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. \(\blacksquare\)

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

\[
    2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + (-1) \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \quad \blacksquare
\]

\(^1\)We say a list of vectors \( (\mathbf{v}_1, \ldots, \mathbf{v}_k) \) is *linearly dependent* (or “\( \mathbf{v}_1, \ldots, \mathbf{v}_k \) are linearly dependent”) if either (i) some vector in the list appears more than once, or (ii) the set \( \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \) 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 of its columns that (as we’ll see) is linearly independent.

**Algorithm 1** Greedy algorithm for CR factorization

**Input:** $A = [a_1, \ldots, a_n]$, an $m \times n$ matrix.

1. Initialize $C$ to the empty list of $m$-vectors.
2. for $k = 1, \ldots, n$ do
3. If $a_k$ is not in $\text{CS}(C)$, then append $a_k$ to the end of $C$.
4. end for
5. return $C$.

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

$$A = \begin{bmatrix}
\uparrow & \uparrow & \uparrow & \uparrow \\
a_1 & a_2 & a_3 & a_4 \\
\downarrow & \downarrow & \downarrow & \downarrow
\end{bmatrix} = \begin{bmatrix}
1 & 2 & 3 & 4 \\
2 & 4 & 4 & 4 \\
3 & 6 & 5 & 4
\end{bmatrix}.$$

- Initially: $C$ is the empty list.
- Iteration $k = 1$: $a_1 \notin \text{CS}(C)$, so $a_1$ is appended to $C$. At the end of this iteration,
  $$C = \begin{bmatrix}
\uparrow \\
a_1 \\
\downarrow
\end{bmatrix} = \begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix}.$$
- Iteration $k = 2$: $a_2 = 2a_1$, so there is no change to $C$.
- Iteration $k = 3$: $a_3 \notin \text{CS}(C)$, so $a_3$ is appended to $C$. At the end of this iteration,
  $$C = \begin{bmatrix}
\uparrow & \uparrow \\
a_1 & a_3 \\
\downarrow & \downarrow
\end{bmatrix} = \begin{bmatrix}
1 & 3 \\
2 & 4 \\
3 & 5
\end{bmatrix}.$$
- Iteration $k = 4$: $a_4 = 2a_3 - 2a_1$, so there is no change to $C$.  

\footnote{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 \( a_k \) was the \( i \)th column appended to \( C \), then the \( k \)th column of \( R \) has a 1 as its \( i \)th component and 0’s elsewhere.

- If \( a_k \) was not appended to \( C \), then the \( k \)th column of \( R \) reveals how to express \( a_k \) as a linear combination of the vectors among \( a_1, \ldots, a_{k-1} \) 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 \( Bx = By \), then \( x = y \).*

**Proof.** Let \( B = [b_1, \ldots, b_k] \) be matrix whose columns are \( k \) linearly independent vectors. Suppose \( Bx = By \) for some \( x = (x_1, \ldots, x_k) \) and \( y = (y_1, \ldots, y_k) \). Then \( B(x - y) = 0 \), meaning

\[
(x_1 - y_1)b_1 + \cdots + (x_k - y_k)b_k = 0.
\]

Suppose for sake of contradiction that \( x \neq y \). Then \( x_i \neq y_i \) for some \( i \); without loss of generality, assume \( i = 1 \). We can thus “solve for \( b_1 \)”: \[ b_1 = \frac{-x_2 - y_2}{x_1 - y_1} b_2 - \cdots - \frac{x_k - y_k}{x_1 - y_1} b_k, \]
so \( b_1 \) is a linear combination of the other \( b_i \)’s, a contradiction of the linear independence of \( \{b_1, \ldots, b_k\} \). Hence we conclude that \( x = y \). \( \square \)

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

\[
\begin{bmatrix}
2 \\
4 \\
7
\end{bmatrix} = 2 \begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix} + 0 \begin{bmatrix}
3 \\
4 \\
5
\end{bmatrix} = \begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix} \begin{bmatrix}
2 \\
4 \\
5
\end{bmatrix}.
\]

\[
\begin{bmatrix}
4 \\
4 \\
4
\end{bmatrix} = -2 \begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix} + 2 \begin{bmatrix}
3 \\
4 \\
5
\end{bmatrix} = \begin{bmatrix}
1 \\
2 \\
3
\end{bmatrix} \begin{bmatrix}
-2 \\
2 \\
4
\end{bmatrix}.
\]

Therefore,

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

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, $\text{CS}(A) = \text{CS}(C)$.

3 Reduced row echelon form

The matrix $R$ described above has a property called \textit{reduced row echelon form}. It is a special case of a property called \textit{row echelon form}.

- We say a matrix is in \textit{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 \textit{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 \textit{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 entries.

(It is typical to drop the all-zeros rows of a matrix in REF or RREF.)
Example of a matrix in REF.
\[
\begin{bmatrix}
2 & 4 & 10 & 16 \\
0 & 0 & 5 & 10 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]
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.
\[
\begin{bmatrix}
1 & 2 & 0 & -2 \\
0 & 0 & 1 & 2
\end{bmatrix}
\]

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 \(d \in \{0, \ldots, n\}\), 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 \(j\)th row of the matrix has a leading entry in the \(k\)th column, then all other non-zero rows of the matrix have 0’s in their \(k\)th entries, and hence the \(j\)th row is not in the span of the other non-zero rows.

\[\square\]

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 $\text{span}(\mathcal{E})$, 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. \qed
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 $A$. 

$\square$
A Growth Theorem

Theorem 5 (Growth Theorem). Let $S$ be a set of vectors, and let $v$ be a vector not in $S$.

- If $v \in \text{span}(S)$, then $S \cup \{v\}$ is linearly dependent and
  \[
  \text{span}(S) = \text{span}(S \cup \{v\}).
  \]

- If $v \notin \text{span}(S)$, then
  \[
  \text{span}(S) \subsetneq \text{span}(S \cup \{v\});
  \]
  and if, additionally, $S$ is linearly independent, then so is $S \cup \{v\}$.

Proof. Assume $v \in \text{span}(S)$. Then $v$ can be written as a linear combination of other vectors $\{v_1, \ldots, v_m\} \subseteq S$, say,

$$v = a_1v_1 + \cdots + a_mv_m.$$

This means that $S \cup \{v\}$ is linearly dependent. Now consider any vector $u \in \text{span}(S \cup \{v\})$. This means $u$ can be written as a linear combination of $\{u_1, \ldots, u_n\} \subseteq S \cup \{v\}$, say,

$$u = b_1u_1 + \cdots + b_nu_n.$$

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

$$u = b_1u_1 + \cdots + b_{n-1}u_{n-1} + b_n(a_1v_1 + \cdots + a_mv_m),$$

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

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

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

\[\Box\]

## B Removal Theorem

**Theorem 6** (Removal Theorem). Let \( S \) be a set of vectors.

- If \( S \) is linearly dependent, then there is a vector \( \mathbf{v} \in S \) such that

\[
\text{span}(S \setminus \{ \mathbf{v} \}) = \text{span}(S).
\]

- If \( S \) is linearly independent, then every proper subset \( S' \subset S \) is linearly independent and

\[
\text{span}(S') \subsetneq \text{span}(S).
\]

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

\[
\mathbf{v} = a_1 \mathbf{v}_1 + \cdots + a_m \mathbf{v}_m.
\]

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

\[
\mathbf{u} = b_1 \mathbf{u}_1 + \cdots + b_n \mathbf{u}_n.
\]
We may assume that the \( u_i \)'s are distinct (else use a linear combination of fewer vectors from \( S \)). If, say, \( u_n = v \), then we can still write

\[
u = b_1 u_1 + \cdots + b_{n-1} u_{n-1} + b_n (a_1 v_1 + \cdots + a_m v_m),
\]

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

Now instead assume \( S \) is linearly independent. Consider any proper subset \( S' \subset S \), and take any \( v \in S \setminus S' \). First, \( S' \) is linearly independent since otherwise there exists a vector in \( S' \) (and hence in \( S \)) that can be written as a linear combination of other vectors in \( S' \) (which are also in \( S \)). Next, suppose for sake of contradiction that \( v \) can be written as a linear combination of \( \{v_1, \ldots, v_m\} \subseteq S' \), say,

\[
v = a_1 v_1 + \cdots + a_m v_m,
\]

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

C  Exchange Theorem

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

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

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

**Proof.** Let \( m = |E| \) and \( n = |W| \). The proof is by induction on \( n \). If \( n = 0 \), then clearly \( n \leq m \), and we can take \( F = E \) 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 \( \mathcal{W} \subseteq \text{span}(\mathcal{E}) \) is a set of \( n+1 \) linearly independent vectors from \( \text{span}(\mathcal{E}) \), then \( n+1 \leq m \), and there exists a subset \( \mathcal{F} \subseteq \mathcal{E} \) with \( |\mathcal{F}| = m - (n+1) \) such that \( \text{span}(\mathcal{E}) = \text{span}(\mathcal{W} \cup \mathcal{F}) \).

So let \( \mathcal{W} = \{w_1, \ldots, w_{n+1}\} \subseteq \text{span}(\mathcal{E}) \) be \( n+1 \) linearly independent vectors from \( \text{span}(\mathcal{E}) \). The subset \( \mathcal{W}^- = \{w_1, \ldots, w_n\} = \mathcal{W} \setminus \{w_{n+1}\} \) 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 \( \mathcal{F}^+ = \{f_1, \ldots, f_{m-n}\} \subseteq \mathcal{E} \) with \( |\mathcal{F}^+| = m - n \) such that

\[
\text{span}(\mathcal{E}) = \text{span}((\mathcal{W}^- \cup \mathcal{F}^+)). \tag{1}
\]

Since \( w_{n+1} \in \text{span}(\mathcal{E}) = \text{span}(\mathcal{W}^- \cup \mathcal{F}^+) \) as per (1), we have

\[
w_{n+1} = a_1w_1 + \cdots + a_nw_n + b_1f_1 + \cdots + b_{m-n}f_{m-n} \tag{2}
\]

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 \( w_{n+1} \) as a linear combination of \( w_1, \ldots, w_n \), which is impossible since \( w_1, \ldots, w_{n+1} \) 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 \( f_1 \)” in (2)

\[
f_1 = b_1^{-1}w_{n+1} - b_1^{-1}a_1w_1 - \cdots - b_1^{-1}a_nw_n - b_1^{-1}b_2f_2 - \cdots - b_1^{-1}b_{m-n}f_{m-n}. \tag{3}
\]

It is the vector \( f_1 \) that will be “replaced” by \( w_{n+1} \).

Define

\[
\mathcal{F} = \{f_2, \ldots, f_{m-n}\} = \mathcal{F}^+ \setminus \{f_1\},
\]

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

\[
f_1 \in \text{span}(\{w_1, \ldots, w_{n+1}, f_2, \ldots, f_{m-n}\}) = \text{span}(\mathcal{W} \cup \mathcal{F}).
\]

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

\[
\mathcal{W}^- \cup \mathcal{F}^+ = \mathcal{W}^- \cup \{f_1\} \cup \mathcal{F} \subseteq \text{span}(\mathcal{W} \cup \mathcal{F}). \tag{4}
\]

Recalling (1) from the “inductive hypothesis”, we have \( \text{span}(\mathcal{E}) = \text{span}(\mathcal{W}^- \cup \mathcal{F}^+) \), which means that every vector in \( \text{span}(\mathcal{E}) \) 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 $\text{span}(\mathcal{E}) \subseteq \text{span}(\mathcal{W} \cup \mathcal{F})$. Since the $\mathcal{W} \cup \mathcal{F} \subseteq \text{span}(\mathcal{E})$, it follows that
\[
\text{span}(\mathcal{E}) = \text{span}(\mathcal{W} \cup \mathcal{F})
\]
as claimed.

We have thus completed the proof of the “inductive step”, so the overall claim follows by the principle of mathematical induction. \qed

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 $[e_1, \ldots, e_m]$ and $[w_1, \ldots, w_n]$.

1. if $n > m$ then
2. return FAIL (“$[w_1, \ldots, w_n]$ is not linearly independent.”)
3. end if
4. Initialize $F = [f_1, \ldots, f_m] = [e_1, \ldots, e_m]$.
5. for $k = 1, \ldots, n$ do
6. Find scalars $a_1, \ldots, a_{k-1}$ and $b_1, \ldots, b_{m-k+1}$ such that
\[
w_k = a_1w_1 + \cdots + a_{k-1}w_{k-1} + b_1f_1 + \cdots + b_{m-k+1}f_{m-k+1}.
\]
7. if no such scalars are found then
8. return FAIL (“$w_k \notin \text{span}(\{e_1, \ldots, e_m\})$.”)
9. else if $b_1 = \cdots = b_{m-k+1} = 0$ then
10. return FAIL (“$[w_1, \ldots, w_k]$ is not linearly independent.”)
11. end if
12. Pick any $i \in \{1, \ldots, m - k + 1\}$ such that $b_i \neq 0$.
13. Discard $f_i$, and re-number the remaining vectors $F = [f_1, \ldots, f_{m-k}]$.
14. end for
15. return $F$. 

12