Module overview

Concepts from mathematical analysis are introduced and reinterpreted as data analysis procedures, with a particular focus on the structure of the real numbers.

- Number systems: $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{Q}_p$
- Rigid body motion, data sets
- Approximation in $\mathbb{R}$: interpolation, least-squares, min-max
- Approximation in $\mathbb{R}^d$, $d \in \mathbb{N}$
- Data transformations: interpolation, integration, differentiation
- Data reduction: least-squares, min-max
- Data representation: functional analysis

$\rightarrow$ Set theory definition (Zermelo-Fraenkel) of $\mathbb{N}$: $0 = \{\} = \emptyset$, $1 = \{0\} = \{\emptyset\}$, $2 = \{0, 1\} = \{\emptyset, \{\emptyset\}\}, 3 = \{0, 1, 2\} = \{\emptyset, \{\emptyset, \{\emptyset\}\}\}, \ldots$

$\rightarrow$ Peano axiomatic definition of $(\mathbb{N}, S(n), =)$:

1. $0 \in \mathbb{N}$
2. $\forall n \in \mathbb{N}, n = n$, reflexive
3. $\forall m, n \in \mathbb{N}, m = n \Rightarrow n = m$, symmetric
4. $\forall m, n, p \in \mathbb{N}, m = n \wedge n = p \Rightarrow m = p$, transitive
5. $\forall a, b,\ a = b \wedge b \in \mathbb{N} \Rightarrow a \in \mathbb{N}$
6. $\forall n \in \mathbb{N}, S(n) \in \mathbb{N}$
7. $\forall m, n \in \mathbb{N}, m = n \Leftrightarrow S(m) = S(n)$, injective
8. $\forall n \in \mathbb{N}, S(n) \neq 0$
9. $0 \in K \wedge (\forall n \in \mathbb{N} \wedge n \in K \Rightarrow S(n) \in K) \Rightarrow \mathbb{N} \subseteq K$, induction

Introduce a shorthand notation for n repeated compositions of successor function

$$\text{Addition:}\, \forall m,n \in \mathbb{N},\ m+n = \underbrace{S(S(\ldots S(m)))}_{n\,\text{compositions}} = \underbrace{S \circ S \circ \ldots \circ S(m)}_{n\,\text{compositions}}$$

Introduce the inverse of the successor function $P(n) = S^{-1}(n)$.

Note that $\mathbb{N}$ is closed under S, (axiom 6), $n \in \mathbb{N} \Rightarrow S(n) \in \mathbb{N}$, but not under $P(n)$ due to axiom 8. Similar to addition, repeated composition of P receives a shorthand notation

$$\text{Subtraction:}\, \forall m,n \in \mathbb{N},\ m-n = \underbrace{P(P(\ldots P(m)))}_{n\,\text{compositions}} = \underbrace{P \circ P \circ \ldots \circ P(m)}_{n\,\text{compositions}}$$

The problem that now arises is to define the set under which P is closed. The closure of $\mathbb{N}$ under P (subtraction) is $\mathbb{Z}$, the set of integers.

- An essential mathematical concept is to establish that seemingly different objects, e.g., sets A, B are actually instances of a single category. The standard mathematical technique is to establish an isomorphism (one-to-one mappings) between elements of A, B
- Questions:
  - Are the even naturals the same as the naturals? Yes:

$$E = \{2n, n \in \mathbb{N}\}$$

  - Are there as many integers as naturals? (Equivalently, are the integers countable?) Yes:

$$g(n) = \left[\frac{n}{2}\right](-1)^n, n \in \mathbb{N}$$

with $[x] = \mathrm{ceil}(n)$

The naturals embody the idea of counting forward, the integers that of counting forward and backward. Besides counting, data analysis often requires comparison.

Consider $\mathbb{Z} \times \mathbb{Z} \backslash \{0\}$. The rationals $\mathbb{Q}$ are the equivalence classes with respect to $=$ within $\mathbb{Z} \times \mathbb{Z} \backslash \{0\}$. The rationals are countable

$$
\begin{array}{c|ccccc}
p \backslash q & 0 & 1 & 2 & 3 & \ldots \\
\hline
1 & \frac{0}{1} & \frac{1}{1} & \frac{2}{1} & \frac{3}{1} & \ldots \\
2 & \frac{0}{2} & \frac{1}{2} & \frac{2}{2} & \frac{3}{2} & \\
3 & \frac{0}{3} & \frac{1}{3} & \frac{2}{3} & \frac{3}{3} & \\
\vdots & & & & & \ddots
\end{array}
$$

Addition, division embody the concept of counting $\frac{p}{q} + \frac{m}{n}$ and comparison $\frac{p}{q} / \frac{m}{n}$ of rationals. $(\mathbb{Q}, +, \times)$ forms an algebraic structure known as a *field*.

Similar to definition of $\mathbb{N}$, there exists an axiomatic definition of $\mathbb{R}$, but the approach presented here is based on extending the idea of completion.

A 1-to-1 mapping of $\mathbb{N}$ to a subset of rationals is a *sequence* $a_n = \frac{p_n}{q_n}, n \in \mathbb{N}, p_n \in \mathbb{Z}$, $q_n \in \mathbb{N} \backslash \{0\}$. Sequences are different paths in the table $\mathbb{Z} \times \mathbb{N} \backslash \{0\}$.

A *Cauchy sequence* is whose elements become arbitrarily "closer". Define "closeness" of a_n, a_{n+p} by subtraction and absolute value, $|a_{n+p} - a_n|$. A Cauchy sequence satisfies: $\forall \varepsilon > 0$, $\exists N_\varepsilon \in \mathbb{N}$, such that $n > N_\varepsilon, p \in \mathbb{N} \Rightarrow |a_{n+p} - a_n| < \varepsilon$.

The *limit* of a sequence a is a number that satisfies: $\forall \varepsilon > 0$, $\exists N_\varepsilon \in \mathbb{N}$, such that $n > N_\varepsilon, |a_n - a| < \varepsilon$. We state $a_n \to a$ as $n \to \infty$, or $\lim_{n\to\infty} a_n = a$.

There exist sequences of rationals that do not converge to a rational

$$\left(1 + \frac{1}{n}\right)^n \to e, \frac{F_{n+1}}{F_n} \to \frac{1+\sqrt{5}}{2}, F_{n+2} = F_{n+1} + F_n, F_0 = F_1 = 1$$

$\mathbb{R}$ is the completion of $\mathbb{Q}$ w.r.t. the distance ("closeness" concept) $|a - b|$.

Suppose that closeness between rationals is now defined differently, not using addition but multiplication. Consider $p \in \mathbb{N}$, a prime number. Any rational $x \in \mathbb{Q}$ can be written as

$$x = \frac{r}{s} p^a$$

The p-adic norm is defined as

$$|x|_p = p^{-a}$$

Example. $x = \frac{24}{55} = 2^3 \cdot 3^1 \cdot 5^{-1} \cdot 11^{-1}$. $|x|_2 = 2^{-3}$, $|x|_3 = 3^{-1}$, $|x|_5 = 5$, $|x|_{11} = 11$.

The p-adic numbers $\mathbb{Q}_p$ are the completion of $\mathbb{Q}$ w.r.t. $||_p$ concept of closeness.

A point mass has no dimension, and its position in $\mathbb{R}^3$ is given the coordinates $X = (x, y, z)$ or $X = (x_1, x_2, x_3)$. A rigid body has extent, and maintains relative position between its constituents. In addition to position, a rigid body has orientation $\Theta = (\theta, \varphi, \psi)$ or $\Theta = (\theta_1, \theta_2, \theta_3)$.

A trajectory is a succession of positions. The trajectory of a rigid body is given by $(X(t), \Theta(t))$ with t a label for successive positions. If $t \in \mathbb{R}$, $(X, \Theta): \mathbb{R} \to \mathbb{R}^6$.

Newton's laws relate successive positions to forces $F = (f_1, f_2, f_3)$ and moments (torques) $L = (l_1, l_2, l_3)$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(m\frac{\mathrm{d}X}{\mathrm{d}t}\right) = F, \frac{\mathrm{d}}{\mathrm{d}t}\left(I\frac{\mathrm{d}\Theta}{\mathrm{d}t}\right) = L$$

with m, I denoting the inertia w.r.t. changes in rectilinear or rotational motion.

- Define a rigid body data set from positions of a "smart-phone" carrying human
- Such devices carry embedded accelometers that measure $a = \ddot{X}, \varepsilon = \ddot{\Theta}$
- Simple questions within this module:
  - Can we quantify how much information is recorded?
  - Can we reconstruct trajectories from (a, ε)?
  - Does the reconstruction of trajectories have the same information content as the data?
  - Can we represent the positions or trajectories more econmically?
- Complex questions within this module:
  - Can we identify what the data represents? (e.g., climbing, walking, ...)
  - Can we identify the particular human?

The approximation problem: Given $x(t)$, $x\colon \mathbb{R}\to\mathbb{R}$, find $y(t)$, $y\colon\mathbb{R}\to\mathbb{R}$ that is "easier to compute", and "is close" to x, $x\cong y$. Approximation criteria:

Interpolation. x given by data set $\mathcal{X}=\{(t_i,x(t_i)),i=1,\dots,m\}$, find y such that $y(t_i)=x(t_i)$. To ensure simple computation of the approximation, express y through a linear combination of basis functions $\mathcal{B}=\{b_1(t),\dots,b_n(t)\}$:

$y(t)=\sum_{j=1}^n c_j b_j(t)$, leading to linear system $\sum_{j=1}^n b_j(t_i)\,c_j=x(t_i)$, $\mathbf{B}\,\mathbf{c}=\mathbf{x}$. To satisfy all interpolation conditions, set $n=m$.

Least-squares. As above, but seek data compression $n\ll m$, by relaxing approximation condition to $\mathbf{B}\,\mathbf{c}\cong\mathbf{x}$, implemented as $\min_{\mathbf{c}}\|\mathbf{B}\,\mathbf{c}-\mathbf{x}\|_2$. Recall that the 2-norm of a vector $\mathbf{x}\in\mathbb{R}^m$ is $\|\mathbf{x}\|_2=\left(\sum_{i=1}^m x_i^2\right)^{1/2}$

Min-max. As above, but seek data compression through a different choice of norm, $\min_{\mathbf{c}}\|y(t;\mathbf{c})-x(t)\|_\infty$, where the inf-norm of a function $x\colon\mathbb{R}\to\mathbb{R}$ is

$\|x\|_\infty=\sup_t|x(t)|\cong\max_{i\in\{1,\dots,m\}}|x(t_i)|$.

Consider now the problem of approximating $u(\boldsymbol{x})$, $u\colon\mathbb{R}^d\to\mathbb{R}$. The *multivariate* approximation problem is to find $v(\boldsymbol{x})$, $v\colon\mathbb{R}^d\to\mathbb{R}$, such that $v\cong u$ and v is "easy to compute".

The formalism is identical to $d=1$:

Interpolation. $\sum_{j=1}^{n} b_j(\boldsymbol{x}_i)\, c_j = u(\boldsymbol{x}_i)$, $i=1,\ldots,m$, $\boldsymbol{x}_i\in\mathbb{R}^d$

Least-squares. $\min_{\mathbf{c}} \|\mathbf{B}\,\mathbf{c}-\mathbf{u}\|_2$, $\mathbf{c}\in\mathbb{R}^n$

Min-max. $\min_{\mathbf{c}} \|\mathbf{B}\,\mathbf{c}-\mathbf{u}\|_\infty$, $\mathbf{c}\in\mathbb{R}^n$

Consider that k samples are required along each variable direction, the total number of data points is $m=k^d$, and rapidly increases with d, known as the curse of dimensionality. Sample placement becomes a crucial consideration.

In matrix form interpolation is stated as $\mathbf{B}\mathbf{c} = \mathbf{I}\mathbf{x}$, with $\mathbf{c}, \mathbf{x} \in \mathbb{R}^m$, $\mathbf{B}, \mathbf{I} \in \mathbb{R}^{m \times m}$.

Remark 1. There is no data compression, simply a transformation from one basis ($\mathbf{I}$) to another ($\mathbf{B}$). Interpolation is akin to faithful translation, the same information is conveyed in two different languages.

Remark 2. The lack of data compression is a direct consequence of the interpolation criterion $y(t_i) = x(t_i)$, that imposes the strong condition of equality at all data points. This suggests alternative approaches:

→ Replace equality by a different equivalence relation

→ Replace data representation by real numbers by alternative constructs

Remark 3. The interpolation operator $\mathbf{B}^{-1}$ is a square matrix, $\mathbf{B}^{-1} \in \mathbb{R}^{m \times m}$, a linear operator between spaces of equal dimension, a feature of non-compressive data transformations.

Data transformations: integration

Consider $a(t)$ specified by data $\mathcal{A} = \{(t_i = ih, a(t_i)), i = 0, \ldots, m\}$.

The running sum of data values, $v_j = h\sum_{i=1}^{j} a(t_i)$, $j = 1, \ldots, m$, is an equivalent representation that contains the same amount of data, and obtained by approximation of the integral

$$v(t_i) = \int_0^{t_i} a(\tau)\,\mathrm{d}\tau + v_0,$$

through an upper Darboux sum (rectangle quadrature rule), using samples at $\tau \in \{h, 2h, \ldots ih\}$. This linear, non-compressive data transformation can again be expressed through a matrix multiplication

$$\mathbf{v} = h\begin{pmatrix} 1 & 0 & \ldots & 0 \\ 1 & 1 & \ldots & 0 \\ & & \ddots & \\ 1 & 1 & \ldots & 1 \end{pmatrix}\mathbf{a} = \mathbf{Q}\mathbf{a}.$$

Data transformations: differentiation

Consider $v(t)$ specified by data $\mathcal{V} = \{(t_i = ih, v(t_i)), i = 0, \ldots, m\}$.

The set of data value differences, $a_i = h^{-1}(v_i - v_{i-1})$, $i = 1, \ldots, m$, is an equivalent representation that contains the same amount of data, and obtained by approximation of the derivative $v'(t_i) = a_i$, through a finite difference formula.

This linear, non-compressive data transformation can again be expressed through a matrix multiplication

$$\mathbf{v} = \mathbf{Q}\mathbf{a} \Rightarrow \mathbf{a} = \mathbf{D}\mathbf{v} = \mathbf{Q}^{-1}\mathbf{v}.$$

$$\mathbf{D} = \mathbf{Q}^{-1} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ & & & \ddots & & \\ & & & & 1 & 0 \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{pmatrix}$$

Consider now the least squares problem

$$\min_{\mathbf{c}} \|\mathbf{B}\,\mathbf{c} - \mathbf{x}\|_2, \mathbf{x} \in \mathbb{R}^m, \mathbf{c} \in \mathbb{R}^n, n < m, \mathbf{B} \in \mathbb{R}^{m \times n},$$

a data-compressive procedure, usually with $n \ll m$. The non-square matrix $\mathbf{B}$ is not invertible, since invertability would imply no data compression. The same role is now played by the (Moore-Penrose) pseudo-inverse, $\mathbf{B}^+ \in \mathbb{R}^{n \times m}$. Pseudo-inverse computation:

- if $\mathbf{B}$ is an orthogonal matrix (i.e., $\mathbf{B}^T\mathbf{B} = \mathbf{I}_n$), $\mathbf{B}^+ = \mathbf{B}^T$.
- for general $\mathbf{B}$, introduce the singular value decomposition, $\mathbf{B} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, with $\mathbf{U} \in \mathbb{R}^{m \times m}, \mathbf{V} \in \mathbb{R}^{n \times n}, \boldsymbol{\Sigma} \in \mathbb{R}_+^{m \times n}$, $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$, $\mathbf{U}, \mathbf{V}$ orthogonal. The pseudo-inverse is given by

$$\mathbf{B}^+ = \mathbf{V}\boldsymbol{\Sigma}^+\mathbf{U}^T, \boldsymbol{\Sigma}^+ = \mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0).$$

The general min-max problem $\min_{\mathbf{c}} \| y(t;\mathbf{c}) - x(t) \|_\infty$ is quite difficult to solve in general, but the solution $y(t;\mathbf{c})$ is closely approximated by a linear combination of Chebyshev polynomials $\mathbf{y} = \mathbf{T}\,\mathbf{c}$, with $\mathbf{c} \in \mathbb{R}^n$, the solution of the problem $\min_{\mathbf{c}} \|\mathbf{T}\,\mathbf{c} - \mathbf{x}\|_\infty$.

The Chebyshev polynomials are defined by

$$T_0(t) = 1, T_1(t), T_{n+1}(t) = 2t\,T_n(t) - T_{n-1}(t).$$

Within the set Π_n of all monic polynomials defined on a finite interval (standardized as $[-1,1]$) the monic Chebyshev polynomials $2^{1-n}\,T_n(t)$, have the property of minimal inf-norm, i.e., they are the solution of

$$\min_{p \in \Pi_n} \|p\|_\infty, \|T_n\|_\infty = 2^{1-n}.$$

Consider now the problem of ascertaining data content of a set of trajectories $\mathcal{T}_m = \{\boldsymbol{x}_1(t), \ldots, \boldsymbol{x}_m(t)\}$, $\boldsymbol{x}: \mathbb{R} \to \mathbb{R}^d$, e.g., as given by the measured accelerations of m phone-carrying humans. The analysis point of view would be to think of small differences between individuals and try to define calculus operations. This can be done by considering the set of all such trajectories $\mathcal{T} = \{\boldsymbol{x}(t)\}$, and associate a scalar "label" to each trajectory, through the mathematical construct of a functional $\ell: \mathcal{T} \to \mathbb{R}$. Small changes in the trajectory would change the value of the label

$$\delta\ell = \ell(\boldsymbol{x} + \delta\boldsymbol{x}) - \ell(\boldsymbol{x}).$$

The functional ℓ is typically approximated by a multivariate function $f: \mathbb{R}^{dN} \to \mathbb{R}$ with N the number of trajectory samples

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N).$$