Chapter 3

Subspaces and rank

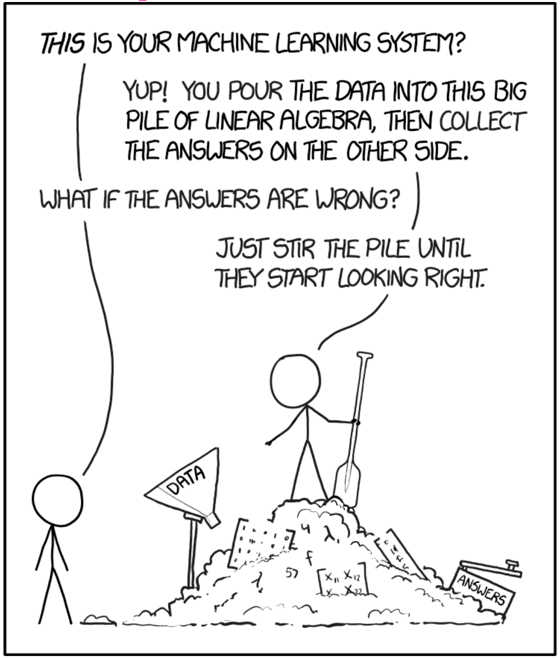
Contents (class version)

3.0 Introduction

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 (and beyond).

Source material for this chapter includes [1, §2.2-2.4, 3.4, 3.1, 3.5, 5.1].

From https://xkcd.com/1838



3.1 Subspaces

L§2.2

Define [1, Def. 2.6]. For a **vector space** $\mathcal{V}$ defined on a field $\mathbb{F}$, a nonempty subset $\mathcal{S} \subseteq \mathcal{V}$ is called a **subspace** or **linear subspace** of $\mathcal{V}$ iff

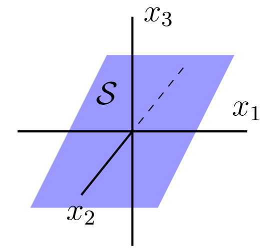
- $\mathcal{S}$ is closed under vector addition:
- $\mathcal{S}$ is closed under scalar multiplication:

Fact. A subspace $\mathcal{S}$ always includes the zero vector $\mathbf{0}$.

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

Laub [1, p. 9] uses the notation $\mathcal{S} \subseteq \mathcal{V}$ to indicate that $\mathcal{S}$ 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 $\mathbf{0}$. (But keep in mind that $\{\mathbf{0}\}$ and $\mathcal{V}$ are also subspaces.)



Example. $\mathcal{S} = \{\mathbf{0}\}$ where $\mathbf{0} \in \mathcal{V}$ for some vector space $\mathcal{V}$.
This is the most minimalist and uninteresting subspace.

Example. $\mathcal{S} = \{\alpha \mathbf{1}_N \,:\, \alpha \in \mathbb{C}\} \subseteq \mathbb{C}^N$. We will see shortly that here $\mathcal{S} = \mathsf{span}(\mathbf{1}_N)$.

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

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

$\boxed{\textbf{??}}$

Example. The set of vectors orthogonal to some vector $\boldsymbol{v} \in \mathcal{V}$, *i.e.*, $\mathcal{S} = \{\boldsymbol{x} \in \mathcal{V} \,:\, \langle \boldsymbol{x}, \boldsymbol{v} \rangle = 0\}$.
This set is a subspace because it is closed under vector addition and scalar multiplication::

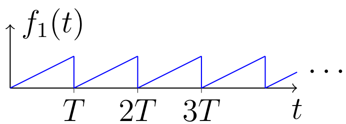
- $\boldsymbol{x}, \boldsymbol{z} \in \mathcal{S} \implies \langle \boldsymbol{x} + \boldsymbol{z}, \boldsymbol{v} \rangle = \langle \boldsymbol{x}, \boldsymbol{v} \rangle + \langle \boldsymbol{z}, \boldsymbol{v} \rangle = 0 + 0 = 0 \implies \boldsymbol{x} + \boldsymbol{z} \in \mathcal{S}$
- $\boldsymbol{x} \in \mathcal{S} \implies \langle \alpha \boldsymbol{x}, \boldsymbol{v} \rangle = \alpha \langle \boldsymbol{x}, \boldsymbol{v} \rangle = \alpha 0 = 0 \implies \alpha \boldsymbol{x} \in \mathcal{S}, \ \forall \alpha \in \mathbb{F}$
  Later in the chapter we will see that $\mathcal{S} = (\mathsf{span}(\boldsymbol{v}))^{\perp}$.

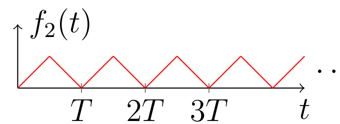
Periodic functions as a subspace (A signal processing ex (Read)

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.

The vector space $\mathcal{V}$ above is **infinite dimensional** because $\forall k \in \mathbb{Z}, \ \mathrm{e}^{\imath 2\pi k t/T} \in \mathcal{V}$, and that set is **linearly independent**. In fact it is a basis, and that fact is the foundation of **Fourier series**. ♦♦

Span

L§2.3

Define. Given a set of vectors $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N\}$ in a vector space $\mathcal{V}$ over field $\mathbb{F}$, the **span** of those vectors is

$$\mathsf{span}(\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N\}) \triangleq \qquad (3.1)$$

This set is also called the **linear span** or **hull**. (Caution: it differs from the **convex hull** of a set.)

Often we will collect the vectors into a matrix $\boldsymbol{U} = [\boldsymbol{u}_1 \ \ldots \ \boldsymbol{u}_N]$ and define $\mathsf{span}(\boldsymbol{U}) \triangleq \mathsf{span}(\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N\})$, although the usual mathematical definition is that the argument of $\mathsf{span}(\cdot)$ is a set of vectors, not a matrix.

Exercise. Verify that $\mathsf{span}(\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N\})$ is a **subspace** of the vector space $\mathcal{V}$.

Example. For $\mathcal{V} = \mathbb{R}^3$, if $\boldsymbol{u}_1 = (1, 0, 1)$ and $\boldsymbol{u}_2 = (1, 0, -1)$ then $\mathsf{span}(\{\boldsymbol{u}_1, \boldsymbol{u}_2\})$ is the entire (x, z) plane.

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

$$\begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_N \end{bmatrix} = d_1 \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{bmatrix} + \cdots + d_N \begin{bmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & 1 \end{bmatrix} =$$

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 $\mathcal{S}$ is a (possibly uncountably infinite) subset of a vector space $\mathcal{V}$ over field $\mathbb{F}$, the span of $\mathcal{S}$ consists of *every* (finite by definition) **linear combination** of elements in $\mathcal{S}$:

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

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

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

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

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

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

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

Example. Consider the vector space $\mathcal{V}$ of all polynomials. Now consider the (**countably infinite**) set of monomials with even powers: $\mathcal{S} = \left\{x^k \ : \ k = 0, 2, 4, \ldots\right\}$. The span of this set $\mathcal{S}$ is the subspace of polynomials having terms with even powers, *i.e.*, the subspace of even polynomials where $p(-x) = p(x)$.

Linear independence

Define. A set of vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N$ 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

$$\sum_{n=1}^{N} \alpha_n \boldsymbol{u}_n = \boldsymbol{0}.$$

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 $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N$ in a vector space is **linearly independent** iff for any scalars $\alpha_1, \ldots, \alpha_N \in \mathbb{F}$:

$$\sum_{n=1}^{N} \alpha_n \boldsymbol{u}_n = \boldsymbol{0} \Longrightarrow$$

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 $\{\boldsymbol{e}_1 \boldsymbol{e}_1', \ldots, \boldsymbol{e}_N \boldsymbol{e}_N'\}$ is linearly independent.

Any set of **orthonormal** vectors is a **linearly independent** set.

Any set of *nonzero* **orthogonal** vectors is a **linearly independent** set.

Example. The collection of vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3\}$ where = $\boldsymbol{v}_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\boldsymbol{v}_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is an **orthogonal** set, but the set is **linearly dependent** because $7\boldsymbol{v}_1 + 0\boldsymbol{v}_2 + 0\boldsymbol{v}_3 = \boldsymbol{0}$.

Generalization to infinite sets (Read) ♦♦

Recall that $\mathcal{S} - \{\boldsymbol{x}\}$ denotes the set $\mathcal{S}$ with the vector $\boldsymbol{x}$ removed.

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

$$\exists \boldsymbol{x} \in \mathcal{S} \text{ s.t. } \boldsymbol{x} \in \mathsf{span}(\mathcal{S} - \{\boldsymbol{x}\}),$$

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

Example. Consider the vector space of all polynomials. Let $\mathcal{S}$ denote the (countably infinite) set of all **monomials**, *i.e.*, $\mathcal{S} = \{f(x) = x^n \;:\; n = 0, 1, \ldots\}$. One can show by elementary algebra [wiki] that $\mathcal{S}$ is linearly independent. (One cannot write a 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. L§2.3

Define [1, Def. 2.14]. A set of vectors $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots\}$ in (a vector space or subspace) $\mathcal{V}$ is a **basis** for $\mathcal{V}$ iff

-
-

The vectors in a basis are called **basis vectors**. In general bases are not unique; nearly all vector spaces have multiple bases; the only exception is the trivial vector space $\mathcal{V} = \{\mathbf{0}\}$. Thus, generally a basis is *not unique*.

Example. The **standard basis** for $\mathbb{F}^N$ is the set of N unit vectors $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_N\}$ because the set is linearly independent and if $\boldsymbol{x} \in \mathbb{F}^N$ we can write $\boldsymbol{x} = \sum_{n=1}^N x_n \boldsymbol{e}_n$.

Example. The **standard basis** for $\mathbb{F}^{2\times 2}$ is the following set of four 2×2 matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

because the set is linearly independent we can express any 2×2 matrix as a linear combination of these four matrices.

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.

Bases are not unique in general.

Example. If $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots\}$ is a basis for $\mathcal{V}$, then $\{-\boldsymbol{b}_1, -\boldsymbol{b}_2, \ldots\}$ is also a basis for $\mathcal{V}$,

---

Fact. If $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N\}$ is a basis for $\mathcal{V}$ for $N \in \mathbb{N}$, then every $\boldsymbol{v} \in \mathcal{V}$ has a *unique* representation of the form

$$\boldsymbol{v} = \sum_{n=1}^{N} \alpha_n \boldsymbol{b}_n = \boldsymbol{B}\boldsymbol{\alpha}, \qquad \begin{aligned} \boldsymbol{\alpha} &= [\alpha_1 \ \ldots \ \alpha_N]^\top \\ \boldsymbol{B} &= [\boldsymbol{b}_1 \ \ldots \ \boldsymbol{b}_N]. \end{aligned} \tag{3.3}$$

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

A basis is a generalization of the usual concept of a **coordinate system**.

--- (Read)

Proof of uniqueness of (3.3).
For any $\boldsymbol{v} \in \mathcal{V}$, there *exists* some coordinates $(\alpha_1, \alpha_2, \ldots)$ by definition of **span**.
To prove uniqueness, use contradiction. Suppose the representation were not unique, *i.e.*, suppose there exists coefficients $(\beta_1, \beta_2, \ldots)$ such that $\boldsymbol{v} = \sum_n \beta_n \boldsymbol{b}_n$, where at least one β_n differs from α_n. Then $\mathbf{0} = \boldsymbol{v} - \boldsymbol{v} = \sum_n \beta_n \boldsymbol{b}_n - \sum_n \alpha_n \boldsymbol{b}_n = \sum_n (\beta_n - \alpha_n) \boldsymbol{b}_n$. But because at least one coefficient in that sum is nonzero, that would imply that the set $\{\boldsymbol{b}_n\}$ is linearly dependent, contradicting the definition of a basis.

Fundamental properties of bases

Basis vectors are always nonzero, because of the linear independence condition in the definition.

Fact. Every subspace in a vector space has a **basis**.

Proof for $\mathcal{S} \subseteq \mathbb{F}^N$. (Read)

If $\mathcal{S} = \{\mathbf{0}_N\}$ then the empty set $\emptyset$ is a basis for $\mathcal{S}$ because $\mathsf{span}(\emptyset) = \mathbf{0}$. See p. 3.8.

Now consider a nonzero subspace $\mathcal{S}$. There is a nonzero vector $\boldsymbol{b}_1 \in \mathcal{S}$. If $\mathsf{span}(\{\boldsymbol{b}_1\}) = \mathcal{S}$, then $\{\boldsymbol{b}_1\}$ is a basis for $\mathcal{S}$. Otherwise there is a nonzero vector $\boldsymbol{b}_2 \in \mathcal{S}$ that is not in $\mathsf{span}(\boldsymbol{b}_1)$. If $\mathsf{span}(\{\boldsymbol{b}_1, \boldsymbol{b}_2\}) = \mathcal{S}$, then $\{\boldsymbol{b}_1, \boldsymbol{b}_2\}$ is a basis for $\mathcal{S}$. Otherwise there is a nonzero vector $\boldsymbol{b}_3 \in \mathcal{S}$ that is not in $\mathsf{span}(\{\boldsymbol{b}_1, \boldsymbol{b}_2\})$. If $\mathsf{span}(\{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}) = \mathcal{S}$, then $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3\}$ is a basis for $\mathcal{S}$. Otherwise continue as above. At each step, the vectors in the set are linearly independent. The process will terminate in $k \leq N$ steps, because any $N + 1$ vectors in $\mathbb{F}^N$ are linearly dependent. The resulting set $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k\}$ is linearly independent and $\mathsf{span}(\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k\}) = \mathcal{S}$.

For more general cases see stackexchange. $\square$

Fact. Every **basis** for a subspace has the same **cardinality** (*i.e.*, number of elements, possibly infinite).

This fact is called the **dimension theorem for vector spaces**. That link shows a proof.

Example. The set of complex exponential signals $\left\{ e^{\imath 2\pi k t/T}, \, t \in \mathbb{R} \, : \, -K \leq k \leq K \right\}$ 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 $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N\}$ with $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots\}$ and allow an infinite sum in (3.3). Dealing rigorously with infinite sums requires care beyond the scope of EECS 551; see EECS 600! ♦♦

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

Example. The following vector space is important in signal processing.

The vector space of sinusoids of frequency ν is: $\mathcal{V} = \{A\cos(2\pi\nu t + \phi), \, t \in \mathbb{R} \, : \, A, \phi \in \mathbb{R}\}$.
Is $\mathcal{S} = \{3\cos(2\pi\nu t), \; 5\sin(2\pi\nu t), \; 7\cos(2\pi\nu t - \pi/4)\}$ a basis for $\mathcal{V}$? Hint: [wiki]

3.
A: Yes
B: No, because $\mathcal{S}$ has linear dependence
C: No, because $\mathcal{S}$ does not span $\mathcal{V}$
D: Both B & C.

??

Warm-up questions

4. Let $\mathcal{V} = \mathbb{R}^3$ and $\mathcal{S} \triangleq \{(1,1,1),\ (1,0,0)\}$. What is $\mathrm{span}(\mathcal{S})$?

A: $\mathbf{0}$ B: a line C: a plane D: all of $\mathbb{R}^3$ E: None of these **??**

5. Upper triangular matrices are a subspace of $\mathbb{F}^{N \times N}$. (?)

A: True B: False **??**

6. A basis for such upper triangular matrices in vector space $\mathbb{F}^{N \times N}$ contains how many vectors?

A: N B: N^2 C: $N^2/2$ D: $N(N-1)$ E: None of these **??**

Dimension

L§2.3

Define. The **dimension** of a subspace $\mathcal{S}$ is the number of elements in any **basis** for $\mathcal{S}$.

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), per p. 3.14.

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

Example. $\dim(\mathbb{F}^{M \times N}) = MN$. Use the canonical or standard basis: $\{\boldsymbol{e}_m \boldsymbol{e}_n' : m = 1, \ldots, M, \ n = 1, \ldots, N\}$.

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

What is the dimension of the trivial vector space $\mathcal{V} = \{\mathbf{0}\}$? $\operatorname{span}(\emptyset) = \mathbf{0}$ and $\dim(\emptyset) = 0$. See p. 3.8.

7. What is the dimension of the subspace in $\mathbb{F}^{N \times N}$ of $N \times N$ diagonal matrices?
A: 1 B: N C: $2N$ D: N^2 E: ∞ $\boxed{??}$

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**.

8. What is the dimension of the vector space of all polynomials?
A: 0 B: 1 C: infinite D: undefined $\boxed{??}$

Dimension of a span and bases (Read)

Fact. If $\mathcal{S} = \mathrm{span}(\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N\})$ then $\dim(\mathcal{S}) \leq N$.

The basis reduction theorem states: either $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N\}$ is a **basis** for $\mathcal{S} = \mathrm{span}(\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_N\})$, or one can remove $N - \dim(\mathcal{S})$ appropriately chosen vectors from that set to form a **basis** for $\mathcal{S}$.

The **basis extension theorem** states that every **linearly independent** set of M vectors in a finite-dimensional subspace $\mathcal{S}$ can be extended (with $\dim(\mathcal{S}) - M$ additional vectors) to form a basis of $\mathcal{S}$.
For a proof, see [2, Thm. 5.3.7].

Fact. If $\mathcal{S}$ and $\mathcal{T}$ are both subspaces of a finite-dimensional vector space $\mathcal{V}$, then

$$\mathcal{S} \subseteq \mathcal{T} \Longrightarrow \dim(\mathcal{S}) \leq \dim(\mathcal{T}).$$

For a proof that uses the **basis extension theorem** see [2, Thm. 5.4.4].

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

Sums and intersections of subspaces

L§2.4

Define. If $\mathcal{S}, \mathcal{T} \subseteq \mathcal{V}$ then

- the **sum** of two subspaces is defined as

$$\mathcal{S} + \mathcal{T} = $$

- the **intersection** of two subspaces is defined as

$$\mathcal{S} \cap \mathcal{T} = $$

Exercise. Verify that $\mathcal{S} + \mathcal{T}$ and $\mathcal{S} \cap \mathcal{T}$ are both subspaces of $\mathcal{V}$. (See [1, Theorem 2.22].)

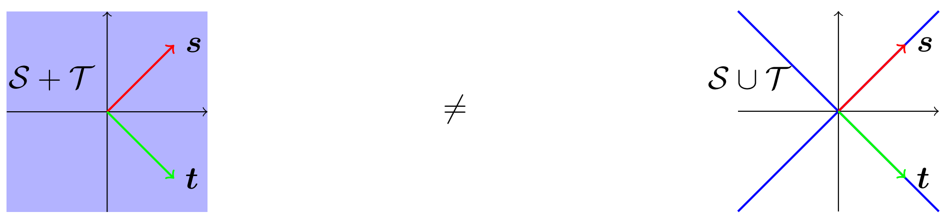
Example. If $\mathcal{S}$ is the subspace of upper Hessenberg matrices in $\mathbb{R}^{N\times N}$ and $\mathcal{T}$ is the subspace of lower Hessenberg matrices in $\mathbb{R}^{N\times N}$ then $\mathcal{S}+\mathcal{T}=\mathbb{R}^{N\times N}$.

9. Considering the same $\mathcal{S}$ and $\mathcal{T}$, what is $\mathcal{S}\cap\mathcal{T}$?

A: $\emptyset$ B: diagonal matrices C: tridiagonal matrices D: $\mathbb{R}^{N\times N}$ E: None of these **??**

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

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



We may discuss **union of subspaces** [3–5] later in the course.

10. A union of subspaces is a subspace. (?)

A: True B: False **??**

Direct sum of subspaces

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

Define. For subspaces $\mathcal{S}$ and $\mathcal{T}$ of vector space $\mathcal{V}$, we write the subspace sum $\mathcal{S} + \mathcal{T}$ as a **direct sum**

$$\mathcal{S} \oplus \mathcal{T} \text{ iff }$$

In this case we say $\mathcal{S}$ and $\mathcal{T}$ are **complements** of each other in the subspace $\mathcal{S} + \mathcal{T}$.

Example. $\mathcal{V} = \mathbb{R}^2$ and $\mathcal{S} = \mathsf{span}(\boldsymbol{s}),\ \boldsymbol{s} = (1, 1)$ and $\mathcal{T} = \mathsf{span}(\boldsymbol{t}),\ \boldsymbol{t} = (1, -1)$. Then $\mathcal{V} = \mathcal{S} \oplus \mathcal{T}$.

For $\boldsymbol{s}, \boldsymbol{t} \in \mathcal{V}$ with $\boldsymbol{s}, \boldsymbol{t} \neq \boldsymbol{0}$ and $\boldsymbol{t} \neq \alpha \boldsymbol{s}$ for all α, let $\mathcal{S} = \mathsf{span}(\boldsymbol{s})$ and $\mathcal{T} = \mathsf{span}(\boldsymbol{t})$.
Is $\mathsf{span}(\{\boldsymbol{s}, \boldsymbol{t}\}) = \mathcal{S} \oplus \mathcal{T}$?

A: Always
11. B: If and only if $\boldsymbol{s}$ and $\boldsymbol{t}$ are orthogonal
C: If and only if $\boldsymbol{s}$ and $\boldsymbol{t}$ 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 $\mathcal{U} = \mathcal{S} \oplus \mathcal{T}$ then:

- every $\boldsymbol{u} \in \mathcal{U}$ can be written uniquely in the form $\boldsymbol{u} = \boldsymbol{s} + \boldsymbol{t}$ for some $\boldsymbol{s} \in \mathcal{S}$ and $\boldsymbol{t} \in \mathcal{T}$,
-

To prove uniqueness, suppose $\boldsymbol{u} = \boldsymbol{s}_1 + \boldsymbol{t}_1 = \boldsymbol{s}_2 + \boldsymbol{t}_2$.
Then $\underbrace{\boldsymbol{s}_1 - \boldsymbol{s}_2}_{\in \mathcal{S}} = \underbrace{\boldsymbol{t}_2 - \boldsymbol{t}_1}_{\in \mathcal{T}}$. But $\mathcal{S} \cap \mathcal{T} = \boldsymbol{0} \Longrightarrow \boldsymbol{s}_1 - \boldsymbol{s}_2 = \boldsymbol{0}$ and $\boldsymbol{t}_2 - \boldsymbol{t}_1 = \boldsymbol{0} \Longrightarrow$ the representation is unique.

This property gives us a tool to help quantify dimension.

Example. In the previous example, $\mathsf{dim}(\mathcal{S}) = \mathsf{dim}(\mathcal{T}) = 1$ and $\mathsf{dim}(\mathcal{V}) = 2$.

(Read)

More generally, if we have any two subspaces in a vector space $\mathcal{V}$, then [1, Thm. 2.27]:

$$\mathsf{dim}(\mathcal{S} + \mathcal{T}) = \mathsf{dim}(\mathcal{S}) + \mathsf{dim}(\mathcal{T}) - \mathsf{dim}(\mathcal{S} \cap \mathcal{T}) . \tag{3.4}$$

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.)

Example. In $\mathcal{V} = \mathbb{R}^3$, if $\mathcal{S}$ is the $x{-}y$ plane and $\mathcal{T}$ is the $y{-}z$ plane,
then $\mathsf{dim}(\mathcal{S} + \mathcal{T}) = 3$, $\mathsf{dim}(\mathcal{S}) = \mathsf{dim}(\mathcal{T}) = 2$, and $\mathsf{dim}(\mathcal{S} \cap \mathcal{T}) = 1$.

Orthogonal complement of a subspace

L§3.4

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

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

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

$\mathcal{S}^{\perp}$ is itself a **subspace** of $\mathcal{V}$

$$(\mathcal{S}^{\perp})^{\perp} = \mathcal{S} \tag{3.5}$$

$$\mathcal{S} \oplus \mathcal{S}^{\perp} = \mathcal{V} \text{ because } \mathcal{S} \cap \mathcal{S}^{\perp} = \{\mathbf{0}\} \tag{3.6}$$

$$\mathsf{dim}(\mathcal{S}) + \mathsf{dim}\big(\mathcal{S}^{\perp}\big) = \mathsf{dim}(\mathcal{V}) \tag{3.7}$$

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

Example. Recall that in any vector space $\mathcal{V}$, the point $\{\mathbf{0}\}$ is a subspace. Clearly $\{\mathbf{0}\}^{\perp} = \mathcal{V}$ and $\mathcal{V}^{\perp} = \{\mathbf{0}\}$.

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

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

Caution. The **orthogonal complement** is *not* the same as the ordinary set **complement**.

Linear transformations

(Read) L§3.1

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

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

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

Matrix-vector multiplication as a linear transformation

Example. Let $\mathcal{V} = \mathbb{C}^N$ and $\mathcal{W} = \mathbb{C}^M$ and $\boldsymbol{A} \in \mathbb{C}^{M \times N}$.
Consider the transformation defined by $\boldsymbol{y} = L(\boldsymbol{x}) \iff \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, *i.e.*, $\boldsymbol{x} \mapsto \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.
This transformation is **linear**.
Proof. $L(\alpha \boldsymbol{x} + \beta \boldsymbol{z}) = \boldsymbol{A}(\alpha \boldsymbol{x} + \beta \boldsymbol{z}) = \alpha \boldsymbol{A}\boldsymbol{x} + \beta \boldsymbol{A}\boldsymbol{z} = \alpha L(\boldsymbol{x}) + \beta L(z)$,
where $\alpha, \beta \in \mathbb{C}$ are arbitrary as are $\boldsymbol{x}, \boldsymbol{z} \in \mathbb{C}^N$. □

For more examples of linear transformations, see [1, Ex. 3.2].

Range of a matrix

L§3.4

Define. The **range** of a matrix $\boldsymbol{A} = [\boldsymbol{a}_1 \ \ldots \ \boldsymbol{a}_N] \in \mathbb{F}^{M \times N}$, also known as its **column space**, is the span of its columns:

Equivalently: because the matrix-vector product is a linear combination of the columns:

$$\boldsymbol{A}\boldsymbol{x} = \sum_{n=1}^{N} \boldsymbol{a}_n x_n.$$

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 $\boldsymbol{A}$ is the span of its rows: $\mathcal{R}(\boldsymbol{A}')$.

Example. For $\boldsymbol{A} = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$, we have

$$\mathcal{R}(\boldsymbol{A}) = \mathsf{span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \end{bmatrix} \right\}\right) = \mathsf{span}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right) = \left\{ \alpha \begin{bmatrix} 1 \\ 2 \end{bmatrix} \ : \ \alpha \in \mathbb{F} \right\}$$

and $\mathcal{R}^{\perp}(\boldsymbol{A}) = \mathsf{span}\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right)$.

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

A: $\emptyset$ B: Usually $\mathbb{R}^{N}$. C: Always $\mathbb{R}^{N}$. D: Usually $\mathbb{R}^{N \times N}$. E: Always $\mathbb{R}^{N \times N}$.

??

Practical use in JULIA. JULIA has a `range` method, but it is unrelated to the **range** of a matrix! So there is no built-in function for $\mathcal{R}(\boldsymbol{A})$, but it is easy to write one using an SVD. (See section starting on p. 4.65).

Range and matrix multiplication (Read)

Often we are interested in the range of matrix products. Clearly:

$$\boldsymbol{A} \in \mathbb{F}^{M \times N}, \quad \boldsymbol{B} \in \mathbb{F}^{N \times K} \Longrightarrow \mathcal{R}(\boldsymbol{A}\boldsymbol{B}) \subseteq \mathcal{R}(\boldsymbol{A}), \tag{3.8}$$

because $\mathcal{R}(\boldsymbol{A}\boldsymbol{B}) = \left\{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} = \boldsymbol{B}\boldsymbol{z},\ \boldsymbol{z} \in \mathbb{F}^K\right\} \subseteq \left\{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \in \mathbb{F}^N\right\}$.

If $\boldsymbol{A}$ is a $M \times N$ matrix and $\boldsymbol{B}$ is a $N \times N$ **invertible matrix** then $\mathcal{R}(\boldsymbol{A}\boldsymbol{B}) = \mathcal{R}(\boldsymbol{A})$.

Proof. Using (3.8) it suffices to show that $\mathcal{R}(\boldsymbol{A}) \subseteq \mathcal{R}(\boldsymbol{A}\boldsymbol{B})$.

If $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A})$ then there is some $\boldsymbol{x} \in \mathbb{F}^N$ such that $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. Because $\boldsymbol{B}$ is invertible, we can define $\boldsymbol{z} = \boldsymbol{B}^{-1}\boldsymbol{x}$ for which $\boldsymbol{A}\boldsymbol{B}\boldsymbol{z} = \boldsymbol{A}\boldsymbol{B}(\boldsymbol{B}^{-1}\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ so $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A}\boldsymbol{B})$. $\square$

However, invertibility of $\boldsymbol{B}$ is not a necessary condition.

Using concepts introduced later, one can show that a more general sufficient condition is for $\boldsymbol{B}$ to have full **row rank**. But full row rank of $\boldsymbol{B}$ is also not a necessary condition.

Example. If $\boldsymbol{A} = \boldsymbol{0}$, then $\mathcal{R}(\boldsymbol{A}) = \mathcal{R}(\boldsymbol{A}\boldsymbol{B})$ for any suitably sized matrix $\boldsymbol{B}$.

A HW problem may explore necessary and sufficient conditions on $\boldsymbol{B}$, in terms of properties of $\boldsymbol{A}$, such that $\mathcal{R}(\boldsymbol{A}\boldsymbol{B}) = \mathcal{R}(\boldsymbol{A})$.

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. L§3.5

Define. For any $M \times N$ matrix $\boldsymbol{A}$:
- **column rank** of $\boldsymbol{A} \triangleq$
- **row rank** of $\boldsymbol{A} \triangleq$

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

Proof. (Read)

Let r denote the column rank of $\boldsymbol{A} = [\boldsymbol{a}_1 \; \dots \; \boldsymbol{a}_N]$. Because r is the column rank, there exists a basis $\boldsymbol{V} = [\boldsymbol{v}_1 \; \dots \; \boldsymbol{v}_r]$ such that one can write every vector in $\mathcal{R}(\boldsymbol{A})$ as a linear combination of the columns of $\boldsymbol{V}$. Each column of $\boldsymbol{A}$ is in $\mathcal{R}(\boldsymbol{A})$, thus we can express each column of $\boldsymbol{A}$ as a linear combination of the columns of $\boldsymbol{V}$, *i.e.*,

$$\begin{aligned} \boldsymbol{a}_1 &= c_{11}\boldsymbol{v}_1 + \cdots + c_{r1}\boldsymbol{v}_r \\ &\vdots \\ \boldsymbol{a}_N &= c_{1N}\boldsymbol{v}_1 + \cdots + c_{rN}\boldsymbol{v}_r, \end{aligned}$$

where the c_{ij} values denote the coordinates or coefficients w.r.t. the basis $\boldsymbol{V}$.

In matrix form we get a sum-of-outer-products form of matrix multiplication:

$$\underbrace{\boldsymbol{A}}_{M \times N} = \underbrace{\begin{bmatrix} \boldsymbol{v}_1 & \dots & \boldsymbol{v}_r \end{bmatrix}}_{M \times r} \underbrace{\begin{bmatrix} c_{11} & \dots & c_{1N} \\ \vdots & & \vdots \\ c_{r1} & \dots & c_{rN} \end{bmatrix}}_{r \times N} = \begin{bmatrix} \boldsymbol{v}_1 & \dots & \boldsymbol{v}_r \end{bmatrix} \begin{bmatrix} - & \boldsymbol{c}_1^\top & - \\ & \vdots & \\ - & \boldsymbol{c}_r^\top & - \end{bmatrix} = \boldsymbol{V}\boldsymbol{C} = \sum_{k=1}^r \boldsymbol{v}_k \boldsymbol{c}_k^\top .$$

Now the mth row of $\boldsymbol{A}$ is a linear combination of the rows of $\boldsymbol{C}$:

$$\boldsymbol{A}_{[m,:]} = \boldsymbol{e}_m' \boldsymbol{A} = \sum_{k=1}^r (\boldsymbol{e}_m' \boldsymbol{v}_k) \boldsymbol{c}_k^\top = \sum_{k=1}^r v_{mk} \boldsymbol{c}_k^\top .$$

This construction holds for any row of $\boldsymbol{A}$ (or linear combinations thereof)

$$\Longrightarrow \mathcal{R}(\boldsymbol{A}') = \text{ row space of } \boldsymbol{A} \subseteq \mathcal{R}\left(\boldsymbol{C}^\top\right)$$
$$\Longrightarrow \text{row rank of } \boldsymbol{A} \leq r = \text{ column rank of } \boldsymbol{A}.$$

Because $\boldsymbol{A}$ was arbitrary, the same argument applies to its transpose, so:

$$\text{row rank of } \boldsymbol{A}' \leq \text{ column rank of } \boldsymbol{A}'$$
$$\Longrightarrow \text{column rank of } \boldsymbol{A} \leq \text{ row rank of } \boldsymbol{A}$$
$$\Longrightarrow \text{column rank of } \boldsymbol{A} = \text{ row rank of } \boldsymbol{A}.$$

□

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:

$$\mathsf{rank}(\boldsymbol{A}) \triangleq \mathsf{dim}(\mathcal{R}(\boldsymbol{A})) .$$

Corollary:

$$\boldsymbol{A} \in \mathbb{F}^{M \times N} \Longrightarrow \qquad (3.9)$$

Proof. (Read)

$\mathsf{rank}(\boldsymbol{A})$ = row rank of $\boldsymbol{A} \leq$ # rows $= M$
$\mathsf{rank}(\boldsymbol{A})$ = col rank of $\boldsymbol{A} \leq$ # cols $= N$
$\Longrightarrow \mathsf{rank}(\boldsymbol{A}) \leq \min(M, N)$ □

Example. For $\boldsymbol{A} = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$, we showed in a previous example that

$$\mathcal{R}(\boldsymbol{A}) = \mathsf{span}\left(\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \end{bmatrix} \right\}\right) = \mathsf{span}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} \right).$$

Thus $\mathsf{rank}(\boldsymbol{A}) = 1$.

14. What is the rank of $\begin{bmatrix} 5 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 7 \end{bmatrix}$?

A: 0 B: 1 C: 2 D: 3 E: 4 **??**

Practical use in JULIA: `rank(A)`. It involves a threshold on the singular values. (HW)

Rank of a matrix product

L§3.5

Theorem. Multiplying matrices never increases rank:

$$\boldsymbol{A} \in \mathbb{F}^{M \times N}, \quad \boldsymbol{B} \in \boldsymbol{F}^{N \times K} \Longrightarrow 0 \leq \operatorname{rank}(\boldsymbol{A}\boldsymbol{B}) \leq \qquad (3.10)$$

Proof. (Read)

$$\boldsymbol{A}\boldsymbol{B} = \begin{bmatrix} | & & | \\ \boldsymbol{a}_1 & \ldots & \boldsymbol{a}_N \\ | & & | \end{bmatrix} \begin{bmatrix} - & \boldsymbol{b}_1^\top & - \\ & \vdots & \\ - & \boldsymbol{b}_N^\top & - \end{bmatrix} = \boldsymbol{a}_1 \boldsymbol{b}_1^\top + \cdots + \boldsymbol{a}_N \boldsymbol{b}_N^\top$$

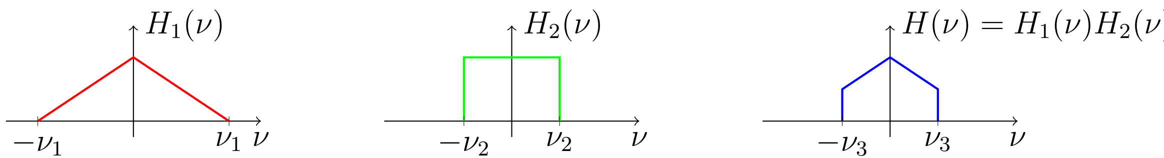
$\Longrightarrow$ every column of $\boldsymbol{A}\boldsymbol{B}$ is a linear combination of columns of $\boldsymbol{A}$ (see also (1.7))
$\Longrightarrow \mathcal{R}(\boldsymbol{A}\boldsymbol{B}) \subseteq \mathcal{R}(\boldsymbol{A})$
$\Longrightarrow \operatorname{rank}(\boldsymbol{A}\boldsymbol{B}) \leq \operatorname{rank}(\boldsymbol{A})$ using the inequality on p. 3.18.

Similarly, because $(\boldsymbol{A}\boldsymbol{B})' = \boldsymbol{B}'\boldsymbol{A}'$, applying the same logic we have
$\operatorname{rank}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{rank}((\boldsymbol{A}\boldsymbol{B})') = \operatorname{rank}(\boldsymbol{B}'\boldsymbol{A}') \leq \operatorname{rank}(\boldsymbol{B}') = \operatorname{rank}(\boldsymbol{B})$.
Combining both inequalities yields the first inequality in (3.10). The second inequality follows from (3.9). □

- Caution: in general: $\operatorname{rank}(\boldsymbol{A}\boldsymbol{B}) \neq \operatorname{rank}(\boldsymbol{B}\boldsymbol{A})$, even if the sizes are compatible.
- If $N = 1$ and $\boldsymbol{A}$ and $\boldsymbol{B}$ are nonzero (sufficient conditions), then both upper bounds in (3.10) are achieved.
- Challenge. Find necessary conditions on $\boldsymbol{A}$,$\boldsymbol{B}$ for when one of the upper bounds in (3.10) is achieved.

$\boldsymbol{AB}$ is a composition of two linear transforms.
In words, (3.10) implies that composition cannot enlarge subspace dimension.

Example. DSP analogy: cascade of two filters with band-limited frequency responses.



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

---

15. For $\boldsymbol{u} \in \mathbb{C}^M$, $\boldsymbol{v} \in \mathbb{C}^N$, what is the {minimum, maximum} possible rank of **outer product** $\boldsymbol{u}\boldsymbol{v}'$?
A: $0, 1$ B: $1, 1$ C: $0, \min(M, N)$ D: $1, \min(M, N)$ E: None of these $\boxed{\textbf{??}}$

Other properties of rank

There is also a lower bound for the **rank** of a **matrix product**, called **Sylvester's rank inequality**:

$$\boldsymbol{A} \in \mathbb{F}^{M \times N}, \quad \boldsymbol{B} \in \boldsymbol{F}^{N \times K} \Longrightarrow \operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B}) - N \leq \operatorname{rank}(\boldsymbol{A}\boldsymbol{B}) \,.$$

Furthermore, rank is **subadditive**:

$$\boldsymbol{A}, \boldsymbol{B} \in \mathbb{F}^{M \times N} \Longrightarrow 0 \leq \operatorname{rank}(\boldsymbol{A} + \boldsymbol{B}) \leq \operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B}) \,.$$

A tighter inequality is

$$\boldsymbol{A}, \boldsymbol{B} \in \mathbb{F}^{M \times N} \Longrightarrow 0 \leq \operatorname{rank}(\boldsymbol{A} + \boldsymbol{B}) \leq \min(\operatorname{rank}(\boldsymbol{A}) + \operatorname{rank}(\boldsymbol{B}), M, N).$$

Warm-up question(s)

16. If $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_K$ are sized appropriately to allow multiplication, then $\mathsf{rank}(\boldsymbol{A}_1 \boldsymbol{A}_2 \cdots \boldsymbol{A}_K) \leq \min(\mathsf{rank}(\boldsymbol{A}_1), \ldots, \mathsf{rank}(\boldsymbol{A}_K))$.

A: True B: False **??**

17. For $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{C}^M$ and $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{C}^N$, what is the {minimum, maximum} possible rank of the $M \times N$ matrix $3\boldsymbol{u}_1\boldsymbol{v}_1' + 5\boldsymbol{u}_2\boldsymbol{v}_2'$?

A: $0, 1$ B: $1, 2$ C: $0, \min(M, N)$ D: $1, \min(M, N)$ E: None of these **??**

18. If $\mathcal{V} = \mathbb{F}^6$ and $\boldsymbol{x}$ and $\boldsymbol{y}$ are two linearly independent vectors in $\mathcal{V}$, then what is the dimension of $(\mathsf{span}(\{\boldsymbol{x}, \boldsymbol{y}\}))^\perp$?

A: 1 or less B: 2 C: 3 D: 4 E: 5 or more **??**

Unitary invariance of rank / eigenvalues / singular values

Theorem. If $\boldsymbol{A} \in \mathbb{F}^{M \times N}$ then multiplying by a **unitary** matrix on the left or right does not change the rank:
- $\boldsymbol{Q} \in \mathbb{F}^{M \times M}$ and $\boldsymbol{Q}$ unitary $\Longrightarrow$
- $\boldsymbol{Q} \in \mathbb{F}^{N \times N}$ and $\boldsymbol{Q}$ unitary $\Longrightarrow$

Rank and eigenvalues

Corollary. If $\boldsymbol{A}$ is Hermitian (or **normal**) with unitary eigendecomposition $\boldsymbol{A} = \boldsymbol{V} \boldsymbol{\Lambda} \boldsymbol{V}'$ then, because $\boldsymbol{V}$ is unitary (and so is $\boldsymbol{V}'$), we have the important property:

$$\operatorname{rank}(\boldsymbol{A}) = \qquad = \text{ number of nonzero eigenvalues of } \boldsymbol{A}.$$

19. If $\boldsymbol{A}$ is square, then $\operatorname{rank}(\boldsymbol{A})$ = number of nonzero eigenvalues of $\boldsymbol{A}$.

A: True B: False $\boxed{??}$

Rank and SVD

- By unitary invariance, if $\boldsymbol{A}$ has SVD $\boldsymbol{A} = \boldsymbol{U\Sigma V'}$, then $\mathsf{rank}(\boldsymbol{A}) = \mathsf{rank}(\boldsymbol{\Sigma})$.
- $\mathsf{rank}(\boldsymbol{\Sigma}) = r =$ # of nonzero singular values of $\boldsymbol{A}$, because $\mathcal{R}(\boldsymbol{\Sigma}) = \mathsf{span}(\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_r\})$

Thus the SVD expression for a matrix $\boldsymbol{A}$ having rank r where $r \leq \min(M, N)$ simplifies:

$$\boldsymbol{A} \in \mathbb{F}^{M \times N} \Longrightarrow \boldsymbol{A} = \boldsymbol{U\Sigma V'} = \sum_{k=1}^{\min(M,N)} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k' = \boxed{}$$

When we write $\sum_{k=1}^{\min(M,N)}$ then some of the σ_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 $\boldsymbol{A}$, then all the σ_k values used in the sum are nonzero.

By convention, any empty sum (when $r = 0$) evaluates to zero, in this case $\mathbf{0}_{M \times N}$.

This convention holds in JULIA as well: `sum(ones(0))` evaluates to `0` because `ones(0)` is an empty `Float64` variable (a 0-dimensional vector).

Exercise. Determine the convention for an empty product. **??**

3.3 Nullspace of a matrix

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

Nullspace or kernel L§3.4

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 $\boldsymbol{A}$ is

$$\mathcal{N}(\boldsymbol{A}) = \mathsf{ker}(\boldsymbol{A}) \triangleq$$

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

Exercise. Verify that $\mathcal{N}(\boldsymbol{A})$ is indeed a **subspace**.
Thus using the subspace font "$\mathcal{N}$" is appropriate.

Example. For $\boldsymbol{A} = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$, we have $\mathcal{N}(\boldsymbol{A}) = \mathsf{span}\left(\begin{bmatrix} -3 \\ 1 \end{bmatrix}\right)$ and $\mathcal{N}^{\perp}(\boldsymbol{A}) = \mathsf{span}\left(\begin{bmatrix} 1 \\ 3 \end{bmatrix}\right)$.

20. If $\boldsymbol{A} \in \mathbb{C}^{M \times N}$, then $\mathcal{N}(\boldsymbol{A})$ 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. $\boxed{\textbf{??}}$

Practical use in JULIA: `nullspace(A)`. Like `rank`, this involves a threshold on the singular values.

Properties of null space

(Read)

- $\mathcal{N}(\boldsymbol{A}) = \{\mathbf{0}\} \iff \boldsymbol{A}$ has linearly independent columns
- $\mathcal{N}(\boldsymbol{A}) = \mathbb{F}^N \iff \boldsymbol{A} = \mathbf{0}_{M \times N}$
- $\mathcal{N}(\boldsymbol{B}) \subseteq \mathcal{N}(\boldsymbol{AB})$
  Proof. $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{B}) \Longrightarrow \boldsymbol{Bx} = \mathbf{0} \Longrightarrow \boldsymbol{ABx} = \mathbf{0} \Longrightarrow \boldsymbol{x} \in \mathcal{N}(\boldsymbol{AB})$
- If $\mathcal{N}(\boldsymbol{A}) = \{\mathbf{0}\}$, *i.e.*, if $\boldsymbol{A}$ has full column rank, then $\mathcal{N}(\boldsymbol{B}) = \mathcal{N}(\boldsymbol{AB})$.
  This is not a necessary condition, but it is adequate for our purposes.

Decomposition theorem for matrices

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

$$\qquad\qquad = \mathbb{F}^N$$

$$\qquad\qquad = \mathbb{F}^M.$$

In other words, every "input" vector $\boldsymbol{x} \in \mathbb{F}^N$ can be decomposed uniquely as $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{x}_1$, where $\boldsymbol{x}_0 \in \mathcal{N}(\boldsymbol{A})$ so $\boldsymbol{Ax}_0 = \mathbf{0}$ and $\boldsymbol{x}_1 \in \mathcal{N}^\perp(\boldsymbol{A})$.

Likewise, every vector $\boldsymbol{y} \in \mathbb{F}^M$ can be decomposed uniquely as $\boldsymbol{y} = \boldsymbol{y}_1 + \boldsymbol{y}_0$, where $\boldsymbol{y}_1 \in \mathcal{R}(\boldsymbol{A})$, so $\boldsymbol{Ax}_1 = \boldsymbol{y}_1$ for some $\boldsymbol{x}_1 \in \mathbb{F}^N$, and $\boldsymbol{y}_0 \in \mathcal{R}^\perp(\boldsymbol{A})$.

The above statement was for an arbitrary $\boldsymbol{y} \in \mathbb{F}^M$.
If we have an "output vector" $\boldsymbol{y} = \boldsymbol{Ax}$, then $\boldsymbol{y} \in \mathcal{R}(\boldsymbol{A})$ by definition and $\boldsymbol{y}_0 = \mathbf{0}$.

Relationships between null space and range of a matrix

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

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

Proof: If $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{A})$ then $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, so $\forall \boldsymbol{y}$ we have $\boldsymbol{y}'(\boldsymbol{A}\boldsymbol{x}) = 0 \Longrightarrow \boldsymbol{x}'(\boldsymbol{A}'\boldsymbol{y}) = 0 \Longrightarrow \boldsymbol{x} \in \mathcal{R}^{\perp}(\boldsymbol{A}')$.
Conversely, if $\boldsymbol{x} \in \mathcal{R}^{\perp}(\boldsymbol{A}')$ then $\forall \boldsymbol{y}$, $0 = \boldsymbol{x}'(\boldsymbol{A}'\boldsymbol{y})$.
Take $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ and we have $\boldsymbol{x}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} = 0$ so $\|\boldsymbol{A}\boldsymbol{x}\| = 0$ which implies $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, so $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{A})$.
Thus using (3.5): $\mathcal{N}(\boldsymbol{A}) = \mathcal{R}^{\perp}(\boldsymbol{A}') \Longrightarrow \mathcal{N}^{\perp}(\boldsymbol{A}) = \mathcal{R}(\boldsymbol{A}')$.
The proof of the second equality is left to HW (*cf.* [1, Problem 3.6]). □

Corollary:

$$\begin{aligned} \dim\left(\mathcal{N}^{\perp}(\boldsymbol{A})\right) &= \dim(\mathcal{R}(\boldsymbol{A}')) = \dim(\mathcal{R}(\boldsymbol{A})) = \operatorname{rank}(\boldsymbol{A}) \\ \dim\left(\mathcal{R}^{\perp}(\boldsymbol{A})\right) &= \dim(\mathcal{N}(\boldsymbol{A}')). \end{aligned} \tag{3.11}$$

Nullity

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

$$\mathrm{nullity}(\boldsymbol{A}) \triangleq$$

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

$$= N.$$

Proof. Because $\mathbb{F}^N = \mathcal{N}(\boldsymbol{A}) \oplus \mathcal{N}^\perp(\boldsymbol{A})$ we have from (3.6), (3.7) and (3.11):
$N = \dim(\mathcal{N}(\boldsymbol{A})) + \dim\left(\mathcal{N}^\perp(\boldsymbol{A})\right) = \dim(\mathcal{N}(\boldsymbol{A})) + \mathrm{rank}(\boldsymbol{A})$. □

---

21. If $\boldsymbol{u} \in \mathbb{C}^M - \{\mathbf{0}\}$ and $\boldsymbol{v} \in \mathbb{C}^N - \{\mathbf{0}\}$, what is the nullity of their **outer product**, *i.e.*, $\mathrm{nullity}(\boldsymbol{u}\boldsymbol{v}')$?

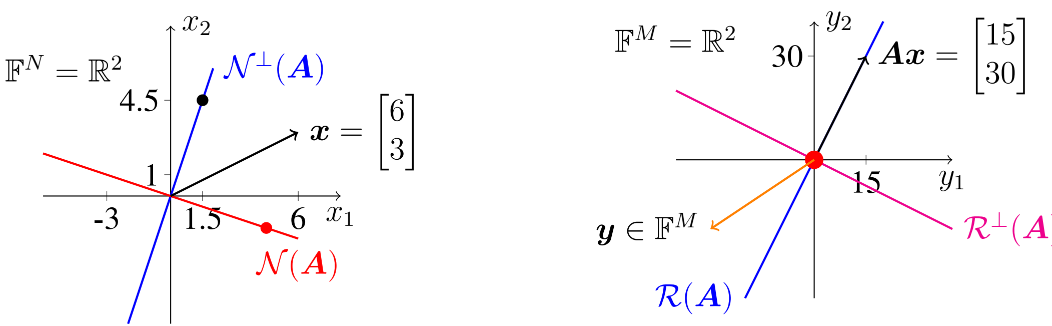
A: 0 B: 1 C: $N - 1$ D: N E: M **??**

3.4 The four fundamental spaces

Now we unify the **null space** and **range space** (and their orthogonal complements) for a general matrix $\boldsymbol{A}$. We start by revisiting a simple 2×2 example.

Example. For $\boldsymbol{A} = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$, recall we showed previously that:

$$\mathcal{N}(\boldsymbol{A}) = \operatorname{span}\left(\begin{bmatrix} -3 \\ 1 \end{bmatrix}\right), \quad \mathcal{R}(\boldsymbol{A}) = \operatorname{span}\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right),$$
$$\mathcal{N}^{\perp}(\boldsymbol{A}) = \operatorname{span}\left(\begin{bmatrix} 1 \\ 3 \end{bmatrix}\right), \quad \mathcal{R}^{\perp}(\boldsymbol{A}) = \operatorname{span}\left(\begin{bmatrix} 2 \\ -1 \end{bmatrix}\right).$$



The characteristics illustrated in the previous diagram hold in general. L§3.5

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

$$
\begin{array}{ccccc}
 & \text{“input”} & & \text{“output”} & \\
\text{dim} & \mathbb{F}^N & & \mathbb{F}^M & \text{dim} \\
\hline
r & \boxed{\mathcal{N}^{\perp}(\boldsymbol{A}) = \mathcal{R}(\boldsymbol{V}_r)} & \xrightarrow{\boldsymbol{A}} & \boxed{\mathcal{R}(\boldsymbol{A}) = \mathcal{R}(\boldsymbol{U}_r)} & r \\
 & \oplus \{\mathbf{0}\} & & \{\mathbf{0}\} \oplus & \\
N - r & \boxed{\mathcal{N}(\boldsymbol{A}) = \mathcal{R}(\boldsymbol{V}_0)} & & \boxed{\mathcal{R}^{\perp}(\boldsymbol{A}) = \mathcal{R}(\boldsymbol{U}_0)} & M - r \quad \text{(“unreachable”)}
\end{array}
$$

(The next page defines the matrices $\boldsymbol{U}_r$, $\boldsymbol{U}_0$, $\boldsymbol{V}_r$, $\boldsymbol{V}_0$.)

There is no standard notation for these four matrices.
Some books use $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$; I have also seen $\boldsymbol{V}_{\|}$ and $\boldsymbol{V}_{\perp}$ [6].

L§5.1

When $\boldsymbol{A} \in \mathbb{F}^{M \times N}$ has rank $0 < r < \min(M, N)$, we can partition the SVD components as follows:

$$\boldsymbol{A} = \sum_{k=1}^{\min(M,N)} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k' = \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k' = \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}' =$$

where $\boldsymbol{\Sigma}_r$ is $r \times r$ and contains the *nonzero* singular values of $\boldsymbol{A}$ along its diagonal.

Here $\boldsymbol{U}_r = \begin{bmatrix} \boldsymbol{u}_1 & \dots & \boldsymbol{u}_r \end{bmatrix}$, $\boldsymbol{U}_0 = \begin{bmatrix} \boldsymbol{u}_{r+1} & \dots & \boldsymbol{u}_M \end{bmatrix}$, $\boldsymbol{V}_r = \begin{bmatrix} \boldsymbol{v}_1 & \dots & \boldsymbol{v}_r \end{bmatrix}$, $\boldsymbol{V}_0 = \begin{bmatrix} \boldsymbol{v}_{r+1} & \dots & \boldsymbol{v}_M \end{bmatrix}$.
In the corner case where $r = 0$, we have $\boldsymbol{A} = \boldsymbol{0}$, $\boldsymbol{\Sigma} = \boldsymbol{0}$, $\boldsymbol{U} = \boldsymbol{U}_0$, $\boldsymbol{V} = \boldsymbol{V}_0$, and there is no $\boldsymbol{U}_r$ or $\boldsymbol{V}_r$ or $\boldsymbol{\Sigma}_r$.
In the tall case where $r = N \leq M$, we have $\boldsymbol{A} = \boldsymbol{U}_N \boldsymbol{\Sigma}_N \boldsymbol{V}$ because $\boldsymbol{V} = \boldsymbol{V}_r = \boldsymbol{V}_N$, and there is no $\boldsymbol{V}_0$.
In the wide case where $r = M \leq N$, we have $\boldsymbol{A} = \boldsymbol{U} \boldsymbol{\Sigma}_M \boldsymbol{V}_M$ because $\boldsymbol{U} = \boldsymbol{U}_r = \boldsymbol{U}_M$, and there is no $\boldsymbol{U}_0$.

22. What is the size of the lower right $\boldsymbol{0}$ 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

??

23. If $\boldsymbol{x} = \boldsymbol{V}_0 \boldsymbol{z}$ for some $\boldsymbol{z} \in \mathbb{F}^{N-r}$, then what is $\boldsymbol{A}\boldsymbol{x}$?
A: $\boldsymbol{0}$ B: $\boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{z}$ C: $\boldsymbol{U}_r \boldsymbol{V}_r \boldsymbol{z}$ D: $\boldsymbol{U}_0 \boldsymbol{V}_r \boldsymbol{z}$ E: None of these

??

Anatomy of the SVD

There are two cases of the above partitioning to consider in more detail: when $\boldsymbol{A}$ is **tall** or **wide**.

SVD of a tall matrix

When $\boldsymbol{A}$ is "**tall**" or "thin," *i.e.*, $M > N \Longrightarrow r \leq N < M$, then we can simplify:

$$\underbrace{\boxed{\boldsymbol{A}}}_{M \times N} = \boldsymbol{U \Sigma V}' = \underbrace{\left[\begin{array}{c|c} \boldsymbol{U}_r & \boldsymbol{U}_0 \end{array}\right]}_{M \times r \,|\, M \times (M-r)} \underbrace{\left[\begin{array}{c} \boldsymbol{\Sigma}_r \\ \hline \boldsymbol{0} \end{array}\right]}_{M \times r} \underbrace{\boldsymbol{V}_r'}_{r \times N} = \qquad (3.12)$$

24. What is the size of the lower $\boldsymbol{0}$ 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 $\boldsymbol{U \Sigma V}' = \boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{V}_r'$, we do not mean that the individual terms match (they do not!); we mean that the overall product matches.

If $r = N$ then $\mathcal{N}(\boldsymbol{A}) = \boldsymbol{0}$ and $\boldsymbol{V}_r = \boldsymbol{V}_N = \boldsymbol{V}$ and there is no $\boldsymbol{V}_0$.

SVD of a wide matrix

When $\boldsymbol{A}$ is **wide**, *i.e.*, $M < N \Longrightarrow r \leq M < N$, then we can simplify:

$$\underbrace{\boxed{\boldsymbol{A}}}_{M \times N} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}' = \underbrace{\boldsymbol{U}_r}_{M \times r} \underbrace{\left[\begin{array}{c|c} \boldsymbol{\Sigma}_r & \boldsymbol{0} \end{array}\right]}_{r \times N} \underbrace{\left[\begin{array}{c} \boldsymbol{V}_r' \\ \hline \boldsymbol{V}_0' \end{array}\right]}_{N \times N} \begin{array}{l} \} \; r \times N \\ \\ \} \; (N-r) \times N \end{array} =$$

If $r = M$ then $\dim(\mathcal{R}(\boldsymbol{A})) = M$ and $\boldsymbol{U}_r = \boldsymbol{U}_M = \boldsymbol{U}$ and there is no $\boldsymbol{U}_0$ and all of $\mathbb{F}^M$ is reachable, *i.e.*, $\mathcal{R}(\boldsymbol{A}) = \mathbb{F}^M$.

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

Note the symmetry between the above two "compact" SVD representations for the tall and wide cases, as there must be because if $\boldsymbol{A}$ is tall then $\boldsymbol{A}'$ is wide.

Practical use of SVD (Read)

The above cases are somewhat (but not exactly) related to the `(U,s,V) = svd(A, full=false)` command in JULIA, which is the default option for `svd`.

- `(U,s,V) = svd(A, full=true)` returns the **full SVD**, where `U` is $M \times M$ and `V` is $N \times N$ and `s` is a vector of length $\min(M, N)$.
- `(U,s,V) = svd(A, full=false)` or just `(U,s,V) = svd(A)` returns the **economy SVD**, where `U` is $M \times \min(M, N)$ and `U * Diagonal(s) * V'` equals $\boldsymbol{A}$ to within numerical precision.

In class, when we speak of the **compact SVD**, we mean $\boldsymbol{A} = \boldsymbol{U}_r \boldsymbol{\Sigma}_r \boldsymbol{V}_r'$ as described in these pages.

Size (memory) comparison: **compact** $\leq$ **economy** $=$ **thin** $\leq$ **full**.

There is no built-in JULIA (or MATLAB) command for computing the **compact SVD**!
Computing the compact SVD when $r > 0$ requires multiple JULIA commands, such as:

```
U,s,V = svd(A)
r = rank(Diagonal(s))
Ur = U[:,1:r]; Vr = V[:,1:r]; sr = s[1:r]
```

(This trick avoids a redundant SVD of $\boldsymbol{A}$)

after which we can recover $\boldsymbol{A}$ to within numerical precision by

```
Ur * Diagonal(sr) * Vr'
```

<u>Exercise</u>. Make a small tall matrix like `A = [4 2; 2 1; 0 0]` and experiment with the SVD options above and look at numerical effects by evaluating the recovery error `A - U * Diagonal(s) * V'`

For a **tall** matrix $\boldsymbol{A}$, when we use the (default) `full=false` option of JULIA's SVD, *i.e.*,
`(U,s,V) = svd(A)`
or
`(U,s,V) = svd(A, full=false)`
then the output components correspond to:

$$\underbrace{\underbrace{\begin{bmatrix} \boldsymbol{U}_{:,1} & \dots & \boldsymbol{U}_{:,N} \end{bmatrix}}_{M \times N}}_{\texttt{U}} \underbrace{\underbrace{\begin{bmatrix} \boldsymbol{\Sigma}_r & \boldsymbol{0}_{(N-r)\times r} \\ \boldsymbol{0}_{(N-r)\times(N-r)} & \boldsymbol{0}_{(N-r)\times(N-r)} \end{bmatrix}}_{N \times N}}_{\texttt{Diagonal(s)}} \underbrace{\underbrace{\boldsymbol{V}'}_{N \times N}}_{\texttt{V'}},$$

where `s` $= (\sigma_1, \dots, \sigma_r, 0, \dots, 0)$ has length $N \leq M$ in the tall case.
In practice, the last $N - r$ elements of the vector `s` will not be exactly zero, because of finite numerical precision.

Caution: for a tall $\boldsymbol{A}$, `U` is $M \times N$ whereas $\boldsymbol{U}$ is $M \times M$, and `s` is a N vector whereas $\boldsymbol{\Sigma}$ is $M \times N$.

Example. For a concrete example, see the rank-1 case on p. 3.60.

Exercise. Use the fact that $\mathcal{R}(\boldsymbol{A}) = \mathcal{R}(\boldsymbol{U}_r)$, to modify the **compact SVD** code above to write a function `matrixrange(A)` that returns an orthonormal basis for the range of a given matrix, akin to how `nullspace` returns a basis for its nullspace.

SVD of finite differences (discrete derivative) (Read)

Example. The **derivative** of $f(t) = \sin(\omega t)$ is $\dot{f}(t) = \omega \cos(\omega t)$ and the second derivative is $\ddot{f}(t) = -\omega^2 \sin(\omega t)$. Consider an analog system whose inputs are twice differentiable signals, and whose output is the second derivative of the input:

$$f(t) \to \boxed{\text{second derivative}} \to \ddot{f}(t).$$

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

Consider the $(N-1) \times N$ matrix $\boldsymbol{C}$ that performs a **finite difference** operation when multiplying a vector:

$$\boldsymbol{C} \triangleq \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ & & \ddots & \ddots & & \\ 0 & \dots & 0 & -1 & 1 & 0 \\ 0 & \dots & 0 & 0 & -1 & 1 \end{bmatrix}, \text{ so } \boldsymbol{C}\boldsymbol{x} = \begin{bmatrix} x_2 - x_1 \\ \vdots \\ x_N - x_{N-1} \end{bmatrix} \text{ for } \boldsymbol{x} \in \mathbb{C}^N. \tag{3.13}$$

The $N \times N$ Gram matrix $\boldsymbol{C}'\boldsymbol{C}$ is almost Toeplitz and the related $N-1 \times N-1$ 2nd-difference matrix $\boldsymbol{C}\boldsymbol{C}'$

is exactly Toeplitz:

$$\boldsymbol{C}'\boldsymbol{C} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & -1 & 2 & -1 \\ 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix} \qquad \boldsymbol{C}\boldsymbol{C}' = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ & & \ddots & \ddots & \ddots & \\ 0 & \dots & 0 & -1 & 2 & -1 \\ 0 & 0 & \dots & 0 & -1 & 2 \end{bmatrix}$$

These three important matrices arise in many signal and image processing applications (see HW02 and EECS 556) 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 an economy SVD of $\boldsymbol{C}$ involves the **discrete cosine transform** (**DCT**) and **discrete sine transform** (**DST**) as follows [7]:

$$\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}' = \mathrm{DST}\,\boldsymbol{\Sigma}\,\mathrm{DCT}'. \tag{3.14}$$

So for this important matrix, the left and right singular vectors turn out to form matrices that are themselves important in signal processing. The facts that $\boldsymbol{C}'\boldsymbol{C} = \mathrm{DCT}\,\boldsymbol{\Sigma}^2\,\mathrm{DCT}'$ and $\boldsymbol{C}\boldsymbol{C}' = \mathrm{DST}\,\boldsymbol{\Sigma}^2\,\mathrm{DST}'$ are discrete analogues of the fact that the second derivative of a sinusoid is a sinusoid. The following code verifies (3.14).

```
using LinearAlgebra
N = 6
C = diagm(0 => -ones(Int,N-1), 1 => ones(Int,N-1))[1:(N-1),:] # diff

h = pi / N # from strang:06:bts
U = [sin(j*k*h) for j=1:(N-1), k=1:(N-1)] # DST
for i=1:(N-1); U[:,i] ./= norm(U[:,i]); end
V = [cos((2*j-1)*k*h/2) for j=1:N, k=1:(N-1)] # DCT
for i=1:(N-1); V[:,i] ./= norm(V[:,i]); end
@show maximum(abs.(U' * C * V .* (1 .- I(N-1)))) # verify
```

Synthesis view of matrix decomposition

(Read)

- Eigen-decomposition of square matrix (when it is exists, *e.g.*, $\boldsymbol{A}$ Hermitian):

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}' = \sum_i \lambda_i \boldsymbol{q}_i \boldsymbol{q}_i'$$

  with $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I}$

- SVD of a $M \times N$ matrix:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}' = \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k'$$

  where $\sigma_k > 0$ for $k = 1, \ldots, r =$ rank, and $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{I}_M$ and $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_N$.

In both cases we can "synthesize" the matrix using a sum of rank-1, outer-product terms.

3.5 Orthogonal bases

In signal processing and beyond, bases consisting of orthogonal vectors are particularly important.

Define. A collection of vectors $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots\}$ in a vector space $\mathcal{V}$ is called a **orthogonal basis** for $\mathcal{V}$ iff

- $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots\}$ is a **basis** for $\mathcal{V}$, *i.e.*,
  - $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots\}$ is **linearly independent**
  - $\mathsf{span}(\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots\}) = \mathcal{V}$
- The basis vectors are **orthogonal**, *i.e.*, $\langle \boldsymbol{b}_n, \boldsymbol{b}_m \rangle = 0$ 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 $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots\}$ are nonzero orthogonal vectors, then they are also linearly independent.

Proof (by contradiction). Suppose $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots\}$ 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 $\mathbf{0} = \sum_{n=1}^N \boldsymbol{v}_n c_n$. Pick any $m \in \{1, \ldots, N\}$ and we have $0 = \boldsymbol{v}_m' \mathbf{0} = \boldsymbol{v}_m' \sum_{n=1}^N \boldsymbol{v}_n c_n = \boldsymbol{v}_m' \boldsymbol{v}_m c_m = \|\boldsymbol{v}_m\|_2^2 \, c_m \Longrightarrow c_m = 0$ because $\boldsymbol{v}_m$ is nonzero. Those holds for every m, contradicting the assumption that not all c_n are zero. □

Thus the linear independence condition in the definition above is implied by the orthogonality condition, at least for nonzero vectors. So an equivalent definition of an orthogonal basis of $\mathcal{V}$ is a set of nonzero vectors $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots\}$ that are orthogonal and that span $\mathcal{V}$.

The above definition is general enough to accommodate infinite-dimensional vector spaces.
In finite dimensions the situation is simpler:

The N columns of any orthogonal matrix $\boldsymbol{V} \in \mathbb{R}^{N \times N}$ are an **orthonormal basis** for $\mathbb{R}^N$.

The N columns of any unitary matrix $\boldsymbol{V} \in \mathbb{C}^{N \times N}$ are an **orthonormal basis** for $\mathbb{C}^N$.

Proof.
- The orthogonality condition ensures linear independence by the above theorem.
- For any vector $\boldsymbol{x} \in \mathbb{F}^N$ we have $\boldsymbol{x} = \boldsymbol{I}\boldsymbol{x} = (\boldsymbol{V}\boldsymbol{V}')\,\boldsymbol{x} = \underbrace{\boldsymbol{V}}_{\text{basis}}\underbrace{(\boldsymbol{V}'\boldsymbol{x})}_{\text{coef.}} \in \mathsf{span}(\boldsymbol{V}) \Longrightarrow \mathsf{span}(\boldsymbol{V}) = \mathbb{F}^N$.

Example. The harmonic discrete complex exponential signals $\{\boldsymbol{q}_k : k = 1, \ldots, N\}$ are an **orthonormal basis** for the vector space $\mathbb{C}^N$, where $[\boldsymbol{q}_k]_n = s(k-1, n-1)$ and $s(k, n) = \frac{1}{\sqrt{N}} \mathrm{e}^{\imath 2\pi k n/N}$. This basis corresponds to the (orthonormal) discrete Fourier transform (**DFT**). In matrix form with $w_N = \mathrm{e}^{\imath 2\pi/N}$:

$$\boldsymbol{Q} = [\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N] = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & \ldots & 1 & \ldots & 1 \\ \vdots & & & & \vdots \\ 1 & \ldots & w_N^{(k-1)(n-1)} & \ldots & w_N^{(N-1)(n-1)} \\ \vdots & & & & \vdots \\ 1 & \ldots & w_N^{(k-1)(N-1)} & \ldots & w_N^{(N-1)(N-1)} \end{bmatrix}$$

Exercise: Show that $\langle \boldsymbol{q}_k,\, \boldsymbol{q}_l \rangle = 0$ for $k \neq l$.

Finding coordinates in an orthogonal basis (Read)

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

$$\boldsymbol{x} = \boldsymbol{I}\boldsymbol{x} = \sum_{n=1}^{N} \boldsymbol{e}_n x_n = \underbrace{\left[\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \right]}_{\text{standard basis}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \\ x_N \end{bmatrix}}_{\hookrightarrow \text{ coordinates}}.$$

This is the most trivial orthogonal basis.

Now suppose $\boldsymbol{Q} \in \mathbb{F}^{N \times N}$ is an **orthogonal matrix** (or **unitary matrix**) and hence a basis for $\mathbb{F}^N$.
By definition of an **orthogonal matrix** (or **unitary matrix**): $\boldsymbol{Q}\boldsymbol{Q}' = \boldsymbol{I}$, hence

$$\boldsymbol{x} = \boldsymbol{I}\boldsymbol{x} = (\boldsymbol{Q}\boldsymbol{Q}')\,\boldsymbol{x} = \underbrace{\boldsymbol{Q}}_{\text{basis}} \underbrace{(\boldsymbol{Q}'\boldsymbol{x})}_{\text{coord.}} = \underbrace{\boldsymbol{Q}}_{\text{basis}} \underbrace{\boldsymbol{\alpha}}_{\hookrightarrow \text{ coordinates}}, \quad \text{where } \boldsymbol{\alpha} = \boldsymbol{Q}'\boldsymbol{x} \in \mathbb{F}^N$$

$$= \sum_{n=1}^{N} \underbrace{\boldsymbol{q}_n}_{\substack{\text{basis} \\ \text{vector}}} \underbrace{\alpha_n}_{\hookrightarrow \text{ coordinate!}}, \qquad \text{where } \boldsymbol{Q} = [\boldsymbol{q}_1 \ \ldots \ \boldsymbol{q}_N].$$

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

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

- For a basis in general we need $\boldsymbol{\alpha} = \boldsymbol{Q}^{-1}\boldsymbol{x}$, requiring matrix inversion that needs $O(N^3)$ computation.
- For an orthogonal basis, we need $\boldsymbol{\alpha} = \boldsymbol{Q}'\boldsymbol{x}$, 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 $\boldsymbol{Q}^{-1}$ exists for a general basis? A basis is a linearly independent set.

Note that the definition of an **orthogonal basis** is in terms of a set of vectors, but often we collect those vectors into a matrix and call the matrix an orthogonal basis. One must be careful though about the terminology. If $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_N\}$ is an **orthogonal basis** for $\mathbb{F}^N$, the matrix $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_1 & \ldots & \boldsymbol{b}_N \end{bmatrix}$ is not necessarily an **orthogonal matrix**.

Example. The vectors $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \end{bmatrix} \right\}$ form an orthogonal basis for $\mathbb{R}^2$, but $\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}$ is not an orthogonal matrix.

3.6 Spotting eigenvectors

For some matrices, one can find eigenvectors "by inspection." Recognizing these cases is a useful skill.

Example. Consider the symmetric **outer product**: $\boldsymbol{A} = \boldsymbol{z}\boldsymbol{z}'$.
Observe that

$$\boldsymbol{A}\boldsymbol{z} = (\boldsymbol{z}\boldsymbol{z}')\boldsymbol{z} = \boldsymbol{z}(\boldsymbol{z}'\boldsymbol{z}) = \underbrace{(\boldsymbol{z}'\boldsymbol{z})}_{\text{scalar}}\,\boldsymbol{z}.$$

Thus $\boldsymbol{z}$ is an eigenvector of $\boldsymbol{A}$ with eigenvalue that is the norm squared of $\boldsymbol{z}$: $\lambda = \boldsymbol{z}'\boldsymbol{z} = \|\boldsymbol{z}\|_2^2$.

This symmetric matrix has rank 1 and it has one nonzero eigenvalue. HW

Example. Consider the $N \times N$ matrix with $N = 2n$ where

$$A = \left[\begin{array}{ccc|ccc} 1 & \dots & 1 & & & \\ \vdots & & \vdots & & \mathbf{0} & \\ 1 & \dots & 1 & & & \\ \hline & & & 9 & \dots & 9 \\ & \mathbf{0} & & \vdots & & \vdots \\ & & & 9 & \dots & 9 \end{array}\right] = \left[\begin{array}{c} \mathbf{1}_n \\ \hline \mathbf{0}_n \end{array}\right] \left[\begin{array}{c|c} \mathbf{1}_n' & \mathbf{0}_n' \end{array}\right] + \left[\begin{array}{c} \mathbf{0}_n \\ \hline 3\mathbf{1}_n \end{array}\right] \left[\begin{array}{c|c} \mathbf{0}_n' & 3\mathbf{1}_n' \end{array}\right].$$

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

$$u_1 = \left[\begin{array}{c} \frac{1}{\sqrt{n}}\mathbf{1}_n \\ \hline \mathbf{0}_n \end{array}\right], \quad u_2 = \left[\begin{array}{c} \mathbf{0}_n \\ \hline \frac{1}{\sqrt{n}}\mathbf{1}_n \end{array}\right].$$

25. What is the corresponding eigenvalue λ_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? $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \boldsymbol{e}_1 \boldsymbol{e}_2'$ 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 $\boldsymbol{A} = \boldsymbol{b}\boldsymbol{c}'$ with $\boldsymbol{b} \neq \boldsymbol{0}_M$ and $\boldsymbol{c} \neq \boldsymbol{0}_N$. Clearly:

$$\boldsymbol{A} = \boldsymbol{b}\,\boldsymbol{c}' =$$

where $\boldsymbol{U}_0$ is a $M \times (M-1)$ matrix with orthonormal columns that span the subspace $\mathsf{span}^\perp(\boldsymbol{b})$ and $\boldsymbol{V}_0$ is a $N \times (N-1)$ matrix with orthonormal columns that span the subspace $\mathsf{span}^\perp(\boldsymbol{c})$.

Is it unique? No, because we could use $-\boldsymbol{b}$ and $-\boldsymbol{c}$, and there are many choices for $\boldsymbol{U}_0$ and $\boldsymbol{V}_0$.

If $\boldsymbol{A}$ is **tall** $(M > N)$, then the economy SVD would be similar to the full SVD except `U` involves only the first $N-1$ columns of $\boldsymbol{U}_0$ and the inner matrix would be $N \times N$ and all zeros except for σ_1 in the upper left.

Matrix-vector products and the SVD (Read)

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'\boldsymbol{x} \\ &= \boldsymbol{U}\boldsymbol{\Sigma}\underbrace{(\boldsymbol{V}'\boldsymbol{x})}_{\hookrightarrow \text{ coordinates of } \boldsymbol{x} \text{ in term of basis } \boldsymbol{V}} \end{aligned}$$

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

$$\boldsymbol{A}\boldsymbol{x} = \sum_{k=1}^{r} \underbrace{\sigma_k}_{\text{gain}} \underbrace{\alpha_k}_{\text{coor.}} \underbrace{\boldsymbol{u}_k}_{\hookrightarrow \text{ left singular vector (basis vector)}}.$$

In particular, if $\boldsymbol{A} \in \mathbb{F}^{M \times N}$ and $\boldsymbol{x} = \boldsymbol{v}_n$, the nth right singular vector, for some $n \in \{1, \ldots, N\}$, then

$$\boldsymbol{\alpha} = \boldsymbol{V}'\boldsymbol{x} = \boldsymbol{e}_n \Longrightarrow \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{v}_n = \sigma_n \boldsymbol{u}_n.$$

Warm-up question(s)

26. If $M \times N$ matrix $\boldsymbol{A}$ has SVD $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}'$, then $\mathcal{R}(\boldsymbol{A}) = \mathcal{R}(\boldsymbol{U})$.

A: True B: False $\boxed{??}$

3.7 Application: Signal classification by nearest subspace

In **classification** problems we have a test vector $\boldsymbol{v} \in \mathcal{V}$ that we want to classify. One classification method define subsets $\mathcal{S}_1, \ldots, \mathcal{S}_C$ of $\mathcal{V}$ for each of the C classes (usually learned from labeled training data), and then classifies $\boldsymbol{v}$ by finding the closest subset. (This relates to the **k-nearest neighbors algorithm** for $k = 1$.)

Projection onto a set

The process of "finding the closest subset" involves first finding the nearest point to $\boldsymbol{v}$ in each set.

Define. The nearest point to $\boldsymbol{v} \in \mathcal{V}$ in a set $\mathcal{S} \subseteq \mathcal{V}$, also called the **projection** of $\boldsymbol{v}$ onto $\mathcal{S}$, is

$$\hat{\boldsymbol{v}} = \mathcal{P}_{\mathcal{S}}(\boldsymbol{v}) \triangleq \qquad (3.15)$$

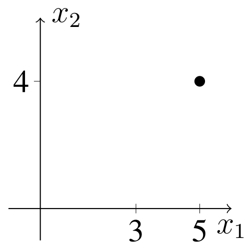
(By default we use the 2-norm here, though one could also use other norms.)
(Usually we focus on convex sets; more on that in later chapters.)

Having defined the **projection** operation $\mathcal{P}_{\mathcal{S}}$, the mathematical expression for the nearest-subset classifier is

$$\hat{c} = \underset{c \in \{1,\ldots,C\}}{\arg\min} \qquad (3.16)$$

27. Let $\mathcal{V} = \mathbb{R}^2$ and $\mathcal{S} = \{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1 = 3\}$. Determine $\mathcal{P}_{\mathcal{S}}(\boldsymbol{v})$ when $\boldsymbol{v} = (5, 4)$.

A: $(3, 3)$ B: $(3, 4)$ C: $(4, 4)$ D: $(5, 4)$ E: $(5, 3)$ **??**



Properties of a projection operation

- For a subset $\mathcal{S} \subseteq \mathcal{V}$, we have $\mathcal{P}_{\mathcal{S}} : \mathcal{V} \mapsto \mathcal{S}$.
- $\mathcal{P}_{\mathcal{S}}(\mathcal{P}_{\mathcal{S}}(\boldsymbol{v})) = \mathcal{P}_{\mathcal{S}}(\boldsymbol{v}), \ \forall \boldsymbol{v} \in \mathcal{V}$, *i.e.*, $\mathcal{P}_{\mathcal{S}}$ is an **idempotent** operation:

$$\mathcal{P}_{\mathcal{S}} \circ \mathcal{P}_{\mathcal{S}} = \mathcal{P}_{\mathcal{S}}. \tag{3.17}$$

In fact this second property is the defining property of a **projection**.

Example. For the above example, $\mathcal{P}_{\mathcal{S}}$

Nearest point in a subspace

Often we model each class using a set $\mathcal{S}_c$ that is a **subspace** of $\mathcal{V}$. Thus for classification we need to find the point in a subspace $\mathcal{S} \subseteq \mathcal{V}$ that is nearest to a test vector $\boldsymbol{v} \in \mathcal{V}$.

The **projection** onto a **subspace** is a **linear operation**. (See proof in Ch. 4.)

Specifically, when $\mathcal{S}$ is a subspace of N-dimensional vector space $\mathcal{V} = \mathbb{F}^N$, then there is a $N \times N$ matrix $\boldsymbol{P}_{\mathcal{S}}$ such that:

$$\mathcal{P}_{\mathcal{S}}(\boldsymbol{v}) = \qquad, \qquad \forall \boldsymbol{v} \in \mathcal{V}.$$

Note the difference in notation and meaning between the projection operation $\mathcal{P}$ and the projection matrix $\boldsymbol{P}$.

Because $\mathcal{P}_{\mathcal{S}}$ is an **idempotent** operation in general, per (3.17), when $\mathcal{S}$ is a subspace and the projection matrix $\boldsymbol{P}_{\mathcal{S}}$ is an **idempotent** matrix: HW

$$\forall \boldsymbol{v} \in \mathcal{V} : \mathcal{P}_{\mathcal{S}}(\mathcal{P}_{\mathcal{S}}(\boldsymbol{v})) = \mathcal{P}_{\mathcal{S}}(\boldsymbol{v}) \Longrightarrow \forall \boldsymbol{v} \in \mathcal{V} : \boldsymbol{P}_{\mathcal{S}}^2 \boldsymbol{v} = \boldsymbol{P}_{\mathcal{S}} \boldsymbol{v} \Longrightarrow \boldsymbol{P}_{\mathcal{S}}^2 = \boldsymbol{P}_{\mathcal{S}}.$$

To be specific, if $\mathcal{S}$ is a K-dimensional subspace in a N-dimensional vector space $\mathcal{V}$, and if $\boldsymbol{B}$ is a $N \times K$ matrix whose K columns form a **basis** for $\mathcal{S}$, then the $N \times N$ projection matrix is

$$\boldsymbol{P}_{\mathcal{S}} = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'.$$

Instead of proving this equality, we focus on orthonormal bases that avoid needing any matrix inverse.

Subspace with orthonormal basis

The process of finding the nearest point in a subspace is easy when we have an **orthonormal basis** for it.

Suppose $\boldsymbol{U} \in \mathbb{F}^{M \times K}$ has orthonormal columns, *i.e.*, $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{I}_K$, and define the subspace $\mathcal{S} = \mathcal{R}(\boldsymbol{U})$, for which $\boldsymbol{U}$ is an orthonormal basis.
Claim:

$$\hat{\boldsymbol{v}} = \mathcal{P}_{\mathcal{S}}(\boldsymbol{v}) = \mathcal{P}_{\mathcal{R}(\boldsymbol{U})}(\boldsymbol{v}) = \qquad (3.18)$$

Because of this property, we define the $M \times M$ **projection matrix** $\boldsymbol{P}_{\mathcal{S}} \triangleq \boldsymbol{U}\boldsymbol{U}'$, so that $\mathcal{P}_{\mathcal{R}(\boldsymbol{U})}(\boldsymbol{v}) = \boldsymbol{P}_{\mathcal{S}}\boldsymbol{v}$.

Proof. (Read)
Clearly $\hat{\boldsymbol{v}} \in \mathcal{S}$ by definition. Now consider any other point $\boldsymbol{s} \in \mathcal{S} = \mathcal{R}(\boldsymbol{U}) = \boldsymbol{U}\boldsymbol{w}$, for some coefficients $\boldsymbol{w} \in \mathbb{F}^K$. We can define $\boldsymbol{e} \triangleq \boldsymbol{s} - \hat{\boldsymbol{v}} \in \mathcal{R}(\boldsymbol{U}) \Longrightarrow \boldsymbol{e} = \boldsymbol{U}\boldsymbol{z}$ for some $\boldsymbol{z} \in \mathbb{F}^K$. Now

$$\begin{aligned}
\|\boldsymbol{v} - \boldsymbol{s}\|_2^2 = \|\boldsymbol{v} - (\hat{\boldsymbol{v}} + \boldsymbol{e})\|_2^2 = \|(\boldsymbol{I} - \boldsymbol{U}\boldsymbol{U}')\boldsymbol{v} - \boldsymbol{U}\boldsymbol{z}\|_2^2 &= \|(\boldsymbol{I} - \boldsymbol{U}\boldsymbol{U}')\boldsymbol{v}\|_2^2 + \|\boldsymbol{U}\boldsymbol{z}\|_2^2 \\
&\geq \|(\boldsymbol{I} - \boldsymbol{U}\boldsymbol{U}')\boldsymbol{v}\|_2^2 = \|\boldsymbol{v} - \hat{\boldsymbol{v}}\|_2^2,
\end{aligned}$$

because $(\boldsymbol{I} - \boldsymbol{U}\boldsymbol{U}')\boldsymbol{v} \perp \boldsymbol{U}\boldsymbol{z}$, showing that $\hat{\boldsymbol{v}}$ minimizes the distance. □

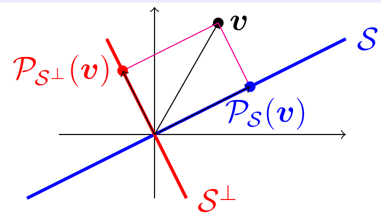
Ch. 4 further discusses the **orthogonality principle**: $(\boldsymbol{I} - \boldsymbol{U}\boldsymbol{U}')\boldsymbol{v} \perp \boldsymbol{U}\boldsymbol{z}\, \forall \boldsymbol{z} \implies \boldsymbol{v} - \mathcal{P}_{\mathcal{S}}(\boldsymbol{v}) \perp \mathcal{S}$.

Fact. If $\mathcal{S}$ is a **subspace** in vector space $\mathcal{V}$, then

$$\mathcal{P}_{\mathcal{S}^\perp} = \mathcal{I} - \mathcal{P}_{\mathcal{S}}, \qquad \textit{i.e.}, \qquad \mathcal{P}_{\mathcal{S}^\perp}(\boldsymbol{v}) = \boldsymbol{v} - \mathcal{P}_{\mathcal{S}}(\boldsymbol{v}),\ \forall \boldsymbol{v} \in \mathcal{V}, \tag{3.19}$$

where $\mathcal{I}$ denotes the **identity function** on $\mathcal{V}$, *i.e.*, $\mathcal{I}(\boldsymbol{v}) = \boldsymbol{v},\ \forall \boldsymbol{v} \in \mathcal{V}$.
Put another way related to (3.6) and (3.7):

$$\boldsymbol{v} = \mathcal{P}_{\mathcal{S}}(\boldsymbol{v}) + \mathcal{P}_{\mathcal{S}^\perp}(\boldsymbol{v}).$$



In particular, if $\mathcal{S}$ has **orthonormal basis** $\boldsymbol{U}$, then using (3.18):

$$\mathcal{P}_{\mathcal{S}^\perp}(\boldsymbol{v}) = \boldsymbol{v} - \mathcal{P}_{\mathcal{S}}(\boldsymbol{v}) = \boldsymbol{v} - \boldsymbol{P}_{\mathcal{S}}\boldsymbol{v} = \boldsymbol{v} - \boldsymbol{U}\boldsymbol{U}'\boldsymbol{v} = \underbrace{(\boldsymbol{I} - \boldsymbol{U}\boldsymbol{U}')}_{\boldsymbol{P}_{\mathcal{S}^\perp}}\boldsymbol{v}.$$

In this case we define the matrix analogue of (3.19) to be:

$$\boldsymbol{P}_{\mathcal{S}}^{\perp} \triangleq \boldsymbol{P}_{\mathcal{S}^\perp} = \boldsymbol{I} - \boldsymbol{P}_{\mathcal{S}} = \boldsymbol{I} - \boldsymbol{U}\boldsymbol{U}', \tag{3.20}$$

which is four notations for the projection matrix for the orthogonal complement of a subspace having orthonormal basis $\boldsymbol{U}$.

Importantly, using (3.20): $\boldsymbol{P}_{\mathcal{S}}^{\perp}\boldsymbol{P}_{\mathcal{S}} = (\boldsymbol{I} - \boldsymbol{U}\boldsymbol{U}')(\boldsymbol{U}\boldsymbol{U}') = \boldsymbol{U}\boldsymbol{U}' - \boldsymbol{U}\boldsymbol{U}'\boldsymbol{U}\boldsymbol{U}' = \boldsymbol{U}\boldsymbol{U}' - \boldsymbol{U}\boldsymbol{U}' = \boldsymbol{0}$.

Exercise. Verify that $\boldsymbol{P}_{\mathcal{S}}$ and $\boldsymbol{P}_{\mathcal{S}}^{\perp}$ are **orthogonal projection** matrices, as discussed in Ch. 4:

- $\boldsymbol{P}_{\mathcal{S}}^2 = \boldsymbol{P}_{\mathcal{S}}$
- $\left(\boldsymbol{P}_{\mathcal{S}}^{\perp}\right)^2 = \boldsymbol{P}_{\mathcal{S}}^{\perp}$
- $\boldsymbol{P}_{\mathcal{S}}' = \boldsymbol{P}_{\mathcal{S}}$ and $\left(\boldsymbol{P}_{\mathcal{S}}^{\perp}\right)' = \boldsymbol{P}_{\mathcal{S}}^{\perp}$
- $\boldsymbol{P}_{\mathcal{S}}\boldsymbol{P}_{\mathcal{S}}^{\perp} = \boldsymbol{0}$.

Machine learning application: Signal classification by nearest subspace

Consider an application like handwritten digit recognition where we have $C = 10$ classes of digits, and training data for each class. One method for classification (that will be explored in a discussion task) is to determine subspaces $\mathcal{S}_1, \ldots, \mathcal{S}_C$ for each class using the SVD methods detailed in Ch. 6 (related to PCA) with corresponding orthonormal bases $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_C$ for each class. Then we perform nearest-subspace classification of test data $\boldsymbol{v}$ by applying (3.16) with (3.18) as follows:

$$\hat{c} = \underset{c \in \{1,\ldots,C\}}{\arg\min}\, d_c(\boldsymbol{v}), \quad d_c(\boldsymbol{v}) = $$

This classification method is quite easy to implement.

28. If $\mathcal{V} = \mathbb{R}^M$ and each orthonormal basis matrix $\boldsymbol{U}_c$ is $M \times K$, how many multiplies are needed to classify one test vector $\boldsymbol{v} \in \mathcal{V}$?

A: KM B: CKM C: $2CKM$ D: $2C(KM)^2$ E: $C(2K+1)M$ $\boxed{??}$

Optimization preview

The projection operation (3.15) is one example (of many in these notes) of an **optimization** problem. In generic notation, a typical optimization problem is

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x} \in \mathcal{X}}{\arg\min}\, f(\boldsymbol{x}),$$

where:
- $\mathcal{X}$ is a set of allowable values of $\boldsymbol{x}$, *e.g.*, $\mathbb{F}^N$,
- $f(\cdot)$ is a **cost function** that quantifies how "undesirable" any given argument $\boldsymbol{x}$ is, so we want to minimize it, and
- $\hat{\boldsymbol{x}}$ is the (or a) minimizing (best) value for $\boldsymbol{x}$.

Often the set $\mathcal{X}$ is **convex set** and often f is a **convex function**.

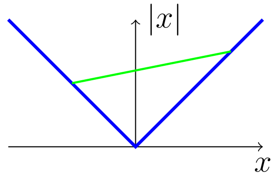
Define. A function $f : \mathcal{V} \mapsto \mathbb{R}$ is called a **convex function** on vector space $\mathcal{V}$ iff for every $\boldsymbol{x}, \boldsymbol{z} \in \mathcal{V}$, and every $\alpha \in [0, 1]$,

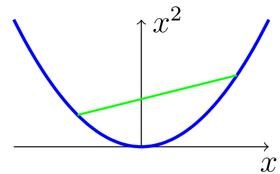
$$f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{z}) \leq \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{z}).$$

Define. A function $f : \mathcal{V} \mapsto \mathbb{R}$ is called a **strictly convex function** on vector space $\mathcal{V}$ iff for every $\boldsymbol{x}, \boldsymbol{z} \in \mathcal{V}$, and every $\alpha \in (0, 1)$,

$$f(\alpha \boldsymbol{x} + (1-\alpha)\boldsymbol{z}) < \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{z}).$$

Example. $f(x) = |x|$ is **convex** and $g(x) = x^2$ is **strictly convex**





Example. A HW problem shows that $f(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2$ and $\sigma_1(\boldsymbol{X}) : \mathbb{F}^{M \times N} \mapsto \mathbb{R}$ are convex functions.

29. The cost functions $\{(3.15), (3.16)\}$ are **convex functions** ?

A: F,F B: F,T C: T,F D: T,T $\boxed{\textbf{??}}$

Save this one for later after defining **convex set** in Ch. 4:

30. The allowable sets "$\mathcal{X}$" ((3.15), (3.16)) are **convex sets** ?

A: F,F B: F,T C: T,F D: T,T $\boxed{\textbf{??}}$

3.8 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 (and the orthogonal complements thereof) to components of its SVD like V_0 and U_r.
A concrete application is signal classification by nearest subspace methods.

A concept related to **rank** that arises in compressed sensing theory is the **spark** of a matrix. ♦♦

For a futuristic (?) demonstration of the importance of subspaces, see:
https://www.youtube.com/watch?v=H4qkodI6rSM

Bibliography

[1] A. J. Laub. *Matrix analysis for scientists and engineers*. Soc. Indust. Appl. Math., 2005 (cit. on pp. 3.3, 3.4, 3.12, 3.13, 3.17, 3.19, 3.22, 3.23, 3.24, 3.29, 3.40, 3.41, 3.42, 3.44).

[2] I. Lankham, B. Nachtergaele, and A. Schilling. *Linear algebra as an introduction to abstract mathematics*. World Scientific, 2016 (cit. on p. 3.18).

[3] Y. M. Lu and M. N. Do. "Sampling signals from a union of subspaces [A new perspective for the extension of this theory]". In: *IEEE Sig. Proc. Mag.* 25.2 (Mar. 2008), 41–7 (cit. on p. 3.20).

[4] T. Blumensath. "Sampling and reconstructing signals from a union of linear subspaces". In: *IEEE Trans. Info. Theory* 57.7 (July 2011), 4660–71 (cit. on p. 3.20).

[5] R. Vidal. "Subspace clustering". In: *IEEE Sig. Proc. Mag.* 28.2 (Mar. 2011), 52–68 (cit. on p. 3.20).

[6] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig. "ESPIRiT-an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA". In: *Mag. Res. Med.* 71.3 (Mar. 2014), 990–1001 (cit. on p. 3.44).

[7] G. Strang. "Bringing the SVD to life". In: *SIAM News* 39.1 (Jan. 2006) (cit. on p. 3.51).