ENGG 5781: Matrix Analysis and Computations **2016-17 Second Term**

Lecture 1: Basic Concepts

Instructor: Wing-Kin Ma

This note is not a supplementary material for the main slides. I will write notes such as this one when certain concepts cannot be put on slides. This time, the aim is to elaborate upon subspace concepts at a more fundamental level.

1 Subspace and Linear Independence

1.1 Subspace

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

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

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 $\text{span}\{\mathbf{a}_1, \ldots \mathbf{a}_n\}$ is a subspace for any given $\mathbf{a}_1, \ldots \mathbf{a}_n$. In the literature, span is commonly used to represent a subspace. For example, we can represent $\mathbb{R}^m$ by

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

In fact, any subspace can be written as a span:

Theorem 1.1 *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 $\mathcal{S} = \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$.*

In the literature you would notice that we take the result in Theorem 1.1 for granted—in a way that is almost like a common sense and without elaboration. There is an easy way to prove Theorem 1.1, but the proof requires some result in linear independence. We relegate the proof to the later part of this note.

1.2 Linear Independence

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

$$\sum_{i=1}^n \alpha_i \mathbf{a}_i \neq \mathbf{0}, \quad \text{for all } \boldsymbol{\alpha} \in \mathbb{R}^n \text{ with } \boldsymbol{\alpha} \neq \mathbf{0}.$$

Otherwise, it is called *linearly dependent*. A subset $\{\mathbf{a}_{i_1}, \ldots, \mathbf{a}_{i_k}\}$, where $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ with $i_j \neq i_l$ for any $j \neq l$, 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 irreducibly 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

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

We also have the following results which are very basic and we again take them almost for granted:

1. Let $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\} \subset \mathbb{R}^m$ be a linearly independent vector set. Suppose $\mathbf{y} \in \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$. Then the coefficient $\boldsymbol{\alpha}$ for the representation
$$\mathbf{y} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i$$
is unique; i.e., there does *not* exist a $\boldsymbol{\beta} \in \mathbb{R}^n$, $\boldsymbol{\beta} \neq \boldsymbol{\alpha}$, such that $\mathbf{y} = \sum_{i=1}^{n} \beta_i \mathbf{a}_i$.
2. If $\{\mathbf{a}_1, \ldots \mathbf{a}_n\} \subset \mathbb{R}^m$ is linearly independent, then $n \leq m$ must hold.

The first result is simple: if there exists a $\boldsymbol{\beta} \neq \boldsymbol{\alpha}$ such that $\mathbf{y} = \sum_{i=1}^{n} \beta_i \mathbf{a}_i$, then we have $\sum_{i=1}^{n} (\alpha_i - \beta_i)\mathbf{a}_i = \mathbf{y} - \mathbf{y} = \mathbf{0}$, which contradicts the linear independence of $\{\mathbf{a}_1, \ldots, \mathbf{a}_m\}$. For the second result, I show you in the proof in [2]. The proof is done by induction. Suppose $m = 1$. Then it is easy to see that $n \leq 1$ must hold. Next suppose $m \geq 2$. We need to show that $n \leq m$. Partition the $\mathbf{a}_i$'s as

$$\mathbf{a}_i = \begin{bmatrix} \mathbf{b}_i \\ c_i \end{bmatrix}, \quad i = 1, \ldots, n,$$

where $\mathbf{b}_i \in \mathbb{R}^{m-1}$, $c_i \in \mathbb{R}$. If $c_1 = \ldots = c_n = 0$, then the linear independence of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ implies that $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ is also linearly independent. By induction, we know that for any collection of n linearly independent vectors in $\mathbb{R}^{m-1}$, it must hold that $n \leq m - 1$. It follows that $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ must satisfy $n \leq m - 1$ too; hence, we get $n \leq m$. If $c_i \neq 0$ for some i, we need more work. Assume w.l.o.g. that $c_n \neq 0$ (we can always reshuffle the ordering of $\mathbf{a}_1, \ldots, \mathbf{a}_n$). For any $\alpha_1, \ldots, \alpha_{n-1} \in \mathbb{R}$, choose α_n as follows

$$\alpha_n = -\frac{1}{c_n} \left(\sum_{i=1}^{n-1} \alpha_i c_i \right).$$

Then we see that

$$\sum_{i=1}^{n} \alpha_i \mathbf{a}_i = \begin{bmatrix} \sum_{i=1}^{n-1} \alpha_i \mathbf{b}_i - \sum_{i=1}^{n-1} \frac{c_i}{c_n} \alpha_i \mathbf{b}_n \\ 0 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n-1} \alpha_i \left(\mathbf{b}_i - \frac{c_i}{c_n} \mathbf{b}_n \right) \\ 0 \end{bmatrix}$$

Let $\tilde{\mathbf{b}}_i = \mathbf{b}_i - (c_i/c_n)\mathbf{b}_n$ for ease of explanation. From the above equation, we see that the linear independence of $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ implies that $\{\tilde{\mathbf{b}}_1, \ldots, \tilde{\mathbf{b}}_{n-1}\}$ is linearly independent. By induction, we know that $\{\tilde{\mathbf{b}}_1, \ldots, \tilde{\mathbf{b}}_{n-1}\}$ must satisfy $n - 1 \leq m - 1$; hence, we get $n \leq m$. The proof is complete.

1.3 Orthogonality and Orthonormality

A collection of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^m$ is said to be *orthogonal* if $\mathbf{a}_i^T \mathbf{a}_j = 0$ for all i, j with $i \neq j$, and *orthonormal* if $\mathbf{a}_i^T \mathbf{a}_j = 0$ for all i, j with $i \neq j$ and $\|\mathbf{a}_i\|_2 = 1$ for all i. Some basic results arising from such a definition are as follows.

1. An orthogonal or orthonormal collection of vectors is also linearly independent.
2. Let $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{R}^m$ be an orthonormal set of vectors. Suppose $\mathbf{y} \in \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. Then the coefficient $\boldsymbol{\alpha}$ for the representation

$$\mathbf{y} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i$$

   is uniquely given by $\alpha_i = \mathbf{a}_i^T \mathbf{y}$, $i = 1, \ldots, n$.
3. Let $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subset \mathbb{R}^m$ be a linearly independent set of vectors. There exists an orthonormal set of vectors $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$ such that

$$\text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \text{span}\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}.$$

The first and second results above are easy to verify. The proof of the third result is constructive and is a consequence of the Gram-Schmidt procedure; see the main slides for the proof.

1.4 Proof of Theorem 1.1

We now prove Theorem 1.1. Suppose that the subspace $\mathcal{S}$ does not equal $\{\mathbf{0}\}$; the case of $\mathcal{S} = \{\mathbf{0}\}$ is trivial to prove. Let n be a positive integer and $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathcal{S}$. We have

$$\text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} \subseteq \mathcal{S},$$

and the reason is that any linear combination of vectors in $\mathcal{S}$ also lies in $\mathcal{S}$. Now we wish to show that there exists a linearly independent collection of $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathcal{S}$ such that $\mathcal{S} \subseteq \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. We adopt a constructive proof.

First, pick any nonzero vector $\mathbf{a}_1 \in \mathcal{S}$. If any $\mathbf{y} \in \mathcal{S}$ can be written as $\mathbf{y} = \alpha_1 \mathbf{a}_1$ for some α_1, then we finish. If not, consider the following recursive process. Suppose that we have previously picked a linearly independent collection of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_{k-1} \in \mathcal{S}$, but there exists $\mathbf{y} \in \mathcal{S}$ such that $\mathbf{y} \neq \sum_{i=1}^{k-1} \alpha_i \mathbf{a}_i$ for any α_i's. Note that we have the case of $k = 1$ already. Then we pick $\mathbf{a}_k$ as any $\mathbf{y} \in \mathcal{S}$ that satisfies $\mathbf{y} \neq \sum_{i=1}^{k-1} \alpha_i \mathbf{a}_i$ for any α_i's. Clearly, $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are linearly independent. If any $\mathbf{y} \in \mathcal{S}$ can be written as $\mathbf{y} = \sum_{i=1}^{k} \alpha_i \mathbf{a}_i$ for some α_i's, then it means $\mathcal{S} \subseteq \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ and we finish. Otherwise, we increase k by one and repeat the above steps.

The question is whether the recursive process above will stop; if yes, our proof is complete. Suppose that the process has reached an iteration number of $k > m$. On one hand, the process says that $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is linearly independent. On the other hand, if $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is linearly independent, then $k \leq m$ must hold (a take-it-for-granted fact in linear independence). Hence, by contradiction, the process must stop; $k = m$ is the largest possible iteration number. The proof is complete.

2 Basis and Dimension

Let 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 shown:

1. A subspace may have more than one basis.
2. A subspace always has an orthonormal basis (if it does not equal $\{\mathbf{0}\}$).
3. 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 $\dim \mathbb{R}^m = \dim \text{span}\{\mathbf{e}_1, \ldots, \mathbf{e}_m\} = m$. 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 $\dim \text{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = k$. We have the following properties for subspace dimension.

1. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^m$ be subspaces. If $\mathcal{S}_1 \subseteq \mathcal{S}_2$, then $\dim \mathcal{S}_1 \leq \dim \mathcal{S}_2$.
2. Let $\mathcal{S}$ be a subspace of $\mathbb{R}^m$. We have $\dim \mathcal{S} = m$ if and only if $\mathcal{S} = \mathbb{R}^m$.
3. Let $\mathcal{S}_1, \mathcal{S}_2 \subseteq \mathbb{R}^m$ be subspaces. We have $\dim(\mathcal{S}_1 + \mathcal{S}_2) \leq \dim \mathcal{S}_1 + \dim \mathcal{S}_2$; note that the notation $\mathcal{X} + \mathcal{Y} = \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in \mathcal{X}, \mathbf{y} \in \mathcal{Y}\}$ denotes the sum of two subsets $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^m$.

3 Projections onto Subspaces

Let $\mathcal{S} \subseteq \mathbb{R}^m$ be a nonempty closed set, and let $\mathbf{y} \in \mathbb{R}^m$ be any given vector. A *projection* of $\mathbf{y}$ onto $\mathcal{S}$ is any solution to the following problem

$$\min_{\mathbf{z} \in \mathcal{S}} \ \|\mathbf{z} - \mathbf{y}\|_2^2.$$

In words, a projection of $\mathbf{y}$ onto $\mathcal{S}$ is a point in $\mathcal{S}$ that is closest to $\mathbf{y}$ in the Euclidean sense. In general, there may be more than one such closest point. If every $\mathbf{y} \in \mathbb{R}^m$ has only one projection of $\mathbf{y}$ onto $\mathcal{S}$, we will use the following notation

$$\Pi_{\mathcal{S}}(\mathbf{y}) = \arg\min_{\mathbf{z} \in \mathcal{S}} \ \|\mathbf{z} - \mathbf{y}\|_2^2.$$

to denote the projection of $\mathbf{y}$ onto $\mathcal{S}$.

We are interested in projections onto subspaces. Such concepts play a crucial role in linear algebra and matrix analysis. Consider the following theorem:

Theorem 1.2 *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 $\mathbf{z} \in \mathcal{S}$. Thus, we can write $\Pi_{\mathcal{S}}(\mathbf{y}) = \arg\min_{\mathbf{z} \in \mathcal{S}} \ \|\mathbf{z} - \mathbf{y}\|_2^2$.*

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

$$\mathbf{y}_s \in \mathcal{S}, \qquad \mathbf{z}^T(\mathbf{y}_s - \mathbf{y}) = 0, \quad \textit{for all } \mathbf{z} \in \mathcal{S}. \tag{1}$$

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, we should mention that there always exists a vector in $\mathcal{S}$ at which the minimum of $\|\mathbf{z} - \mathbf{y}\|_2^2$ over all $\mathbf{z} \in \mathcal{S}$ is attained; in this claim we only need $\mathcal{S}$ to be closed. This result is shown by applying the Weierstrass theorem, and readers are referred to [1, proof of Proposition B.11] for details.

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 $\mathbf{z} - \bar{\mathbf{y}}_s = \mathbf{z}$. 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 1.2 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 1.2 to subspaces.

4 Orthogonal Complements

Let $\mathcal{S}$ a nonempty subset 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 direct consequence of the projection theorem in Theorem 1.2.

Theorem 1.3 *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: Let us rephrase the problem in Statement 1 as follows: Given a vector $\mathbf{y} \in \mathbb{R}^m$, find a vector $\mathbf{y}_s$ such that $\mathbf{y}_s \in \mathcal{S}$ and $\mathbf{y} - \mathbf{y}_s \in \mathcal{S}^\perp$. This problem is exactly the same as (1) in Theorem 1.2. Thus, by Theorem 1.2 the solution is uniquely given by $\mathbf{y}_s = \Pi_{\mathcal{S}}(\mathbf{y})$. Also, we have $\mathbf{y}_c = \mathbf{y} - \Pi_{\mathcal{S}}(\mathbf{y}) \in \mathcal{S}^\perp$.

To show Statement 2, let us consider the following problem: Given a vector $\mathbf{y} \in \mathbb{R}^m$, find a vector $\bar{\mathbf{y}}_c$ such that $\bar{\mathbf{y}}_c \in \mathcal{S}^\perp$ and $\mathbf{y} - \bar{\mathbf{y}}_c \in (\mathcal{S}^\perp)^\perp$, or equivalently,

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

By Theorem 1.2, the solution is uniquely given by $\bar{\mathbf{y}}_c = \Pi_{\mathcal{S}^\perp}(\mathbf{y})$. On the other hand, the above conditions are seen to satisfy if we choose $\bar{\mathbf{y}}_c = \mathbf{y} - \Pi_{\mathcal{S}}(\mathbf{y}) \in \mathcal{S}^\perp$. It follows that $\Pi_{\mathcal{S}^\perp}(\mathbf{y}) = \mathbf{y} - \Pi_{\mathcal{S}}(\mathbf{y})$. ■

Armed with Theorem 1.3, we can easily prove the following results.

Property 1.1 *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]

[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 $\mathcal{S}$.

Proof: Statement 1 is merely a consequence of the first statement in Theorem 1.3. For Statement 2, let us assume $\mathcal{S}$ does not equal either $\{\mathbf{0}\}$ or $\mathbb{R}^m$; the cases of $\{\mathbf{0}\}$ and $\mathbb{R}^m$ are trivial. Let $\{\mathbf{u}_1, \ldots, \mathbf{v}_k\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_l\}$ be orthonormal bases of $\mathcal{S}$ and $\mathcal{S}^\perp$, respectively; here note that $k = \dim \mathcal{S}, l = \dim \mathcal{S}^\perp$. We can write

$$\mathcal{S} + \mathcal{S}^\perp = \text{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{v}_1, \ldots, \mathbf{v}_l\}.$$

Also, from the definition of orthogonal complements, it is immediate that $\mathbf{u}_i^T \mathbf{v}_j = 0$ for all i, j. Hence, $\{\mathbf{u}_1, \ldots, \mathbf{u}_k, \mathbf{v}_1, \ldots, \mathbf{v}_l\}$ is an orthonormal basis for $\mathcal{S} + \mathcal{S}^\perp$, and consequently we have $\dim(\mathcal{S} + \mathcal{S}^\perp) = k + l$. Moreover, since $\mathcal{S} + \mathcal{S}^\perp = \mathbb{R}^m$, we also have $\dim(\mathcal{S} + \mathcal{S}^\perp) = m$. The result $m = \dim \mathcal{S} + \dim \mathcal{S}^\perp$ therefore holds.

The proof of Statement 3 is as follows. One can verify from the orthogonal complements definition that $\mathbf{y} \in \mathcal{S}$ implies $\mathbf{y} \in (\mathcal{S}^\perp)^\perp$. On the other hand, any $\mathbf{y} \in (\mathcal{S}^\perp)^\perp$ satisfies

$$\begin{aligned} \mathbf{y} &= \Pi_{(\mathcal{S}^\perp)^\perp}(\mathbf{y}) \\ &= \mathbf{y} - \Pi_{\mathcal{S}^\perp}(\mathbf{y}) = \mathbf{y} - (\mathbf{y} - \Pi_{\mathcal{S}}(\mathbf{y})) \\ &= \Pi_{\mathcal{S}}(\mathbf{y}), \end{aligned}$$

where the first equation above is due to the equivalence $\mathbf{y} \in \mathcal{S} \Longleftrightarrow \mathbf{y} = \Pi_{\mathcal{S}}(\mathbf{y})$ (which is easy to verify), and the second equation above is due to the second statement of Theorem 1.3. The equality $\mathbf{y} = \Pi_{\mathcal{S}}(\mathbf{y})$ shown above implies that $\mathbf{y} \in \mathcal{S}$. ■

Let us give an example on the application of Property 1.1. It is known that

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

The above result can be shown by Property 1.1. First, we use $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{A}^T)^\perp$. Second, from the second result in Property 1.1 we have

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

Third, it is known that $\dim \mathcal{R}(\mathbf{A}^T) = \text{rank}(\mathbf{A}^T) = \text{rank}(\mathbf{A})$. Combining these results together gives $\dim \mathcal{N}(\mathbf{A}) = n - \text{rank}(\mathbf{A})$.

References

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

[2] S. Boyd and L. Vandenberghe. *Vectors, Matrices, and Least Squares (Working Title).* 2017. Available online at `http://stanford.edu/class/ee103/mma.pdf`.