Mathematics III

Semester 2018-I

Francisco Rosales Marticorena

Faculty of Economics and Finance

Table of Contents:

1 Vectors in $\mathbb{R}^n$
2 Matrices in $\mathbb{R}^{n \times m}$
3 Gaußian Elimination
4 Vector Spaces
5 Linear Transformations
6 Diagonalization
7 Quadratic Forms
8 Block Matrices
9 Complex Numbers
10 Taylor Polys & Taylor Series
11 Topics on Calculus
12 Comparative Statics
13 Ordinary Differential Equations
14 Difference Equations

Vectors in $\mathbb{R}^n$

Definition 1.1 (Vector in $\mathbb{R}^n$)
A vector in the $\mathbb{R}^n$ space is a list of values that represents a magnitude and a direction in such space.

- In this class: we will consider vectors in $\mathbb{R}^n$ whose initial point is the origin, hence a vector only represents magnitude
- Notation: we will represent vectors $\boldsymbol{v} \in \mathbb{R}^n$ as column matrices, i.e

$$\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}_{n \times 1} .$$

We will also use $\boldsymbol{v} = (v_1, \ldots, v_n)^\top$.

Notation:

- scalars: lowercase letters, e.g. $\alpha, \beta, \gamma, \ldots$
- vectors: lowercase bolded letters, e.g. $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}, \ldots$
- matrices: uppercase bolded letters, e.g. $\boldsymbol{A}, \boldsymbol{B}, \ldots$
- number sets: naturals ($\mathbb{N}$), integers ($\mathbb{Z}$), reals ($\mathbb{R}$), complex ($\mathbb{C}$)

Special Vectors:

Consider the following vectors in $\mathbb{R}^n$:

1. **Null vector:**
$$\mathbf{0} = (0, \ldots, 0)^\top$$
2. **Sum vector:**
$$\boldsymbol{s} = (1, \ldots, 1)^\top$$
In fact for some $\boldsymbol{u} \in \mathbb{R}^n$, $\boldsymbol{u}^\top \boldsymbol{s} = \sum_{i=1}^n u_i$
3. **Canonical vector:**
$$\boldsymbol{e}_j = \boldsymbol{I}_j,$$
where $\boldsymbol{I}_j$ denotes the j-th column of the identity matrix. Also called selector vector, since $\boldsymbol{u}^\top \boldsymbol{e}_j = u_j$

Vector Addition and Scalar Multiplication

Consider $\kappa, x_i \in \mathbb{R}$, and $\boldsymbol{a}_i, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$.

- **Addition:** $\boldsymbol{w} = \boldsymbol{u} + \boldsymbol{v}$, where $w_i = u_i + v_i$
- **Scalar Multiplication:** $\boldsymbol{v} = \kappa \boldsymbol{u}$, where $v_i = \kappa u_i$

Moreover, vector

$$\boldsymbol{v} = x_1 \boldsymbol{a}_1 + x_2 \boldsymbol{a}_2 + \cdots + x_m \boldsymbol{a}_m = \sum_{i=1}^{m} x_i \boldsymbol{a}_i,$$

is called a linear combination of vectors $\boldsymbol{a}_i$. We stress that

$$\boldsymbol{v} = \sum_{i=1}^{m} x_i \boldsymbol{a}_i = \boldsymbol{A}\boldsymbol{x},$$

where $\boldsymbol{x} = (x_1, \ldots, x_m)^\top$ and $\boldsymbol{A} = [\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \ldots \boldsymbol{a}_m]$ is an $n \times m$ matrix.

Properties:

Consider $\lambda, \kappa \in \mathbb{R}$ and $\mathbf{0}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$

1. $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$
2. $\boldsymbol{u} + \mathbf{0} = \boldsymbol{u}$
3. $\boldsymbol{u} + (-\boldsymbol{u}) = \mathbf{0}$
4. $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$
5. $\lambda(\boldsymbol{u} + \boldsymbol{v}) = \lambda\boldsymbol{u} + \lambda\boldsymbol{v}$
6. $(\lambda + \kappa)\boldsymbol{u} = \lambda\boldsymbol{u} + \kappa\boldsymbol{u}$
7. $(\lambda\kappa)\boldsymbol{u} = \lambda(\kappa\boldsymbol{u})$
8. $1\boldsymbol{u} = \boldsymbol{u}$

Inner Product

Definition 1.2 (Inner Product)
Consider $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. The inner product between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is denoted by $\boldsymbol{u} \cdot \boldsymbol{v}$ or $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ and is defined as

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u}^\top \boldsymbol{v} = \sum_{i=1}^{n} u_i v_i.$$

The inner product is sometimes called dot product.

Properties:

Consider $\lambda \in \mathbb{R}$ and $\mathbf{0}, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$

1. $\langle (\boldsymbol{u} + \boldsymbol{v}), \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle$
2. $\langle \lambda \boldsymbol{u}, \boldsymbol{v} \rangle = \lambda \langle \boldsymbol{u}, \boldsymbol{v} \rangle$
3. $\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle$
4. $\langle \boldsymbol{u}, \boldsymbol{u} \rangle \geq 0$ and $\langle \boldsymbol{u}, \boldsymbol{u} \rangle = 0$, iff $\boldsymbol{u} = \mathbf{0}$

Norms

A vector norm, denoted $\|\cdot\|$ is a scalar that measures the magnitude of a vector. In this class we will use the Euclidean norm, defined as

$$\|\boldsymbol{u}\|_2 = \left(\sum_{i=1}^{n} u_i^2\right)^{\frac{1}{2}}, \quad \boldsymbol{u} \in \mathbb{R}^n,$$

which is also called L_2 norm. Other well-known norms are:

- L_1 norm: $\|\boldsymbol{u}\|_1 = \sum_{i=1}^{n} |u_i|$
- L_∞ norm: $\|\boldsymbol{u}\|_\infty = \max_{i=1,\ldots,n} |u_i|$
- L_p norm: $\|\boldsymbol{u}\|_p = \left(\sum_{i=1}^{n} u_i^p\right)^{\frac{1}{p}}$, $1 \leq p \leq \infty$

Hereafter we will use $\|\cdot\| := \|\cdot\|_2$ only.

Example 1.1
Show that $\|\boldsymbol{Q}\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$, $\boldsymbol{x} \in \mathbb{R}^n$ *and* $\boldsymbol{Q} \in \mathbb{R}^{n \times n}$ *(see section II) such that* $\boldsymbol{Q}^{-1} = \boldsymbol{Q}^\top$.

From the defintion of the norm,

$$\|\boldsymbol{Q}\boldsymbol{x}\|_2 = (\boldsymbol{x}^\top \boldsymbol{Q}^\top \boldsymbol{Q}\boldsymbol{x})^{1/2} = (\boldsymbol{x}^\top \boldsymbol{Q}^{-1} \boldsymbol{Q}\boldsymbol{x})^{1/2} = (\boldsymbol{x}^\top \boldsymbol{x})^{1/2} = \|\boldsymbol{x}\|_2.$$

Example 1.2

Show that $\|\boldsymbol{x}\|_\infty \le \|\boldsymbol{x}\|_2 \le \sqrt{n}\|\boldsymbol{x}\|_\infty$*, where* $\|\boldsymbol{x}\|_\infty = \max_{1\le i\le n}|x_i|$ *and* $\boldsymbol{x} \in \mathbb{R}^n$*.*

For $\|\boldsymbol{x}\|_\infty \le \|\boldsymbol{x}\|_2$, let $x_k = \max_{1\le i\le n} x_i$, then

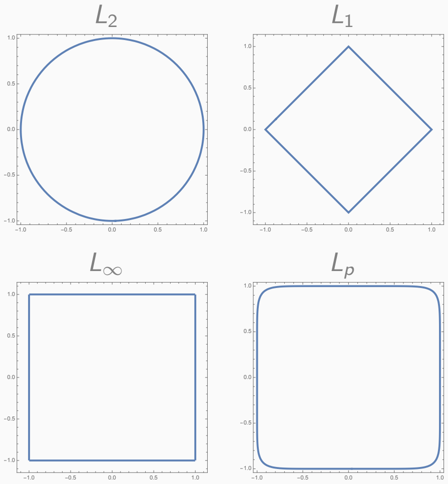
$$\left(\sum_{i=1}^{n} x_i^2\right)^{1/2} = \left(\sum_{i=1, i\neq k}^{n} x_i^2 + x_k^2\right)^{1/2}, \text{ and } \left(\sum_{i=1, i\neq k}^{n} x_i^2 + x_k^2\right) \ge x_k^2 = |x_k|^2$$

which assures that $\|\boldsymbol{x}\|_2 \ge \|\boldsymbol{x}\|_\infty$. For $\|\boldsymbol{x}\|_2 \le \sqrt{n}\|\boldsymbol{x}\|_\infty$, note that

$$\left(\sum_{i=1}^{n} x_i^2\right)^{1/2} \le \left(\sum_{i=1}^{n} x_k^2\right)^{1/2} = \left(n x_k^2\right)^{1/2} = \sqrt{n}|x_k|,$$

then $\|\boldsymbol{x}\|_2 \le \sqrt{n}\|\boldsymbol{x}\|_\infty$.

Unit spheres:



A vector that lies in the unit circle is called unit vector. Any $\boldsymbol{u} \in \mathbb{R}^n$ under $\|\cdot\|_2$ can be turned into a unit vector letting $\boldsymbol{v} = \boldsymbol{u}/(\boldsymbol{u}^\top \boldsymbol{u})^{1/2}$.

Properties:

For $\lambda \in \mathbb{R}$ and $\boldsymbol{u} \in \mathbb{R}^n$, it follows that

1. $\|\boldsymbol{u}\|^2 = \boldsymbol{u}^\top \boldsymbol{u}$
2. $\|\boldsymbol{u}\| \geq 0$ and $\|\boldsymbol{u}\| = 0$, iff $\boldsymbol{u} = \mathbf{0}$
3. $\|\lambda \boldsymbol{u}\| = |\lambda| \|\boldsymbol{u}\|$
4. $\| - \boldsymbol{u}\| = \|\boldsymbol{u}\|$
5. $\|\boldsymbol{u} \pm \boldsymbol{v}\| = 0$, iff $\boldsymbol{u} = \mp \boldsymbol{v}$
6. $\|\boldsymbol{u} \pm \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2 \pm 2\boldsymbol{u}^\top \boldsymbol{v}$

Example 1.3
The parallelogram law states that "the sum of the squares of the lengths of the four sides of a parallelogram equals the sum of the squares of the lengths of the two diagonals". Argue that this is right.

Angle between two Vectors

Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$.The angle θ between $\boldsymbol{u}$ and $\boldsymbol{v}$ can be computed as

$$\theta = \arccos\left(\frac{\boldsymbol{u}^\top \boldsymbol{v}}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}\right), \|\boldsymbol{u}\| > 0, \|\boldsymbol{v}\| > 0,$$

where we have used $\cos(\alpha \mp \beta) = \cos(\alpha)\cos(\beta) \pm \sin(\alpha)\sin(\beta)$.

- The sign of $\boldsymbol{u}^\top \boldsymbol{v}$ is that of $\cos(\theta)$. Thus $\boldsymbol{u}^\top \boldsymbol{v} > 0$ if θ is acute, and $\boldsymbol{u}^\top \boldsymbol{v} < 0$ if θ is obtuse.
- If $\theta = 0$, then $\cos(\theta) = 1 \Rightarrow \boldsymbol{u}^\top \boldsymbol{v} = \|\boldsymbol{u}\|\|\boldsymbol{v}\|$
- If $\theta = \pi$, then $\cos(\theta) = -1 \Rightarrow \boldsymbol{u}^\top \boldsymbol{v} = -\|\boldsymbol{u}\|\|\boldsymbol{v}\|$
- If $\theta = \pi/2$, then $\cos(\theta) = 0 \Rightarrow \boldsymbol{u}^\top \boldsymbol{v} = 0$, $\|\boldsymbol{u} \pm \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 \pm \|\boldsymbol{v}\|^2$

Important results:

Theorem 1.1 (Cauchy-Schwarz inequality)

For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, it holds that

$$(\boldsymbol{u}^\top \boldsymbol{v})^2 \leq \|\boldsymbol{u}\|^2 \|\boldsymbol{v}\|^2$$

Proof.
let $\alpha \in \mathbb{R}$,

$$0 \leq \langle \alpha \boldsymbol{u} + \boldsymbol{v}, \alpha \boldsymbol{u} + \boldsymbol{v} \rangle = \alpha^2 \|\boldsymbol{u}\|^2 + 2\alpha \langle \boldsymbol{u}, \boldsymbol{v} \rangle + \|\boldsymbol{v}\|^2 = a\alpha^2 + b\alpha + c,$$

where $a = \|\boldsymbol{u}\|^2$, $b = 2\langle \boldsymbol{u}, \boldsymbol{v} \rangle$, and $c = \|\boldsymbol{v}\|^2$. Note that the resulting quadratic polynomial cannot have two real roots since $0 \leq a\alpha^2 + b\alpha + c$. Hence $b^2 - 4ac \leq 0$ and $\langle \boldsymbol{u}, \boldsymbol{v} \rangle^2 \leq \|\boldsymbol{u}\|^2 \|\boldsymbol{v}\|^2$. ■

Theorem 1.2 (Triangle inequality (Minkowski's inequality))
For $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, it holds that

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

Proof.

$$\|\boldsymbol{u}+\boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + 2\langle \boldsymbol{u}, \boldsymbol{v}\rangle + \|\boldsymbol{v}\|^2 \leq \|\boldsymbol{u}\|^2 + 2\|\boldsymbol{u}\|\|\boldsymbol{v}\| + \|\boldsymbol{v}\|^2,$$

which follows from theorem 1.1, leading to

$$\|\boldsymbol{u}+\boldsymbol{v}\|^2 \leq (\|\boldsymbol{u}\| + \|\boldsymbol{v}\|)^2, \quad \text{or} \quad \|\boldsymbol{u}+\boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|.$$

$\blacksquare$

Example 1.4

Simulate the path a player makes in the pitch. Build a simple model to accomplish the task.

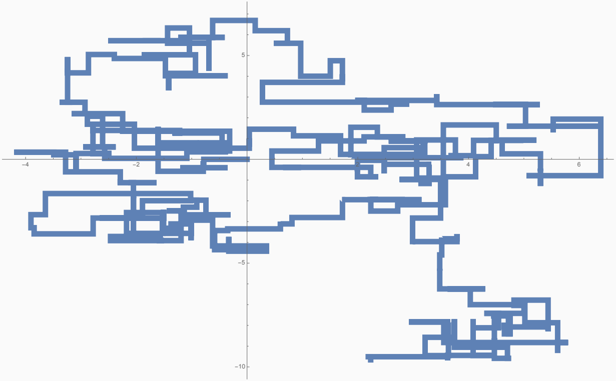
- Consider the case of a player who, for each moment t in time, decides
  1. if he/she moves horizontally (left or right) and vertically (up or down)
  2. how much he/she wants to move at time t (call this δ_t)

For example if he/she starts at the origin and moves horizontally in the first moment and vertically on the second, we obtain that:

$$\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \delta_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \delta_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

i.e a linear combination of selector vectors.

Sample path:



Example 1.5 (Assessment question (2015-II))
Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, and let the angle between them be denoted by $\theta_{xy} \in [0, \pi/2]$. Consider $\beta \in \mathbb{R}$, and define $\boldsymbol{e}(\beta) = \boldsymbol{y} - \beta \boldsymbol{x}$.

- *Find β that minimizes $\|\boldsymbol{e}(\beta)\|^2$.*
- *Let $\hat{\beta} = \arg\min_\beta \|\boldsymbol{e}(\beta)\|^2$. Denote $\hat{\boldsymbol{y}} = \hat{\beta}\boldsymbol{x}$ and $\hat{\boldsymbol{e}} = \boldsymbol{e}(\hat{\beta}) = \boldsymbol{y} - \hat{\boldsymbol{y}}$. Show that $\|\hat{\boldsymbol{e}}\|^2 = \|\boldsymbol{y}\|^2 - \|\hat{\boldsymbol{y}}\|^2$.*
- *Denote $\rho = \|\hat{\boldsymbol{y}}\|^2 / \|\boldsymbol{y}\|^2$. What is the relation between ρ and θ_{xy}?*

Matrices in $\mathbb{R}^{n\times m}$

Definition 2.1 (Matrix)
A matrix of dimension $m \times n$ is a rectangular array of numbers, with m rows and n columns:

$$\boldsymbol{A} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

- Sometimes we write $\boldsymbol{A} = [a_{i,j}]$.
- If $m = 1$, then $\boldsymbol{A}$ is a row matrix. If $n = 1$ then $\boldsymbol{A}$ is a column matrix. If $m = n = 1$, then $\boldsymbol{A}$ is a scalar.

Matrix Addition and Scalar Multiplication

Let $\lambda \in \mathbb{R}$ and $\boldsymbol{A} = [a_{i,j}]$, $\boldsymbol{B} = [b_{i,j}]$, $\boldsymbol{C} = [c_{i,j}]$, be all $m \times n$ matrices

- **Addition:** $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$ where $c_{i,j} = a_{i,j} + b_{i,j}$
- **Scalar multiplication:** $\boldsymbol{B} = \lambda \boldsymbol{A}$ where $b_{i,j} = \lambda a_{i,j}$

Properties:

Let $\alpha, \beta \in \mathbb{R}$, $\mathbf{0}$ be a $m \times n$ matrix of zeroes, and $\boldsymbol{A} = [a_{i,j}]$, $\boldsymbol{B} = [b_{i,j}]$, $\boldsymbol{C} = [c_{i,j}]$, be all $m \times n$ matrices. It is easy to show that

1. $(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C})$
2. $\boldsymbol{A} + \mathbf{0} = \mathbf{0} + \boldsymbol{A} = \boldsymbol{A}$
3. $\boldsymbol{A} + (-\boldsymbol{A}) = \mathbf{0}$
4. $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$
5. $\alpha(\boldsymbol{A} + \boldsymbol{B}) = \alpha\boldsymbol{A} + \alpha\boldsymbol{B}$
6. $(\alpha + \beta)\boldsymbol{A} = \alpha\boldsymbol{A} + \beta\boldsymbol{A}$
7. $(\alpha\beta)\boldsymbol{A} = \alpha(\beta\boldsymbol{A})$
8. $1\boldsymbol{A} = \boldsymbol{A}$

Matrix Multiplication

Let $\boldsymbol{A}$ be a $m \times n$ matrix, $\boldsymbol{B}$ be a $p \times q$ matrix, and consider $n = p$. The entries of the matrix resulting from multiplying $\boldsymbol{A}$ and $\boldsymbol{B}$ are given by

$$c_{i,j} = \sum_{k=1}^{p} a_{i,k} b_{k,j}$$

We stress that the product $\boldsymbol{AB}$ is not defined if $n \neq p$.

The following representations are also useful:

$$\begin{aligned} c_{i,j} &= \langle (\boldsymbol{A}^\top)_i, \boldsymbol{B}_j \rangle, \\ \boldsymbol{C}_j &= \boldsymbol{A}\boldsymbol{B}_j = \sum_{k=1}^{p} b_{k,j} \boldsymbol{A}_k, \end{aligned}$$

where $\boldsymbol{M}_\ell$ denotes the ℓ-th column of matrix $\boldsymbol{M}$. Hence, each col of $\boldsymbol{C}$ is a linear combination of cols of $\boldsymbol{A}$.

Properties:

Let $\lambda \in \mathbb{R}$ and $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ be conformable matrices

1. $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$
2. $\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$
3. $(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$
4. $\lambda(\boldsymbol{AB}) = (\lambda\boldsymbol{A})\boldsymbol{B} = \boldsymbol{A}(\lambda\boldsymbol{B})$

But we stress that:

- $\boldsymbol{AB} \neq \boldsymbol{BA}$
- $\boldsymbol{AB} = \mathbf{0} \not\Rightarrow \boldsymbol{B} = \mathbf{0}$ or $\boldsymbol{A} = \mathbf{0}$
- $\boldsymbol{AB} = \boldsymbol{AC} \not\Rightarrow \boldsymbol{B} = \boldsymbol{C}$

Example 2.1

Say you have a picture as a $n \times n$ *matrix* $\boldsymbol{A}$ *in a grey scale, i.e. with entries* $a_{i,j} \in [0, 1]$, *where* $a_{i,j} = 0$ *means black and* $a_{i,j} = 1$ *means white.*

1. *How would you add light to the picture? (assume that for any transformed matrix* $\tilde{\boldsymbol{A}}$ *it holds that* $\tilde{a}_{i,j} \leq 0$ *means black and* $\tilde{a}_{i,j} \geq 0$ *means white.)*
2. *How can you rotate the picture horizontally by pre or post multiplying* $\boldsymbol{A}$ *by another matrix?*

- Define $\tilde{\boldsymbol{A}} = \lambda \boldsymbol{A}$, for $\lambda > 0$.

$1\boldsymbol{A}$

$1.5\boldsymbol{A}$

$2\boldsymbol{A}$

$2.5\boldsymbol{A}$

$3\boldsymbol{A}$

$3.5\boldsymbol{A}$

- Define $\tilde{\boldsymbol{A}} = \boldsymbol{AB}$, where $\boldsymbol{B}_j = \boldsymbol{I}_{n-j+1}$.

Transpose of a Matrix

The transpose of a matrix, written $\boldsymbol{A}^\top$ is obtained by writing the columns of $\boldsymbol{A}$, in order, as rows. That is, if $\boldsymbol{B} = \boldsymbol{A}^\top$ then $b_{i,j} = a_{j,i}$.

Properties:

Let $\lambda \in \mathbb{R}$ and $\boldsymbol{A}$ and $\boldsymbol{B}$ be conformable matrices.

- $(\boldsymbol{A} + \boldsymbol{B})^\top = \boldsymbol{A}^\top + \boldsymbol{B}^\top$
- $(\boldsymbol{A}^\top)^\top = \boldsymbol{A}$
- $(\lambda \boldsymbol{A})^\top = \lambda \boldsymbol{A}^\top$
- $(\boldsymbol{A}\boldsymbol{B})^\top = \boldsymbol{B}^\top \boldsymbol{A}^\top$

Square Matrices

If $\boldsymbol{A}$ is a $m \times n$ matrix and $m = n$ then $\boldsymbol{A}$ is a square matrix and $a_{i,i}, i = 1, \ldots, n$ are its diagonal entries.

Example 2.2
The n-dimensional identity matrix is given by

$$\boldsymbol{I}_n = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Sometimes one writes $\boldsymbol{I}_n = [\delta_{i,j}]$, where $\delta_{i,j}$ is the Kronecker delta function

$$\delta_{i,j} = \delta(i,j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Non-singular Matrices

Definition 2.2 (Inverse of a Matrix)
An $n \times n$ matrix $\boldsymbol{A}$ is invertible (non-singular) if there exists an $n \times n$ matrix $\boldsymbol{B}$ such that $\boldsymbol{B}\boldsymbol{A} = \boldsymbol{I}_n$

Example 2.3

To find the inverse involves solving for x_i in

$$\underbrace{\begin{bmatrix} x_1 & x_2 \\ x_3 & x_4 \end{bmatrix}}_{\boldsymbol{B}} \cdot \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\boldsymbol{A}} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{I}_2}.$$

Remark 2.1

The inverse of $\boldsymbol{A}$ can be obtained by pre-multiplying a sequence of elementary matrices to the original matrix, i.e.

$$\boldsymbol{E}_k \dots \boldsymbol{E}_2 \boldsymbol{E}_1 \boldsymbol{A} = \boldsymbol{I}, \quad \boldsymbol{E}_k \dots \boldsymbol{E}_2 \boldsymbol{E}_1 = \boldsymbol{B}.$$

Remember that an elementary matrix is matrix that differs from the identity matrix in one elementary row operation (row switching, row multiplication, row addition).

In this particular case, one can find:

$$\underbrace{\begin{bmatrix} -\frac{1}{a} & 0 \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{E}_8} \cdot \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & \frac{a}{bc} \end{bmatrix}}_{\boldsymbol{E}_7} \cdot \underbrace{\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{E}_6} \cdot \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & \frac{bc}{a} \end{bmatrix}}_{\boldsymbol{E}_5} \cdot \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & -\frac{a}{ad-bc} \end{bmatrix}}_{\boldsymbol{E}_4} \cdot$$
$$\underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}}_{\boldsymbol{E}_3} \cdot \underbrace{\begin{bmatrix} -c & 0 \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{E}_2} \cdot \underbrace{\begin{bmatrix} \frac{1}{a} & 0 \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{E}_1} \cdot \underbrace{\begin{bmatrix} a & b \\ c & d \end{bmatrix}}_{\boldsymbol{A}} = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\boldsymbol{I}_2}$$

Many times is more practical to apply Gaußian elimination directly on:

$$\left[\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right] \sim \left[\begin{array}{cc|cc} 1 & 0 & x_1 & x_2 \\ 0 & 1 & x_3 & x_4 \end{array}\right] = \left[\begin{array}{cc|cc} 1 & 0 & \frac{d}{ad-bc} & -\frac{b}{ad-bc} \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{array}\right]$$

The previous is actually a system with four unknowns and four equations.

$$\begin{bmatrix} a & c & 0 & 0 \\ b & d & 0 & 0 \\ 0 & 0 & a & c \\ 0 & 0 & b & d \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \Rightarrow \left[\begin{array}{cccc|c} a & c & 0 & 0 & 1 \\ b & d & 0 & 0 & 0 \\ 0 & 0 & a & c & 0 \\ 0 & 0 & b & d & 1 \end{array}\right] \sim$$

$$\left[\begin{array}{cccc|c} 1 & \frac{c}{a} & 0 & 0 & \frac{1}{a} \\ b & d & 0 & 0 & 0 \\ 0 & 0 & 1 & \frac{c}{a} & 0 \\ 0 & 0 & \frac{b}{d} & 1 & \frac{1}{d} \end{array}\right] \sim \left[\begin{array}{cccc|c} 1 & \frac{c}{a} & 0 & 0 & \frac{1}{a} \\ 0 & 1 & 0 & 0 & -\frac{b}{ad-bc} \\ 0 & 0 & 1 & \frac{c}{a} & 0 \\ 0 & 0 & 0 & 1 & \frac{a}{ad-bc} \end{array}\right] \sim$$

$$\left[\begin{array}{cccc|c} 1 & \frac{c}{a} & 0 & 0 & \frac{1}{a} \\ 0 & 1 & 0 & 0 & -\frac{b}{ad-bc} \\ 0 & 0 & 1 & \frac{c}{a} & 0 \\ 0 & 0 & 0 & 1 & \frac{a}{ad-bc} \end{array}\right] \sim \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & \frac{d}{ad-bc} \\ 0 & 1 & 0 & 0 & -\frac{b}{ad-bc} \\ 0 & 0 & 1 & 0 & -\frac{c}{ad-bc} \\ 0 & 0 & 0 & 1 & \frac{a}{ad-bc} \end{array}\right]$$

Gaußian Elimination

This section is included for completeness. For full proofs on the theorems stated here, see:

Javier Zuñiga (2013). *Precálculo*. Universidad del Pacífico.

Basic Definitions

- Many problems in linear algebra reduce to

$$\underbrace{\begin{bmatrix} a & b & \cdots & a_{1,n} \\ c & d & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}}_{\boldsymbol{A}} \cdot \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}}_{\boldsymbol{b}} \tag{1}$$

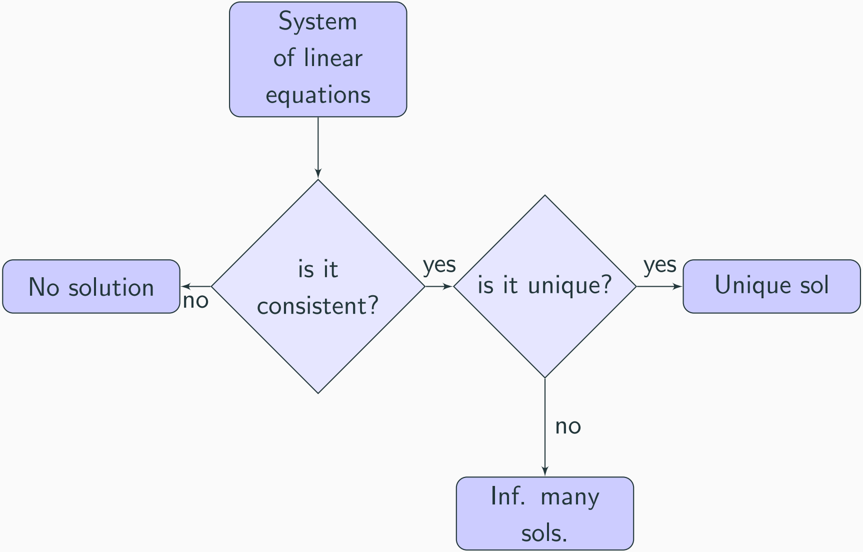
where $\boldsymbol{A}$ is a matrix of coefficients $a_{i,j} \in \mathbb{R}$ and $\boldsymbol{b}$ is a vector of constants $b_i \in \mathbb{R}$, for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, m$.

- Note that we can write (1) as

$$\sum_{j=1}^{n} a_{i,j} x_j = b_i \quad \text{or} \quad \sum_{j=1}^{n} x_j \boldsymbol{A}_j = \boldsymbol{b},$$

  where $\boldsymbol{A}_\ell \in \mathbb{R}^m$ denotes the ℓ-th column of $\boldsymbol{A}$ and $i = 1, 2, \ldots, n$.
- Linear systems can be classified in homogeneous (if $\boldsymbol{b} = \mathbf{0}$) or nonhomogeneous (if $\boldsymbol{b} \neq \mathbf{0}$).

- Linear systems can (also) be classified as



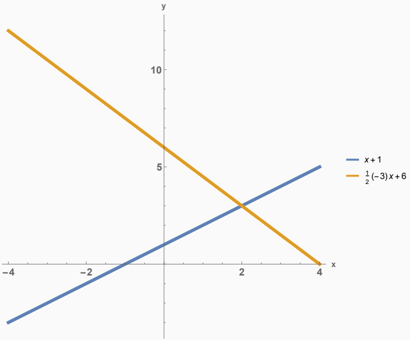
Theorem 3.1

Any system $\mathcal{L}$ of linear equations as in (1) has either

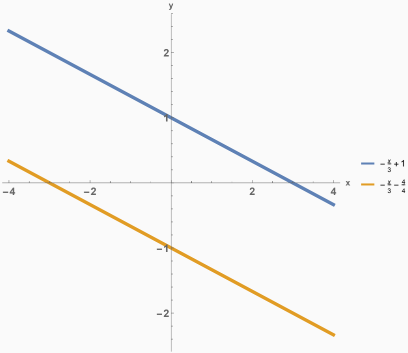
1. *a unique solution*
2. *no solution*
3. *an infinite number of solutions*

Proof.
Omitted. ■

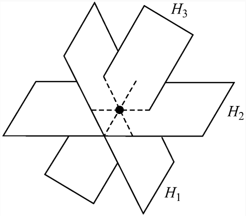
(a)



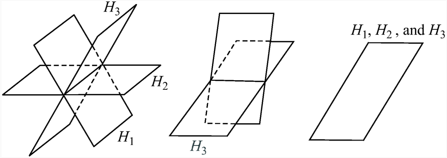
(b)



(c)

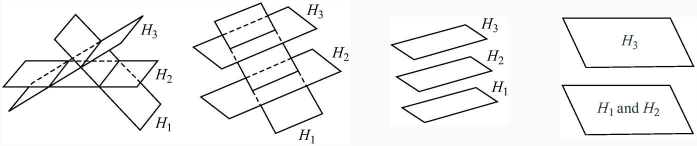


(*a*) Unique solution



(i) (ii) (iii)

(*c*) Infinite number of solutions



(i) (ii) (iii) (iv)

(*b*) No solutions

Definition 3.1 (Echelon form)
A matrix is said to be in echelon form if

1. *all zero rows, if any, are at the bottom of the matrix.*
2. *each leading non-zero entry in a row (pivot) is to the right of the leading non-zero entry in the preceding row.*

Example 3.1

The following matrices are in echelon form and their pivots are in red

$$\boldsymbol{A} = \begin{bmatrix} 2 & 3 & 2 & 0 & 4 & 5 & -6 \\ 0 & 0 & 1 & 1 & -3 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{C} = \begin{bmatrix} 0 & 1 & 3 & 0 & 0 & 4 \\ 0 & 0 & 0 & 1 & 0 & -3 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}.$$

Definition 3.2 (Row Canonical form)
A matrix is said to be in row canonical form if it is in echelon form and

1. *each pivot entry is one.*
2. *each pivot is the only non-zero entry in its column.*

Example 3.2
The matrices from example 3.1 can be written in row canonical form after applying a sequence of row operations.

$$\boldsymbol{A} = \begin{bmatrix} \color{red}{1} & 3/2 & 0 & -1 & 5 & 0 & -19/6 \\ 0 & 0 & \color{red}{1} & 1 & -3 & 0 & -2/3 \\ 0 & 0 & 0 & 0 & 0 & \color{red}{1} & 1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \color{red}{1} & 2 & 0 \\ 0 & 0 & \color{red}{1} \\ 0 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{C} = \begin{bmatrix} 0 & \color{red}{1} & 3 & 0 & 0 & 4 \\ 0 & 0 & 0 & \color{red}{1} & 0 & -3 \\ 0 & 0 & 0 & 0 & \color{red}{1} & 2 \end{bmatrix}.$$

Consider the augmented matrix $\boldsymbol{M} = [\boldsymbol{A} \quad \boldsymbol{b}]$, with $\boldsymbol{A}$ and $\boldsymbol{b}$ as in (1). It can be shown that the following row operations preserve the solution of the system:

1. row interchange: $\boldsymbol{M}_{i,\cdot} \leftrightarrow \boldsymbol{M}_{j,\cdot}$
2. row scaling: $k\boldsymbol{M}_{i,\cdot} \rightarrow \boldsymbol{M}_{i,\cdot}$
3. row addition: $k\boldsymbol{M}_{i,\cdot} + \boldsymbol{M}_{j\cdot} \rightarrow \boldsymbol{M}_{j,\cdot}$,

where "$\leftrightarrow$" reads interchanges and "$\rightarrow$" reads replaces.

Theorem 3.2
Every matrix $\boldsymbol{A}$ is row eq. to <u>some</u> matrix in echelon form, and is row eq. to a <u>unique</u> matrix in row canonical form.

Proof.
Omitted. ∎

Definition 3.3 (Rank)

The rank of a matrix $\boldsymbol{A}$ *denoted* $\operatorname{rank}(\boldsymbol{A})$ *is equal to the* *number of pivots* *in an echelon form of* $\boldsymbol{A}$.

Example 3.3
Find the rank for matrices in example 3.1

$$
\boldsymbol{A} = \begin{bmatrix} 2 & 3 & 2 & 0 & 4 & 5 & -6 \\ 0 & 0 & 1 & 1 & -3 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 6 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}
$$

$$
\boldsymbol{C} = \begin{bmatrix} 0 & 1 & 3 & 0 & 0 & 4 \\ 0 & 0 & 0 & 1 & 0 & -3 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{bmatrix}, \quad \boldsymbol{D} = [\boldsymbol{B} \quad \boldsymbol{C}].
$$

Trivially, $\operatorname{rank}(\boldsymbol{A}) = 3$, $\operatorname{rank}(\boldsymbol{B}) = 2$, $\operatorname{rank}(\boldsymbol{C}) = 3$ *and* $\operatorname{rank}(\boldsymbol{D}) = 3$.

Gaußian Elimination

Gaußian elimination is an algorithm that takes any matrix $\boldsymbol{M}$ (as input) and returns an its row canonical form (as output). It uses:

1. Forward Elimination: takes $\boldsymbol{M}$ and returns the echelon form of $\boldsymbol{M}$
2. Backward Elimination: takes the echelon form of $\boldsymbol{M}$ and returns the row canonical form of $\boldsymbol{M}$

In fact, solving $\boldsymbol{Ax} = \boldsymbol{b}$ is equivalent to finding the row canonical form of $\boldsymbol{M} = [\boldsymbol{A} \quad \boldsymbol{b}]$. The justification follows from these facts:

1. any elementary row operation on $\boldsymbol{M}$ is equivalent to applying the corresponding operation on the system itself.
2. the system has a solution iff the echelon form of the augmented matrix does not have a row of the form $(0, \ldots, 0, b_i)$ with $b_i \neq 0$.
3. in the row canonical form of $\boldsymbol{M}$ (excluding zero rows) the coefficient of each basic variable is a pivot and it is equal to one, and it is the only non-zero entry in its respective column; hence the free variable form of the solution of the system of linear equations is obtained by simply transferring the free variables to the other side.

Forward Elimination

Step 1

1. find the first column with a non-zero entry. Call it j_1
2. interchange rows so that $a_{1,j} \neq 0$
3. use a_{1,j_1} as pivot to obtain zeros below a_{1,j_1}
   1. set $m = -\frac{a_{i,j_1}}{a_{1,j_1}}$
   2. for $i > 1$ do $\boldsymbol{M}_{i,\cdot} \to m\boldsymbol{M}_{1,\cdot} + \boldsymbol{M}_{i,\cdot}$

Step 2 repeat step 1 with the submatrix formed by all the rows excluding the first row. Let j_2 denote the first column in the subsystem with a non-zero entry. Hence $a_{2,j_2} \neq 0$

Step 3 to r repeat step 1 until the remaining submatrix has only zero rows.

Note that in the resulting echelon form matrix, the pivots will be $a_{1,j_1}, \ldots, a_{r,j_r}$, where r is the rank of $\boldsymbol{M}$.

Backward Elimination

Step 1

1. use row scaling so that the last pivot is one, $\frac{1}{a_{r,j_r}}\boldsymbol{M}_{r,\cdot} \to \boldsymbol{M}_{r,\cdot}$
2. use $a_{r,j_r} = 1$ to obtain zeroes above the last pivot, $-a_{i,j_r}\boldsymbol{M}_{r,\cdot} + \boldsymbol{M}_{i,\cdot} \to \boldsymbol{M}_{i,\cdot}$

Step 2 repeat step 1 for $\boldsymbol{M}_{r-1,\cdot}, \ldots, \boldsymbol{M}_{2,\cdot}$

Step r use row scaling so that the pivot in the first row is one.

Example 3.4

Solve the following system

$$
\begin{aligned}
x_1 + x_2 - 2x_3 + 4x_4 &= 5 \\
2x_1 + 2x_2 - 3x_3 + x_4 &= 3 \\
3x_1 + 3x_2 - 4x_3 - 2x_4 &= 1
\end{aligned}
\tag{2}
$$

$$
\begin{aligned}
\boldsymbol{M} &= \begin{bmatrix} 1 & 1 & -2 & 4 & 5 \\ 2 & 2 & -3 & 1 & 3 \\ 3 & 3 & -4 & -2 & 1 \end{bmatrix} \\
&\sim \begin{bmatrix} 1 & 1 & -2 & 4 & 5 \\ 0 & 0 & 1 & -7 & -7 \\ 0 & 0 & 2 & -14 & -14 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 0 & -10 & -9 \\ 0 & 0 & 1 & -7 & -7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}
\end{aligned}
$$

Hence the system has <u>infinite solutions</u>, with free variables x_2, x_4*; and pivot variables* x_1, x_3.

Example 3.5
Solve the following system

$$\begin{aligned} x_1 + x_2 - 2x_3 + 3x_4 &= 4 \\ 2x_1 + 3x_2 + 3x_3 - x_4 &= 3 \\ 5x_1 + 7x_2 + 4x_3 + x_4 &= 5 \end{aligned} \qquad (3)$$

$$\begin{aligned} \boldsymbol{M} &= \begin{bmatrix} 1 & 1 & -2 & 3 & 4 \\ 2 & 3 & 3 & -1 & 3 \\ 5 & 7 & 4 & 1 & 5 \end{bmatrix} \\ &\sim \begin{bmatrix} 1 & 1 & -2 & 3 & 4 \\ 0 & 1 & 7 & -7 & -5 \\ 0 & 2 & 14 & -14 & 15 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & -2 & 3 & 4 \\ 0 & 1 & 7 & -7 & -5 \\ 0 & 0 & 0 & 0 & -5 \end{bmatrix} \end{aligned}$$

Hence the system has <u>no solution</u>. It contains a degenerate equation.

Example 3.6

Solve the following system

$$\begin{aligned} x_1 + 2x_2 + 1x_3 &= 3 \\ 2x_1 + 5x_2 - 1x_3 &= -4 \\ 3x_1 - 2x_2 - 1x_3 &= 5 \end{aligned} \qquad (4)$$

$$\begin{aligned} \boldsymbol{M} &= \begin{bmatrix} 1 & 2 & 1 & 3 \\ 2 & 5 & -1 & -4 \\ 3 & -2 & -1 & 5 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & -3 & -10 \\ 0 & -8 & -4 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & -3 & -10 \\ 0 & 0 & -28 & -84 \end{bmatrix} \\ &\sim \begin{bmatrix} 1 & 2 & 1 & 3 \\ 0 & 1 & -3 & -10 \\ 0 & 0 & 1 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \end{bmatrix} \end{aligned}$$

Hence the system has a <u>unique solution</u>.

Existence and Uniqueness

Theorem 3.3

Consider a system of equations with n unknowns s.t.: $\boldsymbol{M} = [\boldsymbol{A} \quad \boldsymbol{b}]$, *then*

1. *the system has a solution iff* $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{M})$
2. *the solution is unique iff* $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{M}) = n$

Proof.

Omitted. ∎

Theorem 3.4

The square system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ *has a unique solution iff* $\boldsymbol{A}$ *is invertible. In such cases* $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$ *is the unique solution of the system.*

Proof.

Omitted. ∎

Theorem 3.5

The following propositions are equivalent:

- $\det(\boldsymbol{A}) \neq 0$
- $\boldsymbol{A}$ *is invertible*
- $\boldsymbol{A}$ *is full rank*
- *the columns of* $\boldsymbol{A}$ *are linearly independent*
- *the rows of* $\boldsymbol{A}$ *are linearly independent*
- $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ *has unique solution at* $\boldsymbol{x} = \boldsymbol{0}$

Proof.

Omitted. ■

Vector Spaces

Definition

Definition 4.1 (Vector space)

Let $\mathcal{V}$ be a set closed under:

- **Vector addition:** *if $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{V}$, then $\boldsymbol{a} + \boldsymbol{b} \in \mathcal{V}$*
- **Scalar multiplication:** *if $\boldsymbol{a} \in \mathcal{V}$, $k \in \mathbb{R}$, then $k\boldsymbol{a} \in \mathcal{V}$,*

then $\mathcal{V}$ is called a vector space if for $k, \kappa \in \mathbb{R}$ and $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{w} \in \mathcal{V}$:

1. $(\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{w} = \boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{w})$
2. $\exists \mathbf{0} \in \mathcal{V}$ *s.t.* $\boldsymbol{b} + \mathbf{0} = \boldsymbol{b}$, *for each* $\boldsymbol{b} \in \mathcal{V}$
3. *For each* $\boldsymbol{b} \in \mathcal{V}$, $\exists -\boldsymbol{b} \in \mathcal{V}$ *s.t.* $\boldsymbol{b} + (-\boldsymbol{b}) = \mathbf{0}$
4. $\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{b} + \boldsymbol{a}$
5. $k(\boldsymbol{b} + \boldsymbol{w}) = k\boldsymbol{b} + k\boldsymbol{w}$
6. $(k + \kappa)\boldsymbol{b} = k\boldsymbol{b} + \kappa\boldsymbol{b}$
7. $(k\kappa)\boldsymbol{b} = k(\kappa\boldsymbol{b})$
8. $1\boldsymbol{a} = \boldsymbol{a}$

Example 4.1
The following are vector spaces:

- $\mathbb{R}^n$: *space of n-dimensional vectors*
- $\mathbb{R}^{m\times n}$: *space of $m \times n$ dimensional matrices*
- $\mathcal{P}_n(x)$: *the space of polynomials of degree $\leq n$, i.e.*

$$\mathcal{P}_n(x) = \left\{ p(x) = \sum_{i=0}^{s} a_i x^i, a_i \in \mathbb{R} : s \leq n \right\}$$

Example 4.2

For the following sets

$$\begin{aligned}\mathcal{V} &= \{\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{a} \in \mathbb{R}^n : \boldsymbol{a}^\top \boldsymbol{x} = 0\} \\ \mathcal{V}^* &= \{\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{a} \in \mathbb{R}^n, \delta \neq 0 : \boldsymbol{a}^\top \boldsymbol{x} = \delta\}\end{aligned}$$

Show that $\mathcal{V}$ is a vector space, while $\mathcal{V}^*$ is not.

Linear Combinations

Definition 4.2 (Linear combination)
Let $\mathcal{V}$ be a vector space, a vector $\boldsymbol{b}$ is a linear combination (lc) of vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathcal{V}$ if there exist scalars $x_1, \ldots, x_m$ s.t.

$$\boldsymbol{b} = x_1 \boldsymbol{a}_1 + \cdots + x_m \boldsymbol{a}_m = \sum_{i=1}^{m} x_i \boldsymbol{a}_i$$

Definition 4.3 (Linear combination)
We say $\boldsymbol{b}$ is a linear combination of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathcal{V}$ if there is a solution to the equation $\boldsymbol{b} = x_1 \boldsymbol{a}_1 + \cdots + x_m \boldsymbol{a}_m$, where x_i are unknown scalars.

Remark 4.1
The question of expressing a given vector $\boldsymbol{b} \in \mathbb{R}^n$ as a lc of $\boldsymbol{a} \ldots, \boldsymbol{a}_m \in \mathbb{R}^n$ is equivalent to the question of solving $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.

Such a system may have

- *unique solution*
- *many solutions*
- *no solution.*

The no-solution case means that $\boldsymbol{b}$ cannot be written as lc of the $\boldsymbol{a}$'s.

Example 4.3

Express $\boldsymbol{b} = (3, 7, -4)^\top$ *as a lc of*

$$\boldsymbol{a}_1 = (1, 2, 3)^\top, \quad \boldsymbol{a}_2 = (2, 3, 7)^\top \quad \text{and } \boldsymbol{a}_3 = (3, 5, 6)^\top$$

The augmented matrix for $\boldsymbol{Ax} = \boldsymbol{b}$ *reads*

$$\boldsymbol{M} = \begin{bmatrix} 1 & 2 & 3 & 3 \\ 2 & 3 & 5 & 7 \\ 3 & 7 & 6 & -4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -4 \\ 0 & 0 & 1 & 3 \end{bmatrix}$$

Gaußian elimination leads to: $\boldsymbol{b} = 2\boldsymbol{a}_1 - 4\boldsymbol{a}_2 + 3\boldsymbol{a}_3$.

Example 4.4
Express the polynomial $q(x) = 3x^2 + 5x - 5$ *as a lc of*

$$p_1(x) = x^2 + 2x + 1, \qquad p_2(x) = 2x^2 + 5x + 4 \quad \textit{and} \quad p_3(x) = x^2 + 3x + 6$$

The augmented matrix for $\boldsymbol{Ax} = \boldsymbol{b}$ *reads*

$$\boldsymbol{M} = \left[\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 2 & 5 & 3 & 5 \\ 1 & 4 & 6 & -5 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -2 \end{array}\right]$$

Gaußian elimination leads to: $q(x) = 3p_1(x) + p_2(x) - 2p_3(x)$.

Subspaces

Definition 4.4 (Subspace)
Let $\mathcal{W} \subset \mathcal{V}$. Then $\mathcal{W}$ is a subspace of $\mathcal{V}$ if it is itself a vector space.

Remark 4.2
Formally, one would need to show that $\mathcal{W} \subset \mathcal{V}$ and that the conditions in definition 4.1 hold. In addition, note that any vector space has always two trivial subspaces, i.e. $\mathcal{V}, \mathbf{0} \subset \mathcal{V}$ and $\mathcal{V} \sqcup \mathbf{0} = \mathcal{V}$.

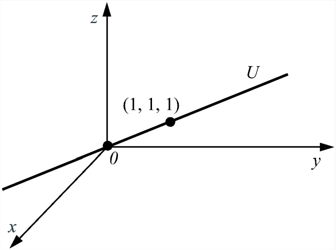
Example 4.5

1. *Show that set* $\mathcal{U} \subset \mathbb{R}^3$

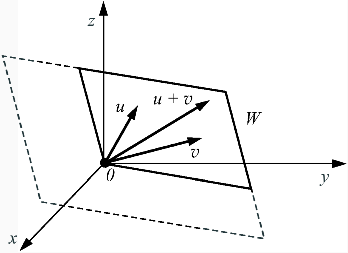
$$\mathcal{U} = \{(x_1, x_2, x_3) : x_1 = x_2 = x_3, x_i \in \mathbb{R}\}$$

   is a subspace of $\mathbb{R}^3$*.*

2. *Show that any plane that goes through the origin in* $\mathbb{R}^3$ *is a subspace of* $\mathbb{R}^3$*.*



(*a*)



(*b*)

Example 4.6
Show that

$$\mathcal{W} = \{\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n : \boldsymbol{a}^\top \boldsymbol{x} = 0, \boldsymbol{b}^\top \boldsymbol{x} = 0\},$$

is a vector subspace of $\mathcal{V}$, as defined in example 4.2.

Example 4.7
Let $\mathcal{V} = \mathbb{R}^{n\times n}$,

1. *Show that $\mathcal{W} \subset \mathcal{V}$ is a subspace of $\mathcal{V}$, where $\mathcal{W}$ is the set of all upper triangular matrices $n \times n$.*
2. *Show that $\mathcal{W} \subset \mathcal{V}$ is a subspace of $\mathcal{V}$, where $\mathcal{W}$ is the set of all symmetric matrices $n \times n$.*

Example 4.8
Let $\mathcal{P}(x)$ represent the space of all polynomials, then

1. *Show that $\mathcal{P}_n(x)$, is a subspace of $\mathcal{P}(x)$, where $\mathcal{P}_n(x)$ denotes the space of polynomials of degree at most n.*
2. *Show that $\mathcal{Q}(x)$ is a subspace of $\mathcal{P}(x)$, where $\mathcal{Q}(x)$ denotes the space of polynomials with only even powers of x.*

Theorem 4.1 (Intersection of subspaces)

The intersection of any number of subspaces of a vector space $\mathcal{V}$ is a subspace of $\mathcal{V}$.

Proof.

Without loss of generality, let $\mathcal{U}$ and $\mathcal{W}$ be subspaces of vector space $\mathcal{V}$. Now suppose $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{U} \cap \mathcal{W}$. Then $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{U}$ and $\boldsymbol{a}, \boldsymbol{b} \in \mathcal{W}$. Further, since $\mathcal{U}$ and $\mathcal{W}$ are subspaces, for any $\alpha, \beta \in \mathbb{R}$, $\alpha \boldsymbol{a} + \beta \boldsymbol{b} \in \mathcal{U}$ and $\alpha \boldsymbol{a} + \beta \boldsymbol{b} \in \mathcal{W}$. Thus $\alpha \boldsymbol{a} + \beta \boldsymbol{b} \in \mathcal{U} \cap \mathcal{W}$, and $\mathcal{U} \cap \mathcal{W}$ is subspace of $\mathcal{V}$. ∎

Definition 4.5 (Nullspace of a Matrix)
Let $\boldsymbol{A}$ be a $m \times n$ matrix, the Nullspace of matrix $\boldsymbol{A}$, is defined as

$$null(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \mathbf{0}\}.$$

Remark 4.3
Note that $null(\boldsymbol{A})$ is a subspace of $\mathbb{R}^n$ since i) $null(\boldsymbol{A}) \subset \mathbb{R}^n$; and ii) Given $\boldsymbol{a}, \boldsymbol{b} \in null(\boldsymbol{A})$ and $\alpha, \beta \in \mathbb{R}$ it holds that

$$\boldsymbol{A}(\alpha\boldsymbol{a} + \beta\boldsymbol{b}) = \alpha(\boldsymbol{A}\boldsymbol{a}) + \beta(\boldsymbol{A}\boldsymbol{b}) = \mathbf{0}_m,$$

which implies that $(\alpha\boldsymbol{a} + \beta\boldsymbol{b}) \in null(\boldsymbol{A})$.

Linear span

Definition 4.6 (Linear span)
Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathcal{V}$. The collection of all lc of such vectors, denoted by

$$\text{span}(\boldsymbol{a}_i),$$

is called linear span of $\boldsymbol{a}_1 \ldots, \boldsymbol{a}_m$.

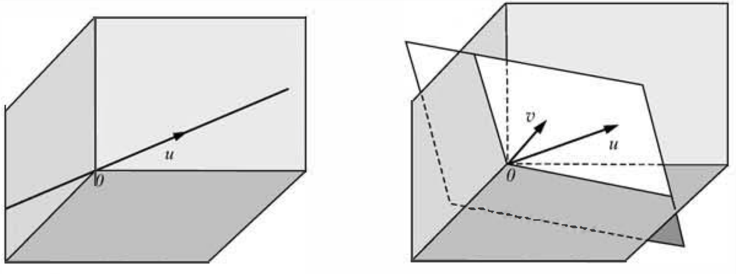
Remark 4.4
Note that span($\boldsymbol{a}_i$) is a subspace of $\mathcal{V}$. By construction, if $\mathcal{V}$ is a vector space, then span($\boldsymbol{a}_i$) $\subset \mathcal{V}$, and for $\boldsymbol{a}, \boldsymbol{b} \in$ span(a_i),

$$\begin{aligned}
\boldsymbol{a} &= x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_m\boldsymbol{a}_m \\
\boldsymbol{b} &= y_1\boldsymbol{a}_1 + y_2\boldsymbol{a}_2 + \cdots + y_m\boldsymbol{a}_m \\
\alpha\boldsymbol{a} + \beta\boldsymbol{b} &= \{(\alpha x_1 + \beta y_1)\boldsymbol{a}_1 + \cdots + (\alpha x_m + \beta y_m)\boldsymbol{a}_m\} \in \textit{span}(\boldsymbol{a}_i),
\end{aligned}$$

for some $\alpha, \beta \in \mathbb{R}$.

Example 4.9
Consider vector space $\mathcal{V} = \mathbb{R}^3$,

1. *Let $\boldsymbol{u} \neq \mathbf{0}, \boldsymbol{u} \in \mathcal{V}$. Then span$(\boldsymbol{u})$ consists of all scalar multiples of $\boldsymbol{u}$. Geometrically span$(\boldsymbol{u})$ is the line through the origin $\mathcal{O}$ and the end point $\boldsymbol{u}$.*
2. *Let $\boldsymbol{u}, \boldsymbol{a} \in \mathcal{V}$ that are not multiples of each other. Then span$(\boldsymbol{u}, \boldsymbol{a})$ is the plane through the origin $\mathcal{O}$ and the end points $\boldsymbol{u}$ and $\boldsymbol{v}$.*
3. *Consider $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$. Then, span$(\boldsymbol{e}_i) = \mathbb{R}^3$.*



(a) (b)

Definition 4.7 (Range of a matrix)
Let $\boldsymbol{A}$ be a $m \times n$ matrix, the range of matrix $\boldsymbol{A}$, is defined as

$$range(\boldsymbol{A}) = \{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n\},$$

i.e., the space spanned by the columns of $\boldsymbol{A}$ (recall that $\boldsymbol{y} = \sum_{i=1}^{m} x_i \boldsymbol{a}_i$). The range of $\boldsymbol{A}$ is also called the column space of $\boldsymbol{A}$

Remark 4.5
Note that $range(\boldsymbol{A})$ is a subspace of $\mathbb{R}^m$, since i) $range(\boldsymbol{A}) \subset \mathbb{R}^m$ and ii) Given $\boldsymbol{a} = \boldsymbol{A}\boldsymbol{x}_1$ and $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}_2$, such that $\boldsymbol{a}, \boldsymbol{b} \in range(\boldsymbol{A})$, for $\alpha, \beta \in \mathbb{R}$

$$\alpha\boldsymbol{a} + \beta\boldsymbol{b} = \boldsymbol{A}(\alpha\boldsymbol{x}_1 + \beta\boldsymbol{x}_2),$$

which implies that $(\alpha\boldsymbol{a} + \beta\boldsymbol{b}) \in range(\boldsymbol{A})$.

Definition 4.8 (Linear dependence)
$\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathcal{V}$ are linearly dependent if there exist scalars $x_1, \ldots, x_m$, not all of them 0, such that

$$x_1\boldsymbol{a}_1 + \cdots + x_m\boldsymbol{a}_m = \mathbf{0}.$$

Otherwise we say the vectors are linearly independent.

Remark 4.6
Note that if there exists some $x_j \neq 0$, we can write

$$\frac{x_1}{x_j}\boldsymbol{a}_1 + \cdots + \frac{x_{j-1}}{x_j}\boldsymbol{a}_j + \frac{x_j}{x_j}\boldsymbol{a}_j + \frac{x_{j+1}}{x_j}\boldsymbol{a}_j + \cdots + \frac{x_m}{x_j}\boldsymbol{a}_m = \mathbf{0}$$

$$\sum_{i=1, i\neq j}^{m} x_i\boldsymbol{a}_i = \boldsymbol{a}_j,$$

which means that we can write $\boldsymbol{a}_j$ as a linear combination of the others.

Remark 4.7
Let $\mathcal{S} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m\}$, note that

1. *If $\boldsymbol{a}_j = \mathbf{0}$ for some j, then the vectors must be ld since*

   $$\exists \alpha \neq 0 : 0\boldsymbol{a}_1 + 0\boldsymbol{a}_2 + \cdots + 0\boldsymbol{a}_{j-1} + \alpha\boldsymbol{a}_j + 0\boldsymbol{a}_{j+1} + \cdots + 0\boldsymbol{a}_m = \mathbf{0}$$

2. *If $m = 1$, then $\boldsymbol{a}_1$ by itself is li.*
3. *If $\boldsymbol{a}_i = \alpha\boldsymbol{a}_j$, then the vectors must be ld. Without loss of generality, let $i = 1$, $j = 2$ and note that*

   $$1\boldsymbol{a}_1 + (-\alpha)\boldsymbol{a}_2 + 0\boldsymbol{a}_3 + \cdots + 0\boldsymbol{a}_m = \mathbf{0}$$

4. *Two vectors $\boldsymbol{a}_i$ and $\boldsymbol{a}_j$ are ld iff one is a multiple of the other.*
5. *If $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m\}$ is a li set, then any rearrangement of the vectors is li as well.*
6. *If $\mathcal{S}$ is li, then any $\mathcal{W} \subset \mathcal{S}$ is li as well.*

Example 4.10

1. *Show that the vectors*

$$\boldsymbol{u} = (1,1,0)^\top, \boldsymbol{a} = (1,3,2)^\top \text{ and } \boldsymbol{w} = (4,9,5)^\top$$

do <u>not</u> form a li set.

2. *Show that*

$$\boldsymbol{u} = (1,2,3)^\top, \boldsymbol{a} = (2,5,7)^\top \text{ and } \boldsymbol{w} = (1,3,5)^\top$$

are li.

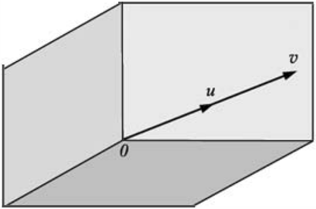
3. *Let $\mathcal{V}$ be the vector space of functions $f_i : \mathbb{R} \to \mathbb{R}$. Show that the functions $f_1(x) = \sin(x)$, $f_2(x) = e^x$ and $f_3(x) = x^2$ are li.*

4. *Show that every set $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{R}^n$, of mutually orthogonal non-null vectors vectors, is li.*

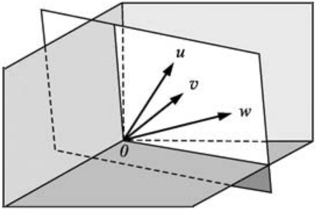
Linear dependence in $\mathbb{R}^3$:

1. If $\boldsymbol{a}_1, \boldsymbol{a}_2 \in \mathbb{R}^3$ lie in the same line, then they are ld.
2. If $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3 \in \mathbb{R}^3$ lie on a plane, then they are ld.
3. If $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3 \in \mathbb{R}^2$, then they are ld.

More generally, a set of $m > n$ vectors in $\mathbb{R}^n$, is necessarily ld.



(*a*) *u* and *v* are linearly dependent



(*b*) *u*, *v*, and *w* are linearly dependent

Linear dependence and linear combinations:

Both notions are closely related.

1. Assume $\boldsymbol{a}_k$ is a lc of the other vectors in $\boldsymbol{a}_1 \ldots, \boldsymbol{a}_m$

$$\begin{aligned} \boldsymbol{a}_k &= x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_{k-1}\boldsymbol{a}_{k-1} + x_{k+1}\boldsymbol{a}_{k+1} + \cdots + x_m\boldsymbol{a}_m \\ \mathbf{0} &= x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_{k-1}\boldsymbol{a}_{k-1} + (-1)\boldsymbol{a}_k + x_{k+1}\boldsymbol{a}_{k+1} + \cdots + x_m\boldsymbol{a}_m \end{aligned}$$

   implying that the vectors are ld.

2. Assume the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ are ld

$$\begin{aligned} y_1\boldsymbol{a}_1 &+ y_2\boldsymbol{a}_2 + \cdots + y_m\boldsymbol{a}_m = \mathbf{0}, \text{for some } y_k \neq 0 \\ \boldsymbol{a}_k &= y_k^{-1}y_1\boldsymbol{a}_1 + \cdots + y_k^{-1}y_{k-1}\boldsymbol{a}_{k-1} + y_k^{-1}y_{k+1}\boldsymbol{a}_{k+1} + y_k^{-1}y_m\boldsymbol{a}_m, \end{aligned}$$

   implying that there is one vector that is a lc of the others

Linear dependence and echelon matrices:

Consider echelon matrix $\boldsymbol{A}$, with pivots in red, and rows $\boldsymbol{r}_\ell$

$$\boldsymbol{A} = \begin{bmatrix} 0 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0 & 0 & 4 & 3 & 2 & 3 & 4 \\ 0 & 0 & 0 & 0 & 7 & 8 & 9 \\ 0 & 0 & 0 & 0 & 0 & 6 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- r_1 cannot be expressed as a lc of r_2, r_3, r_4
- The nonzero rows of a matrix in echelon form are linearly independent.

Theorem 4.2
Let $\boldsymbol{A}$ be a $m \times n$ matrix, with columns $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_m \in \mathbb{R}^m$, then if

- $m < n$, *the cols of $\boldsymbol{A}$ are ld*
- $m = n$, *the cols of $\boldsymbol{A}$ are li iff* $\operatorname{rank}(\boldsymbol{A}) = n$
- $m > n$, *the cols of $\boldsymbol{A}$ are li iff* $\operatorname{rank}(\boldsymbol{A}) = n$

Proof.
Omitted. ■

Remark 4.8
Note that for square matrices ($m = n$), the condition of $\operatorname{rank}(\boldsymbol{A}) = n$ *is equivalent to* $\det(\boldsymbol{A}) \neq 0$, *which means that matrix $\boldsymbol{A}$ must be invertible.*

Basis and Dimension

Definition 4.9 (Basis)

A set $\mathcal{S} = \{\boldsymbol{a}_1, \dots, \boldsymbol{a}_n\}$ is a basis of $\mathcal{V}$ if $\mathcal{S}$ is li and every $\boldsymbol{b} \in \mathcal{V}$ can be written as a lc of $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$.

Definition 4.10 (Basis)

A set $\mathcal{S} = \{\boldsymbol{a}_1, \dots, \boldsymbol{a}_n\}$ is a basis of $\mathcal{V}$ if every $\boldsymbol{b} \in \mathcal{V}$ can be written uniquely as a lc of $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$.

Remark 4.9
We will use the expression "$\mathcal{S}$ spans $\mathcal{V}$" if every $\boldsymbol{b} \in \mathcal{V}$ can be written as a lc of $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$,

Theorem 4.3

Definitions 4.9 and 4.10 are equivalent.

Proof.

Consider the statements: [1] $\mathcal{S} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$ is li and $\mathcal{S}$ spans $\mathcal{V}$; and [2] every $\boldsymbol{b} \in \mathcal{V}$ can be written uniquely as a lc of $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$

- Suppose [1] holds, then $\boldsymbol{b} = x_1\boldsymbol{a}_1 + \cdots + x_n\boldsymbol{a}_n$ and $\boldsymbol{b} = y_1\boldsymbol{a}_1 + \cdots + y_n\boldsymbol{a}_n$, hence

$$\mathbf{0} \quad = \quad (x_1 - y_1)\boldsymbol{a}_1 + \cdots + (x_n - y_n)\boldsymbol{a}_n,$$

  but since $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$ is a li set, the last equality only holds for $x_i = y_i$. Thus [1] implies [2].
- Suppose [2] holds, then $\mathcal{S}$ spans $\mathcal{V}$. Now assume $\exists z_i \neq 0 : \mathbf{0} = z_1\boldsymbol{a}_1 + \cdots + z_n\boldsymbol{a}_n$, but by [2] the representation $\mathbf{0} = 0\boldsymbol{a}_1 + \cdots + 0\boldsymbol{a}_n$ is unique. Thus [2] implies [1].

■

Special cases:

Definition 4.11 (Orthogonal basis in $\mathbb{R}^n$)
A basis $\mathcal{S} = \boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ *is orthogonal if* $\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle = 0$ *for* $i \neq j$

Definition 4.12 (Orthonormal basis in $\mathbb{R}^n$)
A basis $\mathcal{S} = \boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ *is orthonormal if* $\langle \boldsymbol{a}_i, \boldsymbol{a}_j \rangle = \delta_{ij}$

Example 4.11

Show that the basis $\mathcal{S} = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3\}$, where

$$\boldsymbol{a}_1 = (1, 1, 0)^\top, \boldsymbol{a}_2 = (-1, 1, 2)^\top, \text{ and } \boldsymbol{a}_3 = (1, -1, 1)^\top,$$

is an orthogonal basis in $\mathbb{R}^3$.

Lemma 4.1 (Steinitz)

Suppose $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$ *spans* $\mathcal{V}$*, and suppose* $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m\}$ *is li. Then* $m \leq n$*, and* $\mathcal{V}$ *is spanned by a set of the form:*

$$\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m, \boldsymbol{a}_{\ell_1}, \ldots, \boldsymbol{a}_{\ell_{n-m}}\},$$

where, $\boldsymbol{a}_{\ell_i}$ *is used to indicate that vectors* $\boldsymbol{a}_i$ *can be in a different order.*

Proof.

Since $\{\boldsymbol{a}_i\}$ spans $\mathcal{V}$ we have that $\{\boldsymbol{b}_1, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$, also spans $\mathcal{V}$ and is ld because one of the vectors is a lc of the preceding vectors. This vector cannot be $\boldsymbol{b}_1$, so it must be some $\boldsymbol{a}_j$. Thus we obtain the spanning set

$$\{\boldsymbol{b}_1, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_{j-1}, \boldsymbol{a}_{j+1}, \ldots, \boldsymbol{a}_n\}.$$

Now we repeat the argument by adding vector $\boldsymbol{b}_2$ and removing vector $\boldsymbol{a}_k$ to obtain the spanning set

$$\{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_{j-1}, \boldsymbol{a}_{j+1}, \ldots, \boldsymbol{a}_{k-1}, \boldsymbol{a}_{k+1}, \ldots, \boldsymbol{a}_n\}.$$

We repeat the argument with $\boldsymbol{b}_3$ and so forth. If $m \leq n$ we end up with a set of the form

$$\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m, \boldsymbol{a}_{\ell_1}, \ldots, \boldsymbol{a}_{\ell_{n-m}}\}.$$

If $m > n$ we run into a contradiction. ∎

Theorem 4.4
Let $\mathcal{V}$ be a vector space such that one basis has m elements and another basis has n elements. Then $m = n$.

Proof.

Suppose $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$ and $\{\boldsymbol{b}_1, \ldots\}$ are two bases of $\mathcal{V}$. By lemma 4.1, because $\{\boldsymbol{a}_i\}$ spans $\mathcal{V}$, then $\{\boldsymbol{b}_i\}$ must contain n or less elements, otherwise the vectors would be ld. If $\{\boldsymbol{b}_i\}$ contains less than n elements, then $\{\boldsymbol{a}_i\}$ would be ld. Thus the cardinality of $\{\boldsymbol{b}_i\}$ is exactly n. ■

Definition 4.13 (Dimension of a vector space)
Let $\mathcal{S} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n\}$ *be a basis of* $\mathcal{V}$. *The dimension of* $\mathcal{V}$*, denoted* *dim*$(\mathcal{V})$*, is the number of elements in the basis, i.e.* $\dim(\mathcal{V}) = \#\mathcal{S} = n$.

- The vector space $\{\mathbf{0}\}$ is defined to have dimension 0.
- Suppose $\mathcal{V}$ does not have a finite basis, then $\mathcal{V}$ is said to be of infinite dimension.

Example 4.12
Consider the following vectors in $\mathbb{R}^n$

$$\boldsymbol{e}_1 = (1, 0, \dots, 0)^\top, \boldsymbol{e}_2 = (0, 1, \dots, 0)^\top, \dots, \boldsymbol{e}_n = (0, 0, \dots, 1)^\top.$$

These vectors are li and span $\mathbb{R}^n$*. Equivalently, any vector* $\boldsymbol{b} \in \mathbb{R}^n$ *is a lc of the above vectors:*

$$\boldsymbol{b} = a_1 \boldsymbol{e}_1 + a_2 \boldsymbol{e}_2 + \cdots + a_n \boldsymbol{e}_n.$$

Accordingly the vectors form a basis in $\mathbb{R}^n$ *called the* <u>*standard*</u> *basis of* $\mathbb{R}^n$*. Thus* $\mathbb{R}^n$ *has dimension* n*.*

Example 4.13

The following matrices form a basis for the vector space $\mathbb{R}^{2\times 3}$*:*

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

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

More generally, in the vector space $\mathbb{R}^{m\times n}$ *of all* $m \times n$ *matrices, let* $\boldsymbol{E}_{ij}$ *denote the matrix with the* (i, j)*-entry equal to* 1 *and* 0*'s elsewhere. All such matrices form a basis of* $\mathbb{R}^{m\times n}$ *called the* *standard* *basis of* $\mathbb{R}^{m\times n}$

Example 4.14

1. *Consider $\mathcal{P}_n(x)$ and the set*

$$\mathcal{S} = \{1, x, x^2, \ldots, x^n\}$$

of $n+1$ monomials is a basis of $\mathcal{P}_n(x)$. In particular, any polynomial of degree $\leq n$ can be expressed uniquely as a lc of the elements in the basis. Therefore $dim(\mathcal{P}_n(x)) = n+1$.

2. *Consider $\mathcal{P}(x)$, and the set*

$$\mathcal{S} = \{f_1(x), f_2(x), \ldots, f_m(x)\},$$

where f_ℓ denotes a polynomial of degree ℓ. Then any polynomial of degree $> m$ cannot be written as a lc of the elements in $\mathcal{S}$, thus $\mathcal{S}$ is not a basis of $\mathcal{P}(x)$. In fact $dim(\mathcal{P}(x)) = \infty$.

Theorem 4.5

Let $\mathcal{V}$ be a vector space of finite dimension n. Then:

1. *Any set of $n+1$ vectors (or more) in $\mathcal{V}$ are ld.*
2. *Any li set $\mathcal{A} = \{\boldsymbol{a}_i\}_{i=1}^n$ with n elements is a basis of $\mathcal{V}$.*
3. *Any spanning set $\mathcal{S} = \{\boldsymbol{a}_i\}_{i=1}^n$ of $\mathcal{V}$ with n elements is a basis of $\mathcal{V}$.*

Proof.

Suppose $\mathcal{B} = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ is a basis of $\mathcal{V}$.

1. Because $\mathcal{B}$ spans $\mathcal{V}$, any $n+1$ vectors or more are ld.
2. Since elements from $\mathcal{B}$ can be adjoined to $\mathcal{A}$ to form a spanning set of $\mathcal{V}$ with n elements. Because $\mathcal{A}$ has already n elements, $\mathcal{A}$ itself is a spanning set of $\mathcal{V}$. Thus $\mathcal{S}$ is a basis of $\mathcal{V}$.
3. Suppose $\mathcal{S}$ is ld. Then $\mathcal{S} \setminus \boldsymbol{a}_k$ is ss of $\mathcal{V}$ with $n-1$ elements. But $\mathcal{B}$ has n elements. Thus $\mathcal{S}$ must be li, and it is a basis of $\mathcal{V}$.

■

Theorem 4.6

Suppose $\mathcal{S}$ spans a vector space $\mathcal{V}$, then:

1. *Any maximum number of li vectors in $\mathcal{S}$ form a basis of $\mathcal{V}$*
2. *Suppose one deletes from $\mathcal{S}$ every vector that is a lc of preceding vectors in $\mathcal{S}$. Then the remaining vectors form a basis of $\mathcal{V}$*

Proof.

1. Suppose $\{\boldsymbol{a}_i\}_{i=1}^{n}$ is a maximum li subset of $\mathcal{S}$, and suppose $\boldsymbol{b} \in \mathcal{S}$. Accordingly $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n, \boldsymbol{b}\}$ is ld because $\boldsymbol{b} \in \text{span}(\boldsymbol{a}_i)$ and hence $\mathcal{S} \subseteq \text{span}(\boldsymbol{a}_i)$. This leads to $\mathcal{V} = \text{span}(\mathcal{S}) \subseteq \text{span}(\boldsymbol{a}_i)$. Thus li set $\{\boldsymbol{a}\}_{i=1}^{n}$ spans $\mathcal{V}$ and it is a basis of $\mathcal{V}$.
2. The remaining vectors form a maximum li subset of $\mathcal{S}$, hence by [1], it is a basis of $\mathcal{V}$.

∎

Theorem 4.7

Let $\mathcal{V}$ be a vector space of finite dimension and let $\mathcal{S} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r\}$ be a set of linearly independent vectors in $\mathcal{V}$, then $\mathcal{S}$ is part of a basis in $\mathcal{V}$; that is, $\mathcal{S}$ may be extended to a basis of $\mathcal{V}$.

Proof.

Suppose $\mathcal{B} = \{\boldsymbol{b}_i\}_{i=1}^n$ is a basis of $\mathcal{V}$. Note that $\mathcal{V}$ is also spanned by

$$\mathcal{S} \cup \mathcal{B} = \{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_r, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$$

By theorem 4.6, we can delete each vector that is a lc of the preceding vectors to obtain a basis $\mathcal{B}'$ of $\mathcal{V}$. Because $\mathcal{S}$ is li, no $\boldsymbol{a}_i$ is a lc of the preceding vectors. Thus $\mathcal{B}'$ contains every vector in $\mathcal{S}$, and $\mathcal{S}$ is part of the basis $\mathcal{B}'$ of $\mathcal{V}$. ∎

Dimensions and subspaces:

The following theorem gives the basic relation between the dimension of a vector space and the dimension of a subspace.

Theorem 4.8

Let $\mathcal{W}$ be a subspace of an n-dimensional vector space $\mathcal{V}$, then $dim(\mathcal{W}) \leq n$. In particular if $dim(\mathcal{W}) = n$, then $\mathcal{W} = \mathcal{V}$.

Example 4.15

Let $\mathcal{W}$ be a subspace of $\mathbb{R}^3$. Note that dim$(\mathbb{R}^3) = 3$. Theorem 4.8 tells us that dim$(\mathcal{W})$ can only be $0, 1, 2$ or 3. The following cases apply:

1. *if dim$(\mathcal{W}) = 0$, then $\mathcal{W} = \{0\}$, is a point at the origin $\mathcal{O}$.*
2. *if dim$(\mathcal{W}) = 1$, then $\mathcal{W}$ is a line that passes through the origin $\mathcal{O}$.*
3. *if dim$(\mathcal{W}) = 2$, then $\mathcal{W}$ is a plane that passes through the origin $\mathcal{O}$.*
4. *if dim$(\mathcal{W}) = 3$, $\mathcal{W}$ is the entire space $\mathbb{R}^3$.*

Example 4.16

1. *Let* $\mathcal{M} \subset \mathbb{R}^{2\times 2}$ *be a vector subspace*

$$\mathcal{M} = \{\boldsymbol{M} \in \mathbb{R}^{2\times 2} : \boldsymbol{M} = \boldsymbol{M}^\top\}.$$

   what is the dimension of $\mathcal{M}$*?*

2. *Let* $\mathcal{M} \subset \mathbb{R}^{3\times 3}$ *be a vector subspace*

$$\mathcal{M} = \{\boldsymbol{M} \in \mathbb{R}^{3\times 3} : \boldsymbol{M} = -\boldsymbol{M}^\top\}.$$

   what is the dimension of $\mathcal{M}$*?*

Let $\boldsymbol{A}$ be a $m \times n$ matrix, and $\boldsymbol{A}_\ell \in \mathbb{R}^n$ denote its columns. Denote $\text{colsp}(\boldsymbol{A})$ the subspace of $\mathbb{R}^m$ spanned by the columns of $\boldsymbol{A}$.

Definition 4.14 (Rank)

The rank of a matrix, written $\text{rank}(\boldsymbol{A})$ *is equal to the maximum number of li cols in* $\boldsymbol{A}$*, or equivalently, the dimension of the column space of* $\boldsymbol{A}$

$$\text{rank}(\boldsymbol{A}) = dim(colsp(\boldsymbol{A})).$$

Remark 4.10

Note that the rank is the number of elements in a basis of $colsp(\boldsymbol{A})$*, and that this number equals to the number of pivots in an echelon form of* $\boldsymbol{A}$*. See definition 3.3.*

Application to find a basis for $\mathcal{W} = \text{span}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r)$:

Frequently we are given a list of vectors $\mathcal{S} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r\}, \boldsymbol{u}_i \in \mathbb{R}^n$ and we want to find a basis for the subspace $\mathcal{W}$ of $\mathbb{R}^n$ spanned by the given vectors, that is a basis of

$$\mathcal{W} = \text{span}(\mathcal{S}) = \text{span}(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r).$$

One method to accomplish this goal is the casting-out algorithm

Algorithm 1 (Casting-out algorithm)

1. *Form matrix* $\boldsymbol{M}$*, whose columns are the given vectors*
2. *Row reduce* $\boldsymbol{M}$ *to echelon form* $\tilde{\boldsymbol{M}}$
3. *For each column* $\tilde{\boldsymbol{M}}_k$ *without a pivot, delete (cast-out) the vector* $\boldsymbol{u}_k$ *from the list* $\mathcal{S}$ *of given vectors.*
4. *Output the remaining vectors in* $\mathcal{S}$ *(which correspond to columns with pivots).*

Example 4.17

Let $\mathcal{W}$ be a subspace of $\mathbb{R}^5$ spanned by

$$\boldsymbol{u}_1 = (1, 2, 1, 3, 2)^\top, \boldsymbol{u}_2 = (1, 3, 3, 5, 3)^\top, \boldsymbol{u}_3 = (3, 8, 7, 13, 8)^\top$$
$$\boldsymbol{u}_4 = (1, 4, 6, 9, 7)^\top \text{ and } \boldsymbol{u}_5 = (5, 13, 13, 25, 19)^\top.$$

Find a basis for $\mathcal{W}$ consisting of the original given vectors and find dim($\mathcal{W}$).

Example 4.18
Let $\mathcal{W}$ be a subspace of $\mathbb{R}^4$ spanned by

$$\boldsymbol{u}_1 = (1, -2, 5, -3)^\top, \boldsymbol{u}_2 = (2, 3, 1, -4)^\top, \boldsymbol{u}_3 = (3, 8, -3, -5)^\top$$

1. *Find a basis and dimension of $\mathcal{W}$.*
2. *Extend the basis of $\mathcal{W}$ to a basis of $\mathbb{R}^4$.*

Example 4.19 (Assessment Question (2015-II))
Consider the following vectors in $\mathbb{R}^4$:

$$\boldsymbol{b}_1 = (1,0,1,0)^\top, \boldsymbol{b}_2 = (1,0,0,-1)^\top, \boldsymbol{b}_3 = (1,1,-1,-1)^\top, \boldsymbol{b}_4 = (1,1,1,1)^\top,$$

and let $\mathcal{A} = \{\boldsymbol{x} \in \mathbb{R}^4 \,:\, x_1 = x_2 + x_3 + \alpha x_4\}$ be a subspace of $\mathbb{R}^4$

- *Find $\bar{\alpha} = \alpha$ s.t. $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, \boldsymbol{b}_4\}$ spans $\mathcal{A}$*
- *Is $\{\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3, \boldsymbol{b}_4\}$ a basis of $\mathcal{A}$ given $\bar{\alpha}$?*
- *Find a basis of subspace $\mathcal{A}$ and compute dim$(\mathcal{A})$*

Linear Transformations

Special matrices:

Let $k \in \mathbb{N}$ and $\boldsymbol{A} \in \mathbb{R}^{n \times n}$

- diagonal: if $a_{i,j} = 0$ for $i \neq j$
- upper triangular: if $a_{ij} = 0$ for $i > j$
- lower triangular: if $a_{ij} = 0$ for $i < j$
- symmetric: if $\boldsymbol{A} = \boldsymbol{A}^\top$
- antisymmetric: if $\boldsymbol{A} = -\boldsymbol{A}^\top$
- nilpotent: if $\boldsymbol{A}^k = \mathbf{0}$,
- idempotent: if $\boldsymbol{A}^2 = \boldsymbol{A}$
- normal: if $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{A}^\top\boldsymbol{A}$
- orthogonal: if $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{I}_n$

Determinant

Definition 5.1 (Minor)
Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. The minor of $\boldsymbol{A}$ at (i, j) is the matrix $\boldsymbol{M}_{i,j} \in \mathbb{R}^{n-1 \times n-1}$ resulting after removing the i-th row and j-th column from $\boldsymbol{A}$.

Definition 5.2 (Determinant)
Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. The determinant of $\boldsymbol{A}$ is given by Laplace's formula

$$\det(\boldsymbol{A}) = \begin{cases} a & \text{if } n = 1, \\ \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \det(\boldsymbol{M}_{i,j}) & \text{if } n > 1 \end{cases}$$

for any fixed row $i = 1, \ldots, n$

Remark 5.1
The formula holds wrt any fixed column as well in the sense that:

$$\det(\boldsymbol{A}) = \sum_{i=1}^{n} (-1)^{i+j} a_{i,j} \det(\boldsymbol{M}_{i,j}), \quad n > 1,$$

Properties:

Let $\lambda \in \mathbb{R}$ and $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. It holds that

1. $\det(\lambda \boldsymbol{A}) = \lambda^n \det(\boldsymbol{A})$
2. $\det(\boldsymbol{A}^\top) = \det(\boldsymbol{A})$
3. $\det(\boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{A}) \det(\boldsymbol{B})$
4. If $\boldsymbol{A}$ is upper (or lower) triangular, then $\det(\boldsymbol{A}) = \prod_{i=1}^n a_{i,i}$

But we stress that:

$$\det(\boldsymbol{A} + \boldsymbol{B}) \neq \det(\boldsymbol{A}) + \det(\boldsymbol{B})$$

Example 5.1

Prove the 3rd Property.

Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$. First we recap some well known facts:

1. If $\boldsymbol{B}$ is obtained by interchanging two rows of $\boldsymbol{A}$, then

$$\det(\boldsymbol{B}) = -\det(\boldsymbol{A})$$

2. if $\boldsymbol{B}$ is obtained by adding a scalar multiple of one row to another row of $\boldsymbol{A}$, then

$$\det(\boldsymbol{B}) = \det(\boldsymbol{A})$$

3. if $\boldsymbol{B}$ is obtained by multiplying a row of $\boldsymbol{A}$ by a scalar $\lambda \in \mathbb{R}$, then

$$\det(\boldsymbol{B}) = \lambda \det(\boldsymbol{A})$$

Denote an elementary matrix $\boldsymbol{E} \in \mathbb{R}^{n \times n}$. We first show that $\det(\boldsymbol{E}\boldsymbol{B}) = \det(\boldsymbol{E})\det(\boldsymbol{B})$.

- If $\boldsymbol{E}$ is obtained by switching to rows of $\boldsymbol{I}$, then $\det(\boldsymbol{E}) = -1$ and

$$\det(\boldsymbol{E}\boldsymbol{B}) = -\det(\boldsymbol{B}) = \det(\boldsymbol{E})\det(\boldsymbol{B})$$

- If $\boldsymbol{E}$ is obtained by adding a scale multiple of a row to another, then $\det(\boldsymbol{E}) = 1$ and

$$\det(\boldsymbol{E}\boldsymbol{B}) = \det(\boldsymbol{B}) = \det(\boldsymbol{E})\det(\boldsymbol{B})$$

- If $\boldsymbol{E}$ is obtained by multiplying a row of $\boldsymbol{I}$ by λ, then $\det(\boldsymbol{E}) = \lambda$ and

$$\det(\boldsymbol{E}\boldsymbol{B}) = \lambda\det(\boldsymbol{B}) = \det(\boldsymbol{E})\det(\boldsymbol{B})$$

Hence, $\det(\boldsymbol{E}\boldsymbol{B}) = \det(\boldsymbol{E})\det(\boldsymbol{B})$.

Now note that

$$\det(\boldsymbol{E}_1\boldsymbol{E}_2\ldots\boldsymbol{E}_k\boldsymbol{B}) = \det(\boldsymbol{E}_1)\det(\boldsymbol{E}_2\ldots\boldsymbol{E}_k\boldsymbol{B}) = \det(\boldsymbol{E}_1)\ldots\det(\boldsymbol{E}_k)\det(\boldsymbol{B})$$

and consider $\det(\boldsymbol{AB})$ for:

1. $\boldsymbol{A}$ non-singular. In this case we can always write $\boldsymbol{A} = \boldsymbol{E}_1\ldots\boldsymbol{E}_k$, and it follows that $\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$.
2. $\boldsymbol{A}$ singular. In this case $\boldsymbol{A}$ would be row equivalent to a matrix that has all zeros in a row and $\det(\boldsymbol{A}) = 0$. Thus $\boldsymbol{AB}$ would be singular as well because if $\boldsymbol{AB}(\boldsymbol{AB})^{-1} = \boldsymbol{I}$, then $\boldsymbol{B}(\boldsymbol{AB})^{-1} = \boldsymbol{A}^{-1}$, which leads to a contradiction. Hence, $\det(\boldsymbol{AB}) = 0\det(\boldsymbol{B}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$.

Example 5.2
Let $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$. *Shown that* $\det(\boldsymbol{A})$ *is the area of the parallelepiped formed by the columns* $\boldsymbol{A}_j$. *What is the area formed by the rows of* $\boldsymbol{A}$*?*

Example 5.3
Show that the determinant of a matrix in $\mathbb{R}^{2\times 2}$ *with columns in the Euclidean unit circle is less or equal to one. Hint:*
$\sin(\alpha \pm \beta) = \sin(\alpha)\cos(\beta) \pm \cos(\alpha)\sin(\beta)$.

Consider the representation

$$\boldsymbol{A} = \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} \cos(\alpha) & \cos(\beta) \\ \sin(\alpha) & \sin(\beta) \end{bmatrix},$$

with unit norm columns. Then, the determinant

$$\det(\boldsymbol{A}) = \sin(\beta)\cos(\alpha) - \cos(\beta)\sin(\alpha) = -\sin(\alpha - \beta),$$

and since $-1 \leq \sin(\alpha - \beta) \leq 1$, it holds that $-1 \leq -\sin(\alpha - \beta) \leq 1$.

Trace

Definition 5.3 (Trace)
The trace of $\boldsymbol{A} = [a_{i,j}] \in \mathbb{R}^{n \times n}$*, written* $\operatorname{tr}(\boldsymbol{A})$ *is the sum of the diagonal elements of* $\boldsymbol{A}$*,*

$$\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{i,i}$$

One can also write $\operatorname{tr}(\boldsymbol{A}) = \langle \boldsymbol{s}, \operatorname{diag}(\boldsymbol{A}) \rangle$, where $\boldsymbol{s}$ is the sum vector and $\operatorname{diag}(\boldsymbol{A}) = (a, \ldots, a_{n,n})^\top$.

Properties:

Let $\lambda \in \mathbb{R}$ and $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{n \times n}$

1. $\operatorname{tr}(\boldsymbol{A} + \boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}) + \operatorname{tr}(\boldsymbol{B})$
2. $\operatorname{tr}(\lambda \boldsymbol{A}) = \lambda \operatorname{tr}(\boldsymbol{A})$
3. $\operatorname{tr}(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{A}^\top)$
4. $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{A})$

Definition 5.4 (Linear transformation)
A linear transformation is a linear function F that maps elements from $\mathbb{R}^n$ into itself, i.e.

$$F : \mathbb{R}^n \to \mathbb{R}^n.$$

Linear transformation F has a matrix representation wrt basis $\mathcal{B}$:

$$\boldsymbol{F}_{\mathcal{B}} = [(F \circ \boldsymbol{b}_1)_{\mathcal{B}}, \quad \dots \quad (F \circ \boldsymbol{b}_n)_{\mathcal{B}}],$$

where $(F \circ \boldsymbol{b}_i)_{\mathcal{B}}$ is the column vector that refers to the coordinates of $F \circ \boldsymbol{b}_i$ when expressed as an expansion of basis $\mathcal{B}$.

Example 5.4
Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by

$$F(x, y) = (2x + 3y, 4x - 5y)$$

.

1. *Find the matrix representation of F wrt a basis with elements: $\boldsymbol{u}_1 = (1, 2)^\top$ and $\boldsymbol{u}_2 = (2, 5)^\top$.*
2. *Find the matrix representation of F wrt the canonical basis of $\mathbb{R}^2$.*

Remark 5.2
In what follows we will assume the transformation F is written wrt the canonical basis.

Consider a parallelotope $\mathcal{P}$ described by vertices

$$\mathcal{U} = [\boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \dots \quad \boldsymbol{v}_n],$$

and parallelotope $\mathcal{P}'$ described by vertices

$$\mathcal{U}' = [\boldsymbol{A}(\boldsymbol{v}_1) \quad \boldsymbol{A}(\boldsymbol{v}_2) \quad \dots \quad \boldsymbol{A}(\boldsymbol{v}_n)].$$

A transformation $\boldsymbol{A}$ is linear if

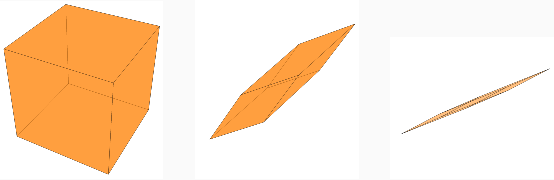
$$\boldsymbol{A}(\mathcal{U}) = \mathcal{U}',$$

i.e. if parallelotopes are mapped into parallelotopes.

Example 5.5

Consider $\mathcal{P}$ described by the canonical basis of $\mathbb{R}^3$; $\mathcal{P}'$ described by $\boldsymbol{A}_1 = (1,3,1)^\top$, $\boldsymbol{A}_2 = (2,1,1)^\top$ and $\boldsymbol{A}_3 = (1,1,2)^\top$; and $\mathcal{P}''$ described by $\boldsymbol{B}_1 = (1/2,1,1/2)^\top$, $\boldsymbol{B}_2 = (2,1,1)^\top$ and $\boldsymbol{B}_3 = (1,1/2,1/2)^\top$.

$\mathcal{P}$ $\mathcal{P}'$ $\mathcal{P}''$



$\boldsymbol{I}(\boldsymbol{v})$ $\boldsymbol{A}(\boldsymbol{v})$ $(\boldsymbol{B}\boldsymbol{A})(\boldsymbol{v})$

This interpretation makes immediate that $(\boldsymbol{B}\boldsymbol{A})(\boldsymbol{v}) = \boldsymbol{B}(\boldsymbol{A}(\boldsymbol{v}))$

Definition 5.5 (Inverse of a Transformation)
The inverse $\boldsymbol{A}^{-1}$ of a linear transformation $\boldsymbol{A}$ is the transformation that applied either before or after $\boldsymbol{A}$ cancels the effect of $\boldsymbol{A}$, i.e.

$$(\boldsymbol{A}^{-1}\boldsymbol{A})(\boldsymbol{v}) = \boldsymbol{v} = (\boldsymbol{A}\boldsymbol{A}^{-1})(\boldsymbol{v})$$

- If a transformation $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *squeezes* a parallelotope into a degenerate parallelotope, $\boldsymbol{A}^{-1}$ does not exist. In general:

$$\begin{aligned} \text{rank}(\boldsymbol{A}) &= \text{dim}(\text{range}(\boldsymbol{A})) \leq n \\ \text{rank}(\boldsymbol{B}\boldsymbol{A}) &\leq \min\{\text{rank}(\boldsymbol{A}), \text{rank}(\boldsymbol{B})\}. \end{aligned}$$

- A linear transformation $\boldsymbol{A}$ is invertible if it is full-rank. In this case it maps a basis into another basis.

Definition 5.6 (Rotation)
A linear transformation $\boldsymbol{A}$ *is a rotation in the Euclidean space* $\mathbb{R}^n$ *if it does not alter the norm of any vector* $\boldsymbol{v} \in \mathbb{R}^n$*.*

$$\|\boldsymbol{A}\boldsymbol{x}\| = \|\boldsymbol{x}\|$$

Example 5.6
Show all orthogonal matrices are rotations.

Since $\boldsymbol{A}^\top = \boldsymbol{A}^{-1}$, using the definition of the norm:

$$\|\boldsymbol{A}\boldsymbol{x}\| = (\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x})^{1/2} = (\boldsymbol{x}^\top \boldsymbol{A}^{-1} \boldsymbol{A}\boldsymbol{x})^{1/2} = \|\boldsymbol{x}\|.$$

Example 5.7
Show that $\det(\boldsymbol{A}) = \pm 1$, *where* $\boldsymbol{A}$ *is an orthogonal matrix.*

$$
\begin{aligned}
1 &= \det(\boldsymbol{I}) \\
&= \det(\boldsymbol{A}^{-1}\boldsymbol{A}) \\
&= \det(\boldsymbol{A}^{\top}\boldsymbol{A}) \\
&= \det(\boldsymbol{A}^{\top})\det(\boldsymbol{A}) \\
&= \det(\boldsymbol{A})^2
\end{aligned}
$$

Thus, $\det(\boldsymbol{A}) = \pm 1$.

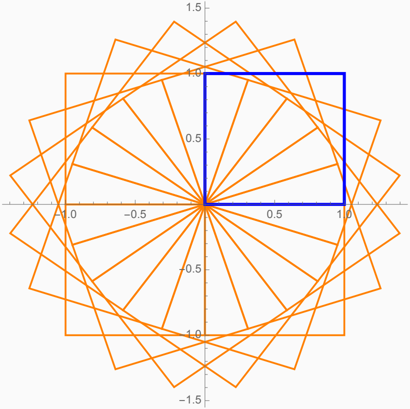
Example 5.8

Any rotation $\boldsymbol{A}_\theta \in \mathbb{R}^{2\times 2}$ can be written as:

$$\boldsymbol{A}_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix},$$

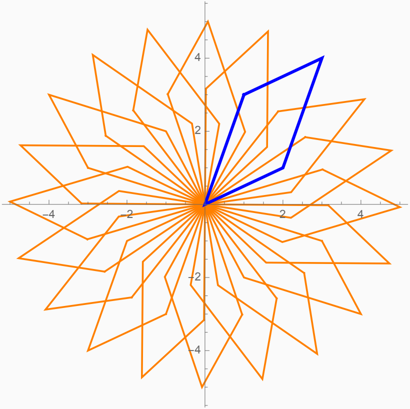
and represents a counterclockwise rotation of angle θ.

Identity matrix



$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Arbitrary matrix



$$\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$$

Definition 5.7 (Similarity transformation)
Two matrices $\boldsymbol{A}, \tilde{\boldsymbol{A}} \in \mathbb{R}^{n \times n}$ *are similar linear transformations if there exists some invertible matrix* $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ *such that:*

$$\tilde{\boldsymbol{A}} = \boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}.$$

Two similar linear transformations may have very different matrix representations $\boldsymbol{A}$ and $\tilde{\boldsymbol{A}}$. Nevertheless, they must share many properties.

Determinant Invariance

Consider parallelotope $\mathcal{P}$ described by vertices of a set of li vectors in $\mathbb{R}^n$. Matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ maps $\mathcal{P}$ into $\mathcal{P}'$. In doing so it stretches and turns $\mathcal{U}$ modifying its volume in some factor called the determinant.

Example 5.9
Let $\mathcal{U}$ be a parallelogram described by vectors $\boldsymbol{u}_1 = (1,0)^\top, \boldsymbol{u}_2 = (0,1)^\top$ and let

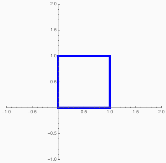
$$\boldsymbol{A} = \begin{bmatrix} 1/2 & 3/2 \\ 1 & 1 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} 1/2 & 1 \\ 1 & 2 \end{bmatrix}$$

It follows that

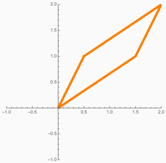
$$\begin{aligned} vol(\mathcal{P}') &= |\det(\boldsymbol{A})| vol(\mathcal{P}) = vol(\mathcal{P}) \\ vol(\mathcal{P}'') &= |\det(\boldsymbol{B})| vol(\mathcal{P}) = 0 \end{aligned}$$

$\boldsymbol{A}$ Mapping

$\mathcal{P}$

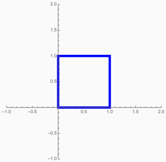


$\mathcal{P}'$

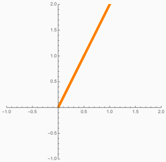


$\boldsymbol{B}$ Mapping

$\mathcal{P}$



$\mathcal{P}''$



- The transformation $\boldsymbol{A}$ is not invertible if and only of $\mathcal{P}'$ is degenerate, i.e. if

$$\text{vol}(\mathcal{P}') = \det(\boldsymbol{A})\text{vol}(\mathcal{P}) = 0 \quad \Leftrightarrow \quad \det(\boldsymbol{A}) = 0$$

- Moreover, for the composite transformation it holds that

$$\det(\boldsymbol{B}\boldsymbol{A}) = \det(\boldsymbol{B})\det(\boldsymbol{A}).$$

- The determinant is an invariant for similar linear transformations

$$\det(\tilde{\boldsymbol{A}}) = \det(\boldsymbol{B}^{-1})\det(\boldsymbol{A})\det(\boldsymbol{B}) = \det(\boldsymbol{A}).$$

Trace Invariance

Definition 5.8
The trace of $\boldsymbol{A} = [a_{i,j}] \in \mathbb{R}^{n \times n}$*, written* $\mathrm{tr}(\boldsymbol{A})$ *is the sum of the diagonal elements of* $\boldsymbol{A}$*,*

$$\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{i,i}$$

It is easy to show the circular property of the trace

$$\mathrm{tr}(\boldsymbol{ABC}) = \mathrm{tr}(\boldsymbol{BCA}) = \mathrm{tr}(\boldsymbol{CAB}),$$

which shows that the trace is an invariant for similarity transformations

$$\mathrm{tr}(\tilde{\boldsymbol{A}}) = \mathrm{tr}(\boldsymbol{B}^{-1}\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{ABB}^{-1}) = \mathrm{tr}(\boldsymbol{A}).$$

Diagonalization

Definition 6.1 (Diagonalizable transformation)
Matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ *is diagonalizable if it is similar to a diagonal matrix:*

$$\boldsymbol{D} = \boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B},$$

for some invertible matrix $\boldsymbol{B} \in \mathbb{R}^{n \times n}$*, and diagonal* $\boldsymbol{D} \in \mathbb{R}^{n \times n}$*.*

Definition 6.2 (Eigenvalues and Eigenvectors)
Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ *be diagonalizable. A scalar* λ *is called an eigenvalue of* $\boldsymbol{A}$ *if there exists vector a non-zero vector* $\boldsymbol{v}$ *such that*

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}.$$

Any vector $\boldsymbol{v}$ *satisfying this relation is called eigenvector of* $\boldsymbol{A}$ *belonging to the eigenvalue* λ*.*

Remark 6.1

- *If* $\boldsymbol{v}$ *is an eigenvector of* $\boldsymbol{A}$, *then so is any multiple of* $\boldsymbol{v}$.
- *Each eigenvalue* λ *has a set of vectors associated to it. The set* $\mathcal{E}_\lambda$ *of all such eigenvectors is a subspace of* $\mathbb{R}^n$.

Remark 6.2
An eigenvector is a vector that is not rotated by the transformation, but only scaled by it.

Theorem 6.1

Matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is similar to a diagonal matrix $\boldsymbol{\Lambda}$ (is diagonalizable) iff $\boldsymbol{A}$ has n li eigenvectors. In this case the diagonal elements of λ are the corresponding eigenvalues and $\boldsymbol{\Lambda} = \boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V}$.

Proof.
($\Rightarrow$) If $\boldsymbol{A}$ is diagonalizable, then it has a diagonal factorization

$$\boldsymbol{\Lambda} = \boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V},$$

where $\boldsymbol{V}$ is non-singular. Moreover,

$$\begin{aligned} \boldsymbol{A}\boldsymbol{V} &= \boldsymbol{V}\boldsymbol{\Lambda} \\ [\boldsymbol{A}\boldsymbol{v}_1, \dots, \boldsymbol{A}\boldsymbol{v}_n] &= [\lambda_1 \boldsymbol{v}_1, \dots, \lambda_n \boldsymbol{v}_n], \end{aligned}$$

that is, $\boldsymbol{v}_i$ are the eigenvectors corresponding to the eigenvalues λ_i of $\boldsymbol{A}$. In addition, this eigenvectors $\{\boldsymbol{v}_i\}$ are li, since they correspond to the columns of $\boldsymbol{V}$. ■

Remark 6.3

- *Easy to evaluate polynomials* $f(x) = \sum_{i=1}^{n} c_i x^i$, *of* $\boldsymbol{A}$*:*

$$
\begin{aligned}
f(\boldsymbol{A}) &= f(\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}) \\
&= \sum_{i=1}^{n} c_i \underbrace{\left(\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\right)\ldots\left(\boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1}\right)}_{i \text{ times}} \\
&= \sum_{i=1}^{n} c_i \left(\boldsymbol{V}\boldsymbol{\Lambda}^i\boldsymbol{V}^{-1}\right) \\
&= \boldsymbol{V}\left(\sum_{i=1}^{n} c_i \boldsymbol{\Lambda}^i\right)\boldsymbol{V}^{-1} = \boldsymbol{V} f(\boldsymbol{\Lambda})\boldsymbol{V}^{-1}.
\end{aligned}
$$

Example 6.1
Let

$$\boldsymbol{A} = \begin{bmatrix} 3 & 1 \\ 2 & 2 \end{bmatrix}, \quad \boldsymbol{v}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \quad \textit{and} \quad \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

Find the corresponding eigenvalues. Is $\boldsymbol{A}$ *diagonalizable? Find* $\boldsymbol{A}^4$*.*

Example 6.2

Show that if $\boldsymbol{A}$ has an eigenvalue λ, and corresponding eigenvector $\boldsymbol{v}$; then $\tilde{\boldsymbol{A}}$ has the same eigenvalue, with eigenvector $\boldsymbol{w} = \boldsymbol{B}^{-1}\boldsymbol{v}$.

Lösung.

$$\begin{aligned} \tilde{\boldsymbol{A}}\boldsymbol{w} &= \boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B}\boldsymbol{w} \\ &= \boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{v} \\ &= \lambda\boldsymbol{B}^{-1}\boldsymbol{v} \\ &= \lambda\boldsymbol{w} \end{aligned}$$

■

Computation of eigenvalues:

To compute an eigenvalue λ, one should solve for

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}.$$

It reduces to find the roots of the characteristic polynomial:

$$P(\lambda) = \det(\boldsymbol{A} - \lambda_i \boldsymbol{I}) = 0.$$

Each λ_i has an algebraic multiplicity m_i, which is the multiplicity of λ_i as a root of $P(\lambda)$. Moreover, since $P(\lambda)$ allows the factorization

$$P(\lambda) = (\lambda - \lambda_1)^{m_1} \dots (\lambda - \lambda_r)^{m_r}$$

it is clear that: $\sum_{i=1}^{r} m_i = n$.

Computation of eigenvectors:

- Each eigenvalue has a set of eigenvectors $\mathcal{E}_{\lambda_i}$ corresponding to it.
- $\operatorname{span}(\mathcal{E}_{\lambda_i})$ is a vector subspace and its dimension is the geometric multiplicity, i.e. $e_i = \dim(\mathcal{E}_i)$ of λ_i.
- In practical terms, $\mathcal{E}_{\lambda_i}$ can be obtained for each λ_i by finding the nullspace $\operatorname{null}(\boldsymbol{A} - \lambda_i \boldsymbol{I})$.
- For any eigenvalue λ_i it holds that $1 \leq e_i \leq m_i$.

Example 6.3
Show that eigenvalues of $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$ are not necessarily real.

One can show that

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = \lambda^2 - \lambda \operatorname{tr}(\boldsymbol{A}) + \det(\boldsymbol{A}),$$

with known solutions

$$\lambda = \frac{1}{2}\left(\operatorname{tr}(\boldsymbol{A}) \pm \sqrt{\operatorname{tr}(\boldsymbol{A})^2 - 4\det(\boldsymbol{A})}\right).$$

This shows that in general $\lambda \in \mathbb{C}$, and that $\lambda \in \mathbb{R}$ is a special case corresponding to $\operatorname{tr}(\boldsymbol{A})^2 \geq 4\det(\boldsymbol{A})$.

Remark 6.4
λ depends on invariants $\det(\cdot)$ and $\operatorname{tr}(\cdot)$, and it is an invariant itself. That is, the eigenvalues of matrix $\boldsymbol{A}$ and a similarity transformation of $\boldsymbol{A}$, say $\tilde{\boldsymbol{A}}$, are the same.

Algorithm 2 (Diagonalization)
Input: $\boldsymbol{A} \in \mathbb{R}^{n \times n}$

1. *Find the characteristic polynomial $P(\lambda)$ of $\boldsymbol{A}$.*
2. *Find the roots of $P(\lambda)$ to obtain eigenvalues of $\boldsymbol{A}$.*
3. *Repeat [1] and [2] for each eigenvalue λ of $\boldsymbol{A}$.*
   1. *Form matrix $\boldsymbol{M} = \boldsymbol{A} - \lambda \boldsymbol{I}$.*
   2. *Find a basis for the solution space of $\boldsymbol{M}\boldsymbol{v} = \boldsymbol{0}$.*
4. *Consider the collection $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$ of all the eigenvectors from [3].*
   1. *If $m \neq n$, then $\boldsymbol{A}$ is not diagonalizable.*
   2. *If $m = n$, then $\boldsymbol{A}$ is diagonalizable, and for $\boldsymbol{V} = [\boldsymbol{v}_1 \quad \ldots \quad \boldsymbol{v}_n]$*

$$\boldsymbol{\Lambda} = \boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V},$$

where λ_i is the eigenvalue corresponding to eigenvector $\boldsymbol{v}_i$.

Output: Either "Non-diagonalizable" or "Diagonalizable" as $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^{-1}$.

Example 6.4

Apply the diagonalization algorithm to

$$\boldsymbol{A} = \begin{bmatrix} 4 & 2 \\ 3 & -1 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 5 & -1 \\ 1 & 3 \end{bmatrix}, \quad \textit{and} \quad \boldsymbol{C} = \begin{bmatrix} 3 & -5 \\ 2 & -3 \end{bmatrix},$$

Theorem 6.2 (Diagonalization criteria)

If $\boldsymbol{A}$ is an $n \times n$ matrix and has n different eigenvalues, then $\boldsymbol{A}$ is diagonalizable.

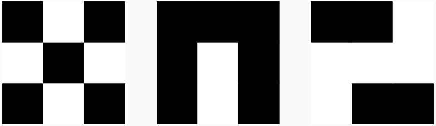
Proof.
Trivially, n different eigenvalues lead to n different eigenvectors, which are, by construction, linearly independent. Hence, there exists a similar matrix $\boldsymbol{\Lambda}$ such that $\boldsymbol{\Lambda} = \boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V}$. ■

Remark 6.5
The converse of Theorem 6.2 does <u>not</u> hold.

Example 6.5

Are the following matrices diagonalizable?

$\boldsymbol{A}$ $\boldsymbol{B}$ $\boldsymbol{C}$



.

Hint: Remember that for a 3×3 *matrix* $\boldsymbol{Q}$,

$$\det(\boldsymbol{Q} - \lambda \boldsymbol{I}) = \lambda^3 - \mathrm{tr}(\boldsymbol{Q})\lambda^2 + (\mathrm{cof}(\boldsymbol{Q},1,1) + \mathrm{cof}(\boldsymbol{Q},2,2) + \mathrm{cof}(\boldsymbol{Q},3,3))\lambda - \det(\boldsymbol{Q})$$

Theorem 6.3 (Spectral Theorem)

Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ be a symmetric diagonalizable matrix. Then it has factorization $\boldsymbol{A} = \boldsymbol{V\Lambda V}^\top$, $\lambda_i \in \mathbb{R}$.

Proof.
If $\boldsymbol{A}$ is symmetric and diagonalizable, then

$$\boldsymbol{A} = \boldsymbol{V\Lambda V}^{-1} \quad \text{and} \quad \boldsymbol{A}^\top = \boldsymbol{V\Lambda V}^{-1}.$$

Transposing the first equation, we obtain

$$\boldsymbol{V}^{-\top}\boldsymbol{\Lambda V}^\top = \boldsymbol{V\Lambda V}^{-1} \quad \to \quad \boldsymbol{\Lambda} = \underbrace{\boldsymbol{V}^\top\boldsymbol{V}}_{\tilde{\boldsymbol{V}}^{-1}}\boldsymbol{\Lambda}\underbrace{\boldsymbol{V}^{-1}\boldsymbol{V}^{-\top}}_{\tilde{\boldsymbol{V}}},$$

i.e., $\boldsymbol{\Lambda}$ is similar to itself for some non-singular $\tilde{\boldsymbol{V}}$, which must be the identity, hence $\boldsymbol{V}^\top = \boldsymbol{V}^{-1}$ and one can write $\boldsymbol{A} = \boldsymbol{V\Lambda V}^\top$. ∎

Example 6.6

Compute the eigenvalues and eigenvectors of

$$\boldsymbol{A} = \begin{bmatrix} \frac{9}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{11}{4} \end{bmatrix}$$

Solving for $\det(\boldsymbol{A} - \lambda \boldsymbol{I})$ we obtain

$$\lambda_1 = 3, \quad \lambda_2 = 2.$$

The corresponding eigenvectors solve:

$$\begin{bmatrix} \frac{9}{4} - 3 & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{11}{4} - 3 \end{bmatrix} \boldsymbol{v}_1 = \mathbf{0} \quad \text{and} \quad \begin{bmatrix} \frac{9}{4} - 2 & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{11}{4} - 2 \end{bmatrix} \boldsymbol{v}_2 = \mathbf{0},$$

leading to: $\boldsymbol{v}_1 = (1, \sqrt{3})^\top$ and $\boldsymbol{v}_2 = (-\sqrt{3}, 1)^\top$, or $\tilde{\boldsymbol{v}}_1 = (1/2, \sqrt{3}/2)^\top$ and $\tilde{\boldsymbol{v}}_2 = (-\sqrt{3}/2, 1/2)^\top$.

Geometric Interpretation:

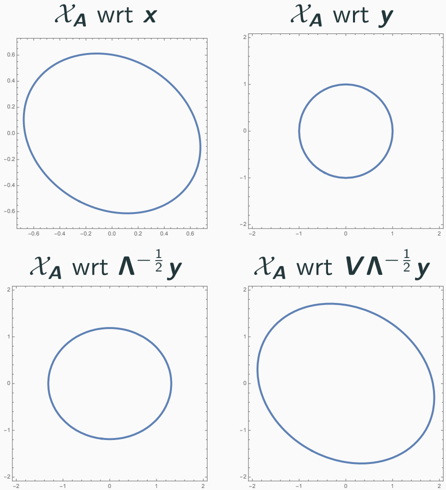
- Let $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top$. Consider the ellipsoid:

$$\begin{aligned}
\mathcal{X}_{\boldsymbol{A}} &= \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} = 1\} \\
&= \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{x}^\top (\boldsymbol{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}^\top)\boldsymbol{x} = 1\} \\
&= \{\boldsymbol{x} \in \mathbb{R}^n : \underbrace{\boldsymbol{x}^\top \boldsymbol{V}\boldsymbol{\Lambda}^{1/2}}_{\boldsymbol{y}^\top}\underbrace{\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}^\top \boldsymbol{x}}_{\boldsymbol{y}} = 1\} \\
&= \{\boldsymbol{y} \in \mathbb{R}^n : \boldsymbol{y}^\top \boldsymbol{y} = 1\}
\end{aligned}$$

- Hence $\mathcal{X}_{\boldsymbol{A}}$ turns into the equation of the unit sphere by the transformation:

$$\begin{aligned}
\boldsymbol{y} &= \boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{V}^\top \boldsymbol{x} \\
\boldsymbol{x} &= \boldsymbol{V}\boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{y}.
\end{aligned}$$

Consider the data from example 6.6:



Example 6.7 (The $25,000,000,000 eigenvector: the linear algebra behind Google)
Google's success derives in large part from its PageRank algorithm, which ranks the importance of webpages according to an eigenvector of a weighted link matrix.

Source:

`https://www.rose-hulman.edu/~bryan/googleFinalVersionFixed.pdf`

Properties and Special Cases

Special cases:

1. $\text{eig}(\boldsymbol{A}^\top) = \text{eig}(\boldsymbol{A})$.
2. $\text{eig}(\boldsymbol{A}^{-1}) = \frac{1}{\text{eig}(\boldsymbol{A})}$.
3. $\text{eig}(\boldsymbol{A}^k) = \text{eig}(\boldsymbol{A})^k$.
4. If $\boldsymbol{A}^2 = \boldsymbol{A}$, then $\text{eig}(\boldsymbol{A}) = 0$ or 1 and $\text{rank}(\boldsymbol{A}) = \text{tr}(\boldsymbol{A})$.
5. If $\boldsymbol{A}^2 = \boldsymbol{I}$, then $\text{eig}(\boldsymbol{A}) = \pm 1$.
6. If $\boldsymbol{A}^\top = \boldsymbol{A}^{-1}$, then $\text{eig}(\boldsymbol{A}) = \pm 1$.

Properties and Special Cases

Properties:

1. The rank of $\boldsymbol{A}$ is equal to the number of non-zero eigenvalues.
2. If $\boldsymbol{A}$ is singular, at least one of its eigenvalues is zero.
3. If $\boldsymbol{A}$ is non-singular, then all of its eigenvalues are different from zero.

Example 6.8

Show that the linear regression model $\boldsymbol{y} \approx \boldsymbol{X}\beta$, *has a linear estimator* $\hat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$, *where* $\boldsymbol{H}$ *is a projection matrix.*

Note that in the linear regression model $\boldsymbol{y} \in \mathbb{R}^n$ is approximated by n vecs $\boldsymbol{X}_\ell \in \mathbb{R}^n$ using $\boldsymbol{\beta} \in \mathbb{R}^m$ optimized at

$$\hat{\boldsymbol{\beta}} = \arg\min_{\beta} \|\boldsymbol{y} - \boldsymbol{X}\beta\| = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y},$$

with solution

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \underbrace{\boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top}_{\boldsymbol{H}} \boldsymbol{y},$$

where $\mathrm{eig}(\boldsymbol{H})$ are 1 or 0. Hence $\boldsymbol{H}$ is a projection matrix, i.e. $\boldsymbol{H}^2 = \boldsymbol{H}$.

Quadratic Forms

Definition 7.1 (Quadratic Form)
Any polynomial function $Q : \mathbb{R}^n \to \mathbb{R}$, of n variables which is homogeneous of degree 2. In general:

$$\begin{aligned} Q(x_1, \ldots, x_2) &= \sum_{i=1}^{n} \sum_{j=1}^{n} a_{i,j} x_i x_j \\ &= \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} \end{aligned}$$

for a rep. in sum or matrix form, where $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{x} \in \mathbb{R}^n$.

Remark 7.1

Without loss of generality, it can be assumed that $\boldsymbol{A}$ is symmetric, hence

$$Q = \sum_{i=1}^{n} \lambda_i z_i^2,$$

where $\boldsymbol{z} = \boldsymbol{V}^\top \boldsymbol{x}$ and $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top$.

Definite Matrices

Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{A} = \boldsymbol{A}^\top$ st $Q = \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}$.

Definition 7.2 (Semi-positive & positive definite matrices)
Matrix $\boldsymbol{A}$ is semi-positive definite ($\boldsymbol{A} \succeq 0$) if

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \geq 0 \quad \textit{for all} \quad \boldsymbol{x} \neq \boldsymbol{0}$$

It is positive definite if the inequality is strict ($\boldsymbol{A} \succ 0$).

Definition 7.3 (Semi-negative & negative definite matrices)
Matrix $\boldsymbol{A}$ is semi-negative definite ($\boldsymbol{A} \preceq 0$) if

$$\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \leq 0 \quad \textit{for all} \quad \boldsymbol{x} \neq \boldsymbol{0}$$

It is negative definite if the inequality is strict ($\boldsymbol{A} \prec 0$).

Remark 7.2
We say that $\boldsymbol{A}$ *is indefinite if it does not belong to any of the cases.*

From remark 7.1, it follows that:

1. if $\lambda_i \geq 0$ ($\lambda_i > 0$), then $\boldsymbol{A} \succeq 0$ ($\boldsymbol{A} \succ 0$)
2. if $\lambda_i \leq 0$ ($\lambda_i < 0$), then $\boldsymbol{A} \preceq 0$ ($\boldsymbol{A} \prec 0$)

Definition 7.4 (Principal Minor)
The principal minor of an $n \times n$ matrix $\boldsymbol{A}$ is defined as

$$\boldsymbol{D}_k = \det\left(\begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,k} \\ a_{2,1} & a_{2,2} & \dots & a_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k,1} & a_{k,2} & \dots & a_{k,k} \end{bmatrix}\right)$$

Theorem 7.1
Let $\boldsymbol{A}$ be a symmetric $n \times n$ matrix. Then for all $k = 1, \dots, n$:

1. $\boldsymbol{A} \succ 0$ ($\boldsymbol{A} \succeq 0$) *iff* $\boldsymbol{D}_k > 0$ ($\boldsymbol{D}_k \geq 0$)
2. $\boldsymbol{A} \prec 0$ ($\boldsymbol{A} \preceq 0$) *iff* $(-1)^k \boldsymbol{D}_k > 0$ ($(-1)^k \boldsymbol{D}_k \geq 0$)

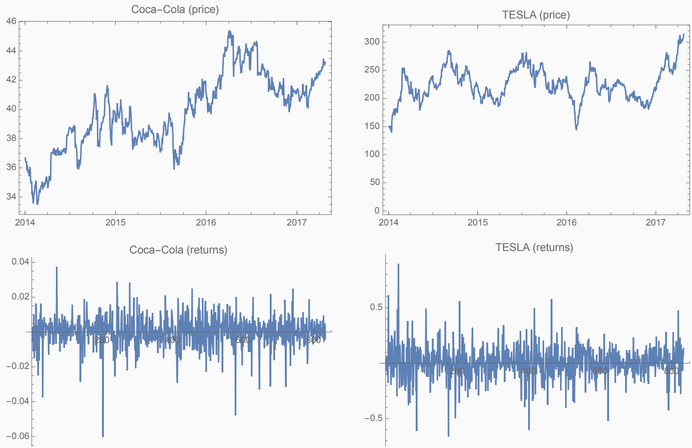
Proof.
Omitted. ∎

Example 7.1 (Midterm question)
True or False:

1. *If* $\boldsymbol{A} \prec 0$ *and* $\boldsymbol{B}$ *is singular, then* $\boldsymbol{B}^\top \boldsymbol{A}\boldsymbol{B} \prec 0$.
2. *If* $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ *and* $\boldsymbol{B} \succeq 0$, *then* $\det(\boldsymbol{I} + \boldsymbol{B}) \geq 1$.

Example 7.2

Consider an agent who invests his wealth in n risky assets.



BUYER'S GUIDE

2017 TESLA MODEL S P100D FIRST TEST: A NEW RECORD - 0-60 MPH IN 2.28 SECONDS!

The Model S P100D sets a new record (and accelerates like a real jerk)

We all understand acceleration. It's the rate of change of velocity. This 4,891-pound Tesla Model S P100D does it best, reaching 30, 40, 50, and 60 mph from a standstill more quickly than any other production vehicle we've ever tested, full stop. **In our testing, no production car has ever cracked 2.3 seconds from 0 to 60 mph. But Tesla has, in 2.275507139 seconds.**

The Tesla does not hold the advantage forever, though, because higher speeds give the advantage to horsepower over instant torque. The Ferrari LaFerrari hits 70 mph a tenth of a second quicker; the Porsche 918 and McLaren P1 pull ahead at 80 mph, and these hypercars all continue to pull away at higher speeds. But around town, everybody has long since lifted off the accelerator pedal.

Source:www.motortrend.com

- Denote $Z_{\text{P}} = \boldsymbol{w}_{\text{P}}^{\top}\boldsymbol{z}$ the return of portfolio P, where $\boldsymbol{w}$ represents the holdings of the N-dimensional vector of risky assets returns.
- Statistics of Z_{P}:

$$
\begin{aligned}
\mathbb{E}\left[Z_{\text{P}}\right] &= \mathbb{E}\left[\boldsymbol{w}_{\text{P}}^{\top}\boldsymbol{z}\right] = \boldsymbol{w}_{\text{P}}^{\top}\mathbb{E}\left[\boldsymbol{z}\right] = \boldsymbol{w}_{\text{P}}^{\top}\boldsymbol{\mu} =: \mu_{\text{P}} \\
\operatorname{var}\left[Z_{\text{P}}\right] &= \mathbb{E}\left[\left(Z_{\text{P}} - \mathbb{E}[Z_{\text{P}}]\right)\left(Z_{\text{P}} - \mathbb{E}[Z_{\text{P}}]\right)^{\top}\right] \\
&= \mathbb{E}\left[\boldsymbol{w}_{\text{P}}^{\top}\left(\boldsymbol{z} - \boldsymbol{\mu}\right)\left(\boldsymbol{z} - \boldsymbol{\mu}\right)^{\top}\boldsymbol{w}_{\text{P}}\right] = \boldsymbol{w}_{\text{P}}^{\top}\boldsymbol{\Sigma}\boldsymbol{w}_{\text{P}} =: \sigma_{\text{P}}^{2},
\end{aligned}
$$

  with dist. $Z_{\text{P}} \sim \mathcal{N}(\mu_{\text{P}}, \sigma_{\text{P}}^{2})$, and Sharpe ratio of portfolio: $\text{SR}_{\text{P}} = \frac{\mu_{\text{P}}}{\sigma_{\text{P}}}$.
- The investor's problem is to allocate

$$
\hat{\boldsymbol{w}}_{\text{P}} = \underset{\boldsymbol{w}_{\text{P}}}{\arg\max}\left\{\mathcal{U}(\boldsymbol{w}_{\text{P}}; \hat{\mu}, \hat{\boldsymbol{\Sigma}}) : \boldsymbol{w}_{\text{P}} \in \mathcal{C}\right\},
$$

  where $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ are sample estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ resp; $\mathcal{U}(\cdot)$ is a convex function and $\mathcal{C}$ is a convex.

Properties:

Let $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{A} = \boldsymbol{A}^\top$ st $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top$.

1. If $\boldsymbol{A}$ is semidefinite (positive or negative), then it is singular.
2. If $\boldsymbol{A} \succ 0$ ($\boldsymbol{A} \succeq 0$) then $-A \prec 0$ ($-A \preceq 0$).
3.
   - if $\boldsymbol{A} \succ 0$ ($\boldsymbol{A} \prec 0$), then $a_{i,i} > 0$ ($a_{i,i} < 0$).
   - if $\boldsymbol{A} \succeq 0$ ($\boldsymbol{A} \preceq 0$), then $a_{i,i} \geq 0$ ($a_{i,i} \leq 0$).
4.
   - if $\boldsymbol{A} \succ 0$ ($\boldsymbol{A} \prec 0$), then $\lambda_i > 0$ ($\lambda_i < 0$).
   - if $\boldsymbol{A} \succeq 0$ ($\boldsymbol{A} \preceq 0$), then $\lambda_i \geq 0$ ($\lambda_i \leq 0$).
5.
   - if $\lambda_i > 0$ ($\lambda_i < 0$), then $\boldsymbol{A} \succ 0$ ($\boldsymbol{A} \prec 0$).
   - if $\lambda_i \geq 0$ ($\lambda_i \leq 0$), then $\boldsymbol{A} \succeq 0$ ($\boldsymbol{A} \preceq 0$).
6.
   - if $\boldsymbol{A} \succ 0$, then $\det(\boldsymbol{A}) > 0$ and $\mathrm{tr}(\boldsymbol{A}) > 0$.
   - if $\boldsymbol{A} \succeq 0$, then $\det(\boldsymbol{A}) = 0$ and $\mathrm{tr}(\boldsymbol{A}) \geq 0$.
   - if $\boldsymbol{A} \prec 0$, then $\det(\boldsymbol{A}) < 0$ for n odd, $\det(\boldsymbol{A}) > 0$ ow, and $\mathrm{tr}(\boldsymbol{A}) < 0$.
   - if $\boldsymbol{A} \preceq 0$, then $\det(\boldsymbol{A}) = 0$ and $\mathrm{tr}(\boldsymbol{A}) \leq 0$.

Properties: (Cont.)

Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{A} = \boldsymbol{A}^\top$ st $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top$.

7. If $\boldsymbol{A} \succ 0$ ($\boldsymbol{A} \prec 0$), then $\boldsymbol{A}^{-1} \succ 0$ ($\boldsymbol{A}^{-1} \prec 0$).
8. If $\boldsymbol{A} = \boldsymbol{B}^\top \boldsymbol{B}$. If $\boldsymbol{B}$ is full-rank (is not full-rank), then $\boldsymbol{A} \succ 0$ ($\boldsymbol{A} \succeq 0$).
9. If $\boldsymbol{A} \succ 0$, then $\exists \boldsymbol{B} \succ 0$ st $\boldsymbol{B}^2 = \boldsymbol{A}$ is called the square root of $\boldsymbol{A}$. If $\boldsymbol{A} \succeq 0$, then $\exists \boldsymbol{B} \succeq 0$.
10. $\boldsymbol{B}^\top \boldsymbol{A} \boldsymbol{B}$, where $\boldsymbol{B}$ is a full-rank square matrix, is defined as $\boldsymbol{A}$.

Example 7.3 (Assessment question)
Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{A} \neq \mathbf{0}$, $\boldsymbol{A} \succeq 0$ *and* $\theta > 0$

1. *Show that* $\operatorname{eig}(\boldsymbol{A} + \theta \boldsymbol{I}_n)$ *correspond to* $\operatorname{eig}(\boldsymbol{A})$ *shifted by* θ.
2. *Show that* $\operatorname{eig}(\boldsymbol{A}(\boldsymbol{A} + \theta \boldsymbol{I}_n)^{-1})$ *are real numbers between* 0 *and* 1.

Block Matrices

Block Matrices

A block matrix is a matrix whose entries are matrices themselves, denoted $\boldsymbol{A} = [\boldsymbol{A}_{i,j}]$. Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, then one can write it in block form as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{1,1} & \boldsymbol{A}_{1,2} & \cdots & \boldsymbol{A}_{1,s} \\ \boldsymbol{A}_{2,1} & \boldsymbol{A}_{2,2} & \cdots & \boldsymbol{A}_{2,s} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}_{r,1} & \boldsymbol{A}_{r,2} & \cdots & \boldsymbol{A}_{r,s} \end{bmatrix},$$

where $\boldsymbol{A}_{i,j} \in \mathbb{R}^{m_i \times n_j}$ and $m = \sum_{i=1}^{r} m_i$ and $n = \sum_{j=1}^{s} n_j$.

Operations:

Let $\lambda \in \mathbb{R}$ and suppose $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{C}$ are block matrices with the same number of row and column blocks and suppose that corresponding blocks have the same size.

- **Addition:** $\boldsymbol{A} + \boldsymbol{B} = [\boldsymbol{A}_{i,j} + \boldsymbol{B}_{i,j}]$
- **Scalar multiplication:** $\lambda \boldsymbol{A} = [\lambda \boldsymbol{A}_{i,j}]$
- **Product:** $\boldsymbol{AB} = \boldsymbol{C}$ for $\boldsymbol{C}_{i,j} = \sum_{k=1}^{s} \boldsymbol{A}_{i,k} \boldsymbol{B}_{k,j}$, where s is the number of block columns in $\boldsymbol{A}$ (and the number of block rows in $\boldsymbol{B}$).
- **Transpose:** $\boldsymbol{A}^\top = [\boldsymbol{A}_{j,i}^\top]$

(more) Operations:

For the following two operations, let:

$$\begin{aligned} \boldsymbol{C}_1 &= \boldsymbol{A}_{1,1} - \boldsymbol{A}_{1,2}\boldsymbol{A}_{2,2}^{-1}\boldsymbol{A}_{2,1}, \quad \det(\boldsymbol{A}_{2,2}) \neq 0 \\ \boldsymbol{C}_2 &= \boldsymbol{A}_{2,2} - \boldsymbol{A}_{2,1}\boldsymbol{A}_{1,1}^{-1}\boldsymbol{A}_{1,2}, \quad \det(\boldsymbol{A}_{1,1}) \neq 0 \end{aligned}$$

- **Determinant:**

$$\det(\boldsymbol{A}) = \det(\boldsymbol{A}_{2,2})\det(\boldsymbol{C}_1) = \det(\boldsymbol{A}_{1,1})\det(\boldsymbol{C}_2)$$

- **Inverse:**

$$\boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{A}_{1,1}^{-1} + \boldsymbol{A}_{1,1}^{-1}\boldsymbol{A}_{1,2}\boldsymbol{C}_2^{-1}\boldsymbol{A}_{2,1}\boldsymbol{A}_{1,1}^{-1} & -\boldsymbol{C}_1^{-1}\boldsymbol{A}_{1,2}\boldsymbol{A}_{2,2}^{-1} \\ -\boldsymbol{A}_{2,2}^{-1}\boldsymbol{A}_{2,1}\boldsymbol{C}_1^{-1} & \boldsymbol{A}_{2,2}^{-1} + \boldsymbol{A}_{2,2}^{-1}\boldsymbol{A}_{2,1}\boldsymbol{C}_1^{-1}\boldsymbol{A}_{1,2}\boldsymbol{A}_{2,2}^{-1} \end{bmatrix},$$

where $\det(\boldsymbol{C}_1) \neq 0$ and $\det(\boldsymbol{C}_2) \neq 0$.

Definition 8.1
Let $\boldsymbol{A} = [a_{ij}] \in \mathbb{R}^{m \times n}$ *y* $\boldsymbol{B} \in \mathbb{R}^{r \times s}$*. The Kronecker product of* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *is the partitioned matrix*

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{1,1}\boldsymbol{B} & a_{1,2}\boldsymbol{B} & \cdots & a_{1,n}\boldsymbol{B} \\ a_{2,1}\boldsymbol{B} & a_{2,2}\boldsymbol{B} & \cdots & a_{2,n}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}\boldsymbol{B} & a_{m,2}\boldsymbol{B} & \cdots & a_{m,n}\boldsymbol{B}, \end{bmatrix},$$

where $\boldsymbol{A} \otimes \boldsymbol{B} \in \mathbb{R}^{mr \times ns}$*.*

Properties:

Let $\lambda \in \mathbb{R}$, $\boldsymbol{A} \in \mathbb{R}^{m \times m}$, $\boldsymbol{B} \in \mathbb{R}^{n \times n}$, then:

1. $\lambda \otimes \boldsymbol{A} = \lambda \boldsymbol{A}$
2. $\boldsymbol{A} \otimes (\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{A} \otimes \boldsymbol{C}$ and $(\boldsymbol{A} + \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{C}$
3. $\boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C}) = (\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C}$
4. $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{AC} \otimes \boldsymbol{BD}$
5. $(\boldsymbol{A} \otimes \boldsymbol{B})^\top = \boldsymbol{A}^\top \otimes \boldsymbol{B}^\top$

We stress that:

$$\boldsymbol{A} \otimes \boldsymbol{B} \neq \boldsymbol{B} \otimes \boldsymbol{A}$$

From the previous, it follows that:

6. $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$
7. $\operatorname{tr}(\boldsymbol{A} \otimes \boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}) \operatorname{tr}(\boldsymbol{B})$
8. $\det(\boldsymbol{A} \otimes \boldsymbol{B}) = \det(\boldsymbol{A})^{\operatorname{rank}(\boldsymbol{B})} \det(\boldsymbol{B})^{\operatorname{rank}(\boldsymbol{A})}$
9. $\{\operatorname{eig}(\boldsymbol{A} \otimes \boldsymbol{B})\} = \{\operatorname{eig}(\boldsymbol{A})\operatorname{eig}(\boldsymbol{B})^\top\}$

where $\{\kappa_i\}$ denotes the set of values κ_i, that is, the values in no particular order or structure,

Example 8.1
Let,

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \boldsymbol{B} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \quad \text{and} \quad \boldsymbol{C} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

and compute the following

$$\boldsymbol{A} \otimes \boldsymbol{C} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{B} \otimes \boldsymbol{C} = \begin{bmatrix} a & a & b & b \\ 0 & a & 0 & b \\ c & c & d & d \\ 0 & c & 0 & d \end{bmatrix}$$

$$\boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{C} = \begin{bmatrix} a+1 & a+1 & b & b \\ 0 & a+1 & 0 & b \\ c & c & d+1 & d+1 \\ 0 & c & 0 & d+1 \end{bmatrix}$$

Definition 8.2 (vec operator)

Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, the vec *operator stacks the columns of $\boldsymbol{A}$ in one vector as:*

$$\operatorname{vec}\left(\begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix}\right) = \begin{bmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \\ \vdots \\ \boldsymbol{A}_n \end{bmatrix},$$

with dimensionality $m \times n$ and where $\boldsymbol{A}_\ell$ denotes the ℓ-th column of $\boldsymbol{A}$.

Properties:

Let $\lambda \in \mathbb{R}$, $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times n}$, then:

1. $\operatorname{vec}(\lambda \boldsymbol{A}) = \lambda \operatorname{vec}(\boldsymbol{A})$
2. $\operatorname{vec}(\boldsymbol{A} + \boldsymbol{B}) = \operatorname{vec}(\boldsymbol{A}) + \operatorname{vec}(\boldsymbol{B})$

From the previous, it can be shown that:

3. $\operatorname{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}^\top \otimes \boldsymbol{A}) \operatorname{vec}(\boldsymbol{B})$
4. $\operatorname{tr}(\boldsymbol{AB}) = \operatorname{vec}(\boldsymbol{A}^\top)^\top \operatorname{vec}(\boldsymbol{B}) = \operatorname{vec}(\boldsymbol{B}^\top)^\top \operatorname{vec}(\boldsymbol{A})$

Example 8.2

Show that given certain $\boldsymbol{A}, \boldsymbol{B}$ *and* $\boldsymbol{C}$*, the Lyapunov equation*

$$\boldsymbol{AX} + \boldsymbol{XB} = \boldsymbol{C},$$

can be solved for $\text{vec}(\boldsymbol{X})$*.*

Hint: Remember that:

- $\text{vec}(\boldsymbol{A} + \boldsymbol{B}) = \text{vec}(\boldsymbol{A}) + \text{vec}(\boldsymbol{B})$
- $\text{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}^\top \otimes \boldsymbol{A})\,\text{vec}(\boldsymbol{B})$

Example 8.3

If $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *is symmetric, the redundant entries of* $\boldsymbol{A}$ *can be omitted by stacking its columns with the* vech *operator. For the* $\boldsymbol{A} \in \mathbb{R}^{2\times 2}$ *case,*

$$\mathrm{vech}\left(\begin{bmatrix} a_{1,1} & a \\ a & a_{2,2} \end{bmatrix}\right) = \begin{bmatrix} a_{1,1} \\ a \\ a_{2,2} \end{bmatrix},$$

with dimensionality $n(n-1)/2$*.*

Complex Numbers

Motivation

Example 9.1
Solve for x in

$$\begin{aligned} x^2 - 4 &= 0 &(5)\\ x^2 - 4/9 &= 0 &(6)\\ x^2 - 2 &= 0 &(7)\\ x^2 + 1 &= 0 &(8) \end{aligned}$$

Remark 9.1
The number sets have the relation: $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$

Definition 9.1 (Complex number)
Any number that can be written as

$$z = a + bi, \quad a, b \in \mathbb{R}, \quad i = \sqrt{-1},$$

is a complex number. We will also use

$$\begin{aligned} Re(z) &= a \\ Im(z) &= b, \end{aligned}$$

to refer to the real and imaginary parts of z respectively. The set of all complex numbers is denoted

$$\mathbb{C} = \{z = a + bi \ : \ a, b \in \mathbb{R}\}.$$

Definition 9.2 (Conjugate)
Let $z = a + bi \in \mathbb{C}$, the conjugate of z is denoted $\bar{z}$ and defined as

$$\bar{z} = a - bi.$$

Definition 9.3 (Module)
Let $z = a + bi \in \mathbb{C}$, the module of z is denoted $|z|$ and defined as

$$|z| = \sqrt{a^2 + b^2}.$$

Operations:

Let $z = a + bi$ and $w = c + di$,

- $z + w = (a + c) + (b + d)i$
- $z \cdot w = (ac - bd) + (ad + bc)i$
- $z^{-1} = (a - bi)/(a^2 + b^2) = \bar{z}/|z|^2$
- Let $w \neq 0$, $z \cdot w^{-1} = z \cdot \bar{w}/|w|^2$

Remark 9.2
Note that for $z = a + bi$ and $w = c + di$, $w \neq 0$ it holds that

$$\frac{z}{w} = \left(\frac{ac + bd}{c^2 + d^2}\right) + \left(\frac{bc - ad}{c^2 + d^2}\right) i.$$

Module properties:

Let $z, w \in \mathbb{C}$, $\lambda \in \mathbb{R}$

- $|\lambda z| = \lambda |z|$
- $|z| = |\bar{z}|$
- $|z|^2 = z \cdot \bar{z}$
- $|z \cdot w| = |z||w|$
- $|z/w| = |z|/|w|$

Remark 9.3
Note that if $z = a + bi$ and $b = 0$, i.e. $z \in \mathbb{R}$, then $|z| = abs(z)$.

Example 9.2
Let $z = a + bi$, $a, b \in \mathbb{R}$. Find the region given by $|\bar{z} + i| < 3$.

Definition 9.4 (Complex polynomial)
The n-th degree polynomial

$$P_n(z) = a_n z^n + \cdots + a_1 z + a_0, \quad a_n \neq 0$$

is called a complex polynomial if $a_0, a_1, \ldots, a_n \in \mathbb{C}$. *It is called a real polynomial if* $a_0, a_1, \ldots, a_n \in \mathbb{R}$.

Theorem 9.1

$P_n(z) = 0, n \geq 1$ has at least one solution.

Proof.
Omitted ∎

Corolary 9.1
$P_n(z) = 0, n \geq 1$ has n solutions (including algebraic multiplicities).

Example 9.3

Let $P_n(x)$ denote a real polynomial. Show that if z is a root of $P_n(x)$ then $\bar{z}$ is also a root of $P_n(x)$. Hint: note that $\bar{z}^k = \overline{z^k}$

Polar Representation

Definition 9.5 (Polar representation)
The polar representation of $z = a + bi \in \mathbb{C}$, is given by

$$z = a + bi = r\cos(\theta) + r\sin(\theta)i,$$

where r is the modulus of z, and θ is called argument of z.

Remark 9.4
Since sine and cosine are periodic functions, if θ is an argument of z then $\theta + 2k\pi, k \in \mathbb{Z}$ is also an argument of z. If $\theta \in]-\pi, \pi]$, then it is called principal value of the argument.

Example 9.4
Consider the complex numbers

$$z = \cos(\alpha) + \sin(\alpha)i \quad \textit{and} \quad w = \cos(\beta) + \sin(\beta)i, \quad |w| \neq 0, \alpha, \beta \in \mathbb{R}$$

Compute $z \cdot w$ and z/w. Hint: remember that:

$$\begin{aligned} \sin(\alpha \pm \beta) &= \sin(\alpha)\cos(\beta) \pm \cos(\alpha)\sin(\beta) \\ \cos(\alpha \pm \beta) &= \cos(\alpha)\cos(\beta) \mp \sin(\alpha)\sin(\beta) \end{aligned}$$

Euler's Formula

Theorem 9.2 (Euler's formula)

$$e^{i\theta} = \cos(\theta) + i\sin(\theta), \quad \theta \in \mathbb{R}.$$

Proof.
By power series expansion. Will see it in the next two classes. ∎

Remark 9.5
Note that If $z = r\cos(\theta) + r\sin(\theta)i$, then $z = re^{i\theta}$. Thus:

- $e^{i\theta} = e^{i(\theta+2k\pi)}$
- $e^{i\pi/2} = i$
- $e^{0i} = e^{2\pi i} = 1$
- $e^{i\pi} = -1$

Operations (using Euler's formula):

Let $z = re^{i\theta}$, $w = se^{i\varphi}$,

- $\bar{z} = re^{-i\theta}$
- $z \cdot w = rse^{i(\theta+\varphi)} = rs(\cos(\theta + \varphi) + i \sin(\theta + \varphi))$.
- $\frac{z}{w} = \frac{r}{s} e^{i(\theta-\varphi)} = \frac{r}{s}(\cos(\theta - \varphi) + i \sin(\theta - \varphi))$.

Example 9.5

Show that the hint in example 9.3 holds true.

Example 9.6 (Assessment question)

1. *Using Euler's formula, show that* $\cos(\theta) = \dfrac{e^{i\theta} + e^{-i\theta}}{2}$.
2. *Using the result in [1], find* $y = \arccos(a), a > 1$. *Moreover, show that if* $a = 1$ *then* $y = 0$. *Hint: if* $y = \arccos(a)$, *then* $\cos(y) = a$.

Theorem 9.3 (De Moivre)

$$(\cos(\theta) + i\sin(\theta))^n = \cos(n\theta) + i\sin(n\theta), \quad \theta \in \mathbb{R}, n \in \mathbb{Z}.$$

Proof.
Trivial from Euler's formula on the LHS. ∎

By the fundamental theorem of algebra, the equation $x^n - z = 0$ has n complex solutions. Then $z^{1/n}$ (the n-th root of z) take the values:

$$z = r^{1/n} e^{i(\theta+2k\pi)/n}, \quad k = 0, \ldots, n-1$$

Vectors in $\mathbb{C}^n$

Let $\boldsymbol{z}, \boldsymbol{w}$ denote vectors $\boldsymbol{z} = (z_1, \ldots, z_n)^\top$, $\boldsymbol{w} = (w_1, \ldots, w_n)^\top$ with entries $z_i, w_i \in \mathbb{C}$. These vectors belong to the vector space $\mathbb{C}^n$.

Operations:

Let $\boldsymbol{z}, \boldsymbol{w} \in \mathbb{C}^n$, and $\lambda \in \mathbb{C}$

- $\boldsymbol{z} + \boldsymbol{w} = (z_1 + w_1, \ldots, z_n + w_n)^\top$
- $\lambda \boldsymbol{z} = (\lambda z_1, \ldots, \lambda z_n)^\top$
- $\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \boldsymbol{z}^\top \bar{\boldsymbol{w}}$

Remark 9.6
Note that for $\boldsymbol{z}, \boldsymbol{w} \in \mathbb{C}^n$, $\langle \boldsymbol{z}, \boldsymbol{w} \rangle \neq \langle \boldsymbol{w}, \boldsymbol{z} \rangle$. *However:* $\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \overline{\langle \boldsymbol{w}, \boldsymbol{z} \rangle}$

Example 9.7

Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{C}^3$*,*

$$\boldsymbol{u} = (2 + 3i, 4 - i, 3 + 5i)^\top, \quad \boldsymbol{v} = (3 - 4i, 5i, 4 - 2i)^\top.$$

Compute $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ *and* $\|\boldsymbol{u}\|$*.*

Example 9.8

Find the eigenvalues and eigenvectors of

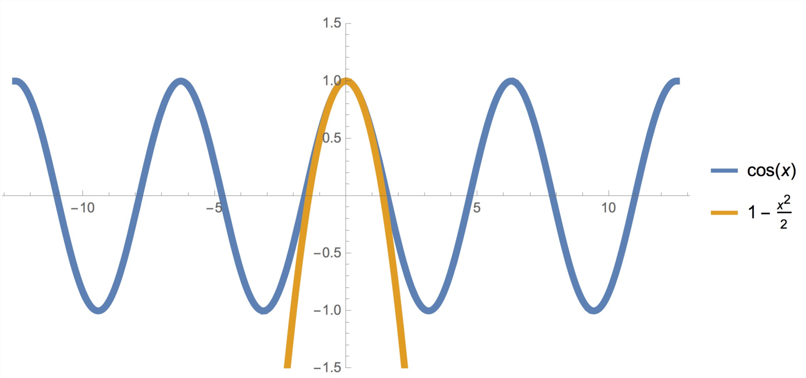
$$\boldsymbol{A} = \begin{bmatrix} 3 & -5 \\ 2 & -3 \end{bmatrix}.$$

Eigen decomposition leads to:

$$\boldsymbol{A} = \begin{bmatrix} 3+i & 3-i \\ 2 & 2 \end{bmatrix} \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \begin{bmatrix} -\frac{i}{2} & \frac{1}{4}+\frac{3i}{4} \\ \frac{i}{2} & \frac{1}{4}-\frac{3i}{4} \end{bmatrix}$$

Taylor Polys & Taylor Series

Motivation



Conjecture: $1 - \dfrac{x^2}{2} \approx \cos(x)$ at $x = 0$.

Example 10.1

Find the c_i in $p_2(x) = c_0 + c_1 x + c_2 x^2$ st $p(x) \approx \cos(x)$ at $x = 0$.

$$\begin{aligned}
\cos(x) \quad &\rightarrow \quad p(x) = c_0 + c_1 x + c_2 x^2 \\
\cos'(x) = -\sin(x) \quad &\rightarrow \quad p'(x) = c1 + 2c_2 x \\
\cos''(x) = -\cos(x) \quad &\rightarrow \quad p''(x) = 2c_2
\end{aligned}$$

$$\begin{aligned}
\cos(0) = 1 \quad &\rightarrow \quad p(0) = c_0 \\
\cos'(0) = 0 \quad &\rightarrow \quad p'(0) = c_1 \\
\cos''(0) = -1 \quad &\rightarrow \quad p''(0) = 2c_2
\end{aligned}$$

Matching derivatives is a good idea.

We are interested in the system:

$$\begin{bmatrix} 1 & x & x^2 \\ 0 & 1 & 2x \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} \cos(x) \\ \cos'(x) \\ \cos''(x) \end{bmatrix},$$

at $x = 0$. The augmented matrix evaluated at this point reads:

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

Hence, $c_0 = 1, c_1 = 0, c_2 = -1/2$. The conjecture is correct.

Example 10.2

Find the c_i in $p_3(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3$ st $p(x) \approx \cos(x)$ at $x = 0$.

$$\begin{aligned}
\cos(x) &\rightarrow & p(x) &= c_0 + c_1 x + c_2 x^2 + c_3 x^3 \\
\cos'(x) = -\sin(x) &\rightarrow & p'(x) &= c1 + 2c_2 x + 3c_3 x^2 \\
\cos''(x) = -\cos(x) &\rightarrow & p''(x) &= 2c_2 + 6c_3 x \\
\cos'''(x) = \sin(x) &\rightarrow & p'''(x) &= 6c_3
\end{aligned}$$

$$\begin{aligned}
\cos(0) = 1 &\rightarrow & p(0) &= c_0 \\
\cos'(0) = 0 &\rightarrow & p'(0) &= c_1 \\
\cos''(0) = -1 &\rightarrow & p''(0) &= 2c_2 \\
\cos'''(0) = 0 &\rightarrow & p'''(0) &= 6c_3
\end{aligned}$$

$$\begin{bmatrix} 1 & x & x^2 & x^3 \\ 0 & 1 & 2x & 3x^2 \\ 0 & 0 & 2 & 6x \\ 0 & 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} \cos(x) \\ \cos'(x) \\ \cos''(x) \\ \cos'''(x) \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & -1 \\ 0 & 0 & 0 & 6 & 0 \end{bmatrix}.$$

Hence, $c_0 = 1, c_1 = 0, c_2 = -1/2, c_3 = 0.$

Remark 10.1
The polynomial

$$p(x) = 1 - \frac{x^2}{2}$$

is not only the best quadratic approx. of $\cos(x)$*. It is also the best cubic approx.*

Example 10.3

Find the c_i in $p_4(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4$ st $p(x) \approx \cos(x)$ at $x = 0$.

$$\begin{aligned}
\cos(x) &\rightarrow p(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 \\
\cos'(x) = -\sin(x) &\rightarrow p'(x) = c1 + 2c_2 x + 3c_3 x^2 + 4c_4 x^3 \\
\cos''(x) = -\cos(x) &\rightarrow p''(x) = 2c_2 + 6c_3 x + 12c_4 x^2 \\
\cos'''(x) = \sin(x) &\rightarrow p'''(x) = 6c_3 + 24c_4 x \\
\cos''''(x) = \cos(x) &\rightarrow p''''(x) = 24c_4
\end{aligned}$$

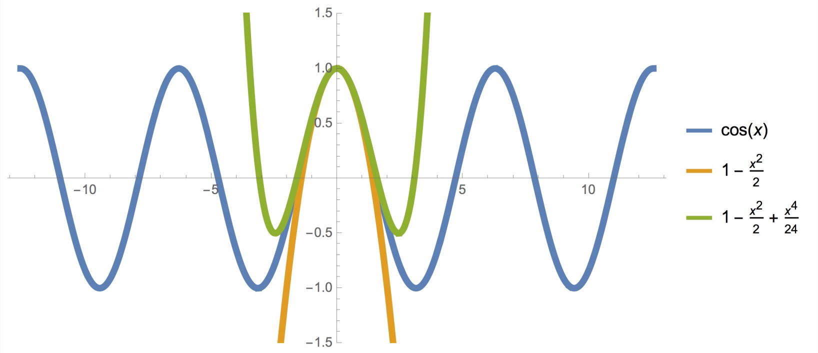
$$\begin{aligned}
\cos(0) = 1 &\rightarrow p(0) = c_0 \\
\cos'(0) = 0 &\rightarrow p'(0) = c_1 \\
\cos''(0) = -1 &\rightarrow p''(0) = 2c_2 \\
\cos'''(0) = 0 &\rightarrow p'''(0) = 6c_3 \\
\cos''''(0) = 1 &\rightarrow p'''(0) = 24c_4
\end{aligned}$$

$$\begin{bmatrix} 1 & x & x^2 & x^3 & x^4 \\ 0 & 1 & 2x & 3x^2 & 4x^3 \\ 0 & 0 & 2 & 6x & 12x^2 \\ 0 & 0 & 0 & 6 & 24x \\ 0 & 0 & 0 & 0 & 24 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix} = \begin{bmatrix} \cos(x) \\ \cos'(x) \\ \cos''(x) \\ \cos'''(x) \\ \cos''''(x) \end{bmatrix}, \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & -1 \\ 0 & 0 & 0 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 24 & 1 \end{array}\right].$$

Hence, $c_0 = 1, c_1 = 0, c_2 = -1/2, c_3 = 0, c_4 = 1/24.$

Remark 10.2

This fourth degree polynomial outperforms the quadratic approximation from the conjecture. But there is a complexity trade-off.



Remark 10.3
Note that we increased precision in a neighbourhood around $x = 0$ using $p_4(x)$ instead of $p_2(x)$ by using derivative information at $x = 0$ only.

Remark 10.4
Note that c_0, c_1 and c_2 are the same in $p_2(x)$ and $p_4(x)$. This is because the augmented matrices have a particular pattern:

$$p_2(x) \quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & -1 \end{array}\right] \quad p_3(x) \quad \left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & -1 \\ 0 & 0 & 0 & 6 & 0 \end{array}\right] \quad p_4(x) \quad \left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & -1 \\ 0 & 0 & 0 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 24 & 1 \end{array}\right].$$

This happens because we evaluate the systems at $x = 0$.

Hence, if we were to approx. $\cos(x)$ at $x = x_0$, we would write:

$$p_4(x) = c_0 + c_1(x - x_0) + c_2(x - x_0)^2 + c_3(x - x_0)^3 + c_4(x - x_0)^4$$

Remark 10.5
Compare the resulting fourth degree approximation

$$p_4(x) = 1 - \frac{1}{2}x^2 + \frac{1}{24}x^4 = 1\frac{1}{1}x^0 + 0\frac{1}{1}x^1 + -1\frac{1}{2}x^2 + 0\frac{1}{6}x^3 + 1\frac{1}{24}x^4$$

and the corresponding augmented matrix

$$\left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & -1 \\ 0 & 0 & 0 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 24 & 1 \end{array}\right].$$

In fact, we can go one step further and write

$$p_4(x) = 1\frac{1}{0!}x^0 + 0\frac{1}{1!}x^1 + -1\frac{1}{2!}x^2 + 0\frac{1}{3!}x^3 + 1\frac{1}{4!}x^4$$

Definition 10.1 (Taylor polynomial)
Given an n-times differentiable function $f : \mathcal{I} \to \mathbb{R}$, in the open interval $\mathcal{I} \subset \mathbb{R}$, the Taylor polynomial of f at $x_0 \in \mathcal{I}$ is defined as

$$p_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k,$$

with approximation error:

$$R_n(x) = f(x) - p_n(x)$$

Example 10.4

Compute the Taylor polynomials of the following functions at $x_0 = 0$

1. $f(x) = e^x$
2. $g(x) = \frac{1}{1-x}$
3. $h(x) = x^2 + 3x + 1$

The corresponding results read:

1. $1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \cdots + \frac{x^n}{n!} + O\left(x^{n+1}\right)$
2. $1 + x + x^2 + x^3 + x^4 + \cdots + x^n + O\left(x^{n+1}\right)$
3. $1 + 3x + x^2$

Theorem 10.1 (Lagrange's Formula for R_n)
If $f : \mathcal{I} \to \mathbb{R} \in \mathcal{C}^{n+1}$ and $x_0, x \in \mathcal{I}$, there exists $x_c \in]x_0, x[$ such that

$$R_n(x) = \frac{f^{(n+1)}(x_c)}{(n+1)!}(x - x_0)^{n+1}.$$

Proof.
Omitted ∎

Corolary 10.1
If $f : \mathcal{I} \to \mathbb{R} \in \mathcal{C}^{n+1}$, then

$$\lim_{x \to x_0} \frac{R_n(x)}{|x - x_0|^n} = 0.$$

Definition 10.2 (Sequence)
Let $x : \mathbb{N} \to \mathbb{R}$, $n \in \mathbb{N}$ and define $x(n) =: x_n \in \mathbb{R}$. The ordered list

$$\{x_1, x_2, x_3, \dots\}$$

is called a sequence, and is denoted by $\{x_n\}_{n\in\mathbb{N}}$.

Example 10.5

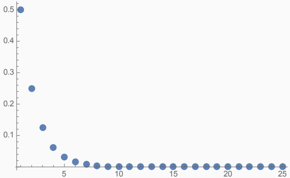
- *sequence of even numbers: $\{2n\}_{n\in\mathbb{N}}$*
- *sequence of odd numbers: $\{2n-1\}_{n\in\mathbb{N}}$*
- *sequence of prime numbers.*
- *sequence of Fibonacci numbers.*

Remark 10.6
We consider sequences indexed by values starting at 1, unless ow stated.

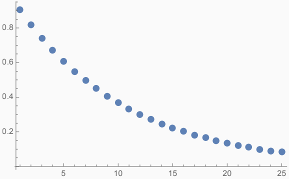
Example 10.6

Consider the following sequences:

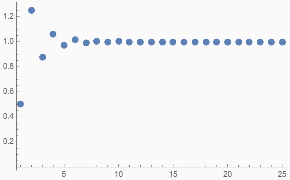
$x_n = 1/2^n$



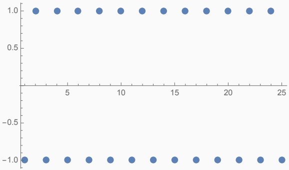
$x_n = e^{-0.1n}$



$x_n = 1 + (-1/2)^n$



$x_n = (-1)^n$



Remark 10.7
Sometimes the elements of a sequence $\{x_n\}_{n\in\mathbb{N}}$ can be written as a *recurrence.*

Example 10.7
Consider sequence $\{2^n\}_{n\in\mathbb{N}}$, and note that $x_n = 2^n = 2^{n-1}2$. Hence

$$\begin{aligned} x_n &= 2^n \\ x_n &= 2x_{n-1}. \end{aligned}$$

The first equality is an explicit expression of x_n in terms of n. The second equality is an expression of x_n in terms of its previous value x_{n-1}.

Example 10.8 (Final Exam (2017-I))

What is the maximum number of pizza slices that one can get by making n cuts?

Example 10.9

The recurrence $x_n = x_{n-1} + x_{n-2}$ with $x_1 = 1$ and $x_2 = 1$ provides the Fibonacci sequence. The numbers in the sequence are called <u>Fibonacci numbers</u>.

Definition 10.3 (Sequence of partial sums of $\{x_n\}_{n\in\mathbb{N}}$)
The sequence of partial sums of $\{x_n\}_{n\in\mathbb{N}}$ is

$$\{\underbrace{x_1}_{s_1}, \underbrace{x_1 + x_2}_{s_2}, \underbrace{x_1 + x_2 + x_3}_{s_3}, \dots\}$$

and is denoted by $\{s_n\}_{n\in\mathbb{N}}$

Definition 10.4 (Series of $\{x_n\}_{n\in\mathbb{N}}$)
Let $\{s_n\}_{n\in\mathbb{N}}$ denote the sequence of partial sums of $\{x_n\}_{n\in\mathbb{N}}$, then

$$\lim_{n\to\infty} s_n = \sum_{n=1}^{\infty} x_n,$$

is called the <u>series</u> associated to $\{x_n\}_{n\in\mathbb{N}}$.

Definition 10.5 (Convergence of $\{x_n\}_{n\in\mathbb{N}}$)
If $\sum_{n=1}^{\infty} x_n$ is a unique scalar, the series converges, otherwise it diverges.

Definition 10.6 (Ratio test)
Given the series $\sum_{n=1}^{\infty} x_n$, the ratio

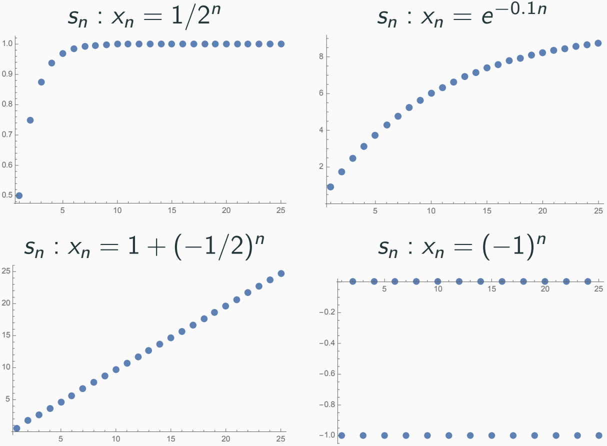
$$L = \lim_{n\to\infty} \left| \frac{x_{n+1}}{x_n} \right|,$$

provides a convergence criteria. If $L < 1$, the series converges, otherwise it diverges.

Example 10.10
Compute the ratio test for the sequences in example 10.6.

Example 10.11

Sequences of partial sums for the sequences in example 10.6.



Example 10.12

Use the ratio test to determine if the series of $\{1/n!\}_{n=0}^{\infty}$ converges.

Remark 10.8

Note that number e is defined by the (convergent) series associated to the sequence in example 10.12

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} = \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \dots$$

Definition 10.7 (Sequence of functions)
Analog to the notion of sequence of numbers, we can define sequences of functions as the list $f_1, f_2, \ldots$. indexed in the naturals, denoted $\{f_n\}_{n \in \mathbb{N}}$.

Definition 10.8 (Pointwise limit of $\{f_n\}_{n \in \mathbb{N}}$)
The pointwise limit of a sequence of functions $\{f_n\}_{n \in \mathbb{N}}$ with domain $\mathcal{D}$ at $x \in \mathcal{D}$ is denoted:

$$f(x) = \lim_{n \to \infty} f_n(x).$$

Example 10.13
Consider $f_n(x) = x^n$, $x \in [0, 1]$ and $n \in \mathbb{N}$. Compute the pointwise limit.

The sequence of functions reads:

$$x, x^2, x^3, \ldots$$

Hence, if $x \in [0, 1)$, $\lim_{n\to\infty} x^n = 0$. In general, the pointwise limit reads:

$$f(x) = \begin{cases} 0, & \text{if } x \in [0, 1) \\ 1, & \text{if } x = 1. \end{cases}$$

Definition 10.9 (Series of $\{f_n\}_{n\in\mathbb{N}}$)
The series of $\{f_n\}_{n\in\mathbb{N}}$ is defined as

$$\sum_{n=1}^{\infty} f_n(x),$$

which is the pointwise limit of the sequence of partial sums of $\{f_n\}_{n\in\mathbb{N}}$.

Definition 10.10 (Power series)
Let $\{f_n\}_{n\in\mathbb{N}}$ denote a sequence of functions where $f_n(x) = a_n(x - x_0)^n$. The series associated to $\{f_n\}_{n\in\mathbb{N}}$

$$\sum_{n=1}^{\infty} a_n(x - x_0)^n,$$

is called power series.

Definition 10.11 (Radius of convergence)
The radius of convergence of a power series is the radius of the largest disk in which the series converges. It is given by:

$$r = \lim_{n\to\infty} \left| \frac{a_n}{a_{n+1}} \right|,$$

and provides a convergence criteria. In $|x - x_0| < r$, the series <u>*converges*</u>*. It* <u>*diverges*</u> *otherwise.*

Definition 10.12 (Taylor series)
Let $f : \mathcal{I} \to \mathbb{R}$, $f \in \mathcal{C}^\infty$. The Taylor series associated to f in x_0 is the power series

$$T_f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n,$$

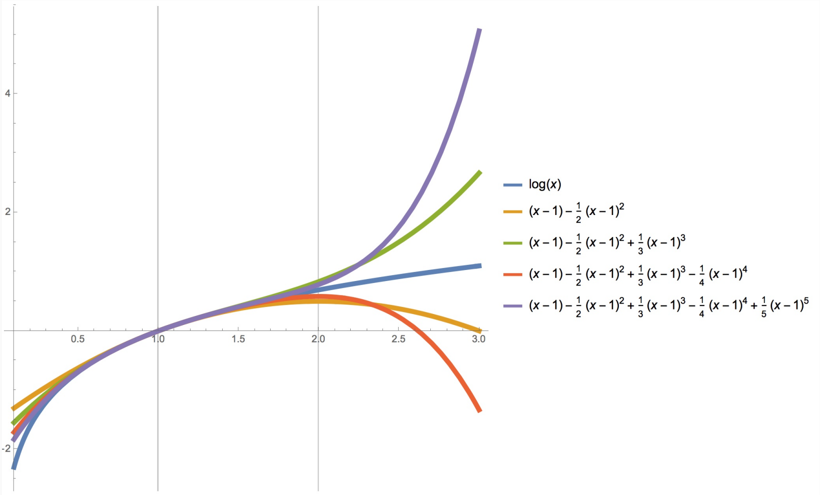
whose convergence radius is given by

$$r = \lim_{n\to\infty} \left| \frac{f^{(n)}(x_0)}{f^{(n+1)}(x_0)}(n+1) \right|.$$

Remark 10.9
If $x_0 = 0$, the Taylor series f is called <u>MacLaurin</u> series.

Example 10.14

Compute the MacLaurin series of $\ln(x)$ *around* $x_0 = 1$*, and determine the radius of convergence.*



Example 10.15

Find the MacLaurin series of the following functions and indicate their convergence intervals:

1. $e^{i\theta}$
2. $\sin(\theta)$
3. $\cos(\theta)$

Example 10.16
Show that Euler's formula holds by using the results in exercise 10.15,

Definition 10.13 (Analytic function)
Function $f : \mathcal{I} \to \mathbb{R}$ is analytic in $x_0 \in \mathcal{I}$, if

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n,$$

for all x near x_0.

Properties of analytic functions:

Let $f(x) = \sum_{n=0}^{\infty} a_n(x-a)^n$, $g(x) = \sum_{n=0}^{\infty} b_n(x-a)^n$, and for $\lambda \in \mathbb{R}$:

- $f(x) + g(x) = \sum_{n=0}^{\infty}(a_n + b_n)(x-a)^n$
- $\lambda f(x) = \sum_{n=0}^{\infty} \lambda \cdot a_n(x-a)^n$
- $f'(x) = \sum_{n=1}^{\infty} n a_n(x-a)^{n-1}$
- $\int f(x)dx = \sum_{n=0}^{\infty} \frac{a_n}{n+1}(x-a)^{n+1} + C_0$

Remark 10.10

Consider the Taylor approximation of a function wrt more than one variable. For $f : \mathbb{R}^n \to \mathbb{R}$, and $x_0 \in \mathbb{R}^n$, one can write

- $p_1(x) = f(x_0) + \nabla f(x_0)(x - x_0)$
- $p_2(x) = f(x_0) + \nabla f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^\top Hf(x_0)(x - x_0)$
- $p_3(x) = \dots$

Topics on Calculus

Definition 11.1 (Transformation)
A transformation $\mathbb{R}^n \to \mathbb{R}^m$ is a mapping

$$F : \mathbb{R}^n \to \mathbb{R}^m$$

that takes vector $\boldsymbol{x} \in \mathbb{R}^n$ and maps it onto $\boldsymbol{F}(\boldsymbol{x}) \in \mathbb{R}^m$.

Remark 11.1
We reserve the word function for the case $m = 1$.

In other words, a transformation is a multiple function application

$$\boldsymbol{F}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \ldots, f_m(\boldsymbol{x}))^\top$$

to vector $\boldsymbol{x} \in \mathbb{R}^n$, for functions $f_i : \mathbb{R}^n \to \mathbb{R}, i = 1, \ldots, m$.

Definition 11.2 (Functions in $\mathcal{C}^q$)
A function $f : \mathbb{R}^n \to \mathbb{R}$ with arguments $x_1, \ldots, x_n$ belongs to $\mathcal{C}^q$ if $\partial^j f / \partial x_i^j$ is continuous for $j = 0, 1, \ldots, q$ and $i = 1, 2, \ldots, n$.

Remark 11.2
All unspecified functions in this class will be assumed to belong to $\mathcal{C}^1$, unless otherwise stated.

Example 11.1

A company uses for inputs (x, y, z, m) to produce three products (U, V, W). The production function is given by:

$$(U, V, W) = F(x, y, z, m) = (\ln(x^2 y) + z, (y^3 zm)^{1/2}, e^m z^2 x + y^2).$$

Definition 11.3 (Gradient of a function)
Let $f : \mathbb{R}^n \to \mathbb{R}$, mapping elements $\boldsymbol{x} \in \mathbb{R}^n$ onto $y \in \mathbb{R}$, for $y = f(\boldsymbol{x})$, upon differentiation

$$dy = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(\boldsymbol{x}) dx_i = \nabla \boldsymbol{f}(\boldsymbol{x}) \cdot d\boldsymbol{x}$$

where $dy \in \mathbb{R}$, $d\boldsymbol{x} \in \mathbb{R}^n$ and the gradient of function f at $\boldsymbol{x}$ is:

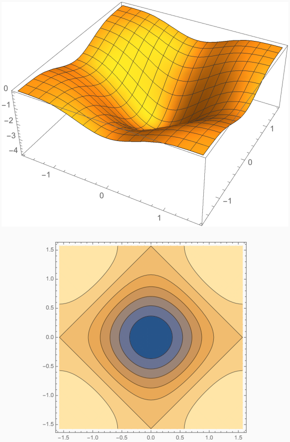
$$\nabla \boldsymbol{f}(\boldsymbol{x}) = \left(\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \ldots, \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \right) = \left(f_{x_1}(\boldsymbol{x}), \ldots, f_{x_n}(\boldsymbol{x}) \right),$$

defined, by convention, as a row vector.

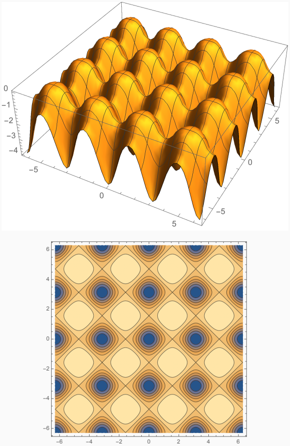
Example 11.2

Consider $f(x,y) = -\left(\cos(x)^2 + \cos(y)^2\right)^2$, *compute* $\nabla \boldsymbol{f}$ *and interpret.*

$x, y \in [-\pi/2, \pi/2]$



$x, y \in [-2\pi, 2\pi]$



Definition 11.4 (Directional derivative)
Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\boldsymbol{x}, \boldsymbol{u} \in \mathbb{R}^n$. The directional derivative of function $f(\cdot)$ at $\boldsymbol{x}$ wrt direction $\boldsymbol{u}$ is

$$\partial_{\boldsymbol{u}} f(\boldsymbol{x}) = \lim_{h \to 0} \frac{f(\boldsymbol{x} + h\boldsymbol{u}) - f(\boldsymbol{x})}{h}.$$

Remark 11.3
Note that if $f \in \mathcal{C}^1$ then

- $\partial_{\boldsymbol{u}} f(\boldsymbol{x}) = \nabla f(\boldsymbol{x}) \cdot \boldsymbol{u}$.
- $\partial_{\boldsymbol{e}_i} f(\boldsymbol{x}) = \dfrac{\partial f}{\partial x_i}(\boldsymbol{x}), \quad i = 1, \ldots, n.$

Definition 11.5 (Jacobian matrix)
Let transformation $F : \mathbb{R}^n \to \mathbb{R}^m$. The Jacobian matrix of F at $\boldsymbol{x}$ is

$$\boldsymbol{DF}(\boldsymbol{x}) = \begin{bmatrix} \nabla \boldsymbol{f}_1(\boldsymbol{x}) \\ \vdots \\ \nabla \boldsymbol{f}_m(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{x}) & \frac{\partial f_1}{\partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial f_1}{\partial x_n}(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\boldsymbol{x}) & \frac{\partial f_m}{\partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial f_m}{\partial x_n}(\boldsymbol{x}) \end{bmatrix}.$$

The derivative $\boldsymbol{y}$ is then $d\boldsymbol{y} = \boldsymbol{DF}(\boldsymbol{x}) \cdot d\boldsymbol{x}$.

Remark 11.4

- *If $m = n$, $\boldsymbol{DF}(\boldsymbol{x}) \in \mathbb{R}^{n \times n}$ and $\det(\boldsymbol{DF}(\boldsymbol{x}))$ is the Jacobian of F at $\boldsymbol{x}$.*
- *The approximate variation in F from $\boldsymbol{x}_0$ to $\boldsymbol{x}$ is the discrete analog $\Delta \boldsymbol{y} = \boldsymbol{DF}(\boldsymbol{x}_0) \cdot \Delta \boldsymbol{x}$, where $\Delta \boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_0$.*

Example 11.3

Using the information in example 11.1, compute $\boldsymbol{DF}(\boldsymbol{x})$ *and find the marginal productivity of input* z *in the production of* U*,* V *and* W*, at the current input usage* $(x, y, z, m)^\top = (1, 1, 1, 1)^\top$*.*

Definition 11.6 (Hessian matrix)
The gradient ∇f is a transformation $G : \mathbb{R}^n \to \mathbb{R}^n$, of the form

$$\boldsymbol{G}(\boldsymbol{x}) = \nabla \boldsymbol{f}(\boldsymbol{x})^\top = \begin{bmatrix} f_{x_1}(\boldsymbol{x}) \\ \vdots \\ f_{x_n}(\boldsymbol{x}) \end{bmatrix}.$$

The Hessian of function $f(\cdot)$ is the Jacobian of the gradient $\nabla f^\top$,

$$\boldsymbol{Hf}(\boldsymbol{x}) = \boldsymbol{DG}(\boldsymbol{x}) = \begin{bmatrix} \nabla \boldsymbol{f}_{x_1}(\boldsymbol{x}) \\ \nabla \boldsymbol{f}_{x_2}(\boldsymbol{x}) \\ \vdots \\ \nabla \boldsymbol{f}_{x_n}(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} f_{x_1x_1}(\boldsymbol{x}) & f_{x_1x_2}(\boldsymbol{x}) & \cdots & f_{x_1x_n}(\boldsymbol{x}) \\ f_{x_2x_1}(\boldsymbol{x}) & f_{x_2x_2}(\boldsymbol{x}) & \cdots & f_{x_2x_n}(\boldsymbol{x}) \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_nx_1}(\boldsymbol{x}) & f_{x_nx_2}(\boldsymbol{x}) & \cdots & f_{x_nx_n}(\boldsymbol{x}), \end{bmatrix}$$

where $f_{x_ix_j}(\boldsymbol{x}) = \dfrac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x})$.

Example 11.4
Using the information in example 11.2 and using the Hessian matrix determine if $f(0,0)$ is a minimum.

FOC:

$$\begin{aligned} f_x &= 4\sin(x)\cos(x)\left(\cos^2(x)+\cos^2(y)\right) \\ f_y &= 4\cos(y)(\cos(x)^2+\cos(y)^2)\sin(y) \end{aligned}$$

SOC:

$$\begin{aligned} f_{xx} &= 2(\cos(2x)(\cos(2y)+2)+\cos(4x)) \\ f_{yy} &= 2(\cos(2y)(\cos(2x)+2)+\cos(4y)) \\ f_{xy} &= -8\sin(x)\cos(x)\sin(y)\cos(y) \end{aligned}$$

Since $f_{xx}(0,0)=f_{yy}(0,0)=8$, $f_{xy}(0,0)=0$, and $\det(\boldsymbol{Hf}(0,0))=64$, it follows that $f(x,y)$ is a minimum at (0,0).

Generalized Chain Rule

Let $F : \mathbb{R}^n \to \mathbb{R}^m$ and $G : \mathbb{R}^m \to \mathbb{R}^p$. The composition of G with F is denoted $G \circ F : \mathbb{R}^n \to \mathbb{R}^p$

$$(G \circ F)(\boldsymbol{x}) = G(F(\boldsymbol{x})).$$

Theorem 11.1 (Generalized Chain Rule)
Let $F : \mathbb{R}^n \to \mathbb{R}^m$ and $G : \mathbb{R}^m \to \mathbb{R}^p$. Then

$$D(G \circ F)(\boldsymbol{x}) = \boldsymbol{DG}(\boldsymbol{F}(\boldsymbol{x})) \cdot \boldsymbol{DF}(\boldsymbol{x}).$$

Proof.
Omitted. ∎

Example 11.5

Consider the information in example 11.1. Additionally, suppose that U, V and W are inputs themselves for product Q, st. $Q = \alpha U + \beta V + \gamma W$. Compute $D(Q \circ F)(x, y, z, m)$ at $(x, y, z, m)^\top = (1, 1, 1, 1)^\top$.

Implicit Transformations

Definition 11.7 (Implicit Transformation)

Let $G : \mathbb{R}^n \to \mathbb{R}^m$ denote a transformation. In particular, the expressions

$$\begin{aligned} \boldsymbol{y} &= G(\boldsymbol{x}), \quad \boldsymbol{y} \in \mathbb{R}^m, \boldsymbol{x} \in \mathbb{R}^n && (9) \\ F(\boldsymbol{x}, \boldsymbol{y}) &= \boldsymbol{y} - G(\boldsymbol{x}) = 0, \quad \boldsymbol{y} \in \mathbb{R}^m, \boldsymbol{x} \in \mathbb{R}^n, && (10) \end{aligned}$$

are equivalent. In expression (9) transformation $G(\cdot)$ is given explicitly, while in (10), $G(\cdot)$ is given as an implicit transformation inside of

$$F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m \quad \Leftrightarrow \quad F : \mathbb{R}^{n+m} \to \mathbb{R}^m,$$

which is a different transformation in terms of two vector variables.

Remark 11.5

Going from $G(\cdot)$ to $F(\cdot)$ is trivial, but going from $F(\cdot)$ to $G(\cdot)$ is not.

Remark 11.6

Given $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^m$, the Jacobian reads:

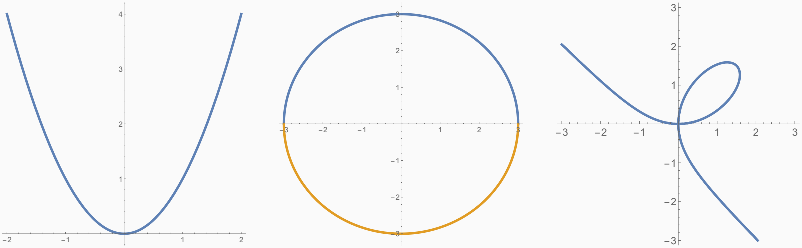
$$\boldsymbol{DF}(\boldsymbol{x}, \boldsymbol{y}) = \begin{bmatrix} \nabla \boldsymbol{f}_1(\boldsymbol{x}, \boldsymbol{y}) \\ \vdots \\ \nabla \boldsymbol{f}_m(\boldsymbol{x}, \boldsymbol{y}) \end{bmatrix} = \left[\boldsymbol{D_x F}(\boldsymbol{x}, \boldsymbol{y}) \quad \boldsymbol{D_y F}(\boldsymbol{x}, \boldsymbol{y}) \right],$$

where $\boldsymbol{D_x F}(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{m \times n}$, and $\boldsymbol{D_y F}(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{m \times m}$.

Example 11.6

Consider the transformations

$$
\begin{aligned}
F_1(x,y) &= y - x^2 &\rightarrow\quad & G_1(x) = x^2 \\
F_2(x,y) &= x^2 + y^2 - 9 &\rightarrow\quad & G_{1,2}(x) = \pm\sqrt{9 - x^2}, \quad x \in (-3,3) \\
F_3(x,y) &= x^3 - 3xy + y^3 &\rightarrow\quad & \textit{more involved.}
\end{aligned}
$$



Theorem 11.2 (Implicit Transformation)

Let $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$*, and define the set*

$$\mathcal{K} = \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{n+m} \ : \ F(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}\}.$$

If

$$\det(\boldsymbol{D_y F}(\boldsymbol{x}, \boldsymbol{y})) \neq 0, \quad (\boldsymbol{x}, \boldsymbol{y}) \in (\mathcal{U} \times \mathcal{V}) \subseteq \mathcal{K},$$

then there exists an implicit function $G(\cdot)$ *where*

$$\begin{aligned} &G : \mathcal{U} \to \mathbb{R}^m \quad \textit{such that} \\ &G(\boldsymbol{x}) = \boldsymbol{y} \textit{ and } F(\boldsymbol{x}, G(\boldsymbol{x})) = \boldsymbol{0}, \quad \boldsymbol{x} \in \mathcal{U}, \boldsymbol{y} \in \mathcal{V} \end{aligned}$$

Sketch of proof.

Let $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$, and define the set

$$\mathcal{K} = \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{n+m} \ : \ F(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}\}.$$

Upon differentiation

$$\boldsymbol{D_x F}(\boldsymbol{x}, \boldsymbol{y}) \cdot d\boldsymbol{x} + \boldsymbol{D_y F}(\boldsymbol{x}, \boldsymbol{y}) \cdot d\boldsymbol{y} = \boldsymbol{0}, \quad (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{K}$$

which can be re-written as:

$$d\boldsymbol{y} = (\boldsymbol{D_y F}(\boldsymbol{x}, \boldsymbol{y}))^{-1} \cdot (-\boldsymbol{D_x F}(\boldsymbol{x}, \boldsymbol{y}) \cdot d\boldsymbol{x}),$$

only if there exists some domain where the former inverse exists, i.e.

$$\det(\boldsymbol{D_y F}(\boldsymbol{x}, \boldsymbol{y})) \neq 0, \quad (\boldsymbol{x}, \boldsymbol{y}) \in (\mathcal{U} \times \mathcal{V}) \subseteq \mathcal{K}.$$

■

Example 11.7

The Jacobians of transformations

$$\begin{aligned} F_1(x,y) &= y - x^2 \\ F_2(x,y) &= x^2 + y^2 - 9, \end{aligned}$$

are respectively

$$\begin{aligned} \boldsymbol{DF}_1(\boldsymbol{x},\boldsymbol{y}) &= \begin{bmatrix} \boldsymbol{D_x F}_1(\boldsymbol{x},\boldsymbol{y}) & \boldsymbol{D_y F}_1(\boldsymbol{x},\boldsymbol{y}) \end{bmatrix} = \begin{bmatrix} -2x & 1 \end{bmatrix} \\ \boldsymbol{DF}_2(\boldsymbol{x},\boldsymbol{y}) &= \begin{bmatrix} \boldsymbol{D_x F}_2(\boldsymbol{x},\boldsymbol{y}) & \boldsymbol{D_y F}_2(\boldsymbol{x},\boldsymbol{y}) \end{bmatrix} = \begin{bmatrix} 2x & 2y \end{bmatrix}, \end{aligned}$$

in the first case $\det(\boldsymbol{D_y F}(\boldsymbol{x}_{\mathcal{K}}, \boldsymbol{y}_{\mathcal{K}}))$ *is always non-zero, while in the second case it is non-zero only if* $y \neq 0$*, i.e. if* $y > 0$ *or* $y < 0$*. In fact*

$$\begin{aligned} G_1(x) &= +\sqrt{9-x^2}, \quad x \in (-3,3) \quad \rightarrow \quad y = G_1(x) > 0 \\ G_2(x) &= -\sqrt{9-x^2}, \quad x \in (-3,3) \quad \rightarrow \quad y = G_2(x) < 0 \end{aligned}$$

Cramer's Rule

Sometimes the dimensionality of the system is large and quantity $\boldsymbol{y}_j$ is of interest. In such cases one can use Cramer's rule:

$$d\boldsymbol{y}_j = \frac{\det\left([\boldsymbol{D_y F}(\boldsymbol{x}_{\mathcal{K}}, \boldsymbol{y}_{\mathcal{K}})]_j\right)}{\det\left(\boldsymbol{D_y F}(\boldsymbol{x}_{\mathcal{K}}, \boldsymbol{y}_{\mathcal{K}})\right)},$$

where $[\boldsymbol{D_y F}(\boldsymbol{x}_{\mathcal{K}}, \boldsymbol{y}_{\mathcal{K}})]_j$ is the matrix resulting from replacing the j-th column of $\boldsymbol{D_y F}(\boldsymbol{x}_{\mathcal{K}}, \boldsymbol{y}_{\mathcal{K}})$ by $(-\boldsymbol{D_x F}(\boldsymbol{x}_{\mathcal{K}}, \boldsymbol{y}_{\mathcal{K}}) \cdot d\boldsymbol{x})$.

Comparative Statics

Consider a system of equations with variables

$$\begin{aligned} \boldsymbol{z} &= \{x_1, x_2, \ldots, y_1, y_2, \ldots\} \\ &= \{\boldsymbol{x}, \boldsymbol{y}\} \end{aligned}$$

called exogeneous ($\boldsymbol{x}$) and endogenous ($\boldsymbol{y}$). Economists are interested in writting $\boldsymbol{y}$ in terms of $\boldsymbol{x}$. But that is <u>not</u> always possible.

Remark 12.1
Writing $\boldsymbol{y}$ in terms of $\boldsymbol{x}$ is only feasible if, for the sys. rep. $F(\boldsymbol{x}, \boldsymbol{y}) = 0$, the conditions of theorem 11.2 hold, i.e. $\det(\boldsymbol{D_y F}(\boldsymbol{x}, \boldsymbol{y})) \neq 0$.

Remark 12.2
A necessary condition to write $\boldsymbol{y}$ in terms of $\boldsymbol{x}$ is that the number of endogeneous variables matches the number of equations in the system.

Example 12.1
Consider the following economic model

$$\begin{aligned} &\textit{Demand:} && Q^D = D(P, Y) && D_P < 0, D_Y \geq 0 \\ &\textit{Supply:} && Q^S = S(P, A) && S_P > 0, S_A \geq 0 \\ &\textit{Equilibrium:} && Q \equiv Q^D = Q^S, && \end{aligned}$$

where Y and A are assumed exogenous.

1. *Find the equilibrium condition in one equation. Verify the conditions for the existence of and implicit function for P in terms of Y and A.*
2. *Assuming the previous conditions hold, compute the effect on P and Q of an increment in Y, and an increment in A.*
3. *Repeat the previous analysis for a two equation model.*

Example 12.2

Consider the following macroeconomic model

$$
\begin{aligned}
Y &= C(Y) + G + X(Y^*, q) - M(Y, q), && 0 < C_Y < 1,\\
Y^* &= C^*(Y^*) + G^* + X^*(Y, q) - M^*(Y^*, q), && 0 < C^*_{Y^*} < 1,\\
X^*(Y, q) &= M(Y, q), && 0 < M_Y < 1,\ M_q < 0,\\
M^*(Y^*, q) &= X(Y^*, q), && 0 < X_{Y^*} < 1,\ X_q > 0,
\end{aligned}
$$

where G, G^* and q are exogeneous variables.

1. *What the conditions for the existence of implicit functions in the endogeneous variables wrt the exogeneous variables?*
2. *Assuming the former conditions hold, compute the effect on Y and Y^* given an increment in G, keeping G^* and q fixed.*
3. *¿Under which conditions an increment in q does not have an effect on world income $Y + Y^*$?*

Ordinary Differential Equations

Example 13.1 (Lascaux caves)

In the year 1940 a group of boys was hiking in the vicinity of a town in France called Lascaux. They suddenly became aware that their dog had disappeared. In the ensuing search he was found in a deep hole from he was unable to climb out. When one of the boys lowered himself into the hole to help extricate the dog, he made a startling discovery. The hole was in fact an ancient cave. On the walls of the cave there were marvellous paintings of stags, wild horses and cattle. In addition to the wall paintings, there were also found the charcoal remains of a fire. The problem of interest is to determine: how long ago the cave inhabitants lived there?

Consider that:

1. Charcoal is burnt wood.
2. Organic matter changes with time.
3. All living organisms contain two isotopes of carbon: C^{12} and C^{14}.
4. After death, C^{14} is progressively lost because of radiation.
5. The rate of decomposition of C^{14} is proportional to its level.
6. 99.876% of C^{14} present at death will remain after 10 years.
7. 85.5% of C^{14} present at death had decomposed by the time of the discovery.

C-14 model:

The dynamic behaviour of C^{14} can be represented by:

$$\frac{dx}{dt} = -kx, \quad k > 0,$$

where x denotes the amount of C^{14}, t is time and k is a constant.

The solution reads:

$$x(t) = e^{-kt}H, \quad H = x(0),$$

where H represents the amount of C^{14} at the moment of death $t = 0$.
We are interested in finding the elapsed time until the discovery $t = T$:

$$T = \ln(x(T)H^{-1})^{-1/k}$$

Given the model

$$x(t) = e^{-kt}H, \quad H = x(0),$$

and since 99.876% of x remains after 10 years of death,

$$x(10) = 0.99876H = e^{-10k}H \quad \Rightarrow \quad k = 0.000124077.$$

Similarly, using the fact that 85.5% of C^{14} present at death had decomposed at the time of the discovery $t = T$, it follows that

$$\begin{aligned} T &= \ln(x(T)H^{-1}) \times (-1/k) \\ &= \ln((1 - 0.855)HH^{-1}) \times (-1/k) \\ &= \ln(0.145) \times (-1/0.000124077) \end{aligned}$$

i.e. the cave inhabitants lived there approx. 15500 years ago.

Remark 13.1

The method we have described for determining the age of an archeological remain is known as the carbon-14 test.

Remark 13.2

The C-14 model is actually a first order ordinary differential equation (ODE).

Example 13.2 (Newton's law)

Mass times acceleration equals force, $ma = f$, where m is the particle mass, $a = d^2x/dt^2$ is the particle acceleration, and f is the force acting on the particle. Hence Newton's law is the differential equation

$$m\frac{d^2x}{dt^2}(t) = f\left(t, x(t), \frac{dx}{dt}(t)\right),$$

where the unknown is $x(t)$ – the position of the particle in space at the time t. As we see above, the force may depend on time, on the particle position in space, and on the particle velocity.

Remark 13.3

Example 13.2 shows a <u>second order</u> ODE.

Example 13.3 (Heat equation)

The temperature T in a solid material changes in time and in three space dimensions – labeled by $\boldsymbol{x} = (x, y, z)$ – according to the equation

$$\frac{\partial T}{\partial t}(t, \boldsymbol{x}) = k \left(\frac{\partial^2 T}{\partial x^2}(t, \boldsymbol{x}) + \frac{\partial^2 T}{\partial y^2}(t, \boldsymbol{x}) + \frac{\partial^2 T}{\partial z^2}(t, \boldsymbol{x}) \right), \quad k > 0$$

where k is a positive constant representing thermal properties of the material.

Remark 13.4

Example 13.3 shows a PDE (not ODE). It is a <u>first order</u> PDE in time and a <u>second order</u> PDE in space.

Basics

<u>Classification:</u>

- Autonomous: t is not explicitly in the equation.
- Linear: is linear in x and its derivatives.
- Homogeneous: if $x(t)$ is sol, then $c \cdot x(t), c \neq 0$ is sol.

Example 13.4
Classify the following ODEs:

- $\ddot{x} + 3\dot{x} + 2x = t^2$
- $x^{(3)} + (\dot{x})^2 - 1 = 0$
- $\ddot{x} + 2t^2\dot{x} - tx = e^t$
- $\dot{x} - e^t x = 0$
- $3\ddot{x} - \dot{x} + x = 0$
- $\dot{x} = f(t)$

Definition 13.1 (First order ODE)
A first order ODE on the unknown x is

$$\dot{x} = f(t, x(t)),$$

where f is a given function and $\dot{x} = dx/dt$.

Example 13.5
In particular, a linear first order ODE reads

$$\dot{x} = a(t)x + b(t).$$

where $a(t)$ is called coefficient and $b(t)$ is called term. If both are constants we call the equation constant coefficient & constant term equation (CC&CT).

Example 13.6
Show that $x(t) = e^{2t} - 3/2$ *is solution of the equation*

$$\dot{x} = 2x + 3$$

Example 13.7
Find the differential equation $\dot{x} = f(x)$ *satisfied by*

$$x(t) = 4e^{2t} + 3$$

Theorem 13.1 (Linear CC&CT Equations)

The linear ODE

$$\dot{x} = ax + b, \quad a, b \in \mathbb{R}, a \neq 0$$

has solutions

$$x(t) = He^{at} - \frac{b}{a}, \quad H \in \mathbb{R}.$$

Proof.
Case 1: $b = 0$, trivially $x(t) = He^{at}$. Case 2: $b \neq 0$, take $\tilde{x} = x + b/a$, so that $\dot{\tilde{x}} = a\tilde{x}$. By the result in case 1, it follows that $\tilde{x}(t) = He^{at}$ and

$$x(t) = He^{at} - \frac{b}{a}$$

■

Remark 13.5

We solved $\dot{x} = ax$, by transforming it into a total derivative,

$$\ln(x)' = a \quad \Rightarrow \quad (\ln(x) - at)' = 0 \quad \Rightarrow \quad \psi'(t, x(t)) = 0,$$

with respect to a potential function $\psi(t, x(t)) = \ln(x(t)) - at$. We solved $\dot{x} = ax + b$ using the potential function

$$\psi(t, x(t)) = \ln\left(x(t) + \frac{b}{a}\right) - at.$$

Proof (revisited).
Consider $\dot{x} - ax = b$ and multiply both sides by $\mu : \dot{\mu} = -a\mu$, hence

$$\mu\dot{x} - a\mu x = \mu b \quad \Rightarrow \quad \mu\dot{x} + \dot{\mu}x = \mu b \quad \Rightarrow \quad (\mu x)' = \mu b$$

Function μ is called an integrating factor. Clearly $\mu(t) \propto e^{-at}$. Hence

$$(e^{-at}x)' = e^{-at}b \quad \Rightarrow \quad \left(e^{-at}x + \frac{b}{a}e^{-at}\right)' = 0,$$

with potential function $\psi(t, x(t)) = \left(x + \frac{b}{a}\right) e^{-at}$. Integration leads to

$$x(t) = He^{at} - \frac{b}{a}.$$

$\blacksquare$

Example 13.8
Find all the solutions to the constant coefficient equation $\dot{x} = 2x + 3$.

Remark 13.6
The method used in the revisited proof of theorem 13.1 is called the integrating factor method.

Theorem 13.2 (CC&CT IVP)

The initial value problem:

$$\dot{x} = ax + b, \quad x(t_0) = x_0,$$

has a unique solution given by:

$$x(t) = \left(x_0 + \frac{b}{a}\right) e^{a(t-t_0)} - \frac{b}{a}$$

Proof.
Trivial. Notice that at $t = t_0$, $H = (x_0 + b/a)e^{-at_0}$ and replace. ∎

Example 13.9
Find the solution $x(t)$ *of* $\dot{x} = 2x + 3$ *with* $x(0) = 1$.

Definition 13.2 (Steady state)
Given $\dot{x} = f(x(t), t)$, the quantity

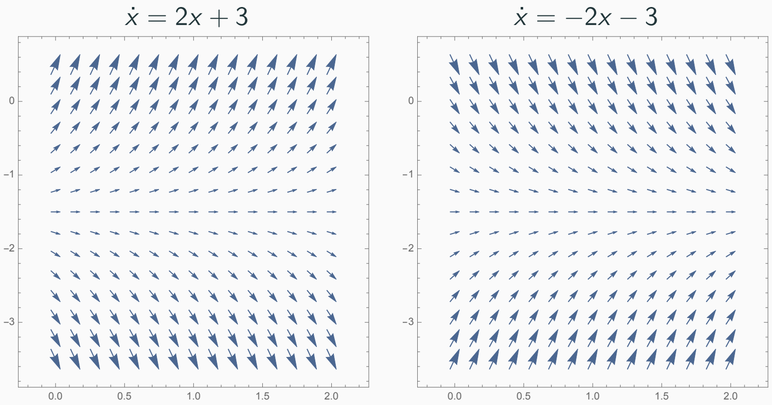
$$x_{ss} = \lim_{t \to \infty} x(t),$$

is called steady state as long as the limit is finite.

Remark 13.7
If a steady state $x_{ss} \in \mathbb{R}$ exists, we say $x(t)$ is <u>convergent</u>. Otherwise we say it is <u>divergent</u>.

Remark 13.8
If x_{ss} is provided ahead, we call that ODE a <u>terminal value problem</u>.

Direction fields:



Remark 13.9
The unique solution of the IVP for $a < 0$ can be written as

$$x(t) = (x_0 - x_{ss})\, e^{a(t-t_0)} + x_{ss}, \quad x_{ss} = -\frac{b}{a}.$$

In the economic jargon $x_0 - x_{ss}$ is the initial disequilibrium and x_{ss} is the long term equilibrium.

Remark 13.10
In general, constant coefficient equation methods are <u>not</u> useful for varying coefficient equations. But, in some specific cases, they are.

Example 13.10
Consider the varying coefficient equation

$$\dot{x} = a(t)x + b(t), \quad b(t)/a(t) = k, a(t) \neq 0$$

where $k \in \mathbb{R}$. Note that for $\tilde{x} = x + k$, it holds that $\dot{\tilde{x}} = a(t)\tilde{x}$

$$x(t) = He^{A(t)dt} - \frac{b(t)}{a(t)},$$

where $\int a(t)dt = A(t) + const$ and $H = e^{const}$.

Theorem 13.3 (VC&VT)

The ordinary differential equation

$$\dot{x} = a(t)x + b(t), \quad a, b \in \mathcal{C},$$

has infinetly many solutions given by

$$x(t) = He^{A(t)} + e^{A(t)} \int e^{-A(t)} b(t) dt,$$

where $\int a(t)dt = A(t) + const.$ *Moreover, the function* $e^{-A(t)} = \mu(t)$ *is called integrating factor of the equation.*

Proof.
Consider $\dot{x} - a(t)x = b(t)$ and multiply both sides by $\mu : \dot{\mu} = -a(t)\mu$, hence

$$\mu\dot{x} - a(t)\mu x = \mu b(t) \quad \Rightarrow \quad \mu\dot{x} + \dot{\mu}x = \mu b(t) \quad \Rightarrow \quad (\mu x)' = \mu b(t)$$

Function μ is called an <u>integrating factor</u>. Clearly $\mu(t) \propto e^{-A(t)}$. Hence

$$\left(e^{-A(t)}x\right)' = e^{-A(t)}b(t) \quad \Rightarrow \quad \left(e^{-A(t)}x - \int e^{-A(t)}b(t)dt + H\right)' = 0,$$

with potential function $\psi(t, x(t)) = e^{-A(t)}x - \int e^{-A(t)}b(t)dt + H$. Integration leads to

$$x(t) = He^{A(t)} + e^{A(t)} \int e^{-A(t)}b(t)dt.$$

■

Linear Varying Term Equations

Theorem 13.4 (Varying Terms)

The ordinary differential equation

$$\dot{x} = ax + b(t), \quad a \in \mathbb{R}, b \in \mathcal{C}, a \neq 0$$

has infinitely many solutions given by

$$x(t) = He^{at} + e^{at} \int e^{-at} b(t) dt.$$

Proof.
Trivial. Special case of theorem 13.3. ∎

Remark 13.11

It can be shown that the solution of $\dot{x} = ax + b(t)$ in theorem 13.4 can be written as the sum of two linearly independent funcs $x(t) = x_c(t) + x_p(t)$:

$$\begin{aligned} x_c(t) &= He^{at} \\ x_p(t) &= e^{at} \int e^{-at} b(t) dt. \end{aligned}$$

$x_c(t)$ is called complementary (or homogeneous) solution and $x_p(t)$ is called particular (or inhomogeneous) solution.

Definition 13.3 (Linear independence btw funcs)
Functions $x_1, x_2 : \mathcal{I} \to \mathbb{R}$ are ld on $\mathcal{I}$ iff $\exists c$ st $x_1(t) = cx_2(t)$ for all $t \in \mathcal{I}$. Otherwise the two functions are called li.

Definition 13.4 (Wronksian)
The Wronskian of the differentiable functions x_1, x_2 is the function

$$W_{x_1,x_2}(t) = \det\left(\begin{bmatrix} x_1(t) & \dot{x}_1(t) \\ x_2(t) & \dot{x}_2(t) \end{bmatrix}\right)$$

Theorem 13.5 (Wronksian)
If the functions $x_1, x_2 : \mathcal{I} \to \mathbb{R}$ are ld, then their Wronskian function vanishes on $\mathcal{I}$, i.e. $W_{x_1,x_2}(t) = 0$.

Proof.
Trivial. Consider $x_2(t) = cx_1(t)$ and use the Wronskian's definition. ■

Algorithm 3 (Varying terms)

Input: $\dot{x} = ax + b(t)$, $a \in \mathbb{R}, b \in \mathcal{C}, a \neq 0$. *E.g.* $b(t) = b$.

1. *Conjecture* $x_c(t) = He^{rt}$, $H \neq 0$, *so that* $\dot{x}_c = Hre^{rt}$. *Replace in the homogeneous equation*

$$\dot{x}_c = ax_c + 0 \quad \Rightarrow \quad Hre^{rt} = Hae^{rt} \quad \Rightarrow \quad He^{rt}P(r) = 0,$$

   where $P(r) = (r - a)$, *which is satisfied at* $r = a$, *so* $x_c(t) = He^{at}$.

2. *Conjecture* $x_p(t)$ *has the same form as* $b(t)$, *i.e.* $x_p(t) = k$, *so that* $\dot{x}_p = 0$. *Replace in the inhomogeneous equation*

$$\dot{x}_p = ax_p + b \quad \Rightarrow \quad 0 = ak + b \quad \Rightarrow \quad k = -\frac{b}{a},$$

   and conclude that $x_p(t) = -b/a$.

Output: $x(t) = x_c(t) + x_p(t) = He^{at} - b/a$.

Remark 13.12
Note that the first step in algorithm 3 is trivial for first order ODEs. However the method is generalizable to ODEs with arbitrary order.

Remark 13.13
In algorithm 3, $P(r)$ is a polynomial in r, called the characteristic polynomial of the ODE. Its degree is the same as the order of the ODE.

Use algorithm 3 to find the function x solution of the ODE

$$\dot{x} = ax + b(t), \quad a \in \mathbb{R}, b \in \mathcal{C}, a \neq 0,$$

for the following cases:

1. $b(t) = b_0 + b_1 t + b_2 t^2$
2. $b(t) = b_1 \cos(\theta t) + b_2 \sin(\theta t)$
3. $b(t) = be^{\theta t}$, $b \neq 0$

The complementary solution for all cases is the same. Since the ODE is of first order, it has the trivial solution

$$x_c(t) = He^{at}, \quad H \neq 0.$$

The solutions for the particular cases follow

1. Polynomial

$$x_p(t) = -\left(\frac{2b_2 + ab_1 + a^2 b_0}{a^3}\right) - \left(\frac{2b_2 + ab_1}{a^2}\right) t - \frac{b^2}{a} t^2$$

2. Trigonometric

$$x_p(t) = -\left(\frac{ab_1 + \theta b_2}{a^2 + \theta^2}\right) \cos(\theta t) + \left(\frac{\theta b_1 - ab_2}{a^2 + \theta^2}\right) \sin(\theta t)$$

3. Exponential

$$\begin{aligned} x_p(t) &= \left(\frac{b}{\theta - a}\right) e^{\theta t}, \quad \theta \neq a \\ x_p(t) &= bte^{\theta t}, \quad \theta = a \end{aligned}$$

Remark 13.14
Non-linear ODEs are harder to solve than linear ODEs. Two exceptions are the Bernoulli equation, and separable equations.

Definition 13.5 (Bernoulli equation)
The Bernoulli equation is

$$\dot{x} = p(t)x + q(t)x^n,$$

where p, q are known functions, and $n \in \mathbb{R}$.

Remark 13.15
If $n \neq 0, 1$ the equation is non-linear. If $n = 2$ the equation is called logistic equation.

Theorem 13.6 (Bernoulli)

The function x is a solution of the Bernoulli equation

$$\dot{x} = p(t)x + q(t)x^n, \quad n \neq 1,$$

iff the function $v = 1/x^{n-1}$ is solution of the linear differential equation

$$\dot{v} = -(n-1)p(t)v - (n-1)q(t),$$

Proof.

Let $v = x^{-(n-1)}$, thus $-\dot{v}/(n-1) = \dot{x}/x^n$. Dividing the Bernoulli equation by x^n we obatin

$$\frac{\dot{x}}{x^n} = \frac{p(t)}{x^{n-1}} + q(t) \quad \Rightarrow \quad \dot{v} = -(n-1)p(t)v - (n-1)q(t)$$

$\blacksquare$

Example 13.11

Find every non-zero solution x of

$$\dot{x} = x + 2x^5$$

Definition 13.6 (Separable equations)
A separable differential equation is an ODE that can be written as

$$h(x)\dot{x} = g(t),$$

where h and g are known functions.

Example 13.12

- $\dot{x} = \frac{t^2}{1-x^2}$
- $\dot{x} + x^2 \cos(2t) = 0$
- $\dot{x} = a(t)x$
- $\dot{x} = e^x + \cos(t)$
- $\dot{x} = a_0 x + b_0$
- $\dot{x} = a(t)x + b(t)$

Remark 13.16
Linear ODEs are separable iff $a(t)/b(t)$ is a constant.

Theorem 13.7 (Separable equation)
If $h, g \in \mathcal{C}$, $h \neq 0$, then

$$h(x)\dot{x} = g(t),$$

has infinitely many solutions satisfying

$$H(x(t)) = G(t) + const, \quad const \in \mathbb{R},$$

where H and G are the antiderivatives of h and g respectively.

Proof.

$$\int \left(h(x)\frac{dx}{dt} - g(t) \right) dt = 0 \quad \Rightarrow \quad \int (H(x(t)) - G(t))' = 0.$$

Re-ordering the terms $H(x(t)) = G(t) + const$. ∎

Example 13.13
Find all solutions x to the differential equation

$$-\frac{\dot{x}}{x^2} = \cos(2t)$$

Direct integration leads to:

$$\int \left(-\frac{1}{x^2}\frac{dx}{dt} - \cos(2t)\right) dt = 0 \quad \Rightarrow \quad \int \left(\frac{1}{x} - \frac{\sin(2t)}{2}\right)' = 0.$$

Re-ordering the terms leads to

$$x(t) = \frac{2}{\sin(2t) + 2H}$$

Example 13.14
Find all solutions x to the differential equation

$$\dot{x} = \frac{t^2}{1 - x^2}.$$

Direct integration leads to

$$x - \frac{x^3}{3} = \frac{t^3}{3} + const.$$

Remark 13.17
When the solution is given in terms of an algebraic equation, we say that it is given in implicit form.

Definition 13.7 (Implicit form solution)

A function x is a solution in implicit form of $h(x)\dot{x} = g(t)$ iff function x is a solution of the algebraic equation

$$H(x) = G(t) + const,$$

where H and G are the antiderivatives of h and g respectively. In the case H is invertible, the solution x is given in explicit form iff is written as

$$x(t) = H^{-1}\left(G(t) + const\right).$$

Example 13.15

Find the solution of the initial value problem below in explicit form,

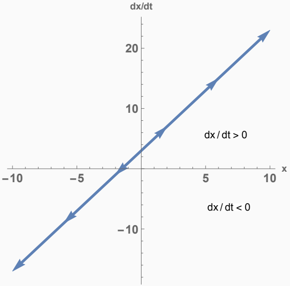
$$\dot{x} = \frac{2-t}{1+x}, \quad x(0) = 1.$$

Definition 13.8 (Phase diagram)
The phase diagram of an autonomous ODE $\dot{x} = f(x)$ is the plot of $\dot{x}$ with respect to x. In this context function $f(\cdot)$ is called phase curve.

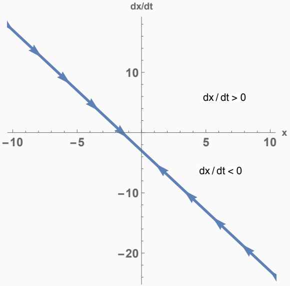
Example 13.16

Compute the phase plot of the ODE $\dot{x} = ax + b$.

$\dot{x} = 2x + 3$



$\dot{x} = -2x - 3$



Remark 13.18
Given $x^* \in \{x : f(x) = 0\}$*, the slope of* $f(x)$ *at* x^* *provides qualitative info. Namely, if* $f'(x^*) < 0$*, then the ODE has a stable equilibrium at* x^**.Otherwise the equilibrium is unstable.*

Example 13.17
Compute the phase diagram of

$$\dot{x} = (x+1)^2 - 16$$

Example 13.18
Compute the phase diagram of

$$\dot{x} = \frac{1}{2}x - x^2$$

Example 13.19 (Solow's growth model)

Consider a macroeconomic model with equations

$$\dot{K} = sQ(K, L) \quad \text{and} \quad \frac{\dot{L}}{L} \;=\; \lambda, \quad \text{for} \quad s, \lambda > 0$$

i.e. capital increases prop to production, and labor increases at an exponential rate.

Given $k = K/L$*, write an ODE for* $\dot{k}$*, considering a production func homogeneous of degree 1, and draw its phase diagram.*

1. Write production in terms of $k = K/L$:

$$\frac{1}{L}Q = \frac{1}{L}f(K,L) = f\left(\frac{K}{L},1\right) = \phi(k), \quad k = K/L.$$

2. Write an ODE for k given $K = kL$:

$$\dot{K} = \dot{k}L + \dot{L}k \quad \Rightarrow \quad \dot{k} = \frac{1}{L}\dot{K} - \frac{\dot{L}}{L}k \quad \Rightarrow \quad \dot{k} = s\phi(k) - \lambda k$$

3. compute the sign of $\phi_k(k)$ and $\phi_{kk}(k)$:

$$\begin{aligned}
\frac{dQ}{dK} &= \frac{d\phi(k)}{dK}L = \phi_k(k)\frac{dk}{dK}L = \phi_k(k) \quad \Rightarrow \phi_k(k) > 0 \\
\frac{d^2Q}{dK^2} &= \frac{d\phi_k(k)}{dK} = \phi_{kk}(k)\frac{dk}{dK} = \phi_{kk}(k)\frac{1}{L} \quad \Rightarrow \phi_{kk}(k) < 0
\end{aligned}$$

Example 13.20
Draw the phase diagram for the Solow model at example 13.19 considering a Cobb-Douglas[1] *production function:* $Q(K, L) = K^\alpha L^{1-\alpha}$.

In this case we obtain the Bernoulli equation:

$$\dot{k} = s\phi(k) - \lambda k \quad \Rightarrow \quad \dot{k} = sk^\alpha - \lambda k,$$

which allows for linearization using $v = 1/k^{\alpha-1}$, leading to:

$$\begin{aligned} v(t) &= He^{-\lambda(1-\alpha)t} + \frac{s}{\lambda} \\ k(t) &= \left(\frac{1}{He^{-\lambda(1-\alpha)t} + \frac{s}{\lambda}}\right)^{\frac{1}{\alpha-1}} \end{aligned}$$

[1]Note that this function is homogeneous of degree α. The qualitative features of the solution in example 13.19 are applicable to this case as well.

Definition 13.9 (Second order linear differential equation)
A second order linear differential equation in the unknown x is

$$\ddot{x} + a_1(t)\dot{x} + a_0(t)x = b(t),$$

where $a_1, a_0 : \mathcal{I} \to \mathbb{R}$ are given functions on the interval $\mathcal{I} \subset \mathbb{R}$

Theorem 13.8 (Initial value problem)

The initial value problem

$$\ddot{x} + a_1(t)\dot{x} + a_0(t)x = b(t), \quad x(t_0) = x_0, \quad \dot{x}(t_0) = x_1,$$

where $a_0, a_1, b : \mathcal{I} \to \mathbb{R}$ and $a_0, a_1, b \in \mathcal{C}$, $t_0 \in \mathcal{I}$, and $x_0, x_1 \in \mathbb{R}$ has a solution and it is unique.

Proof.
Outside the scope of this course. See Picard-Lindelöf Theorem. ∎

Algorithm 4 (Varying Terms Revisited)

Input: $\ddot{x} + a_1\dot{x} + a_0 x = b(t)$, $a_0, a_1 \in \mathbb{R}, b \in \mathcal{C}, a \neq 0$. *E.g.* $b(t) = b$.

1. *Conjecture* $x_c(t) = He^{rt}$, $H \neq 0$, *so that* $\dot{x}_c = Hre^{rt}$ *and* $\ddot{x}_c = Hr^2 e^{rt}$. *Replace in the homogeneous equation*

$$\ddot{x} + a_1\dot{x} + a_0 x = 0 \quad \Rightarrow \quad He^{rt} P(r) = 0,$$

where $P(r) = r^2 + a_1 r + a_0$, *has roots which are i) real and different; ii) real and equal; or iii) complex conjugate.*

2. *Conjecture* $x_p(t)$ *has the same form as* $b(t)$, *i.e.* $x_p(t) = k$, *so that* $\dot{x}_p = \ddot{x} = 0$. *Replace in the inhomogeneous equation*

$$\ddot{x}_p + a_1\dot{x}_p + a_0 x_p = b \quad \Rightarrow \quad a_0 k = b \quad \Rightarrow \quad k = b/a_0,$$

and conclude that $x_p(t) = b/a_0$.

Output: $x(t) = x_c(t) + x_p(t) = He^{at} - b/a$.

Homogeneous Solution

Remark 13.19
We focus on the homogeneous sol, since the non-homogeneous sol can be found by the methods presented for first order ODEs.

Remark 13.20
The solution of the problem $\ddot{x} + a_1\dot{x} + a_0 x = b(t)$, $a_0, a_1 \in \mathbb{R}, b \in \mathcal{C}$ *can be computed using algorithm 4. However in this case one faces three cases for* $x_c(t)$ *depending on the roots of* $P(r) = r^2 + a_1 r + a_0$. *Namely*

$$r = \frac{-a_1 \pm \sqrt{a_1^2 - 4a_0}}{2}.$$

- Case 1: roots are real and different for $a_1^2 - 4a_0 > 0$.

$$x_c(t) = x_{c,1}(t) + x_{c,2}(t) = H_1 e^{r_1 t} + H_2 e^{r_2 t},$$

  which *converges* if $r_1 < 0$ and $r_2 < 0$.
- Case 2: roots real and equal if $a_1^2 - 4a_0 = 0$.

$$x_c(t) = x_{c,1}(t) + x_{c,2}(t) = H_1 e^{rt} + H_2 t e^{rt},$$

  which *converges* if $r < 0$.

- Case 3: roots complex conjugate if $a_1^2 - 4a_0 < 0$. Consider

$$r = p \pm qi = -\frac{a_1}{2} \pm \frac{\sqrt{a_1^2 - 4a_0}}{2}, \quad p = -\frac{a_1}{2}, q = \frac{\sqrt{-a_1^2 + 4a_0}}{2}$$

and denote $r_1 = p + qi$ and $r_2 = p - qi$. Hence

$$\begin{aligned} x_{c,1}(t) &= H_1 e^{(p+qi)t} = H_1 e^{pt} e^{(qt)i} = H_1 e^{pt} \{\cos(qt) + i\sin(qt)\} \\ x_{c,2}(t) &= H_2 e^{(p-qi)t} = H_2 e^{pt} e^{(-qt)i} = H_2 e^{pt} \{\cos(qt) - i\sin(qt)\} \end{aligned}$$

Hence:

$$\begin{aligned} x_c(t) &= x_{c,1}(t) + x_{c,2}(t) \\ &= e^{pt} \{(H_1 + H_2)\cos(qt) + (H_1 - H_2) i \sin(qt)\} \\ &= G_1 e^{pt} \cos(qt) + G_2 e^{pt} \sin(qt), \end{aligned}$$

which has an oscillating pattern, and *converges* for $p < 0$.

Example 13.21
Find the solution $x(t)$ of

$$\ddot{x} + \dot{x} = 4t^2$$

Example 13.22
Write the system

$$\begin{aligned} \dot{x}_1 &= a_{1,1}x_1 + a_{1,2}x_2 + b_1(t) \\ \dot{x}_2 &= a_{2,1}x_1 + a_{2,2}x_2 + b_2(t), \end{aligned}$$

as a second order ODE on $x_1(t)$.

Writing $\dot{x}_2(t)$ in terms of $x_1(t)$ in the second equation and replacing the result in the differential of the first one leads to the result:

$$\ddot{x}_1 - (a_{1,1} + a_{2,2})\dot{x}_1 + (a_{1,1}a_{2,2} - a_{1,2}a_{2,1})x_1 = \dot{b}_1(t) - a_{2,2}b_1(t) + a_{1,2}b_2(t).$$

Definition 13.10 (Arbitrary order linear differential equation)

An arbitrary order linear differential equation in the unknown x is

$$x^{(n)} + a_{n-1}(t)x^{(n-1)} + a_{n-2}(t)x^{(n-2)} + \cdots + a_0(t)x = b(t)$$

where $a_0, \ldots, a_{n-1}, b : \mathcal{I} \to \mathbb{R}$ are continuous functions.

Remark 13.21
We will study the case when functions a_i take constant values.

Remark 13.22
We focus on the homogeneous sol, since the non-homogeneous sol can be found by the methods presented for first order ODEs.

Homogeneous Solution

- Objective: to find n linearly independent functions whose linear combination is the complementary solution $x_c(t)$.
- Linear independence: verified by the Wronksian

$$W_{\boldsymbol{x}} = \det\left(\begin{bmatrix} x_1 & \dot{x}_1 & \ddot{x}_1 & \dots & x_1^{(n-1)} \\ x_2 & \dot{x}_2 & \ddot{x}_2 & \dots & x_2^{(n-1)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ x_n & \dot{x}_n & \ddot{x}_n & \dots & x_n^{(n-1)} \end{bmatrix}\right),$$

  where the dependence on t was omitted for notational ease.
- Characteristic polynomial: has the form

$$P(r) = r^n + a_{n-1}r^{n-1} + a_{n-2}r^{n-2} + \cdots + a_0 = 0.$$

  whose zeros can be found using, e.g. Ruffini's method.

- Three types of roots can be found for the characteristic polynomial:
  1. Real and different
  2. Real and equal
  3. Complex conjugate.
- In general, the complementary solution reads:

$$x_c(t) = \underbrace{\sum_{i=1}^{m_1} H_i e^{r_i t}}_{\text{real diff}} + \underbrace{\sum_{i=1}^{m_2} G_i e^{rt} t^{i-1}}_{\text{real eq}} + \underbrace{\sum_{i=1}^{m_3} e^{p_i t} \left[F_{1,i} \cos(q_i t) + F_{2,i} \sin(q_i t)\right]}_{\text{comp conj}},$$

  where the number of the real diff, real eq and comp conj roots is m_1, m_2 and m_3 resp, and $n = m_1 + m_2 + 2m_3$.

Example 13.23

Find the solution $x(t)$ of

$$x^{(5)} + 4x^{(4)} - 5x^{(3)} - x^{(2)} - 4\dot{x} + 5x = 0$$

Remark 13.23

The convergence of $x(t)$ depends on the convergence of $x_p(t)$ and $x_c(t)$. The convergence of $x_p(t)$ depends on the form of $b(t)$, while the convergence of $x_c(t)$ depends on the sign of the real part of $P(t)$'s roots.

Theorem 13.9 (Routh-Hurwitz)

The real part of the roots in the polynomial

$$P(r) = r^n + a_{n-1} r^{n-1} + a_{n-2} r^{n-2} + \cdots + a_0,$$

is negative iff the n determinants of the following sequence

$$\det(a_{n-1}), \det\left(\begin{bmatrix} a_{n-1} & 1 \\ a_{n-3} & a_{n-2} \end{bmatrix}\right), \det\left(\begin{bmatrix} a_{n-1} & 1 & 0 \\ a_{n-3} & a_{n-2} & a_{n-1} \\ a_{n-5} & a_{n-4} & a_{n-3} \end{bmatrix}\right) \ldots$$

are all positive.

Proof.

Outside the scope of this course. ∎

Example 13.24

Verify that the solution $x(t)$ in example 13.23 does not converge using theorem 13.9.

Definition 13.11 (Linear first order differential systems)

A linear first order differential system is a system of the form

$$\begin{aligned}
\dot{x}_1 &= a_{11}(t)x_1 + a_{12}x_2(t) + \cdots + a_{1n}(t)x_n + b_1(t) \\
\dot{x}_2 &= a_{21}(t)x_1 + a_{12}x_2(t) + \cdots + a_{1n}(t)x_n + b_2(t) \\
&\vdots \\
\dot{x}_n &= a_{n1}(t)x_1 + a_{n2}(t)x_2 + \cdots + a_{nn}(t)x_n + b_n(t),
\end{aligned}$$

or $\dot{\boldsymbol{x}} = \boldsymbol{A}(t)\boldsymbol{x} + \boldsymbol{b}(t)$, *with solutions* $\boldsymbol{x}(t)$ *denoted orbits.*

Remark 13.24

In this section we will assume that

- $\boldsymbol{A}(t) = \boldsymbol{A}$
- $\boldsymbol{A}$ *is* 2×2.
- $\boldsymbol{A}$ *is non-singular*
- $\boldsymbol{A}$ *is diagonalizable*
- $\boldsymbol{b}(t) = \boldsymbol{b}$

Example 13.25

Write the (arbitrary order) linear ODE

$$x^{(n)} + a_{n-1}x^{(n-1)} + a_{n-2}x^{(n-2)} + \cdots + a_0 x = b(t)$$

as a linear first order ODE wrt an n-dimensional vector.

Denote $x_1 = x$, $x_2 = \dot{x}_1$, $\ldots$, $x_n = \dot{x}_{n-1}$ so that

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_{n-1} \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ -a_0 & -a_1 & \ldots & -a_{n-2} & a_{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ b(t) \end{bmatrix}$$

Remark 13.25

This matrix representation is called the companion form of the ODE.

Remark 13.26

From remark 13.24, it is obvious that the system

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \quad \leftrightarrow \quad \dot{\boldsymbol{z}} = \boldsymbol{\Lambda}\boldsymbol{z} + \boldsymbol{w},$$

where $\boldsymbol{z} = \boldsymbol{V}^{-1}\boldsymbol{x}$ *and* $\boldsymbol{w} = \boldsymbol{V}^{-1}\boldsymbol{b}$. *For the new system* $\boldsymbol{z}(t) = \boldsymbol{z}_c(t) + \boldsymbol{z}_p$ *we obtain*

- *Particular solution*

$$\boldsymbol{z}_p = -\boldsymbol{\Lambda}^{-1}\boldsymbol{w}.$$

- *Complementary solution*

$$\boldsymbol{z}_c(t) = \begin{bmatrix} z_{c,1}(t) \\ z_{c,2}(t) \end{bmatrix} = \begin{bmatrix} c_1 e^{r_1 t} \\ c_2 e^{r_2 t} \end{bmatrix} = \begin{bmatrix} e^{r_1 t} & 0 \\ 0 & e^{r_2 t} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

Remark 13.27
Note that given the ansatz for the complementary solution

$$\boldsymbol{x}_c(t) = \begin{bmatrix} x_{c,1}(t) \\ x_{c,2}(t) \end{bmatrix} = \begin{bmatrix} H_1 e^{rt} \\ H_2 e^{rt} \end{bmatrix} = \boldsymbol{v} e^{rt},$$

where $\boldsymbol{v}$ *is a 2d vector. It follows that*

$$\dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} \quad \rightarrow \quad r\boldsymbol{v}e^{rt} = \boldsymbol{A}\boldsymbol{v}e^{rt} \quad \rightarrow \quad \boldsymbol{A}\boldsymbol{v} = r\boldsymbol{v},$$

meaning that r *is an eigenvalue of* $\boldsymbol{A}$ *and* $\boldsymbol{v}$ *its corresponding eigenvector.*

Remark 13.28
Remember that eigenvectors are not defined uniquely. i.e. if $\boldsymbol{v}$ *is an eigenvector of* $\boldsymbol{A}$*, any* $\boldsymbol{v}^* \propto \boldsymbol{v}$ *is also an eigenvector of* $\boldsymbol{A}$*.*

Remark 13.29
Remember that the characteristic polynomial

$$P(r) = r^2 - \mathrm{tr}(\boldsymbol{A})r + \det(\boldsymbol{A}) = 0,$$

has roots given by

$$r = \frac{1}{2}\left[\mathrm{tr}(\boldsymbol{A}) \pm \sqrt{\mathrm{tr}(\boldsymbol{A})^2 - 4\det(\boldsymbol{A})}\right].$$

Like before, the solution depends on the nature of the roots in the characteristic polynomial, e.g. real and different, etc.

Case 1: Eigenvalues are real and different:

Consider $r_1, r_2 \in \mathbb{R}$, and $r_1 \neq r_2$. The solution for $\boldsymbol{z}(t)$ in remark 13.26 is of use for this case since $\boldsymbol{z}_c(t) = \boldsymbol{V}^{-1}\boldsymbol{x}_c(t)$. Hence

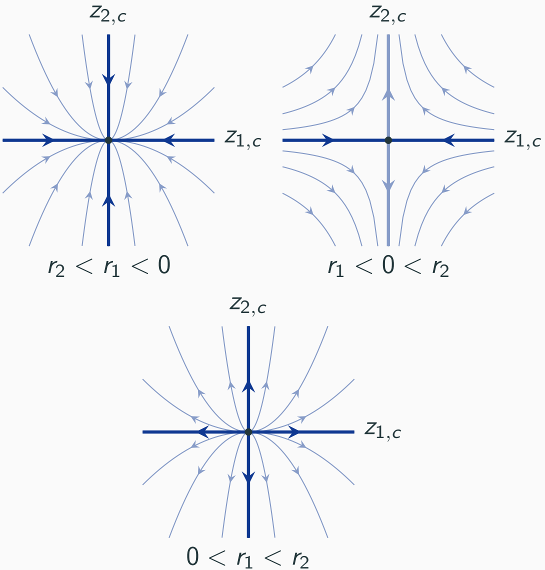
$$
\begin{aligned}
\boldsymbol{x}_c(t) = \begin{bmatrix} x_{c,1}(t) \\ x_{c,2}(t) \end{bmatrix} &= \underbrace{[\boldsymbol{v}_1 \quad \boldsymbol{v}_2]}_{\boldsymbol{V}} \underbrace{\begin{bmatrix} e^{r_1 t} & 0 \\ 0 & e^{r_2 t} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}}_{\boldsymbol{z}_c(t)} \\
&= c_1 e^{r_1 t}\boldsymbol{v}_1 + c_2 e^{r_2 t}\boldsymbol{v}_2
\end{aligned}
$$

For case 1, we face the following cases for $\boldsymbol{x}_c(t)$:

1. Converges: if $r_1 < 0$ and $r_2 < 0$.
   The origin is a stable node
2. Diverges: if $r_1 > 0$ and $r_2 > 0$.
   The origin is an unstable node
3. Diverges: if $r_1 < 0 < r_2$. For $c_1 \neq 0, c_2 = 0$
   The origin is a saddle point.
   The line $x_{c,2} = (v_{2,1}/v_{1,1})x_{c,1}$ is a saddle path.

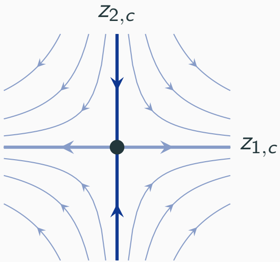
Phase Diagram

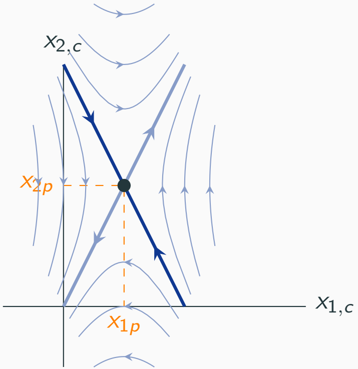
Consider the $\boldsymbol{z}_c(t)$ system:



The $\boldsymbol{x}(t)$ system is a linear transformation of the $\boldsymbol{z}_c(t)$ system:

$$\boldsymbol{x}(t) = \boldsymbol{x}_c(t) + \boldsymbol{x}_p = \underbrace{\boldsymbol{V}\boldsymbol{z}_c(t)}_{\text{rotates/streaches}} + \underbrace{\boldsymbol{x}_p}_{\text{translades}}$$



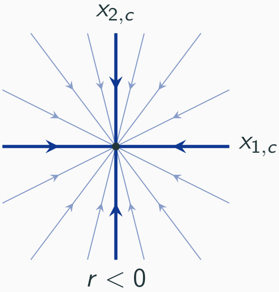


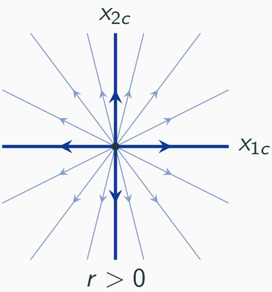
Case 2: Eigenvalues real repeated:

$$\boldsymbol{x}_c(t) = c_1 e^{rt} \boldsymbol{v}_1 + c_2 e^{rt} \boldsymbol{v}_2,$$

where we assumed that geo. mult. of r is 2, so that $\boldsymbol{A}$ is diagonalizable.

1. converges: if $r < 0$. The origin is called proper stable node.
2. diverges: if $r > 0$. The origin is called proper unstable node.





Remark 13.30
The case geo. mult. of r equal to 1 corresponds to a non-diagonalizable matrix $\boldsymbol{A}$. We will not study this case, but will mention that the system produces proper stable and improper unstable nodes for $r < 0$ and $r > 0$ respectively.

Case 3: Eigenvalues complex conjugate:

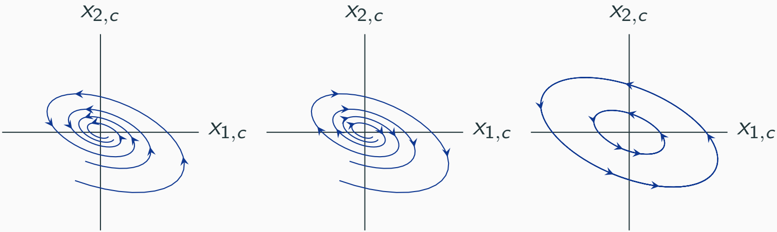
Denote eigenvalues and eigenvectors

$$
\begin{aligned}
r_1 = p + qi \quad \text{with} \quad \boldsymbol{v}_1 &= (\alpha_1 + \alpha_2 i, \beta_1 + \beta_2 i)^\top \\
r_2 = p - qi \quad \text{with} \quad \boldsymbol{v}_2 &= (\alpha_1 - \alpha_2 i, \beta_1 - \beta_2 i)^\top.
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{x}_c(t) &= e^{pt} \begin{bmatrix} \alpha_1 + \alpha_2 i & \alpha_1 - \alpha_2 i \\ \beta_1 + \beta_2 i & \beta_1 - \beta_2 i \end{bmatrix} \begin{bmatrix} e^{(qt)i} & 0 \\ 0 & e^{-(qt)i} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \\
&= e^{pt} \begin{bmatrix} \alpha_1 & -\alpha_2 \\ \beta_1 & -\beta_2 \end{bmatrix} \begin{bmatrix} c_1 & c_2 \\ -c_2 & c_1 \end{bmatrix} \begin{bmatrix} \cos(qt) \\ \sin(qt) \end{bmatrix}
\end{aligned}
$$

The system

1. Converges: if $p < 0$. The origin is called a stable spiral.
2. Diverges: if $p > 0$. The origin is called an unstable spiral.
3. Diverges: if $p = 0$. The origin is called a vortex.



$p < 0$ $\quad$ $p > 0$ $\quad$ $p = 0$

Summary

Remark 13.31
Remember that the trace and determinant are invariants, i.e.

$$\begin{aligned} \operatorname{tr}(\boldsymbol{A}) &= \operatorname{tr}(\boldsymbol{\Lambda}) = r_1 + r_2 \\ \det(\boldsymbol{A}) &= \det(\boldsymbol{\Lambda}) = r_1 \times r_2. \end{aligned}$$

- If $\operatorname{di}(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{A})^2 - 4\det(\boldsymbol{A}) < 0$, the roots are complex and the equilibrium is a spiral or a vortex.
- If $\operatorname{di}(\boldsymbol{A}) > 0$, the roots are real.
  - If $\det(\boldsymbol{A}) > 0$, r_1 and r_2 have the same sign.
    - If $\operatorname{tr}(\boldsymbol{A}) < 0$, the equilibrium is an stable node.
    - If $\operatorname{tr}(\boldsymbol{A}) > 0$, the equilibrium an unstable node.
  - If $\det(\boldsymbol{A}) < 0$, r_1 and r_2 have opposite signs and therefore the equilibrium is a saddle point.

Equilibria in the two variable system

Equilibrium	Roots are...		$\det(\boldsymbol{A})$	$\mathrm{tr}(\boldsymbol{A})$	$\mathrm{di}(\boldsymbol{A})$
Saddle point	Real	$r_1 < 0, r_2 > 0$	< 0		> 0
Stable node	Real	$r_1, r_2 < 0$	> 0	< 0	> 0
Unstable node	Real	$r_1, r_2 > 0$	> 0	> 0	> 0
Vortex	Imaginary	$p = 0$	> 0	$= 0$	< 0
Stable spiral	Comp conj	$p < 0$	> 0	< 0	< 0
Unstable spiral	Comp conj	$p > 0$	> 0	> 0	< 0

Example 13.26

Find the solution $\boldsymbol{x}(t)$ *to the system*

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -a & c \\ c & -a \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} b \\ 0 \end{bmatrix}$$

where $a > 0$, $c > 0$ *and* $a \neq c$.

Definition 13.12 (Autonomous system)

The autonomous system of variables $x_1(t)$ and $x_2(t)$ follows

$$\begin{aligned} \dot{x}_1 &= f(x_1, x_2) \\ \dot{x}_2 &= g(x_1, x_2), \end{aligned}$$

where $f, g : \mathbb{R}^2 \to \mathbb{R}$. Equivalently one can write

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(x_1, x_2)$$

where $\boldsymbol{F} : \mathbb{R}^2 \to \mathbb{R}^2$ is a transformation.

Definition 13.13 (Steady state)
A pair $\boldsymbol{x}_0 \in \mathbb{R}^2$ st $\boldsymbol{F}(\boldsymbol{x}_0) = 0$.

Example 13.27

Find all the steady states of the two-dimensional (decoupled) system

$$\begin{aligned}\dot{x}_1 &= -x_1 + x_1^3 \\ \dot{x}_2 &= -2x_2\end{aligned}$$

Example 13.28

Find all the steady states of the two-dimensional system with feedback

$$\begin{aligned}\dot{x}_1 &= x_1^2 + x_2^2 - 1 \\ \dot{x}_2 &= x_1 x_2\end{aligned}$$

Example 13.29
Compute the first order Taylor expansion around $\boldsymbol{x}_0$ *for the system in definition 13.12*

$$\begin{aligned}\dot{\boldsymbol{x}} &= \boldsymbol{F}(\boldsymbol{x}) \\ &\approx \boldsymbol{F}(\boldsymbol{x}_0) + \boldsymbol{DF}(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) = \boldsymbol{DF}(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0),\end{aligned}$$

where $\boldsymbol{DF}(\boldsymbol{x}_0)$ denotes the Jacobian matrix of $\boldsymbol{F}(\boldsymbol{x}) : \mathbb{R}^2 \to \mathbb{R}^2$ at $\boldsymbol{x}_0$.

Definition 13.14 (Linearization)
The linearization of the <u>*non-linear*</u> *system* $\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x})$ *at a steady state* $\boldsymbol{x}_0$ *is the* <u>*linear system*</u>

$$\dot{\boldsymbol{u}} = \boldsymbol{DF}(\boldsymbol{x}_0)\boldsymbol{u},$$

where $\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{x}_0$.

Definition 13.15 (Hyperbolic steady state)
A steady state $\boldsymbol{x}_0$ *is called hyperbolic if the real part of all eigenvalues of* $\boldsymbol{DF}(\boldsymbol{x}_0)$ *is nonzero.*

Example 13.30
Find the linearized system at every steady state in example 13.27. Are the steady states hyperbolic?

Theorem 13.10 (Grobman-Hartman)

Let $\dot{\boldsymbol{u}} = \boldsymbol{DF}(\boldsymbol{x}_0)\boldsymbol{u}$ denote the linearization of $\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x})$ at hyperbolic steady state $\boldsymbol{x}_0$. Then, there is a neighborhood of $\boldsymbol{x}_0$ where all the solutions of the linear system can be transformed into solutions of the nonlinear system by a continuous, invertible, transformation.

Proof.
Outside the scope of the course. ■

Remark 13.32
Theorem 13.10 means that the orbits near a hyperbolic $\boldsymbol{x}_0$ in the non-linear system are "similar" to the orbits in the linearized system, so we can use the latter to study the qualitative behaviour of the former.

Algorithm 5 (Phase diagram)
Input: linear/non-linear system.

1. *Draw locus* $f(x, y) = 0$ *and* $g(x, y) = 0$ *in the* (x, y) *plane.*
2. *Draw steady states* $\boldsymbol{x}_0 = (x_0, y_0)$ *at the crossings of the loci.*
3. *Draw direction field by evaluating the dynamics of the system around* $\boldsymbol{x}_0$*, e.g. compute* $\partial\dot{x}/\partial y$ *and* $\partial\dot{y}/\partial x$.

Output: phase diagram

Example 13.31
Draw the phase diagram of the system

$$\begin{aligned}\dot{x} &= -ax + cy + b \\ \dot{y} &= cx - ay,\end{aligned}$$

where $a > 0, c > 0, a \neq c$.

1. The loci of the system reads

$$\begin{aligned}\dot{x} &= 0 \quad \rightarrow \quad y = \frac{a}{c}x - \frac{b}{c} \\ \dot{y} &= 0 \quad \rightarrow \quad y = \frac{c}{a}x.\end{aligned}$$

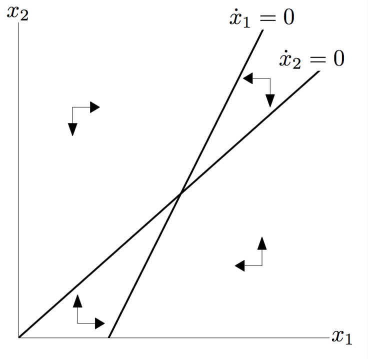
2. The steady states correspond to the solution of

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = -\begin{bmatrix} -a & c \\ c & -a \end{bmatrix}^{-1} \begin{bmatrix} b \\ 0 \end{bmatrix} = \frac{b}{a^2 - c^2} \begin{bmatrix} a \\ c \end{bmatrix}$$

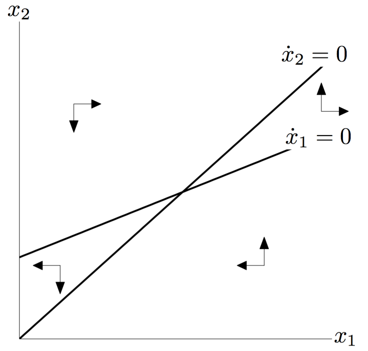
3. From the system we know that

$$\begin{aligned} \dot{x} &= -ax + cy + b \quad \rightarrow \quad \frac{\partial \dot{x}}{\partial x} < 0, \frac{\partial \dot{x}}{\partial y} > 0 \\ \dot{y} &= cx - ay \quad \rightarrow \quad \frac{\partial \dot{y}}{\partial x} > 0, \frac{\partial \dot{y}}{\partial y} < 0 \end{aligned}$$

$c < a$

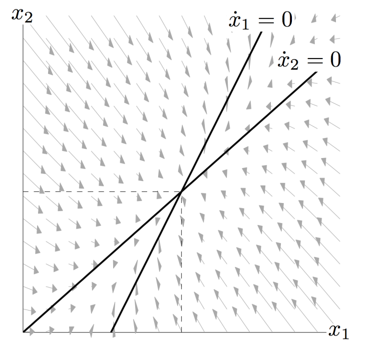


$c > a$

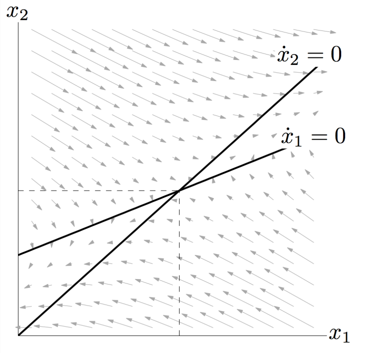


Direction fields:

$c < a$



$c > a$



Example 13.32 (Assessment Question (2015-2))
Consider the model

$$\begin{aligned} \frac{\dot{x}}{x} &= 1 - \frac{\alpha x + y}{(1+\alpha)K} \\ \frac{\dot{y}}{y} &= r\left(1 - \frac{\alpha y + x}{(1+\alpha)K}\right) \end{aligned}$$

where $x > 0$,$y > 0$,$K > 0$, $r > 0$ *and* $\alpha > 1$.

1. Find the steady states of the system.

The loci of the system reads:

$$\begin{aligned}\dot{x} &= 0 \quad \rightarrow \quad y = -\alpha x + (1+\alpha)K \\ \dot{y} &= 0 \quad \rightarrow \quad y = -\frac{1}{\alpha}x + \left(\frac{1+\alpha}{\alpha}\right)K.\end{aligned}$$

The solution is then $(x_0, y_0)^\top = (K, K)^\top$.

2. Draw the phase diagram and determine if the equilibrium is stable.

$$
\begin{aligned}
\frac{\partial \dot{x}}{\partial y} &= -\frac{x}{(1+\alpha)K} < 0 \\
\frac{\partial \dot{y}}{\partial x} &= -\frac{ry}{(1+\alpha)K} < 0,
\end{aligned}
$$

from the results is clear that the steady state is spiralling (outwards or inwards)

3 Linearize the system and verify if the equilibrium is stable.

The Jacobian reads:

$$\boldsymbol{DF}(\boldsymbol{x}) = \frac{1}{(1+\alpha)K}\begin{bmatrix}(1+\alpha)K - 2\alpha x - y & -x \\ -ry & (1+\alpha)rK - 2\alpha ry - xr\end{bmatrix}$$

Let $\boldsymbol{DF}(\boldsymbol{x}_0) = \boldsymbol{A}$, then

$$\boldsymbol{A} \;=\; -\frac{1}{1+\alpha}\begin{bmatrix}\alpha & 1 \\ r & \alpha r\end{bmatrix} \quad \rightarrow \quad \mathrm{di}(\boldsymbol{A}) = \frac{(\alpha r - \alpha)^2 + 4r}{(1+\alpha)^2} > 0.$$

And because

$$\mathrm{tr}(\boldsymbol{A}) = -\frac{\alpha(r+1)}{1+\alpha}, \quad \text{and} \quad \det(\boldsymbol{A}) = \frac{r(\alpha-1)}{\alpha+1},$$

the steady state is a stable node.

Difference Equations

Definitions

Definition 14.1 (k-th Difference operator)

The k-th difference of x_t is defined recursively as

$$\Delta^k(x_t) = \Delta(\Delta^{k-1}(x_t)),$$

for $\Delta x_t = x_t - x_{t-1}$ and $\Delta^0 = 1$.

Definition 14.2 (Difference equation)

The n-th order difference equation (DE) is an equation of the form

$$f(\Delta^n x_t, \Delta^{n-1} x_t, \ldots, x_t) = b_t,$$

whose unknown is the sequence x_t itself. The order of the DE is the highest difference operator order in the equation.

Example 14.1

- $\Delta x_t = t^2 + 1 \rightarrow$ *order one*
- $\Delta x_t - q x_{t-1} = b \rightarrow$ *order one*
- $\Delta^2 x_t - a_1 \Delta x_t - a_0 x_t = b_t \rightarrow$ *order two*

Definition 14.3 (Difference equation)

The n-th order difference equation (DE) is an equation that relates a sequence x_t and its lags $x_{t-1}, \ldots, x_{t-n}$, i.e $g(x_t, x_{t-1}, x_{t-2}, x_{t-n}) = b_t$.

Theorem 14.1
Definitions 14.2 and 14.3 are equivalent.

Proof.
Trivial. Note that $\Delta^k x_t$ in definition 14.2 is a function of $x_t, \ldots, x_{t-k}$, hence the LHS reads:

$$
\begin{aligned}
& f(\Delta^n x_t, \Delta^{n-1} x_t, \ldots, x_t) = \\
& f(f_n(x_{t-1}, x_{t-2}, \ldots, x_{t-n}), f_{n-1}(x_{t-1}, x_{t-2}, \ldots, x_{t-n+1}), \cdots + f_0(x_t)) = \\
& g(x_t, x_{t-1}, x_{t-2}, \ldots, x_{t-n}) = b_t,
\end{aligned}
$$

which leads to definition 14.3. ∎

Example 14.2
Re-write the DEs in example 14.1 according to definition 14.3.

Remark 14.1
Hereafter we will use definition 14.3.

Definition 14.4 (Characterization of DEs)

1. *Order: is the largest difference between the subindices of x in the equation.*
2. *Autonomous: if t does not appear explicitly in the equation.*
3. *Linear: if $x_t, x_{t-1}, \dots, x_{t-n}$ appear in linear form.*
4. *Homogeneous: if there are no zero-order terms.*
5. *Stable: if $\lim_{t\to\infty} x_t = x_{ss}$, and x_{ss} is a finite constant.*

First order Linear DEs

Definition 14.5 (First order linear DE)
The linear first order DE is defined as

$$x_t - a x_{t-1} = w_t, \quad a \in \mathbb{R},$$

where w_t is a sequence depending on time.

Theorem 14.2

Given the known value x_τ, the linear first order DE has a solution

$$\begin{aligned} x_{\tau+s} &= a^s x_\tau + \sum_{j=1}^{s} a^{s-j} w_{\tau+j}, \quad t > \tau \\ x_{\tau-s} &= a^{-s} x_\tau - \sum_{j=0}^{s-1} a^{j-s} w_{\tau-j}, \quad t < \tau \end{aligned}$$

where $s \geq 1$.

Proof.
Solving backwards ($t > \tau$), we obtain

$$\begin{aligned}
x_{\tau+1} &= a x_\tau + w_{\tau+1} \\
x_{\tau+2} &= a x_{\tau+1} + w_{\tau+2} = a^2 x_\tau + a w_{\tau+1} + w_{\tau+2} \\
&\vdots \\
x_{\tau+s} &= a^s x_\tau + \sum_{j=1}^{s} a^{s-j} w_{\tau+j},
\end{aligned}$$

while solving forward ($t < \tau$), we obtain

$$\begin{aligned}
x_{\tau-1} &= a^{-1}(x_\tau - w_\tau) \\
x_{\tau-2} &= a^{-1}(x_{\tau-1} - w_{\tau-1}) = a^{-2} x_\tau - a^{-2} w_\tau - a^{-1} w_{\tau-1} \\
&\vdots \\
x_{\tau-s} &= a^{-s} x_\tau - \sum_{j=0}^{s-1} a^{j-s} w_{\tau-j}.
\end{aligned}$$

■

Example 14.3

Compute the solution of

$$x_t - ax_{t-1} = w, \quad w \in \mathbb{R}.$$

Hint: Remember that

$$\sum_{j=0}^{n} ar^j = a\left(\frac{r^{n+1}-1}{r-1}\right).$$

Replacing in the previous solutions

$$x_t = \underbrace{\left(x_\tau - \frac{w}{1-a}\right) a^{-\tau}}_{H} a^t + \frac{w}{1-a} = Ha^t + \frac{w}{1-a}$$

Remark 14.2
It is illustrative to compare the previous solution with the solution of the first order linear ODE with constant coefficients

$$\begin{aligned} \dot{x} &= ax + b \quad \rightarrow \quad He^{at} - \frac{b}{a} \\ x &= ax_{t-1} + w \quad \rightarrow \quad Ha^t + \frac{w}{1-a}. \end{aligned}$$

- *Particular solution:* for ODEs we assign $\dot{x} = 0$, while for DEs we assign $x_t = x_{t-1}$.
- *Complementary solution:* for ODEs and DEs the root is "a" and it provides information about the convergence of the equation. It can be obtain from the ansatz of the form $x_{ct} = Hr^t$.

Remark 14.3
Note that the magnitude of a in example 14.3 provides a convergence criteria. Namely, if $|a| < 1$,

$$\lim_{t\to\infty} x_t = \frac{w}{1-a}.$$

In addition, note that if $|a| > 1$, the former limit might be infinity ($a > 0$), or might not even exist ($a < 0$).

Example 14.4 (Cobweb model)
Consider the supply and demand model

$$D_t = \alpha - \beta P_t \quad y \quad S_t = -\gamma + \delta P_t^e,$$

where $\alpha, \beta, \gamma, \delta > 0$. *Find* P_t *implied by the equilibrium condition* $D_t = S_t$. *Describe the behaviour of this trayectory and indicate the conditions for it to be stable if*

1. *expectations are static, i.e.* $P_t^e = P_{t-1}$,
2. *expectations adapt, i.e.* $P_t^e = (1-\eta)P_{t-1}^e + \eta P_{t-1}$, *where* $\eta \in]0,1]$.

Example 14.5

Compute the solution of

$$x_t - a x_{t-1} = w^t, \quad a \in \mathbb{R}.$$

Replacing in the backward (or forward) sol in theorem 14.2:

$$x_t = \underbrace{\left(x_\tau - \frac{w^{\tau+1}}{w-a} \right) a^{-\tau}}_{H} a^t + \left(\frac{w}{w-a} \right) w^t = H a^t + \left(\frac{w}{1-a} \right) w^t$$

Remark 14.4
The backward and forward solutions may <u>*not*</u> *always deliver the same results. For details see section 6.2.4 in Winkelried's lecture notes.*

Definition 14.6 (Arbitrary order DE)

The arbitrary order DE is defined as

$$x_t + a_{n-1}x_{t-1} + \cdots + a_1 x_{t-n+1} + a_0 x_{t-n} = w_t, \quad a_i \in \mathbb{R},$$

where w_t is a sequence depending on time.

Remark 14.5
Note that definition 14.6 resembles the definition of an arbitrary order mixed coefficient ODE, whose solution was based on two ansatz.

Homogeneous solution:

Analog to the ODE case, the ansatz $x_{c,t} = Hr^t$, leads to

$$\begin{aligned} 0 &= Hr^t + a_{n-1}r^{t-1} + \cdots + a_1 r^{t-n+1} + a_0 r^{t-n} \\ &= Hr^{t-n}\underbrace{\left(r^n + a_{n-1}r^{n-1} + \cdots + a_1 r + a_0\right)}_{P(r)} = 0, \end{aligned}$$

with roots of $P(r)$ providing info about the convergence of the sol.

Remark 14.6
Note that one can use the Wronskian to verify if the n complementary solutions $x_{c,t,i}, i = 1, \ldots, n$ are linearly independent.

Theorem 14.3

In general, the homogeneous solution of the arbitrary order DE in definition 14.6 can be written as

$$x_t = \sum_{i=1}^{m_1} H_i r_i^t + \sum_{i=1}^{m_2} + G_i t^{i-1} r^t + \sum_{i=1}^{m_3} \rho_i^t \left(F_{i,1} \cos(\theta_i t) + F_{i,2} \sin(\theta_i t)\right),$$

where the n roots of the characteristic polynomial contain roots that are: real and different (m_1), real and equal (m_2) and complex conjugate (m_3), so that $n = m_1 + m_2 + 2m_3$.

Proof.
Trivial. See the case for higher order ODEs. ∎

Remark 14.7
The convergence of $x_{c,t}$ depends on the condition $|Re(r_i)| < 1$, for $i = 1, \ldots, n$ in theorem 14.3. A practical way to check for convergence is given in the next theorem.

Theorem 14.4 (Schur)
The condition $|Re(r_i)| < 1$, for $i = 1, \ldots, n$ holds iff the sequence

$$\det\left(\begin{bmatrix} a_n & a_0 \\ a_0 & a_n \end{bmatrix}\right), \det\left(\begin{bmatrix} a_n & 0 & a_0 & a_1 \\ a_{n-1} & a_n & 0 & a_0 \\ a_0 & 0 & a_n & a_{n-1} \\ a_1 & a_0 & 0 & a_n \end{bmatrix}\right), \ldots$$

contains only positive values.

Proof.
Outside theste scope of this course. ■

Example 14.6

Find the solution complementary solution $x_{c,t}$ of

1. $x_t = (x_{t+1} + x_{t-1})/2.$
2. $x_{t+2} + 4x_t = 0, \quad x_0 = 1, x_1 = 0.$

with corresponding solutions

1. $x_t = H_1 + H_2 t$
2. $x_t = 2^t \cos(\pi t/2)$

Inhomogeneous solution:

Analog to the ODE case, the particular solution $x_{p,t}$ can be obtained by the method of undetermined coefficients. Particular cases follow.

w_t	$x_{p,t}$
w const	k const
n-degree poly	n-degree poly
uw^t	kw^t
$u_1 \cos(st) + u_2 \sin(st)$	$k_1 \cos(st) + k_2 \sin(st)$
$u_1 w^t \cos(st) + u_2 w^t \sin(st)$	$k_1 w^t \cos(st) + k_2 w^t \sin(st)$

Example 14.7

Find the particular solution $x_{p,t}$ of

1. $x_{t+2} - bx_{t+1} + dx_t = ac^t$.
2. $x_{t+2} + x_{t+1} + x_t = 3t^2$.

with corresponding solutions

1. $x_{p,t} = ac^t/(c^2 - bc + k)$
2. $x_{p,t} = 1/3 - 2t + t^2$

Example 14.8 (Final exam (2015-2))

Consider a market $D_t = \alpha - \beta P_t$ and $S_t = -\gamma + \delta P_t^e$, $\alpha, \beta, \gamma, \delta > 0$. Moreover consider that the producers accumulate inventories by $I_t = \phi(P_{t+1}^e - P_t)$, $\phi > 0$ st the equilibrium condition reads $S_t = D_t + (I_t - I_{t-1})$. Let $P_t^e = P_t$.

1. *Compute the steady states for $\bar{P}$, $\bar{I}$ and $\bar{D}$.*
2. *Show that the market price trajectory is given by*

$$P(t) = C_1 \lambda^t + C_2 \lambda^{-t} + \bar{P},$$

   where $\bar{P}$ is a steady state value, C_1 and C_2 are arbitrary, and $\lambda > 1$.

3. *Let $P(0) = P_0$ and $I(0) = I_0$. Compute the trajectory of $I(t)$ and indicate the condition so that the system is stable.*

Nonlinear DEs

Definition 14.7 (Phase diagram)
The phase diagram of an autonomous DE is the graph of x_t as a function of x_{t-1} given by

$$x_t = f(x_{t-1}).$$

Moreover the steady state of the DE is a value x_{ss} such that $x_{ss} = f(x_{ss})$.

Definition 14.8 (Linearization of a nonlinear DE)
The first order Taylor expansion of the DE at x_{ss} is given by

$$x_t = f(x_{t-1}) \approx f(x_{ss}) + f'(x_{ss})(x_{t-1} - x_{ss}) = f'(x_{ss})x_{t-1} + (1 - f'(x_{ss}))x_{ss},$$

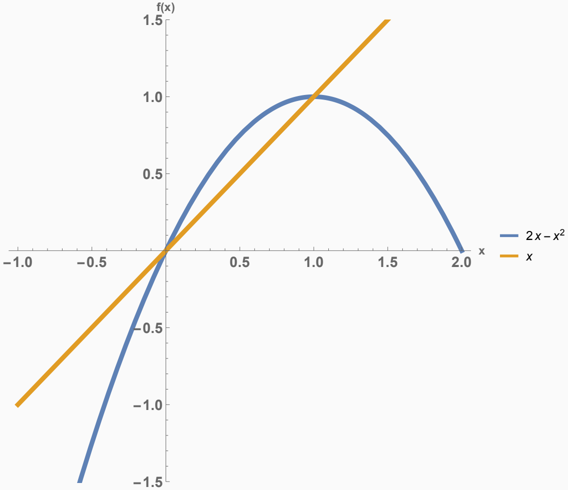
whose stability depends on $f'(x_{ss})$.

Example 14.9
Study the following systems $x_t = f(x_{t-1})$

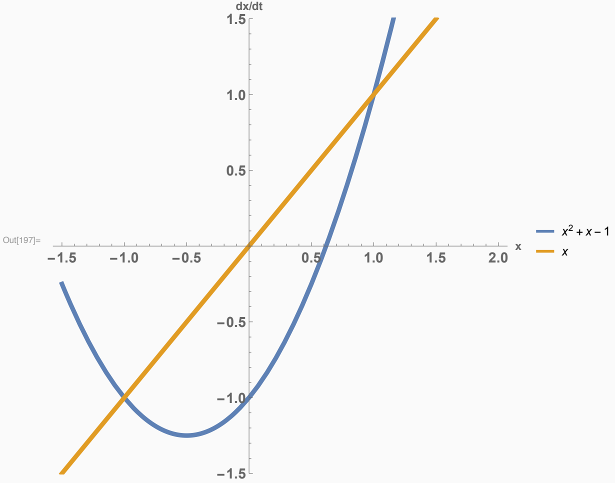
1. $f(x) = 2x - x^2.$
2. $f(x) = x^2 + x - 1.$
3. $f(x) = \sqrt[3]{x}.$

$$f(x) = 2x - x^2$$

Out[196]=

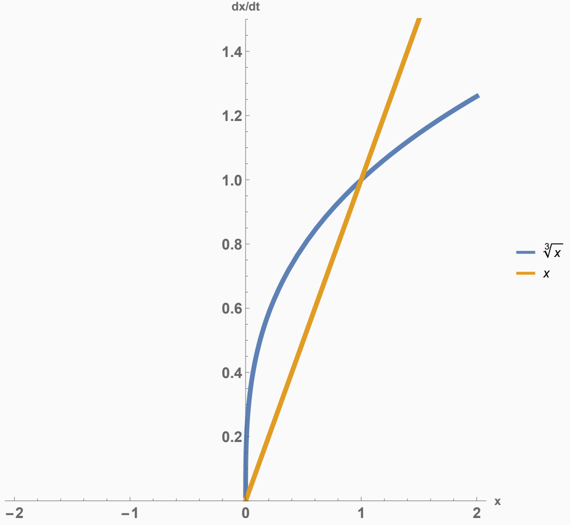


$$f(x) = x^2 + x - 1$$



$$f(x) = \sqrt[3]{x}$$

Out[198]=



Definition 14.9 (Linear first order differential systems)

A linear first order differential system is a system of the form

$$\begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} x_{1,t-1} \\ x_{2,t-1} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

or, in matrix form

$$\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{x}_{t-1} + \boldsymbol{w} \quad \rightarrow \quad \boldsymbol{z}_t = \boldsymbol{\Lambda}\boldsymbol{z}_t + \boldsymbol{\omega},$$

provided that $\boldsymbol{A}$ *is diagonalizable, and* $\boldsymbol{z}_t = \boldsymbol{V}^{-1}\boldsymbol{x}_t$ *and* $\boldsymbol{\omega}_t = \boldsymbol{V}^{-1}\boldsymbol{w}$.

Remark 14.8
The system $\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{x}_{t-1} + \boldsymbol{w}$ *where* $\boldsymbol{x}_t, \boldsymbol{w}_t$ *are 2d vectors,* $\boldsymbol{A} = [a_{ij}]$ *with* $\det(\boldsymbol{I} - \boldsymbol{A}) \neq 0$ *has a particular solution* $\boldsymbol{x}_{p,t} = (\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{w}$

Homogeneous Solution

Remark 14.9
Note that given the ansatz for the complementary solution

$$\boldsymbol{x}_{c,t} = \begin{bmatrix} x_{c,1,t} \\ x_{c,2,t} \end{bmatrix} = \begin{bmatrix} H_1 r^t \\ H_2 r^t \end{bmatrix} = \boldsymbol{v} r^t,$$

where $\boldsymbol{v}$ *is a 2d vector. It follows that*

$$\boldsymbol{x}_{c,t} = \boldsymbol{A}\boldsymbol{x}_{c,t-1} \quad \to \quad \boldsymbol{v} r^t = \boldsymbol{A}\boldsymbol{v} r^t r^{-1} \quad \to \quad \boldsymbol{A}\boldsymbol{v} = r\boldsymbol{v},$$

meaning that r *is an eigenvalue of* $\boldsymbol{A}$ *and* $\boldsymbol{v}$ *its corresponding eigenvector.*

Remark 14.10
Remember that eigenvectors are not defined uniquely. i.e. if $\boldsymbol{v}$ *is an eigenvector of* $\boldsymbol{A}$*, any* $\boldsymbol{v}^* \propto \boldsymbol{v}$ *is also an eigenvector of* $\boldsymbol{A}$*.*

Remark 14.11
Remember that the characteristic polynomial

$$P(r) = r^2 - \mathrm{tr}(\boldsymbol{A})r + \det(\boldsymbol{A}) = 0,$$

has roots given by

$$r = \frac{1}{2}\left[\mathrm{tr}(\boldsymbol{A}) \pm \sqrt{\mathrm{tr}(\boldsymbol{A})^2 - 4\det(\boldsymbol{A})}\right].$$

Like before, the solution depends on the nature of the roots in the characteristic polynomial, e.g. real and different, etc.

Case 1: Eigenvalues are real and different:

$$
\begin{aligned}
\boldsymbol{x}_{c,t} = \begin{bmatrix} x_{c,1,t} \\ x_{c,2,t} \end{bmatrix} &= \underbrace{[\boldsymbol{v}_1 \quad \boldsymbol{v}_2]}_{\boldsymbol{V}} \underbrace{\begin{bmatrix} {r_1}^t & 0 \\ 0 & {r_2}^t \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}}_{\boldsymbol{z}_{c,t}} \\
&= c_1 {r_1}^t \boldsymbol{v}_1 + c_2 {r_2}^t \boldsymbol{v}_2
\end{aligned}
$$

Case 2: Eigenvalues real repeated:

$$
\boldsymbol{x}_{c,t} = c_1 r^t \boldsymbol{v}_1 + c_2 r^t \boldsymbol{v}_2,
$$

where we assumed that geo. mult. of r is 2, so that $\boldsymbol{A}$ is diagonalizable.

Case 3: Eigenvalues complex conjugate:

$$\begin{aligned} r_1 = p + qi = \rho e^{\theta i} \quad \text{with} \quad \boldsymbol{v}_1 &= (\alpha_1 + \alpha_2 i, \beta_1 + \beta_2 i)^\top \\ r_2 = p - qi = \rho e^{-\theta i} \quad \text{with} \quad \boldsymbol{v}_2 &= (\alpha_1 - \alpha_2 i, \beta_1 - \beta_2 i)^\top. \end{aligned}$$

$$\begin{aligned} \boldsymbol{x}_{c,t} &= \rho^t \begin{bmatrix} \alpha_1 + \alpha_2 i & \alpha_1 - \alpha_2 i \\ \beta_1 + \beta_2 i & \beta_1 - \beta_2 i \end{bmatrix} \begin{bmatrix} e^{(\theta t) i} & 0 \\ 0 & e^{-(\theta t) i} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \\ &= \rho^t \begin{bmatrix} \alpha_1 & -\alpha_2 \\ \beta_1 & -\beta_2 \end{bmatrix} \begin{bmatrix} c_1 & c_2 \\ -c_2 & c_1 \end{bmatrix} \begin{bmatrix} \cos(\theta t) \\ \sin(\theta t) \end{bmatrix} \end{aligned}$$

Equilibria in the two variable system

Equilibrium	Roots are...	
Saddle point	Real	$0 < \lvert r_1 \rvert < 1, \lvert r_2 \rvert > 1$
Stable node	Real	$0 < \lvert r_1 \rvert, \lvert r_2 \rvert < 1$
Unstable node	Real	$\lvert r_1 \rvert, \lvert r_2 \rvert > 1$
Vortex	Imaginary	$\rho = 1$
Stable spiral	Comp conj	$\rho < 1$
Unstable spiral	Comp conj	$\rho > 1$

Example 14.10 (Final exam (2016-2))
Consider the system

$$\begin{aligned} x_t &= c(x_{t-1} + y_{t-1}) \\ y_t &= c(y_{t-1} + z_{t-1}) \\ z_t &= c(z_{t-1} + x_{t-1}) \end{aligned}$$

where c is a constant.

1. *For which values of c is the system stable? Hint: remember that $a^3 + b^3 = (a+b)(a^2 - ab + b^2)$*
2. *Suppose $c = 1/2$. Describe the path of x_t.*

Definition 14.10 (Lag operator)
The lag operator is defined as the linear operator

$$L\,x_t = x_{t-1}.$$

Remark 14.12
Note that the composition of the operator reads

$$L^k\,x_t = x_{t-k},$$

where $k \in \mathbb{Z}$. That is, e.g. $L^0\,x_t = x_t$ or $L^{-k}\,x_t = x_{t+k}$, where the latter is called lead operator.

Remark 14.13
L is closed under addition and scalar multiplication.

$$L\ (x_t + y_t) = L\ x_t + L\ y_t$$
$$L\ (\alpha x_t) = \alpha L\ x_t.$$

Remark 14.14
Note the relation between the lag operator and the difference operator:

$$\begin{aligned}
\Delta x_t &= x_t - x_{t-1} = (1-L)x_t \\
\Delta^2 x_t &= (1-L)x_t - (1-L)x_{t-1} = (1-L)^2 x_t \\
&\vdots \\
\Delta^k x_t &= (1-L)^k x_t
\end{aligned}$$

Operations:

1. If $f(x) = \sum_{k=0}^{\infty} a_k x^k$, then $f(L) = \sum_{k=0}^{\infty} a_k L^k$.
2. If $f(L) = \sum_{k=0}^{\infty} a_k L^k$ and $g(L) = \sum_{k=0}^{\infty} b_k L^k$,

   $$(\alpha f + \beta g)(L) = \sum_{k=0}^{\infty} (\alpha a_k + \beta b_k) L^k \quad \text{and} \quad (f \cdot g)(L) = \sum_{k=0}^{\infty} c_k L^k,$$

   where $c_k = \sum_{j=0}^{k} a_j b_{k-j}$.
3. If $g(L)$ is st $g(L) \cdot f(L) = 1$, we call $g(L)$ the "inverse" of $f(L)$.

Definition 14.11 (Linear difference equation)

A linear difference equation can be written as

$$f(L)x_t = w_t,$$

where w_t is a sequence of time.

Remark 14.15

Note that since f is invertible, definition 14.11 is equivalent to $x_t = f(L)^{-1}w_t$, which in fact delivers the particular solution of the DE.

Definition 14.12 (Linear first order DE)

The linear first order DE is $f(L)x_t = w_t$, with $f(L) = (1 - aL)$.

Theorem 14.5

The particular solution of the DE in definition 14.12 is given by

$$x_{p,t} = \begin{cases} \sum_{k=0}^{\infty} a^k w_{t-k}, & \text{if} \quad |a| < 1 \\ -\sum_{k=1}^{\infty} a^{-k} w_{t+k}, & \text{if} \quad |a| > 1. \end{cases}$$

Proof.
For $|a| < 1$, remark 14.15 leads to

$$x_{p,t} = \frac{1}{1 - aL} w_t = \left(\sum_{k=0}^{\infty} a^k L^k \right) w_t = \sum_{k=0}^{\infty} a^k w_{t-k}.$$

If, on the other hand, $|a| > 1$, the previous approx. does not hold.
Consider $\alpha = 1/a$

$$\begin{aligned} x_{p,t} &= \frac{1}{1 - aL} w_t = -\frac{\alpha L^{-1}}{1 - \alpha L^{-1}} w_t = -\alpha L^{-1} (1 + \alpha L^{-1} + \alpha^2 L^{-2} + \ldots) w_t \\ &= -\left(\sum_{k=1}^{\infty} \alpha^k L^{-k} \right) w_t = -\sum_{k=1}^{\infty} a^{-k} w_{t+k}. \end{aligned}$$

■

Definition 14.13 (Linear second order DE)
The linear second order DE can be written as $f(L)x_t = w_t$, *for* $f(L) = 1 - a_1 L - a_2 L^2$.

Remark 14.16
Note that one can write $f(L) = (1 - r_1 L)(1 - r_2 L)$ *st.* $r_1 + r_2 = a_1$ *and* $r_1 r_2 = -a_2$, *which can be either real or complex. Hence we can re-write de DE as*

$$x_t = f(L)^{-1} w_t = \frac{1}{(1 - r_1 L)(1 - r_2 L)} w_t,$$

which has different forms depending on the nature of the roots.

Case 1: roots are real and different:

Given $x_{p,t} = f(L)^{-1} w_t$, we obtain

$$\frac{1}{f(L)} = \frac{1}{(1 - r_1 L)(1 - r_2 L)} = \frac{1}{r_1 - r_2} \left(\frac{r_1}{1 - r_1 L} - \frac{r_2}{1 - r_2 L} \right)$$

Hence:

$$x_{p,t} = \begin{cases} \frac{1}{r_1 - r_2} \sum_{j=0}^{\infty} \left(r_1^{j+1} - r_2^{j+1} \right) w_{t-j}, & \text{if} \quad |r_1| < 1, |r_2| < 1 \\ \frac{1}{r_1 - r_2} \sum_{j=1}^{\infty} \left((1/r_2)^{j-1} - (1/r_1)^{j-1} \right) w_{t+j}, & \text{if} \quad |r_1| > 1, |r_2| > 1 \\ \frac{r_1}{r_1 - r_2} \sum_{j=0}^{\infty} r_1^j w_{t-j} + \frac{r_2}{r_1 - r_2} \sum_{j=1}^{\infty} (1/r_2)^j w_{t+j}, & \text{if} \quad |r_1| < 1, |r_2| > 1 \end{cases}$$

Case 2: roots are complex conjugate:

Given $x_{p,t} = f(L)^{-1} w_t$, remember that given s,

$$\begin{aligned} r_1^s &= \rho^s[\cos(\theta s) + i \sin(\theta s)] \\ r_2^s &= \rho^s[\cos(\theta s) - i \sin(\theta s)] \end{aligned}$$

Which implies that:

$$\begin{aligned} r_1 - r_2 &= 2i\rho \sin(\theta) \\ r_1^{j+1} - r_2^{j+1} &= 2i\rho^{j+1} \sin(\theta(j+1)) \\ r_2^{1-j} - r_1^{1-j} &= -2i\rho^{1-j} \sin(\theta(1-j)) \end{aligned}$$

$$x_{p,t} = \begin{cases} \sum_{j=0}^{\infty} \left(\rho^j \frac{\sin(\theta(j+1))}{\sin(\theta)} \right) w_{t-j}, & \text{if} \quad \rho < 1 \\ -\sum_{j=1}^{\infty} \left(\rho^{-j} \frac{\sin(\theta(1-j))}{\sin(\theta)} \right) w_{t+j}, & \text{if} \quad \rho > 1 \end{cases}$$