Linear dependence

COMS 3251 Fall 2022 (Daniel Hsu)

1 Linear dependence

We say a set of vectors is *linearly dependent* if there is some vector in the set that can be expressed as a linear combination of the others. If a set of vectors is not linearly dependent, we say it is *linearly independent*.[1]

Examples.

1. The set $\{(1,0,0),(0,1,0),(2,2,0)\}$ is linearly dependent, because the third vector is twice the sum of the first two.
2. The set $\{(1,0,0),(1,1,0)\}$ is linearly independent; there is no way to write either vector as a scaling of the other.
3. The empty set is (trivially) linearly independent.
4. Any set containing $\mathbf{0}$ (the empty sum) is linearly dependent. ∎

Equivalent definition: A set of vectors $\mathcal{S}$ is *linearly dependent* if $\mathbf{0}$ can be written as a "not-all-zeros" linear combination of a non-empty subset of $\mathcal{S}$; i.e., for some distinct $\mathbf{v}_1,\dots,\mathbf{v}_k \in \mathcal{S}$ with $k \geq 1$, and some $c_1,\dots,c_k \in \mathbb{R}$ not all equal to 0,

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k \;=\; \mathbf{0}.$$

This version doesn't "blame" any individual vector for the linear dependence.

Example. The set $\{(1,0,0),(0,1,0),(2,2,0)\}$ is linearly dependent because

$$2\begin{bmatrix}1\\0\\0\end{bmatrix} + 2\begin{bmatrix}0\\1\\0\end{bmatrix} + (-1)\begin{bmatrix}2\\2\\0\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix}.$$ ∎

[1] We say a list of vectors $(\mathbf{v}_1,\dots,\mathbf{v}_k)$ is *linearly dependent* (or "$\mathbf{v}_1,\dots,\mathbf{v}_k$ are linearly dependent") if either (i) some vector in the list appears more than once, or (ii) the set $\{\mathbf{v}_1,\dots,\mathbf{v}_k\}$ is linearly dependent. If neither (i) nor (ii) holds, it is *linearly independent*. By "k linearly independent vectors", we mean a linearly independent list of k distinct vectors.

2 CR factorization

The following algorithm takes as input an $m \times n$ matrix A and returns a subset[2] of its columns that (as we'll see) is linearly independent.

Algorithm 1 Greedy algorithm for CR factorization

Input: $A = [\mathbf{a}_1, \dots, \mathbf{a}_n]$, an $m \times n$ matrix.

```
Initialize C to the empty list of m-vectors.
for k = 1, ..., n do
    If a_k is not in CS(C), then append a_k to the end of C.
end for
return C.
```

Example. Consider the execution of Algorithm 1 on the following matrix:

$$A = \begin{bmatrix} \uparrow & \uparrow & \uparrow & \uparrow \\ \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 \\ \downarrow & \downarrow & \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 4 & 4 \\ 3 & 6 & 5 & 4 \end{bmatrix}.$$

- Initially: C is the empty list.
- Iteration $k = 1$: $\mathbf{a}_1 \notin \mathsf{CS}(C)$, so $\mathbf{a}_1$ is appended to C. At the end of this iteration,
$$C = \begin{bmatrix} \uparrow \\ \mathbf{a}_1 \\ \downarrow \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$
- Iteration $k = 2$: $\mathbf{a}_2 = 2\mathbf{a}_1$, so there is no change to C.
- Iteration $k = 3$: $\mathbf{a}_3 \notin \mathsf{CS}(C)$, so $\mathbf{a}_3$ is appended to C. At the end of this iteration,
$$C = \begin{bmatrix} \uparrow & \uparrow \\ \mathbf{a}_1 & \mathbf{a}_3 \\ \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 3 & 5 \end{bmatrix}.$$
- Iteration $k = 4$: $\mathbf{a}_4 = 2\mathbf{a}_3 - 2\mathbf{a}_1$, so there is no change to C. ∎

[2] The algorithm returns a list of columns (in the form of a matrix). However, it will be guaranteed that the columns in the list are distinct.

Let d be the number of m-vectors in C at the end of Algorithm 1, so C is an $m \times d$ matrix. Later, we'll see that the number d is a fundamental property of the matrix A.

Throughout the execution of Algorithm 1, the vectors in C are, by construction, linearly independent (cf. Theorem 5, the Growth Theorem). If a column of A is not appended to C, then it is a linear combination of the previous columns that were appended to C.

Therefore, alongside the execution of Algorithm 1 (or in another loop over the columns of A), we can construct a $d \times n$ matrix R such that, for each $k = 1, \ldots, n$:

- If $\mathbf{a}_k$ was the ith column appended to C, then the kth column of R has a 1 as its ith component and 0's elsewhere.
- If $\mathbf{a}_k$ was not appended to C, then the kth column of R reveals how to express $\mathbf{a}_k$ as a linear combination of the vectors among $\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}$ that were appended to C. By Theorem 1, there is only one choice for this column of R.

Theorem 1 (Unique Representations Theorem)**.** *If the columns of a matrix B are linearly independent, and $B\mathbf{x} = B\mathbf{y}$, then $\mathbf{x} = \mathbf{y}$.*

Proof. Let $B = [\mathbf{b}_1, \ldots, \mathbf{b}_k]$ be matrix whose columns are k linearly independent vectors. Suppose $B\mathbf{x} = B\mathbf{y}$ for some $\mathbf{x} = (x_1, \ldots, x_k)$ and $\mathbf{y} = (y_1, \ldots, y_k)$. Then $B(\mathbf{x} - \mathbf{y}) = \mathbf{0}$, meaning

$$(x_1 - y_1)\,\mathbf{b}_1 + \cdots + (x_k - y_k)\,\mathbf{b}_k \;=\; \mathbf{0}.$$

Suppose for sake of contradiction that $\mathbf{x} \neq \mathbf{y}$. Then $x_i \neq y_i$ for some i; without loss of generality, assume $i = 1$. We can thus "solve for $\mathbf{b}_1$":

$$\mathbf{b}_1 \;=\; -\frac{x_2 - y_2}{x_1 - y_1}\,\mathbf{b}_2 - \cdots - \frac{x_k - y_k}{x_1 - y_1}\,\mathbf{b}_k,$$

so $\mathbf{b}_1$ is a linear combination of the other $\mathbf{b}_i$'s, a contradiction of the linear independence of $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$. Hence we conclude that $\mathbf{x} = \mathbf{y}$. □

The matrix R from above is a transcript for the execution of Algorithm 1 on input A. It also shows how to "reproduce" A via matrix multiplication:

$$A \;=\; CR.$$

This is called the *CR factorization* of A.

Continuing the previous example. For the columns of A that were not included in C, we have

$$\begin{bmatrix}2\\4\\7\end{bmatrix} = 2\begin{bmatrix}1\\2\\3\end{bmatrix} + 0\begin{bmatrix}3\\4\\5\end{bmatrix} = \begin{bmatrix}1&3\\2&4\\3&5\end{bmatrix}\begin{bmatrix}2\\0\end{bmatrix},$$
$$\begin{bmatrix}4\\4\\4\end{bmatrix} = -2\begin{bmatrix}1\\2\\3\end{bmatrix} + 2\begin{bmatrix}3\\4\\5\end{bmatrix} = \begin{bmatrix}1&3\\2&4\\3&5\end{bmatrix}\begin{bmatrix}-2\\2\end{bmatrix}.$$

Therefore,

$$A = \begin{bmatrix}1&3\\2&4\\3&5\end{bmatrix}\begin{bmatrix}1&2&0&-2\\0&0&1&2\end{bmatrix} = CR, \quad \text{where} \quad R = \begin{bmatrix}1&2&0&-2\\0&0&1&2\end{bmatrix}. \quad \blacksquare$$

From the CR factorization $A = CR$, we see that every linear combination of the columns of A is a linear combination of the columns of C. In other words, $\mathsf{CS}(A) = \mathsf{CS}(C)$.

3 Reduced row echelon form

The matrix R described above has a property called *reduced row echelon form*. It is a special case of a property called *row echelon form*.

- We say a matrix is in *row echelon form (REF)* if:
  - any all-zeros row appears below all non-zero rows; and
  - for any non-zero row, the left-most non-zero entry—which is called the *leading entry* (a.k.a. *pivot*) for the row—is in a column that is strictly to the right of the columns that contain leading entries of any previous rows.
- We say a matrix is in *reduced row echelon form (RREF)* if:
  - the matrix is in REF;
  - every leading entry is equal to 1; and
  - the column containing a leading entry has 0's in all other entires.

(It is typical to drop the all-zeros rows of a matrix in REF or RREF.)

Example of a matrix in REF.

$$\begin{bmatrix} \underline{2} & 4 & 10 & 16 \\ 0 & 0 & \underline{5} & 10 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

This matrix has two non-zero rows. The leading entry of each non-zero row is underlined. The leading entry for the first row is in the first column. The leading entry for the second row is in the third column. ∎

Example of a matrix in RREF.

$$\begin{bmatrix} \underline{1} & 2 & 0 & -2 \\ 0 & 0 & \underline{1} & 2 \end{bmatrix}$$ ∎

Our discussion about Algorithm 1 has established the following theorem.

Theorem 2. *The execution of Algorithm 1 on a matrix $A \in \mathbb{R}^{m \times n}$ produces, for some $d \in \{0, \dots, n\}$, a matrix $C \in \mathbb{R}^{m \times d}$ of d linearly independent columns of A; furthermore, there exists a matrix $R \in \mathbb{R}^{d \times n}$ in RREF, without any all-zeros rows, such that $A = CR$.*

A remarkable property of matrices in RREF is the following theorem.

Theorem 3. *The non-zero rows of a matrix in RREF are linearly independent.*

Proof. If the jth row of the matrix has a leading entry in the kth column, then all other non-zero rows of the matrix have 0's in their kth entries, and hence the jth row is not in the span of the other non-zero rows. □

4 Rank

Recall that d denotes the number of linearly independent columns picked out by the execution of Algorithm 1 on A. We'll see next that this number d is a fundamental quantity associated with A.

Theorem 3 implies that the d rows of the aforementioned matrix R are linearly independent. Since $A = CR$, every row of A is a linear combination of the d rows of R.

In fact, it turns out that A must also have d linearly independent rows.

Theorem 4. *For any non-negative integer k and any matrix A, the following statements are equivalent:*

- *A has at least k linearly independent columns.*
- *A has at least k linearly independent rows.*

Proof. Since we can interchange the roles of rows and columns, it suffices to prove that if A has at least k linearly independent rows, then A has at least k linearly independent columns.

So assume A has at least k linearly independent rows. Now consider the execution of Algorithm 1 on A. Say it produces a matrix C with d linearly independent columns; let R be the $d \times n$ matrix in RREF such that $A = CR$ as guaranteed by Theorem 2. Since the rows of A are linear combinations of the d rows of R, and there are at least k linearly independent rows of A (by assumption), it must be that $d \geq k$ (Fact 1). Since C contains d linearly independent columns of A, it follows that A has at least k linearly independent columns. □

Fact 1. *Let $\mathcal{E}$ and $\mathcal{W}$ be finite sets of vectors. If $\mathcal{W}$ is a linearly independent subset of* $\operatorname{span}(\mathcal{E})$*, then* $|\mathcal{W}| \leq |\mathcal{E}|$.

Theorem 4 is a fundamental theorem of linear algebra, tying together the columns of a matrix and the rows of a matrix, which *a priori* may otherwise seem to not have anything to do with each other.

Define the rank of a matrix A to be the (maximum) number of linearly independent columns in A. By Proposition 1 (below), this number is the same as the number of column vectors in C returned by the execution of Algorithm 1 on input A. And by Theorem 4, it is also the (maximum) number of linearly independent rows of A.

Proposition 1. *If a matrix A contains k linearly independent columns, then the execution of Algorithm 1 on input A will produce a matrix C containing k linearly independent columns of A.*

Proof. Suppose A has k linearly independent columns. By Theorem 4, A has k linearly independent rows. The claim is now proved using the same argument from (the second paragraph of) the proof of Theorem 4. □

Corollary 1. *The rank of a matrix A is equal to all of the following:*

- *the number of columns of A in C returned by the execution of Algorithm 1 on input A,*
- *the number of linearly independent columns of A, and*
- *the number of linearly independent rows of A.*

Proof. Apply Theorem 4 and Proposition 1, each with k being the number of linearly independent columns of A. □

A Growth Theorem

Theorem 5 (Growth Theorem)**.** *Let $\mathcal{S}$ be a set of vectors, and let $\mathbf{v}$ be a vector not in $\mathcal{S}$.*

- *If $\mathbf{v} \in \operatorname{span}(\mathcal{S})$, then $\mathcal{S} \cup \{\mathbf{v}\}$ is linearly dependent and*

$$\operatorname{span}(\mathcal{S}) \; = \; \operatorname{span}(\mathcal{S} \cup \{\mathbf{v}\}).$$

- *If $\mathbf{v} \notin \operatorname{span}(\mathcal{S})$, then*

$$\operatorname{span}(\mathcal{S}) \; \subsetneq \; \operatorname{span}(\mathcal{S} \cup \{\mathbf{v}\});$$

  and if, additionally, $\mathcal{S}$ is linearly independent, then so is $\mathcal{S} \cup \{\mathbf{v}\}$.

Proof. Assume $\mathbf{v} \in \operatorname{span}(\mathcal{S})$. Then $\mathbf{v}$ can be written as a linear combination of other vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_m\} \subseteq \mathcal{S}$, say,

$$\mathbf{v} \; = \; a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m.$$

This means that $\mathcal{S} \cup \{\mathbf{v}\}$ is linearly dependent. Now consider any vector $\mathbf{u} \in \operatorname{span}(\mathcal{S} \cup \{\mathbf{v}\})$. This means $\mathbf{u}$ can be written as a linear combination of $\{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subseteq \mathcal{S} \cup \{\mathbf{v}\}$, say,

$$\mathbf{u} \; = \; b_1\mathbf{u}_1 + \cdots + b_n\mathbf{u}_n.$$

We may assume that the $\mathbf{u}_i$'s are distinct (else use a linear combination of fewer vectors from $\mathcal{S} \cup \{\mathbf{v}\}$). If, say, $\mathbf{u}_n = \mathbf{v}$, then we can still write

$$\mathbf{u} \; = \; b_1\mathbf{u}_1 + \cdots + b_{n-1}\mathbf{u}_{n-1} + b_n(a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m),$$

which is a linear combination of vectors from $\mathcal{S}$. Hence we have $\mathbf{u} \in \operatorname{span}(\mathcal{S})$. So we conclude that $\operatorname{span}(\mathcal{S} \cup \{\mathbf{v}\}) \subseteq \operatorname{span}(\mathcal{S})$. Since we clearly also have $\operatorname{span}(\mathcal{S}) \subseteq \operatorname{span}(\mathcal{S} \cup \{\mathbf{v}\})$, it follows that $\operatorname{span}(\mathcal{S}) = \operatorname{span}(\mathcal{S} \cup \{\mathbf{v}\})$.

Now assume $\mathbf{v} \notin \operatorname{span}(\mathcal{S})$. Clearly $\mathbf{v} \in \operatorname{span}(\mathcal{S} \cup \{\mathbf{v}\})$. So $\operatorname{span}(\mathcal{S}) \neq \operatorname{span}(\mathcal{S} \cup \{\mathbf{v}\})$. Since we clearly also have $\operatorname{span}(\mathcal{S}) \subseteq \operatorname{span}(\mathcal{S} \cup \{\mathbf{v}\})$, it follows that $\operatorname{span}(\mathcal{S}) \subsetneq \operatorname{span}(\mathcal{S} \cup \{\mathbf{v}\})$.

Finally, assume both that $\mathbf{v} \notin \operatorname{span}(\mathcal{S})$ and that $\mathcal{S}$ is linearly independent. Suppose for the sake of contradiction that $\mathcal{S} \cup \{\mathbf{v}\}$ is linearly dependent. By assumption, $\mathbf{v}$ is not in $\mathcal{S}$, and $\mathbf{v}$ is not a linear combination of vectors in $\mathcal{S}$. So the linear dependence of $\mathcal{S} \cup \{\mathbf{v}\}$ implies that there is a vector $\mathbf{u} \in \mathcal{S}$ that is not equal to $\mathbf{v}$, but can be written as a linear combination of $\mathbf{v}$ and some $\{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subseteq \mathcal{S} \setminus \{\mathbf{u}\}$, say,

$$\mathbf{u} \;=\; b_0\mathbf{v} + b_1\mathbf{u}_1 + \cdots + b_n\mathbf{u}_n.$$

If $b_0 = 0$, then we have expressed a vector in $\mathcal{S}$ as a linear combination of other vectors in $\mathcal{S}$, a contradiction of the assumption that $\mathcal{S}$ is linearly independent. If $b_0 \neq 0$, then we can "solve for $\mathbf{v}$" and write

$$\mathbf{v} \;=\; b_0^{-1}\mathbf{u} - b_0^{-1}b_1\mathbf{u}_1 - \cdots - b_0^{-1}b_n\mathbf{u}_n,$$

which expresses $\mathbf{v}$ as a linear combination of vectors from $\mathcal{S}$, a contradiction of the assumption that $\mathbf{v} \notin \operatorname{span}(\mathcal{S})$. Therefore, we conclude that $\mathcal{S} \cup \{\mathbf{v}\}$ is linearly independent. □

B Removal Theorem

Theorem 6 (Removal Theorem)**.** *Let $\mathcal{S}$ be a set of vectors.*

- *If $\mathcal{S}$ is linearly dependent, then there is a vector $\mathbf{v} \in \mathcal{S}$ such that*

$$\operatorname{span}(\mathcal{S} \setminus \{\mathbf{v}\}) \;=\; \operatorname{span}(\mathcal{S}).$$

- *If $\mathcal{S}$ is linearly independent, then every proper subset $\mathcal{S}' \subsetneq \mathcal{S}$ is linearly independent and*

$$\operatorname{span}(\mathcal{S}') \;\subsetneq\; \operatorname{span}(\mathcal{S}).$$

Proof. Assume $\mathcal{S}$ is linearly dependent. Therefore, there exists $\mathbf{v} \in \mathcal{S}$ such that $\mathbf{v}$ can be written as a linear combination of other vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_m\} \subseteq \mathcal{S} \setminus \{\mathbf{v}\}$, say,

$$\mathbf{v} \;=\; a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m.$$

Now consider any vector $\mathbf{u} \in \operatorname{span}(\mathcal{S})$. This means $\mathbf{u}$ can be written as a linear combination of $\{\mathbf{u}_1, \dots, \mathbf{u}_n\} \subseteq \mathcal{S}$, say,

$$\mathbf{u} \;=\; b_1\mathbf{u}_1 + \cdots + b_n\mathbf{u}_n.$$

We may assume that the $\mathbf{u}_i$'s are distinct (else use a linear combination of fewer vectors from $\mathcal{S}$). If, say, $\mathbf{u}_n = \mathbf{v}$, then we can still write

$$\mathbf{u} \;=\; b_1\mathbf{u}_1 + \cdots + b_{n-1}\mathbf{u}_{n-1} + b_n(a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m),$$

which is a linear combination of vectors from $\mathcal{S} \setminus \{\mathbf{v}\}$. Hence we have $\mathbf{u} \in \operatorname{span}(\mathcal{S} \setminus \{\mathbf{v}\})$. So we conclude that $\operatorname{span}(\mathcal{S}) \subseteq \operatorname{span}(\mathcal{S} \setminus \{\mathbf{v}\})$. Since we clearly also have $\operatorname{span}(\mathcal{S} \setminus \{\mathbf{v}\}) \subseteq \operatorname{span}(\mathcal{S})$, it follows that $\operatorname{span}(\mathcal{S} \setminus \{\mathbf{v}\}) = \operatorname{span}(\mathcal{S})$.

Now instead assume $\mathcal{S}$ is linearly independent. Consider any proper subset $\mathcal{S}' \subsetneq \mathcal{S}$, and take any $\mathbf{v} \in \mathcal{S} \setminus \mathcal{S}'$. First, $\mathcal{S}'$ is linearly independent since otherwise there exists a vector in $\mathcal{S}'$ (and hence in $\mathcal{S}$) that can be written as a linear combination of other vectors in $\mathcal{S}'$ (which are also in $\mathcal{S}$). Next, suppose for sake of contradiction that $\mathbf{v}$ can be written as a linear combination of $\{\mathbf{v}_1, \dots, \mathbf{v}_m\} \subseteq \mathcal{S}'$, say,

$$\mathbf{v} \;=\; a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m,$$

then there exists a vector in $\mathcal{S}$ (namely $\mathbf{v}$) that can be written as a linear combination of other vectors in $\mathcal{S} \setminus \{\mathbf{v}\}$ (which happen to be in $\mathcal{S}'$). This conclusion contradicts the assumption that $\mathcal{S}$ is linearly independent. Hence $\mathbf{v}$ is not in $\operatorname{span}(\mathcal{S}')$. Since $\mathbf{v} \in \mathcal{S} \subseteq \operatorname{span}(\mathcal{S})$, it must be that $\operatorname{span}(\mathcal{S}') \neq \operatorname{span}(\mathcal{S})$. And since we clearly have $\operatorname{span}(\mathcal{S}') \subseteq \operatorname{span}(\mathcal{S})$, it must be that $\operatorname{span}(\mathcal{S}') \subsetneq \operatorname{span}(\mathcal{S})$. □

C Exchange Theorem

Theorem 7, given below, is an elaboration of Fact 1.

Theorem 7 (Exchange Theorem). *Let $\mathcal{E}$ and $\mathcal{W}$ be finite sets of vectors. If $\mathcal{W}$ is a linearly independent subset of* $\operatorname{span}(\mathcal{E})$, *then*

- $|\mathcal{W}| \leq |\mathcal{E}|$, *and*
- *there is a subset $\mathcal{F} \subseteq \mathcal{E}$ with $|\mathcal{F}| = |\mathcal{E}| - |\mathcal{W}|$ such that* $\operatorname{span}(\mathcal{E}) = \operatorname{span}(\mathcal{W} \cup \mathcal{F})$.

Proof. Let $m = |\mathcal{E}|$ and $n = |\mathcal{W}|$. The proof is by induction on n. If $n = 0$, then clearly $n \leq m$, and we can take $\mathcal{F} = \mathcal{E}$ to establish the rest of the claim.

Now assume, as the "inductive hypothesis", that the claim holds for a particular value of $n \geq 0$. To complete the "inductive step", we show that if $\mathcal{W} \subseteq \mathrm{span}(\mathcal{E})$ is a set of $n+1$ linearly independent vectors from $\mathrm{span}(\mathcal{E})$, then $n+1 \leq m$, and there exists a subset $\mathcal{F} \subseteq \mathcal{E}$ with $|\mathcal{F}| = m-(n+1)$ such that $\mathrm{span}(\mathcal{E}) = \mathrm{span}(\mathcal{W} \cup \mathcal{F})$.

So let $\mathcal{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_{n+1}\} \subseteq \mathrm{span}(\mathcal{E})$ be $n+1$ linearly independent vectors from $\mathrm{span}(\mathcal{E})$. The subset $\mathcal{W}^- = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\} = \mathcal{W} \setminus \{\mathbf{w}_{n+1}\}$ is also linearly independent (by Theorem 6, the Removal Theorem). By the "inductive hypothesis", we have $n \leq m$, and also there exists a subset $\mathcal{F}^+ = \{\mathbf{f}_1, \ldots, \mathbf{f}_{m-n}\} \subseteq \mathcal{E}$ with $|\mathcal{F}^+| = m-n$ such that

$$\mathrm{span}(\mathcal{E}) \;=\; \mathrm{span}(\mathcal{W}^- \cup \mathcal{F}^+). \tag{1}$$

Since $\mathbf{w}_{n+1} \in \mathrm{span}(\mathcal{E}) = \mathrm{span}(\mathcal{W}^- \cup \mathcal{F}^+)$ as per (1), we have

$$\mathbf{w}_{n+1} \;=\; a_1\mathbf{w}_1 + \cdots + a_n\mathbf{w}_n + b_1\mathbf{f}_1 + \cdots + b_{m-n}\mathbf{f}_{m-n} \tag{2}$$

for some scalars $a_1, \ldots, a_n, b_1, \ldots, b_{m-n}$. If $n = m$ or $b_1 = \cdots = b_{m-n} = 0$, then (2) expresses $\mathbf{w}_{n+1}$ as a linear combination of $\mathbf{w}_1, \ldots, \mathbf{w}_n$, which is impossible since $\mathbf{w}_1, \ldots, \mathbf{w}_{n+1}$ are linearly independent by assumption. Hence, we must have $n+1 \leq m$ (as claimed), and also that $b_i \neq 0$ for some $i \in \{1, \ldots, m-n\}$. Without loss of generality, assume that $b_1 \neq 0$, and then "solve for $\mathbf{f}_1$" in (2):

$$\mathbf{f}_1 \;=\; b_1^{-1}\mathbf{w}_{n+1} - b_1^{-1}a_1\mathbf{w}_1 - \cdots - b_1^{-1}a_n\mathbf{w}_n - b_1^{-1}b_2\mathbf{f}_2 - \cdots - b_1^{-1}b_{m-n}\mathbf{f}_{m-n}. \tag{3}$$

It is the vector $\mathbf{f}_1$ that will be "replaced" by $\mathbf{w}_{n+1}$.

Define

$$\mathcal{F} \;=\; \{\mathbf{f}_2, \ldots, \mathbf{f}_{m-n}\} \;=\; \mathcal{F}^+ \setminus \{\mathbf{f}_1\},$$

which has $|\mathcal{F}| = m-n-1 = m-(n+1)$ vectors. From (3), we see that

$$\mathbf{f}_1 \;\in\; \mathrm{span}(\{\mathbf{w}_1, \ldots, \mathbf{w}_{n+1}, \mathbf{f}_2, \ldots, \mathbf{f}_{m-n}\}) \;=\; \mathrm{span}(\mathcal{W} \cup \mathcal{F}).$$

Since we also clearly have $\mathcal{W}^- \cup \mathcal{F} \subseteq \mathrm{span}(\mathcal{W} \cup \mathcal{F})$, it follows that

$$\mathcal{W}^- \cup \mathcal{F}^+ \;=\; \mathcal{W}^- \cup \{\mathbf{f}_1\} \cup \mathcal{F} \;\subseteq\; \mathrm{span}(\mathcal{W} \cup \mathcal{F}). \tag{4}$$

Recalling (1) from the "inductive hypothesis", we have $\mathrm{span}(\mathcal{E}) = \mathrm{span}(\mathcal{W}^- \cup \mathcal{F}^+)$, which means that every vector in $\mathrm{span}(\mathcal{E})$ is a linear combination of the vectors in $\mathcal{W}^- \cup \mathcal{F}^+$; and by (4), each of the vectors in $\mathcal{W}^- \cup \mathcal{F}^+$ is a linear combination of the vectors in $\mathcal{W} \cup \mathcal{F}$. From this argument, we obtain $\mathrm{span}(\mathcal{E}) \subseteq \mathrm{span}(\mathcal{W} \cup \mathcal{F})$. Since the $\mathcal{W} \cup \mathcal{F} \subseteq \mathrm{span}(\mathcal{E})$, it follows that

$$\mathrm{span}(\mathcal{E}) \;=\; \mathrm{span}(\mathcal{W} \cup \mathcal{F})$$

as claimed.

We have thus completed the proof of the "inductive step", so the overall claim follows by the principle of mathematical induction. □

The proof of Theorem 7 also justifies Algorithm 2, given below, which finds the subset $\mathcal{F}$ as guaranteed under the conditions of Theorem 7.

Algorithm 2 Exchange algorithm

Input: Two lists of distinct vectors $[\mathbf{e}_1, \dots, \mathbf{e}_m]$ and $[\mathbf{w}_1, \dots, \mathbf{w}_n]$.

if $n > m$ **then**
  return FAIL ("$[\mathbf{w}_1, \dots, \mathbf{w}_n]$ is not linearly independent.")
end if
Initialize $F = [\mathbf{f}_1, \dots, \mathbf{f}_m] = [\mathbf{e}_1, \dots, \mathbf{e}_m]$.
for $k = 1, \dots, n$ **do**
  Find scalars $a_1, \dots, a_{k-1}$ and $b_1, \dots, b_{m-k+1}$ such that

$$\mathbf{w}_k \;=\; a_1\mathbf{w}_1 + \cdots + a_{k-1}\mathbf{w}_{k-1} + b_1\mathbf{f}_1 + \cdots + b_{m-k+1}\mathbf{f}_{m-k+1}.$$

  if no such scalars are found **then**
    return FAIL ("$\mathbf{w}_k \notin \mathrm{span}(\{\mathbf{e}_1, \dots, \mathbf{e}_m\})$.")
  else if $b_1 = \cdots = b_{m-k+1} = 0$ **then**
    return FAIL ("$[\mathbf{w}_1, \dots, \mathbf{w}_k]$ is not linearly independent.")
  end if
  Pick any $i \in \{1, \dots, m-k+1\}$ such that $b_i \neq 0$.
  Discard $\mathbf{f}_i$, and re-number the remaining vectors $F = [\mathbf{f}_1, \dots, \mathbf{f}_{m-k}]$.
end for
return F.