Linear Algebra: A Concise Review

These notes contain a summary of what I see as the major results in linear algebra. They are self-contained — we will prove all of the statements made below from first principles — but are not meant as a replacement for a full course in linear algebra. They are merely meant to reinforce the ideas that are important and show how they tie together mathematically.

You can find a detailed exposition of the material below in any one of a number of introductory texts. All linear algebra textbooks seem to say the same things, but the one that says them the best (in my opinion) is Gilbert Strang's *Linear Algebra and its Applications*, now in its 4th edition. Strang's lectures are also available on YouTube; I highly recommend them if you feel you need a refresh of this material, or want to see a legendary instructor in action[1].

We will assume familiarity with basic definitions and terminology from linear algebra including linear vector space, linear combinations, orthogonal, dot/inner product, and matrix.

The discussion below is terse, and is presented as a series of definitions and propositions. We will box the major results so they can be quickly referenced.

We start with a definition (which is probably not needed) of the fundamental mathematical construct of our study.

Definition 1 *The space $\mathbb{R}^N$ is simply the collection of all vectors (i.e. lists of numbers arranged in a column) with N elements that are real-valued (and finite).*

Any $\boldsymbol{x} \in \mathbb{R}^N$ can be written as a linear combination of the standard unit vector $\boldsymbol{e}_j$:

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots x_N\boldsymbol{e}_N, \tag{1}$$

where

$$e_n[j] = \begin{cases} 1, & j = n, \\ 0, & j \neq n. \end{cases}$$

1 Span and linear independence

Definition 2 *Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ be vectors in $\mathbb{R}^N$. Their* **span** *is the subset of $\mathbb{R}^N$ formed by taking all linear combinations between them:*

$$\text{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\} = \left\{ \boldsymbol{x} \in \mathbb{R}^N \ : \ \boldsymbol{x} = \sum_{k=1}^{K} \alpha_k \boldsymbol{v}_k \ \ \textit{for some} \ \ \alpha_1, \ldots, \alpha_K \in \mathbb{R} \right\}.$$

Definition 3 *Let $\mathcal{V} = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_K\}$ be a set of vectors in $\mathbb{R}^N$. We say that the vectors in $\mathcal{V}$ are* **linearly independent** *if there is no way to write one of the vectors as a linear combination*

[1] See also https://news.mit.edu/2019/gil-strang-still-going-strong-online-and-print-0508.

of the others. Equivalently,

$$\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_N \boldsymbol{v}_K = \boldsymbol{0} \quad \textit{if and only if} \quad \alpha_1 = \alpha_2 = \cdots = \alpha_K = 0.$$

If $\alpha_1 \boldsymbol{v}_1 + \cdots + \alpha_N \boldsymbol{v}_K = \boldsymbol{0}$ *with one or more of the* α_k *non-zero, then the vectors in* $\mathcal{V}$ *are called* linearly dependent.

We use these definitions to prove two basic facts. The first is that every set of vectors contains a subset of linearly independent vectors with the same span; the second is that a set of orthonormal vectors can be computed that also have the same span.

Proposition 1 *Let* $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ *be an arbitrary set of vectors in* $\mathbb{R}^N$*. There is a subset of these vectors that are linearly independent and have the same span.*

Proof Form the set $\mathcal{V}'$ in the following way. Starting with $\mathcal{V}'$ empty, for each $k = 1, \ldots, K$ add $\boldsymbol{v}_k$ to $\mathcal{V}'$ if $\mathcal{V}' \cup \{\boldsymbol{v}_k\}$ is a linearly independent set of vectors. By construction, all of the vectors left out of $\mathcal{V}'$ can be written as linear combinations of the vectors that are in $\mathcal{V}'$, so the span in unchanged. Also by construction, the vectors in $\mathcal{V}'$ are linearly independent. ■

Proposition 2 *Let* $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ *be a set of linearly independent vectors in* $\mathbb{R}^N$*. There is a set of* K *normalized, mutually orthogonal vectors* $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K$ *that have the same span:*

$$\boldsymbol{u}_k^{\mathrm{T}} \boldsymbol{u}_j = \begin{cases} 1, & k = j, \\ 0, & k \neq j. \end{cases} \quad \textit{and} \quad \mathrm{Span}\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K\} = \mathrm{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}.$$

Proof (You will prove this on the homework.) ■

Of course, we don't really need the $\{\boldsymbol{v}_k\}$ to be linearly independent in Proposition 2, as the previous proposition shows that we can simply extract a linearly independent set and start from that. In this case, there will be a smaller number of $\boldsymbol{u}_k$ than $\boldsymbol{v}_k$.

2 Bases for $\mathbb{R}^N$

Definition 4 *We call the vectors* $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ *a* **basis** *for* $\mathbb{R}^N$ *if:*

1. $\mathrm{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\} = \mathbb{R}^N$*, and*
2. *the* $\{\boldsymbol{v}_k\}$ *are linearly independent.*

We will see below that to meet this condition, we must have $K = N$. Part 1 of the definition above means that every $\boldsymbol{x} \in \mathbb{R}^N$ can be written as a linear combination of the $\{\boldsymbol{v}_k\}$. Part 2 implies that this linear combination is unique; since this may or may not be obvious, we will prove it quickly right now.

Proposition 3 *Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ be a basis for $\mathbb{R}^N$. Then there is exactly one set of $\alpha_1, \ldots, \alpha_K$ such that*

$$\boldsymbol{x} = \sum_{k=1}^{K} \alpha_k \boldsymbol{v}_k.$$

Proof That there is at least one set of such $\{\alpha_k\}$ follows from the definition of linear span. Suppose, though that there are also $\beta_1, \ldots, \beta_K$ such that $\boldsymbol{x} = \sum_{k=1}^{K} \beta_k \boldsymbol{v}_k$. Then

$$\sum_{k=1}^{K} \alpha_k \boldsymbol{v}_k - \sum_{k=1}^{K} \beta_k \boldsymbol{v}_k = \boldsymbol{0} \quad \Rightarrow \quad \sum_{k=1}^{K} (\alpha_k - \beta_k) \boldsymbol{v}_k = \boldsymbol{0}.$$

Since the $\boldsymbol{v}_k$ are linearly independent, it must be true that $\beta_k = \alpha_k$ for all $k = 1, \ldots, K$. ■

The rest of this section is devoted to proving the following fundamental result:
A set of vectors is a basis for $\mathbb{R}^N$ if and only if it consists of N linearly independent vectors.
We break the result into three parts, first will prove that if we have a set of N linearly independent vectors, then it is a basis. We then show that if the set contains fewer than N vectors, then there are things in $\mathbb{R}^N$ that are not in their span, and if it contains more than N vectors, then they cannot be linearly independent.

The hardest part of all of this is the next proposition. The proof is long, so we defer it to the Technical Details section at the end. It is, however, very informative, constructive, and worth studying.

Proposition 4 *Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N$ be a set of linearly independent vectors in $\mathbb{R}^N$. Then the $\{\boldsymbol{v}_n\}$ are a basis for $\mathbb{R}^N$; for every $\boldsymbol{x} \in \mathbb{R}^N$ there exists $\alpha_1, \ldots, \alpha_N \in \mathbb{R}$ such that*

$$\boldsymbol{x} = \sum_{n=1}^{N} \alpha_n \boldsymbol{v}_n.$$

Proof (See Appendix A.) ■

Given Proposition 4, all of the remaining results in this section follow from short arguments.

Proposition 5 *Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ be a set of vectors in $\mathbb{R}^N$. If $K > N$, then the $\{\boldsymbol{v}_k\}$ are not linearly independent.*

Proof Consider the first N vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N$. If these vectors are not linearly independent, then the proposition follows. If they are linearly independent, then by Proposition 4 they form a basis for $\mathbb{R}^M$ and there exists $\alpha_1, \ldots, \alpha_N$ not all equal to zero such that

$$\boldsymbol{v}_{N+1} = \sum_{k=1}^{N} \alpha_k \boldsymbol{v}_k.$$

Thus $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{N+1}\}$ (and hence $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$) are not linearly independent. ■

Proposition 6 *Let $\boldsymbol{A}$ be an $M \times N$ matrix with $M < N$. Then there is at least one $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{x} \neq \boldsymbol{0}$, such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$.*

Proof Apply Proposition 5 on the columns of $\boldsymbol{A}$. ■

Proposition 7 *Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ for $K < N$ be a set of linearly independent vectors in $\mathbb{R}^N$. Then there exists a set of $N - K$ vectors $\boldsymbol{z}_{K+1}, \ldots, \boldsymbol{z}_N$ such that*

$$\boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{v}_k = 0, \quad \forall\ i = K+1, \ldots, N, \quad k = 1, \ldots, K,$$
$$\boldsymbol{z}_i^{\mathrm{T}} \boldsymbol{z}_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \quad \forall\ i, j = K+1, \ldots, N.$$

Proof Apply Proposition 6 with $\boldsymbol{v}_1^{\mathrm{T}}, \ldots, \boldsymbol{v}_K^{\mathrm{T}}$ as the rows of a $K \times N$ matrix $\boldsymbol{A}$. For the $\boldsymbol{x}$ returned, take $\boldsymbol{z}_{K+1} = \boldsymbol{x}/\|\boldsymbol{x}\|_2$. Repeat with $\boldsymbol{v}_1^{\mathrm{T}}, \ldots, \boldsymbol{v}_K^{\mathrm{T}}, \boldsymbol{z}_{K+1}^{\mathrm{T}}$ as the rows of a $K+1 \times N$ matrix, etc. ■

Proposition 8 *Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ be a set of vectors in $\mathbb{R}^N$ with $K < N$. Then the $\{\boldsymbol{v}_k\}$ do not span $\mathbb{R}^N$: there exists at least one vector $\boldsymbol{z} \in \mathbb{R}^N$ such that $\boldsymbol{z} \notin \mathrm{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$.*

Proof Whether or not the vectors are linearly independent, Propositions 1 and 7 tell us that there is at least one $\boldsymbol{z} \neq 0$ (and in fact $\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z} = 1$) such that $\boldsymbol{z}^{\mathrm{T}}\boldsymbol{v}_k = 0$ for all $k = 1, \ldots, K$. But supposing that $\boldsymbol{z}$ is in the span of the $\{\boldsymbol{v}_k\}$ leads to a contradiction. For if there exists $\alpha_1, \ldots, \alpha_K$ such that $\boldsymbol{z} = \sum_{k=1}^K \alpha_k \boldsymbol{v}_k$, then

$$\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z} = \boldsymbol{z}^{\mathrm{T}} \left(\sum_{k=1}^K \alpha_k \boldsymbol{v}_k \right) = \sum_{k=1}^K \alpha_k\, \boldsymbol{z}^{\mathrm{T}} \boldsymbol{v}_k = 0.$$

So it must be the case that $\boldsymbol{z} \notin \mathrm{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$. ■

We now summarize the main result in this section with the following:

Proposition 9 *Every basis for $\mathbb{R}^N$ consists of exactly N vectors.*

Proof Combine Propositions 4, 5 and 8. ■

By definition, those vectors have to be linearly independent. Of course, not all sets of N vectors are a basis — if the vectors are linearly dependent, they are not.

3 Subspaces of $\mathbb{R}^N$ and dimension

Definition 5 *A* **subspace** *$\mathcal{S}$ of $\mathbb{R}^N$ is a subset of $\mathbb{R}^N$ that is closed under addition and scalar multiplication. That is*

$$\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S} \quad \Rightarrow \quad \alpha\boldsymbol{x} + \beta\boldsymbol{y} \in \mathcal{S} \ \text{ for all } \ \alpha, \beta \in \mathbb{R}.$$

It should be clear that for any set of vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K \in \mathbb{R}^N$ that $\text{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$ is a subspace.

Definition 6 *Let $\mathcal{S}$ be a subspace of $\mathbb{R}^N$. We call the set of vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$ a* **basis** *for $\mathcal{S}$ if*

1. $\text{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\} = \mathcal{S}$, *and*
2. *the $\{\boldsymbol{v}_k\}$ are linearly independent.*

It is clear that if we start with K linearly independent vectors $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$ and generate $\mathcal{S} = \text{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$ then $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$ are a basis for $\mathcal{S}$.

But what if we are just handed a subspace $\mathcal{S}$ of $\mathbb{R}^N$? Does every subspace have a basis? The answer is "yes", and is codified in the next proposition.

Proposition 10 *Every subspace $\mathcal{S}$ of $\mathbb{R}^N$ has a basis.*

Proof [2] Since $\mathcal{S} \subset \mathbb{R}^N$, we have already seen that we cannot choose more than N linearly independent vectors from $\mathcal{S}$. Let K be the largest integer such that there are K linearly independent vectors in $\mathcal{S}$, and let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ be one such set of vectors. Let $\boldsymbol{x}$ be any non-zero vector in $\mathcal{S}$. By definition of K, we know that the set of $K+1$ vectors $\{\boldsymbol{x}, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$ is linearly dependent, so there exists $\alpha_1, \ldots, \alpha_K$ not all equal to zero such that

$$\alpha_1 \boldsymbol{v}_1 + \cdots \alpha_K \boldsymbol{v}_K + \boldsymbol{x} = \boldsymbol{0} \quad \Rightarrow \quad \boldsymbol{x} = -\sum_{k=1}^{K} \alpha_k \boldsymbol{v}_k.$$

Thus $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ span $\mathcal{S}$ (and are linearly independent by construction). ■

The proof above should make it clear that every basis for a subspace $\mathcal{S}$ has the same number of elements. We call this number the dimension of $\mathcal{S}$.

Definition 7 *The* **dimension** *of a subspace $\mathcal{S}$ is the maximal number of linearly independent vectors in $\mathcal{S}$.*

We close this section with two complementary statements. The first might be thought of as a way to test if a collection of vectors is a basis for a subspace.

Proposition 11 *Let $\mathcal{S}$ be a subspace of $\mathbb{R}^N$, and let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ be a set of linearly independent vectors in $\mathcal{S}$. If the only vector orthogonal to all of the $\{\boldsymbol{v}_k\}$ is $\boldsymbol{0}$, then the $\{\boldsymbol{v}_k\}$ form a basis for $\mathcal{S}$.*

Proof (Homework.) ■

Proposition 12 *Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K$ be vectors in $\mathbb{R}^N$, and let $\boldsymbol{x} \in \text{Span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$. If $\boldsymbol{x}$ is orthogonal to all of the $\boldsymbol{v}_n$, then $\boldsymbol{x} = \boldsymbol{0}$:*

$$\boldsymbol{v}_k^{\mathrm{T}} \boldsymbol{x} = 0, \ k = 1, \ldots, K \quad \Rightarrow \quad \boldsymbol{x} = \boldsymbol{0}.$$

[2] I lifted this very nice argument from the class notes of David Speyer from the University of Michigan.

Proof Write

$$\boldsymbol{x} = \sum_{k=1}^{K} \alpha_k \boldsymbol{v}_k.$$

Then

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = \sum_{k=1}^{K} \alpha_k \boldsymbol{v}_k^{\mathrm{T}}\boldsymbol{x} = 0,$$

and since $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = 0 \Leftrightarrow \boldsymbol{x} = \boldsymbol{0}$, the proposition holds. ■

4 Matrices and the four fundamental subspaces

Definition 8 *The* **row space** *of an $M \times N$ matrix $\boldsymbol{A}$ is the subspace of $\mathbb{R}^N$ spanned by the rows of $\boldsymbol{A}$. We will denote this space* $\mathrm{Row}(\boldsymbol{A})$.

Definition 9 *The* **column space** *of an $M \times N$ matrix $\boldsymbol{A}$ is the subspace of $\mathbb{R}^M$ spanned by the columns of $\boldsymbol{A}$. We will denote this space* $\mathrm{Col}(\boldsymbol{A})$.

Note that by definition, $\mathrm{Row}(\boldsymbol{A}) = \mathrm{Col}(\boldsymbol{A}^{\mathrm{T}})$.

Definition 10 *The* **null space** *of an $M \times N$ matrix $\boldsymbol{A}$ is the subspace of $\mathbb{R}^N$ that $\boldsymbol{A}$ maps to the zero vector in $\mathbb{R}^M$:*

$$\mathrm{Null}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^N \ : \ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}.$$

If it is not already obvious to you that that $\mathrm{Null}(\boldsymbol{A})$ is a subspace, then realize that if $\boldsymbol{A}\boldsymbol{x}_1 = \boldsymbol{0}$ and $\boldsymbol{A}\boldsymbol{x}_2 = \boldsymbol{0}$, then

$$\boldsymbol{A}(\alpha\boldsymbol{x}_1 + \beta\boldsymbol{x}_2) = \alpha\boldsymbol{A}\boldsymbol{x}_1 + \beta\boldsymbol{A}\boldsymbol{x}_2 = \boldsymbol{0}.$$

Definition 11 *The* **left null space** *of an $M \times N$ matrix is* $\mathrm{Null}(\boldsymbol{A}^{\mathrm{T}})$*, the subspace of $\mathbb{R}^M$ that $\boldsymbol{A}^{\mathrm{T}}$ maps to the zero vector in $\mathbb{R}^N$.*

Below, we prove the following fundamental relationships between these spaces.

- The dimensions of $\mathrm{Row}(\boldsymbol{A})$ and $\mathrm{Col}(\boldsymbol{A})$ are equal. This dimension is called the **rank** of the matrix $\boldsymbol{A}$.
- If an $M \times N$ matrix has rank R (and so $\mathrm{Row}(\boldsymbol{A})$ has dimension R), then $\mathrm{Null}(\boldsymbol{A})$ has dimension $N - R$. Moreover, $\mathrm{Row}(\boldsymbol{A})$ and $\mathrm{Null}(\boldsymbol{A})$ are **orthogonal complements** in $\mathbb{R}^N$: every vector in $\mathrm{Row}(\boldsymbol{A})$ is orthogonal to every vector in $\mathrm{Null}(\boldsymbol{A})$, and every vector in $\mathbb{R}^N$ can be written as a linear combination of a vector in $\mathrm{Row}(\boldsymbol{A})$ and a vector in $\mathrm{Null}(\boldsymbol{A})$.
- Similarly, $\mathrm{Col}(\boldsymbol{A})$ and $\mathrm{Null}(\boldsymbol{A}^{\mathrm{T}})$ are orthogonal complements in $\mathbb{R}^M$.

To prove the first statement above, we start with a critical intermediate result.

Proposition 13 *Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_K$ be linearly independent vectors in $\mathbb{R}^N$, and define*

$$\mathcal{S} = \text{Span}\left\{ \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{x} \\ \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{x} \\ \vdots \\ \boldsymbol{a}_K^{\mathrm{T}}\boldsymbol{x} \end{bmatrix}, \ \boldsymbol{x} \in \mathbb{R}^N \right\}.$$

Then $\mathcal{S} = \mathbb{R}^K$.

Proof It should be clear that $\mathcal{S}$ is a subspace, so it is enough to find K linearly independent vectors in $\mathcal{S}$. We will show that K such vectors can be generated by taking $\boldsymbol{x} = \boldsymbol{a}_k, \ k = 1, \ldots, K$ above. Set

$$\boldsymbol{v}_1 = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_1 \\ \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_K^{\mathrm{T}}\boldsymbol{a}_1 \end{bmatrix}, \quad \boldsymbol{v}_2 = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_2 \\ \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_K^{\mathrm{T}}\boldsymbol{a}_2 \end{bmatrix}, \quad \cdots \quad , \boldsymbol{v}_K = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_K \\ \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_K \\ \vdots \\ \boldsymbol{a}_K^{\mathrm{T}}\boldsymbol{a}_K \end{bmatrix},$$

and let $\alpha_1, \ldots, \alpha_K \in \mathbb{R}$ be such that

$$\sum_{k=1}^{K} \alpha_k \boldsymbol{v}_k = \boldsymbol{0}.$$

Then

$$\begin{aligned} &\sum_{k=1}^{K} \alpha_k \boldsymbol{a}_j^{\mathrm{T}} \boldsymbol{a}_k = 0, \quad \text{for } j = 1, \ldots, K \\ \Rightarrow \ &\boldsymbol{a}_j^{\mathrm{T}} \left(\sum_{k=1}^{K} \alpha_k \boldsymbol{a}_k \right) = 0, \quad \text{for } j = 1, \ldots, K \\ \Rightarrow \ &\sum_{k=1}^{K} \alpha_k \boldsymbol{a}_k = \boldsymbol{0}, \end{aligned}$$

where the last assertion is an application of Proposition 12. Since the $\boldsymbol{a}_k$ are linearly independent, this means that $\alpha_k = 0, \ k = 1, \ldots, K$. Hence the $\{\boldsymbol{v}_k\}$ are linearly independent as well. We complete the proof by applying Proposition 4. ■

If the rows of $\boldsymbol{A}$ are linearly independent, then Proposition 13 proves shows that the dimensions of $\text{Row}(\boldsymbol{A})$ and $\text{Col}(\boldsymbol{A})$ are equal. For the general case, a little more work needs to be done, which we again defer to the Technical Details section at the end.

Proposition 14 *Let $\boldsymbol{A}$ be an $M \times N$ matrix. Then the dimensions of* $\text{Row}(\boldsymbol{A})$ *and* $\text{Col}(\boldsymbol{A})$ *are the same. This dimension is called the* **rank** *of $\boldsymbol{A}$.*

Proof (See Appendix B.) ■

Note that it is always true that $\text{rank}(\boldsymbol{A}) \leq \min(M, N)$.

The following proposition is another cornerstone of linear algebra. It says that any matrix $\boldsymbol{A}$ partitions $\mathbb{R}^N$ into two subspaces, and these subspaces are completely orthogonal to one another.

Proposition 15 *Let $\boldsymbol{A}$ be an $M \times N$ matrix of rank R. The subspaces* $\mathrm{Row}(\boldsymbol{A})$ *and* $\mathrm{Null}(\boldsymbol{A})$ *are* **orthogonal complements***: $\mathrm{Row}(\boldsymbol{A})$ has dimension R, $\mathrm{Null}(\boldsymbol{A})$ has dimension $N - R$,*

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = 0, \quad \text{for all } \boldsymbol{u} \in \mathrm{Null}(\boldsymbol{A}),\ \boldsymbol{v} \in \mathrm{Row}(\boldsymbol{A}),$$

and every $\boldsymbol{x} \in \mathbb{R}^N$ can be written as

$$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2, \quad \boldsymbol{x}_1 \in \mathrm{Row}(\boldsymbol{A}), \quad \boldsymbol{x}_2 \in \mathrm{Null}(\boldsymbol{A}).$$

Proof We know that we can extract R linearly independent rows from $\boldsymbol{A}$; call these $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_R \in \mathbb{R}^N$. Applying Proposition 7 with these rows gives us $N - R$ linearly independent (and in fact orthonormal) vectors $\boldsymbol{z}_{R+1}, \ldots, \boldsymbol{z}_N$ that are orthogonal to each of these R rows and hence to all of the rows of $\boldsymbol{A}$ — it is clear that anything in the span of these $N - R$ vectors is in the null space of $\boldsymbol{A}$. Since $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r, \boldsymbol{z}_{R+1}, \ldots, \boldsymbol{z}_N\}$ i sa set of N linearly independent vectors in $\mathbb{R}^N$, they form a basis for $\mathbb{R}^N$, thus every $\boldsymbol{x} \in \mathbb{R}^N$ can be written

$$\boldsymbol{x} = \underbrace{\sum_{n=1}^{R} \alpha_n \boldsymbol{a}_n}_{\in \mathrm{Row}(\boldsymbol{A})} + \underbrace{\sum_{n=R+1}^{N} \alpha_n \boldsymbol{z}_n}_{\in \mathrm{Null}(\boldsymbol{A})}.$$

The two terms above are also orthogonal to one another, since all of the $\boldsymbol{z}_n$ are orthogonal to all of the $\boldsymbol{a}_n$.

It remains to show that vectors in the span of $\{\boldsymbol{z}_{R+1}, \ldots, \boldsymbol{z}_N\}$ are the only vectors in $\mathbb{R}^N$ in $\mathrm{Null}(\boldsymbol{A})$. Applying $\boldsymbol{A}$ to the expression above yields

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} &= \sum_{n=1}^{R} \alpha_n \boldsymbol{A}\boldsymbol{a}_n + \sum_{n=R+1}^{N} \alpha_n \boldsymbol{A}\boldsymbol{z}_n \\ &= \sum_{n=1}^{R} \alpha_n \boldsymbol{A}\boldsymbol{a}_n, \end{aligned}$$

since $\boldsymbol{A}\boldsymbol{z}_n = \boldsymbol{0}$. By Proposition 13, the vectors $\boldsymbol{A}\boldsymbol{a}_n$ in the first sum are linearly independent, and so $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ if and only if $\alpha_n = 0$ for $n = 1, \ldots, R$, i.e. if and only if $\boldsymbol{x} \in \mathrm{Span}(\{\boldsymbol{z}_{R+1}, \ldots, \boldsymbol{z}_N\})$. ■

The proof above covers a lot, but there is actually a very simple way to argue that the row space is orthogonal to the null space. If $\boldsymbol{u} \in \mathrm{Null}(\boldsymbol{A})$, then $\boldsymbol{A}\boldsymbol{u} = \boldsymbol{0}$, and if $\boldsymbol{v} \in \mathrm{Row}(\boldsymbol{A})$, there exists at least one vector $\boldsymbol{w} \in \mathbb{R}^M$ such that $\boldsymbol{v} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{w}$. Then

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = \boldsymbol{u}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{w} = (\boldsymbol{A}\boldsymbol{u})^{\mathrm{T}}\boldsymbol{w} = \boldsymbol{0}^{\mathrm{T}}\boldsymbol{w} = \boldsymbol{0}.$$

Any $M \times N$ matrix also partitions $\mathbb{R}^M$ in a similar manner, which is the content of our next proposition.

Proposition 16 *Let $\boldsymbol{A}$ be an $M \times N$ matrix of rank R. The subspaces* $\mathrm{Col}(\boldsymbol{A})$ *and* $\mathrm{Null}(\boldsymbol{A}^{\mathrm{T}})$ *are* **orthogonal complements***: $\mathrm{Col}(\boldsymbol{A})$ has dimension R, $\mathrm{Null}(\boldsymbol{A}^{\mathrm{T}})$ has dimension $N - R$,*

$$\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v} = 0, \quad \text{for all } \boldsymbol{u} \in \mathrm{Null}(\boldsymbol{A}^{\mathrm{T}}),\ \boldsymbol{v} \in \mathrm{Col}(\boldsymbol{A}),$$

and every $\boldsymbol{x} \in \mathbb{R}^M$ can be written as

$$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2, \quad \boldsymbol{x}_1 \in \mathrm{Col}(\boldsymbol{A}), \quad \boldsymbol{x}_2 \in \mathrm{Null}(\boldsymbol{A}^{\mathrm{T}}).$$

5 Systems of linear equations: existence and uniqueness

This section is devoted to solving systems of linear equations. Given an $M \times N$ matrix $\boldsymbol{A}$ and a vector $\boldsymbol{b} \in \mathbb{R}^M$, the statements below establish conditions for the existence of a $\boldsymbol{x} \in \mathbb{R}^N$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, and if such an $\boldsymbol{x}$ exists, whether it is unique (i.e. it is the only vector in $\mathbb{R}^N$ that maps to $\boldsymbol{b}$ through $\boldsymbol{A}$).

Proposition 17 *Let $\boldsymbol{A}$ be a square $N \times N$ matrix. If $\operatorname{rank}(A) = N$, then $\boldsymbol{A}$ is invertible: for every $\boldsymbol{b} \in \mathbb{R}^N$ there exists exactly one $\boldsymbol{x} \in \mathbb{R}^N$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. This $\boldsymbol{x}$ is found by applying a fixed matrix $\boldsymbol{A}^{-1}$ to $\boldsymbol{b}$, so $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$.*

Proof Since $\operatorname{rank}(\boldsymbol{A}) = N$, the columns of $\boldsymbol{A}$ are a basis for $\mathbb{R}^N$. Thus every $\boldsymbol{b} \in \mathbb{R}^N$ can be written in exactly one way as a linear combination of the columns of $\boldsymbol{A}$, so we collect the coefficients for this linear combination into the vector $\boldsymbol{x}$. The proof of Proposition 4 shows that these coefficients are linear combinations of the elements of $\boldsymbol{b}$, with weights that can be computed from only the entries of $\boldsymbol{A}$. These weights can be collected into a matrix which we call $\boldsymbol{A}^{-1}$. ■

Proposition 18 *The $N \times N$ matrix $\boldsymbol{A}^{-1}$ from Proposition 17 obeys $\boldsymbol{A}^{-1}\boldsymbol{A} = \mathbf{I}$ and $\boldsymbol{A}\boldsymbol{A}^{-1} = \mathbf{I}$, where $\mathbf{I}$ is the $N \times N$ identity matrix.*

Proof We first establish that if $\boldsymbol{M}\boldsymbol{x} = \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^N$, then $\boldsymbol{M} = \mathbf{I}$. For if $\boldsymbol{M} = \mathbf{I} + \boldsymbol{G}$ and $\boldsymbol{M}\boldsymbol{x} = \boldsymbol{x} + \boldsymbol{G}\boldsymbol{x} = \boldsymbol{x}$ for all $\boldsymbol{x}$, then $\boldsymbol{G}\boldsymbol{x} = \mathbf{0}$ for all $\boldsymbol{x}$, and so $\boldsymbol{G} = \mathbf{0}$. Now for any $\boldsymbol{x}, \boldsymbol{b}$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, we know from Proposition 17 that $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$. Thus $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x}$ for all $\boldsymbol{x} \in \mathbb{R}^N$, and $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{b}$ for all $\boldsymbol{b} \in \mathbb{R}^N$. ■

Proposition 19 *Let $\boldsymbol{A}$ be an $M \times N$ matrix. Then the system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has at least one solution if and only if $\boldsymbol{b} \in \operatorname{Col}(\boldsymbol{A})$.*

Proof Since every linear combination of the columns of $\boldsymbol{A}$ can be induced by applying it to a vector $\boldsymbol{x}$, the proposition pretty much follows from the definition of $\operatorname{Col}(\boldsymbol{A})$. ■

Proposition 20 *If $\operatorname{rank}(\boldsymbol{A}) < M$, then there are (many) $\boldsymbol{b} \in \mathbb{R}^M$ for which there is no solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.*

Proof By Proposition 16 any $\boldsymbol{b} \in \mathbb{R}^M$ can be written as

$$\boldsymbol{b} = \boldsymbol{b}_1 + \boldsymbol{b}_2, \quad \text{where} \quad \boldsymbol{b}_1 \in \operatorname{Col}(\boldsymbol{A}), \quad \boldsymbol{b}_2 \in \operatorname{Null}(\boldsymbol{A}^{\mathrm{T}}).$$

If $\boldsymbol{b}_2 \neq 0$, then $\boldsymbol{b} \notin \operatorname{Col}(\boldsymbol{A})$ and there is no solution. If $\operatorname{rank}(\boldsymbol{A}) = r < M$, then there will be many vectors whose $\boldsymbol{b}_2$ component is non-zero. ■

Note that when $\operatorname{rank}(\boldsymbol{A}) < M$, the vector $\boldsymbol{b}$ has to be very special for there to be a $\boldsymbol{x}$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. If, for example, the entries of $\boldsymbol{b}$ are drawn independently from a zero-mean Gaussian distribution, then there will be no solution with probability 1. Also note that the result holds whether the matrix is "fat" ($M < N$), "skinny" ($M > N$), or "square" ($M = N$) — it depends on the *rank* of the matrix (rather than just the number of rows) in relation to the number of columns.

Proposition 21 *If* $\text{rank}(A) = N \leq M$*, then if a solution to* $\boldsymbol{Ax} = \boldsymbol{b}$ *exists (meaning that* $\boldsymbol{b} \in \text{Col}(\boldsymbol{A})$*) it is unique.*

Proof Since $\text{rank}(A) = N$, the columns of $\boldsymbol{A}$ are a basis for $\text{Col}(\boldsymbol{A})$. Thus every vector in $\text{Col}(\boldsymbol{A})$ (every $\boldsymbol{b}$ for which there is at least one solution to $\boldsymbol{Ax} = \boldsymbol{b}$) space can be written as a linear combination of the columns of $\boldsymbol{A}$ in exactly one way. ■

Proposition 22 *If* $\text{rank}(A) = M < N$*, then for every* $\boldsymbol{b}$ *there are an infinite number of* $\boldsymbol{x}$ *such that* $\boldsymbol{Ax} = \boldsymbol{b}$*.*

Proof Since $\text{Col}(\boldsymbol{A}) = \mathbb{R}^M$, every $\boldsymbol{b}$ is in the column space of $\boldsymbol{A}$, so by Proposition 19 we know that there is at least one $\boldsymbol{x}$ such that $\boldsymbol{Ax} = \boldsymbol{b}$. Since $\text{rank}(A) < N$, we also know that $\text{Null}(\boldsymbol{A})$ has dimension at least one (meaning it contains an infinity of vectors). A $\boldsymbol{x} + \boldsymbol{x}_0$, $\boldsymbol{x}_0 \in \text{Null}(\boldsymbol{A})$ will also satisfy $\boldsymbol{A}(\boldsymbol{x} + \boldsymbol{x}_0) = \boldsymbol{b}$, there are an infinite number of solutions. ■

You can also combine two of the propositions to show that if $\text{rank}(\boldsymbol{A}) < M < N$, then $\boldsymbol{Ax} = \boldsymbol{b}$ has either no solutions (when $\boldsymbol{b} \notin \text{Col}(\boldsymbol{A})$) or an infinite number of solutions ($\boldsymbol{b} \in \text{Col}(\boldsymbol{A})$).

A Technical Details: Proof of Proposition 4

It is clear that every vector is $\mathbb{R}^N$ is in the span of the unit vectors $\{\boldsymbol{e}_n\}$, where

$$e_n[i] = \begin{cases} 1, & i = n, \\ 0, & i \neq n \end{cases}.$$

So we will prove the proposition by showing that each of the $\boldsymbol{e}_n$ are in the span of the $\{\boldsymbol{v}_n\}$.

We do this by induction on N. For $N = 1$, the proposition is self-evident. We can also treat $N = 2$ explicitly. It is easy to check that by taking

$$\alpha_1 = \frac{v_2[2]}{v_1[1]v_2[2] - v_1[2]v_2[1]}, \quad \alpha_2 = \frac{-v_1[2]}{v_1[1]v_2[2] - v_1[2]v_2[1]},$$

we have $\boldsymbol{e}_1 = \alpha_1\boldsymbol{v}_1 + \alpha_2\boldsymbol{v}_2$, and by taking

$$\beta_1 = \frac{-v_2[1]}{v_1[1]v_2[2] - v_1[2]v_2[1]}, \quad \beta_2 = \frac{v_1[1]}{v_1[1]v_2[2] - v_1[2]v_2[1]},$$

we have $\boldsymbol{e}_2 = \beta_1\boldsymbol{v}_1 + \beta_2\boldsymbol{v}_2$. The linear independence of the $\boldsymbol{v}_1, \boldsymbol{v}_2$ ensures that $v_1[1]v_2[2] \neq v_1[2]v_2[1]$, and so all four of these coefficients are well-defined.

To make the induction argument clear, let's see how the result for $N = 2$ implies the result for $N = 3$. We arrange the three linearly independent vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$ as columns in a 3×3 matrix

$$\boldsymbol{V} = \begin{bmatrix} | & | & | \\ \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 \\ | & | & | \end{bmatrix} = \begin{bmatrix} v_1[1] & v_2[1] & v_3[1] \\ v_1[2] & v_2[2] & v_3[2] \\ v_1[3] & v_2[3] & v_3[3] \end{bmatrix}.$$

We start by showing how $\boldsymbol{e}_1$ can be written as a linear combination. Consider the 2×2 matrix in the lower right hand corner of the above

$$\begin{bmatrix} \boldsymbol{v}_2' & \boldsymbol{v}_3' \end{bmatrix} = \begin{bmatrix} v_2[2] & v_3[2] \\ v_2[3] & v_3[3] \end{bmatrix}.$$

If $\boldsymbol{v}_2'$ and $\boldsymbol{v}_3'$ are not linearly independent, then there is an a_2, a_3 such that $a_2\boldsymbol{v}_2' + a_3\boldsymbol{v}_3' = \boldsymbol{0}$, and so

$$a_2\boldsymbol{v}_2 + a_3\boldsymbol{v}_3 = \begin{bmatrix} c \\ 0 \\ 0 \end{bmatrix},$$

where we know $c \neq 0$ since otherwise $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ would not be linearly independent (violating one of the premises of the theorem). So in this case $\boldsymbol{e}_1 = (a_2/c)\boldsymbol{v}_2 + (a_3/c)\boldsymbol{v}_3$.

If $\boldsymbol{v}_2'$ and $\boldsymbol{v}_3'$ are linearly independent, then we know from the $N = 2$ case that the vectors

$$\boldsymbol{u}_2 = \begin{bmatrix} c_2 \\ 1 \\ 0 \end{bmatrix}, \quad \boldsymbol{u}_3 = \begin{bmatrix} c_3 \\ 0 \\ 1 \end{bmatrix}$$

are in the span of $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$ for some c_2, c_3. Then

$$\boldsymbol{v}_1 - v_1[2]\boldsymbol{u}_2 - v_1[3]\boldsymbol{u}_3 = \begin{bmatrix} v_1[1] - v_1[2]c_2 - v_1[3]c_3 \\ 0 \\ 0 \end{bmatrix}.$$

First, note that the vector above is in the span of $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$, as $\boldsymbol{u}_2$ and $\boldsymbol{u}_3$ are in the span of $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$. We also know that $v_1[1] - v_1[2]c_2 - v_1[3]c_3 \neq 0$, as otherwise $\boldsymbol{v}_1, \boldsymbol{v}_2$, and $\boldsymbol{v}_3$ would not be linearly independent. Thus

$$\boldsymbol{e}_1 = (1/c)\boldsymbol{v}_1 - (v_1[2]/c)\boldsymbol{u}_2 - (v_1[3]/c)\boldsymbol{u}_3, \quad c = v_1[1] - v_1[2]c_2 - v_1[3]c_3.$$

The argument for $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ being in $\text{Span}(\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\})$ follows exactly the same flow. To show $\boldsymbol{e}_2$ is in the span, we examine the 2×2 submatrix created by eliminating the second row and column of $\boldsymbol{V}$

$$\begin{bmatrix} v_1[1] & v_3[1] \\ v_1[3] & v_3[3] \end{bmatrix},$$

and if its columns are not linearly independent, then $c\boldsymbol{e}_2$ is in the span for a $c \neq 0$, and hence $\boldsymbol{e}_2$ is in the span. If the columns are linear independent, then again we apply the $N = 2$ case to show that

$$\boldsymbol{w}_1 = \begin{bmatrix} 1 \\ c_1 \\ 0 \end{bmatrix}, \quad \boldsymbol{w}_3 = \begin{bmatrix} 0 \\ c_3 \\ 1 \end{bmatrix}$$

are both in the span of $\boldsymbol{v}_1, \boldsymbol{v}_3$, and so

$$\boldsymbol{e}_2 = -v_2[1](1/c)\boldsymbol{w}_1 + (1/c)\boldsymbol{v}_2 - (v_2[3]/c)\boldsymbol{w}_3, \quad c = v_2[2] - v_2[1]c_1 - v_2[3]c_2.$$

For $\boldsymbol{e}_3$, we start with the 2×2 matrix in the upper left hand corner of $\boldsymbol{V}$ ... we leave the rest of the argument to the reader.

The argument for $\mathbb{R}^N$ is simply a generalization of the above. To show that $\boldsymbol{e}_j \in \text{Span}(\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\})$, we define the index set $\mathcal{I}_j = \{i \in \mathbb{N} \mid 1 \leq i \leq N,\ i \neq j\}$, and the $N-1 \times N-1$ matrix $\boldsymbol{V}'$ by extracting the jth row and column of the $N \times N$ matrix $\boldsymbol{V}$; the columns of $\boldsymbol{V}'$ are labeled $\boldsymbol{u}_i$ for $i \in \mathcal{I}_j$. If the $\{\boldsymbol{u}_i\}_{i\in\mathcal{I}_j}$ are not linearly independent, then

$$\boldsymbol{0} \in \text{Span}(\{\boldsymbol{u}_i\}_{i\in\mathcal{I}_j}) \ \Rightarrow\ c\boldsymbol{e}_j \in \text{Span}(\{\boldsymbol{v}_i\}_{i\in\mathcal{I}_j}) \ \text{ for some } c \neq 0,$$

and hence $\boldsymbol{e}_j \in \text{Span}(\{\boldsymbol{v}_i,\ i = 1, \ldots, N\})$.

If the $\{\boldsymbol{u}_i\}_{i\in\mathcal{I}_j}$ are linearly independent, then by induction, there exist $\{\boldsymbol{w}_i \in \mathbb{R}^N,\ i \in \mathcal{I}_j\}$ such that

$$w_i[n] = \begin{cases} 1, & i = n, \\ c_i, & i = j, \\ 0, & i \neq n, i \neq j. \end{cases}$$

for some $\{c_i\}$. Then

$$\boldsymbol{v}_j - \sum_{i\in\mathcal{I}_j} v_j[i]\boldsymbol{w}_i = c\boldsymbol{e}_j, \quad c = v_j[j] - \sum_{i\in\mathcal{I}_j} v_j[i]c_i.$$

Again, we know $c \neq 0$, as otherwise the $\{\boldsymbol{v}_n\}$ would not be linearly independent. Thus

$$\boldsymbol{e}_j \in \text{Span}\left(\{\boldsymbol{v}_j\} \cup \{\boldsymbol{w}_i,\ i \in \mathcal{I}_j\}\right) \subset \text{Span}\left(\{\boldsymbol{v}_n,\ n = 1, \ldots, N\}\right).$$

B Technical Details: Proof of Proposition 14

We will assume without loss of generality that $M \leq N$ as we can just as easily apply the argument below on $\boldsymbol{A}^{\mathrm{T}}$ with the row and column spaces changing roles. The proof relies heavily on Proposition 13, as if we use $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M$ to denote the rows of $\boldsymbol{A}$, then

$$\text{Col}(\boldsymbol{A}) = \text{Span}\left\{ \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{x} \\ \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{x} \\ \vdots \\ \boldsymbol{a}_M^{\mathrm{T}}\boldsymbol{x} \end{bmatrix},\quad \boldsymbol{x} \in \mathbb{R}^N \right\}.$$

In fact, if the dimension R of $\text{Row}(\boldsymbol{A})$ is M, then the proposition follows immediately from Proposition 13. So we just need to handle the case where $r < M$.

If $R < M$, then Proposition 1 above tells us that there is a subset of R rows of $\boldsymbol{A}$ that span $\text{Row}(\boldsymbol{A})$; without loss of generality, we will assume that these are the first R rows $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_R$.

We will now argue, similar to Proposition 13, that

$$\boldsymbol{v}_1 = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_1 \\ \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_R^{\mathrm{T}}\boldsymbol{a}_1 \\ - \\ \boldsymbol{a}_{R+1}^{\mathrm{T}}\boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_M^{\mathrm{T}}\boldsymbol{a}_1 \end{bmatrix}, \ldots, \boldsymbol{v}_R = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{a}_R \\ \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_R^{\mathrm{T}}\boldsymbol{a}_R \\ - \\ \boldsymbol{a}_{R+1}^{\mathrm{T}}\boldsymbol{a}_R \\ \vdots \\ \boldsymbol{a}_M^{\mathrm{T}}\boldsymbol{a}_R \end{bmatrix},$$

is a basis for Col($\boldsymbol{A}$). The reason for the explicit separation between the first R components of the vectors above and the last $M-R$ will be clear below. Now let $\boldsymbol{y}$ be a vector in Col($\boldsymbol{A}$); we can write

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}}\boldsymbol{x} \\ \boldsymbol{a}_2^{\mathrm{T}}\boldsymbol{x} \\ \vdots \\ \boldsymbol{a}_R^{\mathrm{T}}\boldsymbol{x} \\ - \\ \boldsymbol{a}_{R+1}^{\mathrm{T}}\boldsymbol{x} \\ \vdots \\ \boldsymbol{a}_M^{\mathrm{T}}\boldsymbol{x} \end{bmatrix} \quad \text{for some} \quad \boldsymbol{x} \in \mathbb{R}^N.$$

From Proposition 13 above, we know that there exists $\alpha_1, \ldots, \alpha_R$ so that top part of $\boldsymbol{y}$ can be written as a linear combination of the top parts of the $\boldsymbol{v}_n,\ n = 1, \ldots, R$; if

$$\boldsymbol{y}' = \sum_{n=1}^{R} \alpha_n \boldsymbol{v}_n,$$

then $y'[n] = y[n]$ for $n = 1, \ldots, R$. We need to show that the bottom entries are equal as well.

Each $\boldsymbol{a}_\ell$ for $R+1 \le \ell \le M$ can be written as a linear combination of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_R$:

$$\boldsymbol{a}_\ell = \sum_{n=1}^{R} \beta_{\ell,n} \boldsymbol{a}_n.$$

Because of this, the bottom entries of every vector in Col($\boldsymbol{A}$) can be written as a fixed linear combination of the first R entries. With $\boldsymbol{y}$ as a particular example, for $R+1 \le \ell \le M$,

$$y[\ell] = \boldsymbol{a}_\ell^{\mathrm{T}}\boldsymbol{x} = \left(\sum_{n=1}^{R} \beta_{\ell,n}\boldsymbol{a}_n\right)^{\mathrm{T}} \boldsymbol{x} = \sum_{n=1}^{R} \beta_{\ell,n}\boldsymbol{a}_n^{\mathrm{T}}\boldsymbol{x} = \sum_{n=1}^{R} \beta_{\ell,n} y[n].$$

Since $\boldsymbol{y}'$ is also in Col($\boldsymbol{A}$), we also have

$$y'[\ell] = \sum_{n=1}^{R} \beta_{\ell,n} y'[n] = \sum_{n=1}^{R} \beta_{\ell,n} y[n] = y[\ell],$$

because we already know that the first R entries of $\boldsymbol{y}$ and $\boldsymbol{y}'$ must match. Thus $\boldsymbol{y}' = \boldsymbol{y}$, and the $\{\boldsymbol{v}_n,\ n = 1, \ldots, R\}$ are a basis for Col($\boldsymbol{A}$), meaning that the dimension of Col($\boldsymbol{A}$) must be R.