Chapter 3

Subspaces and rank

Contents (class version)

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.0 Introduction</td>
<td>3.2</td>
</tr>
<tr>
<td>3.1 Subspaces</td>
<td>3.3</td>
</tr>
<tr>
<td>Span</td>
<td>3.6</td>
</tr>
<tr>
<td>Linear independence</td>
<td>3.8</td>
</tr>
<tr>
<td>Basis</td>
<td>3.10</td>
</tr>
<tr>
<td>Dimension</td>
<td>3.13</td>
</tr>
<tr>
<td>Sums and intersections of subspaces</td>
<td>3.14</td>
</tr>
<tr>
<td>Direct sum of subspaces</td>
<td>3.16</td>
</tr>
<tr>
<td>Dimensions of sums of subspaces</td>
<td>3.17</td>
</tr>
<tr>
<td>Orthogonal complement of a subspace</td>
<td>3.18</td>
</tr>
<tr>
<td>Linear transformations</td>
<td>3.19</td>
</tr>
<tr>
<td>Range of a matrix</td>
<td>3.20</td>
</tr>
<tr>
<td>3.2 Rank of a matrix</td>
<td>3.21</td>
</tr>
<tr>
<td>Rank of a matrix product</td>
<td>3.24</td>
</tr>
<tr>
<td>3.3 Nullspace and the SVD</td>
<td>3.27</td>
</tr>
</tbody>
</table>
An important operation in signal processing and machine learning is dimensionality reduction. There are many such methods, but the starting point is usually linear methods that map data to a lower-dimensional set called a subspace. The notion of dimension is quantified by rank. This chapter reviews subspaces, span, dimension, rank and null space. These linear algebra concepts may seem to lack signal processing context initially, but they are crucial to thoroughly understanding the SVD, a primary tool for the rest of the course.

The reading for this chapter is [1, §2.2-2.4, 3.4, 3.1, 3.5, 5.1].
Define [1, Def. 2.6]. For a vector space $\mathcal{V}$ defined on a field $\mathbb{F}$, a nonempty subset $S \subseteq \mathcal{V}$ is called a subspace or linear subspace of $\mathcal{V}$ iff

- $S$ is closed under vector addition: $u, v \in S \implies u + v \in S$
- $S$ is closed under scalar multiplication: $v \in S$ and $\alpha \in \mathbb{F} \implies \alpha v \in S$

Fact. A subspace $S$ always includes the zero vector $0$.

Proof. Because a field $\mathbb{F}$ always includes the scalar $0$, and because $S$ is nonempty, it contains some vector $v$. Because $S$ is closed under scalar multiplication, $S$ contains the vector $0v$ which is the zero vector. $\square$

Laub [1, p. 9] uses the notation $S \subseteq \mathcal{V}$ to indicate that $S$ is a subspace of $\mathcal{V}$, although the more usual meaning of the symbol $\subseteq$ is to denote a subset.

When visualizing subspaces, think of lines or planes or hyperplanes going through the origin $0$. 
Example. $S = \{0\}$ where $0 \in \mathcal{V}$ for some vector space $\mathcal{V}$.
This is the most minimalist and uninteresting subspace.

Example. The subspace of symmetric matrices $S = \{A \in \mathbb{R}^{N \times N} : A \text{ is symmetric}\}$.
Here $\mathcal{V} = \mathbb{R}^{N \times N}$ and $S \subseteq \mathcal{V}$.
It is easy to verify that $S$ is closed under vector addition and scalar multiplication.

Example. The subset of orthogonal matrices $S = \{A \in \mathbb{R}^{N \times N} : A' A = I\}$.
Is this a subspace of $\mathbb{R}^{N \times N}$?
A: Y  B: N  C: Insufficient information

Example. $S = \{\alpha 1_N : \alpha \in \mathbb{C}\} \subseteq \mathbb{C}^N$. We will see shortly that here $S = \text{span}(1_N)$.
Periodic functions as a subspace

Let $\mathcal{V}$ denote the vector space consisting of all 1D periodic functions having period $T$ for some $T \neq 0$.

In other words, if $f \in \mathcal{V}$, then $f(t + T) = f(t) \forall t \in \mathbb{R}$.

One can verify that $\mathcal{V}$ is indeed a vector space, i.e., it is closed under vector addition and scalar multiplication.

Now consider the set $\mathcal{S}$ of 1D functions having period $T > 0$ that are band-limited to maximum frequency $KT$ for some $K \in \mathbb{N}$.

Exercise. Verify that $\mathcal{S}$ is a subspace in $\mathcal{V}$, i.e., $\mathcal{S} \subseteq \mathcal{V}$.

Is the set of all 1D periodic functions a vector space?

No, it is not closed under summation because the sum of two periodic functions with different periods whose ratio is irrational is aperiodic.
Span

Define. Given a set of vectors \( \{u_1, \ldots, u_N\} \) in a vector space \( V \) over field \( \mathbb{F} \), the span of those vectors is

\[
\text{span}(\{u_1, \ldots, u_N\}) \triangleq \{ \sum_{n=1}^{N} \alpha_n u_n : \alpha_n \in \mathbb{F} \}\]  

(3.1)

This set is also called the linear span or hull.

Often we will collect the vectors into a matrix \( U = [u_1 \ldots u_N] \) and define \( \text{span}(U) \triangleq \text{span}(\{u_1, \ldots, u_N\}) \), although the usual mathematical definition is that the argument of \( \text{span}(\cdot) \) is a set of vectors, not a matrix.

Exercise. Verify that \( \text{span}(\{u_1, \ldots, u_N\}) \) is a subspace of the vector space \( V \).

Example. For \( V = \mathbb{R}^3 \), if \( u_1 = (1, 0, 1) \) and \( u_2 = (1, 0, -1) \) then \( \text{span}(\{u_1, u_2\}) \) is the entire \((x, z)\) plane.

Example. For \( V = \mathbb{R}^{N \times N} \), the vector space of \( N \times N \) matrices, \( \text{span}(\{e_1 e'_1, \ldots, e_N e'_N\}) \) is the subspace of all \( N \times N \) diagonal matrices because

\[
\begin{bmatrix}
  d_1 \\
  \cdots \\
  d_N
\end{bmatrix} = d_1 \begin{bmatrix}
  1 \\
  \cdots \\
  0
\end{bmatrix} + \cdots + d_N \begin{bmatrix}
  0 \\
  \cdots \\
  0
\end{bmatrix} = \begin{bmatrix}
  d_1 \\
  \cdots \\
  d_N
\end{bmatrix}
\]
Span of infinite collection of vectors

The definition of span in (3.1) is for a finite set of vectors. Some infinite-dimensional vector spaces are also important, such as the space of $T$-periodic functions above. Another example is the vector space of all polynomials. To work with such vector spaces, we use the following more general definition of span.

Define. If $S$ is a (possibly uncountably infinite) subset of a vector space $V$ over field $F$, the span of $S$ consists of all (finite by definition) linear combinations of elements in $S$:

$$\text{span}(S) \triangleq \left\{ \mathbf{x} \in V : \mathbf{x} = \sum_{n=1}^{N} \alpha_n \mathbf{u}_n, \; \mathbf{u}_n \in S, \; \alpha_n \in F, \; N \in \mathbb{N} \right\}. \quad (3.2)$$

Define. The span of the empty set is the zero vector: $\text{span}(\emptyset) = \mathbf{0}$.

These definitions ensure that the span of any set (empty, finite, or infinite) is a subspace.

Example. Let $V = \mathbb{R}^3$ and consider the following (uncountably infinite) set:

$$S = \{\alpha(1, 1, 1) : \alpha \in \mathbb{R}\} \cup \{(1, 0, 0)\}.$$

In words: $S$ is a line (through the origin) and another point not on that line. What is $\text{span}(S)$?

A: a line  B: a plane  C: all of $\mathbb{R}^3$  D: None of these
**Linear independence**

Define. A set of vectors $u_1, \ldots, u_N$ is **linearly dependent** iff there exists a tuple of coefficients $\alpha_1, \ldots, \alpha_N \in \mathbb{F}$, not all of which are zero, where

$$\sum_{n=1}^{N} \alpha_n u_n = 0.$$ 

In words, a set of vectors is **linearly dependent** iff any one of the vectors is in the **span** of the other vectors. This latter interpretation generalizes to infinite dimensional vector spaces.

A set of vectors that is not linearly dependent is called a **linearly independent set**. L§2.3

Define. A set of vectors $u_1, \ldots, u_N$ in a vector space is **linearly independent** iff for any scalars $\alpha_1, \ldots, \alpha_N \in \mathbb{F}$:

$$\sum_{n=1}^{N} \alpha_n u_n = 0 \implies \alpha_1 = \ldots = \alpha_N = 0.$$ 

In other words, no **linear combination** of the vectors is zero, except when all the coefficients are zero.

**Example.** In $\mathbb{R}^{N \times N}$, the set of matrices $\{e_1e_1', \ldots, e_N e_N'\}$ is linearly independent.
Generalization to infinite sets

Recall that \( S - \{x\} \) denotes the set \( S \) with the vector \( x \) removed.

Define. A (possibly uncountably infinite) set of vectors \( S \) in a vector space is **linearly dependent** iff

\[
\exists x \in S \text{ s.t. } x \in \text{span}(S - \{x\}),
\]

otherwise the set is called **linearly independent**.

Example. Consider the vector space of all polynomials. Let \( S \) denote the (countably infinite) set of all **monomials**, i.e.,

\[
S = \{ f(x) = x^n : n = 0, 1, \ldots \}.
\]

One can show by elementary algebra [wiki] that \( S \) is linearly independent. (One cannot write monomial like \( x^5 \) as a linear combination of other monomials.)
Basis

Now we use the concepts of linear independence and span to define a particularly important concept: basis.

Define [1, Def. 2.14]. A set of vectors \{b_1, \ldots, b_N\} in (a vector space or subspace) \(\mathcal{V}\) is a basis for \(\mathcal{V}\) iff

- \{\(b_1, \ldots, b_N\)\} is a linearly independent set,
- \(\text{span}(\{b_1, \ldots, b_N\}) = \mathcal{V}\).

In general bases are not unique; nearly all vector spaces have multiple bases; the only exception is the trivial vector space \(\mathcal{V} = \{0\}\). In other words, generally a basis is not unique.

Example. If \{\(b_1, \ldots, b_N\)\} is a basis for \(\mathcal{V}\), then \{-\(b_1, \ldots, -b_N\}\} is also a basis for \(\mathcal{V}\).

Example. Euclidean space \(\mathbb{R}^N\) has many interesting bases used in signal processing, including those based on the DFT, the discrete cosine transform (DCT), and orthogonal wavelet transform (OWT), among others.

Fact. If \{\(b_1, \ldots, b_N\)\} is a basis for \(\mathcal{V}\), then every \(v \in \mathcal{V}\) has a unique representation of the form

\[ v = \sum_{n=1}^{N} \alpha_n b_n = B\alpha, \quad \alpha = [\alpha_1 \ldots \alpha_N]^\top, \quad B = [b_1 \ldots b_N]. \] (3.3)

See [1, p. 12]. The coefficients \(\{\alpha_n \in \mathbb{F}\}\) are the coordinates of \(v\) with respect to the basis \(B\).

A basis is a generalization of the usual concept of a coordinate system.
Sketch of proof of uniqueness of (3.3).
For any $v \in \mathcal{V}$, there exists some coordinates $(\alpha_1, \ldots, \alpha_N)$ by definition of \textit{span}.
To prove uniqueness, use contradiction. Suppose the representation were not unique, \textit{i.e.}, suppose there exists $(\beta_1, \ldots, \beta_N)$ such that $v = \sum_{n=1}^{N} \beta_n b_n$, where at least one $\beta_n$ differs from $\alpha_n$. Then $0 = v - v = \sum_{n=1}^{N} \beta_n b_n - \sum_{n=1}^{N} \alpha_n b_n = \sum_{n=1}^{N} (\beta_n - \alpha_n) b_n$. But because at least one coefficient in that sum is nonzero, that would imply that the set \{\(b_n\)\} is linearly dependent, contradicting the definition of a basis.
Example. The set of complex exponential signals \( \{ e^{i2\pi kt/T}, t \in \mathbb{R} : -K \leq k \leq K \} \) is a basis for all \( T \)-periodic signals that are band-limited with maximum frequency \( K/T \).

This fact about sinusoids (not proved here) is the foundation for additive synthesis musical sound generation.

The definition of basis above also generalizes to infinite dimensional spaces. Simply replace \( \{ b_1, \ldots, b_N \} \) with \( \{ b_1, b_2, \ldots \} \) and allow an infinite sum in (3.3). Dealing rigorously with infinite sums requires care ♦♦ beyond the scope of ECE 510.

Example. The set of monomials is a basis for the vector space of all polynomials.

The vector space of sinusoids of frequency \( \nu \) is: \( V = \{ A \cos(2\pi \nu t + \phi), t \in \mathbb{R} : A, \phi \in \mathbb{R} \} \).

Is \( S = \{ 3 \cos(2\pi \nu t), 5 \sin(2\pi \nu t), 7 \cos(2\pi \nu t - \pi/4) \} \) a basis for \( V \)?

A: Yes
B: No, because \( S \) has linear dependence
C: No, because \( S \) does not span \( V \)
D: Both B & C.

??
Define. The **dimension** of a subspace $S$ is the number of elements in any basis for $S$.

This definition is well defined because every subspace has a basis and even though that basis is not unique in general, every basis has the same number of elements (possibly infinite).

**Example.** $\dim(\mathbb{R}^N) = N$. Use the canonical or standard basis: $\{e_n : n = 1, \ldots, N\}$.

See [1, Ex. 2.20] for further examples.

**What is the dimension of the subspace of $N \times N$ diagonal matrices?**

A: 1  
B: $N$  
C: $2N$  
D: $N^2$  
E: $\infty$  

Define. There are two types of subspaces (and vector spaces).

- A finite-dimensional (sub)space has a basis with $\dim \in \mathbb{N}$.
- Otherwise we call the (sub)space **infinite dimensional**

What is the dimension of the trivial vector space $\mathcal{V} = \{0\}$? $\text{span}(\emptyset) = 0$ and $\dim(\emptyset) = 0$.

**What is the dimension of the vector space of all polynomials?**

A: 0  
B: 1  
C: infinite  
D: undefined  

??
**Dimension of a span**

Fact. If \( S = \text{span}(\{ \mathbf{u}_1, \ldots, \mathbf{u}_N \}) \) then \( \dim(S) \leq N \).

Exercise. Prove using the definition of dimension and basis.

To further discuss dimension, we first need to discuss subspace sums.

---

**Sums and intersections of subspaces**

Define. If \( S, T \subseteq V \) then

- the **sum** of two subspaces is defined as
  \[
  S + T = \{ s + t : s \in S, t \in T \}
  \]

- the **intersection** of two subspaces is defined as
  \[
  S \cap T = \{ v \in V : v \in S, v \in T \}
  \]

Exercise. Verify that \( S + T \) and \( S \cap T \) are both subspaces of \( V \). (See [1, Theorem 2.22].)
Example. If $S$ is the subspace of upper Hessenberg matrices in $\mathbb{R}^{N \times N}$ and $T$ is the subspace of lower Hessenberg matrices in $\mathbb{R}^{N \times N}$ then $S + T = \mathbb{R}^{N \times N}$.

Considering the same $S$ and $T$, what is $S \cap T$?
A: $\emptyset$  B: diagonal matrices  C: tridiagonal matrices  D: $\mathbb{R}^{N \times N}$  E: None of these

Caution: $S + T$ is not the same as the union of subspaces $S \cup T$.

Example. Consider $\mathcal{V} = \mathbb{R}^2$ and $S = \text{span}(s)$, $s = (1, 1)$ and $T = \text{span}(t)$, $t = (1, -1)$. Then $S + T = \mathbb{R}^2$ because $\{s, t\}$ spans $\mathbb{R}^2$ whereas the subspace union is just two lines:

We may discuss union of subspaces [2–4] later in the course.

Is a union of subspaces always a subspace?
A: Yes  B: No  C: Neither
Direct sum of subspaces

Now we define a particularly useful version of subspace sum. L§2.4

Define. We write subspace \( U \) as a **direct sum**

\[
U = S \oplus T
\]

of subspaces \( S \) and \( T \), all in a common vector space \( V \), iff

- \( S \cap T = 0 \)
- \( S + T = U \).

In which case we say \( S \) and \( T \) are **complements** of each other in \( U \).

Example. \( V = \mathbb{R}^2 \) and \( S = \text{span}(s), \ s = (1, 1) \) and \( T = \text{span}(t), \ t = (1, -1) \).

Then \( V = S \oplus T \).

For \( s, t \neq 0 \) and \( t \neq \alpha s \) for all \( \alpha \), let \( S = \text{span}(s) \) and \( T = \text{span}(t) \). Is \( \text{span}\{s, t\} = S \oplus T \)?

A: Always
B: If and only if \( s \) and \( t \) are orthogonal
C: If and only if \( s \) and \( t \) are orthonormal
D: Never
E: None of the above.

??
Dimensions of sums of subspaces

Yet another definition that is crucial for understanding the SVD.

[1, Theorem 2.26] If \( U = S \oplus T \) then

- Every \( u \in U \) can be written uniquely in the form \( u = s + t \) for some \( s \in S \) and \( t \in T \).

To prove uniqueness, suppose \( u = s_1 + t_1 = s_2 + t_2 \).

Then \( s_1 - s_2 = t_2 - t_1 \). But \( S \cap T = 0 \implies s_1 - s_2 = 0 \) and \( t_2 - t_1 = 0 \implies \) the representation is unique.

This property gives us a tool to help quantify dimension.

Example. In the previous example, \( \dim(S) = \dim(T) = 1 \) and \( \dim(V) = 2 \).

More generally, if we have any two subspaces in a vector space \( V \), then [1, Thm. 2.27]:

\[
\dim(S + T) = \dim(S) + \dim(T) - \dim(S \cap T).
\]  

(3.4)

So the direct sum equation above is a special case of this equality.
(We use this fact later in a proof about low-rank decomposition.)
Orthogonal complement of a subspace

Define. For a subspace $S$ of a vector space $V$, the **orthogonal complement** of $S$ is the subset of vectors in $V$ that are orthogonal to every vector in $S$:

$$S^\perp = \{ v \in V : \langle s, v \rangle = v^t s = 0, \ \forall s \in S \}.$$

Key properties of orthogonal complements when $V$ is finite dimensional (like $\mathbb{R}^N$ or $\mathbb{C}^N$) [1, Theorem 3.11]:

1. $S^\perp$ is itself a subspace of $V$
2. $(S^\perp)^\perp = S$  \hspace{1cm} (3.5)
3. $S \oplus S^\perp = V$  \hspace{1cm} (3.6)
4. $\dim(S) + \dim(S^\perp) = \dim(V)$  \hspace{1cm} (3.7)

Example. For $V = \mathbb{R}^3$, if $S = \text{span}\{(1, 1, 0), (1, -1, 0)\}$ then $S^\perp = \text{span}\{(0, 0, 1)\}.$

In $\mathbb{R}^3$, if $S$ is a line through the origin, then what geometric shape is $S^\perp$?

A: empty set  
B: point  
C: line  
D: plane  
E: $\mathbb{R}^3$  
??
Linear transformations

Define. Let $V$ and $W$ be vector spaces on a common field $F$. A function $L : V \mapsto W$ is a linear transformation or linear map or linear transform \cite[Def. 3.1]{1} iff

$$L(\alpha u + \beta v) = \alpha L(u) + \beta L(v), \quad \forall \alpha, \beta \in F \text{ and } \forall u, v \in V.$$  

Example. Consider $V = \mathbb{R}^2$ and let $W$ denote the space of $T$-periodic functions on $\mathbb{R}$. Construct $L(\cdot)$ by $s = L([a \ b]^T) \iff s(t) = a \cos(2\pi t/T) + b \sin(2\pi 5t/T + \pi/4)$.

Exercise. Verify that $L(\cdot)$ is a linear transformation.

Matrix-vector multiplication as a linear transformation

Example. Let $V = \mathbb{C}^N$ and $W = \mathbb{C}^M$ and $A \in \mathbb{C}^{M \times N}$. Consider the transformation defined by $y = L(x) \iff y = Ax$, i.e., $x \mapsto y = Ax$.

This transformation is linear.

Proof. $L(\alpha x + \beta z) = A(\alpha x + \beta z) = \alpha Ax + \beta Az = \alpha L(x) + \beta L(z)$, where $\alpha, \beta \in \mathbb{C}$ are arbitrary as are $x, z \in \mathbb{C}^N$.

For more examples of linear transformations, see \cite[Ex. 3.2]{1}.
**Range of a matrix**

Define. The **range** of a matrix $A = [a_1 \ldots a_N]$, also known as its **column space**, is the span of its columns:

$$\text{span}\{a_1, \ldots, a_N\}.$$

Equivalently:

$$\mathcal{R}(A) = \{Ax : x \in \mathbb{F}^N\}.$$

The range of a matrix in $\mathbb{F}^{M \times N}$ is a subspace of $\mathbb{F}^M$.

Define. The **row space** of a matrix $A$ is the span of its rows: $\mathcal{R}(A')$.

If $D$ is a diagonal matrix in $\mathbb{R}^{N \times N}$, what is $\mathcal{R}(D)$?

A: $\emptyset$  B: Usually $\mathbb{R}^N$.  C: Always $\mathbb{R}^N$.  D: Usually $\mathbb{R}^{N \times N}$.  E: Always $\mathbb{R}^{N \times N}$.
3.2 Rank of a matrix

The preceding material about subspaces applied to vectors in general vector spaces. Now we specialize to a concept that is specific to matrices: the rank of a matrix.

Define. For any $M \times N$ matrix $A$:

- **column rank** of $A \triangleq \dim(\mathcal{R}(A)) = \# \text{ of linearly independent columns of } A$
- **row rank** of $A \triangleq \dim(\mathcal{R}(A')) = \# \text{ of linearly independent rows of } A$

Theorem. For any $M \times N$ matrix $A$, its **row rank** = its **column rank**.

Proof. Let $r$ denote the column rank of $A = [a_1 \ldots a_N]$. Because $r$ is the column rank, there exists a basis $V = [v_1 \ldots v_r]$ such that one can write every vector in $\mathcal{R}(A)$ as a linear combination of the columns of $V$. Each column of $A$ is in $\mathcal{R}(A)$, thus we can express each column of $A$ as a linear combination of the columns of $V$, i.e.,

$$a_1 = c_{11}v_1 + \cdots + c_{r1}v_r$$

$$\vdots$$

$$a_N = c_{11}v_1 + \cdots + c_{r1}v_r,$$

where the $c_{ij}$ values denote the coordinates or coefficients w.r.t. the basis $V$. 
In matrix form we get a sum-of-outer-products form of matrix multiplication:

\[
\begin{bmatrix}
A
\end{bmatrix}_{M \times N} = \begin{bmatrix} v_1 & \ldots & v_r \end{bmatrix} \begin{bmatrix} c_{11} & \ldots & c_{1N} \\
\vdots & & \vdots \\
c_{r1} & \ldots & c_{rN} \end{bmatrix} = \begin{bmatrix} v_1 & \ldots & v_r \end{bmatrix} \begin{bmatrix} -c_1^\top & - \\
\vdots & & \vdots \\
-c_r^\top & - \end{bmatrix} = VC = \sum_{k=1}^{r} v_k c_k^\top.
\]

Now the \( m \)th row of \( A \) is a linear combination of the rows of \( C \):

\[
A_{[m,:]} = e_m' A = \sum_{k=1}^{r} (e_m' v_k) c_k^\top = \sum_{k=1}^{r} v_{mk} c_k^\top.
\]

This construction holds for any row of \( A \) (or linear combinations thereof)

\[
\implies \mathcal{R}(A^\prime) = \text{row space of } A \subseteq \mathcal{R}(C^\top)
\]

\[
\implies \text{row rank of } A \leq r = \text{column rank of } A.
\]

Because \( A \) was arbitrary, the same argument applies to its transpose, so:

\[
\text{row rank of } A^\prime \leq \text{column rank of } A^\prime
\]

\[
\implies \text{column rank of } A \leq \text{row rank of } A
\]

\[
\implies \text{column rank of } A = \text{row rank of } A.
\]

For an alternate proof, see [1, Theorem. 3.17].
Rank definition summary

Because the row rank and column rank of a matrix are always identical, we generally simply speak of the rank of a matrix without the “row” or “column” qualifier. And it suffices to define:

\[ \text{rank}(A) \triangleq \text{dim} (\mathcal{R}(A)). \]

Corollary:

\[ A \in \mathbb{R}^{M \times N} \implies \]

\[ \text{rank}(A) \leq \min(M, N) \]  \hfill (3.8)

Proof. \( \text{rank}(A) = \text{row rank of } A \leq \# \text{ rows} = M \)

\( \text{rank}(A) = \text{col rank of } A \leq \# \text{ cols} = N \)

\( \implies \text{rank}(A) \leq \min(M, N) \)

\[ \square \]
Rank of a matrix product

Theorem. Multiplying matrices never increases rank:

\[ \text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B)) \]  \hspace{1cm} (3.9)

Proof.

\[ AB = \begin{bmatrix} a_1 & \cdots & a_N \end{bmatrix} \begin{bmatrix} - & b_1^T & - \\ \vdots \\ - & b_N^T & - \end{bmatrix} = a_1 b_1^T + \cdots + a_N b_N^T \]

\[ \implies \text{every column of } AB \text{ is a linear combination of columns of } A \]
\[ \implies \mathcal{R}(AB) \subseteq \mathcal{R}(A) \]
\[ \implies \text{rank}(AB) \leq \text{rank}(A) \]

Similarly, because \((AB)' = B'A'\), applying the same logic we have
\[ \text{rank}(AB) = \text{rank}((AB)') = \text{rank}(B'A') \leq \text{rank}(B') = \text{rank}(B) . \]

Combining both inequalities yields (3.9).

\[ \square \]

\( AB \) is a composition of two linear transforms.
\[ \implies \text{Composition cannot enlarge subspace dimension.} \]

Caution: in general: \( \text{rank}(AB) \neq \text{rank}(BA) \).
Example. DSP analogy: cascade of two filters with band-limited frequency responses.

\[ H(\nu) = H_1(\nu)H_2(\nu) \]

where \([\nu_3, \nu_3] = [\nu_1, \nu_1] \cap [\nu_2, \nu_2]\). Composing filters cannot recover lost frequencies.

For \(u \in \mathbb{C}^M\) and \(v \in \mathbb{C}^N\), what is the minimum and maximum possible rank of \(uv'\)?

A: 0,1  B: 1,1  C: 0,\min(M,N)  D: 1,\min(M,N)  E: None of these

Other properties of rank

There is also a lower bound for the rank of a product, called **Sylvester’s rank inequality**: if \(A\) is an \(m \times n\) matrix and \(B\) is \(n \times k\), then

\[ \text{rank}(A) + \text{rank}(B) - n \leq \text{rank}(AB). \]

Furthermore, rank is subadditive:

\[ \text{rank}(A + B) \leq \text{rank}(A) + \text{rank}(B). \]
Unitary invariance of rank

Theorem. If \( A \in \mathbb{F}^{M \times N} \) then multiplying by a unitary matrix on the left or right does not change the rank:

- \( Q \in \mathbb{F}^{M \times M} \) and \( Q \) unitary \( \implies \) rank \((QA)\) = rank \((A)\)
- \( Q \in \mathbb{F}^{N \times N} \) and \( Q \) unitary \( \implies \) rank \((AQ)\) = rank \((A)\)

Rank and eigenvalues

Corollary. If \( A \) is Hermitian with eigendecomposition \( A = V \Lambda V' \) then because \( V \) is unitary:

\[
\text{rank}(A) = \text{rank}(V \Lambda V') = \text{rank}(\Lambda) = \text{number of nonzero eigenvalues of } A.
\]
3.3 Nullspace and the SVD

To fully understand an SVD of a matrix, we need both range and nullspace (and orthogonal complements thereof).

**Nullspace or kernel**

The set of vectors that yield zero when multiplied by a matrix is often important.

Define. The null space or kernel of a $M \times N$ matrix $A$ is

$$ \mathcal{N}(A) = \ker(A) \triangleq \{ x \in \mathbb{F}^N : Ax = 0 \}.$$  

Clearly we always have $0 \in \mathcal{N}(A)$.

Exercise. Verify that $\mathcal{N}(A)$ is indeed a subspace.

Thus using the subspace font “$\mathcal{N}$” is appropriate.

If $A \in \mathbb{C}^{M \times N}$, then $\mathcal{N}(A)$ is a subspace of what vector space?

- A: $\mathbb{C}^M$
- B: $\mathbb{C}^N$
- C: $\mathbb{R}^N$
- D: $\mathbb{R}^M$
- E: None of these.
**Decomposition theorem for matrices**

If $A \in \mathbb{F}^{M \times N}$ then the input and output spaces of $A$ satisfy [1, Theorem 3.14]:

$$\begin{align*}
N(A) \oplus N(\perp A) &= \mathbb{F}^N \\
R(A) \oplus R(\perp A) &= \mathbb{F}^M.
\end{align*}$$

In other words, every “input” vector $x \in \mathbb{F}^N$ can be decomposed uniquely as $x = x_0 + x_1$, where $x_0 \in N(A)$ so $Ax_0 = 0$ and $x_1 \in N(\perp A)$.

Likewise, every vector $y \in \mathbb{F}^M$ can be decomposed uniquely as $y = y_1 + y_0$, where $y_1 \in R(A)$, so $Ax_1 = y_1$ for some $x_1 \in \mathbb{F}^N$, and $y_0 \in R(\perp A)$.

The above statement was for an arbitrary $y \in \mathbb{F}^M$.

If we have an “output vector” $y = Ax$, then $y \in R(A)$ by definition and $y_0 = 0$. 
Relationships between null space and range of a matrix

[1, Theorem 3.12]. For any matrix $A$, its null space and range are related by:

$$\mathcal{N}^\perp(A) = \mathcal{R}(A')$$
$$\mathcal{R}^\perp(A) = \mathcal{N}(A')$$

Proof: If $x \in \mathcal{N}(A)$ then $Ax = 0$, so $\forall y$ we have $y'(Ax) = 0 \implies x'(A'y) = 0 \implies x \in \mathcal{R}^\perp(A')$.

Conversely, if $x \in \mathcal{R}^\perp(A')$ then $\forall y$, $0 = x'(A'y)$.

Take $y = Ax$ and we have $\|Ax\| = 0$ which implies $Ax = 0$, so $x \in \mathcal{N}(A)$.

Thus using (3.5): $\mathcal{N}(A) = \mathcal{R}^\perp(A') \implies \mathcal{N}^\perp(A) = \mathcal{R}(A')$.

The proof of the second equality is left to HW (cf. [1, Problem 3.6]).

Corollary:

$$\dim(\mathcal{N}^\perp(A)) = \dim(\mathcal{R}(A')) = \dim(\mathcal{R}(A)) = \text{rank}(A)$$
$$\dim(\mathcal{R}^\perp(A)) = \dim(\mathcal{N}(A')).$$
Nullity

Define. The nullity of a matrix is the dimension of its null space:

$$\text{nullity}(A) \triangleq \dim(N(A))$$

Rank plus nullity property \[1, \text{Corollary 3.18}\]. If $A \in \mathbb{F}^{M \times N}$ then

$$\dim(N(A)) + \dim(N^\perp(A)) = \dim(N(A)) + \text{rank}(A) = N.$$ 

Proof. Because $\mathbb{F}^N = N(A) \oplus N^\perp(A)$ we have from (3.6), (3.7) and (3.10):

$$N = \dim(N(A)) + \dim(N^\perp(A)) = \dim(N(A)) + \text{rank}(A).$$

If $u \in \mathbb{C}^M - \{0\}$ and $v \in \mathbb{C}^N - \{0\}$, what is nullity($uv'$)?

A: 0
B: 1
C: $N - 1$
D: $N$
E: $M$

??
The four fundamental spaces related to a matrix $A$

Now we unify the null space and range space (and their orthogonal complements) for a general matrix $A$. L§3.5

The following diagram summarizes the fundamental theorem of linear algebra for a matrix $A \in \mathbb{F}^{M \times N}$ where we think of $A$ as a mapping from $\mathbb{F}^N$ to $\mathbb{F}^M$. See [1, Section 3.5] and [1, Theorem 5.1].

```
<table>
<thead>
<tr>
<th>dim $\mathbb{F}^N$</th>
<th>“input”</th>
<th>“output”</th>
<th>dim $\mathbb{F}^M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r$</td>
<td>$\mathcal{N}^\perp(A) = \mathcal{R}(V_r)$</td>
<td>$A$</td>
<td>$\mathcal{R}(A) = \mathcal{R}(U_r)$</td>
</tr>
<tr>
<td></td>
<td>$\oplus {0}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N - r$</td>
<td>$\mathcal{N}(A) = \mathcal{R}(V_0)$</td>
<td></td>
<td>$\mathcal{R}^\perp(A) = \mathcal{R}(U_0)$</td>
</tr>
</tbody>
</table>
```
**A** has rank \( r \leq \min(M, N) \), so we can partition the SVD components as follows:

\[
A = U \Sigma V' =
\]

where \( \Sigma_r \) is \( r \times r \) and contains the *nonzero* singular values of \( A \) along its diagonal.

What is the size of the lower right 0 above?

A: \( M \times N \)  
B: \( M \times (N - r) \)  
C: \( (M - r) \times N \)  
D: \( (M - r) \times (N - r) \)  
E: None of these

??

**Relationship of rank with SVD**

Recall SVD expression for a matrix \( A \) having rank \( r \) where \( r \leq \min(M, N) \):

\[
A \in \mathbb{C}^{M \times N} \implies A = U \Sigma V' = \sum_{k=1}^{\min(M,N)} \sigma_k u_k v_k' = \sum_{k=1}^r \sigma_k u_k v_k'.
\]

When we write \( \sum_{k=1}^{\min(M,N)} \) then some of the \( \sigma_k \) values may be zero, namely \( \sigma_{r+1}, \ldots, \sigma_{\min(M,N)} \).

When we write \( \sum_{k=1}^r \) where \( r \) is the rank of \( A \), then all the \( \sigma_k \) values used in the sum are nonzero.

By the unitary invariance of rank, \( \text{rank}(A) = \text{rank}(\Sigma) = \text{number of nonzero singular values} \).
Anatomy of the SVD

For $A \in \mathbb{F}^{M \times N}$ with rank $r$, in general we can write:

$$A = U \Sigma V' = \sum_{k=1}^{\min(M,N)} \sigma_k u_k v_k' = \sum_{k=1}^{r} \sigma_k u_k v_k'$$

where $\Sigma$ is $r \times r$ and contains the \textit{nonzero} singular values of $A$.

There are two main cases worth considering in more detail.

These cases are somewhat (but not exactly) related to the \texttt{svd(A,'econ')} in MATLAB.

In class, when we speak of the \textit{thin SVD} we mean $A = U_r \Sigma_r V_r'$ as described in these pages.
A is “tall” or “thin,” i.e., $M > N \implies r \leq N < M$:

\[
\begin{align*}
A &= U \Sigma V' = \\
&= \begin{bmatrix} U_r & U_0 \end{bmatrix} \\
&\begin{bmatrix} \Sigma_r \\ 0 \end{bmatrix} V'_r
\end{align*}
\]

(3.11)

What is the size of the lower $0$ above?

A: $M \times N$  B: $M \times (N - r)$  C: $(M - r) \times N$  D: $(M - r) \times (N - r)$  E: None of these

Caution. When we write $U \Sigma V' = U_r \Sigma_r V'_r$ we do not mean that the individual terms match (they do not!), we mean that the overall product matches.

If $r = N$ then $\mathcal{N}(A) = 0$ and there is no $V_0$. 
**A** is wide, *i.e.*, $M < N \implies r \leq M < N$:

$$
\begin{bmatrix}
A
\end{bmatrix}_{M \times N} = U \Sigma V' = 
\begin{bmatrix}
U_r & \Sigma_r & 0
\end{bmatrix}
\begin{bmatrix}
V'_r \\
V'_0
\end{bmatrix}
\begin{bmatrix}
\Sigma_r \\
V'_r
\end{bmatrix}_{r \times N}^T
\begin{bmatrix}
\Sigma_r \\
V'_r
\end{bmatrix}_{r \times N}^T_{N \times N}
\begin{bmatrix}
r \times N
(N - r) \times N
\end{bmatrix}
$$

If $r = M$ then $\dim(\mathcal{R}(A)) = M$ and all of $\mathbb{F}^M$ is reachable.

However, note that $\dim(\mathcal{N}(A)) = N - r > M - r \geq 0$, so a wide $A$ has a nontrivial null space.

Note the symmetry between the above two "compact" SVD representations, as there must be because if $A$ is tall then $A'$ is wide.
SVD of finite differences (discrete derivative)

Example. It is well known that the derivative of \( f(t) = \sin(\omega t) \) is \( \dot{f}(t) = \omega \cos(\omega t) \) and the second derivative is \( \ddot{f}(t) = -\omega^2 \sin(\omega t) \). In other words, think of an analog system whose inputs are differentiable signals, and whose output is the second derivative of the input:

\[
\begin{align*}
  f(t) \rightarrow \text{second derivative} \rightarrow \ddot{f}(t).
\end{align*}
\]

The eigenfunctions of this system include any signal of the form \( A \cos(\omega t + \phi) \) or \( A \sin(\omega t + \phi) \) or \( A e^{i\omega t + \phi} \), all of which have eigenvalue \(-\omega^2\). We now explore the matrix-vector analog of this.

Consider the \((N + 1) \times N\) matrix \( C \) that performs a finite difference operation:

\[
C \triangleq \begin{bmatrix}
1 & 0 & 0 & 0 & \ldots & 0 \\
0 & -1 & 1 & 0 & \ldots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \ldots & 0 & 0 & -1 & 1 \\
0 & \ldots & 0 & 0 & 0 & 1 \\
\end{bmatrix}, \quad \text{so} \quad Cx = \begin{bmatrix}
x_1 \\
x_2 - x_1 \\
\vdots \\
x_N - x_{N-1} \\
x_N \\
\end{bmatrix}.
\]

This important matrix arises in many signal and image processing applications and in every field of engineering that uses a finite-element method (FEM) based on finite differences to approximate differential equations arising from physics.
One can show using trigonometric identities that the SVD of $C$ involves the discrete cosine transform (DCT) and discrete sine transform (DST) as follows [5]:

$$C = U\Sigma V' = \text{DCT} \Sigma \text{DST}' .$$

So for this important matrix, the left and right singular vectors turn out to form matrices that are themselves important in signal processing. The fact that $C'C = \text{DST} \Sigma^2 \text{DST}'$ is analogous to the fact that the second derivative of a sinusoid is a sinusoid.
Synthesis view of matrix decomposition

- Eigen-decomposition of square matrix (when it exists, e.g., \(A\) Hermitian):

\[
A = \mathbf{Q} \Lambda \mathbf{Q}' = \sum_i \lambda_i \mathbf{q}_i \mathbf{q}_i'
\]

with \(\mathbf{Q}'\mathbf{Q} = \mathbf{I}\)

- SVD of a \(M \times N\) matrix:

\[
A = \mathbf{U} \Sigma \mathbf{V}' = \sum_{k=1}^r \sigma_k \mathbf{u}_k \mathbf{v}_k'
\]

where \(\sigma_k > 0\) for \(k = 1, \ldots, r = \text{rank}\), and \(\mathbf{U}'\mathbf{U} = \mathbf{I}_M\) and \(\mathbf{V}'\mathbf{V} = \mathbf{I}_N\).

In both cases we can “synthesize” the matrix using a sum of rank-1, outer-product terms.
3.4 Orthogonal bases

In signal processing and beyond, of the many bases for a vector space \( V \), the ones that have orthogonal basis vectors are particularly important.

Define. A collection of vectors \( \{b_1, b_2, \ldots\} \) in a vector space \( V \) is called a **orthogonal basis** for \( V \) iff

- \( \{b_1, b_2, \ldots\} \) is a **basis** for \( V \), i.e.,
  - \( \{b_1, b_2, \ldots\} \) is **linearly independent**
  - \( \text{span}(\{b_1, b_2, \ldots\}) = V \)
- The basis vectors are **orthogonal**, i.e., \( \langle b_n, b_m \rangle = 0 \) for \( n \neq m \).

The following theorem shows that the above definition *almost* has a redundancy. (read)

**Theorem (orthogonality implies linear independence for nonzero vectors).** If \( \{v_1, v_2, \ldots\} \) are nonzero orthogonal vectors, then they are also linearly independent.

**Proof (by contradiction).** Suppose \( \{v_1, v_2, \ldots\} \) are orthogonal, nonzero, and also linearly dependent. Then there exists \( N \in \mathbb{N} \) and \( c_1, \ldots, c_N \in \mathbb{F} \), not all equal to zero, such that \( 0 = \sum_{n=1}^{N} v_n c_n \). Pick any \( m \in \{1, \ldots, N\} \) and we have \( 0 = v'_m 0 = v'_m \sum_{n=1}^{N} v_n c_n = v'_m v_m c_m = \|v_m\|_2^2 c_m \Rightarrow c_m = 0 \) because \( v_m \) is nonzero. Those holds for every \( m \), contradicting the assumption that not all \( c_n \) are zero. \( \square \)

Thus the linear independence condition in the definition above is implied by the orthogonality condition, at least for nonzero vectors.
The above definition is general enough to accommodate infinite-dimensional vector spaces. In finite dimensions the situation is simpler:

Any orthogonal matrix $V \in \mathbb{R}^{N \times N}$ is a basis for $\mathbb{R}^N$.

Any unitary matrix $V \in \mathbb{C}^{N \times N}$ is a basis for $\mathbb{C}^N$.

Proof.
- The orthogonality condition ensures linear independence by the above theorem.
- For any vector $x \in \mathbb{F}^N$ we have $x = Ix = (VV')x = \sum (V'x) \in \text{span}(V) \implies \text{span}(V) = \mathbb{F}^N$.

Finding coordinates in an orthogonal basis

For $x \in \mathbb{F}^N$, its elements are coordinates in the standard basis aka canonical basis:

$x = Ix = \sum_{n=1}^{N} e_n x_n = \begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 1 & 0 \\
0 & 0 & \cdots & 0 & 1
\end{bmatrix}\begin{bmatrix}
x_1 \\
x_2 \\
\vdots \\
x_{N-1} \\
x_N
\end{bmatrix}.$

This is the most trivial orthogonal basis.
Now suppose $Q \in \mathbb{F}^{N \times N}$ is an orthogonal matrix (or unitary matrix) and hence a basis for $\mathbb{F}^N$.

By definition of an orthogonal matrix (or unitary matrix): $QQ' = I$, hence

$$x = Ix = (QQ')x = Q (Q'x) = Q \alpha,$$

where $\alpha = Q'x \in \mathbb{F}^N$.

Alternate perspective: to write $x$ in the basis $Q$, we want $x = Q\alpha$, so we need the coordinate vector to be $\alpha = Q^{-1}x = Q'x$.

It is important to appreciate the convenience of an orthogonal basis for finding coefficients (coordinates).

- For a basis in general we need $\alpha = Q^{-1}x$, requiring matrix inversion that needs $O(N^3)$ computation.
- For an orthogonal basis, we need $\alpha = Q'x$, which is simply matrix multiplication that needs $O(N^2)$ computation in general. For some bases (like DFT, DCT, OWT) it can be just $O(N \log N)$.

How do we know that the inverse $Q^{-1}$ exists for a general basis?
For some matrices, one can find eigenvectors “by inspection.” Recognizing these is a useful skill.

Example. Consider the symmetric outer product: \( A = zz' \).

Observe that

\[
Az = (zz')z = z(z'z) = (z'z)z.\]

Thus \( z \) is an eigenvector of \( A \) with eigenvalue that is the norm squared of \( z \):

\[
\lambda = z'z = \|z\|^2.
\]

This symmetric matrix has rank 1 and it has one nonzero eigenvalue.
Example. Consider the $N \times N$ matrix with $N = 2n$ where

$$A = \begin{bmatrix}
1 & \cdots & 1 \\
\vdots & \ddots & \vdots \\
1 & \cdots & 1 \\
0 & \cdots & 0 \\
9 & \cdots & 9 \\
0 & \cdots & 0 \\
9 & \cdots & 9
\end{bmatrix} = \begin{bmatrix}
1_n \\
0_n \\
0_n \\
0_n \\
31_n
\end{bmatrix} = \begin{bmatrix}
1_n' \\
0_n' \\
0_n' \\
0_n' \\
31_n'
\end{bmatrix} + \begin{bmatrix}
1_n \\
0_n \\
0_n \\
0_n \\
31_n
\end{bmatrix}.$$ 

By inspection one can build on the previous example to see that two of the eigenvectors are:

$$u_1 = \begin{bmatrix}
\frac{1}{\sqrt{n}}1_n \\
0_n
\end{bmatrix}, \quad u_2 = \begin{bmatrix}
0_n \\
\frac{1}{\sqrt{n}}1_n
\end{bmatrix}.$$ 

What is the corresponding eigenvalue $\lambda_1$?

A: 1  B: $\sqrt{n}$  C: $n$  D: $1/n$  E: $1/\sqrt{n}$

The second non-zero eigenvalue is 9 times larger. All the other eigenvalues are zero.

This symmetric matrix has rank 2 and it has two nonzero eigenvalues.

As noted previously, in general for symmetric matrices, rank $=$ number of (possibly non-distinct) nonzero eigenvalues.
Example. What about asymmetric matrices? \( A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = e_1 e'_2 \) has \( \lambda_1 = \lambda_2 = 0 \) but \( r = 1 \).

**SVD by inspection**

For some matrices, one can find an SVD by inspection.

Example. Consider the \( M \times N \) rank-1 outer-product matrix \( A = bc' \) with \( b \neq 0_M \) and \( c \neq 0_N \). Clearly:

\[
A = bc' = \begin{bmatrix} \|b\| \|c\| & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \|b\| & 0 \\ 0 & \|c\| \end{bmatrix}
\]

where \( U_0 \) is a \( M \times (M - 1) \) matrix with orthonormal columns that span the subspace \( \text{span}^\perp(b) \) and \( V_0 \) is a \( N \times (N - 1) \) matrix with orthonormal columns that span the subspace \( \text{span}^\perp(c) \).

Is it unique? No, because we could use \(-b\) and \(-c\), and there are many choices for \( U_0 \) and \( V_0 \).
Matrix-vector products and the SVD

\[ Ax = U \Sigma V' x \]
\[ = U \Sigma (V' x) \]
\[ \mapsto \text{coordinates of } x \text{ in term of basis } V \]

Define the coordinates (coefficients) to be \( \alpha = V' x \). Then expanding the matrix product:

\[ Ax = \sum_{k=1}^{r} \sigma_k \alpha_k u_k \]
\[ \mapsto \text{left singular vector (basis vector)} \]

In particular, if \( A \in \mathbb{R}^{M \times N} \) and \( x = v_n \), the \( n \)th right singular vector, for some \( n \in \{1, \ldots, N\} \), then

\[ \alpha = V' x = e_n \implies Ax = Av_n = \sigma_n u_n. \]
Summary

This chapter has a lot of general concepts in it (subspace, basis, dimension, null space, rank).
All of the definitions lead to the key result: “four fundamental subspaces” portion that relates the null space
and range of a matrix to components of its SVD like $V_0$ and $U_r$.

For a futuristic (?) demonstration of the importance of subspaces, see:
https://www.youtube.com/watch?v=H4qkdI6rSM

Bibliography


