Contents

Conservation Laws

Conservative PDEs are usually derived from constitutive (physical) laws that *conserve* certain quantities $\mathbf{u}$ e.g. mass, momentum, density, heat, energy, population, particles, cars,. . .
PDEs in conservative form are so called because conservation laws can always be written in conservative form.

Definition 1.1 Scalar Conservation Law:

$$\frac{\partial}{\partial t}u + \operatorname{div}_{\mathbf{x}} f\left(u(\mathbf{x},t),\mathbf{x}\right) = s\left(u(\mathbf{x},t),\mathbf{x},t\right) \qquad \text{in}\tilde{\Omega} := \boldsymbol{\Omega}\times]0,T[$$

f: flux of conserved quantity u
s: production/source term (1.1)

Definition 1.2 1D Conservation Law:

$$u_t + \frac{\partial}{\partial x} f\left(u(x,t),x\right) = s\left(u(x,t),x,t\right) \qquad \text{in } \tilde{\Omega} := \Omega\times]0,T[\tag{1.2}$$

Definition 1.3 1D inviscide Conservation Law:

$$\begin{aligned} u_t + f\left(u(x,t),x\right)_x &= 0 \qquad \text{in } \tilde{\Omega} := \Omega\times]0,T[\\ u(0,x) &= u_0(x) \end{aligned} \tag{1.3}$$

1. Examples

1.1. Transport Equation

Definition 1.4 Transport Equation $f = au$:

$$\begin{aligned} u_t + a(x,t)u_x &= 0 \\ u(x,0) &= \phi(x) \end{aligned} \tag{1.4}$$

1.2. Traffic Flow

1.3. Burgers Equation

Definition 1.5 (Inviscid) Burgers Equation $f = \left(\frac{u^2}{2}\right)$:

$$\begin{aligned} u_t + uu_x &= 0 \\ u(x,0) &= \phi(x) \end{aligned} \tag{1.5}$$

Corollary 1.1 Conservative Formulation [proof 8.4]:

$$\begin{aligned} u_t + \left(\frac{u^2}{2}\right)_{\mathbf{x}} &= 0 \\ u(x,0) &= \Phi(x) \end{aligned} \tag{1.6}$$

1.3.1. Exploding Gradient Problem

Lemma 1.1 [proof 8.5]
Exploding Gradients:
The Burgers equation with smooth initial data $u_0(x) \in \mathcal{C}^1$ and at least one point x_i s.t. $u_0'(x_i) < 0$ will lead to a discontinuity/shockwave[def. 2.3] at a critical time t_{crit}:

$$\text{if } \exists x_i : u_0'(x_i) < 0 \tag{1.7}$$

$$\implies \quad \exists\text{shockwave} \quad \text{at} \quad t_{\text{crit}} = -\frac{1}{\min_{x\in\mathbb{R}} u_0'(x)}$$

Explanation 1.1 (Exploding Gradient Problem).

$$u_x \mapsto +\infty \qquad \text{with time } t$$

thus $\quad u_t + f'(u)u_x = 0 \quad$ is meaningless→ Weak Solutions

1.4. Riemann Problem

Definition 1.6 Riemann Problem: Is an initial value problem of a conservation law with picewise initial data with a single discontinuity of the form:

$$u_t + f(u)_x = 0 \qquad u_0 = \begin{cases} U_R & \text{if } x > 0 \\ U_L & \text{if } x < 0 \end{cases} \tag{1.8}$$

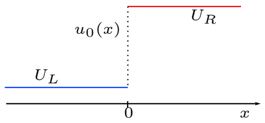


Figure 1: $U_L < U_R$

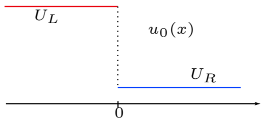


Figure 2: $U_L > U_R$

1.5. Method of Characteristics

For a general introduction to the method of characteristics have a look at Section 2.

Definition 1.7 [proof 8.2]
Characteristic Equations for Conservation Laws:
A curve $\Gamma := (\gamma(\tau),\tau) : [0,T] \mapsto \mathbb{R}\times]0,T[$ in (x,t)-plane is a *characteristic curve* for the conservation law eq. (1.3), if:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\gamma(\tau) = f'\left(u(\gamma(\tau),\tau)\right) \qquad 0 \leqslant \tau \leqslant T \tag{1.9}$$

Corollary 1.2 u **is constant along Characteristics:**

$$u(\gamma(\tau),\tau) = u(\gamma(0),0) = u_0(\gamma_0) \tag{1.10}$$

Proposition 1.1 [proof 8.3]
General Solution for scalar Conservation Laws:
The general solution ofeq. (1.3) is given in terms of the inital condition maybe a non-linear equations:

$$u(x,t) = u_0\left(x - f'\left(u(x,t)\right)t\right) \tag{1.11}$$

Corollary 1.3
Riemann Problem Solution and Inital Data:
In case of a Riemann problem we see that the solution is given by a propagation of the inital data:

$$\begin{aligned} u(x,t) &= \begin{cases} U_L & \text{if } x - f'(u_0)\,t < 0 \\ U_R & \text{if } x - f'(u_0)\,t > 0 \end{cases} && (1.12) \\ &= \begin{cases} U_L & \text{if } x < f'(u_0)\,t \\ U_R & \text{if } x > f'(u_0)\,t \end{cases} && (1.13) \end{aligned}$$

Definition 1.8 [proof 8.3]
Characteristics for Sclar Conservation Laws:
The characteristics going through $(x_0,0)$ are given by the straight lines:

$$x(t) = x_0 + f'\left(u_0(x_0)\right)t \tag{1.14}$$

Problem

We have seen by lemma 1.1 that there exist problems where the spatial gradient explodes $\iff$ we have discontinuous or multivalued solutions s.t. eq. (1.11) and eq. (1.3) are not even well defined.

Weak Solutions

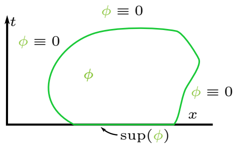
Problems

1. Riemann problems[def. 1.6] may lead to discontinuous solutions u – example 9.1.
2. Even smooth/continuous initial data may lead to discontinuous solutions u – example 9.2

⇒ need a formula that is well defined even in the case of discontinuous ⇒ integral form!

Definition 2.1 Test function $\phi \in \mathcal{C}_C^1\left(\mathbb{R}\times[0,T]\right)$:

Are smooth, compactly supported functions, that are easy to work with.
Idea: use some test functions ϕ that has nicer properties than u and shift the derivative from u to ϕ by using integration by parts.



Definition 2.2 Weak Solutions [proof 8.6]:
For $u_0 \in L^\infty(\mathbb{R})$, $u : \mathbb{R}\times]0,T[\mapsto \mathbb{R}$ is a weak solution of eq. (1.3) if:

$$\int_{-\infty}^{\infty}\int_{0}^{T}\left(u\phi_t + f(u)\phi_x\right)\mathrm{d}x\,\mathrm{d}t + \int_{-\infty}^{\infty} u_0(x)\phi(x,0)\,\mathrm{d}x = 0$$

$$\wedge\, u :\in L^\infty(\mathbb{R}\times]0,T[) \qquad \forall\phi \in C_0^\infty(\mathbb{R}\times[0,T[),\ \phi(\cdot,T) = 0 \tag{2.1}$$

Note

Recall L^∞ bounded but not necessarily differentiable functions i.e. step fucntions.

Explanation 2.1. *Derivatives of u are gone ⇒ we do no longer have the exploding gradient problem lemma 1.1.*

Definition 2.3 Shock Γ:
Let $x = \gamma(t) \in \mathcal{C}^1\left(\mathbb{R}_+\right)$ be a smooth curve along the (x,t)-plane and $u \in L^\infty\left(\mathbb{R}\times\mathbb{R}_+\right)$ a weak solution[def. 2.2] of the scalar conservation laweq. (1.3). Then a discontinuity of u along $\gamma(t)$ is called a *shock*:

$$U(x,t) = \begin{cases} U^-(x,t) & \text{if } x < \gamma(t) \qquad U^- \in \mathcal{C}^1\left(\Gamma^-\right) \\ U^+(x,t) & \text{if } x > \gamma(t) \qquad U^+ \in \mathcal{C}^1\left(\Gamma^+\right) \end{cases}$$

$$\begin{aligned} \Gamma &:= \left\{(x,t) \in \mathbb{R}\times\mathbb{R}_+ \,|x = \gamma(t)\right\} \\ \Gamma^+ &:= \left\{(x,t) \in \mathbb{R}\times\mathbb{R}_+ \,|x > \gamma(t)\right\} \\ \Gamma^- &:= \left\{(x,t) \in \mathbb{R}\times\mathbb{R}_+ \,|x < \gamma(t)\right\} \end{aligned} \tag{2.2}$$

1. The Rankine-Hugoniot Condition

Definition 2.4 [example 9.4],[proof 8.7]
Rankine-Hugoniot Condition: Is a condition on the *shock-speed* $s(t) = \gamma'(t)$ of a shock[def. 2.3] i.e. how fast the shock-wave travels:

$$s(t)\left(u^+(t) - u^-(t)\right) = f\left(u^+(t)\right) - f\left(u^-(t)\right) \tag{2.3}$$

Corollary 2.1 Shock Speed:
Is the speed of a shock[def. 2.3] i.e. the speed of the traveling discontinuity:

$$s(t) = \gamma'(t) = \frac{f\left(u^+(t)\right) - f\left(u^-(t)\right)}{u^+(t) - u^-(t)} \tag{2.4}$$

Explanation 2.2.
The location of a discontinuity which is initially located at $x_0 = 0$ is given by:

$$x(t) = x_0 + s(t)t = x_0 + \gamma'(t)t$$

Theorem 2.1
Necessary Conditions for Weak Solutions and Shocks:
Given a shock wave Γ[def. 2.3] u is a weak solution of eq. (1.2) if and only if:

1. u^- and u^+ are classical solutions of eq. (1.2).
2. the shock speed $s(t) = \gamma'(t)$ satisfies the RH-condition eq. (2.3) at any discontinuities $x = \gamma(t)$.

1.1. Shock Waves

Definition 2.5 [proof 8.8]
Shock Wave Solution for the Riemann Problem:
For conservation laws with a *monotonic* flux function f and Riemann data [def. 1.6]:

$$u_t + f(u)_x = 0 \qquad u_0 = \begin{cases} U_L & \text{if } x < 0 \\ U_R & \text{if } x > 0 \end{cases}$$

The solution is given by

$$u(x,t) = \begin{cases} U_L & \text{if } x < \gamma'(t)t \\ U_R & \text{if } x > \gamma'(t)t \end{cases} \tag{2.5}$$

Definition 2.6 Types of Shocks:
There exist two types of shocks:

1. **Colliding Shocks**: are weak solutions where the inital data flows into the shock.
2. **Emenating Shocks**: are weak solutions where the inital data flows out of the shock.

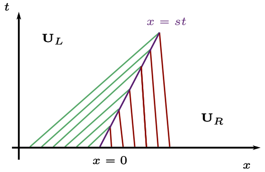


Figure 3: Colliding Shocks

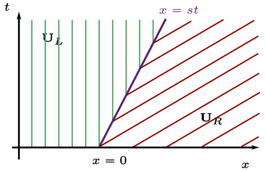


Figure 4: Emenating Shocks

1.1.1. Lax-Oleinik Entropy Condition

Problem

Emenating shocks admit infinitely many weak solutions:

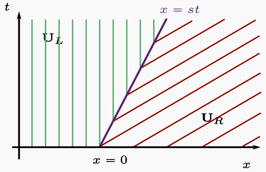


Figure 5: Emenating Shock 1

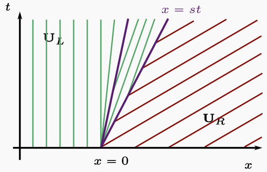


Figure 6: Emenating Shock 2

Thus we allow only for *colliding shocks* ⇒ Lax-Oleinik Entropy Condition.

Proposition 2.1 (Convex Functions)
Lax-Oleinik Entropy Condition:
The characteristics of a general *scalar conservation law* with *monotonic* f function have to flow into the shock:

$$f'\left(u^-(t)\right) > s(t) > f'\left(u^+(t)\right) \tag{2.6}$$

The characteristic equations $x(t)$ are plotted in terms of the $x - t$-plane thus the axes and slopes are turned around.

Corollary 2.2 Categorization by f:

$$f\begin{cases}\text{convex}\\ \text{concave}\end{cases} \implies f'\begin{cases}\text{increasing}\\ \text{decreasing}\end{cases} \Rightarrow \text{iff} \begin{cases} U_L > U_R \\ U_L < U_R \end{cases}$$

$$\implies \qquad \text{physical/colliding shock}$$

Explanation 2.3.

- *For an evolution equation the flow of information should come from the initial data.*
- *We want to require that information flows into and not out from a shock.*

Corollary 2.3 (Burgers Equation)
Lax-Oleinik Entropy Condition: Characteristics of the Burgers equation have to flow into the shock and not emanate at it:

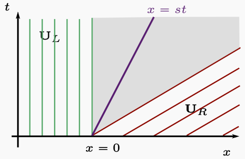
$$u^-(t) > s(t) > u^+(t) \tag{2.7}$$

1.2. Rarefaction Waves

We have seen that an emanating shock:

1. admits infinitely many solutions
2. does not make any sense from a physical standpoint

Question can we find a unique solution for the emanating shocks that does not flow out of the shocks?



Definition 2.7 [example 9.7],[proof 8.9]
Rarefaction Wave Solution for the Riemann Problem:
Let f be a monotonic flux function s.t.:

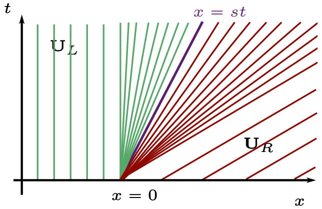
$$f \in \mathcal{C}^2(\mathbb{R}) \text{ is strictly } \begin{cases} \text{convex} \\ \text{concave} \end{cases} \quad \text{and} \quad \begin{matrix} U_L < U_R \\ U_L > U_R \end{matrix} \tag{2.8}$$

Then the solution of the Riemann problem:

$$u_t + f(u)_x = 0 \qquad u_0 = \begin{cases} U_L & \text{if } x < 0 \\ U_R & \text{if } x > 0 \end{cases}$$

ion is given by a rarefaction wave:

$$u(x,t) = \begin{cases} u_L & \\ \left(f'\right)^{-1}\left(\frac{x}{t}\right) & \text{if} \\ u_R & \end{cases} \quad \begin{matrix} x \leqslant \min \overbrace{\left\{f'(u_L), f'(u_R)\right\}}^{:=a} t \\ at < x \leqslant bt \\ x > \underbrace{\max\left\{f'(u_L), f'(u_R)\right\}}_{:=b} t \end{matrix} \tag{2.9}$$



Corollary 2.4 Lax Olenik Entropy Condition: Equation (2.9) satisfies the Lax-Olenik entropy conditionproposition 2.1:

$$f'\left(u^-(t)\right) = s(t) = f'\left(u^+(t)\right) \tag{2.10}$$

2. Entropy Solutions

The Lax-Olenek entropy condition is based on the heuristic that information emanates from initial date, now we want to derive an entropy condition from a mathematical standpoint.

Proposition 2.2 Viscous Approximation: Is a parabolic *convection-diffusion equation* of the form:

$$u_t^\epsilon + f\left(u_t^\epsilon\right)_x = \epsilon u_{xx}^\epsilon \qquad \epsilon > 0 \tag{2.11}$$

Idea

In the limit $\epsilon \to 0$ we recover the inviscide scalar conservation law eq. (1.3). Thus we can study eq. (2.11) in order to study eq. (1.3).

Definition 2.8 Vanishing Viscosity Solution: Is a weak solution u that is the limit of solutions $u = \lim_{\epsilon \to 0} u^\epsilon$ of the viscous equationeq. (2.11).

Definition 2.9 Entropy Pair (s, q):
The pair (s, q) is called entropy pair, where s is any *strictly convex function*[def. 15.26]. Then the entropy pair is defined by the relation:

$$q(u) = \int_0^u f'(\eta)s'(\eta)\,\mathrm{d}\eta \quad \Longrightarrow \quad q' = s'f' \tag{2.12}$$

s : entropy function $\qquad q$: entropy flux

Definition 2.10 Entropy Condition [proof 8.10]:
Any vanishing viscosity solution[def. 2.8] u satisfies:

$$s(u)_t + q(u)_x \leqslant 0 \tag{2.13}$$

Corollary 2.5 [proof 8.11]
Kruzkov's Entropy Condition: Is an entropy condition that holds for weak-solutions:

$$\int_{\mathbb{R}}\int_{\mathbb{R}_+} s(u(x,t))\,\phi_t(x,t) + q(u(x,t))\,\phi_x \,\mathrm{d}x\,\mathrm{d}t + \int_{\mathbb{R}} s(u_0(x))\phi(x,0)\,\mathrm{d}x \geqslant 0 \tag{2.14}$$

$$\forall \phi \in \mathcal{C}_C^1\left(\mathbb{R} \times \mathbb{R}_+\right), \phi \geqslant 0$$

Definition 2.11 Entropy Solution:
A function $u \in L^\infty(\mathbb{R}, \mathbb{R}_+)$ is an entropy solution of the inviscide scalar conservation law eq. (1.3) iff:

1. u is a *weak solution*[def. 2.2] of eq. (1.3).
2. u satisfies the entropy conditioneq. (2.13)/eq. (2.14) for all entropy pairs[def. 2.9] (s, q)

Law 2.1 2nd Laws Of Thermodynamics [proof 8.12]:
The total (mathematical) entropy s decreases in time:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}} s\left(u^\epsilon(x,t)\right) \leqslant 0 \quad \forall \text{ strict. Convex} \tag{2.15}$$

$$\iff \int_{\mathbb{R}} s\left(u^\epsilon(x,t)\right)\mathrm{d}x \leqslant \int_{\mathbb{R}} s\left(u_0(x)\right)\mathrm{d}x \qquad \forall t \tag{2.16}$$

Note: mathematical entropy

The mathematical entropy is defined as the negative physical definition of the entropy $s^{\text{math}} = -s^{\text{phys}} \Rightarrow$ decreases.

2.1. Properties of Entropy Solutions

Property 2.1: Entropy solutions[def. 2.11] for *strictly convex*[def. 15.26] flux function f satisfies the Lax-Oleinik entropy conditioneq. (2.6).

Property 2.2: Entropy solutions are unique.

2.1.1. L^p-bound on entropy solutions

L2-Norm

Property 2.3 L2-Norm:

$$S(u) = u^2 \quad \Longrightarrow \quad \int_{\mathbb{R}}^2 u(x,t)\,\mathrm{d}x \leqslant \int_{\mathbb{R}} u_0^2(x)\,\mathrm{d}x \quad \forall t \tag{2.17}$$

L1-Norm

Property 2.4 L1-Norm:

$$S(u) = |u| \quad \Longrightarrow \quad \int_{\mathbb{R}} |u(x,t)|\,\mathrm{d}x \leqslant \int_{\mathbb{R}} |u_0(x)|\,\mathrm{d}x \quad \forall t \tag{2.18}$$

Lp-Norm

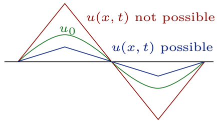
Property 2.5 Lp-Norm:

$$\|u(\cdot,t)\|_{L^p} \leqslant \|u_0\|_{L^p} \qquad \forall 1 \leqslant p \leqslant \infty \tag{2.19}$$

2.1.2. Maximum Principle

Principle 2.1 [proof 8.13]
Maximum Principle: Equation (1.3) attains its maximums on the boundary or its a constant:

$$\max(u(x,t)) \leqslant \max(0, \max u_0(x)) \tag{2.20}$$
$$\min(u(x,t)) \geqslant \min(0, \min u_0(x)) \tag{2.21}$$



2.1.3. Total Variation Diminishing

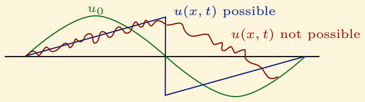
Definition 2.12 Total Variation: If g is differentiable $g \in \mathcal{C}^1([a, b])$ the total variation is defined as:

$$\|g\|_{\mathrm{TV}([a,b])} = \int_a^b \left|\frac{\mathrm{d}g}{\mathrm{d}x}\right|\mathrm{d}x \tag{2.22}$$

Explanation 2.4. *Its a measure on how much a function varies/fluctuates within a interval* $[a, b]$.

Theorem 2.2 [proof 8.14]
Total Variation Diminishing (TVD):
The total variation of an entropy solutions diminished with time:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}} |u_x^\epsilon(\cdot,t)|\,\mathrm{d}x \leqslant 0 \tag{2.23}$$



Corollary 2.6 :

$$\int_{\mathbb{R}} |u_x^\epsilon(\cdot,t)|\,\mathrm{d}x \leqslant \int_{\mathbb{R}} \left|u_x^0\right|\mathrm{d}x \tag{2.24}$$

Corollary 2.7 [proof 8.15]
Total Variation Diminishing in Time: The total time variation is bounded by the space variation:

$$\int_{\mathbb{R}} |u_t^\epsilon(\cdot,t)|\,\mathrm{d}x \leqslant C \int_{\mathbb{R}} \left|u_x^\epsilon(\cdot,t)\right|\mathrm{d}x \tag{2.25}$$

2.1.4. Monotonicity Preservation

Property 2.6
Conversation Laws are Monotonicity Preserving:
If U and V are *entropy solutiuons*[def. 2.11] of eq. (1.3) with initial data U_0 and V_0 then it holds:

$$U_0(x) \leqslant V_0(x) \quad \forall x \quad \Longrightarrow \quad U(x,t) \leqslant V(x,t) \quad \forall x, t \tag{2.26}$$

Finite Volume Methods

From the previous sections we have seen that the solution of conservation laws[def. 1.1] are non-continuous s.t. point values may not be well defined. A solution to this remedy is to work with averages, which are well defined for any integrable function and thus also for solutions of conservation laws.

Definition 3.1 Finite Volume Scheme Grid:

Space Discretization $[x_L, x_R]$

$$x_j := x_L + \left(j + \frac{1}{2}\right)\Delta x \qquad \Delta x := \frac{x_R - x_L}{N+1} \tag{3.1}$$

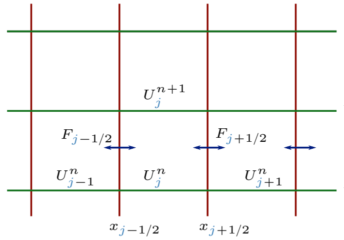
$$= \frac{x_{j-1/2} + x_{j+1/2}}{2} \tag{3.2}$$

$$x_{j\pm\frac{1}{2}} := x_j \pm \Delta x/2 = \begin{cases} x_L + j\Delta x & - \\ x_L + (j+1)\Delta x & + \end{cases} \tag{3.3}$$

$$j \in \{1, \ldots, N+1\}$$

Time Discretization $[0, T]$

$$t^n := n\Delta t \tag{3.4}$$



Definition 3.2 Control Volumes/Cells:
Are the cells defined over the meshpoints x_j of the grid:

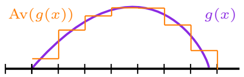
$$\mathcal{C}_j := \left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right) \tag{3.5}$$

Definition 3.3
General Evolution Equation/Scheme:
A $(2p+1)$-point scheme is an update formula that propagates our conserved quantity of interest U in time and relies on $(2p+1)$ cell averages:

$$U_j^{n+1} = H\left(U_{j-p}^n, \ldots, U_{j+p}^n\right) \tag{3.6}$$

Definition 3.4 Cell Averages:
Are averages calculated over the cells[def. 3.2] of a grid[def. 3.1]

$$U_j^n :\approx \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} u\left(x, t^n\right) \mathrm{d}x \tag{3.7}$$

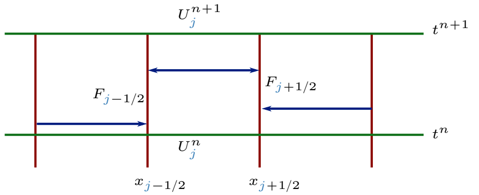


Corollary 3.1 Initial Cell Averages:
Are the initial cell averages for $t = 0$, given by the inital data:

$$U_j^0 :\approx \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} u_0(x)\,\mathrm{d}x \tag{3.8}$$

Definition 3.5 [proof 8.16]
Integrated (Boundary) Fluxes:
Is the flux of our quantity of interest u over left and right boundary of the cells:

$$F_{j\pm\frac{1}{2}}^n := \int_{t_n}^{t_{n+1}} f\left(u\left(x_{j\pm\frac{1}{2}}\right), t\right) \mathrm{d}t \tag{3.9}$$



Definition 3.6 Numerical Fluxes $F_{j-1/2}, F_{j+1/2}$**:**
Are the *exact* integrated or *approximate* numerical fluxes across the boundaries.

Definition 3.7 [proof 8.17]
Finite Volume Scheme:
discretize conservation laws and calculate cell averages[def. 3.4] iteratively by integrating conservation laws[def. 1.1] over the domain $\left[x_{j-1/2}, x_{j+1/2}\right) \times \left[t^n, t^{n+1}\right)$:

$$U_j^{n+1} = U_j^n - \frac{\Delta t}{\Delta x}\left(F_{j+1/2}^n - F_{j-1/2}^n\right)$$

$$U_j^0 = \frac{1}{\Delta x}\int_{x_{j+1/2}}^{x_{j+1/2}} U_0(x)\,\mathrm{d}x \tag{3.10}$$

Corollary 3.2 [proof 8.18]
Finite Volume Schemes in Incremental Form:

$$U_j^{n+1} = U_j^n + C_{j+1/2}^n\left(U_{j+1}^n - U_j^n\right) - D_{j-1/2}^n\left(U_j^n - U_{j-1}^n\right)$$

For the FVM the coefficients are: (3.11)

$$C_{j+1/2}^n := \frac{\Delta t}{\Delta x}\left(\frac{F(u_j, u_j) - F(u_j, u_{j+1})}{u_{j+1} - u_j}\right) \tag{3.12}$$

$$D_{j-1/2}^n := \frac{\Delta t}{\Delta x}\left(\frac{F(u_j, u_j) - F(u_j, u_{j-1})}{u_j - u_{j-1}}\right) \tag{3.13}$$

if F is lipschitz[def. 15.21] in both arguments this is equal to:

$$C_{j+1/2}^n = -\frac{\Delta t}{\Delta x}\frac{\partial F}{\partial b}\left(u_j^n, \cdot\right) \tag{3.14}$$

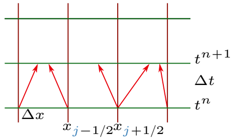
$$D_{j+1/2}^n = -\frac{\Delta t}{\Delta x}\frac{\partial F}{\partial a}\left(\cdot, u_j^n\right) \tag{3.15}$$

1. Properties of Schemes

1.1. The CFL Condition

Definition 3.8 CFL Condition:

$$\max_j \left|f'\left(U_j^n\right)\right|\frac{\Delta t}{\Delta x} \leqslant \frac{1}{2} \tag{3.16}$$



Explanation 3.1. *Enforces that that neighbouring waves in a cell do not inersect each other:*

$$\mathrm{CFL} := \max_j \left|f'\left(U_j^n\right)\right|\Delta t \leqslant \underbrace{\frac{1}{2}\Delta x}_{\textit{half the cell width}} \tag{3.17}$$

Corollary 3.3
The CFL condition can be used to calculate Δt:

$$\Delta t = \mathrm{CFL}\frac{\Delta x}{\max_j \left|f'\left(U_j^n\right)\right|} \tag{3.18}$$

1.2. Conservative Schemes

Definition 3.9 Conservative Schemes:
Are schemes[def. 3.3] that conserve our conservative quantities U over time:

$$\sum_j U_j^{n+1} = \sum_j U_j^n \tag{3.19}$$

Explanation 3.2. *This is nothing else but the fundamental property of our analytic conservation laws.*

Corollary 3.4 FVS are conservative:
FVM schemes[def. 3.7] are conservative.

Note

Finite difference schemes i.e. upwind are usually not conservative⇒blow up.

1.3. Consistent Schemes

Definition 3.10 Consistent Schemes:
A $2p+1$ point scheme[def. 3.3] with numerical Fluxes

$$F_{j+1/2}^n = F\left(U_{j-p+1}^n, \ldots, U_{j+p}^n\right) \tag{3.20}$$

$$F_{j-1/2}^n = F\left(U_{j-p}^n, \ldots, U_{j+p-1}^n\right) \tag{3.21}$$

is consistent if the analytical flux function f is consistent with the numerical flux F that is:

$$F(U, \ldots, U) = f(u) \tag{3.22}$$

Explanation 3.3. *This basically states that if the left and right states are consistent/have the same value then our approximation should do nothing and be equal to the real flux.*

Corollary 3.5 Consitency for FVM:
A FVM[def. 3.7] method is consistent iff for its numerical flux functions it holds that:

$$F(a, a) = f(a) \tag{3.23}$$

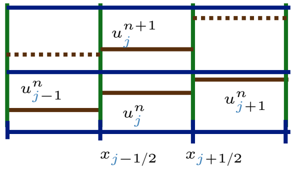
Note

Most of the schemes that we see in the next chapter are consistent and conservative.

1.4. Discrete Maximum Principle

Principle 3.1 Discrete Maximum Principle:
Is the discrete form of principle 2.1:

$$\min\left(U_{j-1}^n, U_j^n, U_{j+1}^n\right) \leqslant U_j^{n+1} \leqslant \max\left(U_{j-1}^n, U_j^n, U_{j+1}^n\right) \tag{3.24}$$



Explanation 3.4. *The previous conserved quantities U^n corresponds to the initial data U_0 of the next Riemann problem.*

Property 3.1 [proof 8.21]
Consistent Monotone Three Point Schemes: Consistent Monotone Three Point Schemes of the form:

$$U_j^{n+1} = H\left(U_{j-1}^n, U_j^n, U_j^n\right) \tag{3.25}$$

satisfy the discrete maximum principle 3.1.

1.5. Discrete Total Variation Diminishing Property

Definition 3.11 Discrete Total Variation: Let g be a function defined on $[a, b]$ then the total variation of g is given by:

$$\|g\|_{\mathrm{TV}([a,b])} = \sup_{\mathcal{P}} \sum_{j=1}^{N-1}\left|g\left(x_{j+1}\right) - g\left(x_j\right)\right| \tag{3.26}$$

where the supremum is taken over all paritions $\mathcal{P} := \{a = x_1 < x_2 < \cdots < x_N = b\}$

Definition 3.12
Discrete Total Variation Diminishing (TVD):
Is the discrete form of theorem 2.2:

$$\left\|U^{n+1}\right\|_{TV(\mathbb{R})} := \sum_j \left|U_{j+1}^{n+1} - U_j^{n+1}\right| \leqslant \sum_j \left|U_{j+1}^n - U_j^n\right| \tag{3.27}$$

Bounded Variation

Definition 3.13 Bounded Variation:

$$\|g\|_{\mathrm{BV}([a,b])} = \|g\|_{L^1([a,b])} + \|g\|_{\mathrm{TV}([a,b])} \tag{3.28}$$

Explanation 3.5. *The total variation[def. 30.18] is only a semi-norm as the TV of any constant function is zero. ⇒ definition of bounded variation makes this a real norm.*

Definition 3.14
Bounded Variation Function Space **BV:**

$$\mathrm{BV}(\mathbb{R}) := \left\{g \in L^1(\mathbb{R}) : \|g\|_{\mathrm{BV}(\mathbb{R})} < \infty\right\} \tag{3.29}$$

1.5.1. Harten's Lemma

Lemma 3.1 [proof 8.22]
Harten's Lemma:
A scheme in incremental form[eq. (5.12)]

$$U_j^{n+1} = U_j^n + C_{j+1/2}^n\left(U_{j+1}^n - U_j^n\right) - D_{j-1/2}^n\left(U_j^n - U_{j-1}^n\right) \tag{3.30}$$

1. with coefficients satisfying:
$$C_{j+1/2}^n, D_{j+1/2}^n \geqslant 0 \quad \text{and} \quad C_{j+1/2}^n + D_{j+1/2}^n \leqslant 1 \quad \forall n, j \tag{3.31}$$
is TVD[eq. (3.27)]
2. with coefficients satisfying:
$$C_{j+1/2}^n, D_{j+1/2}^n \geqslant 0 \quad \text{and} \quad C_{j+1/2}^n + D_{j-1/2}^n \leqslant 1 \quad \forall n, j \tag{3.32}$$
Satisfies: $$\left\|U^{n+1}\right\|_{L^\infty} \leqslant \left\|U^n\right\|_{L^\infty} \qquad \forall n \tag{3.33}$$

1.6. Monotonicity Preserving Schemes

Definition 3.15 Monotone Scheme:
A numerical scheme [def. 3.3] is monotone if the update function H is non-decreasing[def. 15.12] in each of its arguments:

$$\begin{aligned} &a \mapsto H(a, \ldots) && \uparrow \text{ when inceas. } a \text{ and fixing all others} \\ &b \mapsto H(\ldots, b \ldots) && \uparrow \text{ when inceas. } b \text{ and fixing all others} \\ &c \mapsto H(\ldots, c, \ldots) && \uparrow \text{ when inceas. } c \text{ and fixing all others} \end{aligned} \tag{3.34}$$

if H is Lipschitz continuous[def. 15.21] this equals to:

$$\frac{\partial H}{\partial a}, \frac{\partial H}{\partial b}, \frac{\partial H}{\partial c}, \ldots \geqslant 0 \tag{3.35}$$

Corollary 3.6 [proof 8.19]
Monotonicity Preservation: Monotone Schemes[def. 3.15] are *monotonicity preserving* Property 2.6.

$$U_0(x) \leqslant V_0(x) \quad \forall x \quad \Longrightarrow \quad U(x, t) \leqslant V(x, t) \quad \forall x, t \tag{3.36}$$

1.6.1. Monotone Finite Volume Methods

Corollary 3.7 [example 9.3][proof 8.20]
CFL Condition for FVS :
A FVS[def. 3.7] with *monotone* locally Lipschitz continuous[def. 15.21] two-point flux $F(a, b)$ imposes the following CFL (eq. (3.16)) type condition:

$$\left|\frac{\partial F}{\partial a}(v, w)\right| + \left|\frac{\partial F}{\partial b}(u, v)\right| \leqslant 2\max_{a,b}\left(\left|\frac{\partial F}{\partial a}(v, w)\right|, \left|\frac{\partial F}{\partial b}(u, v)\right|\right) \leqslant \frac{\Delta x}{\Delta t} \qquad \forall u, v, w \tag{3.37}$$

Corollary 3.8 [proof 8.20]
Monotone Finite Volume Method:
A Finite Volume Method[def. 3.7] is *montone* iff:

① The flux functions are monotone increasing/decreasing:

$$a \mapsto F(a, b) \quad \text{is } \textbf{non-decreasing} \text{ for fixed } a \tag{3.38}$$

$$b \mapsto F(a, b) \quad \text{is } \textbf{non-increasing} \text{ for fixed } b \tag{3.39}$$

if F is lipschitz continuous[def. 15.21]:

$$\frac{\partial F(a, \cdot)}{\partial a} \geqslant 0 \tag{3.40}$$

$$\frac{\partial F(\cdot, b)}{\partial b} \leqslant 0 \tag{3.41}$$

② it fulfills the CFL-type condition[cor. 3.7]

2. MCC Schemes

Definition 3.16
Monotone Consistent Conservative (MCC) Schemes:
MCC schemes satisfy:

① The Entropy Condition[def. 2.10]

② The Discrete Maximum Principle 3.1

③ The Discrete TVD Property[def. 3.12]

Corollary 3.9
(MCC) Schemes and Entropy Solutions:
MCC schemes will converge to the real entropy solution[def. 2.11] as $\Delta x, \Delta t \to 0$

Summary

	Monotone	Consistent	Conservative	
Godunov	✓	✓	✓	
Roe	✗	✓	✓	
LxF	✓	✓	✓	
EO	✓	✓	✓	
Rusanov	✓	✓	✓	
Central	✗	✓	✓	

Finite Volume Methods Schemes

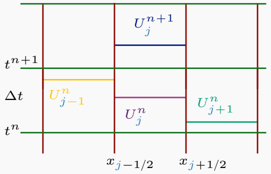
1. Exact Riemann Solvers

1.1. Godunov Method

Problem

The finite volume method**??** requires us to calculate the integrated fluxes eq. (8.6) but those depend again on the unknown solution U.

However Gudonuv noticed that the cell **averages** are constant in each cell $\mathcal{C}_j$ for each time level s.t. each *cell interface* $x_{j+1/2}$ defines a Riemann problem.



Definition 4.1 FVM Riemann Problem:

$$U_t + f(U)_x = 0 \tag{4.1}$$

$$U(x,t^n) = \begin{cases} U_j^n & \text{if } x < x_{j+1/2} \\ U_{j+1}^n & \text{if } x > x_{j+1/2} \end{cases} \tag{4.2}$$

Corollary 4.1 [proof 8.23]
Scaled Gudunov Riemann Problem:
For $U_j(x,t) = U_j\left(\frac{x-x_{j+1/2}}{t-t^n}\right)$ the Riemann problem[def. 4.1] becomes the standard Riemann problem:

$$u(x,0) = \begin{cases} U_j^n & \text{if } x < 0 \\ U_{j+1}^n & \text{if } x > 0 \end{cases} \tag{4.3}$$

Definition 4.2 Godunov Flux:

$$F_{j+1/2}^n\left(U_j^n, U_{j+1}^n\right) = \begin{cases} \min\limits_{U_j^n \leqslant \theta \leqslant U_{j+1}^n} f(\theta) & \text{if } U_j^n \leqslant U_{j+1}^n \\ \max\limits_{U_{j+1}^n \leqslant \theta \leqslant U_j^n} f(\theta) & \text{if } U_j^n > U_{j+1}^n \end{cases} \tag{4.4}$$

Corollary 4.2 Godunov Flux for convex functions: For convex functions f with $\alpha := \min f(\theta)$ it holds:

$$F_{j+1/2}^n\left(U_j^n, U_{j+1}^n\right) = \max\left(f\left(\max\left(U_j^n, \alpha\right)\right), f\left(\min\left(U_{j+1}^n, \alpha\right)\right)\right)$$

Cons
- Solving Equation (4.4) many times for each timestep can become extremely expensive.

2. Approximate Riemann Solvers

2.1. Linearized Riemann Solvers/Roe Schemes

Solving the exact Riemann problemeq. (4.2) can become very expensive. Thus we want to find an approximate solution by *linearizing* non-linear flux functions f.

Definition 4.3
Approximate Riemann Problem[proof 8.24] **[Linear Transport Equation]:**
Is the Riemann problemeq. (4.2) with linearized flux function:

$$u_t + \hat{A}_{j+\frac{1}{2}} u_x = 0 \tag{4.5}$$

$$u(x,t^n) = \begin{cases} U_j^n & \text{if } x < x_{j+1/2} \\ U_{j+1}^n & \text{if } x > x_{j+1/2} \end{cases} \tag{4.6}$$

2.1.1. Arithmetic Roe Average

Definition 4.4 Arithmetic Average:

$$\hat{A}_{j+\frac{1}{2}} = f'\left(\frac{U_j^n + U_{j+1}^n}{2}\right) \tag{4.7}$$

Pros
- Simple
- Consistent

Cons
- Does not satisfy RH condition

2.1.2. Murman Roe Scheme

Definition 4.5 Roe Average: Directly approximate $f'(u)$ using finite differences:

$$\hat{A}_{j+\frac{1}{2}} = \begin{cases} \frac{f\left(U_{j+1}^n\right) - f\left(U_j^n\right)}{U_{j+1}^n - U_j^n} & \text{if } U_{j+1}^n \neq U_j^n \\ f'\left(U_j^n\right) & \text{if } U_{j+1}^n = U_j^n \end{cases} \tag{4.8}$$

Explanation 4.1. *If $u_{j+1}^n = u_j^n$ we don't want to divide by zero.*

Definition 4.6 Roe Flux:
Solving eq. (4.6) with eq. (4.8) leads to the Roe flux:

$$F_{j+1/2}^n = F^{\text{Roe}}\left(U_j^n, U_{j+1}^n\right) = \begin{cases} f\left(U_j^n\right) & \text{if } \hat{A}_{j+\frac{1}{2}} \geqslant 0 \\ f\left(U_{j+1}^n\right) & \text{if } \hat{A}_{j+\frac{1}{2}} < 0 \end{cases} \tag{4.9}$$

Pros
- is simpler in comparison to Godunov scheme
- approximates shock/non-entropy solutions

Cons
- fails at Rarefactions as it does not take into account non-linear bi-directional propagation of information

2.2. Central/HLL Schemes

Harten-Lax-van-Lear **1974**

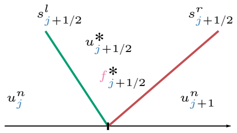
The *Roe-Scheme* fails at resolving rarefaction, this is due to the *linearization* of the *Riemann problem* which leads to a *single wave* solution that travels either to the left or right, depending on the sign of the *Roe average* $\hat{A}_{j+\frac{1}{2}}$.

Problem: the exact solution for a rarefaction can lead to waves traveling in both directions.
Idea: approximate the solution by two waves traveling in opposite directions with speeds s_j^r and s_j^l.

Definition 4.7 [proof 8.25]
Central Schemes:

$$F_{j+1/2}^n = F\left(U_j^n, U_{j+1}^n\right)$$
$$= f_{j+1/2}^{*}$$

+FVM eq. (3.10)



$$f_{j+1/2}^{*} = \frac{s_{j+1/2}^r f\left(U_j^n\right) - s_{j+1/2}^l f\left(U_{j+1}^n\right) + s_{j+1/2}^r s_{j+1/2}^l \left(U_{j+1}^n - U_j^n\right)}{s_{j+1/2}^r - s_{j+1/2}^l} \tag{4.10}$$

The left $s_{j+1/2}^l$ and right $s_{j+1/2}^r$ speeds have to be specified and depend on the scheme.

Corollary 4.3 $-s_{j+1/2}^l = s_{j+1/2}^r =: s_{j+1/2}$
Symmetric Waves:
For anti-symmetric speeds we obtain:

$$f_{j+1/2}^{*} = \frac{f\left(U_j^n\right) + f\left(U_{j+1}^n\right)}{2} - \frac{s_{j+1/2}}{2}\left(U_{j+1}^n - U_j^n\right) \tag{4.11}$$

2.2.1. Lax-Friedrichs Scheme

Definition 4.8 Lax Friedrichs Scheme: Chooses the wave speeds s.t. waves from neighboring Riemann problems do not interact with each other:

$$s_{j+1/2}^l = -\frac{\Delta x}{2\Delta t} \qquad s_{j+1/2}^r = \frac{2\Delta x}{\Delta t} \tag{4.12}$$

with eq. (4.11) it follows:

$$F_{j+1/2}^n = F^{\text{LxF}}\left(U_j^n, U_{j+1}^n\right) \tag{4.13}$$
$$= \frac{f\left(U_j^n\right) + f\left(U_{j+1}^n\right)}{2} - \frac{\Delta x}{2\Delta t}\left(U_{j+1}^n - U_j^n\right)$$

Explanation 4.2. *LxF makes sure that waves do not interfere with each other, that is each wave can maximally travel a distance of $\Delta x = \left|\frac{\Delta t}{s_{j+1/2}^l}\right|$ i.e. to the next interface until we the next time point.*

Pros
- Easy to implement

Cons
- Does not take into account the local speeds
- Is not the most accurate
- Uses always an additional unnecessary grid point

2.2.2. Rusanov Scheme

Definition 4.9
Rusanov/Local-Lax-Friedrichs Scheme:
Takes also into account the local speeds of the waves:

$$s_{j+1/2} = \max\left(|f'\left(U_j^n\right)|, |f'\left(U_{j+1}^n\right)|\right) \tag{4.14}$$

with eq. (4.11) and $s_{j+1/2}^r = s_{j+1/2} = -s_{j+1/2}^l$ it follows:

$$F_{j+1/2}^n = F^{\text{Rus}}\left(U_j^n, U_{j+1}^n\right) \tag{4.15}$$
$$= \frac{f\left(U_j^n\right) + f\left(U_{j+1}^n\right)}{2} \tag{4.16}$$
$$- \frac{\max\left(|f'\left(U_j^n\right)|, |f'\left(U_{j+1}^n\right)|\right)}{2}\left(U_{j+1}^n - U_j^n\right)$$

2.2.3. Enquist-Osher Flux

Definition 4.10
Engquist Osher Scheme:
Is related to[def. 4.9] but is kind of a continuous version:

$$F_{j+1/2}^n = F^{\text{EO}}\left(U_j^n, U_{j+1}^n\right) \tag{4.17}$$
$$= \frac{f\left(U_j^n\right) + f\left(U_{j+1}^n\right)}{2} - \frac{1}{2}\int_{U_j^n}^{U_{j+1}^n} \left|f'(\theta)\right| \mathrm{d}\theta$$

Corollary 4.4 Engquist Oshner for Convex Functions:
For convex functions f with a single minimum $\alpha := \min f(\theta)$ it holds:

$$F^{\text{EO}}\left(U_j^n, U_{j+1}^n\right) = f^{+}\left(U_j^n\right) + f^{-}\left(U_{j+1}^n\right) \tag{4.18}$$

$$f^{+}(u) := f(\max(u, \alpha))$$
$$f^{-}(u) := f(\min(u, \alpha))$$

Higher Order Schemes

Goal

Design higher-order (2^{nd})-order schemes which are stable that is:

1. TVD[def. 3.12].
2. fullfils the Max Principleprinciple 3.1.

and reduce the error/are more accurate.

Definition 5.1 Truncation Error τ**:**
The truncation error w.r.t. $u_j^{n+1} = H\left(u_{j-1}^n, u_j^n, u_{j+1}^n\right)$ is defined as:

$$\tau := u\left(x_j, t^{n+1}\right) - H\left(u\left(x_{j-1}, t^n\right), u\left(x_j, t^n\right), u\left(x_{j+1}, t^n\right)\right) \tag{5.1}$$

Definition 5.2 Order of Scheme:
The order of a scheme is defined:

$$q: \quad \max_{j,n} \left|\tau_j^n\right| \leqslant C\Delta t^{q+1} \tag{5.2}$$

1. Lax-Wendroff Scheme 1961

Definition 5.3 [proof 8.26]
Lax-Wendroff Scheme:

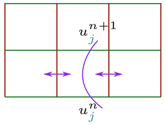
$$\begin{aligned} u_j^{n+1} = u_j^n &- \frac{\Delta t}{2\Delta x}\left(f\left(u_{j+1}^n\right) - f\left(u_{j-1}^n\right)\right) \\ &+ \frac{\Delta t^2}{2\Delta x^2}\left[a_{j+1/2}^n\left(f\left(u_{j+1}^n\right) - f\left(u_j^n\right)\right)\right. \\ &\left. - a_{j-1/2}^n\left(f\left(u_j^n\right) - f\left(u_{j-1}^n\right)\right)\right] \end{aligned} \tag{5.3}$$

$$f'(u)\left(x_{j+1/2}\right) =: a_{j+1/2}^n = f'\left(\frac{u_j^n + u_{j+1}^n}{2}\right)$$

Corollary 5.1 As a Finite Volume Scheme:

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x}\left(F_{j+1/2}^n - F_{j-1/2}^n\right)$$

$$F_{j+1/2}^n = F_{j+1/2}^n\left(u_j^n, u_{j+1}^n\right)$$



$$= \frac{f\left(u_j^n\right) + f\left(u_{j+1}^n\right)}{2} - \frac{\Delta t}{\Delta x} a_{j+1/2}^n\left(f\left(u_{j+1}^n\right) - f\left(u_j^n\right)\right)$$

Pros

- Formally 2^{nd}-order accurate
- Is Consistent
- Conservative

Cons

- Comes with oscillations
- Is not monotone
- Is not TVD
- Does not fulfill the discrete maximum principle

2. REA-Algorithms

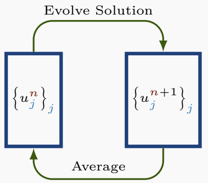
2.1. Reconstruction

Definition 5.4 Averaging Operator:

$$\operatorname{Avg}(g) = \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} g(x)\,\mathrm{d}x \qquad \text{if } x_{j-1/2} \leqslant x \leqslant x_{j+1/2} \tag{5.4}$$

Interpretation of Gurdonuv Type Schemes

$$\text{Given } \left\{u_j^n\right\}_j \xrightarrow[\text{Rieman Problem Sol.}]{\text{Evolve (approx.)}} \text{Average}$$

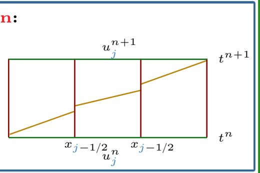


$$\left\{u_j^n\right\}_j \leftarrow \operatorname{Avg}\left(\left\{u_j^{n+1}\right\}_j\right)$$

Definition 5.5 Reconstruction:

Replacing cell-averages[def. 3.4] by *piecewise-linears*:

$p^n(x) = p_j^n$
if $x_{j-1/2} \leqslant x \leqslant x_{j+1/2}$
$p_j^n(x) := a_j^n x + b^n$



Definition 5.6 REA Algorithm:

$$\text{R–E–A–R–E–A–R–E–A} \tag{5.5}$$

1. Reconstruction: at time t^n we know the approximate cell averages u_j^n and realize them by some functions:
$$u(x, t^n) = p_j^n(x) \qquad x_{j-1/2} \leqslant x \leqslant x_{j+1/2}$$
2. Evolution: the reconstruction function is evolved in time by solving the Riemann problem either exactly or approximately:
$$u(x, t^n) \overset{\text{evolve}}{\mapsto} u(x, t^{n+1})$$
3. Averaging: we average the solutions at the next time step t^{n+1} over each control volume.

Corollary 5.2 Evolution is TVD:
We have seen that all Riemann solver (apart from Roe-Scheme) are TVD[def. 3.12]

Corollary 5.3 Averaging is TVD:
Given a function $f \in \operatorname{Lip}(\Omega)$ then it holds that the average is TVD[def. 3.12]:

$$\operatorname{TV}\left(\operatorname{Av}\left(f\right)\right) \leqslant \operatorname{TV}(f) \qquad \operatorname{Av}(f)_j := \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} f(x)\,\mathrm{d}x \tag{5.6}$$

Lemma 5.1 Piecewise Constant Averaging:
If we replace the exact solutions with piecewise constant averages then it holds for the error:

$$\|g - \operatorname{Avg}(g)\|_{L^1} \leqslant C\Delta X = \mathcal{O}(\Delta x) \qquad g \in L^1(\Omega) \tag{5.7}$$

Definition 5.7 Generalized Riemann Problem:

$$u_t + f(u)_x = 0 \tag{5.8}$$

$$u(x, t^n) = p^n(x) \tag{5.9}$$

Cons

- Hard to solve exactly! (except for $f(u) = au$)

2.2. Approximate Reconstruction

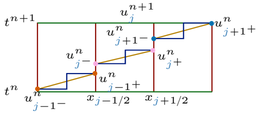
Definition 5.8 Approximate Reconstruction:
Approximate *piecewise-linears* of the cell-averages[def. 3.4] by two a simpler problem:

$$p^n(x) = \left\{p_j^n(x)\right\}_j$$

$$p_j^n(x) = a_j^n x + b_j^n$$

$$u_{j+}^n = p_j^n\left(x_{j+1/2}\right)$$

$$u_{j-}^n = p_j^n\left(x_{j-1/2}\right)$$



Corollary 5.4 Linear Approximate Reconstruction:

$$u_{j\pm}^n = p_j^n\left(x_{j\pm1/2}\right) = \underbrace{u_j^n}_{\text{midpoint}} \pm \overbrace{\frac{\Delta x}{2}}^{\text{distance to boundary}} \sigma_j^n \tag{5.10}$$

Definition 5.9 FVM Evolution and Averaging:

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x}\left(F\left(u_{j+}^n, u_{j+1-}^n\right) - F\left(u_{j-1+}^n, u_{j-}^n\right)\right) \tag{5.11}$$

Corollary 5.5 [proof 8.29]
FVM Evolution and Averaging in Incremental Form:

$$U_j^{n+1} = U_j^n + C_{j+1/2}^n\left(U_{j+1}^n - U_j^n\right) - D_{j-1/2}^n\left(U_j^n - U_{j-1}^n\right)$$

$$c_{j+1/2}^n = \frac{\Delta t}{\Delta x}\frac{f\left(u_{j+}^n, u_{j-}^n\right) - f\left(u_{j+}^n, u_{j+1-}^n\right)}{u_{j+1}^n - u_j^n}$$

$$d_{j+1/2}^n = \frac{\Delta t}{\Delta x}\frac{f\left(u_{j+1+}^n, u_{j+1-}^n\right) - f\left(u_{j+}^n, u_{j-1-}^n\right)}{u_{j+1}^n - u_j^n} \tag{5.12}$$

Lemma 5.2
TVD REA Scheme:
A FVM REA[def. 5.6] scheme is TVD iff *construction*, *averaging* and *evolution* are all TVD.

We know that evolution[cor. 5.2] and averaging[cor. 5.3] is TVD thus we need to find a *reconstruction* that is TVD.

Lemma 5.3 [proof 8.30]
TVD REA scheme:
A REA[def. 5.6] scheme is TVD iff:

1. eq. (5.26) satisfies the CFL condition eq. (3.37)
2. $T_1, T_2 \geqslant 0$
3. $T_1 + T_2 \leqslant 2$

$$T_1 := \frac{U_{j+1-}^n - U_{j-}^n}{U_{j+1}^{n+1} - U_j^n} \qquad T_2 := \frac{U_{j+1+}^n - U_{j+}^n}{U_{j+1}^{n+1} - U_j^n} \tag{5.13}$$

2.2.1. Constraints

1. Conservation:
$$\frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} p_j^n\,\mathrm{d}x = u_j^n$$
$$\overset{\text{proof 8.27}}{\Longrightarrow} \quad \int_D p^n(x)\,\mathrm{d}x = \int_D u_0(x)\,\mathrm{d}x$$
$$\overset{\text{proof 8.28}}{\Longrightarrow} \quad p_j^n = u_j^n + \sigma_j^n\left(x - x_j\right)$$
2. TVD: how would we choose the *slope* σ_j^n?
Obvious choices would be:
   - Forward Differences:
   $$\sigma_j^n = \frac{u_{j+1}^n - u_j^n}{\Delta x}$$
   - Backward Differences:
   $$\sigma_j^n = \frac{u_j^n - u_{j-1}^n}{\Delta x}$$
   - Central Differences:
   $$\sigma_j^n = \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x}$$

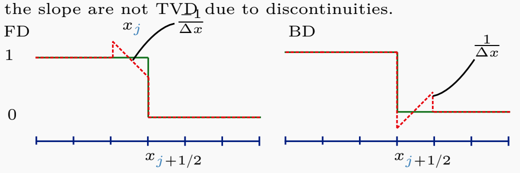
Problem: schemes using this slopes are unstable, satisfy neither TVD nor-discrete maximum principle preserving.

2.3. Limiters

We have seen that schemes using simple finite differences for the reconstructions slope σ_j^n are unstable and we know that the evolution and averaging operations are TVD[cor. 5.2] thus we need to ensure that the reconstruction is TVD as well:

$$\operatorname{TV}\left(p^n\right) \leqslant \operatorname{TV}\left(u^n\right)$$

The problem is that schemes using simple finite differences for the slope are not TVD due to discontinuities.



2.3.1. Minmod Limiter

Definition 5.10 Minmod Limiter: Compare the upwind- and downwind slope and checks if they have the same sign. If yes, it sets the slope to the smaller one otherwise it sets the slope to zero.

$$\sigma_j^n = \operatorname{minmod}\left(\frac{u_{j+1}^n - u_j^n}{\Delta x}, \frac{u_j^n - u_{j-1}^n}{\Delta x}\right) \tag{5.14}$$

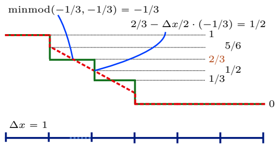
$$\operatorname{minmod}\left(a_1, \ldots, a_n\right) \tag{5.15}$$

$$= \begin{cases} \operatorname{sign}(a_1)\min_{1\leqslant k\leqslant n}\left(|a_k|\right) & \text{if } \operatorname{sign}(a_1) = \cdots = \operatorname{sign}(a_n) \\ 0 & \text{otherwise} \end{cases}$$

Corollary 5.6 [proof 8.32]
Minmod is TVD:

If the reconstrution p^n uses a min-mod limiter, then:

$$\operatorname{TV}\left(p^n\right) \leqslant \operatorname{TV}\left(u^n\right) \tag{5.16}$$



2.3.2. Superbee Limiter

Definition 5.11 [Roe 1981]
Superbee Limiter:

$$\sigma_j^n = \operatorname{maxmod}\left(\sigma_j^L, \sigma_j^R\right) \tag{5.17}$$

$$\sigma_j^L = \operatorname{minmod}\left(\frac{u_{j+1}^n - u_j^n}{\Delta x}, 2\frac{u_j^n - u_{j-1}^n}{\Delta x}\right)$$

$$\sigma_j^R = \operatorname{minmod}\left(2\frac{u_{j+1}^n - u_j^n}{\Delta x}, \frac{u_j^n - u_{j-1}^n}{\Delta x}\right)$$

$$\operatorname{maxmod}\left(a_1, \ldots, a_n\right) \tag{5.18}$$

$$= \begin{cases} \operatorname{sign}(a_1)\max_{1\leqslant k\leqslant n}\left(|a_k|\right) & \text{if } \operatorname{sign}(a_1) = \cdots = \operatorname{sign}(a_n) \\ 0 & \text{otherwise} \end{cases}$$

Corollary 5.7 Superbee is TVD: If the reconstrution p^n uses a superbee-mod limiter, then:

$$\operatorname{TV}\left(p^n\right) \leqslant \operatorname{TV}\left(u^n\right) \tag{5.19}$$

2.3.3. MC Limiter

Definition 5.12 [Vanleer 1987]
Monotonized Central (MC):

$$\sigma_j^n = \operatorname{minmod}\left(2\frac{u_{j+1}^n - u_j^n}{\Delta x}, 2\frac{u_j^n - u_{j-1}^n}{\Delta x}, \frac{u_{j+1}^n - u_{j-1}^n}{2\Delta x}\right)$$

$$\operatorname{minmod}\left(a_1, \ldots, a_n\right) \tag{5.20}$$

$$= \begin{cases} \operatorname{sign}(a_1)\min_{1\leqslant k\leqslant n}\left(|a_k|\right) & \text{if } \operatorname{sign}(a_1) = \cdots = \operatorname{sign}(a_n) \\ 0 & \text{otherwise} \end{cases}$$

Corollary 5.8 MC is TVD: If the reconstrution p^n uses a mc-mod limiter, then:

$$\operatorname{TV}\left(p^n\right) \leqslant \operatorname{TV}\left(u^n\right) \tag{5.21}$$

2.4. TVD REA Schemes

Lemma 5.4 [example 8.32],[proof 8.31]
TVD FVM REA Scheme:
A three point FVM REA[def. 5.6] scheme is TVD iff:

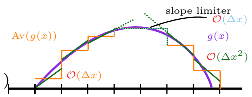
(1) eq. (5.26) satisfies the CFL condition eq. (3.37)

(2) and the following condition is satisfied:

$$-2 \leqslant \frac{\delta^n_{j+1} - \delta^n_j}{u^n_{j+1} - u^n_j} \leqslant 2 \qquad \delta_j := \sigma^n_j \Delta x \tag{5.22}$$

Proposition 5.1 Order of Accuracy:
Given $g(x) \in \mathcal{C}^2$ and g is monotone (no extreme) and not slope limited then it holds for[def. 5.9]:

$$\|g(x) - p_n(x)\|_{L^\infty} \approx \mathcal{O}(\Delta x^2) \tag{5.23}$$



If we require TVD slope limiters however we will have again be first order accuracy at the regions of slope limiters/local extrema:

$$\|g(x) - p_n(x)\|_{L^\infty} \approx \mathcal{O}(\Delta x) \tag{5.24}$$

Note

Also have a look at *expected order of convergence* (EOC) or rate of convergence[def. 23.10] for numerical experiments.

3. Higher Order Time Schemes

3.1. Semi-Discrete Schemes

Definition 5.13 [example 9.8]
Semi-Discrete FVM: Is a time-continuous but space-discrete formulation of[def. 5.9]:

$$\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = \mathscr{L}(\mathbf{u}) \tag{5.25}$$

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} u_j(t) &= \mathscr{L}(\mathbf{u}_j) \qquad \text{rate of change} \\ &=: -\frac{1}{\Delta x}\left(F\left(u^n_{j+}, u^n_{j+1-}\right) - F\left(u^n_{j-1+}, u^n_{j-}\right)\right) \end{aligned} \tag{5.26}$$

Definition 5.14 Semi-discrete Cell Averages:

$$U^n_j :\approx \frac{1}{\Delta x}\int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} u(x,t)\,\mathrm{d}x \tag{5.27}$$

Definition 5.15 **(SSP) RK Schemes**
Strong Stability Preserving Runge-Kutta Methods:
Are Runge-Kutta methods that preserve the TVD property eq. (3.27).

Summary what we need

(1) Mesh/Grid

(2) Numerical Flux $F(u, v)$ (consistent/monotone)

(3) Reconstruction: given $\{u_j\}$ output $\left\{u^{\pm}_j\right\}$

$$u^{\pm}_j = u_j \pm \sigma_j \frac{\Delta x}{2}$$

(4) Slope Limiters for the slope σ_j

(5) SSP-RK scheme

3.1.1. Heun's Method

Definition 5.16
Heun's Method **(SSP-RK2):**
Applies forward Euler twice and averages them to obtain a 2nd-order method:

$$\mathbf{U}^* = \mathbf{U}^n + \Delta t\,\mathscr{L}(\mathbf{U}^n) \tag{5.28}$$

$$\mathbf{U}^{**} = \mathbf{U}^* + \Delta t\,\mathscr{L}\left(\mathbf{U}^*\right) \tag{5.29}$$

$$\mathbf{U}^{n+1} = \frac{\mathbf{U}^n + \mathbf{U}^{**}}{2} \tag{5.30}$$

Properties

Property 5.1 [proof 8.33]
TVD: Heun's Method is TVD.

Property 5.2 [proof 8.34]
2nd Order Accuracy:
Heun's Method is second order accurate.

Systems of Conservation Laws

Definition 6.1 Systems of Conservation Law:

$$\mathbf{u}_t + f(\mathbf{u}(\mathbf{x},t),\mathbf{x})_{\mathbf{x}} = s(\mathbf{u}(\mathbf{x},t),\mathbf{x},t) \qquad \text{in } \tilde{\Omega} := \Omega\times]0,T[\tag{6.1}$$

1. Linear System of Conservation Laws

Definition 6.2 [examples 9.9 and 9.10 and ??]
Linear System of Conservation Laws:

$$\mathbf{u}_t + \mathbf{A}\mathbf{u}_{\mathbf{x}} = s(\mathbf{u}(\mathbf{x},t),\mathbf{x},t) \qquad \text{in } \tilde{\Omega} := \Omega\times]0,T[$$
$$\mathbf{u} = [u_1 \quad u_2 \cdots\cdots u_m]^\intercal \qquad \mathbf{u} = [f_1 \quad f_2 \cdots\cdots f_m]^\intercal \tag{6.2}$$

Corollary 6.1
Linear Sys. of Cons. Laws with Variable Coefficients:

$$\mathbf{u}_t + (\mathbf{A}(\mathbf{x},t)\mathbf{u})_{\mathbf{x}} = s(\mathbf{u}(\mathbf{x},t),\mathbf{x},t) \qquad \text{in } \tilde{\Omega} := \Omega\times]0,T[\tag{6.3}$$

Corollary 6.2 [proof 8.35]
Linearizing Systems of Conservation Laws:
Equation (6.1) can be linearized into eq. (6.2).

1.1. Types of Linear Systems

Definition 6.3 Hyperbolic System:
The linear systemeqs. (6.2) and (6.3) are called *hyperbolic* if the matrix $\mathbf{A}$ is diagonalizable and has m real eigenvalues:

$$\text{spectrum}(\mathbf{A})(\mathbf{x},t) = \{\lambda(\mathbf{x},t)_1,\ldots,\lambda(\mathbf{x},t)_m\} \in \mathbb{R} \qquad \forall \mathbf{x},t \tag{6.4}$$

Corollary 6.3 Strictly Hyperbolic System:
The linear systemeqs. (6.2) and (6.3) is called *strictly hyperbolic* if it is *hyperbolic*[def. 6.3] and all eigenvalues are distinct:

$$\textit{eq. } (6.4) \qquad + \qquad \lambda_1 \neq \lambda_2 \neq \ldots \neq \lambda_m \tag{6.5}$$

1.2. Properties of Schemes

1.2.1. Discrete Total Variation Diminishing Property

Definition 6.4
Discrete Total Variation Diminishing (TVD): Hyperbolic linear systems of conservation laws

$$\left\|\mathbf{U}^{n+1}\right\|_{TV(\mathbb{R})} := \sum_j \left\|\mathbf{U}^{n+1}_{j+1} - \mathbf{U}^{n+1}_j\right\| \leqslant \sum_j \left\|\mathbf{U}^n_{j+1} - \mathbf{U}^n_j\right\|$$
$$\leqslant \sum_j \sum_p^m \left|U^{p,n}_{j+1} - U^{p,n}_j\right| \tag{6.6}$$

2. Decoupling of Linear Systems

Proposition 6.1 [proof 8.36]
Decoupled hyperbolic lin. Cons. Law:
Hyperbolic linear systems of conservation laws[def. 6.2] can be decoupled into m linear equations:

$$\mathbf{W}_t + \Lambda\mathbf{W}_x = 0 \quad \Longleftrightarrow \quad W^p_t + \lambda_p W^p_x = 0 \quad \forall p = 1,\ldots,m$$
$$\mathbf{W} = \mathbf{R}^{-1}\mathbf{U} \qquad \mathbf{R} = [\mathbf{r}_1 \cdots\cdots \mathbf{r}_p] \qquad \mathbf{A}\mathbf{r}_j = \lambda_j \mathbf{r} \tag{6.7}$$

Corollary 6.4 [proof 6.1]
Solution of hyp. lin. cons. laws:

$$W^p(x,t) = W^p_0(x - \lambda_p t) \qquad \mathbf{W}_0(x) = \mathbf{R}^{-1}\mathbf{U}_0(x) \tag{6.8}$$
$$\mathbf{U}(x,t) = \mathbf{R}\mathbf{W}(x,t) \tag{6.9}$$

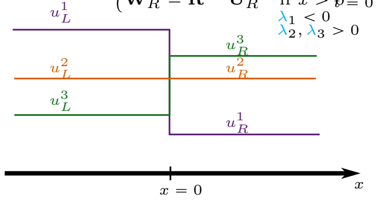
Proof 6.1 Solution of hyp. lin. cons. law:

2.0.1. Riemann Problems

Definition 6.5 Decoupled Riemann Problem:
Splits the original Riemann data in multiple problems:

$$\mathbf{W}_t + \Lambda\mathbf{W}_x = 0$$

$$\mathbf{W}_0(x) = \begin{cases} \mathbf{W}_L = \mathbf{R}^{-1}\mathbf{U}_L & \text{if } x < 0 \\ \mathbf{W}_R = \mathbf{R}^{-1}\mathbf{U}_R & \text{if } x > 0 \end{cases} \tag{6.10}$$



Corollary 6.5
Riemann Problem for hyp. lin. cons. law: The solution of a Riemann problem of a hyperbolic[def. 6.3] linear conservation laweq. (6.10) is given by:

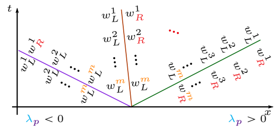
$$W^p(x,t) = W^p_0(x - \lambda_p t) = \begin{cases} W^p_L & \text{if } \lambda_p t < 0 \\ W^p_R & \text{if } \lambda_p t > 0 \end{cases} \tag{6.11}$$

Explanation 6.1.

- *λ_p speed of the wave*
- *$\lambda_p t$ is called the p-th wave*

Corollary 6.6 [proof 8.37]
Jumps: The Riemann problem of a linear system of conservation laws[cor. 6.5] decomposed into m jumps s.t. we obtain m waves/solutions:

$$\mathbf{U}_R - \mathbf{U}_L = \sum_{p=1}^m \alpha^p r_p \tag{6.12}$$



$\alpha^p := (\mathbf{W}_R - \mathbf{W}_L)$: strength of the p-th wave
r_p: direction of the characteristics

2.1. FVM Scheme

Definition 6.6 [proof 8.40]
Finite Volume Scheme for Linear Systems:

① **Reconstruction:**

$$\mathbf{U}(x,t^n) = p^n_j(x) \overset{i.e.}{=} \begin{cases} \mathbf{U}^n_j & \text{p.w. const} \\ \mathbf{U}^n_j = \mathbf{U}^n_j \pm \frac{\Delta x}{2}\sigma^n_j & \text{linear} \end{cases}$$
$$x_{j-1/2} \leqslant x \leqslant x_{j+1/2}$$

② **Evolution:** by solving Riemann problems:

$$\mathbf{U}_t + \mathbf{A}\mathbf{U}_x = 0$$
$$\mathbf{U}(x,t^n) = \begin{cases} \mathbf{U}^n_j & \text{if } x < x_{j+1/2} \\ \mathbf{U}^n_{j+1} & \text{if } x > x_{j+1/2} \end{cases}$$

③ **Averaging:**

$$\mathbf{U}^{n+1}_j = \mathbf{U}^n_j - \frac{\Delta t}{\Delta x}\left(\mathbf{F}^n_{j+1/2} - \mathbf{F}^n_{j-1/2}\right)$$
$$\mathbf{F}^n_{j\pm1/2} = \mathbf{F}\left(\mathbf{U}^n_j, \mathbf{U}^n_{j+1}\right) = \mathbf{A}_{j\pm1/2}\mathbf{U}(x_{j\pm1/2}, t^n)$$

2.1.1. CFL Condition

Definition 6.7 CFL Condition System of Cons. Laws:
The wave speed is given by $\lambda_{\max} := \max_{1\leqslant p\leqslant m}|\lambda_p|$ s.t. it follows from eq. (3.16):

$$\lambda_{\max} \leqslant \frac{\Delta x}{\Delta t}\frac{1}{2} \tag{6.13}$$

2.1.2. Exact Fluxes

Godunov Flux

Definition 6.8 [proof 8.38]
Godunov Flux:

$$\mathbf{F}^n_{j+1/2} = \mathbf{A}\mathbf{U}_{j+1/2} \tag{6.14}$$
$$= \frac{1}{2}\mathbf{A}\left(\mathbf{U}^n_j + \mathbf{U}^n_{j+1}\right) - \frac{1}{2}\mathbf{R}|\Lambda|\mathbf{R}^{-1}\left(\mathbf{U}^n_{j+1} - \mathbf{U}^n_j\right)$$

Property 6.1 [proof 8.39]
Total Variation Bounded (TVB): Godunov flux for systems of scalar conservation laws is total variation bounded:

$$\text{TV}(\mathbf{U}^{n+1}) \leqslant \|\mathbf{R}\|\|\mathbf{R}^{-1}\|\text{TV}(\mathbf{U}^n) \tag{6.15}$$

Note

It is not TVD as we do not know what the condition numbers $\|\mathbf{R}\|\|\mathbf{R}^{-1}\|$ are.

Godunov Flux is the

2.1.3. Approximate Fluxes

Central Fluxes

Definition 6.9 Lax Friedrichs Scheme:
Chooses the wave speeds s.t. waves from neighboring Riemann problems do not interact with each other:

$$s^l_{j+1/2} = -\frac{\Delta x}{2\Delta t} \qquad s^r_{j+1/2} = \frac{2\Delta x}{\Delta t} \tag{6.16}$$

with eq. (4.11) it follows:

$$\mathbf{F}^n_{j+1/2} = \mathbf{F}^{\text{LxF}}\left(\mathbf{U}^n_j, \mathbf{U}^n_{j+1}\right) \tag{6.17}$$
$$= \frac{1}{2}\mathbf{A}\left(\mathbf{U}^n_j + \mathbf{U}^n_{j+1}\right) - \frac{\Delta x}{2\Delta t}\left(\mathbf{U}^n_{j+1} - \mathbf{U}^n_j\right)$$

Definition 6.10
Rusanov/Local-Lax-Friedrichs Scheme:
Takes into account the local speeds λ_p of the waves (and not only the grid):

$$s_{j+1/2} = \max|\Lambda| \tag{6.18}$$

with eq. (4.11) and $s^r_{j+1/2} = s_{j+1/2} = -s^l_{j+1/2}$ it follows:

$$\mathbf{F}^n_{j+1/2} = \mathbf{F}^{\text{Rus}}\left(u^n_j, u^n_{j+1}\right) \tag{6.19}$$
$$= \frac{1}{2}\mathbf{A}\left(\mathbf{U}^n_j + \mathbf{U}^n_{j+1}\right) \tag{6.20}$$
$$- \frac{\lambda_{\max}}{2}\left(\mathbf{U}^n_{j+1} - \mathbf{U}^n_j\right)$$

3. Higher Order Schemes

Goal

Design a 2nd-order TVB-stable scheme.

3.1. Reconstruction

Definition 6.11 Conservative Variables:
Are the variables $\mathbf{U}$ used to write a system in conservative form.

Definition 6.12 Primitive Variables: Are the variables, that make up the conservative variables.

Definition 6.13 Characteristic Variabels:
Are the variables of the decoupled linear system $\mathbf{W} = \mathbf{R}^{-1}\mathbf{U}$ are called the Characteristic Variables.

Definition 6.14 Primitive Reconstruction:
Apply limiters section 3 componentwise to the primitive variables $\mathbf{U}_j$.

Pros	**Cons**
• Easy to apply	• Does not necessarily lead to TVBProperty 6.1 stable reconstruction section 1 scheme.

Definition 6.15 Characteristic Reconstruction: Apply limiterssection 3 componentwise to the characteristic variables $\mathbf{W}_j$.

$$\gamma^n_j = \text{limiter}\left(\mathbf{W}^n_{j-1}, \mathbf{W}^n_j, \mathbf{W}^n_{j+1}\right) \quad \Longrightarrow \quad \sigma^n_j = \mathbf{R}\gamma^n_j \tag{6.21}$$

Corollary 6.7 : Scheme ③ [def. 5.9] is:

- is 2nd-order accurate in space formally.
- is TVB-stable if σ is defined by Characteristic reconstruction.

3.2. Higher Order in Time

Proposition 6.2
Heun's Method for Systems of Conservation Laws:
Given a system of conservation laws the following scheme:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{U}_j(t) = -\frac{\Delta t}{\Delta x}\left(\mathbf{F}\left(\mathbf{U}^n_{j+}, \mathbf{U}^n_{j+1-}\right) - \mathbf{F}\left(\mathbf{U}^n_{j-1+}, \mathbf{U}^n_{j-}\right)\right)$$
$$-\frac{\Delta t}{\Delta x}\left(\mathbf{F}_{j+1/2}(t) - \mathbf{F}_{j-1/2}\right) =: \mathscr{L}(\mathbf{U}(t))_j$$
$$\mathbf{U}^+_j = p(x_{j+1/2}) \qquad \mathbf{U}^-_j = p(x_{j-1/2})$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{U}(t) = \mathscr{L}(\mathbf{U}(t)) \qquad \mathbf{U}(t) := [\cdots \quad \mathbf{U}_{j-1} \quad \mathbf{U}_j \quad \mathbf{U}_{j+1} \quad \cdots]$$

with Heun's Method[def. 5.16] is 2nd-order in time.

Non-Linear Systems of Conservation Laws

Definition 7.1
Nonlinear Systems of Conservation Laws:

$$\begin{aligned}\partial_t \mathbf{U} + \partial_x \mathrm{f}(\mathbf{U}) &= \mathbf{0} \\ \mathbf{U}(x,0) &= \mathbf{U}_0(x)\end{aligned} \qquad \begin{aligned}&\mathbf{U}: \mathbb{R}\times\mathbb{R}_+ \to \mathcal{U} \in \mathbb{R}^m \\ &\mathbf{U} \in L^\infty(\mathbb{R}\times[0,T];\mathcal{U}) \\ &\mathrm{f}: \mathcal{U} \to \mathbb{R}^m \quad \text{(nonlinear)}\end{aligned} \tag{7.1}$$

Definition 7.2 Admissible Set $\mathcal{U}$**:**
Is the domain of admissible values that make sense in a physical way.

Definition 7.3 j**-th Wave Family:** The j-th wave family of nonlinear systems of conservation laws[def. 7.1] is defined as the eigenvalue-eigenvector pair of the Jaccobian $\mathrm{f}'(\mathbf{U})$:

$$\{\lambda_j(\mathbf{U}), \mathbf{r}_j(\mathbf{U})\} \tag{7.2}$$

Definition 7.4 [example 9.12]
Hyperbolic Nonlinear Systems of Conservation Laws:
A nonlinear scalar conservation laweq. (7.1) is *hyperbolic* if the Jaccobian[def. 16.6] $\mathrm{f}'(\mathbf{U})$ has:

① *real eigenvalues* $\iff$ spectrum $(\mathrm{f}'(\mathbf{U})) \in \mathbb{R}$:

$$\lambda\left(\mathrm{f}'(\mathbf{U})\right) = \{\lambda_1(\mathbf{U}) \leqslant \lambda_2(\mathbf{U}) \leqslant \ldots \leqslant \lambda_m(\mathbf{U})\} \in \mathbb{R}$$

② Linearly independent eigenvectors:

$$r_1(\mathbf{U}), r_2(\mathbf{U}), \ldots, r_m(\mathbf{U}) \tag{7.3}$$

Definition 7.5 [example 9.13]
Strictly Hyperbolic Non. Lin. Sys. of Conservation Laws: Is a hyperbolic Nonlinear Systems of Conservation Laws with distinct *real eigenvalues*:

$$\lambda\left(\mathrm{f}'(\mathbf{U})\right) = \{\lambda_1(\mathbf{U}) < \lambda_2(\mathbf{U}) < \ldots < \lambda_m(\mathbf{U})\} \in \mathbb{R}$$

Corollary 7.1 Diagonalizability: A Hyperbolic Nonlinear System of Conservation laws has a diagonalizable Jacobian matrix $\mathrm{f}'(\mathbf{U})$:

$$\mathrm{f}'(\mathbf{U}) = \mathbf{R}(\mathbf{U})\,\Lambda(\mathbf{U})\mathbf{R}(\mathbf{U})^{-1} \tag{7.4}$$

$$\Lambda(\mathbf{U}) := \operatorname{diag}(\lambda_1(\mathbf{U}), \ldots, \lambda_m(\mathbf{U}))$$

$$\mathbf{R}(\mathbf{U}) := \left[\mathbf{r}_1(\mathbf{U})\cdots\cdots\mathbf{r}_m(\mathbf{U})\right]$$

Definition 7.6 [example 9.12]
Genuinely Nonlinear Wave Family: A *hyperbolic systems*[def. 7.4] j^{th}*-wave family* is *genuinely nonlinear* iff:

$$\nabla\lambda_j(\mathbf{U})\cdot\mathbf{r}_j(\mathbf{U}) \neq 0 \qquad \forall\mathbf{U}\in\mathcal{U}, \qquad j \in \{1,\ldots,m\} \tag{7.5}$$

Explanation 7.1. *Corresponds to a notion of convexity.*

Definition 7.7
Linearly Degenerat Wave Family: A *hyperbolic systems*[def. 7.4] j^{th}*-wave family* is *linearly degenerated* iff:

$$\nabla\lambda_j(\mathbf{U})\cdot\mathbf{r}_j(\mathbf{U}) = 0 \qquad \forall\mathbf{U}\in\mathcal{U}, \qquad j \in \{1,\ldots,m\} \tag{7.6}$$

Explanation 7.2. *Linearly to a notion of convexity.*

1. Weak Solutions

Definition 7.8 [proof 8.41]
Weak Solution for 7.1:
$\mathbf{U} \in L^\infty(\mathbb{R}\times\mathbb{R}_+)$ is a weak solution of [def. 7.1] iff:

$$\int_{\mathbb{R}_+}\int_{\mathbb{R}} \mathbf{U}\partial_t\phi + \mathrm{f}(\mathbf{U})\,\partial_x\phi + \int_{\mathbb{R}} \mathbf{U}_0(x)\phi(x,0)\,\mathrm{d}x = 0 \tag{7.7}$$

$$\forall\phi \in \mathcal{C}_c^\infty(\mathbb{R}\times[0,\infty))$$

1.1. The Rankine-Hugoniot Condition

Definition 7.9 [proof 8.7]
Rankine-Hugoniot Condition: Is a condition on the *shock-speed* $s(t) = \gamma'(t)$ of a shock[def. 2.3] i.e. how fast the shock-wave travels:

$$s(t)\left(\mathbf{U}^+(t) - \mathbf{U}^-(t)\right) = \mathrm{f}\left(\mathbf{U}^+(t)\right) - \mathrm{f}\left(\mathbf{U}^-(t)\right) \tag{7.8}$$

$$\mathbf{U}^+ = \lim_{\mathbf{x}\to\gamma^+(t)} \mathbf{U}(\mathbf{x},t) \qquad \mathbf{U}^- = \lim_{\mathbf{x}\to\gamma^-(t)} \mathbf{U}(\mathbf{x},t)$$

Corollary 7.2 Unknowns vs. Equations:
- **Unknown's:** $\mathbf{U}^+, \mathbf{U}^- \in \mathbb{R}^m$, $s(t) \in \mathbb{R} \implies 2m+1$
- **Equations:** $\mathrm{f}(\mathbf{U}) \in \mathbb{R}^m \iff$ Equation (7.8)$\in \mathbb{R}^m \implies m$

Corollary 7.3 Relationship to Weak Solutions:
If $\mathbf{U}$ is a $\mathcal{C}^1_{pw}$ function with only jump-type discontinuities, the following statemnts are equivalent:
- $\mathbf{U}$ is a weak solution[def. 7.8] of the conservation law[def. 7.1].
- $\mathbf{U}$ is a classical solution whenever it is $\mathcal{C}^1$, and satisfies the Rankine-Hugoniot condition[def. 7.9] across every discontinuity $\mathbf{x} \to \gamma(t)$.

2. Simple Solutions

Definition 7.10
Riemann Problem for Sys. of Non-linear Cons. Laws:

$$\partial_t\mathbf{U} + \partial_x\mathrm{f}(\mathbf{U}) = \mathbf{0}$$
$$\mathbf{U}(x,0) = \mathbf{U}_0(x) = \begin{cases} U_R & \text{if } x > 0 \\ U_L & \text{if } x < 0 \end{cases} \tag{7.9}$$

Recall

For Riemann problems of scalar conservation laws we obtain different solutions:
① Shock Solutions[def. 2.5]
② Rarefaction Solutions[def. 2.7]
we now study solutions of non-linear systems of conservation laws eq. (7.36).

Definition 7.11 [proof 8.42]
Eigenvalue Problem for Non-lin. sys. of cons. laws: Is the problem we need to solve in order to find solutions to non-linear systems of conservation laws[def. 7.1]:

$$\mathrm{f}'(\mathbf{v}(\xi))\,\mathbf{v}'(\xi) = \xi\mathbf{v}'(\xi) \qquad \begin{aligned}\mathbf{v}'(\xi) &= \mathbf{r}_j(\mathbf{v}(\xi)) \\ \xi &= \lambda_j(\mathbf{v}(\xi))\end{aligned} \qquad j \in \{1,\ldots,m\} \tag{7.10}$$

Definition 7.12 [proof 8.43]
Simple ODE: Is the shifted problem eq. (8.20) with initial conditions at zero:

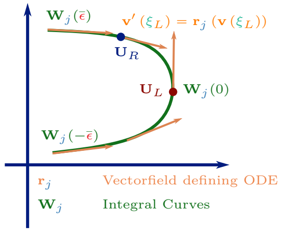
$$\begin{aligned}\mathbf{W}'(\epsilon) &= \mathbf{r}_j(\mathbf{W}(\epsilon)) \\ \mathbf{W}_j(0) &= \mathbf{U}_L\end{aligned} \qquad \epsilon = \xi - \lambda_j(\mathbf{U}_L) \tag{7.11}$$

Note: Piccard-Lindeloef Theorem

Recall from analysis If $\mathbf{r}_p(\mathbf{W}_p(t))$ is Lipschitz continuous[def. 15.21] then eq. (7.11) has a solution for $\epsilon \in [0-\bar{\epsilon}, 0+\bar{\epsilon}]$.

Explanation 7.3 (Integral Curves).

The solution of equation eq. (7.11) is given by integral curves that are tangent to the eigenvectors $\mathbf{r}_p(\mathbf{W}_p(t))$ *of the wave families.*



$\mathbf{W}_j(\bar{\epsilon})$
$\mathbf{v}'(\xi_L) = \mathbf{r}_j(\mathbf{v}(\xi_L))$
$\mathbf{U}_R$
$\mathbf{U}_L$
$\mathbf{W}_j(0)$
$\mathbf{W}_j(-\bar{\epsilon})$
$\mathbf{r}_j$ Vectorfield defining ODE
$\mathbf{W}_j$ Integral Curves

2.1. Contact Discontinuities

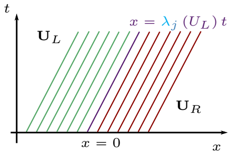
Lemma 7.1 Existence Contact Discontinuity:
Let the j-th wave family[def. 7.3] be *linear degenerate*[def. 7.7] and let $\mathbf{U}_L \in \mathcal{U}$. Then by the Piccard-Lindeloef Theorem?? there exists an *integral curve* solving eq. (7.11):

$$\mathcal{C}_j(\mathbf{U}_L) = \left\{\mathbf{W}_j\left(\epsilon^*\right) \in \mathbb{R}^m : \epsilon^* \in [-\bar{\epsilon}, \bar{\epsilon})\right\} \tag{7.12}$$

if $\mathbf{U}_R \in \mathcal{C}_j(\mathbf{U}_L)$ then there exists a *contact discontinuity solution*[def. 7.13] $\mathbf{U}$ to the Riemann problem eq. (7.36).

Definition 7.13 [proof 8.44]
Contact Discontinuity Solution:
If lemma 7.1 is satisfied then the solution of eq. (7.36) is given by:

$$\mathbf{U}(x,t) = \begin{cases} \mathbf{U}_L & \frac{x}{t} < \lambda_j(\mathbf{U}_L) \\ \mathbf{U}_R & \lambda_j(\mathbf{U}_R) < \frac{x}{t} \end{cases} \tag{7.13}$$



Explanation 7.4. *Appear in gas genomics when a with a discontinuity in mass density but not in the pressure or velocity, in comparison to real shocks, which move faster than the gas itself due to a discontinuity in pressure.*

Definition 7.14 [proof 8.45]
Rankine-Hugoniot Condition:
A contact discontinuity solution[def. 7.13] fulfills the Rankine-Hugoniot Condition:

$$f(\mathbf{U}_R) - f(\mathbf{U}_L) = s(\mathbf{U}_R - \mathbf{U}_L) \qquad s := \lambda_j(\mathbf{U}_R) = \lambda_j(\mathbf{U}_L) \tag{7.14}$$

2.2. Rarefactions

Lemma 7.2 Existence Rarefaction Solution:
Let the j-th wave family[def. 7.3] be *genuinely nonlinear*[def. 7.6] and let $\mathbf{U}_L \in \mathcal{U}$. Then by the Piccard-Lindeloef Theorem?? there exists an *integral curve* solving eq. (7.11):

$$\mathcal{R}_j(\mathbf{U}_L) = \left\{\mathbf{W}_j\left(\epsilon^*\right) \in \mathbb{R}^m : \epsilon^* \in [0, \bar{\epsilon})\right\} \tag{7.15}$$

if $\mathbf{U}_R \in \mathcal{R}_j(\mathbf{U}_L)$ then there exists a rarefaction [def. 2.7],[def. 7.15] $\mathbf{U}$ to the Riemann problemeq. (7.36).

Note: Lipschitz Boundaries

We exclude $-\bar{\epsilon}$ i.e. use $[0, \bar{\epsilon})$ as integration boundaries because for the rarefaction solution we have different eigenvalues and in this case the right eigenvalue could be larger than the left eigenvalue, which wouldn't make sense:

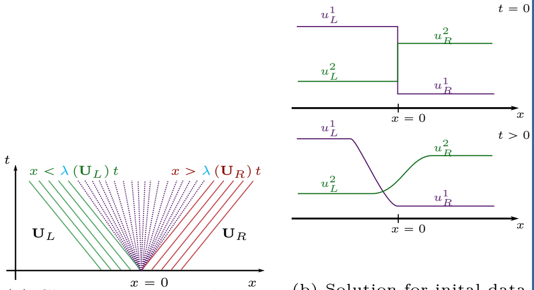
$$\lambda_j(\mathbf{U}_R) = \epsilon + \lambda_j(\mathbf{U}_L) < \lambda_j(\mathbf{U}_L) \qquad \lightning$$

Proposition 7.1 [proof 8.46]
Rarefaction and GNL wave families: Rarefaction solutions of non-linear systems of conservation laws[def. 7.1] exist if the wave families are *genuinely nonlinear*[def. 7.6]:

$$\nabla\lambda_j(\mathbf{v}(\xi))^\intercal\,\mathbf{r}_j(\mathbf{v}(\xi)) = 1 \qquad \forall j \in \{1,\ldots,m\} \tag{7.16}$$

Definition 7.15 [proof 8.46]
Rarefaction Solution:
If lemma 7.2 is satisfied then the solution of eq. (7.36) is given by:

$$\mathbf{U}(x,t) = \begin{cases} \mathbf{U}_L & \frac{x}{t} < \lambda_j(\mathbf{U}_L) \\ \mathbf{W}_j\left(\frac{x}{t} - \lambda_j(\mathbf{U}_L)\right) & \lambda_j(\mathbf{U}_L) < \frac{x}{t} < \lambda_j(\mathbf{U}_R) \\ \mathbf{U}_R & \lambda_j(\mathbf{U}_R) < \frac{x}{t} \end{cases} \tag{7.17}$$



(a) Characteristics splitting the solution in three regsion. The RH-condition eq. (7.8) is trivaly fulfilled

(b) Solution for inital data and later point; both waves travel with the same speed, one to the right the other to the left.

Note

The eigenvectors $\mathbf{r}_j(\mathbf{v}(\xi))$ for a gnl family can always be rescaled s.t. eq. (7.16) equals to 1.

2.3. Shock Waves

We have seen:
- Smooth genuinely non-linear solutions – Rarefactions
- Discontinuous linear degenerate solutions – Contact Discontinuous

but what about genuinely non-linear discountinuties – real shocks?

Definition 7.16 Hugoniot Locus:

$$\mathcal{H}(\mathbf{U}_L) = \Big\{\mathbf{U}_R \in \mathcal{U} : \exists s \in \mathbb{R} \text{ s.t.} \quad f(\mathbf{U}_R) - f(\mathbf{U}_L) = s(\mathbf{U}_R - \mathbf{U}_L)\Big\} \tag{7.18}$$

Notes

- The set of the Hugoniot Locus consist of all $\mathbf{U}_R \in \mathcal{U}$ s.t:

$$\mathbf{U}(x,t) = \begin{cases} \mathbf{U}_L & \frac{x}{t} < s \\ \mathbf{U}_R & s < \frac{x}{t} \end{cases}$$

- The set of contact discontinuities is a subset of the Hugoniot Locus i.e. $\mathcal{C}_j(\mathbf{U}_L) \in \mathcal{H}(\mathbf{U}_L)$

Lemma 7.3 : Assume a strictly hyperbolic[def. 7.5] nonlinear scalar conservation laweq. (7.1) with $\mathbf{U} \in \mathbf{U}_L$ then there exist m curves passing through $\mathbf{U}_L$:

$$\mathcal{H}(\mathbf{U}_L) = \mathcal{H}_1(\mathbf{U}_L) \cup \cdots \cup \mathcal{H}_m(\mathbf{U}_L) \tag{7.19}$$

Definition 7.17 [proof 8.47]
Shock Wave ODE:

$$\mathbf{W}_j'(0) = \mathbf{r}_j(\mathbf{U}_L) \qquad \mathbf{W}_j(0) = \mathbf{U}_L \qquad \forall j = 1,\ldots,m \tag{7.20}$$

2.4. Entrop Conditions

The entropy conditions based on the Lax-Olenek entropy condition must of course also be satisfied for non-linear scalar conservation laws.

Proposition 7.2 Viscous Approximation:
Is a parabolic *convection-diffusion equation* of the form:

$$\begin{aligned}\partial_t \mathbf{U} + \partial_x \mathrm{f}(\mathbf{U}) &= \nu \partial_{xx}\mathbf{U} \qquad \mathbf{U} : \mathbb{R}\times\mathbb{R}_+ \to \mathcal{U} \in \mathbb{R}^m\\ \mathbf{U}(x,0) &= \mathbf{U}_0(x) \qquad \mathbf{U} \in L^\infty(\mathbb{R}\times[0,T];\mathcal{U}) \\ &\qquad \mathrm{f} : \mathcal{U} \to \mathbb{R}^m \quad \text{(nonlinear)}\end{aligned} \tag{7.21}$$

Definition 7.18 Vanishing Viscosity Solution:
In the limit $\epsilon \to 0$ we recover the inviscide non-linear scalar conservation laws. Thus we can study proposition 7.2 for $\epsilon \to 0$ in order to study small scale effects.

Definition 7.19 [examples 9.14 and 9.15]
Entropy Pair (s, q)**:**
The pair (s, q) is called entropy pair, where S is any *strictly convex function*[def. 15.26]. Then the entropy pair is defined by the relation:

$$q(\mathbf{U}) = \int_{\mathbf{0}}^{\mathbf{U}} \mathrm{f}'(\gamma)\mathbf{s}'(\gamma)\,\mathrm{d}\gamma \qquad \Longrightarrow \qquad \mathbf{q}'(\mathbf{U})^\intercal = \mathbf{s}'(\mathbf{U})^\intercal\mathrm{f}'(\mathbf{U}) \tag{7.22}$$

Entropy function $\mathbf{s}$ $\quad \mathbf{s} : \mathcal{U} \subset \mathbb{R}^m \to \mathbb{R}$, strictly convex?? 20.2
Entropy flux $\mathbf{q}$ $\quad \mathbf{q} : \mathcal{U} \subset \mathbb{R}^m \to \mathbb{R}$

Note

For most physical nonlinear hyperbolic systems, there exists only one entropy, whereas for scalar conservation laws there exist a pair for any convex entropy function s.

Definition 7.20 [proof 8.48]
Entropy Condition:
Any vanishing viscosity solution[def. 2.8] u satisfies:

$$\partial_t s(\mathbf{U}) + \partial_x q(\mathbf{U}) \leqslant \mathbf{0} \tag{7.23}$$

Corollary 7.4 similar to [proof 8.11]
Kruzkov's Entropy Condition:
A solution $\mathbf{U}$ ofeq. (7.1) is a weak solution if it satisfies the Kruzkov's Entropy Condition for all entropy pairs[def. 7.19] (s, q):

$$\begin{aligned}\int_{\mathbb{R}}\int_{\mathbb{R}_+} s(\mathbf{U}(x,t))\,\phi_t(x,t) + q(\mathbf{U}(x,t))\,\phi_x\,\mathrm{d}x\,\mathrm{d}t \\ + \int_{\mathbb{R}} s(\mathbf{U}_0(x))\phi(x,0)\,\mathrm{d}x \geqslant 0\end{aligned} \tag{7.24}$$

$$\forall \phi \in \mathcal{C}^1_C(\mathbb{R}\times\mathbb{R}_+),\ \phi \geqslant 0$$

Definition 7.21 Entropy Solution:
A *weak solution*[def. 2.2] of eq. (7.1) $\mathbf{U} \in L^\infty(\mathbb{R},\mathbb{R}_+)$ is an entropy solution of the inviscide non-linear system of scalar conservation laws eq. (7.1) iff $\mathbf{U}$ satisfies the entropy condition eq. (7.24) for all entropy pairs[def. 2.9] (s, q)

2.4.1. Lax Entropy Condition

Definition 7.22 [proof 8.49],[proof 8.50]
Entropy Dissipation: States that the entropy across a discontinuity can only decrease (in a mathematical sense):

$$\left(q\left(U^+\right) - q\left(U^-\right)\right) - s\left(s\left(U^+\right) - s\left(U^-\right)\right) \leqslant 0 \tag{7.25}$$

Definition 7.23 Entropy Solution Equivalence:
Let $\mathbf{U} \in \mathcal{C}^1$ with jump discontinuities across smooth curves, then the following statements are equal:

- $\mathbf{U}$ is an entropy solution[def. 7.21] of eq. (7.1)
- $\mathbf{U}$
  - is a classical solution of eq. (7.1), whenever $\mathbf{U} \in \mathcal{C}^1$
  - fulfills the entropy dissipation equation eq. (7.25) for all entropy pair (s, q)

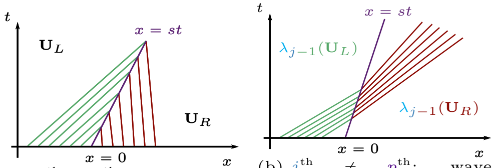
Proposition 7.3 [proof 8.49]
Contact Discontinuity Entropy:
There is no entropy dissipation across *contact discontinuities*[def. 7.13]:

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}E(\epsilon) \equiv 0 \qquad E(\epsilon) \equiv 0 \tag{7.26}$$

Proposition 7.4 [proof 8.50]
Lax Entropy Condition: For *genuinely nonlinear strictly hyperbolic systems*[cor. 6.3] of conservation laws it holds:

$$\lambda_p(\mathbf{U}_R) < s < \lambda_p(\mathbf{U}_L) \tag{7.27}$$

$$\lambda_{p-1}(\mathbf{U}_L) < s < \lambda_{p+1}(\mathbf{U}_R) \tag{7.28}$$



(a) $j^{\text{th}} = p^{\text{th}}$: wave family impingings on shock

(b) $j^{\text{th}} \neq p^{\text{th}}$: wave families go through on shock

Corollary 7.5 : Equation eq. (7.28) can be rewritten as:

$$\lambda_j(\mathbf{U}_L) < s \quad \lambda_j(\mathbf{U}_R) < s \qquad 1 \leqslant i \leqslant j-1 \tag{7.29}$$

and corresponds to characteristics that have both smaller speeds then the discontinuity.

Lemma 7.4 Lax Entropy Solution:
Let the j-th wave family be *genuinely nonlinear*[def. 7.6] and let $\mathbf{U}_L \in \mathcal{U}$. Then there exists a curve:

$$\mathcal{S}_j(\mathbf{U}_L) = \Big\{\mathbf{W}_j(\epsilon) : \epsilon \in [-\bar{\epsilon}, 0]; \tag{7.30}$$

$$f(\mathbf{W}_j(\epsilon)) - f(\mathbf{U}_L) = s(\mathbf{W}_j(\epsilon) - \mathbf{U}_L)\Big\} \tag{7.31}$$

emanating from $\mathbf{U}_L$.
If $\mathbf{U}_R \in \mathcal{S}_j(\mathbf{U}_L)$ then there exists an entropy solution figs. 8a and 8b:

$$\mathbf{U}(x,t) = \begin{cases}\mathbf{U}_L & \frac{x}{t} < s\\ \mathbf{U}_R & s < \frac{x}{t}\end{cases} \tag{7.32}$$

Explanation 7.5. *We require the negative integral curve i.e.* $-\bar{<}\epsilon \leqslant 0$ *s.t. the entropy condition is fulfiled, which leads in turn to the figures figs. 8a and 8b, depending on the wave family.*

Lemma 7.5 Entropy Solution:
Assume a *strictly hyperbolic* non-linear scalar system of conservation laws[def. 7.5] with only *genuinely non-linear* or *linear degenerate* wave families. Then $\mathbf{U}$ is an entropy solution of[def. 7.1] if and only if at every jump $\exists j \in \{1,\dots,m\}$:

- the j-th wave family is linear degenerate [def. 7.7] $\Longrightarrow$ contact discontinuit proposition 7.3 and [def. 7.13].
- the j-th wave family is genuinely nonlinear, and the *Lax entropy condition* holds eqs. (7.27) and (7.28) $\Rightarrow$ lemma 7.4.

2.5. Summary

In the previous section we considered *strictly hyperbolic*[cor. 6.3] Riemann problems for systems of scalar conservation laws[def. 7.10]. We have seen that if *each* wave family is either *linear degenerate*[def. 7.7] or *genuinely-nonlinear* eq. (7.5) then there exist m curves $\mathcal{W}_1(\mathbf{U}_L),\dots,\mathcal{W}_m(\mathbf{U}_L)$ through $\mathbf{U}_L$ and if $\mathbf{U}_R$ lies in any of these curves then the riemann problem can be solved with a simple solution:

$$\mathcal{W}(\mathbf{U}_L) = \mathcal{W}_1(\mathbf{U}_L) \cup \dots \cup \mathcal{W}_m(\mathbf{U}_L) \tag{7.33}$$

$$\mathcal{W}(\mathbf{U}_L) = \begin{cases}\mathcal{W}_j = \mathcal{C}_j(\mathbf{U}_L) & \text{if the } j\text{-th wave family is } \textit{linearly degenerate}\\ \mathcal{W}_j = \mathcal{S}_j(\mathbf{U}_L) \cup \mathcal{R}_j(\mathbf{U}_L) & \text{if the } j\text{-th wave family is } \textit{genuinely non-linear}\end{cases}$$

① If $\mathbf{U}_R \in \mathcal{R}_p(\mathbf{U}_L) \cup \mathcal{C}_j(\mathbf{U}_L)$:

- If $(\lambda_p, \mathbf{r}_p)$ genuinely nonlinear ⇒ rarefaction
- If $(\lambda_p, \mathbf{r}_p)$ linearly degenerate ⇒ contact discontinuity

② If $\mathbf{U}_R \in \mathcal{H}_p(\mathbf{U}_L)_{[-\bar{\epsilon},0]} \cup \mathcal{C}_j(\mathbf{U}_L) = \mathcal{S}_p(\mathbf{U}_L) \cup \mathcal{C}_j(\mathbf{U}_L)$:

- If $(\lambda_p, \mathbf{r}_p)$ genuinely nonlinear ⇒ shocks
- If $(\lambda_p, \mathbf{r}_p)$ linearly degenerate ⇒ contact discontinuity

Each of the curves $\mathcal{R}_p(\mathbf{U}_L), \mathcal{C}_p(\mathbf{U}_L)$ and $\mathcal{R}_p(\mathbf{U}_L)$ can be paremeterized by some function:

$$\mathbf{W}_j(\mathbf{U}_L, \epsilon) \qquad \epsilon \in \begin{cases}(-\bar{\epsilon}, \bar{\epsilon})\\ (-\bar{\epsilon}, 0]\\ [0, \bar{\epsilon})\end{cases} \qquad \bar{\epsilon}(\mathbf{U}_L) > 0$$

- *Contact Discontinuity Integral Curves:*
$$\mathcal{C}_j(\mathbf{U}_L) = \left\{\mathbf{W}_j\left(\epsilon^*\right) \in \mathbb{R}^m : \epsilon^* \in [-\bar{\epsilon}, \bar{\epsilon})\right\}$$
- *Rarefaction Integral Curves* $\mathcal{R}_p(\mathbf{U}_L)$:
$$\mathcal{R}_p(\mathbf{U}_L) = \Big\{\mathbf{W}_p(\epsilon) : \frac{\mathrm{d}\mathbf{W}_p(t)}{\mathrm{d}t} = \mathbf{r}_p(\mathbf{W}_p(\epsilon)),\ \mathbf{W}_p(0) = \mathbf{U}_L, \epsilon \in [0, \bar{\epsilon}]\Big\}$$
- *Hugoniot Locus*:
$$\mathcal{S}_p(\mathbf{U}_L) = \Big\{\mathbf{W}_j(\epsilon) : \epsilon \in [-\bar{\epsilon}, 0];\ f(\mathbf{W}_j(\epsilon)) - f(\mathbf{U}_L) = s(\mathbf{W}_j(\epsilon) - \mathbf{U}_L)\Big\}$$

For any $\mathbf{U} \in \mathbf{W}_j(\mathbf{U}_L, \epsilon) \in \mathcal{W}(\mathbf{U}_L)$, there exist then a *simple solution* $\mathbf{u}_j(\mathbf{U}_L, \epsilon; x, t)$ that is either of the formula eqs. (7.13), (7.17) and (7.31) depending whether $\mathbf{U}_R$ lies in $\mathcal{C}_j(\mathbf{U}_L)$, $\mathcal{S}_j$, $\mathcal{R}_j$.

3. General Riemann Problems

What if the wave families of the Riemann problem are neither linear degenerate or genuinely non-linear?

Finite Volume Method

Definition 7.24
FVM RP for Sys. of Non-linear Cons. Laws:

$$\begin{aligned}\partial_t \mathbf{U} + \partial_x \mathrm{f}(\mathbf{U}) &= \mathbf{0}\\ \mathbf{U}(x, t^n) &= \begin{cases}\mathbf{U}_j & \text{if } x < x_{j+1/2}\\ \mathbf{U}_{j+1} & \text{if } x > x_{j+1/2}\end{cases}\end{aligned} \tag{7.34}$$

Definition 7.25
Finite Volume Scheme: For non-linear scalar systems of conservation laws it holds:

$$\mathbf{U}_j^{n+1} = \mathbf{U}_j^n - \frac{\Delta t}{\Delta x}\left(\mathbf{F}^n_{j+1/2} - \mathbf{F}^n_{j-1/2}\right) \qquad \forall j, n \tag{7.35}$$

$$\mathbf{U}_j^0 = \frac{1}{\Delta x}\int_{x_{j+1/2}}^{x_{j+1/2}} \mathbf{U}_0(x)\,\mathrm{d}x \qquad \mathbf{F}^n_{j+1/2} = \mathrm{f}(\mathbf{u}(0))$$

4. Linearized Riemann Solvers/Roe Schemes

Definition 7.26 [proof 8.51]
Locally Linearized Riemann Problem Approximation:

$$\begin{aligned}\mathbf{U}_t + \mathbf{A}^n_{j+1/2}\mathbf{U}_x &= \mathbf{0}\\ \mathbf{U}(x, t^n) &= \begin{cases}\mathbf{U}_j & \text{if } x < x_{j+1/2}\\ \mathbf{U}_{j+1} & \text{if } x > x_{j+1/2}\end{cases}\end{aligned} \tag{7.36}$$

4.1. Properties of linear Approximations

Property 7.1 Strict Hyperbolicity:
$\mathbf{A}^n_{j+1/2} \in \mathbb{R}^{m\times m}$ should be strictly hyperbolic[cor. 6.3].

Property 7.2 Consistency:
$\mathbf{A}^n_{j+1/2} = \mathbf{A}^n_{j+1/2}\left(\mathbf{u}^n_j, \mathbf{u}^{n+1}_j\right)$ should be consistent:

$$\mathbf{A}^n_{j+1/2}(\mathbf{u}, \mathbf{u}) = \mathrm{f}'(\mathbf{u}) \tag{7.37}$$

Explanation 7.6. *If the left and right states are consistent/have the same value then our approximation should do nothing and be equal to the real flux.*

Property 7.3 [proof 8.53]
Roes Criterion: Isolated Discontinuities should be preserved exactly by our approximation:

$$\mathrm{f}\left(\mathbf{u}^n_{j+1}\right) - \mathrm{f}\left(\mathbf{u}^n_j\right) = \mathbf{A}^n_{j+1/2}\left(\mathbf{u}^n_{j+1} - \mathbf{u}^n_j\right) \tag{7.38}$$

4.2. Choices for the linearized flux

4.2.1. Arithmetic Average

Definition 7.27 Arithmetic Average:

$$\mathbf{A}^n_{j+\frac{1}{2}} = \mathrm{f}'\left(\frac{\mathbf{U}^n_j + \mathbf{U}^n_{j+1}}{2}\right) \tag{7.39}$$

Pros	Cons
• Simple • Satisfies eq. (7.37).	• Does not satisfy eq. (7.38).

4.2.2. Roe Matrices

Definition 7.28 [proof 8.52]
Roe Matrices $\mathbf{A}^n_{j+1/2}$**:**
Are matrices that satisfy the properties 7.1 to 7.3 and **??**

$$\mathbf{A}^n_{j+1/2} = \int_0^1 \mathrm{f}'\left(\mathbf{u}^n_j + \tau\left(\mathbf{u}^n_{j+1} - \mathbf{u}^n_j\right)\right)\mathrm{d}\tau \tag{7.40}$$

Problem

Equation (7.40) is not easy to calculate and in general not possible to calculate in general.

Proposition 7.5 [examples 9.16 and 9.17]
Roe Matrix:
We derive the row matrix by solving eq. (7.38):

$$[[\mathrm{f}]] = \mathbf{A}[[\mathbf{u}]] \iff \mathrm{f}\left(\mathbf{u}^n_{j+1}\right) - \mathrm{f}\left(\mathbf{u}^n_j\right) = \mathbf{A}^n_{j+1/2}\left(\mathbf{u}^n_{j+1} - \mathbf{u}^n_j\right)$$

using a clever change of variables depenindg on the underlying problem:

$$\mathbf{Z} : \mathbf{U} \mapsto \mathbf{Z}(U) \qquad \mathbf{Z} \in \mathcal{U} \subset \mathbb{R}^m \tag{7.41}$$

Explanation 7.7. *When writing down eq. (7.38) we often arrive at rational equations. By a cleaver change of variables we can transform those equations into polynomial equations, which are much easier to solve.*

Formula 7.1 Useful Identities:

$$\bar{a} := \frac{a_l + a_r}{2} \qquad [\![a]\!] := a_r - a_l \tag{7.42}$$

$$[\![ab]\!] = \bar{b}[\![a]\!] + \bar{a}[\![b]\!] \tag{7.43}$$

$$[\![a^2]\!] = 2\bar{a}[\![a]\!] \tag{7.44}$$

$$[\![a^4]\!] = 4\overline{a^2}\bar{a}[\![a]\!] \tag{7.45}$$

4.3. Schemes

4.3.1. Roe's Scheme

Definition 7.29 [proof 8.38]
Roe Flux:

$$\begin{aligned}\mathbf{F}^n_{j+1/2} =& \mathbf{A}\mathbf{U}_{j+1/2} = \frac{\mathbf{A}\left(\mathbf{U}^n_j + \mathbf{U}^n_{j+1}\right)}{2} \\ &- \frac{1}{2}\mathbf{R}^n_{j+1/2}\left|\mathbf{\Lambda}^n_{j+1/2}\right|\left(\mathbf{R}^n_{j+1/2}\right)^{-1}\left(\mathbf{U}^n_{j+1} - \mathbf{U}^n_j\right)\end{aligned} \tag{7.46}$$

$$\mathbf{R}^n_{j+1/2} = \left[\mathbf{r}^{1,n}_{j+1/2}\cdots\cdots\cdot\mathbf{r}^{m,n}_{j+1/2}\right]$$

$$\left|\mathbf{\Lambda}^n_{j+1/2}\right| = \left[\left|\lambda^{1,n}_{j+1/2}\right|\cdots\cdots\cdot\left|\lambda^{m,n}_{j+1/2}\right|\right]$$

$\mathbf{r}^{p,n}_{j+1/2}, \lambda^{p,n}_{j+1/2}$-pth eienvector pair of $\mathbf{A}^n_{j+1/2}$.

Explanation 7.8.
$\mathbf{F}^n_{j+1/2}$ = *Average Flux/Central Scheme + Numerical Diffusion*

- *Central differences in space is unconditionally unstable.*
- *Diffusion term helps to stabilize the computation*
- *pth-component of Numerical Diffusion* $\propto \left|\lambda^{p,n}_{j+1/2}\right|$
- $\left|\lambda^{p,n}_{j+1/2}\right| \propto$ *average* $\left(\lambda^{p,n}_j, \lambda^{p,n}_{j+1}\right)$ *where* $\lambda^{p,n}_j$*-pth eigenvalue of* $\mathbf{f}'(\mathbf{U}^n_j)$

Definition 7.30 Roe Scheme:

$$\textit{Equation } (7.35) + \textit{Equation } (7.46) \tag{7.47}$$

Pros	**Cons**
• Great at approximating shocks • Approximates *Linear* systems of conservation laws exactly	• Fails at transonic rarefactions • Computationally expensive as we eigenvaluedecomposition

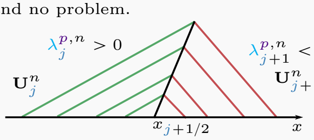
4.3.2. Harten's Entropy Fix

If p-th component of *Numerical Diffusion* in eq. (7.46) is zero that is if $\left|\lambda^{p,n}_{j+1/2}\right| \propto \text{average}\left(\lambda^{p,n}_j, \lambda^{p,n}_{j+1}\right) = 0$, then there exists nothing to stabilize, leading to instability in the p-th component.
When is the p-th component of *Numerical Diffusion* zero?
The problem arises in the p-th component if:

$$\text{sign}\left(\lambda^{p,n}_j\right) \neq \text{sign}\left(\lambda^{p,n}_{j+1}\right) \quad \text{and} \quad \left|\lambda^{p,n}_j\right| \approx \left|\lambda^{p,n}_{j+1}\right|$$

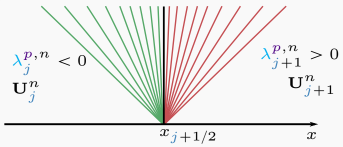
Case I: $\lambda^{p,n}_j > 0$ **and** $\lambda^{p,n}_{j+1} < 0$

By the Lax-entropy condition we obtain a shock wave. Thus information will only be taken from one side thus we have no averaging and no problem.



Case II $\lambda^{p,n}_j < 0$ **and** $\lambda^{p,n}_{j+1} > 0$

Here we cross a zero at some point. Thus the Roe scheme can fail due to averaging of positive and negative eigenvalues s.t. the diffusion term becomes zero and we end up with blow up at some point.



Definition 7.31 Roe Flux with Harten's Entropy Fix:
Makes sure that the numerical flux term does not reach zero and thus avoid blow up:

$$\begin{aligned}\mathbf{F}^n_{j+1/2} =& \mathbf{A}\mathbf{U}_{j+1/2} = \frac{\mathbf{A}\left(\mathbf{U}^n_j + \mathbf{U}^n_{j+1}\right)}{2} \\ &- \frac{1}{2}\mathbf{R}^n_{j+1/2}\left|\mathbf{\Lambda}^n_{j+1/2}\right|_\epsilon\left(\mathbf{R}^n_{j+1/2}\right)^{-1}\left(\mathbf{U}^n_{j+1} - \mathbf{U}^n_j\right)\end{aligned} \tag{7.48}$$

$$\mathbf{R}^n_{j+1/2} = \left[\mathbf{r}^{1,n}_{j+1/2}\cdots\cdots\cdot\mathbf{r}^{m,n}_{j+1/2}\right]$$

$$\left|\mathbf{\Lambda}^n_{j+1/2}\right|_\epsilon = \left[\left|\lambda^{1,n}_{j+1/2}\right|_\epsilon\cdots\cdots\cdot\left|\lambda^{m,n}_{j+1/2}\right|_\epsilon\right]$$

$\mathbf{r}^{p,n}_{j+1/2}, \lambda^{p,n}_{j+1/2}$-pth eienvector pair of $\mathbf{A}^n_{j+1/2}$.

$$|\lambda|_\epsilon = \begin{cases}|\lambda| & \text{if } |\lambda| \geqslant \epsilon \\ \frac{\lambda^2+\epsilon^2}{2\epsilon} & \text{if } |\lambda| \leqslant \epsilon\end{cases} \qquad |\cdot|_\epsilon : \mathbb{R} \mapsto \mathbb{R} \tag{7.49}$$

Pros	**Cons**
• Great at approximating shocks • Approximates *Linear* systems of conservation laws exactly	• Computationally expensive as we eigenvaluedecomposition • We do not know the right size for ϵ

Note

Rarely used in practive nowadays.

5. Central/HLL Schemes

Ami (H)arten–Peter (L)ax–Bram van (L)eer 1779-80

We have seen that the Roe schemeeq. (7.46) can very expensive. Another idea by is to approximate the m waves/discontinuities by only $2 \leqslant l \leqslant m$ waves/discontinuities and hope that they are enough to approximate our solution.

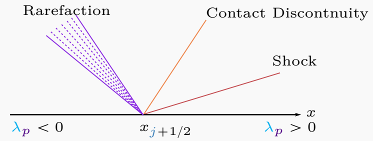


Figure 9: Example of possible waves that we might have to approximate

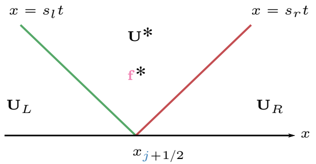
5.1. Two Wave Solver

Definition 7.32 [proof 8.25]
Central Flux:

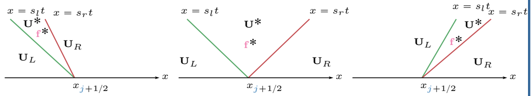
$$F^n_{j+1/2} = F\left(\mathbf{U}^n_j, \mathbf{U}^n_{j+1}\right) = \begin{cases} f\left(\mathbf{U}^n_j\right) & \text{if } s^{l,n}_{j+1/2} \geqslant 0 \\ f^*_{j+1/2} & \text{if } s^{l,n}_{j+1/2} < 0 < s^{r,n}_{j+1/2} \\ f\left(\mathbf{U}^n_{j+1}\right) & \text{if } s^{r,n}_{j+1/2} \leqslant 0\end{cases}$$

$$\begin{aligned}f^*_{j+1/2} &= \\ &= \frac{s^r_{j+1/2}f\left(\mathbf{U}^n_j\right) - s^l_{j+1/2}f\left(\mathbf{U}^n_{j+1/2}\right) + s^r_{j+1/2}s^l_{j+1/2}\left(\mathbf{U}^n_{j+1} - \mathbf{U}^n_j\right)}{s^r_{j+1/2} - s^l_{j+1/2}}\end{aligned} \tag{7.50}$$

The left $s^{l,n}_{j+1/2}$ and right $s^{r,n}_{j+1/2}$ speeds have to be specified and depend on the scheme.



Explanation 7.9. *Depending on our wave speeds we either take the exact left* $f\left(\mathbf{U}^n_j\right)$*, right* $f\left(\mathbf{U}^n_{j+1}\right)$ *flux or the approximate intermediate flux* $f^*_{j+1/2} \approx f\left(\mathbf{U}^*\right)$ *which is derived/approximated by conservation.*



Corollary 7.6 $-s^l_{j+1/2} = s^r_{j+1/2} =: s_{j+1/2}$
Symmetric Waves:
For anti-symmetric speeds we obtain:

$$f^*_{j+1/2} = \frac{f\left(\mathbf{U}^n_j\right) + f\left(\mathbf{U}^n_{j+1}\right)}{2} - \frac{s_{j+1/2}}{2}\left(\mathbf{U}^n_{j+1} - \mathbf{U}^n_j\right) \tag{7.51}$$

5.1.1. Lax-Friedrich's Scheme

Definition 7.33 Lax Friedrichs Scheme:
Chooses the wave speeds s.t. waves from neighboring Riemann problems do not interact with each other:

$$s^{l,n}_{j+1/2} = -\frac{\Delta x}{2\Delta t} \qquad s^{r,n}_{j+1/2} = \frac{2\Delta x}{\Delta t} \tag{7.52}$$

with eq. (7.51) it follows:

$$\begin{aligned}F^n_{j+1/2} &= F^{\text{LxF}}\left(\mathbf{U}^n_j, \mathbf{U}^n_{j+1}\right) \\ &= \frac{f\left(\mathbf{U}^n_j\right) + f\left(\mathbf{U}^n_{j+1}\right)}{2} - \frac{\Delta x}{2\Delta t}\left(\mathbf{U}^n_{j+1} - \mathbf{U}^n_j\right)\end{aligned} \tag{7.53}$$

Explanation 7.10. *LxF makes sure that waves do not interfere with each other, that is each wave can maximally travel a distance of* $\Delta x = \left|\frac{\Delta t}{s^l_{j+1/2}}\right|$ *i.e. to the next interface until we the next time point.*

Pros	**Cons**
• Easy to implement	• Does not take into account the local speeds • Is not the most accurate • Uses always an additional unnecessary grid point

5.1.2. Rusanov Scheme 1954

Definition 7.34
Rusanov/Local-Lax-Friedrichs Scheme:
Takes also into account the local speeds of the waves:

$$s^{r,n}_{j+1/2} = -s^{l,n}_{j+1/2} = \max\left(\max_p\left|\lambda^{n,p}_j\right|, \left|\lambda^{n,p}_{j+1}\right|\right) \tag{7.54}$$

$\lambda^{p,n}_j/\lambda^{p,n}_{j+1}$ is the p-th eigenvalue of $\mathbf{f}'\left(\mathbf{U}^n_j\right)/\mathbf{f}'\left(\mathbf{U}^n_{j+1}\right)$ with eq. (7.51) and $s^r_{j+1/2} = s_{j+1/2} = -s^l_{j+1/2}$ it follows:

$$\begin{aligned}F^n_{j+1/2} =& F^{\text{Rus}}\left(\mathbf{U}^n_j, \mathbf{U}^n_{j+1}\right) \\ =& \frac{f\left(\mathbf{U}^n_j\right) - f\left(\mathbf{U}^n_{j+1/2}\right)}{2} \\ &- \frac{1}{2}\max\left(\max_p\left|\lambda^{n,p}_j\right|, \left|\lambda^{n,p}_{j+1}\right|\right)\left(\mathbf{U}^n_{j+1} - \mathbf{U}^n_j\right)\end{aligned} \tag{7.55}$$

Pros	**Cons**
• Easy to implement • Takes into account local information	• Is still a symmetric scheme i.e. problem when all waves go in one direction/are unidirectional.

5.1.3. HLL

Definition 7.35
HLL original:
Approximates the wave cone to capture everything:

$$s^{l,n}_{j+1/2} = \min\left(\lambda^{1,n}_j, \lambda^{1,n}_{j+1}\right) \qquad s^{r,n}_{j+1/2} = \max\left(\lambda^{m,n}_j, \lambda^{m,n}_{j+1}\right) \tag{7.56}$$

Pros	**Cons**
• Takes into account local information • No longer symmetric, thus can capture unidirectional wavs	• Is still an approximation consisting just of three waves i.e. already for three waves it will no longer model the middle wave.

5.1.4. Einfeldt

Definition 7.36
Einfeldt Scheme:
Is a more refined version of the HLL scheme:

$$s^{l,n}_{j+1/2} = \min_p \min\left(\lambda^{p,n}_j, \hat{\lambda}^{p,n}_{j+1}\right) \tag{7.57}$$

$$s^{r,n}_{j+1/2} = \max_p \max\left(\lambda^{p,n}_j, \hat{\lambda}^{p,n}_{j+1}\right) \tag{7.58}$$

$\hat{\lambda}^{p,n}_{j+1}$ is the p-th eigenvalue of the Roe-matrix $\mathbf{A}^2_{j+1/2}$ (**??** 4.2.2).

Pros	**Cons**
• Takes into account local information • No longer symmetric, thus can capture unidirectional wavs	• Is still an approximation consisting just of three waves i.e. already for three waves it will no longer model the middle wave.

5.2. Three Wave Solver

For many problems such as the euler equation, the general solution may depend on three different types of solution waves-fig. 9 thus two wave solver may be not accurate to capture such solutions.

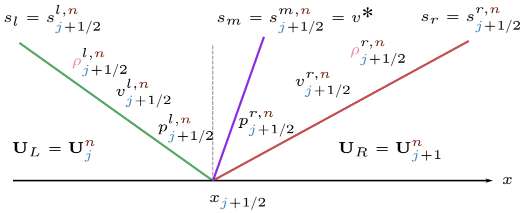
5.2.1. HLL-3/HLLC Solver

Definition 7.37 [proof 8.54]

HLL-3/HLL-C(enter):

$$F^{n}_{j+1/2} = F\left(\mathbf{U}^{n}_{j}, \mathbf{U}^{n}_{j+1}\right) = \begin{cases} F\left(\mathbf{U}^{n}_{j}\right) & \text{if} \quad 0<s^{l,n}_{j+1/2} \\ F^{l,n}_{j+1/2} & \text{if} \quad s^{l,n}_{j+1/2}\leqslant 0<s^{*,n}_{j+1/2} \\ F^{r,n}_{j+1/2} & \text{if} \quad s^{*,n}_{j+1/2}\leqslant 0<s^{r,n}_{j+1/2} \\ F\left(\mathbf{U}^{n}_{j}\right) & \text{otherwise} \end{cases} \tag{7.59}$$

$$F^{\theta,n}_{j+1/2} = F\left(\mathbf{U}^{n}_{j+k}\right) + s^{\theta,n}_{j+1/2}\left(\mathbf{U}^{\theta,n}_{j+1/2} - \mathbf{U}^{n}_{j+k}\right) \qquad k \in \{0, 1\}$$



$x(t) = s^{l,n}_{j+1/2} \cdot t \qquad x(t) = s^{m,n}_{j+1/2} \cdot t \qquad x(t) = s^{r,n}_{j+1/2} \cdot t$

$$v^{*,n}_{j+1/2} = s^{m,n}_{j+1/2}$$

$$\frac{\rho^{j}_{j+1} v^{n}_{j+1}\left(s^{r,n}_{j+1/2} - v^{n}_{j+1}\right) - \rho^{n}_{j} v^{n}_{j}\left(s^{l,n}_{j+1/2} - v^{n}_{j}\right) - \left(p^{j}_{j} - p^{n}_{j}\right)}{\rho^{n}_{j+1}\left(s^{r,n}_{j+1/2} - v^{n}_{j+1}\right)\rho^{n}_{j}\left(s^{l,n}_{j+1/2} - v^{n}_{j}\right)}$$

$$\rho^{l,n}_{j+1/2} = \frac{\rho^{n}_{j}(v^{n}_{j} - s^{l,n}_{j+1/2})}{(v^{*}_{j+1/2} - s^{l,n}_{j+1/2})} \rho^{r,n}_{j+1/2} = \frac{\rho^{n}_{j+1}(v^{n}_{j+1} - s^{r,n}_{j+1/2})}{(v^{*}_{j+1/2} - s^{r,n}_{j+1/2})}$$

$$p^{*,n}_{j+1/2} = p^{n}_{j+k} + \rho^{n}_{j+k}\left(v^{n}_{j+k} - v^{*,n}_{j+1/2}\right)\left(v^{n}_{j+k} - s^{\alpha,n}_{j+1/2}\right)$$

Note

The third component of the RH condition will in general not be satisfied and we define the flux over either of the intermediate components $F^{\theta,n}_{j+1/2} \approx F\left(\mathbf{U}^{\theta,n}_{j+1/2}\right)$

Proofs

1. Conservation Laws

Proof 8.1 Integral Surfaceexplanation 19.3:

$$\mathbf{n}\cdot\mathbf{v}=0 \Leftrightarrow a\mathbf{u}_x+b\mathbf{u}_y-c=0 \Leftrightarrow eq.\ (19.13) \tag{8.1}$$

Proof 8.2 Characteristic Equations[def. 1.7]:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}u\left(\gamma(\tau),\tau\right) &\overset{\text{C.R}}{=} u_t\left(\gamma(\tau),\tau\right)+u_x\left(\gamma(\tau),\tau\right)\frac{\mathrm{d}\gamma(\tau)}{\mathrm{d}\tau}\\ &\overset{\text{eq. (1.9)}}{=} u_t\left(\gamma(\tau),\tau\right)+u_x\left(\gamma(\tau),\tau\right)f'\left(\gamma(\tau),\tau\right)\\ &\overset{\text{eq. (1.3)}}{=} 0\end{aligned}$$

Proof 8.3 proposition 1.1:

$$\begin{aligned}u_t+f(u)u_x&=0\\ u(x,0)&=u_0(x)\end{aligned}$$

$$\textbf{ODEs}\quad\begin{cases}\dfrac{\mathrm{d}t}{\mathrm{d}r}=1\Rightarrow \mathrm{d}t=\underline{\mathrm{d}r}\\[2mm] \dfrac{\mathrm{d}\gamma}{\mathrm{d}\tau}=\underline{\dfrac{\mathrm{d}x}{\mathrm{d}r}}=\underline{\dfrac{\mathrm{d}x}{\mathrm{d}t}}=f'\left(u(x,t)\right)\overset{\text{eq. (1.10)}}{=}\text{const}\\[2mm] \dfrac{\mathrm{d}u}{\mathrm{d}r}=0\end{cases}$$

$$\textbf{I.C.}\qquad t_s(0)=0\qquad x_s(0)=s\qquad u_s(0)=u_0(s)$$

$$t_s(r)=r+C_1(s)\xrightarrow{t_s(0)=0}t=r$$

$$\frac{\mathrm{d}u}{\mathrm{d}r}=\frac{\mathrm{d}u}{\mathrm{d}t}=0\Rightarrow u_s(r)=C_2(s)\xrightarrow{u_s(0)=u_0(s)}u_s(r)=u_0(s)$$

From eq. (1.10) we know that $u(x,t)$ is constant along our characteristics and thus also $f'\left(u_0(x,t)\right)$ must be constant along them:

$$\frac{\mathrm{d}x}{\mathrm{d}r}=f'\left(u(x,t)\right)\qquad\Longrightarrow\qquad\int\mathrm{d}x=\int f(u(x,t))\,\mathrm{d}r$$

$$x(r)=C_3(s)+f'(u(x,t))r\xrightarrow{x_s(0)=s}x(r)=\underline{s}+f'(u(x,t))$$

thus we have found the general solution characteristics:

$$\lambda_s(r)=\left(r\quad s+f'(u(x,t))\quad u_0(s)\right)$$

Again, from eq. (1.10) we know that $u(x,t)$ is constant along our characteristics and thus the solution is given by:

$$u(x,t)=u_0\left(\underline{s}\right)=u_0\left(x-f'(u(x,t)t\right)$$

Proof 8.4 Conservative Form Burgers Equation [cor. 1.1]:

$$\frac{\partial}{\partial x}\frac{1}{2}u(\mathbf{x},t)^2=\frac{2}{2}u(\mathbf{x},t)_{\mathbf{x}}u(\mathbf{x},t)$$

Proof 8.5 Exploding Gradient Problemlemma 1.1:
Evolution of Spatial Gradients along Characteristics

$$\begin{aligned}u_t+uu_x&=0\\ u(x,0)&=u_0(x)\end{aligned}$$

Consider the problem for solving for the spatial gradients $v:=u_x$:

$$\frac{\partial}{\partial x}(\cdot)\Rightarrow\qquad (u_x)_t+u\,(u_x)_x+u_x\cdot u_x=0$$

$$\boxed{\begin{aligned}v_t+uv_x&=\underline{-v^2}\\ v(x,0)=v_0(x)&=u_0'(x)\end{aligned}} \tag{8.2}$$

$$\begin{aligned}&\textbf{ODEs } u\qquad \frac{\mathrm{d}x}{\mathrm{d}t}=\underline{u}\left(x(t),t\right)\\ &\qquad\qquad\quad \frac{\mathrm{d}v\left(x(t),t\right)}{\mathrm{d}t}\overset{\text{C.R.}}{=}v_t+v_x\underline{\frac{\mathrm{d}x}{\mathrm{d}t}}=\underline{v_t+v_xu}=-v^2\end{aligned}$$

$$\textbf{ODEs } v\qquad \frac{\mathrm{d}v}{\mathrm{d}t}=-v^2\qquad\qquad v(0)=v_0$$

$$-\int\frac{1}{v^2}\,\mathrm{d}v=\int\mathrm{d}t\qquad\Longrightarrow\qquad\frac{1}{v}=t+C$$

$$v(x,0)=\frac{1}{C}=v_0\qquad\Longrightarrow\qquad C=\frac{1}{v_0}$$

$$v(x,t)=\frac{1}{t+\frac{1}{v_0}}=\frac{1}{\frac{1}{v_0}(1+v_0t)}=\frac{v_0(x)}{1+v_0(x)t}=\frac{u_0'(x)}{1+u_0'(x)\underline{t}}$$

$$\text{If}\ \begin{cases}u_0'(t)>0\\ u_0'(t)<0\end{cases}\Rightarrow\ \begin{cases}v(t) & \text{well behaved}\\ v(t)\to\infty & \text{as } \underline{t}\to-\frac{1}{u_0'(x)}\end{cases}$$

Thus soon as we have a negative gradient for the initial data we will run into blow up at some time.

2. Weak Solutions

Proof 8.6 Weak Solution[def. 2.2]:
We first multiply by a test function $\phi\in\mathcal{C}_0^1\left(\mathbb{R}\times\mathbb{R}_+\right)$ and integrate over space and time:

$$\overbrace{\int_{-\infty}^{\infty}\underbrace{\int_0^{\infty}u_t\phi\,\mathrm{d}t}_{I_{1a}}\,\mathrm{d}x}^{I_1}+\overbrace{\int_0^{\infty}\underbrace{\int_{-\infty}^{\infty}f(u)_x\phi\,\mathrm{d}x}_{I_{2a}}\,\mathrm{d}t}^{I_2}=0$$

$\partial\Omega$

$$\begin{aligned}I_{1a}:\quad\int_0^{+\infty}u_t\phi\,\mathrm{d}t&\overset{\text{eq. (17.6)}}{=}u(x,\infty)\underbrace{\phi(x,\infty)}_{\equiv 0}-\underbrace{u(x,0)}_{u_0(x)}\phi(x,0)\\ &\qquad-\int_0^{\infty}u\phi_t\,\mathrm{d}t\end{aligned}$$

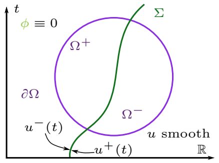
$$I_1=-\int_{-\infty}^{\infty}\int_0^{+\infty}u\phi_t\,\mathrm{d}t\,\mathrm{d}x-\int_{-\infty}^{\infty}u_0(x)\phi(x,0)\,\mathrm{d}x$$

$$\begin{aligned}I_{2a}:\quad\int_{-\infty}^{+\infty}f(u)_x\phi\,\mathrm{d}x&\overset{\text{eq. (17.6)}}{=}f\left(u(\infty,t)\right)\underbrace{\phi(\infty,t)}_{\equiv 0}\\ &\qquad-f\left(u(-\infty,t)\right)\underbrace{\phi(-\infty,t)}_{\equiv 0}-\int_{-\infty}^{\infty}f(u)\phi_x\,\mathrm{d}x\end{aligned}$$

$$I_2=\int_0^{+\infty}\int_{-\infty}^{\infty}f(u)\phi_x\,\mathrm{d}x\,\mathrm{d}t$$

$$\int\limits_{-\infty}^{\infty}\int\limits_0^{\infty}\left(u\phi_t+f(u)\phi_x\right)\mathrm{d}x\,\mathrm{d}t+\int\limits_{-\infty}^{\infty}u_0(x)\phi(x,0)\,\mathrm{d}x=0$$

Proof 8.7 Rankine-Hugoniot Condition[def. 2.4]:
Lets consider a shock-wave[def. 2.3]/discontinuity given by a curve:



$$\begin{aligned}&\Sigma=\left\{(x,t)\in\left(\mathbb{R}\times\mathbb{R}_+\right):x=\sigma(t)\right\}\\ &\Sigma=(\sigma(t),t)\ \forall t\\ &\text{s.t.}\quad u^{\pm}(t):=\lim_{h\to 0}u\left(\sigma(t)\pm ht\right)\\ &\qquad\quad u^+(t)\neq u^-(t)\end{aligned}$$

Now we choose a test function $\phi\in\mathcal{C}_C^1(\Omega)$ and $\sup(\phi)\subset\Omega$. We know that u is a *weak solution* of $\Omega\subseteq\mathbb{R}\times\mathbb{R}_+$:

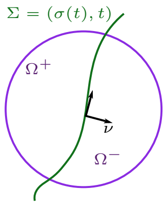
$$\int_{\Omega}\left(u\phi_t+f(u)\phi_x\right)\mathrm{d}x\,\mathrm{d}t+\int\limits_{\mathbb{R}}u_0(x)\underbrace{\phi(x,0)}_{\sup(\phi)\subset\Omega\Rightarrow\equiv 0}\mathrm{d}x=0$$

$$\int_{\Omega}\left(u\phi_t+f(u)\phi_x\right)\mathrm{d}x\,\mathrm{d}t=0$$

$$\underbrace{\int_{\Omega_-}\left(u\phi_t+f(u)\phi_x\right)\mathrm{d}x\,\mathrm{d}t}_{I_1}+\underbrace{\int_{\Omega_+}\left(u\phi_t+f(u)\phi_x\right)\mathrm{d}x\,\mathrm{d}t}_{I_2}=0$$

using I.B.P. and the fact that $\phi\equiv 0$ on $\partial\Omega$ we obtain:

$$\begin{aligned}I_1&=\int_{\Omega_-}\operatorname{grad}\phi\begin{bmatrix}f(u)\\ u\end{bmatrix}\mathrm{d}\Omega\\ &\overset{\text{eq. (17.7)}}{=}-\int_{\Omega_-}\operatorname{div}_{x,t}\begin{bmatrix}f(u)\\ u\end{bmatrix}\phi\,\mathrm{d}\Omega+\int_{\partial\Omega_-}\begin{bmatrix}f\left(u^+\right)\\ u^+\end{bmatrix}\boldsymbol{\nu}\phi\,\mathrm{d}\Sigma\\ &=-\int_{\Omega_-}\left(u_t+f(u)_x\right)\phi\,\mathrm{d}x\,\mathrm{d}t\\ &\quad+\int_{\Sigma}\left(u^+(t)\phi\nu_t^++f\left(u^+(t)\right)\phi\nu_x^+\right)\phi\,\mathrm{d}\Sigma\end{aligned}$$



Where the line measure of the “inner boundary” is given by σ and the unit normal of the line is given by:

$$\text{Tangent}=\begin{pmatrix}\sigma'(t)\\ 1\end{pmatrix}$$

$$\nu=\begin{pmatrix}\nu_x\\ \nu t\end{pmatrix}=\begin{pmatrix}\frac{-1}{\sqrt{1+\sigma'(t)}}\\ \frac{\sigma'(t)}{\sqrt{1+\sigma'(t)}}\end{pmatrix}$$

$-\nu$ is the unit normal vector of Ω^+ s.t. it follows:

$$\begin{aligned}I_1+I_2=&-\int_{\Omega_-\cup\Omega_+}\overbrace{\left(u_t+f(u)_x\right)}^{=0}\phi\,\mathrm{d}x\,\mathrm{d}t\\ &+\int_{\Sigma}\Big[\left(u^+(t)-u^+(t)\right)\nu_t\\ &+\left(f(u^+(t))-f(u^-(t))\right)\nu_x\Big]\phi(\sigma(t),t)\,\mathrm{d}\Sigma\qquad\forall\phi\end{aligned}$$

$$\Rightarrow\quad\left(u^+(t)-u^+(t)\right)\nu_t+f\left(u^+(t)\right)-f\left(u^-(t)\right)\nu_x=0$$

$$\begin{aligned}\frac{\sigma'(t)}{1+\sigma'(t)}&\left(u^+(t)-u^+(t)\right)\\ &-\frac{1}{1+\sigma'(t)}f\left(u^+(t)\right)-f\left(u^-(t)\right)=0\end{aligned}$$

$$f\left(u^+(t)\right)-f\left(u^-(t)\right)=\sigma'(t)\left(u^+(t)-u^+(t)\right)$$

Proof 8.8 Shock Wave Solution[def. 2.5]:
We know that in the absence of discontinuities the solution is given byeq. (1.14) – that is the inital data is propagated. However at the discontinuity $f'(u)$ is no longer well defined but we can resolve this issue by using the shock speed of the Rankine-Hugoniot conditioneq. (2.4) as a substitute.

Proof 8.9 Rarefaction Waves: The solution of the conservation laweq. (1.2) is invariant to the scaling of the input parameters:

$$\begin{aligned}&u(x,t)\text{ solves eq. (1.2)}\\ \Longrightarrow\quad &w(x,t):=u(\lambda x,\lambda t)\text{ solves eq. (1.2)}\qquad\lambda\neq 0\end{aligned}$$

thus it is natural to assume self-similarity – i.e. a solution $v(\xi)$ that only depends on the ratio $\xi:=x/t$:

$$u(x,t)=v\left(\frac{x}{t}\right)=v(\xi)$$

$$\begin{aligned}\xi_t&=\frac{-x}{t^2} & \xi_x&=\frac{1}{t}\\ u_t&=v'(\xi)\xi_t=v'(\xi)\frac{-x}{t^2} & u_x&=v'(\xi)\xi_x=v'(\xi)\frac{1}{t}\end{aligned}$$

$$f(u)_x=f'(u)u_x=f'\left(v(\xi)\right)v'(\xi)\xi_x=\frac{1}{t}f'\left(v(\xi)\right)v'(\xi)$$

Plug it into eq. (1.2): $\quad 0=\underline{u_t}+\underline{f(u)_x}=\underline{u_t}+\underline{f'(u)u_x}$

$$\begin{aligned}0&=\underline{v'(\xi)\frac{-x}{t^2}}+\underline{\frac{1}{t}f'\left(v(\xi)\right)v'(\xi)}\\ &=\underline{v'(\xi)\frac{-\xi}{t}}+\underline{\frac{1}{t}f'\left(v(\xi)\right)v'(\xi)}\qquad\Big|\cdot t\\ \Longrightarrow&\quad\left(f'\left(v(\xi)\right)-\xi\right)v'=0\end{aligned}$$

Thus either $v'=0$ or in the non-trivial case it follows that for convex f (invertible):

$$\begin{aligned}\left(f'\left(v(\xi)\right)-\xi\right)v'&=0\qquad\Big|/v'\\ f'\left(v(\xi)\right)&=\xi\\ v\left(\frac{x}{t}\right)&=\left(f'\right)^{-1}\left(\frac{x}{t}\right)\end{aligned} \tag{8.3}$$

From this it follows that:

$$u(x,t):=\left(f'\right)^{-1}\left(\frac{x}{t}\right)\quad\text{is a smooth solution of eq. (1.2)}$$

Proof 8.10 Entropy Condition: We first multiply eq. (2.11) by $s'(u^{\epsilon})$:

$$\begin{aligned}&S'(u^{\epsilon})u_t^{\epsilon}+S'(u^{\epsilon})f'(u^{\epsilon})u_x^{\epsilon}=\epsilon S'(u^{\epsilon})u_{xx}^{\epsilon}\\ \Rightarrow\quad&\partial_tS(u^{\epsilon})+q'(u^{\epsilon})u_x^{\epsilon}=\epsilon\underline{S'(u^{\epsilon})u_{xx}^{\epsilon}}\end{aligned}$$

$$\text{with } S(u)_{xx}=\left(S'(u)u_x\right)_x\overset{P.R.}{=}S''(u)u_x^2+\underline{S'(u)u_{xx}}$$

$$\Rightarrow\quad\partial_tS(u^{\epsilon})+\partial_xq(u^{\epsilon})=\epsilon S(u^{\epsilon})_{xx}-\epsilon\underbrace{S''(u^{\epsilon})}_{\geqslant 0\text{ convex}}\underbrace{(u_x^{\epsilon})^2}_{\geqslant 0}$$

$$\Rightarrow\quad\boxed{\partial_tS(u^{\epsilon})+\partial_xq(u^{\epsilon)}\leqslant\epsilon S(u^{\epsilon})_{xx}} \tag{8.4}$$

thus the vanishing viscosity solution $u=\lim_{\epsilon\to 0}u^{\epsilon}$ satisfieseq. (2.13).

Proof 8.11 Entropy Condition for Distributions[cor. 2.5]: Let $\phi \in \mathcal{C}^1_C\left(\mathbb{R}\times\mathbb{R}_+\right), \phi \geqslant 0$. Integrate eq. (8.4) and multiply it by ϕ:

$$\int\limits_{\mathbb{R}\times\mathbb{R}_+} S(u^\epsilon)_t\phi + q(u^\epsilon)_x\phi \,\mathrm{d}x\,\mathrm{d}t \leqslant \epsilon \int\limits_{\mathbb{R}_+} \underbrace{\int\limits_{\mathbb{R}} S(u^\epsilon)_{xx}\phi \,\mathrm{d}x\,\mathrm{d}t}_{I_c}$$

$$I_c \overset{eq.\ (17.6)}{=} \underbrace{\phi(x,t)s\left(u^\epsilon\right)_x\Big|_{-\infty}^{\infty}}_{0} - \int_{-\infty}^{\infty}\phi_x(x,t)s\left(u^\epsilon\right)_x \mathrm{d}x$$

$$\overset{eq.\ (17.6)}{=} \underbrace{\phi(x,t)_x s\left(u^\epsilon\right)\Big|_{-\infty}^{\infty}}_{0} - \int_{-\infty}^{\infty}\phi_{xx}(x,t)s\left(u^\epsilon\right)\mathrm{d}x$$

$$\overbrace{\int_{-\infty}^{\infty}\underbrace{\int_0^{\infty} S\left(u^\epsilon\right)_t \phi \,\mathrm{d}t}_{I_{1a}}\mathrm{d}x}^{I_1} + \overbrace{\int_0^{\infty}\underbrace{\int_{-\infty}^{\infty} q\left(u^\epsilon\right)_x \phi \,\mathrm{d}x}_{I_{2a}}\mathrm{d}t}^{I_2} \leqslant \int_{\mathbb{R}_+} I_c$$

$$I_{1a}: \quad \int_0^{+\infty} S\left(u^\epsilon\right)_t \phi \overset{\text{eq. }(17.6)}{=} S\left(u^\epsilon(x,\infty)\right)\underbrace{\phi(x,\infty)}_{\equiv 0}$$

$$- \underbrace{S\left(u^\epsilon(x,0)\right)}_{S(u_0(x))}\phi(x,0) - \int_0^{\infty} S\left(u^\epsilon\right)\phi_t \,\mathrm{d}t$$

$$I_1 = -\int_{\mathbb{R}\times\mathbb{R}_+} S\left(u^\epsilon\right)\phi_t\,\mathrm{d}t - \int_{-\infty}^{\infty} S\left(u_0(x)\right)\phi(x,0)\,\mathrm{d}x$$

$$I_{2a}: \quad \int_{-\infty}^{+\infty} q\left(u^\epsilon\right)_x \phi\,\mathrm{d}x \overset{\text{eq. }(17.6)}{=} q\left(u^\epsilon(\infty,t)\right)\underbrace{\phi(\infty,t)}_{\equiv 0}$$

$$- q\left(u^\epsilon(-\infty,t)\right)\underbrace{\phi(-\infty,t)}_{\equiv 0} - \int_{-\infty}^{\infty} q\left(u^\epsilon\right)\phi_x\,\mathrm{d}x$$

$$I_2 = -\int_0^{+\infty}\int_{-\infty}^{\infty} q\left(u^\epsilon\right)\phi_x\,\mathrm{d}x\,\mathrm{d}t$$

$$\Longrightarrow \lim_{\epsilon\to 0}\int\limits_{\mathbb{R}\times\mathbb{R}_+}\left(S\left(u^\epsilon\right)\phi_t + q\left(u^\epsilon\right)\phi_x\right)\mathrm{d}x\,\mathrm{d}t$$

$$+ \int\limits_{-\infty}^{\infty} S\left(u_0(x)\right)\phi(x,0)\,\mathrm{d}x$$

$$\geqslant \epsilon \underbrace{\int\limits_{\mathbb{R}\times\mathbb{R}_+}^{\infty} \phi_{xx}(x,t)s\left(u^\epsilon\right)\mathrm{d}x\,\mathrm{d}t}_{0}$$

Proof 8.12 2nd law of thermodynamicslaw 2.1: Integrate eq. (8.4) in space:

$$\int_{\mathbb{R}}\partial_t S(u^\epsilon)\,\mathrm{d}x + \int_{\mathbb{R}}\partial_x q(u^\epsilon)\,\mathrm{d}x \leqslant \epsilon\int_{\mathbb{R}} S(u^\epsilon)_{xx}\,\mathrm{d}x$$

$$\partial_t\int_{\mathbb{R}} S(u^\epsilon)\,\mathrm{d}x + \underbrace{\left[q(u^\epsilon(\infty,t)) - q(u^\epsilon(-\infty,t))\right]}_{0}$$

$$\leqslant \epsilon\underbrace{\left[s(u^\epsilon(\infty,t))_x - s(u^\epsilon(-\infty,t))_x\right]}_{0}$$

Note

$u^\epsilon(\infty,t) = u^\epsilon(-\infty,t)$ for periodic B.C. or zero otherwise.

Proof 8.13 Maximum Principle 2.1:

1. Assume eq. (1.3) attains a strict maximum at its interior $\left(x*, t*\right)$

$$u_t^\epsilon\left(x*,t*\right)\equiv 0 \qquad u_x^\epsilon\left(x*,t*\right)\equiv 0 \qquad u_{xx}^\epsilon\left(x*,t*\right) < 0$$

Now define the sum of all the termseq. (2.11), which are supposed to equal zero if u solves this equation:

$$R\left(x*,t*\right) := \underbrace{u_t^\epsilon\left(x*,t*\right)}_{=0} + f'\left(u^\epsilon\left(x*,t*\right)\right)\underbrace{u_x^\epsilon\left(x*,t*\right)}_{=0}$$

$$-\underbrace{\epsilon u_{xx}^\epsilon\left(x*,t*\right)}_{<0}$$

But $R\left(x*,t*\right) < 0$ and not 0 – a contradiction, thus the maximums cannot be inside the interior.

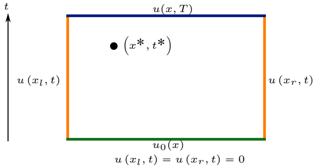
2. Now assume u attains a strict maximum at $\left(x*, T\right)$ the time horizon boundary:

$$u_x^\epsilon\left(x*,T*\right)\equiv 0 \qquad u_{xx}^\epsilon\left(x*,T*\right) < 0$$

for the time derivative we can define the backward in time derivative:

$$u_t^\epsilon(x,T) = \lim_{h\to 0}\frac{u_t^\epsilon(x,T) - u_t^\epsilon(x,T-h)}{h} > 0$$

Thus $R(x,T) > 0$ again a contradiction.
Note: as $u_t^\epsilon(x,T-h)$ is inside the interior and we already know that the interior has no maximum.



Proof 8.14 Total Variation Diminishing theorem 2.2: Lets $v^\epsilon = u_x^\epsilon$ and differentiate eq. (2.11):

$$u_{tx}^\epsilon + \left(f'\left(u^\epsilon\right)u_x^\epsilon\right)_x = \epsilon u_{xxx}^\epsilon$$

$$u_{xt}^\epsilon \overset{eq.\ (16.2)}{=} -f''\left(u^\epsilon\right)\left(u_x^\epsilon\right)^2 - f'\left(u^\epsilon\right)u_{xx}^\epsilon + \epsilon u_{xxx}^\epsilon$$

$$v_t^\epsilon = -f''\left(u^\epsilon\right)\left(v^\epsilon\right)^2 - f'\left(u^\epsilon\right)v_x^\epsilon + \epsilon v_{xx}^\epsilon \tag{8.5}$$

Now we define the test function:

$$\phi(v) = \eta(v) = |v| \qquad \eta'(v) = \operatorname{sign}(v) \qquad \eta''(v) = 2\delta_{\{v=0\}}$$

and multiply eq. (8.5) by $n'\left(v^\epsilon\right)$

$$\eta'\left(v^\epsilon\right)v_t^\epsilon = -f'\left(u^\epsilon\right)\eta'\left(v^\epsilon\right)v_x^\epsilon - f''\left(u^\epsilon\right)\eta'\left(v^\epsilon\right)\left(v^\epsilon\right)^2$$
$$+\epsilon\eta'\left(v^\epsilon\right)v_{xx}^\epsilon$$

$$\partial_t\left(v^\epsilon\right) = -f'\left(u^\epsilon\right)\partial_x\left(v^\epsilon\right) - f''\left(u^\epsilon\right)\eta'\left(v^\epsilon\right)\left(v^\epsilon\right)^2$$
$$+\epsilon\eta'\left(v^\epsilon\right)v_{xx}^\epsilon$$

$$\int\limits_{\mathbb{R}}\partial_t\left(v^\epsilon\right) = -\overbrace{\int\limits_{\mathbb{R}} f'\left(u^\epsilon\right)\partial_x\left(v^\epsilon\right)\mathrm{d}x}^{I)} - \int\limits_{\mathbb{R}} f''\left(u^\epsilon\right)\eta'\left(v^\epsilon\right)\left(v^\epsilon\right)^2\mathrm{d}x$$

$$+\epsilon\underbrace{\int\limits_{\mathbb{R}}\eta'\left(v^\epsilon\right)v_{xx}^\epsilon\,\mathrm{d}x}_{II)}$$

$$I) \overset{eq.\ (17.6)}{=} f'\left(u^\epsilon\right)\eta\left(v^\epsilon\right)\Big|_{-\infty}^{\infty} - \int_{\mathbb{R}}\left(f'\left(u^\epsilon\right)\right)_x\eta\left(u^\epsilon\right)\mathrm{d}x$$

$$= \overbrace{f'\left(u^\epsilon\right)\left|u^\epsilon\right|}^{u_x(\partial\Omega,t)=0\Rightarrow=0}\Big|_{-\infty}^{\infty} - \int_{\mathbb{R}} f''\left(u^\epsilon\right)u_x^\epsilon\eta\left(u^\epsilon\right)\mathrm{d}x$$

$$= -\int_{\mathbb{R}} f''\left(u^\epsilon\right)v^\epsilon\eta\left(u^\epsilon\right)\mathrm{d}x$$

$$II) \overset{eq.\ (17.6)}{=} \overbrace{\eta'\left(v^\epsilon\right)v_x^\epsilon\Big|_{-\infty}^{\infty}}^{=0} - \int_{\mathbb{R}}\left(\eta'\left(v^\epsilon\right)\right)_x v_x^\epsilon\,\mathrm{d}x$$

$$= -\int_{\mathbb{R}}\eta''\left(v^\epsilon\right)\left(v_x^\epsilon\right)^2\mathrm{d}x$$

$$\Longrightarrow \frac{\mathrm{d}}{\mathrm{d}t}\int\limits_{\mathbb{R}}\eta\left(v^\epsilon\right)\mathrm{d}x = +\int_{\mathbb{R}} f''\left(u^\epsilon\right)v^\epsilon\eta\left(u^\epsilon\right)\mathrm{d}x$$

$$-\int\limits_{\mathbb{R}} f''\left(u^\epsilon\right)\eta'\left(v^\epsilon\right)\left(v^\epsilon\right)^2\mathrm{d}x$$

$$-\epsilon\int_{\mathbb{R}}\eta''\left(v^\epsilon\right)\left(v_x^\epsilon\right)^2\mathrm{d}x$$

$$= +\int_{\mathbb{R}}\underbrace{\left[v^\epsilon\eta\left(u^\epsilon\right) - \eta'\left(v^\epsilon\right)\left(v^\epsilon\right)^2\right]}_{=0} f''\left(u^\epsilon\right)\mathrm{d}x$$

$$\underbrace{-2\epsilon\int_{x:v^\epsilon=0}\left(v_x^\epsilon\right)^2\mathrm{d}x}_{\leqslant 0}$$

Thus it follows that:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\limits_{\mathbb{R}}\eta\left(v^\epsilon\right)\mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}t}\int\limits_{\mathbb{R}}\left|u_x^\epsilon\right|\mathrm{d}x \leqslant 0$$

Proof 8.15 TVD in time[cor. 2.7]: From eq. (1.3) we have:

$$u_t^\epsilon = -f'\left(u^\epsilon\right)u_x^\epsilon$$

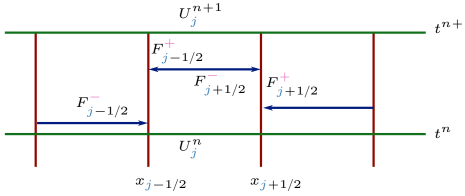
$$\Longrightarrow \left|u_t^\epsilon\right| \overset{eq.\ (20.89)}{\leqslant} \left|f'\left(u^\epsilon\right)\right|\left|u_x^\epsilon\right| \overset{\substack{eq.\ (2.20)\\ f\ \text{convex}}}{=} C\left|u_x^\epsilon\right|$$

$$\Longrightarrow \int_{\mathbb{R}}\left|u_t^\epsilon(\cdot,t)\right|\mathrm{d}x \leqslant C\int_{\mathbb{R}}\left|u_x^\epsilon(\cdot,t)\right|\mathrm{d}x$$

3. Finite Volume Methods

Proof 8.16 Integrated Boundary Fluxes[def. 3.5]:
The values of the flux at the boundary points $x_{j\pm1/2}$ may not be continuous, thus we take the values of the fluxes inside the cell over which we are integrating and proof afterwards, that in fact they are continuous:

$$\bar{F}^{n,\pm}_{j\pm\frac{1}{2}} := \int_{t_n}^{t_{n+1}} f\left(u\left(x^{\pm}_{j\pm\frac{1}{2}}\right), t\right) \mathrm{d}t \tag{8.6}$$



$$\bar{F}^{n,\pm}_{j+\frac{1}{2}} := \int_{t_n}^{t_{n+1}} f\left(u\left(x^{\pm}_{j+\frac{1}{2}}, t\right)\right) \mathrm{d}t$$

Either u

- is continuous at the boundary:
$$u\left(x^{+}_{j+\frac{1}{2}}, t\right) = u\left(x^{-}_{j+\frac{1}{2}}, t\right)$$
- or is a *stationary* (for $t = 0$) shock at the boundaries $x_{j+1/2}$ and thus has to fullfil the RH conditioneq. (2.3) with $s(t) = 0$:
$$f\left(u\left(x^{-}_{j+\frac{1}{2}}, t\right)\right) = f\left(u\left(x^{+}_{j+\frac{1}{2}}, t\right)\right)$$

thus it follow that the fluxes over the boundaries are conserved/continuous quantities:

$$\bar{F}^{n,+}_{j+\frac{1}{2}} = \int_{t_n}^{t_{n+1}} f\left(u\left(x^{+}_{j+\frac{1}{2}}, t\right)\right) \mathrm{d}t \tag{8.7}$$

$$= \int_{t_n}^{t_{n+1}} f\left(u\left(x^{-}_{j+\frac{1}{2}}, t\right)\right) \mathrm{d}t = \bar{F}^{n,-}_{j+\frac{1}{2}} = F_{j+\frac{1}{2}} \tag{8.8}$$

Proof 8.17 Finite Volume Methods[def. 3.7]:

$$\int_{t^n}^{t^{n+1}} \int_{x_{j-1/2}}^{x_{j+1/2}} \text{eq. (1.2)}$$

$$\int_{x_{j-1/2}}^{x_{j+1/2}} \int_{t^n}^{t^{n+1}} u_t \,\mathrm{d}t\,\mathrm{d}x + \int_{t^n}^{t^{n+1}} \int_{x_{j-1/2}}^{x_{j+1/2}} f_x(u)\,\mathrm{d}x\,\mathrm{d}t = 0$$

$$\overset{\text{eq. (15.14)}}{\iff} \int_{x_{j-1/2}}^{x_{j+1/2}} U\left(x, t^{n+1}\right) \mathrm{d}x - \int_{x_{j-1/2}}^{x_{j+1/2}} U\left(x, t^n\right) \mathrm{d}x$$
$$= \int_{t^n}^{t^{n+1}} f\left(U\left(x_{j+1/2}, t\right)\right) \mathrm{d}t - \int_{t^n}^{t^{n+1}} f\left(U\left(x_{j-1/2}, t\right)\right) \mathrm{d}t$$

the result follow immediately from the definitions[def's. 3.4, 3.5]

Proof 8.18 FVM Incremental Form[cor. 3.2]:

$$\text{Equation (3.10)} + \overbrace{\frac{\Delta t}{\Delta x} F\left(U_j, U_j\right) - \frac{\Delta t}{\Delta x} F\left(U_j, U_j\right)}^{=0}$$
$$\Longrightarrow U_j^n + \frac{\Delta t}{\Delta x}\left(F\left(U_j, U_j\right) - F^n_{j+1/2}\right) - \frac{\Delta t}{\Delta x}\left(F\left(U_j, U_j\right) - F^n_{j-1/2}\right)$$
$$\Longrightarrow U_j^n + \frac{\Delta t}{\Delta x}\left(F\left(U_j, U_j\right) - F^n_{j+1/2}\right) \frac{U_{j+1} - U_j}{U_{j+1} - U_j}$$
$$- \frac{\Delta t}{\Delta x}\left(F\left(U_j, U_j\right) - F^n_{j-1/2}\right) \frac{U_j - U_{j-1}}{U_j - U_{j-1}}$$

Proof 8.19 Monotonicity Preserving Schemes[cor. 3.6]: Assume $u_j^n \leqslant v_j^n, \forall j$ and H is monotone:

$$u_j^{n+1} = H\left(u^n_{j-1}, u^n_j, u^n_{j+1}\right)$$
$$u_j^{n+1} \overset{eq.\ (3.34)}{\leqslant} H\left(v^n_{j-1}, u^n_j, u^n_{j+1}\right)$$
$$u_j^{n+1} \overset{eq.\ (3.34)}{\leqslant} H\left(v^n_{j-1}, v^n_j, u^n_{j+1}\right)$$
$$u_j^{n+1} \overset{eq.\ (3.34)}{\leqslant} H\left(v^n_{j-1}, v^n_j, v^n_{j+1}\right)$$
$$\Longrightarrow \quad u_j^{n+1} \leqslant v_j^{n+1}$$

Proof 8.20 Monotone FVS eq. (3.37):

$$H(x, y, z) = y - \frac{\Delta t}{\Delta x}\left(F(x, z) - F(x, y)\right)$$
$$\frac{\partial H}{\partial x} = \frac{\Delta t}{\Delta x}\frac{\partial F}{\partial a}(x, y) \overset{\text{mon. non-dec.}}{\geqslant} 0$$
$$\frac{\partial H}{\partial z} = -\frac{\Delta t}{\Delta x}\frac{\partial F}{\partial b}(y, z) \overset{\text{mon. non-inc.}}{\geqslant} 0$$
$$\frac{\partial H}{\partial y} = 1 - \frac{\Delta t}{\Delta x}\frac{\partial F}{\partial a} - \frac{\Delta t}{\Delta x}\frac{\partial F}{\partial b} \overset{!}{\geqslant} 0$$
$$\Longrightarrow \quad \left|\frac{\partial F}{\partial a}\right| + \left|\frac{\partial F}{\partial b}\right| \leqslant \frac{\Delta t}{\Delta x} \tag{8.9}$$

Proof 8.21 Property 3.1:
Let $\bar{U}_j^n = \max\left(U^n_{j-1}, U^n_j, U^n_j\right)$ and let H be a monotone update function:

$$U_j^{n+1} = H\left(U^n_{j-1}, U^n_j, U^n_j\right) \overset{eq.\ (3.34)}{\leqslant} H\left(\bar{U}^n_j, U^n_j, U^n_j\right)$$
$$\overset{eq.\ (3.34)}{\leqslant} H\left(\bar{U}^n_j, \bar{U}^n_j, U^n_j\right)$$
$$\overset{eq.\ (3.34)}{\leqslant} H\left(\bar{U}^n_j, \bar{U}^n_j, \bar{U}^n_j\right)$$
$$\overset{eq.\ (3.22)}{=} \bar{U}^n_j$$
$$= \max\left(U^n_{j-1}, U^n_j, U^n_j\right)$$

Similar for min.

Proof 8.22 Harten's Lemma: From eq. (5.12) we can define U_{j+1}

$$u^{n+1}_{j+1} = u^n_{j+1} + C^n_{j+3/2}\left(u^n_{j+2} - u^n_{j+1}\right) - D^n_{j+1/2}\left(u^n_{j+1} - u^n_j\right)$$

From this and eq. (5.12) it follows $u^{n+1}_{j+1} - u^n_{j+1}$:

$$u^{n+1}_{j+1} - u^n_{j+1} = \left(1 - C^n_{j+1/2} - D^n_{j+1/2}\right)\left(u^n_{j+1} - u^n_j\right)$$
$$+ C^n_{j+3/2}\left(u^n_{j+2} - u^n_{j+1}\right) + D^n_{j-1/2}\left(u^n_j - u^n_{j-1}\right)$$

Assuming:

$$C^n_{j+1/2}, D^n_{j+1/2} \geqslant 0 \qquad C^n_{j+1/2} + D^n_{j+1/2} \leqslant 1 \qquad \forall j$$

with this and Equation (20.89) it follows:

$$\left|u^{n+1}_{j+1} - u^n_j\right| \leqslant \overbrace{\left(1 - C^n_{j+1/2} - D^n_{j+1/2}\right)}^{0\leqslant} \underline{\left|u^n_{j+1} - u^n_j\right|}$$
$$+ C^n_{j+3/2}\left|u^n_{j+2} - u^n_{j+1}\right| + D^n_{j-1/2}\left|u^n_j - u^n_{j-1}\right|$$

we can analogously define from this:

$$\left|u^{n+1}_{j+2} - u^{n+1}_j\right| \leqslant \left(1 - C^n_{j+3/2} - D^n_{j+3/2}\right)\left|u^n_{j+2} - u^n_{j+1}\right|$$
$$+ C^n_{j+5/2}\left|u^n_{j+3} - u^n_{j+2}\right| + D^n_{j+1/2}\underline{\left|u^n_{j+1} - u^n_j\right|}$$
$$\left|u^n_j - u^n_{j-1}\right| \leqslant \left(1 - C^n_{j-1/2} - D^n_{j-1/2}\right)\left|u^n_j - u^n_{j-1}\right|$$
$$+ C^n_{j+1/2}\underline{\left|u^n_{j+1} - u^n_j\right|} + D^n_{j-3/2}\left|u^n_{j-1} - u^n_{j-2}\right|$$

summing this three, solving for $\underline{\left|u^n_{j+1} - u^n_j\right|}$ leads to

$$\sum\left|u^{n+1}_{j+1} - u^{n+1}_j\right| \leqslant \sum\left|u^n_{j+1} - u^n_j\right|$$

Proof 8.23 Godunov Scheme??: We assume a *self-similar* solution and want to have the Riemann problem at zero thus we subtract the offset $x_{j+1/2}, t^n$:

$$U_j(x, t) = U_j\left(\frac{x - x_{j+1/2}}{t - t^n}\right) \tag{8.10}$$

Next we are only interested in the flux at the boundary $x_{j+1/2}$ s.t. we obtain:

$$F_{j+\frac{1}{2}} = \int_{t_n}^{t_{n+1}} f\left(u\left(x_{j+\frac{1}{2}}, t\right)\right) \mathrm{d}t$$
$$= \int_{t_n}^{t_{n+1}} f\left(U\left(\frac{x_{j+1/2} - x_{j+1/2}}{t - t^n}\right)\right) \mathrm{d}t = \Delta t f\left(U\left(0\right)\right)$$

where U is the solution of the standard Riemann problem:

$$u_t + f\left(u\right)_x = 0 \tag{8.11}$$

$$u(x, 0) = \begin{cases} U^n_j & \text{if } x < 0 \\ U^n_{j+1} & \text{if } x > 0 \end{cases} \tag{8.12}$$

Proof 8.24 Linearized Riemann Problem[def. 4.3]:

$$f\left(u\right) = f\left(u^n_j\right) + f'\left(\theta^n_{j+\frac{1}{2}}\right)\left(u - u^n_j\right) \qquad \theta^n_{j+\frac{1}{2}} \in \left[u^n_j, u^n_{j+1}\right]$$

$$\Longrightarrow \qquad f'\left(u\right)_x \approx f'\left(\theta^n_{j+\frac{1}{2}}\right) u_x := \hat{A}_{j+\frac{1}{2}} u_x \tag{8.13}$$

Where $\hat{A}_{j+\frac{1}{2}}\left(\theta^n_{j+\frac{1}{2}}\right) = f'\left(\theta^n_{j+\frac{1}{2}}\right)$ is a constant state around which the nonlinear flux function is linearized.

The question that remains is at which point $\left(\theta^n_{j+\frac{1}{2}}\right) \in \left[u^n_j, u^n_{j+1}\right]$ should we evaluate $\hat{A}_{j+\frac{1}{2}}$.

Proof 8.25 Central Scheme[def. 4.7]:

$$u(x, t) = \begin{cases} u^n_j & \text{if } x < s_{j+1/2}t \\ u^n_{j+1/2} & \text{if } s^l_{j+1/2}t < x < s^r_{j+1/2}t \\ u^n_{j+1} & x > s^r_{j+1/2}t \end{cases}$$

By local conservation using the RH-conditioneq. (2.3) we can determine the middle state:

$$f\left(u^n_{j+1}\right) - f^{*}_{j+1/2} = s^r_{j+1/2}\left(u^n_{j+1} - u^{*}_{j+1/2}\right) \tag{8.14}$$

$$f\left(u^n_j\right) - f^{*}_{j+1/2} = s^l_{j+1/2}\left(u^n_j - u^{*}_{j+1/2}\right) \tag{8.15}$$

$$eq.\ (8.15) * s^l_{j+1/2} + eq.\ (8.15)/s^r_{j+1/2}$$

$$\Rightarrow s^l_{j+1/2}s^R_{j+1/2}\left(u^n_{j+1} - u^n_j\right)$$
$$= s^l_{j+1/2}f\left(u^n_{j+1}\right) + s^R_{j+1/2}f\left(u^n_j\right) + \left(s^l_{j+1/2} - s^R_{j+1/2}\right) f^{*}_{j+1/2}$$

4. Higher Order Schemes

Proof 8.26
Lax-Wndroff[def. 5.3]: Based on *Cauchy-Kovalevskaya Procedure.* Given:

$$u_t + f\left(u\right)_x = 0$$
$$u(x, 0) = u_0$$

Idea: replace temporal derivatives with spatial derivatives:

$$\underline{u_t} = -f(u)_x$$

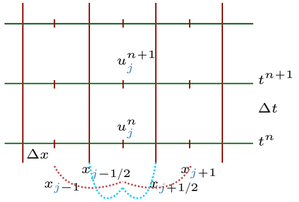
$$\underline{u_{tt}} = -f(u)_{xt} \overset{\text{C.R}}{=} -\left(f'(u)\underline{u_t}\right)_x = \left(f'(u)f(u)_x\right)$$

$\Rightarrow$ finite difference scheme $u^n_j \approx u\left(x_j, t^n\right)$ but we want to find u^{n+1}_j

Idea: use 2^{nd}-order Taylor expansion:

$$u^{n+1}_j \approx u\left(x_j, t^{n+1}\right) = u\left(x_j, t^n + \Delta t\right)$$
$$= u\left(x_j, t^n\right) + \Delta t\underline{u_t}\left(x_j, t^n\right) + \frac{\Delta t^2}{2}\underline{u_{tt}}\left(x_j, t^n\right)$$
$$+ \mathcal{O}\left(\Delta t^3\right)$$

This terms can now be approximated using central differences:



$$u^{n+1}_j \approx u^n_j - \Delta t \underline{f(u)_x\left(x_j, t^n\right)}$$
$$+ \frac{\Delta t^2}{2}\underline{\left(f'(u)f(u)_x\right)_x\left(x_j, t^n\right)}$$

$$\underline{f(u)_x\left(x_j, t^n\right)} \approx \frac{f\left(u^n_{j+1}\right) - f\left(u^n_{j-1}\right)}{2\Delta x}$$

$$\underline{\left(f'(u)f(u)_x\right)_x\left(x_j, t^n\right)} \approx$$
$$\approx \frac{\underline{f'(u)}\,\underline{f(u)_x}\left(x_{j+1/2}\right) - f'(u)f(u)_x\left(x_{j-1/2}\right)}{\Delta x}$$

$$\underline{f'(u)}\left(x_{j+1/2}\right) = a^n_{j+1/2} = f'\left(\frac{u^n_j + u^n_{j+1}}{2}\right)$$

$$\underline{f(u)}\left(x_{j+1/2}\right) \approx \frac{f\left(u^n_{j+1}\right) - f\left(u^n_j\right)}{\Delta x}$$

Note

The gitter points are chosen such that in the end we can use u_{j-1}, u_j, u_{j+1}

Proof 8.27 Sum of integrals:

Proof 8.28 Conservation and reconstruction: We calculate the flux at the interfaces x_j thus we need to recover the true value:

$$p^n_j\left(x_j\right) = u^n_j \tag{8.16}$$

Proof 8.29
FVM Evolution and Averaging Incremental Form[cor. 5.5]:
Add and subtract $F\left(U^n_{j+},u^n_{j-}\right)$ fromeq. (5.26) and divide and multiply by $U^n_j - U^n_{j-1}$:

$$U^{n+1}_j = U^n_j + \overbrace{+\frac{\Delta t}{\Delta x}\left[\frac{F\left(U^n_{j+},u^n_{j-}\right) - F\left(U^n_{j+},u^n_{j+1-}\right)}{u^n_{j+1}-u^n_j}\right]}^{C^n_{j+1/2}}\left(U^n_{j+1}-U^n_j\right) \underbrace{-\frac{\Delta t}{\Delta x}\left[\frac{F\left(U^n_{j+1+},u^n_{j+1-}\right) - F\left(U^n_{j+},u^n_{j-1-}\right)}{u^n_{j+1}-u^n_j}\right]}_{D^n_{j+1/2}}\left(U^n_j-U^n_{j-1}\right)$$

Proof 8.30 TVD FVM scheme$^{\text{lemma 5.3}}$:
We need to show that evolution and averaging eq. (5.26) is TVD i.e. fullfils hartens lemma eq. (3.31):

$$c^n_{j+1/2} = \frac{\Delta t}{\Delta x}\frac{F\left(u^n_{j+},\underline{u^n_{j-}}\right) - F\left(u^n_{j+},\underline{u^n_{j+1-}}\right)}{u^n_{j+1}-u^n_j}$$

$$\overset{\text{Lips. Cont.}}{=} \frac{\Delta t}{\Delta x}\frac{\partial F}{\partial b}\left(u^n_{j+},\cdot\right)\left(\frac{\underline{u^n_{j-}} - \underline{u^n_{j+1-}}}{u^n_{j+1}-u^n_j}\right)$$

$$:= \frac{\Delta t}{\Delta x}\frac{\partial F}{\partial b}\left(u^n_{j+},\cdot\right)\cdot(-T_1) \overset{\substack{1.T_1\geqslant 0\\ 2.eq.\ (3.41)}}{\geqslant} 0$$

$$d^n_{j-1/2} = \frac{\Delta t}{\Delta x}\frac{f\left(\underline{u^n_{j+1+}},u^n_{j+1-}\right) - f\left(\underline{u^n_{j+}},u^n_{j-1-}\right)}{u^n_{j+1}-u^n_j}$$

$$\overset{\text{Lips. Cont.}}{=} \frac{\Delta t}{\Delta x}\frac{\partial F}{\partial a}\left(\cdot,u^n_{j+1-}\right)\left(\frac{\underline{u^n_{j+1+}} - \underline{u^n_{j+}}}{u^n_{j+1}-u^n_j}\right)$$

$$:= \frac{\Delta t}{\Delta x}\frac{\partial F}{\partial a}\left(\cdot,u^n_{j+1-}\right)\cdot(T_2) \overset{\substack{1.T_2\geqslant 0\\ 2.eq.\ (3.40)}}{\geqslant} 0$$

next wee need to show that $c^n_{j+1/2} + d^n_{j+1/2} \leqslant 1$ of eq. (3.31) if fullfiled:

$$\begin{aligned} c^n_{j+1/2} + d^n_{j+1/2} &= \frac{\Delta t}{\Delta x}\left(-\frac{\partial f}{\partial b}\left(u^n_{j+},\cdot\right)\right)T_1 \\ &\quad + \frac{\Delta t}{\Delta x}\left(\frac{\partial f}{\partial a}\left(\cdot,u^n_{j+1-},\cdot\right)\right)T_2 \\ &\leqslant \frac{\Delta t}{\Delta x}\max_{a,b}\left(\left|\frac{\partial F}{\partial a}\right|,\left|\frac{\partial F}{\partial b}\right|\right)(T_1+T_2) \\ &\overset{\text{eq. (3.37)}}{\leqslant} \frac{1}{2}(T_1+T_2) \leqslant 1 \\ &\implies T_1 + T_2 \leqslant 2 \end{aligned}$$

Proof 8.31 TVD FVM REA Scheme: Fromeq. (5.10) we know:

$$\begin{aligned} u^n_{j+} &= u^n_j + \frac{\sigma^n_j}{2}\Delta x = u^n_j + \frac{\delta^n_j}{2} \\ u^n_{j-} &= u^n_j - \frac{\sigma^n_j}{2}\Delta x = u^n_j - \frac{\delta^n_j}{2} \\ T_1 &= \frac{u^n_{j+1} - \frac{\delta^n_{j-1}}{2} - u^n_j + \frac{\delta^n_j}{2}}{u^n_{j+1}-u^n_j} \\ &= 1 - \frac{1}{2}\left[\frac{\delta^n_{j+1}-\delta^n_j}{u^n_{j+1}-u^n_j}\right] \\ T_2 &= 1 + \frac{1}{2}\left[\frac{\delta^n_{j+1}-\delta^n_j}{u^n_{j+1}-u^n_j}\right] \\ &\implies T_1 + T_2 \equiv 2 \end{aligned}$$

and the rest follows from the condition that $T_1, T_2 \geqslant 0$

Proof 8.32 TVD Minmod Limiter[cor. 5.6],lemma 5.4:

$$\frac{\delta^n_j}{u_{j+1}-u^n_j} = \frac{\Delta x\sigma^n_j}{u_{j+1}-u^n_j} = \frac{\operatorname{minmod}\left(u^n_{j+1}-u^n_j,u^n_j-u^n_{j-1}\right)}{u_{j+1}-u^n_j}$$

$$\operatorname{sign}(u^n_{j+1}-u^n_j) \neq \operatorname{sign}(u^n_j-u^n_{j-1}) \implies \sigma^n_j = 0$$

$$\operatorname{sign}(u^n_{j+1}-u^n_j) = \operatorname{sign}(u^n_j-u^n_{j-1}) = \pm 1$$

$$\begin{aligned} \implies \frac{\delta^n_j}{u_{j+1}-u^n_j} &= \operatorname{minmod}\left(\frac{u_{j+1}-u^n_j}{u_{j+1}-u^n_j},\frac{u_{j+1}-u^n_j}{u_{j+1}-u^n_j}\right) \\ &= \operatorname{minmod}\Bigg(1,\underbrace{\frac{u_{j+1}-u^n_j}{u_{j+1}-u^n_j}}_{\geqslant 0}\Bigg) \leqslant 1 \end{aligned}$$

$$\implies 0 \leqslant \frac{\delta^n_j}{u_{j+1}-u^n_j} \leqslant 1 \qquad 0 \leqslant \frac{\delta^n_{j+1}}{u_{j+1}-u^n_j} \leqslant 1$$

$$\implies -1 \leqslant \frac{\delta^n_{j+1}-\delta^n_j}{u_{j+1}-u^n_j} \leqslant 1 \tag{8.17}$$

Proof 8.33 Heun's Method TVD[def. 5.16]:
From Harten's Lemmaeq. (5.12) we know that F.E. is TVD s.t.

$$\operatorname{TV}\left(U^*\right) \leqslant \operatorname{TV}\left(U^n\right)$$

$$\operatorname{TV}\left(U^{**}\right) \leqslant \operatorname{TV}\left(U^*\right)$$

$$\implies \operatorname{TV}\left(U^{**}\right) \leqslant \operatorname{TV}\left(U^*\right) \leqslant \operatorname{TV}\left(U^n\right)$$

$$\begin{aligned} \operatorname{TV}\left(U^{n+1}\right) &= \operatorname{TV}\left(\frac{U^n + U^{**}}{2}\right) \\ &\overset{\operatorname{TV}(au+bv) \leqslant a\operatorname{TV}(u) + b\operatorname{TV}(v)}{\leqslant} \frac{1}{2}\operatorname{TV}\left(U^n\right) + \frac{1}{2}\operatorname{TV}\left(U^{**}\right) \\ &\leqslant \frac{1}{2}\operatorname{TV}\left(U^n\right) + \frac{1}{2}\operatorname{TV}\left(U^n\right) = \operatorname{TV}\left(U^n\right) \\ \operatorname{TV}\left(U^{n+1}\right) &\leqslant \operatorname{TV}\left(U^n\right) \end{aligned}$$

Proof 8.34 Heuristic Heun's Method 2nd Order[def. 5.16]:
We take a linear ODE:

$$u_t = au \quad \overset{\text{exac. sol}}{\Longrightarrow} \quad u_{n+1} = u_n\mathrm{e}^{a\Delta t}$$

with the discretization $U_n := u(t_n)$ for our scheme it follows:

$$\begin{aligned} U^* &= U_n + a\Delta t U_n \\ U^{**} &= U^* + a\Delta t U^* \\ & U_n + a\Delta t U_n + a\Delta t U_n + a^2\Delta t^2 U_n \\ & U_n + 2a\Delta t U_n + a^2\Delta t^2 U_n \\ U_{n+1} &= \frac{1}{2}\left(U^n + U^{**}\right) = U_n + a\Delta t U_n + \frac{1}{2}a^2\Delta t^2 U_n \\ &= U_n\left(1 + a\Delta t + \frac{1}{2}a^2\Delta t^2\right) \end{aligned}$$

for a Taylor expansion of the exact solution it holds:

$$U_{n+1} = u_n e^{a\Delta t} = U_n\left(1 + a\Delta t + \frac{1}{2}a^2\Delta t^2 + \frac{1}{6}a^3\Delta t^3 + \ldots\right)$$

$$\implies \tau_n = \left|u_{n+1} - U_{n+1}\right| = \mathcal{O}(\Delta t^3) \implies \text{2nd order}$$

5. Systems of Conservation Laws

Proof 8.35 Linearizing Conservation Laws[cor. 6.2]:
Let $\bar{\mathbf{u}}(\mathbf{x},t) \in \mathbb{R}^m$ a solution of eq. (6.1) and define $\hat{\mathbf{u}}(\mathbf{x},t) := \mathbf{u} - \bar{\mathbf{u}}(\mathbf{x},t)$ s.t.:

$$\begin{aligned}(\mathbf{u} - \bar{\mathbf{u}}(\mathbf{x},t))_t + (f(\mathbf{u}) - f(\bar{\mathbf{u}}))_{\mathbf{x}} &= 0\\ \hat{\mathbf{u}}_t + (f(\mathbf{u}) - f(\bar{\mathbf{u}}))_{\mathbf{x}} &= 0\end{aligned}$$

$f(\mathbf{u}) - f(\bar{\mathbf{u}}$ can be approximated by a Taylor expansion:

$$\underline{f(\mathbf{u}) - f(\bar{\mathbf{u}})} = f'(\bar{\mathbf{u}})\,(\mathbf{u} - \bar{\mathbf{u}}) + \mathcal{O}(\|\mathbf{u} - \bar{\mathbf{u}}\|^2)$$

for small perturbations/step sizes δ $\mathbf{u} + \delta = \bar{\mathbf{u}}$ it holds that $\mathcal{O}(\|\mathbf{u} - \bar{\mathbf{u}}\|^2) \ll 1$:

$$\Longrightarrow \hat{\mathbf{u}}_t + \left(\underline{f'(\bar{u})\hat{u})}\right)_x =: \hat{\mathbf{u}}_t + (\mathbf{A}(\mathbf{x},t)\hat{u}))_x\ 0$$

Proof 8.36 Decoupled hyperbolic lin. Cons. Law.proposition 6.1:

$$\begin{aligned}\mathbf{U}_t + \mathbf{A}\mathbf{U}_x &= 0 &&\\ \mathbf{U}_t + \mathbf{R}\Lambda\mathbf{R}^{-1}\mathbf{U}_x &= 0 && \textit{Equation } (20.124)\\ (\mathbf{R}^{-1}\mathbf{U})_t + \mathbf{R}^{-1}\mathbf{R}\Lambda(\mathbf{R}^{-1}\mathbf{U})_x &= 0 && \text{Multiplying by } \mathbf{R}^{-1}\\ \mathbf{W}_t + \Lambda\mathbf{W}_x &= 0 && \mathbf{W} := \mathbf{R}^{-1}\mathbf{U}\end{aligned}$$

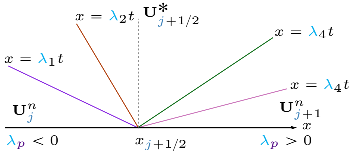
Proof 8.37 Jump Decomposition[cor. 6.6]:

$$\begin{aligned}\mathbf{U}_R - \mathbf{U}_L = \mathbf{R}(\mathbf{W}_R - \mathbf{W}_L) &= \sum_{p=1}^{m} \left(W_R^p - W_L^p\right) r_p\\ &:= \sum_{p=1}^{m} \alpha^p r_p\end{aligned}$$

Proof 8.38 Godunov Flux Systems of Cons. Laws.[def. 6.8]:
Idea we split Equation (6.12):

$$\mathbf{U}_R - \mathbf{U}_L = \sum_{p=1}^{m} \alpha^p r_p$$

into positive and negative jumps:



And then multiply by $\mathbf{A}$:

$$\mathbf{A}\mathbf{U}^n_{j+1/2} = \mathbf{A}\mathbf{U}_j + \mathbf{A} \sum_{p:\lambda_p<0}^{m} \alpha^p_{j+1/2} r_p \qquad (8.18)$$

$$\mathbf{A}\mathbf{U}^n_{j+1/2} = \mathbf{A}\mathbf{U}_{j+1} - \mathbf{A} \sum_{p:\lambda_p\geqslant 0}^{m} \alpha^p_{j+1/2} r_p \qquad (8.19)$$

$$\begin{aligned}8.18 &= \mathbf{A}\mathbf{U}_j + \sum_{p:\lambda_p<0}^{m} \alpha^p_{j+1/2}\lambda_p r_p \qquad r_p \text{ eigenv. of } \mathbf{A}\\ &= \mathbf{A}\mathbf{U}_j + \sum_{p=1}^{m} \lambda_p^- \alpha^p_{j+1/2} r_p\\ 8.19 &= \mathbf{A}\mathbf{U}_{j+1} - \sum_{p=1}^{m} \lambda_p^+ \alpha^p_{j+1/2} r_p\end{aligned}$$

$$\begin{aligned}\frac{1}{2}(8.18+8.19) &= \mathbf{A}\mathbf{U}^n_{j+1/2}\\ &= \frac{1}{2}\left(\mathbf{A}\mathbf{U}_j^n + \mathbf{A}\mathbf{U}_{j+1}^n - \sum_{p=1}^{m} (\lambda_p^+ - \lambda_p^-)\alpha^p_{j+1/2} r_p\right)\\ &= \frac{1}{2}\mathbf{A}\left(\mathbf{U}_j^n + \mathbf{U}_{j+1}^n\right) - \frac{1}{2}\sum_{p=1}^{m} \underline{|\lambda_p|}\,\underline{\alpha^p_{j+1/2}}\,\underline{r_p}\\ &= \frac{1}{2}\mathbf{A}\left(\mathbf{U}_j^n + \mathbf{U}_{j+1}^n\right) - \frac{1}{2}\underline{\mathbf{R}}\,\underline{|\Lambda|}\,\underline{\mathbf{R}^{-1}\left(\mathbf{U}_{j+1}^n - \mathbf{U}_j^n\right)}\end{aligned}$$

Notes

- $a^+ := \max(a,0)$ $\qquad a^- := \min(a,0)$
- $a = a^+ + a^-$ $\qquad a^- - a^+ = |a|$
- $|\Lambda| = \text{diag}(|\lambda_1|,\ldots,|\lambda_m|)$

Proof 8.39 Godunov TVB Property 6.1:

$$\begin{aligned}\text{TV}(\mathbf{U}^{n+1}) &= \sum_j \left\|\mathbf{U}^{n+1}_{j+1} - \mathbf{U}^{n+1}_j\right\|\\ &= \sum_j \left\|\mathbf{R}\mathbf{W}^{n+1}_{j+1} - \mathbf{R}\mathbf{W}^{n+1}_j\right\|\\ &= \sum_j \left\|\mathbf{R}(\mathbf{W}^{n+1}_{j+1} - \mathbf{W}^{n+1}_j)\right\|\\ &\leqslant \|\mathbf{R}\| \sum_j \left\|\mathbf{W}^{n+1}_{j+1} - \mathbf{W}^{n+1}_j\right\|\end{aligned}$$

we know that w^p solves the linear transport eq. s.t it holds:

$$\sum_j |w^{p,n+1}_{j+1} - w^{p,n+1}_j| \leqslant \sum_j |w^{p,n}_{j+1} - w^{p,n}_j|$$

$$\begin{aligned}\Longrightarrow \|\mathbf{R}\| \sum_j \left\|\mathbf{W}^{n+1}_{j+1} - \mathbf{W}^{n+1}_j\right\| &\leqslant \|\mathbf{R}\| \sum_j \left\|\mathbf{W}^n_{j+1} - \mathbf{W}^n_j\right\|\\ &= \|\mathbf{R}\| \sum_j \left\|\mathbf{R}^{-1}\mathbf{U}^n_{j+1} - \mathbf{R}^{-1}\mathbf{U}^n_j\right\|\\ &\leqslant \|\mathbf{R}\|\|\mathbf{R}^{-1}\| \sum_j \left\|\mathbf{U}^n_{j+1} - \mathbf{U}^n_j\right\|\end{aligned}$$

Proof 8.40 Exact Flux for conservation laws:

$$\begin{aligned}\mathbf{F}^n_{j+1/2} = \mathbf{F}\left(\mathbf{U}_j^n, \mathbf{U}_{j+1}^n\right) &= \frac{1}{\Delta t}\int_{t^n}^{t^{n+1}} \mathbf{f}\left(\mathbf{U}(x_{j+1/2}, t)\right) \mathrm{d}t\\ &= \frac{1}{\Delta t}\mathbf{A}\mathbf{U}_{j+1/2}\int_{t^n}^{t^{n+1}} \mathrm{d}t = \mathbf{A}_{j+1/2}\mathbf{U}_{j+1/2}\int_{t^n}\end{aligned}$$

6. Non-linear Systems of Conservation Laws

Proof 8.41 Weaks Solutions[def. 7.8]:
Multiply [def. 7.1] by a test function $\phi \in \mathcal{C}_0^1(\mathbb{R}\times\mathbb{R}_+)$ and integrate over space and time:

$$\int_{\mathbb{R}_+}\int_{\mathbb{R}} \phi\partial_t\mathbf{U} + \phi\partial_x\mathbf{f}(\mathbf{U})\,\mathrm{d}x\,\mathrm{d}t = 0$$

exactly as in [proof 8.6] but now with vector valued functions.

Proof 8.42 Eigenvalue Equation Conservation Laws[def. 7.11]:
The solution of the conservation law[def. 7.1] is invariant to the scaling of the input parameters:

$$\begin{aligned}&\mathbf{U}(x,t) \text{ solves eq. (7.1)}\\ \implies\quad &\mathbf{w}(x,t) := \mathbf{U}(\lambda x,\lambda t) \text{ solves eq. (7.1)} \qquad \lambda\neq 0\end{aligned}$$

thus it is natural to assume self-similarity – i.e. a solution $\mathbf{v}(\xi)$ that only depends on the ration $\xi := x/t$:

$$\mathbf{U}(x,t) = v\left(\frac{x}{t}\right) = \mathbf{v}(\xi)$$

$$\xi_t = \frac{-x}{t^2} \qquad\qquad \xi_x = \frac{1}{t}$$

$$\mathbf{U}_t = \mathbf{v}'(\xi)\xi_t = \mathbf{v}'(\xi)\frac{-x}{t^2} \qquad \mathbf{U}_x = \mathbf{v}'(\xi)\xi_x = \mathbf{v}'(\xi)\frac{1}{t}$$

$$\mathbf{f}(\mathbf{U})_x = \mathbf{f}'(\mathbf{U})\,\mathbf{U}_x = \mathbf{f}'(\mathbf{v}(\xi))\,\mathbf{v}'(\xi)\xi_x = \frac{1}{t}\mathbf{f}'(\mathbf{v}(\xi))\,\mathbf{v}'(\xi)$$

Plug it into **??**: $\qquad 0 = \underline{\partial_t\mathbf{U}} + \underline{\partial_x\mathbf{f}(\mathbf{U})}$

$$\begin{aligned}0 &= \underline{\mathbf{v}'(\xi)\frac{-x}{t^2}} + \underline{\frac{1}{t}\mathbf{f}'(\mathbf{v}(\xi))\,\mathbf{v}'(\xi)}\\ &= \underline{\mathbf{v}'(\xi)\frac{-\xi}{t}} + \underline{\frac{1}{t}\mathbf{f}'(\mathbf{v}(\xi))\,\mathbf{v}'(\xi)} \qquad\qquad \Big|\cdot t\\ \implies\quad & \mathbf{f}'(\mathbf{v}(\xi))\,\mathbf{v}'(\xi) = \xi\mathbf{v}'(\xi)\end{aligned}$$

Thus either $\mathbf{v}(\xi)' = 0$ or in the non-trivial case it follows that $\mathbf{v}(\xi)'$ is an eigenvector of the Jacobian $\mathbf{f}'(\mathbf{v}(\xi))$ with corresponding eigenvalue ξ:

$$\mathbf{f}'(\mathbf{v}(\xi))\,\mathbf{v}'(\xi) = \xi\mathbf{v}'(\xi) \qquad \begin{aligned}\mathbf{v}'(\xi) &= \mathbf{r}_j(\mathbf{v}(\xi))\\ \xi &= \lambda_j(\mathbf{v}(\xi))\end{aligned} \qquad j\in\{1,\dots,m\} \tag{8.20}$$

Proof 8.43 Simple ODE[def's. 7.12, 7.15]:
From eq. (8.20):

$$\mathbf{v}'(\xi) = \mathbf{r}_j(\mathbf{v}(\xi)) \qquad\qquad \xi = \lambda_j\left(\underline{\mathbf{v}(\xi)}\right) \tag{8.21}$$

we see that if:

$$\underline{\mathbf{v}(\xi_L)} = \mathbf{U}_L \quad\text{and}\quad \underline{\mathbf{v}(\xi_R)} = \mathbf{U}_R \quad \text{for some } \xi_L,\xi_R\in\mathbb{R}$$

then it must hold that:

$$\xi_L = \lambda_j(\mathbf{U}_L) \qquad\qquad \xi_R = \lambda_j(\mathbf{U}_R)$$

from which it follows that:

$$\mathbf{U}(x,t) = \begin{cases}\mathbf{U}_L & \frac{x}{t} < \lambda_j(\mathbf{U}_L) = \xi_L\\ \mathbf{v}_j\left(\frac{x}{t}\right) & \lambda_j(\mathbf{U}_L) < \frac{x}{t} < \lambda_j(\mathbf{U}_R)\\ \mathbf{U}_R & \xi_R = \lambda_j(\mathbf{U}_R) < \frac{x}{t}\end{cases} \tag{8.22}$$

now we need to take care of the initial condition.
We know that

$$\mathbf{v}(\xi_L) = \mathbf{U}_L \qquad\iff\qquad \xi_L = \mathbf{U}_L \tag{8.23}$$

but we do not know what $\xi_L = \lambda_j(\mathbf{U}_L)$ is.
Idea: we re-parameterize eq. (8.20) in terms of a new variable ϵ s.t. that eq. (8.23) is satisfied at $\xi = 0$ and $\mathbf{v}(\xi_L)$:

$$\epsilon := \xi_L - \lambda_j(\mathbf{U}_L) \quad \overset{\substack{\text{if } \epsilon = 0\\ \xi_L=\lambda_j(\mathbf{U}_L)}}{\Longrightarrow} \quad \mathbf{W}(\epsilon)\Big|_{\epsilon=0} = U_L$$

Proof 8.44 Contact Discontinuity[def. 7.13]:
We are looking at eq. (8.21) and differentiate $\lambda(\mathbf{W}_j(\epsilon))$:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\epsilon}\lambda(\mathbf{W}_j(\epsilon)) &= \nabla\lambda(\mathbf{W}_j(\epsilon))\,\mathbf{W}_j'(\epsilon) = \nabla\lambda(\mathbf{W}_j(\epsilon))\,\mathbf{r}_j(\mathbf{W}(\epsilon))\\ &= 0 \qquad (\text{eq. (7.6)})\end{aligned}$$

$$\implies \quad \int_0^\epsilon \nabla\lambda(\mathbf{W}_j(\epsilon))\,\mathrm{d}\epsilon = 0$$

$$\implies \quad \underline{\lambda(\mathbf{W}_j)} = \lambda(\mathbf{W}_j(0)) \overset{\text{eq. (7.11)}}{=} \lambda(\mathbf{U}_L) \qquad \forall\epsilon\in(-\bar{\epsilon},\bar{\epsilon})$$

We know that $\lambda(\mathbf{W}_j) = \lambda(\mathbf{U}_L)$, thus if $\exists\epsilon\in(-\bar{\epsilon},\bar{\epsilon})$ s.t. $\mathbf{U}_R = \mathbf{W}_j(\epsilon)$ then it holds:

$$\lambda(\mathbf{W}_j) = \lambda(\mathbf{U}_L) = \lambda(\mathbf{U}_R) = \text{const}$$

Thus the middle rarefaction solution in eq. (8.22) cannot exist.

Proof 8.45 RH condition for contact discontinuities[def. 7.14]:
We want to proof a RH condition. From [proof 8.44] we know that:

$$\lambda(\mathbf{W}_j) = \lambda(\mathbf{U}_L) \overset{\text{if }\exists\epsilon:\mathbf{U}_R=\mathbf{W}_j(\epsilon)}{=} \lambda(\mathbf{U}_R) = \text{const}$$

let us differentiate $f(\mathbf{W}_j) - \lambda_j(\mathbf{W}_j)\,\mathbf{W}_j$:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\epsilon}\left(f(\mathbf{W}_j) - \underline{\lambda_j(\mathbf{W}_j)}\mathbf{W}_j\right) &= \frac{\mathrm{d}}{\mathrm{d}\epsilon}\left(f(\mathbf{W}_j) - \lambda_j(\mathbf{W}_j)\,\mathbf{W}_j\right)\\ &= f'(\mathbf{W}_j)\,\mathbf{W}_j' - \lambda_j(\mathbf{W}_j)\,\mathbf{W}_j'\\ &= \left(f'(\mathbf{W}_j) - \lambda_j(\mathbf{W}_j)\right)\mathbf{r}_j\\ &= (\lambda(\mathbf{W}_j) - \lambda_j(\mathbf{W}_j))\,\mathbf{r}_j\\ &= (\lambda(\mathbf{W}_j) - \lambda_j(\mathbf{W}_j))\,\mathbf{r}_j = 0 \qquad \forall\epsilon\in(-\bar{\epsilon},\bar{\epsilon})\end{aligned}$$

Thus:

$$f(\mathbf{W}_j) - \underline{\lambda_j(\mathbf{W}_j)}\mathbf{W}_j = \text{const} \qquad \forall\epsilon\in(-\bar{\epsilon},\bar{\epsilon})$$

Thus it must hold that:

$$\begin{gathered}f(\mathbf{U}_L) - \lambda_j(\mathbf{U}_L)\,\mathbf{U}_L = f(\mathbf{U}_R) - \lambda_j(\mathbf{U}_R)\,\mathbf{U}_R\\ f(\mathbf{U}_R) - f(\mathbf{U}_L) = s(\mathbf{U}_R - \mathbf{U}_L)\\ s := \lambda_j(\mathbf{U}_R) = \lambda_j(\mathbf{U}_L)\end{gathered}$$

Proof 8.46
Rarefaction sol. of non-linear sys. of conser. laws$^{prop.\ 7.1}$:
Differentiate eq. (8.20) w.r.t. ξ:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\xi}\xi = \frac{\mathrm{d}}{\mathrm{d}\xi}\lambda_j(\mathbf{v}(\xi))\\ &= \nabla\lambda_j(\mathbf{v}(\xi))^\intercal\,\mathbf{v}'(\xi)\\ &= \nabla\lambda_j(\mathbf{v}(\xi))^\intercal\,\mathbf{r}_j(\mathbf{v}(\xi)) \qquad (\text{eq. (8.20)})\\ &= c = 1 \qquad (\text{eq. (7.5)} + \text{rescaling } \mathbf{r}_j)\end{aligned}$$

Thus in comparison to the contact discontinuity **we do not have** the condition that $\lambda(\mathbf{U}_L) = \lambda(\mathbf{U}_R) = \text{const}$.

Proof 8.47 Shock Wave ODE: We want to find another expression for the shock speed in eq. (7.18). Idea we use the mean value theoremtheorem 16.1:

$$M(\mathbf{U}_L,\mathbf{U}) = \int_0^1 f'(\tau\mathbf{U}_L + (\tau-1)\mathbf{U})\,\mathrm{d}\tau = \frac{f(\mathbf{U}) - f(\mathbf{U}_L)}{\mathbf{U}-\mathbf{U}_L}$$

Thus we obtain the equation:

$$\begin{aligned}\mathcal{H}(\mathbf{U}_L) = \Big\{\mathbf{U}\in\mathcal{U} : \exists s\in\mathbb{R} \text{ s.t.}\\ M(\mathbf{U}_L,\mathbf{U})(\mathbf{U}-\mathbf{U}_L) = s(\mathbf{U}-\mathbf{U}_L)\Big\}\end{aligned} \tag{8.24}$$

Thus we obtain an equation with $m+1$ unknown's $(\mathbf{U}_L, s)$, where $(\mathbf{U}-\mathbf{U}_L)$ must be an eigenvector of $M(\mathbf{U}_L,\mathbf{U})$.
By the *Implicit Function Theorem***??** theorem we know that eq. (7.18) must have m curves $\{\mathbf{W}_j\}_{j=1}^m$:

$$\begin{aligned}f(\mathbf{W}_j(\epsilon)) - f(\mathbf{U}_L) &= s(\mathbf{W}_j(\epsilon) - \mathbf{U}_L)\\ \mathbf{W}_j(0) &= \mathbf{U}_L\end{aligned} \qquad \forall j = 1,\dots,m \tag{8.25}$$

Dividing by ϵ and taking the limit leads to:

$$\frac{f(\mathbf{W}_j(\epsilon)) - f(\mathbf{U}_L)}{\epsilon} = s\frac{(\mathbf{W}_j(\epsilon) - \mathbf{U}_L)}{\epsilon}$$

$$\lim_{\epsilon\to 0} \qquad f'(\mathbf{W}_j(0))\,\mathbf{W}_j'(0) = s\mathbf{W}_j'(0)$$

$$s = \lambda_j(\mathbf{U}_L) \qquad\qquad \mathbf{W}_j'(0) = \mathbf{r}_j(\mathbf{U}_L)$$

Proof 8.48 Entropy Cond. Non-lin. Systems[def. 7.20]:
Similar to [proof 8.10] but from *stric convexity* it follows that the Hessian[def. 16.8] matrix $\mathbf{s}''(\mathbf{U})$ is positive definite[def. 20.73].

Proof 8.49
Entropy Dissipation Contact Discontinuity[def. 7.22]:
At contact discontinuities it holds:

$$\mathbf{W}_j'(\epsilon) = \mathbf{r}_j(\mathbf{W}(\epsilon)) \qquad\qquad \mathbf{W}_j(0) = \mathbf{U}_L$$

$$\lambda_j(\mathbf{W}(\epsilon)) = \lambda_j(\mathbf{U}_L) = s$$

$$E(\epsilon) := q(\mathbf{W}_j(\epsilon)) - q(\mathbf{U}_L) - \lambda_j\left(s(\mathbf{W}_j(\epsilon)) - s(\mathbf{U}_L)\right)$$

$$\begin{aligned}E(\epsilon)' =& q'(\mathbf{W}_j(\epsilon))\,\mathbf{W}_j'(\epsilon) - \lambda_j(\mathbf{U}_L)\,s'(\mathbf{W}_j(\epsilon))\,\mathbf{W}_j'(\epsilon)\\ &\mathbf{s}'(\mathbf{W}_j(\epsilon))^\intercal\,\mathbf{f}'(\mathbf{W}_j(\epsilon))\,\mathbf{W}_j'(\epsilon) - \lambda_j(\mathbf{U}_L)\,s'(\mathbf{W}_j(\epsilon))\,\mathbf{W}\\ &\mathbf{s}'(\mathbf{W}_j(\epsilon))^\intercal\left[\mathbf{f}'(\mathbf{W}_j(\epsilon))\,\mathbf{W}_j'(\epsilon) - \lambda_j(\mathbf{U}_L)\,\mathbf{W}_j'(\epsilon)\right]\\ &\mathbf{s}'(\mathbf{W}_j(\epsilon))^\intercal\Big[\underbrace{\mathbf{f}'(\mathbf{W}_j(\epsilon))\,\mathbf{r}_j(\epsilon) - \lambda_j(\mathbf{U}_L)\,\mathbf{W}_j'(\epsilon)}_{\text{eigenvalue equation} \implies \equiv 0}\Big]\end{aligned}$$

Thus it follows that:

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}E(\epsilon) \equiv 0 \qquad\implies\qquad E(\epsilon) = E(\mathbf{U}_L) \overset{E(\mathbf{U}_L)\equiv 0}{=} 0 \tag{8.26}$$

Proof 8.50
Entropy Dissipation Genuinely Nonlinear[def. 7.22]:
Consider a genuinely non-linear wave family $(\lambda_j,\mathbf{r}_j)$ and define:

$$E(\epsilon) := q(\mathbf{W}_j(\epsilon)) - q(\mathbf{U}_L) - \lambda_j\left(s(\mathbf{W}_j(\epsilon)) - s(\mathbf{U}_L)\right)$$

together with the RH condition it follows through tedious computation that:

$$E(\epsilon) < 0 \quad \text{for } \epsilon \text{ small} \qquad\iff\qquad \lambda_j(\mathbf{U}_R) < s < \lambda_j(\mathbf{U}_L)$$

for *strictly hyperbolic systems*[cor. 6.3] one can also deduce for small ϵ that:

$$\lambda_{j-1}(\mathbf{U}_L) < s < \lambda_{j+1}(\mathbf{U}_R) \tag{8.27}$$

Proof 8.51 Locally Linearized Riemann Problem[def. 7.26]:
We locally $[\mathbf{U}_j^n, \mathbf{U}_{j+1}^n]$ approximate $\mathbf{f}_x$ using Taylor:

$$\begin{aligned}\mathbf{f}(\mathbf{u}) &\overset{eq.\ (15.56)}{=} \mathbf{f}\left(\mathbf{u}_j^n\right) + \mathbf{f}'(\theta)\left(\mathbf{u} - \mathbf{u}_j^n\right) \qquad \theta\in[\mathbf{U}_j^n,\mathbf{U}_{j+1}^n]\\ \mathbf{f}(\mathbf{u})_x &= \mathbf{f}'(\theta)\,\mathbf{u}_x := \mathbf{A}\left(\mathbf{u}_j^n,\mathbf{u}_{j+1}^n\right)\mathbf{u}_x\end{aligned}$$

Proof 8.52 Roe Matrix[def. 7.28]: We use the mean value theorem eq. (16.1) to relate eq. (7.38) and the RH condition[def. 7.9]:

$$\begin{aligned}\mathbf{f}\left(\mathbf{U}_{j+1/2}^n\right) - \mathbf{f}\left(\mathbf{U}_j^n\right)\\ &= \underline{\int_0^1 \mathbf{f}'\left(\mathbf{u}_j^n + \tau\left(\mathbf{u}_{j+1}^n - \mathbf{u}_j^n\right)\right)}\left(\mathbf{U}_j^n - \mathbf{U}_{j+1}^n\right)\mathrm{d}\tau\\ \mathbf{f}\left(\mathbf{U}_{j+1/2}^n\right) - \mathbf{f}\left(\mathbf{U}_j^n\right) &= \underline{\mathbf{A}_{j+1/2}^n}\left(\mathbf{U}_j^n - \mathbf{U}_{j+1}^n\right)\end{aligned}$$

Proof 8.53 Roes Criterion – Property 7.3:
We assume that the exact solution of the non-linearized Riemann problem[def. 7.24] is given by a single discontinuity i.e. a *shock wave* or a *contact discontinuity* s.t. the exact solution is given by:

$$\mathbf{U}(x,t) = \begin{cases}\mathbf{U}_j^n & x < x_{j+1/2} + s_{j+1/2}^n(t-t^n)\\ \mathbf{U}_{j+1}^n & x > x_{j+1/2} + s_{j+1/2}^n(t-t^n)\end{cases}$$

and must satisfy the Rankine Heuginote condition**??**:

$$\mathbf{f}\left(\mathbf{U}_{j+1}^n(t)\right) - \mathbf{f}\left(\mathbf{U}_j^n(t)\right) = s_{j+1/2}^n\left(\mathbf{U}_j^n(t) - \mathbf{U}_{j+1}^n(t)\right)$$

Plugging in Roes Criterioneq. (7.38) leads to:

$$\mathbf{A}_{j+1/2}^n\left(\mathbf{u}_j^n,\mathbf{u}_j^{n+1}\right) = s_{j+1/2}^n\left(\mathbf{U}_j^n(t) - \mathbf{U}_{j+1}^n(t)\right) \tag{8.28}$$

This implies that $\left(\mathbf{u}_j^n,\mathbf{u}_j^{n+1}\right)$ is an eigenvector of the matrix $\mathbf{A}_{j+1/2}^n$ and $s_{j+1/2}^n$ is the corresponding eigenvalue. Thus in order for equation eq. (8.28) to hold we need to require:

$$\exists p\in\{1,\dots,m\} : \qquad \begin{aligned}s_{j+1/2}^n &= \lambda_{j+1/2}^{p,n}\\ \underline{\left(\mathbf{u}_j^n - \mathbf{u}_j^{n+1}\right)} &= \mathbf{r}_{j+1/2}^{p,n}\end{aligned}$$

$$\implies \quad \left(\mathbf{u}_j^n - \mathbf{u}_j^{n+1}\right) \overset{\text{eq. (6.12)}}{=} \sum_{l=1}^m \mathbf{W}_{j+1/2}^{l,n}\mathbf{r}_{j+1/2}^{l,n} \overset{!}{=} \underline{\mathbf{r}_{j+1/2}^{l,p}}$$

$$\implies \qquad\qquad \mathbf{u} \text{ is a solution}$$

Proof 8.54 HLL-3/HLLC[def. 7.37]:

1. We have seen in example 9.13 that first and third wave families are genuinely non-linear while the second wave family is linear degenerate and thus results in a contact discontinuity.
From this it follows that the pressure and the velocity are constant across the second discontinuity and that only the denity changes:
$v^{l,n}_{j+1/2} = v^{r,n}_{j+1/2} = v^{*,n}_{j+1/2} \qquad p^{l,n}_{j+1/2} = p^{r,n}_{j+1/2} = p^{*,n}_{j+1/2}$

2. Moreover from example 9.13 we also know that the speed of the second contact discontinuity is equal to is eigenvalue which is equal to the velocity:
$$s^{m,n}_{j+1/2} = v^*$$

Thus we can write the euler equations in terms of the conservative variables as:

$$\begin{aligned}\partial_t \rho + \partial_x(\rho v) &= 0\\ \partial_t(\rho v) + \partial_x\left(\rho v^2 + p\right) &= 0 \qquad (8.29)\\ \partial_t E + \partial_x\left((E+p)v\right) &= 0\end{aligned}$$

$$E = \frac{p}{\gamma - 1} + \frac{1}{2}\rho v^2 \qquad \gamma > 1: \text{ heat capacity ratio} \qquad (8.30)$$

The compressible euler equations can be written as conservation law:

$$\mathbf{U} = \begin{pmatrix}\rho\\ m\\ E\end{pmatrix} = \begin{pmatrix}\rho\\ mv\\ \frac{p}{\gamma-1} + \frac{1}{2}\rho v^2\end{pmatrix} \qquad \mathbf{f}(\mathbf{U}) = \begin{pmatrix}\rho v\\ \rho v^2 + p\\ (E+p)v\end{pmatrix}$$

$$\mathbf{U}^{\alpha,n}_{j+1/2} = \begin{pmatrix}\rho^{\alpha,n}_{j+1/2}\\ \rho^{\alpha,n}_{j+1/2} v^{*,j}_{j+1/2}\\ \frac{p^{*,n}_{j+1/2}}{\gamma-1} + \frac{1}{2}\rho^{\alpha,n}_{j+1/2}\left(v^{*,j}_{j+1/2}\right)^2\end{pmatrix} \qquad \alpha \in \{l, r\}$$

Now we use conservation/the RH condition[def. 7.9]:

$$\mathbf{F}\left(\mathbf{U}^{\theta,n}_{j+1/2}\right) - \mathbf{F}\left(\mathbf{U}^{\theta,n}_{j+k}\right) = s^{\theta,n}_{j+1/2}\left(\mathbf{U}^{\theta,n}_{j+1/2} - \mathbf{U}_{j+k}\right)$$
$$k \in \{0, 1\} \qquad (8.31)$$

We begin with the left l and right r discontinuity for the first component of the Euler equations.

$$\begin{aligned}\rho^{l,n}_{j+1/2}(v^* - s^{l,n}_{j+1/2}) &= \rho^n_j(v^n_j - s^{l,n}_{j+1/2})\\ \rho^{r,n}_{j+1/2}(v^* - s^{r,n}_{j+1/2}) &= \rho^n_{j+1}(v^n_{j+1} - s^{r,n}_{j+1/2})\end{aligned} \qquad (8.32)$$

Thus it follows:

$$\rho^{l,n}_{j+1/2} = \frac{\rho^n_j(v^n_j - s^{l,n}_{j+1/2})}{(v^*_{j+1/2} - s^{l,n}_{j+1/2})} \rho^{r,n}_{j+1/2} = \frac{\rho^n_{j+1}(v^n_{j+1} - s^{r,n}_{j+1/2})}{(v^*_{j+1/2} - s^{r,n}_{j+1/2})}$$

Next we look use either the left or right discontinuity with the second component of eq. (8.31) and use again the RH condition[def. 7.9]:

$$\begin{aligned}&\rho^{*,n}_{j+1/2}\left(v^{*,n}_{j+1/2}\right)^2 + p^{*,n}_{j+1/2} - \rho^n_j(v^n_j)^2 - p^n_j\\ &= s^{m,n}_{j+1/2}\left(\rho^{*,n}_{j+1/2}v^{*,n}_{j+1/2} - \rho^{*,n}_j v^{*,n}_j\right)\end{aligned}$$

With eq. (8.32) we can solve for $p^{*,n}_{j+1/2}$:

$$p^{*,n}_{j+1/2} = p^n_{j+k} + \rho^n_{j+k}\left(v^n_{j+k} - v^{*,n}_{j+1/2}\right)\left(v^n_{j+k} - s^{\alpha,n}_{j+1/2}\right)$$
$$\alpha \in \{l, r\} \qquad k \in \{0, 1\}$$

next we need to find a find an expression for $v^{*,n}_{j+1/2}$, we do this by using conservation over all three waves:

$$\begin{aligned}\mathbf{F}\left({}^n_{j+1}\right) - \mathbf{F}\left(\mathbf{U}^n_j\right) =&s^{r,n}_{j+1}\left(\mathbf{U}^n_{j+1/2} - \mathbf{U}^{r,n}_{j+1/2}\right)\\ &+ s^{m,n}_{j+1/2}\left(\mathbf{U}^{r,n}_{j+1/2} - \mathbf{U}^{l,n}_{j+1/2}\right)\\ &+ s^{l,n}_{j+1/2}\left(\mathbf{U}^{l,n}_{j+1/2} - \mathbf{U}^n_j\right)\end{aligned}$$

Proof 8.55: we compare the second component:

$$\begin{aligned}&\rho^n_{j+1}v^2_{j+1} + p^n_{j+1} - \cancel{\rho^{r,n}_{j+1/2}v^*_{j+1/2}} - p^{*,n}_{j+1/2}\\ &\cancel{\rho^{r,n}_{j+1/2}\left(v^{*,n}_{j+1/2}\right)^2} + p^{*,n}_{j+1/2} - \cancel{\rho^{*,n}_{j+1/2}\left(v^{*,n}_{j+1/2}\right)^2} - p^{*,n}_{j+1/2}\\ &\cancel{\rho^{l,n}_{j+1/2}\left(v^{*,n}_{j+1/2}\right)^2} + p^{*,n}_j - \rho^n_j\left(v^n_j\right)^2 - p^n_j\\ =&s^{r,n}_{j+1/2}\left(\rho^n_{j+1}v^n_{j+1} - \rho^{r,n}_{j+1/2}v^{*,n}_{j+1/2}\right)\\ &v^{*,n}_{j+1/2}\left(\rho^{r,n}_{j+1/2}v^{*,n}_{j+1/2} - \rho^{l,n}_{j+1/2}v^{*,n}_{j+1/2}\right)\\ &s^{l,n}_{j+1/2}\left(\rho^{l,n}_{j+1/2}v^{*,n}_{j+1/2} - \rho^n_{j+1}v^n_{j+1}\right)\end{aligned}$$

From this it follows:

$$\begin{aligned}&\underline{s^{r,n}_{j+1/2}\rho^n_{j+1}v^n_{j+1} - s^{r,n}_{j+1/2}\rho^{r,n}_{j+1/2}v^{*,n}_{j+1/2}}\\ &\underline{+\rho^{r,n}_{j+1/2}\left(v^{*,n}_{j+1/2}\right)^2 - \rho^{l,n}_{j+r/2}\left(v^{*,n}_{j+1/2}\right)^2}\\ &\underline{+s^{l,n}_{j+1/2}\rho^{l,n}_{j+1/2}v^{*,n}_{j+1/2} - s^{l,n}_{j+1/2}\rho^n_j v^n_j}\\ &= \rho^n_{j+1}\left(v^n_{j+1}\right)^2 + p^n_{j+1} - \rho^n_j\left(v^n_j\right)^2 - p^n_j\\ &\underline{v^{*,n}_{j+1/2}\rho^{r,n}_{j+1/2}\left(v^{*,n}_{j+1/2} - s^{r,n}_{j+1/2}\right)}\\ &\underline{-v^{*,n}_{j+1/2}\rho^{l,n}_{j+1/2}\left(v^{*,n}_{j+1/2} - s^{l,n}_{j+1/2}\right)}\\ &= \rho^n_{j+1}\left(v^n_{j+1}\right)^2 + p^n_{j+1} - \rho^n_j\left(v^n_j\right)^2 - p^n_j\\ &\underline{-s^{r,n}_{j+1/2}\rho^n_{j+1}v^n_{j+1} + s^{l,n}_{j+1/2}\rho^n_j v^n_j}\end{aligned}$$

pluggin in $\rho^{l,n}_{j+1/2}$ and $\rho^{r,n}_{j+1/2}$ on the lhs leads to:

$$v^{*,n}_{j+1/2}\left(\rho^n_{j+1}\left(v^n_{j+1} - s^{r,n}_{j+1/2}\right) - \rho^n_j\left(v^n_j - s^{l,n}_{j+1/2}\right)\right) = \text{——}\shortparallel\text{——}$$

From this it follows:

$$v^{*,n}_{j+1/2} =$$
$$\frac{\rho^j_{j+1}v^n_{j+1}\left(s^{r,n}_{j+1/2} - v^n_{j+1}\right) - \rho^n_j v^n_j\left(s^{l,n}_{j+1/2} - v^n_j\right) - \left(p^j_j - p^n_j\right)}{\rho^n_{j+1}\left(s^{r,n}_{j+1/2} - v^n_{j+1}\right)\rho^n_j\left(s^{l,n}_{j+1/2} - v^n_j\right)}$$

Examples

Example 9.1
Burgers Equation Riemann Problem:

$$\underline{u_t + uu_x} = 0$$

$$u(x,0) = u_0(x) = \begin{cases} 1 & \text{if } x < 0 \\ 0 & \text{if } x > 0 \end{cases}$$

ODEs $$\begin{cases} \dfrac{\mathrm{d}t}{\mathrm{d}r} = 1 \Rightarrow \mathrm{d}t = \underline{\mathrm{d}r} \\ \dfrac{\mathrm{d}x}{\underline{\mathrm{d}r}} = \dfrac{\mathrm{d}x}{\underline{\mathrm{d}t}} = u \\ \dfrac{\mathrm{d}u}{\mathrm{d}r} = 0 \end{cases}$$

$$\frac{\mathrm{d}u\,(x(t),t)}{\mathrm{d}t} \overset{\text{C.R.}}{=} u_t\,(x(t),t) + u_x\,(x(t),t)\,\underline{\frac{\mathrm{d}x(t)}{\mathrm{d}t}}$$
$$= \underline{u_t\,(x(t),t) + u_x\,(x(t),t)\,u} = 0$$

thus u is constant along the projectd characteristics and it holds that $u(x(t),t) = u(x(0),0) = u_0(x_0)$.

$x(t) = x(r)$.
Lets look at the inital data and the *projected characteristcs*:

$$\frac{\mathrm{d}u}{\mathrm{d}r} = \frac{\mathrm{d}u}{\mathrm{d}t} = 0 \qquad \Longrightarrow \qquad u(x,t) = \underline{C}$$

$$\underline{\frac{\mathrm{d}x(t)}{\mathrm{d}t}} = u\,(x(t),t) = \underline{C} \qquad \Longrightarrow \qquad \int_{x_0}^{x_t} \mathrm{d}x(t) = \int_0^t C\,\mathrm{d}t$$

$$x(t) = x_0 + Ct = x_0 + ut \qquad x_0 = x(t) - ut$$

thus we have found the general solution:

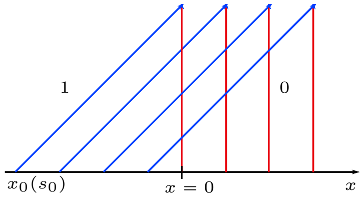
$$u(x,t) = u_0\,(x_0) = u_0\,(x - ut)$$

now lets look at the initial conditions for u_0:

$$\underline{\frac{\mathrm{d}x(t)}{\mathrm{d}t}}\bigg|_{t=0} = u\,(x(t),t)\Big|_{t=0} = u_0(x_0) = \begin{cases} 1 & \text{if } x < 0 \\ 0 & \text{if } x > 0 \end{cases}$$

$$\Longrightarrow \quad \begin{cases} \int x(t)\,\mathrm{d}x = \int 1\,\mathrm{d}t & \Rightarrow x(t) = x_0 + t \quad \text{if } x < 0 \\ \int \dfrac{\mathrm{d}x(t)}{\mathrm{d}t}\,\mathrm{d}t = \int 0\,\mathrm{d}t & \Rightarrow x(t) = x_0 \quad \text{if } x > 0 \end{cases}$$

$$\Longrightarrow \qquad x(t) = \begin{cases} x_0 + t & \text{if } x < 0 \\ x_0 & \text{if } x > 0 \end{cases}$$



For $x < 0$ we obtain the solutions:

$$u(x,t) = u_0\,(x_0) = u_0\,(x - ut) \tag{9.1}$$

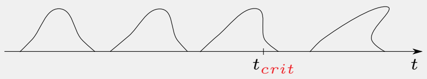
But for $x > 0$ we have intersecting project. characteristics i.e. a multivalued function that cannot be inverted and thus have no unique solution.

Note

The characteristic ODEs are ODEs and thus equations of one independent variable.
The ODE $u(r) = u(t)$ is not to be confused with our solution $u(x,t)$.

Physical Interpretation

At the singularity (shockwave) $t = t_{crit}$ the faster characteristic (taller part of a wave) will overtake the slower one (shorter part of wave), causing the wave to break.

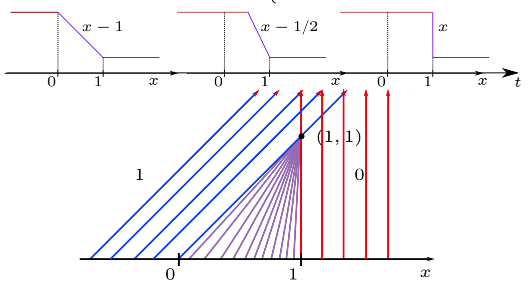


Thus their exists a physical meaning after t_{crit} where the classical solution does not exist anymore. **Question**: If there exists no strong solution is there a way to find another solution? ⇒ weak solutions.

Example 9.2
Burgers Equation Continuous Initial Data:

$$\underline{u_t + uu_x} = 0$$

$$u(x,0) = u_0(x) = \begin{cases} 1 & \text{if } x < 0 \\ 1 - x & \text{if } 0 \leqslant x \leqslant 1 \\ 0 & \text{if } x > 0 \end{cases}$$



Thus even for smooth initial data we will get intersection after the point $(1,1)$.

Example 9.3 Monotoni-city LxF[cor. 3.7]**:** Consider the LxF scheme[def. 4.8]:

$$F(a,b) = \frac{1}{2}\,(f(a) + f(b)) - \frac{\Delta x}{2\Delta x}\,(b - a)$$

$$\frac{\partial f}{\partial a} = \frac{1}{2}f'(a) + \frac{1}{2}\frac{\Delta x}{\Delta t} = \frac{1}{2}\left(\frac{\Delta x}{\Delta t} + f'(a)\right) \overset{!}{\geqslant} 0$$
$$\frac{\partial f}{\partial b} = \frac{1}{2}f'(b) - \frac{1}{2}\frac{\Delta x}{\Delta t} = -\frac{1}{2}\left(\frac{\Delta x}{\Delta t} - f'(b)\right) \overset{!}{\geqslant} 0$$

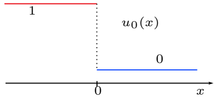
In order for both of this equations to hold it follows the CFL condition:

$$\left|f'(x)\right| \leqslant \frac{\Delta x}{\Delta t}$$

Example 9.4 RH for Riemann Problem[def. 2.4]**:**

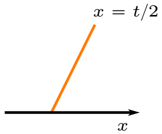
$$u_t + \left(\frac{u^2}{2}\right)_x = 0 \tag{9.2}$$

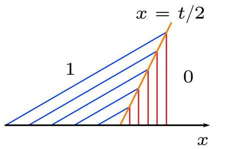
$$u_0 = \begin{cases} 1 & \text{if } x < 0 \\ 0 & \text{if } x > 0 \end{cases} \tag{9.3}$$

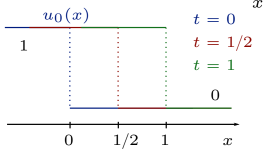


$$s(t) = \sigma'(t) = \frac{f\left(u^-(t)\right) - f\left(u^+(t)\right)}{u^-(t) - u^+(t)} = \frac{f(1) - f(0)}{1 - 0}$$
$$= \frac{\frac{1}{2} - 0}{1} = \frac{1}{2} \quad \Rightarrow \quad \sigma(t) = \frac{t}{2}$$

$$u = \begin{cases} 1 & \text{if } x < \frac{t}{2} \\ 0 & \text{if } x > \frac{t}{2} \end{cases} \tag{9.4}$$





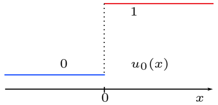


Thus we found a weak solution, where the characteristics are colliding on a traveling discontinuity/shockwave[def. 2.3].

Example 9.5 RK for Riemann Problem emanating:

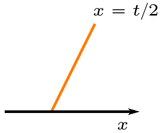
$$u_t + \left(\frac{u^2}{2}\right)_x = 0 \tag{9.5}$$

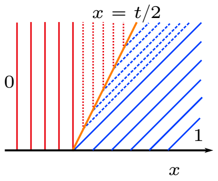
$$u_0 = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \end{cases} \tag{9.6}$$

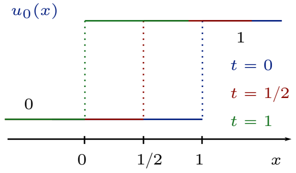


$$s(t) = \sigma'(t) = \frac{f\left(u^-(t)\right) - f\left(u^+(t)\right)}{u^-(t) - u^+(t)} = \frac{f(0) - f(1)}{0 - 1}$$
$$= \frac{-\frac{1}{2} - 0}{-1} = \frac{1}{2} \quad \Rightarrow \quad \sigma(t) = \frac{t}{2}$$

$$u(x,t) = \begin{cases} 0 & \text{if } x < \frac{t}{2} \\ 1 & \text{if } x > \frac{t}{2} \end{cases} \tag{9.7}$$





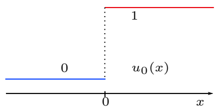


Problem we now get an area with characteristics emanating from the shock, thus we cannot track them back to the initial data.
This region of outflowing characteristics may in fact be filled in several ways seeexample 9.6

Example 9.6 RK for Riemann Problem emanating:

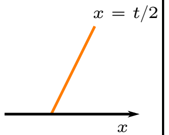
$$u_t + \left(\frac{u^2}{2}\right)_x = 0 \tag{9.8}$$

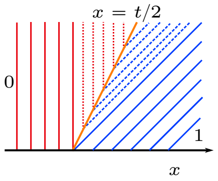
$$u_0 = \begin{cases} 0 & x < \frac{1}{4}t \\ \frac{1}{2} \quad \text{if} & \frac{1}{4}t < x < \frac{3}{4}t \\ 1 & x > \frac{3}{4}t \end{cases} \tag{9.9}$$

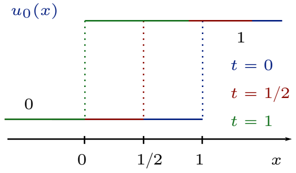


$$s(t) = \sigma'(t) = \frac{f\left(u^-(t)\right) - f\left(u^+(t)\right)}{u^-(t) - u^+(t)} = \frac{f(0) - f(1)}{0 - 1}$$
$$= \frac{-\frac{1}{2} - 0}{-1} = \frac{1}{2} \quad \Rightarrow \quad \sigma(t) = \frac{t}{2}$$

$$u(x,t) = \begin{cases} 0 & \text{if } x < \frac{t}{2} \\ 1 & \text{if } x > \frac{t}{2} \end{cases} \tag{9.10}$$







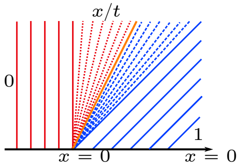
This solution obviously also fullfils the previous problem. **problem**: we thus can construct arbitrary many weak solutions by using the rh conditioneq. (2.3) with different inermediate states.

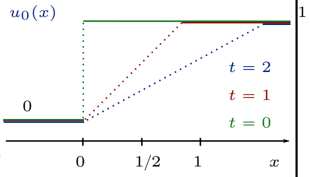
Example 9.7 Riemann Rarefaction[def. 2.7]**:**

$$u(x,0) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 0 \end{cases} \qquad u_t\left(\frac{u^2}{2}\right)_x = 0$$

Thus $f'(u) = u \implies (f')^{-1}(u) = u$ s.t.:

$$u(x,t) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x}{t} & \text{if } x > 0 \\ 1 & \text{if } x > 0 \end{cases}$$





- Thus after the after a small time period our solution will be piecewise/lipschitz continious $\implies u^- = u^+ \implies f\left(u^+\right) = f\left(u^-\right) \implies$ RH conditioneq. (2.3) will be automatically satisfied.
- From this it also follows that the Lax-entropy conditioneq. (2.6) is fullfiled.

$$f\left(u^+\right) = s(t) = f\left(u^-\right)$$

Example 9.8 $f(u) = au, \quad a > 0$
Why do we need Semi-Disc. FVS[def. 5.13]**:** Consider the upwind flux $F(u,u) = au$ then it follows for the FVM[def. 5.9]

$$u_j^{n+1} = u_j^n - \frac{a\Delta t}{\Delta x}\left(u_{j+}^n - u_{j-1+}^n\right) \tag{9.11}$$

and$\sigma_j^n \in \{\text{minmod,MC,superbee}\}$ it follows for the truncation error[def. 5.1]:

$$\|\tau_j^n\| \approx \mathcal{O}(\Delta x^3) + \mathcal{O}(\Delta t^2) \overset{eq.\ (3.17)}{\approx} \mathcal{O}(\Delta x^3) + \mathcal{O}(\Delta x^2)$$

thus schemes seem to be 2nd order may actually be first order due to the time-discretization.

Example 9.9 Wave Equation: The wave equation:

$$\overbrace{u_{tt}}^{\text{acceleration}} \quad -c^2 \overbrace{u_{xx}}^{\text{strain}} = 0$$

can be rewritten as a first-order system of equations by using the *change of variables*:

$$v := u_t \qquad\qquad w := -cu_x$$

$$u_{tt} - c\,(cu_{xx}) = 0 \qquad \Longrightarrow \qquad v_t + cw_x$$

we can find a second equations to obtain a system:

$$w_t = -cu_{xt} = -c\,(u_t)_x = -cv_x$$

Hence it follows:

$$\begin{aligned} v_t + cw_x &= 0 \\ w_t + cv_x &= 0 \end{aligned} \iff \mathbf{u}_t + \mathbf{A}\mathbf{u}_x = 0 \quad \mathbf{u} := \begin{bmatrix} v \\ w \end{bmatrix}, \mathbf{A} = \begin{bmatrix} 0 & c \\ c & 0 \end{bmatrix} \tag{9.12}$$

Example 9.10 Linearized Euler Equations:

Example 9.11 Laplace's Equations:

$$\Delta\mathbf{u} = 0 \qquad \Longrightarrow \qquad u_{tt} + u_{xx} = 0$$

can be rewritten as a first-order system of equations by using the *change of variables* similary toexample 9.9 but with $c = 1$ and a changed sign:

$$v := u_t \qquad\qquad w := u_x$$

$$u_{tt} + u_{xx} = 0 \qquad \Longrightarrow \qquad v_t + w_x$$

we can find a second equations to obtain a system:

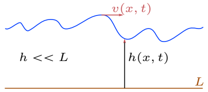
$$w_t = u_{xt} = (u_t)_x = v_x$$

Hence it follows:

$$\begin{aligned} v_t + w_x &= 0 \\ w_t - v_x &= 0 \end{aligned} \iff \mathbf{u}_t + \mathbf{A}\mathbf{u}_x = 0 \quad \mathbf{u} := \begin{bmatrix} v \\ w \end{bmatrix}, \mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \tag{9.13}$$

Example 9.12 Shallow Water Equations:

$$\begin{aligned} \partial_t h + \partial_x(hv) &= 0 \\ \partial_t(hv) + \partial_x\left(\frac{1}{2}gh^2 + hv^2\right) &= 0 \end{aligned} \tag{9.14}$$



$v(x,t)$: horizontal velocity of water column at x.
With $m := hv$ eq. (9.14) can be rewritten as non-linear scalar conservation laweq. (7.1):

$$\mathbf{U} = \begin{pmatrix} h \\ m \end{pmatrix} \qquad \mathbf{f}(\mathbf{U}) = \begin{pmatrix} m \\ \frac{1}{2}gh^2 + \frac{m^2}{h} \end{pmatrix} \tag{9.15}$$

$$\mathbf{f}'(\mathbf{U}) = \begin{pmatrix} 0 & 1 \\ gh & \frac{2m}{h} \end{pmatrix} \qquad \left|\begin{pmatrix} 0 & 1 \\ gh & \frac{2m}{h} \end{pmatrix}\right| = gh$$

$$\lambda_{1/2}(\mathbf{f}'(\mathbf{U})) \overset{eq.\ (20.77)}{\overset{\text{tr}=0}{=}} v \mp c \qquad c := \sqrt{gh}$$

$$\left(\mathbf{f}'(\mathbf{U}) - \lambda_j\right)\mathbf{r}_j(\mathbf{U}) = 0 \quad \Longrightarrow \mathbf{r}_{1/2}(\mathbf{U}) = \begin{pmatrix} 1 \\ v \mp c \end{pmatrix}$$

- Assuming that $h > 0$ we find that $\mathcal{U} = \left\{(h,m) \in \mathbb{R}^2 : h > 0\right\}$ s.t. eq. (9.14) is *hyperbolic.*
- moreover we find that both wave families of eq. (9.14) are *genuinely nonlinear*[def. 7.4]:

$$\nabla\lambda_{1/2}(\mathbf{U}) \cdot \mathbf{r}_{1/2}(\mathbf{U}) = \mp\frac{3}{2}\sqrt{\frac{g}{h}}$$

Example 9.13 Compressible Euler Equations:

$$\partial_t\rho + \partial_x(\rho v) = 0 \tag{9.16}$$

$$\partial_t(\rho v) + \partial_x\left(\rho v^2 + p\right) = 0 \tag{9.17}$$

$$\partial_t E + \partial_x((E+p)v) = 0 \tag{9.18}$$

The pressure p and the total energy E are related by the equation of state:

$$E = \frac{p}{\gamma - 1} + \frac{1}{2}\rho v^2 \qquad \gamma > 1 : \text{ heat capacity ratio} \tag{9.19}$$

The compressible euler equations can be written as conservation law:

$$\mathbf{U} = \begin{pmatrix} \rho \\ m \\ E \end{pmatrix} \qquad \mathbf{f}(\mathbf{U}) = \begin{pmatrix} \rho v \\ \rho v^2 + p \\ (E+p)v \end{pmatrix}$$

$$\begin{aligned} \lambda_1 &= v - c \\ \lambda_2 &= v \\ \lambda_3 &= v + c \end{aligned} \qquad c = \sqrt{\frac{\gamma p}{\rho}} \qquad v = \frac{m}{\rho}$$

For non-antimatter the pressure has to be positive thus admissible set is given by:

$$\mathcal{U} = \left\{(p, m, E) : p > 0 \iff E > \frac{m^2}{2\rho}\right\}$$

and the euler equations are thus a *strictly hyperbolic* system[def. 7.5].

$$\begin{aligned} \mathbf{r}_1 &= \begin{pmatrix} 1 & v - c & H - vc \end{pmatrix}^\intercal \\ \mathbf{r}_2 &= \begin{pmatrix} 1 & v & \frac{v^2}{2} \end{pmatrix}^\intercal \\ \mathbf{r}_3 &= \begin{pmatrix} 1 & v + c & H + vc \end{pmatrix}^\intercal \end{aligned} \qquad H = \frac{E+p}{\gamma} \text{ Enthalpy}$$

The second wave family is *linearly degenerated*:

$$\nabla\lambda_2 \cdot \mathbf{r}_2 = \begin{pmatrix} -\frac{m}{\rho^2} \\ \frac{1}{\rho} \\ 0 \end{pmatrix}^\intercal \mathbf{r}_2 = -\frac{m}{\rho^2} + \frac{v}{\rho} = -\frac{v}{\rho} + \frac{v}{\rho} = 0$$

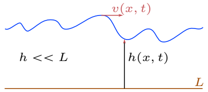
while the first and third wave family are *genuinely non-linear*.

Note

$$E = \frac{p}{\gamma - 1} + \frac{1}{2}\rho v^2 \implies p = ()\gamma - 1)\left(E - \frac{m^2}{2\rho}\right)$$

Example 9.14 Shallow Water Equations Entropy Pair[def. 7.19]**:**

$$\begin{aligned} \partial_t h + \partial_x(hv) &= 0 \\ \partial_t(hv) + \partial_x\left(\frac{1}{2}gh^2 + hv^2\right) &= 0 \end{aligned} \tag{9.20}$$



$v(x,t)$: horizontal velocity of water column at x.
With $m := hv$ eq. (9.14) can be rewritten as non-linear scalar conservation laweq. (7.1):

$$\mathbf{U} = \begin{pmatrix} h \\ m \end{pmatrix} \qquad \mathbf{f}(\mathbf{U}) = \begin{pmatrix} m \\ \frac{1}{2}gh^2 + \frac{m^2}{h} \end{pmatrix} \tag{9.21}$$

We now define the energy of a state $\mathbf{U} \in \mathcal{U} = \left\{(h,m) \in \mathbb{R}^2 : h > 0\right\}$ as the sum of the potential and kinetic energy:

$$s(\mathbf{U}) = \frac{1}{2}gh^2 + \frac{1}{2}hv^2$$

- Assuming that $h > 0$ we see that $s(\mathbf{U})$ is strictly convex is *hyperbolic.*
- if we define $q(\mathbf{U}) = h^2v + \frac{1}{3}hv^3$ it is straight forward to see that s, q is an entropy pair.

Example 9.15 Compressible Euler Equations[def. 7.19]**:**

$$\partial_t\rho + \partial_x(\rho v) = 0 \tag{9.22}$$

$$\partial_t(\rho v) + \partial_x\left(\rho v^2 + p\right) = 0 \tag{9.23}$$

$$\partial_t E + \partial_x((E+p)v) = 0 \tag{9.24}$$

The pressure p and the total energy E are related by the equation of state:

$$E = \frac{p}{\gamma - 1} + \frac{1}{2}\rho v^2 \qquad \gamma > 1 : \text{ heat capacity ratio} \tag{9.25}$$

The compressible Euler equations can be written as conservation law:

$$\mathbf{U} = \begin{pmatrix} \rho \\ m \\ E \end{pmatrix} \qquad \mathbf{f}(\mathbf{U}) = \begin{pmatrix} \rho v \\ \rho v^2 + p \\ (E+p)v \end{pmatrix}$$

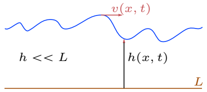
The *thermodynamic entropy* is defined as:

$$s(\mathbf{U}) = -\frac{\gamma S}{\gamma - 1} \qquad S - \log\left(\frac{p}{\rho^\gamma}\right) \text{ specifyc entropy} \tag{9.26}$$

If we define $q(\mathbf{U}) = -\frac{\gamma v S}{\gamma - 1}$ then s, q is an entropy pair.

Example 9.16
Roe Matrix for Shallow Water Equation??:

$$\begin{aligned} \partial_t h + \partial_x(hv) &= 0 \\ \partial_t(hv) + \partial_x\left(\frac{1}{2}gh^2 + hv^2\right) &= 0 \end{aligned} \tag{9.27}$$



$v(x,t)$: horizontal velocity of water column at x and the moment is $m = hv$:

$$\mathbf{U} = \begin{pmatrix} h \\ m \end{pmatrix} \qquad \mathbf{f}(\mathbf{U}) = \begin{pmatrix} m \\ \frac{1}{2}gh^2 + \frac{m^2}{h} \end{pmatrix} \tag{9.28}$$

Thus we have:

$$\llbracket\mathbf{U}\rrbracket = \begin{pmatrix} \llbracket h\rrbracket \\ \llbracket hv\rrbracket \end{pmatrix} \qquad \llbracket\mathbf{f}\rrbracket = \begin{pmatrix} \llbracket hv\rrbracket \\ \llbracket \frac{1}{2}gh^2 + hv^2\rrbracket \end{pmatrix} \tag{9.29}$$

From eq. (7.38) it follows:

$$\begin{pmatrix} \llbracket hv\rrbracket \\ \llbracket \frac{1}{2}gh^2 + hv^2\rrbracket \end{pmatrix} \overset{!}{=} \begin{pmatrix} A_{11}\llbracket h\rrbracket + A_{12}\llbracket hv\rrbracket \\ A_{21}\llbracket h\rrbracket + A_{22}\llbracket hv\rrbracket \end{pmatrix}$$

We see that $A_{11} = 0$ and $A_{12} = 1$ in order for the first equation to hold.
In order to solve the second rational equation we use the approach proposition 7.5 and define:

$$z_1 = \sqrt{h} \qquad\qquad z_2 = \sqrt{h}v$$

$$\begin{aligned} \Longrightarrow \qquad h &= z_1^2 \qquad & hv &= z_1z_2 \\ \Longrightarrow \qquad h^2 &= z^4 \qquad & hv^2 &= z_2^2 \end{aligned}$$

Thus it follows for the second equation:

$$\left\llbracket \frac{1}{2}gz_1^4 + z_2^2 \right\rrbracket \overset{!}{=} A_{21}\llbracket z_1^2\rrbracket + A_{22}\llbracket z_1z_2\rrbracket$$

Using the identities from proposition 7.5 we obtain:

$$2\bar{z}_2 + 2g\overline{z_1^2}\bar{z}_1\llbracket z_1\rrbracket = 2A_{21}\bar{z}_1\llbracket z_1\rrbracket + A_{22}\bar{z}_2\llbracket z_1\rrbracket + A_{22}\bar{z}_1\llbracket z_2\rrbracket$$

By comparing $\llbracket\cdot\rrbracket$ terms we find:

$$A_{22}\bar{z}_1 = 2\bar{z}_2 \qquad \Longrightarrow \qquad \underline{A_{22}} = \frac{2\bar{z}_2}{\bar{z}_1}$$

$$2A_{21}\bar{z}_1 + A_{22}\bar{z}_2 = 2g\overline{z_1^2}\bar{z}_1 \qquad \Longrightarrow \qquad A_{21} = g\overline{z_1^2} - \underline{A_{22}}\frac{\bar{z}_2}{2\bar{z}_1}$$

$$A_{22} = \frac{2\bar{z}_2}{\bar{z}_1} \qquad\qquad A_{21} = g\overline{z_1^2} - \left(\frac{\bar{z}_2}{\bar{z}_1}\right)^2$$

Plugging in z_1 and z_2 leads to the Roe matrix:

$$\mathbf{A}_{j+1/2}^n = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ g\bar{h} - \hat{v}^2 & 2\hat{v} \end{pmatrix} \tag{9.30}$$

with the Roe Averages defined as:

$$\bar{h} := \frac{h_j^n + h_{j+1}^n}{2} \qquad \hat{v} = \frac{\sqrt{h_j}v_j^n + \sqrt{h_{j+1}^n}v_{j+1}^n}{\sqrt{h_j^n} + \sqrt{h_{j+1}^n}} \tag{9.31}$$

Thus the Roe matrix is exactly equal to the Jaccobian of $\mathbf{f}'(\mathbf{U})$ **but** evaluate at the *Roe Averages.*

Example 9.17
Roe Matrix for Euler Equation??:

Math Appendix

Set Theory

Definition 11.1 Set $A = \{1, 3, 2\}$**:** is a well-defined group of distinct items that are considered as an object in its own right. The arrangement/order of the objects does not matter but each member of the set must be unique.

Definition 11.2 Empty Set $\{\}/\varnothing$**:** is the unique set having no elements/cardinality[def. 11.5] zero.

Definition 11.3 Multiset/Bag: Is a set-like object in which multiplicity[def. 11.4] matters, that is we can have multiple elements of the same type.
I.e. $\{1, 1, 2, 3\} \neq \{1, 2, 3\}$

Definition 11.4 Multiplicity: The multiplicity n_a of a member a of a multiset[def. 11.3] $\mathcal{S}$ is the number of times it appears in that set.

Definition 11.5 Cardinality $|S|$**:** Is the number of elements that are contained in a set.

Definition 11.6 The Power Set $\mathcal{P}(S)/2^S$**:** The power set of any set S is the set of all subsets of S, including the empty set and S itself. The cardinality of the power set is 2^S is equal to $2^{|S|}$.

1. Closure

Definition 11.7 Closure: A set is *closed* under an operation Ω if performance of that operations onto members of the set always produces a member of that set.

2. Open vs. Closed Sets

Definition 11.8 Open Sets:

- **Euclidean Spaces:**
  A subset $U \in \mathbb{R}$ is open, if for every $x \in U$ it exists $\epsilon(x)\,\mathbb{R}_+$ s.t. a point $y \in \mathbb{R}$ belongs to U if:
  $$\|x - y\|_2 < \epsilon(x) \tag{11.1}$$
- **Metric Spaces**[def. 20.63]: a Subset U of a metric space (M, d) is open if:
  $$\exists \epsilon > 0: \quad \text{if} \quad d(x, y) < \epsilon \qquad \forall y \in M, \forall x \in U \qquad \Longrightarrow \quad y \in U \tag{11.2}$$
- **Toplogical Spaces**[def. 22.2]: Let (X, τ) be a topological space. A set A is said to be open if it is contained in τ.

Definition 11.9 Closed Set: Is the complement of an open set[def. 11.8].

Definition 11.10 Bounded Set: A set $S \subset \mathbb{R}^n$is *bounded* if there exists a constant K s.t. the absolute value of every component of every element of S is less or equal to K.

3. Number Sets

3.1. The Real Numbers $\mathbb{R}$

3.1.1. Intervals

Definition 11.11 Closed Interval $[a, b]$**:** The closed interval of a and b is the set of all real numbers that are within a and b, including a and b:
$$[a, b] = \{x \in \mathbb{R} \,|\, a \leqslant x \leqslant b\} \tag{11.3}$$

Definition 11.12 Open Interval (a, b)**:** The open interval of a and b is the set of all real numbers that are within a and b:
$$(a, b) = \{x \in \mathbb{R} \,|\, a < x \leqslant <\} \tag{11.4}$$

3.2. The Rational Numbers $\mathbb{Q}$

Example 11.1 Power Set/Cardinality of $S = \{x, y, z\}$**:** The subsets of S are:
$\{\varnothing\}$, $\{x\}$, $\{y\}$, $\{z\}$, $\{x, y\}$, $\{x, z\}$, $\{y, z\}$, $\{x, y, z\}$ and hence the power set of S is $\mathcal{P}(S) = \{\{\varnothing\}, \{x\}, \{y\}, \{z\}, \{x, y\}, \{x, z\}, \{y, z\}, \{x, y, z\}\}$ with a cardinality of $|S| = 2^3 = 8$.

4. Set Functions

4.1. Submoduluar Set Functions

Definition 11.13 Submodular Set Functions: A submodular function $f : 2^\Omega \mapsto \mathbb{R}$ is a function that satisifies:
$$f(A \cup \{x\}) - f(A) \geqslant f(B \cup \{X\}) - F(B) \qquad \begin{matrix} \forall A \subseteq B \subset \Omega \\ \{x\} \in \Omega \backslash B \end{matrix} \tag{11.5}$$

Explanation 11.1 (Definition 11.13). *Addaing an element x to the the smaller subset A yields at least as much information/value gain as adding it to the larger subset B.*

Definition 11.14 Montone Submodular Function: A *monotone* submodular function is a submodular function[def. 11.13] that satisifies:
$$f(A) \leqslant f(B) \qquad \forall A \subseteq B \subseteq \Omega \tag{11.6}$$

Explanation 11.2 (Definition 11.14). *Adding more elements to a set will always increase the information/value gain.*

4.2. Complex Numbers

Definition 11.15 Complex Conjugate $\bar{z}$**:** The complex conjugate of a complex number $z = x + iy$ is defined as:
$$\bar{z} = x - iy \tag{11.7}$$

Corollary 11.1 Complex Conjugate Of a Real Number: The complex conjugate of a real number $x \in \mathbb{R}$ is x:
$$\bar{x} = x \qquad \Longrightarrow \qquad x \in \mathbb{R} \tag{11.8}$$

Formula 11.1 Euler's Formula:
$$e^{\pm ix} = \cos x \pm i \sin x \tag{11.9}$$

Formula 11.2 Euler's Identity:
$$e^{\pm i} = -1 \tag{11.10}$$

Note

$$e^{n} = 1 \Leftrightarrow n = \mathrm{i}\, 2\pi k, \qquad k \in \mathbb{N} \tag{11.11}$$

Sequences&Series

Definition 12.1 Index Set: Is a set[def. 11.1] A, whose members are labels to another set S. In other words its members index member of another set. An index set is build by enumerating the members of S using a function f s.t.
$$f : A \mapsto S \qquad A \in \mathbb{N} \tag{12.1}$$

Definition 12.2 Sequence $(a_n)_{n \in A}$**:** A sequence is an by an *index set A enumerated* multiset[def. 11.3] (repetitions are allowed) of objects in which *order does matter*.

Definition 12.3 Series: is an infinite ordered set of terms combined together by addition.

1. Types of Sequences

1.1. Arithmetic Sequence

Definition 12.4 Arithmetic Sequence: Is a sequence where the *difference* between two consecutive terms constant i.e. $(2, 4, 6, 8, 10, 12, \ldots)$.
$$t_n = t_0 + nd \qquad d \text{ :difference between two terms} \tag{12.2}$$

1.2. Geometric Sequence

Definition 12.5 Geometric Sequence: Is a sequence where the *ratio* between two consecutive terms constant i.e. $(2, 4, 8, 16, 32, \ldots)$.
$$t_n = t_0 \cdot r^n \qquad r \text{ :ratio between two terms} \tag{12.3}$$

Property 12.1 Sum of Geometric Sequence:
$$\sum_{k=1}^{n} ar^{k-1} = \frac{a(1 - r^n)}{1 - r} \tag{12.4}$$

2. Converging Sequences

2.1. Pointwise Convergence

Definition 12.6 $\lim_{n \to \infty} f_n = f$ **pointwise Pointwise Convergence[?]:** Let (f_n) be a sequence of functions with the same domain[def. 15.8] and codomain[def. 15.9]. The sequence is said to convergence pointwise to its *pointwise limit function f* if it satisfies:
$$|\lim_{n \to \infty} f_n(x) - f(x)| = 0 \qquad \forall x \in \mathtt{dom}(f_i) \tag{12.5}$$

2.2. Uniform Convergence

Definition 12.7 $\lim_{n \to \infty} f_n = f$ **uniform/**$f_n \overset{\infty}{\rightrightarrows} f$ **Uniform Convergence[?]:** Let (g_n) be a sequence of functions with the same domain[def. 15.8] and codomain[def. 15.9]. The sequence is said to convergence uniformly to its *pointwise limit function f* if it satisfies:
$$\exists \epsilon > 0 : \exists n \geqslant 1 \quad \sup_{x \in \mathtt{dom}(f_i)} \quad |g_n(x) - f(x)| < \epsilon \qquad \forall x \in \mathtt{dom}(f_i) \tag{12.6}$$

Note

Uniform convergence is characterized by the uniform norm**??**, and is stronger than pointwise convergence.

Toplogy

Definition 13.1 Topological Space[?] (X, τ)**:** Is an ordered pair (X, τ), where X is a set and τ is a topology[def. 22.1] on X.

Definition 13.2 Topological Space[?] (X, τ)**:** Is an ordered pair (X, τ), where X is a set and τ is a topology[def. 22.1] on X.

1. Weak Topologies

Definition 13.3 Weak Topology $\mathcal{C}(\mathcal{K}; \mathbb{R})$**:** Is the coresests topology s.t all cont. linear functionals w.r.t. to the strong topology are continuous.
Neighbourhood Basis:
$$\{f \,||l_1| < \epsilon_1, \ldots, |l_n| < \epsilon_n, \forall \epsilon_i, \forall n, \forall \text{lin. functions} f\} \tag{13.1}$$

Note

The weak closure:

- is usually larger as the uniform closure, as for the weak closure there are many more convergence sequences
- is easier to calculate than the uniform closure

2. Compact Space

Corollary 13.1 Euclidean Space: In the euclidean case, a set $X \in \mathbb{R}$ is compact iff:

- it is closed[def. 11.9]
- bounded

3. Closure

Definition 13.4 Closure of a Set[?] $\mathrm{cl}_{X,\tau}(S)/\bar{S}$**:** The closure of a subset S of a toplogical space[def. 22.2] (X, τ) is defined equivilantly by:

- Is the union of S and its boundary ∂S.
- is the set S together with its limit points.

Note

If the topological space X, τ is clear from context, then the closure of a set S is often written simply as $\bar{S}$.

Corollary 13.2 Uniform Closure $\bar{\cdot}^{\|\cdot\|_\infty}$**:** The uniform closure of a set of functions A is *the space of all functions that can be approximated* by a sequence (f_n) of uniformly-converging functions from A. [def. 12.7] functions

Corollary 13.3 Weak Closure:

Logic

1. Boolean Algebra

1.1. Basic Operations

Definition 14.1 Conjunction/AND $\wedge$:

Definition 14.2 Disjunction/OR $\vee$:

Definition 14.3 Negation/NOT $\neg$:

1.1.1. Expression as Integer

If the truth values $\{0,1\}$ are interpreted as integers then the basic operations can be represent with basic arithmetic operations.

$$\begin{aligned} x \wedge y &= xy = \min(x,y) \\ x \vee y &= x + y = \max(x,y) \\ \neg x &= 1 - x \\ x \oplus y &= (x+y)\cdot(\neg x + \neg y) = x \cdot \neg y + \neg x \cdot y \end{aligned}$$

Note: non-linearity of XOR

$$(x+y)\cdot(\neg x + \neg y) = -x^2 - y^2 - 2xy + 2x + 2y$$

1.2. Boolean Identities

Property 14.1 Idempotence:

$$x \wedge x \equiv x \qquad \text{and} \qquad x \vee x \equiv x \tag{14.1}$$

Property 14.2 Identity Laws:

$$x \wedge \text{true} \equiv x \qquad \text{and} \qquad x \vee \text{false} \equiv x \tag{14.2}$$

Property 14.3 Zero Law's:

$$x \wedge \text{false} \equiv \text{false} \qquad \text{and} \qquad x \vee \text{true} \equiv \text{true} \tag{14.3}$$

Property 14.4 Double Negation:

$$\neg\neg x \equiv x \tag{14.4}$$

Property 14.5 Complementation:

$$x \wedge \neg x \equiv \text{false} \qquad \text{and} \qquad x \vee \neg x \equiv \text{true} \tag{14.5}$$

Property 14.6 Commutativity:

$$x \vee y \equiv y \vee x \qquad \text{and} \qquad x \wedge y \equiv y \wedge x \tag{14.6}$$

Property 14.7 Associativity:

$$(x \vee y) \vee z \equiv x \vee (y \vee z) \tag{14.7}$$
$$(x \wedge y) \vee z \equiv x \vee (y \wedge z) \tag{14.8}$$

Property 14.8 Distributivity:

$$x \vee (y \wedge z) \equiv (x \vee y) \wedge (x \vee z) \tag{14.9}$$
$$x \wedge (y \vee z) \equiv (x \wedge y) \vee (x \wedge z) \tag{14.10}$$

Property 14.9 De Morgan's Laws:

$$\neg(x \vee z) \equiv (\neg x \wedge \neg y) \tag{14.11}$$
$$\neg(x \wedge z) \equiv (\neg x \vee \neg y) \tag{14.12}$$

Note

The algebra axioms come in pairs that can be obtained by interchanging $\wedge$ and $\vee$.

1.3. Normal Forms

Definition 14.4 Literal [example 14.1]:
Literals are atomic formulas or their negations

Definition 14.5 Negation Normal Form (NNF): A formula F is in negation normal form is the negation operator $\neg$ is only applied to literals[def. 14.4] and the only other operators are $\wedge$ and $\vee$.

Definition 14.6 Conjunctive Normal Form (CNF): An boolean algebraic expression F is in CNF if it is a *conjunction* of *clauses*, where each clause is a disjunction of *literals*[def. 14.4] $L_{i,j}$:

$$F_{\text{CNF}} = \bigwedge_{i=1}^{n} \left(\bigvee_{j=1}^{m_i} L_{i,j} \right) \tag{14.13}$$

Definition 14.7 Disjunctive Normal Form (DNF): An boolean algebraic expression F is in DNF if it is a *disjunction* of *clauses*, where each clause is a conjunction of *literals*[def. 14.4] $L_{i,j}$:

$$F_{\text{DNF}} = \bigvee_{i=1}^{n} \left(\bigwedge_{j=1}^{m_i} L_{i,j} \right) \tag{14.14}$$

Note

- true is a CNF with no clause and a single literal.
- false is a CNF with a single clause and no literals

1.3.1. Transformation to CNF and DNF

DNF

Algorithm 14.1:

① Using *De Morgan's laws*Property 14.9 and double negationProperty 14.4 transform F into *Negation Normal Form*[def. 14.5]:

$\neg\neg x$	by	x
$\neg(x \wedge y)$	by	$(\neg x \vee \neg y)$
$\neg(x \vee y)$	by	$(\neg x \wedge \neg y)$
$\neg$true	by	false
$\neg$false	by	true

② Using distributive lawsProperty 14.8 substitute all:

$x \wedge (y \vee z)$	by	$(x \wedge y) \vee (x \wedge z)$
$(y \vee z) \wedge x$	by	$(y \wedge x) \vee (z \wedge x)$
$x \wedge$ true	by	true
true $\wedge\ x$	by	true

③ Using the identityProperty 14.2 and zero laws Property 14.3 remove true from any cause and delete all clauses containing false.

Note

For the CNF form simply use duality for step 2 and 3 i.e. swap $\wedge$ and $\vee$ and true and false.

Using Truth Tables [example 14.2]

To obtain a DNF formula from a truth table we need to have a *conjunctive*[def. 14.3] for each row where F is true.

2. Examples

Example 14.1 Literals:
Boolean literals: $x, \neg y, s$
Not boolean literals: $\neg\neg x$, $(x \wedge y)$

Example 14.2 DNF from truth tables:

x	y	z	F
0	0	0	1
0	0	1	0
0	1	0	0
0	1	1	1
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

Need a conjunction of:

- $(\neg x \wedge \neg y \wedge \neg z)$
- $(\neg x \wedge y \wedge z)$
- $(x \wedge \neg y \wedge \neg z)$
- $(x \wedge y \wedge z)$

$$(\neg x \wedge \neg y \wedge \neg z) \wedge (\neg x \wedge y \wedge z) \wedge (x \wedge \neg y \wedge \neg z) \wedge (x \wedge y \wedge z)$$

Calculus and Analysis

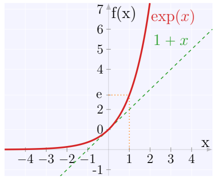
1. Functional Analysis

1.1. Elementary Functions

1.1.1. Exponential Numbers

Definition 15.1 Exponential Function $\exp : \mathbb{K} \mapsto \mathbb{K}$:

$$\exp(x) = \mathrm{e}^{x} = \sum_{n=0}^{\infty} \frac{x}{n!} = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^{n} \tag{15.1}$$



Definition 15.2 Exponential/Euler Number e:

$$\mathrm{e} = \sum_{n=0}^{\infty} \frac{1}{n!} = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^{n} = 2.7182 \tag{15.2}$$

Properties Defining the Expeontial Function

Property 15.1:

$$\exp(x+y) = \exp(x) + \exp(y) \tag{15.3}$$

Property 15.2:

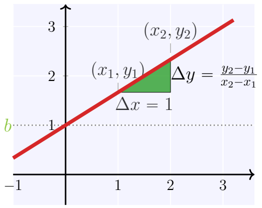
$$\exp(x) \leqslant 1 + x \tag{15.4}$$

1.1.2. Affine Linear Functions

Definition 15.3 Affine Linear Function $f(x) = ax + b$:
An affine linear function are functions that can be defined by a scaling $s_a(x) = ax$ plus a translation $t_b(x) = x + b$:

$$M = \{f : \mathbb{R} \mapsto \mathbb{R} | f(x) = (s_a \circ t_b)(x) = ax + b, \quad a, b \in \mathbb{R}\} \tag{15.5}$$

$$f(x) = ax + b$$
$$f(0) = b$$
$$f'(x) = a$$



Formula 15.1 [proof 15.1]
Linear Function from Point and slope $f(x_0) = y_1$:
Given a point (x_1, y_1) and a slope a we can derive:

$$f(x) = a \cdot (x - x_0) + y_0 = ax + (y_1 - ax_0) \tag{15.6}$$

Formula 15.2 Linear Function from two Points:

$$f(x) = a \cdot (x - x_p) + y_p = ax + (y_p - ax_p) \tag{15.7}$$
$$a = \frac{y_1 - y_0}{x_1 - x_0} \qquad p = \{ \text{ or } 2\}1$$

1.1.3. Polynomials

Definition 15.4 Polynomial: A function $\mathcal{P}_n : \mathbb{R} \mapsto \mathbb{R}$ is called *Polynomial*, if it can be represented in the form:

$$\mathcal{P}_n(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1}x^{n-1} + a_n x^n \tag{15.8}$$

Corollary 15.1 Degree n-of a Polynomial $\deg(\mathcal{P}_n)$: the *degree* of the polynomial is the highest exponent of the variable x, among all non-zero coefficients $a_i \neq 0$.

Definition 15.5 Monomial: Is a polynomial with only one term.

Cubic Polynomials

Definition 15.6 Cubic Polynomials: Are polynomials of degree[cor. 15.1] 3 and have four coefficients:

$$f(x) = a_3x^3 + a_2x^2 + a_1x + a_0 \tag{15.9}$$

1.2. Functional Compositions

Definition 15.7 Functional Compositions $f \circ g$:
Let $f : A \mapsto B$ and $g : D \mapsto C$ be to mappings s.t. $\texttt{codom}(f) \subseteq D$ then we can define a composition function $(f \circ g) A \mapsto D$ as:

$$h(\mathbf{x}) = (g \circ f)(\mathbf{x}) = g(f(\mathbf{x})) \qquad \text{with} \qquad \mathbf{x} \in A \tag{15.10}$$

Corollary 15.2 Nested Functional Composition:

$$F_{k:1}(\mathbf{x}) = (F_k \circ \cdots \circ F_1)(\mathbf{x}) = F_k\Big(F_{k-1} \circ \cdots \circ \big(F_1(\mathbf{x})\big)\Big) \tag{15.11}$$

2. Proofs

Proof 15.1 formula 15.1:

$$f(x_0) = y_0 = ax_0 + b \qquad \Rightarrow \qquad b = y_0 - ax_0$$

Theorem 15.1
First Fundamental Theorem of Calculus:
Let f be a continuous real-valued function defined on a closed interval $[a,b]$.
Let F be the function defined $\forall x \in [a,b]$ by:

$$F(X) = \int_a^x f(t)\,\mathrm{d}t \tag{15.12}$$

Then it follows:

$$F'(x) = f(x) \qquad \forall x \in (a,b) \tag{15.13}$$

Theorem 15.2
Second Fundamental Theorem of Calculus:
Let f be a real-valued function on a closed interval $[a,b]$ and F an antiderivative of f in $[a,b]$: $F'(x) = f(x)$, then it follows if f is Riemann integrable on $[a,b]$:

$$\int_a^b f(t)\,\mathrm{d}t = F(b) - F(a) \qquad \Longleftrightarrow \qquad \int_a^x \frac{\partial}{\partial x} F(t)\,\mathrm{d}t = F(x) \tag{15.14}$$

Definition 15.8 Domain of a function $\texttt{dom}(\cdot)$:
Given a function $f : \mathcal{X} \to \mathcal{Y}$, the set of all possible input values $\mathcal{X}$ is called the domain of f – $\texttt{dom}(f)$.

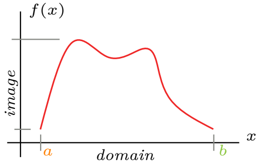
Definition 15.9
Codomain/target set of a function $\texttt{codom}(\cdot)$:
Given a function $f : \mathcal{X} \to \mathcal{Y}$, the codaomain of that function is the set $\mathcal{Y}$ into which all of the output of the function is **constrained** to fall.

Definition 15.10 Image (Range) of a function: $f[\cdot]$

Given a function $f : \mathcal{X} \to \mathcal{Y}$, the image of that function is the set to which the function can actually map:

$$\{y \in \mathcal{Y} | y = f(x), \quad \forall x \in \mathcal{X}\} := f[\mathcal{X}] \tag{15.15}$$

Evaluating the function f at each element of a given subset A of its domain $\texttt{dom}(f)$ produces a set called the *image of A under (or through) f*.
The image is thus a subset of a function's codomain.



Misnomer Range: The term Range is ambiguous s.t. certain books refer to it as codomain and other as image.

Definition 15.11 Inverse Image/Preimage $f^{-1}(\cdot)$:
Let $f : X \mapsto Y$ be a function, and A a subset set of its codomain Y.
Then the preimage of A under f is the set of all elements of the domain X, that map to elements in A under f:

$$f^{-1}(A) = \{x \subseteq X : f(x) \subseteq A\} \tag{15.16}$$

Example 15.1 :
Given $f : \mathbb{R} \to \mathbb{R}$
defined by $f : x \mapsto x^2 \iff f(x) = x^2$
$\texttt{dom}(f) = \mathbb{R}$, $\texttt{codom}(f) = \mathbb{R}$ **but** its image is $f[\mathbb{R}] = \mathbb{R}_+$.

Image (Range) of a subset

The image of a subset $A \subseteq \mathcal{X}$ under f is the subset $f[A] \subseteq \mathcal{Y}$ defined by:

$$f[A] = \{y \in \mathcal{Y} | y = f(x), \quad \forall x \in A\} \tag{15.17}$$

Note: Range

The term range is ambiguous as it may refer to the image or the codomain, depending on the definition.
However, modern usage almost always uses range to mean image.

Definition 15.12 (strictly) Increasing Functions:
A function f is called monotonically increasing/increasing/non-decreasing if:

$$x \leqslant y \iff f(x) \leqslant f(y) \qquad \forall x, y \in \texttt{dom}(f) \tag{15.18}$$

And **strictly increasing** if:

$$x < y \iff f(x) < f(y) \qquad \forall x, y \in \texttt{dom}(f) \tag{15.19}$$

Definition 15.13 (strictly) Decreasing Functions: A function f is called monotonically decreasing/decreasing or non-increasing if:

$$x \geqslant y \quad \iff \quad f(x) \geqslant f(y) \quad \forall x, y \in \texttt{dom}(f) \tag{15.20}$$

And *strictly* decreasing if:

$$x > y \quad \iff \quad f(x) > f(y) \quad \forall x, y \in \texttt{dom}(f) \tag{15.21}$$

Definition 15.14 Monotonic Function: A function f is called monotonic iff either f is increasing or decreasing.

Definition 15.15 Linear Function: A function $L : \mathbb{R}^n \mapsto \mathbb{R}^m$ is linear if and only if:

$$L(\mathbf{x} + \mathbf{y}) = L(\mathbf{x}) + L(\mathbf{y})$$
$$L(\alpha\mathbf{x}) = \alpha L(\mathbf{x}) \qquad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \quad \alpha \in \mathbb{R}$$

Corollary 15.3 Linearity of Differentiation: The derivative of **any** linear combination of functions equals the same linear combination of the derivatives of the functions:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(af(x) + bg(x)\right) = a\frac{\mathrm{d}}{\mathrm{d}x}f(x) + b\frac{\mathrm{d}}{\mathrm{d}x}g(x) \qquad a, b \in \mathbb{R} \tag{15.22}$$

Definition 15.16 Quadratic Function: A function $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ is quadratic if it can be written in the form:

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\intercal\mathbf{A}\mathbf{x} + \mathbf{b}^\intercal\mathbf{x} + c \tag{15.23}$$

3. Norms

3.1. Infinity/Supremum Norm

Definition 15.17 Infinity/Supremum Norm:

$$\|f\|_\infty := \sup_{x \in \texttt{dom}(f)} |f(x)| \tag{15.24}$$

Note

In order to make this a proper norm one usually considers *bounded functions* s.t.:

$$\|f\|_\infty \leqslant M < \infty$$

Corollary 15.4 Ininity Norm induced Metric: The infinty norm naturally induces a metric[def. 20.62]:

$$d := (f, g) := \|f - g\|_\infty \tag{15.25}$$

4. Smoothness

Definition 15.18 Smoothness of a Function $\mathcal{C}^k$: **Given** a function $f : \mathcal{X} \to \mathcal{Y}$, the function is said to be of class k if it is differentiable up to order k **and** continuous, on its entire domain:

$$f \in \mathcal{C}^k(\mathcal{X}) \quad \iff \quad \exists f', f'', \ldots, f^{(k)} \text{ continuous} \tag{15.26}$$

Note

- P.w. continuous ≠ continuous.
- A function of that is k times differentiable must at least be of class $\mathcal{C}^{k-1}$.
- $\mathcal{C}^m(\mathcal{X}) \subset \mathcal{C}^{m-1}, \ldots \mathcal{C}^1 \subset \mathcal{C}^0$
- Continuity is implied by the differentiability of all **derivatives** of up to order $k - 1$.

4.0.1. Continuous Functions

Definition 15.19 Continuous Function $\mathcal{C}^0$**:** Functions that do not have any jumps or peaks.

4.0.2. Piece wise Continuous Functions

Definition 15.20 Piecewise Linear Functions $\mathcal{C}^0_{\mathbf{pw}}$**:**

4.0.3. Continously Differentiable Function

Corollary 15.5 Continuously Differentiable Function $\mathcal{C}^1$**:** Is the class of functions that consists of all differentiable functions whose derivative is continuous.
Hence a function $f : \mathcal{X} \to \mathcal{Y}$ of the class must satisfy:

$$f \in \mathcal{C}^1(\mathcal{X}) \quad \iff \quad f' \text{ continuous} \tag{15.27}$$

4.0.4. Smooth Functions

Corollary 15.6 Smooth Function $\mathcal{C}^\infty$: Is a function $f : \mathcal{X} \to \mathcal{Y}$ that has derivatives infinitely many times differerntiable.

$$f \in \mathcal{C}^\infty(\mathcal{X}) \quad \iff \quad f', f'', \ldots, f^{(\infty)} \tag{15.28}$$

4.1. Lipschitz Continuous Functions

Often functions are not differentiable but we still want to state something about the rate of change of a function ⇒ hence we need a weaker notion of differentiablility.

Definition 15.21 Lipschitz Continuity: A Lipschitz continuous function is a function f whose rate of change is bound by a Lipschitz Constant L:

$$|f(\mathbf{x}) - f(\mathbf{y})| \leqslant L\|\mathbf{x} - \mathbf{y}\| \qquad \forall \mathbf{x}, \mathbf{y}, \qquad L > 0 \tag{15.29}$$

Note

This property is useful as it allows us to conclude that a small perturbation in the input (i.e. of an algorithm) will result in small changes of the output ⇒ tells us something about robustness.

4.1.1. Lipschitz Continuous Gradient

Definition 15.22 Lipschitz Continuous Gradient: A *continuously differentiable* function $f : \mathbb{R}^d \mapsto \mathbb{R}$ has L-*Lipschitz continuous gradient* if it satisfies:

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leqslant L\|\mathbf{x} - \mathbf{y}\| \qquad \forall \mathbf{x}, \mathbf{y} \in \texttt{dom}(f), \qquad L > 0 \tag{15.30}$$

if $f \in \mathcal{C}^2$, this is equivalent to:

$$\nabla^2 f(\mathbf{x}) \leqslant L\mathbf{I} \qquad \forall \mathbf{x} \in \texttt{dom}(f), \qquad L > 0 \tag{15.31}$$

Lemma 15.1 Descent Lemma [Poorfs 15.5,??]: If a function $f : \mathbb{R}^d \mapsto \mathbb{R}$ has *Lipschitz continuous gradient* eq. (15.30) over its domain, then it holds that:

$$|f(\mathbf{x}) - f(\mathbf{y}) - \nabla f(\mathbf{y})^\intercal(\mathbf{x} - \mathbf{y})| \leqslant \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2 \tag{15.32}$$

Note

If f is twice differentiable then the largest eigenvalue of the Hessian (Definition 16.8) of f is uniformly upper bounded by L

4.2. L-Smooth Functions

Definition 15.23 L-Smoothness: A L-smooth function is a function $f : \mathbb{R}^d \mapsto \mathbb{R}$ that satisfies:

$$f(\mathbf{x}) \leqslant f(\mathbf{y}) + \nabla f(\mathbf{y})^\intercal(\mathbf{x} - \mathbf{y}) + \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2$$

with $\forall \mathbf{x}, \mathbf{y} \in \texttt{dom}(f), \quad L > 0 \quad (15.33)$

If f is a twice differentiable this is equivalent to:

$$\nabla^2 f(\mathbf{x}) \leqslant L\mathbf{I} \qquad L > 0 \tag{15.34}$$

Theorem 15.3 [proof 15.6]
L-Smoothness of convex functions: A *convex* and L-Smooth function ([def. 15.23]) has a *Lipschitz continuous gradient*eq. (15.30) thus it holds that:

$$f(\mathbf{x}) \leqslant f(\mathbf{y}) + \nabla f(\mathbf{y})^\intercal(\mathbf{x} - \mathbf{y}) \leqslant \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|^2 \tag{15.35}$$

Note

L-smoothnes is a weaker condition than L-Lipschitz continuous gradients

5. Convexity and Concavity

Definition 15.24 Convex Functions: A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if it satisfies:

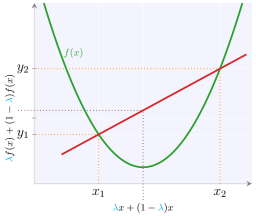
$$f(\lambda x + (1 - \lambda)y) \leqslant \lambda f(x) + (1 - \lambda)f(y) \tag{15.36}$$
$$\forall \lambda \in [0, 1] \qquad \forall x, y \in \texttt{dom}(f)$$

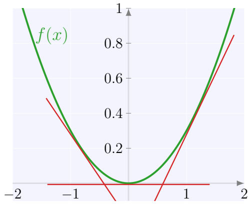
If f is a differentiable function this is equivalent to:

$$f(x) \geqslant f(y) + \nabla f(y)^\intercal(x - y) \qquad \forall x, y \in \texttt{dom}(f) \tag{15.37}$$

If f is a twice differentiable function this is equivalent to:

$$\nabla^2 f(x) \geqslant 0 \qquad \forall x, y \in \texttt{dom}(f) \tag{15.38}$$





Definition 15.25 Concave Functions: A function $f : \mathbb{R}^n \to \mathbb{R}$ is concave if it satisfies:

$$f(\lambda x + (1 - \lambda)y) \geqslant \lambda f(x) + (1 - \lambda)f(y) \qquad \begin{array}{l}\forall x, y \in \texttt{dom}(f)\\ \forall \lambda \in [0, 1]\end{array} \tag{15.39}$$

Corollary 15.7 Convexity → global minimima: Convexity implies that all local minima (if they exist) are global minima.

5.1. Properties

Property 15.3 Monotonicity of the Derivative:

$$\text{If } f : \mathbb{R} \mapsto \mathbb{R} \text{ is } \begin{cases} \text{convex} & f'(a) < f'(b) \\ \text{concave} & f'(a) > f'(b) \end{cases} \qquad a < b, \quad a, b \in \mathbb{R} \tag{15.40}$$

5.1.1. Properties that preserve convexity

Property 15.4 Non-negative weighted Sums: Let f be a convex function then $g(x)$ is convex as well:

$$g(x) = \sum_{i=1}^{n} \alpha_i f_i(x) \qquad \forall \alpha_j > 0$$

Property 15.5 Composition of Affine Mappings: Let f be a convex function then $g(x)$ is convex as well:

$$g(x) = f(\mathbf{A}x + b)$$

Property 15.6 Pointwise Maxima: Let f be a convex function then $g(x)$ is convex as well:

$$g(x) = \max_i\{f_i(x)\}$$

5.2. Strict Convexity/Concavity

Definition 15.26 Stricly Convex Functions: A function $f : \mathbb{R}^n \to \mathbb{R}$ is strictly convex if it satisfies:

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y) \qquad \begin{array}{l}\forall x, y \in \texttt{dom}(f)\\ \forall \lambda \in [0, 1]\end{array}$$

If f is a differentiable function this is equivalent to:

$$f(x) > f(y) + \nabla f(y)^\intercal(x - y) \qquad \forall x, y \in \texttt{dom}(f) \tag{15.41}$$

If f is a twice differentiable function this is equivalent to:

$$\nabla^2 f(x) > 0 \qquad \forall x, y \in \texttt{dom}(f) \tag{15.42}$$

Intuition

- Convexity implies that a function f is bound by/below a linear interpolation from x to y and strong convexity that f is strictly bound/below.
- eq. (15.41) implies that $f(x)$ is above the tangent $f(x) + \nabla f(x)^\intercal(y - x)$ for all $x, y \in \texttt{dom}(f)$
- ?? implies that $f(x)$ is flat or curved upwards

Corollary 15.8 Strict Convexity → Uniqueness: Strict convexity implies a unique minimizer ⟺ at most one global minimum.

Corollary 15.9 : A twice differentiable function of one variable $f : \mathbb{R} \to \mathbb{R}$ is convex on an interval $\mathcal{X} = [a, b]$ if and only if its second derivative is non-negative on that interval $\mathcal{X}$:

$$f''(x) \geqslant 0 \qquad \forall x \in \mathcal{X} \tag{15.43}$$

5.3. Strong Convexity/Concavity

Definition 15.27 μ-Strong Convexity: Let $\mathcal{X}$ be a Banach space over $\mathbb{K} = \mathbb{R}, \mathbb{C}$. A function $f : \mathcal{X} \to \mathbb{R}$ is called strongly convex iff the following equation holds:

$$f\left(tx + (1 - t)y\right) \leqslant tf(x) + (1 - t)f(y) - \frac{t(1 - t)}{2}\mu\|x - y\|$$
$$\forall x, y \in \mathcal{X}, \qquad t \in [0, 1], \qquad \mu > 0$$

If $f \in \mathcal{C}^1 \iff f$ is differentiable, this is equivalent to:

$$f(y) \geqslant f(x) + \nabla f(x)^\intercal(y - x) + \frac{\mu}{2}\|y - x\|_2^2 \tag{15.44}$$

If $f \in \mathcal{C}^2 \iff f$ is twice differentiable, this is equivalent to:

$$\nabla^2 f(x) \geqslant \mu\mathbf{I} \qquad \forall x, y \in \mathcal{X} \qquad \mu > 0 \tag{15.45}$$

Corollary 15.10
Strong Convexity implies Strict Convexity:

Property 15.7:

$$f(\mathbf{y}) \leqslant f(\mathbf{y}) + \nabla f(\mathbf{y})^\intercal(\mathbf{x} - \mathbf{y}) + \frac{1}{2\mu}\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_2^2 \tag{15.46}$$

Intuition

Strong convexity implies that a function f is lower bounded by its second order (quadratic) approximation, rather then only its first order (linear) approximation.

Size of μ

The parameter μ specifies how strongly the bounding quadratic function/approximation is.

Proof 15.2: eq. (15.45) analogously to **Proof** eq. (15.34)

Note

If f is twice differentiable then the smallest eigenvalue of the Hessian ([def. 16.8]) of f is uniformly lower bounded by μ **Hence** strong convexity can be considered as the analogous to smoothness

Example 15.2 Quadratic Function: A quadratic function eq. (15.23) is convex if:

$$\nabla^2_{\mathbf{x}}\text{eq. } (15.23) = \mathbf{A} \geqslant 0 \tag{15.47}$$

Corollary 15.11 :

Strong convexity ⇒ Strict convexity ⇒ Convexity

Functions

Even Functions: have rotational symmetry with respect to the origin.
⇒**Geometrically**: its graph remains unchanged after reflection about the y-axis.

$$f(-x) = f(x) \tag{15.48}$$

Odd Functions: are symmetric w.r.t. to the y-axis.
⇒**Geometrically**: its graph remains unchanged after rotation of 180 degrees about the origin.

$$f(-x) = -f(x) \tag{15.49}$$

Theorem 15.4 Rules:
Let f be even and f odd respectively.

$g =: f \cdot f$ is even $\qquad$ $g =: f \cdot f$ is even
$g =: f \cdot f$ is odd $\qquad$ the same holds for division

Examples

Even: $\cos x$, $|x|$, c, x^2, x^4,... $\exp(-x^2/2)$.
Odd: $\sin x$, $\tan x$, x, x^3, x^5,....

$$\begin{aligned} x\text{-Shift}: \quad & f(x - c) \Rightarrow \text{shift to the right} \\ & f(x + c) \Rightarrow \text{shift to the left} && (15.50) \\ y\text{-Shift}: \quad & f(x) \pm c \Rightarrow \text{shift up/down} && (15.51) \end{aligned}$$

Proof 15.3: **eq. (15.50)** $f(x_n - c)$ we take the x-value at x_n but take the y-value at $x_o := x_n - c$
⇒ we shift the function to x_n.

Euler's formula

$$e^{\pm ix} = \cos x \pm i \sin x \tag{15.52}$$

Euler's Identity

$$e^{\pm i} = -1 \tag{15.53}$$

Note

$$e^{n} = 1 \Leftrightarrow n = \mathrm{i}\, 2\pi k, \qquad k \in \mathbb{N} \tag{15.54}$$

Corollary 15.12 Every norm is a convex function: By using definition [def. 15.24] and the triangular inequality it follows (with the exception of the L0-norm):

$$\|\lambda x + (1-\lambda)y\| \leqslant \lambda\|x\| + (1-\lambda)\|y\|$$

5.4. Taylor Expansion

Definition 15.28 Taylor Expansion:

$$T_n(x) = \sum_{i=0}^{n} \frac{1}{n!} f^{(i)}(x_0) \cdot (x - x_0)^{(i)} \tag{15.55}$$

$$= f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2} f''(x_0)(x - x_0)^2 + \mathcal{O}(x^3) \tag{15.56}$$

Definition 15.29 Incremental Taylor:
Goal: evaluate $T_n(x)$ (eq. (15.56)) at the point $x_0 + \Delta x$ in order to propagate the function $f(x)$ by $h = \Delta x$:

$$T_n(x_0 \pm h) = \sum_{i=0}^{n} \frac{h^i}{n!} f^{(i)}(x_0) i^{-1} \tag{15.57}$$

$$= f(x_0) \pm h f'(x_0) + \frac{h^2}{2} f''(x_0) \pm f'''(x_0)(h)^3 + \mathcal{O}(h^4)$$

Note

If we chose Δx small enough it is sufficient to look only at the first two terms.

Definition 15.30 Multidimensional Taylor: Suppose $X \in \mathbb{R}^n$ is open, $\mathbf{x} \in X$, $f : X \mapsto \mathbb{R}$ and $f \in \mathcal{C}^2$ then it holds that

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + \nabla_{\mathbf{x}} f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_0)^{\mathsf{T}} H(\mathbf{x} - \mathbf{x}_0) \tag{15.58}$$

Definition 15.31 Argmax: The argmax of a function defined on a set D is given by:

$$\arg\max_{x \in D} f(x) = \{x | f(x) \geqslant f(y), \forall y \in D\} \tag{15.59}$$

Definition 15.32 Argmin: The argmin of a function defined on a set D is given by:

$$\arg\min_{x \in D} f(x) = \{x | f(x) \leqslant f(y), \forall y \in D\} \tag{15.60}$$

Corollary 15.13 Relationship $\arg\min \leftrightarrow \arg\max$**:**

$$\arg\min_{x \in D} f(x) = \arg\max_{x \in D} -f(x) \tag{15.61}$$

Property 15.8 Argmax Identities:

1. **Shifting**:
$$\forall \lambda \text{ const} \qquad \arg\max f(x) = \arg\max f(x) + \lambda \tag{15.62}$$
2. **Positive Scaling**:
$$\forall \lambda > 0 \text{ const} \qquad \arg\max f(x) = \arg\max \lambda f(x) \tag{15.63}$$
3. **Negative Scaling**:
$$\forall \lambda < 0 \text{ const} \qquad \arg\max f(x) = \arg\min \lambda f(x) \tag{15.64}$$
4. **Positive Functions**:
$$\forall \arg\max f(x) > 0, \forall x \in \mathtt{dom}(f)$$
$$\arg\max f(x) = \arg\min \frac{1}{f(x)} \tag{15.65}$$
5. **Stricly Monotonic Functions**: for all strictly monotonic increasing functions[def. 15.12] g it holds that:
$$\arg\max g(f(x)) = \arg\max f(x) \tag{15.66}$$

Definition 15.33 Max: The maximum of a function f defined on the set D is given by:

$$\max_{x \in D} f(x) = f(x^*) \quad \text{with} \quad \forall x^* \in \arg\max_{x \in D} f(x) \tag{15.67}$$

Definition 15.34 Min: The minimum of a function f defined on the set D is given by:

$$\min_{x \in D} f(x) = f(x^*) \quad \text{with} \quad \forall x^* \in \arg\min_{x \in D} f(x) \tag{15.68}$$

Corollary 15.14 Relationship $\min \leftrightarrow \max$**:**

$$\min_{x \in D} f(x) = -\max_{x \in D} -f(x) \tag{15.69}$$

Property 15.9 Max Identities:

1. **Shifting**:
$$\forall \lambda \text{ const} \qquad \max\{f(x) + \lambda\} = \lambda + \max f(x) \tag{15.70}$$
2. **Positive Scaling**:
$$\forall \lambda > 0 \text{ const} \qquad \max \lambda f(x) = \lambda \max f(x) \tag{15.71}$$
3. **Negative Scaling**:
$$\forall \lambda < 0 \text{ const} \qquad \max \lambda f(x) = \lambda \min f(x) \tag{15.72}$$
4. **Positive Functions**:
$$\forall \arg\max f(x) > 0, \forall x \in \mathtt{dom}(f) \qquad \max \frac{1}{f(x)} = \frac{1}{\min f(x)} \tag{15.73}$$
5. **Stricly Monotonic Functions**: for all strictly monotonic increasing functions[def. 15.12] g it holds that:
$$\max g(f(x)) = g(\max f(x)) \tag{15.74}$$

Definition 15.35 Supremum: The supremum of a function defined on a set D is given by:

$$\sup_{x \in D} f(x) = \{y | y \geqslant f(x), \forall x \in D\} = \min_{y | y \geqslant f(x), \forall x \in D} y \tag{15.75}$$

and is the smallest value y that is equal or greater $f(x)$ for any x ⟺ smallest upper bound.

Definition 15.36 Infinmum: The infinmum of a function defined on a set D is given by:

$$\inf_{x \in D} f(x) = \{y | y \leqslant f(x), \forall x \in D\} = \max_{y | y \leqslant f(x), \forall x \in D} y \tag{15.76}$$

and is the biggest value y that is equal or smaller $f(x)$ for any x ⟺ largest lower bound.

Corollary 15.15 Relationship $\sup \leftrightarrow \inf$**:**

$$\in_{x \in D} f(x) = -\sup_{x \in D} -f(x) \tag{15.77}$$

Note

The supremum/infinmum is necessary to handle unbound function that seem to converge and for which the max/min does not exist as the argmax/argmin may be empty.
E.g. consider $-e^x/e^x$ for which the max/min converges toward 0 but will never reached s.t. we can always choose a bigger x ⇒ there exists no argmax/argmin ⇒ need to bound the functions from above/below ⟺ infinmum/supremum.

Definition 15.37 Time-invariant system (TIS): A function f is called time-invariant, if shifting the input in time leads to the same output shifted in time by the same amount.

$$y(t) = f(x(t), t) \xrightarrow[\forall \tau]{\text{time-invariance}} y(t - \tau) = f(x(t - \tau), t) \tag{15.78}$$

Definition 15.38 Inverse Function $g = f^{-1}$**:**
A function g is the inverse function of the function $f : A \subset \mathbb{R} \to B \subset \mathbb{R}$ if

$$f(g(x)) = x \qquad \forall x \in \mathtt{dom}(g) \tag{15.79}$$

and

$$g(f(u)) = u \qquad \forall u \in \mathtt{dom}(f) \tag{15.80}$$

Property 15.10
Reflective Property of Inverse Functions: f contains (a, b) if and only if f^{-1} contains (b, a).
The line $y = x$ is a symmetry line for f and f^{-1}.

Theorem 15.5 The Existence of an Inverse Function:
A function has an inverse function if and only if it is one-to-one.

Corollary 15.16 Inverse functions and strict monotonicity: If a function f is strictly monotonic [def. 15.14] on its entire domain, then it is one-to-one and therefore has an inverse function.

6. Special Functions

6.1. The Gamma Function

Definition 15.39 The gamma function $\Gamma(\alpha)$**:** Is extension of the factorial function (**??**) to the real and complex numbers (with a positive real part):

$$\Gamma(z) = \int_0^\infty x^{z-1} \mathrm{e}^{-x} \, \mathrm{d}x \qquad \Re(z) > 0 \tag{15.81}$$

$$\Gamma(n) \quad \overset{n \in \mathbb{N}}{\Longleftrightarrow} \quad \Gamma(n) = (n-1)!$$

7. Proofs

Proof 15.4: lemma 15.1 for $\mathcal{C}^1$ functions:
Let $g(t) \equiv f(\mathbf{y} + t(\mathbf{x} - \mathbf{y}))$ from the FToC (theorem 15.2) we know that:

$$\int_0^1 g'(t) \, \mathrm{d}t = g(1) - g(0) = f(\mathbf{x}) - f(\mathbf{y})$$

It then follows from the reverse:

$$\begin{aligned} & |f(\mathbf{x}) - f(\mathbf{y}) - \nabla f(\mathbf{y})^{\mathsf{T}}(\mathbf{x} - \mathbf{y})| \\ \overset{\substack{\text{Chain. R} \\ \text{FToC}}}{=} & \left| \int_0^1 \nabla f(\mathbf{y} + t(\mathbf{x} - \mathbf{y}))^{\mathsf{T}}(\mathbf{x} - \mathbf{y}) \, \mathrm{d}t - \nabla f(\mathbf{y})^{\mathsf{T}}(\mathbf{x} - \mathbf{y}) \right| \\ = & \left| \int_0^1 (\nabla f(\mathbf{y} + t(\mathbf{x} - \mathbf{y})) - \nabla f(\mathbf{y}))^{\mathsf{T}} (\mathbf{x} - \mathbf{y}) \, \mathrm{d}t \right| \\ = & \left| \int_0^1 (\nabla f(\mathbf{y} + t(\mathbf{x} - \mathbf{y})) - \nabla f(\mathbf{y}))^{\mathsf{T}} (\mathbf{x} - \mathbf{y}) \, \mathrm{d}t \right| \\ \overset{\text{C.S.}}{\leqslant} & \left| \int_0^1 \|\nabla f(\mathbf{y} + t(\mathbf{x} - \mathbf{y})) - \nabla f(\mathbf{y})\| \cdot \|\mathbf{x} - \mathbf{y}\| \, \mathrm{d}t \right| \\ \overset{\text{eq. (15.30)}}{=} & \left| \int_0^1 L\|\mathbf{y} + t(\mathbf{x} - \mathbf{y}) - \mathbf{y}\| \cdot \|\mathbf{x} - \mathbf{y}\| \, \mathrm{d}t \right| \\ = & \left| L\|\mathbf{x} - \mathbf{y}\|^2 \int_0^1 t \, \mathrm{d}t \right| = \frac{L}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 \end{aligned}$$

Proof 15.5: **??** for $\mathcal{C}^2$ functions:

$$f(\mathbf{y}) \overset{\text{Taylor}}{=} f(\mathbf{x}) + \nabla f(\mathbf{x})^{\mathsf{T}}(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^{\mathsf{T}} \nabla^2 f(z)(\mathbf{y} - \mathbf{x})$$

Now we plug in $\nabla^2 f(\mathbf{x})$ and recover eq. (15.33):

$$f(\mathbf{y}) \leqslant f(\mathbf{x}) + \nabla f(\mathbf{x})^{\mathsf{T}}(\mathbf{y} - \mathbf{x}) + \frac{1}{2}(\mathbf{y} - \mathbf{x})^{\mathsf{T}} L(\mathbf{y} - \mathbf{x})$$

Proof 15.6: theorem 15.3:
With the definition of convexity for a differentiable function (eq. (15.41)) it follows

$$f(x) - f(y) + \nabla f(y)^{\mathsf{T}}(x - y) \geqslant 0$$

$$\Rightarrow |f(x) - f(y) + \nabla f(y)^{\mathsf{T}}(x - y)|$$

$$\overset{\text{if eq. (15.41)}}{=} f(x) - f(y) + \nabla f(y)^{\mathsf{T}}(x - y)$$

with lemma 15.1 and [def. 15.23] it follows theorem 15.3

Differential Calculus

1. Mean Value Theorem

Theorem 16.1 Mean Value Theorem: Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous function, differentiable on the open interval (a, b), with $a < b$. Then there exist some $c \in (a, b)$ s.t.

$$f'(c) = \frac{f(b) - f(a)}{b - a} = \frac{1}{b - mca} \int_a^b f(x)\,\mathrm{d}x \qquad (16.1)$$

2. The Product Rule

Rule 16.1 (Product /Leibniz Rule).
Let u, v be two differentiable functions $u, v \in \mathcal{C}^1$ then it holds that:

$$\frac{\mathrm{d}\big(u(x)v(x)\big)}{\mathrm{d}x} = (uv)' = u'v + v'u \qquad (16.2)$$

3. The Chain Rule

Formula 16.1 Generalized Chain Rule:
Let $\mathbf{F} : \mathbb{R}^n \mapsto \mathbb{R}^k$ and $\mathbf{G} : \mathbb{R}^k \mapsto \mathbb{R}^m$ be to general maps then it holds:

$$\underbrace{\partial\left(\mathbf{G} \circ \mathbf{F}\right)}_{\mathbb{R}^n \mapsto \mathbb{R}^{m \times n}} = \underbrace{\left(\partial \mathbf{G} \circ \mathbf{F}\right) \cdot \partial \mathbf{F}}_{\mathbb{R}^n \mapsto \left(\mathbb{R}^{m \times k} \cdot \mathbb{R}^{k \times n}\right)} \qquad \begin{aligned} \partial F &: \mathbb{R}^n \mapsto \mathbb{R}^{k \times n} \\ \partial G &: \mathbb{R}^k \mapsto \mathbb{R}^{m \times k} \end{aligned} \qquad (16.3)$$

4. Directional Derivative

5. Partial Differentiation

Definition 16.1 Paritial Derivative:
Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a real valued function, its partial derivative $\partial_i f : \mathbb{R}^n \mapsto \mathbb{R}$ is defined as the directional derivative**??** along the coordinate axis of one of its variables:

$$\partial_i f(\mathbf{x}) = \frac{\partial f}{\partial x_i} = D_{x_i} f = \lim_{h \to 0} \frac{f(\mathbf{x}, x_i \leftarrow x_i + h) - f(\mathbf{x})}{h}$$
$$= \lim_{h \to 0} \frac{f(x_1, \dots, x_i + h, \dots, x_n) - f(x_1, \dots, x_i, \dots, x_n)}{h} \qquad (16.4)$$

5.1. The Gradient

5.1.1. The Nabla Operator

Definition 16.2 Nabla Operator/Del ∇**:**
Given a cartesian coordinate system $\mathbb{R}^n$ with coordinates $x_1, \dots, x_n$ and associated unit vectors $\hat{\mathbf{e}}_1, \dots, \hat{\mathbf{e}}_n$ its *del* operator is defined as:

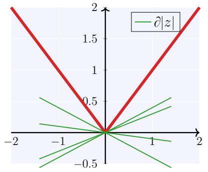
$$\nabla = \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \tilde{\mathbf{e}}_i = \begin{bmatrix} \frac{\partial}{\partial x_1}(\mathbf{x}) \\ \frac{\partial}{\partial x_2}(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_n}(\mathbf{x}) \end{bmatrix} \qquad (16.5)$$

Definition 16.3 Gradient:
Given a *scalar valued* function $f : \mathbb{R}^n \mapsto \mathbb{R}$ its gradient $\nabla f : \mathbb{R}^n \mapsto \mathbb{R}^n$ is defined as vector $\mathbb{R}^n$ of the partial derivatives[def. 16.1] w.r.t. all coordinate axes:

$$\operatorname{grad} f(\mathbf{x}) := \nabla f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \frac{\partial f}{\partial x_2}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{bmatrix} = \left(\frac{\partial f}{\partial \mathbf{x}}\right)^{\mathsf{T}} \qquad (16.6)$$

5.1.2. The Subderivative

Definition 16.4 [proof 16.1]
Subgradient g**:**



Let $f : R^n \mapsto \mathbb{R}$ be a continuous (not necessarily differentiable) function.
$g \in \mathbb{R}^n$ is a subgradient of f at a point $\mathbf{x}_0 \in \mathbb{R}^n$ if it satisfies:

$$g : f(\mathbf{x}) - f(\mathbf{x}_0) \geqslant \mathbf{g}^{\mathsf{T}}(\mathbf{x} - \mathbf{x}_0) \qquad (16.7)$$

Definition 16.5 [example 16.1]
Subderivative $\partial f(\mathbf{x}_0)$**:**
Let $f : R^n \mapsto \mathbb{R}$ be a continuous (not necessarily differentiable) function. The subdifferential of f at a point $\mathbf{x}_0 \in \mathbb{R}^n$ is defined as the set of all possible subgradients[def. 16.4] g:

$$\partial f(\mathbf{x}_0)\left\{g : f(\mathbf{x}) - f(\mathbf{x}_0) \geqslant \mathbf{g}^{\mathsf{T}}(\mathbf{x} - \mathbf{x}_0) \quad \forall \mathbf{x} \in \mathbb{R}^n\right\} \qquad (16.8)$$

Heuristic

We can guess the sub derivative at a point by looking at all the slopes that are smaller then the graph.

5.2. The Jacobian

Definition 16.6
Jacobian/Jacobi Matrix $\mathbf{Df}, \mathbf{J_f}$**:**
Given a *vector valued* function

$$\mathbf{f} : \mathbb{R}^n \mapsto \mathbb{R}^m \qquad \text{its derivative} \qquad \mathbf{J_f} : \mathbb{R}^n \mapsto \mathbb{R}^{m \times n}$$

with components $\partial_{ij}\mathbf{f} = \partial_i f_j : \mathbb{R}^n \mapsto \mathbb{R}$ is a vector valued function defined as:

$$\mathbf{J}(\mathbf{f}(\mathbf{x})) = \mathbf{J_f}(\mathbf{x}) = \mathbf{Df} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\mathbf{x}) = \frac{\partial(f_1, \dots, f_m)}{\partial(x_1, \dots, x_n)}(\mathbf{x}) \qquad (16.9)$$

$$= \begin{bmatrix} \frac{\partial \mathbf{f}}{\partial x_1} \cdots\cdots \frac{\partial \mathbf{f}}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^{\mathsf{T}} \mathbf{f}_1 \\ \vdots \\ \nabla^{\mathsf{T}} f_m \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \frac{\partial f_1}{\partial x_2}(\mathbf{x}) & \cdots\cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{x}) & & \ddots & \frac{\partial f_2}{\partial x_n}(\mathbf{x}) \\ \vdots & & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}) & \frac{\partial f_m}{\partial x_2}(\mathbf{x}) & & \frac{\partial f_m}{\partial x_n}(\mathbf{x}) \end{bmatrix}$$

Explanation 16.1. *Rows of the Jaccobian are transposed gradients*[def. 16.3] *of the component functions $f_1, \dots, f_m$.*

Corollary 16.1 :

6. Second Order Derivatives

Definition 16.7 Second Order Derivative $\frac{\partial^2}{\partial x_i \partial x_j}$**:**

Theorem 16.2
Symmetry of second derivatives/Schwartz's Theorem:
Given a continuous and twice differentiable function $f : \mathbb{R}^n \mapsto \mathbb{R}$ then its second order partial derivatives commute:

$$\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j} = \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}$$

6.1. The Hessian

Definition 16.8 Hessian Matrix:
Given a function $f : \mathbb{R} \mapsto \mathbb{R}^n$ its Hessian$\in \mathbb{R}^{n \times n}$ is defined as:

$$\mathbf{H}(\mathbf{f})(\mathbf{x}) = \mathbf{H}_f(\mathbf{x}) = \mathbf{J}(\nabla \mathbf{f}(\mathbf{x}))^T \qquad (16.10)$$

$$= \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) \cdots\cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{x}) \cdots\cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}) \end{bmatrix}$$

and it corresponds to the Jacobian of the Gradient.
Due to the differentiability and theorem 16.2 it follows that the Hessian is (if it exists):

- Symmetric
- Real

Corollary 16.2 Eigenvector basis of the Hessian: Due to the fact that the Hessian is real and symmetric we can decompose it into a set of real eigenvalues and an orthogonal basis of eigenvectors $\{(\lambda_1, \mathbf{v}_1), \dots, \lambda_n, \mathbf{v}_n\}$.
Not let $\mathbf{d}$ be a directional unit vector then the second derivative in that direction is given by:

$$\mathbf{d}^{\mathsf{T}}\mathbf{H}\mathbf{d} \quad \iff \quad \mathbf{d}^{\mathsf{T}} \sum_{i=1}^{n} \lambda_i \mathbf{v}_i \quad \overset{\text{if } \mathbf{d}=\mathbf{v}_j}{\iff} \quad \mathbf{d}^{\mathsf{T}} \lambda_j \mathbf{v}_j$$

- The eigenvectors that have smaller angle with $\mathbf{d}$ have bigger weight/eigenvalues
- The minimum/maximum eigenvalue determines the minimum/maximum second derivative

7. Extrema

Definition 16.9 Critical/Stationary Point: Given a function $f : \mathbb{R}^n \mapsto \mathbb{R}$, that is differentiable at a point $\mathbf{x}_0$ then it is called a critical point if the functions derivative vanishes at that point:

$$f'(\mathbf{x}_0) = 0 \qquad \iff \qquad \nabla_{\mathbf{x}} f(\mathbf{x}_0) = 0$$

Corollary 16.3 Second Derivative Test $f : \mathbb{R} \mapsto \mathbb{R}$:
Suppose $f : \mathbb{R} \mapsto \mathbb{R}$ is twice differentiable at a stationary point x [def. 16.9] then it follows that:

- $f''(x) > 0 \iff$ $f'(x + \epsilon) > 0$ slope points uphill; $f'(x - \epsilon) < 0$ slope points downhill; $f(x)$ is a local minimum
- $f''(x) < 0 \iff$ $f'(x + \epsilon) > 0$ slope points downhill; $f'(x - \epsilon) < 0$ slope points uphill; $f(x)$ is a local maximum

$\epsilon > 0$ sufficiently small enough

Corollary 16.4 Second Derivative Test $f : \mathbb{R}^n \mapsto \mathbb{R}$:
Suppose $f : \mathbb{R}^n \mapsto \mathbb{R}$ is twice differentiable at a stationary point $\mathbf{x}$ [def. 16.9] then it follows that:

- If $\mathbf{H}$ is p.d $\iff \forall \lambda_i > 0 \in \mathbf{H} \rightarrow f(\mathbf{x})$ is a local min.
- If $\mathbf{H}$ is n.d $\iff \forall \lambda_i < 0 \in \mathbf{H} \rightarrow f(\mathbf{x})$ is a local max.
- If $\exists \lambda_i > 0 \in \mathbf{H}$ and $\exists \lambda_i < 0 \in \mathbf{H}$ then $\mathbf{x}$ is a local maximum in one cross section of f but a local minimum in another
- If $\exists \lambda_i = 0 \in \mathbf{H}$ and all other eigenvalues have the same sign the test is inclusive as it is inconclusive in the cross section corresponding to the zero eigenvalue.

Note

If $\mathbf{H}$ is positive definite for a minima $\mathbf{x}^*$ of a *quadratic* function f then this point must be a global minimum of that function.

8. Proofs

Proof 16.1: Definition 16.4 $f(\mathbf{x}) \geqslant f(\mathbf{x}_0) + \mathbf{g}^{\mathsf{T}}(\mathbf{x} - \mathbf{x}_0) \quad \forall \mathbf{x} \in \mathbb{R}^n$ corresponds to a line (see formula 15.1) at the point $\mathbf{x}_0$ with slope $\mathbf{g}^{\mathsf{T}}$.
Thus we search for all lines with smaller slope then function graph.

9. Examples

Example 16.1 Subderivatives Absolute Value Function
$|x|$**:** $f : \mathbb{R} \mapsto \mathbb{R}$ with $f(x) = |x|$ at the point $x = 0$ it holds:

$$f(x) - f(0) \geqslant gx \qquad \Longrightarrow \qquad \text{the interval } [-1; 1]$$

For $x \neq 0$ the subgradient is equal to the gradient. Thus it follows for the subderivatives/differentials:

$$\partial|x| = \begin{cases} -1 & \text{if } x < 0 \\ [-1, 1] & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases}$$

Integral Calculus

Theorem 17.1 Important Integral Properties:

Addition $$\int\limits_{a}^{b} f(x)\,\mathrm{d}x = \int\limits_{a}^{c} f(x)\,\mathrm{d}x + \int\limits_{c}^{b} f(x)\,\mathrm{d}x \tag{17.1}$$

Reflection $$\int\limits_{a}^{b} f(x)\,\mathrm{d}x = -\int\limits_{b}^{a} f(x)\,\mathrm{d}x \tag{17.2}$$

Translation $$\int\limits_{a}^{b} f(x)\,\mathrm{d}x \overset{u:=x\pm c}{=} \int\limits_{a\pm c}^{b\pm c} f(x\mp c)\,\mathrm{d}x \tag{17.3}$$

f **Odd** $$\int\limits_{-a}^{a} f(x)\,\mathrm{d}x = 0 \tag{17.4}$$

f **Even** $$\int\limits_{-a}^{a} f(x)\,\mathrm{d}x = 2\int\limits_{0}^{a} f(x)\,\mathrm{d}x \tag{17.5}$$

Proof 17.1: **eqs. (17.4) and (17.5)**

$$\begin{aligned}
I := \int\limits_{-a}^{a} f(x)\,\mathrm{d}x &= \int\limits_{-a}^{0} f(x)\,\mathrm{d}x + \int\limits_{0}^{a} f(x)\,\mathrm{d}x \\
&\overset{\substack{t=-x\\ dt=-dx}}{=} -\int\limits_{a}^{0} f(-x)\,\mathrm{d}x + \int\limits_{0}^{a} f(x)\,\mathrm{d}x \\
&= \int\limits_{0}^{a} f(-x) + f(x)\,\mathrm{d}x = \begin{cases} 0 & \text{if} \quad f \quad \text{odd} \\ 2I & \text{if} \quad f \quad \text{even} \end{cases}
\end{aligned}$$

Definition 17.1 Integration by Parts:

$$\int_{a}^{b} u\,\mathrm{d}v = uv\Big|_{a}^{b} - \int_{a}^{b} v\,\mathrm{d}u \tag{17.6}$$

1. Integral Theorems

1.1. Greens Identities

Theorem 17.2 Greens First Identity:
Let $\bar{\Omega} = \Omega \cup \partial\Omega$, for all vector fields $\mathbf{j} \in \left(\mathcal{C}^1_{\mathrm{pw}}(\bar{\Omega})\right)^d$ and scalar functions $v \in \mathcal{C}^1_{\mathrm{pw}}\left(\bar{\Omega}\right)$ it holds:

$$\int_{\Omega} \mathbf{j}^{\mathsf{T}} \operatorname{grad} v\,\mathrm{d}\mathbf{x} = -\int_{\Omega} \operatorname{div} \mathbf{j} v\,\mathrm{d}\mathbf{x} + \int_{\partial\Omega} \mathbf{j}^{\mathsf{T}} \mathbf{n} v\,\mathrm{d}S \tag{17.7}$$

Differential Equations

Definition 17.2 [??]
Differential Operator:
A differential operator $\mathscr{L}$ is a mapping of a suitable function space onto another function space, involving only values of the function argument and its derivatives in the same point:

$$\mathscr{L} : C^n(\Omega) \mapsto C^k(\Omega), \quad k < n$$

Note: $\mathscr{L}$ is a differential operator of order $k - n$.

Definition 17.3 Linear Differential Operator:
Is a differential operator $\mathscr{L}$ that satisfies:

$$\mathscr{L}(\alpha u + \beta v) = \alpha\,\mathscr{L}(u) + \beta\,\mathscr{L}(v) \qquad \forall \alpha, \beta \in \mathbb{R} \tag{17.8}$$

Ordinary Differential Quations

Partial Differential Equations (PDE)s

Definition 19.1 Partial Differential Equation:
Let $\mathbf{u} = \mathbf{u}(x_1, \dots, x_n) : \mathbb{R}^k \mapsto \mathbb{R}^l$ be an unknown function depending on $\mathbf{x} = (x_1 \cdots\cdots x_k)$ and let f be a known function.
The known function $\mathscr{F}$, depending on differentials of the unknown function $\mathbf{u}$ is called a Partial Differential equation:

$$\mathscr{F}\left(\mathbf{u}, \frac{\partial \mathbf{u}}{\partial x_1}, \dots, \frac{\partial \mathbf{u}^n}{\partial x_i^j, \dots, \partial x_j^r}, f\right) = \mathscr{F}(\mathbf{u}, D\mathbf{u}, \dots, D^n\mathbf{u}, f) = \mathbf{0}$$

or

$$\mathscr{L}(\mathbf{u}) = f \quad \text{in } \Omega \tag{19.1}$$

Corollary 19.1 Dependent Variables:

$$\mathbf{u} : \mathbb{R}^k \mapsto \mathbb{R}^l \tag{19.2}$$

Corollary 19.2 Independent Variables:

$$\mathbf{x} = (x_1 \cdots\cdots x_k) \tag{19.3}$$

Definition 19.2 Order n:
Is the highest partial derivative that appears in a PDE.

1. Algebraic Types

1.1. Linearity

Definition 19.3 [??]
Linear PDEs:
A linear PDE naturally defines a linear operator [def. 17.3]. A linear PDE must be linear reagarding the unkown function $\mathbf{u}$. **In other words** all dependent variables $\mathbf{u}$ and their corresponding derivatives depend only on the independent variables $x_1, x_2, \dots, x_m$:

$$a(x,y)\mathbf{u}_x + b(x,y)\mathbf{u}_y + c(x,y)\mathbf{u} = d(x,y) \tag{19.4}$$

Definition 19.4 [??]
Semilinear PDEs:
Are PDEs whose coefficients of the highest order n-terms are functions depending only on the independent variables **but** not onto the dependent variables $\mathbf{u}$ or their derivatives.
Thus the PDE is linear regarding to the highest order terms:

$$a(x,y)\mathbf{u}_x + b(x,y)\mathbf{u}_y = c(x,y,\mathbf{u}) \tag{19.5}$$

Definition 19.5 [??]
Quasilinear PDEs:
Are PDEs whose coefficients of the highest order (n) terms are functions only depending on the independent variables **and** on the dependent variables $\mathbf{u}$ and their derivatives up to an order $m < n$, that is smaller than the highest order terms n:

$$a(x,y,\mathbf{u})\mathbf{u}_x + b(x,y,\mathbf{u})\mathbf{u}_y = c(x,y,\mathbf{u}) \tag{19.6}$$

Definition 19.6 [??]
Fully Non-linear PDEs:
Are PDEs where all terms of the highest order n are non-linear:

$$a(x,y,\mathbf{u},\mathbf{u}')\mathbf{u}_x + b(x,y,\mathbf{u},\mathbf{u}')\mathbf{u}_y = c(x,y,\mathbf{u}) \tag{19.7}$$

Note: $\neg$(**Quasilinear** $\Leftrightarrow$ **Fully Nonlinear**)

1.2. Homogenity

Definition 19.7 Homogenuous $\mathscr{L}(\mathbf{u}) = 0$:
All terms depend on $\mathbf{u}$ or on derivatives of $\mathbf{u}$.

Definition 19.8 Non-Homogenuous $\mathscr{L}(\mathbf{u}) = f$:
Their exists non-zero terms f that do not depend on $\mathbf{u}$ or on derivatives of $\mathbf{u}$.

1.3. Constant Coefficients

Definition 19.9 PDEs with Constant Coefficients:
Is a PDE whose coefficients $a, b, c, \dots$ are constants i.e. independent variables.

1.4. 2nd-Order Linear PDEs in two variables

Definition 19.10
2nd-Order Linear PDEs in two Variables:

$$\mathscr{L}(\mathbf{u}) = a\mathbf{u}_{xx} + 2b\mathbf{u}_{xy} + c\mathbf{u}_{yy} + d\mathbf{u}_x + e\mathbf{u}_y + f\mathbf{u} = g \tag{19.8}$$

where $a, b, \dots, g$ are functions depending on x and y.

Definition 19.11 Principal Part: Is the operator $\mathscr{L}_0$, that consists of the second-(=highest) order parts of $\mathscr{L}$:

$$\mathscr{L}_2(\mathbf{u}) := a\mathbf{u}_{xx} + 2b\mathbf{u}_{xy} + c\mathbf{u}_{yy}$$

Definition 19.12 PDEs Discriminante: Is defined by:

$$\underline{\delta(\mathscr{L})} := -\det\begin{pmatrix} a & b \\ b & c \end{pmatrix} = b^2 - ac \tag{19.9}$$

Explanation 19.1. *It turns out that many fundamental properties of the solution of eq. (19.8) are determined by its principal part, or rather by the sign of the discriminant $\delta(\mathscr{L})$.*

Definition 19.13 [??]
Parabolic PDEs: Let[def. 19.10] be a PDE defined on $\Omega \subset \mathbb{R}^2$, then the PDE is called hyperbolic if:

$$\delta(\mathscr{L}) = b^2 - ac = 0 \tag{19.10}$$

Definition 19.14 [??]
Hyperbolic PDEs: Let[def. 19.10] be a PDE defined on $\Omega \subset \mathbb{R}^2$, then the PDE is called hyperbolic if:

$$\delta(\mathscr{L}) = b^2 - ac > 0 \tag{19.11}$$

Definition 19.15 [??]
Parabolic PDEs: Let[def. 19.10] be a PDE defined on $\Omega \subset \mathbb{R}^2$, then the PDE is called elliptic if:

$$\delta(\mathscr{L}) = b^2 - ac < 0 \tag{19.12}$$

Explanation 19.2.
The reason for this categorization are normal quadratic equations in two variables:

$$Ax^2 + By^2 + Cxy + Dx + Ey + f = 0$$

If $B^2 - 4AC = 0 \Leftrightarrow$ *the equation is a parabola.*
If $B^2 - 4AC > 0 \Rightarrow$ *the equation is a hyperbola.*
If $B^2 - 4AC < 0 \Rightarrow$ *the equation is an ellipse.*

2. Method Of Characteristics

Is a method that makes use of geometrical aspects in order to solve **1st-order PDEs** with two variables by constructing integral surfaces and can be used to solve PDEs of the type:

$$\begin{aligned}
&\text{Linear:} && a(x,y)\mathbf{u}_x + b(x,y)\mathbf{u}_y = c(x,y) && (19.13)\\
&\text{Semilin.:} && a(x,y)\mathbf{u}_x + b(x,y)\mathbf{u}_y = c(x,y,\mathbf{u}) && (19.14)\\
&\text{Quasilin.:} && a(x,y,\mathbf{u})\mathbf{u}_x + b(x,y,\mathbf{u})\mathbf{u}_y = c(x,y,\mathbf{u}) && (19.15)
\end{aligned}$$

Formula 19.1 Method of Characteristics:
$x := x(r;s) \qquad y := y(r;s) \qquad z := u(r;s)$
Parameter.: $\lambda(r;s) := x(r;s)\mathbf{e}_x + y(r;s)\mathbf{e}_y + z(r;s)\mathbf{e}_z$

$$\frac{\partial \lambda}{\partial r}(r;s) = (a, b, c)$$

E.g. $v := v(x(r;s), y(r;s), z(r;s))$

$$\frac{\partial x}{\partial r}(r;s) = \dot{x} = a(\lambda_s(r))$$
$$\frac{\partial y}{\partial r}(r;s) = \dot{y} = b(\lambda_s(r))$$
$$\frac{\partial z}{\partial r}(r;s) = \dot{z} = c(\lambda_s(r))$$

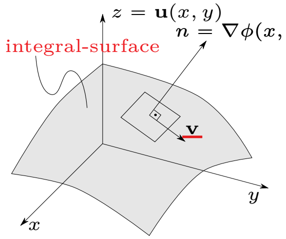
Compact:
$\dot{x} = a(x,y,u) \qquad \dot{y} = b(x,y,u) \qquad \dot{u} = c(x,y,u)$
I.C.: $x(0;s) = x_0(s) \quad y_0(0;s) = y_0(s) \quad u(0;s) = u_0(s)$

Definition 19.16 Integral Surface ϕ:
An function $\phi : \mathbb{R}^3 \mapsto \mathbb{R}$ is a an *integral surface* of a vector field $\mathbf{V} : \mathbb{R}^3 \mapsto \mathbb{R}^3$ if ϕ is a surface that has in every point a tangent plane containing a vector $\mathbf{v} = \begin{pmatrix} a & b & c \end{pmatrix}$ of $\mathbf{V}$.

Corollary 19.3 PDEs and Integral Surfaces:
The solution of a PDE $\mathbf{u}(x,y)$ can be thought of as an integral surface:

$$z = u(x,y) \qquad \text{or implicitly} \qquad \phi(x,y,z) = u(x,y) - z \tag{19.16}$$

Explanation 19.3 ([proof 8.1]
Integral Surface and PDEs).
The solution $\mathbf{u}(x,y)$ of eq. (19.13) can be sought of as an surface $z = \mathbf{u}(x,y)$ in $\mathbb{R}^3$ or in implicit form $\phi(x,y,z) := \mathbf{u}(x,y) - z$.



Let:$\underline{\mathbf{n}(x,y)} := \operatorname{grad}\phi = \begin{pmatrix}\phi_x\\ \phi_y\\ \phi_z\end{pmatrix} = \begin{pmatrix}\mathbf{u}_x\\ \mathbf{u}_y\\ -1\end{pmatrix}$ **and**

Let $\mathbf{V} := \begin{pmatrix} a(x,y)\\ b(x,y)\\ c(x,y)\end{pmatrix}$ *be a vector field* $\mathbb{R}^3 \mapsto \mathbb{R}^3$ *and*

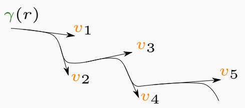
$$\underline{\mathbf{n}(x,y)} := \operatorname{grad}\phi = \begin{pmatrix}\phi_x\\ \phi_y\\ \phi_z\end{pmatrix} = \begin{pmatrix}\mathbf{u}_x\\ \mathbf{u}_y\\ -1\end{pmatrix}$$

Idea: *we can rewrite eq. (19.13) as:*

$$\left\langle \begin{pmatrix} a & b & c\end{pmatrix}^\top, \nabla\phi(x,y,z)\right\rangle = \left\langle \begin{pmatrix} a(x,y)\\ b(x,y)\\ c(x,y)\end{pmatrix}, \underline{\begin{pmatrix}\mathbf{u}_x\\ \mathbf{u}_y\\ -1\end{pmatrix}}\right\rangle = 0$$

Geometric Interpretation:
$\mathbf{v}$ *is orthogonal to the normal* $\underline{\mathbf{n}}$ *for all points* $(x, y, \mathbf{u}(x,y))$.
Hence *every vector* $\mathbf{v} = \begin{pmatrix} a & b & c\end{pmatrix}^\top$ *lies in the tangent plane containing* ϕ.
Consequently *in order to find a surface* ϕ *(and thus also a solution* $\mathbf{u}$*), we need to search for* ϕ *s.t. the vector* $\mathbf{v}$ *lies in the tangent plane for every possible point of* ϕ.

Idea



We first simplify the task and start by constructing/finding integral curves λ and then we construct the integral surface ϕ out of this curves.

3. Linear Equations

Definition 19.17
Characteristic/Integral Curve $\lambda_s(r) = \lambda(r;s)$:
Given a vector field $\mathbf{V}$ an integral curve $\lambda(r)$ of that vector field, is a curve parameterized by parameter r:

$$\lambda(r) := x(r)\mathbf{e}_x + y(r)\mathbf{e}_y + z(r)\mathbf{e}_z = \begin{pmatrix} x(r)\\ y(r)\\ z(r)\end{pmatrix} \tag{19.17}$$

s.t. at each point r of the curve a vector $\mathbf{v}$ of the vector field:

$$\mathbf{v} = \begin{pmatrix} a(x(r), y(r))\\ b(x(r), y(r))\\ c(x(r), y(r))\end{pmatrix} \in \mathbf{V} \tag{19.18}$$

is tangent to the curve:

$$\frac{\mathrm{d}\lambda(r)}{\mathrm{d}r} = \mathbf{V}(\lambda(r)) = \begin{pmatrix} \underline{a(x(r), y(r))}\\ \underline{b(x(r), y(r))}\\ \underline{c(x(r), y(r))}\end{pmatrix} = \begin{pmatrix} a(\lambda(r))\\ b(\lambda(r))\\ c(\lambda(r))\end{pmatrix} \tag{19.19}$$

Definition 19.18 Characteristic Equations:
The set of ordinary differential equations of a PDE arising from Equation (19.19) are called *characteristic equations*:

$$\frac{\partial x(r)}{\partial r} = \dot{x} = \underline{a(\lambda(r))} = a(r) \tag{19.20}$$
$$\frac{\partial y(r)}{\partial r} = \dot{y} = \underline{b(\lambda(r))} = b(r) \tag{19.21}$$
$$\frac{\partial z(r)}{\partial r} = \dot{z} = \underline{c(\lambda(r))} = c(r) \tag{19.22}$$

Problem: in order to get a unique solution we need to specify initial conditions.
Idea: If a characteristc has an arbitrary point in common with the integral surface ϕ then the whole characteristic λ will lie in the integral surface.

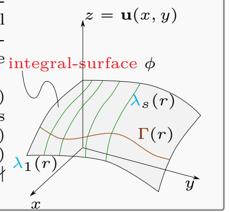
Proof 19.1: **Let**: $\phi(\lambda(r)) = u(x(r), y(r)) - z(r)$

$$\Rightarrow \frac{\mathrm{d}\phi}{\mathrm{d}r} = u_x \underline{\frac{\mathrm{d}x}{\mathrm{d}r}} + u_y \underline{\frac{\mathrm{d}y}{\mathrm{d}r}} - 1\,\underline{\frac{\mathrm{d}z}{\mathrm{d}r}} =$$
$$= \begin{pmatrix} u_x\\ u_y\\ -1\end{pmatrix}\begin{pmatrix} a(x(r), y(r))\\ b(x(r), y(r))\\ c(x(r), y(r))\end{pmatrix} = \begin{pmatrix} u_x\\ u_y\\ -1\end{pmatrix}\dot{\lambda}(r) = 0$$

Thus: $\phi(\lambda(r_0)) = 0 \iff \phi(\lambda(r)) = 0, \ \forall r$

Definition 19.19
Characteristic (Curve) $\lambda_s(r) = \lambda(r;s)$:
is an integral curve of the vector field $\mathbf{V}$ that is uniequely determined by a parameter s.

Consequence: For every characteristic s we need to specify one inital point on the integral surface in order to have all the characteristics lie within the integralsurface.
Idea: we define another curve $\Gamma(s)$ on the integralsurface that transvers all the characteristic curves $\lambda_s(r)$ transversal (=angle beetween $\Gamma(s)$ and $\lambda_s(r)$ is never zero $\Leftrightarrow \Gamma(s) \nparallel \lambda_s(r)$).



Definition 19.20 Inital Condition: $s \mapsto \Gamma(s), \ \Gamma : \mathbb{R} \mapsto \mathbb{R}^3$

$$\lambda_s(r) = \begin{pmatrix} x_s(r)\\ y_s(r)\\ z_s(r)\end{pmatrix}, \quad \Gamma(s) = \begin{pmatrix} \underline{x_0(s)}\\ \underline{y_0(s)}\\ \underline{z_0(s)}\end{pmatrix} \qquad \lambda_s(0) \overset{!}{=} \Gamma(s)$$

$$\Rightarrow \underline{x_s}(0) = \underline{x_0(s)} \qquad \underline{y_s}(0) = \underline{y_0(s)} \qquad \underline{z_s}(0) = \underline{z_0(s)}$$

Definition 19.21
Projected Characteristic Curves $\gamma(\tau)$:
Are curves in the plane of the independent variables of our PDE, along which u is constant or satisfies certain conditions. If u is constant along $g(\tau)$ then the initial data is simply propagated along those characteristic curves:

$$\frac{\mathrm{d}}{\mathrm{d}\gamma} u(\gamma(\tau), \tau) = 0 \iff u(\gamma(\tau), \tau) = u_0(\gamma(\tau)) \tag{19.23}$$

Hint: If the PDE is linear, then the two first characteristics do not depend on u and can be solved directly, u will then be constant along those characteristics:

$a(x,y)\mathbf{u}_x + b(x,y)\mathbf{u}_y = c(x,y)$

$$\frac{\mathrm{d}x}{\mathrm{d}r} = a \qquad \frac{\mathrm{d}y}{\mathrm{d}r} = b \qquad \frac{\mathrm{d}u}{\mathrm{d}r} = c \quad \textbf{implies} \quad \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{b(x,y)}{a(x,y)}$$

Hint: If we divide the PDE by a we have to solve a PDE less, beacause the first ODE will allways be:

$\dot{x} = 1 \Rightarrow \qquad x = r \Rightarrow \qquad x_s(r) = x_0(s)$

4. Quasilinear Equations

Solving Quasilinear Equations

$$a(x,y,u)\mathbf{u}_x \quad + b(x,y,u)\mathbf{u}_y \quad = c(x,y,u)$$

$$u_{|\Gamma}(r,s) = \phi(s)$$

$$\frac{\mathrm{d}x}{\mathrm{d}r} = a(x,y,u) \quad \frac{\mathrm{d}y}{\mathrm{d}r} = b(x,y,u) \quad \frac{\mathrm{d}u}{\mathrm{d}r} = c(x,y,u)$$

$$x_s(0) = x_0(s) \quad y_s(0) = y_0(s) \quad z_s(0) = \phi(s)$$

Results

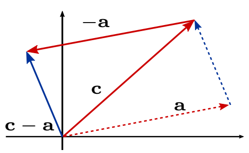
Now the projected characteristic curves may depend on u as well as on x,y. **Thus** the first two characteristics are no longer decoupled form the third one.

1. We may get projected characteristic curves crossing themselfs.
2. u is no longer constant along the projected characteristic curves, rather the PDE reduces to an ODE satisfying certain conditions along this curves.

Linear Algebra

1. Vectors

Definition 20.1 Vector Substraction:



$$\mathbf{b} = \mathbf{c} - \mathbf{a} \qquad (20.1)$$

2. Linear Systems of Equations

2.1. Gaussian Elimination

2.1.1. Rank

Definition 20.2 Matrix Rank $\mathfrak{rank}$:
The ranks of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is defined as the the dimension[def. 20.11] of the vector space spaned[def. 20.7] by its row or column vectors:

$$\begin{aligned} \mathfrak{rank}(\mathbf{A}) &= \dim\left(\{\mathbf{a}_{:,1}, \dots, \mathbf{a}_{:,n}\}\right) \\ &= \dim\left(\{\mathbf{a}_{1,:}, \dots, \mathbf{a}_{m,:}\}\right) \\ &\overset{\text{def. } 20.48}{=} \dim(\mathfrak{R}(\mathbf{A})) \end{aligned} \qquad (20.2)$$

Corollary 20.1 :

- The column-and row-ranks of a square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ are equal.
- The rank of a non-symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is limited by the smaller dimension:

$$\mathfrak{rank}(\mathbf{A}) \leqslant \min\{n, m\} \qquad (20.3)$$

Property 20.1 Rank of Matrix Product: Let $\mathbf{A} \in \mathbb{R}^{m,n}$ and $\mathbf{B} \in \mathbb{R}^{n,p}$ then the rank of the matrix product is limited:

$$\mathfrak{rank}(\mathbf{AB}) \leqslant \min\{\mathfrak{rank}(\mathbf{A}), \mathfrak{rank}(\mathbf{B})\} \qquad (20.4)$$

3. Sparse Linear Systems

Definition 20.3 Sparse Matrix $\mathbf{A} \in \mathbb{K}^{m,n}, m, n \in \mathbb{N}_{>0}$:
A matrix $\mathbf{A}$ is sparse if:

$$\text{nnz}(\mathbf{A}) \ll mn \qquad \mathbf{A} \in \mathbb{K}^{m,n}, m, n \in \mathbb{N}_{>0} \qquad (20.5)$$

$$\text{nnz} := \#\left\{(i, j) \in \{1, \dots, m\} \times \{1, \dots, n\} : a_{i,j} \neq 0\right\}$$

4. Vector Spaces

4.1. Vector Space

Definition 20.4 Vector Space: TODO

4.2. Vector Subspace

Definition 20.5 Vector Subspaces:
A non-empty subset U of a $\mathbb{K}$-vector space $\mathcal{V}$ is called a subspace of $\mathcal{V}$ if it satisfies:

$$\mathbf{u}, \mathbf{v} \in U \implies \mathbf{u} + \mathbf{v} \in U \qquad (20.6)$$

$$\mathbf{u} \in U \implies \lambda\mathbf{u} \in U \qquad \forall\lambda \in \mathbb{K} \qquad (20.7)$$

Definition 20.6 Linearcombination:
Let $X = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subset \mathcal{V}$ be a non-empty and finite subset of vectors of an $\mathbb{K}$-vector space $\mathcal{V}$. A *linear combination* of X is a combination of the vectors defined as:

$$\mathbf{v} = \sum_{i=1}^{n} \lambda_i \mathbf{v}_i = \alpha_1\mathbf{v}_1 + \dots + \alpha_n\mathbf{v}_n \qquad \alpha_i \in \mathbb{K} \qquad (20.8)$$

Definition 20.7
Span/Linear Hull $\langle X \rangle$:
Is the set of all possible linear combinations[def. 20.6] of finite set $X = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subset \mathcal{V}$ of a $\mathbb{K}$ vector space $\mathcal{V}$:

$$\langle X \rangle = \text{span}(X) = \left\{\mathbf{v} \middle| \sum_{i=1}^{n} \alpha_i\mathbf{v}_i, \forall\alpha_i \in \mathbb{K}\right\} \qquad (20.9)$$

Definition 20.8 Generating Set: A *generating set* of vectors $X = \{\mathbf{v}_1, \dots, \mathbf{v}_m\} \in \mathcal{V}$ of a vector spaces $\mathcal{V}$ is a set of vectors that *span*[def. 20.7] $\mathcal{V}$:

$$\text{span}(\mathbf{v}_1 \dots, \mathbf{v}_m) = \mathcal{V} \qquad (20.10)$$

Explanation 20.1 (Definition 20.8).

The generating set of vector space (or set of vectors) $\mathcal{V} \overset{i.e.}{=} \mathbb{R}^n$ *is a subset* $X = \{\mathbf{v}_1, \dots, \mathbf{v}_m\} \subset \mathcal{V}$ *s.t. every element of* $\mathcal{V}$ *can be produced by* $\text{span}(X)$.

Definition 20.9 Linear Independence: A set of vector $\{\mathbf{v}_1, \dots, \mathbf{v}_n\} \in \mathcal{V}$ is called linear independent if the satisfy:

$$\mathbf{v} = \sum_{i=1}^{n} \lambda_i\mathbf{v}_i = \mathbf{0} \iff \alpha_1 = \dots = \alpha_n = 0 \qquad (20.11)$$

Corollary 20.2 : A set of vector $\{\mathbf{x}_1, \dots, \mathbf{x}_n\} \in \mathcal{V}$ is called linear independent, if for every subset $X = \mathbf{x}_1, \dots, \mathbf{x}_m \subseteq \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ it holds that:

$$\langle X \rangle \subsetneq \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \qquad (20.12)$$

4.3. Basis

Definition 20.10 Basis $\mathfrak{B}$:
A subset $\mathfrak{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of a $\mathbb{K}$-vector space $\mathcal{V}$ is called a basis of $\mathcal{V}$ if:

$$\langle\mathfrak{B}\rangle = \mathcal{V} \quad \text{and} \quad \mathfrak{B} \text{ is a linear independent generating set} \qquad (20.13)$$

Corollary 20.3 : The unit vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ build a standard basis of the $\mathbb{R}^n$.

Corollary 20.4 Basis Representation:
Let $\mathfrak{B}$ be a basis of a $\mathbb{K}$-vector space $\mathcal{V}$, then it holds that every vector $\mathbf{v} \in \mathcal{V}$ can be represented as a linear combination[def. 20.6] of $\mathfrak{B}$ by a unique set of coefficients α_i:

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i\mathbf{b}_i \qquad \begin{matrix} \alpha_1, \dots, \alpha_n \in \mathbb{K} \\ \mathbf{b}_1, \dots, \mathbf{b}_n \in \mathfrak{B} \end{matrix} \qquad (20.14)$$

4.3.1. Dimensionality

Definition 20.11 Dimension of a vector space $\dim(\mathcal{V})$:
Let $\mathcal{V}$ be a vector space. The dimension of $\mathcal{V}$ is defined as the number of necessary basis vectors $\mathfrak{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ in order to span $\mathcal{V}$:

$$\dim(\mathcal{V}) := |\mathfrak{B}| = n \in \mathbb{N}_0 \qquad (20.15)$$

Corollary 20.5 : n-linearly independent vectors of a $\mathbb{K}$-vector space $\mathcal{V}$ with finite dimension n constitute a basis.

Note

If $\mathcal{V}$ is infinite $\dim(\mathcal{V}) = \infty$.

4.4. Affine Subspaces

Definition 20.12 Affine Subspaces: Given a $\mathbb{K}$-vector space $\mathcal{V}$ of dimension $\dim(\mathcal{V}) \geqslant 2$ a sub vector space[def. 20.5] U of $\mathcal{V}$ defined as:

$$\mathcal{W} := \mathbf{v} + U = \{\mathbf{v} + \mathbf{x} | \mathbf{x} \in U\} \qquad \mathbf{v} \in \mathcal{V} \qquad (20.16)$$

Corollary 20.6 Direction: The sub vector spaces U are called *directions* of $\mathcal{V}$ and it holds:

$$\dim(\mathcal{W}) := \dim(U) \qquad (20.17)$$

4.4.1. Hyperplanes

Definition 20.13 Hyperplane $\mathcal{H}$:
A hyperplane is a $d-1$ dimensional subspace of an d-dimensional ambient space that can be specified by the hess normal form[def. 20.14]:

$$\mathcal{H} = \left\{\mathbf{x} \in \mathbb{R}^d \middle| \hat{\mathbf{n}}^\intercal\mathbf{x} - d = 0\right\} \qquad (20.18)$$

Corollary 20.7 Half spaces: A hyperplane $\mathcal{H} \in \mathbb{R}^{d-1}$ separates its d-dimensional ambient space into two half spaces:

$$\mathcal{H}^+ = \left\{x \in \mathbb{R}^d \middle| \tilde{\mathbf{n}}^\intercal\mathbf{x} + b > 0\right\} \qquad (20.19)$$

$$\mathcal{H}^- = \left\{x \in \mathbb{R}^d \middle| \tilde{\mathbf{n}}^\intercal\mathbf{x} + b < 0\right\} = \mathbb{R}^d - \mathcal{H}^+ \qquad (20.20)$$

Notes

Hyperplanes in $\mathbb{R}^2$ are lines and hyperplanes in R^3 are lines.

Hess Normal Form

Definition 20.14 Hess Normal Form:
Is an equation to describe hyperplanes[def. 20.13] in $\mathbb{R}^d$:

$$\mathbf{r}^\intercal\tilde{\mathbf{n}} - d = 0 \iff \tilde{\mathbf{n}}^\intercal(\mathbf{r} - \mathbf{r}_0) \qquad \mathbf{r}_0 := \mathbf{r}^\intercal d \geqslant 0 \qquad (20.21)$$

where all points described by the vector $\mathbf{r} \in \mathbb{R}^d$, *that satisfy* this equations lie on the hyperplane.

Note

The direction of the unit normal vector is usually chosen s.t. $\mathbf{r}^\intercal\tilde{\mathbf{n}} \geqslant 0$.

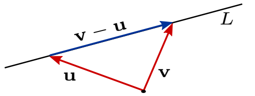
4.4.2. Lines

Definition 20.15 Lines: Lines are a set[def. 11.1] of the form:

$$L = \mathbf{u} + \mathbb{K}\mathbf{v} = \{\mathbf{u} + \lambda\mathbf{v} | \lambda \in \mathbb{K}\} \qquad \mathbf{u}, \mathbf{v} \in \mathcal{V}, \mathbf{v} \neq 0 \qquad (20.22)$$

Two Point Formula

Definition 20.16 Two Point Formula:



$$L = \mathbf{u} + \mathbb{K}\mathbf{v} \qquad (20.23)$$

4.4.3. Planes

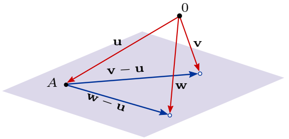
Definition 20.17 Planes: Planes are sets defined as:

$$E = \mathbf{u} + \mathbb{K}\mathbf{v} + \mathbb{K}\mathbf{w} = \{\mathbf{u} + \lambda\mathbf{v} + \mu\mathbf{w} | \lambda, \mu \in \mathbb{K}\} \qquad (20.24)$$

$$\mathbf{u}, ,\mathbf{w} \in \mathcal{V} \qquad \text{s.t. } \mathbf{v}, \mathbf{u} \neq \mathbf{0} \qquad \text{and} \qquad \mathbf{v}, \mathbf{w} \text{ lin. indep.}$$

Parameterform

Definition 20.18 Two Point Formula:

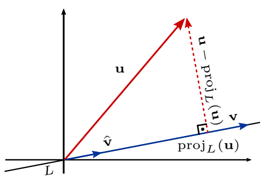


$$\begin{aligned} E &= \mathbf{u} + \mathbb{K}(\mathbf{v} - \mathbf{u}) \\ &\quad + \mathbb{K}(\mathbf{w} - \mathbf{u}) \end{aligned} \qquad (20.25)$$

4.4.4. Minimal Distance of Vector Subspaces

Projections in 2D

Definition 20.19 2D Vector Projection [Proof 20.17,20.18]:



$$\begin{aligned} \mathbf{u}_\mathbf{v} &= \text{proj}_L(\mathbf{u}) \\ &= u_v\tilde{\mathbf{v}} = \left(\mathbf{u}^\intercal\tilde{\mathbf{v}}\right)\tilde{\mathbf{v}} \\ &= \frac{\mathbf{u}^\intercal\mathbf{v}}{\|\mathbf{v}\|^2}\mathbf{v} = \frac{\mathbf{u}^\intercal\mathbf{v}}{\mathbf{v}^\intercal\mathbf{v}}\mathbf{v} \end{aligned} \qquad (20.26)$$

Corollary 20.8 [proof 20.8]
2D Projection Matrix **P:** Is the matrix that satisfies:

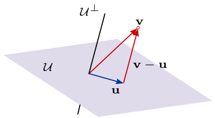
$$\mathbf{Pu} = \text{proj}_L(\mathbf{u}) \qquad \mathbf{P} = \frac{\mathbf{vv}^\intercal}{\mathbf{v}^T\mathbf{v}} = \frac{\mathbf{vv}^\intercal}{\|\mathbf{v}\|^2} \qquad (20.27)$$

Proof 20.1: [Corollary 20.8]

$$\frac{1}{\mathbf{v}^\intercal\mathbf{v}}\mathbf{u}^\intercal\mathbf{vv} = \frac{1}{\mathbf{v}^\intercal\mathbf{v}}\mathbf{v}\left(\mathbf{v}^\intercal\mathbf{u}\right) = \frac{1}{\mathbf{v}^\intercal\mathbf{v}}\left(\mathbf{vv}^\intercal\right)\mathbf{u}$$

General Projections

Definition 20.20 [proof 20.19]
General Vector Projection:
Is the orthogonal projection $\mathbf{u}$ of a vector $\mathbf{v}$ onto a sub-vector space $\mathcal{U}$



$$\mathbf{u} = \sum_{i=1}^{n} \alpha_i\mathbf{b}_i \qquad (20.28)$$

$$\mathbf{AA}^\intercal\alpha_i = \mathbf{A}^\intercal\mathbf{v} \qquad \mathbf{A} = \begin{pmatrix} \mathbf{b}_1 \\ \vdots \\ \vdots \\ \mathbf{b}_n \end{pmatrix}$$

where $\mathfrak{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ is a basis of the vector subspace $\mathcal{U}$.

Theorem 20.1 Projection Theorem: Let $\mathcal{U}$ a sub vector space of a finite euclidean vector space $\mathcal{V}$. Then there exists for every vector $\mathbf{v} \in \mathcal{V}$ a vector $\mathbf{u} \in \mathcal{U}$ obtained by an *orthogonal*[def. 20.65] projection

$$p : \begin{cases} \mathcal{V} \to \mathcal{U} \\ \mathbf{v} \mapsto \mathbf{u} \end{cases} \qquad (20.29)$$

the vector $u' := \mathbf{v} - \mathbf{u}$ representing the distance between $\mathbf{u}$ and $\mathbf{v}$ and is minimal:

$$\|\mathbf{u}'\| = \|\mathbf{v} - \mathbf{u}\| \leqslant \|\mathbf{v} - \mathbf{w}\| \qquad \forall\mathbf{w} \in \mathcal{U} \qquad \mathbf{u}' \in \mathcal{U}^\perp \qquad (20.30)$$

4.5. Affine Subspaces

4.6. Planes

https://math.stackexchange.com/questions/1485509/show-that-two-planes-are-parallel-and-find-the-distance-between-them

5. Matrices

Special Kind of Matrices

5.1. Symmetric Matrices

Definition 20.21 Symmetric Matrices: A matrix $\mathbf{A} \in \mathbb{K}^{n\times n}$ is called *symmetric* if it satisfies:

$$\mathbf{A} = \mathbf{A}^\mathsf{T} \tag{20.31}$$

Property 20.2 [proof ??]
Eigenvalues of real symmetric Matrices: The eigenvalues of a real symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ are real:

$$\text{spectrum}(\mathbf{A}) \in \{\mathbb{R}_{\geqslant 0}\}_{i=1}^{n} \tag{20.32}$$

Property 20.3 [proof ??]
Orthogonal Eigenvector basis: Eigenvectors of real symmetric matrices with distinct eigenvalues are orthogonal.

Corollary 20.9
Eigendecomposition Symmetric Matrices: If $\mathbf{A} \in \mathbb{R}^{n,n}$ is a real *symmetric*[def. 20.21] matrix then its eigenvectors are *orthogonal* and its eigen-decomposition[def. 20.84] is given by:

$$\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^\mathsf{T} \tag{20.33}$$

5.2. Orthogonal Matrices

Definition 20.22 Orthogonal Matrix: A real valued square matrix $\mathbf{Q} \in \mathbb{R}^{n\times n}$ is said to be orthogonal if its row vectors (and respectively its column vectors) build an orthonormal[def. 20.66] basis:

$$\langle \mathbf{q}_{:i}, \mathbf{q}_{:j}\rangle = \delta_{ij} \qquad \text{and} \qquad \langle \mathbf{q}_{i:}, \mathbf{q}_{j:}\rangle = \delta_{ij} \tag{20.34}$$

This is exactly true if the inverse of $\mathbf{Q}$ equals its transpose:

$$\mathbf{Q}^{-1} = \mathbf{Q}^\mathsf{T} \qquad \Longleftrightarrow \qquad \mathbf{Q}\mathbf{Q}^\mathsf{T} = \mathbf{Q}^\mathsf{T}\mathbf{Q} = \mathbf{I}_n \tag{20.35}$$

Attention: *Orthogonal* matrices are sometimes also called *orthonormal matrices.*

5.3. Hermitian Matrices

Definition 20.23 Conjugate Transpose $\mathbf{A}^\mathsf{H}/\mathbf{A}^*$
Hermitian Conjugate/Adjoint Matrix:
The conjugate transpose of a matrix $\mathbf{A} \in \mathbb{C}^{m\times n}$ is defined as:

$$\mathbf{A}^\mathsf{H} := \left(\overline{\mathbf{A}^\mathsf{T}}\right) = \overline{\mathbf{A}}^\mathsf{T} \qquad \Longleftrightarrow \qquad \mathbf{a}^\mathsf{H}_{i,j} = \bar{\mathbf{a}}_{j,i} \qquad \begin{array}{c} 1 \leqslant i \leqslant n \\ 1 \leqslant j \leqslant m \end{array} \tag{20.36}$$

Definition 20.24
Hermitian/Self-Adjoint Matrices $\mathbf{A} = \mathbf{A}^\mathsf{H}$**:**
A hermitian matrix is complex square matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ who is equal to its own *conjugate transpose*[def. 20.23]:

$$\mathbf{A} = \mathbf{A}^\mathsf{H} = \overline{\mathbf{A}}^\mathsf{T} \qquad \Longleftrightarrow \qquad \mathbf{a}_{i,j} = \bar{\mathbf{a}}_{j,i} \qquad i \in \{1,\dots,n\} \tag{20.37}$$

Corollary 20.10 : [def. 20.23] implies that $\mathbf{A}$ must be a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$.

Corollary 20.11 Real Hermitian Matrices: From [cor. 11.1] it follows:

$$\mathbf{A} \in \mathbb{R}^{n\times n} \ \textit{hermitian} \qquad \Longrightarrow \qquad \mathbf{A} \ \textit{real symmetric}^{[def.\ 20.21]} \tag{20.38}$$

Property 20.4 [proof 20.15]
Eigenvalues of Hermitan Matrices: The eigenvalues of a hermitian matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ are real:

$$\text{spectrum}(\mathbf{A}) \in \{\mathbb{R}_{\geqslant 0}\}_{i=1}^{n} \tag{20.39}$$

Property 20.5 [proof 20.16]
Orthogonal Eigenvector basis: Eigenvectors of hermitian matrices with distinct eigenvalues are orthogonal.

Corollary 20.12
Eigendecomposition Symmetric Matrices: If $\mathbf{A} \in \mathbb{C}^{n,n}$ is a hermitian matrix[def. 20.24] then its eigendecomposition[def. 20.84] is given by:

$$\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^\mathsf{H} \tag{20.40}$$

5.4. Unitary Matrices

Definition 20.25 Unitary Matrix $\mathbf{U}\mathbf{U}^\mathsf{H}$**:**
is a complex square matrix $\mathbf{U} \in \mathbb{C}^{n\times n}$ whose inverse[def. 20.39] is equal to its *conjugate transpose*[def. 20.23]:

$$\mathbf{U}^\mathsf{H}\mathbf{U} = \mathbf{U}\mathbf{U}^\mathsf{H} = \mathbf{I} \tag{20.41}$$

Corollary 20.13 Real Unitary Matrix: A real matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ that is unitary is an *orthogonal matrix*[def. 20.22].

Property 20.6
Preservation of Euclidean Norm [proof 20.14]:
Orthogonal and unitary matrices $\mathbf{Q} \in \mathbb{K}^{n,n}$ do not affect the 2-norm:

$$\|\mathbf{Q}\mathbf{x}\|_2 = \|\mathbf{x}\|_2 \qquad \forall \mathbf{x} \in \mathbb{K}^n \tag{20.42}$$

5.5. Similar Matrices

Definition 20.26 Similar Matrices: Two square matrices $\mathbf{A} \in \mathbb{K}^{n\times n}$ and $\mathbf{B} \in \mathbb{K}^{n\times n}$ are called *similar* if there exists a invertible matrix $\mathbf{S} \in \mathbb{K}^{n\times n}$ s.t.:

$$\exists \mathbf{S}: \qquad \mathbf{B} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S} \tag{20.43}$$

Corollary 20.14
Similarity Transformation/Conjugation:
The mapping:

$$\mathbf{A} \mapsto \mathbf{S}^{-1}\mathbf{A}\mathbf{S} \tag{20.44}$$

is called *similarity transformation*

Corollary 20.15
Eigenvalues of Similar Matrices [proof 20.13]:

If $\mathbf{A} \in \mathbb{K}^{n\times n}$ has the eigenvalue-eigenvector pairs $\{\{\lambda_i, \mathbf{v}_i\}\}_{i=1}^n$ then its *conjugate*eq. (20.44) $\mathbf{B}$ has the same eigenvalues with transformed eigenvectors:

$$\{\{\lambda_i, \mathbf{u}_i\}\}_{i=1}^n \qquad \mathbf{u}_i := \mathbf{S}^{-1}\mathbf{v}_i \tag{20.45}$$

5.6. Skew Symmetric Matrices

Definition 20.27
Skey Symmetric/Antisymmetric Matrices:

$$\mathbf{A}^\mathsf{T} = -\mathbf{A} \tag{20.46}$$

5.7. Triangular Matrix

Definition 20.28 Triangular Matrix: An upper (lower) triangular matrix, is a matrix whose element's below (above) the main diagonal are all zero:

$$\begin{pmatrix} l_{11} & & \\ l_{21} & l_{22} & \\ \vdots & \vdots & \ddots \\ l_{n1} & l_{n2} \cdots\cdots & l_{nn} \end{pmatrix} \qquad \begin{pmatrix} u_{11} & u_{12} \cdots\cdots & u_{1n} \\ & u_{22} \cdots\cdots & u_{2n} \\ & \ddots & \vdots \\ & & u_{nn} \end{pmatrix}$$

Figure 10: Lower Tri. Mat. Figure 11: Upper Tri. Mat.

5.7.1. Unitriangular Matrix

Definition 20.29 Unitriangular Matrix: An upper (lower) unitriangular matrix, is a upper (lower) triangular matrix[def. 20.28] whose diagonal elements are all ones.

5.7.2. Strictly Triangular Matrix

Definition 20.30 Strictly Triangular Matrix: An upper (lower) strictly triangular matrix, is a upper (lower) triangular matrix[def. 20.28] whose diagonal elements are all zero.

5.8. Block Partitioned Matrices

Definition 20.31 Block Partitioned Matrix:
A matrix $\mathbf{M} \in \mathbb{R}^{k+l,k+l}$ can be partitioned into a *block partitioned matrix*:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \qquad \mathbf{A} \in \mathbb{R}^{k,k}, \mathbf{B} \in \mathbb{R}^{k,l}, \mathbf{C} \in \mathbb{R}^{l,k}, \mathbf{D} \in \mathbb{R}^{l,l} \tag{20.47}$$

Definition 20.32 Block Partitioned Linear System:
A linear system $\mathbf{M}\mathbf{x} = \mathbf{b}$ with $\mathbf{M} \in \mathbb{R}^{k+l,k+l}$ and $\mathbf{x}, \mathbf{b} \in \mathbb{R}^{k+l}$ can be partitioned into a *block partitioned system*:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix} \qquad \begin{array}{r} \mathbf{A} \in \mathbb{R}^{k,k}, \mathbf{B} \in \mathbb{R}^{k,l}, \mathbf{C} \in \mathbb{R}^{l,k}, \mathbf{D} \in \mathbb{R}^{l,l} \\ \mathbf{x}_1, \mathbf{b}_1 \in \mathbb{R}^{k}, \mathbf{x}_2, \mathbf{b}_2 \in \mathbb{R}^{l} \end{array} \tag{20.48}$$

5.8.1. Schur Complement

Definition 20.33 Schur Complement: Given a block partitioned matrix[def. 20.31] $\mathbf{M} \in \mathbb{R}^{k+l,k+l}$ its Schur complements are given by:

$$\mathbf{S}_A = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B} \qquad \mathbf{S}_D = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} \tag{20.49}$$

5.8.2. Inverse of Block Partitioned Matrix

Definition 20.34 proof 20.3
Inverse of a Block Partitioned Matrix:
Given a block partitioned matrix[def. 20.31] $\mathbf{M} \in \mathbb{R}^{k+l,k+l}$ its inverse $\mathbf{M}^{-1}$ can be partitioned as well:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \qquad \mathbf{M}^{-1} = \begin{bmatrix} \widetilde{\mathbf{A}} & \tilde{\mathbf{B}} \\ \tilde{\mathbf{C}} & \widetilde{\mathbf{D}} \end{bmatrix} \tag{20.50}$$

$$\widetilde{\mathbf{A}} = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{S}_\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1} \qquad \tilde{\mathbf{C}} = -\mathbf{S}_\mathbf{A}^{-1}\mathbf{C}\mathbf{A}^{-1}$$

$$\tilde{\mathbf{B}} = -\mathbf{A}^{-1}\mathbf{B}\mathbf{S}_\mathbf{A}^{-1} \qquad \widetilde{\mathbf{D}} = \mathbf{S}_\mathbf{A}^{-1}$$

where $\mathbf{S}_\mathbf{A} = \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ is the Schur complement of $\mathbf{A}$.

5.9. Properties of Matrices

5.9.1. Square Root of p.s.d. Matrices

Definition 20.35 Square Root:

5.9.2. Trace

Definition 20.36 Trace: The trace of an $\mathbf{A} \in \mathbb{R}^{n\times n}$ matrix is defined as:

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \cdots + a_{nn} \tag{20.51}$$

Property 20.7 Trace of a Scalar:

$$\operatorname{tr}(\mathbb{R}) = \mathbb{R} \tag{20.52}$$

Property 20.8 Trace of Transpose:

$$\operatorname{tr}\left(\mathbf{A}^\mathsf{T}\right) = \operatorname{tr}(\mathbf{A}) \tag{20.53}$$

Property 20.9 Trace of multiple Matrices:

$$\operatorname{tr}(\mathbf{ABC}) = \operatorname{tr}(\mathbf{BCA}) = \operatorname{tr}(\mathbf{CBA}) \tag{20.54}$$

6. Matrices and Determinants

6.1. Determinants

6.1.1. Laplace/Cofactor Expansion

Definition 20.37 Minor:

Definition 20.38 Cofactors:

Properties

Property 20.10 Determinant times Scalar $\det(\alpha\mathbf{A})$**:**
Given a matirx $\mathbf{A} \in \mathbb{R}^{n\times n}$ it holds:

$$\det(\alpha \cdot \mathbf{A}) = \alpha^n\mathbf{A} \tag{20.55}$$

6.2. Inverese of Matrices

Definition 20.39 Inverse Matrix $\mathbf{A}^{-1}$**:**

6.2.1. Invertability

Definition 20.40
Singular/Non-Invertible Matrix $\det(\mathbf{A}) = 0$**:**
A square matrix $\mathbf{A} \in \mathbb{K}^{n\times n}$ is singular or non-invertible if it satisfies the following and equal conditions:

- $\det(\mathbf{A}) = 0$
- $\dim(\mathbf{A}) < n$
- $\nexists \mathbf{B} : \mathbf{B} = \mathbf{A}^{-1}$
- $\mathbf{A}\mathbf{x} = \mathbf{b}$ has either
  - no solution $\mathbf{x}$
  - infinitely many solutions $\mathbf{x}$

Transformations And Mapping

7. Linear & Affine Mappings/Transformations

7.1. Linear Mapping

Definition 20.41
Linear Mapping: A linear mapping, function or transformation is a map $l : V \mapsto W$ between two $\mathbb{K}$-vector spaces[def. 20.4] V and W if it satisfies:

$$l(\mathbf{x}+\mathbf{y}) = l(\mathbf{x}) + l(\mathbf{y}) \qquad \text{(Additivity)} \tag{20.56}$$

$$l(\alpha\mathbf{x}) = \alpha l(\mathbf{x}) \quad \forall\alpha \in \mathbb{K} \qquad \text{(Homogenitivity)} \tag{20.57}$$

$$\forall \mathbf{x}, \mathbf{y} \in V$$

Proposition 20.1 [proof 20.8]
Equivalent Formulations: Definition 20.41 is equivalent to:

$$l(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha l(\mathbf{x}) + \beta l(\mathbf{y}) \qquad \begin{matrix}\forall \alpha, \beta \in \mathbb{K}\\ \forall \mathbf{x}, \mathbf{y} \in V\end{matrix} \tag{20.58}$$

Corollary 20.16 Superposition Principle:
Definition 20.41 is also known as the superposition principle: *"the net response caused by two or more signals is the sum of the responses that would have been caused by each signal individually."*

Corollary 20.17 [proof 20.10]
A linear mapping $\iff$ Ax:
For every matrix $\mathbf{A} \in \mathbb{K}^{m\times n}$ the map:

$$l_{\mathbf{A}} : \begin{cases} \mathbb{K}^n & \rightarrow & \mathbb{K}^m \\ \mathbf{x} & \mapsto & \mathbf{Ax} \end{cases} \tag{20.59}$$

is a *linear map* and every linear map l can be represented by a matrix vector product:

$$l \text{ is linear} \iff \exists \mathbf{A} \in \mathbb{K}^{n\times m} : f(x) = \mathbf{Ax} \quad \forall \mathbf{x} \in \mathbb{K}^m \tag{20.60}$$

Principle 20.1 [proof 20.9]
Principle of linear continuation: A linear mapping $l : \mathcal{V} \mapsto \mathcal{W}$ is determined by the image of the basis $\mathfrak{B}$ of $\mathcal{V}$:

$$l(\mathbf{v}) = \sum_{i=1}^{n} \beta_i l(b_i) \qquad \mathfrak{B}(\mathcal{V}) = \{b_1, \dots, b_n\} \tag{20.61}$$

Property 20.11 [proof 20.11]
Compositions of linear mappings are linear $f \circ g$: Let g, f be linear functions mapping from $\mathcal{V}$ to $\mathcal{W}$ (i.e. matching) then it holds that $f \circ g$ is a linear[def. 20.41].

Definition 20.42 Level Sets:

7.2. Affine Mapping

Definition 20.43 Affine Transformation/Map:
Let $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$ then:

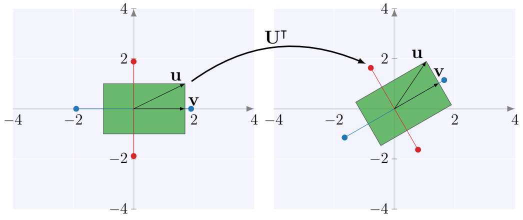
$$\mathbf{Y} = \mathbf{Ax} + \mathbf{b} \tag{20.62}$$

is called an affine transformation of $\mathbf{x}$.

7.3. Orthogonal Transformations

Definition 20.44 Orthogonal Transformation:
A linear transformation $T : \mathcal{V} \mapsto \mathcal{V}$ of an inner product space[def. 20.76] is an orthogonal transformation if preserves the inner product:

$$T(\mathbf{u}) \cdot T(\mathbf{v})\mathbf{u} \cdot \mathbf{v} \qquad \forall \mathbf{u}, \mathbf{v} \in \mathcal{V} \tag{20.63}$$



Corollary 20.18 Orthogonal Matrix Transformation:
An orthogonal matrix[def. 20.22] $\mathbf{Q}$ provides an orthogonal transformation:

$$(\mathbf{Qu})^\intercal(\mathbf{Qv}) = \mathbf{uv} \tag{20.64}$$

Explanation 20.2 (Improper Rotations).
Orthogonal transformations in two or three dimensional euclidean space[def. 20.44] represent improper rotations:

- *Stiff Rotations*
- *Reflections*
- *Reflections+Rotations*

Corollary 20.19 Preservation of Orthogonality: Orthogonal transformation preserver orthogonality.

Corollary 20.20 [proof 20.6]
Preservation of Norm:
An orthogonal transformation $\mathbf{Q} : \mathcal{V} \mapsto \mathcal{V}$ preservers the *length/norm*:

$$\|\mathbf{u}\|_{\mathcal{V}} = \|\mathbf{Qu}\|_{\mathcal{V}} \tag{20.65}$$

Corollary 20.21 Preservation of Angle:
An orthogonal transformation T preservers the *angle*[def. 20.64] of its vectors:

$$\angle(\mathbf{u}, \mathbf{v}) = \angle(T(\mathbf{u}), T(\mathbf{v})) \tag{20.66}$$

7.4. Kernel & Image

7.4.1. Kernel

Definition 20.45 Kernel/Null Space $\mathtt{N}/\varphi^{-1}(\{0\})$**:**
Let φ be a linear mapping[def. 20.41] between two a $\mathbb{K}$-vector spaces $\varphi : \mathcal{V} \mapsto \mathcal{W}$.
The *kernel* of φ is defined as:

$$\mathtt{N}(\varphi) := \varphi^{-1}(\{\mathbf{0}\}) = \{\mathbf{v} \in \mathcal{V} \mid \varphi(\mathbf{v}) = \mathbf{0}\} \subseteq \mathcal{V} \tag{20.67}$$

Definition 20.46 Right Null Space $\mathtt{N}(\mathbf{A})$**:**
If $\varphi = \mathbf{A} =\in \mathbb{K}^{m\times n}$ then the eq. (20.67) is equal to:

$$\mathtt{N}(\mathbf{A}) = \varphi_{\mathbf{A}}^{-1}(\{0\}) = \left\{\mathbf{v} \in \mathbb{K}^n \mid \mathbf{Av} = \mathbf{0}\right\} \in \mathbb{K}^m \tag{20.68}$$

Definition 20.47 Left Null Space $\mathtt{N}(\mathbf{A}^\intercal)$**:**
If $\varphi = \mathbf{A} =\in \mathbb{K}^{m\times n}$ then the *left* null space is defined as:

$$\mathtt{N}(\mathbf{A}^\intercal) = \varphi_{\mathbf{A}^\intercal}^{-1}(\{0\}) = \left\{\mathbf{v} \in \mathbb{K}^m \mid \mathbf{A}^\intercal\mathbf{v} = \mathbf{0}\right\} \in \mathbb{K}^n \tag{20.69}$$

Note

The term *left* null space stems from the fact that:

$$(\mathbf{A}^\intercal\mathbf{x})^\intercal = \mathbf{0} \qquad \text{is equal to} \qquad \mathbf{x}^\intercal\mathbf{A} = \mathbf{0}$$

7.4.2. Image

Definition 20.48 Image/Range $\mathfrak{R}/\varphi$**:**
Let φ be a linear mapping[def. 20.41] between two a $\mathbb{K}$-vector spaces $\varphi : \mathcal{V} \mapsto \mathcal{W}$.
The *imgae* of φ is defined as:

$$\mathfrak{R}(\varphi) := \varphi(\mathcal{V}) = \{\varphi(\mathbf{v}) \mid \mathbf{v} \in \mathcal{V}\} \subseteq \mathcal{W} \tag{20.70}$$

Definition 20.49 Column Space $\mathbf{Ax}$**:**
If $\varphi = \mathbf{A} = (\mathbf{c}_1 \cdots\cdots \mathbf{c}_n) \in \mathbb{K}^{m\times n}$ then eq. (20.70) is equal to:

$$\begin{aligned}\mathfrak{R}(\mathbf{A}) = \varphi_{\mathbf{A}}(\mathbb{K}^n) &= \left\{\mathbf{Ax}\middle|\forall\mathbf{x} \in \mathbb{K}^n\right\} = \left\langle(\mathbf{c}_1 \cdots\cdots \mathbf{c}_n)\right\rangle \\ &= \left\{\mathbf{v}\middle| \sum_{i=1}^{n} \alpha_i\mathbf{c}_i, \forall\alpha_i \in \mathbb{K}\right\}\end{aligned} \tag{20.71}$$

Definition 20.50 Row Space $\mathbf{A}^\intercal\mathbf{x}$**:**
If $\varphi = \mathbf{A} = (\mathbf{r}_1^\intercal \cdots\cdots \mathbf{r}_m^\intercal) \in \mathbb{K}^{m\times n}$ then the column space is defined as:

$$\begin{aligned}\mathfrak{R}(\mathbf{A}^\intercal) = \varphi_{\mathbf{A}}(\mathbb{K}^m) &= \left\{\mathbf{A}^\intercal\mathbf{x}\middle|\forall\mathbf{x} \in \mathbb{K}^m\right\} = \left\langle(\mathbf{r}_1 \cdots\cdots \mathbf{r}_m)\right\rangle \\ &= \left\{\mathbf{v}\middle| \sum_{i=1}^{m} \alpha_i\mathbf{r}_i, \forall\alpha_i \in \mathbb{K}\right\}\end{aligned} \tag{20.72}$$

From orthogonality it follows $x \in \mathfrak{R}(\mathbf{A})$, $y \in \mathtt{N}(\mathbf{A}) \Rightarrow x^\top y = 0$.

Corollary 20.22 Orthogonality [proof 20.12]**:**
The *right* (*left*) null space[def. 20.45] is *orthogonal*[def. 20.65] to the *row*[def. 20.50] (*column*[def. 20.49]) space:

$$\mathtt{N}(\mathbf{A}) \perp \mathfrak{R}(\mathbf{A}^\intercal) \qquad \text{and} \qquad \mathtt{N}(\mathbf{A}^\intercal) \perp \mathfrak{R}(\mathbf{A}) \tag{20.73}$$

7.4.3. Rank Nullity Theorem

Theorem 20.2 Rank-Nullity theorem:
Let $\mathcal{V}$ be a finite vector space and let φ be a linear mapping $\varphi : \mathcal{V} \mapsto \mathcal{W}$ then it holds:

$$\dim(\mathcal{V}) = \dim \underbrace{\left(\varphi^{-1}(\{\mathbf{0}\})\right)}_{\text{Kernel}} + \dim \underbrace{(\varphi(\mathcal{V}))}_{\text{Image}} \tag{20.74}$$

Corollary 20.23 Representation as Standardbases:
For every linear mapping $\varphi : \mathbb{K}^n \mapsto \mathbb{K}^m$ there exists a matrix $\mathbf{A}$ that represents this mapping:

$$\varphi = \varphi_{\mathbf{A}} = (\varphi(\mathbf{e}_1) \cdots\cdots \varphi(\mathbf{e}_n)) \in \mathbb{K}^{m\times n} \tag{20.75}$$

where

8. Eigenvalues and Vectors

Definition 20.51 Eigenvalues: Given a square matrix $\mathbf{A} \in \mathbb{K}^{n,n}$ the eigenvalues

Definition 20.52 Spectrum: The spectrum of a square matrix $\mathbf{A} \in \mathbb{K}^{n\times n}$ is the set of its eigenvalues[def. 20.51]:

$$\text{spectrum}(\mathbf{A}) = \lambda(\mathbf{A}) = \{\lambda_1, \dots, \lambda_n\} \tag{20.76}$$

Formula 20.1 Eigenvalues of a 2x2 matrix: Given a 2x2-matrix $\mathbf{A}$ its eigenvalues can be calculated by:

$$\{\lambda_1, \lambda_2\} \in \frac{\mathtt{tr}(\mathbf{A}) \pm \sqrt{\mathtt{tr}(\mathbf{A})^2 - 4\det(\mathbf{A})}}{2} \tag{20.77}$$

$$\text{with} \qquad \mathtt{tr}(\mathbf{A}) = a + d \qquad \det(\mathbf{A}) = ad - bc$$

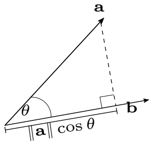
9. Vector Algebra

9.1. Dot/Standard Scalar Product

Definition 20.53 Scalar Projection a_b**:**

The scalar projection of a vector **a** onto a vector **b** is the *scalar* magnitude of the shadow/projection of the vector **a** onto **b**:

$$a_b = \|\mathbf{a}\| \cos \theta_{a,b} = \mathbf{a}\tilde{\mathbf{b}} \qquad (20.78)$$



Definition 20.54 [proof 20.4]
Standard Scalar/Dot Product:
Given two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ the standard scalar product is defined as:

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\mathsf{T}\mathbf{v} = \langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} u_i v_i = u_1 v_1 + \cdots + u_n v_n$$
$$= \|a\|\|b\| \cos \theta = u_v \hat{\mathbf{v}} = v_u \hat{\mathbf{u}} \quad \theta \in [0, \pi] \qquad (20.79)$$

Explanation 20.3 (Geometric Interpretation).
It is the magnitude of one vector times the magnitude of the shadow/scalar projection of the other vector.
Thus the dot product tells you:
1. *How much are two vectors pointing into the same direction*
2. *With what magnitude*

Property 20.12 Orthogonal Direction $\perp$**:**
For $\theta \in [-\pi, \pi/2]$ rad $\cos\theta = 0$ and it follows:
$$\mathbf{u} \cdot \mathbf{v} = 0 \qquad \iff \qquad \mathbf{u} \perp \mathbf{v} \qquad (20.80)$$

Note: Perpendicular

Perpendicular corresponds to orthogonality of two lines.

Property 20.13 Maximizing Direction:
For $\theta = 0$ rad $\cos\theta = 1$ and it follows:
$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\|\|\mathbf{v}\| \qquad (20.81)$$

Property 20.14 Minimizing Direction:
For $\theta = \pi$rad $\cos\theta = -1$ and it follows:
$$\mathbf{u} \cdot \mathbf{v} = -\|\mathbf{u}\|\|\mathbf{v}\| \qquad (20.82)$$

Definition 20.55 Vector Projecion:

9.2. Cross Product

9.3. Outer Product

Definition 20.56 Outer Product $\mathbf{u}\mathbf{v}^\mathsf{T} = \mathbf{u} \otimes \mathbf{v}$**:**
Given two vectors $\mathbf{u} \in \mathbb{K}^m$, $\mathbf{v} \in \mathbb{K}^n$ their outer product is defined as:

$$\mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^{\mathsf{H}} = [\mathbf{u}_1 \cdots\cdots \mathbf{u}_m] \begin{bmatrix} \bar{\mathbf{v}}_1 \\ \vdots \\ \bar{\mathbf{v}}_n \end{bmatrix} \qquad (20.83)$$
$$= \begin{bmatrix} \mathbf{u}_1 \odot \bar{\mathbf{v}}_1 \\ \vdots \\ \mathbf{u}_m \odot \bar{\mathbf{v}}_n \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1\bar{\mathbf{v}}_1 & \mathbf{u}_1\bar{\mathbf{v}}_2 \cdots\cdots \mathbf{u}_1\bar{\mathbf{v}}_n \\ \mathbf{u}_2\bar{\mathbf{v}}_1 & \mathbf{u}_2\bar{\mathbf{v}}_2 \cdots\cdots \mathbf{u}_2\bar{\mathbf{v}}_n \\ \vdots & \vdots \quad \ddots \quad \vdots \\ \mathbf{u}_m\bar{\mathbf{v}}_1 & \mathbf{u}_m\bar{\mathbf{v}}_2 \cdots\cdots \mathbf{u}_m\bar{\mathbf{v}}_n \end{bmatrix}$$

Proposition 20.2 [proof 20.5]
Rank of Outer Product: The outer product of two vectors is of rank one:
$$\mathfrak{rank}(\mathbf{u} \otimes \mathbf{v}) = 1 \qquad (20.84)$$

9.4. Vector Norms

Definition 20.57 Norm $\|\cdot\|_{\mathcal{V}}$**:**
Let $\mathcal{V}$ be a vector space over a field F, a norm on $\mathcal{V}$ is a map:
$$\|\cdot\|_{\mathcal{V}} : \mathcal{V} \mapsto \mathbb{R}_+ \qquad (20.85)$$
that satisfies:
$$\|\mathbf{x}\|_{\mathcal{V}} = 0 \iff \mathbf{x} = 0 \qquad \text{(Definitness)} \qquad (20.86)$$
$$\|\alpha\mathbf{x}\|_{\mathcal{V}} = |\alpha|\|\mathbf{x}\|_{\mathcal{V}} \qquad \text{(Homogenity)} \qquad (20.87)$$
$$\|\mathbf{x} + \mathbf{y}\|_{\mathcal{V}} \leqslant \|\mathbf{x}\|_{\mathcal{V}} + \|\mathbf{y}\|_{\mathcal{V}} \qquad \text{(Triangular Inequality)} \qquad (20.88)$$
$$\alpha \in \mathbb{K} \qquad \forall \mathbf{x}, \mathbf{y} \in \mathcal{V}$$

Explanation 20.4 (Definition 20.57).
A norm is a measures of the size of its argument.

Corollary 20.24 Normed vector space: Is a vector space $\mathcal{V}$ over a field F, on which a norm $\|\cdot\|_{\mathcal{V}}$ can be defined.

9.4.1. Cauchy Schwartz

Definition 20.58 [proof 20.21]
Cauchy Schwartz Inequality:
$$|\mathbf{u}^\mathsf{T}\mathbf{v}| \leqslant \|\mathbf{u}\|\|\mathbf{v}\| \qquad (20.89)$$

9.4.2. Triangular Inequality

Definition 20.59 [proof 20.22]
Triangular Inequality: States that the length of the sum of two vectors is lower or equal to the sum of their individual lengths:
$$\|\mathbf{u} + \mathbf{v}\| \leqslant \|\mathbf{u}\| + \|\mathbf{v}\| \qquad (20.90)$$

Corollary 20.25 Reverse Triangular Inequality:
$$-\|\mathbf{x} - \mathbf{y}\|_{\mathcal{V}} \leqslant \|\mathbf{x}\|_{\mathcal{V}} - \|\mathbf{y}\|_{\mathcal{V}} \leqslant \|\mathbf{x} - \mathbf{y}\|_{\mathcal{V}}$$
$$\text{resp.} \qquad \big|\|\mathbf{x}\|_{\mathcal{V}} - \|\mathbf{y}\|_{\mathcal{V}}\big| \leqslant \|\mathbf{x} - \mathbf{y}\|_{\mathcal{V}}$$

9.5. Distances

Definition 20.60
Distance Function/Measure $d : S \times S \mapsto \mathbb{R}_+$**:**
Let S be a set, a distance functions is a mapping d that satisfies:
$$d(x, x) = 0 \qquad \text{(Zero Identity Distance)} \qquad (20.91)$$
$$d(x, y) = d(y, x) \qquad \text{(Symmetry)} \qquad (20.92)$$
$$d(x, z) \leqslant d(x, y) + d(y, z) \qquad \text{(Triangular Identiy)} \qquad (20.93)$$
$$\forall x, y, z \in S$$

Explanation 20.5 (Definition 20.60).
Is measuring the distance between two things.

9.5.1. Contraction

Definition 20.61 Contraction: Given a metric space (M, d) is a mapping $f : M \mapsto M$ that satisfies:
$$d(f(x), f(y)) \leqslant \lambda d(x, y) \qquad \lambda \in [0, 1) \qquad (20.94)$$

9.6. Metrics

Definition 20.62 Metric $d : S \times S \mapsto \mathbb{R}_+$**:**
Is a distance measure[def. 20.60] that additionally satisfies the identity of indiscernibles:
$$d(x, y) = 0 \iff x = y \qquad \forall x, y \in S$$

Corollary 20.26 Metric→Norm: Every norm $\|\cdot\|_{\mathcal{V}}$ on a vector space $\mathcal{V}$ over a field F induces a metric by:
$$d(x, y) = \|x - y\|_{\mathcal{V}} \qquad \forall x, y \in \mathcal{V}$$
metric induced by norms additionally satisfy: $\forall x, y \in \mathcal{V}$, $\alpha \in F \subseteq \mathbb{K}$ $K = \mathbb{R}$ or $\mathbb{C}$
1. Homogenity/Scaling: $d(\alpha x, \alpha y)_{\mathcal{V}} = |\alpha| d(x, y)_{\mathcal{V}}$
2. Translational Invariance: $d(x + \alpha, y + \alpha) = d(x, y)$

Conversely not every metric induces a norm **but** if a metric d on a vector space $\mathcal{V}$ satisfies the properties then it induces a norm of the form:
$$\|\mathbf{x}\|_{\mathcal{V}} := d(\mathbf{x}, 0)_{\mathcal{V}}$$

Note

Similarity measure is a much weaker notion than a metric as triangular inequality does not have to hold.
Hence: If a is similar to b and b is similar to c it does not imply that a is similar to c.

Note

(bilinear form$\xrightarrow{\text{induces}}$)
inner product$\xrightarrow{\text{induces}}$norm$\xrightarrow{\text{induces}}$metric.

9.6.1. Metric Space

Definition 20.63 Metric Space (M, d)**:**
A *metric space* is a pair (M, d) of a set M and a metric[def. 20.62] d defined on M:
$$d : M \times M \mapsto \mathbb{R}_+ \qquad (20.95)$$

10. Angles

Definition 20.64 Angle between Vectors $\angle(\mathbf{u}, \mathbf{v})$**:** Let $\mathbf{u}, \mathbf{v} \in \mathbb{K}^n$ be two vectors of an inner product space[def. 20.76] $\mathcal{V}$. The angle $\alpha \in [0, \pi]$ between $\mathbf{u}, \mathbf{v}$ is defined by:
$$\angle(\mathbf{u}, \mathbf{v}) := \alpha \qquad \cos\alpha = \frac{\mathbf{u}^\mathsf{T}\mathbf{v}}{\|\mathbf{u}\|\|\mathbf{v}\|} \qquad \begin{matrix} \mathbf{u}, \mathbf{v} \in \mathcal{V} \\ \alpha \in [0, \pi] \end{matrix} \qquad (20.96)$$

11. Orthogonality

Definition 20.65 Orthogonal Vectors: Let $\mathcal{V}$ be an inner-product space[def. 20.76]. A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \in \mathcal{V}$ is called *orthogonal* iff:
$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0 \qquad \forall i \neq \qquad (20.97)$$

11.1. Orthonormality

Definition 20.66 Orthonormal Vectors: Let $\mathcal{V}$ be an inner-product space[def. 20.76]. A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_n, \ldots\} \in \mathcal{V}$ is called *orthonormal* iff:
$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \qquad \forall i, j \qquad (20.98)$$

12. Special Kind of Vectors

12.1. Binary/Boolean Vectors

Definition 20.67
Binary/Boolean Vectors/Bit Maps $\mathbb{B}^n$**:** Are vectors that contain only zero or one values:
$$\mathbb{B}^n = \{0, 1\}^n \qquad (20.99)$$

Definition 20.68
R-Sparse Boolean Vectors $\mathbb{B}^n_r$**:**
Are boolean vectors that contain exact r one values:
$$\mathbb{B}^n_r = \left\{ \mathbf{x} \in \{0, 1\}^n : \mathbf{x}^\mathsf{T}\mathbf{x} = \sum_{i=1}^{n} \mathbf{x} = r \right\} \qquad (20.100)$$

12.2. Probablistic Vectors

Definition 20.69 Probabilistic Vectors: Are vectors that represent probabilities and satisfy:
$$\left\{ \mathbf{x} \in [0, 1]^n : \sum_{i=1}^{n} x_i = 1 \right\} \qquad (20.101)$$

13. Vector Spaces and Measures

13.1. Bilinear Forms

13.2. Quadratic Forms

13.2.1. Min/Max Value

Corollary 20.27 [proof 20.20]
Extreme Value: The minimum/maximum of a quadratic form?? with a quadratic matrix $\mathbf{A} \in \mathbb{R}^{n,n}$ is given by the eigenvector corresponding to the smallest/largest eigenvector of **A**:
$$\mathbf{v}_1 \in \arg\min_{\mathbf{x}^\mathsf{T}\mathbf{x}=1} \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} \qquad \mathbf{v}_1 \in \arg\max_{\mathbf{x}^\mathsf{T}\mathbf{x}=1} \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} \qquad (20.102)$$

Note

$$(\mathbf{Q}^\mathsf{T}\tilde{\mathbf{n}})^\mathsf{T}\mathbf{Q}^\mathsf{T}\tilde{\mathbf{n}} = \tilde{\mathbf{n}}^\mathsf{T}\mathbf{Q}\mathbf{Q}^\mathsf{T}\tilde{\mathbf{n}} = \tilde{\mathbf{n}}^\mathsf{T}\tilde{\mathbf{n}} = 1$$

13.2.2. Skew Symmetric Matirx

Corollary 20.28
Quadratic Form of Skew Symmetric matrix: The quadratic form of a skew symmetric matrix[def. 20.27] vanishes:
$$\alpha = \mathbf{x}^\mathsf{T}\mathbf{A}_{\text{skew}}\mathbf{x} = \left(\mathbf{x}^\mathsf{T}\mathbf{A}^\mathsf{T}_{\text{skew}}\mathbf{x}\right)^\mathsf{T} = \left(\mathbf{x}^\mathsf{T}\mathbf{A}_{\text{skew}}\mathbf{x}\right)^\mathsf{T} = -\alpha \qquad (20.103)$$
Which can only hold iff $\alpha = 0$.

13.3. Inner Product – Generalization of the dot product

Definition 20.70 Bilinear Form/Functional:
Is a mapping $a : \mathcal{V} \times \mathcal{V} \mapsto F$ on a field of scalars $F \subseteq \mathbb{K}$, $K = \mathbb{R}$ or $\mathbb{C}$ that satisfies:
$$a(\alpha u + \beta v, w) = \alpha a(u, w) + \beta a(v, w)$$
$$a(u, \alpha v + \beta w) = \alpha a(u, v) + \beta a(u, w)$$
$$\forall u, v, w \in \mathcal{V}, \qquad \forall \alpha, \beta \in \mathbb{K}$$
Thus: a is linear w.r.t. each argument.

Definition 20.71 Symmetric bilinear form: A bilinear form a on $\mathcal{V}$ is symmetric if and only if:
$$a(u, v) = a(v, u) \qquad \forall u, v \in \mathcal{V}$$

Definition 20.72 Positive (semi) definite bilinear form:
A symmetric bilinear form a on a vector space $\mathcal{V}$ over a field F is positive defintie if and only if:
$$a(u, u) > 0 \qquad \forall u \in \mathcal{V} \backslash \{0\} \qquad (20.104)$$
$$\text{And positive semidefinte} \iff \geqslant \qquad (20.105)$$

Corollary 20.29 Matrix induced Bilinear Form:
For finite dimensional inner product spaces $\mathcal{X} \in \mathbb{K}^n$ any *symmetric* matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ induces a bilinear form:
$$a(\mathbf{x}, \mathbf{x}') = \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x}' = (\mathbf{A}\mathbf{x}')\mathbf{x},$$

Definition 20.73 Positive (semi) definite Matrix $\succ$**:**
A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is positive defintie if and only if:
$$\mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} > 0 \qquad \iff \qquad \mathbf{A} \succ 0 \quad \forall \mathbf{x} \in \mathbb{R}^n \backslash \{0\} \qquad (20.106)$$
$$\text{And } \textbf{positive semidefinte} \iff \geqslant \qquad (20.107)$$

Corollary 20.30 [proof 20.2]
Eigenvalues of positive (semi) definite matrix:
A positive definite matrix is a matrix where every eigenvalue is *strictly* positive and positive semi definite if every eigenvalue is *positive*.
$$\forall \lambda_i \in \text{eigenv}(\mathbf{A}) > 0 \qquad (20.108)$$
$$\text{And positive semidefinte} \iff \geqslant \qquad (20.109)$$

Note

Positive definite matrices are often assumed to be symmetric but that is not necessarily true.

Proof 20.2: **??** 20.2 (for real matrices):
Let **v** be an eigenvector of **A** then it follows:
$$0 \overset{??\ 20.2}{<} \mathbf{v}^\mathsf{T}\mathbf{A}\mathbf{v} = \mathbf{v}^\mathsf{T}\lambda\mathbf{v} = \|\mathbf{v}\|\lambda$$

Corollary 20.31 Positive Definiteness and Determinant: The determinant of a positive definite matrix is always positive. **Thus** a positive definite matrix is always *nonsingular*

Definition 20.74 Negative (semi) definite Matrix $\prec$:
A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is negative defintie if and only if:

$$\mathbf{x}^\intercal \mathbf{A}\mathbf{x} < 0 \iff \mathbf{A} \prec 0 \quad \forall \mathbf{x} \in \mathbb{R}^n \backslash \{0\} \qquad (20.110)$$

$$\text{And } \textbf{negative semidefinte} \iff \preccurlyeq \qquad (20.111)$$

Theorem 20.3 Sylvester's criterion: Let $\mathbf{A}$ be *symmetric/Hermitian* matrix and denote by $\mathbf{A}^{(k)}$ the $k \times k$ upper left sub-matrix of $\mathbf{A}$.
Then it holds that:

- $$\mathbf{A} \succ 0 \iff \det\left(\mathbf{A}^k\right) > 0 \qquad k = 1, \dots, n \qquad (20.112)$$
- $$\mathbf{A} \prec 0 \iff (-1)^k \det\left(\mathbf{A}^k\right) > 0 \qquad k = 1, \dots, n \qquad (20.113)$$
- $\mathbf{A}$ is indefinite if the first $\det\left(\mathbf{A}^k\right)$ that breaks both of the previous patterns is on the wrong side.
- Sylvester's criterion is inconclusive ($\mathbf{A}$ can be anything of the previous three) if the first $\det\left(\mathbf{A}^k\right)$ that breaks both patterns is 0.

14. Inner Products

Definition 20.75 Inner Product: Let $\mathcal{V}$ be a vector space over a field $F \in \mathbb{K}$ of scalars. An inner product on $\mathcal{V}$ is a map:

$$\langle\cdot,\cdot\rangle : \mathcal{V} \times \mathcal{V} \mapsto F \subseteq \mathbb{K} \qquad K = \mathbb{R} \text{ or } \mathbb{C} \qquad (20.114)$$

that satisfies: $\forall x, y, z \in \mathcal{V}, \quad \alpha, \beta \in F$

1. (Conjugate) Stmmetry: $\langle x, y\rangle = \overline{\langle x, y\rangle}$.
2. Linearity in the first argument:
$$\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle$$
3. Positve-definiteness:
$$\langle x, x\rangle \geqslant 0 : x = 0 \iff \langle x, x\rangle = 0$$

Definition 20.76 Inner Product Space $(\mathcal{V}, \langle\cdot,\cdot\rangle_{\mathcal{V}})$: Let $F \in \mathbb{K}$ be a field of scalars.
An inner product space $\mathcal{V}$ is a vetor space over a field F together with an an **inner product** $\langle\cdot,\cdot\rangle_{\mathcal{V}}$).

Corollary 20.32 Inner product↦S.p.d. Bilinear Form: Let $\mathcal{V}$ be a vector space over a field $F \in \mathbb{K}$ of scalar.
An **inner product** on $\mathcal{V}$ is a positive definite symmetric bilinear form on $\mathcal{V}$.

Example: scalar prodct

Let $a(u, v) = u^\intercal \mathbf{I} v$ then the standard scalar product can be defined in terms of a bilinear form vice versa the standard scalar product induces a bilinear form.

Note

Inner products must be positive definite by defintion $\langle \mathbf{x}, \mathbf{x}\rangle \geqslant 0$, whereas bilinear forms must not.

Corollary 20.33 Inner product induced norm $\langle\cdot,\cdot\rangle_{\mathcal{V}} \to \|\cdot\|_{\mathcal{V}}$: Every inner product $\langle\cdot,\cdot\rangle_{\mathcal{V}}$ induces a norm of the form:

$$\|\mathbf{x}\|_{\mathcal{V}} = \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle} \qquad \mathbf{x} \in \mathcal{V}$$

Thus We can define function spaces by their associated norm $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ and inner product spaces lead to normed vector spaces and vice versa.

Corollary 20.34 Energy Norm: A *s.p.d.* bilinear form $a : \mathcal{V} \times \mathcal{V} \mapsto F$ induces an energy norm:

$$\|\mathbf{x}\|_a := (a(\mathbf{x}, \mathbf{x}))^{\frac{1}{2}} = \sqrt{a(\mathbf{x}, \mathbf{x})} \qquad \mathbf{x} \in \mathcal{V}$$

15. Matrix Algebra

16. Matrix Norms

16.1. Operator Norm

Definition 20.77 Operator/Induced Norm:
Let $\|\cdot\|_\mu : \mathbb{K}^m \mapsto \mathbb{R}$ and $\|\cdot\|_\nu : \mathbb{K}^n \mapsto \mathbb{R}$ be vector norms.
The operator norm is defined as:

$$\|\mathbf{A}\|_{\mu,\nu} := \sup_{\substack{\mathbf{x}\in\mathbb{K}^n \\ \mathbf{x}\neq 0}} \frac{\|\mathbf{A}\mathbf{x}\|_\mu}{\|\mathbf{x}\|_\nu} = \sup_{\|\mathbf{x}\|_\nu = 1} \|\mathbf{A}\mathbf{x}\|_\mu \qquad \|\cdot\|_\mu : \mathbb{K}^m \mapsto \mathbb{R} \qquad (20.115)$$

Explanation 20.6 (Definition 20.77). *Is a measure for the largest factor by which a matrix* $\mathbf{A}$ *can stretch a vector* $\mathbf{x} \in \mathbb{R}^n$.

16.2. Induced Norms

Corollary 20.35 Induced Norms: Let $\|\cdot\|_p : \mathbb{K}^{m \times n} \mapsto \mathbb{R}$ defined as:

$$\|\mathbf{A}\|_p := \sup_{\substack{\mathbf{x}\in\mathbb{K}^n \\ \mathbf{x}\neq 0}} \frac{\|\mathbf{A}\mathbf{x}\|_p}{\|\mathbf{x}\|_p} = \sup_{\|\mathbf{y}\|_p = 1} \|\mathbf{A}\mathbf{y}\|_p \qquad (20.116)$$

Explanation 20.7 ([Corollary 20.35]).
Induced norms are matrix norms induced by vector norms as we:

- *Only work with vectors* $\mathbf{A}\mathbf{x}$
- *And use the normal p-vector norms* $\|\cdot\|_p$

Note supremum

The set of vectors $\{\mathbf{y} | \|\mathbf{y}\| = 1\}$ is compact, thus if we consider finite matrices the supremum is attained and we may replace it by the max.

16.3. Induced Norms

16.3.1. 1-Norm

Definition 20.78 Column Sum Norm $\|\mathbf{A}\|_1$:

$$\|\mathbf{A}\|_1 = \sup_{\substack{\mathbf{x}\in\mathbb{K}^n \\ \mathbf{x}\neq 0}} \frac{\|\mathbf{A}\mathbf{x}\|_1}{\|\mathbf{x}\|_1} = \max_{1\leqslant j\leqslant n} \sum_{i=1}^{m} |a_{ij}| \qquad (20.117)$$

16.3.2. ∞-Norm

Definition 20.79 Row Sum Norm $\|\mathbf{A}\|_\infty$:

$$\|\mathbf{A}\|_\infty = \sup_{\substack{\mathbf{x}\in\mathbb{K}^n \\ \mathbf{x}\neq 0}} \frac{\|\mathbf{A}\mathbf{x}\|_\infty}{\|\mathbf{x}\|_\infty} = \max_{1\leqslant i\leqslant m} \sum_{j=1}^{n} |a_{ij}| \qquad (20.118)$$

16.3.3. Spectral Norm L2-Norm

Spectral Radius & Singular Value

Definition 20.80 Spectral Radius $\rho(\mathbf{A})$:
The spectral radius is defined as the largest eigenvalue of a matrix:

$$\rho(\mathbf{A}) = \max\{\lambda | \lambda \in \operatorname{eigenval}(\mathbf{A})\} \qquad (20.119)$$

Definition 20.81 Singular Value σ_i:
Given a matrix $\mathbf{A} \in \mathbb{K}^{m \times n}$ its n real and positive singular values are defined as:

$$\sigma(\mathbf{A}) := \left\{\left\{\sqrt{\lambda_i}\right\}_{i=1}^{n} | \lambda_i \in \operatorname{eigenval}\left(\mathbf{A}^\intercal\mathbf{A}\right)\right\} \qquad (20.120)$$

Spectral Norm

Definition 20.82 L2/Spectral Norm $\|\mathbf{A}\|_2$:

$$\|\mathbf{A}\|_2 = \sup_{\substack{\mathbf{x}\in\mathbb{K}^n \\ \|\mathbf{x}\|_2 = 1}} \|\mathbf{A}\mathbf{x}\|_2 = \max_{\|\mathbf{x}\|_2 = 1} \sqrt{\mathbf{x}^\intercal\mathbf{A}^\intercal\mathbf{A}\mathbf{x}} \qquad (20.121)$$

$$= \max_{\|\mathbf{x}\|_2 = 1} \sqrt{\rho\left(\mathbf{A}^\intercal\mathbf{A}\right))} =: \sigma_{\max}(\mathbf{A}) \qquad (20.122)$$

16.4. Energy Norm

16.5. Forbenius Norm

Definition 20.83 Forbenius Norm $\|\mathbf{A}\|_F$:
The *Forbenius norm* $\|\cdot\|_F : \mathbb{K}^{m \times n} \mapsto \mathbb{R}$ is defined as:

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{i,j}^2|} = \sqrt{\operatorname{tr}\left(\mathbf{A}\mathbf{A}^\mathsf{H}\right)} \qquad (20.123)$$

16.6. Distance

17. Decompositions

17.1. Eigen/Spectral decomposition

Definition 20.84 $\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}$,[proof 20.25]
Eigendecomposition/ Spectral Decomposition :
Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be a *diagonalizable* square matrix and define by $\mathbf{X} = [\mathbf{x}_1 \cdots\cdots \mathbf{x}_n] \in \mathbb{R}^{n \times n}$ a non-singular matrix whose column vectors are the eigenvectors of $\mathbf{A}$ with associated eigenvalue matrix $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. Then $\mathbf{A}$ can be represented as:

$$\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1} \qquad (20.124)$$

Proposition 20.3 Diagonalization: If non of $\mathbf{A}$ eigenvalues are zero it can be diagonalized:

$$\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathbf{\Lambda} \qquad (20.125)$$

Proposition 20.4 Existence:

$$\exists \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1} \iff \mathbf{A} \text{ diagonalizable} \qquad (20.126)$$

17.2. QR-Decompositions

17.3. Singular Value Decomposition

Definition 20.85
Singular Value Decomposition (SVD) $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{H}$:
For any matrix $\mathbf{A} \in \mathbb{K}^{m,n}$ there exist unitary matrices[def. 20.25]

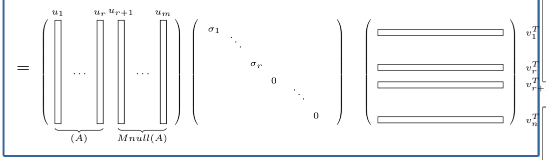
$$\mathbf{U} \in \mathbb{K}^{m,m} \qquad \mathbf{V} \in \mathbb{K}^{n,n}$$

and a (generalized) digonal matrix:

$$\mathbf{\Sigma} \in \mathbb{R}^{m,n} \qquad p := \min\{m, n\}$$
$$\mathbf{\Sigma} = \operatorname{gendiag}(\sigma_1, \dots, \sigma_p) \in \mathbb{R}^{m,n}$$

such that:

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\mathsf{H} \qquad (20.127)$$



17.3.1. Eigenvalues

Proposition 20.5 [proof 20.23]:
The eigenvalues of a matrix $\mathbf{A}^\intercal\mathbf{A}$ are positive.

Proposition 20.6 [proof 20.24]
Similarity Transformation: The unitary matrix $\mathbf{V}$ provides a *similarity transformation*[cor. 20.14] of $\mathbf{A}^\intercal\mathbf{A}$ into a diagonal matrix $\mathbf{\Sigma}^\intercal\mathbf{\Sigma}$:

$$\mathbf{\Sigma}^\intercal\mathbf{\Sigma} \mapsto \mathbf{V}^\mathsf{H}\mathbf{A}^\intercal\mathbf{A}\mathbf{V} \qquad (20.128)$$

Corollary 20.36 eigenval($\mathbf{A}^\intercal\mathbf{A}$) = eigenval($\Sigma^\intercal\Sigma$):
From proposition 20.6 and [cor. 20.15] it follows that:

$$\operatorname{eigenval}(\mathbf{A}^\intercal\mathbf{A}) = \operatorname{eigenval}(\Sigma^\intercal\Sigma) \qquad (20.129)$$

$$\implies \quad \|\mathbf{A}\|_2 = \sqrt{\rho\left(\mathbf{A}^\intercal\mathbf{A}\right)} = \sqrt{\lambda_{\max}} = \sigma_{\max}$$

Note

λ and *singularvalue* corresponds to the eigenvalues/singular-values of $\mathbf{A}^\intercal\mathbf{A}$ and not $\mathbf{A}$

17.3.2. Best Lower Rank Approximation

Theorem 20.4 Eckart Yound Theorem: Given a matrix $\mathbf{X} \in \mathbb{K}^{m,n}$ the *reduced* SVD $\mathbf{X}$ defined as:

$$\mathbf{U}_k := [\mathbf{u}_{:,1} \cdots\cdots \mathbf{u}_{:,k}] \in \mathbb{K}^{m,k}$$
$$\mathbf{X}_k := \mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^\mathsf{H} \qquad \mathbf{\Sigma}_k = \operatorname{diag}(\sigma_1, \dots, \sigma_k) \in \mathbb{R}^{k,k}$$
$$k \leqslant \min\{m, n\}$$
$$\mathbf{V}_k = [\mathbf{v}_{:,1} \cdots\cdots \mathbf{v}_{:,k}] \in \mathbb{K}^{n,k}$$

provides the best lower k rank approximation of $\mathbf{X}$:

$$\min_{\mathbf{Y}\in\mathbb{K}^{n,m} : \operatorname{rank}(\mathbf{Y}) \leqslant k} \|\mathbf{X} - \mathbf{Y}\|_F = \|\mathbf{X} - \mathbf{X}_k\|_F \qquad (20.130)$$

18. Matric Calculus

18.1. Derivatives

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{b}^\top\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^\top\mathbf{b}) = \mathbf{b}$$
$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^\top\mathbf{x}) = 2\mathbf{x}$$

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{A}\mathbf{x} = \mathbf{A} \qquad (20.131)$$

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^\intercal\mathbf{A}\mathbf{x} = \left(\mathbf{A} + \mathbf{A}^\intercal\right)\mathbf{x} \qquad (20.132)$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{b}^\top\mathbf{A}\mathbf{x}) = \mathbf{A}^\top\mathbf{b} \qquad \frac{\partial}{\partial \mathbf{X}}(\mathbf{c}^\top\mathbf{X}\mathbf{b}) = \mathbf{c}\mathbf{b}^\top \qquad \frac{\partial}{\partial \mathbf{x}}(\|\mathbf{x} - \mathbf{b}\|_2) = \frac{\mathbf{x}-\mathbf{b}}{\|\mathbf{x}-\mathbf{b}\|_2}$$

$$\frac{\partial}{\partial \mathbf{x}}(\|\mathbf{x}\|_2^2) = \frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^\top\mathbf{x}) = 2\mathbf{x} \qquad \frac{\partial}{\partial \mathbf{X}}(\|\mathbf{X}\|_F^2) = 2\mathbf{X}$$

$$\frac{\partial}{\partial \mathbf{x}}||\mathbf{x}||_1 = \frac{\mathbf{x}}{|\mathbf{x}|}$$

$$\frac{\partial}{\partial \mathbf{x}}(\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2) = \mathbf{2}(\mathbf{A}^\top\mathbf{A}\mathbf{x} - \mathbf{A}^\top\mathbf{b}) \qquad \frac{\partial}{\partial \mathbf{X}}(|\mathbf{X}|) = |\mathbf{X}| \cdot \mathbf{X}^{-1}$$

$$\frac{\partial}{\partial x}(\mathbf{Y}^{-1}) = -\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\mathbf{Y}^{-1}$$

19. Proofs

Proof 20.3: [def. 20.34]

$$\mathbf{M}\mathbf{M}^{-1} = \begin{bmatrix} \mathbf{I}_{k,k} & \mathbf{0}_{k,l} \\ \mathbf{0}_{l,k} & \mathbf{I}_{l,l} \end{bmatrix} \qquad (20.133)$$

19.1. Vector Algebra

Proof 20.4 Definition 20.54:

$$\textbf{(1)}: \underline{\|a - b\|} \overset{\text{eq. (21.19)}}{=} \|a\|^2 + \|b\|^2 - 2\|a\|\|b\|\cos\theta$$
$$\textbf{(2)}: \underline{\|a - b\|} = (a - b)(a - b) = \|a\|^2 + \|b\|^2 - 2(ab)$$
$$\underline{\|a - b\|} = \underline{\|a - b\|} \quad \Rightarrow \quad ab = \|a\|\|b\|\cos\theta$$

Proof 20.5 Proposition 20.2: The outer product of $\mathbf{u}$ with $\mathbf{v}$ corresponds to a scalar multiplication of $\mathbf{v}$ with elements u_i thus the rank must be that of $\mathbf{v}$, which is a vector and hence of rank 1

$$\mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^\mathsf{H} = \begin{bmatrix} \mathbf{u}_1 \odot \bar{\mathbf{v}}_1 \\ \vdots \\ \mathbf{u}_m \odot \bar{\mathbf{v}}_n \end{bmatrix}$$

19.2. Mappings

Proof 20.6: Corollary 20.20

$$\|\mathbf{Q}\mathbf{x}\|^2 = (\mathbf{Q}\mathbf{x})^\intercal\mathbf{Q}\mathbf{x} = \mathbf{x}^\intercal\mathbf{Q}^\intercal\mathbf{Q}\mathbf{x} = \mathbf{x}^\intercal\mathbf{x} = \|\mathbf{x}\|^2$$

Proof 20.7: Corollary 20.21 Follows immediately from definition 20.64 in combination with eqs. (20.63) and (20.65).

Proof 20.8: Proposition 20.1:

$$\implies \quad l(\alpha\mathbf{x} + \beta\mathbf{y}) \overset{20.56}{=} l(\alpha\mathbf{x}) + l(\beta\mathbf{y}) \overset{20.57}{=} \alpha l(\mathbf{x}) + \beta l(\mathbf{y})$$
$$l(\alpha\mathbf{x} + \mathbf{0}) = \alpha l(\mathbf{x})$$
$$\impliedby \quad l(1\mathbf{x} + 1\mathbf{y}) = l(\mathbf{x}) + l(\mathbf{y})$$

Proof 20.9 principle 20.1:
Every vector $\mathbf{v} \in \mathcal{V}$ can be represented by a basis eq. (20.14) of $\mathcal{V}$. With *homogentity*eq. (20.57) and *additivity*eq. (20.56) it follows for the image of all $\mathbf{v} \in \mathcal{V}$:

$$l(\mathbf{v}) = l(\alpha_1 b_1 + \cdots + \alpha_n b_n) = l\alpha_1(b_1) + \cdots + l(\alpha_n) b_n \qquad (20.134)$$

$\Rightarrow$ the image of the basis of $\mathcal{V}$ determines the linear mapping.

Proof 20.10 Proof [Corollary 20.17]:

$$\implies \quad l_\mathbf{A}(\alpha\mathbf{x} + \mathbf{y}) = \mathbf{A}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\mathbf{A}\mathbf{x} + \beta\mathbf{A}\mathbf{y} = \alpha l(\mathbf{x}) + \beta l(\mathbf{y})$$

$\impliedby$ Let $\mathfrak{B}$ be a standard normal basis of $\mathcal{V}$ with eq. (20.134):

$$l(\mathbf{x}) = \sum_{i=1}^{n} x_i l(\mathbf{e}_i) = \sum_{i=1}^{n} x_i \mathbf{A}_{:,i} = \mathbf{A}\mathbf{x} \qquad \mathbf{A}_{:,i} := \mathbf{l}(\mathbf{e_i}) \in \mathbb{R}^n$$

Proof 20.11 Proof Property 20.11:

$$(g \circ f)(\alpha\mathbf{x}) = g(f(\alpha\mathbf{x})) = g(\alpha f(\mathbf{x})) = \alpha(g \circ f)(\mathbf{x})$$
$$(g \circ f)(\mathbf{x} + \mathbf{y}) = g(f(\mathbf{x} + \mathbf{y})) = g(f(\mathbf{x}) + f(\mathbf{y}))$$
$$= (g \circ f)(\mathbf{x}) + (g \circ f)(\mathbf{y})$$

or even simpler as every linear form can be represented by a matrix product:

$$f(y) = \mathbf{A}\mathbf{y} \qquad g(z) = \mathbf{B}\mathbf{z} \quad \Rightarrow \quad (f \circ g)(\mathbf{x}) = \mathbf{A}\mathbf{B}\mathbf{x} := \mathbf{C}\mathbf{x}$$

Proof 20.12: [Corollary 20.22] Let $\mathbf{y} \in \mathtt{N}(\mathbf{A})$ ($\mathbf{z} \in \mathtt{N}(\mathbf{A}^\intercal)$) then it follows:

$$\begin{aligned} &\mathtt{N}(\mathbf{A}) \perp \mathfrak{R}(\mathbf{A}^\intercal) \qquad && (\mathbf{A}^\intercal\mathbf{x})^\intercal\mathbf{y} = \mathbf{x}^\intercal\mathbf{A}\mathbf{y} = \mathbf{x}^\intercal\mathbf{0} = 0 \\ &\mathtt{N}(\mathbf{A}^\intercal) \perp \mathfrak{R}(\mathbf{A}) \qquad && (\mathbf{A}\mathbf{x})^\intercal\mathbf{z} = \mathbf{x}^\intercal\mathbf{A}^\intercal\mathbf{z} = \mathbf{x}^\intercal\mathbf{0} = 0 \end{aligned}$$

19.3. Special Matrices

Proof 20.13 [Corollary 20.15]: Let $\mathbf{u} = \mathbf{S}^{-1}\mathbf{v}$ then it follows:

$$\mathbf{S}^{-1}\mathbf{A}\mathbf{S}\mathbf{u} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}\mathbf{v} = \lambda\mathbf{S}^{-1}\mathbf{v} = \lambda\mathbf{u}$$

Proof 20.14 Property 20.6:

$$\|\mathbf{Q}\mathbf{x}\|_2^2 = (\mathbf{Q}\mathbf{x})^\intercal\mathbf{Q}\mathbf{x} = \mathbf{x}^\intercal\mathbf{Q}^\intercal\mathbf{Q}\mathbf{x} = \|\mathbf{x}\|_2^2$$

Proof 20.15: Property 20.4

Let $\mathbf{A} \in \mathbb{K}^{n\times n}$ be a hermitian matrix[def. 20.24] and let $\lambda \in \mathbb{K}$ be an eigenvalue of $\mathbf{A}$ with corresponding eigenvector $\mathbf{v} \in \mathbb{K}^n$:

$$\lambda(\bar{\mathbf{v}}^\intercal\mathbf{v}) = \bar{\mathbf{v}}^\intercal\lambda\mathbf{v} = \bar{\mathbf{v}}^\intercal\mathbf{A}\mathbf{v} = \overline{(\bar{\mathbf{v}}^\intercal\mathbf{A}\mathbf{v})} = \overline{\mathbf{A}\mathbf{v}}^\intercal\mathbf{v} = \bar{\lambda}\,(\bar{\mathbf{v}}^\intercal\mathbf{v})$$

$$\lambda\,(\bar{\mathbf{v}}^\intercal\mathbf{v}) = \bar{\lambda}\,(\bar{\mathbf{v}}^\intercal\mathbf{v})$$

1. $\bar{\mathbf{v}}\mathbf{v} = \sum_{i=1}^n |v_i|^2 > 0$ as $\mathbf{v} \neq \mathbf{0}$
2. $\lambda = \bar{\lambda}$ which can only hold for $\lambda \in \mathbb{R}$ (Equation (11.8))

Proof 20.16: **??**

19.4. Vector Spaces

Proof 20.17 Definition 20.19: We know that $\text{proj}_L(\mathbf{u})$ must be a vector times a certain magnitude:

$$\text{proj}_L(\mathbf{u}) = \alpha\tilde{\mathbf{v}} \qquad \alpha \in \mathbb{K} \tag{20.135}$$

the magnitude follows from the scalar projection[def. 20.53] in the direction of $\mathbf{v}$ which concludes the derivation.

Proof 20.18 Definition 20.19 (via orthogonality): We know that $\mathbf{u} - \text{proj}_L(\mathbf{u})$ must be orthogonal[def. 20.65] to $\mathbf{v}$

$$(\mathbf{u} - \text{proj}_L(\mathbf{u}))^\intercal\mathbf{v} = (\mathbf{u} - \alpha\mathbf{v})^\intercal\mathbf{v} = 0 \Rightarrow \qquad \alpha = \frac{\mathbf{u}^\intercal\mathbf{v}}{\mathbf{v}^\intercal\mathbf{v}}$$

Proof 20.19: Definition 20.20 Let $\mathfrak{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ a basis of $\mathcal{U}$ s.t. by [cor. 20.4]:

$$\mathbf{u} = \sum_{i=1}^n \alpha_i\mathbf{b}_i$$

the coefficients $\{\alpha_i\}_{i=1}^n$ need to be determined. We know that:

$$\mathbf{v} - \mathbf{u} \perp \mathbf{b}_1, \ldots, \mathbf{v} - \mathbf{u} \perp \mathbf{b}_n$$

$$\Longrightarrow \qquad \left(\mathbf{v} - \sum_{i=1}^n \alpha_i\mathbf{b}_i\right)\cdot\mathbf{b}_j = 0 \qquad j = 1, \ldots, n$$

this linear system of equations can be rewritten as:

$$(\mathbf{b}_1 \cdots\cdots \mathbf{b}_n)\begin{pmatrix}\mathbf{b}_1\\ \vdots \\ \mathbf{b}_n\end{pmatrix}\begin{pmatrix}\alpha_1\\ \vdots \\ \alpha_n\end{pmatrix} = \begin{pmatrix}\mathbf{b}_1\\ \vdots \\ \mathbf{b}_n\end{pmatrix}\mathbf{v}$$

Proof 20.20: Corollary 20.27

Let $\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\intercal$ be the eigendecomposition[cor. 20.12] of $\mathbf{A}$ then it follows:

$$\begin{aligned} \min_{\tilde{\mathbf{n}}^\intercal\tilde{\mathbf{n}}=1} \tilde{\mathbf{n}}^\intercal\mathbf{A}\tilde{\mathbf{n}} &= \min_{\|\tilde{\mathbf{n}}\|=1} \tilde{\mathbf{n}}^\intercal(\mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\intercal)\tilde{\mathbf{n}} \\ &= \min_{\|\tilde{\mathbf{n}}\|=1} (\mathbf{Q}^\intercal\tilde{\mathbf{n}})^\intercal\boldsymbol{\Lambda}(\mathbf{Q}^T\tilde{\mathbf{n}}) \\ &= \min_{\mathbf{x}=1} \mathbf{x}^\intercal\boldsymbol{\Lambda}\mathbf{x} \qquad \mathbf{x} := \mathbf{Q}^\intercal\tilde{\mathbf{n}} \\ &= \min_{\mathbf{x}=1} \sum_{i=1}^n \mathbf{x}_i^2\Lambda_{ii} = \min_{\mathbf{x}=1} \sum_{i=1}^n \mathbf{x}_i^2\lambda_i \end{aligned}$$

Thus in order to obtain the minimum value we need to choose the eigenvector that leads to the smallest eigenvalue.

19.5. Norms

Proof 20.21: **??** 20.21

$$|\mathbf{u}\cdot\mathbf{v}| \overset{\text{eq. (20.79)}}{=} \|\mathbf{u}\|\|\mathbf{v}\||\cos\theta| \leqslant \|\mathbf{u}\|\|\mathbf{v}\|$$

Proof 20.22: Definition 20.59

$$\|\mathbf{u}+\mathbf{v}\|^2 = (\mathbf{u}+\mathbf{v})(\mathbf{u}+\mathbf{v}) = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2(\mathbf{u}\cdot\mathbf{v})$$

from cauchy schwartz we know:

$$\mathbf{u}\cdot\mathbf{v} \leqslant |\mathbf{u}\cdot\mathbf{v}| \overset{\text{eq. (20.89)}}{\leqslant} \|\mathbf{u}\|\|\mathbf{v}\|$$

$$\|\mathbf{u}+\mathbf{v}\|^2 \leqslant \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2(\|\mathbf{u}\|\|\mathbf{v}\|) = (\|\mathbf{u}\| + \|\mathbf{v}\|)^2$$

19.6. Decompositions

19.6.1. Symmetric - Antisemitic

Definition 20.86 Symmetric - Antisymmetric Decomposition: Any matrix $\mathbf{A} \in \mathbb{K}^{n\times n}$ can be decomposed into the sum of a *symmetric matrix*[def. 20.21] $\mathbf{A}^{\text{sym}}$ and a *skew-symmetric matrix***??** $\mathbf{A}^{\text{skes}}$:

$$\mathbf{A} = \mathbf{A}^{\text{sym}} + \mathbf{A}^{\text{skew}} \qquad \begin{aligned} \mathbf{A}^{\text{sym}} &= \frac{1}{2}\left(\mathbf{A} + \mathbf{A}^{\mathsf{H}}\right) \\ \mathbf{A}^{\text{skew}} &= \frac{1}{2}\left(\mathbf{A} - \mathbf{A}^{\mathsf{H}}\right) \end{aligned} \tag{20.136}$$

19.6.2. SVD

Proof 20.23 [Corollary 20.5]: $\mathbf{B} := \mathbf{A}^\intercal\mathbf{A}$ corresponds to a *symmetric positive definite* form[def. 20.73]:

$$\mathbf{x}^\intercal\mathbf{B}\mathbf{x} = \mathbf{x}^\intercal\mathbf{A}^\intercal\mathbf{A}\mathbf{x} = \|\mathbf{A}\mathbf{x}\|_2^2 > 0$$

thusProposition 20.6 follows immediately form [Corollary 20.2].

Proof 20.24 Proposition 20.6:

$$\mathbf{A}^\intercal\mathbf{A} \overset{\text{SVD}}{=} \left(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}\right)^{\mathsf{H}}\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}} = \mathbf{V}\boldsymbol{\Sigma}^{\mathsf{H}}\underbrace{\mathbf{U}^{\mathsf{H}}\mathbf{U}}_{\mathbf{I}_m}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}} = \mathbf{V}\boldsymbol{\Sigma}^{\mathsf{H}}\boldsymbol{\Sigma}\mathbf{V}^{\mathsf{H}}$$

$$\Longrightarrow \qquad \mathbf{V}^{\mathsf{H}}\mathbf{A}^\intercal\mathbf{A}\mathbf{V} = \boldsymbol{\Sigma}^\intercal\boldsymbol{\Sigma}$$

19.6.3. Eigendecomposition

Proof 20.25 Definition 20.84:

$$\mathbf{A}\mathbf{X} = [\lambda_1\mathbf{x}_1 \cdots\cdots \lambda_n\mathbf{x}_n] = \mathbf{X}\boldsymbol{\Lambda}$$

Geometry

Corollary 21.1 Affine Transformation in 1D: Given: numbers $x \in \hat{\Omega}$ with $\hat{\Omega} = [a, b]$
The affine transformation of $\phi : \hat{\Omega} \to \Omega$ with $y \in \Omega = [c, d]$ is defined by:

$$y = \phi(x) = \frac{d-c}{b-a}(x-a) + c \tag{21.1}$$

Proof 21.1: [cor. 21.1] By [def. 20.43] we want a function $f : [a, b] \to [c, d]$ that satisfies:

$$f(a) = c \qquad \textbf{and} \qquad f(b) = d$$

additionally $f(x)$ has to be a linear function ([def. 15.15]), that is the output scales the same way as the input scales.
Thus it follows:

$$\frac{d-c}{b-a} = \frac{f(x)-f(a)}{x-a} \qquad \iff \qquad f(x) = \frac{d-c}{b-a}(x-a) + c$$

Trigonometry

0.1. Trigonometric Functions

0.1.1. Sine

Definition 21.1 Sine:

$$\sin\alpha = \frac{\text{opposite}}{\text{hypotenuse}} = \frac{a}{c} \tag{21.2}$$

0.1.2. Cosine

Definition 21.2 Cosine:

$$\cos\alpha\alpha = \frac{\text{adjacent}}{\text{hypotenuse}} = \frac{b}{c} \tag{21.3}$$

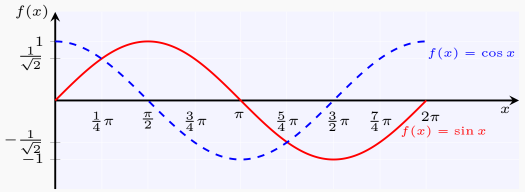
0.1.3. Tangens

Definition 21.3 Tangens:

$$\cos\alpha\alpha = \frac{\text{opposite}}{\text{adjacent}} = \frac{a}{b} = \frac{a/c}{b/c} = \frac{\sin\alpha}{\cos\alpha} \tag{21.4}$$

0.1.4. Trigonometric Functions and the Unit Circle

Sine and Cosine



$$\cos x \overset{(15.52)}{=} \frac{1}{2}\left[e^{ix} + e^{-ix}\right] \tag{21.5}$$

$$\sin x \overset{(15.52)}{=} \frac{1}{2i}\left[e^{ix} - e^{-ix}\right] = -\frac{i}{2}\left[e^{ix} - e^{-ix}\right] \tag{21.6}$$

Note

Using theorem 21.1 if follows:

$$\cos(\alpha \pm \pi) = -\cos\alpha \qquad \text{and} \qquad \sin(\alpha \pm \pi) = -\sin\alpha \tag{21.7}$$

0.1.5. Sinh

Definition 21.4 Sinh:

$$\sinh x \overset{(eq.\ (15.52))}{=} \frac{1}{2}\left[e^{x} - e^{-x}\right] = -\mathrm{i}\sin(\mathrm{i}\,x) \tag{21.8}$$

Property 21.1: $\sinh x = 0$ has a unique root at $x = 0$.

0.1.6. Cosh

Definition 21.5 Cosh:

$$\cosh x \overset{(15.52)}{=} \frac{1}{2}\left[e^{x} + e^{-x}\right] = \cos(\mathrm{i}\,x) \tag{21.9}$$

(21.10)

Property 21.2: $\cosh x$ is strictly positive.

Proof 21.2:

$$e^{x} = \cosh x + \sinh x \qquad e^{-x} = \cosh x - \sinh x \tag{21.11}$$

0.2. Addition Theorems

Theorem 21.1 Addition Theorems:

$$\sin(\alpha \pm \beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta \tag{21.12}$$

$$\cos(\alpha \pm \beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta \tag{21.13}$$

0.3. Werner Formulas

Werner Formulas

$$\sin\alpha\cos\beta = \frac{1}{2}\left[\sin(\alpha+\beta) + \sin(\alpha-\beta)\right] \tag{21.14}$$

$$\sin\alpha\sin\beta = \frac{1}{2}\left[\cos(\alpha-\beta) - \cos(\alpha+\beta)\right] \tag{21.15}$$

$$\cos\alpha\cos\beta = \frac{1}{2}\left[\cos(\alpha+\beta) + \cos(\alpha-\beta)\right] \tag{21.16}$$

Note

Using theorem 21.1 if follows:

$$\cos(\alpha \pm \pi) = -\cos\alpha \qquad \text{and} \qquad \sin(\alpha \pm \pi) = -\sin\alpha \tag{21.17}$$

0.4. Law of Cosines

Law 21.1 Law of Cosines [proof 21.3]**:**
relates the three side of a *general* triangle to each other.

$$a^2 = b^2 + c^2 - 2bc\cos\theta_{b,c} \tag{21.18}$$

Law 21.2 Law of Cosines for Vectors [proof 21.4]**:**
relates the length of vectors to each other.

$$\|\mathbf{a}\|^2 = \|\mathbf{c}-\mathbf{b}\|^2 = \|\mathbf{b}\|^2 + \|\mathbf{c}\|^2 - 2\|\mathbf{b}\|\|\mathbf{c}\|\cos\theta_{\mathbf{b},\mathbf{c}} \tag{21.19}$$

Law 21.3 Pythagorean theorem: special case of **??** for right triangle:

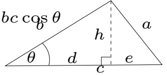
$$a^2 = b^2 + c^2 \tag{21.20}$$

1. Proofs

Proof 21.3: Law 21.1 From the defintion of the sine and cosine we know that:

$$\sin\theta = \frac{h}{b} \Rightarrow \underline{h} \qquad \text{and} \qquad \cos\theta = \frac{d}{b} \Rightarrow \underline{d}$$

$$\begin{aligned} \underline{e} &= c - \underline{d} = c - b\cos\theta \\ a^2 &= \underline{e}^2 + \underline{h}^2 = c^2 - 2cb\cos\theta + b^2\cos^2\theta + b^2\sin^2\theta \\ &= c^2 + b^2 - 2bc\cos\theta \end{aligned}$$



Proof 21.4: Law 21.2 Notice that $\mathbf{c} = \mathbf{a} + \mathbf{b} \Rightarrow \mathbf{a} = \mathbf{c} - \mathbf{b}$ and we can either use **??** 21.3 or notice that:

$$\begin{aligned} \|\mathbf{c}-\mathbf{b}\|^2 &= (\mathbf{c}-\mathbf{b})\cdot(\mathbf{c}-\mathbf{b}) \\ &= \mathbf{c}\cdot\mathbf{c} - 2\mathbf{c}\cdot\mathbf{b} + \mathbf{b}\cdot\mathbf{b} \\ &= \|\mathbf{c}\|^2 + \|\mathbf{b}\|^2 - 2\left(\|\mathbf{c}\|\|\mathbf{b}\|\cos\theta\right) \end{aligned}$$

Topology

Definition 22.1 Topology of set τ**:**
Let X be a set. A collection τ of open**??** subsets of X is called *topology* of X if it satisfies:

- $\varnothing \in \tau$ and $X \in \tau$
- Any finite or infinite union of subsets of τ is contained in τ:
$$\{U_i : i \in \mathbf{I}\} \subseteq \tau \qquad \Longrightarrow \qquad \cup_{i\in\mathbf{I}} U_i \in \tau \tag{22.1}$$
- The intersection of a finite number of elements of τ also belongs to τ:
$$\{U_i\}_{i=1}^{n} \in \tau \qquad \Longrightarrow \qquad U_1 \cap \cdots \cap U_n \in \tau \tag{22.2}$$

Definition 22.2 Topological Space[?] (X, τ)**:**
Is an ordered pair (X, τ), where X is a set and τ is a topology[def. 22.1] on X.

Numerical Methods

1. Machine Arithmetic's

1.1. Machine Numbers

Definition 23.1 Institute of Electrical and Electronics Engineers (IEEE): Is a engineering associations that defines a standard on how computers should treat machine numbers in order to have certain guarantees.

Definition 23.2 Machine/Floating Point Numbers $\mathbb{F}$: Computers are only capable to represent a *finite, discrete* set of the real numbers $\mathbb{F} \subset \mathbb{R}$

1.1.1. Floating Point Arithmetic's $x\tilde{\Omega}y = \mathbf{fl}(x\Omega y)$

Corollary 23.1 Closure: Machine numbers $\mathbb{F}$ are not *closed*[def. 11.7] under basic arithmetic operations:

$$\mathbb{F}\,\Omega\,\mathbb{F} \mapsto \not{\mathbb{F}} \qquad \Omega = \{+,-,*,/\} \tag{23.1}$$

Note

Corollary 23.1 provides a problem as the computer can only represent floating point number $\mathbb{F}$.

Definition 23.3 Floating Point Operation $\tilde{\Omega}$: Is a basic arithmetic operation that obtains a number $x \in \mathbb{F}$ by applying a function rd:

$$\mathbb{F}\,\tilde{\Omega}\,\mathbb{F} \mapsto \mathbb{F} \qquad \begin{aligned}\tilde{\Omega} &:= \mathrm{rd}\circ\Omega \\ \Omega &= \{+,-,*,/\}\end{aligned} \tag{23.2}$$

Definition 23.4 Rounding Function rd: Given a real number $x \in \mathbb{R}$ the rounding function replaces it by the nearest machine number $\tilde{x} \in \mathbb{F}$. If this is ambiguous (there are two possibilities), then it takes the larger one:

$$\mathrm{rd} : \begin{cases}\mathbb{R} \mapsto \mathbb{F} \\ x \mapsto \max \operatorname*{arg\,min}_{\tilde{x}\in\mathbb{F}} |x - \tilde{x}|\end{cases} \tag{23.3}$$

Consequence

Basic arithmetic rules such as associativity do no longer hold for operations such as addition and subtraction.

Axiom 23.1 Axiom of Round off Analysis: Let $x, y \in \mathbb{F}$ be (normalized) floats and assume that $x\tilde{\Omega}y \in \mathbb{F}$ (i.e. no over/underflow). Then it holds that:

$$\begin{aligned} x\tilde{\Omega}y &= (x\Omega y)(1+\delta) && \Omega = \{+,-,*,/\} \\ \tilde{f}(x) &= f(x)(1+\delta) && f \in \{\exp, \sin, \cos, \log, \dots\} \end{aligned} \tag{23.4}$$

with $|\delta| < \mathrm{EPS}$

Explanation 23.1 (axiom 23.1). *gives us a guarantee that for any two floating point numbers $x, y \in \mathbb{F}$, any operation involving them will give a floating point result which is within a factor of $1+\delta$ of the true result $x\Omega y$.*

Definition 23.5 Overflow: Result is bigger then the biggest representable floating point number.

Definition 23.6 Underflow: Result is smaller then the smaller representable floating point number i.e. to close to zero.

1.2. Roundoff Errors

Log-Sum-Exp Trick

The sum exponential trick is at trick that helps to calculate the log-sum-exponential in a robust way by avoiding over/underflow. The log-sum-exponential[def. 23.7] is an expression that arises frequently in machine learning i.e. for the cross entropy loss or for calculating the evidence of a posterior prediction.
The root of the problem is that we need to calculate the exponential $\exp(x)$, this comes with two different problems:

- If x is large (i.e. 89 for single precision floats) then $\exp(x)$ will lead to overflow
- If x is very negative $\exp(x)$ will lead to underflow/0. This is not necessarily a problem but if $\exp(x)$ occurs in the denominator or the logarithm for example this is catastrophic.

Definition 23.7 Log sum Exponential:

$$\mathrm{LogSumExp}(x_1, \dots, x_n) := \log\left(\sum_{i=1}^n \mathrm{e}^{x_i}\right) \tag{23.5}$$

Formula 23.1 Log-Sum-Exp Trick:

$$\log\left(\sum_{i=1}^n \mathrm{e}^{x_i}\right) = a + \log\sum_{i=1}^n \mathrm{e}^{x_i - a} \qquad a := \max_{i\in\{1,\dots,n\}} x_i \tag{23.6}$$

Explanation 23.2 (formula 23.1). *The value a can be any real value but for robustness one usually chooses the max s.t.*

- *The leading digits are preserved by pulling out the maximum a*
- *Inside the log only zero or negative numbers are exponentiated, so there can be no overflow.*
- *If there is underflow inside the log we know that at least the leading digits have been returned by the max.*

Proof 23.1:

$$\begin{aligned}\mathrm{LSE} &= \log\Big(\sum_{i=1}^n e^{x_i}\Big) = \log\Big(\sum_{i=1}^n e^{x_i - a}e^{a}\Big) \\ &= \log\Big(e^{a}\sum_{i=1}^n e^{x_i - a}\Big) = \log\Big(\sum_{i=1}^n e^{x_i - a}\Big) + \log(e^{a}) \\ &= \log\Big(\sum_{i=1}^n e^{x_i - a}\Big) + a\end{aligned}$$

Definition 23.8 Partition Π: Given an interval $[0, T]$ a sequence of values $0 < t_0 < \cdots < t_n < T$ is called a partition $\Pi(t_0, \dots, t_n)$ of this interval.

2. Convergence

2.1. O-Notation

2.1.1. Small $o(\cdot)$ Notation

Definition 23.9 Little o Notation:

$$f(n) = o(g(n)) \iff \lim_{n\to\infty}\frac{f(n)}{g(n)} = 0 \tag{23.7}$$

2.1.2. Big $\mathcal{O}(\cdot)$ Notation

3. Rate Of Convergence

Definition 23.10 Rate of Convergence: Is a way to measure the rate of convergence of a sequence $\{\mathbf{x}^{(k)}\}_k \in \mathbb{R}^n$ to a value to $\mathbf{x}^*$. Let $\rho \in [0,1]$ be the *rate of convergence* and define:

$$\lim_{k\mapsto\infty}\frac{\left\|\mathbf{x}^{k+1} - \mathbf{x}^*\right\|}{\left\|\mathbf{x}^{k} \quad - \mathbf{x}^*\right\|} = \rho \tag{23.8}$$

$$\iff \lim_{k\mapsto\infty}\left\|\mathbf{x}^{k+1} - \mathbf{x}^*\right\| \leqslant \rho\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\| \qquad \forall k \in \mathbb{N}_0$$

Definition 23.11 Linear/Exponential Convergence: A sequence $\{\mathbf{x}^{(k)}\}_k \in \mathbb{R}^n$ converges *linearly* to $\mathbf{x}^*$ if in the asymptotic limit $k \to \infty$ if it satisfies:

$$\rho \in (0,1) \qquad \forall k \in \mathbb{N}_0 \tag{23.9}$$

Definition 23.12 Superlinear Convergence: A sequence $\{\mathbf{x}^{(k)}\}_k \in \mathbb{R}^n$ converges *superlinear* to $\mathbf{x}^*$ if in the asymptotic limit $k \to \infty$ if it satisfies:

$$\rho = 1 \tag{23.10}$$

Definition 23.13 Sublinear Convergence: A sequence $\{\mathbf{x}^{(k)}\}_k \in \mathbb{R}^n$ converges *sublinear* to $\mathbf{x}^*$ if in the asymptotic limit $k \to \infty$ if it satisfies:

$$\rho = 0 \iff \left\|\mathbf{x}^{k+1} - \mathbf{x}^*\right\| = o\left(\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|\right) \tag{23.11}$$

Definition 23.14 Logarithmic Convergence: A sequence $\{\mathbf{x}^{(k)}\}_k \in \mathbb{R}^n$ converges *logarithmically* to $\mathbf{x}^*$ if it converges *sublinear*[def. 23.13] and additoinally satisfies

$$\rho = 0 \iff \left\|\mathbf{x}^{k+2} - \mathbf{x}^{k+1}\right\| = o\left(\left\|\mathbf{x}^{k+1} - \mathbf{x}^{k}\right\|\right) \tag{23.12}$$

Exponetial Convergence

Linear convergence is sometimes called exponential convergence. This is due to the fact that:

1. We often have expressions of the form:
$$\left\|\mathbf{x}^{k+1} - \mathbf{x}^*\right\| \leqslant \underbrace{(1-\alpha)}_{:=\rho}\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|$$
2. and that $(1-\alpha) = \exp(-\alpha)$ from which follows that:
$$\text{eq. (23.13)} \iff \left\|\mathbf{x}^{k+1} - \mathbf{x}^*\right\| \leqslant \mathrm{e}^{-\alpha}\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|$$

Definition 23.15 Convergence of order p: In order to distinguish *superlinear convergence* we define the order of convergence.
A sequence $\{\mathbf{x}^{(k)}\}_k \in \mathbb{R}^n$ converges superlinear with order $p \in \{2, \dots\}$ to $\mathbf{x}^*$ if it satisfies:

$$\lim_{k\mapsto\infty}\frac{\left\|\mathbf{x}^{k+1} - \mathbf{x}^*\right\|}{\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|^p} = C \qquad C < 1 \tag{23.13}$$

Definition 23.16 Exponential Convergence: A sequence $\{\mathbf{x}^{(k)}\}_k \in \mathbb{R}^n$ converges exponentially with rate ρ to $\mathbf{x}^*$ if in the asymptotic limit $k \to \infty$ it satisfies:

$$\left\|\mathbf{x}^{k+1} - \mathbf{x}^*\right\| \leqslant \rho^k\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\| \qquad \rho < 1 \tag{23.14}$$

$$\left\|\mathbf{x}^{k+1} - \mathbf{x}^*\right\| \in o\left(\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|\right) \tag{23.15}$$

4. Linear Systems of Equations

4.1. Direct Methods

4.1.1. LU-Decomposition

Definition 23.17 LU Decomposition:

4.1.2. Symmetric Matrices

LDL-Decomposition

4.1.3. Symmetric Positive Definite Matrices

For linear systems with s.p.d.[def. 20.73] matrices $\mathbf{A}$ the LU-decomposition[def. 23.17] simplifies to the Cholesky Decomposition[def. 23.18].

Cholesky Decomposition

Definition 23.18 Cholesky Decomposition: Let $\mathbf{A}$ be a s.p.d.[def. 20.73] then it can be factorized into:

$$\mathbf{A} = \mathbf{G}\mathbf{G}^\intercal \qquad \text{with} \qquad \mathbf{G} := \mathbf{L}\mathbf{D}^{1/2} \tag{23.16}$$

4.2. Iterative Methods

5. Iterative Methods for Non-linear Systems

Definition 23.19 General Non-linear System of Equations (NLSE) F: Is a system of non-linear equations F (that do **not** satisfy linearity??):

$$F :\subseteq \mathbb{R}^n \mapsto \mathbb{R}^n \qquad \text{seek to find} \qquad \mathbf{x} \in \mathbb{R}^n : F(\mathbf{x}) = \mathbf{0} \tag{23.17}$$

Definition 23.20 Stationary m-point Iteration ϕ_F: Let $n, m \in \mathbb{R}$ and let $U \subseteq (R^n)^m = \mathbb{R}^n \times \cdots \times \mathbb{R}^n$ be a set. The function $\phi : U \mapsto \mathbb{R}^n$, called ($m$-point) iteration function is an iterative algorithm that produces an iterative sequence $\left(\mathbf{x}^{(k)}\right)_k$ of approximate solutions to eq. (23.17), using the m most recent iterates:

$$\mathbf{x}^{(k)} = \phi_F\left(\mathbf{x}^{(k-1)}, \dots, \mathbf{x}^{(k-m)}\right) \tag{23.18}$$

$$\text{Inital Guess} \qquad \mathbf{x}^{(0)}, \dots, \mathbf{x}^{(m-1)}$$

Note

Stationary as ϕ does no explicitly depend on k.

Definition 23.21 Fixed Point $\mathbf{x}^*$: Is a point $\mathbf{x}^*$ for which the sequence does not change anymore:

$$\mathbf{x}^* = \phi_F\left(\mathbf{x}^{(k-1)}, \dots, \mathbf{x}^{(k-m)}\right) \quad \text{with} \quad \begin{aligned}\mathbf{x}^{(k-1)} &= \mathbf{x}^* \\ &\vdots \\ \mathbf{x}^{(k-m)} &= \mathbf{x}^*\end{aligned} \tag{23.19}$$

5.0.1. Convergence

Question

Does the sequence $\left(\mathbf{x}^{(k)}\right)_k$ converge to a limit:

$$\lim_{k\to\infty}\mathbf{x}^{(k)} = \mathbf{x}^* \tag{23.20}$$

5.0.2. Consistency

Definition 23.22 Consistent m-point Iterative Method: A *stationary* m-point method[def. 23.20] is *consistent* with a non-lineary system of equations[def. 23.19] F iff:

$$F\left(\mathbf{x}^*\right) \iff \phi_F\left(\mathbf{x}^*, \dots, \mathbf{x}^*\right) = \mathbf{x}^* \tag{23.21}$$

5.0.3. Speed of Convergence

5.1. Fixed Point Iterations $m = 1$

Definition 23.23 Fixed Point Iteration: Is a 1-point method $\phi_F : U \subset \mathbb{R}^n \mapsto \mathbb{R}^n$ that seeks a fixed point $\mathbf{x}^*$ to solve $F(\mathbf{x}) = 0$:

$$\mathbf{x}^{(k+1)} = \phi_F\left(\mathbf{x}^{(k)}\right) \qquad \text{Inital Guess: } \mathbf{x}^{(0)} \tag{23.22}$$

Corollary 23.2 Consistency: If ϕ_F is *continuous* and $\mathbf{x}^* = \lim_{k\to\infty} x^{(k)}$ then $\mathbf{x}^*$ is a fixed point[def. 23.21] of ϕ.

Algorithm 23.1 Fixed Point Iteration:

```
    Input: Inital Guess: $\mathbf{x}^{(0)}$
1:  Rewrite $F(\mathbf{x}) = 0$ into a form of $\mathbf{x} = \phi_F(\mathbf{x})$
                                        ▷ There exist many ways
2:  for $k = 1, \dots, T$ do
3:      Use the fixed point method:
            $\mathbf{x}^{(k+1)} = \phi_F\left(\mathbf{x}^{(k)}\right)$        (23.23)
4:  end for
```

6. Numerical Quadrature

Definition 23.24 Order of a Quadrature Rule: The order of a quadrature rule $\mathcal{Q}_n : \mathcal{C}^0([a, b]) \to \mathbb{R}$ is defined as:

$$\mathrm{order}(\mathcal{Q}_n) := \max\left\{n \in \mathbb{N}_0 : \mathcal{Q}_n(p) = \in_a^b p(t)\,\mathrm{d}t \quad \forall p \in \mathcal{P}_n\right\} + 1 \tag{23.24}$$

Thus it is the maximal degree+1 of polynomials (of degree maximal degree) $\mathcal{P}_{\text{maximal degree}}$ for which the quadrature rule yields exact results.

Note

Is a quality measure for quadrature rules.

6.1. Composite Quadrature

Definition 23.25 Composite Quadrature: **Given** a mesh $\mathcal{M} = \{a = x_0 < x_1 < \ldots < x_m = b\}$ apply a Q.R. $\mathcal{Q}_n$ to each of the mesh cells $I_j := [x_{j-1}, x_j] \quad \forall j = 1, \dots, m \mathrel{\widehat{=}}$ p.w. Quadrature:

$$\int_a^b f(t)\,\mathrm{d}t = \sum_{j=1}^m \int_{x_{j-1}}^{x_j} f(t)\,\mathrm{d}t = \sum_{j=1}^m \mathcal{Q}_n(f_{I_j}) \tag{23.25}$$

Lemma 23.1 Error of Composite quadrature Rules:
Given a function $f \in \mathcal{C}^k([a,b])$ with integration domain:

$$\sum_{i=1}^{m} h_i = |b-a| \qquad \text{for } \mathcal{M} = \{x_j\}_{j=1}^{m}$$

Let: $h_{\mathcal{M}} = \max_j |x_j, x_{j-1}|$ be the mesh-width
Assume an equal number of quadrature nodes for each interval $I_j = [x_{j-1}, x_j]$ of the mesh $\mathcal{M}$ i.e. $n_j = n$.
Then the error of a quadrature rule $\mathcal{Q}_n(f)$ of order q is given by:

$$\begin{aligned}\epsilon_n(f) &= \mathcal{O}\left(n^{-\min\{k,q\}}\right) = \mathcal{O}\left(h_{\mathcal{M}}^{\min\{k,q\}}\right) && \text{for } n \to \infty \\ &\overset{[\text{cor. } 15.6]}{=} \mathcal{O}\left(n^{-q}\right) = \mathcal{O}\left(h_{\mathcal{M}}^{q}\right) && \text{with } h_{\mathcal{M}} = \frac{1}{n}\end{aligned} \tag{23.26}$$

Definition 23.26 Complexity W: Is the number of function evaluations $\triangleq$ number of quadrature points.

$$W(\mathcal{Q}(f)_n) = \#\text{f-eval} \triangleq n \tag{23.27}$$

Lemma 23.2 Error-Complexity $W(\epsilon_n(f))$: Relates the complexity to the quadrature error.
Assuming and quadrature error of the form :

$$\epsilon_n(f) = \mathcal{O}(n^{-q}) \qquad \Longleftrightarrow \qquad \epsilon_n(f) = cn^{-q} \qquad c \in \mathbb{R}_+$$

the error complexity is algebraic (**??**) and is given by:

$$W(\epsilon_n(f)) = \mathcal{O}(\epsilon_n^{1/q}) = \mathcal{O}\left(\sqrt[q]{\epsilon_n}\right) \tag{23.28}$$

Proof 23.2: lemma 23.2: **Assume**: we want to reduce the error by a factor of ϵ_n by increasing the number of quadrature points $n_{\text{new}} = a \cdot n_{\text{old}}$.
Question: what is the additional effort (#f-eval) needed in order to achieve this reduction in error?

$$\frac{c \cdot n_n^q}{c \cdot n_o^q} = \frac{1}{\epsilon_n} \qquad \Rightarrow \qquad n_n = n_o \cdot \sqrt[q]{\epsilon_n} = \mathcal{O}(\sqrt[q]{\epsilon_n}) \tag{23.29}$$

6.1.1. Simpson Integration

Definition 23.27 Simpson Integration:

Optimization

Definition 24.1 Fist Order Method: A first-order method is an algorithm that chooses the k-th iterate in

$$\mathbf{x}_0 + \operatorname{span}\{\nabla f(\mathbf{x}_0), \dots \nabla f(\mathbf{x}_{k-1})\} \qquad \forall k = 1, 2, \dots \tag{24.1}$$

Note

Gradient descent is a first order method

1. Linear Optimization

1.1. Polyhedra

Definition 24.2 Polyhedron: Is a set $P \in \mathbb{R}^n$ that can be described by the *finite* intersection of m closed *half spaces*??:

$$P = \{\mathbf{x} \in \mathbb{R}^n \,|\mathbf{A}\mathbf{x} \leqslant \mathbf{b}\} = \{\mathbf{x} \in \mathbb{R}^n \,|\mathbf{a}_j \mathbf{x} \leqslant b_j, j = 1, \dots, m\}$$

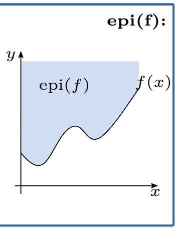
$$\mathbf{A} \in \mathbb{R}^{m \times n} \qquad \mathbf{b} \in \mathbb{R}^m \tag{24.2}$$

1.1.1. Polyhedral Function

Definition 24.3 Epigraph/Subgraph **epi(f):**

The epigraph of a function $f \in \mathbb{R}^n \mapsto \mathbb{R}$ is defined as the set of point that lie above its gaph:

$$\operatorname{epi}(f) := \{(\mathbf{x}, y) \in \mathbb{R}^n \,|y \geqslant f(\mathbf{x})\} \subseteq \mathbb{R}^{n+1} \tag{24.3}$$



Definition 24.4 Polyhedral Function: A function f is *polyhedral* if its epigraph $\operatorname{epi}(f)$[def. 24.3] is a polyhedral set[def. 24.2]:

$$f \text{ is polyhedral} \iff \operatorname{epi}(f) \text{ is polyhedral} \tag{24.4}$$

2. Lagrangian Optimization Theory

Definition 24.5 (Primal) Constraint Optimization: **Given** an optimization problem with domain $\Omega \subseteq \mathbb{R}^d$:

$$\begin{aligned} &\min_{\mathbf{w}\in\Omega} f(\mathbf{w}) && \\ \textbf{s.t.} \quad & g_i(\mathbf{w}) \leqslant 0 && 1 \leqslant i \leqslant k \\ & h_j(\mathbf{w}) = 0 && 1 \leqslant j \leqslant m \end{aligned}$$

Definition 24.6 Lagrange Function:

$$\mathscr{L}(\alpha, \beta, \mathbf{w}) := f(\mathbf{w}) + \alpha\mathbf{g}(\mathbf{w}) + \beta\mathbf{h}(\mathbf{w}) \tag{24.5}$$

Extremal Conditions

$\nabla \mathscr{L}(\mathbf{x}) \overset{!}{=} 0$ Extremal point $\mathbf{x}^*$

$\frac{\partial}{\partial \beta} \mathscr{L}(\mathbf{x}) = h(\mathbf{x}) \overset{!}{=} 0$ Constraint satififsfaction

For the inequality constraints $g(\mathbf{x}) \leqslant 0$ we distinguish two situations:

Case I : $g(\mathbf{x}^*) < 0$ switch const. off

Case II : $g(\mathbf{x}^*) \geqslant 0$ optimze using active eq. constr.

$\frac{\partial}{\partial \alpha} \mathscr{L}(\mathbf{x}) = g(\mathbf{x}) \overset{!}{=} 0$ Constraint satififsfaction

Definition 24.7 Lagrangian Dual Problem: Is given by:

$$\begin{aligned} \text{Find} \quad & \max \theta(\alpha, \beta) = \inf_{\mathbf{w}\in\Omega} \mathscr{L}(\mathbf{w}, \alpha, \beta) && \\ \textbf{s.t.} \quad & \alpha_i \geqslant 0 && 1 \leqslant i \leqslant k \end{aligned}$$

Solution Strategy

1. Find the extremal point $\mathbf{w}^*$ of $\mathscr{L}(\mathbf{w}, \alpha, \beta)$:
$$\left.\frac{\partial \mathscr{L}}{\partial \mathbf{w}}\right|_{\mathbf{w}=\mathbf{w}^*} \overset{!}{=} 0 \tag{24.6}$$
2. Insert $\mathbf{w}^*$ into $\mathscr{L}$ and find the extremal point β^* of the resulting dual Lagrangian $\theta(\alpha, \beta)$ for the active constraints:
$$\left.\frac{\partial \theta}{\partial \beta}\right|_{\beta=\beta^*} \overset{!}{=} 0 \tag{24.7}$$
3. Calculate the solution $\mathbf{w}^*(\beta^*)$ of the constraint minimization problem.

Value of the Problem

Value of the problem: the value $\theta(\alpha^*, \beta^*)$ is called the value of problem (α^*, β^*).

Theorem 24.1 Upper Bound Dual Cost: Let $\mathbf{w} \in \Omega$ be a feasible solution of the primal problem [def. 24.5] and (α, β) a feasible solution of the respective dual problem [def. 24.7]. Then it holds that:

$$f(\mathbf{w}) \geqslant \theta(\alpha, \beta) \tag{24.8}$$

Proof 24.1:

$$\begin{aligned} \theta(\alpha, \beta) &= \inf_{\mathbf{u}\in\Omega} \mathscr{L}(\mathbf{u}, \alpha, \beta) \leqslant \mathscr{L}(\mathbf{w}, \alpha, \beta) \\ &= f(\mathbf{w}) + \sum_{i=1}^{k} \underbrace{\alpha_i}_{\geqslant 0} \underbrace{g_i(\mathbf{w})}_{\leqslant 0} + \sum_{j=}^{m} \beta_j \underbrace{h_j(\mathbf{w})}_{=0} \\ &\leqslant f(\mathbf{w}) \end{aligned}$$

Corollary 24.1 Duality Gap Corollary: The value of the dual problem is upper bounded by the value of the primal problem:

$$\sup\{\theta(\alpha, \beta) : \alpha \geqslant 0\} \leqslant \inf\{f(\mathbf{w}) : \mathbf{g}(\mathbf{w}) \leqslant 0, \mathbf{h}(\mathbf{w}) = 0\} \tag{24.9}$$

Theorem 24.2 Optimality: The triple $(\mathbf{w}^*, \alpha^*, \beta^*)$ is a saddle point of the Lagrangian function for the primal problem, if and only if its components are optimal solutions of the primal and dual problems and if there is no duality gap, that is, the primal and dual problems having the same value:

$$f(\mathbf{w}^*) = \theta(\alpha^*, \beta^*) \tag{24.10}$$

Definition 24.8 Convex Optimization: **Given**: a convex function f and a convex set S solve:

$$\begin{aligned} &\min f(\mathbf{x}) \\ \text{s.t.} \quad & \mathbf{x} \in S \end{aligned} \tag{24.11}$$

Often S is specified using linear inequalities:

$$\text{e.g.} \qquad S = \left\{\mathbf{x} \in \mathbb{R}^d : \mathbf{A}\mathbf{x} \leqslant \mathbf{b}\right\}$$

Theorem 24.3 Strong Duality: Given an convex optimization problem:

$$\begin{aligned} &\min_{\mathbf{w}\in\Omega} f(\mathbf{w}) && \\ \textbf{s.t.} \quad & g_i(\mathbf{w}) \leqslant 0 && 1 \leqslant i \leqslant k \\ & h_j(\mathbf{w}) = 0 && 1 \leqslant j \leqslant m \end{aligned}$$

where g_i, h_i can be written as affine functions: $y(\mathbf{w}) = \mathbf{A}\mathbf{w} - b$.

Then it holds that the duality gap is zero and we obtain an optimal solution.

Theorem 24.4 Kuhn-Tucker Conditions: Given an optimization problem with convex domain $\Omega \subseteq \mathbb{R}^d$,

$$\begin{aligned} &\min_{\mathbf{w}\in\Omega} f(\mathbf{w}) && \\ \textbf{s.t.} \quad & g_i(\mathbf{w}) \leqslant 0 && 1 \leqslant i \leqslant k \\ & h_j(\mathbf{w}) = 0 && 1 \leqslant j \leqslant m \end{aligned}$$

with $f \in C^1$ convex and g_i, h_i affine.

Necessary and sufficient conditions for a normal point $\mathbf{w}^*$ to be an optimum are the existence of α^*, β^* s.t.:

$$\frac{\partial \mathscr{L}(\mathbf{w}, \alpha, \beta)}{\partial \mathbf{w}} \overset{!}{=} 0 \qquad \frac{\partial \mathscr{L}(\mathbf{w}^*, \alpha, \beta)}{\partial \beta} \overset{!}{=} 0 \tag{24.12}$$

under the condtions that:

- $\forall i_1, \dots, k \qquad \alpha_i^* g_i(\mathbf{w}^*) = 0$, s.t.:
  - Inactive Constraint: $g_i(\mathbf{w}^*) < 0 \to \alpha_i = 0$.
  - Active Constraint:
    $g_i(\mathbf{w}^*) \nless 0 \to \alpha_i \geqslant 0 \quad \text{s.t.} \quad \alpha_i^* g_i(\mathbf{w}^*) = 0$

Consequence

We may become very sparce problems, if a lot of constraints are not actice $\iff$ $\alpha_i = 0$.

Only a few points, for which $\alpha_i > 0$ may affact the decision surface.

Combinatorics

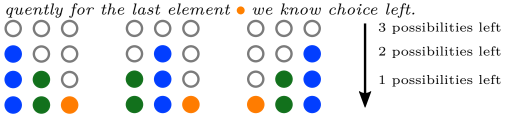
1. Permutations

Definition 25.1 Permutation: A n-Permutation is the (re)*arrangement* of n elements of a set[def. 11.1] $\mathcal{S}$ of size $n = |\mathcal{S}|$ into a sequences[def. 12.2].

Definition 25.2 Number of Permutations of a Set $n!$:
Let $\mathcal{S}$ be a set[def. 11.1] $n = |\mathcal{S}|$ *distinct* objects. The number of permutations of $\mathcal{S}$ is given by:

$$P_n(\mathcal{S}) = n! = \prod_{i=0}^{n-1}(n-i) = n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot 1 \tag{25.1}$$

Explanation 25.1. *If we have i.e. three distinct elements {●, ●, ●} For the first element ● that we arrange we have three possible choices where to put it. However this reduces the number of possible choices for the second element ● to only two. Consequently for the last element ● we know choice left.*



Definition 25.3
Number of Permutations of a Multiset:
Let $\mathcal{S}$ be a multi set[def. 11.3] with $n = |\mathcal{S}|$ total and k *distinct* objects. Let n_j be the multiplicity[def. 11.4] of the member $j \in \{1, \ldots, k\}$ of the multiset $\mathcal{S}$. The permutation of $\mathcal{S}$ is given by:

$$P_{n_1,\ldots,n_k}(\mathcal{S}) = \frac{n!}{n_1!\cdot\ldots\cdot n_k} \qquad \text{s.t.} \quad \sum_{j=1}^{k} n_j \leqslant n \qquad k < n \tag{25.2}$$

Note

We need to divide by the permutations as sequence/order does not change if we exchange objects of the same kind (e.g. red ball by red ball) ⇒ less possibilities to arrange the elements uniquely.

2. Combinations

Definition 25.4 k-Combination:
A k-combination of a set $\mathcal{S}$ of size $n = \mathcal{S}$ is a subset $\mathcal{S}_k$ (order does not matter) of $k = |\mathcal{S}_k|$ *distinct* elements, *chosen* from $\mathcal{S}$.

Definition 25.5 Number of k-Combinations $C_{n,k}$:
The number of k-combinations of a set $\mathcal{S}$ of size $n = \mathcal{S}$ is given by:

$$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{25.3}$$

3. Variation

Definition 25.6 Variation:
A k-variation of a set $\mathcal{S}$ of size $n = \mathcal{S}$ is

1. a selection/combination[def. 25.4] of a subset $\mathcal{S}_k$ (order does not matter) of k-*distinct* elements $k = |\mathcal{S}_k|$, *chosen* from $\mathcal{S}$
2. and an k arrangement/permutation[def. 25.2] of that subset $\mathcal{S}_k$ (with or without repetition) into a sequence[def. 12.2]

Definition 25.7
Number of Variations without repetitions V^n_k:
Let $\mathcal{S}$ be a set[def. 11.1] $n = |\mathcal{S}|$ *distinct* objects from which we choose k elements. The number of variations of size $k = |\mathcal{S}_k|$ of the set $\mathcal{S}$ *without repetitions* is given by:

$$V^n_k(\mathcal{S}) = \binom{n}{k}k! = \frac{n!}{(n-k)!} \tag{25.4}$$

Note

Sometimes also denotes as P^n_k.

Definition 25.8
Number of Variations with repetitions $\bar{V}^n_k$:
Let $\mathcal{S}$ be a set[def. 11.1] $n = |\mathcal{S}|$ *distinct* objects from which we choose k elements. The number of variations of size $k = |\mathcal{S}_k|$ of the set $\mathcal{S}$ from which we *choose and always return* is given by:

$$\bar{V}^n_k(\mathcal{S}) = n^k \tag{25.5}$$

Stochastics

Definition 25.9 Stochastics: Is a collective term for the areas of *probability theory* and *statistics*.

Definition 25.10 Statistics: Is concerned with the analysis of data/experiments in order to draw conclusion of the underlying governing models that describe these experiments.

Definition 25.11 Probability: Is concerned with the quantification of the uncertainty of random experiments by use of statistical models. Hence it is the opposite of statistics.

Definition 25.12 Probability: Probability is the measure of the likelihood that an event will occur in a Random Experiment. Probability is quantified as a number between 0 and 1, where, loosely speaking, 0 indicates impossibility and 1 indicates certainty.

Note: Stochastics vs. Stochastic

Stochasticss is a noun and is a collective term for the areas of probability theory and statistics, while stochastic is a *adjective*, describing that a certain phenomena is governed by uncertainty i.e. a process.

Probability Theory

Definition 26.1 Probability Space $W = \{\Omega, \mathcal{F}, \mathbb{P}\}$:
Is the unique triple $\{\Omega, \mathcal{F}, \mathbb{P}\}$, where Ω is its sample space, $\mathcal{F}$ its σ-algebra of events, and $\mathbb{P}$ its probability measure.

Definition 26.2 [example 26.1]
Sample Space Ω:
Is the set of all possible outcomes (elementary events [cor. 26.5]) of an experiment.

Definition 26.3 [example 26.2]
Event A:
An "event" is a subset of the sample space Ω and is a property which can be observed to hold or not to hold *after* the experiment is done.
Mathematically speaking not every subset of Ω is an event and has an associated probability.
Only those subsets of Ω that are part of the corresponding σ-algebra $\mathcal{F}$ are events and have their assigned probability.

Corollary 26.1 : If the outcome ω of an experiment is in the subset A, then the event A is said to "have occured".

Corollary 26.2 Complement Set A^C:
is the contrary event of A.

Corollary 26.3 The Union Set $A \cup B$:
Let A, B be two events. The event "A or B" is interpreted as the union of both.

Corollary 26.4 The Intersection Set $A \cap B$:
Let A, B be two events. The event "A and B" is interpreted as the intersection of both.

Corollary 26.5 The Elementary Event ω:
Is a "singleton", i.e. a subset $\{\omega\}$ containing a single outcome ω of Ω.

Corollary 26.6 The Sure Event Ω:
Is equal to the sample space as it contains all possible elementary events.

Corollary 26.7 The Impossible Event $\varnothing$:
The impossible event i.e. nothing is happening is denoted by the empty set.

Definition 26.4 The Family of All Events $\mathcal{A}/2^{\Omega}$:
The set of all subset of the sample space Ω called family of all events is given by the power set of the sample space $\mathcal{A} = 2^{\Omega}$ (for finite sample spaces).

Definition 26.5 Probability $\mathbb{P}(A)$:
Is a number associated with every A, that measures the likelihood of the event to be realized "a priori". The bigger the number the more likely the event will happen.
1. $0 \leqslant \mathbb{P}(A) \leqslant 1$
2. $\mathbb{P}(\Omega) = 1$
3. If $A \cap B = \varnothing$ then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

Note

We can think of the probability of an event A as the limit of the "frequency" of repeated experiments:

$$\mathbb{P}(A) = \lim_{n\to\infty} \frac{\delta_n(A)}{n} \qquad \text{where} \qquad \delta(A) = \begin{cases} 1 \text{ if } \omega \in A \\ 0 \text{ if } \omega \notin A \end{cases}$$

1. Sigma Algebras

Definition 26.6 [Proof 26.3]
Sigma Algebra σ:
A set $\mathcal{F}$ of subsets of Ω is called a σ-algebra on Ω if the following properties apply
- $\Omega \in \mathcal{F}$ and $\varnothing \in \mathcal{F}$
- If $A \in \mathcal{F}$ then $\Omega \backslash A = A^C \in \mathcal{F}$:
  The complementary subset of A is also in Ω.
- For all $A_i \in \mathcal{F} : \bigcup_{i=1} A_i \in \mathcal{F}$

Explanation 26.1 ([def. 26.6]). *The σ-algebra determines what events we can measure, it represents all of the possible events of the experiment that we can detect.*
Thus the sigma algebra is a mathematical construct that tells us how much information we obtain once we conduct some experiment.

Corollary 26.8 $\mathcal{F}_{\min}$: $\mathcal{F} = \{\varnothing, \Omega\}$ is the simplest σ-algebra, telling us only if an event happened $\omega \in \Omega$ happened or not but not which one.

Corollary 26.9 $\mathcal{F}_{\max}$: $\mathcal{F} = 2^{\Omega}$ consists of all subsets of Ω and thus corresponds to full information i.e. we know if and which event happened.

Definition 26.7 Measurable Space $\{\Omega, \mathcal{F}\}$:
Is the pair of a set and sigma algebra i.e. a sample space and sigma algebra $\{\Omega, \mathcal{F}\}$.

Corollary 26.10 $\mathcal{F}$-measurable Event $A_i \in \mathcal{F}$:
The measurable events A_i of $\mathcal{F}$ are called *$\mathcal{F}$-measurable* or *measurable sets*.

Definition 26.8 [Example 26.4]
Sigma Algebra generated by a subset of Ω $\sigma(\mathcal{C})$:
Let $\mathcal{C}$ be a class of subsets of Ω. The σ-algebra generated by $\mathcal{C}$, denoted by $\sigma(\mathcal{C})$, is the *smallest* sigma algebra $\mathcal{F}$ that included all elements of $\mathcal{C}$.

Definition 26.9 [Example 26.5]
Borel σ-algebra $\mathcal{B}(\mathbb{R})$:
The Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is the smallest σ-algebra containing all open intervals in $\mathbb{R}$. The sets in contained in $\mathcal{B}(\mathbb{R})$ are called Borel sets.
The extension to the multi-dimensional case, $\mathcal{B}(\mathbb{R}^n)$, is straightforward.
For all real numbers $a, b \in \mathbb{R}$, $\mathcal{B}(\mathbb{R})$ contains various sets.

Why do we need Borel Sets

So far we only looked at atomic events ω, with the help of sigma algebras we are now able to measure continuous events s.a. $[0, 1]$.

Definition 26.10 Borel Set:

Corollary 26.11 Generating Borel σ-Algebra[Proof 26.1]:
The Borel σ-algebra of $\mathbb{R}$ is generated by intervals of the form $(-\infty, a]$, where $a \in \mathbb{Q}$ ($\mathbb{Q}$ =rationals).

Definition 26.11 ($\mathbb{P}$)-trivial Sigma Algebra:
is a σ-algebra $\mathcal{F}$ for which each event has a probability of zero or one:

$$\mathbb{P}(A) \in \{0, 1\} \qquad \forall A \in \mathcal{F} \tag{26.1}$$

Interpretation

A trivial sigma algebra means that all events are almost surely constant and that there exist no non-trivial information.
An example of a trivial sigma algebra is $\mathcal{F}_{\min} = \{\Omega, \varnothing\}$.

2. Measures

Definition 26.12 Measure μ:
A measure defined on a measurable space $\{\Omega, \mathcal{F}\}$ is a function/map:

$$\mu : \mathcal{F} \mapsto [0, \infty] \tag{26.2}$$

for which holds:
- $\mu(\varnothing) = 0$
- countable additivity [def. 26.13]

Definition 26.13 Countable/σ-Additive Function:
Given a function μ defined on a σ-algebra $\mathcal{F}$.
The function μ is said to be countable additive if for every countable sequence of pairwise disjoint elements $(F_i)_{i\geqslant 1}$ of $\mathcal{F}$ it holds that:

$$\mu\left(\bigcup_{i=1}^{\infty} F_i\right) = \sum_{i=1}^{\infty} \mu(F_i) \quad \text{for all} \quad F_j \cap F_k = \varnothing \quad \forall j \neq k \tag{26.3}$$

Corollary 26.12 Additive Function: A function that satisfies countable additivity, is also additive, meaning that for every $F, G \in \mathcal{F}$ it holds:

$$F \cap G = \varnothing \quad \Longrightarrow \quad \mu(F \cup G) = \mu(F) + \mu(G) \tag{26.4}$$

Explanation 26.2. *If we take two events that cannot occur simultaneously, then the probability that at least one of the events occurs is just the sum of the measures (probabilities) of the original events.*

Definition 26.14 [Example 26.6]
Equivalent Measures $\mu \sim \nu$:
Let μ and ν be two measures defined on a measurable space[def. 26.7] $(\Omega, \mathcal{F})$. The two measures are said to be equivalent if it holds that:

$$\mu(A) > 0 \iff \nu(A) > 0 \qquad \forall A \subseteq \mathcal{F} \tag{26.5}$$

this is equivalent to μ and ν having equivalent null sets:

$$\mathcal{N}_\mu = \mathcal{N}_\nu \qquad \begin{aligned} \mathcal{N}_\mu &= \{A \in \mathcal{A} | \mu(A) = 0\} \\ \mathcal{N}_\nu &= \{A \in \mathcal{A} | \nu(A) = 0\} \end{aligned} \tag{26.6}$$

Definition 26.15 Measure Space $\{\mathcal{F}, \Omega, \mu\}$:
The triplet of sample space, sigma algebra and a measure is called a measure space.

2.1. Borel Measures

Definition 26.16 Borel Measure: A Borel Measure is any *measure*[def. 26.12] μ defined on the Borel σ-algebra[def. 26.9] $\mathcal{B}(\mathbb{R})$.

2.1.1. The Lebesgue Measure

Definition 26.17 Lebesgue Measure on $\mathcal{B}$ λ:
Is the Borel measure[def. 26.16] defined on the measurable space $\{\mathbb{R}, \mathcal{B}(\mathbb{R})\}$ which assigns for every half-open interval $(a, b]$ interval its length:

$$\lambda((a, b]) := b - a \tag{26.7}$$

Corollary 26.13 Lebesgue Measure of Atomitcs:

- The Lebesgue measure of a set containing only one point must be zero:
  $$\lambda(\{a\}) = 0 \tag{26.8}$$
- The Lebesgue measure of a set containing countably many points $A = \{a_1, a_2 \ldots, a_n\}$ must be zero:
  $$\lambda(A) + \sum_{i=1}^{n} \lambda(\{a_i\}) = 0 \tag{26.9}$$
- The Lebesgue measure of a set containing uncountably many points $A = \{a_1, a_2 \ldots, \}$ can be either zero, positive and finite or infinite.

3. Probability/Kolomogorov's Axioms 1931

One problem we are still having is the range of μ, by standardizing the measure we obtain a well defined measure of events.

Axiom 26.1 Non-negativity: The probability of an event is a non-negative real number:

$$\text{If } A \in \mathcal{F} \qquad \text{then} \qquad \mathbb{P}(A) \geqslant 0 \tag{26.10}$$

Axiom 26.2 Unitairity: The probability that at least one of the elementary events in the entire sample space Ω will occur is equal to one:

$$\text{The certain event} \qquad \mathbb{P}(\Omega) = 1 \tag{26.11}$$

Axiom 26.3 σ-additivity: If $A_1, A_2, A_3, \ldots \in \mathcal{F}$ are mutually disjoint, then:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \tag{26.12}$$

Corollary 26.14 : As a consequence of this it follows:

$$\mathbb{P}(\varnothing) = 0 \tag{26.13}$$

Corollary 26.15 Complementary Probability:

$$\mathbb{P}(A^C) = 1 - \mathbb{P}(A) \qquad \text{with} \qquad A^C = \Omega - A \tag{26.14}$$

Definition 26.18 Probability Measure $\mathbb{P}$:
a probability measure is function $\mathbb{P} : \mathcal{F} \mapsto [0, 1]$ defined on a σ-algebra $\mathcal{F}$ of a sample space Ω that satisfies the probability axioms.

4. Conditional Probability

Definition 26.19 Conditional Probability: Let A, B be events, with $\mathbb{P}(B) \neq 0$. Then the conditional probability of the event A given B is defined as:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \qquad \mathbb{P}(B) \neq 0 \tag{26.15}$$

5. Independent Events

Theorem 26.1
Independent Events: Let A, B be two events. A and B are said to be independent iffy:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \qquad \begin{aligned} \mathbb{P}(A|B) &= \mathbb{P}(A), & \mathbb{P}(B) > 0 \\ \mathbb{P}(B|A) &= \mathbb{P}(B), & \mathbb{P}(A) > 0 \end{aligned} \tag{26.16}$$

Note

The requirement of no impossible events follows from [def. 26.19]

Corollary 26.16 Pairwise Independent Evenest:
A finite set of events $\{A_i\}_{i=1}^{n} \in \mathcal{A}$ is *pairwise independent* if every pair of events is independent:

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cap \mathbb{P}(A_j) \qquad i \neq j, \quad \forall i, j \in \mathcal{A} \tag{26.17}$$

Corollary 26.17 Mutal Independent Evenest:
A finite set of events $\{A_i\}_{i=1}^{n} \in \mathcal{A}$ is *mutal independent* if every event A_j is independent of any intersection of the other events:

$$\mathbb{P}\left(\bigcap_{i=i}^{k} B_i\right) = \prod_{i=1}^{k} \mathbb{P}(B_i) \qquad \begin{aligned} &\forall \{B_i\}_{i=1}^{k} \subseteq \{A_i\}_{i=1}^{n} \\ &k \leqslant n, \qquad \{A_i\}_{i=1}^{n} \in \mathcal{A} \end{aligned} \tag{26.18}$$

6. Product Rule

Law 26.1 Product Rule: Let A, B be two events then the probability of both events occurring simultaneously is given by:

$$\mathbb{P}(A \cap B) = \mathbb{P}(B|A)\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) \tag{26.19}$$

Law 26.2
Generalized Product Rule/Chain Rule: is the generalization of the product rule?? to n events $\{A_i\}_{i=1}^{n}$

$$\mathbb{P}\left(\bigcap_{i=i}^{k} E_i\right) = \prod_{k=1}^{n} \mathbb{P}\left(E_k \middle| \bigcap_{i=i}^{k-1} E_i\right) = \qquad (26.20)$$

$$= \mathbb{P}(E_n|E_{n-1} \cap \ldots \cap E_1) \cdot \mathbb{P}(E_{n-1}|E_{n-2} \cap \ldots \cap E_1) \cdots$$
$$\cdots \mathbb{P}(E_3|E_2 \cap E_1)\mathbb{P}(E_2|E_1)\mathbb{P}(E_1)$$

7. Law of Total Probability

Definition 26.20 Complete Event Field: A complete event field $\{A_i : i \in I \subseteq \mathbb{N}\}$ is a countable or finite partition of Ω that is the partitions $\{A_i : i \in I \subseteq \mathbb{N}\}$ are a *disjoint union* of the sample space:

$$\bigcup_{i\in I} A_i = \Omega \qquad A_i \cap A_j = \varnothing \qquad i \neq j, \forall i, j \in I \qquad (26.21)$$

Theorem 26.2
Law of Total Probability/Partition Equation:
Let $\{A_i : i \in I\}$ be a complete event field[def. 26.20] then it holds for $B \in \mathcal{B}$:

$$\mathbb{P}(B) = \sum_{i\in I} \mathbb{P}(B|A_i)\mathbb{P}(A_i) \qquad (26.22)$$

8. Bayes Theorem

Law 26.3 Bayes Rule: Let A, B be two events s.t. $\mathbb{P}(B) > 0$ then it holds:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \qquad \mathbb{P}(B) > 0 \qquad (26.23)$$

follows directly fromeq. (26.19).

Theorem 26.3 Bayes Theorem: Let $\{A_i : i \in I\}$ be a complete event field[def. 26.20] and $B \in \mathcal{B}$ a random event s.t. $\mathbb{P}(B) > 0$, then it holds:

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\sum_{i\in I} \mathbb{P}(B|A_i)\mathbb{P}(A_i)} \qquad (26.24)$$

proof ?? 26.2

Distributions on $\mathbb{R}$

9.1. Distribution Function

Definition 26.21 Distribution Function of $\mathbb{P}$ F**:**
The *distribution function* F induced by a a probability measure $\mathbb{P}$ on $(\mathbb{R}, \mathcal{B})$ is the function:

$$F(x) = \mathbb{P}((\infty, x]) \qquad (26.25)$$

Theorem 26.4 : A function F is the distribution function of a (unique) probability on $(\mathbb{R}, \mathcal{B})$ iff:

- F is non-decreasing
- F is right continuous
- $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to+\infty} F(x) = 1$

Corollary 26.18 : A probability $\mathbb{P}$ is uniquely determined by a distribution function F
That is if there exist another probability $\mathbb{Q}$ s.t.

$$G(x) = \mathbb{Q}((-\infty, x])$$

and if $F = G$ then it follows $\mathbb{P} = \mathbb{Q}$.

9.2. Random Variables

A random variable X is a function/map that determines a quantity of interest based on the outcome $\omega \in \Omega$ of a random experiment. Thus X is not really a variable in the classical sense but a variable with respect to the outcome of an experiment. Its value is determined in two steps:

1. The outcome of an experiment is a random quantity $\omega \in \Omega$
2. The outcome ω determines (possibly various) quantities of interests $\iff$ *random variables*

Thus a random variable X, defined on a probability space $\{\Omega, \mathcal{F}, \mathbb{P}\}$ is a mapping from Ω into another space $\mathcal{E}$, usually $\mathcal{E} = \mathbb{R}$ or $\mathcal{E} = \mathbb{R}^n$:

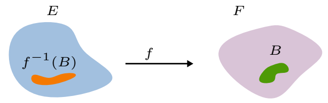
$$X : \Omega \mapsto \mathcal{E} \qquad \omega \mapsto X(\omega)$$

Let now $E \in \mathcal{E}$ be a quantity of interest, in order to quantify its probability we need to map it back to the original sample space Ω:

$$\underbrace{\mathbb{P}_X(E)}_{\text{Probability for an event in } E} = \mathbb{P}(\{\omega : X(\omega) \in E\}) = \mathbb{P}(X \in E) = \overbrace{\mathbb{P}\left(X^{-1}(E)\right)}^{\text{Probability for an event in } \Omega}$$

Definition 26.22 $\mathcal{E}$-measurable function: Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be two measurable spaces. A function $f : E \mapsto F$ is called measurable (relative to $\mathcal{E}$ and $\mathcal{F}$) if

$$\forall B \in \mathcal{F} : \qquad f^{-1}(B) = \{\omega \in \mathcal{E} : f(\omega) \in B\} \in \mathcal{E} \qquad (26.26)$$



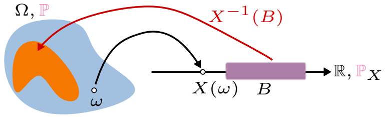
Interpretation

The pre-image[def. 15.11] of B under f i.e. $f^{-1}(B)$ maps all values of the target space F back to the sample space $\mathcal{E}$ (for all possible $B \in \mathcal{F}$).

Definition 26.23 Random Variable: A real-valued random variable (vector) X, defined on a probability space $\{\Omega, \mathcal{E}, \mathbb{P}\}$ is an $\mathcal{E}$-measurable function mapping, if it maps its sample space Ω into a target space $(F, \mathcal{F})$:

$$X : \Omega \mapsto \mathcal{F} \quad (\mathcal{F}^n) \qquad (26.27)$$

Since X is $\mathcal{E}$-measurable it holds that $X^{-1} : \mathcal{F} \mapsto \mathcal{E}$



Corollary 26.19 : Usually $F = \mathbb{R}$, which usually amounts to using the Borel σ-algebra $\mathcal{B}$ of $\mathbb{R}$.

Corollary 26.20 Random Variables of Borel Sets: Given that we work with Borel σ-algebras then the definition of a random variable is equivalent to (due to [cor. 26.11]):

$$\begin{aligned} X^{-1}(B) &= X^{-1}((-\infty, a]) \\ &= \{\omega \in \Omega : X(\omega) \leqslant a\} \in \mathcal{E} \quad \forall a \in \mathbb{R} \end{aligned} \qquad (26.28)$$

Definition 26.24
Realization of a Random Variable $x = X(\omega)$**:** Is the value of a random variable that is actually observed after an experiment has been conducted. In order to avoid confusion lower case letters are used to indicate actual observations/realization of a random variable.

Corollary 26.21 Indicator Functions $I_A(\omega)$**:**
An important class of measurable functions that can be used as r.v. are indicator functions:

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases} \qquad (26.29)$$

We know that a probability measure $\mathbb{P}$ on $\mathbb{R}$ is characterized by the quantities $\mathbb{P}((-\infty, a])$. Thus the quantities.

Corollary 26.22 : Let $(F, \mathcal{F}) = (\mathbb{R}, \mathcal{B})$ and let $(E, \mathcal{E})$ be an arbitrary measurable space. Let X be a real value function on E.
Then it holds that X is measurable *if and only if*

$$\{X \leqslant a\} = \{\omega : X(\omega) \leqslant a\} = X^{-1}((-\infty, a]) \in \mathcal{E}, \forall a \in \mathbb{R}$$

or $$\{X < a\} \in \mathcal{E}.$$

Explanation 26.3 ([cor. 26.22]). *A random variable is a function that is measurable if and only if its distribution function is defined.*

9.3. The Law of Random Variables

Definition 26.25 Law/Distribution of X $\mathscr{L}(X)$**:**
Let X be a r.v. on $\{\Omega, \mathcal{F}, \mathbb{P}\}$, with values in $(E, \mathcal{E})$, then the *distribution/law* of X is defined as:

$$\mathbb{P} : \mathcal{B} \mapsto [0, 1] \qquad (26.30)$$
$$\mathbb{P}^X(B) = \mathbb{P}\{X \in B\} = \mathbb{P}(\omega : X(\omega) \in B) \qquad \forall b \in \mathcal{E}$$

Note

- Sometimes $\mathbb{P}^X$ is also called the *image* of $\mathbb{P}$ by X
- The law can also be written as:
$$\mathbb{P}^X(B) = \mathbb{P}(X^{-1}(B)) = (\mathbb{P} \circ X^{-1})(B)$$

Theorem 26.5 : The law/distribution of X is a probability measure $\mathbb{P}$ on $(E, \mathcal{E})$.

Definition 26.26
(Cumulative) Distribution Function F_X**:**
Given a real-valued r.v. then its *cumulative distribution function* is defined as:

$$F_X(x) = \mathbb{P}^X((-\infty, x]) = \mathbb{P}(X \leqslant x) \qquad (26.31)$$

Corollary 26.23 : The distribution of $\mathbb{P}^X$ of a real valued r.v. is entirely characterized by its cumulative distribution function F_X[def. 26.33].

Property 26.1:

$$\mathbb{P}(X > x) = 1 - F_X(x) \qquad (26.32)$$

Property 26.2: Probability of $X \in [a, b]$

$$\mathbb{P}(a < X \leqslant B) = F_X(b) - F_X(a) \qquad (26.33)$$

9.4. Probability Density Function

Definition 26.27 Continuous Random Variable: Is a r.v. for which a probability density function f_X exists.

Definition 26.28 Probability Density Function: Let X be a r.v. with associated cdf F_X. If F_X is continuously integrable for all $x \in \mathbb{R}$ then X has a *probability density* f_X defined by:

$$F_X(x) = \int_{-\infty}^{x} f_X(y)\,\mathrm{d}y \qquad (26.34)$$

or alternatively:

$$f_X(x) = \lim_{\epsilon\to0} \frac{\mathbb{P}(x \leqslant X \leqslant x + \epsilon)}{\epsilon} \qquad (26.35)$$

Corollary 26.24 $\mathbb{P}(X = b) = 0, \qquad \forall b \in \mathbb{R}$**:**

$$\mathbb{P}(X = b) = \lim_{a\to b} \mathbb{P}(a < X \leqslant b) = \lim_{a\to b} \int_a^b f(x) = 0 \qquad (26.36)$$

Corollary 26.25 : From [cor. 26.24] it follows that the exact borders are not necessary:

$$\begin{aligned} \mathbb{P}(a < X < b) &= \mathbb{P}(a \leqslant X < b) \\ &= \mathbb{P}(a < X \leqslant b) = \mathbb{P}(a \leqslant X <\leqslant b) \end{aligned}$$

Corollary 26.26 :

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1 \qquad (26.37)$$

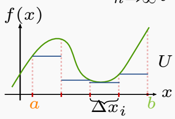
Notes

- Often the cumulative distribution function is referred to as “cdf” or simply *distribution function.*
- Often the probability density function is referred to as “pdf” or simply *density.*

9.5. Lebesgue Integration

Problems of Riemann Integration

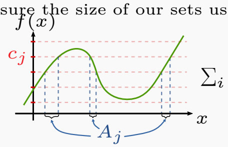
- Difficult to extend to higher dimensions – general domains of definitions $f : \Omega \mapsto \mathbb{R}$
- Depends on continuity
- Integration of limit processes require strong uniform convergence in order to integrate limit processes
$$\lim_{n\to\infty} \int f(x)\,\mathrm{d}x \overset{\text{str. u.c.}}{=} \int \lim_{n\to\infty} f(x)\,\mathrm{d}x$$



$$U(p) = \sum_i \sup(f(x_i)) \cdot \Delta x_i \xrightarrow{n\to\infty} \int f\,\mathrm{d}x$$

Idea

Partition domain by function values of equal size i.e. values that lie within the same sets/have the same value A_j build up the partitions w.r.t. to the variable x.
Problem: we do not know how big those sets/partitions on the x-axis will be.
Solution: we can use the measure μ of our measure space $\{\Omega, \mathcal{A}, \mu\}$ in order to obtain the size of our sets $A_j \Rightarrow$ we do not have to care anymore about discontinuities, as we can measure the size of our sets using our measure.



$$\sum_i c_i \cdot \mu(A_i) \xrightarrow{n\to\infty} \int f\,\mathrm{d}\mu$$

Definition 26.29 Lebesgue Integral:

$$\lim_{n\to\infty} \sum_{i=1}^{n} c_i\mu(A_i) = \int_\Omega f\,\mathrm{d}\mu \qquad \begin{matrix} f(x) \approx c_i \\ \forall x \in A_i \end{matrix} \qquad (26.38)$$

Definition 26.30
Simple Functions (Random Variables): A r.v. X is called simple if it takes on only a finite number of values and hence can be written in the form:

$$X = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i} \qquad a_i \in \mathbb{R} \qquad \mathcal{A} \ni A_i = \begin{cases} 1 & \text{if } \{X = a_i\} \\ 0 & \text{else} \end{cases} \qquad (26.39)$$

9.6. Independent Random Variables

We have seen that two events A and B are independent if knowledge that B has occurred does not change the probability that A will occur theorem 26.1.
For two random variables X, Y we want to know if knowledge of Y leaves the probability of X, to take on certain values unchanged.

Definition 26.31 Independent Random Variables:
Two real valued random variables X and Y are said to be independent iff:

$$\mathbb{P}(X \leqslant x|Y \leqslant y) = \mathbb{P}(X \leqslant x) \qquad \forall x, y \in \mathbb{R} \qquad (26.40)$$

which amounts to:

$$\begin{aligned} F_{X,Y}(x, y) &= \mathbb{P}(\{X \leqslant x\} \cap \{Y \leqslant y\}) = \mathbb{P}(X \leqslant x, Y \leqslant y) \\ &= F_X(x)F_Y(y) \quad \forall x, y \in \mathbb{R} \end{aligned} \qquad (26.41)$$

or alternatively iff:

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \qquad \forall A, B \in \mathcal{B} \quad (26.42)$$

Note

If the joint distribution $F_{X,Y}(x,y)$ can be factorized into two functions of x and y then X and Y are independent.

Definition 26.32
Independent Identically Distributed:

10. Product Rule

Law 26.4 Product Rule: Let X, Y be two random variables then their jo

Law 26.5
Generalized Product Rule/Chain Rule:

11. Change Of Variables Formula

Formula 26.1
(Scalar Discret) Change of Variables: Let X be a discret rv $X \in \mathcal{X}$ with pmf p_X and define $Y \in \mathcal{Y}$ as $Y = g(x)$ s.t. $\mathcal{Y} = \{y|y = g(x), \forall x \in \mathcal{X}\}$. **Where** g is an arbitrary strictly monotonic ([def. 15.14]) function.
Let: $\mathcal{X}_y = x_i$ be the set of all $x_i \in \mathcal{X}$ s.t. $y = g(x_i)$.
Then the pmf of Y is given by:

$$p_Y(y) = \sum_{x_i \in \mathcal{X}_y} p_X(x_i) = \sum_{x \in \mathcal{Y}:g(x)=y} p_X(x) \tag{26.43}$$

see proof **??** 26.3

Formula 26.2
(Scalar Continuous) Change of Variables:
Let $X \sim f_X$ be a continuous r.v. and let g be an arbitrary strictly monotonic[def. 15.14] function.
Define a new r.v. Y as

$$\mathcal{Y} = \{y|y = g(x), \forall x \in \mathcal{X}\} \tag{26.44}$$

then the pdf of Y is given by:

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = f_X(x)\left|\frac{d}{dy}\left(g^{-1}(y)\right)\right| \tag{26.45}$$

$$= f_X(x)\frac{1}{\left|\frac{dy}{dx}\right|} = \frac{f_X(g^{-1}(y))}{\left|\frac{dg}{dx}(g^{-1}(y))\right|} \tag{26.46}$$

Formula 26.3
(Continuous) Change of Variables:
Let $X = \{X_1, \dots, X_n\} \sim f_X$ be a continuous random vector and let g be an arbitrary strictly monotonic[def. 15.14] function

$$g : \mathbb{R}^n \mapsto \mathbb{R}^m$$

Define a new r.v. Y as

$$\mathcal{Y} = \{\mathbf{y}|\mathbf{y} = g(\mathbf{x}), \forall \mathbf{x} \in \mathcal{X}\} \tag{26.47}$$

and let $h(\mathbf{x}) := g(\mathbf{x})^{-1}$ then the pdf of Y is given by:

$$\begin{aligned} f_Y(\mathbf{y}) &= f_X(x_1, \dots, x_n) \cdot |J| \\ &= f_X(h_1(\mathbf{y}), \dots, h_n(\mathbf{y})) \cdot |J| \\ &= f_X(\mathbf{y})|\det D_{\mathbf{x}} h(\mathbf{x})|\Big|_{\mathbf{x}=\mathbf{y}} \\ &= f_X(g^{-1}(\mathbf{y}))\left|\det\left(\frac{\partial g}{\partial \mathbf{x}}\right)\right|^{-1} \end{aligned} \tag{26.48}$$

where $J = \det Dh$ is the Jaccobian[def. 16.6].
See also proof **??** 26.6 and example 26.8

Note

A monotonic function is required in order to satisfy inevitability.

Probability Distributions on $\mathbb{R}^n$

13. Joint Distribution

Definition 26.33
Joint (Cumulative) Distribution Function $F_{\mathbf{X}}$**:**
Let $\mathbf{X} = (X_1 \cdots\cdots X_n)$ be a random vector in $\mathbb{R}^n$, then its *cumulative distribution function* is defined as:

$$\begin{aligned} F_{\mathbf{X}}(\mathbf{x}) &= \mathbb{P}^X((-\infty, \mathbf{x}]) = \mathbb{P}(\mathbf{X} \leqslant \mathbf{x}) \\ &= \mathbb{P}(X_1 \leqslant x_1, \dots X_n \leqslant x_n) \end{aligned} \tag{26.49}$$

Definition 26.34 Joint Probability Distribution:
Let $\mathbf{X} = (X_1 \cdots\cdots X_n)$ be a random vector in $\mathbb{R}^n$ with associated cdf $F_{\mathbf{X}}$. If $F_{\mathbf{X}}$ is continuously integrable for all $\mathbf{x} \in \mathbb{R}$ then $\mathbf{X}$ has a *probability density* f_X defined by:

$$F_X(x) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f_{\mathbf{X}}(y_1, \dots, y_n)\, dy_1\, dy_n \tag{26.50}$$

or alternatively:

$$f_{\mathbf{X}}(\mathbf{x}) = \lim_{\epsilon \to 0} \frac{\mathbb{P}(x_1 \leqslant X_1 \leqslant x_1 + \epsilon, \dots, x_n \leqslant X_n \leqslant x_n + \epsilon)}{\epsilon} \tag{26.51}$$

13.1. Marginal Distribution

Definition 26.35 Marginal Distribution:

14. The Expectation

Definition 26.36 Expectation:

$$\mathbb{E}[X] = \int_\Omega X(\omega)\mathbb{P}(d\omega) = \int_\Omega X\, d\mathbb{P} \tag{26.52}$$

Corollary 26.27 Expectation of simple r.v.:
If X is a simple[def. 26.30] r.v. its *expectation* is given by:

$$\mathbb{E}[X] = \sum_{i=1}^{n} a_i \mathbb{P}(A_i) \tag{26.53}$$

14.1. Properties

14.1.1. Linear Operators

14.1.2. Quadratic Form

Definition 26.37 proof 26.7
Expectation of a Quadratic Form:
Let $\boldsymbol{\epsilon} \in \mathbb{R}^n$ be a random vector with $\mathbb{E}[\boldsymbol{\epsilon}] = \boldsymbol{\mu}$ and $\mathbb{V}[\boldsymbol{\epsilon}] = \Sigma$:

$$\mathbb{E}\left[\boldsymbol{\epsilon}^\intercal \mathbf{A} \boldsymbol{\epsilon}\right] = \mathrm{tr}(\mathbf{A}\Sigma) + \boldsymbol{\mu}^\intercal \mathbf{A} \boldsymbol{\mu} \tag{26.54}$$

14.2. The Jensen Inequality

Theorem 26.6 Jensen Inequality: Let X be a random variable and g some function, then it holds:

$$\begin{matrix} g(\mathbb{E}[X]) \leqslant \mathbb{E}[g(X)] & \multirow{2}{*}{if} & g \text{ is convex}^{[\text{def. } 15.24]} \\ g(\mathbb{E}[X]) \geqslant \mathbb{E}[g(X)] & & g \text{ is concave}^{[\text{def. } 15.25]} \end{matrix} \tag{26.55}$$

14.3. Law of the Unconscious Statistician

Law 26.6 Law of the Unconscious Statistician:
Let $X \in \mathcal{X}$,$Y \in \mathcal{Y}$ be random variables where Y is defined as:

$$\mathcal{Y} = \{y|y = g(x), \forall x \in \mathcal{X}\}$$

then the expectation of Y can be calculated in terms of X:

$$\mathbb{E}_Y[y] = \mathbb{E}_X[g(x)] \tag{26.56}$$

Consequence

Hence if we p_X we do not have to first calculate p_Y in order to calculate $\mathbb{E}_Y[y]$.

14.4. Properties

14.5. Law of Iterated Expectation (LIE)

Law 26.7 [proof 26.8]
Law of Iterated Expectation (LIE):

$$\mathbb{E}[X] = \mathbb{E}_Y \mathbb{E}[X|Y] \tag{26.57}$$

14.6. Hoeffdings Bound

Definition 26.38 Hoeffdings Bound:
Let $\mathbf{X} = \{X_i\}_{i=1}^n$ be i.i.d. random variables strictly bounded by the interval $[a, b]$ then it holds:

$$\mathbb{P}(|\mu_{\mathbf{X}} - \mathbb{E}[X]| \geqslant \epsilon) \leqslant 2\exp\left(\frac{-2n^2\epsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \overset{[0,1]}{=} 2e^{-2n\epsilon^2} \tag{26.58}$$

Explanation 26.4. *The difference of the expectation from the empirical average to be bigger than ϵ is upper bound in probability.*

15. Moment Generating Function (MGF)

Definition 26.39 Moment of Random Variable: The i-th moment of a random variable X is defined as (if it exists):

$$m_i := \mathbb{E}\left[X^i\right] \tag{26.59}$$

Definition 26.40 ψ_X
Moment Generating Function (MGF):

$$\psi_X(t) = \mathbb{E}\left[e^{tX}\right] \qquad t \in \mathbb{R} \tag{26.60}$$

Corollary 26.28 Sum of MGF: The moment generating function of a sum of n independent variables $(X_j)_{1 \leqslant j \leqslant n}$ is the product of the moment generating functions of the components:

$$\psi_{S_n}(t) = \psi_{X_1}(t) \cdots \psi_{X_n}(t) \qquad S_n := X_1 + \dots X_n \tag{26.61}$$

Corollary 26.29 : The i-th moment of a random variable is the i-th derivative of its associated moment generating function evaluated zero:

$$\mathbb{E}\left[X^i\right] = \psi_X^{(i)}(0) \tag{26.62}$$

16. The Characteristic Function

Transforming probability distributions using the Fourier transform is a handy tool in probability in order to obtain properties or solve problems in another space before transforming them back.

Definition 26.41 $\hat{\mu}$
Fourier Transformed Probability Measure:

$$\hat{\mu} = \int e^{i\langle u,x\rangle}\mu(dx) \tag{26.63}$$

Corollary 26.30 : As $e^{i\langle u,x\rangle}$ can be rewritten using formulaeqs. (11.9) and (11.10) it follows:

$$\hat{\mu} = \int \cos(\langle u, x\rangle)\,\mu(dx) + i\int \sin(\langle u, x\rangle)\,\mu(dx) \tag{26.64}$$

where $x \mapsto \cos(\langle x, u\rangle)$ and $x \mapsto \sin(\langle x, u\rangle)$ are both bounded and Borel i.e. Lebesgue integrable.

Definition 26.42 Characteristic Function φ_X**:** Let $\mathbf{X}$ be an $\mathbb{R}^n$-valued random variable. Its characteristic function $\varphi_{\mathbf{X}}$ is defined on $\mathbb{R}^n$ as:

$$\varphi_{\mathbf{X}}(u) = \int e^{i\langle \mathbf{u},\mathbf{x}\rangle}\mathbb{P}^X(d\mathbf{x}) = \widehat{\mathbb{P}^X}(\mathbf{u}) \tag{26.65}$$

$$= \mathbb{E}\left[e^{i\langle \mathbf{u},\mathbf{x}\rangle}\right] \tag{26.66}$$

Corollary 26.31 : The characteristic function φ_X of a distribution always exists as it is equal to the Fourier transform of the probability measure, which always exists.

Note

This is an advantage over the moment generating function.

Theorem 26.7 : Let μ be a probability measure on $\mathbb{R}^n$. Then $\hat{\mu}$ is a bounded continuous function with $\hat{\mu}(0) = 1$.

Theorem 26.8 Uniqueness Theorem: The Fourier Transform $\hat{\mu}$ of a probability measure μ on $\mathbb{R}^n$ *characterizes* μ. That is, if two probability measures on $\mathbb{R}^n$ admit the same Fourier transform, they are equal.

Corollary 26.32 : Let $\mathbf{X} = (X_1, \dots, X_n)$ be an $\mathbb{R}^n$-valued random variable. Then the real valued r.v.'s $(X_j)_{1 \leqslant j \leqslant n}$ are independent if and only if:

$$\varphi_X(u_1, \dots, u_n) = \prod_{j=1}^{n} \varphi_{X_j}(u_j) \tag{26.67}$$

Proofs

Proof 26.1: [cor. 26.11]: Let $\mathcal{C}$ denote all open intervals.
Since every open set in $\mathbb{R}$ is the countable union of open intervals[def. 11.12], it holds that $\sigma(\mathcal{C})$ is the Borel σ-algebra of $\mathbb{R}$.
Let $\mathcal{D}$ denote all intervals of the form $(-\infty, a]$, $a \in \mathbb{Q}$.
Let $a, b \in \mathcal{C}$, and let

- $(a_n)_{n>1}$ be a sequence of rationals *decreasing* to a and
- $(b_n)_{n>1}$ be a sequence of rationals *increasing strictly* to b

$$(a, b) = \cup_{n=1}^{\infty}(a_n, b_n] = \cup_{n=1}^{\infty}\left((-\infty, b_n] \cap (-\infty, a_n]^C\right)$$

Thus $\mathcal{C} \subset \sigma(\mathcal{D})$, whence $\sigma(\mathcal{C}) \subset \sigma(\mathcal{D})$ **but** as each element of $\mathcal{D}$ is a closed subset, $\sigma(\mathcal{D})$ must also be contained in the Borel sets $\mathcal{B}$ with

$$\mathcal{B} = \sigma(\mathcal{C}) \subset \sigma((D) \subset \mathcal{B}$$

Proof 26.2: theorem 26.3 Plug eq. (26.22) into the denominator and eq. (16.2) into the nominator and then use [def. 26.19]:

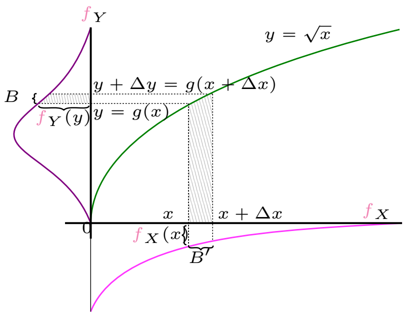
$$\frac{\mathbb{P}(B|A_j)\mathbb{P}(A_j)}{\sum_{i \in I}\mathbb{P}(B|A_i)\mathbb{P}(A_i)} = \frac{\mathbb{P}(B \cap A_j)}{\mathbb{P}(B)} = \mathbb{P}(A_j|B)$$

Proof 26.3: **??**:

$$Y = g(X) \iff \mathbb{P}(Y = y) = \mathbb{P}(x \in \mathcal{X}_y) = p_Y(y)$$

Proof 26.4: **??** (non-formal): The probability contained in a differential area must be invariant under a change of variables that is:

$$|f_Y(y)\, dy| = |f_x(x)\, dx|$$



Proof 26.5: **??** from CDF:

$$\mathbb{P}(Y \leqslant y) = \mathbb{P}(g(X) \leqslant y) = \begin{cases} \mathbb{P}(X \leqslant g^{-1}(y)) & \text{if } g \text{ is increas.} \\ \mathbb{P}(X \geqslant g^{-1}(y)) & \text{if } g \text{ is decreas.} \end{cases}$$

If g is monotonically increasing:

$$F_Y(y) = F_X(g^{-1}(y))$$

$$f_Y(y) = \frac{d}{dy}F_X(g^{-1}(y)) = f_X(x) \cdot \frac{d}{dy}g^{-1}(y)$$

If g is monotonically decreasing:

$$F_Y(y) = 1 - F_X(g^{-1}(y))$$

$$f_Y(y) = \frac{d}{dy}F_X(g^{-1}(y)) = -f_X(x) \cdot \frac{d}{dy}g^{-1}(y)$$

Proof 26.6: **??**: **Let** $B = [x, x + \Delta x]$ and $B' = [y, y + \Delta y] = [g(x), g(x + \Delta x)]$ we know that the probability of equal events is equal:

$$y = g(x) \qquad \Rightarrow \qquad \mathbb{P}(y) = \mathbb{P}(g(x)) \text{ (for disc. rv.)}$$

Now lets consider the probability for the continuous r.v.s:

$$\mathbb{P}(X \in B) = \int_x^{x+\Delta x} f_X(t)\,\mathrm{d}t \xrightarrow{\Delta x \to 0} |\Delta x \cdot f_x(x)|$$

For y we use Taylor (**??**)

$$\begin{aligned} g(x + \Delta x) &\overset{\text{eq. (15.56)}}{=} g(x) + \frac{\mathrm{d}g}{\mathrm{d}x}\Delta y && \text{for } \Delta x \to 0 \\ &= y + \Delta y && \text{with } \Delta y := \frac{\mathrm{d}g}{\mathrm{d}x} \cdot \Delta x \end{aligned} \tag{26.68}$$

Thus for $\mathbb{P}(Y \in B')$ it follows:

$$\begin{aligned} \mathbb{P}(X \in B') &= \int_y^{y+\Delta y} f_Y(t)\,\mathrm{d}t \xrightarrow{\Delta y \to 0} |\Delta y \cdot f_Y(y)| \\ &= \left|\frac{\mathrm{d}g}{\mathrm{d}x}(x)\Delta x \cdot f_Y(y)\right| \end{aligned}$$

Now we simply need to related the surface of the two pdfs:

$$B = [x, x + \Delta x] \overset{\text{same surfaces}}{\propto} [y, y + \Delta y] = B'$$

$$\mathbb{P}(Y \in B) = \mathbb{P}(X \in B')$$

$$\overset{\Delta y \to 0}{\iff} |f_Y(y) \cdot \Delta y| = \left|f_Y(y) \cdot \frac{\mathrm{d}g}{\mathrm{d}x}(x)\Delta x\right| = |f_X(x) \cdot \Delta x|$$

$$f_Y(y)\left|\cdot \frac{\mathrm{d}g}{\mathrm{d}x}(x)\right||\Delta x| = f_X(x) \cdot |\Delta x|$$

$$\Rightarrow f_Y(y) = \frac{f_X(x)}{\left|\frac{\mathrm{d}g}{\mathrm{d}x}(x)\right|} = \frac{f_X(g^{-1}(y))}{\left|\frac{\mathrm{d}g}{\mathrm{d}x}g^{-1}(y)\right|}$$

Proof 26.7: [def. 26.37]

$$\begin{aligned} \mathbb{E}\left[\boldsymbol{\epsilon}^\intercal \mathbf{A} \boldsymbol{\epsilon}\right] &\overset{\text{eq. (20.52)}}{=} \mathbb{E}\left[\mathtt{tr}(\boldsymbol{\epsilon}^\intercal \mathbf{A} \boldsymbol{\epsilon})\right] \\ &\overset{\text{eq. (20.54)}}{=} \mathbb{E}\left[\mathtt{tr}(\mathbf{A} \boldsymbol{\epsilon}\boldsymbol{\epsilon}^\intercal)\right] \\ &= \mathtt{tr}\left(\mathbb{E}\left[\mathbf{A}\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\intercal\right]\right) \\ &= \mathtt{tr}\left(\mathbf{A}\mathbb{E}\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^\intercal\right]\right) \\ &= \mathtt{tr}\left(\mathbf{A}\left(\Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^\intercal\right)\right) \\ &= \mathtt{tr}\left(\mathbf{A}\Sigma\right) + \mathtt{tr}\left(\mathbf{A}\boldsymbol{\mu}\boldsymbol{\mu}^\intercal\right) \\ &\overset{\text{eq. (20.52)}}{=} \mathtt{tr}\left(\mathbf{A}\Sigma\right) + \mathbf{A}\boldsymbol{\mu}\boldsymbol{\mu}^\intercal \end{aligned}$$

Proof 26.8: law 26.7

$$\begin{aligned} \mathbb{E}[X] &= \sum_x x \cdot \mathrm{p}_X(x) = \sum_x x \cdot \sum_y \mathrm{p}_{X,Y}(x, y) \\ &= \sum_x x \cdot \sum_y \mathrm{p}_{X|Y}(x|y) \cdot \mathrm{p}_Y(y) \\ &= \sum_y \mathrm{p}_Y(y) \cdot \sum_x x \cdot \mathrm{p}_{X|Y}(x|y) \\ &= \sum_y \mathrm{p}_Y(y) \cdot \mathbb{E}[X|Y] = \mathbb{E}_Y\left[\mathbb{E}[X|Y]\right] \end{aligned}$$

Examples

Example 26.1 :

- Toss of a coin (with head and tail): $\Omega = \{H, T\}$.
- Two tosses of a coin: $\Omega = \{HH, HT, TH, TT\}$
- A cubic die: $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$
- The positive integers: $\Omega = \{1, 2, 3, \ldots\}$
- The reals: $\Omega = \{\omega | \omega \in \mathbb{R}\}$

Example 26.2 :

- Head in coin toss $A = \{H\}$
- Odd number in die roll: $A = \{\omega_1, \omega_3, \omega_5, \}$
- The integers smaller five: $A = \{1, 2, 3, 4\}$

Example 26.3 : If the sample space is a die toss $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$, the sample space may be that we are only told whether an even or odd number has been rolled:

$$\mathcal{F} = \{\varnothing, \{\omega_1, \omega_3, \omega_5\}, \{\omega_2, \omega_4, \omega_6\}\}$$

Example 26.4 : If we are only interested in the subset $A \in \Omega$ of our experiment, then we can look at the corresponding generating σ-algebra $\sigma(A) = \left\{\varnothing, A, A^{\mathrm{C}}, \Omega\right\}$.

Example 26.5 :

- open half-lines: $(-\infty, a)$ and (a, ∞),
- union of open half-lines: $(a, b) = (-\infty, a) \cup (b, \infty)$,
- closed interval: $[a, b] = \overline{(-\infty, \cup a) \cup (b, \infty)}$,
- closed half-lines: $(-\infty, a] = \bigcup_{n=1}^{\infty}[a - n, a]$ and $[a, \infty) = \bigcup_{n=1}^{\infty}[a, a + n]$,
- half-open and half-closed $(a, b] = (-\infty, b] \cup (a, \infty)$,
- every set containing only one real number: $\{a\} = \bigcap_{n=1}^{\infty}(a - \frac{1}{n}, a + \frac{1}{n})$,
- every set containing finitely many real numbers: $\{a_1, \ldots, a_n\} = \bigcup_{k=1}^{n} a_k$.

Example 26.6 Equivalent (Probability) Measures:

$$\Omega = \{1, 2, 3\} \qquad \begin{aligned} \mathbb{P}(\{1, 2, 3\}) &= \{2/3, 1/6, 1/6\} \\ \tilde{\mathbb{P}}(\{1, 2, 3\}) &= \{1/3, 1/3, 1/3\} \end{aligned}$$

Example 26.7 :

Example 26.8 ??: Let $X, Y \overset{\text{ind.}}{\sim} \mathcal{N}(0, 1)$.
Question: proof that:

$$U = X + Y \qquad V = X - 1$$

are indepdent and normally distributed:

$$h(u, v) = \begin{cases} h_1(u, v) = \underline{\frac{u+v}{2}} \\ h_2(u, v) = \underline{\frac{u-v}{2}} \end{cases} \qquad J = \det\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} = -\frac{1}{2}$$

$$\begin{aligned} f_{U,V} &= f_{X,Y}(\underline{x}, \underline{y}) \cdot \frac{1}{2} \\ &\overset{\text{indp.}}{=} f_X(\underline{x}) \cdot f_X(\underline{y}) \\ &= \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{\underline{x}^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{\underline{y}^2}{2}} \\ &= \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\left\{\left(\underline{\frac{u+v}{2}}\right)^2 + \left(\underline{\frac{u-v}{2}}\right)^2 / 2\right\}} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{2}} \mathrm{e}^{-\frac{u^2}{4}} \cdot \frac{1}{\sqrt{2\pi}\sqrt{2}} \mathrm{e}^{-\frac{v^2}{4}} \end{aligned}$$

Thus U, V are independent r.v. distributed as $\mathcal{N}(0, 2)$.

Statistics

The probability that a discret random variable x is equal to some value $\bar{x} \in \mathcal{X}$ is:

$$\mathrm{p}_x(\bar{x}) = \mathbb{P}(x = \bar{x})$$

Definition 27.1 Almost Surely $\mathbb{P}$-(a.s.):
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. An event $\omega \in \mathcal{F}$ happens almost surely iff

$$\mathbb{P}(\omega) = 1 \qquad \iff \qquad \omega \text{ happens a.s.} \tag{27.1}$$

Definition 27.2 Probability Mass Function (PMF):

Definition 27.3 Discrete Random Variable (DVR): The set of possible values $\bar{x}$ of $\mathcal{X}$ is countable of finite.

$$\mathcal{X} = \{0, 1, 2, 3, 4, \ldots, 8\} \qquad \mathcal{X} = \mathbb{N} \tag{27.2}$$

Definition 27.4 Probability Density Function (PDF):
Is real function $f : \mathbb{R}^n \to [0, \infty)$ that satisfies:

Non-negativity: $$f(x) \geqslant 0, \quad \forall x \in \mathbb{R}^n \tag{27.3}$$

Normalization: $$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \overset{!}{=} 1 \tag{27.4}$$

Must be integrable (27.5)

Note: why do we need probability density functions

A continuous random variable X can realise an infinite count of real number values within its support B
(as there are an infinitude of points in a line segment).
Thus we have an infinitude of values whose sum of probabilities must equal one.
Thus these probabilities must each be zero otherwise we would obtain a probability of ∞. As we can not work with zero probabilities we use the next best thing, infinitesimal probabilities (defined as a limit).
We say they are almost surely equal to zero:

$$\mathbb{P}(X = x) = 0 \qquad \textbf{a.s.}$$

To have a sensible measure of the magnitude of these infinitesimal quantities, we use the concept of probability density, which yields a probability mass when integrated over an interval.

Definition 27.5 Continuous Random Variable (CRV):
A real random variable (rrv) X is said to be (absolutely) continuous if there exists a pdf ([def. 27.4]) f_X s.t. for any subset $B \subset \mathbb{R}$ it holds:

$$\mathbb{P}(X \in B) = \int_B f_X(x)\,\mathrm{d}x \tag{27.6}$$

Property 27.1 Zero Probability: If X is a continuous rrv ([def. 27.5]), then:

$$\mathbb{P}(X = a) = 0 \qquad \forall a \in \mathbb{R} \tag{27.7}$$

Property 27.2 Open vs. Closed Intervals: For any real numbers a and b, with $a < b$ it holds:

$$\begin{aligned} \mathbb{P}(a \leqslant X \leqslant b) = \mathbb{P}(a \leqslant X < b) &= \mathbb{P}(a < X \leqslant b) \\ &= \mathbb{P}(a < X < b) \end{aligned} \tag{27.8}$$

$\iff$ including or not the bounds of an interval does not modify the probability of a continuous rrv.

Note

Changing the value of a function at finitely many points has no effect on the value of a definite integral.

Corollary 27.1 : In particular for any real numbers a and b with $a < b$, letting $B = [a, b]$ we obtain:

$$\mathbb{P}(a \leqslant X \leqslant b) = \int_a^b f_x(x)\,\mathrm{d}x$$

Proof 27.1: Property 27.1:

$$\begin{aligned} \mathbb{P}(X = a) &= \lim_{\Delta x \to 0} \mathbb{P}(X \in [a, a + \Delta x]) \\ &= \lim_{\Delta x \to 0} \int_a^{a+\Delta x} f_X(x)\,\mathrm{d}x = 0 \end{aligned}$$

Proof 27.2: Property 27.2:

$$\begin{aligned} \mathbb{P}(a \leqslant X \leqslant b) = \mathbb{P}(a \leqslant X < b) &= \mathbb{P}(a < X \leqslant b) \\ &= \mathbb{P}(a < X < b) = \int_a^b f_X(x)\,\mathrm{d}x \end{aligned}$$

Definition 27.6 Support of a probability density function: The support of the density of a pdf $f_X(.)$ is the set of values of the random variable X s.t. its pdf is non-zero:

$$\mathtt{supp}(()f_X) := \{x \in \mathcal{X} | f(x) > 0\} \tag{27.9}$$

Note: this is not a rigorous definition.

Theorem 27.1 RVs are defined by a PDFs: A probability density function f_X completely determines the distribution of a continuous real-valued random variable X.

Corollary 27.2 Identically Distributed: From theorem 27.1 it follows that to RV X and Y that have exactly the same pdf follow the same distribution.
We say X and Y are identically distributed.

0.1. Cumulative Distribution Fucntion

Definition 27.7 Cumulative distribution function (CDF): Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
The (cumulative) distribution function of a real-valued random variable X is the function given by:

$$F_X(x) = \mathbb{P}(X \leqslant x) \qquad \forall x \in \mathbb{R}$$

Property 27.3:

Monotonically Increasing $$x \leqslant y \iff F_X(x) \leqslant F_X(y) \quad \forall x, y \in \mathbb{R} \tag{27.10}$$

Upper Limit $$\lim_{x \to \infty} F_X(x) = 1 \tag{27.11}$$

Lower Limit $$\lim_{x \to -\infty} F_X(x) = 0 \tag{27.12}$$

Definition 27.8 CDF of a discret rv X: Let X be discret rv with pdf p_X, then the CDF of X is given by:

$$F_X(x) = \mathbb{P}(X \leqslant x) = \sum_{t=-\infty}^{x} \mathrm{p}_X(t)$$

Definition 27.9 CDF of a continuous rv X: Let X be continuous rv with pdf f_X, then the CDF of X is given by:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t \qquad \iff \qquad \frac{\partial F_X(x)}{\partial x} = f_X(x)$$

Lemma 27.1 Probability Interval: Let X be a continuous rrv with pdf f_X and cumulative distribution function F_X, then it holds that:

$$\mathbb{P}(a \leqslant X \leqslant b) = F_X(b) - F_X(a) \tag{27.13}$$

Proof 27.3: [def. 27.9]:

$$F_X(x) = \mathbb{P}(X \leqslant x) = \mathbb{P}(X \in (-\infty, x)) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t$$

Proof 27.4: lemma 27.1:

$$\mathbb{P}(a \leqslant X \leqslant b) = \mathbb{P}(X \leqslant b) - \mathbb{P}(X \leqslant a)$$

or by the fundamental theorem of calculus (theorem 15.2):

$$\mathbb{P}(a \leqslant X \leqslant b) = \int_a^b f_X(t)\,\mathrm{d}t = \int_a^b \frac{\partial F_X(t)}{\partial t}\,\mathrm{d}t = [F_X(t)]|_a^b$$

Theorem 27.2 A continuous rv is fully characterized by its CDF: A cumulative distribution function completely determines the distribution of a continuous real-valued random variable.

1. Key figures

1.1. The Expectation

Definition 27.10 Expectation (disc. case):

$$\mu_X := \mathbb{E}_x[x] := \sum_{\bar{\mathbf{x}} \in \mathcal{X}} \bar{\mathbf{x}}\mathrm{p}_x(\bar{\mathbf{x}}) \tag{27.14}$$

Definition 27.11 Expectation (cont. case):

$$\mathbb{E}_x[x] := \int_{\bar{\mathbf{x}} \in \mathcal{X}} \bar{\mathbf{x}} f_x(\bar{\mathbf{x}})\,\mathrm{d}\bar{\mathbf{x}} \tag{27.15}$$

Law 27.1 Expectation of independent variables:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \tag{27.16}$$

Property 27.4 Translation and scaling: If $\mathbf{X} \in \mathbb{R}^n$ and $\mathbf{Y} \in \mathbb{R}^n$ are random vectors, and $a, b, a \in \mathbb{R}^n$ are constants then it holds:

$$\mathbb{E}[a + b\mathbf{X} + c\mathbf{Y}] = a + b\mathbb{E}[\mathbf{X}] + c\mathbb{E}[\mathbf{Y}] \tag{27.17}$$

Thus $\mathbb{E}$ is a linear operator ([def. 15.15]).

Note: Expectation of the expectation

The expectation of a r.v. X is a constant hence with Property 27.6 it follows:

$$\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X] \tag{27.18}$$

Property 27.5 Matrix×Expectation: If $\mathbf{X} \in \mathbb{R}^n$ is a randomn vector and $\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times m}$ are constant matrices then it holds:

$$\mathbb{E}[\mathbf{AXB}] = \mathbf{A}\mathbb{E}[(\mathbf{XB})] = \mathbf{A}\mathbb{E}[\mathbf{X}]\mathbf{B} \tag{27.19}$$

Proof 27.5: eq. (27.24):

$$\mathbb{E}[XY] = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{X,Y}(x,y)xy \overset{??}{=} \sum_{x\in\mathcal{X}} p_X(x)x \sum_{y\in\mathcal{Y}} p_Y(y)y = \mathbb{E}[X]\,\mathbb{E}[Y]$$

Definition 27.12 Autocorrelation/Crosscorelation $\gamma(t_1, t_2)$: Describes the covariance ([def. 27.16]) between the two values of a stochastic process $(\mathbf{X}_t)_{t\in T}$ at different time points t_1 and t_2.

$$\gamma(t_1,t_2) = \mathrm{Cov}\left[\mathbf{X}_{t_1}, \mathbf{X}_{t_2}\right] = \mathbb{E}\left[\left(\mathbf{X}_{t_1} - \mu_{t_1}\right)\left(\mathbf{X}_{t_2} - \mu_{t_2}\right)\right] \tag{27.20}$$

For zero time differences $t_1 = t_2$ the autocorrelation functions equals the variance:

$$\gamma(t,t) = \mathrm{Cov}\left[\mathbf{X}_t, \mathbf{X}_t\right] \overset{\text{eq. (27.35)}}{=} = \mathbb{V}\left[\mathbf{X}_t\right] \tag{27.21}$$

Notes

- **Hence** the autocorrelation describes the correlation of a function or signal with itself at a previous time point.
- **Given** a random time dependent variable $\mathbf{x}(t)$ the autocorrelation function $\gamma(t, t-\tau)$ describes how *similar* the time translated function $\mathbf{x}(t-\tau)$ and the original function $\mathbf{x}(t)$ are.
- If there exists some relation between the values of the time series that is non-random then the autocorrelation is non-zero.
- The autocorrelation is maximized/most similar for no translation $\tau = 0$ at all.

2. Key Figures

2.1. The Expectation

Definition 27.13 Expectation (disc. case):

$$\mu_X := \mathbb{E}_x[x] := \sum_{\bar{\mathbf{x}}\in\mathcal{X}} \bar{\mathbf{x}} p_x(\bar{\mathbf{x}}) \tag{27.22}$$

Definition 27.14 Expectation (cont. case):

$$\mathbb{E}_x[x] := \int_{\bar{\mathbf{x}}\in\mathcal{X}} \bar{\mathbf{x}} f_x(\bar{\mathbf{x}})\,\mathrm{d}\bar{\mathbf{x}} \tag{27.23}$$

Law 27.2 Expectation of independent variables:

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y] \tag{27.24}$$

Property 27.6 Translation and scaling: If $\mathbf{X} \in \mathbb{R}^n$ and $\mathbf{Y} \in \mathbb{R}^n$ are random vectors, and $a, b, a \in \mathbb{R}^n$ are constants then it holds:

$$\mathbb{E}[a + b\mathbf{X} + c\mathbf{Y}] = a + b\mathbb{E}[\mathbf{X}] + c\mathbb{E}[\mathbf{Y}] \tag{27.25}$$

Thus $\mathbb{E}$ is a linear operator[def. 15.15].

Property 27.7
Affine Transformation of the Expectation:
If $\mathbf{X} \in \mathbb{R}^n$ is a randomn vector, $\mathbf{A} \in \mathbb{R}^{m\times n}$ a constant matrix and $b \in \mathbb{R}^n$ then it holds:

$$\mathbb{E}[\mathbf{A}\mathbf{X} + b] = \mathbf{A}\mu + \mathbf{b} \tag{27.26}$$

Note: Expectation of the expectation

The expectation of a r.v. X is a constant hence with Property 27.6 it follows:

$$\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X] \tag{27.27}$$

Property 27.8 Matrix×Expectation: If $\mathbf{X} \in \mathbb{R}^n$ is a randomn vector and $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times m}$ are constant matrices then it holds:

$$\mathbb{E}[\mathbf{AXB}] = \mathbf{A}\mathbb{E}[(\mathbf{XB})] = \mathbf{A}\mathbb{E}[\mathbf{X}]\,\mathbf{B} \tag{27.28}$$

Proof 27.6: eq. (27.24):

$$\mathbb{E}[XY] = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p_{X,Y}(x,y)xy \overset{??}{=} \sum_{x\in\mathcal{X}} p_X(x)x \sum_{y\in\mathcal{Y}} p_Y(y)y = \mathbb{E}[X]\,\mathbb{E}[Y]$$

2.2. The Variance

Definition 27.15 Variance $\mathbb{V}[X]$: The variance of a random variable X is the expected value of the squared deviation from the expectation of X ($\mu = \mathbb{E}[X]$).
It is a measure of how much the actual values of a random variable X fluctuate around its executed value $\mathbb{E}[X]$ and is defined by:

$$\mathbb{V}[X] := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \overset{\text{see ?? 27.7}}{=} \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 \tag{27.29}$$

2.2.1. Properties

Property 27.9 Variance of a Constant: If $a \in \mathbb{R}$ is a constant then it follows that its expected value is deterministic ⇒ we have no uncertainty ⇒ no variance:

$$\mathbb{V}[a] = 0 \qquad \textbf{with} \qquad a \in \mathbb{R} \tag{27.30}$$

see shift and scaling for proof **??** 27.8

Property 27.10 Shifting and Scaling:

$$\mathbb{V}[a + bX] = a^2\sigma^2 \qquad \textbf{with} \qquad a \in \mathbb{R} \tag{27.31}$$

see **??** 27.8

Property 27.11 [proof 27.9]
Affine Transformation of the Variance:
If $\mathbf{X} \in \mathbb{R}^n$ is a randomn vector, $\mathbf{A} \in \mathbb{R}^{m\times n}$ a constant matrix and $b \in \mathbb{R}^n$ then it holds:

$$\mathbb{V}[\mathbf{A}\mathbf{X} + b] = \mathbf{A}\mathbb{V}[\mathbf{X}]\,\mathbf{A}^\intercal \tag{27.32}$$

Definition 27.16 Covariance: The Covariance is a measure of how much two or more random variables vary linearly with each other.

$$\begin{aligned}\mathrm{Cov}[X,Y] &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]\\ &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]\end{aligned} \tag{27.33}$$

see **??** 27.10

Definition 27.17 Covariance Matrix: The variance of a k-dimensional random vector $\mathbf{X} = \begin{pmatrix}X_1 & \dots & X_k\end{pmatrix}$ is given by a p.s.d. eq. (20.107) matrix called Covariance Matrix.
The Covariance is a measure of how much two or more random variables vary linearly with each other and the Variance on the diagonal is again a measure of how much a variable varies:

$$\begin{aligned}\mathbb{V}[\mathbf{X}] :=& \Sigma(\mathbf{X}) := \mathrm{Cov}[\mathbf{X},\mathbf{X}] := \\ =& \mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - \mathbb{E}[\mathbf{X}])^\intercal\right]\\ =& \mathbb{E}\left[\mathbf{X}\mathbf{X}^\intercal\right] - \mathbb{E}[\mathbf{X}]\,\mathbb{E}[\mathbf{X}]^\intercal \in [-\infty, \infty]\end{aligned} \tag{27.34}$$

$$= \begin{bmatrix} \mathbb{V}[X_1] & \cdots & \mathrm{Cov}[X_1, X_k] \\ \vdots & \ddots & \\ \mathrm{Cov}[X_k, X_1] & \cdots & \mathbb{V}[X_k] \end{bmatrix}$$

$$= \begin{bmatrix} \mathbb{E}[(X_1 - \mu_1)(X_1 - \mu_1)] & \cdots & \mathbb{E}[(X_1 - \mu_1)(X_k - \mu_k)] \\ \vdots & \ddots & \\ \mathbb{E}[(X_k - \mu_k)(X_1 - \mu_1)] & \cdots & \mathbb{E}[(X_k - \mu_k)(X_k - \mu_k)] \end{bmatrix}$$

Note: Covariance and Variance

The variance is a special case of the covariance in which two variables are identical:

$$\mathrm{Cov}[X,X] = \mathbb{V}[X] \equiv \sigma^2(X) \equiv \sigma_X^2 \tag{27.35}$$

Property 27.12 Translation and Scaling:

$$\mathrm{Cov}(a + bX, c + dY) = bd\mathrm{Cov}(X,Y) \tag{27.36}$$

Property 27.13
Affine Transformation of the Covariance:
If $\mathbf{X} \in \mathbb{R}^n$ is a randomn vector, $\mathbf{A} \in \mathbb{R}^{m\times n}$ a constant matrix and $b \in \mathbb{R}^n$ then it holds:

$$\mathrm{Cov}[\mathbf{A}\mathbf{X} + b] = \mathbf{A}\mathbb{V}[\mathbf{X}]\,\mathbf{A}^\intercal = \mathbf{A}\Sigma(\mathbf{X})\mathbf{A}^\intercal \tag{27.37}$$

Definition 27.18 Correlation Coefficient: Is the standardized version of the covariance:

$$\begin{aligned}\mathrm{Corr}[\mathbf{X}] :=& \frac{\mathrm{Cov}[\mathbf{X}]}{\sigma_{X_1}\cdots\sigma_{X_k}} \in [-1, 1] \\ =& \begin{cases} +1 & \text{if } Y = aX + b \text{ with } a > 0, b \in \mathbb{R} \\ -1 & \text{if } Y = aX + b \text{ with } a < 0, b \in \mathbb{R} \end{cases}\end{aligned} \tag{27.38}$$

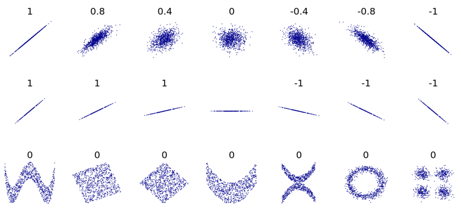


Figure 12: Several sets of (x, y) points, with their correlation coefficient

Law 27.3 Translation and Scaling:

$$\mathrm{Corr}(a + bX, c + dY) = \mathrm{sign}(b)\mathrm{sign}(d)\mathrm{Cov}(X,Y) \tag{27.39}$$

Note

- The correlation/covariance reflects the noisiness and direction of a linear relationship (top row fig. 12), **but** not the slope of that relationship (middle row fig. 12) nor many aspects of nonlinear relationships (bottom row)
- The set in the center of fig. 12 has a slope of 0 but in that case the correlation coefficient is undefined because the variance of Y is zero.
- Zero covariance/correlation $\mathrm{Cov}(X,Y) = \mathrm{Corr}(X,Y) = 0$ implies that there does not exist a **linear** relationship between the random variables X and Y.

Difference Covariance&Correlation

1. Variance is affected by scaling and covariance not **??** and law 27.3.
2. Correlation is dimensionless, whereas the unit of the covariance is obtained by the product of the units of the two RV variables.

Law 27.4 Covariance of independent RVs: The covariance/correlation of two independent variable's (**??**) is zero:

$$\begin{aligned}\mathrm{Cov}[X,Y] &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]\\ &\overset{\text{eq. (27.24)}}{=} = \mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[X]\,\mathbb{E}[Y] = 0\end{aligned}$$

Zero covariance/correlation⇏ independence

$$\mathrm{Cov}(X,Y) = \mathrm{Corr}(X,Y) = 0 \not\Rightarrow p_{X,Y}(x,y) = p_X(x)p_Y(y)$$

For example: let $X \sim \mathcal{U}([-1, 1])$ and let $Y = X^2$.

1. Clearly X and Y are dependent
2. **But** the covariance/correlation between X and Y is non-zero:

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{Cov}(X, X^2) = \mathbb{E}\left[X\cdot X^2\right] - \mathbb{E}[X]\,\mathbb{E}\left[X^2\right]\\ &= \mathbb{E}\left[X^3\right] - \mathbb{E}[X]\,\mathbb{E}\left[X^2\right] \underset{\text{eq. (27.52)}}{\overset{\text{eq. (27.63)}}{=}} 0 - 0\cdot\mathbb{E}\left[X^2\right]\end{aligned}$$

⇒ the relationship between Y and X must be non-linear.

Definition 27.19 Quantile: Are specific values q_α in the range[def. 15.10] of a random variable X that are defined as the value for which the cumulative probability is less then q_α with probability $\alpha \in (0, 1)$:

$$q_\alpha : \mathbb{P}(X \leqslant x) = F_X(q_\alpha) = \alpha \quad \xrightarrow{F \text{ invert.}} \quad q_\alpha = F_X^{-1}(\alpha) \tag{27.40}$$

3. Proofs

Proof 27.7: eq. (27.29)

$$\begin{aligned}\mathbb{V}[X] &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\right]\\ &\overset{\text{Property 27.6}}{=} \mathbb{E}\left[X^2\right] - 2\mathbb{E}[X]\,\mathbb{E}[X] + \mathbb{E}[X]^2 = \mathbb{E}\left[X^2\right] - \mu^2\end{aligned}$$

Proof 27.8: Property 27.10

$$\begin{aligned}\mathbb{V}[a + bX] &= \mathbb{E}\left[(a + bX - \mathbb{E}[a + bX])^2\right]\\ &= \mathbb{E}\left[\left(\cancel{a} + bX - \cancel{a} - b\mathbb{E}[X]\right)^2\right]\\ &= \mathbb{E}\left[(bX - b\mathbb{E}[X])^2\right]\\ &= \mathbb{E}\left[b^2(X - \mathbb{E}[X])^2\right]\\ &= b^2\mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = b^2\sigma^2\end{aligned}$$

Proof 27.9: Property 27.11

$$\begin{aligned}\mathbb{V}(\mathbf{A}\mathbf{X} + b) &= \mathbb{E}\left[(\mathbf{A}\mathbf{X} - \mathbb{E}[\mathbf{X}\mathbf{A}])^2\right] + 0 =\\ &= \mathbb{E}\left[(\mathbf{A}\mathbf{X} - \mathbb{E}[\mathbf{A}\mathbf{X}])(\mathbf{A}\mathbf{X} - \mathbb{E}[\mathbf{A}\mathbf{X}])^\intercal\right]\\ &= \mathbb{E}\left[\mathbf{A}(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{A}(\mathbf{X} - (\mathbb{E}[\mathbf{X}]))^\intercal\right]\\ &= \mathbb{E}\left[\mathbf{A}(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - (\mathbb{E}[\mathbf{X}])^\intercal\mathbf{A}^\intercal\right]\\ &= \mathbf{A}\mathbb{E}\left[(\mathbf{X} - \mathbb{E}[\mathbf{X}])(\mathbf{X} - (\mathbb{E}[\mathbf{X}])^\intercal\right]\mathbf{A}^\intercal = \mathbf{A}\mathbb{V}[\mathbf{X}]\,\mathbf{A}^\intercal\end{aligned}$$

Proof 27.10: eq. (27.33)

$$\begin{aligned}\mathrm{Cov}[X,Y] &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]\\ &= \mathbb{E}[XY - X\mathbb{E}[Y] - \mathbb{E}[X]\,Y + \mathbb{E}[X]\,\mathbb{E}[Y]]\\ &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[X]\,\mathbb{E}[Y] + \mathbb{E}[X]\,\mathbb{E}[Y]\\ &= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]\end{aligned}$$

Discrete Distributions

Definition 27.20 Multivariate Distribution: the variate refers to the number of input variables i.e. a m-variate distribution has m-input variables whereas a uni-variate distribution has only one.

Dimensional vs. Multivariate

The dimension refers to the number of dimensions we need to embed the function. If the variables of a function are independent than the dimension is the same as the number of inputs but the number of input variables can also be less.

4.1. Bernoulli Distribution $\mathrm{Bern}(p)$

Definition 27.21 Bernoulli Trial: Is a random experiment with exactly two possible outcomes, success (1) and failure (0), in which the probability of success/failure is constant in every trial i.e. independent trials.

Definition 27.22 Bernoulli Distribution $X \sim \mathrm{Bern}(p)$: X is a binary variable i.e. can only attain the values 0 (failure) or 1 (success) with a parameter p that signifies the success probability:

$$p(x; p) = \begin{cases} p & \text{for } x = 1 \\ 1 - p & \text{for } x = 0 \end{cases} \iff \begin{cases} \mathbb{P}(X = 1) = p \\ \mathbb{P}(X = 0) = 1 - p \end{cases}$$
$$= p^x \cdot (1 - p)^{1-x} \quad \text{for } x \in \{0, 1\}$$

$$\mathbb{E}[X] = p \quad (27.41) \qquad \mathbb{V}[X] = p(1 - p) \quad (27.42)$$

4.2. Multinoulli/Categorical Distribution $\mathrm{Cat}(n, p)$

Definition 27.23
Multinulli/Categorical Distribution $X \sim \mathrm{Cat}(\mathbf{p})$:
Is the generalization of the Bernoulli distribution[def. 27.22] to a sample space[def. 26.2] of k individual items $\{c_1, \dots, c_c\}$ with probabilities $\mathbf{p} = \{p_1, \dots, p_k\}$:

$$p(x = c_i|\mathbf{p}) = p_i \qquad \iff \qquad p(x|\mathbf{p}) = \prod_i^k p_i^{\delta[x=c_i]}$$

$$\sum_{j=1}^{k} p_j = 1 \qquad p_j \in [0, 1] \qquad \forall j = 1, \dots, k \qquad (27.43)$$

$$\mathbb{E}[X] = \mathbf{p} \qquad \mathbb{V}[X]_{i,j} = \Sigma_{i,j} = \begin{cases} p_i(1 - p_i) & \text{if } i = j \\ -p_i p_j & \text{if } i \neq j \end{cases}$$

Corollary 27.3
One-hot encoded Categorical Distribution:
If we encode the k categories by a *sparse vectors*[def. 20.68] with norm one:

$$\mathbb{B}_r^n = \left\{ \mathbf{x} \in \{0, 1\}^n : \mathbf{x}^\intercal \mathbf{x} = \sum_{i=1}^{n} \mathbf{x} = 1 \right\}$$

s.t. $\quad \mathbf{x}_j = \mathbf{e}_j \quad \iff \quad \mathbf{x} = c_j$

then we can rewrite eq. (27.43) as:

$$p(\mathbf{x}|\mathbf{p}) = \prod_i^k \mathbf{x}_i \cdot p_i \qquad \sum_{j=1}^{k} p_j = 1 \qquad (27.44)$$

4.3. Binomial Distribution $\mathcal{B}(n, p)$

Definition 27.24 Binomial Coefficient: The binomial coefficient occurs inside the binomial distribution**??** and signifies the different combinations/order that x out of n successes can happen.

Definition 27.25 Binomial Distribution [proof **??**]**:**
Models the probability of exactly X success given a fixed number *n-Bernoulli experiments*[def. 27.21], where the probability of success of a single experiment is given by p:

$$p(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

n :nb. of repetitions
x :nb. of successes
p :probability of success

$$\mathbb{E}[X] = np \quad (27.45) \qquad \mathbb{V}[X] = np(1 - p) \quad (27.46)$$

Note: Binomial Coefficient

The Binomial Coefficient corresponds to the permutation of two classes and not the variations as it seems from the formula.
Lets consider a box of n balls consisting of black and white balls. If we want to know the probability of drawing first x white and then $n - x$ black balls we can simply calculate:

$$\underbrace{(p \cdots p)}_{\text{x-times}} \cdot \underbrace{(q \cdots q)}_{n-x\text{-times}} = p^x q^{n-x}$$

4.4. Geometric Distribution $\mathrm{Geom}(p)$

Definition 27.26 Geometric Distribution $\mathrm{Geom}(p)$:
Models the probability of the number X of Bernoulli trials[def. 27.21] *until the first success*

$$p(x) = p(1 - p)^{x-1}$$

x :nb. of repetitions *until first success*
p :success probability *of single Bernoulli experiment*

$$F(x) = \sum_{i=1}^{x} p(1 - p)^{i-1} \overset{\text{eq. (12.4)}}{=} 1 - (1 - p)^x$$

$$\mathbb{E}[X] = \frac{1}{p} \quad (27.47) \qquad \mathbb{V}[X] = \frac{1 - p}{p^2} \quad (27.48)$$

Notes

- $\mathbb{E}[X]$ is the mean waiting time until the first success
- the number of trials x in order to have at least one success with a probability of $p(x)$:
$$x \geqslant \frac{p(x)}{1 - p}$$
- $\log(1 - p) \approx -p$ for small p

4.5. Poisson Distribution $\mathrm{Pois}(\lambda)$

Definition 27.27 Poisson Distribution: Is an extension of the binomial distribution, where the realization x of the random variable X may attain values in $\mathbb{Z}_{\geqslant 0}$.
It expresses the probability of a given number of events X occurring in a fixed interval if those events occur independently of the time since the last event.

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!} \qquad \begin{matrix} \lambda > 0 \\ x \in \mathbb{Z}_{\geqslant 0} \end{matrix} \qquad (27.49)$$

Event Rate λ: describes the average number of events in a single interval.

$$\mathbb{E}[X] = \lambda \quad (27.50) \qquad \mathbb{V}[X] = \lambda \quad (27.51)$$

Continuous Distributions

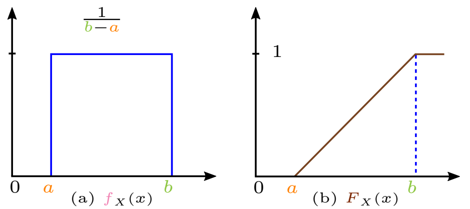
5.1. Uniform Distribution $\mathcal{U}(a, b)$

Definition 27.28 Uniform Distribution $\mathcal{U}(a, b)$:
Is probability distribution, where all intervals of the **same** length on the distribution's support[def. 27.6] $\mathrm{supp}(\mathcal{U}[a, b]) = [a, b]$ are equally probable/likely.

$$f(x) = \frac{1}{b - a} \mathbb{1}_{x \in [a;b)} = \begin{cases} \frac{1}{b-a} = \text{const} & \text{if} & a \leqslant x \leqslant b \\ 0 & & \text{else} \end{cases} \qquad (27.52)$$

$$F(x) = \begin{cases} 0 & & x < a \\ \frac{x-a}{b-a} & \text{if} & a \leqslant x \leqslant b \\ 1 & & x > b \end{cases} \qquad (27.53)$$

$$\mathbb{E}[X] = \frac{a + b}{2} \qquad \mathbb{V}(X) = \frac{(b - a)^2}{12} \qquad (27.54)$$



(a) $f_X(x)$ (b) $F_X(x)$

5.2. Exponential Distribution $\exp(\lambda)$

Definition 27.29 Exponential Distribution $X \sim \exp(\lambda)$:
Is the continuous analogue to the geometric distribution [def. 27.26].
It describes the probability $f(x; \lambda)$ that a continuous Poisson process (i.e., a process in which events occur continuously and independently at a constant average rate) will succeed/change state after a time interval x.

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{if} & x \geqslant 0 \\ 0 & & x < 0 \end{cases} \qquad (27.55)$$

$$F(x; \lambda) = \begin{cases} 1 - e^{-\lambda x} & \text{if} & x \geqslant 0 \\ 0 & & x < 0 \end{cases} \qquad (27.56)$$

$$\mathbb{E}[X] = \frac{1}{\lambda} \qquad \mathbb{V}(X) = \frac{1}{\lambda^2} \qquad (27.57)$$

5.3. Laplace Distribution

Definition 27.30 Laplace Distribution:

Laplace Distibution $\qquad f(\mathbf{x}; \mu, \sigma) = \frac{1}{2\sigma} \exp\left(-\frac{|\mathbf{x} - \mu|}{\sigma}\right)$ (27.58)

5.4. The Normal Distribution $\mathcal{N}(\mu, \sigma)$

Definition 27.31 Normal Distribution $\mathbf{X} \sim \mathcal{N}(\mu, \sigma^2)$:
Is a symmetric distribution where the population parameters μ, σ^2 are equal to the expectation and variance of the distribution:

$$\mathbb{E}[X] = \mu \qquad \mathbb{V}(X) = \sigma^2 \qquad (27.59)$$

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right\} \qquad (27.60)$$

$$F(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left\{-\frac{1}{2}\left(\frac{u - \mu}{\sigma}\right)^2\right\} \mathrm{d}u \qquad (27.61)$$
$$x \in \mathbb{R} \quad \text{or} \quad -\infty < x < \infty$$

$$\varphi_X(u) = \exp\left\{iu\mu - \frac{u^2\sigma^2}{2}\right\} \qquad (27.62)$$

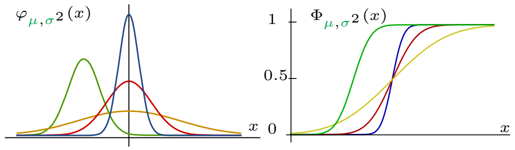


Figure 14: $\mu = 0$, $\sigma^2 = 0.2$ $\quad$ $\mu = 0$, $\sigma^2 = 1.0$ $\quad$ $\mu = 0$, $\sigma^2 = 5.0$ $\quad$ $\mu = -2$, $\sigma^2 = 0.5$

Property 27.14: $\mathbb{P}_X(\mu - \sigma \leqslant x \leqslant \mu + \sigma) = 0.66$

Property 27.15: $\mathbb{P}_X(\mu - 2\sigma \leqslant x \leqslant \mu + 2\sigma) = 0.95$

5.5. The Standard Normal distribution $\mathcal{N}(0, 1)$

Historic Problem: the cumulative distribution eq. (27.61) does not have an analytical solution and numerical integration was not always computationally so easy. So how should people calculate the probability of x falling into certain ranges $\mathbb{P}(x \in [a, b])$?
Solution: use a standardized form/set of parameters (by convention) $\mathcal{N}_{0,1}$ and tabulate many different values for its cumulative distribution $\phi(x)$ s.t. we can transform all families of Normal Distributions into the standardized version $\mathcal{N}(\mu, \sigma^2) \xrightarrow{z} \mathcal{N}(0, 1)$ and look up the value in its table.

Definition 27.32
Standard Normal Distribution $\mathbf{X} \sim \mathcal{N}(0, 1)$:

$$\mathbb{E}[X] = 0 \qquad \mathbb{V}(X) = 1 \qquad (27.63)$$

$$f(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \qquad (27.64)$$

$$F(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}u^2} \mathrm{d}u \qquad (27.65)$$
$$x \in \mathbb{R} \quad \text{or} \quad -\infty < x < \infty$$

$$\psi_X(u) = e^{\frac{u^2}{2}} \qquad \varphi_X(u) = e^{-\frac{u^2}{2}} \qquad (27.66)$$

Corollary 27.4
Standard Normal Distribution Notation: As the standard normal distribution is so commonly used people often use the letter Z in order to denote its the *standard* normal distribution and its α-quantile[def. 27.19] is then denoted by:

$$z_\alpha = \Phi^{-1}(\alpha) \qquad \alpha \in (0, 1) \qquad (27.67)$$

5.5.1. Calculating Probabilities

Property 27.16 Symmetry: Let $z > 0$

$$\mathbb{P}(Z \leqslant z) = \Phi(z) \qquad (27.68)$$
$$\mathbb{P}(Z \leqslant -z) = \Phi(-z) = 1 - \Phi(z) \qquad (27.69)$$
$$\mathbb{P}(-a \leqslant Z \leqslant b) = \Phi(b) - \Phi(-a) = \Phi(b) - (1 - \Phi(a))$$
$$\overset{a=b=z}{=} 2\Phi(z) - 1 \qquad (27.70)$$

5.5.2. Linear Transformations of Normal Dist.

Proposition 27.1 Linear Transformation [proof 27.12]: Let X be a normally distributed random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, then the linear transformed r.v. Y given by the *affine transformation* $Y = a + bX$ with $a \in \mathbb{R}$,$b \in \mathbb{R}_+$ follows:

$$Y \sim \mathcal{N}\left(a + b\mu, b^2\sigma^2\right) \iff f_Y(y) = \frac{1}{|b|} f_X\left(\frac{y-a}{b}\right) \tag{27.71}$$

Proposition 27.2 Standardization [proof 27.13]: Let X be a normally distributed random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, then there exists a linear transformation $Z = a + bX$ s.t. Z is a standard normally distributed random variable:

$$X \sim \mathcal{N}(\mu, \sigma^2) \xrightarrow{Z = \frac{X-\mu}{\sigma}} Z \sim \mathcal{N}(0,1) \tag{27.72}$$

Note

If we know how many standard deviations our distribution is away from our target value then we can characterize it fully by the standard normal distribution.

Proposition 27.3 [proof 27.14]
Standardization of the CDF: Let $F_X(X)$ be the cumulative distribution function of a normally distributed random variable $X \sim \mathcal{N}(\mu, \sigma^2)$, then the cumulative distribution function $\Phi_Z(z)$ of the standardized random normal variable $Z \sim \mathcal{N}(0,1)$ is related to $F_X(X)$ by:

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \tag{27.73}$$

6. The Multivariate Normal distribution

Definition 27.33 Multivariate Normal/Gaussian: An $\mathbb{R}^n$-valued random variable $\mathbf{X} = \begin{pmatrix} X_1 & \ldots, X_n \end{pmatrix}$ is *Multivariate Gaussian/Normal* if every linear combination of its components is a (one-dimensional) Gaussian:

$$\exists \mu, \sigma : \quad \mathscr{L}\left(\sum_{i=1}^{n} \alpha_i X_j\right) = \mathcal{N}(\mu, \sigma^2) \quad \forall \alpha_i \in \mathbb{R} \tag{27.74}$$

(possible degenerated $\mathcal{N}(0,0)$ for $\forall \alpha_j = 0$)

Note

- **Joint** vs. **multivariate**: a joint normal distribution can be a multivariate normal distribution or a product of univariate normal distributions **but**
- Multivariate refers to the number of variables that are placed as inputs to a function.

Definition 27.34
Multivariate Normal distribution $\mathbf{X} \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:
A k-dimensional random vector

$$\mathbf{X} = \begin{pmatrix} X_1 & \ldots & X_n \end{pmatrix}^\intercal \quad \text{with} \quad \boldsymbol{\mu} = \begin{pmatrix} \mathbb{E}[\mathbf{x}_1] & \ldots & \mathbb{E}[\mathbf{x}_k] \end{pmatrix}^\intercal$$

and $k \times k$ **p.s.d.**covariance matrix:

$$\boldsymbol{\Sigma} := \mathbb{E}[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\intercal] = [\mathrm{Cov}[\mathbf{x}_i, \mathbf{x}_j], 1 \leqslant i,j \leqslant k]$$

follows a k-dim multivariate normal/Gaussian distribution if its law[def. 26.25] satisfies:

$$f_{\mathbf{X}}(X_1, \ldots, X_k) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{27.75}$$

$$= \underbrace{\frac{1}{\sqrt{(2\pi)^k \det(\boldsymbol{\Sigma})}}}_{\text{Normalisation}} \exp\left(-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu})^\intercal \boldsymbol{\Sigma}^{-1}(\mathbf{X}-\boldsymbol{\mu})\right)$$

$$\varphi_{\mathbf{X}}(\mathbf{u}) = \exp\left\{i\mathbf{u}^\intercal\boldsymbol{\mu} - \frac{1}{2}\mathbf{u}\Sigma\mathbf{u}\right\} \tag{27.76}$$

6.1. Joint Gaussian Distributions

Definition 27.35 Jointly Gaussian Random Variables: Two random variables X, Y both scalars or vectors, are said to be jointly Gaussian if the joint vector random variable $\mathbf{Z} = \begin{bmatrix} X & Y \end{bmatrix}^\intercal$ is again a GRV.

Property 27.17 proof 27.16
Joint Independent Gaussian Random Variables: Let $X_1, \ldots, X_n$ be $\mathbb{R}$-valued *independent* random variables with laws $\mathcal{N}\left(\mu_i, \sigma_i^2\right)$. Then the law of $\mathbf{X} = \begin{pmatrix} X_1 & \ldots & X_n \end{pmatrix}$ is a (multivariate) Gaussian distribution $\mathbf{X} \sim \mathcal{N}(\mu, \Sigma)$ with:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix} \quad \text{and} \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \tag{27.77}$$

Corollary 27.5 Quadratic Form:
If $\mathbf{x}$ and $\mathbf{y}$ are both independent GRVs

$$\mathbf{x} \sim \mathcal{N}(\mu_x, \Sigma_x) \qquad \mathbf{y} \sim \mathcal{N}(\mu_y, \Sigma_y)$$

then they are jointly Gaussian[def. 27.35] given by:

$$\mathrm{p}(\mathbf{x}, \mathbf{y}) = \mathrm{p}(x)\mathrm{p}(y) \tag{27.78}$$

$$\propto \exp\left(-\frac{1}{2}\left\{(\mathbf{x}-\boldsymbol{\mu}_x)^\intercal \boldsymbol{\Sigma}_x^{-1}(\mathbf{x}-\boldsymbol{\mu}_x) + (\mathbf{y}-\boldsymbol{\mu}_y)^\intercal \boldsymbol{\Sigma}_y^{-1}(\mathbf{y}-\boldsymbol{\mu}_y)\right\}\right)$$

$$= \exp\left(-\frac{1}{2}\begin{bmatrix}(\mathbf{x}-\boldsymbol{\mu}_x)^\intercal & (\mathbf{y}-\boldsymbol{\mu}_y)^\intercal\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_x^{-1} & 0 \\ 0 & \boldsymbol{\Sigma}_y^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{x}-\boldsymbol{\mu}_x \\ \mathbf{y}-\boldsymbol{\mu}_y\end{bmatrix}\right)$$

$$\triangleq \exp -\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_z)^\intercal \boldsymbol{\Sigma}_z^{-1}(\mathbf{z}-\boldsymbol{\mu}_z)$$

Property 27.18
Marginal Distribution of Multivariate Gaussian: Let $\mathbf{X} = \begin{pmatrix} X_1 & \ldots & X_n \end{pmatrix}^\intercal \sim \mathcal{N}(\mu, \Sigma)$ be a an $\mathbb{R}^n$ valued Gaussian and let $V = \{1, 2, \ldots, n\}$ be the index set of its variables. The k-variate marginal distribution of the Gaussian indexed by a subset of the variables:

$$A = \{i_1, \ldots, i_k\} \qquad i_j \in V \tag{27.79}$$

is given by:

$$\mathbf{X} = \begin{pmatrix} X_{i_1} & \ldots & X_{i_k} \end{pmatrix}^\intercal \sim \mathcal{N}(\mu_A, \Sigma_{AA}) \tag{27.80}$$

$$\Sigma = \begin{bmatrix} \sigma^2_{i_1,i_1} & \cdots & \sigma^2_{i_1,i_k} \\ \vdots & \ddots & \vdots \\ \sigma^2_{i_k,i_1} & \cdots & \sigma^2_{i_k,i_k} \end{bmatrix} \quad \text{and} \quad \mu = \begin{bmatrix} \mu_{i_1} \\ \mu_{i_2} \\ \vdots \\ \mu_{i_k} \end{bmatrix}$$

6.2. Conditional Gaussian Distributions

Property 27.19 Conditional Gaussian Distribution: Let $\mathbf{X} = \begin{pmatrix} X_1 & \ldots & X_n \end{pmatrix}^\intercal \sim \mathcal{N}(\mu, \Sigma)$ be a an $\mathbb{R}^n$ valued Gaussian and let $V = \{1, 2, \ldots, n\}$ be the index set of its variables. Suppose we take two disjoint subsets of V:

$$A = \{i_1, \ldots, i_k\} \qquad B = \{j_1, \ldots, j_m\} \qquad i_l, j_{l'} \in V$$

then the conditional distribution of the random vector $\mathbf{X}_A$, conditioned on $\mathbf{X}_B$ given by $\mathrm{p}(\mathbf{X}_A | \mathbf{X}_B = \mathbf{x}_B)$ is:

$$\mathbf{X}_A = \begin{pmatrix} X_{i_1} & \ldots & X_{i_k} \end{pmatrix}^\intercal \sim \mathcal{N}\left(\mu_{A|B}, \Sigma_{A|B}\right) \tag{27.81}$$

$$\boxed{\begin{aligned} \mu_{A|B} &= \mu_A + \Sigma_{AB}\Sigma_{BB}^{-1}(\mathbf{x}_B - \mu_B) \\ \Sigma_{A|B} &= \Sigma_{AA} - \Sigma_{AB}\Sigma_{BB}^{-1}\Sigma_{BA} \end{aligned}}$$

Note

Can be proofed using the matrix inversion lemma but is a very tedious computation.

Corollary 27.6
Conditional Distribution of Joint Gaussian's: Let $\mathbf{X}$ and $\mathbf{Y}$ be jointly Gaussian random vectors:

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{C}^\intercal & \mathbf{B} \end{bmatrix}\right) \tag{27.82}$$

then the *marginal* distribution of $\mathbf{x}$ conditioned on $\mathbf{y}$ can be written as: $X \sim \mathcal{N}\left(\mu_{X|Y}, \Sigma_{X|Y}\right)$

$$\boxed{\begin{aligned} \boldsymbol{\mu}_{X|Y} &= \boldsymbol{\mu}_X + \mathbf{C}\mathbf{B}^{-1}(\mathbf{y} - \boldsymbol{\mu}_Y) \\ \Sigma_{X|Y} &= \mathbf{A} - \mathbf{C}\mathbf{B}^{-1}\mathbf{C}^\intercal \end{aligned}} \tag{27.83}$$

6.3. Transformations

Property 27.20 Multiples of Gaussian's $\mathbf{A}X$: Let $\mathbf{X} = \begin{pmatrix} X_1 & \ldots & X_n \end{pmatrix}^\intercal \sim \mathcal{N}(\mu, \Sigma)$ be a an $\mathbb{R}^n$ valued Gaussian and let $\mathbf{A} \in \mathbb{R}^{d \times n}$ then it follows:

$$Y = \mathbf{A}X \in \mathbb{R} \qquad Y \sim \mathcal{N}\left(\mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A}^\intercal\right) \tag{27.84}$$

Property 27.21 Affine Transformation of GRVs: Let $\mathbf{y} \in \mathbb{R}^n$ be GRV, $\mathbf{A} \in \mathbb{R}^{d \times n}$, $b \in \mathbb{R}^d$ and let $\mathbf{x}$ be defined by the affine transformation[def. 20.43]:

$$\mathbf{x} = \mathbf{A}\mathbf{y} + b \qquad \mathbf{A} \in \mathbb{R}^{d \times n}, b \in \mathbb{R}^d$$

Then $\mathbf{x}$ is a GRV (see **??** 27.15).

Property 27.22 Linear Combination of jointly GRVs: Let $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$ two jointly GRVs, and let $\mathbf{z}$ be defined as:

$$\mathbf{z} = \mathbf{A}_x\mathbf{x} + \mathbf{A}_y\mathbf{y} \qquad \mathbf{A}_x \in \mathbb{R}^{d \times n}, \mathbf{A}_x \in \mathbb{R}^{d \times m}$$

Then $\mathbf{z}$ is GRV (see **??** 27.17).

Definition 27.36 Gaussian Noise: Is statistical noise having a probability density function (PDF) equal to that of the normal/Gaussian distribution.

6.4. Gamma Distribution $\Gamma(x, \alpha, \beta)$

Definition 27.37 Gamma Distribution $X \sim \Gamma(x, \alpha, \beta)$**:** Is a widely used distribution that is related to the exponential distribution, Erlang distribution, and chi-squared distribution as well as Normal distribution:

$$\mathrm{f}(x; \alpha, \beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \mathrm{e}^{-\beta x} & \text{if} \quad x > 0 \\ 0 & \phantom{\text{if}} \quad x \leqslant 0 \end{cases} \tag{27.85}$$

$$\Gamma(\alpha) \overset{\text{eq. (15.81)}}{=} \int_0^\infty t^{\alpha-1}\mathrm{e}^{-t}\,\mathrm{d}t \tag{27.86}$$

with $\alpha, \beta \in \mathbb{R}_{>0}$

6.5. Chi-Square Distribution χ^2_k

6.6. Student's t-distribution

Definition 27.38 Student' t-distribution:

6.7. Delta Distribution

Definition 27.39 The delta function $\delta(\mathbf{x})$:
The delta/dirac function $\delta(\mathbf{x})$ is defined by:

$$\int_{\mathbb{R}} \delta(\mathbf{x}) f(\mathbf{x})\,\mathrm{d}\mathbf{x} = f(0)$$

for any integrable function f on $\mathbb{R}$.
Or alternativly by:

$$\delta(x - x_0) = \lim_{\sigma \to 0} \mathcal{N}(x | x_0, \sigma) \tag{27.87}$$

$$\approx \infty \mathbb{1}_{\{x = x_0\}} \tag{27.88}$$

Property 27.23 Properties of δ:

- **Normalization**: The delta function integrates to 1:
  $$\int_{\mathbb{R}} \delta(x)\,\mathrm{d}x = \int_{\mathbb{R}} \delta(x) \cdot c_1(x)\,\mathrm{d}x = c_1(0) = 1$$
  where $c_1(x) = 1$ is the constant function of value 1.
- **Shifting**:
  $$\int_{\mathbb{R}} \delta(x - x_0) f(x)\,\mathrm{d}x = f(x_0) \tag{27.89}$$
- **Symmetry**: $\int_{\mathbb{R}} \delta(-x) f(x)\,\mathrm{d}x = f(0)$
- **Scaling**: $\int_{\mathbb{R}} \delta(\alpha x) f(x)\,\mathrm{d}x = \frac{1}{|\alpha|} f(0)$

Note

- In mathematical terms δ is not a function but a **gernalized function**.
- We may regard $\delta(x - x_0)$ as a density with all its probability mass centered at the signle point x_0.
- Using a box/indicator function s.t. its surface is one and its width goes to zero, instead of a normaldistribution eq. (27.87) would be a non-differentiable/discret form of the dirac measure.

Definition 27.40 Heaviside Step Function:

$$H(x) := \frac{\mathrm{d}}{\mathrm{d}x} \max\{x, 0\} \quad x \in \mathbb{R}_{\neq 0} \tag{27.90}$$

or alternatively:

$$H(x) := \int_{-\infty}^{x} \delta(s)\,\mathrm{d}s \tag{27.91}$$

Proofs

Proof 27.11 Definition 27.25: Consider a sequence of n random $\{X_i\}_{i=1}^n$ Bernoulli experiments[def. 27.22] with success probability p.
Define the r.v. Y_n to be the sum of the n Bernoulli variables:

$$Y_n = \sum_{i=1}^{n} X_i \qquad n \in \mathbb{N}$$

i.e. the total number of successes. Now lets calculate the probability density function f_n of Y_n. First let $(x_1 \cdots\cdots x_n) \in \{0,1\}^n$ and let $y = \sum_{i=1}^n x_i$ a bit sting of zeros and ones, with one occuring y times.

$$\mathbb{P}((X_1, X_2, \cdots, X_n) = (x_1, x_2, \ldots, x_n))$$

$$= \underbrace{(\mathrm{p}\cdots\mathrm{p})}_{\mathrm{y}} \cdot \underbrace{(q \cdots q)}_{n-y\text{-times}} = \mathrm{p}^y(1-\mathrm{p})^{n-y}$$

However we need to take into account that there exists further realization $\mathbf{X} = \mathbf{x}$, that correspond to different orders of the elements in our two classes $\{0,1\}$ which leads to $\frac{n!}{y!(n-y)!} = \binom{n}{y}$:

$$f_n(y) = \binom{n}{y}\mathrm{p}^y(1-\mathrm{p})^{n-y} \qquad y \in \{0, 1, \ldots, n\}$$

Proof 27.12: proposition 27.1: Let X be normally distributed with $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$F_Y(y) \overset{y>0}{=} \mathbb{P}_Y(Y \leqslant y) = \mathbb{P}(a + bX \leqslant y) = \mathbb{P}_X\left(X \leqslant \frac{y-a}{b}\right)$$

$$= F_X\left(\frac{y-a}{b}\right)$$

$$F_Y(y) \overset{y<0}{=} \mathbb{P}_Y(Y \leqslant y) = \mathbb{P}(a + bX \leqslant y) = \mathbb{P}_X\left(X \geqslant \frac{y-a}{b}\right)$$

$$= 1 - F_X\left(\frac{y-a}{b}\right)$$

Differentiating both expressions w.r.t. y leads to:

$$f_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} = \begin{cases} \frac{1}{b}\frac{\mathrm{d}F_X\left(\frac{y-a}{b}\right)}{\mathrm{d}y} \\ \frac{1}{-b}\frac{\mathrm{d}F_X\left(\frac{y-a}{b}\right)}{\mathrm{d}y} \end{cases} = \frac{1}{|b|} f_X(x)\left(\frac{y-a}{b}\right)$$

eq. (27.71)).
in order to prove that $Y \sim \mathcal{N}\left(a + b\mu, b^2\sigma^2\right)$ we simply plug f_X in the previous expression:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma|b|}\exp\left\{-\frac{1}{2}\left(\frac{\frac{y-a}{b} - \mu}{\sigma}\right)^2\right\}$$

$$= \frac{1}{\sqrt{2\pi}\sigma|b|}\exp\left\{-\frac{1}{2}\left(\frac{y - (a + b\mu)}{\sigma|b|}\right)^2\right\}$$

Proof 27.13: proposition 27.2: Let X be normally distributed with $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$Z := \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma} = aX + b \quad \text{with } a = \frac{1}{\sigma}, b = -\frac{\mu}{\sigma}$$

$$\overset{\text{eq. (27.71)}}{\sim} \mathcal{N}\left(a\mu + b, a^2\sigma^2\right) \sim \mathcal{N}\left(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{\sigma^2}{\sigma^2}\right) \sim \mathcal{N}(0,1)$$

Proof 27.14: proposition 27.3: Let X be normally distributed with $X \sim \mathcal{N}(\mu, \sigma^2)$:

$$F_X(x) = \mathbb{P}(X \leqslant x) \overset{\substack{-\mu \\ \div\sigma}}{=} \mathbb{P}\left(\frac{X-\mu}{\sigma} \leqslant \frac{x-\mu}{\sigma}\right) = \mathbb{P}\left(Z \leqslant \frac{x-\mu}{\sigma}\right)$$

$$= \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Proof 27.15: Property 27.21 scalar case

Let $y \sim \mathrm{p}(y) = \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$ and define $\mathbf{x} = ay + b \qquad a \in \mathbb{R}_+,\ b \in \mathbb{R}$

Using the Change of variables formula it follows:

$$\mathrm{p}_x(\bar{x}) \overset{\text{eq. (26.46)}}{=} \frac{\mathrm{p}_y(\bar{y})}{\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right|} \qquad \left[\quad \left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = a \quad \right]$$

$$\overset{\bar{y}=\frac{\bar{x}-b}{a}}{=} \frac{1}{a}\frac{1}{\sqrt{2\pi\mu^2}} \exp\left(-\frac{1}{2\sigma^2}\Big(\overbrace{\frac{\bar{x}-b}{a}}^{\bar{y}(\bar{x})} - \mu\Big)^2\right)$$

$$= \frac{1}{\sqrt{2\pi a^2\mu^2}} \exp\left(-\frac{1}{2\sigma^2 a^2}\Big(\bar{x} \underbrace{-b - a\mu}_{\mu_x}\Big)^2\right)$$

Hence $\qquad x \sim \mathcal{N}(\mu_x, \sigma_x^2) = \mathcal{N}(a\mu + b, a^2\sigma^2)$

Note

We can also verify that we have calculated the right mean and variance by:

$$\mathbb{E}[x] = \mathbb{E}[ay + b] = a\mathbb{E}[y] + b = a\mu + b$$
$$\mathbb{V}[x] = \mathbb{V}[ay + b] = a^2\mathbb{V}[y] = a^2\sigma^2$$

Proof 27.16: **??**

$$\mathrm{p}_{\mathbf{X}}(\mathbf{u}) = \prod_{i}^{n} \mathrm{p}_{X_i}(u_i)$$
$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right)$$
$$\varphi_{\mathbf{X}}(\mathbf{u}) = \exp\left\{iu_1\mu_1 - \frac{1}{2}\sigma_1 u_1^2\right\} \cdots \exp\left\{iu_n\mu_n - \frac{1}{2}\sigma_n u_n^2\right\}$$
$$= \exp\left\{ i\sum_{i}^{n} u_n\mu_n - \frac{1}{2}\underset{\cdots\cdots}{\sum_{i}^{n} \sigma_n u_n^2} \right\} = \exp\left\{ i\mathbf{u}^{\intercal}\boldsymbol{\mu} - \frac{1}{2}\underset{\cdots}{\mathbf{u}\Sigma\mathbf{u}} \right\}$$

Proof 27.17: Property 27.22

From Property 27.21 it follows immediately that $\mathbf{z}$ is GRV $\mathbf{z} \sim \mathcal{N}(\mu_z, \Sigma_z)$ with:

$$\mathbf{z} = \mathbf{A}\xi \qquad \textbf{with} \qquad \mathbf{A} = \begin{bmatrix}\mathbf{A}_x & \mathbf{A}_y\end{bmatrix} \text{ and } \xi = \begin{pmatrix}\mathbf{x} & \mathbf{y}\end{pmatrix}$$

Knowing that $\mathbf{z}$ is a GRV it is sufficient to calculate μ_z and Σ_z in order to characterize its distribution:

$$\mathbb{E}[\mathbf{z}] = \mathbb{E}[\mathbf{A}_x x + \mathbf{A}_y y] = \mathbf{A}_x\mu_x + \mathbf{A}_y\mu_y$$
$$\mathbb{V}[\mathbf{z}] = \mathbb{V}[\mathbf{A}\xi] \overset{??}{=} \mathbf{A}\mathbb{V}[\xi]\mathbf{A}^{\intercal}$$
$$= \begin{bmatrix}\mathbf{A}_x & \mathbf{A}_y\end{bmatrix}\begin{bmatrix}\mathbb{V}[x] & \mathrm{Cov}[x,y] \\ \mathrm{Cov}[y,x] & \mathbb{V}[y]\end{bmatrix}\begin{bmatrix}\mathbf{A}_x & \mathbf{A}_y\end{bmatrix}^{\intercal}$$
$$= \begin{bmatrix}\mathbf{A}_x & \mathbf{A}_y\end{bmatrix}\begin{bmatrix}\mathbb{V}[x] & \mathrm{Cov}[x,y] \\ \mathrm{Cov}[y,x] & \mathbb{V}[y]\end{bmatrix}\begin{bmatrix}\mathbf{A}_x^{\intercal} \\ \mathbf{A}_y^{\intercal}\end{bmatrix}$$
$$= \mathbf{A}_x\mathbb{V}[x]\mathbf{A}_x^{\intercal} + \mathbf{A}_y\mathbb{V}[y]\mathbf{A}_y^{\intercal}$$
$$+ \underbrace{\mathbf{A}_y\mathrm{Cov}[y,x]\mathbf{A}_x^{\intercal}}_{=0\text{by independence}} + \underbrace{\mathbf{A}_x\mathrm{Cov}[x,y]\mathbf{A}_y^{\intercal}}_{=0\text{by independence}}$$
$$= \mathbf{A}_x\Sigma_x\mathbf{A}_x^{\intercal} + \mathbf{A}_y\Sigma_y\mathbf{A}_y^{\intercal}$$

Note

Can also be proofed by using the normal definition of [def. 27.15] and tedious computations.

Proof 27.18: Equation (27.43) If $\mathbf{x} = c_i$ i.e. the outcome c_i has occurred then it follows:

$$\prod_{j}^{k} \mathrm{p}_i^{\delta[x=c_i]} = \mathrm{p}_1^0 \cdots \mathrm{p}_i^1 \cdots \mathrm{p}_k^0 = 1 \cdots \mathrm{p}_i \cdots 1 = p(\mathbf{x} = c_i|\mathbf{p})$$

Sampling Methods

1. Sampling Random Numbers

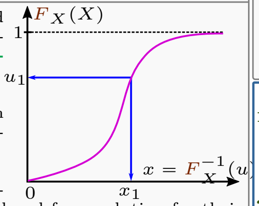
Most math libraries have uniform **random number generator (RNG)** i.e. functions to generate uniformly distributed random numbers $U \sim \mathcal{U}[a, b]$ (eq. (27.52)).
Furthermore repeated calls to these RNG are independent, that is:

$$p_{U_1,U_2}(u_1, u_2) \overset{??}{=} p_{U_1}(u_1) \cdot p_{U_2}(u_2) = \begin{cases} 1 & \text{if } u_1, u_2 \in [a, b] \\ 0 & \text{otherwise} \end{cases}$$

Question: using samples $\{u_1, \ldots, u_n\}$ of these CRVs with uniform distribution, how can we create random numbers with arbitrary discreet or continuous PDFs?

2. Inverse-transform Technique

Idea



Can make use of section 1 and the fact that CDF are increasing functions ([def. 15.12]). **Advantage**:

- Simple to implement
- All discrete distributions can be generated via inverse- transform technique

Drawback:

- Not all continuous distributions can be integrated/have closed form solution for their CDF.
  E.g. Normal-,Gamma-,Beta-distribution.

2.1. Continuous Case

Definition 28.1 One Continuous Variable: Given: a desired continuous pdf f_X and uniformly distributed rn $\{u_1, u_2, \ldots\}$:

1. Integrate the desired pdf f_X in order to obtain the desired cdf F_X:
$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \tag{28.1}$$
2. Set $F_X(X) \overset{!}{=} U$ on the range of X with $U \sim \mathcal{U}[0, 1]$.
3. Invert this equation/find the inverse $F_X^{-1}(U)$ i.e. solve:
$$U = F_X(X) = F_X\Big(\underbrace{F_X^{-1}(U)}_{X}\Big) \tag{28.2}$$
4. Plug in the uniformly distributed rn:
$$x_i = F_X^{-1}(u_i) \qquad \textbf{s.t.} \qquad x_i \sim f_X \tag{28.3}$$

Definition 28.2 Multiple Continuous Variable:
Given: a pdf of multiple rvs $f_{X,Y}$:

1. Use the product rule (**??**) in order to decompose $f_{X,Y}$:
$$f_{X,Y} = f_{X,Y}(x, y) = f_{X|Y}(x|y)f_Y(y) \tag{28.4}$$
2. Use [def. 28.3] to first get a rv for y of $Y \sim f_Y(y)$.
3. Then with this fixed y use [def. 28.3] again to get a value for x of $X \sim f_{X|Y}(x|y)$.

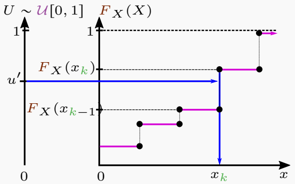
Proof 28.1: [def. 28.3]:
Claim: if U is a uniform rv on $[0, 1]$ then $F_X^{-1}(U)$ has F_X as its CDF.
Assume that F_X is strictly increasing ([def. 15.12]).
Then for any $u \in [0, 1]$ there must exist a **unique** x s.t. $F_X(x) = u$.
Thus F_X must be invertible and we may write $x = \underline{F_X^{-1}(u)}$.
Now let a arbitrary:
$$F_X(a) = \mathbb{P}(\underline{x} \leqslant a) = \mathbb{P}(F_X^{-1}(U) \leqslant a)$$
Since F_X is strictly increasing:
$$\mathbb{P}\left(F_X^{-1}(U) \leqslant a\right) = \mathbb{P}(U \leqslant F_X(a)) \overset{\text{eq. (27.52)}}{=} \int_0^{F_X(a)} 1\,dt = F_X(a)$$

Note

Strictly speaking we may not assume that a CDF is strictly increasing but we as all CDFs are weakly increasing ([def. 15.12]) we may always define an auxiliary function by its infinimum:
$$\hat{F}_X^{-1} := \inf\{x | F_X(X) \geqslant 0\} \qquad u \in [0, 1] \tag{28.5}$$

2.2. Discret Case

Idea



Given: a desired discret pmf p_X s.t. $\mathbb{P}(X = x_i) = p_X(x_i)$ and uniformly distributed rn $\{u_1, u_2, \ldots\}$.
Goal: given a uniformly distributed rn u determine k s.t.:

$$\sum_{i=1}^{k-1} < U \leqslant \sum_{i=1}^{k} \iff F_X(x_{k-1}) < u \leqslant F_X(x_k) \tag{28.6}$$

and return x_k.

Definition 28.3 One Discret Variable:

1. Compute the CDF of p_X ([def. 27.8])
$$F_X(x) = \sum_{t=-\infty}^{x} p_X(t) \tag{28.7}$$
2. Given the uniformly distributed rn $\{u_i\}_{i=1}^n$ find k^i ($\triangleq$ inversion) s.t.:
$$F_X\left(x_{k(i)-1}\right) < u_i \leqslant F_X\left(x_{k(i)}\right) \qquad \forall u_i \tag{28.8}$$

Proof 28.2: **??**: First of all notice that we can always solve for an unique x_k. **Given** a fixed x_k determine the values of u for which:
$$F_X(x_{k-1}) < u \leqslant F_X(x_k) \tag{28.9}$$
Now observe that:
$$u \leqslant F_X(x_k) = F_X(x_{k-1}) + p_X(x_k)$$
$$\Rightarrow F_X(x_{k-1}) < u \leqslant F_X(x_{k-1}) + p_X(x_k)$$
The probability of U being in $(F_X(x_{k-1}), F_X(x_k)]$ is:
$$\mathbb{P}(U \in [F_X(x_{k-1}), F_X(x_k)]) = \int_{F_X(x_{k-1})}^{F_X(x_k)} p_U(t)\,dt$$
$$= \int_{F_X(x_{k-1})}^{F_X(x_k)} 1\,dt = \int_{F_X(x_{k-1})}^{F_X(x_{k-1})+p_X(x_k)} 1\,dt = p_X(x_k)$$
Hence the random variable $x_k \in \mathcal{X}$ has the pdf p_X.

Definition 28.4
Multiple Continuous Variables (Option 1):
Given: a pdf of multiple rvs $p_{X,Y}$:

1. Use the product rule (**??**) in order to decompose $p_{X,Y}$:
$$p_{X,Y} = p_{X,Y}(x, y) = p_{X|Y}(x|y)p_Y(y) \tag{28.10}$$
2. Use **??** to first get a rv for y of $Y \sim p_Y(y)$.
3. Then with this fixed y use **??** again to get a value for x of $X \sim p_{X|Y}(x|y)$.

Definition 28.5
Multiple Continuous Variables (Option 2):
Note: this only works if $\mathcal{X}$ and $\mathcal{Y}$ are finite.
Given: a pdf of multiple rvs $p_{X,Y}$ **let** $N_x = |\mathcal{X}|$ and $N_y = |\mathcal{Y}|$ the number of elements in $\mathcal{X}$ and $\mathcal{Y}$.

Define
$$p_Z(1) = p_{X,Y}(1, 1), p_Z(2) = p_{X,Y}(1, 2), \ldots$$
$$\ldots, p_Z(N_x \cdot N_y) = p_{X,Y}(N_x, N_y)$$
Then simply apply **??** to the auxillary pdf p_Z

1. Use the product rule (**??**) in order to decompose $f_{X,Y}$:
$$f_{X,Y} = f_{X,Y}(x, y) = f_{X|Y}(x|y)f_Y(y) \tag{28.11}$$
2. Use [def. 28.3] to first get a rv for y of $Y \sim f_Y(y)$.
3. Then with this fixed y use [def. 28.3] again to get a value for x of $X \sim f_{X|Y}(x|y)$.

3. Monte Carlo Methods

3.1. Monte Carlo (MC) Integration

Integration methods s.a. Simpson integration[def. 23.27] suffer heavily from the curse of dimensionality.
An n-order[def. 23.24] quadrature scheme $\mathcal{Q}_n$ in 1-dimension is usually of order n/d in d-dimensions.
Idea estimate an integral stochastically by drawing sample from some distribution.

Definition 28.6 Monte Carlo Integration:
$$3 + 4 \tag{28.12}$$

3.2. Rejection Sampling

3.3. Importance Sampling

Descriptive Statistics

1. Populations and Distributions

Definition 29.1 Population $\{x_i\}_{i=1}^N$**:**
is the entire set of entities from which we can draw sample.

Definition 29.2
Families of Probability Distributions p_θ**:**
Are probability distributions that vary only by a set of hyper parameters θ[def. 29.1].

Definition 29.3 [example 29.3]
Population/Statistical Parameter θ**:**
Are the parameters defining families of probability distributions[def. 29.2]

Explanation 29.1 (Definition 29.1). *Such hyper parameters are often characterized by populations following a certain family of distributions with the help of a stastistc. Hence they are called population or statistical parameters.*

1.1. Characteristics of Populations

Definition 29.4 Population Mean: Given a population $\{x_i\}_{i=1}^N$ of size N its variance is defined as:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{29.1}$$

Definition 29.5 Population Variance: Given a population $\{x_i\}_{i=1}^N$ of size N its variance is defined as: $\{x_i\}_{i=1}^N$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2 \tag{29.2}$$

Note

The population variance and mean are equally to the mean derived from the true distribution of the population.

2. Sample Statistics

Definition 29.6 (Sample) Statistic: A statistc is a measuarble function T that assigns a **single** value t to a sample of random variables or population:

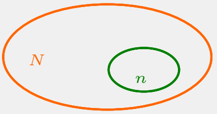
$$t : \mathbb{R}^n \mapsto \mathbb{R} \qquad t = T(X_1, \dots, X_n)$$

E.g. T could be the mean, variance,...

Definition 29.7 Degrees of freedom of a Statistic: Is the number of values in the final calculation of a statistic that are free to vary.

Note

The function itself is independent of the sample's distribution; that is, the function can be stated before realization of the data.



3. Point and Interval Estimation

Assume a population X with a given sample $\{x_i\}_{i=1}^n$ follows some family of distributions:

$$X \sim p_X(;\theta) \tag{29.3}$$

how can we estimate the correct value of the parameter θ or some function of that parameter $\tau(\theta)$?

3.1. Point Estimates

Definition 29.8 (Point) Estimator $\hat{\theta}$**:**
Is a statistic[def. 29.6] that tries estimates an unknown parameter θ of an underlying family of distributions[def. 29.2] for a given sample $\{\mathbf{x}_i\}_{i=1}^n$ of that distribution:

$$\hat{\theta} = t(\mathbf{x}_1, \dots, \mathbf{x}_n) \tag{29.4}$$

Note

The other kind of estimators are interval estimators which do not calculate a statistic **but** an interval of plausible values of an unknown population parameter θ.
The most prevalent forms of interval estimation are:

- Confidence intervals (frequentist method).
- Credible intervals (Bayesian method).

3.1.1. Empirical Mean

Definition 29.9 Sample/Empirical Mean $\bar{x}$**:**
The sample mean is an estimate/statistic of the population mean[def. 29.4] and can be calculated from an observation/sample of the total population $\{x_i\}_{i=1}^n \subset \{x_i\}_{i=1}^N$:

$$\bar{x} = \hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{29.5}$$

Corollary 29.1 [proof 29.1]
Unbiased Sample Mean:
The sample mean estimator is unbiased:

$$\mathbb{E}\left[\hat{\mu}_X\right] = \mu \tag{29.6}$$

Corollary 29.2 [Proof 29.2]
Variance of the Sample Mean:
The variance of the sample mean estimator is given by:

$$\mathbb{V}\left[\hat{\mu}_X\right] = \frac{1}{n}\sigma_X^2 \tag{29.7}$$

3.1.2. Empirical Variance

Definition 29.10 Biased Sample Variance:
The sample variance is an estimate/statistic of the population variance[def. 29.5] and can be calculated from an observation/sample of the total population $\{x_i\}_{i=1}^n \subset \{x_i\}_{i=1}^N$:

$$s_n^2 = \hat{\sigma}_X^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2 \tag{29.8}$$

Definition 29.11 [proof 29.3]
(Unbiased) Sample Variance:
The unbiased form of the sample variance[def. 29.10] is given by:

$$s^2 = \hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \mu)^2 \tag{29.9}$$

Definition 29.12 Bessel's Correction: The factor

$$\frac{n}{n-1} \tag{29.10}$$

is called Bessel's correction. Multiplying the uncorrected population variance eq. (29.8) by this term yields an unbiased estimated of the variance.

Attention:

- The Bessel correction holds for the variance but not for the standard deviation.
- Usually only the unbiased variance is used and sometimes also denoted by s_n^2

3.2. Interval Estimates

Definition 29.13 Interval Estimator $\hat{\theta}$**:**
Is an estimator that tries to bound an unknown parameter θ of an underlying family of distributions[def. 29.2] for a given sample $\{\mathbf{x}_i\}_{i=1}^n$ of that distribution.

Let $\theta \in \Theta$ and define two point statistics[def. 29.6] g and h then an interval estimate is defined as:

$$\mathbb{P}(L_n < \theta < U_n) = \gamma \qquad \begin{matrix} \forall\theta \in \Theta & L_n = g(\mathbf{x}_1, \dots, \mathbf{x}_n) \\ \gamma \in [0,1] & U_n = h(\mathbf{x}_1, \dots, \mathbf{x}_n) \end{matrix} \tag{29.11}$$

Statistical Tests

4. Parametric Hypothesis Testing

Definition 29.14 Parametric Hypothesis Testing:
Hypothesis testing is a statistical procedure in which a hypothesis is tested based on sampled data $X_1, \dots, X_n$.

4.1. Null Hypothesis

Definition 29.15 Null Hypothesis H_0**:**
A null hypothesis H_0 is an *assumption* on a population[def. 29.1] parameter[def. 29.3] θ:

$$H_0 : \theta = \theta_0 \tag{29.12}$$

Note

Often, a null hypothesis cannot be verified, but can only be falsified.

Definition 29.16 Alternative Hypothesis H_A/H_1**:**
The alternative hypothesis H_1 is an *assumption* on a population[def. 29.1] parameter[def. 29.3] θ that is opposite to the null hypothesis.

$$H_A : \theta \begin{cases} > \theta_0 & \text{(one-sided)} \\ < \theta_0 & \text{(one-sided)} \\ \neq \theta_0 & \text{(two-sided)} \end{cases} \tag{29.13}$$

4.2. Test Statistic

The decision on the hypothesis test is based on a sample from the population $X(n) = \{X_1, \dots, X_n\}$ however the decision is usually not based on single sample but a sample statistic[def. 29.6] as this is easier to use.

Definition 29.17 [example 29.4]
Test Statistic/Testing Parameter T**:**
Is a sample statistic[def. 29.6] used for hypothesis tests in order to give evidence for or against a hypothesis:

$$t_n = T(D_n) = T\left(\{X_1, \dots, X_n\}\right) \tag{29.14}$$

4.3. Sampling Distribution

Definition 29.18 $T_{\theta_0}(t)$
Null Distribution/Sampling Distribution under H_0**:**
Let $D_n = \{X_1, \dots X_n\}$ be a random sample from the true population p_{pop} and let $T(D_n)$ be a test statistic of that sample.
The probability distribution of the test statistic under the assumption that the null hypothesis is true is called *sampling distribution*:

$$t \sim T_{\theta_0} = T(t|H_0 \text{ true}) \qquad X_i \sim p_{\text{pop}} \tag{29.15}$$

4.4. The Critical Region

Given a sample $D_n = \{X_1, \dots, X_n\}$ of the true population p_{pop} how should we decide whether the null hypothesis should be rejected or not?
Idea: let $\mathcal{T}$ be the be the set of all possible values that the sample statistic T can map to. Now lets split $\mathcal{T}$ in two disjunct sets $\mathcal{T}_0$ and $\mathcal{T}_1$:

$$\mathcal{T} = \mathcal{T}_0 \cup \mathcal{T}_1 \qquad \mathcal{T}_0 \cap \mathcal{T}_1 = \varnothing$$

- if $t_n = T(X_n) \in \mathcal{T}_0$ we accept the null hypothesis H_0
- if $t_n = T(X_n) \in \mathcal{T}_1$ we reject the null hypothesis for H_1

Definition 29.19 Critical/Rejection Region $\mathcal{T}_1$**:** Is the set of all values of the test statistic[def. 29.17] t_n that causes us to reject the Null Hypothesis in favor for the alternative hypothesis H_A:

$$K = \mathcal{T}_1 = \{\mathcal{T} : H_0 \text{ rejected}\} \tag{29.16}$$

Definition 29.20 Acceptance Region $\mathcal{T}_0$**:** Is the region where we accept the null hypothesis H_0.

$$\mathcal{T}_0 = \{\mathcal{T} : H_0 \text{ accepted}\} \tag{29.17}$$

Definition 29.21 Critical Value c:
Is the value of *the critical region* $c \in \mathcal{T}_1$ which is closest to the *region of acceptance*[def. 29.20]:

4.5. Type I&II Errors

Definition 29.22
False Positive **Type I Error:**
Is the rejection of the null hypothesis H_0, even-tough it is true

$$\begin{aligned} &\text{Test rejects } H_0|H_0 \text{ true} \\ \iff\ & t_n \in \mathcal{T}_1|H_0 \text{ true} \end{aligned} \tag{29.18}$$

Definition 29.23
False Negative **Type II Error:**
Is the acceptance of a null hypothesis H_0, even-tough its false:

$$\begin{aligned} &\text{Test accepts } H_0|H_A \text{ true} \\ \iff\ & t_n \in \mathcal{T}_0|H_A \text{ true} \end{aligned} \tag{29.19}$$

Types of Errors

Decision	H_0 **true**	H_0 **false**
Accept	**TN**	**Type II (FN)**
Reject	**Type I (FP)**	**TP**

4.6. Statistical Significance & Power

Question: how should we choose the split $\{\mathcal{T}_0, \mathcal{T}_1\}$?
The bigger we choose Θ_1 (and thus the smaller Θ_0) the more likely it is to accept the alternative.
Idea: take the position of the adversary and choose Θ_1 so small that $\theta \in \Theta_1$ has only a small *probability* of occurring.

Definition 29.24 [example 29.5]
(Statistical) Significance α**:**
A study's defined significance level α denotes the probability to incur a *Type I Error*[def. 29.22]:

$$\mathbb{P}(t_n \in \mathcal{T}_1|H_0 \text{ true}) = \mathbb{P}(\text{test rejects } H_0|H_0 \text{ true}) \leqslant \alpha \tag{29.20}$$

Definition 29.25 Probability Type II Error β**:**
A test probability to for a *false negative*[def. 29.23] is defined as:

$$\beta(t_n) = \mathbb{P}(t_n \in \mathcal{T}_0|H_1 \text{ true}) = \mathbb{P}(\text{test accepts } H_0|H_1 \text{ true}) \tag{29.21}$$

Definition 29.26 (Statistical) Power $1-\beta$**:**
A study's power $1-\beta$ denotes a tests probability for a *true positive*:

$$1 - \beta(t_n) = \mathbb{P}(t_n \in \mathcal{T}_1|H_1 \text{ true}) \tag{29.22}$$

$$= \mathbb{P}(\text{test rejects } H_0|H_1 \text{ true}) \tag{29.23}$$

Corollary 29.3 Types of Split:
The Critical region is chosen s.t. we incur a Type I Error with probability less than α, which corresponds to the type of the test[def. 29.16]:

$$\begin{array}{llll} & \mathbb{P}(c_2 \leqslant X \leqslant c_1) \leqslant \alpha & \text{two-sided} & \\ \text{or} & \mathbb{P}(c_2 \leqslant X) \leqslant \frac{\alpha}{2} & \text{and} & \mathbb{P}(X \leqslant c_1) \leqslant \frac{\alpha}{2} \\ & \mathbb{P}(c_2 \leqslant X) \leqslant \alpha & \text{one-sided} & \\ & \mathbb{P}(X \leqslant c_1) \leqslant \alpha & \text{one-sided} & \end{array}$$

Truth Decission	H_0 **true**	H_0 **false**
H_0 **accept**	$1-\alpha$	$1-\beta$
H_0 **rejected**	α	β

4.7. P-Value

Definition 29.27 P-Value p**:**
Given a test statistic $t_n = T(X_1, \dots, X_n)$ the p-value $p \in [0,1]$ is the smallest value s.t. we reject the null hypothesis:

$$p := \inf_{\alpha}\{\alpha|t_n \in \mathcal{T}_1\} \qquad t_n = T(X_1, \dots, X_n) \tag{29.24}$$

Explanation 29.2.

- *The smaller the p-value the less likely is an observed statistic* t_n *and thus the higher is the evidence against a null hypothesis.*
- *A null hypothesis has to be rejected if the p-value is bigger than the chosen significance niveau* α.

5. Conducting Hypothesis Tests

1. Select an appropriate test statistic[def. 29.17] T.
2. Define the null hypothesis H_0 and the alternative hypothesis H_1 for T.
3. Find the sampling distribution[def. 29.18] $T_{\theta_0}(t)$ for T, given H_0 true.
4. Chose the significance level α
5. Evaluate the test statistic $t_n = T(X_1, \ldots, X_n)$ for the sampled data.
6. Determine the p-value p.
7. Make a decision (accept or reject H_0)

5.1. Tests for Normally Distributed Data

Let us consider an i.i.d. sample of observations $\{x_i\}_{i=1}^n$, of a normally distributed population $X_{\text{pop}} \sim \mathcal{N}(\mu, \sigma^2)$.
From eqs. (29.6) and (29.7) it follows that the *mean of the sample* is distributed as:

$$\overline{X}_n \sim \mathcal{N}(\mu, \sigma^2/n)$$

thus the mean of the sample $\overline{X}_n$ should equal the mean μ of the population. We now want to test the null hypothesis:

$$H_0 : \mu = \mu_0 \iff \overline{X}_n \sim \mathcal{N}(\mu_0, \sigma^2/n) \tag{29.25}$$

This is obviously only likely if the realization $\bar{x}_n$ is close to μ_0.

5.1.1. Z-Test σ known

Definition 29.28 Z-Test:
For a realization of Z with $\{x_i\}_{i=1}^n$ and mean $\bar{x}_n$:

$$z = \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}$$

we *reject the null hypothesis* $H_0 : \mu = \mu_0$ for the alternative H_A for significance niveau[def. 29.24] α if:

$$\begin{aligned}
|z| \geqslant z_{1-\frac{\alpha}{2}} &\iff z \leqslant z_{\frac{\alpha}{2}} \vee z \geqslant z_{1-\frac{\alpha}{2}} \\
&\iff z \in \mathcal{T}_1 = \left(-\infty, -z_{1-\frac{\alpha}{2}}\right] \cup \left[z_{1-\frac{\alpha}{2}}, \infty\right) \\
z \geqslant z_{1-\alpha} &\iff z \in \mathcal{T}_1 = \left[z_{1-\alpha}, \infty\right) \\
z \leqslant z_\alpha = -z_{1-\alpha} &\iff z \in \mathcal{T}_1 = (-\infty, -z_\alpha] = (\infty, -z_{1-\alpha}]
\end{aligned} \tag{29.26}$$

Notes

- Recall from [def. 27.19] and [cor. 27.4] that:
  $$z_\alpha \overset{\text{i.e. } \alpha=0.05}{=} z_{0.05} = \Phi^{-1}(\alpha) \iff \mathbb{P}(Z \leqslant z_{0.05}) = 0.05$$
- $|z| \geqslant z_{1-\frac{\alpha}{2}}$ which stands for:
  $$\mathbb{P}(Z \leqslant z_{0.05}) + \mathbb{P}(Z \geqslant z_{0.95}) = \mathbb{P}(Z \leqslant -z_{1-0.05}) + \mathbb{P}(Z \geqslant z_{0.95}) = \mathbb{P}(|Z| \geqslant z_{0.95})$$
  can be rewritten as:
  $$z \geqslant z_{1-\frac{\alpha}{2}} \quad \vee \quad -z \geqslant z_{1-\frac{\alpha}{2}} \iff z \leqslant -z_{1-\frac{\alpha}{2}} = z_{\frac{\alpha}{2}}$$
- One usually goes over to the standard normal distribution proposition 27.2 and thus test how far one is away from zero mean ⇒ Z-test.
- We thus inquire a Type I error with probability α and should be small i.e. 1%.

5.1.2. t-Test σ unknown

In reality we usually do not know the true σ of the whole data set and thus calculate it over our sample. This however increases uncertainty and thus our sample does no longer follow a normal distribution but a **t-distribution** wiht $n-1$ degrees of freedom:

$$T \sim t_{n-1} \tag{29.27}$$

Definition 29.29 t-Test:
For a realization of T with $\{x_i\}_{i=1}^n$ and mean $\bar{x}_n$:

$$t = \frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}}$$

we reject the null hypothesis $H_0 : \mu = \mu_0$ for the alternative H_A if:

$$\begin{aligned}
&|t| \geqslant t_{n-1,1-\frac{\alpha}{2}} \\
&\qquad \iff t \in \mathcal{T}_1 = \left(-\infty, -t_{n-1,1-\frac{\alpha}{2}}\right] \cup \left[t_{n-1,1-\frac{\alpha}{2}}, \infty\right) \\
&t \geqslant t_{n-1,1-\alpha} \\
&\qquad \iff t \in \mathcal{T}_1 = \left[t_{n-1,1-\alpha}, \infty\right) \\
&t \leqslant t_{n-1,\alpha} = -t_{n-1,1-\alpha} \\
&\qquad \iff t \in \mathcal{T}_1 = (-\infty, -t_{n-1,\alpha}] = (\infty, -t_{n-1,1-\alpha}]
\end{aligned}$$

Notes

- The t-distribution has fatter tails as the normal distribution ⇒ rare event become more likely
- For $n \to \infty$ the t-distribution goes over into the normal distribution
- The t-distribution gains a degree of foredoom for each sample and loses one for each parameter we are interested in ⇒ n-samples and we are interested in one parameter μ.

5.2. Confidence Intervals

Now we are interested in the opposite of the critical region[def. 29.19] namely the region of plaussible values.

Definition 29.30 Confidence Interval I:
Let $D_n = \{X_1, \ldots, X_n\}$ be a *sample* of observations and T_n a sample statistic of that sample. The confidence interval is defined as:

$$I(D_n) = \{\theta_0 : T_n(D_n) \in \mathcal{T}_0\} = \{\theta_0 : H_0 \text{ is not rejected}\} \tag{29.28}$$

Corollary 29.4 : The confidence interval captures the unkown parameter θ with probability $1-\alpha$:

$$\mathbb{P}_\theta(\theta \in I(D_n)) = \mathbb{P}(T_n(D_n) \in \mathcal{T}_0) = 1 - \alpha \tag{29.29}$$

6. Inferential Statistics

Goal of Inference

1. What is a good guess of the parameters of my model?
2. How do I quantify my uncertainty in the guess?

7. Examples

Example 29.1 ??: Let x be uniformly distributed on $[0,1]$ ([def. 27.28]) with pmf $p_X(x)$ then it follows:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{1}{p_Y(y)} \Rightarrow \mathrm{d}x = \mathrm{d}y p_y(y) \Rightarrow x = \int_{-\infty}^{y} p_y(t)\,\mathrm{d}t = F_Y(x)$$

Example 29.2 ??: Let

Example 29.3 Family of Distributions: The family of normal distribution $\mathcal{N}$ has two parameters $\left\{\mu, \sigma^2\right\}$

Example 29.4 Test Statistic: Lets assume the test statistic follows a normal distribution:

$$T \sim \mathcal{N}(\mu; 1)$$

however we are unsure about the population parameter[def. 29.3] $\theta = \mu$ but assume its equal to θ_0 thus the null-and alternative hypothesis are:

$$H_0 : \mu = \mu_0 \qquad H_1 : \mu \neq \mu_0$$

Example 29.5 Binomialtest:
Given: a manufacturer claims that a maximum of 10% of its delivered components are substandard goods.
In a sample of size $n = 20$ we find $x = 5$ goods that do not fulfill the standard and are skeptical that what the manufacture claims is true, so we want to test:

$$H_0 : p = p_0 = 0.1 \qquad \text{vs.} \qquad H_A : p > 0.1$$

We model the number of number of defective goods using the binomial distribution[def. 27.25]

$$\begin{aligned} X \sim & \mathcal{B}(n, p) \\ \sim & T(n, p) \end{aligned}, n = 20 \qquad \mathbb{P}(X \geqslant x) = \sum_{k=x}^{n} \binom{n}{k} p^k (1-p)^{n-k}$$

from this we find:

$$\mathbb{P}_{p_0}(X \geqslant 4) = 1 - \mathbb{P}_{p_0}(X \leqslant 3) = 0.13$$
$$\mathbb{P}_{p_0}(X \geqslant 5) = 1 - \mathbb{P}_{p_0}(X \leqslant 4) = 0.04 \leqslant \alpha$$

thus the probability that equal 5 or more then 5 parts out of the 20 are rejects is less then 4%.
⇒ throw away null hypothesis for the 5% niveau in favor to the alternative.
⇒ the 5% significance niveau is given by $K = \{5, 6, \ldots, 20\}$

Note

If $x < n/2$ it is faster to calculate $\mathbb{P}(X \geqslant x) = 1 - \mathbb{P}(X \leqslant x - 1)$

8. Proofs

Proof 29.1: [cor. 29.1]:

$$\mathbb{E}[\hat{\mu}_X] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\mathbb{E}\left[\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\mathbb{E}[\underbrace{\mu + \cdots + \mu}_{1,\ldots,n}]$$

Proof 29.2: [cor. 29.2]:

$$\mathbb{V}[\hat{\mu}_X] = \mathbb{V}\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] \overset{\text{Property } 27.10}{=} \frac{1}{n^2}\mathbb{V}\left[\sum_{i=1}^{n} x_i\right]$$
$$\frac{1}{n^2} n \mathbb{V}[X] = \frac{1}{n}\sigma^2$$

Proof 29.3: definition 29.11:

$$\begin{aligned}
\mathbb{E}\left[\hat{\sigma}_X^2\right] =& \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right] \\
=& \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^{n}\left(x_i^2 - 2x_i\bar{x} + \bar{x}^2\right)\right] \\
=& \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + \sum_{i=1}^{n}\bar{x}^2\right] \\
=& \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^{n} x_i^2 - 2n\bar{x} \cdot n\bar{x} + n\bar{x}^2\right] \\
=& \frac{1}{n-1}\mathbb{E}\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right] \\
=& \frac{1}{n-1}\left[\sum_{i=1}^{n}\mathbb{E}\left[x_i^2\right] - n\mathbb{E}\left[\bar{x}^2\right]\right] \\
=& \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n\mathbb{E}\left[\bar{x}^2\right]\right] \\
=& \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2 + \mu^2) - n\left(\frac{1}{n}\sigma^2 + \mu^2\right)\right] \\
=& \frac{1}{n-1}\left[(n\sigma^2 + n\mu^2) - \left(\sigma^2 + n\mu^2\right)\right] \\
=& \frac{1}{n-1}\left[n\sigma^2 - \sigma^2\right] = \frac{1}{n-1}\left[(n-1)\sigma^2\right] = \sigma^2
\end{aligned}$$

Stochastic Calculus

Stochastic Processes

Definition 30.1
Random/Stochastic Process $\{X_t, t \in \mathcal{T} \subseteq \mathbb{R}_+\}$:
An ($\mathbb{R}^d$-valued) stochastic process is a collection of ($\mathbb{R}^d$ *−valued*) random variables X_t on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The index set $\mathcal{T}$ is usually representing time and can be either an interval $[t_1, t_2]$ or a discrete set $\{t_1, t_2, \ldots\}$. Therefore, the random process X can be written as a function:

$$X : \mathcal{T} \subseteq \mathbb{R}_+ \times \Omega \mapsto \mathbb{R}^d \qquad \Longleftrightarrow \qquad (t, \omega) \mapsto X(t, \omega) \tag{30.1}$$

Definition 30.2 Sample path/Trajector/Realization: Is the *stochastic/noise signal* $r(\cdot, \omega)$ on the index set[def. 12.1] $\mathcal{T}$, that we obtain be sampling ω from Ω.

Notation

Even though the r.v. X is a function of two variables, most books omit the argument of the sample space $X(t, \omega) := X(t)$

Corollary 30.1 $\{X_t, t \in \mathcal{T} \subseteq \mathbb{R}_+\} > 0$
Strictly Positive Stochastic Processes: A stochastic process $\{X_t, t \in \mathcal{T} \subseteq \mathbb{R}_+\}$ is called strictly positive if it satisfies:

$$X_t > 0 \qquad \mathbb{P}\text{-a.s.} \qquad \forall t \in \mathcal{T} \tag{30.2}$$

Definition 30.3
Random/Stochastic Chain $\{X_t, t \in \mathcal{T} \subseteq \mathbb{R}_+\}$:
is a collection of random variables defined on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$[def. 26.1]. The random variables are ordered by an associated index set[def. 12.1] $\mathcal{T}$ and take values in the same mathematical *discrete state space*[def. 30.5] S, which must be measurable w.r.t. some σ-algebra[def. 26.6] Σ.
Therefore for a given probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and a measurable space (S, Σ), the random *chain* X is a collection of S-valued random variables that can be written as:

$$X : \mathcal{T} \times \Omega \mapsto S \qquad \Longleftrightarrow \qquad (t, \omega) \mapsto X(t, \omega) \tag{30.3}$$

Definition 30.4 Index/Parameter Set $\mathcal{T}$**:**
Usually represents time and can be either an interval $[t_1, t_2]$ or a discrete set $\{t_1, t_2, \ldots\}$.

Definition 30.5 State Space S**:**
Is the range/possible values of the random variables of a stochastic process[def. 30.1] and must be measurable[def. 26.7] w.r.t. some σ-algebra Σ.

Sample-vs. State Space

Sample space[def. 26.2] hints that we are working with probabilities i.e. probability measures will be defined on our sample space.
State space is used in dynamics, it implies that there is a time progression, and that our system will be in different states as time progresses.

Definition 30.6 Sample path/Trajector/Realization: Is the *stochastic/noise signal* $r(\cdot, \omega)$ on the index set $\mathcal{T}$, that we obtain be sampling ω from Ω.

Notation

Even though the r.v. X is a function of two variables, most books omit the argument of the sample space $X(t, \omega) := X(t)$

1.1. Filtrations

Definition 30.7 Filtration $\mathbb{F} = \{\mathcal{F}_t\}_{t \geqslant 0}$**:**
A collection $\{\mathcal{F}_t\}_{t \geqslant 0}$ of sub *σ-algebras*[def. 26.6] $\{\mathcal{F}_t\}_{t \geqslant 0} \in \mathcal{F}$ is called filtration if it is *increasing*:

$$\mathcal{F}_s \subseteq \mathcal{F}_t \qquad \forall s \leqslant t \tag{30.4}$$

Explanation 30.1 (Definition 30.7). *A filtration describes the flow of information i.e. with time we learn more information.*

Definition 30.8
Filtered Probability Space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geqslant 0}, \mathbb{P})$**:**
A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ together with a filtration $\{\mathcal{F}_t\}_{t \geqslant 0}$ is called a *filtered probability* space.

Definition 30.9 Adapted Process: A stochastic process $\{X_t, t \in \mathcal{T} \subseteq \mathbb{R}_+\}$ is called adapted *to a* filtration $\mathbb{F}$ if:

$$X_t \text{ is } \mathcal{F}_t\text{-measurable} \qquad \forall t \tag{30.5}$$

That is the value of X_t is observable at time t

Definition 30.10 Predictable Process: A stochastic process $\{X_t, t \in \mathcal{T} \subseteq \mathbb{R}_+\}$ is called predictable *w.r.t. a* filtration $\mathbb{F}$ if:

$$X_t \text{ is } \mathcal{F}_{t-1}\text{-measurable} \qquad \forall t \tag{30.6}$$

That is the value of X_t is known at time $t - 1$

Note

The price of a stock will usually be adapted since date k prices are known at date k.
On the other hand the interest rate of a bank account is usually already known at the beginning $k - 1$, s.t. the interest rate r_t ought to be $\mathcal{F}_{k-1}$ measurable, i.e. the process $r = (r_k)_{k=1,\ldots,T}$ should be predictable.

Corollary 30.2 : The amount of information of an adapted random process is increasing see example 30.1.

2. Martingales

Definition 30.11 Martingales: A stochastic process $X(t)$ is a martingale on a *filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geqslant 0}, \mathbb{P})$ if the following conditions hold:

① Given $s \leqslant t$ the best prediction of $X(t)$, with a filtration $\{\mathcal{F}_s\}$ is the current expected value:

$$\forall s \leqslant t \qquad \mathbb{E}\left[X(t) | \mathcal{F}_s\right] = X(s) \quad \text{a.s.} \tag{30.7}$$

② The expectation is finite:

$$\mathbb{E}\left[|X(t)|\right] < \infty \quad \forall t \geqslant 0 \quad X(t) \text{ is } \{\mathcal{F}_t\}_{t \geqslant 0} \text{ adapted} \tag{30.8}$$

Interpretation

- For any $\mathcal{F}_s$-adapted process the best prediction of $X(t)$ is the currently known value $X(s)$ i.e. if $\mathcal{F}_s = \mathcal{F}_{t-1}$ then the best prediction is $X(t - 1)$
- A martingale models fair games of limited information.

Definition 30.12 Auto Covariance $\gamma(t_2 - t_1)$**:**
Describes the covariance[def. 27.16] between two values of a stochastic process $(\mathbf{X}_t)_{t \in \mathcal{T}}$ at different time points t_1 and t_2.

$$\gamma(t_1, t_2) = \operatorname{Cov}\left[\mathbf{X}_{t_1}, \mathbf{X}_{t_2}\right] = \mathbb{E}\left[\left(\mathbf{X}_{t_1} - \mu_{t_1}\right)\left(\mathbf{X}_{t_2} - \mu_{t_2}\right)\right] \tag{30.9}$$

For zero time differences $t_1 = t_2$ the autocorrelation functions equals the variance:

$$\gamma(t, t) = \operatorname{Cov}\left[\mathbf{X}_t, \mathbf{X}_t\right] \overset{\text{eq. (27.35)}}{=} = \mathbb{V}\left[\mathbf{X}_t\right] \tag{30.10}$$

Notes

- **Hence** the autocorrelation describes the correlation of a function or signal with itself at a previous time point.
- **Given** a random time dependent variable $\mathbf{x}(t)$ the autocorrelation function $\gamma(t, t - \tau)$ describes how *similar* the time translated function $\mathbf{x}(t - \tau)$ and the original function $\mathbf{x}(t)$ are.
- If there exists some relation between the values of the time series that is non-random, then the autocorrelation is non-zero.
- The auto covariance is maximized/most similar for no translation $\tau = 0$ at all.

Definition 30.13 Auto Correlation $\rho(t_2 - t_1)$**:**
Is the scaled version of the auto-covariance[def. 30.12]:

$$\rho(t_2 - t_1) = \operatorname{Corr}\left[\mathbf{X}_{t_1}, \mathbf{X}_{t_2}\right] \tag{30.11}$$

$$= \frac{\operatorname{Cov}\left[\mathbf{X}_{t_1}, \mathbf{X}_{t_2}\right]}{\sigma_{X_{t_1}} \sigma_{X_{t_2}}} = \frac{\mathbb{E}\left[\left(\mathbf{X}_{t_1} - \mu_{t_1}\right)\left(\mathbf{X}_{t_2} - \mu_{t_2}\right)\right]}{\sigma_{X_{t_1}} \sigma_{X_{t_2}}}$$

3. Different kinds of Processes

3.1. Markov Process

Definition 30.14 Markov Process: A continuous-time stochastic process $X(t), t \in T$, is called a Markov process if for any finite parameter set $\{t_i : t_i < t_{i+1}\} \in T$ it holds:

$$\mathbb{P}\left(X(t_{n+1}) \in B | X(t_1), \ldots, X(t_n)\right) = \mathbb{P}\left(X(t_{n+1}) \in B | X(t_n)\right)$$

it thus follows for the *transition probability* – the probability of $X(t)$ lying in the set B at time t, given the value x of the process at time s:

$$\mathbb{P}(s, x, t, B) = P(X(t) \in B | X(s) = x) \qquad 0 \leqslant s < t \tag{30.12}$$

Interpretation

In order to predict the future only the current/last value counts.

Corollary 30.3 Transition Density: The transition probability of a continuous distribution p can be calculated via:

$$\mathbb{P}(s, x, t, B) = \int_B \mathrm{p}(s, x, t, y)\, \mathrm{d}y \tag{30.13}$$

3.2. Gaussian Process

Definition 30.15 Gaussian Process: Is a stochastic process $X(t)$ where the random variables follow a Gaussian distribution:

$$X(t) \sim \mathcal{N}\left(\mu(t), \sigma^2(t)\right) \qquad \forall t \in T \tag{30.14}$$

3.3. Diffusions

Definition 30.16 [proof 30.1],[proof 30.2]
Diffusion:
Is a Markov Process[def. 30.14] for which it holds that:

$$\mu(t, X(t)) = \lim_{t \to 0} \frac{1}{\Delta t} \mathbb{E}\left[X(t + \Delta t) - X(t) | X(t)\right] \tag{30.15}$$

$$\sigma^2(t, X(t)) = \lim_{t \to 0} \frac{1}{\Delta t} \mathbb{E}\left[\left(X(t + \Delta t) - X(t)\right)^2 | X(t)\right] \tag{30.16}$$

- $\mu(t, X(t))$ is called **drift**
- $\sigma^2(t, X(t))$ is called **diffusion coefficient**

Interpretation

There exist not discontinuities for the trajectories.

3.4. Brownian Motion/Wiener Process

Definition 30.17
d-**dim standard Brownian Motion/Wiener Process:**
Is an $\mathbb{R}^d$ valued *stochastic process*[def. 30.1] $(W_t)_{t \in \mathcal{T}}$ starting at $\mathbf{x}_0 \in \mathbb{R}^d$ that satisfies:

① **Normal Independent Increments**: the increments are *normally distributed independent random variables*:

$$W(t_i) - W(t_{i-1}) \sim \mathcal{N}\left(0, (t_i - t_{i-1})\mathbb{1}_{d \times d}\right) \qquad \forall i \in \{1, \ldots, T\} \tag{30.17}$$

② **Stationary increments**:
$W(t + \Delta t) - W(t)$ is independent of $t \in \mathcal{T}$

③ **Continuity**: for *a.e.* $\omega \in \Omega$, the function $t \mapsto W_t(\omega)$ is continuous

$$\lim_{t \to 0} \frac{\mathbb{P}\left(|W(t + \Delta t) - W(t)| \geqslant \delta\right)}{\Delta t} = 0 \qquad \forall \delta > 0 \tag{30.18}$$

④ **Start**

$$W(0) := W_0 = 0 \qquad a.s. \tag{30.19}$$

Notation

- In many source the Brownian motion is a synonym for the standard Brownian Motion and it is the same as the Wiener process.
- **However** in some sources the Wiener process is the standard Brownian Motion, while the Brownian motion denotes a general form $\alpha W(t) + \beta$.

Corollary 30.4 $W_t \sim \mathcal{N}(0, \sigma)$ [proof 30.4],[proof 30.5]**:**
The random variable W_t follows the $\mathcal{N}(0, \sigma)$ law

$$\mathbb{E}\left[W(t)\right] = \mu = 0 \tag{30.20}$$

$$\mathbb{V}\left[W(t)\right] = \mathbb{E}\left[W^2(t)\right] = \sigma^2 = t \tag{30.21}$$

3.4.1. Properties of the Wiener Process

Property 30.1 Non-Differentiable Trajectories:
The sample paths of a Brownian motion are not differentiable:

$$\frac{\mathrm{d}W(t)}{t} = \lim_{t \to 0} \mathbb{E}\left[\left(\frac{W(t + \Delta t) - W(t)}{\Delta t}\right)^2\right]$$

$$= \lim_{t \to 0} \frac{\mathbb{E}\left[W(t + \Delta t) - W(t)\right]}{\Delta t} = \lim_{t \to 0} \frac{\sigma^2}{\Delta t} = \infty$$

$\xrightarrow{\text{result}}$ cannot use normal calculus anymore
$\xrightarrow{\text{solution}}$ Ito Calculus see section 31.

Property 30.2 Auto covariance Function:
The auto-covariance[def. 30.12] for a Wiener process

$$\mathbb{E}\left[(W(t) - \mu t)(W(t') - \mu t')\right] = \min(t, t') \tag{30.22}$$

Property 30.3: A standard Brownian motion is a

Quadratic Variation

Definition 30.18 Total Variation: The total variation of a function $f : [a, b] \subset \mathbb{R} \mapsto \mathbb{R}$ is defined as:

$$LV_{[a,b]}(f) = \sup_{\Pi \in \mathcal{S}} \sum_{i=0}^{n_\Pi - 1} |f(x_{i+1}) - f(x_i)| \tag{30.23}$$

$$\mathcal{S} = \left\{\Pi\{x_0, \ldots, x_{n_\Pi}\} : \Pi \text{ is a partition }^{\text{[def. 23.8]}} \text{ of } [a, b]\right\}$$

it is a measure of the (one dimensional) length of a function w.r.t. to the y-axis, when moving alone the function.
Hence it is a measure of the variation of a function w.r.t. to the y-axis.

Definition 30.19
Total Quadratic Variation/"sum of squares":
The total quadratic variation of a function $f : [a, b] \subset \mathbb{R} \mapsto \mathbb{R}$ is defined as:

$$QV_{[a,b]}(f) = \sup_{\Pi \in \mathcal{S}} \sum_{i=0}^{n_\Pi - 1} |f(x_{i+1}) - f(x_i)|^2 \tag{30.24}$$

$$\mathcal{S} = \left\{\Pi\{x_0, \ldots, x_{n_\Pi}\} : \Pi \text{ is a partition }^{\text{[def. 23.8]}} \text{ of } [a, b]\right\}$$

Corollary 30.5 Bounded (quadratic) Variation:
The (quadratic) variation[def. 30.18] of a function is bounded if it is finite:

$$\exists M \in \mathbb{R}_+ : \quad LV_{[a,b]}(f) \leqslant M \quad \left(QV_{[a,b]}(f) \leqslant M\right) \quad \forall \Pi \in \mathcal{S} \tag{30.25}$$

Theorem 30.1 Variation of Wiener Process: Almost surely the total variation of a Brownian motion over a interval $[0, T]$ is infinite:

$$\mathbb{P}\left(\omega : LV(W(\omega)) < \infty\right) = 0 \tag{30.26}$$

Theorem 30.2 [proof 30.6]
Quadratic Variation of standard Brownian Motion:
The quadratic variation of a standard Brownian motion over $[0, T]$ is finite:

$$\lim_{N \to \infty} \sum_{k=1}^{N} \left[W\left(k \frac{T}{N}\right) - W\left((k - 1)\frac{T}{N}\right)\right]^2 = T \quad \text{with probability 1} \tag{30.27}$$

Corollary 30.6 : theorem 30.2 can also be written as:

$$(\mathrm{d}W(t))^2 = \mathrm{d}t \tag{30.28}$$

3.4.2. Lévy's Characterization of BM

Theorem 30.3 [proof 30.7],[proof 30.8]
***d*-dim standard BM/Wiener Process by Paul Lévy:**
An $\mathbb{R}^d$ valued *adapted stochastic process*[def's. 30.1, 30.7] $(W_t)_{t\in\mathcal{T}}$ with the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{R}_+}$, that satisfies:

① **Start**

$$W(0) := W_0 = 0 \qquad a.s. \qquad (30.29)$$

② **Continuous Martingale:** W_t is an a.s. *continuous* martingale[def. 30.11] w.r.t. the filtration $(\mathcal{F}_t)_{t\in\mathcal{T}}$ under $\mathbb{P}$.

③ **Quadratic Variation:**

$$W_t^2 - t \text{ is also an martingale} \iff QV(W_t) = t \qquad (30.30)$$

is a standard Brownian motion[def. 30.24].

Further Stochastic Processes

3.4.3. White Noise

Definition 30.20 Discrete-time white noise: Is a random signal $\{\epsilon_t\}_{t\in T_{\text{discret}}}$ having equal intensity at different frequencies and is defined by:

- Having zero tendencies/expectation (otherwise the signal would not be random):
$$\mathbb{E}[\epsilon * [k]] = 0 \qquad \forall k \in T_{\text{discret}} \qquad (30.31)$$
- Zero autocorrelation[def. 30.13] γ i.e. the signals of different times are in no-way correlated:
$$\begin{aligned}\gamma(\epsilon * [k], \epsilon * [k+n]) &= \mathbb{E}\left[\epsilon * [k]\epsilon * [k+n]^{\intercal}\right]\\ &= \mathbb{V}[\epsilon * [k]]\,\delta_{\text{discret}}[n]\\ &\qquad \forall k, n \in T_{\text{discret}}\end{aligned} \qquad (30.32)$$

With $$\delta_{\text{discret}}[n] := \begin{cases}1 & \text{if } n = 0\\ 0 & \text{else}\end{cases}$$

See proofs

Definition 30.21 Continuous-time white noise: Is a random signal $(\epsilon_t)_{t\in T_{\text{continuous}}}$ having equal intensity at different frequencies and is defined by:

- Having zero tendencies/expectation (otherwise the signal would not be random):
$$\mathbb{E}[\epsilon * (t)] = 0 \qquad \forall t \in T_{\text{continuous}} \qquad (30.33)$$
- Zero autocorrelation[def. 30.13] γ i.e. the signals of different times are in no-way correlated:
$$\gamma(\epsilon * (t), \epsilon * (t+\tau)) = \mathbb{E}\left[\epsilon * (t)\epsilon * (t+\tau)^{\intercal}\right] \qquad (30.34)$$
$$\overset{\text{eq. (27.88)}}{=} \mathbb{V}[\epsilon * (t)]\,\delta(t-\tau) = \begin{cases}\mathbb{V}[\epsilon * (t)] & \text{if } \tau = 0\\ 0 & \text{else}\end{cases}$$
$$\forall t, \tau \in T_{\text{continuous}} \qquad (30.35)$$

Definition 30.22 Homoscedastic Noise: Has constant variability for all observations/time-steps:

$$\mathbb{V}\left[\epsilon_{i,t}\right] = \sigma^2 \qquad \begin{matrix}\forall t = 1,\ldots,T\\ \forall i = 1,\ldots,N\end{matrix} \qquad (30.36)$$

Definition 30.23 Heteroscedastic Noise: Is noise whose variability may vary with each observation/time-step:

$$\mathbb{V}\left[\epsilon_{i,t}\right] = \sigma(i,t)^2 \qquad \begin{matrix}\forall t = 1,\ldots,T\\ \forall i = 1,\ldots,N\end{matrix} \qquad (30.37)$$

3.4.4. Generalized Brownian Motion

Definition 30.24 Brownian Motion:
Let $\{W_t\}_{t\in\mathbb{R}_+}$ be a standard Brownian motion[def. 30.17], and define:

$$X_t = \mu t + \sigma W_t \qquad t \in \mathbb{R}_+ \qquad \begin{matrix}\mu \in \mathbb{R} &: \text{drift parameter}\\ \sigma \in \mathbb{R}_+ &: \text{scale parameter}\end{matrix} \qquad (30.38)$$

then $\{X_t\}_{t\in\mathbb{R}_+}$ is normally distributed with mean μt and variance $t\sigma^2$ $X_t \sim \mathcal{N}\left(\mu t, \sigma^2 t\right)$.

Theorem 30.4 Normally Distributed Increments:
If $W(T)$ is a Brownian motion, then $W(t) - W(0)$ is a normal random variable with mean μt and variance $\sigma^2 t$, where $\mu, \sigma \in \mathbb{R}$. From this it follows that $W(t)$ is distributed as:

$$f_{W(t)}(x) \sim \mathcal{N}(\mu t, \sigma^2 t) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left\{-\frac{(x-\mu t)^2}{2\sigma^2 t}\right\} \qquad (30.39)$$

Corollary 30.7 : More generally we may define the process:

$$t \mapsto f(t) + \sigma W_t \qquad (30.40)$$

which corresponds to a noisy version of f.

Corollary 30.8
Brownian Motion as a Solution of an SDE: A stochastic process X_t follows a BM with drift μ and scale σ if it satisfies the following SDE:

$$\mathrm{d}X(t) = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}W(t) \qquad (30.41)$$
$$X(0) = 0 \qquad (30.42)$$

3.4.5. Geometric Brownian Motion (GBM)

For many processes $X(t)$ it holds that:
- there exists an (exponential) growth
- that the values may not be negative $X(t) \in \mathbb{R}_+$

Definition 30.25 Geometric Brownian Motion:
Let $\{W_t\}_{t\in\mathbb{R}_+}$ be a standard Brownian motion[def. 30.17] the stochastic process $\mathbf{S}_t^1 \triangleq \mathbf{S}^1(t)$ with drift parameter μ and scale σ satisfying the SDE:

$$\begin{aligned}\mathrm{d}\mathbf{S}_t^1 &= \mathbf{S}_t^1\left(\mu\,\mathrm{d}t + \sigma\,\mathrm{d}W_t\right)\\ &= \mu\mathbf{S}_t^1\,\mathrm{d}t + \sigma\mathbf{S}_t^1\,\mathrm{d}W_t\end{aligned} \qquad (30.43)$$

is called geometric Brownian motion and is given by:

$$\mathbf{S}_t^1 = \mathbf{S}_0^1 \exp\left(\sigma W_t + \left(\mu - \frac{1}{2}\sigma^2\right)t\right) \qquad t \in \mathbb{R}_+ \qquad (30.44)$$

Corollary 30.9 Log-normal Returns:
For a geometric BM we obtain log-normal returns:

$$\ln\left(\frac{S_t}{S_0}\right) = \bar{\mu}t + \sigma W(t) \iff \bar{\mu}t + \sigma W(t) \sim \mathcal{N}(\mu t, \sigma^2 t)$$

with $$\bar{\mu} := \mu - \frac{1}{2}\sigma^2 \qquad (30.45)$$

3.4.6. Locally Brownian Motion

Definition 30.26 Locally Brownian Motion:
Let $\{W_t\}_{t\in\mathbb{R}_+}$ be a standard Brownian motion[def. 30.17] a local Brownian motion is a stochastic process $X(t)$ that satisfies the SDE:

$$\mathrm{d}X(t) = \mu\left(X(t), t\right)\mathrm{d}t + \sigma\left(X(t), t\right)\mathrm{d}W(t) \qquad (30.46)$$

Note

A local Brownian motion is an generalization of a geometric Brownian motion.

3.4.7. Ornstein-Uhlenbeck Process

Definition 30.27 Ornstein-Uhlenbeck Process:
Let $\{W_t\}_{t\in\mathbb{R}_+}$ be a standard Brownian motion[def. 30.17] a Ornstein-Uhlenbeck Process or exponentially correlated noise is a stochastic process $X(t)$ that satisfies the SDE:

$$\mathrm{d}X(t) = -aX(t)\,\mathrm{d}t + b\sigma\,\mathrm{d}W(t) \qquad a > 0 \qquad (30.47)$$

3.5. Poisson Processes

Definition 30.28 Rare/Extreme Events: Are events that lead to discontinuous in stochastic processes.

Problem

A Brownian motion is not sufficient as model in order to describe extreme events s.a. crashes in financial market time series. Need a model that can describe such discontinuities/jumps.

Definition 30.29 Poisson Process: A Poisson Process with *rate* $\lambda \in \mathbb{R}_{\geqslant 0}$ is a collection of random variables $X(t)$, $t \in [0,\infty)$ defined on a probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geqslant 0}, \mathbb{P})$, having a discrete *state space* $N = \{0,1,2,\ldots\}$ and satisfies:

1. $X_0 = 0$
2. The increments follow a Poisson distribution[def. 27.27]:
$$\mathbb{P}\left((X_t - X_s) = k\right) = \frac{\lambda(t-s)}{k!}\mathrm{e}^{-\lambda(t-s)} \qquad \begin{matrix}0 \leqslant s < t < \infty\\ \forall k \in \mathbb{N}\end{matrix}$$
3. No correlation of (non-overlapping) increments:
$$\forall t_0 < t_1 < \cdots < t_n : \text{the increments are independent}$$
$$X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}} \qquad (30.48)$$

Interpretation

A Poisson Process is a *continuous-time* process with *discrete, positive* realizations in $\in \mathbb{N}_{\geqslant 0}$

Corollary 30.10 Probability of events: Using Taylor in order to expand the Poisson distribution one obtains:

$$\mathbb{P}\left(X_{(t+\Delta t)} - X_t \neq 0\right) = \lambda\Delta t + o(\Delta t^2) \qquad t \text{ small i.e. } t \to 0 \qquad (30.49)$$

1. Thus the probability of an event happening during Δt is proportional to time period and the rate λ
2. The probability of two or more events to happen *during* Δt is of order $o(\Delta t^2)$ and thus extremely small (as *Deltat* is small).

Definition 30.30 Differential of a Poisson Process: The differential of a Poisson Process is defined as:

$$\mathrm{d}X_t = \lim_{\Delta t \to \mathrm{d}t}\left(X_{(t+\Delta t)} - X_t\right) \qquad (30.50)$$

Property 30.4 Probability of Events for differential: With the definition of the differential and using the previous results from the Taylor expansion it follows:

$$\mathbb{P}\left(\mathrm{d}X_t = 0\right) = 1 - \lambda \qquad (30.51)$$
$$\mathbb{P}\left(|\mathrm{d}X_t| = 1\right) = \lambda \qquad (30.52)$$

Proofs

Proof 30.1: eq. (30.15):
Let by δ denote the displacement of a particle at each step, and assume that the particles start at the center i.e. $x(0) = 0$, then we have:

$$\begin{aligned}\mathbb{E}[x(n)] &= \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} x_i(n)\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[x_i(n-1) \pm \delta]\\ &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[x_i(n-1)]\\ &\overset{\text{induction}}{=} \mathbb{E}\left[x_{n-1}\right] = \ldots\mathbb{E}[x(0)] = 0\end{aligned}$$

Thus in expectation the particles goes nowhere.

Proof 30.2: eq. (30.16):
Let by δ denote the displacement of a particle at each step, and assume that the particles start at the center i.e. $x(0) = 0$, then we have:

$$\begin{aligned}\mathbb{E}\left[x(n)^2\right] &= \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} x_i(n)^2\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[x_i(n-1) \pm \delta]^2\\ &= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[x_i(n-1)^2 \pm 2\delta x_i(n-1) + \delta^2\right]\\ &\overset{\text{ind.}}{=} \mathbb{E}\left[x_{n-1}^2\right] + \delta^2 = \mathbb{E}\left[x_{n-2}^2\right] + 2\delta^2 = \ldots\\ &= \mathbb{E}[x(0)] + n\delta^2 = n\delta^2\end{aligned}$$

as $n = \frac{\text{time}}{\text{step-size}} = \frac{t}{\Delta x}$ it follows:

$$\sigma^2 = \mathbb{E}\left[x^2(n)\right] - \mathbb{E}[x(n)]^2 = \mathbb{E}\left[x^2(n)\right] = \frac{\delta^2}{\Delta x}t \qquad (30.53)$$

Thus in expectation the particles goes nowhere.

Proof 30.3: eq. (30.34):

$$\begin{aligned}\gamma(\epsilon * [k], \epsilon * [k+n]) &= \mathrm{Cov}\left[\epsilon * [k], \epsilon * [k+1]\right]\\ &= \mathbb{E}\left[(\epsilon * [k] - \mathbb{E}[\epsilon * [k]])\left(\epsilon * [k+n] - \mathbb{E}[\epsilon * [k+n]]\right)^{\intercal}\right]\\ &\overset{\text{eq. (30.31)}}{=} \mathbb{E}\left[(\epsilon * [k])(\epsilon * [k+n])\right]\end{aligned}$$

Proof 30.4: [cor. 30.4]:
Since $B_t - B_s$ is the increment over the interval $[s,t]$, it is the same in distribution as the incremeent over the interval $[s-s, t-s] = [0, t-s]$

$$\text{Thus} \qquad B_t - B_s \sim B_{t-s} - B_0$$

but as B_0 is a.s. zero by definition eq. (30.19) it follows:

$$B_t - B_s \sim B_{t-s} \qquad B_{t-s} \sim \mathcal{N}(0, t-s)$$

Proof 30.5: [cor. 30.4]:

$$W(t) = W(t) - \underbrace{W(0)}_{=0} \sim \mathcal{N}(0,t)$$
$$\Rightarrow \qquad \mathbb{E}[X] = 0 \qquad \mathbb{V}[X] = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = t$$

Proof 30.6: theorem 30.2:

$$\begin{aligned}&\sum_{k=0}^{N-1}\left[W(t_k) - W(t_{k-1})\right]^2 && t_k = k\frac{T}{N}\\ &= \sum_{k=0}^{N-1} X_k^2 && X_k \sim \mathcal{N}\left(0, \frac{T}{N}\right)\\ &= \sum_{k=0}^{N-1} Y_k = n\left(\frac{1}{n}\sum_{k=0}^{N-1} Y_k\right) && \mathbb{E}[Y_k] = \frac{T}{N}\\ &\overset{\text{S.L.L.N}}{=} n\frac{T}{n} = T\end{aligned}$$

Proof 30.7: theorem 30.3 ②:

1. first we need to show eq. (30.7): $\mathbb{E}[W_t|\mathcal{F}_s] = W_s$
Due to the fact that W_t is $\mathcal{F}_t$ measurable i.e. $W_t \in \mathcal{F}_t$ we know that:
$$\mathbb{E}[W_t|\mathcal{F}_t] = W_t \qquad (30.54)$$
$$\begin{aligned}\mathbb{E}[W_t|\mathcal{F}_s] &= \mathbb{E}[W_t - W_s + W_s|\mathcal{F}]\\ &= \mathbb{E}[W_t - W_s|\mathcal{F}_s] + \mathbb{E}[W_s|\mathcal{F}_s]\\ &\overset{\text{eq. (30.54)}}{=} \mathbb{E}[W_t - W_s] + W_s\\ &\overset{W_t - W_s \sim \mathcal{N}(0,t-s)}{=} W_s\end{aligned}$$
2. second we need to show eq. (30.8): $\mathbb{E}[|X(t)|] < \infty$
$$\mathbb{E}[|W(t)|]^2 \overset{??}{\leqslant} \mathbb{E}\left[|W(t)|^2\right] = \mathbb{E}\left[W^2(t)\right] = t \leqslant \infty$$

Proof 30.8: theorem 30.3 ③: $W_t^2 - t$ is a martingale?
Using the binomial formula we can write and adding $W_s - W_s$:

$$W_t^2 = (W_t - W_s)^2 + 2W_s(W_t - W_s) + W_s^2$$

using the expectation:

$$\begin{aligned}\mathbb{E}\left[W_t^2|\mathcal{F}_s\right] =&\mathbb{E}\left[(W_t - W_s)^2|\mathcal{F}_s\right] + \mathbb{E}\left[2W_s(W_t - W_s)|\mathcal{F}_s\right]\\ &+ \mathbb{E}\left[W_s^2|\mathcal{F}_s\right]\\ \overset{\text{eq. (30.54)}}{=}& \mathbb{E}\left[(W_t - W_s)^2\right] + 2W_s\mathbb{E}\left[(W_t - W_s)\right] + W_s^2\\ \overset{\text{eq. (30.21)}}{=}& \mathbb{V}[W_t - W_s] + 0 + W_s^2\\ & t - s + W_s^2\end{aligned}$$

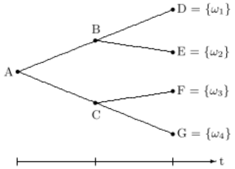
from this it follows that:

$$\mathbb{E}\left[W_t^2 - t|\mathcal{F}_s\right] = W_s^2 - s \qquad (30.55)$$

Examples

Example 30.1 :

Suppose we have a sample space of four elements: $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. At time zero, we do not have any information about which ω has been chosen. At time $T/2$ we know whether we have $\{\omega_1, \omega_2\}$ or $\{\omega_3, \omega_4\}$. At time T, we have full information.



$$\mathcal{F} = \begin{cases} \{\varnothing, \Omega\} & t \in [0, T/2) \\ \{\varnothing, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \Omega\} & t \in [T/2, T) \\ \mathcal{F}_{\max} = 2^{\Omega} & t = T \end{cases} \tag{30.56}$$

Thus, $\mathcal{F}_0$ represents initial information whereas $\mathcal{F}_\infty$ represents full information (all we will ever know). Hence, a stochastic process is said to be defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geqslant 0}, \mathbb{P})$.

Ito Calculus