to provide further coverage on some matrix and linear algebra concepts.

1 Subspace and Linear Independence

A nonempty subset $\mathcal{S}$ of $\mathbb{R}^m$ is called a *subspace* of $\mathbb{R}^m$ if, for any $\alpha, \beta \in \mathbb{R}$,

$$\mathbf{x}, \mathbf{y} \in \mathcal{S} \implies \alpha\mathbf{x} + \beta\mathbf{y} \in \mathcal{S}.$$

Clearly, if a set $\mathcal{S} \subseteq \mathbb{R}^m$ is a subspace, then we have the basic properties that the sum of any two vectors in $\mathcal{S}$ also lies in $\mathcal{S}$ (i.e., $\mathbf{x} + \mathbf{y} \in \mathcal{S}$ for any $\mathbf{x}, \mathbf{y} \in \mathcal{S}$), and that any scalar multiplication of any vector in $\mathcal{S}$ also lies in $\mathcal{S}$ (i.e., $\alpha\mathbf{x} \in \mathcal{S}$ for any $\alpha \in \mathbb{R}$, $\mathbf{x} \in \mathcal{S}$). The aforementioned properties can be extended to finite linear combinations of vectors in $\mathcal{S}$: It can be verified that if $\mathcal{S} \subseteq \mathbb{R}^m$ is a subspace and $\mathbf{a}_1, \ldots \mathbf{a}_n \in \mathcal{S}$, then any linear combination of $\mathbf{a}_1, \ldots \mathbf{a}_n$, i.e., $\sum_{i=1}^n \alpha_i \mathbf{a}_i$, where $\boldsymbol{\alpha} \in \mathbb{R}^n$, also lies in $\mathcal{S}$.

Given a collection of vectors $\mathbf{a}_1, \ldots \mathbf{a}_n \in \mathbb{R}^m$, the *span* of $\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$ is defined as

$$\text{span}\{\mathbf{a}_1, \ldots \mathbf{a}_n\} = \left\{ \mathbf{y} = \sum_{i=1}^n \alpha_i \mathbf{a}_i \;\middle|\; \boldsymbol{\alpha} \in \mathbb{R}^n \right\}.$$

In words, $\text{span}\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$ is the set of all possible linear combinations of the vectors $\mathbf{a}_1, \ldots \mathbf{a}_n$. It is easy to verify that for any $\mathbf{a}_1, \ldots \mathbf{a}_n$, $\text{span}\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$ is a subspace. In the literature, span is commonly used to represent or characterize a subspace. For example, we can represent $\mathbb{R}^m$ by

$$\mathbb{R}^m = \text{span}\{\mathbf{e}_1, \ldots, \mathbf{e}_m\}, \tag{1}$$

where $\mathbf{e}_i = [\ 0, \ldots, 0, 1, 0, \ldots 0\ ]^T \in \mathbb{R}^m$ with the nonzero element being at the ith entry; note that vectors taking such a form are called *unit vectors*. One may wonder whether the converse also holds; i.e., can a subspace $\mathcal{S}$ be always represented by $\mathcal{S} = \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ for some $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$? In fact, this is true: *For every subspace $\mathcal{S} \subseteq \mathbb{R}^m$, there exists a positive integer n and a collection of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ such that $\mathcal{S} = \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$.* The proof is left as a self-practice or assignment problem.

A set of vectors $\{\mathbf{a}_1, \ldots \mathbf{a}_n\} \subset \mathbb{R}^m$ is said to be *linearly independent* if

$$\sum_{i=1}^n \alpha_i \mathbf{a}_i = \mathbf{0},\ \boldsymbol{\alpha} \in \mathbb{R}^n \implies \boldsymbol{\alpha} = \mathbf{0}.$$

Otherwise, it is called *linearly dependent*. A subset $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$, where $i_j \in \{1, 2, \ldots, n\}$ for all j, and $1 \leq k \leq n$, is called a *maximal linearly independent* subset of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ if it is linearly independent and is not contained by any other linearly independent subset of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. From the above definitions, we see that

1. if $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is linearly independent, then any $\mathbf{a}_j$ cannot be a linear combination of the set of the other vectors $\{\mathbf{a}_i\}_{i \in \{1,\ldots n\}, i \neq j}$;

2. if $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is linearly dependent, then there exists a vector $\mathbf{a}_j$ such that it is a linear combination of the other vectors;

3. $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is a maximal linearly independent subset of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ if and only if $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}, \mathbf{a}_j\}$ is linearly dependent for any $j \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$.

Thus, roughly speaking, we may see a linearly independent $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ as a non-redundant or sufficiently different set of vectors, and a maximal linearly independent $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ as an irreducible (and non-redundant) set of vectors for representing the whole vector set $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. It can be easily shown that for any maximal linearly independent subset $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$, we have

$$\text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \text{span}\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}.$$

Given a subspace $\mathcal{S} \subseteq \mathbb{R}^m$ with $\mathcal{S} \neq \{\mathbf{0}\}$, a set of vectors $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\} \subset \mathbb{R}^m$ is called a *basis* for $\mathcal{S}$ if $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ is linearly independent and $\mathcal{S} = \text{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$. Taking $\mathcal{S} = \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ as an example, any maximal linearly independent subset of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is a basis for $\mathcal{S}$. From the definition of bases, the following facts can be derived:

1. a subspace may have more than one basis;

2. all bases for a subspace $\mathcal{S}$ have the same number of elements; i.e., if $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ and $\{\mathbf{c}_1, \ldots, \mathbf{c}_l\}$ are both bases for $\mathcal{S}$, then $k = l$.

Given a subspace $\mathcal{S} \subseteq \mathbb{R}^m$ with $\mathcal{S} \neq \{\mathbf{0}\}$, the *dimension* of $\mathcal{S}$ is defined as the number of elements of a basis for $\mathcal{S}$. Also, by convention, the dimension of the subspace $\{\mathbf{0}\}$ is defined as zero. The notation $\dim \mathcal{S}$ is used to denote the dimension of $\mathcal{S}$. Some examples are as follows. We have $\dim \mathbb{R}^m = m$, and this can be seen from (1). If $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$ is a maximal linearly independent subset of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$, then $\dim \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = k$.

2 Projections onto Subspaces

Given a nonempty closed set $\mathcal{S} \subseteq \mathbb{R}^m$ and a vector $\mathbf{y} \in \mathbb{R}^m$, a *projection* of $\mathbf{y}$ on $\mathcal{S}$ is a vector in $\mathcal{S}$ that is closest to $\mathbf{y}$ in terms of the Euclidean distance. To better describe it, the aforementioned projection can be formulated as a problem

$$\min_{\mathbf{z} \in \mathcal{S}} \ \|\mathbf{z} - \mathbf{y}\|_2.$$

If there exists a vector $\mathbf{y}_s \in \mathcal{S}$ that minimizes $\|\mathbf{z} - \mathbf{y}\|_2$ over all $\mathbf{z} \in \mathcal{S}$, then $\mathbf{y}_s$ is called a projection of $\mathbf{y}$ on $\mathcal{S}$. Note that in the above definition, $\mathcal{S}$ can be any closed set.

We are interested in the case where $\mathcal{S}$ is a subspace. The concepts of projections onto subspaces play a crucial role in linear algebra, matrix analysis, and their applications. In particular, the following theorem is important.

Theorem 1.A *Let $\mathcal{S}$ be a subspace of $\mathbb{R}^m$.*

1. *For every* $\mathbf{y} \in \mathbb{R}^m$*, there exists a unique vector* $\mathbf{y}_s \in \mathcal{S}$ *that minimizes* $\|\mathbf{z}-\mathbf{y}\|_2$ *over all* $\mathbf{z} \in \mathcal{S}$*. The vector* $\mathbf{y}_s$ *is thus the projection of* $\mathbf{y}$ *on* $\mathcal{S}$*, and we denote it as*

$$\Pi_{\mathcal{S}}(\mathbf{y}) = \arg\min_{\mathbf{z}\in\mathcal{S}} \|\mathbf{z}-\mathbf{y}\|_2.$$

2. *Given* $\mathbf{y} \in \mathbb{R}^m$*, we have* $\mathbf{y}_s = \Pi_{\mathcal{S}}(\mathbf{y})$ *if and only if* $\mathbf{y}_s \in \mathcal{S}$ *and*

$$\mathbf{z}^T(\mathbf{y}_s - \mathbf{y}) = 0, \quad \textit{for all } \mathbf{z} \in \mathcal{S}.$$

The above theorem is a special case of the projection theorem in convex analysis and optimization [1, Proposition B.11], which deals with projections onto closed convex sets. In the following we provide a proof that is enough for the subspace case.

Proof: First, consider the existence claim in Statement 1. it can be shown via the Weierstrass theorem that there always exists a vector in $\mathcal{S}$ at which the minimum of $\|\mathbf{z}-\mathbf{y}\|_2$ over all $\mathbf{z} \in \mathcal{S}$ is attained; the details are skipped here, and interested readers are referred to [1, proof of Proposition B.11].

Second, we show the sufficiency of Statement 2. Let $\mathbf{y}_s \in \mathcal{S}$ be a vector that minimizes $\|\mathbf{z}-\mathbf{y}\|_2$ over all $\mathbf{z} \in \mathcal{S}$. Since $\|\mathbf{z}-\mathbf{y}\|_2^2 \geq \|\mathbf{y}_s-\mathbf{y}\|_2^2$ for all $\mathbf{z} \in \mathcal{S}$, and

$$\begin{aligned}\|\mathbf{z}-\mathbf{y}\|_2^2 &= \|\mathbf{z}-\mathbf{y}_s+\mathbf{y}_s-\mathbf{y}\|_2^2 \\ &= \|\mathbf{z}-\mathbf{y}_s\|_2^2 + 2(\mathbf{z}-\mathbf{y}_s)^T(\mathbf{y}_s-\mathbf{y}) + \|\mathbf{y}_s-\mathbf{y}\|_2^2,\end{aligned}$$

we have

$$\|\mathbf{z}-\mathbf{y}_s\|_2^2 + 2(\mathbf{z}-\mathbf{y}_s)^T(\mathbf{y}_s-\mathbf{y}) \geq 0, \qquad \text{for all } \mathbf{z} \in \mathcal{S}.$$

The above equation is equivalent to

$$\|\mathbf{z}\|_2^2 + 2\mathbf{z}^T(\mathbf{y}_s-\mathbf{y}) \geq 0, \qquad \text{for all } \mathbf{z} \in \mathcal{S};$$

the reason is that $\mathbf{z} \in \mathcal{S}$ implies $\mathbf{z}-\mathbf{y}_s \in \mathcal{S}$, and the converse is also true. Now, suppose that there exists a point $\bar{\mathbf{z}} \in \mathcal{S}$ such that $\bar{\mathbf{z}}^T(\mathbf{y}_s-\mathbf{y}) \neq 0$. Then, by choosing $\mathbf{z} = \alpha\bar{\mathbf{z}}$, where $\alpha = -\bar{\mathbf{z}}^T(\mathbf{y}_s-\mathbf{y})/\|\bar{\mathbf{z}}\|_2^2$, one can verify that $\|\mathbf{z}\|_2^2+2\mathbf{z}^T(\mathbf{y}_s-\mathbf{y}) < 0$ and yet $\mathbf{z} \in \mathcal{S}$. Thus, by contradiction, we must have $\mathbf{z}^T(\mathbf{y}_s-\mathbf{y}) = 0$ for all $\mathbf{z} \in \mathcal{S}$.

Third, we show the necessity of Statement 2. Suppose that there exists a vector $\bar{\mathbf{y}}_s \in \mathcal{S}$ such that $\mathbf{z}^T(\bar{\mathbf{y}}_s-\mathbf{y}) = 0$ for all $\mathbf{z} \in \mathcal{S}$. The aforementioned condition can be rewritten as

$$(\mathbf{z}-\bar{\mathbf{y}}_s)^T(\bar{\mathbf{y}}_s-\mathbf{y}) = 0, \qquad \text{for all } \mathbf{z} \in \mathcal{S},$$

where we have used the equivalence $\mathbf{z}-\bar{\mathbf{y}}_s \in \mathcal{S} \Longleftrightarrow \mathbf{z} \in \mathcal{S}$. Now, for any $\mathbf{z} \in \mathcal{S}$, we have

$$\begin{aligned}\|\mathbf{z}-\mathbf{y}\|_2^2 &= \|\mathbf{z}-\bar{\mathbf{y}}_s\|_2^2 + 2(\mathbf{z}-\bar{\mathbf{y}}_s)^T(\bar{\mathbf{y}}_s-\mathbf{y}) + \|\bar{\mathbf{y}}_s-\mathbf{y}\|_2^2 \\ &= \|\mathbf{z}-\bar{\mathbf{y}}_s\|_2^2 + \|\bar{\mathbf{y}}_s-\mathbf{y}\|_2^2 \\ &\geq \|\bar{\mathbf{y}}_s-\mathbf{y}\|_2^2. \end{aligned} \tag{2}$$

The above inequality implies that $\bar{\mathbf{y}}_s$ minimizes $\|\mathbf{z}-\mathbf{y}\|_2$ over all $\mathbf{z} \in \mathcal{S}$. This, together with the previous sufficiency proof, completes the proof of Statement 2. In addition, we note that the equality in (2) holds if and only if $\mathbf{z}-\bar{\mathbf{y}}_s = \mathbf{0}$. This implies that $\bar{\mathbf{y}}_s$ is the only minimizer of $\|\mathbf{z}-\mathbf{y}\|_2$ over all $\mathbf{z} \in \mathcal{S}$. Thus, we also obtain the uniqueness claim in Statement 1. ■.

Theorem 1.A has many implications, e.g., in least squares and orthogonal projections which we will learn in later lectures. In the next section we will consider an application of Theorem 1.A on subspaces.

3 Orthogonal Complements

Given a nonempty subset $\mathcal{S}$ in $\mathbb{R}^m$, the *orthogonal complement* of $\mathcal{S}$ is defined as the set

$$\mathcal{S}_\perp = \left\{ \mathbf{y} \in \mathbb{R}^m \mid \mathbf{z}^T\mathbf{y} = 0 \text{ for all } \mathbf{z} \in \mathcal{S} \right\}.$$

It is easy to verify that $\mathcal{S}_\perp$ is always a subspace even if $\mathcal{S}$ is not. From the definition, we see that

1. any $\mathbf{z} \in \mathcal{S}$, $\mathbf{y} \in \mathcal{S}_\perp$ are orthogonal, i.e., $\mathcal{S}_\perp$ consists of all vectors that are orthogonal to all vectors of $\mathcal{S}$;

2. $\mathcal{S} \cap \mathcal{S}_\perp = \{\mathbf{0}\}$, i.e., except for point $\mathbf{0}$, the sets $\mathcal{S}$ and $\mathcal{S}_\perp$ are non-intersecting.

The following theorem is a consequence of the projection theorem in Theorem 1.A.

Theorem 1.B *Let $\mathcal{S}$ be a subspace of $\mathbb{R}^m$.*

1. *For every $\mathbf{y} \in \mathbb{R}^m$, there exists a unique 2-tuple $(\mathbf{y}_s, \mathbf{y}_c) \in \mathcal{S} \times \mathcal{S}_\perp$ such that*

$$\mathbf{y} = \mathbf{y}_s + \mathbf{y}_c.$$

   Also, such a $(\mathbf{y}_s, \mathbf{y}_c)$ is given by $\mathbf{y}_s = \Pi_{\mathcal{S}}(\mathbf{y}), \mathbf{y}_c = \mathbf{y} - \Pi_{\mathcal{S}}(\mathbf{y})$.

2. *The projection of $\mathbf{y}$ onto $\mathcal{S}_\perp$ is given by $\Pi_{\mathcal{S}_\perp}(\mathbf{y}) = \mathbf{y} - \Pi_{\mathcal{S}}(\mathbf{y})$.*

Proof: The problem in Statement 1 can be rephrased as follows: Given a subspace $\mathcal{S} \subseteq \mathbb{R}^m$ and a vector $\mathbf{y} \in \mathbb{R}^m$, find a vector $\mathbf{y}_s \in \mathcal{S}$ such that $\mathbf{y} - \mathbf{y}_s \in \mathcal{S}_\perp$. By Theorem 1.A, such a $\mathbf{y}_s$ exists and is uniquely given by $\mathbf{y}_s = \Pi_{\mathcal{S}}(\mathbf{y})$.

For Statement 2, recall from Theorem 1.A that we have $\mathbf{y}_c = \Pi_{\mathcal{S}_\perp}(\mathbf{y})$ if and only if $\mathbf{y}_c \in \mathcal{S}_\perp$ and

$$\bar{\mathbf{z}}^T(\mathbf{y}_c - \mathbf{y}) = 0, \quad \text{for all } \bar{\mathbf{z}} \in \mathcal{S}_\perp.$$

Moreover, we see that the above conditions hold if $\mathbf{y}_c = \mathbf{y} - \Pi_{\mathcal{S}}(\mathbf{y})$. Thus, the proof is complete ■

Given two subsets $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^m$, denote $\mathcal{X} + \mathcal{Y} = \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}\}$. Equipped with Theorem 1.B, we can easily prove the following results.

Property 1.A *The following properties hold for any subspace $\mathcal{S} \subseteq \mathbb{R}^m$:*

1. *$\mathcal{S} + \mathcal{S}_\perp = \mathbb{R}^m$;*

2. *$\dim \mathcal{S} + \dim \mathcal{S}_\perp = m$;*

3. *$(\mathcal{S}_\perp)_\perp = \mathcal{S}$.* [1]

Proof: Statement 1 is a direct consequence of Theorem 1.B.1. To prove Statement 2, let $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ and $\{\mathbf{c}_1, \ldots, \mathbf{c}_l\}$ be bases for $\mathcal{S}$ and $\mathcal{S}_\perp$, respectively. Since we can write $\mathcal{S} = \text{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$, $\mathcal{S}_\perp = \text{span}\{\mathbf{c}_1, \ldots, \mathbf{c}_l\}$, it follows from Statement 1 that

$$\mathbb{R}^m = \text{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_k, \mathbf{c}_1, \ldots \mathbf{c}_l\}.$$

[1] We also have the following result: Let $\mathcal{S}$ be any subset (and not necessarily a subspace) in $\mathbb{R}^m$. Then, we have $(\mathcal{S}_\perp)_\perp = \text{span}\, \mathcal{S}$, where span $\mathcal{S}$ is defined as the set of all finite linear combinations of points in $\mathcal{S}$.

Suppose that $\{\mathbf{b}_1, \ldots, \mathbf{b}_k, \mathbf{c}_1, \ldots \mathbf{c}_l\}$ is linearly independent. Then, we have $m = \dim \mathbb{R}^m = \dim \operatorname{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_k, \mathbf{c}_1, \ldots \mathbf{c}_l\} = k + l$. By also noting that $k = \dim \mathcal{S}$, $l = \dim \mathcal{S}_\perp$, we obtain Statement 2. It remains to prove the linear independence of $\{\mathbf{b}_1, \ldots, \mathbf{b}_k, \mathbf{c}_1, \ldots \mathbf{c}_l\}$. Consider the following condition

$$\sum_{i=1}^{k} \alpha_i \mathbf{b}_i = \sum_{i=1}^{l} \beta_i \mathbf{c}_i,$$

for some $\boldsymbol{\alpha} \in \mathbb{R}^k, \boldsymbol{\beta} \in \mathbb{R}^l$. By noting that LHS and RHS of the above equation lie in $\mathcal{S}$ and $\mathcal{S}_\perp$, respectively, and that $\mathcal{S} \cap \mathcal{S}_\perp = \{\mathbf{0}\}$, the above equation implies $\sum_{i=1}^{k} \alpha_i \mathbf{b}_i = \sum_{i=1}^{l} \beta_i \mathbf{c}_i = \mathbf{0}$. Also, since both $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ and $\{\mathbf{c}_1, \ldots, \mathbf{c}_l\}$ are linearly independent, $\sum_{i=1}^{k} \alpha_i \mathbf{b}_i = \sum_{i=1}^{l} \beta_i \mathbf{c}_i = \mathbf{0}$ holds only if $\boldsymbol{\alpha} = \mathbf{0}$, $\boldsymbol{\beta} = \mathbf{0}$. Thus, we have proven that $\{\mathbf{b}_1, \ldots, \mathbf{b}_k, \mathbf{c}_1, \ldots \mathbf{c}_l\}$ is linearly independent.

The proof Statement 3 is as follows. On one hand, one can easily see from the definition that $\mathbf{y} \in \mathcal{S}$ implies $\mathbf{y} \in (\mathcal{S}_\perp)_\perp$. On the other hand, suppose that $\mathbf{y} \in (\mathcal{S}_\perp)_\perp$. Since $\mathbf{y} = \Pi_{(\mathcal{S}_\perp)_\perp}(\mathbf{y})$ (which can be straightforwardly seen from the definition of projections onto subspaces), we use Theorem 1.B.2 to obtain

$$\mathbf{y} = \Pi_{(\mathcal{S}_\perp)_\perp}(\mathbf{y}) = \mathbf{y} - \Pi_{\mathcal{S}_\perp}(\mathbf{y}) = \mathbf{y} - (\mathbf{y} - \Pi_{\mathcal{S}}(\mathbf{y})) = \Pi_{\mathcal{S}}(\mathbf{y}) \in \mathcal{S}.$$

Thus, we have proven the equivalence $\mathbf{y} \in \mathcal{S} \Longleftrightarrow \mathbf{y} \in (\mathcal{S}_\perp)_\perp$. ■

Let us illustrate one application of Property 1.A. Recall that the nullspace of $\mathbf{A} \in \mathbb{R}^{m \times n}$ is denoted by $\mathcal{N}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{0}\}$. As a known result, the dimension of $\mathcal{N}(\mathbf{A})$ is

$$\dim \mathcal{N}(\mathbf{A}) = n - \operatorname{rank}(\mathbf{A}).$$

We can prove the above result by Property 1.A: First, it can be verified that $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{A}^T)_\perp$. Second, by Property 1.A.2, we have

$$n = \dim \mathcal{R}(\mathbf{A}^T) + \dim \mathcal{R}(\mathbf{A}^T)_\perp.$$

Third, by applying $\dim \mathcal{R}(\mathbf{A}^T) = \operatorname{rank}(\mathbf{A}^T) = \operatorname{rank}(\mathbf{A})$ to the above equation, we obtain the desired result.

References

[1] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, Belmont, Mass., U.S.A., 2nd edition, 1999.