Notes on Mathematics

Contents

1 Calculus

1.1 Differentation and Integration

Lemma 1.1 (Simple Calculations)**.**

1. *For* $1 = xx^{-1}$ *the product rule yields* $0 = x^{-1} + x(x^{-1})'$. *Hence*
$$\frac{d}{dx}x^{-1} = -\frac{1}{x^2}$$

2. *Similarly* $x = \sqrt{x}^2$ *and* $1 = 2\sqrt{x}\sqrt{x}'$ *and so*
$$\frac{d}{dx}\sqrt{x} = \frac{1}{2\sqrt{x}}$$

3. *It is*
$$\frac{d}{dx}x^n = nx^{n-1}$$
since via induction the product rule yields
$$\frac{d}{dx}x^n = \frac{d}{dx}xx^{n-1} = x^{n-1} + \frac{d}{dx}x^{n-1} = x^{n-1} + (n-1)x^{n-1} = nx^{n-1}$$

4. *Again, applying the product rule gives*
$$\left(\frac{1}{g}\right)' = \left(\frac{1}{x} \circ g\right)' = -\frac{g'}{g^2}$$
and the quotient rule
$$\left(\frac{f}{g}\right)' = \frac{f'}{g} + f\left(\frac{1}{g}\right)' = \frac{f'}{g} - \frac{fg'}{g^2} = \frac{gf' - fg'}{g^2}$$

5. *Also* $x = f \circ f^{-1}$ *and* $1 = (f^{-1})'f' \circ f^{-1}$. *Thus*
$$(f^{-1})' = \frac{1}{f' \circ f^{-1}}$$
where defined. Especially for $x \neq 0$
$$\log'(x) = \frac{1}{\exp'(\log(x))} = \frac{1}{x}$$

6. $(1-q)(1+q+q^2+\cdots+q^n) = 1-q+q-q^2+q^2-q^3+\cdots+q^{n+1}$ *gives*
$$\sum_{k=0}^{n} q^k = \frac{1-q^{n+1}}{1-q} \quad \textit{and} \quad \sum_{k=m}^{n} q^k = \frac{q^m-q^{n+1}}{1-q}$$

Lemma 1.2 (Exponential Function)**.**

1. *It is*
$$\exp(x+y) = \exp(x)\exp(y)$$
Hence
$$\exp(0) = 1$$
$$\exp(-x) = \exp(x)^{-1}$$
$$\exp(nx) = \exp(x)^n$$

2. *For the derivative*
$$\exp'(x) = \sum_{k=0}^{\infty} \frac{1}{k!}(x^k)' = \sum_{k=0}^{\infty} \frac{1}{k!} k x^{k-1} = \sum_{k=1}^{\infty} \frac{1}{(k-1)!} x^{k-1} = \exp(x)$$

Lemma 1.3 (Sinus and Cosinus)**.**

1. *Sinus and Cosinus power series*
$$\cos(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k!} x^{2k}$$
$$\sin(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} x^{2k+1}$$

2. *Symmetry*
$$\cos(-x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k!} (-x)^{2k} = \cos(x)$$
$$\sin(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} (-x)^{2k+1} = -\sin(x)$$

3. *Derivatives*
$$\cos'(x) = \sum_{k=1}^{\infty} \frac{(-1)^k}{(2k-1)!} x^{2k-1} = \sum_{k=0}^{\infty} \frac{(-1)^{k+1}}{(2k+1)!} x^{2k+1} = -\sin(x)$$
$$\sin'(x) = \sum_{k=0}^{\infty} \frac{(-1)^k}{2k!} x^{2k} = \cos(x)$$

Theorem 1.4 (Fermat Stationary Point)**.** *Let $\Omega \subseteq \mathbb{R}$ be open and $f \in C^1(\Omega)$. If $x^* \in \Omega$ is local extremum then $f'(x^*) = 0$.*

Proof. Assume x^* is the minimum of f in Ω and let $f(x^*) > 0$. Since $f \in C^1(\Omega)$ there exist $\varepsilon, \delta > 0$, so that for $|h| \leq \varepsilon$
$$\frac{f(x^* + h) - f(x^*)}{h} > \delta$$
Pick a negative $h \in [-\varepsilon, 0)$. Then
$$f(x^* + h) < f(x^*) + \delta h < f(x^*)$$
and x^* cannot be the minimum. Analog for maximum with a positive h, then apply to $-f$. □

Theorem 1.5 (Rolle)**.** *Let $f \in C[a,b]$ with $f(a) = f(b)$. If f is differentiable in (a,b) then there exists a $\xi \in (a,b)$ with $f'(\xi) = 0$.*

Proof. Assume f is not constant. Since $[a,b]$ is compact there exists either a global minimum or maximum $\xi \in (a,b)$ and Theorem 1.4 can be applied. $\square$

Theorem 1.6 (Mean Value)**.** *Let $f \in C[a,b]$ be differentiable in (a,b). Then there exists a $\xi \in (a,b)$ with*

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}$$

Proof. Apply Theorem 1.5 to

$$g(x) = f(x) - \frac{f(b) - f(a)}{b - a}(x - a)$$

$\square$

Remark 1.7.

1. *More generally choose any $\varphi \in C^1[a,b]$ with $\varphi(a) = 0$ and $\varphi(b) = f(b) - f(a)$. Set $g(x) = f(x) - \varphi(x)$ to see there is a $\xi \in (a,b)$ with $f'(\xi) = \varphi'(\xi)$.*
2. *Let f be differentiable in (a,b) with $f' = 0$. For $x, y \in (a,b)$*

$$0 = f'(\xi) = \frac{f(y) - f(x)}{y - x}$$

and f is a constant.

3. *Another useful generalization: let $\Omega \subseteq \mathbb{R}^n$ be open and $f \in C^1(\Omega)$. For $x, y \in \Omega$ define $\varphi(t) = f(tx + (1-t)y)$ and apply the chain rule for differentiation*

$$\varphi'(\xi) = \nabla f(\xi x + (1-\xi)y)^T (x - y) = f(x) - f(y)$$

4. *The Cauchy Schwarz inequality then yields*

$$\|f(x) - f(y)\| \leq \|\nabla f(\xi x + (1-\xi)y)\| \|(x - y)\|$$

Theorem 1.8 (Differentiation Theorem)**.** *Let $f \in C[a,b]$ and define*

$$F(x) = \int_a^x f(t)\, dt$$

Then $F \in C^1[a,b]$ with $F'(x) = f(x)$ for $x \in [a,b]$.

Proof. Applying the Mean Value Theorem of Integration gives

$$F(x+h) - F(x) = \int_x^{x+h} f(t)\, dt = f(\xi)h$$

for some $\xi \in (x, x+h)$. $\square$

Theorem 1.9 (Fundamental Theorem of Calculus)**.** *Let $F \in C^1[a,b]$ with $F' = f$ Then*

$$F(b) - F(a) = \int_a^b f(t)\,dt$$

Lemma 1.10 (Integration by Substitution)**.** *Let $I \subseteq \mathbb{R}$ be an interval and $f \in C(I)$. For $\varphi \in C([a,b], I)$ it follows*

$$\int_{\varphi(a)}^{\varphi(b)} f(x)\,dx = \int_a^b f(\varphi(t))\varphi'(t)\,dt$$

Proof. Let $F \in C^1(I)$ with $F' = f$. Then the chain rule for differentiation yields

$$\begin{aligned}\int_{\varphi(a)}^{\varphi(b)} f(x)\,dx &= F(\varphi(b)) - F(\varphi(a)) \\ &= F \circ \varphi(b) - F \circ \varphi(a) \\ &= \int_a^b (F \circ \varphi)'(t)\,dt \\ &= \int_a^b f(\varphi(t))\varphi'(t)\,dt\end{aligned}$$

□

Examples 1.11.

1. *For $\varphi(x) = x^2 + 1$ it is $\varphi(0) = 1$ and $\varphi(2) = 5$. Thus*

$$\int_0^2 x\cos(x^2+1)\,dx = \frac{1}{2}\int_0^2 2x\cos(x^2+1)\,dx = \frac{1}{2}\int_1^5 \cos(t)\,dt = \frac{1}{2}(\sin(5) - \sin(1))$$

2. *Consider $\varphi(x) = \sin(x)$ where $\varphi(0) = 0$ and $\varphi(\pi/2) = 1$. Since $\cos(t) = \sqrt{1 - \sin^2(t)}$ it follows*

$$\int_0^1 \sqrt{1-x^2}\,dx = \int_{\cos(0)}^{\cos(\pi/2)} \sqrt{1-x^2}\,dx = \int_0^{\pi/2} \sqrt{1-\sin^2(t)}\cos(t)\,dt = \int_0^{\pi/2} \cos^2(t)\,dt$$

3. *Let $f \in C[a,b]$ and $\varphi(x) = a + t(b-a)$. Then*

$$\int_a^b f(x)\,dx = (b-a)\int_0^1 f(a + t(b-a))\,dt$$

4. *Let $f(x) = x^n$ and $\varphi(x) = t^m$. As expected*

$$\int_0^1 x^n\,dx = \int_0^1 t^{nm} m t^{m-1}\,dt = m\int_0^1 t^{m(n+1)-1}\,dt = \left[\frac{m}{m(n+1)} t^{m(n+1)}\right]_0^1 = \frac{1}{n+1}$$

1.2 Directional Derivative and Gradients

Lemma 1.12 (Directional Derivative)**.** *Let $\Omega \subseteq \mathbb{R}^n$ be open and $f \in C^1(\Omega)$. Then*

$$\frac{\partial f}{\partial d}(x) = \nabla f(x)^T d$$

for any $d \in \mathbb{R}^n$.

Proof. Let $\varphi(t) = f(x + td)$. Then $\varphi \in C^1[-\varepsilon, \varepsilon]$ for some $\varepsilon > 0$ and the chain rule yields

$$\varphi'(t) = \nabla f(x + td)^T d$$

Hence

$$\varphi'(0) = \lim_{t \to 0} \frac{\varphi(x + td) - \varphi(0)}{t} = \lim_{t \to 0} \frac{f(x + td) - f(x)}{t} = \nabla f(x)^T d$$

□

Remarks 1.13.

1. *Note that by definition the directional derivative is invariant under multiplication with any $\lambda \neq 0$.*
2. *A similar proposition holds under the weaker assumption that d is a only feasable direction for f in x*
3. *For $d = \nabla f(x)/\|\nabla f(x)\|$ it follows that*

   $$\frac{\partial f}{\partial d}(x) = \|\nabla f(x)\| > 0$$

   and for any other $d \in \mathbb{R}^n$ with $\|d\| = 1$ the Cauchy Schwarz inequality yields

   $$|\frac{\partial f}{\partial d}(x)| = |\nabla f(x)^T d| \leq \|\nabla f(x)\| \|d\| = \|\nabla f(x)\|$$

   Hence $\nabla f(x)$ is the direction of the greatest ascent and respectively, $-\nabla f(x)$ is the direction of the greatest descent.

Theorem 1.14 (First Order Necessary Condition)**.** *Let $\Omega \subseteq \mathbb{R}^n$ be open and $f \in C^1(\Omega)$. If $x^* \in \Omega$ is a local minimzer then $\nabla f(x^*) = 0$.*

Proof. Let $h \in \mathbb{R}^n$ and $\delta > 0$ so that $x^* + th \in \Omega$ for all $t \in (-\delta, \delta)$. Then 0 is local minimizer for $\varphi(t) = f(x^* + th)$ and

$$\varphi'(0) = \nabla f(x^*)^T h = 0$$

Now let $h = \nabla f(x^*)$. □

Theorem 1.15 (Banach Fixed-Point Theorem)**.** *Let X be a Banach space and $f \in C(X, X)$ a contraction*

$$\|f(x) - f(y)\| \leq q\|x - y\| \text{ for all } x, y \in X$$

for some $0 < q < 1$. Then there exists a unique fix point $x^* \in X$ with

$$f(x^*) = x^*$$

Furthermore for any $x_0 \in X$ the sequence defined by

$$x_{n+1} = f(x_n)$$

converges aganist x^*.

Proof. Since $\|x_{n+1} - x_n\| = \|f(x_n) - f(x_{n-1})\| \leq q\|x_n - x_{n-1}\|$ it follows, that

$$\|x_{n+1} - x_n\| \leq q^n \|x_1 - x_0\|$$

Furthermore

$$\|x_n - x_m\| \leq \sum_{k=m}^{n} q^k \|x_1 - x_0\| = \frac{q^m - q^{n+1}}{1-q} \|x_1 - x_0\|$$

and (x_n) is a Cauchy sequence. For its limit x^* we have

$$x^* = \lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty} f(x_n) = f(x^*)$$

For any other $y^* \in X$ with $f(y^*) = y^*$ it follows, that

$$\|x^* - y^*\| = \|f(x^*) - f(y^*)\| \leq q\|x^* - y^*\|$$

and therefore $x^* = y^*$.

$\square$

2 Nonlinear Optimization

2.1 Minimization without Constraints

Lemma 2.1 (Gradient Inequality). *Let $M \subseteq \mathbb{R}^n$ be a convex set and $f \in C^1(M)$. Then f is convex if and only if*

$$f(x) \geq f(y) + \nabla f(y)^T (x - y)$$

for all $x, y \in M$.

Proof. Let f be convex and $x, y \in M$. For $0 \leq \lambda \leq 1$ we have

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) = \lambda f(x) - \lambda f(y) + f(y)$$

and

$$f(x) - f(y) \geq \frac{f(\lambda x + (1-\lambda)y) - f(y)}{\lambda} = \frac{f(y + \lambda(x-y)) - f(y)}{\lambda}$$

For $d = x - y$ and $\lambda \to 0$ the term on the right converges to the direction derivative of f in d

$$\frac{\partial f}{\partial d}(y) = \nabla f(y)^T d = \nabla f(y)^T (x - y)$$

Now let $x, y \in M$ and $0 \leq \lambda \leq 1$. For $z = \lambda x + (1-\lambda)y \in M$ it follows that

$$\begin{aligned} \lambda f(x) &\geq \lambda f(z) + \lambda \nabla f(z)^T (x - z) \\ (1-\lambda)f(y) &\geq (1-\lambda)f(z) + (1-\lambda)\nabla f(z)^T (y - z) \end{aligned}$$

Adding the two inequalities gives

$$\begin{aligned} \lambda f(x) + (1-\lambda)f(y) &\geq f(z) + \nabla f(z)^T (\lambda x - \lambda z + (1-\lambda)y - (1-\lambda)z) \\ &= f(z) + \nabla f(z)^T (\lambda x + (1-\lambda)y - z) \\ &= f(z) \end{aligned}$$

□

Exercise 2.2 (Facility Locations). *The facilities are located at:*

$$(3,0), (0,-3), (1,4)$$

Proof. Let

$$\begin{aligned} f(x) &= (x-3)^2 + y^2 + x^2 + (y+3)^2 + (x-1)^2 + (y-4)^2 \\ &= x^2 - 6x + 9 + y^2 + x^2 + y^2 + 6y + 9 + x^2 - 2x + 1 + y^2 - 8y + 16 \\ &= 3x^2 + 3y^2 - 8x - 2y + 35 \end{aligned}$$

Then

$$\nabla f(x,y) = (6x - 8, 6y - 2) \text{ and } \nabla^2 f(x,y) = \begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix} > 0$$

Hence $(4/3, 1/3)$ is the gobal minimum. □

Exercise 2.3 (Convex Functions). *The sum of convex functions is convex.*

Proof. Let $x, y \in M$. Since $\alpha_i > 0$ we have

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \sum_{i=1}^{m} \alpha_i f_i(\lambda x + (1-\lambda)y)) \\ &\leq \sum_{i=1}^{m} \alpha_i \lambda f_i(x) + \sum_{i=1}^{m} \alpha_i (1-\lambda) f_i(y) = \lambda f(x) + (1-\lambda) f(y) \end{aligned}$$

Let $f(x) = x^2$. Then $-f$ is not convex, e.g. $x = 1, y = -1$ and $\lambda = 0.5$.

Exercise 2.4 (Solution of Quadratic Inequality)**.** *Let*

$$f(x) = x^T A x + b^T x + c$$

Proof. The product rule gives

$$\nabla f(x) = x^T A + Ax + b = (A^T + A)x + b = 2Ax + b$$

Thus $\nabla^2 f(x) = 2A > 0$ and f is convex. Hence the level set Γ_{-c} is convex. Since the intersection of convex sets is convex $\Gamma_{-c} \cap \{x \in \mathbb{R}^n : g^T x + h = 0\}$ is convex, too.

Exercise 2.5 (Line Search on Compact Convex Sets)**.** *Let $S \subset \mathbb{R}^n$ be compact and convex. Furthermore let $f \in C^1(S)$ be convex, $x \in S$ and $d \in \mathbb{R}^n$ a descent direction of f in x with $\nabla f(x)^T d < 0$.*

Proof. If $x + \lambda^* d$ is an optimal solution then $\nabla f(x + \lambda^* d)^T d = 0$ according to Theorem 1.14. Let $\nabla f(x + \lambda^* d)^T d = 0$. Then Lemma 2.1 gives

$$f(x + \lambda d) \geq f(x + \lambda^* d) + (\lambda - \lambda^*) \nabla f(x + \lambda^* d)^T d = f(x + \lambda^* d)$$

and $x + \lambda^* d$ is an optimal solution.

Exercise 2.6 (Steepest Descent)**.** *Let*

$$f(x) = \frac{1}{2} x^T A x + b^T x + c$$

where A is symmetrical and positive definite.

Proof. Since $\nabla f(x) = Ax + b$ and $\nabla^2 f(x) = A > 0$ it follows $x^* = -A^{-1} b$. Let v be eigenvector with $Av = \mu v$. For $x_0 = x^* + \theta v$ we have

$$\nabla f(x_0) = Ax^* + \mu\theta v + b = \mu\theta v$$

and for $\lambda \geq 0$

$$\arg\min\{f(x_0 - \lambda \nabla f(x_0)\} = \arg\min\{f(x^* + \theta v - \lambda \mu \theta v)\} = \mu^{-1}$$

Thus

$$x_1 = x_0 - \mu^{-1} \nabla f(x_0) = x^* + \theta v - \mu^{-1} \mu \theta v = x^*$$

and $\nabla f(x_1) = 0$. Hence the algorithm stops after the first iteration. Now let

$$x_0 = x^* + \sum_{i=0}^{m} \theta_i v_i$$

for orthogonal eigenvectors with $Av_i = \mu_i$ and $m \leq n$. Then

$$\nabla f(x_0) = Ax^* + \sum_{i=0}^{m} \mu_i \theta_i v_i + b = \sum_{i=0}^{m} \mu_i \theta_i v_i$$

and

$$x_1 = x_0 - \lambda \sum_{i=0}^{m} \mu_i \theta_i v_i = x^* + \sum_{i=0}^{m} \theta_i v_i - \lambda \sum_{i=0}^{m} \mu_i \theta_i v_i = x^* + \sum_{i=0}^{m} (1 - \lambda \mu_i) \theta_i v_i$$

Since x^* is the minimum we have $\nabla f(x_1) = 0$ iff $\lambda = \mu^{-1}$ for all $0 \leq i \leq m$. □

2.2 One Dimensional Minimization and Direct Search

Definition 2.7 (Unimodal Function)**.** *A function $f : [a, b] \to \mathbb{R}$ is called unimodal if there exists a $\xi \in [a, b]$, so that f is strictly decreasing in $[a, \xi]$ and strictly increasing in $[\xi, b]$.*

In fact ξ is the unique minimum of f in $[a, b]$. According to the definition, for $a \leq x < y \leq b$ we have

$$f(x) > f(y) \text{ for } x, y \in [a, \xi) \text{ and } f(x) < f(y) \text{ for } x, y \in (\xi, b]$$

Thus

$$\xi \in [a, y] \text{ if } f(x) < f(y) \text{ and } \xi \in [x, b] \text{ if } f(x) \geq f(y)$$

Consider now a symmetrical partioning of the interval $[0, 1]$ where two consecutive partionings hold the same ratio respectively:

$$\sigma = 1 - \tau \text{ and } \frac{1}{\tau} = \frac{\tau}{\sigma}$$

Then $1 - \tau = \tau^2$ and solving the quadratic equation $\tau^2 + \tau = 1$ yields

$$\tau = \frac{\sqrt{5} - 1}{2} \approx 0.61803$$

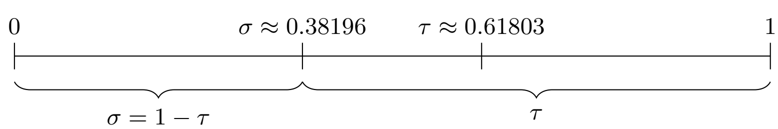

Figure 1: Golden Section

Let now $[a_0, b_0] = [a, b]$ and define

$$[a_{k+1}, b_{k+1}] = \begin{cases} [a_k, y_k] & \text{if } f(x_k) < f(y_k) \\ [x_k, b_k] & \text{if } f(x_k) \geq f(y_k) \end{cases}$$

where

$$x_k = b_k - \tau(b_k - a_k)$$
$$y_k = a_k + \tau(b_k - a_k)$$

It follows that $[a_k, b_k] \supset [a_{k+1}, b_{k+1}]$ is a decreasing series of intervals with

$$(b_{k+1} - a_{k+1}) = \tau(b_k - a_k)$$

where the interval converges to ξ. This leads to the following algorithm:

Algorithm 2.8 (Golden Section Search)**.**

```
"""Basic implementation of the golden section search, this easily can be
improved by storing and resuing the results of the previous iteration
"""

import math


def golden_section_search(f, I, eps=0.00001):
    t = 0.5 * (math.sqrt(5) - 1)
    a, b = I
    while abs(b - a) > eps:
        x, y = b - t * (b - a), a + t * (b - a)
        if f(x) > f(y):
            a = x
        else:
            b = y
    return (a + b) / 2


if __name__ == '__main__':
    p, q, I = 0, 0, (-10, 10)
    p, q, I = -4, 1, (-10, 10)
    f = lambda x: (x + p) ** 2 + q
    x0 = golden_section_search(f, I)
    print(f'arg min f on {I}: {x0}')
```

Algorithm 2.9 (Steepest Descent)**.**

Let $f \in C^1(\mathbb{R}^n)$ and $x_0 \in \mathbb{R}^n$. For $0 < \alpha \leq \beta < 1$ and $\gamma < 1$ let

Exercise 2.10 (Surprising Convergence)**.** *Example for $f \in C^2(\mathbb{R})$ with a sequence of strict local minima converging to a strict local maximum.*

Proof. Let $f \in C[a, b]$ and $\xi \in (a, b)$ so that f is strictly increasing in $(a, \xi]$ and strictly decreasing in $[\xi, b)$. Define

$$g(x) = \int_{\xi - x}^{\xi + x} f(t)\, dt$$

□

3 The Road to Reality

3.1 Hyperbolic Geometry

The ratio between the area A and A' of two similar shapes is given by

$$A' = k^2 A$$

Theorem 3.1 (Pythagoras).

$$a^2 + b^2 = c^2$$

Proof. Let A, B and C be the areas of the three triangles respectively. All triangles are similar, hence

$$B = \frac{b^2}{a^2} A \text{ and } C = \frac{c^2}{b^2} B$$

Since $A + B = C$ it follows that

$$a^2 + b^2 = \frac{b^2 A}{B} + b^2 = \frac{b^2 (A + B)}{B} = \frac{b^2 C}{B} = c^2$$

$\square$

Lemma 3.2 (label and Projective Representation). *The mapping from conformal and projective representation of any point is given by the radial expansion of the following factor*

$$\frac{2R}{R^2 + r^2}$$

Proof. For any point the distance from the origin with regard to the two representations is given by

$$\log \frac{R + r}{R - r} = \frac{1}{2} \log \frac{R + r'}{R - r'} = \log \frac{(R + r')^2}{(R - r')^2}$$

This gives

$$(R - r)^2 (R + r') = (R + r)^2 (R - r') \text{ and } -4R^2 r + 2R^2 r' + 2r^2 r' = 0$$

Hence

$$r' = \frac{2R^2}{R^2 + r^2} r$$

$\square$

3.2 Complex Numbers

Lemma 3.3 (Basic Formulas)**.**

1. *It is*
$$(a+\mathrm{i}b)(c+\mathrm{i}d) = (ac-bd)+\mathrm{i}(ad+bc)$$

2. *Thus*
$$(a+\mathrm{i}b)^2 = (a^2-b^2)+\mathrm{i}2ab$$
and
$$(a+\mathrm{i}b)(a-\mathrm{i}b) = a^2+\mathrm{i}ab-\mathrm{i}ab-\mathrm{i}^2b^2 = a^2+b^2$$

3. *Hence*
$$\frac{a+\mathrm{i}b}{c+\mathrm{i}d} = \frac{(a+\mathrm{i}b)(c-\mathrm{i}d)}{c^2+d^2} = \frac{ac+bd}{c^2+d^2}+\mathrm{i}\frac{bc-ad}{c^2+d^2}$$

4. *For*
$$z = \sqrt{\frac{1}{2}(a+\sqrt{a^2+b^2})}+\mathrm{i}\sqrt{\frac{1}{2}(-a+\sqrt{a^2+b^2})}$$
it follows
$$z^2 = \frac{1}{2}(a+\sqrt{a^2+b^2})-\frac{1}{2}(-a+\sqrt{a^2+b^2})+\mathrm{i}2\sqrt{\frac{1}{4}(\sqrt{a^2+b^2}^2)-a^2} = a+\mathrm{i}b$$

Lemma 3.4 (Binomial Theorem)**.**

1. *For the binomial coefficient Pascal's identity holds*
$$\binom{n}{k-1}+\binom{n}{k} = \binom{n+1}{k}$$

2. *The following equation states the binomial identity*
$$(a+b)^n = \sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k} = \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k$$

3. *For $a=1$ follows*
$$(1+x)^n = \sum_{k=0}^{n}\binom{n}{k}x^k$$

Proof. It is
$$\binom{n}{k}+\binom{n}{k-1} = \frac{n!}{k!(n-k)!}+\frac{n!}{(k-1)!(n-k+1)!} = \frac{n!(n+1-k)+n!k!}{k!(n+1-k)!} = \binom{n+1}{k}$$

Furthermore by using induction

$$\begin{aligned}(a+b)^{n+1} &= \sum_{k=0}^{n}\binom{n}{k}a^{k+1}b^{n-k} + \sum_{k=0}^{n}\binom{n}{k}a^k b^{n+1-k} \\ &= \sum_{k=1}^{n+1}\binom{n}{k-1}a^k b^{n+1-k} + \sum_{k=0}^{n}\binom{n}{k}a^k b^{n+1-k} \\ &= \sum_{k=0}^{n+1}\binom{n+1}{k}a^k b^{n+1-k}\end{aligned}$$

□

3.3 Exponential Function and Logarithms

Exercise 3.5 (Exponential Function)**.** *The Cauchy product yields*

$$\sum_{n=0}^{\infty} a_n \sum_{n=0}^{\infty} b_n = \sum_{n=0}^{\infty}\sum_{k=0}^{n} a_k b_{n-k}$$

if at least one of the series is absolutely convergent. Hence

$$\begin{aligned}\sum_{n=0}^{\infty}\frac{1}{n!}z^n \sum_{n=0}^{\infty}\frac{1}{n!}w^n &= \sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{1}{k!}z^k\frac{1}{(n-k)!}w^{n-k} \\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\sum_{k=0}^{n}\binom{n}{k}z^k w^{n-k} \\ &= \sum_{n=0}^{\infty}\frac{1}{n!}(z+w)^n\end{aligned}$$

Let $t \in \mathbb{R}$. Then

$$\begin{aligned}\mathrm{e}^{\mathrm{i}t} &= \sum_{k=0}^{\infty}\frac{1}{k!}(\mathrm{i}t)^k \\ &= \sum_{k=0}^{\infty}\frac{1}{2k!}(\mathrm{i}t)^{2k} + \sum_{k=0}^{\infty}\frac{1}{(2k+1)!}(\mathrm{i}t)^{2k+1} \\ &= \sum_{k=0}^{\infty}\frac{(-1)^k}{2k!}t^{2k} + \mathrm{i}\sum_{k=0}^{\infty}\frac{(-1)^k}{(2k+1)!}t^{2k+1} \\ &= \cos t + \mathrm{i}\sin t\end{aligned}$$

More generally for $z = \log r + \mathrm{i}t$

$$\mathrm{e}^z = \mathrm{e}^{\log r + \mathrm{i}t} = r\mathrm{e}^{\mathrm{i}t} = r(\cos t + \mathrm{i}\sin t)$$

For $r = 1$ and $t = 2\pi$ this yields

$$\mathrm{e}^{2\pi\mathrm{i}} = \cos 2\pi + \mathrm{i}\sin 2\pi = 1$$

and for $t = 2\pi$ we get

Lemma 3.6 (Euler Equation)**.**

$$\mathrm{e}^{\pi \mathrm{i}} + 1 = 0$$

Exercise 3.7.

1. *If* $\mathrm{e}^z = w$ *then* $z + \pi\mathrm{i}$ *is a logarithm to* $-w$*:* $\mathrm{e}^{z+\pi\mathrm{i}} = \mathrm{e}^z\mathrm{e}^{\pi\mathrm{i}} = -\mathrm{e}^z = -w$.
2. *Since* $\mathrm{e}^{i(s+t)} = \mathrm{e}^{is}\mathrm{e}^{it}$ *it follows*

$$\begin{aligned}\cos(s+t) + \mathrm{i}\sin(s+t) &= (\cos s + \mathrm{i}\sin s)(\cos t + \mathrm{i}\sin t)\\ &= \cos s\cos t - \sin s \sin t + \mathrm{i}(\cos s \sin t + \sin s \cos t)\end{aligned}$$

Hence

$$\begin{aligned}\cos(s+t) &= \cos s\cos t - \sin s\sin t\\ \sin(s+t) &= \cos s\sin t + \sin s\cos t\end{aligned}$$

3. *It is* $\mathrm{e}^{3it} = (\mathrm{e}^{it})^3$ *and thus*

$$\cos 3t + \mathrm{i}\sin 3t = (\cos t + \mathrm{i}\sin t)^3 = \cos^3 t - 3\cos t\sin^2(t) + \mathrm{i}(\cos^2 t\sin t - \sin^3 t)$$

4. *Fun facts*

$$\mathrm{e}^{1-4\pi^2} = \mathrm{e}^{1+(2\mathrm{i}\pi)^2} = \mathrm{e}\mathrm{e}^{2\pi\mathrm{i}}\mathrm{e}^{2\pi\mathrm{i}} = \mathrm{e}$$

and $\mathrm{i} = \mathrm{e}^{\mathrm{i}\pi/2}$ *gives*

$$\mathrm{i}^{\mathrm{i}} = \mathrm{e}^{\mathrm{i}\log\mathrm{i}} = \mathrm{e}^{\mathrm{i}\mathrm{i}\pi/2} = \mathrm{e}^{-\pi/2} \in \mathbb{R}$$

3.4 Complex Analysis

Definition 3.8 (holomorphic Function)**.** *Let* $\Omega \subseteq \mathbb{C}$ *be open. A function* $f : \Omega \to \mathbb{C}$ *is called* differentiable *at* $z \in \Omega$ *if the limit*

$$f'(z) = \lim_{h\to 0} \frac{f(z+h) - f(z)}{h}$$

exists. f *is called* holomorphic *on* Ω *if* f *is complex differentiable at all points of* Ω *and* $f' : \Omega \to \mathbb{C}$ *is called the* derivative *of* f.

Remarks 3.9.

1. f *is differentiable at* $z_0 \in \Omega$ *iff the limit*

$$f'(z_0) = \lim_{z\to z_0} \frac{f(z) - f(z_0)}{z - z_0}$$

exists

2. *If* f *is differentiable at* $z_0 \in \Omega$ *and* $\varepsilon > 0$ *then there exists a small enough environment of* z_0*, so that*

$$|f(z) - f(z_0) - f'(z_0)(z - z_0)| < \varepsilon|z - z_0|$$

Theorem 3.10 (Cauchy Riemann Equations)**.** *Let $f = u + \mathrm{i}v$ be holomorphic. Then f satisfies the Cauchy Riemann equations*

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}$$
$$\frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

Proof. For $h \in \mathbb{R}$ follows

$$\lim_{h \to 0} \frac{f(z+h) - f(z)}{h} = \frac{\partial u}{\partial x}(z) + \mathrm{i}\frac{\partial v}{\partial x}(z)$$

and

$$\lim_{h \to 0} \frac{f(z+\mathrm{i}h) - f(z)}{\mathrm{i}h} = \frac{\partial u}{\mathrm{i}\partial y}(z) + \frac{\partial v}{\partial y}(z) = \frac{\partial v}{\partial y}(z) - \mathrm{i}\frac{\partial u}{\partial y}(z)$$

□

Examples 3.11.

1. *Let $f(z) = z^3$. Then $u(x,y) + \mathrm{i}v(x,y) = x^3 - 3xy^2 + \mathrm{i}(3x^2y - y^3)$ and as expected*

$$\frac{\partial u}{\partial x}(x,y) = x^3 - 3y^2 \quad \text{and} \quad \frac{\partial u}{\partial y}(x,y) = -6xy$$
$$\frac{\partial v}{\partial x}(x,y) = 6xy \quad \text{and} \quad \frac{\partial v}{\partial y}(x,y) = x^3 - 3y^2$$

Lemma 3.12. *Let $D \subseteq \mathbb{C}$ be connected. For arbitrary $z, w \in D$ there exists a polygonal path from z to w.*

Proof. For any path from z to w the image is compact, which can be used to define a finite subcover of disks. Use the center points to define the polygonal path. □

Lemma 3.13. *Let $\gamma : [a,b] \to \mathbb{C}$ a smooth path, $\psi : [c,d] \to [a,b]$ a smooth and increasing bijection and f continous.*

$$\int_\gamma f(z)\,dz = \int_{\gamma\circ\psi} f(z)\,dz$$

Proof. It is

$$\begin{aligned}\int_{\gamma\circ\psi} f(z)\,dz &= \int_c^d f(\gamma\circ\psi(t))(\gamma\circ\psi)'(t))\,dt \\ &= \int_{\psi(a)}^{\psi(b)} f(\gamma(\psi(t))\gamma'(\psi(t))\psi'(t)\,dt \\ &= \int_a^b f(\gamma(s))\gamma'(s)\,ds = \int_\gamma f(z)\,dz\end{aligned}$$

□

Lemma 3.14. *For a smooth path $\gamma : [a, b] \to \mathbb{C}$ define $-\gamma(t) = a + b - t$. Then*

$$\int_{-\gamma} f(z)\, dz = -\int_{\gamma} f(z)\, dz$$

Proof. Using integration by substitution

$$\int_{-\gamma} f(z)\, dz = -\int_a^b f(\gamma(a + b - t))\gamma'(a + b - t))\, dt = \int_b^a f(\gamma(s))\gamma'(s)\, ds = -\int_{\gamma} f(z)\, dz$$

□

In order to use the results from real calculus recall the fact, that for every $z \in \mathbb{C}$ there exists a $t \in [0, 2\pi]$, so that $z = |z|e^{\mathrm{i}t}$ and hence $|z| = ze^{-\mathrm{i}t}$.

Lemma 3.15. *Let $f \in C[a, b]$. Then*

$$\left|\int_a^b f(x)\, dx\right| \leq \int_a^b |f(x)|\, dx$$

Proof. Using the estimation for integrals from real calculus

$$\left|\int_a^b f(x)\, dx\right| = e^{-\mathrm{i}t} \int_a^b f(x)\, dx \leq \int_a^b |e^{-\mathrm{i}t} f(x)|\, dx = \int_a^b |f(x)|\, dx$$

□

Let $\gamma : [a, b] \to \mathbb{C}$ be a smooth path and $a = t_0 < t_1 < \cdots < t_n = b$ a partioning of $[a, b]$. Then

$$\sum_{k=1}^{n} |\gamma(t_k) - \gamma(t_{k-1})| = \sum_{k=1}^{n} \left|\frac{\gamma(t_k) - \gamma(t_{k-1})}{t_k - t_{k-1}}\right| (t_k - t_{k-1}) = \sum_{k=1}^{n} |\gamma'(\xi_k)|(t_k - t_{k-1})$$

yields a reasonable approximation of the length of the path. Hence

Definition 3.16. *For a smooth path $\gamma : [a, b] \to \mathbb{C}$*

$$L(\gamma) = \int_a^b |\gamma'(t)|\, dt$$

is called the length of γ.

Lemma 3.17 (Estimation Lemma)**.** *Let $\gamma : [a, b] \to \mathbb{C}$ be a smooth path. Then*

$$\left|\int_{\gamma} f(z)\, dz\right| \leq L(\gamma) \max_{\gamma[a.b]} f$$

Proof. Using the definition above

$$\left|\int_\gamma f(z)\,dz\right| = \left|\int_a^b f(\gamma(t))\gamma'(t))\,dt\right| \leq \int_a^b |f(\gamma(t))\gamma'(t))|\,dt \leq \max_{\gamma[a.b]} f \int_a^b |\gamma'(t)|\,dt$$

□

Examples 3.18.

1. *Let* $\gamma(t) = t + \mathrm{i}t$. *Then*

$$\int_\gamma z^2\,dz = \int_0^1 (t+\mathrm{i}t)^2(1+\mathrm{i})\,dt = (1+\mathrm{i})\int_0^1 2\mathrm{i}t^2\,dt = \left[(-2+2\mathrm{i})t^2\right]_0^1 = -\frac{2}{3} + \mathrm{i}\frac{2}{3}$$

2. *For* $\gamma(t) = t^2 + \mathrm{i}t$

$$\begin{aligned}\int_\gamma z^2\,dz &= \int_0^1 (t^2+\mathrm{i}t)^2(2+\mathrm{i}t)\,dt = \int_0^1 (2t^5-4t^3) + \mathrm{i}(5t^4 - t^2t)\,dt \\ &= \left[\frac{1}{3}t^6 - t^4\right]_0^1 + \mathrm{i}\left[t^5 - \frac{1}{3}t^3\right]_0^1 = -\frac{2}{3} + \mathrm{i}\frac{2}{3}\end{aligned}$$

3. *And* $\gamma(t) = \mathrm{i} + \mathrm{e}^{\mathrm{i}t}$

$$\begin{aligned}\int_\gamma z^2\,dz &= \int_{3/2\pi}^{2\pi} (\mathrm{i}+\mathrm{e}^{\mathrm{i}t})^2\mathrm{i}\mathrm{e}^{\mathrm{i}t}\,dt = \int_{3/2\pi}^{2\pi} (-1+2\mathrm{i}\mathrm{e}^{\mathrm{i}t}+\mathrm{e}^{2\mathrm{i}t})\mathrm{i}\mathrm{e}^{\mathrm{i}t}\,dt \\ &= \int_{3/2\pi}^{2\pi} -\mathrm{i}\mathrm{e}^{\mathrm{i}t} - 2\mathrm{e}^{2\mathrm{i}t} + \mathrm{i}\mathrm{e}^{3\mathrm{i}t}\,dt = \left[(-\mathrm{e}^{\mathrm{i}t} + \mathrm{i}\mathrm{e}^{2\mathrm{i}t} + \frac{1}{3}\mathrm{e}^{3\mathrm{i}t}\right]_{3/2\pi}^{2\pi} \\ &= \left(-1+\mathrm{i}+\frac{1}{3}\right) - \left(\mathrm{i}-\mathrm{i}+\frac{1}{3}\mathrm{i}\right) = -\frac{2}{3} + \mathrm{i}\frac{2}{3}\end{aligned}$$

4. *Let* $\gamma(t) = \mathrm{e}^{\mathrm{i}t}$ *and* $k \neq -1$. *Then*

$$\int_\gamma z^k\,dz = \int_0^{2\pi} \mathrm{e}^{\mathrm{i}kt}\mathrm{i}\mathrm{e}^{\mathrm{i}t}\,dt = \int_0^{2\pi} \mathrm{i}\mathrm{e}^{\mathrm{i}(k+1)t}\,dt = 0$$

Theorem 3.19. *Let* $D \subseteq \mathbb{C}$ *be a connected domain and* $f \in C(D)$. *Then the following assertions are equivalent*

1. *f has an antiderivative*
2. *For every closd path* γ

$$\int_\gamma f(z)\,dz = 0$$

Proof. Let $F' = f$. Since γ is closed

$$\int_\gamma f(z)\,dz = \int_a^b f(\gamma(t))\gamma'(t)\,dt = \int_a^b (F\circ\gamma)'(t)\,dt = F(\gamma(b)) - F(\gamma(a)) = 0$$

Now fix some arbitrary $a \in D$. For $z \in D$ let γ_z be a path from a to z and define

$$F(z) = \int_{\gamma_z} f(\zeta)\,d\zeta$$

This is well defined since the integral of f vanishes over each closed path. Moreover, since $\gamma_{z+h}+[z+h,z]-\gamma_z$ defines a closed path

$$F(z+h) - F(z) = \int_{\gamma_{z+h}} f(z)\,dz - \int_{\gamma_z} f(z)\,dz = \int_{[z,z+h]} f(z)\,dz = h\int_0^1 f(z+th)\,dt$$

Here the latter integral is continous at 0 with respect to h

$$\left|\int_0^1 f(z+th) - f(z)\,dt\right| \le \int_0^1 |f(z+th) - f(z)|\,dt \le \max_{t\in[0,1]} |f(z+th) - f(z)|$$

□

Corollary 3.20. *The second assertion can be weakend to*

$$\int_{\partial\Delta} f(z)\,dz = 0$$

for every triangle $\Delta \subset D$, where e.g. D is convex or star shaped. Here the antiderivative can directly be defined as

$$F(z) = \int_{[a,z]} f(\zeta)\,d\zeta$$

similar to the real calculus approach. Note, that under this conditions f always has a local antiderivative.

Examples 3.21.

1. *Let $z_0 \in \mathbb{C}$ and $\gamma(t) = z_0 + \mathrm{e}^{\mathrm{i}t}$ for $t \in [0, 2\pi]$. Then*

$$\int_\gamma \frac{1}{z - z_0}\,dz = \int_0^{2\pi} \frac{\mathrm{i}\mathrm{e}^{\mathrm{i}t}}{z_0 + \mathrm{e}^{\mathrm{i}t} - z_0}\,dt = \int_0^{2\pi} \mathrm{i}\,dt = 2\pi\mathrm{i}$$

   and thus $1/(z - z_0)$ has no antiderivative on $\mathbb{C}\setminus\{z_0\}$

2. *Let $z_0 \in \mathbb{C}$ and $z \in D = D_r(z_0)$. Applying Theorem 3.19. to ∂D and a small enough circle around z gives*

$$\int_{\partial D} \frac{1}{\zeta - z}\,d\zeta = \int_{\partial D} \frac{1}{\zeta - z_0}\,d\zeta = 2\pi\mathrm{i}$$

Theorem 3.22 (Goursat)**.** *Let $\Omega \subseteq \mathbb{C}$ be open and f holomorphic on Ω. Then*

$$\int_{\partial\triangle} f(z)\,dz = 0$$

for every triangle $\triangle \subset \Omega$.

Proof. Choose a sequence of triangles $\triangle \supset \triangle_0 \supset \triangle_1 \cdots \supset \triangle_k$ as depicted. Since all the triangles are compact with a vainishing diameter there exists a unique $z_0 \in \Omega$ with $\bigcap \triangle_k = \{z_0\}$. Thus

$$\left|\int_{\partial\triangle} f(z)\,dz\right| \leq 4^k \left|\int_{\partial\triangle_k} f(z)\,dz\right| = 4^k \left|\int_{\partial\triangle_k} f(z) - f(z_0) - f'(z_0)(z-z_0)\,dz\right|$$

Furthermore $L(\partial\triangle) = 2^{-k} L(\partial\triangle_k)$ and

$$|z - z_0| < L(\partial\triangle_k) = 2^{-k} L(\partial\triangle)$$

for any $z \in \triangle_k$. Since f is holomorphic at z_0 for any given $\varepsilon > 0$ there exists a sufficiently large enough k, so that

$$\begin{aligned}\left|\int_{\partial\triangle} f(z)\,dz\right| &\leq 4^k L(\partial\triangle_k) \max_{z\in\triangle_k} |f(z) - f(z_0) - f'(z_0)(z-z_0)| \\ &\leq 4^k L(\partial\triangle_k)\,\varepsilon \max_{z\in\triangle_k} |z - z_0| \\ &\leq L(\partial\triangle)^2\,\varepsilon\end{aligned}$$

□

Corollary 3.23.

1. *A holomorphic function always has a local antiderivative*
2. *A holomorphic function on a star shaped domain has a global antiderivative and*

   $$\int_\gamma f(z)\,dz = 0$$

   for any closed path
3. *The prerequites of Goursat theorem can be weakened to continous and holomorphic with the exception of a finite number of points: adequate partioning of the original triangle*

Theorem 3.24 (Cauchy's Intergral Formula)**.** *Let $\Omega \subseteq \mathbb{C}$ be open and f holomorphic on Ω. Further let $D \subset \Omega$ be a disc. Then*

$$f(z) = \frac{1}{2\pi\mathrm{i}} \int_{\partial D} \frac{f(\zeta)}{\zeta - z}\,dz$$

for $z \in D$.

Proof. For $z \in D$ define

$$h(\zeta) = \frac{f(\zeta) - f(z)}{\zeta - z}$$

for $\zeta \neq z$ and $f'(z)$ for $\zeta = z$. Then h is holomorphic on $D \setminus \{z\}$ and continous at z

$$0 = \int_{\partial D} h(\zeta)\,d\zeta = \int_{\partial D} \frac{f(\zeta)}{\zeta - z}\,d\zeta - f(z)\int_{\partial D} \frac{1}{\zeta - z}\,d\zeta = \int_{\partial D} \frac{f(\zeta)}{\zeta - z}\,d\zeta - 2\pi\mathrm{i} f(z)$$

□

4 Neural Networks

4.1 The Perceptron

Definition 4.1 (Binary Classifiers)**.** *Let $X \subset \mathbb{R}^n$ be the union of two disjoint finite sets $X = M \cup N$.*

1. *A* binary classification problem *is the task to find a mapping $f : X \to \{0,1\}$ with*

$$f(x) = \begin{cases} 1 & \text{for } x \in M \\ 0 & \text{for } x \in N \end{cases}$$

   f then is called a binary classifier *for X*

2. *X is called* separable *if there exists a* weight vector *$w \in \mathbb{R}^n$ and a* bias *$b \in \mathbb{R}$ so that*

$$wx + b > 0 \text{ for } x \in M$$
$$wx + b < 0 \text{ for } x \in N$$

3. *The weight w and the bias b are called* solution to the classification problem. *They implicitely define a binary classifier via*

$$f(x) = \begin{cases} 1 & \text{if } wx + b > 0 \\ 0 & \text{if } wx + b < 0 \end{cases}$$

Examples 4.2.

1. *Let $X = \{0,1\} \times \{0,1\}$ and consider the* and *operator $f(1,1) = 1$ and $f(x,y) = 0$ elsewhere. Then $w = (3,3)$ and $b = -5$ yield a solution to the classification problem $M = f^{-1}(1)$ and $N = f^{-1}(0)$*

2. *Again let $X = \{0,1\} \times \{0,1\}$ and $f(1,0) = f(0,1) = 1$ and $f(0,0) = f(1,1) = 0$, the* xor *operator. Thus for any weight (w_1, w_2) and any bias b*

$$w_1 + b > 0$$
$$w_2 + b > 0$$

$$w_1 + w_2 + b \leq 0$$
$$b \leq 0$$

   Adding two equations respectively shows that there is no solution

3. *The bias can be integrated into the weight vector via $w' = (w,b) \in \mathbb{R}^{n+1}$ and $x' = (x,1) \in \mathbb{R}^{n+1}$. Separability then reduces to*

$$w'x' > 0$$

Geometrical Interpretation

The idea for the perceptron most likely has its origin in a simple geometrical observation. Recall that for $x, y \in \mathbb{R}^n$ the dot product can be expressed as

$$xy = \|x\|\|y\| \cos \alpha$$

where α is the angle between the two vectors. Hence the product is positive if the angle is less than 90° degrees and negative if the angle is between 90° and 180° degrees

$$xy > 0 \text{ for } 0 \leq \alpha < \pi/2 \text{ and } xy < 0 \text{ for } \pi/2 < \alpha \leq \pi$$

Note, that the sign does not dependend on lengths of the vectors, but solely on the angle.

For any two vectors it is easy enough to find a weight that satisfies $wx > 0$ and $wy > 0$. Generally $w = x + y$ is a good guess, but not always correct as shown below in Figure 2.

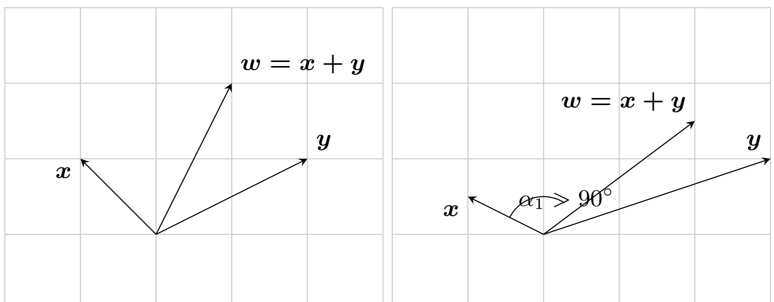

Figure 2: Dot Product and Angle

But, the more similar the lengths of the two vectors are the more likely $x + y$ works. Actually the threshold is tbd. An iterative approach is to increase $w = x + y$ in the direction of vector with the angle greater than 90° degrees.

Examples 4.3.

1. *Let* $x = (-1, 1)$ *and* $y = (6, 1)$*. Then*

$$\begin{array}{lll} w_0 = (5, 2) & w_0 x = -3 < 0 & w_0 y = 32 > 0 \\ w_1 = (4, 3) & w_1 x = -1 < 0 & w_1 y = 27 > 0 \\ w_2 = (3, 4) & w_2 x = 1 > 0 & w_2 y = 22 > 0 \end{array}$$

2. *Let* $x = (4, -6)$ *and* $y = (-10, 5)$*. Then*

$$\begin{array}{lll} w_0 = (-6, -1) & w_0 x = -18 < 0 & w_0 y = 55 > 0 \\ w_1 = (-2, -7) & w_1 x = 34 > 0 & w_1 y = -15 < 0 \\ w_2 = (-12, -2) & w_2 x = -36 < 0 & w_2 y = 110 > 0 \\ w_3 = (-8, -8) & w_3 x = 16 > 0 & w_3 y = 40 > 0 \end{array}$$

Algorithm 4.4 (Weight)**.**

```
import math
from random import randint


def add(v, w):
    return tuple(x + y for x, y in zip(v, w))


def crossprod(v, w):
    return sum(x * y for x, y in zip(v, w))


def norm(v):
    return math.sqrt(crossprod(v, v))


def angle(v, w):
    return math.degrees(math.acos(crossprod(v, w) / (norm(v) * norm(w))))


def weight(x, y):
    w = add(x, y)
    while True:
        if crossprod(w, x) < 0:
            w = add(w, x)
        elif crossprod(w, y) < 0:
            w = add(w, y)
        else:
            return w


def test_weight(x, y):
    print(f'x = {x}, y = {y}')
    i, w = 0, add(x, y)
    while True:
        a, b = crossprod(w, x), crossprod(w, y)
        print(f'    w{i} = {w}, w{i} * x = {a}, w{i} * y = {b}')
        if a < 0:
            w = add(w, x)
        elif b < 0:
            w = add(w, y)
        else:
            break
        i = i + 1


if __name__ == '__main__':
    n = 10
    for _ in range(3):
        x, y = (randint(-n, n), randint(-n, n)), (randint(-n, n), randint(-n, n))
        test_weight(x, y)
```

Algorithm 4.5 (Perceptron)**.**

```
def heaviside(x):
    return 0 if x < 0.0 else 1
```

```python
def crossprod(v, w):
    return sum(x * y for x, y in zip(v, w))


def perceptron(f, M, t=0.1, max_iterations=50):
    n = len(M[0])
    w, b = n * [0.0], 0.0
    for _ in range(max_iterations):
        done = True
        for x, y in M:
            z = heaviside(crossprod(w, x) + b)
            w = [w[i] - t * (z - y) * x[i] for i in range(n)]
            b = b - t * (z - y)
            if not y == z:
                done = False
        if done:
            break
    return lambda x: heaviside(crossprod(w, x) + b)


def and_(x, y):
    return int(x and y)


def or_(x, y):
    return int(x or y)


def xor(x, y):
    return int(x and not y or not x and y)


def nand(x, y):
    return int(not x and not y)


def test_perceptron(f, X):
    M = [(x, f(*x)) for x in X]
    p = perceptron(f, M)
    for x, y in M:
        print(f'{x} -> {p(x)}: {y}')
    print(f'{f.__name__}: {all(p(x) == y for x, y in M)}')


if __name__ == '__main__':
    X = [(0, 0), (0, 1), (1, 0), (1, 1)]
    test_perceptron(or_, X)
    test_perceptron(and_, X)
    test_perceptron(nand, X)
    test_perceptron(xor, X)
```

4.2 The Backtracking Algorithm

Definition 4.6 (Activation Functions).

1. *The* Heaviside *function* $H : \mathbb{R} \to \{0, 1\}$ *is defined as*

$$H(x) = \begin{cases} 1 & \textit{for } x > 0 \\ 0 & \textit{for } x \leq 0 \end{cases}$$

2. *The* sigmoid *function* $\sigma \in C^{\infty}(\mathbb{R})$ *is defined as*

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

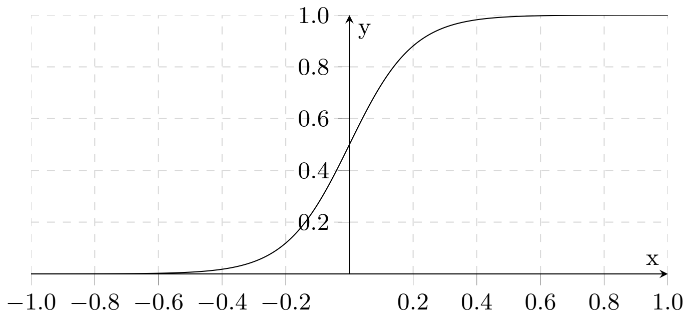

Figure 3: The sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$

Remarks 4.7.

1. *Since the heaviside function is not continous and therefore not differentiable at* 0 *the sigmoid function is often considered its smooth counterpart*

2. *The definition of the sigmoid function yields* $0 < \sigma(x) < 1$ *as well as* $\sigma(x) \to 0$ *for* $x \to -\infty$ *and* $\sigma(x) \to 1$ *for* $x \to \infty$

3. *The quotient rule yields*

$$\sigma'(x) = -\frac{-e^{-x}}{(1 + e^{-x})^2} = \sigma(x)\frac{1 + e^{-x} - 1}{1 + e^{-x}} = \sigma(x)(1 - \sigma(x))$$

and σ *is monotonically increasing over its domain*