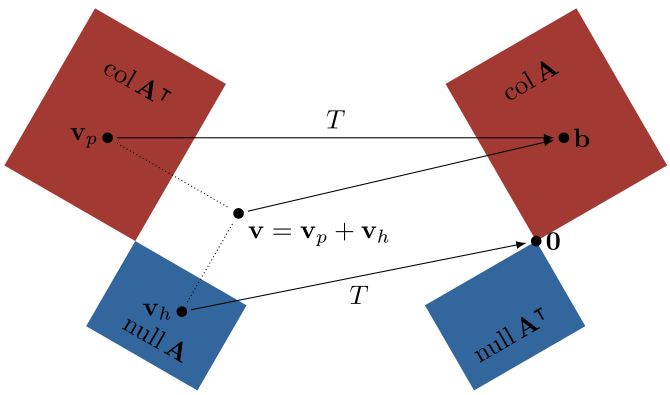
math 54 linear algebra and differential equations

university of california, berkeley



$T : \mathbf{x} \mapsto \mathbf{A}\mathbf{x}$

table of contents

1 Linear equations and transformations

1.1 Linear equations

Let's say we have a system of three variables:

$$x_1 + x_2 + 2x_3 = 4 \tag{1}$$
$$-x_1 + 2x_2 + x_3 = -1 \tag{2}$$
$$x_1 + 2x_2 + 3x_3 = -2 \tag{3}$$

and we want to find a solution x_1, x_2, x_3 to these three equations. We mathematicians are lazy, so we "abbreviate" this equation into an **augmented matrix**:

$$\left[\begin{array}{ccc|c} 1 & 1 & 2 & 4 \\ -1 & 2 & 1 & -1 \\ 1 & 2 & 3 & -2 \end{array}\right]$$

Just like what we can do with equations (1), (2), and (3), we can perform addition of two equations, multiply an equation by a scalar, and switch the order of these equations. We can do the same thing with rows:

Elementary row operations
- addition of a scalar multiple of one row to another ($R_1 \mapsto R_1 + \lambda R_2$)
- scalar multiplication of a row ($R \mapsto \lambda R$)
- interchange rows ($R_1 \leftrightarrow R_2$)

We call the first non-zero number in a row a **pivot**.

We say a matrix is in **row echelon form** if
- all pivots are to the right of the previous pivot ("staircase," hence *échelon*)
- any row of all zeroes are at the bottom

We say a matrix is in **reduced row echelon form** (rref) if
- it is in row echelon form
- all pivots are 1s
- the pivot-containing columns are all zero except for the pivot itself

Theorem Any $n \times k$ linear system will have either:

- no solution
- exactly one solution
- infinitely many solutions

The general strategy is:

- be sure the first entry of the first row is a pivot (interchange rows if you need to)
- make every number below the pivot a zero
- repeat for pivot of second row, pivot of third row, etc. $\implies$ echelon form
- row reduce to make every number above a pivot zero
- make every pivot a 1 $\implies$ reduced row echelon form

$$
\begin{tikzcd}
 & \text{system of linear equations} \arrow[r] & \text{augmented matrix} \arrow[d] & \\
 & & \text{(reduced) echelon form} \arrow[dll, "\text{pivot in output col.}"'] \arrow[d, "\text{no free columns}"] \arrow[dr, "\text{free columns}"] & \\
\text{0 solutions} & & \text{1 solution} & \infty \text{ solutions}
\end{tikzcd}
$$

1.2 Euclidean vectors ($\mathbb{R}^n$)

A **Euclidean vector** $\mathbf{v} \in \mathbb{R}^n$ is an ordered n-tuple of real numbers, and it can be represented by a $n \times 1$ array with real-valued entries:

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \langle v_1, v_2, \ldots, v_n \rangle = (v_1, v_2, \ldots, v_n)$$

The vector $(0, 0, \ldots, 0)$ is what we call the **zero vector**, and we denote this as $\mathbf{0}$. The set of all Euclidean vectors $\mathbb{R}^n$ is called the **n-dimensional Euclidean space**. We begin by citing some rudimentary definitions for vectors in $\mathbb{R}^n$.

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^m$. A **linear combination** is a vector of the form

$$\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_n \mathbf{v}_n \qquad (\text{where } \lambda_1, \ldots \lambda_n \in \mathbb{R})$$

Let $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$. The **span** is the set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_n$:

$$\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} = \{\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n \mid \lambda_1, \ldots \lambda_n \in \mathbb{R}\}$$

Any system of linear equations can again be represented as a coefficient matrix $\mathbf{A}$ multiplied by a vector $\mathbf{x}$ to give an output vector $\mathbf{b}$. If we let $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ and $\mathbf{b} = (b_1, \ldots, b_k)$,

$$\mathbf{A}\mathbf{x} = \mathbf{b} \iff \begin{bmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix} \iff x_1 \begin{bmatrix} | \\ \mathbf{a}_1 \\ | \end{bmatrix} + \cdots + x_n \begin{bmatrix} | \\ \mathbf{a}_n \\ | \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix}$$

1.3 Existence and uniqueness of solutions

From the definition of $\mathbf{A}\mathbf{x} = \mathbf{b}$, we can state the following properties:

- If there does exist a solution $\mathbf{x}$ for $\mathbf{A}\mathbf{x} = \mathbf{b}$, then $\mathbf{b} \in \operatorname{span}\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\}$.
- If there is no solution $\mathbf{x}$ for $\mathbf{A}\mathbf{x} = \mathbf{b}$, then $\mathbf{b} \notin \operatorname{span}\{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n\}$.

If we express this linear system of equations by an augmented matrix, we get the following equivalence:

$$\mathbf{A}\mathbf{x} = \mathbf{b} \iff \begin{bmatrix} \mathbf{A} \mid \mathbf{b} \end{bmatrix}$$

... and all of the same properties of an augmented matrix hold.

Tying this all together, we have the following theorem:

Let $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$ be an $m \times n$ matrix. Then, the following statements are logically equivalent:

- $\mathbf{A}\mathbf{x} = \mathbf{b}$ is consistent
- reduced $\begin{bmatrix} \mathbf{A} \mid \mathbf{b} \end{bmatrix}$ has no pivots in the output column
- $\exists \lambda_1, \ldots, \lambda_n \in \mathbb{R}^n$ such that $\mathbf{b} = \lambda_1 \mathbf{a}_1 + \cdots + \lambda_n \mathbf{a}_n$
- $\mathbf{b} \in \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$

If reduced $\mathbf{A}$ has a pivot in every row, then the augmented matrix $\begin{bmatrix} \mathbf{A} \mid \mathbf{b} \end{bmatrix}$ is *always* consistent for every $\mathbf{b} \in \mathbb{R}^m$. This means that for all $\mathbf{b} \in \mathbb{R}^m$, it is guaranteed that $\mathbf{b} \in \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$. Thus we have the following consequence:

$$\text{reduced } \mathbf{A} \text{ has a pivot in every row} \iff \mathbf{b} \in \operatorname{span}\{\mathbf{a}_1 \dots, \mathbf{a}_n\}$$

This ensures the *existence* of a solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Recall that a system of linear equations can either have no solutions, a unique solution, or infinitely many. How do we differentiate between having a unique solution or infinitely many?

If we think computationally, if the matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ has a unique solution, then the echelon form of $\mathbf{A}$ must have no free columns (i.e. a pivot in every column):

$$\begin{bmatrix} \blacksquare & * & * \\ 0 & \blacksquare & * \\ 0 & 0 & \blacksquare \\ 0 & 0 & 0 \end{bmatrix}$$

Let $\mathbf{A}$ be an $m \times n$ matrix.

$$\mathbf{A}\mathbf{x} = \mathbf{b} \text{ has a unique solution} \iff \text{there is a pivot in every column of reduced } \mathbf{A}$$

Let's start thinking of the case where $\mathbf{A}\mathbf{x} = \mathbf{0}$, such that the equation is always consistent for any matrix $\mathbf{A}$.

$\mathbf{A}\mathbf{x} = \mathbf{0}$ is called a **homogeneous equation**. We call $\mathbf{x} = \mathbf{0}$ the **trivial solution**.

Then we can note that if the reduced form of a matrix $\mathbf{A} = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix}$ has a pivot in every column, then $\mathbf{A}\mathbf{x} = \mathbf{0}$ has the unique solution $\mathbf{x} = \mathbf{0}$. If we let $\mathbf{x} = \begin{bmatrix}x_1 & \cdots & x_n\end{bmatrix}^\mathsf{T}$, then $x_1 = \cdots = x_n = 0$. This means that the linear system $\mathbf{a}_1 x_1 + \cdots + \mathbf{a}_n x_n = \mathbf{0}$ is true only by the trivial solution.

We call this property a special name: linear independence.

A set of vectors $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\} \subseteq \mathbb{R}^k$ is **linearly independent** if and only if $\lambda_1\mathbf{a}_1 + \lambda_2\mathbf{a}_2 + \cdots \lambda_n + \mathbf{a}_n = \mathbf{0}$ is true only when $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$ (unique solution).

Likewise, we call vectors that are *not* linearly independent a linearly dependent set instead:

A set of vectors $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\} \subseteq \mathbb{R}^k$ is **linearly dependent** if and only if $\exists \lambda_i \in \mathbb{R} : 1 \le i \le n$ such that $\lambda_i \neq 0$ and $\lambda_1\mathbf{a}_1 + \lambda_2\mathbf{a}_2 + \cdots + \lambda_n\mathbf{a}_n = \mathbf{0}$.

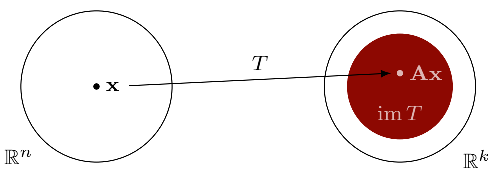
and immediately we have the following consequence:

$$\mathbf{A}\mathbf{x} = \mathbf{b} \text{ has a unique solution} \iff \text{reduced } \mathbf{A} \text{ has a pivot in every column} \iff \{\mathbf{a}_1, \dots, \mathbf{a}_n\} \text{ is linearly independent}$$

We have determined the *uniqueness* of a solution to a linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$.

1.4 Linear transformations $T : \mathbb{R}^n \to \mathbb{R}^k$

Let $\mathbf{x} \in \mathbb{R}^n$. Multiplying $\mathbf{v}$ by a $k \times n$ matrix $\mathbf{A}$ results in a function T between $\mathbb{R}^n$ and $\mathbb{R}^k$:



We call the function T a linear transformation from the domain of $\mathbb{R}^n$ to the codomain of $\mathbb{R}^k$.

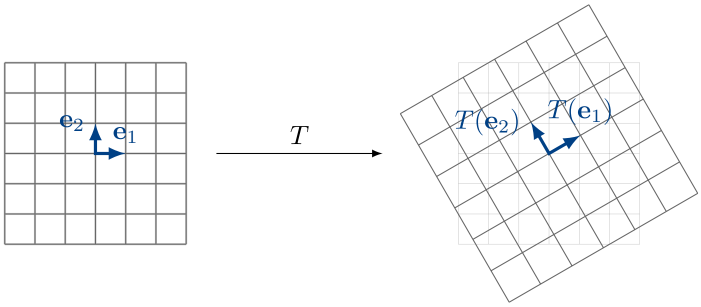
A function $T : \mathbb{R}^n \to \mathbb{R}^k$ is a **linear transformation** if and only if it satisfies the following conditions:

- $\forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^n : T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$
- $\forall \mathbf{u} \in \mathbb{R}^n : \lambda \in \mathbb{R} : T(\lambda \mathbf{u}) = \lambda T(\mathbf{u})$

Let $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ be a $k \times n$ matrix. Notice that the function $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$ is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^k$ because, for any $\mathbf{u} = (u_1, u_2, \dots, u_n) \in \mathbb{R}^n$ and $\mathbf{v} = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$,

- $\mathbf{A}(\mathbf{u} + \mathbf{v}) = \begin{bmatrix} \mathbf{a}_1(u_1 + v_1) & \cdots & \mathbf{a}_n(u_n + v_n) \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 u_1 + \mathbf{a}_1 v_1 & \cdots & \mathbf{a}_n u_n + \mathbf{a}_n v_n \end{bmatrix}$
  $= \begin{bmatrix} \mathbf{a}_1 u_1 & \cdots & \mathbf{a}_n u_n \end{bmatrix} + \begin{bmatrix} \mathbf{a}_1 v_1 & \cdots & \mathbf{a}_n v_n \end{bmatrix} = \mathbf{A}\mathbf{u} + \mathbf{A}\mathbf{v}$
- $\forall \lambda \in \mathbb{R} \quad \mathbf{A}(\lambda \mathbf{u}) = \begin{bmatrix} \mathbf{a}_1 \lambda u_1 & \cdots & \mathbf{a}_n \lambda u_n \end{bmatrix} = \lambda \begin{bmatrix} \mathbf{a}_1 u_1 & \cdots & \mathbf{a}_n u_n \end{bmatrix} = \lambda \mathbf{A}\mathbf{u}$

Using this, we can determine that every matrix transformation is immediately a linear transformation. Likewise, every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^k$ is a transformation that can be represented by a matrix $\mathbf{A}_T$. The key to finding $\mathbf{A}_T$ is to observe what T maps to the standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$ of $\mathbb{R}^n$.



Let $T : \mathbb{R}^n \to \mathbb{R}^k$ be a linear transformation. The **standard matrix** $\mathbf{A}_T$ of T is the matrix of the form

$$\mathbf{A}_T = \begin{bmatrix} T(\mathbf{e}_1) & T(\mathbf{e}_2) & \cdots & T(\mathbf{e}_n) \end{bmatrix}$$

We'll introduce two new concepts to characterize some "special" functions, in motivation to apply these concepts to linear transformations.

Let A and B be two sets. A function $f : A \to B$ is **one-to-one** if and only if (both equivalent definitions)

- **def 1** for all $b \in B$, there exists at most one $a \in A$ such that $f(a) = b$
- **def 2** $\forall a_1, a_2 \in A,\ f(a_1) = f(a_2) \implies a_1 = a_2$

Let A and B be two sets. A function $f : A \to B$ is **onto** if and only if

- **def 1** $\forall b \in B, \exists a \in A$ such that $f(a) = b$
- **def 2** $\operatorname{im} f = B$

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^k$ is one-to-one if and only if $T(\mathbf{x}) = \mathbf{0}$ has the unique solution $\mathbf{x} = \mathbf{0}$.

proof Assume T is not one-to-one. Then, $\exists \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$ where $\mathbf{x}_1 \neq \mathbf{x}_2$ such that $T(\mathbf{x}_1) = T(\mathbf{x}_2)$. Using the linearity of T, $T(\mathbf{x}_1) - T(\mathbf{x}_2) = T(\mathbf{x}_1 - \mathbf{x}_2) = \mathbf{0}$. Since $\mathbf{x}_1 - \mathbf{x}_2 \neq 0$, $T(\mathbf{x}) = \mathbf{0}$ has more than one solution.

Now assume that $T(\mathbf{x}) = \mathbf{0}$ has more than one solution. Then, T is immediately not one-to-one by definition. ■

Let $T : \mathbb{R}^n \to \mathbb{R}^k$ be a linear transformation, and $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$ be the matrix for T.

- T is onto if and only if $\operatorname{span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\} = \mathbb{R}^k$
- T is one-to-one if and only if $\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}$ is linearly independent

proof If T is onto, then $\forall \mathbf{y} \in \mathbb{R}^k$, $\exists \mathbf{x} \in \mathbb{R}^n$ such that $T(\mathbf{x}) = \mathbf{A}\mathbf{x} = \mathbf{y}$. Thus, $\exists \lambda_1, \dots, \lambda_n \in \mathbb{R}^n$ such that $\mathbf{y} = \lambda_1 \mathbf{a}_1 + \cdots + \lambda_n \mathbf{a}_n \in \operatorname{span}\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}$. Thus, $\mathbb{R}^k \subseteq \operatorname{span}\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}$. Since, naturally, $\operatorname{span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\} \subseteq \mathbb{R}^k$, then it must be concluded that $\mathbb{R}^k = \operatorname{span}\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$.

If T is one-to-one, then $T(\mathbf{x}) = \mathbf{0}$ if and only if $\mathbf{x} = \mathbf{0}$. Thus $\mathbf{A}\mathbf{x} = \mathbf{0}$, or equivalently, $\lambda_1 \mathbf{a}_1 + \cdots + \lambda_n \mathbf{a}_n = \mathbf{0}$, has the trivial solution $\lambda_1 = \cdots = \lambda_n = 0$. By definition, $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ is linearly independent. ■

To connect between a linear transformation T and linear independence, we bring up the following definition:

Let $T : \mathbb{R}^k \to \mathbb{R}^n : \mathbf{x} \mapsto \mathbf{Ax}$ be a linear transformation. The **kernel** of T is the set of all vectors $\mathbf{x} \in \mathbb{R}^k$ that map to the zero vector through T. We also call it the **null space** of $\mathbf{A}$.

$$\ker T = \operatorname{null} \mathbf{A} = \{\mathbf{x} \in \mathbb{R}^k \mid T(\mathbf{x}) = \mathbf{Ax} = \mathbf{0}\}$$

So, a "shorthand way" to say that a linear transformation where $T(\mathbf{x}) = \mathbf{0}$ is to say that $\ker T = \{\mathbf{0}\}$. Using this concept, we can present the following system of equivalences:

Let $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$ be a $k \times n$ matrix and $T : \mathbb{R}^n \to \mathbb{R}^k$ via $\mathbf{x} \mapsto \mathbf{Ax}$ be a linear transformation. Then, these statements are all equivalent:

- T is injective (one-to-one)
- $\mathbf{A}$ has a pivot in every column
- $\ker T = \{\mathbf{0}\}$
- $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ is linearly independent

What does a general solution $\mathbf{x}$ to $\mathbf{Ax} = \mathbf{b}$ look like?

Let $\mathbf{A}$ be a $k \times n$ matrix. Then, the general solution $\mathbf{x}$ to $\mathbf{Ax} = \mathbf{b}$ is $\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$, where $\mathbf{Ax}_p = \mathbf{b}$ and $\mathbf{x}_h \in \operatorname{null} \mathbf{A}$.

proof Suppose that $\mathbf{Ax} = \mathbf{b}$ has two particular solutions $\mathbf{x}_p, \mathbf{x} \in \mathbb{R}^n$. Then, using the linearity of $\mathbf{x} \mapsto \mathbf{Ax}$,

$$\begin{aligned} \mathbf{Ax} = \mathbf{Ax}_p &= \mathbf{b} \\ \mathbf{Ax} - \mathbf{Ax}_p &= \mathbf{0} \\ \mathbf{A}\underbrace{(\mathbf{x} - \mathbf{x}_p)}_{\text{let } \mathbf{x}_h} &= \mathbf{0} \end{aligned}$$

Thus, $\mathbf{x}_h = \mathbf{x} - \mathbf{x}_p \implies \mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$, where $\mathbf{x}_h \in \operatorname{null} \mathbf{A}$. ■

1.5 Matrix algebra

In this section, we define three primary matrix operations:

- matrix addition $(\mathbf{A} + \mathbf{B})$
- matrix–scalar multiplication $(\lambda \mathbf{A})$
- matrix multiplication ($\mathbf{AB}$, *a big one*)

Matrix addition: If $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$, and $\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix}$, then $\mathbf{A} + \mathbf{B} := \begin{bmatrix} \mathbf{a}_1 + \mathbf{b}_1 & \cdots & \mathbf{a}_n + \mathbf{b}_n \end{bmatrix}$.

This definition of matrix addition works because, letting the matrix transformations of $\mathbf{A}$ and $\mathbf{B}$ be $T_\mathbf{A}$ and $T_\mathbf{B}$ respectively,

$$\forall \mathbf{x} \in \mathbb{R}^n \quad (T_\mathbf{A} + T_\mathbf{B})\mathbf{x} = (\mathbf{A} + \mathbf{B})\mathbf{x} = \mathbf{Ax} + \mathbf{Bx} = T_\mathbf{A}(\mathbf{x}) + T_\mathbf{B}(\mathbf{x})$$

Matrix–scalar multiplication: If $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$ and $\lambda \in \mathbb{R}$, then $\lambda \mathbf{A} := \begin{bmatrix} \lambda \mathbf{a}_1 & \cdots & \lambda \mathbf{a}_n \end{bmatrix}$.

Likewise, letting the matrix transformation of $\mathbf{A}$ be $T_\mathbf{A}$ and the scaled matrix transformation of $\lambda \mathbf{A}$ be $T_{\lambda \mathbf{A}}$,

$$\forall \mathbf{x} \in \mathbb{R}^n \quad T_{\lambda \mathbf{A}}(\mathbf{x}) = (\lambda \mathbf{A})\mathbf{x} = \lambda(\mathbf{Ax}) = \lambda T_\mathbf{A}(\mathbf{x})$$

Matrix multiplication: Let $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_p \end{bmatrix}$ be a $k \times p$ matrix and $\mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix}$ be a $p \times n$ matrix. Then, $\mathbf{AB} := \begin{bmatrix} \mathbf{Ab}_1 & \cdots & \mathbf{Ab}_n \end{bmatrix}$.

Here's why we define it this way. Letting the transformations of $\mathbf{A}$ and $\mathbf{B}$ be $T_\mathbf{A} : \mathbb{R}^p \to \mathbb{R}^k$ and $T_\mathbf{B} : \mathbb{R}^k \to \mathbb{R}^n$ respectively, then we define $T_{\mathbf{AB}}$ to be the composition of these two linear maps:

$$T_{\mathbf{AB}} := T_\mathbf{A} \circ T_\mathbf{B}$$

which is also a linear transformation from $\mathbb{R}^p \to \mathbb{R}^n$ and thus a matrix transformation. Then, to find the matrix encoding $T_{\mathbf{AB}}$, we simply find its *standard matrix*:

$$\begin{bmatrix}(T_{\mathbf{A}} \circ T_{\mathbf{B}})(\mathbf{e}_1) & \cdots & (T_{\mathbf{A}} \circ T_{\mathbf{B}})(\mathbf{e}_n)\end{bmatrix} = \begin{bmatrix}T_{\mathbf{A}}(\mathbf{b}_1) & \cdots & T_{\mathbf{A}}(\mathbf{b}_n)\end{bmatrix}$$

which results in the definition above.

The **transpose** $\mathbf{A}^\mathsf{T}$ of a $k \times n$ matrix $\mathbf{A}$ is the matrix with entries b_{ij} so that, for each entry a_{ij} in $\mathbf{A}$, $b_{ij} = a_{ji}$. Less confusingly,

$$\begin{bmatrix} a_{11} & \cdots & a_{1k} \\ a_{21} & \cdots & a_{2k} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mm} \end{bmatrix}^\mathsf{T} = \begin{bmatrix} a_{11} & \cdots & a_{m1} \\ a_{12} & \cdots & a_{m2} \\ \vdots & \ddots & \vdots \\ a_{1k} & \cdots & a_{mm} \end{bmatrix}$$

Properties:

- $\mathbf{AB} \neq \mathbf{BA}$
- $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
- $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$
- $\mathbf{I}_n\mathbf{A} = \mathbf{A}\mathbf{I}_n$
- $(\mathbf{A}^\mathsf{T})^\mathsf{T} = \mathbf{A}$
- $(\mathbf{A}+\mathbf{B})^\mathsf{T} = \mathbf{A}^\mathsf{T} + \mathbf{B}^\mathsf{T}$
- $(\mathbf{AB})^\mathsf{T} = \mathbf{B}^\mathsf{T}\mathbf{A}^\mathsf{T}$
- $\mathbf{AB} = \mathbf{0}$ but $\mathbf{A} \neq \mathbf{0}$ and $\mathbf{B} \neq \mathbf{0}$
- $\mathbf{AB} = \mathbf{AC}$ $(\mathbf{A} = \mathbf{0})$ implies $\mathbf{B} \neq \mathbf{C}$

1.6 Invertible matrices

Recall that, for a mapping $f : A \to B$,

- f is one-to-one if there is at most one input for each element in B
- f is onto if there is at least one input for each element in B

An $n \times n$ matrix $\mathbf{A}$ is **invertible** if and only if there is a matrix $\mathbf{A}^{-1}$ such that

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$$

What conditions does a linear mapping $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$ meet for $\mathbf{A}$ to be invertible?

Let $T_{\mathbf{A}}$ be such a linear mapping, and $T_{\mathbf{A}^{-1}}$ be the mapping $\mathbf{x} \mapsto \mathbf{A}^{-1}\mathbf{x}$. Then, by definition of invertible matrix, $\forall \mathbf{x} \in \mathbb{R}^n : T_{\mathbf{A}} \circ T_{\mathbf{A}^{-1}} = T_{\mathbf{A}^{-1}} \circ T_{\mathbf{A}} = I$, where $I : \mathbb{R}^n \to \mathbb{R}^n$ is the identity mapping. Thus, $T_{\mathbf{A}^{-1}} = (T_{\mathbf{A}})^{-1}$, meaning that $T_{\mathbf{A}}$ must be an invertible mapping.

Now, for $T_{\mathbf{A}}$ to be an invertible mapping, $T_{\mathbf{A}}$ must be:

- **one-to-one:** If $T_{\mathbf{A}}$ wasn't one-to-one, then $T_{\mathbf{A}^{-1}}$ is no longer a mapping (there exists some element in the output space that relates to two elements in the input space)
- **onto:** If $T_{\mathbf{A}}$ wasn't onto, then $T_{\mathbf{A}^{-1}}$ is not defined for all of its input space.

Thus, we have the following theorem:

Let $\mathbf{A} = \begin{bmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{bmatrix}$ be an $n \times n$ matrix with associated transformation T. Then, the following are all logically equivalent:

- $\mathbf{A}$ is an invertible matrix
- T is an invertible mapping
- T is one-to-one
  - $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is linearly independent
  - reduced $\mathbf{A}$ has a pivot in every column
- T is onto
  - $\operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\} = \mathbb{R}^n$
  - reduced $\mathbf{A}$ has a pivot in every row

How to find the inverse of an $n \times n$ matrix $\mathbf{A}$?

When $n = 2$,

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \implies \mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

When $n \geq 2$, simply obtain the reduced row echelon form:

$$\begin{bmatrix} \mathbf{A} \mid \mathbf{I}_n \end{bmatrix} \xrightarrow{\text{row reduction}} \begin{bmatrix} \mathbf{I}_n \mid \mathbf{A}^{-1} \end{bmatrix}$$

2 Determinants

The **determinant** of an $n \times n$ matrix $\mathbf{A} = \begin{bmatrix} a_{ij} \end{bmatrix}$ is defined for $n > 1$ as

$$\begin{aligned} \det \mathbf{A} &:= a_{11} \det \mathbf{A}_{11} - a_{12} \det \mathbf{A}_{12} + \cdots \pm a_{1n} \det \mathbf{A}_{1n} \\ &:= \sum_{k=1}^{n} (-1)^{k+1} a_{1k} \det \mathbf{A}_{1k} \end{aligned}$$

where $\mathbf{A}_{ij}$ is the $(i,j)^{\text{th}}$-cofactor of $\mathbf{A}$.

The $(i,j)^{\text{th}}$-cofactor of $\mathbf{A}$ is the matrix formed by deleting the i^{th} row and j^{th} column of $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 4 \\ 0 & 1 & 2 \end{bmatrix} \qquad \implies \qquad \mathbf{A}_{12} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 4 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 0 & 2 \end{bmatrix}$$

$\det \mathbf{A}$ can alternatively be computed as a cofactor expansion across the i^{th} row ...

$$\det \mathbf{A} = (-1)^{i+1} a_{i1} \mathbf{A}_{i1} + (-1)^{i+2} a_{i2} \mathbf{A}_{i2} + \cdots + (-1)^{i+n} a_{in} \mathbf{A}_{in}$$

or as a cofactor expansion down the j^{th} column ...

$$\det \mathbf{A} = (-1)^{j+1} a_{1j} \mathbf{A}_{1j} + (-1)^{j+2} a_{2j} \mathbf{A}_{2j} + \cdots + (-1)^{j+n} a_{nj} \mathbf{A}_{nj}$$

Note that the sign ordering $(-1)^{i+j}$ can be visualized as

$$\begin{bmatrix} + & - & + & \cdots \\ - & + & - & \\ + & - & + & \\ \vdots & & & \ddots \end{bmatrix}$$

2.1 Properties of determinants

Some useful properties:

- $\det \mathbf{AB} = \det \mathbf{A} \det \mathbf{B}$
- $\det \mathbf{A}^{\mathsf{T}} = \det \mathbf{A}$
- If $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & & \\ & & \ddots & \vdots \\ & & & a_{nn} \end{bmatrix}$, then $\det \mathbf{A} = a_{11} a_{22} \cdots a_{nn}$.

Let $\mathbf{A}$ be an $n \times n$ matrix and $\mathbf{E}$ be a matrix corresponding to elementary row operation e. Then,

$$\det \mathbf{EA} = \det \mathbf{E} \det \mathbf{A}$$

where

$$\det \mathbf{E} = \begin{cases} 1 & \text{if } e \text{ is a row replacement} \\ -1 & \text{if } e \text{ is a row interchange} \\ \lambda & \text{if } e \text{ is a row scaling by } \lambda \in \mathbb{R} \end{cases}$$

3 Vector spaces

3.1 Real vector spaces

An **real vector space** $(V, +, \cdot)$ is a nonempty set V such that it has the following properties under the binary operations of scalar addition $(+)$ and scalar multiplication $(\cdot)$:

$$\begin{array}{rll} \forall \mathbf{u}, \mathbf{v} \in V & \mathbf{u} + \mathbf{v} \in V & \text{closure under } + \\ \forall \mathbf{u}, \mathbf{v} \in V & \mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u} & \text{commutativity under } + \\ \forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V & (\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w}) & \text{associativity under } + \\ \exists \mathbf{0} \in V, \forall \mathbf{u} \in V & \mathbf{0} + \mathbf{u} = \mathbf{u} + \mathbf{0} = \mathbf{u} & \text{zero vector} \\ \forall \mathbf{u} \in V, \exists (-\mathbf{u}) \in V & \mathbf{u} + (-\mathbf{u}) = \mathbf{0} & \text{additive inverse} \\ \forall \mathbf{u} \in V, \forall \lambda \in \mathbb{R} & \lambda \mathbf{u} \in V & \text{closure under } \cdot \\ \forall \mathbf{u}, \mathbf{v} \in V, \forall \lambda \in \mathbb{R} & \lambda(\mathbf{u} + \mathbf{v}) = \lambda \mathbf{u} + \lambda \mathbf{v} & \text{distributivity I} \\ \forall \lambda, \mu \in \mathbb{R}, \forall \mathbf{u} \in V & (\lambda + \mu)\mathbf{u} = \lambda \mathbf{u} + \mu \mathbf{u} & \text{distributivity II} \\ \forall \lambda, \mu \in \mathbb{R}, \forall \mathbf{u} \in V & (\lambda \mu)\mathbf{u} = \lambda(\mu \mathbf{u}) & \text{associativity under } \cdot \\ \forall \mathbf{u} \in V & 1\mathbf{u} = \mathbf{u} & \text{multiplicative identity} \end{array}$$

The formal definition of real vector spaces further generalizes the properties of $\mathbb{R}^n$ to any eligible set. Such abstract sets must hold these properties such that a linear transformation can preserve vector addition and scalar multiplication, just like how a matrix transformation $\mathbf{x} \mapsto \mathbf{Ax}$ preserves addition and multiplication.

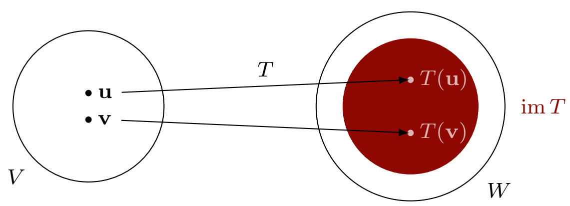
Some common examples of real vector spaces are

$\mathbb{R}^n$	the set of n-tuples of real numbers
$\mathcal{M}_{m \times n}(\mathbb{R})$	the set of $m \times n$ matrices with real number entries
$C(\mathbb{R})$	the set $\{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous on } \mathbb{R}\}$
$C^1(\mathbb{R})$	the set $\{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuously differentiable on } \mathbb{R}\}$
$\mathbb{P}(\mathbb{R})$	the set of all real-valued polynomials
$\mathbb{P}_d(\mathbb{R})$	the set of all real-valued polynomials with degree at most d
$\mathbb{R}^\infty$	the set of sequences of real numbers

We will generalize the concept of linear transformations for abstract vector spaces.

Let V and W be vector spaces. $T : V \to W$ is a **linear transformation** if and only if

$$\begin{array}{rl} \forall \mathbf{u}, \mathbf{v} \in V & T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \\ \forall \mathbf{u} \in V, \forall \lambda \in \mathbb{R} & T(\lambda \mathbf{u}) = \lambda T(\mathbf{u}) \end{array}$$



Let V be a real vector space. A proper subset $U \subset V$ is a **subspace of** V if it satisfies the following conditions:

$\exists \mathbf{0} \in U, \forall \mathbf{u} \in V$	$\mathbf{0} + \mathbf{u} = \mathbf{u} + \mathbf{0} = \mathbf{u}$	zero vector exists
$\forall \mathbf{u}, \mathbf{v} \in U$	$\mathbf{u} + \mathbf{v} \in U$	closed under addition
$\forall \mathbf{u} \in U, \forall \lambda \in \mathbb{R}$	$\lambda \mathbf{u} \in U$	closed under scalar multiplication

3.2 Kernel and range

Let V and W be finite-dimensional vector spaces, and $T : V \to W$ be a linear transformation.

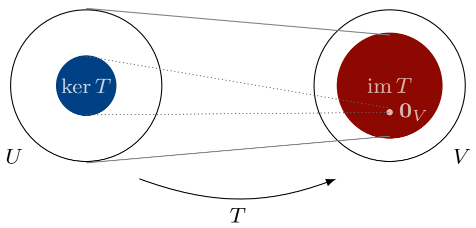
- $\ker T = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}_W\}$ (where $\mathbf{0}_W$ is the zero vector of W)
- $\operatorname{im} T = \{\mathbf{w} \in W \mid \exists \mathbf{v} \in V : T(\mathbf{v}) = \mathbf{w}\}$

If we define $T : \mathbb{R}^n \to \mathbb{R}^m$ to be a matrix transformation encoded by an $m \times n$ matrix $\mathbf{A}$, then we call

- $\operatorname{null} \mathbf{A} = \ker T$, the null space of $\mathbf{A}$
- $\operatorname{col} \mathbf{A} = \operatorname{im} T$, the column space of $\mathbf{A}$

Note that the column space of $\mathbf{A}$ is equal to the span of its columns: $\operatorname{col} \mathbf{A} = \operatorname{span}\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$.

$\ker T$ is a subspace of V, and $\operatorname{im} T$ is a subspace of W.



Let $\mathbf{w} \in W$ and $\exists \mathbf{v}_p \in V$ such that $T(\mathbf{v}_p) = \mathbf{w}$. Then, the general solution to $T(\mathbf{v}) = \mathbf{w}$ is

$$\mathbf{v} = \mathbf{v}_p + \mathbf{v}_h \qquad \text{where } \mathbf{v}_h \in \ker T$$

proof Since $T(\mathbf{v}_p) = \mathbf{w}$, then, for all $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$,

$$\begin{aligned} T(\mathbf{v}_p) - T(\mathbf{v}) &= \mathbf{w} - \mathbf{w} = \mathbf{0}_W \\ T(\mathbf{v}_p - \mathbf{v}) &= \mathbf{0}_W \\ \therefore \mathbf{v_p} - \mathbf{v} &\in \ker T \end{aligned}$$

Let $\mathbf{v}_h = \mathbf{v}_p - \mathbf{v}$. Then, $\mathbf{v}_h \in \ker T$, and $\mathbf{v} = \mathbf{v}_p + \mathbf{v}_h$. ■

A very clarifying example is when we consider the transformation $\dfrac{\mathrm{d}}{\mathrm{d}x} : \mathbb{C}(\mathbb{R}) \to \mathbb{C}(\mathbb{R})$. If a particular solution to the equation $\dfrac{\mathrm{d}}{\mathrm{d}x} \mathbf{f}_p(x) = 2x$ is $\mathbf{f}_p(x) = x^2$, then the general solution is $\mathbf{f}(x) = x^2 + C$ where $C \in \mathbb{R}$. This is because $C \in \ker \dfrac{\mathrm{d}}{\mathrm{d}x}$, as $\forall C \in \mathbb{R}$, $\dfrac{\mathrm{d}}{\mathrm{d}x} C = 0$, which is the zero vector of $\mathbb{C}(\mathbb{R})$.

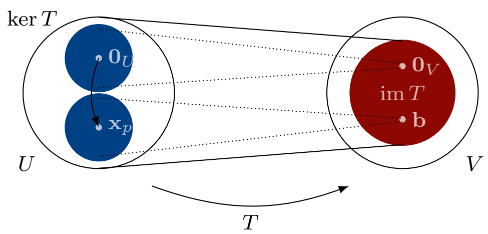


Figure 1: All inputs mapping to $\mathbf{b}$ is simply the vectors of $\ker T$ "translated" by $\mathbf{x}_p$.

3.3 Bases $\mathcal{B}$ and dimension

Recall that

- a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \subset V$ is linearly independent if and only if $\lambda_1 \mathbf{v}_1 + \cdots + \lambda_n \mathbf{v}_n = \mathbf{0}$ has the unique solution $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$.
- the span of $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is the set of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$.

and recall the following property:

$S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is linearly dependent if and only if $\exists \mathbf{v} \in S$ such that $\mathbf{v}$ is a linear combination of the other vectors in S.

We want to generalize results from the $\mathbb{R}^m$ vector space to any abstract vector space, such as:

- $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \in \mathbb{R}^m$ spans $\mathbb{R}^m \implies n \geq m$
- $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \in \mathbb{R}^m$ is linearly independent $\implies n \leq m$

which we derived from the "pivot" arguments specific to $\mathbb{R}^m$.

Right now, think of n as the "dimension" of the $\mathbb{R}^n$ vector space. We will formalize the concept of "dimension" for vector spaces later in this section, but we will introduce a new idea for now.

Let V be a vector space. A set of vectors $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots \mathbf{b}_n\}$ is a **finite basis for** V if it is linearly independent and spans V.

Note that a vector space has infinitely many bases. However, all bases for a vector space do share a common characteristic.

Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for a vector space V. If $B \subseteq V$ consists of more than n vectors, then B is linearly dependent.

proof Let $B = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$, where $p > n$. Thus, the matrix $\left[[\mathbf{u}_1]_{\mathcal{B}}, \ldots, [\mathbf{u}_p]_{\mathcal{B}}\right]$ has at most $n < p$ pivots, meaning that $\{[\mathbf{u}_1]_{\mathcal{B}}, \ldots, [\mathbf{u}_p]_{\mathcal{B}}\}$ is a linearly dependent set. Thus, $\exists \lambda_1, \lambda_2, \ldots, \lambda_p \in \mathbb{R}$ and $\exists i \in \mathbb{Z}$, where $1 \leq i \leq p$, such that $\lambda_i \neq 0$ and

$$\lambda_1[\mathbf{u}_1]_{\mathcal{B}} + \lambda_2[\mathbf{u}_2]_{\mathcal{B}} + \cdots + \lambda_p[\mathbf{u}_p]_{\mathcal{B}} = \mathbf{0}_{\mathbb{R}^n}$$

Since the transformation $[\,\cdot\,]_{\mathcal{B}}$ is linear,

$$\begin{aligned} &\left[\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 + \cdots + \lambda_p \mathbf{u}_p\right]_{\mathcal{B}} = \mathbf{0}_{\mathbb{R}^n} \\ \iff\ &\lambda_1 \mathbf{u}_1 + \lambda_2 \mathbf{u}_2 + \cdots + \lambda_p \mathbf{u}_p = 0\mathbf{b}_1 + \cdots + 0\mathbf{b}_n = \mathbf{0}_V \end{aligned}$$

with at least one $\lambda_i \neq 0$. Thus, $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is linearly dependent. ■

Every basis of a vector space must have the same number of vectors.

proof Let $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_k\}$ be two bases of a vector space V. Since $\mathcal{B}$ is a basis and $\mathcal{C}$ is linearly independent, then $k \leq n$. Likewise, since $\mathcal{C}$ is a basis and $\mathcal{B}$ is linearly independent, then $n \leq k$. Thus, $n = k$. ■

The **dimension** of a vector space V is the number of vectors in a basis $\mathcal{B}$ for V.

We have the following generalizations from $\mathbb{R}^n$ to abstract vector spaces:

- $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ linearly independent $\implies p \leq \dim V$
- $\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_p\} = V \implies p \geq \dim V$
- If $n = \dim V$, then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ linearly independent $\iff \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\} = V$
- $U \subseteq V$ is a subspace $\implies \dim U \leq \dim V$

3.4 Rank and nullity

Let V and W be finite-dimensional vector spaces, and $T : V \to W$ be a linear transformation. Then, the **rank** of T is the dimension of the image of T, and the **nullity** of T is the dimension of the kernel of T.

For an $m \times n$ matrix $\mathbf{A}$,

- $\operatorname{rank} T$ is the number of *pivot* columns in reduced $\mathbf{A}$
- $\operatorname{nullity} T$ is the number of *free* columns in reduced $\mathbf{A}$

Theorem (rank–nullity theorem) Let V and W be finite-dimensional vector spaces and $T : V \to W$ be a linear transformation. Then,

$$\operatorname{rank} T + \operatorname{nullity} T = \dim U$$

By the rank–nullity theorem, we have the following consequences. Suppose that U and V are finite-dimensional vector spaces. Then,

- T is onto $\implies \operatorname{rank} T = \dim V \iff \dim U \geq \dim V$
- T is one-to-one $\implies \operatorname{nullity} T = 0 \iff \operatorname{rank} T = \dim U \leq \dim V$
- T is onto and one-to-one $\implies \operatorname{rank} T = \dim V$, $\operatorname{nullity} T = 0 \implies \dim U = \dim V$

3.5 Bases and coordinate systems $[\,\cdot\,]_{\mathcal{B}}$

Recall that a finite basis $\mathcal{B}$ for an n-dimensional vector space is a set $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ that is linearly independent and spanning the vector space.

Let V be a finite-dimensional vector space, and $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ be a basis for V. Then, let $\mathbf{v} \in V$ such that $\exists \lambda_1, \ldots, \lambda_n \in \mathbb{R} : \mathbf{v} = \lambda_1 \mathbf{b}_1 + \lambda_2 \mathbf{b}_2 + \cdots \lambda_n \mathbf{b}_n$. The **$\mathcal{B}$-coordinate vector** of $\mathbf{v}$ is

$$[\,\mathbf{v}\,]_{\mathcal{B}} \coloneqq \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} \in \mathbb{R}^n$$

Every vector $\mathbf{v} \in V$ has a **unique $\mathcal{B}$-coordinate vector** $[\,\mathbf{v}\,]_{\mathcal{B}}$ because $\mathcal{B}$ is linearly independent. To clarify, let $\lambda_1, \ldots, \lambda_n, \mu_1, \cdots, \mu_n \in \mathbb{R}$ such that $\mathbf{v}$ has two representations:

$$\lambda_1 \mathbf{b}_1 + \lambda_2 \mathbf{b}_2 + \cdots \lambda_n \mathbf{b}_n \quad = \quad \mu_1 \mathbf{b}_1 + \mu_2 \mathbf{b}_2 + \cdots \mu_n \mathbf{b}_n$$

Thus, if we subtract the right-hand side from the left-hand side,

$$(\lambda_1 - \mu_1)\mathbf{b}_1 + (\lambda_2 - \mu_2)\mathbf{b}_2 + \cdots (\lambda_n - \mu_n)\mathbf{b}_n = \mathbf{0}_V$$

But, $\mathcal{B}$ is linearly independent. So, for $i \in \mathbb{Z}$ within $1 \leq i \leq n$, $\lambda_i = \mu_i$. Thus, $[\,\mathbf{v}\,]_{\mathcal{B}}$ is unique.

If $\mathcal{B} \subset \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^n$, we can compute $[\,\mathbf{v}\,]_{\mathcal{B}}$ by solving for the weights $\lambda_1, \cdots \lambda_n$ of $\lambda_1 \mathbf{b}_1 + \cdots + \lambda_n \mathbf{b}_n = \mathbf{0}$. This is equivalent to performing Gaussian elimination to the augmented matrix

$$\begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \mid \mathbf{v} \end{bmatrix} \xrightarrow{\text{row reduction}} \begin{bmatrix} \mathbf{I}_n \mid [\,\mathbf{v}\,]_{\mathcal{B}} \end{bmatrix}$$

We can treat the conversion from $\mathbf{v}$ to its $\mathcal{B}$-coordinate $[\,\mathbf{v}\,]_{\mathcal{B}}$ as a mapping $[\,\mathbf{v}\,]_{\mathcal{B}} : V \to \mathbb{R}^n$ (where $n = \dim V$). Then, from the unique representation theorem presented above, we know that

$[\,\cdot\,]_{\mathcal{B}} : V \to \mathbb{R}^n$ is a linear transformation that is bijective (one-to-one and onto).

and thus we can state that any set of vectors in V and the set of their corresponding $\mathcal{B}$-coordinate vectors share the same following properties:

$\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\}$ linearly independent $\iff \{[\mathbf{v}_1]_{\mathcal{B}}, [\mathbf{v}_2]_{\mathcal{B}}, \ldots, [\mathbf{v}_p]_{\mathcal{B}}\}$ linearly independent
$\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_p\} = V \iff \operatorname{span}\{[\mathbf{v}_1]_{\mathcal{B}}, [\mathbf{v}_2]_{\mathcal{B}}, \ldots, [\mathbf{v}_p]_{\mathcal{B}}\} = \mathbb{R}^n$

3.6 Change of basis $\mathbf{P}_{\mathcal{C}\leftarrow\mathcal{B}}$, $\mathbf{A}_{\mathcal{B},\mathcal{C}}$

What if we want to write how to convert between two bases $\mathcal{B}$ and $\mathcal{C}$ as a matrix?

Changing between two bases, say $\mathcal{B}$ and $\mathcal{C}$ in a vector space V, requires two linear transformations: $[\,\cdot\,]_{\mathcal{C}} : V \to \mathbb{R}^n$ and $[\,\cdot\,]_{\mathcal{B}}^{-1} : \mathbb{R}^n \to V$, where $n = \dim V$.

$$\begin{tikzcd} & V \arrow[dr, "{[\,\cdot\,]_{\mathcal{C}}}"] & \\ \mathbb{R}^n \arrow[ur, "{[\,\cdot\,]_{\mathcal{B}}^{-1}}"] \arrow[rr, "T_{\mathcal{C}\leftarrow\mathcal{B}}"] & & \mathbb{R}^n \end{tikzcd}$$

We define the "conversion" function from $\mathcal{B}$ to $\mathcal{C}$ to be the function composition

$$T_{\mathcal{C}\leftarrow\mathcal{B}} = [\,\cdot\,]_{\mathcal{C}} \circ [\,\cdot\,]_{\mathcal{B}}^{-1}$$

Thus the "standard matrix" of $T_{\mathcal{C}\leftarrow\mathcal{B}}$ is

$$\mathbf{A}_{T_{\mathcal{C}\leftarrow\mathcal{B}}} = \begin{bmatrix} T_{\mathcal{C}\leftarrow\mathcal{B}}(\mathbf{e}_1) & T_{\mathcal{C}\leftarrow\mathcal{B}}(\mathbf{e}_2) & \cdots & T_{\mathcal{C}\leftarrow\mathcal{B}}(\mathbf{e}_n) \end{bmatrix} = \begin{bmatrix} [\mathbf{b}_1]_{\mathcal{C}} & [\mathbf{b}_2]_{\mathcal{C}} & \cdots & [\mathbf{b}_p]_{\mathcal{C}} \end{bmatrix}$$

> Let V and W be vector spaces with bases $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\} \subset V$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_p\} \subset W$. The **change of basis matrix** of T is the standard matrix of the transformation $[\,\cdot\,]_{\mathcal{B}} : V \to \mathbb{R}^p$:
>
> $$\mathbf{P}_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{bmatrix} [\mathbf{b}_1]_{\mathcal{C}} & [\mathbf{b}_2]_{\mathcal{C}} & \cdots & [\mathbf{b}_p]_{\mathcal{C}} \end{bmatrix}$$

What if we want to write linear transformations as a matrix with respect to their bases?

Suppose we have two vector spaces V and W with a linear transformation $T : V \to W$. If we have two bases $\mathcal{B} = \{\mathbf{b}_1, \ldots \mathbf{b}_m\} \subset V$ and $\mathcal{C} \subset W$, we can write a new linear transformation $T_{\mathcal{B},\mathcal{C}} : \mathbb{R}^m \to \mathbb{R}^n$ that takes the input of $\mathcal{B}$-coordinate vectors and outputs $\mathcal{C}$-coordinate vectors:

$$\begin{tikzcd} V \arrow[r, "T"] & W \arrow[d, "{[\,\cdot\,]_{\mathcal{C}}}"] \\ \mathbb{R}^m \arrow[u, "{[\,\cdot\,]_{\mathcal{B}}^{-1}}"] \arrow[r, "T_{\mathcal{B},\mathcal{C}}"] & \mathbb{R}^n \end{tikzcd}$$

$$T_{\mathcal{B},\mathcal{C}} = [\,\cdot\,]_{\mathcal{C}} \circ T \circ [\,\cdot\,]_{\mathcal{B}}^{-1}$$

Thus, if we let $\varepsilon_m = \{\mathbf{e}_1, \mathbf{e}_2, \ldots \mathbf{e}_m\}$, then, for all $i \in \{1, 2, \ldots, m\}$, $\left[T([\mathbf{e}_i]_{\mathcal{B}}^{-1})\right]_{\mathcal{C}} = [T(\mathbf{b}_i)]_{\mathcal{C}}$.

> Let V and W be vector spaces with bases $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p\} \subset V$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_q\} \subset W$. Let $T : V \to W$ be a linear transformation. The **matrix of T with respect to $\mathcal{B}$ and $\mathcal{C}$** is the standard matrix of the transformation $T_{\mathcal{B},\mathcal{C}}$:
>
> $$\mathbf{A}_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} [T(\mathbf{b}_1)]_{\mathcal{C}} & [T(\mathbf{b}_2)]_{\mathcal{C}} & \cdots & [T(\mathbf{b}_p)]_{\mathcal{C}} \end{bmatrix}$$

If we want to use arbitrary bases $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ and $\mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_k\}$ for $\mathbb{R}^n$ and $\mathbb{R}^k$ respectively to express a linear transformation between these bases, recall that

$$\mathbf{P}_{\varepsilon_n\leftarrow\mathcal{B}} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix} \qquad \mathbf{P}_{\mathcal{C}\leftarrow\varepsilon_k} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_k \end{bmatrix}^{-1}$$

Thus, for the standard matrix $\mathbf{A}$ of T, the matrix with respect to $\mathcal{B}$ and $\mathcal{C}$ is

$$\mathbf{A}_{\mathcal{B},\mathcal{C}} = \mathbf{P}_{\mathcal{C}\leftarrow\varepsilon_k}\mathbf{A}\mathbf{P}_{\varepsilon_n\leftarrow\mathcal{B}} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_k \end{bmatrix}^{-1} \mathbf{A} \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix}$$

> Let $\mathbb{R}^n$ and $\mathbb{R}^k$ be vector spaces and $T : \mathbb{R}^n \to \mathbb{R}^k : \mathbf{x} \mapsto \mathbf{A}\mathbf{x}$ be a linear transformation. Then, the matrix of T relative to $\mathcal{B}$ and $\mathcal{C}$ is
>
> $$\mathbf{A}_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_k \end{bmatrix}^{-1} \mathbf{A} \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix}$$

4 Eigenvectors and eigenvalues

> Let V be a real vector space, and $T : V \to V$ be a linear transformation. $\mathbf{v} \in V$ is an **eigenvector** of T with **eigenvalue** $\lambda \in \mathbb{R}$ if and only if
> - $\mathbf{v} \neq \mathbf{0}$
> - $T(\mathbf{v}) = \lambda\mathbf{v}$

Let $\mathbf{A}$ be the matrix encoding T. The equation above can also be written as

$$\begin{aligned} \mathbf{A}\mathbf{v} &= \lambda\mathbf{v} \\ (\mathbf{A} - \lambda\mathbf{I}_n)\mathbf{v} &= \mathbf{0} \\ \iff & \mathbf{v} \in \operatorname{null}(\mathbf{A} - \lambda\mathbf{I}_n) \end{aligned}$$

Let $\mathbf{A}$ be an $n \times n$ matrix. Given an eigenvalue λ of $\mathbf{A}$, the **λ-eigenspace** of $\mathbf{A}$ is the set of all λ-eigenvectors and the zero vector:

$$\operatorname{null}(\mathbf{A} - \lambda\mathbf{I}_n)$$

A real number λ is an eigenvalue if and only if the set of λ-eigenvectors is non-empty. Thus, to find an eigenvalue, we must find a number $\lambda \in \mathbb{R}$ where $\operatorname{null}(\mathbf{A} - \lambda\mathbf{I}_n) \neq \{\mathbf{0}\}$. Thus $\mathbf{A} - \lambda\mathbf{I}_n$ cannot be an invertible matrix, and thus $\det(\mathbf{A} - \lambda\mathbf{I}_n) = 0$. We call the determinant of $\mathbf{A} - \lambda\mathbf{I}_n$ the characteristic polynomial.

Let $\mathbf{A}$ be an $n \times n$ matrix. The **characteristic polynomial** of $\mathbf{A}$ $p_{\mathbf{A}}(x)$ is the nth degree polynomial formed by

$$p_{\mathbf{A}}(x) = \det(\mathbf{A} - x\mathbf{I}_n)$$

The roots of $p_{\mathbf{A}}$ are the eigenvalues of $\mathbf{A}$.

The eigenvalues of an upper triangular matrix are simply the entries of the diagonal.

4.1 Diagonalization

Recall that a linear transformation $T : V \to W$ can be represented by a matrix with respect to two bases $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots \mathbf{b}_n\} \subseteq V$ and $\mathcal{C} \subseteq W$:

$$\mathbf{A}_{\mathcal{B},\mathcal{C}} = \begin{bmatrix} [T(\mathbf{b}_1)]_{\mathcal{C}} & [T(\mathbf{b}_2)]_{\mathcal{C}} & \cdots & [T(\mathbf{b}_n)]_{\mathcal{C}} \end{bmatrix} = \mathbf{P}_{\mathcal{C}\leftarrow\boldsymbol{\varepsilon}_n}\mathbf{A}\mathbf{P}_{\boldsymbol{\varepsilon}_k\leftarrow\mathcal{B}}$$

where $\dim V = n$ and $\dim W = k$, and $\boldsymbol{\varepsilon}_n$ and $\boldsymbol{\varepsilon}_k$ are standard bases for $\mathbb{R}^n$ and $\mathbb{R}^k$ respectively.

If $T : V \to V$ is a linear transformation along the same basis $\mathcal{B}$, then we may represent the matrix of T with respect to $\mathcal{B}$ as

$$\begin{aligned} \mathbf{A}_{\mathcal{B},\mathcal{B}} &= (\mathbf{P}_{\boldsymbol{\varepsilon}_n\leftarrow\mathcal{B}})^{-1}\mathbf{A}\mathbf{P}_{\boldsymbol{\varepsilon}_n\leftarrow\mathcal{B}} \\ &= \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix}^{-1}\mathbf{A}\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \end{bmatrix} \qquad (\text{in } \mathbb{R}^n) \end{aligned}$$

Given a linear transformation $T : V \to V$, the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ corresponding to a basis of eigenvectors $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots \mathbf{b}_n\}$ can be expressed as a **diagonal matrix**:

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

A square matrix $\mathbf{A}$ is **diagonalizable** if there exists an invertible matrix $\mathbf{P}$ such that, for some diagonal matrix $\mathbf{D}$,

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$$

Let $\mathbf{P} = \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_n \end{bmatrix}$. $\mathcal{B} = \{\mathbf{b}_1, \dots \mathbf{b}_n\}$ is a basis of eigenvectors of $\mathbf{A}$ if and only if $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ is a diagonal matrix.

proof If $\mathbf{D} = \mathbf{A}_{\mathcal{B},\mathcal{B}}$ is diagonal, where $\mathbf{A}_{\mathcal{B},\mathcal{B}}$ is the matrix of a linear transformation $T : V \to V$ with respect to $\mathcal{B}$,

Then, for all $i \in \{1, 2, \dots, n\}$, $\left[T(\mathbf{b}_i)\right]_{\mathcal{B}} = \lambda_i\mathbf{e}_i$. This gives us an important relationship between the mapping of $\mathbf{b}_i$ and the diagonal values of $\mathbf{A}_{\mathcal{B},\mathcal{B}}$:

$$T(\mathbf{b}_i) = \lambda_i\mathbf{b}_i$$

meaning that $\mathbf{b}_i$ is an **eigenvector** for T with eigenvalue λ_i. Thus $\mathcal{B}$ is a basis of eigenvectors. ■

$$\mathbf{A} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \qquad \begin{aligned} T(\mathbf{b}_1) &= \lambda_1 \mathbf{b}_1 \\ T(\mathbf{b}_2) &= \lambda_2 \mathbf{b}_2 \\ &\ddots \\ T(\mathbf{b}_n) &= \lambda_1 \mathbf{b}_n \end{aligned}$$

Let λ_i be an eigenvalue of an $n \times n$ matrix $\mathbf{A}$.

- **Algebraic multiplicity** of λ_i = how much $(x - \lambda_i)$ appears in $p_{\mathbf{A}}(x)$
- **Geometric multiplicity** of λ_i = dimension of the λ_i-eigenspace
  = number of linearly independent λ_i-eigenvectors of $\mathbf{A}$
  = $\text{nullity}(\mathbf{A} - \lambda_i \mathbf{I}_n)$

Remark: both multiplicities must be ≥ 1!

A diagonalizable $n \times n$ matrix $\mathbf{A}$ must have n linearly independent eigenvectors to form an eigenbasis. In other words, given eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$,

$$\sum_i \text{nullity}(\mathbf{A} - \lambda_i \mathbf{I}_n) = n$$

Note: Eigenvectors with distinct eigenvalues are linearly independent. Thus, if $\mathbf{A}$ has n distinct eigenvalues, then $\mathbf{A}$ is diagonalizable.

- $\mathbf{A}$ is diagonalizable $\iff$ algebraic multiplicity of λ_i = geometric multiplicity of λ_i for all λ_i
- $\mathbf{A}$ is not diagonalizable $\iff$ algebraic multiplicity of λ_i > geometric multiplicity of λ_i for some λ_i

(think of geometric multiplicity has an ***upper bound*** for diagonalization. if algebraic multiplicity *beats* it, we don't have a diagonalizable matrix!)

5 Orthogonality and the inner product

5.1 Standard inner product $\mathbf{a} \cdot \mathbf{b}$

Given $\mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{R}^n$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$, the **dot product** between $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n$$

Properties: Let $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

- $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$
- $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w}$
- $(\lambda \mathbf{u}) \cdot \mathbf{v} = \lambda(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot (\lambda \mathbf{v})$
- $\mathbf{u} \cdot \mathbf{u} \geq 0$
- $\mathbf{u} \cdot \mathbf{u} = 0 \iff \mathbf{u} = \mathbf{0}$

5.2 Orthogonal sets and matrices

A set of vectors S is **orthogonal** if $\forall \mathbf{u}, \mathbf{v} \in S$, where $\mathbf{u} \neq \mathbf{v}$, $\mathbf{u} \cdot \mathbf{v} = 0$.

An orthogonal set of non-zero vectors in $\mathbb{R}^n$ is linearly independent.

proof Let $S = \{\mathbf{s}_1, \mathbf{s}_2, \dots \mathbf{s}_p\}$ be an orthogonal set of non-zero vectors. Then, if for some $\lambda_1, \lambda_2, \dots \lambda_p$,

$$\begin{aligned} \lambda_1\mathbf{s}_1 + \lambda_2\mathbf{s}_2 + \dots + \lambda_p\mathbf{s}_p &= \mathbf{0} \\ \therefore (\lambda_1\mathbf{s}_1 + \lambda_2\mathbf{s}_2 + \dots + \lambda_p\mathbf{s}_p) \cdot \mathbf{s}_1 &= \mathbf{0} \cdot \mathbf{s}_1 = 0 \\ \lambda_1(\mathbf{s}_1 \cdot \mathbf{s}_1) + \lambda_2(\mathbf{s}_2 \cdot \mathbf{s}_1) + \dots + \lambda_p(\mathbf{s}_p \cdot \mathbf{s}_1) &= 0 \end{aligned}$$

Since $\mathbf{i} \cdot \mathbf{j} = 0$ for $i \neq j$,

$$\lambda_1(\mathbf{s}_1 \cdot \mathbf{s}_1) = 0$$

Since $\mathbf{s}_1 \neq \mathbf{0}$, then $\mathbf{s_1} \cdot \mathbf{s}_1 \neq 0$, and thus $\lambda_1 = 0$. Since $\lambda_1 = \dots = \lambda_p = 0$, S is linearly independent. ■

Let V be a subspace of $\mathbb{R}^n$, and $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \dots \mathbf{b}_p\}$ be an orthogonal basis. Then,

$$\forall \mathbf{v} \in V \qquad \mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{b}_1}{\mathbf{b}_1 \cdot \mathbf{b}_1}\mathbf{b}_1 + \frac{\mathbf{v} \cdot \mathbf{b}_2}{\mathbf{b}_2 \cdot \mathbf{b}_2}\mathbf{b}_2 + \dots + \frac{\mathbf{v} \cdot \mathbf{b}_p}{\mathbf{b}_p \cdot \mathbf{b}_p}\mathbf{b}_p$$

proof Suppose we have $\mathbf{v} \in V$ such that there exists $\lambda_1, \lambda_2 \dots, \lambda_n$ such that

$$\mathbf{v} = \lambda_1\mathbf{b}_1 + \lambda_2\mathbf{b}_2 + \dots + \lambda_n\mathbf{b}_p$$

Then, if we take the inner product of $\mathbf{v}$ and $\mathbf{b}_i$, we get

$$\begin{aligned} \mathbf{v} \cdot \mathbf{b}_i &= (\lambda_1\mathbf{b}_1 + \lambda_2\mathbf{b}_2 + \dots + \lambda_n\mathbf{b}_p) \cdot \mathbf{b}_i \\ &= \lambda_1(\mathbf{b}_1 \cdot \mathbf{b}_i) + \dots + \lambda_i(\mathbf{b}_i \cdot \mathbf{b}_i) + \dots + \lambda_p(\mathbf{b}_p \cdot \mathbf{b}_i) \\ &= \lambda_1(0) + \dots + \lambda_i(\mathbf{b}_i \cdot \mathbf{b}_i) + \dots + \lambda_p(0) \\ &= \lambda_i(\mathbf{b}_i \cdot \mathbf{b}_i) \end{aligned}$$

Because $\mathcal{B}$ is a basis, $\mathbf{b}_i \cdot \mathbf{b_i}$ is nonzero. Thus,

$$\lambda_i = \frac{\mathbf{v} \cdot \mathbf{b}_i}{\mathbf{b}_i \cdot \mathbf{b}_i}$$

■

A set of vectors S is **orthonormal** if it is orthogonal and $\forall \mathbf{v} \in S$, $\|\mathbf{v}\| = 1$.

For example, the standard basis $\boldsymbol{\varepsilon}_n$ of $\mathbb{R}^n$ is the most common orthonormal basis. If a set is orthogonal but not orthonormal, we can "divide" the vectors by their own magnitude (called normalization of the vector):

A **unit vector**, denoted $\hat{\mathbf{u}}$, is a vector with magnitude $\|\hat{\mathbf{u}}\| = 1$.

$$\hat{\mathbf{u}} = \frac{\mathbf{u}}{\|\mathbf{u}\|}$$

An $n \times n$ matrix $\mathbf{U}$ is an **orthogonal matrix** if $\mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{I}_n$.

$\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{bmatrix}$ is an orthogonal matrix if and only if $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is an orthonormal basis.

proof Suppose $\mathbf{U} = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{bmatrix}$ is an orthogonal matrix. Then,

$$\begin{aligned} \mathbf{U}^\mathsf{T}\mathbf{U} &= \begin{bmatrix}\mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_n\end{bmatrix}\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{bmatrix} = \begin{bmatrix}\mathbf{u}_1 \cdot \mathbf{u}_1 & \mathbf{u}_1 \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_1 \cdot \mathbf{u}_n \\ \mathbf{u}_2 \cdot \mathbf{u}_1 & \mathbf{u}_2 \cdot \mathbf{u}_2 & & \mathbf{u}_2 \cdot \mathbf{u}_n \\ \vdots & & \ddots & \vdots \\ \mathbf{u}_n \cdot \mathbf{u}_1 & \mathbf{u}_n \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_n \cdot \mathbf{u}_n\end{bmatrix} \\ &= \begin{bmatrix}1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 1\end{bmatrix} = \mathbf{I}_n \end{aligned}$$

Thus,

- $\forall \mathbf{u}_i, \mathbf{u}_j \in \{\mathbf{u}_1, \dots, \mathbf{u}_n\},\ \mathbf{u}_i \neq \mathbf{u}_j \quad \mathbf{u}_i \cdot \mathbf{u}_j = 0 \implies$ orthogonal
- $\forall \mathbf{u} \in \{\mathbf{u}_1, \dots, \mathbf{u}_n\} \quad \mathbf{u} \cdot \mathbf{u} = 1 \implies$ normalized vectors

By the properties above, $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$ is an orthonormal basis. ■

Let $\mathbf{U}$ be an orthogonal matrix. Then,

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \quad \mathbf{x} \cdot \mathbf{y} = \mathbf{U}\mathbf{x} \cdot \mathbf{U}\mathbf{y}$$

proof

$$\mathbf{U}\mathbf{x} \cdot \mathbf{U}\mathbf{y} = (\mathbf{U}\mathbf{x})^\mathsf{T}\mathbf{U}\mathbf{y} = \mathbf{x}^\mathsf{T}\mathbf{U}^\mathsf{T}\mathbf{U}\mathbf{y} = \mathbf{x}^\mathsf{T}\mathbf{I}_n\mathbf{y} = \mathbf{x}^\mathsf{T}\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$$

■

5.3 Orthogonal complements and projections

The **orthogonal complement** $W^\perp$ of a subspace $W \subset \mathbb{R}^n$ is the set of vectors that are orthogonal to every vector in W.

$$W^\perp := \{\mathbf{v} \in \mathbb{R}^n \mid \forall \mathbf{w} \in W : \mathbf{v} \cdot \mathbf{w} = 0\}$$

A really important theorem that we will use later on (see SVD) is this: let $\mathbf{A}$ be a matrix. Then,

$$(\operatorname{col} \mathbf{A})^\perp = \operatorname{null}(\mathbf{A}^\mathsf{T})$$

proof Suppose $\mathbf{u} \in \operatorname{col} \mathbf{A}$ and $\mathbf{v} \in \operatorname{null} \mathbf{A}^\mathsf{T}$. Then, by definition, $\exists \mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{A}\mathbf{x} = \mathbf{u}$. Therefore $\mathbf{u} \cdot \mathbf{v} = (\mathbf{A}\mathbf{x}) \cdot \mathbf{v} = (\mathbf{A}\mathbf{x})^\mathsf{T}\mathbf{v} = \mathbf{x}^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{v} = \mathbf{x}^\mathsf{T}(\mathbf{A}^\mathsf{T}\mathbf{v}) = \mathbf{x}^\mathsf{T}\mathbf{0} = 0$. Thus $\operatorname{null} \mathbf{A}^\mathsf{T} \subseteq (\operatorname{col} \mathbf{A})^\perp$. We know that $\dim(\operatorname{col} \mathbf{A})^\perp = n - \operatorname{rank} \mathbf{A}$, and by the rank–nullity theorem, $\dim \mathbf{A}^\mathsf{T} = n - \operatorname{rank} \mathbf{A}$. Thus $(\operatorname{col} \mathbf{A})^\perp = \operatorname{null}(\mathbf{A}^\mathsf{T})$. ■

Given a vector $\mathbf{v} \in \mathbb{R}^n$, the **orthogonal projection** of $\mathbf{v}$ on a subspace $W \subset \mathbb{R}^n$ is the unique vector $\operatorname{proj}_W \mathbf{v}$ such that $\mathbf{v} - \operatorname{proj}_W \mathbf{v} \in W^\perp$.

Let $\{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ be an orthogonal basis for W. Then the orthogonal projection is thus constructed

$$\operatorname{proj}_W \mathbf{v} = \frac{\mathbf{v} \cdot \mathbf{b}_1}{\mathbf{b}_1 \cdot \mathbf{b}_1}\mathbf{b}_1 + \frac{\mathbf{v} \cdot \mathbf{b}_2}{\mathbf{b}_2 \cdot \mathbf{b}_2}\mathbf{b}_2 + \cdots + \frac{\mathbf{v} \cdot \mathbf{b}_p}{\mathbf{b}_p \cdot \mathbf{b}_p}\mathbf{b}_p$$

proof Since $\{\mathbf{b}_1, \dots, \mathbf{b}_p\}$ is an orthogonal basis for W, then we have to find weights $\lambda_1, \dots, \lambda_p$ such that $\operatorname{proj}_W \mathbf{v} = \lambda_1\mathbf{b}_1 + \cdots + \lambda_p\mathbf{b}_p$. Since, by definition of orthogonal projection, $\mathbf{v} - \operatorname{proj}_W \mathbf{v} \in W^\perp$. Thus, for all i in $1 \leq i \leq p$,

$$\begin{aligned}(\mathbf{v} - \lambda_1\mathbf{b}_1 - \cdots - \lambda_p\mathbf{b}_p) \cdot \mathbf{b}_i &= 0 \\ \mathbf{v} \cdot \mathbf{b}_i - \lambda_i\mathbf{b}_i \cdot \mathbf{b}_i &= 0 \\ \lambda_i &= \frac{\mathbf{v} \cdot \mathbf{b}_i}{\mathbf{b}_i \cdot \mathbf{b}_i}\end{aligned}$$

■

5.4 Gram–Schmidt algorithm and QR factorization

The Gram–Schmidt algorithm is an algorithm for producing an orthogonal basis for any nonzero subspace of $\mathbb{R}^n$.

Gram–Schmidt orthogonalization
Let $\mathcal{V} = \{\mathbf{v}_1, \mathbf{v}_2 \ldots, \mathbf{v}_p\}$ be a basis of the vector space $\mathbb{R}^p$. To turn $\mathcal{V}$ into an orthogonal basis $\mathcal{B}$, define $\mathcal{U} = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_p\}$ where

$$\begin{aligned}
\mathbf{u}_1 &= \mathbf{v}_1 \\
\mathbf{u}_2 &= \mathbf{v}_2 - \frac{\mathbf{v}_2 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 \\
\mathbf{u}_3 &= \mathbf{v}_3 - \frac{\mathbf{v}_3 \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 - \frac{\mathbf{v}_3 \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 \\
&\vdots \\
\mathbf{u}_p &= \mathbf{v}_p - \frac{\mathbf{v}_p \cdot \mathbf{u}_{p-1}}{\mathbf{u}_{p-1} \cdot \mathbf{u}_{p-1}}\mathbf{u}_{p-1} - \frac{\mathbf{v}_p \cdot \mathbf{u}_{p-2}}{\mathbf{u}_{p-2} \cdot \mathbf{u}_{p-2}}\mathbf{u}_{p-2} - \cdots - \frac{\mathbf{v}_p \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1
\end{aligned}$$

We can equivalently obtain the same results using a process called **QR** factorization. Let **A** and **Q** be the matrices

$$\begin{aligned}
\mathbf{A} &= \big[\text{vectors of any basis } \mathcal{V}\big] = \begin{bmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_p\end{bmatrix} \\
\mathbf{Q} &= \big[\text{orthonormalized basis } \mathcal{U}\big] = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_p\end{bmatrix}
\end{aligned}$$

From how we constructed $\mathcal{U}$ through the Gram–Schmidt algorithm,

$$\begin{aligned}
&\mathbf{u}_k = \mathbf{v}_k - \lambda_{k-1}\mathbf{u}_{k-1} - \lambda_{k-2}\mathbf{u}_{k-2} - \cdots - \lambda_1\mathbf{u}_1 \\
\implies &\mathbf{v}_k \in \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \\
\implies &\mathbf{v}_k = \lambda_{11}\mathbf{u}_1 + \cdots + \lambda_{kk}\mathbf{u}_k + 0\mathbf{u}_{k+1} + \cdots + 0\mathbf{u}_p
\end{aligned}$$

for some coefficients $\lambda_{11}, \ldots, \lambda_{kk} \in \mathbb{R}$. We can store *these coefficients* into a $p \times p$ matrix itself:

$$\mathbf{R} = \begin{bmatrix}
\lambda_{11} & \lambda_{12} & \cdots & \lambda_{1p} \\
0 & \lambda_{22} & \cdots & \lambda_{2p} \\
 & & \ddots & \\
0 & 0 & \cdots & \lambda_{pp}
\end{bmatrix}$$

and thus we have a new method of decomposition:

An $m \times n$ matrix **A** can be factorized as

$$\mathbf{A} = \mathbf{QR}$$

where **Q** is an $m \times n$ matrix whose columns are an orthonormal basis for col **A**, and **R** is an $n \times n$ upper triangular matrix with positive entries on its diagonal.

5.5 Method of least squares

Suppose we have an inconsistent matrix equation $\mathbf{Ax} = \mathbf{b}$, and we want to find the closest approximate vector to $\mathbf{x}$ in col $\mathbf{A}$. To find the best approximation $\hat{\mathbf{x}}$, the vector $\mathbf{A}\hat{\mathbf{x}} \in \operatorname{col}\mathbf{A}$ needs to be as close to $\mathbf{b}$ as possible. We therefore choose the vector $\mathbf{A}\hat{\mathbf{x}} = \operatorname{proj}_{\operatorname{col}\mathbf{A}}\mathbf{b}$.

Let **A** be an $m \times n$ matrix, and $\mathbf{b} \in \mathbb{R}^m$. A **least-square solution** to $\mathbf{Ax} = \mathbf{b}$ is a vector $\hat{\mathbf{x}}$ such that

$$\mathbf{A}\hat{\mathbf{x}} = \operatorname{proj}_{\operatorname{col}\mathbf{A}}\mathbf{b}$$

with the property that $\forall \mathbf{x} \in \mathbb{R}^n$, $\|\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}\| \le \|\mathbf{Ax} - \mathbf{b}\|$.

Note the special case where $\mathbf{Ax} = \mathbf{b}$ is consistent. Then, $\mathbf{b} \in \operatorname{col}\mathbf{A}$, and $\operatorname{proj}_{\operatorname{col}\mathbf{A}}\mathbf{b} = \mathbf{b}$, meaning that $\hat{\mathbf{x}} = \mathbf{x}$ is already a satisfactory vector.

Now let's suppose that $\mathbf{Ax} = \mathbf{b}$ has no solution. By definition, the orthogonal projection $\operatorname{proj}_{\operatorname{col}\mathbf{A}}\mathbf{b}$ of $\mathbf{b}$ is the unique vector such that

$$\mathbf{b} - \operatorname{proj}_{\operatorname{col}\mathbf{A}}\mathbf{b} \in (\operatorname{col}\mathbf{A})^\perp$$

Since $(\operatorname{col}\mathbf{A})^\perp = \operatorname{null}\mathbf{A}^\mathsf{T}$,

$$\begin{aligned}\mathbf{b} - \operatorname{proj}_{\operatorname{col}\mathbf{A}}\mathbf{b} \in (\operatorname{col}\mathbf{A})^\perp &\iff \mathbf{b} - \operatorname{proj}_{\operatorname{col}\mathbf{A}}\mathbf{b} = \mathbf{b} - \mathbf{A}\hat{\mathbf{x}} \in \operatorname{null}\mathbf{A}^\mathsf{T} \\ &\iff \mathbf{A}^\mathsf{T}(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = \mathbf{0} \\ &\iff \mathbf{A}^\mathsf{T}\mathbf{b} - \mathbf{A}^\mathsf{T}\mathbf{A}\hat{\mathbf{x}} = \mathbf{0} \\ &\iff \mathbf{A}^\mathsf{T}\mathbf{b} = \mathbf{A}^\mathsf{T}\mathbf{A}\hat{\mathbf{x}}\end{aligned}$$

$\hat{\mathbf{x}} \in \mathbb{R}^n$ is a least square solution if and only if it is a solution to the **normal equations** to $\mathbf{A}\mathbf{x} = \mathbf{b}$:

$$\mathbf{A}^\mathsf{T}\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^\mathsf{T}\mathbf{b}$$

$$\begin{array}{ccc} \mathbf{A}\mathbf{x} = \mathbf{b} \text{ has a unique least square sol. } \hat{\mathbf{x}}. & \Longleftrightarrow & \mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{x} = \mathbf{A}^\mathsf{T}\mathbf{b} \text{ has a unique solution.} \\ \Updownarrow & & \Updownarrow \\ \text{Columns of } \mathbf{A} \text{ are linearly independent.} & & \text{Columns of } \mathbf{A}^\mathsf{T}\mathbf{A} \text{ are linearly independent.} \end{array}$$

Thus we have the following corollary:

If the columns of $\mathbf{A}$ are linearly independent, then the least squares solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$ is

$$\hat{\mathbf{x}} = (\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T}\mathbf{b}$$

5.6 Inner product spaces $\langle \mathrm{a}, \mathrm{b} \rangle$

Let V be a vector space. An **inner product** on V is a mapping $\langle\,\cdot\,,\cdot\,\rangle : V \times V \to \mathbb{R}$ such that

$$\begin{array}{rl} \forall \mathbf{u},\mathbf{v} \in V & \langle \mathbf{u},\mathbf{v}\rangle = \langle \mathbf{v},\mathbf{u}\rangle \\ \forall \mathbf{u},\mathbf{v},\mathbf{w} \in V & \langle \mathbf{u}+\mathbf{v},\mathbf{w}\rangle = \langle \mathbf{u},\mathbf{w}\rangle + \langle \mathbf{v},\mathbf{w}\rangle \\ \forall \lambda \in \mathbb{R},\ \forall \mathbf{u},\mathbf{v} \in V & \langle \lambda\mathbf{u},\mathbf{v}\rangle = \lambda\,\langle \mathbf{u},\mathbf{v}\rangle \\ \forall \mathbf{v} \in V & \langle \mathbf{v},\mathbf{v}\rangle \geq 0 \\ & \langle \mathbf{v},\mathbf{v}\rangle = 0 \iff \mathbf{v} = \mathbf{0}_V \end{array}$$

An **inner product space** $(V, \langle\cdot,\cdot\rangle)$ is a vector space V coupled with an inner product $\langle\cdot,\cdot\rangle$. The dot product $\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is considered the standard inner product or Euclidean inner product.

Note: A vector space ($\mathbb{R}^n$ in particular) can have multiple inner products.

All inner products on $\mathbb{R}^n$ are of the form $\langle \mathbf{u},\mathbf{v}\rangle = \mathbf{u}^\mathsf{T}\mathbf{A}\mathbf{v}$, where $\mathbf{A}$ is a symmetric matrix with strictly positive eigenvalues (positive definite).

Why symmetric? $\quad \langle \mathbf{u},\mathbf{v}\rangle = \mathbf{u}^\mathsf{T}\mathbf{A}\mathbf{v} \qquad \langle \mathbf{v},\mathbf{u}\rangle = \mathbf{v}^\mathsf{T}\mathbf{A}\mathbf{u}$

$$\therefore \langle \mathbf{u},\mathbf{v}\rangle = (\mathbf{u}^\mathsf{T}\mathbf{A}\mathbf{v})^\mathsf{T} = \mathbf{v}^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{u} = \mathbf{v}^\mathsf{T}\mathbf{A}\mathbf{u} = \langle \mathbf{v},\mathbf{u}\rangle \qquad \Longleftarrow\ \mathbf{A}^\mathsf{T} = \mathbf{A}$$

Why strictly positive eigenvalues? To ensure property 4.

With inner product spaces, we can generalize everything related to orthogonality and the dot product:

Let $\mathbf{u},\mathbf{v}$ be elements of an inner product space V.
- Norm of $\mathbf{u}$: $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u},\mathbf{u}\rangle}$
- Distance between $\mathbf{u}$ and $\mathbf{v}$: $\|\mathbf{u} - \mathbf{v}\|$
- Orthogonal vectors: $\langle \mathbf{u},\mathbf{v}\rangle = 0$

With abstract vector spaces, we have the following new properties:

Theorem (Cauchy–Schwarz inequality) Let $\mathbf{u}$ and $\mathbf{v}$ be elements of an inner product space V. Then,

$$|\langle \mathbf{u},\mathbf{v}\rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

Theorem (triangle inequality) Let $\mathbf{u}$ and $\mathbf{v}$ be elements of an inner product space V. Then,

$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

proof We know that

$$\begin{aligned}
\|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v} \rangle \\
&= \langle \mathbf{u}, \mathbf{u} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\
&= \|\mathbf{u}\| + \|\mathbf{v}\| + 2\langle \mathbf{u}, \mathbf{v} \rangle \\
&\leq \|\mathbf{u}\| + \|\mathbf{v}\| + 2\|\mathbf{u}\|\|\mathbf{v}\| \qquad \text{by the Cauchy–Schwarz inequality} \\
&= \left(\|\mathbf{u}\| + \|\mathbf{v}\|\right)^2
\end{aligned}$$

■

Everything we have developed in $\mathbb{R}^n$ equipped with the standard inner (dot) product can be generalized to the abstract inner product space.

Let $(V, \langle \cdot, \cdot \rangle)$ be an inner product space.
Two vectors $\mathbf{u}, \mathbf{v} \in V$ are **orthogonal** if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.
The **orthogonal complement** $W^\perp$ of a subspace W is the set of all vectors orthogonal to each vector in W: $W^\perp = \{\mathbf{v} \in V \mid \forall \mathbf{w} \in W \ \ \langle \mathbf{v}, \mathbf{w} \rangle = 0\}$.
The **orthogonal projection** of $\mathbf{u}$ onto a subspace W is the unique vector $\operatorname{proj}_W \mathbf{u}$ such that $\mathbf{u} - \operatorname{proj}_W \mathbf{u} \in W^\perp$. Given an orthogonal basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_p\}$ of W, it is equal to

$$\operatorname{proj}_W \mathbf{u} = \frac{\langle \mathbf{u}, \mathbf{w}_1 \rangle}{\langle \mathbf{w}_1, \mathbf{w}_1 \rangle}\mathbf{w}_1 + \frac{\langle \mathbf{u}, \mathbf{w}_2 \rangle}{\langle \mathbf{w}_2, \mathbf{w}_2 \rangle}\mathbf{w}_2 + \cdots + \frac{\langle \mathbf{u}, \mathbf{w_p} \rangle}{\langle \mathbf{w}_p, \mathbf{w}_p \rangle}\mathbf{w}_p$$

The **Gram–Schmidt algorithm** orthogonalizes a set of vectors $\{\mathbf{x}_1, \ldots \mathbf{x}_n\} \subseteq \mathbb{R}^n$ that may not necessarily be orthogonal. Define a new orthogonal basis $\{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ where for all $p \in \{1, 2, \ldots, n\}$,

$$\mathbf{b}_p = \mathbf{x}_p - \frac{\langle \mathbf{x}, \mathbf{b}_{p-1} \rangle}{\langle \mathbf{b}_{p-1}, \mathbf{b}_{p-1} \rangle}\mathbf{b}_{p-1} + \frac{\langle \mathbf{x}, \mathbf{b}_{p-2} \rangle}{\langle \mathbf{b}_{p-2}, \mathbf{b}_{p-2} \rangle}\mathbf{b}_{p-2} + \cdots + \frac{\langle \mathbf{x}, \mathbf{b}_1 \rangle}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle}\mathbf{b}_1$$

6 Eigendecomposition of matrices

The eigendecomposition of a matrix is the factorization of a matrix into a canonical form in terms of its eigenvalues and eigenvectors. We focus on three types of eigendecomposition:

- Diagonalization: $\mathbf{A} = \mathbf{P}\mathbf{A}_{\mathcal{B},\mathcal{B}}\mathbf{P}^{-1}$
- Orthogonal diagonalization/spectral decomposition: $\mathbf{A} = \mathbf{P}\mathbf{A}_{\mathcal{B},\mathcal{B}}\mathbf{P}^\mathsf{T}$
- Singular value decomposition: $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T} = \mathbf{P}_{\boldsymbol{\varepsilon}_k \leftarrow \mathcal{C}}\mathbf{A}_{\mathcal{B},\mathcal{C}}\mathbf{P}_{\mathcal{B} \leftarrow \boldsymbol{\varepsilon}_n}$

6.1 Orthogonal diagonalization of symmetric matrices

Recall that, for a linear transformation $T : V \to V$ whose behavior can be represented by a matrix $\mathbf{A}_{\mathcal{B},\mathcal{B}}$, $\mathbf{A}_{\mathcal{B},\mathcal{B}}$ is a diagonal matrix if and only if $\mathcal{B}$ is an eigenbasis of V. We will extend this notion over when $\mathcal{B}$ is an orthonormal eigenbasis of V. We define some key definitions:

An $n \times n$ matrix $\mathbf{A}$ is **symmetric** if and only if $\mathbf{A}^\mathsf{T} = \mathbf{A}$.

For example, a symmetric matrix may be $\begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$ or $\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$.

An $n \times n$ matrix $\mathbf{A}$ is **orthogonally diagonalizable** if there exists an orthogonal matrix $\mathbf{P}$ such that $\mathbf{P}^\mathsf{T}\mathbf{A}\mathbf{P} = \mathbf{D}$ is a diagonal matrix.

Recall that, for an invertible matrix $\mathbf{P}$, $\mathbf{P}^{-1}\mathbf{AP}$ is diagonal if and only if the columns of $\mathbf{P}$ form an eigenbasis. Suppose that $\mathbf{P}$ is also orthogonal, such that the columns of $\mathbf{P}$ form an orthonormal eigenbasis. Then, $\mathbf{P}^\mathsf{T}\mathbf{AP}$ is diagonal too.

A matrix $\mathbf{A}$ is orthogonally diagonalizable if and only if the columns of $\mathbf{P}$ form an orthonormal eigenbasis.

Let's connect the seemingly distinct ideas of symmetric and orthogonally diagonalizable matrices together.

lemma If a matrix $\mathbf{A}$ is orthogonally diagonalizable, then $\mathbf{A}$ is symmetric.

proof $\mathbf{A}$ is orthogonally diagonalizable if and only if there exists an orthogonal matrix $\mathbf{P}$ such that $\mathbf{P}^\mathsf{T}\mathbf{AP} = \mathbf{D}$ for a diagonal matrix $\mathbf{D}$. Then,

$$\begin{aligned}
\mathbf{P}^\mathsf{T}\mathbf{AP} = \mathbf{D} &\iff \mathbf{A} = \mathbf{PDP}^\mathsf{T} \\
&\iff \mathbf{A}^\mathsf{T} = (\mathbf{PDP}^\mathsf{T})^\mathsf{T} \\
&\iff \mathbf{A}^\mathsf{T} = (\mathbf{P}^\mathsf{T})^\mathsf{T}(\mathbf{PD})^\mathsf{T} \\
&\iff \mathbf{A}^\mathsf{T} = \mathbf{PD}^\mathsf{T}\mathbf{P}^\mathsf{T} = \mathbf{A}
\end{aligned}$$

■

Theorem (spectral theorem) $\mathbf{A}$ is symmetric if and only if $\mathbf{A}$ is orthogonally diagonalizable.

proof ($\implies$, partially) Let $\mathbf{A}$ be a symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \neq \lambda_2$, and $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ be eigenvectors of $\mathbf{A}$ in the λ_1 and λ_2-eigenspaces, respectively. Then, since $\mathbf{A}^\mathsf{T} = \mathbf{A}$,

$$\begin{aligned}
\mathbf{u} \cdot \mathbf{Av} &= \mathbf{u}^\mathsf{T}\mathbf{Av} \\
&= \mathbf{u}^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{v} \\
&= (\mathbf{Au})^\mathsf{T}\mathbf{v} \\
&= \mathbf{Au} \cdot \mathbf{v}
\end{aligned}$$

Using this property, we know that

$$\mathbf{u} \cdot \mathbf{Av} = \mathbf{Au} \cdot \mathbf{v} \implies \mathbf{u} \cdot (\lambda_2 \mathbf{v}) = (\lambda_1 \mathbf{u}) \cdot \mathbf{v} \implies (\lambda_2 - \lambda_1)\mathbf{u} \cdot \mathbf{v} = 0$$

Since $\lambda_1 \neq \lambda_2$, then $\lambda_2 - \lambda_1 \neq 0$. Thus, it must follow that the only factor that is 0 is $\mathbf{u} \cdot \mathbf{v}$. Thus, $\mathbf{u}$ and $\mathbf{v}$ are orthogonal. Thus proves the following lemma:

lemma If a symmetric matrix $\mathbf{A}$ has two eigenvectors $\mathbf{u}$ and $\mathbf{v}$ with two distinct eigenvalues λ_1 and λ_2, then $\mathbf{u}$ and $\mathbf{v}$ are orthogonal.

All that's left to show is that $\mathbf{u}$ and $\mathbf{v}$ form an orthonormal basis, and that $\mathbf{A}$ is diagonalizable, but this is the main hurdle of the proof of the spectral theorem that we do not have the tools to show yet. ■

How do we find an orthonormal basis associated with a symmetric matrix $\mathbf{A}$?

- Find the eigenvalues $\lambda_1, \ldots, \lambda_n$ for $\mathbf{A}$
- Determine a basis for each λ_i-eigenspace $(1 \leq i \leq n)$
- Use the Gram–Schmidt algorithm for eigenvectors within the same λ_i-eigenspace
- Perform normalization of the orthogonal eigenbasis

6.2 Singular value decomposition

Recall that, given a finite-dimensional vector space V with basis $\mathcal{B} \subset V$ and a linear transformation $T : V \to V$ with $\mathcal{B}$-matrix $\mathbf{A}_{\mathcal{B},\mathcal{B}}$.

- $\mathbf{A}_{\mathcal{B},\mathcal{B}}$ is diagonal if and only if $\mathcal{B}$ is an eigenbasis.
- $\mathbf{A}_{\mathcal{B},\mathcal{B}}$ is a symmetric matrix if and only if $\mathcal{B}$ is an orthogonal eigenbasis.

All that's left to show is to find a "good" basis $\mathcal{B}$ for a matrix $\mathbf{A}$ that is not diagonalizable (nor symmetric nor square), such that $\mathbf{A}_{\mathcal{B},\mathcal{B}}$ is "approximately" diagonal.

singular value decomposition I Let $T : \mathbb{R}^n \to \mathbb{R}^k$ be a linear transformation. There exists two orthonormal bases $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_n\} \subset \mathbb{R}^n$ and $\mathcal{C} = \{\mathbf{u}_1, \mathbf{u}_2, \dots \mathbf{u}_k\} \subset \mathbb{R}^k$ such that

$$\mathbf{A}_{\mathcal{B},\mathcal{C}} = \left[\begin{array}{ccc|c} \sigma_1 & & & 0 \\ & \ddots & & \vdots \\ & & \sigma_r & 0 \\ \hline 0 & \cdots & 0 & 0 \end{array}\right]$$

where $r = \operatorname{rank} \mathbf{A}$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

$\mathbf{A}_{\mathcal{B},\mathcal{C}}$ tells us the behavior of T over $\mathbf{v}_i$, an element in $\mathcal{B}$:

$$\begin{aligned} \big[T(\mathbf{v}_i)\big]_{\mathcal{C}} &= \mathbf{A}_{\mathcal{B},\mathcal{C}}[\mathbf{v}_i]_{\mathcal{B}} \\ &= \mathbf{A}_{\mathcal{B},\mathcal{C}}\mathbf{e}_i \\ &= \begin{cases} \lambda_i \mathbf{e}_i & \text{if } i \leq r \\ 0 & \text{if } i > r \end{cases} \\ \therefore T(\mathbf{v}_i) &= \begin{cases} \lambda_i \mathbf{u}_i & \text{if } i \leq r \\ 0 & \text{if } i > r \end{cases} \end{aligned}$$

Since T is linear, there exists a $k \times n$ transformation matrix $\mathbf{A}$ such that $T : \mathbf{x} \mapsto \mathbf{A}\mathbf{x}$. Thus,

$$\begin{aligned} \mathbf{A}_{\mathcal{B},\mathcal{C}} &= \mathbf{P}_{\mathcal{C}\leftarrow\varepsilon_k}\mathbf{A}\mathbf{P}_{\varepsilon_n\leftarrow\mathcal{B}} \\ \iff \mathbf{A} &= \mathbf{P}_{\varepsilon_k\leftarrow\mathcal{C}}\mathbf{A}_{\mathcal{B},\mathcal{C}}\mathbf{P}_{\mathcal{B}\leftarrow\varepsilon_n} \\ &= \big[\mathbf{u}_1 \cdots \mathbf{u}_k\big]\, \mathbf{A}_{\mathcal{B},\mathcal{C}}\, \big[\mathbf{v}_1 \cdots \mathbf{v}_n\big]^{-1} \\ &= \big[\mathbf{u}_1 \cdots \mathbf{u}_k\big]\, \mathbf{A}_{\mathcal{B},\mathcal{C}}\, \big[\mathbf{v}_1 \cdots \mathbf{v}_n\big]^{\mathsf{T}} \quad \text{(since } \mathcal{C} \text{ is orthonormal)} \end{aligned}$$

Let $\mathbf{U} = \big[\mathbf{u}_1 \cdots \mathbf{u}_k\big]$, $\mathbf{\Sigma} = \mathbf{A}_{\mathcal{B},\mathcal{C}}$, and $\mathbf{V} = \big[\mathbf{v}_1 \cdots \mathbf{v}_n\big]$. Then,

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{T}}$$

singular value decomposition II Given a $k \times n$ matrix $\mathbf{A}$, there exists two orthogonal matrices $\mathbf{U}$ and $\mathbf{V}$ such that

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathsf{T}}$$

where $\mathbf{\Sigma} = \left[\begin{array}{ccc|c} \sigma_1 & & & 0 \\ & \ddots & & \vdots \\ & & \sigma_r & 0 \\ \hline 0 & \cdots & 0 & 0 \end{array}\right]$, $r = \operatorname{rank} \mathbf{A}$, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$.

How do we find the singular value decomposition of a $k \times n$ matrix $\mathbf{A}$?

Find $\mathbf{V}$*:*

- Orthogonally diagonalize $\mathbf{A}^{\mathsf{T}}\mathbf{A}$ to obtain an orthonormal eigenbasis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subset \mathbb{R}^n$.
- (*We choose* $\mathbf{A}^{\mathsf{T}}\mathbf{A}$ *since it is symmetric and positive semidefinite.*[1])

Find $\mathbf{\Sigma}$*:*

- Define $\sigma_i := \sqrt{\lambda_i}$ and reorder them (if necessary) such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ (reorder the eigenvector columns in $\mathbf{V}$ accordingly).

[1] A symmetric $n \times n$ matrix $\mathbf{A}$ is positive semidefinite if and only if $\forall \mathbf{x} \in \mathbb{R}^n : \mathbf{x}^{\mathsf{T}}\mathbf{A}\mathbf{x} \geq 0$. In this case, we want the eigenvalues λ_i of $\mathbf{A}$ to be nonnegative so that $\sigma_i = \sqrt{\lambda_i} \in \mathbb{R}$. For an eigenvector $\mathbf{v}_i$ of $\mathbf{A}^{\mathsf{T}}\mathbf{A}$, we have $\lambda_i = \mathbf{v}_i^{\mathsf{T}}\lambda_i\mathbf{v}_i = \mathbf{v}_i^{\mathsf{T}}\left(\mathbf{A}^{\mathsf{T}}\mathbf{A}\right)\mathbf{v}_i \geq 0$. Consequently, we also have $\lambda_i = (\mathbf{A}\mathbf{v}_i)^{\mathsf{T}}\mathbf{A}\mathbf{v}_i = \|\mathbf{A}\mathbf{v}_i\|^2 \iff \|\mathbf{A}\mathbf{v}_i\| = \sqrt{\lambda_i} = \sigma_i$.

- Construct $\mathbf{\Sigma}$ from all singular values σ_i.

Find $\mathbf{U}$*:*

- For all $i \leq r$, define $\mathbf{u}_i = \dfrac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i$ to obtain an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r\} \subset \operatorname{col}\mathbf{A}$
- Extend $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ to an orthonormal basis for $\mathbb{R}^k$: $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$
  - Find a basis $\mathcal{C}'$ for $(\operatorname{col}\mathbf{A})^\perp = \operatorname{null}\mathbf{A}^\mathsf{T}$
  - Orthogonalize $\mathcal{C}'$ using the Gram–Schmidt algorithm
  - Normalize $\mathcal{C}' \implies \{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_k\}$
- Construct $\mathbf{U}$ from the orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$

Putting it all together,

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$$
$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_k \end{bmatrix}$$
$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1 & & & 0 \\ & \ddots & & \vdots \\ & & \sigma_r & \\ 0 & \cdots & \cdots & 0 \end{bmatrix}$$
$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{T}$$

$$\begin{aligned} T_\mathbf{A} &: \mathbb{R}^n \to \mathbb{R}^k \\ &: \mathbf{x} \mapsto \mathbf{A}\mathbf{x} \\ &: \mathbf{v}_i \mapsto \begin{cases} \sigma_i \mathbf{u}_i & i \leq r \\ 0 & i > r \end{cases} \end{aligned}$$

$$\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_r, \underbrace{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n}_{\operatorname{null}\mathbf{A}}\}$$
$$\mathcal{C} = \{\underbrace{\mathbf{u}_1, \ldots, \mathbf{u}_r,}_{\operatorname{col}\mathbf{A}} \ \underbrace{\mathbf{u}_{r+1}, \ldots, \mathbf{u}_k}_{(\operatorname{col}\mathbf{A})^\perp = \operatorname{null}\mathbf{A}^\mathsf{T}}\}$$

7 Fourier series

If you've taken Math 1B, AP Calculus BC, or any Calculus II class, you might recall the Taylor series:

Given a real-valued function $f : \mathbb{R} \to \mathbb{R}$ that is continuous on a real interval $[a, b]$ we want to find an infinite series to express $f(x)$ in terms of monomials ...

$$f(x) = \sum_{n=0}^{\infty} c_n x^n$$

but the unfortunate thing is that f needs to be restricted to be infinitely differentiable on $[a, b]$ for c_n to have a valid value for all $n \geq 0$. Using techniques from real analysis (or something like that), we get that

$$c_n = \frac{f^{(n)}(0)}{n!}$$

where $f^{(n)}(x)$ is the n^{th} derivative of $f(x)$. Thus, the canonical form of the Taylor series, specifically Maclaurin, is

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \cdots$$

We will try to develop a formula for all piecewise continuous functions $f : \mathbb{R} \to \mathbb{R}$ over some symmetric real interval $[-L, L]$. Instead of monomials $\{1, x, x^2, \ldots\}$ as our building blocks, we will use the following trigonometric functions:

$$\cos\left(\frac{\pi}{L}x\right), \sin\left(\frac{\pi}{L}x\right), \cos\left(\frac{2\pi}{L}x\right), \sin\left(\frac{2\pi}{L}x\right), \cos\left(\frac{3\pi}{L}x\right), \sin\left(\frac{3\pi}{L}x\right), \ldots$$

as building blocks. Therefore we will use the series

$$f(x) = c + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{n\pi}{L}x\right) + b_n \sin\left(\frac{n\pi}{L}x\right)\right]$$

as our starting form, and we need to find c, a_n, and b_n such that they are valid for all $x \in (-L, L)$.

Recall that an inner product space is a vector space V equipped with an inner product, which is a binary operation $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ such that

- $\forall \mathbf{u}, \mathbf{v} \in V \quad \langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
- $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in V \quad \langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$
- $\forall \lambda \in \mathbb{R}, \forall \mathbf{u}, \mathbf{v} \in V \quad \langle \lambda\mathbf{u}, \mathbf{v} \rangle = \lambda \langle \mathbf{u}, \mathbf{v} \rangle$

- $\forall \mathbf{u} \in \mathbb{R} \quad \langle \mathbf{u}, \mathbf{u} \rangle \geq 0$
- $\langle \mathbf{u}, \mathbf{u} \rangle = 0 \iff \mathbf{u} = \mathbf{0}_V$

We define the inner product of the vector space of all real-valued functions piecewise continuous on $[-L, L]$ V to be $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ via

$$\forall f, g \in V \quad \big\langle f(x), g(x) \big\rangle \coloneqq \int_{-L}^{L} f(x)g(x)\,\mathrm{d}x$$

Thus we have the following properties: for all nonnegative integers m, n,

$$\begin{aligned}
\left\langle \sin\frac{m\pi}{L}x, \cos\frac{n\pi}{L}x \right\rangle &= 0 \\
\left\langle \sin\frac{m\pi}{L}x, \sin\frac{n\pi}{L}x \right\rangle &= \begin{cases} 0 & \text{if } m \neq n \\ L & \text{if } m = n \end{cases} \\
\left\langle \cos\frac{m\pi}{L}x, \cos\frac{n\pi}{L}x \right\rangle &= \begin{cases} 0 & \text{if } m \neq n \\ L & \text{if } m = n \neq 0 \\ 2L & \text{if } m = n = 0 \end{cases}
\end{aligned}$$

and have the following consequence:

$$\left\{ \cos\frac{0\pi}{L}x, \cos\frac{\pi}{L}x, \sin\frac{\pi}{L}x, \cos\frac{2\pi}{L}x, \sin\frac{2\pi}{L}x, \ldots \right\} \text{ is an orthogonal set}$$

Thus, for any (eligible) f, we can build an infinitely close approximation for f using elements of an orthogonal basis, just like how we constructed an infinitely differentiable function out of the orthogonal[2] basis $\{1, x, x^2, \ldots\}$.

To find valid expressions for c, a_m, and b_m, we take the inner product between $f(x)$ and $\cos(nx)$:

$$\begin{aligned}
\left\langle f(x), \cos\frac{n\pi}{L}x \right\rangle &= \left\langle c\cos 0x + \sum_{m=1}^{\infty} \left[a_m \cos\left(\frac{m\pi}{L}x\right) + b_m \sin\left(\frac{m\pi}{L}x\right) \right], \cos\frac{n\pi}{L}x \right\rangle \\
&= c\left\langle \cos 0x, \cos\frac{n\pi}{L}x \right\rangle + \sum_{m=1}^{\infty} a_m \left\langle \cos\frac{m\pi}{L}x, \cos\frac{n\pi}{L}x \right\rangle + \sum_{n=1}^{\infty} b_n \left\langle \sin\frac{m\pi}{L}x, \cos\frac{n\pi}{L}x \right\rangle \\
&= c\left\langle \cos 0x, \cos\frac{n\pi}{L}x \right\rangle + \sum_{m=1}^{\infty} a_m \left\langle \cos\frac{m\pi}{L}x, \cos\frac{n\pi}{L}x \right\rangle
\end{aligned}$$

When $n = 0$, $\quad \big\langle f(x), \cos 0x \big\rangle = c\,\langle \cos 0x, \cos 0x \rangle + \displaystyle\sum_{m=1}^{\infty} a_m \langle \cos 0x, \cos 0x \rangle$

$$\implies \int_{-L}^{L} f(x)\,\mathrm{d}x = 2Lc + 0 \qquad (\text{since } \langle \cos 0x, \cos 0x \rangle = 0)$$

$$\implies c = \frac{1}{2L}\int_{-L}^{L} f(x)\,\mathrm{d}x$$

When $n > 0$,

$$\begin{aligned}
\big\langle f(x), \cos nx \big\rangle &= c\left\langle \cos 0x, \cos\frac{n\pi}{L}x \right\rangle + \sum_{m=1}^{\infty} a_m \left\langle \cos\frac{m\pi}{L}x, \cos\frac{n\pi}{L}x \right\rangle \\
&= 0 + a_n \left\langle \cos\frac{n\pi}{L}x, \cos\frac{n\pi}{L}x \right\rangle \\
&= a_n L
\end{aligned}$$

$$\implies \int_{-L}^{L} f(x)\cos\frac{n\pi}{L}x\,\mathrm{d}x = a_n L$$

$$\implies a_n = \frac{1}{L}\int_{-L}^{L} f(x)\cos\frac{n\pi}{L}x\,\mathrm{d}x$$

[2]where the inner product is $\big\langle p(x), q(x) \big\rangle \coloneqq \int_0^1 p(x)q(x)\,\mathrm{d}x$

Taking the inner product between $f(x)$ and $\sin(n\pi x/L)$:

$$\begin{aligned}
\left\langle f(x), \sin\frac{n\pi}{L}x \right\rangle &= c\left\langle \cos 0x, \sin\frac{n\pi}{L}x \right\rangle + \sum_{m=1}^{\infty} a_m \left\langle \cos\frac{m\pi}{L}x, \sin\frac{n\pi}{L}x \right\rangle + \sum_{m=1}^{\infty} b_m \left\langle \sin\frac{m\pi}{L}x, \sin\frac{n\pi}{L}x \right\rangle \\
&= 0 + 0 + b_n \left\langle \sin\frac{n\pi}{L}x, \sin\frac{n\pi}{L}x \right\rangle \\
&= b_n L
\end{aligned}$$

Thus we have

$$b_n = \frac{1}{L}\int_{-L}^{L} f(x)\sin\frac{n\pi}{L}x$$

Bringing all of this together, we get a Fourier approximation of a function f piecewise continuous on $[-L, L]$:

Let f be a piecewise continuous function on the real interval $[-L, L]$. The **Fourier series** of f is the trigonometric series

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty}\left[a_n \cos\left(\frac{n\pi}{L}x\right) + b_n \sin\left(\frac{n\pi}{L}x\right)\right]$$

where $a_n = \dfrac{1}{L}\displaystyle\int_{-L}^{L} f(x)\cos\left(\frac{n\pi}{L}x\right)\mathrm{d}x$

$b_n = \dfrac{1}{L}\displaystyle\int_{-L}^{L} f(x)\sin\left(\frac{n\pi}{L}x\right)\mathrm{d}x$

7.1 Convergence of Fourier series

Let $f : \mathbb{R} \to \mathbb{R}$ be a piecewise continuous function on $[-L, L]$. For all $x \in [-L, L]$,

$$\frac{a_0}{2} + \sum_{n=0}^{\infty}\left[a_n \cos\left(\frac{n\pi}{L}x\right) + b_n \sin\left(\frac{n\pi}{L}x\right)\right] = \frac{f(x^-) + f(x^+)}{2}$$

where $f(x^+) = \lim\limits_{h\to 0^+} f(x+h)$

$f(x^-) = \lim\limits_{h\to 0^-} f(x+h)$

Let $f : \mathbb{R} \to \mathbb{R}$ be a piecewise continuous function. The **Fourier sine series** of f on $[0, L]$ is the Fourier series of f_e on $[-L, L]$:

$$f(x) \sim \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi}{L}x\right)$$

where

$$b_n = \frac{2}{L}\int_{0}^{L} f(x)\sin\left(\frac{n\pi}{L}x\right)\mathrm{d}x$$

Let $f : \mathbb{R} \to \mathbb{R}$ be a piecewise continuous function. The **Fourier cosine series** of f on $[0, L]$ is the Fourier series of f_o on $[-L, L]$:

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi}{L}x\right)$$

where

$$a_n = \frac{2}{L}\int_{0}^{L} f(x)\cos\left(\frac{n\pi}{L}x\right)\mathrm{d}x$$

end of linear algebra

8 Differential equations

8.1 Introduction

This course includes a linear algebra-style, surface-level treatment of ordinary differential equations. Some prerequisite knowledge should come from abstract vector spaces, linear transformations coupled with rank and nullity, and bases of vector spaces. Obviously you need calculus knowledge (especially integration by parts) as well.

A **differential equation** is an equation that involves a variable y and its derivatives $y', y'', \dots$ and we can say that a differential equation is of **order** n if it involves up to the nth derivative of the function. A general ordinary differential equation of order n is of the form

$$f(y, y', y'', \dots, y^{(n)}) = 0$$

You may have been exposed to some differential equations already, such as:

- first-order separable: $y' = f(x)g(y) \implies \displaystyle\int \frac{\mathrm{d}y}{g(y)} = \int f(x)\,\mathrm{d}x$
- first-order linear non-homogeneous function-coefficient:

$$y' + P(x)y = Q(x) \implies y = \frac{1}{\mu(x)} \int \mu(x)Q(x)\,\mathrm{d}x \quad \text{where } \mu(x) = e^{\int^x P(t)\,\mathrm{d}t}$$

8.2 Homogeneous second-order linear differential equations

A **homogeneous second-order linear constant-coefficient ordinary differential equation** is a differential equation of the form

$$ay'' + by' + cy = 0 \qquad (a, b, c \in \mathbb{R})$$

Why do we call $ay'' + by' + cy = 0$ linear?

Because the mapping associated with $ay'' + by" + cy$, which is $T : \mathbb{C}^2(\mathbb{R}) \to \mathbb{R} : y \mapsto ay'' + by' + cy$, is a linear transformation.

Thus the general solution set to $ay'' + by' + cy = 0$ is the kernel of T.

Let $\ker T$ be the solution set to $ay'' + by' + cy = 0$. Then, for all $t_0 \in \mathbb{R}$, the mapping $S : \ker T \to \mathbb{R}^2$ defined via $y(t) \mapsto \begin{bmatrix} y(t_0) \\ y'(t_0) \end{bmatrix}$ is a bijective linear transformation.

As a consequence, since S is one-to-one and onto,

- nullity $S = 0$
- rank $S = 2$
- by the rank–nullity theorem, $\dim \ker T = 2$

So, to find the general solution $\ker T$, we just need to find **two** linearly independent solutions (that act as a basis of $\ker T$).

How do we find these solutions? We guess.

Let's *guess* that $ay'' + by' + cy = 0$ has a solution $y = e^{rt}$, where $r \in \mathbb{R}$. Then,

$$\begin{cases} y = e^{rt} \\ y' = re^{rt} \\ y'' = r^2 e^{rt} \end{cases}$$

$$\begin{aligned} \implies a(r^2e^{rt}) + b(re^{rt}) + c(e^{rt}) &= 0 \\ \implies e^{rt}(ar^2 + br + c) &= 0 \end{aligned}$$

Since $e^{rt} > 0$ for all r,

$$ar^2 + br + c = 0$$

The **auxiliary equation** to $ay'' + by' + cy = 0$ is

$$ar^2 + br + c = 0$$

and thus we have the following algorithm of guesses:

Given roots of the auxiliary equation $r_1, r_2 \in \mathbb{C}$,
- $r_1, r_2 \in \mathbb{R}$ and $r_1 \neq r_2$, then $y = c_1 e^{r_1 t} + c_2 e^{r_2 t}$
- $r_1 = r_2 = r$, then $y = c_1 e^{rt} + c_2 t e^{rt}$
- $r_1 = \alpha + i\beta$, $r_2 = \alpha - i\beta$, then $y = c_1 e^{\alpha t} \cos \beta t + c_2 e^{\alpha t} \sin \beta t$

8.3 Non-homogeneous second-order linear differential equations

A **non-homogeneous second-order linear constant-coefficient ordinary differential equation** is an equation of the form

$$ay'' + by' + cy = g(t) \neq 0 \qquad (a, b, c \in \mathbb{R})$$

The **complementary equation** is the homogeneous version of the non-homogeneous second-order linear differential equation:

$$ay'' + by' + cy = 0$$

whose solution we call the **complementary solution**.

Recall back in linear algebra that the equation $\mathbf{A}\mathbf{x} = \mathbf{b}$ has the general solution $\mathbf{x} = \mathbf{v}_p + \mathbf{v}_h$ where $\mathbf{A}\mathbf{v}_p = \mathbf{b}$ (particular solution) and $\mathbf{A}\mathbf{v}_h = \mathbf{0}$.

The differential equations version is that the equation $ay'' + by' + cy = 0$ where $a, b, c \in \mathbb{R}$ has the general solution is composed of the particular solution y_p and y_h.

The **general solution** to $ay'' + by" + c = g(t) \neq 0$ is the solution $y_p + y_h$, where y_p is a particular solution to the equation and y_h is the complementary solution.

proof Just like the linear algebra version, we try to prove this in terms of linear transformations. Let $T : \mathbb{C}^2(\mathbb{R}) \to \mathbb{R} : y \mapsto ay'' + by' + cy$ be a linear transformation, let $y_p(t)$ be a particular solution to $T(y) = g(t)$, and let $y_g(t)$ be the general solution to $T(y) = g(t)$.

$$\begin{aligned} \text{Then, } T(y_p - y_g) &= T(y_p) - T(y_g) \\ &= g(t) - g(t) = 0 \quad \implies y_p - y_g \in \ker T \end{aligned}$$

Let $y_h = y_p - y_g$. Then $y_h \in \ker T$ and $y_g = y_p + y_h$. ■

method of undetermined coefficients
For a differential equation of the form $ay'' + by' + cy = g(t)$ where $g(t)$ is of the form

$$g(t) = p_m(t) e^{\alpha t} \cos \beta t \qquad \text{or} \qquad p_m(t) e^{\alpha t} \sin \beta t$$

then guess that the particular solution y_p is such:

$$y_p(t) = t^s (A_m t^m + \cdots + A_1 t + A_0) e^{\alpha t} \cos \beta t + t^s (A_m t^m + \cdots + A_1 t + A_0) e^{\alpha t} \sin \beta t$$

where $s = 0$ if $e^{\alpha t} \cos \beta t$ is not a complementary solution
$s = 1$ if $e^{\alpha t} \cos \beta t$ is a complementary solution but $t e^{\alpha t}$ is not
$s = 2$ if $e^{\alpha t} \cos \beta t$ and $e^{\alpha t} \cos \beta t$ are both complementary solutions

and evaluate the equation $ay_p'' + by_p' + cy_p = g(t)$ to set up a linear system of coefficients $A_m, \ldots, A_0$. Then, plug them into y_p to find the particular solution.

The **general solution** will be of the form

$$y_p + y_h$$

where y_p is the particular solution found through the method of undetermined coefficients, and y_h is the complementary solution.

8.4 Linear systems of differential equations

Let's define some notation before we get into rewriting linear systems of differential equations as a vector-matrix form ($\mathbf{Ax} = \mathbf{b}$).

A **vector-valued ($\mathbb{R}^n$-valued) function** is a function $\mathbf{x} : \mathbb{R} \to \mathbb{R}^n$ that maps $t \in \mathbb{R}$ to a vector

$$\begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \in \mathbb{R}^n$$

where $x_1, \dots, x_n : \mathbb{R} \to \mathbb{R}$ are all real-valued functions. We notate the image of t as $\mathbf{x}(t)$.

We are concerned with the following functions:

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \qquad \mathbf{x}'(t) \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix} \qquad \mathbf{A}(t) = \begin{bmatrix} a_{11}(t) & \cdots & a_{1n}(t) \\ \vdots & \ddots & \vdots \\ a_{n1}(t) & \cdots & a_{nn}(t) \end{bmatrix}$$

A linear system of n first-order differential equations in $x_1(t), \dots, x_n(t)$ is an equation of the form

$$\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{g}(t)$$

example Take the differential equation $ay'' + by' + cy = 0$. To turn this into a differential equation, we isolate y'' and y' such that we can express them as a vector $\mathbf{x}'(t) = \begin{bmatrix} y'' & y' \end{bmatrix}^\mathsf{T}$ with $\mathbf{x}'(t) = \begin{bmatrix} y' & y \end{bmatrix}^\mathsf{T}$.

We add a "cheeky" extra line here: $y' = y'$ so that the linear system is instead

$$\begin{cases} y'' & = -\dfrac{b}{a}y' - \dfrac{c}{a}y \\ y' & = y' \end{cases}$$

which leads to the matrix representation

$$\begin{bmatrix} y'' \\ y' \end{bmatrix} = \begin{bmatrix} -\frac{b}{a} & -\frac{c}{a} \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y' \\ y \end{bmatrix}$$

■

We look at the special case where $\mathbf{g}'(t) = 0$, such that $\mathbf{x}'(t) = \mathbf{A}(t) + \mathbf{x}(t)$. Immediately, we have the following theorem:

The solution set $\{\mathbf{x}(t) \mid \mathbf{x}'(t) = \mathbf{A}(t) + \mathbf{x}(t)\}$ is a vector space. Assuming that $\mathbf{A}(t)$ has continuous entries at a point $t_0 \in \mathbb{R}$, we have a linear transformation

$$\begin{aligned} T : \{\mathbf{x}(t) \mid \mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t)\} &\to \mathbb{R}^n \\ \mathbf{x}(t) &\mapsto \mathbf{x}(t_0) \end{aligned}$$

which is bijective (one-to-one and onto).

corollary $\dim\{\mathbf{x}(t) \mid \mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t)\} = \dim \mathbb{R}^n = n$

which tells us that we need to find n linearly independent solutions (i.e. finite basis) to determine a solution set.

But the question is: how do we determine whether if $\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)$ are *linearly independent* vector-valued functions or not?

We extend the concept of linear (in)dependence to vector-valued functions like such:

A set of $\mathbb{R}^n$-valued functions $\{\mathbf{x}_i\}_{i=1}^n$ is **linearly dependent** if and only if $\exists \lambda_1, \ldots, \lambda_n$ such that at least one $\lambda_i \neq 0$ and $\forall t \in \mathbb{R} : \lambda_1 \mathbf{x}_1(t) + \ldots + \lambda_n \mathbf{x}_n(t) = 0$.

Recall that a test for linear independence for vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in $\mathbb{R}^n$ is to check whether the matrix $\begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix}$ is invertible, which is true if and only if $\det \begin{bmatrix}\mathbf{v}_1 & \cdots & \mathbf{v}_n\end{bmatrix} \neq 0$.

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be $\mathbb{R}^n$-valued functions. The **Wrońskian** is the $\mathbb{R}^n$-valued function

$$W \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_n\end{bmatrix}(t) = \det \begin{bmatrix}\mathbf{x}_1(t) & \cdots & \mathbf{x}_n(t)\end{bmatrix}$$

and we have the following key proposition:

$$\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\} \text{ is linearly dependent} \implies \forall t \in \mathbb{R},\ W \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_n\end{bmatrix}(t) = 0$$

$$\exists t \in \mathbb{R} \text{ such that } W \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_n\end{bmatrix}(t) \neq 0 \implies \{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\} \text{ is linearly independent}$$

proof If $\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\}$ is linearly dependent, then $\exists \lambda_i \neq 0$ such that $\forall t \in \mathbb{R}$, $\lambda_1 \mathbf{x}_1(t) + \cdots + \lambda_n \mathbf{x}_n(t) = 0$. Thus $\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\}$ is linearly dependent for all $t \in \mathbb{R}$, meaning $\forall t \in \mathbb{R}$, $W \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_n\end{bmatrix}(t) = 0$. ■

warning! The converse of the proposition above is false! $\forall t \in \mathbb{R}$, $W \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_n\end{bmatrix}(t) = 0$ does not necessarily imply that $\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\}$ is linearly dependent.

counterexample $\mathbf{x}_1(t) = \begin{bmatrix}1\\1\end{bmatrix}$ and $\mathbf{x}_2(t) = \begin{bmatrix}t\\t\end{bmatrix}$, meaning $\forall t \in \mathbb{R}$, $W \begin{bmatrix}\mathbf{x}_1(t) & \mathbf{x}_2(t)\end{bmatrix}(t) = \det \begin{bmatrix}1 & t\\1 & t\end{bmatrix} = 0$. But for $\lambda_1 \mathbf{x}_1(t) + \lambda_2 \mathbf{x}_2(t) = 0 \iff \lambda_1 + \lambda_2 t = 0$ for all $t \in \mathbb{R}$, then $\lambda_1 = \lambda_2 = 0$, meaning that $\{\mathbf{x}_1(t), \mathbf{x}_2(t)\}$ is linearly independent. ■

How does this relate to $\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t)$?

Suppose that $\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\}$ are solutions to the equation $\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t)$. Then,

- $\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\}$ linearly dependent $\implies \forall t \in \mathbb{R}$, $W \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_n\end{bmatrix}(t) = 0$
- $\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\}$ linearly independent $\implies \forall t \in \mathbb{R}$, $W \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_n\end{bmatrix}(t) \neq 0$

proof Suppose that $\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\}$ are solutions to the equation $\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t)$. Then, if there exists some $t_0 \in \mathbb{R}$ where $W \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_n\end{bmatrix}(t_0) = 0$, then the vectors $\{\mathbf{x}_i(t_0)\}$ are linearly dependent. Thus, there exists a real number $\lambda_i \neq 0$ such that $\lambda_1 \mathbf{x}_1(t_0) + \cdots + \lambda_n \mathbf{x}_n(t_0) = 0$, by definition of linear dependence.

Since the solution set $\{\mathbf{x}(t) \mid \mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t)\}$ is a vector space, then $\lambda_1 \mathbf{x}_1(t) + \cdots + \lambda_n \mathbf{x}_n(t)$ is a solution with the same initial condition as $\mathbf{0}(t)$. By uniqueness of the initial condition ($\mathbf{x}(t) \mapsto \mathbf{x}(t_0)$ is a bijection), $\lambda_1 \mathbf{x}_1(t) + \cdots + \lambda_n \mathbf{x}_n(t) = \mathbf{0}(t)$.

Thus $\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\}$ is linearly dependent implies $W \begin{bmatrix}\mathbf{x}_1 & \cdots & \mathbf{x}_n\end{bmatrix}(t) = 0$ for all $t \in \mathbb{R}$. ■

A linearly independent set $\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\}$ of solutions to $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$ is called a **fundamental solution set**.

In the non-homogeneous case, i.e. when solving $\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{g}(t)$ where $\mathbf{g}(t) \neq 0$, then

Given a particular solution $\mathbf{x}_p$ to $\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{g}(t)$, the **general solution** to $\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{g}(t)$ is

$$\mathbf{x} = \mathbf{x}_p + \lambda_1 \mathbf{x}_1(t) + \cdots + \lambda_n \mathbf{x}_n(t)$$

where $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ and $\{\mathbf{x}_1(t), \ldots, \mathbf{x}_n(t)\}$ is the fundamental solution set.

8.5 Constant-coefficient linear systems

Recall that a system of k differential equations in n variables can be expressed as

$$\mathbf{x}'(t) = \mathbf{A}(t)\mathbf{x}(t)$$

In this section we will be working with equations of such form where $\mathbf{A}(t)$ is a **matrix with real constant entries**, i.e. $\mathbf{A}(t) = \mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \qquad (a_{11}, \dots, a_{nn} \in \mathbb{R})$$

Given a homogeneous differential equation $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$ where the entries of $\mathbf{A}$ are constant, if $(\lambda, \mathbf{v})$ is an eigenvalue–eigenvector pair, then $\mathbf{x}(t) = e^{\lambda t}\mathbf{v}$ is a solution.

proof Since $(\lambda, \mathbf{v})$ is an eigenvalue–eigenvector pair, then by definition, $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, and $\mathbf{v} \neq \mathbf{0}$. Thus, $\mathbf{x}'(t) = \lambda e^{\lambda t}\mathbf{v}$, and $\mathbf{A}\mathbf{x}(t) = \mathbf{A}e^{\lambda t}\mathbf{v} = \lambda e^{\lambda t}\mathbf{v}$. ∎

To find a completely general solution to a homogeneous differential equation, we need a fundamental solution set (i.e. a **basis** for the solution set of the differential equation).

Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis of eigenvectors of $\mathbf{A}$ with respective eigenvalues $\lambda_1, \dots, \lambda_n$. Then, the set of vector-valued functions $\{\mathbf{x}_1(t), \dots, \mathbf{x}_n(t)\}$ where $\mathbf{x}_i(t) = e^{\lambda_i t}\mathbf{v}_i$ is a fundamental solution set.

proof We already know that $\mathbf{x}(t) = e^{\lambda_i t}\mathbf{v}_i$ is a solution to the differential equation $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$. We employ the Wrońskian:

$$\begin{aligned} W\left[\mathbf{x}_1 \cdots \mathbf{x}_n\right](t) &= \det\left[\mathbf{x}_1(t) \quad \cdots \quad \mathbf{x}_n(t)\right] \\ &= \det\left[e^{\lambda_1 t}\mathbf{v}_1 \quad \cdots \quad e^{\lambda_n t}\mathbf{v}_n\right] \\ &= e^{\lambda_1 t}e^{\lambda_2 t}\cdots e^{\lambda_n t}\det\left[\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n\right] \neq 0 \end{aligned}$$

since $\forall t \in \mathbb{R}$, $e^t > 0$, and $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is a basis. Thus, $\{\mathbf{x}_1(t), \dots, \mathbf{x}_n(t)\}$ is linearly independent and therefore a basis for the solution space of $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$. ∎

corollary If $\mathbf{A}$ is diagonalizable, then a general solution to $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$ can be found.

Given an $n \times n$ matrix $\mathbf{A}$ with eigenvalues of the form $\alpha \pm i\beta$ with eigenvalues $\mathbf{a} + i\mathbf{b}$, then

$$\begin{aligned} \mathbf{x}_1(t) &= (e^{\alpha t}\cos\beta t)\mathbf{a} - (e^{\alpha t}\sin\beta t)\mathbf{b} \\ \mathbf{x}_2(t) &= (e^{\alpha t}\sin\beta t)\mathbf{a} + (e^{\alpha t}\cos\beta t)\mathbf{b} \end{aligned}$$

are real, linearly independent solutions to $\mathbf{x}'(t) = \mathbf{A}\mathbf{x}(t)$.

end of differential equations